Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

IMPORTANCE SAMPLING IN DERIVATIVE SECURITIES PRICING

Yi Su
Michael C. Fu

Robert H. Smith School of Business
University of Maryland
College Park, MD 20742, U.S.A.

ABSTRACT

We formulate the importance sampling problem as a parametric minimization problem under the original measure and use a combination of infinitesimal perturbation analysis (IPA) and stochastic approximation (SA) to minimize the variance of the price estimation. Compared to existing methods, the IPA estimator derived in this paper has significantly smaller estimation variance and doesn't depend on the form of payoff functions and differentiability of the sample path, and thus is more universally applicable and computationally efficient. Under suitable conditions, the objective function is a convex function, the IPA estimator presented is unbiased, and the corresponding stochastic approximation algorithm converges to the true optimal value.

1 INTRODUCTION

Monte Carlo simulation is used for pricing a variety of securities, such as exotic equity options or fixed income securities like mortgage-backed securities. As the complexity of the structure of the financial claims and the dynamics of the underlying assets increases, Monte Carlo simulation often becomes the sole computationally feasible means of security pricing.

The efficiency of Monte Carlo simulations depends on the variance of the estimation. Suppose we estimate the security price p by $\hat{p}$, where $\hat{p}$ is an asymptotically unbiased estimate of p, then by the Central Limit Theorem,

$$\sqrt{N}(\hat{p} - p) \Rightarrow N(0, \sigma_{\hat{p}}^2),$$

where N is the number of simulations and $\sigma_{\hat{p}}^2$ is the variance of estimation. This means that by reducing $\sigma_{\hat{p}}$ by a factor of 10, the number of simulation replications required to obtain the same level of precision will be reduced by a factor of 100. This is the motivation behind a variety of variance reduction techniques (VRT) in Monte Carlo simulations such as control variates, antithetic variate and importance sampling. Examples of successful implementations of control variates for the pricing of financial derivatives include Hull and White (1987, 1988), Turnbull and Wakeman (1991), and Fu, Madan, Wang (1997).

Variance reduction based on importance sampling has not been widely used as other VRTs in pricing financial derivatives until recently. The idea behind importance sampling is to simulate more sample paths on the area that matters; for instance, for a deep out-of-the-money call option, most of the time the payoff from the simulation is 0, so simulating more sample paths with positive payoffs should reduce the variance in the estimation. Mathematically speaking, the fundamental idea behind importance sampling is that under certain regularity conditions, expectation under one probability measure can be expressed as an expectation under another probability measure through the Radon-Nikodym theorem. The right choice of the new probability measure will effectively reduce the variance associated with the estimation.

An early example of importance sampling applied to derivatives pricing is Reider (1993), where increasing the drift substantially decreases the variance in simulations for deep out-of-the-money European call options. Glasserman, Heidelberger, Shahabuddin (1998) apply importance sampling in the Heath, Jarrow, Merton (1992) framework, reporting substantial variance reduction by combining stratified sampling and change of the drift term. Other recent work on applying importance sampling in valuation of financial claims include Andersen (1995) and Boyle, Broadie, Glasserman (1997).

Most closely related to our work is that of Vazquez-Abad and Dufresne (1998), who apply importance sampling combined with control variates to dramatically reduce variance in pricing Asian options. They use gradient estimation and stochastic approximation to find the optimal change of the drift term. We also use gradient-based methods to estimate the optimal importance sampling measure, but our approach differs in one critical aspect. In our setting, the importance sampling problem is transformed into a minimization problem under the original probability measure, which eliminates the dependence between the payoff function and the parameters in the optimization. This leads to a much simpler IPA gradient estimator with significantly smaller estimation variance than the original IPA estimator given in Vazquez-Abad and Dufresne (1998). Perhaps more importantly, since the payoff function is not directly related to the optimization parameters, we do not require differentiability of the payoff function as in the original method, so our method is applicable in much more general setting. If the importance sampling is implemented via a change of the drift term in Brownian motion, then we show that the objective function in our minimization problem is a convex function, establishing the conjecture in Vazquez-Abad and Dufresne (1998). We further prove that our stochastic approximation algorithm a.s. converges to the true global optimization value.

2 FORMULATION AND SETTINGS

We assume the financial market is arbitrage free, so there exists an equivalent probability measure Q (Harrison and Kreps 1979) under which the price at time 0 of a European financial claim $C(T,\omega)$ s.t. $E^Q[C^2(T,\omega)] < +\infty$, where T is the expiration (maturity) date and ω is the sample path of the underlying stochastic process(es), is given by

$$C_0 = E^Q[e^{-\int_0^T r(t,\omega)dt} C(T,\omega)],$$

where Q is called the risk-neutral (martingale) measure and $r(t,\omega)$ is the risk-free interest rate process. We will assume throughout that $r(t,\omega) \geq 0$, i.e., the risk-free interest rate process is non-negative. Defining the present value of the payoff by

$$\hat{C}(T,\omega) = e^{-\int_0^T r(t,\omega)dt} C(T,\omega),$$

we are interested in estimating $C_0 = E^Q[\hat{C}(T,\omega)]$.

Examples of payoff functions $C(T,\omega)$.

$$\begin{array}{rl}
(S_T(\omega) - K)^+ & \text{call,} \\
(K - S_T(\omega))^+ & \text{put,} \\
(T^{-1}\int_0^T S_t dt - K)^+ & \text{continuous Asian,} \\
(S_T - \min\{S_t, 0 \leq t \leq T\})^+ & \text{lookback,} \\
(\max_i\{S_T^i(\omega)\} - K)^+ & \text{basket (max),} \\
(S_T - K)^+ \mathbf{1}\{S_t \leq L, t \in [0,T]\} & \text{barrier (up and out),}
\end{array}$$

where S_t is the stock price at time t (superscripted for the max-option on a basket of stocks), K is the strike price, L is the barrier value for the last example, and $\mathbf{1}\{\cdot\}$ is the indicator function.

A direct estimate for C_0 is obtained by simulating the risk-neutral distribution of the underlying asset(s) and taking the sample mean over replications of $\hat{C}(T,\omega)$. However, by the Radon-Nikodym theorem, if measure Q is absolutely continuous w.r.t some other measure P, then

$$C_0 = E^P\left[\hat{C}(T,\omega)\frac{dQ}{dP}\right],$$

which gives an alternative estimator for simulation under P:

$$\hat{C}(T,\omega)\frac{dQ}{dP}, \tag{1}$$

where $\frac{dQ}{dP}$ is the Jacobian of the measure change, i.e., the Radon-Nikodym derivative. It is a simple mathematical exercise to show that the above estimator is also an unbiased estimator of the option price. However, the new estimator may have different estimation variance, hence the potential for variance reduction.

As a simple example, consider a European call option on a single underlying asset, assuming a constant risk-free interest rate r, i.e.,

$$C(T,\omega) = (S_T - K)^+,$$

for which the price is given by

$$C_0 = E^Q[e^{-rT}(S_T - K)^+].$$

By the Radon-Nikodym theorem, we can also calculate C_0 by

$$C_0 = E^P[e^{-rT}(S_T - K)^+\frac{dQ}{dP}].$$

If we choose

$$p(x) = ce^{-rT}(S_T - K)^+ d(S_T),$$

where $d(S_T)$ is the risk-neutral density of S_T, and c is the normalization constant, then we obtain a zero variance estimator by sampling from distribution $p(x)$. However, $c = 1/C_0$, which requires the full knowledge of the option price, making this zero-variance importance sampling estimator inapplicable in practice.

Now, we concentrate on some more attainable models in practice. We first restrict the new measure to be in a family of probability measure $\{P(\theta,\omega) : \theta \in \Theta\}$, where θ is the parameter and for any $\theta \in \Theta$, measure Q is absolutely continuous w.r.t. $P(\theta)$. We consider the problem of finding the value of θ which gives the best performance in simulation, i.e., the smallest estimation errors in simulation.

The variance of the new estimator (1) is

$$E^P\left[\left(\hat{C}(T,\omega)\frac{dQ}{dP}\right)^2\right]-C_0^2$$
$$=E^P\left[\left(\hat{C}(T,\omega)\right)^2\left(\frac{dQ}{dP}\right)^2\right]-C_0^2.$$

This leads to the following minimization problem in the domain of stochastic optimization:

$$\min_{\theta\in\Theta} V(\theta),$$

where

$$V(\theta)=E^P\left[\left(\hat{C}(T,\omega)\right)^2\left(\frac{dQ}{dP}\right)^2\right]. \tag{2}$$

Remark: Vazquez-Abad and Dufresne (1998) derive the IPA estimator associated with their stochastic optimization problem by directly differentiating the term inside the expectation of (2), which requires derivatives for both $C(T,\omega)$ and $\frac{dQ}{dP}$, since ω, and hence $C(T,\omega)$, clearly depends on θ, in addition to $\frac{dQ}{dP}$. This is because sampling is carried out under P rather than Q. However, this is avoided if the minimization is carried out under the original measure Q, and this is the fundamental difference between our method and their method.

Simple calculation shows that

$$\begin{aligned} V(\theta) &= \int_\Omega \left(\hat{C}(T,\omega)\right)^2\frac{(dQ)^2}{(dP)^2}dP \\ &= \int_\Omega \left(\hat{C}(T,\omega)\right)^2\frac{(dQ)^2}{dP} \\ &= \int_\Omega \left(\hat{C}(T,\omega)\right)^2\frac{dQ}{dP}dQ \\ &= E^Q\left[\left(\hat{C}(T,\omega)\right)^2\frac{dQ}{dP}\right]. \end{aligned}$$

So we only need to find the θ that minimizes

$$V(\theta)=E^Q\left[\hat{C}^2(T,\omega)f(\theta,\omega)\right], \tag{3}$$

where

$$f(\theta,\omega)=\frac{dQ}{dP(\theta)}. \tag{4}$$

The important thing to note is that changing back to the original measure Q eliminates the dependence of $C(T,\omega)$ on θ.

3 STOCHASTIC APPROXIMATION AND IPA

Our approach to minimizing $V(\theta)$ is the same as Vazquez-Abad and Dufresne (1998), in that we use gradient-based stochastic approximation to estimate

$$\theta^*=\arg\min_{\theta\in\Theta} V(\theta),$$

via the following iterative scheme:

$$\theta_{n+1}=\Pi_\Theta(\theta_n-a_n\hat{g}_n), \tag{5}$$

where $\theta_n=((\theta_n)_1,\ldots,(\theta_n)_k)$ represents the nth iterations, $\hat{g}_n$ represents an estimate of the gradient $\nabla V(\theta)$ at θ_n, $\{a_n\}$ is a positive sequence of numbers converging to 0, and Π_Θ denotes a projection on Θ. The difference in our approach is the form of $V(\theta)$ used in deriving the infinitesimal perturbation analysis (IPA) estimator: (3) v.s. (2).

We first make the following assumption.

Assumption 1: $f(\theta,\omega)$ is Q-a.s. piecewise differentiable on Θ.

Differentiating inside the expectation of (3) yields the IPA estimator

$$\hat{C}^2(T,\omega)\frac{\partial f(\theta,\omega)}{\partial\theta}. \tag{6}$$

Under suitable conditions, this IPA estimator is unbiased.

Theorem 1 *(General Unbiasedness) If Assumption 1 holds,*

$$\left|\frac{\partial}{\partial\theta}f(\theta,\omega)\right|<M(\omega) \qquad Q\text{-a.s.;}$$

and there exists a $\delta>0$ *such that*

$$E^Q[C(T,\omega)]^{2+2\delta}<+\infty,\ \ E^Q[M(\omega)]^{1+1/\delta}<+\infty; \tag{7}$$

or

$$E^Q\left[C(T,\omega)^2M(\omega)\right]<+\infty; \tag{8}$$

then (6) is an unbiased estimator of $\frac{\partial}{\partial\theta}V(\theta)$.

Proof. Omitted here due to space considerations. The detailed proofs of this theorem and most of the following results are contained in Su and Fu (2000).

Corollary 1 *(Convexity) If* $f(\theta,\omega)$ *and* $C(T,\omega)$ *satisfy the conditions in Theorem 1 and in addition,*

$$\frac{\partial^2}{\partial\theta^2}f(\theta,\omega)>0 \qquad Q\text{-a.s.,}$$

$$\|\frac{\partial}{\partial\theta}f(\theta+\Delta\theta,\omega)-\frac{\partial}{\partial\theta}f(\theta,\omega)\|<M(\theta,\omega)\|\Delta\theta\|$$

uniformly when $\Delta\theta \to 0$,

$$E^Q[M(\theta,\omega)] < +\infty,$$

then $V(\theta)$ *is a convex function of* θ.

4 CHANGE OF DRIFT IN BROWNIAN MOTION

4.1 Mathematical Framework

Suppose the underlying asset price under the risk-neutral measure Q is an Itô process defined by the following stochastic differential equation:

$$dS_t = \mu(S_t,t)dt + \sigma(S_t,t)d\tilde{W}_t, \tag{9}$$

where $\tilde{W}_t$ is a standard Brownian motion under Q. We define the family of $P(\theta)$ as all the equivalent probability measures w.r.t. Q introduced by changing the drift term of $\tilde{W}_t$ by θ. Then by Girsanov's theorem, we know under $P(\theta)$,

$$dS_t = (\mu(S_t,t) + \theta\sigma(S_t,t))\,dt + \sigma(S_t,t)dW_t, \tag{10}$$

where W_t is a Brownian motion under P, and

$$W_t = \tilde{W}_t - \theta t.$$

The change of measure process is given by

$$\begin{aligned}\frac{dQ}{dP} &= \exp\left(-\theta W_T - \frac{1}{2}\theta^2 T\right)\\ &= \exp\left(-\theta\tilde{W}_T + \frac{1}{2}\theta^2 T\right),\end{aligned}$$

so

$$\frac{\partial f(\theta,\omega)}{\partial\theta} = \left(-\tilde{W}_T + \theta T\right)e^{\left(-\theta\tilde{W}_T + \frac{1}{2}\theta^2 T\right)}. \tag{11}$$

Example: If $\{S_t\}$ follows a geometric Brownian motion, then

$$dS_t = \mu S_t dt + \sigma S_t d\tilde{W}_t,$$

where μ is the drift (mean rate of return) and σ is the volatility (standard deviation rate of return). If we define

$$\lambda = \mu + \theta\sigma,$$

then λ is the mean rate of return of S_t under P. Thus, we can also use the rate of return λ as the parameter, since it is equivalent to θ. The IPA estimator given by (6) in terms of λ is

$$\hat{C}^2(T,\omega)\left(-\frac{\tilde{W}_T}{\sigma} + \frac{\lambda-\mu}{\sigma^2}T\right)e^{\left(-\frac{\lambda-\mu}{\sigma}\tilde{W}_T + \frac{1}{2}\frac{(\lambda-\mu)^2}{\sigma^2}T\right)}. \tag{12}$$

In our computational experiments, we use λ instead of θ to compare with Vazquez-Abad and Dufresne (1998), whose results are expressed in terms of λ.

4.2 Convergence Properties of IPA Estimator

In this section, we present some nice properties for importance sampling applied to a change of drift in Brownian motion.

Theorem 2 *(Unbiasedness under Q) For an asset price process described by the Itô process (9), if*

$$E^Q\left[C(T,\omega)\right]^{2+2\delta} < +\infty,\ \delta > 0,$$

then

$$\hat{C}^2(T,\omega)\left(-\tilde{W}_T + \theta T\right)e^{\left(-\theta\tilde{W}_T + \frac{1}{2}\theta^2 T\right)}$$

is an unbiased estimator of $\frac{\partial}{\partial\theta}V(\theta)$ *under Q.*

Corollary 2 *(Convexity) If the asset price process is given by the Itô process (9), then* $V(\theta)$ *is a convex function.*

For deep out of money options, $C(T,\omega)$ will be 0 most of the time under measure Q, and this could lead to large variance when estimating the gradient. However, we can perform a measure change to obtain a new IPA estimator under P, which is given by

$$\hat{C}^2(T,\omega)\,(-W_T)\exp\left(-2\theta W_T - \theta^2 T\right). \tag{13}$$

Corollary 3 *(Unbiasedness of estimator under* P*) Under the same conditions as in Theorem 2, the new estimator given by (13) is unbiased for* $\frac{\partial}{\partial\theta}V(\theta)$.

In the computational experiments, we use the new IPA estimator given by (13) and call it IPA-Q, because it was derived under Q. In terms of λ, it becomes

$$\hat{C}^2(T,\omega)\left(-\frac{W_T}{\sigma}\right)e^{-\frac{2(\lambda-\mu)}{\sigma}W_T - (\frac{\lambda-\mu}{\sigma})^2 T}. \tag{14}$$

Next, we state a convergence theorem for (5).

Theorem 3 *(Fu 1990) If $\forall\theta \in \Theta$, $\frac{\partial}{\partial\theta}V(\cdot)$ is continuous in θ, $V(\cdot)$ is convex and therefore has a unique minimum $\theta^* \in \Theta$, where Θ is a compact set, and*

$$\theta_{n+1} = \theta_n - a_n g_n(\theta_n),$$
$$\sup_{\theta\in\Theta} E[g_n^2(\theta)] < K < \infty,$$
$$E[g_n(\theta_n)|\mathcal{F}_n] = \frac{\partial}{\partial\theta}V(\theta_n) + \beta_n,$$
$$\text{where } \sum_{j=n}^{\infty} |a_j\beta_j| < \infty,$$
$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty,$$

then $\theta_n \longrightarrow \theta^*$ a.s.

It is easy to verify that the IPA estimator given in (13) satisfies the conditions above and thus strongly converges to the true optimum. The algorithm of applying importance sampling via optimal change of drift in Itô process (9) is as follows.

- Stage I: Optimization stage – Find θ^*.
  Initialization: Set $\theta = \theta_0$.
  Loop: For $n = 1$ to N_1
  - For $i = 1$ to N_2
    - Generate sample path according to (10);
    - Record S_t and W_t;
    - Calculate IPA-Q$_i$ based on (13);
  - $g_n(\theta_n) = \frac{1}{N_2}\sum_{i=1}^{N_2}$ IPA-Q$_i$;
  - $\theta_{n+1} = \theta_n - a_n g_n(\theta_n)$;
  - If $|a_n g_n(\theta_n)| < \epsilon$, exit loop.

  Set $\theta^* = \theta_{n+1}$.
- Stage II: Pricing stage – Simulate at $\theta = \theta^*$.
  For $i = 1$ to N_3
  - Generate sample path according to (10);
  - Record S_t and W_t;
  - Calculate $\hat{C}_i = \hat{C}(T,\omega)\frac{dQ}{dP}$.

  Final price $C_0 = \frac{1}{N_3}\sum_{i=1}^{N_3}\hat{C}_i$.

The algorithm is characterized by the parameters N_1, N_2, N_3, ϵ, and $\{a_n\}$:

$N_1 =$ maximum # of iterations,
$N_2 =$ # replications per iterations,
$N_3 =$ # replications used in pricing stage,
$\epsilon =$ stopping rule precision, and
$a_n =$ step size multiplier of nth iteration.

Remark: An alternative method used in Vazquez-Abad and Dufresne (1998) is to use the sample paths in the optimization stage for estimation, as well.

5 COMPUTATIONAL EXPERIMENTS

5.1 Comparisons Between Two Estimators

We consider Asian options as in Vazquez-Abad and Dufresne (1998), where the underlying stock follows geometric Brownian motion

$$\frac{dS_t}{S_t} = rdt + \sigma d\tilde{W}_t, \tag{15}$$

where r is the risk-free interest rate and σ is the volatility. The payoff function of the option at maturity T is given by

$$C(T) = (A(T) - K)^+, \tag{16}$$

where the average price is defined over the equally spaced discrete time points $N_0 + 1, ..., N$, i.e.,

$$A(T) = \frac{1}{N - N_0}\sum_{i=N_0+1}^{N} S_{\frac{i}{N}T}. \tag{17}$$

We first compare our IPA estimator (IPA-Q) with the initial IPA estimator given in Vazquez-Abad and Dufresne (1998) (denoted henceforth by IPA-VD):

$$2e^{-2rT}(A(T) - K)^+ f^2(\lambda)\times$$
$$\left[(A(T) - K)\left(\frac{W_T}{\sigma} - \frac{(\lambda-r)T}{\sigma^2}\right) + \frac{1}{N-N_0}\sum_{i=N_0+1}^{N}\frac{iT}{N}S_{\frac{iT}{N}}\right],$$

where

$$f(\lambda) = \exp\left(-\frac{\lambda - r}{\sigma}W_T - \frac{(\lambda - r)^2}{2\sigma^2}T\right).$$

The IPA-Q estimator in this case is given by

$$e^{-2rT}\left[A(T) - K)^+\right]^2 f^2(\lambda)\left(-\frac{W_T}{\sigma}\right).$$

The initial stock price is $S_0 = 50$, K=50, σ^2=0.2, r=0.05, T=1.0 year, $N_0 = 0$, and $A(T)$ is a daily average, so that $N = T$.

Table 1 provides 95% confidence intervals based on 50,000 replications, and the final variance ratios are listed in the last column.

Remark: We see that the variance of IPA estimator given in our method is significantly smaller than the variance of estimator given in Vazquez-Abad and Dufresne (1998).

Table 1: Asian Call Options: $S_0 = 50$, K=50, σ^2=0.2, r=0.05, T=1.0 Yr

Derivative Estimation via IPA					
	IPA-VD		IPA-Q		
λ	$\frac{\partial V}{\partial \lambda}$	C.I.	$\frac{\partial V}{\partial \lambda}$	C.I.	VR
0.2	-175.5	15.7	-178.8	4.28	13
0.3	-93.4	9.2	-96.1	2.06	20
0.4	-38.7	7.3	-40.6	1.44	26
0.5	3.89	8.3	3.83	1.89	19
0.6	45.44	12.0	48.39	3.46	12
0.7	94.88	22.2	104.97	7.41	9.0
0.8	168.82	41.6	190.81	16.86	6.1

Table 2: Asian Call Options: $S_0 = 50$, σ^2=0.2, r=0.05, T=1.0 Yr, ϵ=0.001, N_1=20, N_2=50

Convergence Property of IPA							
	IS via IPA-Q				IS via Optimal λ^*		
	Price	C.I.	λ	N_1^*	Price	C.I.	λ^*
K=30	20.407	0.134	0.26	15	20.407	0.135	0.25
K=45	8.320	0.114	0.43	20	8.318	0.115	0.40
K=50	5.675	0.096	0.53	19	5.672	0.096	0.50
K=55	3.713	0.076	0.55	20	3.718	0.076	0.60
K=75	0.575	0.022	0.79	18	0.574	0.022	0.80

The reason is that when calculating IPA estimator, we only need consider the derivative of $\frac{dQ}{dP}$ w.r.t. λ, while Vazquez-Abad and Dufresne (1998) had to consider both the derivative of $\frac{dQ}{dP}$ and the derivative of the payoff function w.r.t. λ.

5.2 Convergence Property

We test the convergence property of our algorithm in this experiment. Again, the parameter used here is λ, and the initial starting value of λ_0 is chosen such that $S_0 = e^{-\lambda_0 T} K$, because then the expected terminal stock price would be at the strike price. We use $N_1 = 20$ iterations and $N_2 = 50$ sample paths in the optimization stage and the stopping criteria $\epsilon = 0.001$, so the total number of simulations used in the optimization stage is less than 1000. We took $a_n = a_0 n^{-0.75}$, where $a_0 = \left|\frac{1}{g_0(\lambda_0)}\right|$. Also, we restrict that in each step $|\Delta\lambda| \leq$0.2. We use N_3 =10,000 simulations in the final estimation stage, where we estimate the option price. In this experiment, the stock prices follow the same geometric Brownian motion as in the last example, $S_0 = 50$, $\sigma^2 = 0.2$, r=0.05, and T=1.0 year with strike prices K=30, 45, 50, 55, 75. The optimal values of λ^* reported are taken from Vazquez-Abad and Dufresne (1998), which are obtained by an extensive brute-force search.

From Table 2, we see that our algorithm converges very fast, coming very close to the optimal value using less than 1000 simulations, where N_1^* is the actual # of iterations used in optimization stage.

5.3 Comparison Between Importance Sampling and Naive Simulations

5.3.1 Asian Options on Partial Average

In this testbed, the stock price again follows geometric Brownian motion as given by (15), with payoff function defined by (16). However, $N_0 \neq 0$. In other words, the average begins at a date N_0 other than at time 0. The other parameter values are S_0=100, σ=0.2, 0.3, r=0.05, 0.09, and T=1.0 year; $A(T)$ is the average daily stock price with the averaging beginning 60 days before the option's maturity date. To test the effect of moneyness on the variance reduction, we consider a range of strike prices: K=100, 110, 120, 130, 140, 150, 160, 170. The algorithm parameter values used are N_1=50, N_2=100, N_3=50,000, $\epsilon = 0.0005$. The other settings are the same as before, and the results are shown in Tables 3 and 4.

As we expect, the computational gains of implementing importance sampling increase dramatically with increasing strike price (more out of the money). For the case of r=0.05, σ=0.2, the variance reduction starts from 7 for the at-the-money call option at K=100 and increases to 173 for the deep out-of-the-money call option at K=170. We also observe an interesting phenomena that as the option price increases with increasing interest rate or volatility, the effectiveness of importance sampling decreases. Our conjecture is that higher values of these parameters increase

Table 3: Asian Call Options on Partial Average: $S_0 = 100$, T=1.0 Yr, ϵ=0.0005, N_1=50, N_2=100, N_3=50000

r=0.05, $\sigma = 0.2$						
	IS via IPA-Q			Naive		
	Price	C.I.	λ	Price	C.I.	VR
K=100	9.747	0.047	0.251	9.840	0.121	6.7
K=110	5.397	0.033	0.308	5.461	0.094	8.2
K=120	2.730	0.020	0.368	2.749	0.067	11
K=130	1.284	0.011	0.430	1.317	0.046	17
K=140	0.575	0.006	0.473	0.544	0.029	25
K=150	0.241	0.003	0.487	0.225	0.019	44
K=160	0.098	0.001	0.501	0.102	0.013	85
K=170	0.038	0.001	0.533	0.040	0.008	173
r=0.09, $\sigma = 0.2$						
K=100	11.732	0.052	0.268	11.839	0.129	6.1
K=110	6.850	0.038	0.335	6.897	0.103	7.3
K=120	3.678	0.025	0.388	3.740	0.078	10
K=130	1.854	0.015	0.441	1.795	0.058	13
K=140	0.867	0.008	0.489	0.846	0.036	19
K=150	0.384	0.004	0.524	0.388	0.025	38
K=160	0.163	0.002	0.548	0.170	0.016	67
K=170	0.067	0.001	0.540	0.068	0.010	106

Table 4: Asian Call Options on Partial Average: $S_0 = 100$, T=1.0 Yr, ϵ=0.0005, N_1=50, N_2=100, N_3=50000

r=0.05, $\sigma = 0.3$						
	IS via IPA-Q			Naive		
	Price	C.I.	λ	Price	C.I.	VR
K=100	13.295	0.068	0.368	13.421	0.185	7.3
K=110	9.103	0.054	0.443	9.119	0.155	8.4
K=120	6.059	0.041	0.507	6.171	0.131	10
K=130	3.985	0.030	0.557	3.885	0.104	12
K=140	2.556	0.022	0.595	2.475	0.084	15
K=150	1.603	0.015	0.629	1.635	0.070	22
K=160	1.006	0.010	0.693	1.012	0.055	30
K=170	0.623	0.007	0.736	0.587	0.041	36
r=0.09, $\sigma = 0.3$						
K=100	15.063	0.074	0.377	15.197	0.193	6.8
K=110	10.571	0.059	0.448	10.597	0.166	7.9
K=120	7.214	0.046	0.525	7.334	0.142	9.5
K=130	4.856	0.035	0.561	4.732	0.115	11
K=140	3.184	0.027	0.603	3.132	0.094	13
K=150	2.046	0.018	0.676	2.105	0.079	19
K=160	1.314	0.013	0.710	1.314	0.063	25
K=170	0.836	0.009	0.764	0.802	0.049	31

the likelihood of options finishing in the money, reducing the power of importance sampling.

5.3.2 Asian-Digital Options

The underlying stock again follows geometric Brownian motion as in (15) with the daily average $A(T)$ given by (17), but the payoff function is given by

$$C(T) = 10 * \mathbf{1}\{A(T) > K\}.$$

Clearly, Vazquez-Abad and Dufresne (1998) is not applicable in this case, since the digital function is not differentiable. The parameter settings in the two-stage algorithm and asset prices are the same as in the case of Asian option on partial average. The results are shown in Tables 5 and 6, and we observe similar behavior as for the other Asian options, although the ratio of variance reduction is somewhat lower, probably due to the smaller range (0 or 1) for the Asian-digital option payoff.

6 CONCLUSION

It is well known that changing the drift in Brownian motion via importance sampling can be used to effectively reduce the estimation error in security pricing. The stochastic optimization approach we present here is capable of finding the optimal change of drift efficiently. In all cases, the computational overhead added is less than 10% of the total computational time, whereas for deep out-of-the-money options, the computational gains range from 10 to 170 times in our simulation experiments, and in all cases, we report significant variance reductions from the simulation results. We also notice that for the equity options, the efficiency of changing the drift term in Brownian motion decreases as the interest rate increases or the volatility increases. The IPA estimator developed here is more widely applicable and has substantially smaller variance than the estimator in Vazquez-Abad and Dufresne (1998), and is also not limited to just changes of drift in Brownian motion. In Su and Fu (2000), a number of other examples are considered, including interest rate derivatives based on the Cox, Ingersoll, and Ross (1985) interest rate model (see also Black, Derman, and Toy 1990).

REFERENCES

Anderson, L. 1995. Efficient techniques for simulation of interest rate models involving non-linear stochastic differential equations. Working paper, General Re Financial Products, New York, New York.

Black, F., Derman, E., and Toy, W. 1990. A one-factor model of interest rates and its application to treasury bond options. *Financial Analysts Journal* 46:33–39.

Boyle, P., Broadie, M., and Glasserman, P. 1997. Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control* 21:1257–1321.

Cox, J., Ingersoll, J., and Ross, A. 1985. A theory of the term structure of interest rates. *Econometrica* 53 (2): 385–408.

Fu, M. 1990. Convergence of the GI/G/1 queue using infinitesimal perturbation analysis. *Journal of Optimization Theory and Applications* 65:149–160.

Fu, M., Madan, D., and Wang, T. 1999. Pricing continuous time Asian options: A comparison of Monte Carlo and Laplace transform inversion methods. *Journal of Computational Finance* 2 (2): 49–74.

Glasserman, P., Heidelberger, P., and Shahabuddin, P. 1998. Importance sampling in the Heath-Jarrow-Morton framework. Working paper, Columbia University, New York, New York.

Harrison, J., and Kreps, D. 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20:381–408.

Heath, D., Jarrow, R., and Merton, A. 1992. Bond pricing and the term structure of interest rates. *Econometrica* 60 (1): 77–105.

Reider, R. 1993. An efficient Monte Carlo technique for pricing options. Working paper, Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania.

Hull, J., and White, A. 1987. The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42:281–300.

Hull, J., and White, A. 1987. The use of the control variate technique in option pricing. *Journal of Financial and Quantitative Analysis* 23:237–251.

Su, Y., and Fu, M. 2000. Optimal importance sampling for derivative securities pricing. Working Paper, University of Maryland, College Park, Maryland.

Turnbull, S., and Wakeman, L. 1991. A quick algorithm for pricing European average options. *Journal of Financial and Quantitative Analysis* 26:377–389.

Vazquez-Abad, F., and Dufresne, D. 1998. Accelerated simulation for pricing Asian options. In *Proceedings of the 1998 Winter Simulation Conference*, ed. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 1493–1500. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

AUTHOR BIOGRAPHIES

YI SU is currently a PhD Candidate in the Robert H. Smith School of Business, at the University of Maryland. He received his B.S. and M.S. in statistics from Peking University, China. His research interests include simulation and mathematical finance, particularly with applications in financial engineering. He is a member of **INFORMS**.

Table 5: Asian Digital Call Options on Partial Average: $S_0 = 100$, T=1.0 Yr, ϵ=0.0005, N_1=50, N_2=100, N_3=50000

r=0.05, $\sigma = 0.2$						
	IS via IPA-Q			Naive		
	Price	C.I.	λ	Price	C.I.	VR
K=100	5.300	0.033	0.149	5.358	0.041	1.6
K=110	3.403	0.028	0.211	3.431	0.040	2.0
K=120	1.947	0.020	0.278	1.988	0.034	3.0
K=130	1.017	0.012	0.344	0.994	0.026	4.4
K=140	0.481	0.007	0.402	0.479	0.018	7.4
K=150	0.213	0.003	0.453	0.211	0.012	13
K=160	0.090	0.002	0.484	0.093	0.008	26
K=170	0.036	0.001	0.531	0.036	0.005	53
r=0.09, $\sigma = 0.2$						
K=100	5.772	0.032	0.164	5.814	0.039	1.4
K=110	3.953	0.030	0.214	3.986	0.040	1.8
K=120	2.420	0.023	0.274	2.465	0.036	2.5
K=130	1.345	0.015	0.356	1.317	0.028	3.5
K=140	0.678	0.009	0.399	0.680	0.021	5.7
K=150	0.318	0.008	0.470	0.322	0.015	9.9
K=160	0.143	0.002	0.499	0.143	0.010	17
K=170	0.061	0.001	0.530	0.055	0.006	27

Table 6: Asian Digital Call Options on Partial Average: $S_0 = 100$, T=1.0 Yr, ϵ=0.0005, N_1=50, N_2=100, N_3=50000

r=0.05, $\sigma = 0.3$						
	IS via IPA-Q			Naive		
	Price	C.I.	λ	Price	C.I.	VR
K=100	4.829	0.033	0.216	4.866	0.042	1.7
K=110	3.549	0.029	0.281	3.578	0.040	2.0
K=120	2.518	0.024	0.318	2.554	0.037	2.5
K=130	1.730	0.018	0.375	1.715	0.032	3.1
K=140	1.148	0.014	0.496	1.142	0.027	4.0
K=150	0.741	0.009	0.508	0.764	0.023	5.8
K=160	0.476	0.007	0.559	0.471	0.018	7.3
K=170	0.299	0.005	0.611	0.296	0.015	10
r=0.09, $\sigma = 0.3$						
K=100	5.094	0.032	0.237	5.145	0.040	1.5
K=110	3.868	0.029	0.283	3.897	0.040	1.8
K=120	2.806	0.025	0.352	2.847	0.037	2.3
K=130	2.001	0.020	0.417	1.975	0.033	2.8
K=140	1.354	0.015	0.480	1.354	0.029	3.6
K=150	0.903	0.011	0.516	0.921	0.024	4.8
K=160	0.590	0.008	0.583	0.587	0.020	6.3
K=170	0.381	0.006	0.608	0.378	0.016	8.4

MICHAEL C. FU is a Professor in the Robert H. Smith School of Business, with a joint appointment in the Institute for Systems Research, at the University of Maryland. He received degrees in mathematics and EE/CS from MIT, and a Ph.D. in applied mathematics from Harvard University. His research interests include simulation and applied probability modeling, particularly with applications towards manufacturing systems, inventory control, and financial engineering. He teaches courses in applied probability, stochastic processes, simulation, computational finance, and operations management, and in 1995 was awarded the Maryland Business School's Allen J. Krowe Award for Teaching Excellence. He is a member of **INFORMS** and **IEEE**, and a Senior Member of **IIE**. He is currently the Simulation Area Editor of *Operations Research*, and serves on the editorial boards of *Management Science*, *INFORMS Journal on Computing*, *IIE Transactions*, and *Production and Operations Management*. He is co-author (with J.Q. Hu) of the book, *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*, which received the INFORMS College on Simulation Outstanding Publication Award in 1998.