6

Monte Carlo Methods for Security Pricing*

Phelim Boyle, Mark Broadie and Paul Glasserman

1 Introduction

In recent years the complexity of numerical computation in financial theory and practice has increased enormously, putting more demands on computational speed and efficiency. Numerical methods are used for a variety of purposes of finance. These include the valuation of securities, the estimation of their sensitivities, risk analysis, and stress testing of portfolios. The Monte Carlo method is a useful tool for many of these calculations, evidenced in part by the voluminous literature of successful applications. For a brief sampling, the reader is referred to the stochastic volatility applications in Duan (1995), Hull and White (1987), Johnson and Shanno (1987), and Scott (1987);[1] the valuation of mortgage-backed securities in Schwartz and Torous (1989); the valuation of path-dependent options in Kemna and Vorst (1990); the portfolio optimization in Worzel et al. (1994); and the valuation of interest-rate derivative claims in Carverhill and Pang (1995). In this paper we focus on recent methodological developments. We review the Monte Carlo approach and describe some recent applications in the finance area.

In modern finance, the prices of the basic securities and the underlying state variables are often modelled as continuous-time stochastic processes. A derivative security, such as a call option, is a security whose payoff depends on one or more of the basic securities. Using the assumption of no arbitrage, financial economists have shown that the price of a generic derivative security can be expressed as the expected value of its discounted payouts. This expectation is taken with respect to a transformation of the original probability measure known as the *equivalent martingale measure* or the *risk-neutral* measure. The book by Duffie (1996) provides an excellent account of this material.

The Monte Carlo method lends itself naturally to the evaluation of security prices represented as expectations. Generically, the approach consists of the following

[1] Wiggins (1987) also studies pricing under stochastic volatility but does not use Monte Carlo simulation.

steps:

- Simulate sample paths of the underlying state variables (e.g., underlying asset prices and interest rates) over the relevant time horizon. Simulate these according to the risk-neutral measure.
- Evaluate the discounted cash flows of a security on each sample path, as determined by the structure of the security in question.
- Average the discounted cash flows over sample paths.

In effect, this method computes a multi-dimensional integral – the expected value of the discounted payouts over the space of sample paths. The increase in the complexity of derivative securities in recent years has led to a need to evaluate high dimensional integrals.

Monte Carlo becomes increasingly attractive compared to other methods of numerical integration as the dimension of the problem increases. Consider the integral of the function $f(x)$ over the d-dimensional unit hypercube. The simple (or crude) Monte Carlo estimate of the integral is equal to the average value of the function f over n points selected at random[2] from the unit hypercube. From the strong law of large numbers this estimate converges to the true value of the integrand as n tends to infinity. In addition, the central limit theorem assures us that the standard error[3] of the estimate tends to zero as $1/\sqrt{n}$. Thus the error convergence rate is independent of the dimension of the problem and this is the dominant advantage of the method over classical numerical integration approaches. The only restriction on the function f is that it should be square integrable, and this is a relatively mild restriction.

Furthermore, the Monte Carlo method is flexible and easy to implement and modify. In addition, the increased availability of powerful computers has enhanced the attractiveness of the method. There are some disadvantages of the method but in recent years progress has been made in overcoming them. One drawback is that for very complex problems a large number of replications may be required to obtain precise results. Different variance reduction techniques have been developed to increase precision. Two of the classical variance reduction techniques are the control variate approach and the antithetic variate method. More recently, moment matching, importance sampling, and conditional Monte Carlo methods have been introduced in finance applications.

Another technique for speeding up the valuation of multidimensional integrals uses deterministic sequences rather than random sequences. These deterministic

---

[2] In standard Monte-Carlo application the n points are usually not truly random but are generated by a deterministic algorithm and are described as *pseudorandom* numbers.

[3] We can readily estimate the variance of the Monte Carlo estimate by using the same set of n random numbers to estimate the expected value of f^2.

sequences are chosen to be more evenly dispersed throughout the region of integration than random sequences. If we use these sequences to estimate multidimensional integrals we can often improve the convergence. Deterministic sequences with this property are known as low-discrepancy sequences or quasi-random sequences. Using this approach one can in theory derive deterministic error bounds, though the practical use of the bounds is problematic. In contrast, standard Monte Carlo yields simple, useful probabilistic error bounds. Although low-discrepancy sequences are well known in computational physics they have only recently been applied in finance problems. There are different procedures for generating such low-discrepancy sequences and these procedures are generally based on number theoretic methods. We describe some of the recent developments in this area. We also discuss applications of this approach to problems in finance and conduct some rough comparisons between standard Monte Carlo methods and two different quasi-random approaches.

Until recently, the valuation of American style options was widely considered outside the scope of Monte Carlo. However Tilley (1993), Barraquand and Martineau (1995), and Broadie and Glasserman (1997), and have proposed approaches to this problem, and there has been other related work as well. We provide a brief survey of the recent research progress in this area.

The layout of the paper is as follows. Variance reduction techniques are described in the next section. The ideas behind the use of low-discrepancy sequences and brief numerical comparisons with standard Monte Carlo methods are given in Section 3. Price sensitivity estimation using simulation is discussed in Section 4. Various approaches to pricing American options using simulation are briefly described in Section 5. Other issues are touched on briefly in Section 6.

2 Variance reduction techniques

In this section, we first discuss the role of variance reduction in meeting the broader objective of improving the computational efficiency of Monte Carlo simulations. We then discuss specific variance reduction techniques and illustrate their application to pricing problems.

2.1 Variance reduction and efficiency improvement

The reduction of variance seems so obviously desirable that the precise argument for its benefit is sometimes overlooked. We briefly review the underlying justification for variance reduction and examine it from the perspective improving computational efficiency.

Suppose we want to compute a parameter θ – for example, the price of a derivative security. Suppose we can generate by Monte Carlo an i.i.d. sequence $\{\hat{\theta}_i, i = 1, 2, \ldots\}$, where each $\hat{\theta}_i$ has expectation θ and variance σ^2. A natural estimator of θ based on n replications is then the sample mean

$$\frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i.$$

By the central limit theorem, for large n this sample mean is approximately normally distributed with mean θ and variance σ^2/n. Probabilistic error bounds in the form of confidence intervals follow readily from the normal approximation, and indicate that the error in the estimator is proportional to $\sigma/\sqrt{n}$. Thus, decreasing the variance σ^2 by a factor of 10, say, while leaving everything else unchanged, does as much for error reduction as increasing the number of samples by a factor of 100.

Suppose, now, that we have a choice between two types of Monte Carlo estimates which we denote by $\{\hat{\theta}_i^{(1)}, i = 1, 2, \ldots\}$ and $\{\hat{\theta}_i^{(2)}, i = 1, 2, \ldots\}$. Suppose that both are unbiased, so that $E[\hat{\theta}_i^{(1)}] = E[\hat{\theta}_i^{(2)}] = \theta$, but $\sigma_1 < \sigma_2$, where $\sigma_j^2 = \text{Var}[\hat{\theta}^{(j)}]$, $j = 1, 2$. From our previous observations it follows that a sample mean of n replications of $\hat{\theta}^{(1)}$ gives a more precise estimate of θ than does a sample mean of n replications of $\hat{\theta}^{(2)}$. But this analysis oversimplifies the comparison because it fails to capture possible differences in the computational effort required by the two estimators. Generating n replications of $\hat{\theta}^{(1)}$ may be more time-consuming than generating n replications of $\hat{\theta}^{(2)}$; smaller variance is not sufficient grounds for preferring one estimator over another.

To compare estimators with different computational requirements as well as different variances, we argue as follows. Suppose the work required to generate one replication of $\hat{\theta}^{(j)}$ is a constant b_j, $j = 1, 2$. (In some problems, the work per replication is stochastic; assuming it is constant simplifies the discussion.) With computing time t, the number of replications of $\hat{\theta}^{(j)}$ that can be generated is $\lfloor t/b_j \rfloor$; for simplicity, we drop the $\lfloor \cdot \rfloor$ and treat the ratios t/b_j as though they were integers. The two estimators available with computing time t are therefore

$$\frac{b_1}{t} \sum_{i=1}^{t/b_1} \hat{\theta}_i^{(1)} \quad \text{and} \quad \frac{b_2}{t} \sum_{i=1}^{t/b_2} \hat{\theta}_i^{(2)}.$$

For large t, these are approximately normally distributed with mean θ and with standard deviations

$$\sigma_1 \sqrt{\frac{b_1}{t}} \quad \text{and} \quad \sigma_2 \sqrt{\frac{b_2}{t}}.$$

Thus, for large t, the first e

Equation (1) provides a putational requirements. I to take the product of vari efficiency as a basis for c ferred only if the variance same argument, a higher-v much less time to generate

In its simplest form, the and Handscomb (1964, p. tended by Glynn and Whi which case each b_j is the the presence of bias.

Equipped with a basis fo now consider specific var widely used techniques i *variates*. We introduce it

Consider the problem option on a no-dividend s simulation, but the exam model, the stock price fo the terminal stock price formula

$$S_T^{(i)} =$$

where S_0 is the current s volatility, T is the option standard normal distribu and see Devroye (1986) Based on n replications with strike K is given by

$$\hat{C} = \frac{1}{n}$$

Thus, for large t, the first estimator should be preferred over the second if

$$\sigma_1^2 b_1 < \sigma_2^2 b_2. \tag{1}$$

Equation (1) provides a sound basis for trading-off estimator variance and computational requirements. In light of the discussion leading to (1), it is reasonable to take the product of variance and work per run as a measure of *efficiency*. Using efficiency as a basis for comparison, the lower-variance estimator should be preferred only if the variance ratio σ_1^2/σ_2^2 is smaller than the work ratio b_2/b_1. By the same argument, a higher-variance estimator may actually be preferable if it takes much less time to generate.

In its simplest form, the principle expressed in (1) dates at least to Hammersley and Handscomb (1964, p.22). More recently, the idea has been substantially extended by Glynn and Whitt (1992). They allow the work per run to be random (in which case each b_j is the *expected* work per run) and also consider efficiency in the presence of bias.

2.2 Antithetic variates

Equipped with a basis for evaluating potential efficiency improvements, we can now consider specific variance reduction techniques. One of the simplest and most widely used techniques in financial pricing problems is the method of *antithetic variates*. We introduce it with a simple example, then generalize.

Consider the problem of computing the Black–Scholes price of a European call option on a no-dividend stock. Of course, there is no need to evaluate this price by simulation, but the example serves as a useful introduction. In the Black–Scholes model, the stock price follows a lognormal diffusion. Independent replications of the terminal stock price under the risk-neutral measure can be generated from the formula

$$S_T^{(i)} = S_0 e^{(r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z_i}, \quad i = 1, \ldots, n, \tag{2}$$

where S_0 is the current stock price, r is the riskless interest rate, σ is the stock's volatility, T is the option's maturity, and the $\{Z_i\}$ are independent samples from the standard normal distribution. See, e.g., Hull (2000) for background on this model, and see Devroye (1986) for methods of sampling from the normal distribution. Based on n replications, a moment-matched estimator of the price of an option with strike K is given by

$$\hat{C} = \frac{1}{n}\sum_{i=1}^{n} C_i \equiv \frac{1}{n}\sum_{i=1}^{n} e^{-rT} \max\{0, S_T^{(i)} - K\}. \tag{3}$$

In this context, the method of antithetic variates[4] is based on the observation that if Z_i has a standard normal distribution, then so does $-Z_i$. The price $\tilde{S}_T^{(i)}$ obtained from (2) with Z_i replaced by $-Z_i$ is thus a valid sample from the terminal stock price distribution. Similarly, each

$$\tilde{C}_i = e^{-rT} \max\{0, \tilde{S}_T^{(i)} - K\}$$

is an unbiased estimator of the option price, as is therefore

$$\hat{C}_{AV} = \frac{1}{n} \sum_{i=1}^{n} \frac{C_i + \tilde{C}_i}{2}.$$

A heuristic argument for preferring $\hat{C}_{AV}$ notes that the random inputs obtained from the collection of antithetic pairs $\{(Z_i, -Z_i)\}$ are more regularly distributed than a collection of $2n$ independent samples. In particular, the sample mean over the antithetic pairs always equals the population mean of 0, whereas the mean over finitely many independent samples is almost surely different from 0. If the inputs are made more regular, it may be hoped that the outputs are more regular as well. Indeed, a large value of $S_T^{(i)}$ resulting from a large Z_i will be paired with a small value of $\tilde{S}_T^{(i)}$ obtained from $-Z_i$.

A more precise argument compares efficiencies. Because C_i and $\tilde{C}_i$ have the same variance,

$$\text{Var}\left[\frac{C_i + \tilde{C}_i}{2}\right] = \frac{1}{2}(\text{Var}[C_i] + \text{Cov}[C_i, \tilde{C}_i]). \tag{4}$$

Thus, we have $\text{Var}[\hat{C}_{AV}] \leq \text{Var}[\hat{C}]$ if $\text{Cov}[C_i, \tilde{C}_i] \leq \text{Var}[C_i]$. However, $\hat{C}_{AV}$ uses twice as many replications as $\hat{C}$, so we must account for differences in computational requirements. If generating the Z_i takes a negligible fraction of the work per replication (which would typically be the case in the pricing of a more elaborate option), then the work to generate $\hat{C}_{AV}$ is roughly double the work to generate $\hat{C}$. Thus, for antithetics to increase efficiency, we require

$$2\,\text{Var}[\hat{C}_{AV}] \leq \text{Var}[\hat{C}],$$

which, in light of (4), simplifies to the requirement that $\text{Cov}[C_i, \tilde{C}_i] \leq 0$.

That this condition is met is easily demonstrated. Define ϕ so that $C_i = \phi(Z_i)$; ϕ is the composition of the mappings from Z_i to the stock price and from the stock price to the discounted option payoff. As the composition of two increasing functions, ϕ is monotone, so by a standard inequality (e.g., Section 2.2 of Barlow

[4] This method was introduced to option pricing in Boyle (1977), where its use was illustrated in the pricing of a European call on a dividend-paying stock.

and Proschan 1975)

$$E[\phi(Z_i)\phi(-Z_i)] \leq E[\phi(Z_i)]E[\phi(-Z_i)], \tag{5}$$

i.e., $\mathrm{Cov}[C_i, \tilde{C}_i] \equiv E[\phi(Z_i)\phi(-Z_i)] - E[\phi(Z_i)]E[\phi(-Z_i)] \leq 0$, and we may conclude that antithetics help.

This argument can be adapted to show that the method of antithetic variates increases efficiency in pricing a European put and other options that depend monotonically on inputs (e.g., Asian options). The notable departure from monotonicity in some barrier options (e.g., a down-and-in call) suggests that the use of antithetics in pricing these options may sometimes be less effective.

In computing confidence intervals with antithetic variates, it is essential that the standard error be estimated using the sample standard deviation of the n averaged pairs $(C_i + \tilde{C}_i)/2$ and not the $2n$ individual observations $C_1, \tilde{C}_1, \ldots, C_n, \tilde{C}_n$. The averaged pairs are independent but the individual observations are not. This is a case (we will see others shortly) in which the use of a variance reduction technique affects the estimation of the standard error and, in particular, requires some "batching" of observations to deal with dependence.

It is worth noting that the method of antithetic variates is by no means restricted to simulations whose only stochastic inputs are standard normal variates. The most primitive stochastic input in most simulations is a sequence $\{U_n\}$ of independent variates uniformly distributed on the unit interval. In this case, $1 - U_n$ has the same distribution as U_n, and the pair $(U_n, 1 - U_n)$ are called antithetic because they exhibit negative dependence. If the simulation output depends monotonically on the input random numbers, then the output obtained from $\{1 - U_1, 1 - U_2, \ldots\}$ will be negatively correlated with that obtained from $\{U_1, U_2, \ldots\}$, resulting in increased efficiency compared with independent replications.

For further general background on antithetic variates and other methods based on correlation induction, see Bratley, Fox, and Schrage (1987), Hammersley and Handscomb (1964), Glynn and Iglehart (1988), and references there. For some examples of application in finance, see Boyle (1977), Clewlow and Carverhill (1994), and Hull and White (1987).

2.3 Control variates

The method of *control variates* is among the most widely applicable, easiest to use, and effective of the variance reduction techniques.[5] Simply put, the principle underlying this technique is "use what you know."

The most straightforward implementation of control variates replaces the evaluation of an unknown expectation with the evaluation of the *difference* between

---

[5] The earliest application of this technique to option pricing is Boyle (1977).

the unknown quantity and another expectation whose value is known. A specific illustration can be found in the analysis of Boyle and Emanuel (1985) and Kemna and Vorst (1990) of Asian options. Let P_A be the price of an option whose payoff depends on the arithmetic average of the underlying asset. Let P_G be the price of an option equivalent in every respect except that a geometric average replaces the arithmetic average. Most options based on averages use arithmetic averaging, so P_A is of much greater practical value; but whereas P_A is analytically intractable, P_G can often be evaluated in closed form. Can knowledge of P_G be leveraged to compute P_A?

It can, through the control variate method. Write $P_A = E[\hat{P}_A]$ and $P_G = E[\hat{P}_G]$, where $\hat{P}_A$ and $\hat{P}_G$ are the discounted option payoffs for a single simulated path of the underlying asset. Then

$$P_A = P_G + E[\hat{P}_A - \hat{P}_G];$$

in other words, P_A can be expressed as the known price P_G plus the expected difference between $\hat{P}_A$ and $\hat{P}_G$. An unbiased estimator of P_A is thus provided by

$$\hat{P}_A^{cv} = \hat{P}_A + (P_G - \hat{P}_G). \tag{6}$$

This representation[6] suggests a slightly different interpretation: $\hat{P}_A^{cv}$ adjusts the straightforward estimator $\hat{P}_A$ according to the difference between the known value P_G and the observed value $\hat{P}_G$. The known error $(P_G - \hat{P}_G)$ is used as a *control* in the estimation of P_A.

If most of the computational effort goes to generating paths of the underlying asset, then the additional work required to evaluate $\hat{P}_G$ along with $\hat{P}_A$ is minor. It therefore seems reasonable to compare variances alone. Since

$$\text{Var}[\hat{P}_A^{cv}] = \text{Var}[\hat{P}_A] + \text{Var}[\hat{P}_G] - 2\,\text{Cov}[\hat{P}_A, \hat{P}_G],$$

this method if effective if the covariance between $\hat{P}_A$ and $\hat{P}_G$ is large. The numerical results of Kemna and Vorst indicate that this is indeed the case. Fu, Madan, and Wang (1998) have investigated the use of other control variates for Asian options, based on Laplace transform values. These appear to be less strongly correlated with the option price.

A closer examination of (6) reveals that this estimator does not make optimal use of the relation between the two option prices. Consider the family of unbiased estimators

$$\hat{P}_A^{\beta} = \hat{P}_A + \beta(P_G - \hat{P}_G), \tag{7}$$

---

[6] To go from (6) to Boyle's (1977) example, let P_G be the price of a European call option on a no-dividend stock and let P_A be the corresponding option price in the presence of dividends.

parameterized by the scalar β. We have

$$\text{Var}[\hat{P}_A^\beta] = \text{Var}[\hat{P}_A] + \beta^2 \, \text{Var}[\hat{P}_G] - 2\beta \, \text{Cov}[\hat{P}_A, \hat{P}_G].$$

The variance-minimizing β is therefore

$$\beta^* = \frac{\text{Cov}[\hat{P}_A, \hat{P}_G]}{\text{Var}[\hat{P}_G]}.$$

Depending on the application, β^* may or may not be close to 1, the implicit value in (6). In using an estimator of the form (6), we forgo an opportunity for greater variance reduction. Indeed, whereas (6) may increase or decrease variance, an estimator based on β^* is guaranteed not to increase variance, and will result in a strict decrease in variance so long as $\hat{P}_A$ and $\hat{P}_G$ are not uncorrelated.

In practice, of course, we rarely know β^* because we rarely know $\text{Cov}[\hat{P}_A, \hat{P}_G]$. However, given n independent replications $\{(P_{Ai}, P_{Gi}), i = 1, \ldots, n\}$ of the pairs $(\hat{P}_A, \hat{P}_G)$ we can estimate β^* via regression. At this point we face a choice. Using all n replications to compute an estimate $\hat{\beta}$ of β^* introduces a bias in the estimator

$$\frac{1}{n} \sum_{i=1}^{n} P_{Ai} + \hat{\beta} \left(P_G - \frac{1}{n} \sum_{i=1}^{n} P_{Gi} \right),$$

and its estimated standard error because of the dependence between $\hat{\beta}$ and the P_{Gi}. Reserving n_1 replications for the estimation of β^* and the remaining $n - n_1$ replications for the sample mean of the P_{Gi} (typically with $n_1 \ll n$) eliminates the bias but may deteriorate the estimate of β^*. Neither issue significantly limits the applicability of the method, because the possible bias vanishes as n increases and because the estimate of β^* need not be very precise to achieve a reduction in variance.

The advantage of working with (7) over (6) becomes even more pronounced when further controls are introduced. For example, when the asset price is simulated under risk-neutral probabilities, the present value $e^{-rT} E[S_T]$ of the terminal price must equal the current price S_0. We can therefore form the estimator

$$\hat{P}_A + \beta_1 (P_G - \hat{P}_G) + \beta_2 (S_0 - e^{-rT} S_T).$$

The variance-minimizing coefficients (β_1^*, β_2^*) are easily found by multiple regression. This optimization step seems particularly crucial in this case; for whereas one might guess that β_1^* is close to 1, it seems unlikely that β_2^* would be. Optimizing over the βs also allows us to exploit controls that are negatively correlated with the option payoff.

For further general background on control variates see Bratley, Fox, and Schrage (1987), Glynn and Iglehart (1988), and Lavenberger and Welch (1981). For examples of control variate applications in finance, see Boyle (1977), Boyle and

Emanuel (1985), Broadie and Glasserman (1996), Carverhill and Pang (1995), Clewlow and Carverhill (1994), Duan (1995), and Kemna and Vorst (1990).

2.4 Moment matching methods

Next we describe a variance reduction technique proposed by Barraquand (1995), who termed it *quadratic resampling*. His technique is based on *moment matching*. As before, we introduce it with the simple example of estimating the European call option price on a single asset and then generalize.

Let Z_i, $i = 1, \ldots, n$, denote independent standard normals used to drive a simulation. The sample moments of the n Z's will not exactly match those of the standard normal. The idea of moment matching is to transform the Z's to match a finite number of the moments of the underlying population. For example, the first moment of the standard normal can be matched by defining

$$\tilde{Z}_i = Z_i - \bar{Z}, \qquad i = 1, \ldots, n, \tag{8}$$

where $\bar{Z} = \sum_{i=1}^{n} Z_i/n$ is the sample mean of the Z's. Note that the $\tilde{Z}_i$'s are normally distributed if the Z_i's are normal. However, the $\tilde{Z}_i$'s are not independent. As before, terminal stock prices are generated from the formula

$$\tilde{S}_T(i) = S_0 e^{(r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}\tilde{Z}_i}, \qquad i = 1, \ldots, n.$$

An unbiased estimator of the call option price is the average of the n values $\tilde{C}_i = e^{-rT} \max(\tilde{S}_T(i) - K, 0)$.

In the standard Monte Carlo method, confidence intervals for the true value C could be estimated from the sample mean and variance of estimator. This cannot be done here since the n values of $\tilde{Z}$ are no longer independent, and hence the values $\tilde{C}_i$ are not independent. This points out one drawback of the moment matching method: confidence intervals are not as easy to obtain.[7] Indeed, for confidence intervals it appears to be necessary to apply moment matching to independent batches of runs and estimate the standard error from the batch means. This reduces the efficacy of the method compared with matching moments across all runs.

Equation (8) showed one way to match the first moment of a distribution with mean zero. If the underlying population does not have a zero mean, transformed Z's could be generated using $\tilde{Z}_i = Z_i - \bar{Z} + \mu_Z$, where μ_Z is the population mean. The idea can easily be extended to match two moments of a distribution. In this case, an appropriate transformation is

$$\tilde{Z}_i = (Z_i - \bar{Z})\frac{\sigma_Z}{s_Z} + \mu_Z, \qquad i = 1, \ldots, n, \tag{9}$$

---

[7] The point is not merely a minor technical issue. The sample variance of the $\tilde{C}_i$'s is usually a poor estimate of $\mathrm{Var}[\tilde{C}_i]$.

(right column, partially cut off)

where s_Z is the sample st... standard deviation. Of cou... estimator of the call option...

Using the transformatio... Z_i's are normal. Hence, ... option value. For most fina... be small. However, the bia... when only the first momen... bias in the moment matchi... in general analytical terms...

The moment matching ... know." In this simple Eu... terminal stock price S_T is... applied to the simulated te... moment, define

where $\mu_{S_T} = S_0 e^{rT}$ and $\bar{S}$... two moments, define

$$\tilde{S}_T ...$$

where $\sigma_{S_T} = S_0\sqrt{e^{2rT}(e^{\sigma...}}$ ... $S_T(i)$'s. Duan and Simona... tive transformation to asse... set of paths.[9] They apply t...

Comparisons of various... this comparison, $n = 100$ ... option price. Standard err... simulation trials were con... The sample standard devi... the standard error of a si... not reported because they... digits reported.

[8] For example, let Z take the valu... $Z = 1$ and $-\$x$ if $Z \neq 1$. The e... by Monte Carlo simulation, dra... to define $\tilde{Z}_i$'s which match the... payoff is $-x$ and the bias is $(1 - ...$ using equation (8) or (9) when ... applying (8) or (9) to uniform ... outside of the relevant domain.

[9] This is equivalent to enforcing ...

where s_Z is the sample standard deviation of the Z_i's and σ_Z is the population standard deviation. Of course, for a standard normal, $\mu_Z = 0$ and $\sigma_Z = 1$. An estimator of the call option price is the average of the n values $\tilde{C}_i$.

Using the transformation (9), the $\tilde{Z}_i$'s are not normally distributed even if the Z_i's are normal. Hence, the corresponding $\tilde{C}_i$ are biased estimators of the true option value. For most financial problems of practical interest, this bias is likely to be small. However, the bias can be arbitrarily large in extreme circumstances (even when only the first moment of the distribution is matched).[8] The dependence and bias in the moment matching method makes it difficult to quantify the improvement in general analytical terms.

The moment matching method is another example of the idea to "use what you know." In this simple European option example, the mean and variance of the terminal stock price S_T is also known. So the moment matching idea could be applied to the simulated terminal stock values $S_T(i)$. In this case, to match the first moment, define

$$\tilde{S}_T(i) = S_T(i) - \bar{S}_T + \mu_{S_T}, \tag{10}$$

where $\mu_{S_T} = S_0 e^{rT}$ and $\bar{S}_T$ is the sample mean of the $S_T(i)$'s. To match the first two moments, define

$$\tilde{S}_T(i) = (S_T(i) - \bar{S}_T)\frac{\sigma_{S_T}}{s_{S_T}} + \mu_{S_T}, \tag{11}$$

where $\sigma_{S_T} = S_0\sqrt{e^{2rT}(e^{\sigma^2 T} - 1)}$ and s_{S_T} is the sample standard deviation of the $S_T(i)$'s. Duan and Simonato (1998) use a related method. They apply a multiplicative transformation to asset prices to enforce the martingale property over a finite set of paths.[9] They apply their method to GARCH option pricing.

Comparisons of various moment matching strategies are given in Table 1. For this comparison, $n = 100$ simulation trials were used to estimate the European call option price. Standard errors were estimated by re-simulation. That is, $m = 10\,000$ simulation trials were conducted, each one based on n replications of the estimator. The sample standard deviation of the m simulation estimates gives an estimate of the standard error of a single simulation estimate. Root-mean-squared errors are not reported because they are identical to the standard errors for the number of digits reported.

---

[8] For example, let Z take the values +1 or −1 with probability one-half. Consider a security which pays +$1 if $Z = 1$ and −x if $Z \neq 1$. The expected payoff of the security is $(1 - x)/2$. To estimate this expected payoff by Monte Carlo simulation, draw n samples Z_i according to the prescribed distribution. Then use equation (8) to define $\tilde{Z}_i$'s which match the first moment. For almost all samples for any large n, the estimated expected payoff is $-x$ and the bias is $(1 + x)/2$. This bias does not decrease as n increases. Care must be taken when using equation (8) or (9) when the support of the random variable of not the entire real line. For example, applying (8) or (9) to uniform or exponential random variables could cause the transformed values to fall outside of the relevant domain.

[9] This is equivalent to enforcing put-call parity.

Table 1. *Standard errors for European call options.*

σ	S_0/K	No variance reduction	MM1 Equation (8)	MM2 Equation (9)	MM1 Equation (10)	MM2 Equation (11)
0.2	0.9	0.24	0.19	0.11	0.19	0.09
	1.0	0.62	0.29	0.09	0.26	0.10
	1.1	0.93	0.19	0.09	0.15	0.11
0.4	0.9	0.80	0.55	0.24	0.51	0.17
	1.0	1.22	0.66	0.19	0.56	0.23
	1.1	1.61	0.63	0.17	0.48	0.28
0.6	0.9	1.40	0.95	0.38	0.84	0.28
	1.0	1.93	1.10	0.31	0.91	0.39
	1.1	2.38	1.13	0.25	0.85	0.49

All results are based on $n = 100$ simulation trials. The option parameters are: $K = 100$, $r = 0.10$, $T = 0.2$, with S_0 and σ varying as indicated. Standard error estimates are based on $m = 10\,000$ simulations.

The results in Table 1 show that matching two moments can reduce the simulation error by a factor ranging from 2 to 10. Matching two moments dominates matching one moment, but there is not a clear choice between transforming the original standard normals using (9) or the terminal stock prices using (11). Further computational results, not included in Table 1, indicate that the improvement factor with moment matching is *essentially constant* as n increases. This may seem counterintuitive, since the moment matching adjustments converge to zero as n increases. But the progressively smaller adjustments are equally important in reducing the estimation error as the number of simulation trials increases. For example, the standard error for $n = 10\,000$ simulation trials is one-tenth of the corresponding number for $n = 100$ reported in Table 1.

The moment matching method can be extended to match covariances. For options that depend on multiple assets, the entire covariance structure is typically a simulation input. Barraquand (1995) suggests a method to match the entire covariance structure and reports error reduction factors ranging from two to several hundred for this method applied to pricing options on the maximum of k assets.

The moment matching procedure could be applied to matching higher order moments as well. In addition to different methods for transforming random outcomes to match specified moments, additional points could be added as another way to match moments.

Whenever a moment is known, it can be used as a control rather than for moment matching. In an appendix, we give a theoretical argument favoring the use of moments as controls rather than for matching.

2.5 *Stratified and Latin hypercube sampling*

Like many variance reduction techniques, stratified sampling seeks to make the inputs to simulation more regular than random inputs. In particular, it forces certain empirical probabilities to match theoretical probabilities, just as moment matching forces empirical moments to match theoretical moments.

Consider, for example, the generation of 100 normal random variates as inputs to a simulation. The empirical distribution of an independent sample $Z_1, \ldots, Z_{100}$ will look only roughly like the normal density; the tails of the distribution – often the most important part – will inevitably be underrepresented. Stratified sampling can be used to force exactly one observation to lie between the $(i-1)^{\text{th}}$ and i^{th} percentile, $i = 1, \ldots, 100$, and thus produce a better match to the normal distribution. One way to implement this generates 100 independent random variates $U_1, \ldots, U_{100}$, uniform on $[0, 1]$ and set $\tilde{Z}_i = N^{-1}((i + U_i - 1)/100)$, $i = 1, \ldots, 100$, where N^{-1} is the inverse of the cumulative normal distribution. This works because $(i + U_i - 1)/100$ falls between the $(i-1)^{\text{th}}$ and i^{th} percentiles of the uniform distribution, and percentiles are preserved by the inverse transform.

Of course, $\tilde{Z}_1, \ldots, \tilde{Z}_{100}$ are highly dependent, complicating the estimation of standard errors. Computing confidence intervals with stratified sampling typically requires batching the runs. For example, with a budget of 100 000 replications we might run 100 independent stratified samples each of size 1000, rather than a single stratified sample of size 100 000. To estimate standard errors we must therefore sacrifice some variance reduction, just as with moment matching.

In principle, this approach applies in arbitrary dimensions. To generate a stratified sample from the d-dimensional unit hypercube, with n strata in each coordinate, we could generate a sequence of vectors $U_j = (U_j^{(1)}, \ldots, U_j^{(d)})$, $j = 1, 2, \ldots$, and then set

$$V_j = \frac{U_j + (i_1, \ldots, i_d)}{n}, \quad i_k = 0, \ldots, n-1, \quad k = 1, \ldots, d.$$

Exactly one V_j will lie in each of the n^d cubes defined by the product of the n strata in each coordinate.

The difficulty in high dimensions is that generating even a single stratified sample of size n^d may be prohibitive unless n is very small. *Latin hypercube sampling* can be viewed as a way of randomly sampling n points of a stratified sample while preserving some of the regularity from stratification. The method was introduced by McKay, Conover, and Beckman (1979) and further analyzed in Stein (1987). It works as follows. Let $\pi_1, \ldots, \pi_d$ be independent random permutations of $\{1, \ldots, n\}$, each uniformly distributed over all $n!$ possible permutations. Set

$$V_j^{(k)} = \frac{U_j^{(k)} + \pi_k(j) - 1}{n}, \quad k = 1, \ldots, d, \quad j = 1, \ldots, n.$$

The randomization ensures that each vector V_j is uniformly distributed over the d-dimensional hypercube. At the same time, the coordinates are perfectly stratified in the sense that exactly one of $V_1^{(k)}, \ldots, V_n^{(k)}$ falls between $(j-1)/n$ and j/n, $j = 1, \ldots, n$, for each dimension $k = 1, \ldots, d$. As before, the dependence introduced by this method implies that standard errors can be estimated only through batching.

These methods can be viewed as part of a hierarchy of methods introducing additional levels of regularity in inputs at the expense of complicating the estimation of errors. Some, like stratified sampling, fix the size the sample while others leave flexibility. The extremes of this hierarchy are straightforward Monte Carlo (completely random) and the low-discrepancy methods (completely deterministic) discussed in Section 3. Owen (1995a, 1995b) discusses these and other methods and introduces a hybrid that combines the regularity of low-discrepancy methods with the simple error estimation of standard Monte Carlo. Shaw (1995) uses an extension proposed by Stein (1987) to handle dependent inputs in a novel approach to estimating value at risk.

2.6 Some numerical comparisons

The variance reduction methods discussed thus far are fairly generic, in the sense that they do not rely on the detailed structure of the security to be priced. This contrasts with the remaining two methods that we discuss – importance sampling and conditional Monte Carlo. These methods must be carefully tailored to each application. It therefore seems appropriate to digress briefly into a numerical comparison of the generic methods on some option pricing problems.

We first examine the performance of these methods in pricing Asian options. The payoff of a discretely sampled arithmetic average Asian option is $\max(\bar{S} - K, 0)$, where $\bar{S} = \sum_{i=1}^{k} S_i/k$, S_i is the asset price at time $t_i = iT/k$, and T is the option maturity. The value of the option is $E[e^{-rT} \max(\bar{S} - K, 0)]$. There is no easily evaluated closed-form expression for this option value. Various formulas to approximate the Asian option price have been developed, but simulation is usually used to test the accuracy of the approximations.

For this Asian option, k random numbers are needed to simulate one option payoff, and nk random numbers are needed in total. Moment matching (MM2, for two moments) was applied k times to the n numbers used to generate each S_i at time t_i. Latin hypercube sampling (LHS) was applied to sample n points from the k-dimensional unit cube. The discretely sampled geometric average Asian price was used as a control variate (see Turnbull and Wakeman 1991 for a closed-form solution for this price). Results appear in Table 2.

The results in Table 2 indicate that matching two moments can reduce the simulation error by a factor ranging from 1 to 10. Using the geometric average Asian

Table 2. *Standar*

σ	K/S_0	No va reduc
0.2	0.9	0.0
	1.0	0.3
	1.1	0.5
0.4	0.9	0.3
	1.0	0.6
	1.1	1.0
0.6	0.9	0.6
	1.0	1.0
	1.1	1.4

All results are based on
average. The option para
varying as indicated. Sta
The geometric average
matching (MM2) was ap
replications.

option price as a control va
and is consistently the mos
Antithetics are consistently

Next we compare these
call options with discrete t
standard call option payoff
$t_i = iT/k$, $i = 1, \ldots, k,$
out if $S_i \le H$ at any tim
a standard call. Moment n
option. Results are given
Table 2, except that the sup

Although it is always risk
we suggest the following
implement, but often leads
is similarly easy to implem
the error estimation is mor
from the same error estima
variate technique can lead t
hinges on finding a good co

Table 2. *Standard errors for arithmetic average Asian options.*

σ	K/S_0	No variance reduction	Antithetic method	Control variate	MM2	LHS
0.2	0.9	0.053	0.052	0.003	0.048	0.049
	1.0	0.344	0.231	0.004	0.162	0.161
	1.1	0.566	0.068	0.006	0.052	0.058
0.4	0.9	0.308	0.297	0.014	0.240	0.248
	1.0	0.694	0.506	0.017	0.352	0.354
	1.1	1.017	0.388	0.021	0.281	0.289
0.6	0.9	0.632	0.583	0.032	0.451	0.455
	1.0	1.052	0.817	0.038	0.566	0.578
	1.1	1.443	0.759	0.047	0.539	0.560

All results are based on $n = 100$ simulation trials with $k = 50$ prices in the average. The option parameters are: $K = 100, r = 0.10, T = 0.2$, with S_0 and σ varying as indicated. Standard error estimates based on $m = 10\,000$ simulations. The geometric average Asian option is used as the control variate. Moment matching (MM2) was applied to the ith price in the average, $i = 1, \ldots, 5$, across replications.

option price as a control variate reduces error by a factor ranging from 20 to 100, and is consistently the most effective method. LHS and MM2 perform similarly. Antithetics are consistently dominated by the other methods.

Next we compare these variance reduction techniques in pricing down-and-out call options with discrete barriers. The payoff of this option at expiration is the standard call option payoff if the asset price S_i exceeds the barrier H at all times $t_i = iT/k$, $i = 1, \ldots, k$, otherwise the payoff is zero. The option is knocked out if $S_i \leq H$ at any time t_i. As a control we use the Black–Scholes price of a standard call. Moment matching and LHS are implemented as with the Asian option. Results are given in Table 3. These are consistent with the pattern in Table 2, except that the superiority of the control variate method is less pronounced.

Although it is always risky to draw conclusions from limited numerical evidence, we suggest the following broad conclusions. The antithetic method is easy to implement, but often leads to only modest error reductions. Moment matching is similarly easy to implement and often leads to significant error reductions, but the error estimation is more difficult and bias is a potential problem. LHS suffers from the same error estimation difficulty but does not introduce bias. The control variate technique can lead to very substantial error reductions, but its effectiveness hinges on finding a good control for each problem.

Table 3. *Standard errors for down-and-out call options with discrete barriers.*

σ	K/S_0	No variance reduction	Antithetic method	Control variate	MM2	LHS
0.2	0.9	0.96	0.44	0.37	0.43	0.39
	1.0	0.62	0.44	0.13	0.31	0.30
	1.1	0.30	0.28	0.03	0.22	0.22
0.4	0.9	1.59	1.15	0.73	0.95	0.88
	1.0	1.22	1.00	0.45	0.76	0.74
	1.1	0.88	0.82	0.26	0.61	0.61
0.6	0.9	2.19	1.83	1.07	1.44	1.36
	1.0	1.86	1.62	0.80	1.25	1.23
	1.1	1.54	1.40	0.58	1.09	1.09

All results are based on $n = 100$ simulation trials. There are $k = 5$ points in the discrete barrier at 95. The other option parameters are: $S_0 = 100$, $r = 0.10$, $T = 0.2$, with K and σ varying as indicated. Standard error estimates are based on $m = 10\,000$ simulations. The standard European call option (Black–Scholes formula) is used as the control variate. Moment matching (MM2) was applied to the ith return, $i = 1, \ldots, 5$, across replications.

2.7 Importance sampling

This technique builds on the observation that an expectation under one probability measure can be expressed as an expectation under another through the use of a *likelihood ratio* or *Radon–Nikodym derivative*. This idea is familiar in finance because it underlies the representation of prices as expectations under a martingale measure. In Monte Carlo, the change of measure is used to try to obtain a more efficient estimator. We present some examples using this technique; for general background see Bratley et al. (1987) or Hammersley and Handscomb (1964).

As a simple example, consider the evaluation of the Black–Scholes price of a call option – i.e., the computation of $e^{-rT}E[\max\{S_T - K, 0\}]$ with S_T as in (2). A straightforward approach generates samples of the terminal value S_T consistent with a geometric Brownian motion having drift r and volatility σ, just as in (2). But we are in fact free to generate S_T consistent with any other drift μ, provided we weight the result with a likelihood ratio. For emphasis, we subscript the expectation operator with the drift parameter. Then

$$E_r[\max\{S_T - K, 0\}] = E_\mu[\max\{S_T - K, 0\}L],$$

where the likelihood ratio L is the ratio of the lognormal densities with parameters

r and μ evaluated at S_T, g

$$L =$$

Indeed, S_T need not even requirement is that the su support of the original m is an *absolute continuity* distribution for S_T whose

Ideally, one would like variance. In the example a S_T from the density

$$f(x$$

where g is the (lognormal) f integrate to 1. The dif method requires knowledg gives some indication of t

Reider (1993) has inve change of drift and volat lute continuity in a discre continuous-time limit.) F tion to have higher drift pricing deep out-of-the-m importance sampling wit put-call parity for indirec importance sampling ide

Andersen (1995) has d for simulating interest ra equation models. We br short rate described, e.g.,

is the price today of a ze In, for example, the Cox-

[10] See, e.g., Hull (1993, Chapter

r and μ evaluated at S_T, given by

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

Indeed, S_T need not even be sampled from a lognormal distribution. The only requirement is that the support of the importance sampling measure contain the support of the original measure so that the likelihood ratio is well-defined; this is an *absolute continuity* requirement. In the example above, this means that any distribution for S_T whose support includes $(0, \infty)$ is admissible.

Ideally, one would like to choose the importance sampling distribution to reduce variance. In the example above, one obtains a *zero-variance* estimator by sampling S_T from the density

$$f(x) = c^{-1} \max\{x - K, 0\} e^{-rT} g(x),$$

where g is the (lognormal) density of S_T and c is a normalizing constant that makes f integrate to 1. The difficulty is that c is the Black–Scholes price itself, so this method requires knowledge of the solution for its implementation. Nevertheless, it gives some indication of the potential gain from importance sampling.

Reider (1993) has investigated the impact of importance sampling based on a change of drift and volatility. (Changing the volatility is consistent with absolute continuity in a discrete-time approximation of a diffusion though not in the continuous-time limit.) He finds that choosing the importance sampling distribution to have higher drift and volatility provides substantial variance reduction in pricing deep out-of-the-money options. He also investigates the combination of importance sampling with antithetic variates and control variates, and the use of put-call parity for indirect estimation. Nielsen (1994) has explored some related importance sampling ideas in sampling from a binomial tree.

Andersen (1995) has developed a powerful application of importance sampling for simulating interest rates and has applied it to nonlinear stochastic differential equation models. We briefly describe his approach. Let r_t be the instantaneous short rate described, e.g., by a diffusion model. Then

$$B(T) = E\left[\exp\left(-\int_0^T r_t \, dt\right)\right]$$

is the price today of a zero-coupon bond with face value \$1, maturing at time T. In, for example, the Cox–Ingersoll–Ross and Vasicek models,[10] $B(T)$ is available

---

[10] See, e.g., Hull (1993, Chapter 15) for background on these models.

in closed form. We may therefore define a new probability measure $\bar{P}$ by setting

$$\bar{P}(A) = E\left[\exp\left(-\int_0^T r_t\, dt - \log B(T)\right)\mathbf{1}_A\right]$$

for any event A, where $\mathbf{1}_A$ denotes the indicator of the event A. Let $\bar{E}$ denote expectation with respect to $\bar{P}$. Then for any random variable X, $E[X] = \bar{E}[XL_T]$ where the likelihood ratio L_T is given by

$$L_T = \exp\left(\int_0^T r_t\, dt + \log B(T)\right).$$

In particular, if we take $X = \exp(-\int_0^T r_t\, dt)$, we know that $E[X] = B(T)$ and therefore $B(T)$ is the expectation under $\bar{E}$ of XL_T; i.e., of

$$\exp\left(-\int_0^T r_t\, dt\right) \times \exp\left(\int_0^T r_t\, dt + \log B(T)\right).$$

But this simplifies to $B(T)$ itself, meaning that we obtain a zero-variance estimator of the bond price by switching to the new probability measure. Moreover, Andersen shows that sample paths of r_t can be generated under $\bar{P}$ simply by applying a change of drift to the original process.

As described above, the method would appear to require knowledge of the solution for its implementation. Nevertheless, the method has two important applications. The first is in the pricing of contingent claims. Because $\bar{P}$ eliminates the variance of bond prices, it should be effective in reducing variance for pricing, e.g., European bond options expiring at time T. Andersen's numerical results bear this out. A second application is in the pricing of bond models with no closed-form solutions: Andersen's results show that the change of drift derived from a tractable model (like CIR or Vasicek) remains effective when applied to an intractable model, and this significantly expands the scope of the method.

Importance sampling is frequently used to make rare events less rare; this is already suggested in Reider's (1994) application to out-of-the-money options. Our next example further highlights this aspect through a new application to barrier options. We consider a knock-in option far from the barrier and use importance sampling to increase the probability of a payout.

Suppose the barrier is monitored at discrete times $n\Delta t$, $n = 0, 1, \ldots, m$, with $\Delta T = T/m$. Set the barrier at $H = S_0 e^{-b}$ and the strike at $K = S_0 e^c$, with $b, c > 0$. A down-and-in call pays $S_T - K$ at time T if $S_T > K$ and $S_{n\Delta t} < H$ for some $n = 1, \ldots, m$. We can write the price of the underlying at monitoring instants as

$$S_{n\Delta t} = S_0 e^{U_n}, \quad U_n = \sum_{i=1}^n X_i,$$

with the X_i i.i.d. normal having mean $(r - \frac{1}{2}\sigma^2)\Delta t$ and variance $\sigma^2 \Delta t$. Let τ be the first time U_n drops below $-b$; then the probability of a payout is $P(\tau < m, U_m > c)$. If b and c are large, this probability is small, and most simulation runs return zero. Through importance sampling, we can increase this probability and thus get more information out of each run.

Consider alternative probability measures P_{μ_1, μ_2} that give U_n a drift of $\mu_1 \Delta t$ until τ and then switch the drift to $\mu_2 \Delta t$. Intuitively, we would like to make $\mu_1 < 0$ to drive the asset price to the barrier and then make $\mu_2 > 0$ to drive it above the strike. For any μ_1, μ_2, we have

$$P(\tau < m, U_m > c) = E_{\mu_1, \mu_2}[L_{\mu_1, \mu_2} \mathbf{1}_{\{\tau < m, U_m > c\}}].$$

The likelihood ratio is given by

$$L_{\mu_1, \mu_2} = \exp(-\theta_1 U_\tau + \psi(\theta_1)\tau - \theta_2(U_m - U_\tau) + \psi(\theta_2)(m - \tau)),$$

where $\theta_i = (\mu_i - r + \frac{1}{2}\sigma^2)/\sigma^2$, $i = 1, 2$, and $\psi(\theta) = (r - \frac{1}{2}\sigma^2)\Delta t\theta + \frac{1}{2}\sigma^2 \Delta t \theta^2$. This follows from algebraic simplification of the product of the ratios of the densities of the X_i under the original and new means.

It remains to choose μ_1, μ_2. Intuitively, most of the variability in L_{μ_1, μ_2} comes from τ (the time of the barrier crossing): for large b, c, in the event of a payout we expect to have $U_\tau \approx -b$ and $U_m \approx c$ so these terms should contribute less variability. If we choose μ_1, μ_2 so that $\psi(\theta_1) = \psi(\theta_2)$, the likelihood ratio simplifies to

$$L_{\mu_1, \mu_2} = \exp(-(\theta_1 - \theta_2)U_\tau - \theta_2 U_m + m\psi(\theta_2)),$$

which depends on τ only through $U_\tau \approx -b$. The condition $\psi(\theta_1) = \psi(\theta_2)$ translates to $\mu_1 = -\mu_2 \equiv -\mu$, so it only remains to choose this drift parameter. We choose it so that the time to traverse the straight line path from 0 to $-b$ and then to c at rate μ equals the number of steps m:

$$\frac{b}{\mu \Delta t} + \frac{(b + c)}{\mu \Delta t} = m;$$

i.e., $\mu = (2b + c)/T$. Interestingly, this change of drift does not depend on the original mean increment $(r - \frac{1}{2}\sigma^2)\Delta t$.

Table 4 illustrates the performance of this method. The computational effort with and without importance sampling is essentially the same, so the efficiency improvement is just the ratio of the variances. The improvement varies widely but shows the potential for dramatic gains from importance sampling, particularly when the barrier is far from the current price of the underlying.[11]

---

[11] The standard errors in the table are all quite small, but so are the associated option values. Hence, the *relative* error without importance sampling is quite significant.

Table 4. *Standard errors for down-and-in calls: importance sampling.*

H	K	No variance reduction	Importance sampling	Efficiency ratio
92	100	0.003 09	0.000 69	20
92	105	0.001 29	0.000 14	85
88	96	0.001 10	0.000 11	96
85	90	0.000 84	0.000 08	116
92	105	0.014 18	0.005 41	7
85	105	0.003 28	0.000 38	75
75	96	0.000 30	0.000 01	1124
75	85	0.001 48	0.000 10	222

All results are based on $n = 100\,000$ simulation trials. The parameters are: $S_0 = 95$, $\sigma = 0.15$, and $r = 0.05$, with the barrier H and strike K varying as indicated. The first four cases have $T = 0.25$ and $m = 50$; the last four have $T = 1$ and $m = 250$.

In recent work, Andersen and Brotherton-Ratcliffe (1996) and Beaglehole, Dybvig, Zhou (1997) show how to eliminate the bias caused by using a simulation at a discrete set of times to price continuous options on extrema, e.g., barrier or lookback options.

2.8 Conditional Monte Carlo

This approach to efficiency improvement exploits the variance reducing property of conditional expectation: for any random variables X and Y, $\mathrm{Var}[E[X|Y]] \leq \mathrm{Var}[X]$, with strict inequality except in trivial cases.[12] In replacing an estimator by its conditional expectation we reduce variance essentially because we are doing part of the integration analytically and leaving less to be done by Monte Carlo.

Hull and White (1987) use this idea to price options with stochastic volatilities. Consider a model in which an asset price and its volatility evolve as follows:

$$dS = rS\,dt + vS\,dW_1$$

$$dv^2 = \alpha v^2\,dt + \xi v^2\,dW_2,$$

with W_1, W_2 independent. Suppose we want to price a standard European call on S. A straightforward approach simulates sample paths of v and S up to time T and averages $\max\{S_T - K, 0\}$ over all paths. An alternative notes that, conditional on the path of v_t in $[0, T]$, the asset price S_t may be treated as having a time-varying

---

[12] This is a direct consequence of Jensen's inequality for conditional expectations.

but deterministic volatility. Thus, conditional on the volatility path, the option can be priced by the Black–Scholes formula:

$$e^{-rT} E[\max\{S_T - K, 0\}|v_t, 0 \le t \le T] = \text{BS}(S_0, K, r, T, \sqrt{V_T}),$$

where

$$V_T = \frac{1}{T} \int_0^T v_t^2 \, dt$$

is the average squared volatility over the path, and $\text{BS}(S, K, T, r, \sigma)$ is the Black–Scholes price of a call with constant volatility σ and the other parameters as indicated. Using this conditional expectation as the estimator is sure to reduce variance and may even reduce computational effort since it obviates simulation of S. It is worth emphasizing that both straightforward Monte Carlo and conditional Monte Carlo would have to be applied to discrete-time approximations of the continuous processes above. Also, the applicability of conditional Monte Carlo in this setting relies on the fact that the evolution of the asset price does not influence the volatility path. See Willard (1997) for an extension to the case of correlated W_1 and W_2.

As a further illustration of the use of conditional Monte Carlo, we give a new illustration in the pricing of a down-and-in call with a discretely monitored barrier. Let $0 = t_0 < t_1 < \cdots < t_m = T$ be the monitoring instants and S_{t_i} the price of the underlying at the ith such instant. The option price is $E[e^{-rT} \max\{S_T - K, 0\}\mathbf{1}_{\{\tau_H \le T\}}]$, where H is the barrier and τ_H is the first monitoring time at which the barrier is breached.

Straightforward simulation generates paths of the underlying and evaluates the estimator

$$e^{-rT} \max\{S_T - K, 0\}\mathbf{1}_{\{\tau_H \le T\}}.$$

Our first alternative conditions on $\{S_0, \ldots, S_{\tau_H}\}$, the path of the underlying until the barrier crossing; i.e.,

$$\begin{aligned}
E[e^{-rT} &\max\{S_T - K, 0\}\mathbf{1}_{\{\tau_H \le T\}}] \\
&= e^{-rT} E[E[\max\{S_T - K, 0\}\mathbf{1}_{\{\tau_H \le T\}}|S_0, \ldots, S_{\tau_H}]] \\
&= e^{-rT} E[\text{BS}(S_{\tau_H}, K, r, T - \tau_H, \sigma)\mathbf{1}_{\{\tau_H \le T\}}].
\end{aligned}$$

This yields the estimator

$$\text{CMC}_1 = e^{-rT} \text{BS}(S_{\tau_H}, K, r, T - \tau_H, \sigma)\mathbf{1}_{\{\tau_H \le T\}}$$

This says: simulate until the barrier is crossed or the option expires; if the barrier was crossed, return the Black–Scholes price starting from price S_{τ_H} with maturity $T - \tau_H$.

Our second alternative conditions one step earlier, at each monitoring instant, evaluating the probability that the barrier will be breached for the first time at the next monitoring instant:

$$
E[e^{-rT}\max(S_T - K, 0)\mathbf{1}_{\{\tau_H \leq T\}}] = e^{-rT}E\left[\max\{S_T - K, 0\}\sum_{n=1}^{m}\mathbf{1}_{\{\tau_H = t_n\}}\right]
$$

$$
= e^{-rT}E\left[\sum_{n=1}^{m}E[\max\{S_T - K, 0\}\mathbf{1}_{\{\tau_H = t_n\}}|S_{t_0}, \ldots, S_{t_{n-1}}]\right]
$$

$$
= e^{-rT}E\left[\sum_{n=0}^{\tau_H - 1}\mathrm{BS2}(S_{t_n}, K, H, r, t_{n+1} - t_n, T - t_n, \sigma)\right]
$$

where $\mathrm{BS2}(S, K, H, r, t, T, \sigma)$ is the price of a down-and-in call that knocks in only if the underlying is below H at time t. We thus arrive at the estimator

$$
\mathrm{CMC}_2 = e^{-rT}\sum_{n=0}^{\tau_H - 1}\mathrm{BS2}(S_{t_n}, K, H, r, t_{n+1} - t_n, T - t_n, \sigma),
$$

with

$$
\mathrm{BS2}(S, K, H, r, t, T, \sigma) = SN_2(a_1, b_1, \rho) - e^{-rT}KN_2(a_2, b_2, \rho)
$$

where $\rho = -\sqrt{t/T}$, N_2 is the bivariate cumulative normal distribution with correlation ρ, and

$$
a_1 = \frac{\log(S/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \quad a_2 = a_1 - \sigma\sqrt{T}
$$

$$
b_1 = \frac{\log(H/S) - (r + \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}}, \quad b_2 = b_1 + \sigma\sqrt{t}.
$$

(The derivation of this formula is fairly standard and therefore omitted.) The CMC_2 estimator can be expected to have lower variance than the CMC_1 estimator because it conditions on *less* information and thus does *more* integration analytically. In fact, CMC_2 is not a conditional Monte Carlo estimator in the strict sense because it conditions on different information at different times, making it more precisely a *filtered* Monte Carlo estimator in the sense of Glasserman (1996).

Because the two estimators above have the same expectation, their difference has mean 0 and can be used as a control variate to form a further estimator

$$
\mathrm{CMC}' = \mathrm{CMC}_1 + \beta(\mathrm{CMC}_2 - \mathrm{CMC}_1).
$$

With β optimized, this has lower variance than either individual estimator.

Numerical results appear in Table 5. As expected, each level of conditioning further reduces variance, and the combined estimator achieves the lowest standard

Table 5. *Comparison of CMC estimators for down-and-in call.*

Method	Standard Error (s)	Computation Time (t)	$s\sqrt{t}$
Base	0.108	0.133	0.039
CMC_1	0.034	0.117	0.012
CMC_2	0.021	3.233	0.038
CMC'	0.014	3.367	0.026

Results based on $n = 10\,000$ replications with $\sigma = 0.4$, $r = 0.10$, $S_0 = K = 100$, $H = 95$, $T = 0.5$, and 10 equally spaced monitoring times.

error of all. However, repeated evaluation of the function BS2 turns out to be time-consuming, making CMC_1 overall the most efficient estimator.

3 Low-discrepancy sequences

For complex problems the performance of the basic Monte Carlo approach may be rather unsatisfactory because the error is $O(1/\sqrt{n})$. We can sometimes improve convergence by using pre-selected deterministic points to evaluate the integral. The accuracy of this approach depends on the extent to which these deterministic points are evenly dispersed throughout the domain of integration. *Discrepancy* measures the extent to which the points are evenly dispersed throughout a region: the more evenly dispersed the points are the lower the discrepancy. Low-discrepancy sequences are often called quasi-random sequences even though they are not at all random.[13] We shall use both terms in this paper.

Low-discrepancy methods have recently been used to tackle a number of problems in finance. These applications are more fully described in papers by Birge (1994), Joy, Boyle, and Tan (1996) and Paskov and Traub (1995); the use of quasi-Monte Carlo is also proposed in Cheyette (1992). In this section we describe how the approach works and review some of the recent applications. The book by Press et al. (1992) provides an intuitive introduction to low-discrepancy sequences and quasi-Monte Carlo methods. Spanier and Maize (1994) provide a recent overview of quasi-random methods and how they can be used to evaluate integrals with medium sized samples. Niederreiter (1992) and Tezuka (1995) provide in-depth analyses of low-discrepancy sequences. Moskowitz and Caflisch (1996) discuss recent developments in improving the convergence of quasi-random Monte Carlo methods. In earlier work, Haselgrove (1961) describes a method for multi-

---

[13] Thus the name quasi-random is very misleading since these sequences are deterministic. However, it seems to be sanctioned by usage.

variate integration that can be applied to security pricing. Haselgrove's method is developed for problems of eight dimensions or less and our numerical experiments suggest that it is competitive with the low-discrepancy sequences investigated in this section for problems of this size.

The basic idea behind the approach is quite intuitive and is readily explained in the one-dimensional case. Suppose we wish to integrate a function $f(x)$ over the interval $[0, 1]$ using a sequence of n points. Rather than pick a random sequence suppose we pick a deterministic sequence of points that are, in some sense, evenly distributed. With this choice, the accuracy of the estimate will be higher than that obtained using the crude Monte Carlo approach. If we use an equally spaced grid we obtain the trapezoidal method of numerical integration which has an error of $O(n^{-1})$. However, the more challenging task is to evaluate multi-dimensional integrals. Without loss of generality we can assume that the domain of integration is contained in the d-dimensional unit hypercube. The advantages of the uniformly spaced grid in the one-dimensional case do not carry over to higher dimensions. The principal reason is that the error bound for the d-dimensional trapezoidal rule is $O(n^{-2/d})$. In addition, if we use an evenly spaced Cartesian grid, we would have to decide the number of points in advance to achieve uniformity. This is restrictive because, in numerical applications, we would like to be able to add points sequentially until some termination criterion is met.

Low-discrepancy sequences have the property that as successive points are added the entire sequence of points still remains more or less evenly dispersed throughout the region. Niederreiter (1992) gives a detailed analysis of the discrepancy of a sequence. Here, we just briefly recall the definition. Suppose we have a sequence of n points $\{x_1, x_2, \ldots, x_n\}$ in the d-dimensional half-open unit cube, $I^d = [0, 1)^d$ and a subset J of I^d. We define

$$D(J; n) = \frac{A(J; n)}{n} - V(J),$$

where $A(J; n)$ is the number of k, $1 \leq k \leq n$, with $x_k \in J$ and $V(J)$ is the volume of J. The discrepancy, D_n, of the sequence is defined to be the supremum of $|D(J; n)|$ over all J. The star discrepancy D_n^*, is obtained by taking the supremum over sets J of the form

$$\prod_{i=1}^{d} [0, u_i).$$

In the one-dimensional case there is a simple explicit form for the (star)[14] discrepancy of a sequence of n points. If we label the points so that, $0 \leq x_1 \leq \cdots \leq$

---

[14] For the rest of the paper we simply use the term discrepancy rather than star discrepancy to refer to D_n^*.

$x_n \leq 1$, then the discrepancy

$$D_n^* =$$

We can see that the star disc attained when

$$J$$

In higher dimensions there i

There are several examp quences proposed by Halton (1988).[15] For these sequenc shown to be

This bound for the discrep the dimension d of the sec accurately in high dimens ridiculously large for reas (1994, p. 23). Furthermor the discrepancy reaches its that for intermediate value that the transition to $O(n^-$ d this will be an enormous

The error in numerical determinstic bound. The points used to evaluate the result is contained in the f

Theorem (Koksma–Hlav $V(f)$ on $[0, 1]^d$ in the Har have

$$\left| \frac{1}{n} \sum_{k=}^{n} \right.$$

---

[15] Interestingly, linear congruentia drive ordinary Monte Carlo – generator; see Niederreiter (197 roughly equal to the total numl Monte Carlo, one prefers inste points required. We thank Peter

[16] For a more complete discussion Niederreiter (1992).

$x_n \leq 1$, then the discrepancy of this sequence is

$$D_n^* = \frac{1}{2n} + \max_{k=1,\ldots,n} \left| x_k - \frac{2k-1}{2n} \right|.$$

We can see that the star discrepancy is at least $1/(2n)$ and that the lowest value is attained when

$$x_k = \frac{2k-1}{2n}, \quad 1 \leq k \leq n.$$

In higher dimensions there is no simple form for the discrepancy of a sequence.

There are several examples of low-discrepancy sequences, including the sequences proposed by Halton (1960), Sobol' (1967), Faure (1982), and Niederreiter (1988).[15] For these sequences the asymptotic form of the star discrepancy has been shown to be

$$D_n^* = O\left(\frac{(\log n)^d}{n} \right).$$

This bound for the discrepancy involves a constant which in general depends on the dimension d of the sequence. These constants are very difficult to estimate accurately in high dimensions. For large values of d the constants "are often ridiculously large for reasonable values of n" according to Spanier and Maize (1994, p. 23). Furthermore for high dimensions it may take a long time before the discrepancy reaches its asymptotic level. Morokoff and Caflisch (1995) note that for intermediate values of n the discrepancy may be $O(\sqrt{n})$. They suggest that the transition to $O(n^{-1}(\log n)^d)$ occurs at around values of $n = e^d$. For large d this will be an enormous number.

The error in numerical integration using a low-discrepancy sequence admits a deterministic bound. The bound reflects both the discrepancy of the sequence of points used to evaluate the integral as well as the regularity of the function. The result is contained in the following theorem.

Theorem (Koksma–Hlawka) *Let $I^d = [0, 1)^d$ and let f have bounded variation $V(f)$ on $[0, 1]^d$ in the Hardy–Krause[16] sense. Then for any $x_1, x_2, \ldots, x_n \in I^d$ we have*

$$\left| \frac{1}{n} \sum_{k=1}^{n} f(x_k) - \int_{I^d} f(u)\, du \right| \leq V(f) D_n^*.$$

---

[15] Interestingly, linear congruential generators – frequently used to generate the pseudo-random numbers that drive ordinary Monte Carlo – produce sets of points with low-discrepancy over the entire period of the generator; see Niederreiter (1976). This suggests the possibility of choosing such a generator with period roughly equal to the total number of points required as a type of quasi-Monte Carlo method. In ordinary Monte Carlo, one prefers instead that the period be many orders of magnitude larger than the number of points required. We thank Peter Hellekalck of the University of Salzburg for this observation.

[16] For a more complete discussion of the Hardy–Krause definition of variation and details on this theorem see Niederreiter (1992).

The error bound provided by this theorem, while it is of theoretical interest, is of little help in most practical situations. The theoretical bound normally overestimates the actual error by a wide margin and $V(f)$ may be difficult to evaluate or even approximate. We have noted that the constants buried in the bounds for the discrepancy are large. Another reason for the coarseness of the bound is that the Koksma–Hlawka theorem does not reflect additional smoothness in f. Intuitively we would expect the approximation to be better as f becomes smoother. In finance applications the payoffs are normally continuous functions of the variables (with some important exceptions – payoffs on digital and barrier options are discontinuous), but may not be sufficiently smooth to have finite variation because of functions like "max" embedded in the payoffs. Hlawka (1971) provides an alternative bound under weaker smoothness requirements.

To date, studies using low-discrepancy sequences in finance applications find that the errors produced are substantially lower than the corresponding errors generated by crude Monte Carlo. Joy, Boyle, and Tan (1996) used Faure sequences to price several complex derivative securities. They found that the quasi-Monte Carlo approach resulted in significantly smaller errors than the standard Monte Carlo approach. They confirmed that the actual error bound (for cases in which it could be computed precisely) was dramatically less than the bound computed from the Koksma–Hlawka inequality. Paskov and Traub (1995) used both Sobol' sequences and Halton sequences to evaluate mortgage-backed security prices. Their work involves the evaluation of integrals with dimensions up to 360; they find that Sobol' sequences are more efficient than Halton sequences and that the quasi-random approach outperforms the standard Monte Carlo approach for these types of problems.[17] Paskov and Traub's results stand in contrast to the claim that is sometimes found in the literature[18] that the superiority of low-discrepancy algorithms vanishes for intermediate values of d around 30. Bratley, Fox, and Niederreiter (1992) conducted practical numerical experiments using low-discrepancy sequences and conclude that standard Monte Carlo is superior to quasi-Monte Carlo for high dimensions, say greater than 12. They used Sobol' and Niederreiter sequences in their tests. They conclude that in high dimensions, "quasi-Monte Carlo seems to offer no practical advantage over pseudo-Monte Carlo because the discrepancy bound for the former is far larger than $\sqrt{n}$ for $n = 2^{30}$, say." (In a personal communication, Fox adds that the crossover probably depends a lot on the sequence.) The reason for the difference between this verdict and the results of the finance applications may be that the integrands typically found in finance applica-

---

[17] Bratley et al. (1992) note that the Niederreiter sequence they tested theoretically beats Sobol' sequences in dimensions higher than seven.

[18] See, for example, Rensburg and Torrie (1993) or Morokoff and Caflisch (1995).

tions behave better than those used by numerical analysts[19] to compare different algorithms. Another important consideration is that financial applications typically involve discounting, and this may effectively reduce dimensionality; for example, some of the 360 months in the life of a mortgage may have little influence on the value of a mortgage-backed security. Nevertheless, the experience of Bratley et al. (1992) serves as a useful caution against assuming that quasi-Monte Carlo will outperform standard Monte Carlo in all situations.

Some theoretical differences among low-discrepancy sequences can be understood through the concepts of (t, m, s)-nets and (t, s)-sequences; these are discussed in detail in Niederreiter (1992). Briefly, an elementary interval in base b in dimension s is a set of the form

$$\prod_{j=1}^{s} \left[\frac{a_j}{b^{k_j}}, \frac{a_j + 1}{b^{k_j}} \right),$$

with k_j, a_j nonnegative integers and $a_j < b^{k_j}$. A (t, m, s)-net (with $0 \le t \le m$) is a set of b^m points in the s-dimensional hypercube such that every elementary interval of volume b^{t-m} contains b^t points. Speaking loosely, this means that the proportion of points in each sufficiently large box equals the volume of the box. Smaller t implies greater uniformity. An infinite sequence forms a (t, s)-sequence if for all $m \ge t$ certain finite subsequences of length b^m form (t, m, s)-nets in base b. Sobol' points are (t, s)-sequences in base 2 and Faure points are $(0, s)$ sequences in prime bases not less than s. Thus, Faure points achieve the smallest value of t, but at the expense of a large base. A smaller base implies that uniformity holds over shorter subsequences.

An important issue in the use of quasi-Monte Carlo concerns the termination criterion, since the Koksma–Hlawka bound is often of little practical value. Various heuristics are available. Birge (1994) suggests that a rough bound may be obtained by tracking the maximum and minimum values over a period that shows equal numbers of increases and decreases. For instance the criterion could be to stop at the first set of two thousand observations in which the number of increases and decreases are within ten percent of each other. He suggests that the maximum and minimum realized values could be used as bounds on the true value. Fox (1986) suggests that we compare the estimate of the integral based on a sample of $2n$ points with the estimate based on n points and stop if the answer lies within some tolerance level. Paskov and Traub (1995) use a similar termination criterion based

---

[19] For example, one of the integrals used by Bratley, Fox, and Niederreiter (1992) was

$$\int_0^1 \cdots \int_0^1 \prod_{k=1}^{d} k \cos(k x_k) dx_1 \cdots dx_d.$$

This integrand is highly periodic for large values of d.

on successive errors: stop when the difference between two consecutive approxi-
mations using $10\,000i$, $i = 1, 2, \ldots, 1000$, sample points falls below some thresh-
old. Owen (1995a, 1995b) proposes a hybrid of Monte Carlo and low-discrepancy
methods which provides error estimates and has good convergence properties. In
addition to these approaches, one can also run standard Monte Carlo at the outset
and use the probabilistic error term to assess when enough low-discrepancy points
have been used in the quasi-random calculation. This benchmarking with standard
Monte Carlo would be useful if the same set of calculations were being carried out
frequently with only slightly different input values. This situation is common in
finance applications. There is often a need to perform the same set of calculations
frequently; e.g., the risk analysis of a book of business at the end of each day.
In these cases one can conduct experiments to see which sets of low-discrepancy
sequences provide the best results. The *right* number of low-discrepancy points
could be determined just once at the outset.

Before leaving this section, we should mention some recent advances and new
techniques to improve the performance of quasi-random Monte Carlo. Niederreiter
and Xing (1996), Tezuka (1994), and Ninomiya and Tezuka (1996) have proposed
new low-discrepancy sequences that appear to have the potential to perform sub-
stantially better than previous methods. We have noted that the efficiency of quasi-
random Monte Carlo improves as the integrand becomes smoother. Moskowitz
and Caflisch (1996) illustrate procedures that can be used for this purpose. It is
sometimes possible to enhance the performance of quasi-random sequences by
reducing the effective dimension of the problem. Moskowitz and Caflisch also
indicate how this can be accomplished in the discretization of a Wiener process
and in the solution of the Feynman–Kac equation. This is relevant for finance
applications since the prices of derivative securities have a Feynman–Kac repre-
sentation. See Acworth, Broadie, and Glasserman (1997), Berman (1996), and
Caflisch, Morokoff, and Owen (1998) for recent work applying low-discrepancy
sequences with alternative constructions of Wiener processes. Spanier and Maize
(1994) discuss a battery of techniques that can be used to improve the performance
of quasi-Monte Carlo methods for relatively small sample sizes.

Next we compare the Monte Carlo method using pseudo-random numbers with
the Faure, Halton, and Sobol' low-discrepancy methods.

3.1 Numerical results

For an initial comparison, we test the methods on the problem of pricing a Eu-
ropean option on a single underlying asset with the usual Black–Scholes assump-
tions. In this framework, the Black–Scholes formula can be evaluated to give the
true option values in order to compare alternative methods. Rather than using

a single option, we evalu
The probability distributio
range of values in practi
root-mean-squared (RMS)

where i is the index of the
value, and $\hat{C}_i$ is the estima

Figure 1 plots RMS r
Monte Carlo method (i.e.,
$O(1/\sqrt{n})$ convergence: e
error by a factor of 10. Th
nates the Monte Carlo met
1000 Monte Carlo points.
tested. Using 192 Sobol'
points.

A major consideration i
time, not just the number o
significantly faster than F
fast as most pseudo-rando
versus computation time c
increases.

A low-discrepancy sequ
certain points in the seque
example, in the Sobol' se
$n = 2^k - 1$ for $k = 1, 2, .$
"favorable" points would
n, the advantage of using
effect can be quite signifi
points 1 through 254 gives
255 gives an RMS error o
initial portion of a low-dis
1 through 63 gives an RM
127 gives an RMS error
was always started at poin
Sobol' points from 64 to 2

[20] The details of the distribution a
[21] We take the first point of the Sol

a single option, we evaluate the methods on a random sample of 500 options. The probability distribution of the parameters is chosen to represent a reasonable range of values in practical applications.[20] The error measure that we use is root-mean-squared (RMS) relative error defined by

$$\text{RMS} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left(\frac{\hat{C}_i - C_i}{C_i} \right)^2}, \tag{12}$$

where i is the index of the $m = 500$ options in the test set, C_i is the true option value, and $\hat{C}_i$ is the estimated option value. The results are given in Figure 1.

Figure 1 plots RMS relative error against the number of points, n. The Monte Carlo method (i.e., using pseudo-random numbers) displays the expected $O(1/\sqrt{n})$ convergence: e.g., increasing n by a factor of 100 decreases the RMS error by a factor of 10. The low-discrepancy method using Faure sequences dominates the Monte Carlo method. Indeed, 129 Faure points gives an error lower than 1000 Monte Carlo points. The Sobol' method is the best of the three methods tested. Using 192 Sobol' points gives an error lower than 10 000 Monte Carlo points.

A major consideration in the comparison of methods is the overall computation time, not just the number of points. The Sobol' sequence numbers can be generated significantly faster than Faure numbers (see, e.g., Bratley and Fox 1988) and as fast as most pseudo-random number methods. Hence, in the important RMS error versus computation time comparison, the relative advantage of the Sobol' method increases.

A low-discrepancy sequence will often have additional uniformity properties at certain points in the sequence (see, e.g., Fox 1986 and Bratley and Fox 1988). For example, in the Sobol' sequence the running average returns to 0.5 at the points $n = 2^k - 1$ for $k = 1, 2, \ldots$. One might expect that choosing n to be one of these "favorable" points would lead to better option price estimates. For large values of n, the advantage of using favorable points becomes negligible, but for small n the effect can be quite significant. Indeed, in the experiment above, using the Sobol' points 1 through 254 gives an RMS error of 10%, while using the points 1 through 255 gives an RMS error of 4%.[21] Better results are often obtained by ignoring an initial portion of a low-discrepancy sequence. For example, using the Sobol' points 1 through 63 gives an RMS error of 13%, while using the Sobol' points 64 through 127 gives an RMS error of 2%. In the results in Figure 1, the Sobol' sequence was always started at point 64, so the label 192 in Figure 1 corresponds to the 192 Sobol' points from 64 to 255. Similarly, the Faure sequence was always started at

---

[20] The details of the distribution are given in Broadie and Detemple (1996).
[21] We take the first point of the Sobol' sequence to be 0.5, not 0.0.
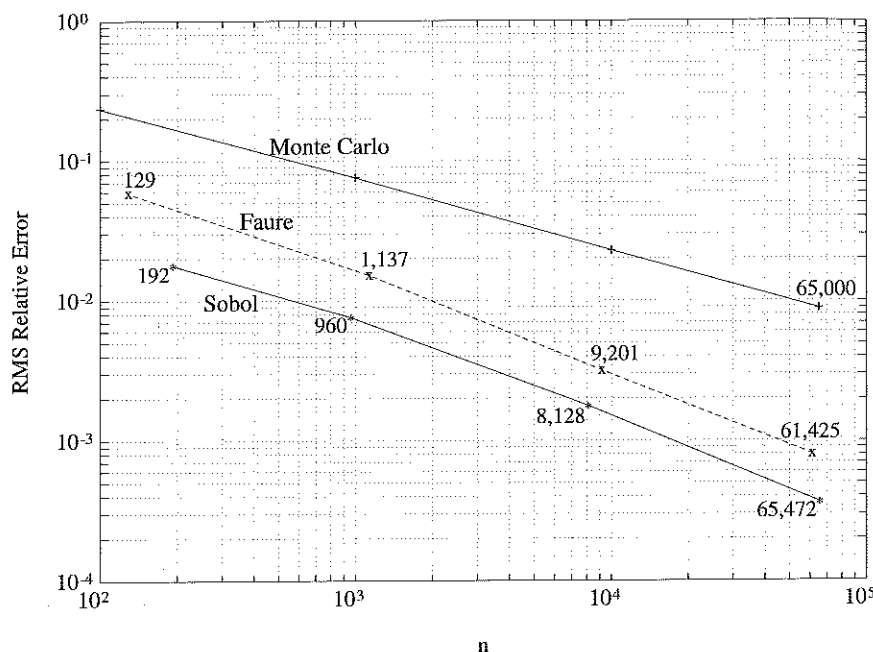
Fig. 1. RMS relative error vs. number of points.

point 16, so the label 129 in Figure 1 corresponds to the 129 Sobol' points from 16 to 144.

3.2 One-dimensional vs. higher dimensional sequences

It is sometimes asserted that low-discrepancy methods can be implemented in existing simulation programs by simply replacing the pseudo-random number generator with a low-discrepancy sequence generator. This naive approach can lead to disastrous results as the following example shows.

Consider pricing a European option on the maximum of two non-dividend paying assets with the parameters: $S_1 = S_2 = K = 100$, $\sigma_1 = \sigma_2 = 0.2$, $\rho = 0.3$, $r = 0.05$, and $T = 1$. Under the usual Black–Scholes assumptions, a formula for the price of the option can be derived (see, e.g., Johnson 1987 or Stulz 1982) and gives a price of 16.442. Running one Monte Carlo simulation with 1000 points (hence 2000 random numbers) gave an estimated price of 16.279 with a standard error of 0.533. Using 2000 one-dimensional low-discrepancy values gave a price estimate of 4.320 using the Sobol' sequence and an estimate of 1.909 using the

Fig. 2. 1000 two-dimensional Faure points.

Faure sequence (starting at point 16). The cause of the problem can be seen by examining Figures 2–5.

Figures 2 and 3 show 1000 two-dimensional Faure and Sobol' points, respectively. The figures illustrate how the sequences fill the two-dimensional space in regular but different ways. By contrast, Figures 4 and 5 show 2000 one-dimensional Faure and Sobol' points, respectively, plotted in two dimensions. The plots are created by taking successive points in the one-dimensional sequence to be the (x, y) coordinates in two-dimensional space. In neither figure are the points filling the two-dimensional space (note that the axes do not extend from 0 to 1) and this explains why the price estimates do not converge to the correct values. Even in the quarter of the unit square where the points fall, the points do not uniformly fill the space. This problem is reminiscent of the well-known "collinearity" or "hyperplane" problem of some pseudo-random number generators, but is even more serious with these low-discrepancy sequences.

A similar problem can occur if a high-dimensional low-discrepancy sequence is used for a problem of low dimension. Figure 6 shows the 49th and 50th dimension of 1000 50-dimensional Faure points. Using the last two dimensions of the 50-dimensional sequence to price a two-dimensional option will give very poor results.

3.3 *Higher dimensional test*

To test the effect of problem dimension, we price options in dimensions $d = 10, 50$, and 100. We price discretely sampled geometric average Asian options, because the problem dimension is easily varied and a closed form solution for the price

P. Boyle, M. Broadie and P. Glasserman
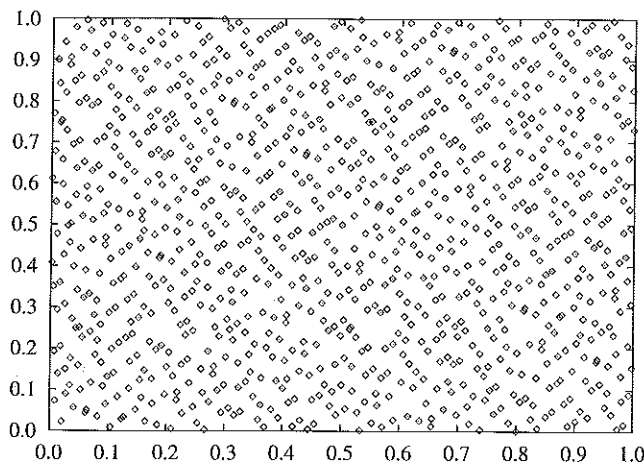
Fig. 3. 1000 two-dimensional Sobol' points.

Fig. 5. 20
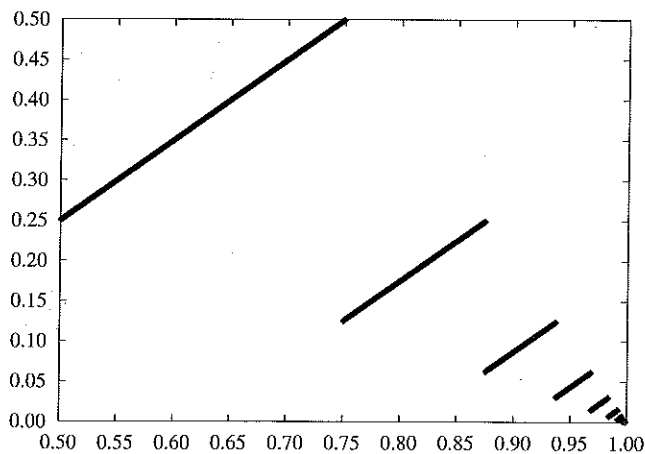
Fig. 4. 2000 one-dimensional Faure points.

Fig. 6. Coordinates 4

is available (see Turnbull and Wakeman 1991). The price of a geometric average Asian option is given by

$$C = E[e^{-rT}(\tilde{S} - K)^+],$$

where $\tilde{S} = (\prod_{i=1}^{d} S_i)^{1/d}$ and S_i is the asset price at time iT/d.

We test standard Monte Carlo, Monte Carlo with antithetic variates, and the low-discrepancy sequences of Faure, Sobol', and Halton.[22] For each dimension, we select 500 option parameters at random, and compute RMS relative error (see

equation 12) for each method
given in Figures 7 and 8, re
100 000 independent pairs of

Results for the Halton sequ
error for standard Monte Car
The antithetic method gives n
terms of RMS error, of the lo
dimension. For this test prob

---

[22] We thank Spassimir Paskov and Joseph Traub for providing their code for the Sobol' sequences.

[23] The details of the distribution are giv
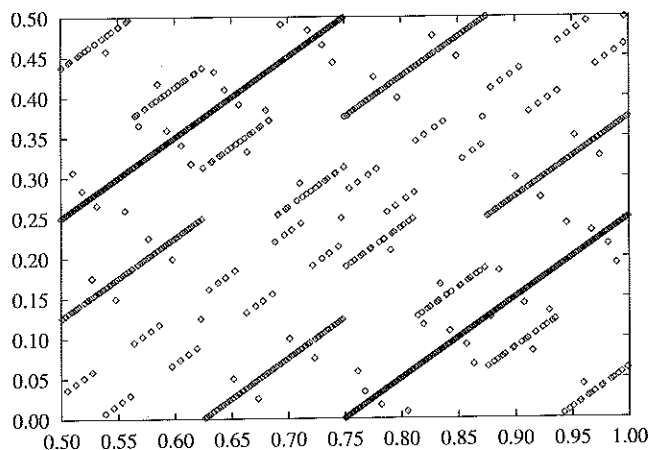
Fig. 5. 2000 one-dimensional Sobol' points.
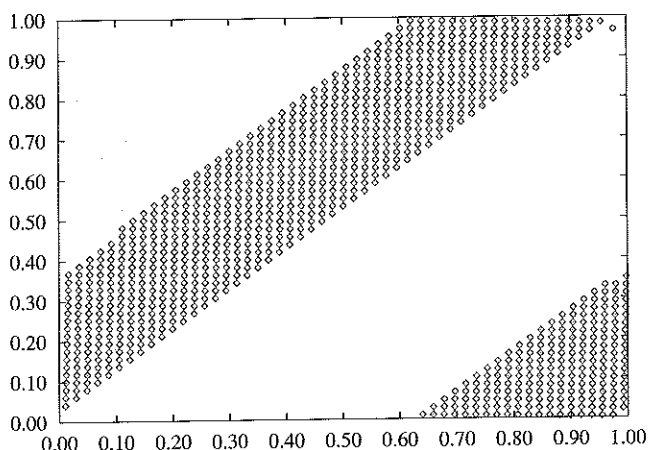
Fig. 6. Coordinates 49 and 50 of 1000 50-dimensional Faure points.

equation 12) for each method.[23] Results for 50 000 and 200 000 sample points are given in Figures 7 and 8, respectively. (The antithetic method uses 25 000 and 100 000 independent pairs of points, respectively.)

Results for the Halton sequence were not competitive and are suppressed. RMS error for standard Monte Carlo is nearly independent of the problem dimension. The antithetic method gives minimal variance reduction. The relative advantage, in terms of RMS error, of the low-discrepancy sequences decreases with the problem dimension. For this test problem, the crossover point is beyond dimension 100.

[23] The details of the distribution are given in Broadie and Detemple (1996).
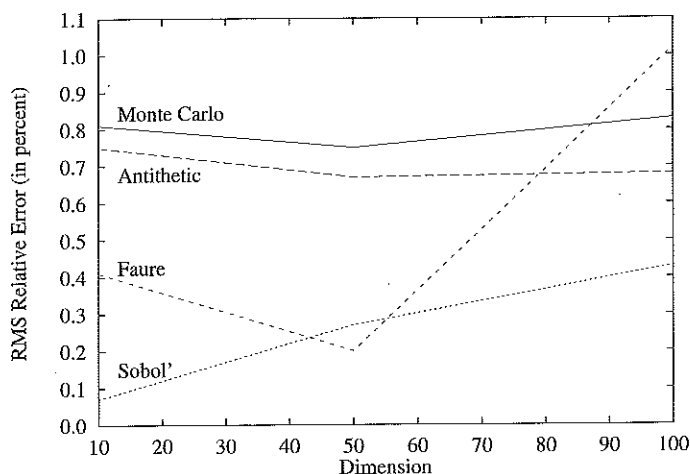
Fig. 7. Results with 50 000 points.
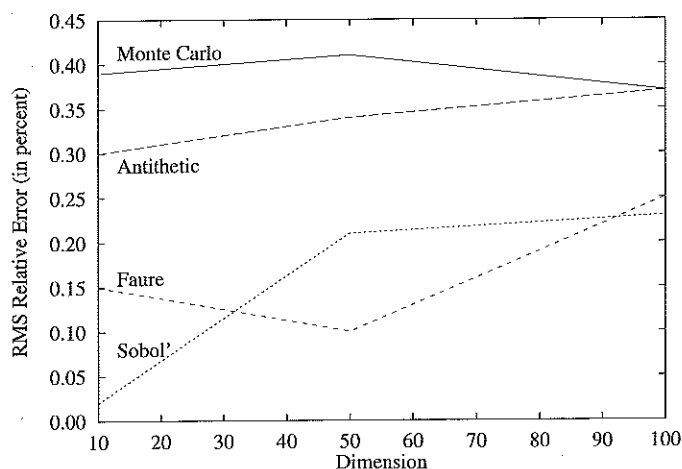
Fig. 8. Results with 200 000 points.

4 Estimating price sensitivities

Most of the discussion in this paper centers on the use of Monte Carlo for pricing
securities. In practice, the evaluation of price *sensitivities* is often as important as
the evaluation of the prices themselves. Indeed, whereas prices for some securities

can be observed in the market, their sensitivities to parameter changes typically cannot and must therefore be computed. Since price sensitivities are important measures of risk, the growing emphasis on risk management systems suggests a greater need for their efficient computation.

The derivatives of a derivative security's price with respect to various model parameters are collectively referred to as *Greeks*, because several of these are commonly referred to with the names of Greek letters.[24] Perhaps the most important of these – and the one to which we give primary attention – is *delta*: the derivative of the price of a contingent claim with respect to the current price of an underlying asset. The delta of a stock option, for example, is the derivative of the option price with respect to the current stock price. An option involving multiple underlying assets has multiple deltas, one for each underlying asset.

In the rest of this section, we discuss various approaches to estimating price sensitivities, especially delta. We begin by examining finite-difference approximations and show that these can be improved through the use of *common random numbers*. We then discuss *direct methods* that estimate derivatives without requiring resimulation at perturbed parameter values.

4.1 Finite-difference approximations

Consider the problem of computing the delta of the Black–Scholes price of a European call; i.e., computing

$$\Delta = \frac{dC}{dS_0},$$

where C is the option price and S_0 is the current stock price. There is, of course, an explicit expression for delta, so simulation is not required, but the example is useful for purposes of illustration. A crude estimate of delta is obtained by generating a terminal stock price

$$S_T = S_0 e^{(r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z} \tag{13}$$

(see (2) for notation) from the current stock price S_0 and a second, independent terminal stock price

$$S_T(\epsilon) = (S_0 + \epsilon)e^{(r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z'} \tag{14}$$

from the perturbed initial price $S_0 + \epsilon$, with Z and Z' independent. For each terminal price, a discounted payoff can be computed like this:

$$\hat{C}(S_0) = e^{-rT}\max\{0, S_T - K\}, \quad \hat{C}(S_0 + \epsilon) = e^{-rT}\max\{0, S_T(\epsilon) - K\}$$

---

[24] See, e.g., Chapter 13 of Hull (2000) for background.

(see (3) for notation). A crude estimate of delta is then provided by the finite-difference approximation

$$\tilde{\Delta} = \epsilon^{-1}[\hat{C}(S_0 + \epsilon) - \hat{C}(S_0)]. \tag{15}$$

By generating n independent replications of S_T and $S_T(\epsilon)$ we can calculate the sample mean of n independent copies of $\tilde{\Delta}$. As $n \to \infty$, this sample mean converges to the true finite-difference ratio

$$\epsilon^{-1}[C(S_0 + \epsilon) - C(S_0)], \tag{16}$$

where $C(\cdot)$ is the option price as a function of the current stock price.

This discussion suggests that to get an accurate estimate of Δ we should make ϵ small. However, because we generated S_T and $S_T(\epsilon)$ independently of each other, we have

$$\text{Var}[\tilde{\Delta}] = \epsilon^{-2}(\text{Var}[\hat{C}(S_0 + \epsilon) + \text{Var}[C(S_0)]]) = O(\epsilon^{-2}),$$

so the variance of $\tilde{\Delta}$ becomes very large if we make ϵ small. To get an estimator that converges to Δ we must let ϵ decrease slowly as n increases, resulting in slow overall convergence. A general result of Glynn (1989) shows that the best possible convergence rate using this approach is typically $n^{-1/4}$. Replacing the forward difference estimator in (15) with the central difference $(2\epsilon)^{-1}[\hat{C}(S_0 + \epsilon) - \hat{C}(S_0 - \epsilon)]$ typically improves the optimal convergence rate to $n^{-1/3}$. These rates should be compared with $n^{-1/2}$, the rate ordinarily expected from Monte Carlo.

Better estimators can generally be improved using the method of *common random numbers*, which, in this context, simply uses the same Z in (13) and (14). Denote by $\hat{\Delta}$ the finite-difference approximation thus obtained. For fixed ϵ, the sample mean of independent replications of $\hat{\Delta}$ also converges to (16). The variance parameter is given by

$$\text{Var}[\hat{\Delta}] = \epsilon^{-2}(\text{Var}[\hat{C}(S_0)] + \text{Var}[\hat{C}(S_0 + \epsilon)] - 2\,\text{Cov}[\hat{C}(S_0), \hat{C}(S_0 + \epsilon)]),$$

because $\hat{C}(S_0)$ and $\hat{C}(S_0 + \epsilon)$ are no longer independent. Indeed, if they are positively correlated, then $\hat{\Delta}$ has smaller variance than $\tilde{\Delta}$. That they are in fact positively correlated follows from the monotonicity of the function mapping Z to $\hat{C}$ by the argument used in our discussion of antithetics in Section 3. Thus, the use of common random numbers reduces the variance of the estimate of delta.

The impact of this variance reduction is most dramatic when ϵ is small. A simple calculation shows that, using common random numbers,

$$|\hat{C}(S_0 + \epsilon) - \hat{C}(S_0)| \le |S_T(\epsilon) - S_T|$$

$$\le \epsilon e^{(r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z}.$$

Because this upper bound ha

$$E[|\hat{C}$$

and therefore that

$$\text{Var}[\epsilon^{-}$$

i.e., the variance of $\hat{\Delta}$ rema that the variance of $\tilde{\Delta}$ incr to estimate Δ (by making numbers. Moreover, this in we may let ϵ decrease faste in faster overall convergenc Perron (1994) shows that a c and that is the best that can on convergence rates using (1992), Glynn (1989), and L

The dramatic success of c fast rate of mean-square co This rate does not apply in a *digital option*[25] paying a fix this option is $C = e^{-rT}BP($

Because $\hat{C}(S_0)$ and $\hat{C}(S_0 +$

$$E[|\hat{C}(S_0 + \epsilon) - \hat{C}(S_0)|^2$$

compared with $O(\epsilon^2)$ for difficult for the digital opti generally. Even in these ca substantial improvement co

Table 6 compares the per central finite-differences wi are compared at four values options discussed above. T errors. The numerical resu are obtained for the standar central differences beat for

[25] Also called a "binary" or "cash-o

Because this upper bound has finite second moment, we may conclude that

$$E[|\hat{C}(S_0 + \epsilon) - \hat{C}(S_0)|^2] = O(\epsilon^2), \tag{17}$$

and therefore that

$$\text{Var}[\epsilon^{-1}\{\hat{C}(S_0 + \epsilon) - \hat{C}(S_0)\}] = O(1);$$

i.e., the variance of $\hat{\Delta}$ remains *bounded* as $\epsilon \to 0$, whereas we saw previously that the variance of $\tilde{\Delta}$ increases at rate ϵ^{-2}. Thus, the more precisely we try to estimate Δ (by making ϵ small) the greater the benefit of common random numbers. Moreover, this indicates that to get an estimator that converges to Δ we may let ϵ decrease faster as n increases than was possible with $\tilde{\Delta}$, resulting in faster overall convergence. An application of Proposition 2 of L'Ecuyer and Perron (1994) shows that a convergence rate of $n^{-1/2}$ can be achieved in this case, and that is the best that can ordinarily be expected from Monte Carlo. For more on convergence rates using common random numbers see Glasserman and Yao (1992), Glynn (1989), and L'Ecuyer and Perron (1994).

The dramatic success of common random numbers in this example relies on the fast rate of mean-square convergence of $\hat{C}(S_0 + \epsilon)$ to $\hat{C}(S_0)$ evidenced by (17). This rate does not apply in all cases. It fails to hold, for example, in the case of a *digital option*[25] paying a fixed amount B if $S_T > K$ and 0 otherwise. The price of this option is $C = e^{-rT}BP(S_T > K)$; the obvious simulation estimator is

$$\hat{C}(S_0) = \mathbf{1}_{\{S_T > K\}}e^{-rT}B.$$

Because $\hat{C}(S_0)$ and $\hat{C}(S_0 + \epsilon)$ differ only when $S_T \leq K < S_T(\epsilon)$, we have

$$E[|\hat{C}(S_0 + \epsilon) - \hat{C}(S_0)|^2] = B^2 e^{-2rT} P(S_T \leq K < S_T(\epsilon))$$
$$= B^2 e^{-2rT} P(S_T \leq K < (1 + \epsilon/S_0)S_T) = O(\epsilon),$$

compared with $O(\epsilon^2)$ for a standard call. As a result, delta estimation is more difficult for the digital option, and a similar argument applies to barrier options generally. Even in these cases, the use of common random numbers can result in substantial improvement compared with differences based on independent runs.

Table 6 compares the performance of four types of delta estimates: forward and central finite-differences with and without common random numbers. The methods are compared at four values of the perturbation parameter ϵ, and applied to the two options discussed above. The values in the table are estimated root mean square errors. The numerical results substantiate the analysis above. Much lower errors are obtained for the standard call than for the digital option, allowing for smaller ϵ; central differences beat forward differences; common random numbers helps, but

---

[25] Also called a "binary" or "cash-or-nothing" option; see Hull (2000, p. 464).

Table 6. *RMS errors for various delta estimation methods.*

| | ϵ | Independent | | Common | |
		Forward	Central	Forward	Central
Standard	10	0.10	0.01	0.100	0.009
Call	1	0.18	0.09	0.012	0.006
Option	0.1	1.78	0.87	0.006	0.006
	0.01	7.47	8.98	0.006	0.006
Digital	20	0.51	0.37	0.51	0.37
Option	10	0.22	0.11	0.21	0.10
	5	0.16	0.07	0.11	0.05
	1	0.67	0.34	0.14	0.10

Root mean square error of delta estimates for two options using four methods with various values of ϵ. Both options have $S_0 = 100$, $K = 100$, $\sigma = 0.40$, $r = 0.10$, and $T = 0.2$. The digital option has $B = 100$. Each entry is computed from 1000 delta estimates, each estimate based on 10 000 replications. The value of delta is 0.580 for the first option and 2.185 for the second.

it helps the standard call more than the digital option. In several cases, the minimal error is obtained using a fairly large ϵ. This reflects the fact that the bias resulting from a large ϵ is sometimes overwhelmed by the large variance resulting from a small ϵ.

Although we have discussed common random numbers in only a limited context, it can easily be applied to a wide range of problems. If all stochastic inputs to a simulation are samples from the normal distribution, then common random numbers can be implemented by using the same samples at two different parameter settings. More generally, if the stochastic inputs are all drawn from a sequence of uniform random variates, then common random numbers can be implemented by using these variates at two different parameter settings.

4.2 Direct estimates

Even with the improvements in performance obtained from common random numbers, derivative estimates based on finite differences still suffer from two shortcomings. They are biased (since they compute difference ratios rather than derivatives) and they require multiple resimulations: estimating sensitivities to d parameter changes requires repeatedly running one simulation with all parameters at their base values and d additional simulations with each of the parameters perturbed.

The computation of 10–50
this represents a significant
required.

Over the last decade, a
timating derivatives by sim
from a single simulation, a
rameter value. Under appr
the derivatives themselves,
focuses on the use of *pathw*
generally called *infinitesim*

The pathwise estimate c
price $\hat{C}$ with respect to S_0.

$$\frac{d\hat{C}}{dS_0}$$

provided the limit exists w
from the same Z, then pro

We have used (13) to get

$$\frac{}{}$$

and

$$\frac{d\hat{C}}{dS_T} = e^{-rT}\frac{}{d}$$

At $S_T = K$, C fails to be d
zero, the random variable

The pathwise derivative
common random numbers
analytically rather than n
because it can be comput
need for a separate simul
expression in (18). The q

[26] Sensitivities to various changes

The computation of 10–50 Greeks[26] for a single security is not unheard of, and this represents a significant computational burden when multiple resimulations are required.

Over the last decade, a variety of *direct* methods have been developed for estimating derivatives by simulation. Direct methods compute a derivative estimate from a single simulation, and thus do not require resimulation at a perturbed parameter value. Under appropriate conditions, they result in unbiased estimates of the derivatives themselves, rather than of a finite-difference ratio. Our discussion focuses on the use of *pathwise* derivatives as direct estimates, based on a technique generally called *infinitesimal perturbation analysis* (see, e.g., Glasserman 1991).

The pathwise estimate of the true delta dC/dS_0 is the derivative of the sample price $\hat{C}$ with respect to S_0. More precisely, it is

$$\frac{d\hat{C}}{dS_0} = \lim_{\epsilon \to 0} \epsilon^{-1} [\hat{C}(S_0 + \epsilon) - \hat{C}(S_0)],$$

provided the limit exists with probability 1. If $\hat{C}(S_0)$ and $\hat{C}(S_0 + \epsilon)$ are computed from the same Z, then provided $S_T \neq K$, we have

$$\begin{aligned}
\frac{d\hat{C}}{dS_0} &= \frac{d\hat{C}}{dS_T}\frac{dS_T}{dS_0} \\
&= e^{-rT}\mathbf{1}_{\{S_T > K\}}\frac{S_T}{S_0}.
\end{aligned} \tag{18}$$

We have used (13) to get

$$\frac{dS_T}{dS_0} = e^{(r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z} = \frac{S_T}{S_0},$$

and

$$\frac{d\hat{C}}{dS_T} = e^{-rT}\frac{d}{dS_T}\max\{0, S_T - K\} = \begin{cases} e^{-rT}, & S_T > K; \\ 0, & S_T < K. \end{cases}$$

At $S_T = K$, C fails to be differentiable; however, since this occurs with probability zero, the random variable $d\hat{C}/dS_0$ is almost surely well defined.

The pathwise derivative $d\hat{C}/dS_0$ can be thought of as a limiting case of the common random numbers finite-difference estimator in which we evaluate the limit analytically rather than numerically. It is a direct estimator of the option delta because it can be computed directly from a simulation starting at S_0 without the need for a separate simulation at a perturbed value S_0. This is evident from the expression in (18). The question remains whether this estimator is unbiased; that

[26] Sensitivities to various changes in the yield curve often account for several of these.

is, whether

$$E\left[\frac{d\hat{C}}{dS_0}\right] = \frac{dC}{dS_0} \equiv \frac{d}{dS_0}E[\hat{C}].$$

The unbiasedness of the pathwise estimate thus reduces to the interchangeability of derivative and expectation. The interchange is easily justified in this case; see Broadie and Glasserman (1996) for this example and conditions for more general cases. Applying the same reasoning used above, we obtain the following pathwise estimators of three other Greeks for the Black–Scholes price:

Rho (dC/dr): $KTe^{-rT}\mathbf{1}_{\{S_T \geq K\}}$

Vega $(dC/d\sigma)$: $e^{-rT}\mathbf{1}_{\{S_T \geq K\}}\frac{S_T}{\sigma}\left(\ln(S_T/S_0) - (r - \frac{1}{2}\sigma^2)T\right)$

Theta $(-dC/dT)$: $re^{-rT}\max(S_T - K, 0) - \mathbf{1}_{\{S_T \geq K\}}e^{-rT}\frac{S_T}{2T}\left(\ln(S_T/S_0)\right.$

$\left. + (r - \frac{1}{2}\sigma^2)T\right).$

Each of these estimators is unbiased.

Of course, Monte Carlo estimators are not required for these derivatives because closed-form expressions are available for each. The Black–Scholes setting is useful for illustration, but the utility of the technique rests on its applicability to more general models. In Broadie and Glasserman (1996), pathwise estimates are derived and studied (both theoretically and numerically) for Asian options and a model with stochastic volatility. For example, the Asian-option delta estimate is simply

$$e^{-rT}\frac{\bar{S}}{S_0}\mathbf{1}_{\{\bar{S}>K\}},$$

where $\bar{S}$ is the average asset price used to determine the option payoff. Evaluating this expression takes *negligible* time compared with resimulating to estimate the option price from a perturbed initial stock price. The pathwise estimate is thus both more accurate and faster to compute than the finite-difference approximation. These advantages extend to a wide class of problems.

As already noted, the unbiasedness of pathwise derivative estimates depends on an interchange of derivative and expectation. In practice, this generally means that the security payoff should be a pathwise continuous function of the parameter in question. The standard call option payoff $e^{-rT}\max\{0, S_T - K\}$ is continuous in each of its parameters. An example where continuity fails is a digital option with payoff $e^{-rT}\mathbf{1}_{\{S_T>K\}}B$, with B the amount received if the stock finishes in the

money.[27] Because of the
simplest form) cannot be a

The problem of disconti
ond derivative of an option
asset. Consider, again, the
for $d\hat{C}/dS_0$ in (18) invol
discontinuous in S_T, preve
get a direct estimator of ga

To address the problem
struct *smoothed* estimators
rive and implement as ord
investigate another techni
ratio method. This metho
rather than the outcome of
the pathwise method over
pathwise method generall

Overviews of these met
and Rubinstein and Shapi
tions see Broadie and Gla

5 Pri

European contingent clai
of the owner. Examples i
of swaps. By contrast, th
on the price path of the u
Many types of American
the-counter market. Exa
shout options, and Ameri
example as "real options"
and Pindyck (1994).

To be concrete, s
$\max_\tau E[e^{-r\tau}h(S_\tau)]$, wh
the payoff at time τ in
$\tau \leq T$. This formula
illustrate the major points

[27] We used this example at the er
numbers is particularly effecti
more effectively.

[28] Though not presented in a Mo
approach.

money.[27] Because of the discontinuity at $S_T = K$, the pathwise method (in its simplest form) cannot be applied to this type of option.

The problem of discontinuities often arises in the estimation of *gamma,* the second derivative of an option price with respect to the current price of an underlying asset. Consider, again, the standard European call option. We have an expression for $d\hat{C}/dS_0$ in (18) involving the indicator $\mathbf{1}_{\{S_T > K\}}$. This shows that $d\hat{C}/dS_0$ is discontinuous in S_T, preventing us from differentiating pathwise a second time to get a direct estimator of gamma.

To address the problem of discontinuities, Broadie and Glasserman (1996) construct *smoothed* estimators. These estimators are unbiased, but not as simple to derive and implement as ordinary pathwise estimators. Broadie and Glasserman also investigate another technique for direct derivative estimation called the *likelihood ratio method.* This method differentiates the probability density of an asset price, rather than the outcome of the asset price itself.[28] The domains of this method and the pathwise method overlap, but neither contains the other. When both apply, the pathwise method generally has lower variance.

Overviews of these methods can be found in Glasserman (1991), Glynn (1987), and Rubinstein and Shapiro (1993). For discussions specific to financial applications see Broadie and Glasserman (1996) and Fu and Hu (1995).

5 Pricing American options by simulation

European contingent claims have cash flows that cannot be influenced by decisions of the owner. Examples include European options, barrier options, and many types of swaps. By contrast, the cash flows of *American* contingent claims depend both on the price path of the underlying asset or assets and the decisions of the owner. Many types of American contingent claims trade on exchanges and in the over-the-counter market. Examples include American options, American swaptions, shout options, and American Asian options. They also arise in other contexts, for example as "real options" in the theory of economic investment described in Dixit and Pindyck (1994).

To be concrete, suppose that we wish to estimate the quantity $\max_\tau E[e^{-r\tau} h(S_\tau)]$, where r is the constant riskless interest rate, $h(S_\tau)$ is the payoff at time τ in state S_τ, and the max is taken over all stopping times $\tau \leq T$. This formulation of the American pricing problem will suffice to illustrate the major points. First, note that the state can be vector-valued and hence

[27] We used this example at the end of Section 3. The settings are related: problems for which common random numbers is particularly effective are generally problems to which the pathwise method can be applied even more effectively.

[28] Though not presented in a Monte Carlo context, the expressions in Carr (1993) are potentially relevant to this approach.

applies to pricing American options on multiple assets. Second, since simulation algorithms are discrete in nature, the continuous-time exercise decision must be approximated by restricting the exercise opportunities to lie in a finite set of times $0 = t_0 < t_1 < \cdots < t_d = T$. This is not always a serious restriction. For example, for a call option on a stock which pays dividends at discrete points in time, it can be shown that early exercise is only optimal just prior to the ex-dividend dates. In other cases, Richardson or other extrapolation techniques can be used to better approximate the price with exercise in continuous time from a finite set of exercise opportunities.[29] However, we now restrict attention to estimating the quantity

$$P \equiv \max_{\tau} E[e^{-r\tau} h(S_\tau)], \tag{19}$$

where the max is taken over all stopping times τ in the set t_i, for $i = 0, \ldots, d$. The need to estimate an optimal stopping time is the crucial distinction between American and European pricing problems.

If the state space is of low dimension, say three or less, a discretization scheme together with a dynamic programming algorithm can often be used to numerically approximate the value in (19). Even in these cases, simulation can be used to estimate the expectation in the recursive step. Simulation-based methods become essential when the dimension of the state space is large.

An obvious simulation-based algorithm for estimating the quantity P in equation (19) is to generate a random path of states S_{t_i}, for $i = 1, \ldots, d$, and form the path estimate

$$\hat{P} = \max_{i=0,\ldots,d} e^{-r t_i} h(S_{t_i}).$$

However, this estimator corresponds to using *perfect foresight*, and so it is biased high. That is, $E[\hat{P}] \geq P$, which follows immediately from the inequality $\max_{i=0,\ldots,d} e^{-r t_i} h(S_{t_i}) \geq e^{-r\tau} h(S_\tau)$. A natural goal would be to develop an alternative unbiased estimator. A negative result in this regard is provided in Broadie and Glasserman (1997): among a large class of estimators, there is no unbiased estimator of P. In particular, the estimators proposed in Tilley (1993), Grant, Vora, and Weeks (1997), and Barraquand and Martineau (1995) are all biased. Unfortunately, they provide no way to estimate the extent of the bias or to correct for the bias in a general setting. Broadie and Glasserman (1997) circumvent this problem by developing two estimators, one biased high and one biased low (but both asymptotically unbiased), which can be used together to form a valid confidence interval for the quantity P. In the remainder of this section, we give brief descriptions of the four methods mentioned and describe some strengths and weaknesses of each.

[29] Geske and Johnson (1984) gave the first financial application of Richardson extrapolation. An extensive treatment of extrapolation techniques is given in Marchuk and Shaidurov (1983).

5.1 Tilley's bundling algorithm

Tilley (1993) sparked considerable interest by demonstrating the potential practicality of applying simulation to pricing American contingent claims. Tilley describes a "bundling procedure" for pricing an American option on a single underlying asset. To estimate P he suggests simulating n paths of asset prices denoted $S_{t_i}(j)$ for $i = 1, \ldots, d$ and $j = 1, \ldots, n$ in the usual way. Next, partition the asset price space and call the paths which fall into a given partition at a fixed time a "bundle." A dynamic programming algorithm is applied to bundles to estimate C. In particular, the estimated option price $P_{t_i}(j)$ at time t_i for path j is the maximum of the immediate exercise value, $h(S_{t_i}(j))$, and the present value of continuing. The latter value is defined to be the average of $e^{-r(t_{i+1}-t_i)} P_{t_{i+1}}(k)$ over all paths k which fall in the bundle containing path j at time t_i. Details of the partitioning are given in Tilley (1993).

In order to implement the algorithm, all paths must be stored so they can be sorted into bundles at each time step. Since simulation typically requires a large number of paths for good estimates, the storage and sorting requirements can be significant. More importantly, the algorithm does not easily generalize to multiple state variables. In higher dimensions, it is not clear how to define the bundles. Even then it is likely that most partitions will contain very few paths and lead to a large bias, or the partitions will be so large that the continuation values are poorly estimated.

Because Tilley's algorithm uses the same paths to estimate the optimal decisions and the value, the estimator tends to be biased high (although the bundling induces an approximation which is difficult to analyze). Tilley introduces a "sharp boundary" variant which reduces the bias, but this variant does not easily generalize to higher dimensions. Carriere (1996) contains further analysis of Tilley's algorithm and suggests a procedure based on spline functions to reduce the bias. It remains to be seen whether the spline procedure is practical for higher dimensional problems. Nevertheless, for single state variable problems, Tilley demonstrated the potential practicality of applying simulation to American-style pricing problems.

5.2 Barraquand and Martineau's stratified state aggregation (SSA) algorithm

Barraquand and Martineau (1995) propose a partitioning algorithm, but unlike Tilley's bundling algorithm, they partition the payoff space instead of the state space. Hence, only a one dimensional space is partitioned at each time step, independent of the number of state variables.[30] Their algorithm works as follows.

---

[30] In fact, they distinguish between partitioning the state space, which they term "stratified state aggregation," and partitioning the payoff space, which they term "stratified state aggregation along the payoff." The latter method is the only one that they test or specify in detail. Hence we focus our discussion on this variant of their method.
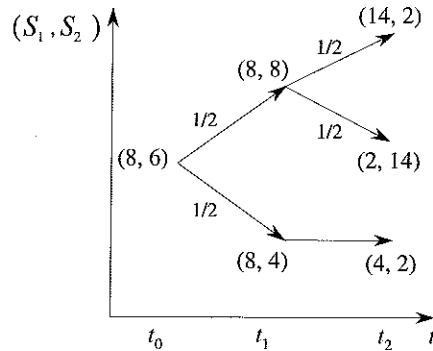
Fig. 9. State evolution.

First, partition the payoff space into K disjoint cells. Then simulate n paths of asset prices denoted $S_{t_i}(j)$ for $i = 1, \ldots, d$ and $j = 1, \ldots, n$ in the usual way. For each payoff cell k at time t_i, record the number of paths, $a_{t_i}(k)$, which fall into the cell. For each pair of cells k and l at consecutive times t_i and t_{i+1}, record the number of paths, $b_{t_i}(k, l)$, which fall into both cells. Also, for each cell k at time t_i, record the sum of the payoff values, $c_{t_i}(k) = \sum h(S_{t_i}(j))$, where the sum is over all paths j which fall into cell k at time t_i. The transition probability from (t_i, k) to (t_{i+1}, l) is approximated by $p_{t_i}(k, l) = b_{t_i}(k, l)/a_{t_i}(k)$. The estimated option price $P_{t_i}(k)$ at time t_i in cell k is the maximum of the immediate exercise value and the present value of continuing. The immediate exercise value is approximated by $c_{t_i}(k)/a_{t_i}(k)$. The present value of continuing is approximated by $e^{-r(t_{i+1}-t_i)} \sum_{l=1}^{K} p_{t_i}(k, l) P_{t_{i+1}}(l)$. This procedure can be applied backwards in time to determine the simulation estimate of the price P.

Details of a payoff space partitioning scheme are given in Barraquand and Martineau (1995). Once a single path is generated and the summary information a, b, and c is recorded, the path can be discarded. Hence the storage requirements with this method are modest: on the order of $K^2 d$. One drawback of this method is a possible lack of convergence, as the following example illustrates.

Figure 9 shows the evolution of two asset prices (S_1, S_2). The option payoff is $h(S_1, S_2) = \max(S_1, S_2)$ and for convenience the riskless rate is taken to be zero. Using the risk-neutral probabilities in Figure 9, the true value of the option at time t_0 is 11, which at time t_1 involves exercise in state $(8, 4)$ but continuing in state $(8, 8)$. When the states are partitioned by their payoffs, these two states are indistinguishable. As seen in the payoff evolution in Figure 10, the best strategy at time t_1 in payoff state 8 is to continue. The apparent value of the option in Figure 10 is 9 ($= (1/2)14 + (1/2)4$). In this example, partitioning the payoff
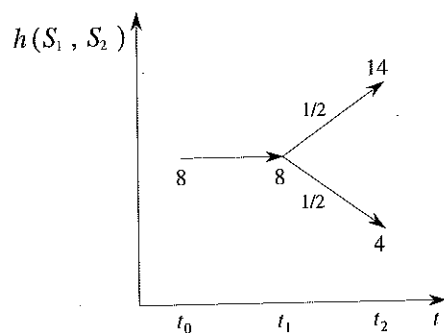
Fig. 10. Payoff evolution.

space leads to a significant underestimate of the option value. Hence, a simulation algorithm based on partitioning the payoff space cannot converge to the correct value. Although this example may seem contrived, Broadie and Detemple (1997) show that the payoff value is not a sufficient statistic for determining the optimal exercise decision for options on the maximum of several assets. Indeed, the payoff process $h(S_t)$ is hardly ever Markovian.

There is currently no way to bound the error in the Barraquand and Martineau method. Without an error estimate, it is difficult to determine the appropriate number of paths to simulate or the appropriate number of partitions to use. Their method can be slightly modified to generate an option price estimate which is biased low as follows. Their procedure gives an exercise strategy based on the immediate exercise payoff. Using this strategy, a new (independent) set of paths can be simulated, and an option value can be estimated under the exercise strategy previously estimated. The resulting option price estimate will be biased low because the exercise policy is not, in general, the optimal policy. With this modification, the average direction of the error is known. Raymar and Zwecher (1997) extend the Barraquand and Martineau approach by basing the exercise decision on a partition of two state-variables, rather than one.

5.3 Broadie and Glasserman's random tree algorithm

Broadie and Glasserman (1997) propose an algorithm based on simulated trees. In order to handle the bias problem, they develop two estimators, one biased high and one biased low, but both convergent and asymptotically unbiased as the computational effort increases. A valid confidence interval for the true value P is obtained by taking the upper confidence limit from the "high" estimator and the lower confidence limit from the "low" estimator. Briefly, their algorithm works as follows.

First, simulate a tree of asset prices (or, more generally, state variables) using b branches at each node. Two paths emanating from a node evolve as independent copies of the state process. The high estimator, Θ, is defined to be the value obtained by the usual dynamic programming algorithm applied to the simulated tree. Then repeat the process for n trees, and compute a point estimate and confidence interval for $E[\Theta]$. A low estimator is obtained by modifying the dynamic programming algorithm at each node. Instead of using all b branches to determine the decision and value, b_1 branches are used to determine the exercise decision, and the remaining $b_2 = b - b_1$ branches are used to determine the continuation value. Their actual low estimator, θ, includes another modification of this procedure which reduces the variance of the estimate. As before, estimates from n trees are combined to give a point estimate and confidence interval for $E[\theta]$. Details of the procedure can be found in Broadie and Glasserman (1997).

For the Θ estimator, all of the branches at a given node are used to determine the optimal decision and the corresponding node value, and this leads to an upward bias, i.e., $E[\Theta] \geq P$. For the θ estimator, the decision and the continuation value are determined from independent information sets. This eliminates the upward bias, but a downward bias occurs, i.e., $E[\theta] \leq P$. The intuition for this result follows. If the correct decision is inferred at a node, the node value estimate would be unbiased. If the incorrect decision is inferred at a node, the node value estimate would be biased low because of the suboptimality of the decision. The expected node value is a weighted average of an unbiased estimate (based on the correct decision) and an estimate which is biased low (based on the incorrect decision). The net effect is an estimate which is biased low. Both estimators are consistent and asymptotically unbiased as b increases.

The computational effort with this algorithm is order nb^d and its main drawback is that d cannot be too large for practical computations. Broadie and Glasserman (1997) give numerical results for options with $d = 4$. As mentioned earlier, to approximate option values with continuous exercise opportunities, some type of extrapolation procedure is required. Special care is necessary to implement extrapolation procedures within a simulation context because of the randomness in the estimates.

5.4 Other developments[31]

Grant, Vora, and Weeks (1997) describe a method specially designed to price American arithmetic Asian options on a single underlying asset. In this application the optimal exercise decision depends on the current asset price and the current

---

[31] More recent developments in pricing American options by simulation include Broadie and Glasserman (1997), Broadie, Glasserman and Ha (2000) and Longstaff and Schwartz (2001).

value of the average. Using repeated simulation runs, they attempt to identify the form of an optimal exercise policy based on these two pieces of information. Once an exercise policy is specified, simulation is used to estimate the option value under this fixed policy. Since the fixed policy is a suboptimal approximation to the optimal stopping rule, their procedure leads to a simulation estimator which is biased low.

GVW perform extensive sensitivity analysis which indicates that their option value estimate is relatively insensitive to deviations in the chosen exercise policy. So it may be that their method gives good option price estimates relative to some accuracy level, but it is not clear how to quantify their error. It is not clear how to improve their estimates to an arbitrary accuracy level as the simulation effort increases. Their procedure is specific to the case of American Asian options and does not at this point constitute a general approach to pricing American contingent claims.

Bossaerts (1989) proposes two estimators of optimal early exercise, a *moment estimator* and a *smooth optimization estimator*, and studies their convergence properties. His method appears to require a parametric representation of the exercise boundary and may therefore face difficulties in higher dimension. The optimization approach described in Fu and Hu (1995) also requires a parametric representation.

Rust (1997)[32] studies the general problem of solving *discrete decision problems*, which include optimal stopping problems as a special case. He develops a Monte Carlo method and shows that it succeeds in breaking the "curse of dimensionality" in these problem. Rust's focus is on computational complexity, but his approach appears to provide a promising direction for finance applications.

5.5 *Summary*

The valuation of securities with American-type features requires the determination of optimal decisions. High dimension versions of these problems arise from multiple state variables and/or path dependencies. Although simulation is a powerful tool for solving some higher dimensional problems, conventional wisdom was that simulation could not be applied to American-style pricing problems. The algorithms described here represent the first attempts to solve these problems that were long thought to be computationally intractable.

6 Further topics

We conclude this paper with a brief mention of two important areas of current work in the application of Monte Carlo methods to finance, not discussed in this article.

---

[32] We thank A. Dixit for pointing us to this reference.

A central numerical issue in simulating interest rates, asset prices with stochastic volatilities, and other complex diffusions is the accurate approximation of stochastic differential equations by discrete-time processes. Kloeden and Platen (1992) discuss a variety of methods for constructing discrete-time approximations with different orders of convergence. Andersen (1995) applies some of these to interest-rate models. In general, decreasing the time increment in a discrete approximation can be expected to give more accurate results, but at the expense of greater computational effort. Duffie and Glynn (1995) analyze this trade-off and characterize asymptotically optimal time steps as the overall computational effort grows.

In this article we have focused almost exclusively on the use of Monte Carlo for pricing. A related, growing area of application is risk management – in particular, the use of Monte Carlo to assess value at risk, credit risk, and related measures. For some examples of recent applications in these areas see Iben and Brotherton-Ratcliffe (1994), Lawrence (1994), Beckström and Campbell (1995) and Glasserman, Heidelberger and Shahabuddin (2000).

Appendix: Moment controls beat moment matching asymptotically

As mentioned in Section 2.4, any time a moment is available for use with moment matching, it can alternatively be used as a control variate. In this appendix, we argue that moment matching is asymptotically equivalent to a control variate technique with suboptimal coefficients, and is therefore dominated by the optimal use of moments as controls. This asymptotic link applies in large samples. A related link between linear and nonlinear control variates is made in Glynn and Whitt (1989), but the current setting does not fit their framework.

Let $Z_1, Z_2, \ldots$ be i.i.d. (not necessarily normal) with mean μ and variance σ^2. Let s denote the sample standard deviation of $Z_1, \ldots, Z_n$ and $\bar{Z}$ their sample mean. Suppose we want to estimate $E[f(Z)]$ for some function f. The standard estimator is $n^{-1} \sum_{i=1}^{n} f(Z_i)$ and the moment matching estimator is $n^{-1} \sum_{i=1}^{n} f(\tilde{Z}_i)$ with $\tilde{Z}_i$ defined in (9). For each i, the scaled difference

$$\sqrt{n}(\tilde{Z}_i - Z_i) = \sqrt{n}\left(\frac{\sigma - s}{s}\right)Z_i - \sqrt{n}[(\sigma \bar{Z}/s) - \mu]$$

converges in distribution, by the central limit theorem for $\bar{Z}$ and s. Thus, $(\tilde{Z}_i - Z_i) = O_p(n^{-1/2})$ (see, e.g., Appendix A of Pollard 1984 for O_p, o_p notation).

Suppose now that, with probability one, f is differentiable at Z_i. Then

$$f(\tilde{Z}_i) = f(Z_i) + f'(Z_i)[\tilde{Z}_i - Z_i] + o_p(n^{-1/2}),$$

suggesting that up to terms $o_p(n^{-1/2})$ the moment matching estimator and standard

---

estimator are related via

$$\frac{1}{n}\sum_{i=1}^{n} f(\tilde{Z}_i) \approx \frac{1}{n}\sum_{i=1}^{n} f($$

$$= \frac{1}{n}\sum_{i=1}^{n} f($$

$$= \frac{1}{n}\sum_{i=1}^{n} f($$

$$+ \left(\frac{1}{n}\sum_{i=1}^{n}\right.$$

$$\equiv \frac{1}{n}\sum_{i=1}^{n} f($$

where $\hat{\beta}_i \to \beta_i, i = 1, 2$, as n

$$\hat{\beta}_1 = E[f'($$

Thus, moment matching is asy

$$\left(\frac{\sigma}{s}\right.$$

as controls (both quantities co
efficients β_1, β_2. In general, t
β_1^*, β_2^*, so moment matching i
method. In addition, the cont
matching itself) because thoug
for finite n. In contrast, the m
$(\bar{Z} - \mu)$ have mean zero for al

Acworth, P., M. Broadie, and P. G
and Quasi Monte Carlo Meth
Monte Methods for Scientific
and P. Zinterhof (eds.), Sprin
Andersen, L., 1995, Efficient Tech
Involving Non-Linear Stocha
Financial Products, New Yor
Andersen, L., and R. Brotherton-R
Barlow, R.E. and F. Proschan, 197
(Holt, Reinhart and Winston,
Barraquand, J., 1995, Numerical V
Securities, Management Scie

estimator are related via

$$\frac{1}{n}\sum_{i=1}^{n} f(\tilde{Z}_i) \approx \frac{1}{n}\sum_{i=1}^{n} f(Z_i) + \frac{1}{n}\sum_{i=1}^{n} f'(Z_i)[\tilde{Z}_i - Z_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n} f(Z_i) + \frac{1}{n}\sum_{i=1}^{n} f'(Z_i)\left[\left(\frac{\sigma}{s} - 1\right)Z_i - \frac{\sigma}{s}\bar{Z} + \mu\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} f(Z_i) + \left(\frac{1}{n}\sum_{i=1}^{n} f'(Z_i)Z_i\right)\left(\frac{\sigma}{s} - 1\right)$$

$$\quad + \left(\frac{1}{n}\sum_{i=1}^{n} f'(Z_i)\right)\left(\mu - \frac{\sigma}{s}\bar{Z}\right)$$

$$\equiv \frac{1}{n}\sum_{i=1}^{n} f(Z_i) + \hat{\beta}_1\left(\frac{\sigma}{s} - 1\right) + \hat{\beta}_2\left(\mu - \frac{\sigma}{s}\bar{Z}\right)$$

where $\hat{\beta}_i \to \beta_i$, $i = 1, 2$, as $n \to \infty$, with

$$\beta_1 = E[f'(Z)Z], \quad \text{and} \quad \beta_2 = E[f'(Z)].$$

Thus, moment matching is asymptotically equivalent to using

$$\left(\frac{\sigma}{s} - 1\right) \quad \text{and} \quad \left(\mu - \frac{\sigma}{s}\bar{Z}\right) \tag{20}$$

as controls (both quantities converge to zero almost surely) with estimates of coefficients β_1, β_2. In general, these do not coincide with the optimal coefficients β_1^*, β_2^*, so moment matching is asymptotically dominated by the control variate method. In addition, the controls in (20) introduce some bias (as does moment matching itself) because though they converge to zero they do not have mean zero for finite n. In contrast, the more natural moment control variates $(s^2 - \sigma^2)$ and $(\bar{Z} - \mu)$ have mean zero for all n and thus introduce no bias.

References

Acworth, P., M. Broadie, and P. Glasserman, 1997, A Comparison of Some Monte Carlo and Quasi Monte Carlo Methods for Option Pricing, in Monte Carlo and Quasi Monte Methods for Scientific Computing, G. Larcher, P. Hellekalek, H. Niederreiter, and P. Zinterhof (eds.), Springer-Verlag, Berlin.

Andersen, L., 1995, Efficient Techniques for Simulation of Interest Rate Models Involving Non-Linear Stochastic Differential Equations, Working paper (General Re Financial Products, New York, NY).

Andersen, L., and R. Brotherton-Ratcliffe, 1996, Exact Exotics, Risk 9, October, 85–89.

Barlow, R.E. and F. Proschan, 1975, Statistical Theory of Reliability and Life Testing (Holt, Reinhart and Winston, New York).

Barraquand, J., 1995, Numerical Valuation of High Dimensional Multivariate European Securities, Management Science 41, 1882–1891.

Barraquand, J. and D. Martineau, 1995, Numerical Valuation of High Dimensional Multivariate American Securities, Journal of Financial and Quantitative Analysis 30, 383–405.

Beaglehole, D., P. Dybvig, and G. Zhou, 1997, Going to Extremes: Correcting Simulation Bias in Exotic Option Valuation, Financial Analysts Journal (Jan/Feb) 62–68.

Beckström, R. and A. Campbell, 1995, An Introduction to VAR (CATS Software, Palo Alto, California).

Berman, L., 1996, Comparison of Path Generation Methods for Monte Carlo Valuation of Single Underlying Derivative Securities, Research Report RC-20570, IBM Research, Yorktown Heights, New York.

Birge, J.R., 1994, Quasi-Monte Carlo Approaches to Option Pricing, Technical Report 94–119 (Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109).

Bossaerts, P., 1989, Simulation Estimators of Optimal Early Exercise, Working paper (Carnegie-Mellon University, Pittsburgh, PA, 15213).

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

Boyle, P. and D. Emanuel, 1985, The Pricing of Options on the Generalized Mean, Working paper (University of Waterloo).

Bratley, P. and B. Fox, 1988, ALGORITHM 659: Implementing Sobol's Quasirandom Sequence Generator, ACM Transactions on Mathematical Software 14, 88–100.

Bratley, P., B.L. Fox, and H. Niederreiter, 1992, Implementation and Tests of Low-Discrepancy Sequences, ACM Transactions on Modelling and Computer Simulation 2, 195–213.

Bratley, P., B.L. Fox, and L. Schrage, 1987, A Guide to Simulation, 2nd Ed. (Springer-Verlag, New York).

Broadie, M. and J. Detemple, 1997, The Valuation of American Options on Multiple Assets, Mathematical Finance 7, 241–286.

Broadie, M. and J. Detemple, 1996, American Option Valuation: New Bounds, Approximations, and a Comparison of Existing Methods, Review of Financial Studies 9, 1211–1250.

Broadie, M. and P. Glasserman, 1996, Estimating Security Price Derivatives by Simulation, Management Science 42, 269–285.

Broadie, M. and P. Glasserman, 1997, Pricing American-Style Securities Using Simulation, Journal of Economic Dynamics and Control 21, 1323–1352.

Broadie, M. and P. Glasserman, 1997, A Stochastic Mesh Method for Pricing High-Dimensional American Options, Working paper, Columbia Business School, New York.

Broadie, M., P. Glasserman, and Z. Ha, 2000, Pricing American Options by Simulation Using a Stochastic Mesh with Optimized Weights, in Probabilistic Constrained Optimization, S. Uryasev, ed., 26–44 (Kluwer, Norwell, Mass.)

Caflisch, R.E., W., Morokoff, and A. Owen, 1998, Valuation of Mortgage Backed Securities Using Brownian Bridges to Reduce Effective Dimension, in Monte Carlo: Methodologies and Applications for Pricing and Risk Management, 301–314 (Risk Publications, London).

Carr, P., 1993, Deriving Derivatives of Derivative Securities, Working paper (Johnson Graduate School of Business, Cornell University).

Carriere, J.F., 1996, Valuation of the Early-Exercise Price for Derivative Securities using Simulations and Splines, Insurance: Mathematics and Economics 19, 19–30.

Carverhill, A. and K. Pang, 1995, Efficient and Flexible Bond Option Valuation in the

Heath, Jarrow and Morton Fra 70–77.

Cheyette, O., 1992, Term Structure Income 2, March, 28–41.

Clewlow, L. and A. Carverhill, 199 Derivatives 2, Winter, 66–74.

Devroye, L., 1986, Non-Uniform R York).

Dixit, A. and R. Pindyck, 1994, Inv Press).

Duan, J.-C., 1995, The GARCH Op

Duan, J.-C. and J.-G. Simonato, 19 Management Science 44, 121

Duffie, D., 1996, Dynamic Asset P Princeton, New Jersey).

Duffie, D. and P. Glynn, 1995, Effi Annals of Applied Probability

Faure H., 1982, Discrépance de Su Dimension s), Acta Arithmeti

Fox, B.L., 1986, ALGORITHM 64 Quasi-Random Sequence Ger 12, 362–376.

Fu, M. and J.Q. Hu, 1995, Sensitiv Pricing, Probability in the Eng

Fu, M., D. Madan, and T. Wong, 1 Comparison of Analytical and Finance 2, 49–74.

Geske, R. and H.E. Johnson, 1984, Journal of Finance 39, 1511–

Glasserman, P., 1991, Gradient Est Publishers, Norwell, Mass).

Glasserman, P., 1993, Filtered Mor 610–634.

Glasserman, P., P. Heideberger, an Techniques for Estimating Va

Glasserman, P. and D.D. Yao, 199 Random Numbers, Managem

Glynn, P.W., 1987, Likelihood Rat of the Winter Simulation Con Diego, California) 366–374.

Glynn, P.W., 1989, Optimization o the Winter Simulation Confer Diego, California) 90–105.

Glynn, P.W. and D.L. Iglehart, 198 Queueing Systems 3, 221–25

Glynn, P.W. and W. Whitt, 1989, I 37, 82–103.

Glynn, P.W. and W. Whitt, 1992, T Operations Research 40, 505–

Grant, D., G. Vora, and D. Weeks,

Heath, Jarrow and Morton Framework, Journal of Fixed Income 5, September, 70–77.

Cheyette, O., 1992, Term Structure Dynamics and Mortgage Valuation, Journal of Fixed Income 2, March, 28–41.

Clewlow, L. and A. Carverhill, 1994, On the Simulation of Contingent Claims, Journal of Derivatives 2, Winter, 66–74.

Devroye, L., 1986, Non-Uniform Random Variate Generation (Springer-Verlag, New York).

Dixit, A. and R. Pindyck, 1994, Investment Under Uncertainty (Princeton University Press).

Duan, J.-C., 1995, The GARCH Option Pricing Model, Mathematical Finance 5, 13–32.

Duan, J.-C. and J.-G. Simonato, 1998, Empirical Martingale Simulation for Asset Prices, Management Science 44, 1218–1233.

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

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

Faure H., 1982, Discrépance de Suites Associées à un Système de Numération (en Dimension s), Acta Arithmetica 41, 337–351.

Fox, B.L., 1986, ALGORITHM 647: Implementation and Relative Efficiency of Quasi-Random Sequence Generators, ACM Transactions on Mathematical Software 12, 362–376.

Fu, M. and J.Q. Hu, 1995, Sensitivity Analysis for Monte Carlo Simulation of Option Pricing, Probability in the Engineering and Information Sciences 9, 417–446.

Fu, M., D. Madan, and T. Wong, 1998, Pricing Continuous Time Asian Options: A Comparison of Analytical and Monte Carlo Methods, Journal of Computational Finance 2, 49–74.

Geske, R. and H.E. Johnson, 1984, The American Put Options Valued Analytically, Journal of Finance 39, 1511–1524.

Glasserman, P., 1991, Gradient Estimation via Perturbation Analysis (Kluwer Academic Publishers, Norwell, Mass).

Glasserman, P., 1993, Filtered Monte Carlo, Mathematics of Operations Research 18, 610–634.

Glasserman, P., P. Heideberger, and P. Shahabuddin, 2000, Variance Reduction Techniques for Estimating Value-at-Risk, Management Science 46, 1349–1365.

Glasserman, P. and D.D. Yao, 1992, Some Guidelines and Guarantees for Common Random Numbers, Management Science 38, 884–908.

Glynn, P.W., 1987, Likelihood Ratio Gradient Estimation: An Overview, in: Proceedings of the Winter Simulation Conference (The Society for Computer Simulation, San Diego, California) 366–374.

Glynn, P.W., 1989, Optimization of Stochastic Systems via Simulation, in: Proceedings of the Winter Simulation Conference (The Society for Computer Simulation, San Diego, California) 90–105.

Glynn, P.W. and D.L. Iglehart, 1988, Simulation Methods for Queues: An Overview, Queueing Systems 3, 221–255.

Glynn, P.W. and W. Whitt, 1989, Indirect Estimation via $L = \lambda W$, Operations Research 37, 82–103.

Glynn, P.W. and W. Whitt, 1992, The Asymptotic Efficiency of Simulation Estimators, Operations Research 40, 505–520.

Grant, D., G. Vora, and D. Weeks, 1997, Path-Dependent Options: Extending the Monte

Carlo Simulation Approach, Management Science 43, 1589–1602.

Halton, J.H., 1960, On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals, Numerische Mathematik 2, 84–90.

Hammersley, J.M. and D.C. Handscomb, 1964, Monte Carlo Methods (Chapman and Hall, London).

Haselgrove, C.B., 1961, A Method for Numerical Integration, Mathematics of Computation 15, 323–337.

Hlawka, E., 1971, Discrepancy and Riemann Integration, in: L. Mirsky, ed., Studies in Pure Mathematics (Academic Press, New York).

Hull, J., 2000, Options, Futures, and Other Derivative Securities, 4th ed. (Prentice-Hall, Englewood Cliffs, New Jersey).

Hull, J. and A. White, 1987, The Pricing of Options on Assets with Stochastic Volatilities, Journal of Finance 42, 281–300.

Iben, B. and R. Brotherton-Ratcliffe, 1994, Credit Loss Distributions and Required Capital for Derivatives Portfolios, Journal of Fixed Income 4, June, 6–14.

Johnson, H., 1987, Options on the Maximum or the Minimum of Several Assets, Journal of Financial and Quantitative Analysis 22, 227–283.

Johnson, H. and D. Shanno, 1987, Option Pricing When the Variance is Changing, Journal of Financial and Quantitative Analysis 22, 143–151.

Joy C., P.P. Boyle, and K.S. Tan, 1996, Quasi-Monte Carlo Methods in Numerical Finance, Management Science 42, 926–938.

Kemna, A.G.Z. and A.C.F. Vorst, 1990, A Pricing Method for Options Based on Average Asset Values, Journal of Banking and Finance 14, 113–129.

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

L'Ecuyer, P. and G. Perron, 1994, On the Convergence Rates of IPA and FDC Derivative Estimators, Operations Research 42, 643–656.

Lavenberg, S.S. and P.D. Welch, 1981, A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations, Management Science 27, 322–335.

Lawrence, D., 1994, Aggregating Credit Exposures: The Simulation Approach, in: Derivative Credit Risk (Risk Publications, London).

Longstaff, F.A. and E.S. Schwartz, 2001, Valuing American Options by Simulation: A Simple Least Squares Approach, Review of Financial Studies 14, 113–148.

Marchuk, G. and V. Shaidurov, 1983, Difference Methods and Their Extrapolations (Springer Verlag, New York).

McKay, M.D., W.J. Conover, and R.J. Beckman, 1979, A Comparison of Three Methods for Selecting Input Variables in the Analysis of Output from a Computer Code, Technometrics 21, 239–245.

Morokoff, W.J. and R.E. Caflisch, 1995, Quasi-Monte Carlo Integration, Journal of Computational Physics, 122, 218–230.

Moskowitz B. and R.E. Caflisch, 1996, Smoothness and Dimension Reduction in Quasi-Monte Carlo Methods, Mathematical and Computer Modeling 23, 37–54.

Niederreiter, H., 1988, Low Discrepancy and Low Dispersion Sequences, Journal of Number Theory 30, 51–70.

Niederreiter, H., 1976, On the Distribution of Pseudo-Random Numbers Generated by the Linear Congruential Method. III, Mathematics of Computation 30, 571–597.

Niederreiter, H., 1992, Random Number Generation and Quasi-Monte Carlo Methods (CBMS-NSF 63, SIAM, Philadelphia, Pa).

Niederreiter, H. and C. Xing, 1996, Low-Discrepancy Sequences and Global Function

Fields with Many Rational F

Nielsen, S., 1994, Importance Sa (Management Science and I

Ninomiya, S., and S. Tezuka, 199 Derivatives, Applied Mathe

Owen, A., 1995a, Monte Carlo V H. Niederreiter and P.J.S. Sl in Scientific Computing (Sp

Owen, A., 1995b, Randomly Per Carlo and Quasi-Monte Car P. Shiue (eds.), 299–317 (Sp

Paskov, S. and J. Traub, 1995, Fa Portfolio Management 22, I

Pollard, D., 1984, Convergence o

Press, W.H., S.A. Teukolsky, W.T in C: The Art of Scientific C

Raymar, S., and M. Zwecher, 19 On the Maximum of Severa

Reider, R., 1993, An Efficient M (Wharton School, Universit

Rubinstein, R. and A. Shapiro, 1

Rust, J., 1997, Using Randomiza 65, 487–516.

Schwartz, E.S. and W.N. Torous Mortgage-Backed Securitie

Scott, L.O., 1987, Option Pricin Estimation, and an Applica 419–438.

Shaw, J., 1995, Beyond VAR an Applications for Pricing an London).

Sobol', I.M., 1967, On the Distr Evaluation of Integrals, US Physics 7, 86–112.

Spanier, J. and E.H. Maize, 199 Relatively Small Samples,

Stein, M., 1987, Large Sample I Sampling, Technometrics 2

Stulz, R.M., 1982, Options on t Journal of Financial Econo

Tezuka, S., 1994, A Generaliza Research Report RTO105 Japan).

Tezuka, S., 1995, Uniform Ran Publishers, Boston).

Tilley, J.A., 1993, Valuing Ame the Society of Actuaries 4

Turnbull, S.M. and L.M. Wake Average Options, Journal

Van Rensberg J. and G.M. Torr

Fields with Many Rational Places, Finite Fields and their Applications 2, 241–273.

Nielsen, S., 1994, Importance Sampling in Lattice Pricing Models, Working paper (Management Science and Information Systems, University of Texas at Austin).

Ninomiya, S., and S. Tezuka, 1996, Toward Real-Time Pricing of Complex Financial Derivatives, Applied Mathematical Finance 3, 1–20.

Owen, A., 1995a, Monte Carlo Variance of Scrambled Equidistribution Quadrature, in: H. Niederreiter and P.J.S. Shiue, eds., Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (Springer-Verlag, Berlin).

Owen, A., 1995b, Randomly Permuted (t, m, s)-Nets and (t, s)-Sequences, in Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, H. Niederreiter and P. Shiue (eds.), 299–317 (Springer-Verlag, New York).

Paskov, S. and J. Traub, 1995, Faster Valuation of Financial Derivatives, Journal of Portfolio Management 22, Fall, 113–120.

Pollard, D., 1984, Convergence of Stochastic Processes, Springer-Verlag, New York.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, 1992, Numerical Recipes in C: The Art of Scientific Computing, 2nd ed. (Cambridge University Press).

Raymar, S., and M. Zwecher, 1997, A Monte Carlo Valuation of American Call Options On the Maximum of Several Stocks, Journal of Derivatives 5 (Fall), 7–24.

Reider, R., 1993, An Efficient Monte Carlo Technique for Pricing Options, Working paper (Wharton School, University of Pennsylvania).

Rubinstein, R. and A. Shapiro, 1993, Discrete Event Systems (Wiley, New York).

Rust, J., 1997, Using Randomization to Break the Curse of Dimensionality, Econometrica 65, 487–516.

Schwartz, E.S. and W.N. Torous, 1989, Prepayment and the Valuation of Mortgage-Backed Securities, Journal of Finance 44, 375–392.

Scott, L.O., 1987, Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application, Journal of Financial and Quantitative Analysis 22, 419–438.

Shaw, J., 1995, Beyond VAR and Stress Testing, in Monte Carlo: Methodologies and Applications for Pricing and Risk Management, 231–244 (Risk Publications, London).

Sobol', I.M., 1967, On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals, USSR Computational Mathematics and Mathematical Physics 7, 86–112.

Spanier, J. and E.H. Maize, 1994, Quasi-Random Methods for Estimating Integrals Using Relatively Small Samples, SIAM Review 36, 18–44.

Stein, M., 1987, Large Sample Properties of Simulations Using Latin Hypercube Sampling, Technometrics 29, 143–151.

Stulz, R.M., 1982, Options on the Minimum or the Maximum of Two Risky Assets, Journal of Financial Economics 10, 161–185.

Tezuka, S., 1994, A Generalization of Faure Sequences and its Efficient Implementation, Research Report RTO105 (IBM Research, Tokyo Research Laboratory, Kanagawa, Japan).

Tezuka, S., 1995, Uniform Random Numbers: Theory and Practice (Kluwer Academic Publishers, Boston).

Tilley, J.A., 1993, Valuing American Options in a Path Simulation Model, Transactions of the Society of Actuaries 45, 83–104.

Turnbull, S.M. and L.M. Wakeman, 1991, A Quick Algorithm for Pricing European Average Options, Journal of Financial and Quantitative Analysis 26, 377–389.

Van Rensberg J. and G.M. Torrie, 1993, Estimation of Multidimensional Integrals: Is

Monte Carlo the Best Method?, Journal of Physics A: Mathematical and General 26, 943–953.

Wiggins, J.B., 1987, Option Values under Stochastic Volatility: Theory and Empirical Evidence, Journal of Financial Economics 19, 351–372.

Willard, G.A., 1997, Calculating Prices and Sensitivities for Path-Dependent Derivative Securities in Multifactor Models, Journal of Derivatives 5 (Fall), 45–61.

Worzel, K.J., C. Vassiadou-Zeniou, and S.A. Zenios, 1994, Integrated Simulation and Optimization Models for Tracking Indices of Fixed-Income Securities, Operations Research 42, 223–233.

Zaremba, S.K., 1968, The Mathematical Basis of Monte Carlo and Quasi-Monte Carlo Methods, SIAM Review 10, 310–314.

Inte