Journal of Computational Finance, Vol.5, No.4, 27-50, 2002

OPTIMAL IMPORTANCE SAMPLING IN SECURITIES PRICING*

Yi Su
Michael C. Fu
Robert H. Smith School of Business
Van Munching Hall
University of Maryland
College Park, MD 20742-1815

September, 2000; revised February 2001

Abstract

To reduce variance in estimating security prices via Monte Carlo simulation, we formulate a parametric minimization problem for the optimal importance sampling measure, which is solved using infinitesimal perturbation analysis (IPA) and stochastic approximation (SA). Compared with existing methods, the IPA estimator we derive is more universally applicable and more computationally efficient. Under suitable conditions, we show that the objective function is a convex function, the IPA estimator is unbiased, and the stochastic approximation algorithm converges to the optimum. Lastly, we demonstrate how combining importance sampling with indirect estimation using put-call parity can lead to further substantial variance reduction.

*This research was supported in part by the National Science Foundation under Grants DMI-9713720 and DMI-9988867 and by the Semiconductor Research Corporation under Grant 97-FJ-491. An earlier version of this manuscript that accepted for presentation at the 2000 Winter Simulation Conference (Su and Fu 2000) did not include the proofs of the theoretical results nor the computational work on fixed income securities and indirect estimation.

1 INTRODUCTION

Since being introduced to the pricing of options by Boyle (1977), Monte Carlo simulation has been 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 or of the dynamics of the underlying assets increase, Monte Carlo simulation often becomes the sole computationally feasible means of security pricing. The efficiency of Monte Carlo simulation 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 the precision of the estimator $\hat{p}$ is generally proportional to the standard error $\sigma_{\hat{p}}/\sqrt{N}$, where N is the number of simulations and $\sigma_{\hat{p}}^2$ is the estimator variance. 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 variance reduction techniques (VRT) in Monte Carlo simulation 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 as widely used as other VRTs in pricing financial derivatives until recently. The idea behind importance sampling is to concentrate simulation on sample paths that contribute most to estimating the expected payoff; for instance, the payoff for a deep out-of-the-money call option will be 0 most of the time, so simulating more sample paths with positive payoffs should reduce the estimation variance. 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 alternative probability measure will lead to effective variance reduction.

An early example of importance sampling applied to derivatives pricing is Reider (1993), where increasing the drift in the underlying geometric Brownian motions substantially decreases the variance in simulations for deep out-of-the-money European call options. Glasserman, Heidelberger, Shahabuddin (1999ab) apply importance sampling in a framework where the underlying processes are derived from normal random vartiables and in the Heath, Jarrow, Morton (1992) framework, reporting substantial variance reduction by combining stratified sampling and a change of drift. 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 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 they do, so our method is applicable in much more general settings. 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 a conjecture in Vazquez-Abad and Dufresne (1998). We further prove that our stochastic approximation algorithm a.s. converges to the true global optimum.

The rest of the paper is organized as follows. The optimal importance sampling problem setting is introduced in the next section. The stochastic approximation approach to solving the optimization problem, the derivation of the IPA estimator, and the proofs of various theoretical properties are presented in Section 3. The specific case of changing the drift in Brownian motion is treated in Section 4, with the detailed algorithm provided. Section 5 describes the numerical results from computational experiments on a testbed of cases, including Asian options and caps/caplets. The idea of indirect estimation using put-call parity is also proposed here and tested on simple vanilla options and Asian options. Section 6 concludes the paper with a brief summary and a direction for pursuing further extensions.

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 $C(T,\omega)$.

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

where $S_t(\omega)$ is the stock price at time t (superscripted for the max-option on a basket of stocks) on sample path ω, K is the strike price, L is the barrier value for the last example, and $\mathbf{1}\{\cdot\}$ is the indicator function. Henceforth, dependence on ω will be understood, so explicit display will be dropped for notational brevity.

The 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 $\hat{C}(T)$. However, by the Radon-Nikodym theorem, if measure Q is absolutely continuous w.r.t. some other measure Q^*, then

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

which gives an alternative estimator for simulation under Q^*:

$$\hat{C}(T)\frac{dQ}{dQ^*}, \tag{1}$$

where $\frac{dQ}{dQ^*}$ is the Jacobian of the measure change, i.e., the Radon-Nikodym derivative. Although estimator (1) is also an unbiased estimator of the option price C_0, the new estimator may have different estimation variance, hence the potential for variance reduction.

In our settings, we restrict the transformed measure to be in a family of probability measures $\{Q^*(\theta) : \theta \in \Theta\}$, where θ is the parameter, Θ is a compact set and for any $\theta \in \Theta$, measure Q is absolutely continuous w.r.t. $Q^*(\theta)$. We consider the problem of finding the value of θ that minimizes the variance of the new estimator (1), which is given by

$$E^{Q^*}\left[\left(\hat{C}(T)\frac{dQ}{dQ^*}\right)^2\right] - C_0^2 = E^{Q^*}\left[\hat{C}^2(T)\left(\frac{dQ}{dQ^*}\right)^2\right] - C_0^2.$$

Since C_0 is a constant, this leads to the following stochastic optimization problem for the second moment of the estimator:

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

where

$$V(\theta) = E^{Q^*}\left[\hat{C}^2(T)\left(\frac{dQ}{dQ^*}\right)^2\right]. \tag{2}$$

Remark: Vazquez-Abad and Dufresne (1998) derive their IPA estimator by directly differentiating the term inside the expectation of (2), which requires derivatives for both $C(T)$ and $\frac{dQ}{dQ^*}$, since the sample path, and hence $C(T)$, clearly depends on θ (as does $\frac{dQ}{dQ^*}$). This is because sampling is carried out under Q^* rather than Q. However, this is avoided if the minimization is carried out under the measure Q, and this is the fundamental difference between our method and theirs.

Simple calculation shows that

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

So we only need to find the θ that minimizes

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

where

$$L(\theta) = \frac{dQ}{dQ^*(\theta)}. \tag{4}$$

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

3 STOCHASTIC APPROXIMATION AND IPA

Our approach to minimizing $V(\theta)$ follows that of Vazquez-Abad and Dufresne (1998), in that we use gradient-based stochastic approximation (SA) 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 $\theta_n = ((\theta_n)_1, ..., (\theta_n)_k)$ 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) vs. (2).

We first make the following assumption.

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

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

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

where $L(\theta)$ is defined by (4). Under suitable conditions, this IPA estimator is unbiased (under measure Q).

Theorem 1 *(General Unbiasedness) If Assumption 1 holds, and $\exists M(\theta)$ s.t.*

$$\|L(\theta+\Delta\theta)-L(\theta)\| < M(\theta)\|\Delta\theta\| \qquad Q\text{-}a.s$$

uniformly when $\Delta\theta \to 0$; *and either*

$$\exists\delta>0, E^Q\left[C(T)\right]^{2+2\delta} < +\infty, \text{ and } E^Q\left[M(\theta)\right]^{1+1/\delta} < +\infty; \tag{7}$$

or

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

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

Proof: If $C(T)$ and $M(\theta)$ satisfy (8), then by the dominated convergence theorem, we know

$$\frac{\partial}{\partial\theta}E^Q\left[\hat{C}^2(T)L(\theta)\right] = E^Q\left[\hat{C}^2(T)\frac{\partial L(\theta)}{\partial\theta}\right].$$

If $C(T)$ and $M(\theta)$ satisfy (7), by Holder inequality,

$$E^Q\left[C^2(T)M(\theta)\right] \le \{E^Q\left[C^{2+2\delta}(T)\right]\}^{\frac{1}{1+\delta}}\{E^Q\left[M(\theta)^{1+1/\delta}\right]\}^{\frac{1+\delta}{\delta}} < +\infty.$$

Then (8) holds. □

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

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

and $\exists M(\theta)$ s.t.

$$\|\frac{\partial}{\partial\theta}L(\theta+\Delta\theta)-\frac{\partial}{\partial\theta}L(\theta)\| < M(\theta)\|\Delta\theta\|$$

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

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

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

Proof: From Theorem 1, we know

$$\frac{\partial}{\partial\theta}E^Q\left[\hat{C}^2(T)L(\theta)\right] = E^Q\left[\hat{C}^2(T)\frac{\partial L(\theta)}{\partial\theta}\right].$$

Since, $E^Q[M(\theta)] < +\infty$, by the dominated convergence theorem

$$\frac{\partial^2}{\partial\theta^2} E^Q\left[\hat{C}^2(T)L(\theta)\right] = E^Q\left[\hat{C}^2(T)\frac{\partial^2 L(\theta)}{\partial\theta^2}\right] \geq 0.$$

□

Although derived under measure Q, implementation of the gradient estimator can also be carried out under an alternative measure such as Q^*, in which case the estimator given by (6) becomes

$$\hat{C}^2(T)\frac{\partial L(\theta)}{\partial\theta}L(\theta), \tag{9}$$

where $L(\theta)$ is defined by (4). This is likely to be advantageous in the same situations in which a change of measure for estimating the price itself is beneficial, since the gradient estimator also contains the term $\hat{C}^2(T)$. Thus, in all of the computational experiments, we use this form of the IPA estimator, which we will call IPA-Q, because it was derived under Q.

4 CHANGE OF DRIFT IN BROWNIAN MOTION

Mathematical Framework

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

$$dS_t = \mu(S_t, t)dt + \sigma(S_t, t)d\tilde{W}_t, \tag{10}$$

where $\tilde{W}_t$ is a standard Brownian motion under Q. We define the family of $Q^*(\theta)$ 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 $Q^*(\theta)$,

$$dS_t = (\mu(S_t, t) + \theta\sigma(S_t, t))\, dt + \sigma(S_t, t)dW_t, \tag{11}$$

where W_t is a Brownian motion under Q^*, and

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

In this case, the change of measure process is available in closed form as a function of W_t, given by

$$L(\theta) = \frac{dQ}{dQ^*(\theta)} = \exp\left(-\theta W_T - \frac{1}{2}\theta^2 T\right) = \exp\left(-\theta\tilde{W}_T + \frac{1}{2}\theta^2 T\right), \tag{12}$$

so

$$\frac{\partial L(\theta)}{\partial \theta} = \left(-\tilde{W}_T + \theta T\right) e^{\left(-\theta \tilde{W}_T + \frac{1}{2}\theta^2 T\right)} = \left(-\tilde{W}_T + \theta T\right) L(\theta) = -W_T L(\theta). \tag{13}$$

This has a particularly nice form, in that $L(\theta)$ depends only on the terminal value of a Brownian motion and not on any values along the path, *independent* of whether or not the payoff function C is path dependent, as in some of the cases we later consider in our numerical experiments: Asian options and caps.

Example: If $\{S_t\}$ follows 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 Q^*. 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 (applying the chain rule $\partial f/\partial\lambda = \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial \lambda}$)

$$\hat{C}^2(T)\left(-\frac{\tilde{W}_T}{\sigma} + \frac{\lambda-\mu}{\sigma^2}T\right) e^{\left(-\frac{\lambda-\mu}{\sigma}\tilde{W}_T + \frac{1}{2}\frac{(\lambda-\mu)^2}{\sigma^2}T\right)}. \tag{14}$$

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.

Even when the change of measure process $L(\theta) = dQ/dQ^*$ is not available in such a convenient closed form as (12) a priori, it can always be calculated in the simulation by taking the appropriate ratio of the input distributions used to generate the random variables driving the underlying processes, e.g., normal densities in the case of processes involving Brownian motion.

Example: Changing the drift term μ in an Itô process defined by (10) to $\lambda = \mu + \theta\sigma$, as in (11), simulating on $[0, T]$ along N equally spaced discrete time points with spacing Δt ($N\Delta t = T$). Paths under Q^* (resp. Q) can be generated using drift λ (resp. μ) and i.i.d. sampling from $N(0, 1)$ for the standard Wiener process. Using the *same* drift λ to generate paths under Q can be achieved by using i.i.d. sampling from $N(-\theta\sqrt{\Delta t}, 1)$, so the direct likelihood ratio is given by

$$L(\theta) = \frac{dQ}{dQ^*} = \frac{\phi(Z_1; -\theta\sqrt{\Delta t})\phi(Z_2; -\theta\sqrt{\Delta t})...\phi(Z_N; -\theta\sqrt{\Delta t})}{\phi(Z_1; 1)\phi(Z_2; 1)...\phi(Z_N; 1)} = \exp\left(-\frac{\theta^2 T}{2} - \theta\sqrt{\Delta t}\sum_{i=1}^{N} Z_i\right),$$

where $\phi(\cdot; m)$ represents the normal density with mean m and standard deviation 1, and $\{Z_i\}$ are i.i.d. $N(0, 1)$. This expression is just the simulation implementation of (12). Note that in discrete time simulation, we could also change volatility.

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 (10), if*

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

then

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

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

Proof: We only need to show

$$\frac{\partial}{\partial\theta}E^Q\left[\hat{C}^2(T)e^{-\theta\tilde{W}_T}\right] = E^Q\left[\hat{C}^2(T)\left(-\tilde{W}_T\right)e^{-\theta\tilde{W}_T}\right].$$

We know $\forall\ \theta'$ and $\epsilon > 0$, $\theta' \in (\theta - \epsilon, \theta + \epsilon)$,

$$\sup_{\theta' \in (\theta-\epsilon, \theta+\epsilon)} \left(-\tilde{W}_T\right) e^{-\theta\tilde{W}_T} \leq C_1 \left|\tilde{W}_T\right| \exp(C_2 \left|\tilde{W}_T\right|),$$

where C_1, C_2 are constants.

However

$$\begin{aligned} & E^Q(C_1 \left|\tilde{W}_T\right|)^{1+1/\delta} \exp[(1 + 1/\delta)(C_2 \left|\tilde{W}_T\right|)] \\ = \; & \frac{1}{\sqrt{2\pi\sigma\sqrt{T}}} \int_R (C_1 |x|)^{1+1/\delta} \exp[(1 + 1/\delta)C_2 |x| - \frac{x^2}{2\sigma\sqrt{T}}]dx < +\infty, \end{aligned}$$

and the result follows from Theorem 1. □

Corollary 2 *(Convexity) If the asset price process is given by the Itô process (10), then $V(\theta)$ is a convex function.*

Proof:

$$\begin{aligned}\frac{\partial^2 L(\theta)}{\partial\theta^2} &= \frac{\partial}{\partial\theta}\left[\left(-\tilde{W}_T+\theta T\right)e^{\left(-\theta\tilde{W}_T+\frac{1}{2}\theta^2 T\right)}\right] \\ &= \left(\left(-\tilde{W}_T+\theta\right)^2+T\right)e^{\left(-\theta\tilde{W}_T+\frac{1}{2}\theta^2 T\right)} > 0.\end{aligned}$$

However,

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2}E^Q\left[\hat{C}^2(T)L(\theta)\right] &= \frac{\partial}{\partial\theta}E^Q\left[\hat{C}^2(T)\left(-\tilde{W}_T+\theta T\right)e^{\left(-\theta\tilde{W}_T+\frac{1}{2}\theta^2 T\right)}\right] \\ &= \frac{\partial}{\partial\theta}\left\{e^{\frac{1}{2}\theta^2 T}E^Q\left[\hat{C}^2(T)\left(-\tilde{W}_T\right)e^{-\theta\tilde{W}_T}\right]\right\} \\ &\quad+\frac{\partial}{\partial\theta}\left\{e^{\frac{1}{2}\theta^2 T}\theta T E^Q\left[\hat{C}^2(T)e^{-\theta\tilde{W}_T}\right]\right\},\end{aligned} \tag{15}$$

but

$$\frac{\partial}{\partial\theta}E^Q\left[\hat{C}^2(T)e^{-\theta\tilde{W}_T}\right] = E^Q\left[\hat{C}^2(T)\left(-\tilde{W}_T\right)e^{-\theta\tilde{W}_T}\right], \tag{16}$$

similarly

$$\frac{\partial}{\partial\theta}E^Q\left[\hat{C}^2(T)\left(-\tilde{W}_T\right)e^{-\theta\tilde{W}_T}\right] = E^Q\left[\hat{C}^2(T)\left(-\tilde{W}_T\right)^2 e^{-\theta\tilde{W}_T}\right]. \tag{17}$$

The result follows by substituting (16) and (17) into (15). □

As discussed at the end of Section 3, it is usually preferable to use the form of the IPA-Q estimator given by (9). For example, for deep out-of-the-money options, $C(T)$ will be 0 most of the time under measure Q, and this could lead to large variance when estimating the gradient. For the Brownian motion setting, the IPA-Q estimator under Q^* becomes

$$\hat{C}^2(T)\left(-W_T\right)\exp\left(-2\theta W_T-\theta^2 T\right). \tag{18}$$

Corollary 3 *(Unbiasedness of estimator under Q^*) Under the same conditions as in Theorem 2, the new estimator given by (18) is unbiased for $\frac{\partial}{\partial\theta}V(\theta)$.*

Proof: By Girsonov's theorem, we know

$$\begin{aligned}E^{Q^*}\left[\hat{C}^2(T)\left(-W_T\right)e^{\left(-2\theta W_T-\theta^2 T\right)}\right] &= E^{Q^*}\left[\hat{C}^2(T)\left(-W_T\right)\left(\frac{dQ}{dQ^*}\right)^2\right] \\ &= E^Q\left[\hat{C}^2(T)\left(-\tilde{W}_T+\theta T\right)e^{\left(-\theta\tilde{W}_T+\frac{1}{2}\theta^2 T\right)}\right] \\ &= \frac{\partial}{\partial\theta}E^Q\left[\hat{C}^2(T)L(\theta)\right].\end{aligned}$$

□

If the underlying asset prices follow a geometric Brownian motion, the IPA-Q estimator in terms of the mean rate of return λ is

$$\hat{C}^2(T)\left(-\frac{W_T}{\sigma}\right)L^2(\lambda), \tag{19}$$

where

$$L(\lambda) = \exp\left(-\frac{\lambda - r}{\sigma}W_T - \frac{(\lambda - r)^2}{2\sigma^2}T\right). \tag{20}$$

In this case, the IPA estimator given in Vazquez-Abad and Dufresne (1998) (denoted henceforth by IPA-VD) can be expressed as

$$2e^{-2rT}C(T)L^2(\lambda)\left[C(T)\left(\frac{W_T}{\sigma} - \frac{\lambda - r}{\sigma^2}T\right) + \frac{\partial C}{\partial \lambda}\right]. \tag{21}$$

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*

$$\begin{aligned} \theta_{n+1} &= \theta_n - a_n g_n(\theta_n), \\ \sup_{\theta\in\Theta} E[g_n^2(\theta)] &< K < \infty, \\ E[g_n(\theta_n) \| \mathcal{F}_n] &= \frac{\partial}{\partial\theta}V(\theta_n) + \beta_n, \end{aligned}$$

where

$$\begin{aligned} \sum_{j=n}^{\infty} |a_j\beta_j| &< \infty, \\ \sum_{n=1}^{\infty} a_n &= \infty, \ \sum_{n=1}^{\infty} a_n^2 < \infty, \end{aligned}$$

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

It is easy to verify that the IPA estimator given in (18) satisfies the conditions above and thus the stochastic approximation (SA) iteration strongly converges to the true optimum. The algorithm for applying importance sampling via an optimal change of drift in Itô process (10) 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 (11);
    - Record S_t and W_t;
    - Calculate IPA-Q_i based on (18);
  - $g_n(\theta_n) = \frac{1}{N_2}\sum_{i=1}^{N_2}$ 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 (11);
  - Record S_t and W_t;
  - Calculate $\hat{C}_i = \hat{C}(T, \omega)L(\theta^*)$.

  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,

ϵ = 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

Comparisons Between Two Estimators

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

$$\frac{dS_t}{S_t} = r dt + \sigma d\tilde{W}_t, \tag{22}$$

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

$$C(T) = (A(T) - K)^+, \tag{23}$$

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

$$A(T) = \frac{1}{N - N_0} \sum_{i=N_0+1}^{N} S_{\frac{i}{N}T}. \tag{24}$$

We first compare our IPA-Q estimator with the IPA-VD estimator for this case, where (21) becomes

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

where $L(\lambda)$ is defined by (20). The IPA-Q estimator in this case, using (19), is given by

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

IPA-Q is much simpler, because it only involves the derivative of $L(\lambda)$, whereas IPA-VD includes both the derivative of $L(\lambda)$ and the derivative of the payoff function w.r.t. λ.

For the computational experiments, 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 half-widths (indicated by the heading "C.I.") based on 50,000 replications and the variance reduction ratios (listed in the last column under the heading "VR"), which are a measure of the degree of computational savings achieved by IPA-Q over IPA-VD, i.e.,

$$VR = \frac{s^2_{IPA-VD}}{s^2_{IPA-Q}} \approx \frac{n_{IPA-VD}}{n_{IPA-Q}},$$

where s^2 is the sample variance and n is the number of simulation replications. For example, in order to achieve the same level of precision as IPA-Q in the first case (λ=0.2), IPA-VD would

	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 1: Asian call options: $S_0 = 50$, K=50, σ^2=0.2, r=0.05, T=1.0 yr.

require approximately 13 times as many simulations or approximately 650,000 replications versus 50,000 replications. Note that since both the old and new estimators are theoretically unbiased, reporting variance is sufficient to compare the performance of the estimators. In all cases, which cover all those reported in Vazquez-Abad and Dufresne (1998), the variance of IPA-Q is significantly smaller than the variance of IPA-VD.

Convergence Property

We test the convergence property of our algorithm in this experiment using the parameter λ. The initial starting value of λ_0 is chosen such that $S_0 = e^{-\lambda_0 T} K$, so that 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 with 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. In this experiment, the stock prices follow the same geometric Brownian motion as in the last example, i.e., $S_0 = 50$, $\sigma^2 = 0.2$, r=0.05, and T=1.0 year, with varying strike prices K=30, 45, 50, 55, 75. The optimal values of λ^* reported are taken from Vazquez-Abad and Dufresne (1998), obtained by an extensive brute-force search.

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

Table 2: Asian call options: $S_0 = 50$, σ^2=0.2, r=0.05, T=1.0yr, ϵ=0.001, N_1=20, N_2=50.

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.

Comparison Between Importance Sampling and Naive Simulations

Asian Options on Partial Average

In this testbed, the stock price again follows geometric Brownian motion as given by (22), with payoff function defined by (23). 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. For naive simulations, we do the simulations under $\lambda = r$.

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

Asian-Digital Options

The underlying stock again follows geometric Brownian motion as in (22), with the daily average $A(T)$ given by (24), but the payoff function 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.

Fixed Income Securities on CIR Model

The Cox-Ingersoll-Ross (1985) CIR model for the short-term interest rate process $\{r(t)\}$ follows the following stochastic differential equation:

$$dr(t) = (\eta - kr(t))dt + \sigma\sqrt{r(t)}d\tilde{W}_t, \tag{25}$$

where $\tilde{W}_t$ is a Brownian motion under the risk-neutral measure Q. Under measure $Q^*(\theta)$, obtained by changing the drift of Brownian motion by θ, we have

$$dr(t) = (\eta + \theta\sigma\sqrt{r(t)} - kr(t))dt + \sigma\sqrt{r(t)}dW_t,$$

where W_t is a Brownian motion under Q^* and $W_t = \tilde{W}_t - \theta t$. Note that since the underlying driving process is again Brownian motion, the change of measure process is given by (12). Since explicit expression of $r(t)$ as a function of θ is unavailable, the IPA-VD estimator is not applicable.

A European caplet option is an option written on the interest rate with the following payoff function at maturity T :

$$C(T) = M(r(T) - K)^+, \tag{26}$$

where K is the strike rate, and M is the nominee amount. The price of the caplet at time 0 is given by

$$C_0 = E^Q \left[\exp\left(-\int_0^T r(t)dt \right) * M(r(T) - K)^+ \right]. \tag{27}$$

Although we use the CIR model here to illustrate the procedure of applying our approach to fixed income derivative securities, our approach is not limited to the CIR model. It is also applicable to some other calibratible models like Black, Derman, Toy (1990).

We apply the two-stage stochastic optimization algorithm to find the optimal θ^* for importance sampling, based on the IPA-Q estimator

$$\exp\left(-\int_0^T r(t)dt \right) C^2(T)(-W_T)L^2(\theta), \tag{28}$$

where $L(\theta)$ is given by (12). For the purpose of estimating the continuous discounting in (27) and (28), the interest rate process $\{r(t)\}$ is simulated for 300 equally spaced time steps on $[0, T]$. The parameter values for the interest rate process are η=0.016, k=0.2, and σ=0.02, with a current spot rate of $r(0)$=0.08. The nominee amount is M=1000, the strike rates are K=0.06, 0.07, 0.08, 0.09, and the maturities are T=0.5, 1.0, 2.0 years. For the optimization algorithm, we use θ_0=0 and $a_0 = \left| \frac{1}{g_0(\theta_0)} \right|$, with the other parameter settings remaining the same as used for the equity options, except when K=0.09, we simply run 50 iterations in the first stage, i.e., let $\epsilon = 0.0$. The simulation results are summarized in Table 7.

A more complicated fixed income security are caps, which are composed of a series of caplets with payoffs of the form (26), and give buyers the protection from the risk of higher interest rates. During the life of the contract, the cap is simply a caplet at each reset date, i.e., the discounted payoff function is given by

$$\sum_{i=1}^{n} D(T_i)M(r(T_i) - K)^+,$$

where T_i is the Ith reset date, and $D(T_i)$ is the price of a zero coupon bond with maturity T_i.

In our computational experiments, we keep the same parameter setting as in the example of caplets. The caps are reset quarterly. The simulation results are summarized in Table 8. The observation from the computational results is quite similar to the case of caplets, and further proves the efficiency of our stochastic approximation method.

Indirect Estimation and Deep in-the-Money Options

From the computational results, we notice that the efficiency of importance sampling decreases with the moneyness of the options. Thus, we are motivated to find more efficient estimators for deep in-the-money call options and propose indirect estimation (e.g., Law and Kelton 2000) using put-call parity:

$$C(S_0, r, \sigma, T, K) = S_0 - K\exp(-rT) + P(S_0, r, \sigma, T, K),$$

where $C(S_0, r, \sigma, T, K)$ is the price of call option and $P(S_0, r, \sigma, T, K)$ is the price of put option. Note that when one contract is deep in the money, the other is deep out of the money, so that the term $S_0 - Ke^{-rT}$ dominates, and we would expect much lower variance than simulating the in-the-money contract directly. We first test this idea on a vanilla European call option for the cases, T=1.0 year; r=0.05, 0.09; σ=0.2, 0.3; S_0 =100. The other parameter setting are the same as the case of partial average Asian call options.

The numerical results (based on 50,000 independent replications) in Table 9 (BS indicates the true Black-Scholes price) indicate that indirect estimation via put-call parity can effectively reduce the estimation variance in pricing in-the-money call options. When combined with importance sampling, we obtain tremendous variance reduction, nearly five orders of magnitude in the best case of our simulation experimental testbed.

To further test the idea of indirect estimation and importance sampling, we apply this idea to the deep in-the-money Asian call options considered earlier. The put-call parity for Asian options is given by

$$C(S_0, r, \sigma, T, K) = \left(E^Q\left[A(T)\right] - K\right)\exp(-rT) + P(S_0, r, \sigma, T, K).$$

In the case of discrete averaging,

$$E^Q\left[A(T)\right] = \frac{1}{N - N_0} S_0 \frac{e^{\frac{N_0}{N}rT} - e^{rT}}{1 - e^{\frac{1}{N}rT}}.$$

The simulation results are summarized in Table 10 and reinforce the conclusions from the previous example.

6 CONCLUSIONS

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 additional computational overhead 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. For deep in-the-money options, we propose an indirect estimation using put-call parity that can lead to further dramatic variance reduction, thus demonstrating that combining VRTs can be a very effective approach. This is also found in Vazquez-Abad and Dufresne (1998), where control variates are combined with importance sampling. Finally, the procedures presented here are widely applicable, and not just limited to processes based on Brownian motion or to any particular form of the security payoff function. However, the present version of the algorithm applies only to European options, so extension to American-style options is a natural avenue for future research, e.g., building on the work in Fu et al. (2000), and McDonald and Schroder (1998).

References

[1] Anderson, L., "Efficient Techniques for Simulation of Interest Rate Models Involving Non-linear Stochastic Differential Equations", working paper, General Re Financial Products, New York, NY, 1995.

[2] Black, F., Derman, E., and Toy, W., "A One-Factor Model of Interest Rates and Its Application to Treasury Bond Options", *Financial Analysts Journal*, 46, 33-39, 1990.

[3] Boyle, P.P., "Options: A Monte Carlo Approach," *Journal of Financial Economics*, **4**, 323-338, 1977.

[4] Boyle, P., Broadie, M., and Glasserman, P., "Monte Carlo Methods for Security Pricing", *Journal of Economic Dynamics and Control*, No. 21, 1257-1321, 1997.

[5] Cox, J., Ingersoll, J., and Ross, A., "A Theory of The Term Structure of Interest Rate", *Econometrica*, Vol. 53, No. 2, 385-408, 1985.

[6] Fu, M., "Convergence of the GI/G/1 Queue Using Infinitesimal Perturbation Analysis", *Journal of Optimization Theory and Applications*, 65, 149-160, 1990.

[7] Fu, M.C., Laprise, S.B., Madan, D.B., Su, Y., and Wu, R., "Pricing American Options: A Comparison of Monte Carlo Simulation Approaches," *Journal of Computational Finance*, 2000, in press.

[8] Fu, M., Madan, D., and Wang, T., "Pricing Continuous Time Asian Options: A Comparison of Monte Carlo and Laplace Transform Inversion Methods", *Journal of Computational Finance*, Vol. 2, No. 2, 49-74, 1999.

[9] Glasserman, P., Heidelberger, P., Shahabuddin, P., "Asymptotically Optimal Importance Sampling and Stratification for Pricing Path-Dependent Options," *Mathematical Finance*, **9**, 117-152, 1999a.

[10] Glasserman, P., Heidelberger, P., Shahabuddin, P., "Importance Sampling in the Heath-Jarrow-Morton Framework," *Journal of Derivatives*, 7, 32-50, 1999b.

[11] Harrison, J., and Kreps, D., "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory*, 20, 381-408, 1979.

[12] Heath, D., Jarrow, R., and Morton, A., "Bond Pricing and the Term Structure of Interest Rates", *Econometrica*, Vol. 60, No. 1, 77-105, 1992.

[13] Hull, J., and White, A., "The Pricing of Options on Assets with Stochastic Volatilities", *Journal of Finance*, 42, 281-300, 1987.

[14] Hull, J., and White, A., "The Use of the Control Variate Technique in Option Pricing", *Journal of Financial and Quantitative Analysis*, 23, 237-251, 1988.

[15] Law, A.M., and Kelton, W.D., *Simulation Modeling and Analysis*, 3rd edition, McGraw-Hill, 2000.

[16] McDonald, R.L. and Schroder, M.D., "A Parity Result for American Options," *Journal of Computational Finance*, 1, 5-13, 1998.

[17] Reider, R., "An Efficient Monte Carlo Techniques For Pricing Options", working paper, Wharton School, University of Pennsylvania, 1993.

[18] Su, Y., and Fu, M.C., "Importance Sampling in Derivative Securities Pricing," *Proceedings of the Winter Simulation Conference*, 587-596, 2000.

[19] Turnbull, S., and Wakeman, L., "A Quick Algorithm for Pricing European Average Options", *Journal of Financial and Quantitative Analysis*, 26, 377-389, 1991.

[20] Vazquez-Abad, F., and Dufresne, D., "Accelerated Simulation for Pricing Asian Options", *Proceedings of the Winter Simulation Conference*, 1493-1500, 1998.

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

Table 3: Asian call options on partial average: $S_0 = 100$, T=1.0yr, ϵ=0.0005, N_1=50, N_2=100, N_3=50000, $\lambda=r$ for Naive.

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

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, λ=r for Naive.

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

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, $\lambda=r$ for Naive.

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

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, $\lambda=r$ for Naive.

	IS via SA/IPA-Q				Naive		
	K	Price	C.I.	θ	Price	C.I.	VR
T=0.5yr	0.06	19.213	0.004	0.270	19.188	0.032	67
	0.07	9.606	0.009	0.500	9.580	0.032	14
	0.08	1.455	0.007	1.450	1.444	0.019	7.5
	0.09	0.006	0.0002	1.45	0.007	0.001	36
T=1yr	0.06	18.444	0.010	0.254	18.416	0.041	16
	0.07	9.251	0.013	0.487	9.221	0.041	9.2
	0.08	1.881	0.009	1.077	1.869	0.025	7.8
	0.09	0.055	0.001	1.234	0.055	0.004	19
T=2yrs	0.06	17.011	0.013	0.254	16.959	0.049	13
	0.07	8.616	0.018	0.408	8.594	0.047	7.2
	0.08	2.233	0.011	0.791	2.222	0.029	7.5
	0.09	0.185	0.002	1.005	0.183	0.008	18

Table 7: Caplets on CIR: r_0=0.08, η=0.016, k=0.2, σ=0.02, M=1000, θ=0 for Naive.

	IS via SA/IPA-Q				Naive		
	K	Price	C.I.	θ	Price	C.I.	VR
T=0.5yr	0.06	38.813	0.017	0.197	38.775	0.051	9.3
	0.07	19.409	0.019	0.392	19.422	0.051	7.3
	0.08	2.545	0.014	1.166	2.542	0.030	4.8
	0.09	0.007	0.0003	1.4355	0.007	0.001	16
T=1yr	0.06	76.053	0.066	0.093	75.997	0.117	3.2
	0.07	38.114	0.053	0.328	38.135	0.117	4.8
	0.08	6.163	0.033	0.837	6.150	0.068	4.2
	0.09	0.088	0.001	1.762	0.093	0.006	28
T=2yr	0.06	146.255	0.146	0.090	146.144	0.270	3.4
	0.07	73.620	0.131	0.258	73.722	0.263	4.0
	0.08	14.604	0.080	0.584	14.711	0.157	3.9
	0.09	0.621	0.007	1.143	0.640	0.027	14

Table 8: Caps on CIR: r_0=0.08, η=0.016, k=0.2, σ=0.02, M=1000, θ=0 for Naive.

	IS+INDIRECT					INDIRECT			Naive		BS	S_0-
	K	Price	C.I.	λ	VR	Price	C.I.	VR	Price	C.I.	Price	Ke^{-rT}
	70	33.540	0.001	-0.378	21267	33.541	0.009	395	33.608	0.175	33.540	33.414
r=.05	80	24.587	0.005	-0.348	1130	24.594	0.023	52	24.756	0.168	24.589	23.902
σ=.2	90	16.703	0.014	-0.258	117	16.682	0.047	11	16.691	0.153	16.700	14.390
	100	10.451	0.028	-0.191	22	10.480	0.076	2.9	10.407	0.130	10.448	4.877
	105	8.028	0.036	-0.164	11	7.988	0.091	1.6	8.070	0.116	8.042	0.121
	70	36.095	0.001	-0.361	63145	36.103	0.006	805	36.250	0.176	36.095	36.025
r=.09	80	27.314	0.003	-0.330	2682	27.320	0.018	92	27.276	0.171	27.314	26.886
σ=.2	90	19.306	0.010	-0.257	238	19.339	0.037	18	19.349	0.159	19.321	17.746
	100	12.696	0.022	-0.175	40	12.671	0.064	4.7	12.757	0.139	12.681	8.607
	109	8.150	0.035	-0.120	11	8.107	0.090	1.7	8.192	0.117	8.163	0.382
	70	34.398	0.007	-0.529	1315	34.410	0.031	69	34.444	0.258	34.395	33.414
r=.05	80	26.455	0.016	-0.425	222	26.508	0.055	20	26.283	0.243	26.463	23.902
σ=.3	90	19.728	0.029	-0.346	57	19.614	0.083	7.2	19.704	0.223	19.696	14.390
	100	14.269	0.045	-0.283	19	14.139	0.113	3.0	14.092	0.198	14.229	4.877
	105	11.992	0.053	-0.260	12	11.924	0.128	2.1	12.040	0.185	11.997	0.121
	70	36.738	0.005	-0.514	2324	36.729	0.026	103	36.860	0.260	36.740	36.024
r=.09	80	28.836	0.013	-0.405	371	28.874	0.047	29	28.675	0.248	28.837	26.886
σ=.3	90	21.940	0.024	-0.332	89	21.925	0.073	10	22.097	0.230	21.939	17.746
	100	16.217	0.039	-0.266	29	16.192	0.102	4.2	16.108	0.208	16.216	8.607
	109	12.107	0.053	-0.221	12	12.040	0.128	2.1	12.156	0.186	12.112	0.382

Table 9: European Calls, S_0=100, T=1yr, λ=r for Naive.

	IS+INDIRECT					INDIRECT			Naive		$E^Q[A(T)]$
	K	Price	C.I.	λ	VR	Price	C.I.	VR	Price	C.I.	-Ke^{-rT}
	70	33.075	0.001	-0.408	22555	33.074	0.007	557	33.186	0.165	32.988
r=.05	80	24.034	0.005	-0.344	889	24.045	0.021	57	24.061	0.158	23.475
σ=.2	90	16.025	0.015	-0.257	88	16.022	0.043	11	16.124	0.144	13.963
	100	9.74	0.030	-0.172	15	9.752	0.072	2.7	9.643	0.119	4.451
	70	35.309	0.001	-0.381	54823	35.311	0.005	956	35.25	0.164	35.259
r=.09	80	26.469	0.004	-0.317	2103	26.473	0.016	102	26.603	0.161	26.120
σ=.2	90	18.402	0.011	-0.218	173	18.419	0.035	18	18.308	0.149	16.981
	100	11.76	0.025	-0.154	27	11.706	0.061	4.4	11.672	0.128	7.841
	70	33.782	0.007	-0.500	1091	33.768	0.027	82	33.589	0.241	32.988
r=.05	80	25.698	0.017	-0.408	180	25.71	0.050	21	25.693	0.228	23.475
σ=.3	90	18.808	0.032	-0.331	43	18.79	0.078	7.0	18.832	0.206	13.963
	100	13.286	0.049	-0.252	14	13.166	0.108	2.8	13.334	0.180	4.451
	70	35.836	0.006	-0.504	1909	35.836	0.023	117	36.029	0.245	35.259
r=.09	80	27.829	0.014	-0.393	280	27.835	0.043	29	27.633	0.231	26.120
σ=.3	90	20.822	0.026	-0.317	66	20.859	0.069	10	20.828	0.214	16.981
	100	15.082	0.042	-0.235	20	15.114	0.097	3.8	14.901	0.189	7.841

Table 10: European Asian Calls, S_0=100, T=1yr, λ=r for Naive.