Noname manuscript No.
(will be inserted by the editor)

Optimal Importance Sampling with Explicit Formulas in Continuous Time

Paolo Guasoni · Scott Robertson

Received: date / Accepted: date

Abstract In the Black-Scholes model, consider the problem of selecting a change of drift which minimizes the variance of Monte Carlo estimators for prices of path-dependent options.

Employing Large Deviations techniques, the asymptotically optimal change of drift is identified as the solution to a one-dimensional variational problem, which may be reduced to the associated Euler-Lagrange differential equation.

Closed-form solutions for geometric and arithmetic average Asian options are provided.

Mathematics Subject Classification (2000) Primary 91B28 · Secondary 60F10, 65C05

Keywords Monte Carlo Methods · Variance Reduction · Importance Sampling · Large Deviations

1 Introduction

Monte Carlo simulation is the method of choice for pricing complex derivatives, such as path dependent options or contracts which rely upon several underlying assets. The main reason for the popularity of this method is ease of implementation, which only requires the ability to generate sample paths of the asset price and evaluating the corresponding derivative payoffs.

From a computational viewpoint, an option pricing problem is usually reduced to the evaluation of the expected payoff $E_P[G]$ under a certain risk-neutral probability P, which is unique in a complete market, and in general it may be chosen according to several optimality criteria. The usual Monte Carlo

The authors acknowledge the support of the National Science Foundation under grants DMS-0532390 (Guasoni) and DGE-0221680 (Robertson) at Boston University.

Boston University, Department of Mathematics and Statistics, 111 Cummington St, Boston, MA 02215 E-mail: guasoni@bu.edu, scottrob@math.bu.edu

estimate of $E_P[G]$ is obtained by the sample average $\bar{G}_n = \frac{1}{n}\sum_{i=1}^n G_i$ from a IID sample $(G_i)_{i=1}^n$ of the payoff. When G is square-integrable, the Central Limit Theorem implies that an asymptotic confidence interval for $E_P[G]$ is given by:

$$\left(\bar{G}_n - q_\alpha \frac{\sigma_n}{\sqrt{n}}, \bar{G}_n + q_\alpha \frac{\sigma_n}{\sqrt{n}}\right)$$

where $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^n (G_i - \bar{G}_n)^2$ is the sample variance of $(G_i)_{i=1}^n$, and q_α is the $(1-\alpha/2)$-quantile of the standard normal distribution.

In practice, many derivative contracts are designed to offer a payout in a specific event of interest to the buyer, and otherwise expire with no value, which means that an event with small probability accounts for most of the option price. In such a situation, this confidence interval can be very unreliable for two reasons. First, the large variance of G may require a prohibitively large number of simulations n for a prescribed accuracy. Second, even a relatively large sample is likely to miss rare but large payoffs, generating a low $\bar{G}_n$ combined with a low σ_n^2. This creates the perverse effect of a deceptively narrow confidence interval, which in most cases grossly misses the true expectation $E_P[G]$. To put it differently, if a small event accounts for a large fraction of the price, then it accounts for an even larger fraction of the variance, and the normal asymptotics are very inaccurate.

Importance sampling is a variance reduction method which addresses this problem by simultaneously changing the probability P and the payoff G as to retain the same expected value, while significantly reducing variance. In a nutshell, if Q is a probability equivalent to P and $H = G\frac{dP}{dQ}$, then:

$$E_P[G] = E_Q[H]$$

Thus, an optimal choice of Q should minimize the variance under the new probability Q of the new payoff H:

$$\mathrm{Var}_Q\left(G\frac{dP}{dQ}\right) = E_P\left[G^2\frac{dP}{dQ}\right] - E_P[G]^2 \tag{1.1}$$

As a matter of fact, $\frac{dQ}{dP} = \frac{G}{E_P[G]}$ achieves zero variance, but unfortunately $E_P[G]$ is the unknown in the first place, so in this generality the problem of selecting an optimal change of measure is ill-posed.

Thus, one has to minimize (1.1) over a subclass of equivalent probabilities, selected as to add little overhead to the simulation of the payoff. But even once such a restriction is made, the variance in (1.1) is very unlikely to have a closed-form solution unless $E_P[G]$ already has one. Hence, in practice one considers an asymptotic approximation of (1.1), and minimizes this quantity over the chosen class of probabilities.

A heuristic approach, which goes back to Siegmund (1976), is to consider a Large Deviations approximation of the variance in (1.1). Then the probability which minimizes the asymptotic variance has proved to be very successful in reducing variance in a broad range of applications. Nevertheless, Glasserman and Wang (1997) showed a number of examples where Large Deviations heuristics fail. The caveat, pointed out by Dupuis and Wang (2004), is that Large Deviations lead to several candidates for optimality, depending on the class of changes of probability considered. If the class is too small, then asymptotic optimality may fail. If it is too large, then the "optimal" change of measure may greatly increase the computational burden of simulation.

Here the main modeling dilemma is between the class of deterministic, or "open-loop", changes of measure and that of adaptive, or "feedback" ones. The former class has negligible impact on simulation speed, but its asymptotic optimality properties are unclear. For the latter class, asymptotic optimality holds in great generality (Dupuis and Wang 2004, 2005), but the overall speed increase is less clear.

To devise efficient importance sampling schemes, Glasserman, Heidelberger and Shahabuddin (1999) considered a discretization of the usual Black-Scholes model on a time grid $t_1, \ldots, t_n$, so that any option payoff can be approximated as a function of n real variables. Then they showed that the variance of (1.1) can be approximated with the Laplace's method for integrals, and studied the problem of minimizing this asymptotic variance through a *deterministic* change of drift. As already mentioned, the choice of a deterministic drift is motivated by its efficient implementation, which translates variance reduction into speed increase. On the other hand, it requires a careful evaluation of the asymptotic optimality properties.

From an applied viewpoint, the crucial question is the calculation of the optimal drift, identified by Glasserman et al. (1999) as the solution to a fixed point problem, which must be solved numerically through an iterative procedure.

This paper takes the approach of Glasserman et al. (1999) to a continuous-time setting, where the optimal deterministic drift in the Black-Scholes model is identified as the solution of a one-dimensional variational problem. In continuous time – and this is the major advantage of this approach – the variational problem reduces to the familiar Euler-Lagrange ODE. In the case of Asian options, considered also by Glasserman et al. (1999), the optimal change of drift even admits closed-form solutions.

From the mathematical viewpoint, the continuous time setting poses a more challenging environment. The Laplace' approximation of integrals, which roughly corresponds to a large deviations principle in $\mathbb{R}^n$, must be replaced by the sample-path Large Deviations result of Schilder (1966) combined with Varadhan's (1966) Integral Lemma, and a number of technical estimates are required to fully justify their application.

At an intuitive level, the main idea of the paper is briefly summarized in section 2 with a heuristic argument, which in spite of its audacity provides the right answer. Section 3 contains the rigorous formulation of the problem and the main results, leaving the proof to the Appendix. If the payoff G is a continuous function of the price path, and satisfies a mild growth condition, then there exists a "candidate optimal" drift which solves a one-dimensional variational problem. This candidate is indeed optimal if it satisfies a given equality, which may involve solving another one-dimensional variational problem.

Section 4 studies in detail the case of Asian Options. For a geometric average call option, the optimal change of drift is simply the parabola (4.5) below, while for an arithmetic average option, the optimal change of drift is given by the more complicated formula (4.12). Although the two expressions seem rather different, the optimal drifts for a geometric and arithmetic average options are in fact very close (Figure 4.1), as their similar payoffs would suggest.

The last observation highlights another important aspect of this methodology. It is developed in the simple Black-Scholes model, but it provides useful guidance also in presence of more complex features, such as stochastic volatility, which are akin to perturbations in the option payoff. Just as the optimal drift of an arithmetic average Asian option is very similar to its geometric counterpart, its optimal drift in a stochastic volatility model will conceivably be close to that in the Black-Scholes model, at least for current volatility parameters and for near expirations.

Some numerical examples follow, illustrating the effectiveness of the method. The best performance is achieved for out-of-the money options, but the variance reduction is six- to ten- fold even for at-the-money strikes.

2 Heuristics

Recall three classic "formulas". The first one is the heuristic representation (due to Feynman) of the Wiener measure on $C\left([0,T];\mathbb{R}\right)$ (for a detailed discussion, see Stroock (1993)):

$$P(dx) = c\exp\left(-\frac{1}{2}\int_0^T \dot{x}_t^2 dt\right)dx$$

The second formula is the Cameron-Martin representation of the change of measure induced by the translation $x \mapsto x+h$ on the Wiener space:

$$\frac{dQ^h}{dP}(x) = \exp\left(\int_0^T \dot{h}_t dx_t - \frac{1}{2}\int_0^T \dot{h}_t^2 dt\right) \tag{2.1}$$

Finally, recall the Laplace' method for the approximation of integrals:

$$\int_{-\infty}^{+\infty} \exp\left(f(x)\right)dx \approx \frac{1}{c}\exp\left(\max_x f(x)\right) \tag{2.2}$$

Combining these formulas to approximate the second moment in (1.1) yields:

$$E_P\left[G^2\frac{dP}{dQ^h}\right] \approx \exp\left(\max_x\left(2F\left(x\right)+\frac{1}{2}\int_0^T(\dot{x}_t-\dot{h}_t)^2dt-\int_0^T \dot{x}_t^2dt\right)\right) \tag{2.3}$$

where $F = \log G$. Thus, minimizing asymptotic variance leads to the problem:

$$\min_h \max_x \left(2F\,(x) + \frac{1}{2}\int_0^T (\dot{x}_t - \dot{h}_t)^2 dt - \int_0^T \dot{x}_t^2 dt \right) \tag{2.4}$$

Swapping the order of optimization, the minimax problem above reduces to:

$$\max_h \left(2F(h) - \int_0^T \dot{h}_t^2 dt \right) \tag{2.5}$$

which is a classical one-dimensional variational problem. Formally, the corresponding Euler-Lagrange becomes:

$$DF(h) + \ddot{h} = 0 \tag{2.6}$$

where DF is understood as a Frechet derivative.

Needless to say, the above argument is heuristic at best and a wild guess at worst. Indeed, it involves derivatives of non-differentiable Brownian paths, it applies Laplace asymptotics in infinite-dimensions, and assumes the validity of a minimax result. In spite of all these issues, up to some mild assumptions on F this characterization of the optimal drift is essentially correct, and the next section (along with the appendix) makes this result precise.

3 Main Result

Assume that the price of the underlying asset S_t follows the Black-Scholes model under some risk-neutral probability P:

$$S_t = S_0 e^{(r-\sigma^2/2)t+\sigma W_t} \tag{3.1}$$

where W_t is a standard Brownian Motion, r the interest rate and σ the volatility. Denote by:

$$\mathbb{W}_T \equiv \{x \in C([0,T], \mathbb{R}), x(0) = 0\}$$

the Wiener space of continuous functions on $[0,T]$ vanishing at zero. This space is endowed with the topology of uniform convergence and with the usual Wiener measure P, defined on the completion of the Borel σ-field $\mathcal{F}_T$, under which the coordinate process $W_t(x) = x_t$ is a standard Brownian Motion with respect to $(\mathcal{F}_t)_{t\in[0,T]}$, the usual augmentation of the natural filtration of W.

In this setting, a derivative contract can be identified with a functional G of the price path $(S_t)_{t\in[0,T]}$, but ease of presentation suggests describing it as a function of the shocks process $(W_t)_{t\in[0,T]}$.

Definition 3.1 A *payoff* is a non-negative functional $G : \mathbb{W}_T \mapsto \mathbb{R}_+$, continuous in the uniform topology.

Example 3.2 Consider the arithmetic average Asian option. Its payoff is given by $(\frac{1}{T}\int_0^T S_t dt - K)^+$, which corresponds to the functional:

$$G(x) = \left(\frac{1}{T} \int_0^T S_0 e^{(r-\sigma^2/2)t + \sigma x_t} dt - K \right)^+$$

Denote by $F = \log G$, taking values in $\mathbb{R} \cup \{-\infty\}$, and define by:

$$\mathbb{H}_T \equiv \left\{ h \in AC\,[0,T] : h(0) = 0, \int_0^T \dot{h}_t^2 dt < \infty \right\}$$

the Cameron-Martin space of absolutely continuous functions with square integrable derivative. For any deterministic drift $h \in \mathbb{H}_T$, consider the stochastic exponential:

$$\mathcal{E}\left(\int_0^{\cdot} \dot{h}_s dW_s \right)_t = \exp\left(\int_0^t \dot{h}_s dW_s - \frac{1}{2} \int_0^t \dot{h}_s^2 ds \right) \tag{3.2}$$

which induces an equivalent probability measure Q^h via the Radon-Nikodym derivative $dQ^h/dP = \mathcal{E}\left(\int_0^{\cdot} \dot{h}_t dW_t\right)_T$. Under Q^h the process $\tilde{W}_t \equiv W_t - h_t$ is a standard Brownian Motion by the classical Cameron-Martin theorem (or simply by Itô's formula). The objective function is the second moment in (1.1):

$$E_P\left[G^2 \frac{dP}{dQ^h} \right] = E_P\left[\exp\left(2F(W) - \int_0^t \dot{h}_s dW_s + \frac{1}{2} \int_0^t \dot{h}_s^2 ds \right) \right]$$

When Monte Carlo simulation is necessary to estimate $E_P[G]$, the above quantity is in general intractable. Instead, as in Glasserman et al. (1999), one considers the small-noise asymptotics:

$$L(h) = \limsup_{\varepsilon \downarrow 0} \varepsilon \log E_P \left[\exp\left(\frac{1}{\varepsilon} \left(2F\left(\sqrt{\varepsilon} W\right) - \int_0^T \sqrt{\varepsilon} \dot{h}_t dW_t + \frac{1}{2} \int_0^T \dot{h}_t^2 dt \right) \right) \right]$$

which correspond to approximating (1.1) with $e^{L(h)}$.

Definition 3.3 An *asymptotically optimal drift* is a solution to the problem:

$$\min_{h \in \mathbb{H}_T} L(h) \tag{3.3}$$

The goal is to find a deterministic expression for $L(h)$, which becomes suitable for optimization. This is possible under the following:

Assumption 3.4 $F : \mathbb{W}_T \mapsto \mathbb{R} \cup \{-\infty\}$ is continuous and satisfies:

$$F(x) \leq K_1 + K_2 \max_{t \in [0,T]} |x_t|^\alpha \tag{3.4}$$

for some constants $K_1, K_2 > 0$ and $\alpha \in (0, 2)$.

Remark 3.5 Condition (3.4) requires that, roughly speaking, the log payoff is subquadratic in the supremum of the log price. This is the case for virtually all options of practical interest.

Theorem 3.6 *Let F satisfy assumption 3.4. Then:*

i) if $h \in \mathbb{H}_T$, and $\dot{h}$ has finite variation, then:

$$L(h) = \sup_{x \in \mathbb{H}_T} \left(2F(x) + \frac{1}{2} \int_0^T (\dot{x}_t - \dot{h}_t)^2 dt - \int_0^T \dot{x}_t^2 dt \right) \tag{3.5}$$

ii) for all $h \in \mathbb{H}_T$, there exists maximizers to both (3.5) and (3.6) below:

$$\sup_{x \in \mathbb{H}_T} \left(2F(x) - \int_0^T \dot{x}_t^2 dt \right) \tag{3.6}$$

iii) if $\hat{h}$ is a solution to (3.6), then $\hat{h}$ is asymptotically optimal if:

$$L(\hat{h}) = 2F(\hat{h}) - \int_0^T \dot{\hat{h}}_t^2 dt \tag{3.7}$$

Furthermore, if (3.7) holds then $\hat{h}$ is the unique solution of (3.6).

Theorem 3.6 yields the following method to find an asymptotically optimal drift. First, find $\hat{h}$ by solving the Euler-Lagrange equation of (3.6). Then, $\hat{h}$ is an asymptotically optimal drift if it has a derivative with finite variation and satisfies the minimax condition (3.7). This is certainly the case when F is a concave functional, and then the standard minimax result for concave-convex functions applies. In general, one has to solve a new variational problem to evaluate $L(\hat{h})$, which also reduces to an Euler-Lagrange ODE.

Once $\hat{h}$ is found, (3.2) and the Cameron-Martin theorem imply that:

$$E_P[G] = E_{Q^{\hat{h}}} \left[\exp \left(F\left(\tilde{W} + \hat{h} \right) - \int_0^T \dot{\hat{h}}_t d\tilde{W}_t - \frac{1}{2} \int_0^T \dot{\hat{h}}_t^2 dt \right) \right] \tag{3.8}$$

where $\tilde{W}$ is a standard Brownian motion under $Q^{\hat{h}}$. Thus, in the new Monte Carlo simulation the drift of S_t changes from r to $r + \sigma \dot{\hat{h}}_t$, while the payoff is rescaled by the factor $\exp\left(- \int_0^T \dot{\hat{h}}_t d\tilde{W}_t - \frac{1}{2} \int_0^T \dot{\hat{h}}_t^2 dt \right)$.

Remark 3.7 In this paper, (3.7) embodies the delicate issue of whether asymptotic optimality holds, and in general the answer depends on the functional considered. In the case of the arithmetic average Asian option, the value $L(\hat{h})$ can only be evaluated numerically, and (3.7) can be established with several significant digits, but not with absolute certainty.

On the contrary, a violation of (3.7) immediately detects a case where asymptotic optimality fails, because a strict inequality, unlike an equality, can be established numerically with sufficient accuracy.

Dupuis and Wang (2004) show that, in a discrete time setting with IID returns, asymptotic optimality holds under very mild conditions for adaptive, i.e. *path-dependent* drifts. The main problem with such drifts is the substantial overhead required by their implementation, which requires the recursive calculation of the change of drift for each simulated path. Thus, variance reduction gains must be split between the decreased number of simulations, and the increased time for each simulation.

By contrast, a deterministic drift requires virtually no overhead, as the same drift is added to all paths, with no extra calculations involved. Considering the substantial variance reduction obtained in the next section with a deterministic drift, the additional performance gain from a path-dependent drift appears unclear.

4 Asian Options

This section employs Theorem 3.6 to find explicit formulas for the asymptotically optimal changes of drift, for geometric and arithmetic average Asian options. As mentioned in the introduction, the relevance of these examples goes beyond the specific model considered, as the asymptotically optimal changes of drift derived under the Black-Scholes assumptions can be very effective (although not optimal anymore) even in more complex models.

For example, the geometric average option considered in the first example allows an explicit solution (Kemna and Vorst 1990) in the Black-Scholes model, so Monte-Carlo is not necessary. However, the introduction of a minimal imperfection such as a dividend may require simulation, which is much more efficient when performed under the optimal drift for the option without dividend.

Similarly, for the arithmetic average option in the second example one may resort to alternative numerical methods, (see for example Geman and Yor (1992), Dufresne (2001), Lyasoff (2006) and the references therein). Again, the latter formulas are not valid under model perturbations, while Monte-Carlo simulation may easily accommodate them, and a change of drift greatly increases its efficiency.

Example 4.1 (Geometric average) Denoting by S_t the asset price at time t, and by K the strike price, the payoff of a Geometric Average Asian Option is:

$$\left(e^{\frac{1}{T}\int_0^T \log S_t dt} - K\right)^+$$

Letting $a = \sigma/T$ and $c = \frac{K}{S_0}\exp\left(-(r - \frac{\sigma^2}{2})\frac{T}{2}\right)$, rewrite this payoff as:

$$G(x) = \frac{K}{c}\left(e^{a\int_0^T x_t dt} - c\right)^+ \tag{4.1}$$

To check Assumption 3.4, note that $F(x) = -\infty$ on the set $G(x) = 0$, while on the set $G(x) > 0$, it is sufficient to choose $\alpha = 1, K_1 = \log\frac{K}{c}, K_2 = aT$.

Also, (3.7) is certainly satisfied since F is concave. Now, rewrite (3.6) as:

$$\max_{x\in\mathbb{H}_T}\left(2\log\left(e^{a\int_0^T x_t dt}-c\right)-\int_0^T \dot{x}_t^2 dt\right) \tag{4.2}$$

The corresponding Euler-Lagrange equation is:

$$\ddot{x}_t=-\alpha \tag{4.3}$$

where

$$\alpha=a\frac{\exp\left(\int_0^T x_t dt\right)}{\exp\left(\int_0^T x_t dt\right)-c} \tag{4.4}$$

Hence, all solutions are of the form:

$$x_t=-\frac{\alpha}{2}t^2+\gamma t \tag{4.5}$$

and therefore belong to $\tilde{\mathbb{H}}_T$. The maximizing solution is found by choosing α and γ that simultaneously solve (4.4) and (4.2). More precisely, substituting (4.5) into (4.4) yields

$$\gamma\left(\alpha\right)=\frac{aT^3\alpha-6\log(\frac{\alpha-a}{c\alpha})}{3aT^2} \tag{4.6}$$

and for this value of $\gamma(\alpha)$, (4.2) is solved by maximizing over $\alpha>a$. The optimal $\hat{\alpha}$ is unique by strict concavity, and is found implicitly via the equation:

$$a\hat{\alpha}T^3+3\log\left(\frac{\hat{\alpha}-a}{c\hat{\alpha}}\right)=0 \tag{4.7}$$

This $\hat{\alpha}$ satisfies $\gamma\left(\hat{\alpha}\right)=\hat{\alpha}T$ and therefore $\hat{x}_t=-\frac{\hat{\alpha}}{2}t^2+\hat{\alpha}Tt$.

Example 4.2 (Arithmetic average) The payoff is now $\left(\frac{1}{T}\int_0^T S_t dt-K\right)^+$, which leads to the functional:

$$G\left(x\right)=d\left(\int_0^T\exp\left(ax_t+bt\right)dt-c\right)^+ \tag{4.8}$$

with $a=\sigma$, $b=r-\frac{1}{2}\sigma^2$, $c=K\frac{T}{S_0}$ and $d=\frac{S_0}{T}$. Furthermore, Assumption 3.4 holds with $\alpha=1$, $K_1=\log d+\log\left(\frac{\exp(bT)-1}{b}\right)$ and $K_2=a$. In this case, the variational problem (3.6) becomes:

$$\max_{x\in\mathbb{H}_T}\left(2\log d+2\log\left(\int_0^T\exp\left(ax_t+bt\right)dt-c\right)-\int_0^T\dot{x}_t^2 dt\right) \tag{4.9}$$

and the Euler-Lagrange equation is:

$$\ddot{x}_t=\lambda\exp\left(ax_t+bt\right) \tag{4.10}$$

where

$$\lambda = -\frac{a}{\int_0^T \exp\left(ax_t + bt\right) dt - c} \tag{4.11}$$

Equation (4.10) admits the family of solutions:

$$x_t = \frac{\alpha - b}{a} t - \frac{2}{a} \log \left(\frac{\exp\left(\alpha t\right) + \gamma}{1 + \gamma} \right) \tag{4.12}$$

Substituting (4.12) into (4.11), the parameters (α, γ) are linked to λ by the condition:

$$\lambda = -\frac{2\gamma\alpha^2}{a\left(1+\gamma\right)^2} \tag{4.13}$$

Eliminating λ from (4.13) and (4.11) yields:

$$\frac{2\gamma\alpha^2}{a\left(1+\gamma\right)^2} = \frac{a\alpha\left(\exp\left(\alpha T\right) + \gamma\right)}{\left(1+\gamma\right)\left(\exp\left(\alpha t\right) - 1\right) - c\alpha\left(\exp\left(\alpha T\right) + \gamma\right)} \tag{4.14}$$

For a fixed α, (4.14) defines a cubic polynomial in γ which yields an explicit solution. Denote by $\hat{\alpha}$ the maximizer, and by $\hat{\gamma}$ and $\hat{x}$ the corresponding parameter and solution.

To check (3.7), maximize the functional:

$$2 \log d + 2 \log \left(\int_0^T \exp\left(ax_t + bt\right) dt - c \right) + \frac{1}{2} \int_0^T (\dot{x}_t - \dot{\hat{x}}_t)^2 dt - \int_0^T \dot{x}_t^2 dt \tag{4.15}$$

over $x \in \mathbb{H}_T$. Here the Euler Lagrange ODE is

$$\ddot{x}_t = 2\lambda \exp\left(ax_t + bt\right) - \ddot{\hat{x}}_t \tag{4.16}$$

where, λ is defined as in (4.11).

This ODE does not admit an explicit solution, except in the trivial case $\lambda = 0$. However, a numerical integration of the Euler-Lagrange equation shows (3.7) holds with several significant digits.

5 Numerical Results

Consider an arithmetic average Asian option with parameters $T = 1, r = 5\%, \sigma = 25\%, S_0 = 50, K = 70$, as in Glasserman et al. (1999), and compare the results of the simulation under three different drifts: the risk-neutral drift, corresponding to usual Monte Carlo simulation, and the asymptotically optimal drifts for an Asian call option of arithmetic and geometric average type. The price paths in absence of random shocks are plotted in Figure 4.1, and show that the arithmetic and geometric drifts are indeed very similar, although the closed form expression is much simpler (in fact, quadratic) in the geometric case.

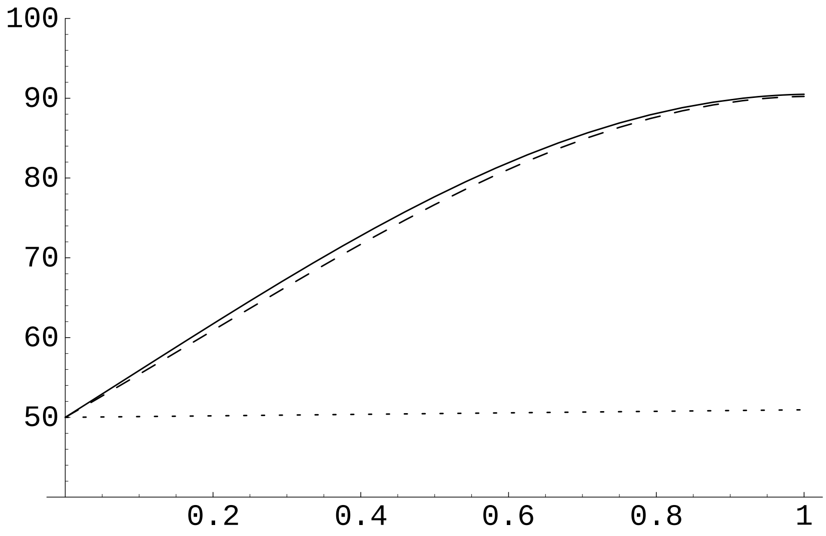

Fig. 4.1 The plots represent the time-evolution (in years) of an asset price (in dollars), in absence of random shocks, under the Black-Scholes model with initial price 50, interest rate 5%, volatility 25%, with risk-neutral drift (dotted line), and asymptotically optimal drifts for an Asian call option with strike 70 of geometric average (solid line) and arithmetic average (dashed line) type.

Table 5.1 shows the results of a typical simulation for an arithmetic average Asian option, for different choices of drift and the sample size. Note that the approximate theoretical price obtained by both the Turnbull and Wakeman (1991) and Levy (1992) formulas is 5.06, and heavily underprices the option. The reason is that both approximations are derived by at-the-money Taylor expansions, which become inaccurate for out-of-the-money strikes.

Since the option is of arithmetic average type, the lowest standard error is obtained with the arithmetic drift, immediately followed by the geometric. Even for small sample sizes, the variance obtained with the optimal arithmetic drift is only slightly lower than that obtained from the geometric drift, which offers a much simpler alternative.

Table 5.2 compares the performance, in terms of variance reduction, of the two drifts across a range of strikes and volatilities. The performance gap increases with the strike, and decreases with volatility. These observations have a common explanation in terms of moneyness, since both a larger strike and a lower volatility cause the option to become more out-of-the-money, and then the role of the drift in reshaping the payoff distribution becomes more critical.

6 Conclusion

Importance sampling can greatly improve the performance of Monte Carlo methods in option pricing, but its success hinges on a change of probability

Sample Size	Arithmetic	Geometric	Risk-Neutral
100,000	6.16 (0.021)	6.21 (0.022)	6.0 (0.22)
20,000	6.26 (0.048)	6.33 (0.050)	5.9 (0.44)
5,000	6.2 (0.095)	6.2 (0.098)	6. (1.1)

Table 5.1 Monte Carlo estimators of an arithmetic Asian option using different changes of drift. Prices are in cents. Parameter values are $T = 1, r = 5\%, S_0 = 50, K = 70, \sigma = 25\%$. The arithmetic and geometric drifts are given by (4.9) and (4.2) respectively. Simulations are performed with a time-increment of 1/252, corresponding to one business day.

Volatility	Strike	Price	(Std. Err.)	Variance Ratios	
				Arithmetic	Geometric
10%	50	191.8	(0.086)	6.56	6.53
	60	0.397	(0.00050)	330	320
15%	50	247.1	(0.12)	7.11	7.04
	60	7.51	(0.0076)	51	50
20%	50	304.0	(0.15)	7.59	7.50
	60	28.00	(0.025)	26.5	26.0
	70	1.063	(0.0013)	310	280
25%	50	361.3	(0.18)	8.03	7.92
	60	60.35	(0.049)	20.1	19.7
	70	6.17	(0.0067)	101	94
30%	50	419.3	(0.22)	8.47	8.34
	60	101.1	(0.078)	17.3	16.9
	70	18.17	(0.018)	56	53
	80	2.75	(0.0033)	260	230
35%	50	477.2	(0.25)	8.96	8.77
	60	147.3	(0.11)	16.0	15.6
	70	38.03	(0.035)	39.5	37.3
	80	8.84	(0.0097)	118	106
	90	1.95	(0.0024)	400	330
40%	50	535.1	(0.28)	9.49	9.27
	60	197.9	(0.14)	15.2	14.8
	70	65.3	(0.057)	31.7	30.1
	80	20.18	(0.021)	73	66
	90	6.09	(0.0070)	210	170

Table 5.2 Variance Reduction Ratios across strikes and volatilities. Parameter values are $T = 1, r = 5\%, S_0 = 50$ and each simulation is performed with 1,000,000 paths. Variance ratios are obtained dividing the variance of the risk-neutral sample by the variance of the geometric and arithmetic samples respectively. Reported option prices and standard errors are from the arithmetic sample, and are quoted in cents. Only significant digits are reported.

(or equivalently, of drift) which is both effective in reducing variance, and parsimonious in computational effort.

Although simulation ultimately takes place in a discrete-time setting as in Glasserman et al. (1999), this paper employs the continuous-time formulation to identify the asymptotically optimal change of drift as the solution of a variational problem. Furthermore, closed-form solutions are derived for Asian options.

The role of the optimal change of drift in reducing variance is twofold: first, it serves as a Girsanov transformation to perform importance sampling. Second, as observed by Glasserman et al. (1999), it provides a very effective "direction" for stratification algorithms, which magnify the performance increase.

The optimal drifts are derived under the Black-Scholes assumptions, but can be employed effectively also in more complex models, where explicit formulas for optimal drifts may not be available.

Finally, this paper considers derivatives on a single asset, but the same methodology could be performed with several assets, studying the problem:

$$\sup_{h\in\mathbb{H}_T^n}\left(2F\left(h\right)-|\dot{h}|^2\right)$$

which leads to a system of Euler-Lagrange ODEs formally equivalent to (2.6).

A Appendix

The proof of Theorem 3.6 is divided into several lemmas. the first one shows the existence of solutions to problems (3.5) and (3.6), using a standard variational argument.

Lemma A.1 *Let F satisfy Assumption 3.4. Then for any $h\in\mathbb{H}_T$ and $M>0$ there exists a maximizer for the problem:*

$$\max_{x\in\mathbb{H}_T}\left(2F\left(x\right)+M\int_0^T(\dot{x}_t-\dot{h}_t)^2dt-2M\int_0^T\dot{x}_t^2dt+(1-2M)\int_0^T\dot{h}_t^2dt\right) \tag{A.1}$$

Proof Recall that if $\dot{g_n}\to\dot{g}$ weakly in $L^2\left[0,T\right]$, then $g_n\to g$ uniformly in $[0,T]$. Since F is continuous in the uniform norm, it follows that it is also weakly continuous. Let $M>0$ and fix $h\in\mathbb{H}_T$. Rewrite (A.1) as

$$\max_{x\in\mathbb{H}_T}\left(2F\left(x\right)-M\|h+x\|^2_{\mathbb{H}_T}+\|h\|^2_{\mathbb{H}_T}\right)$$

As a function of x, $M\|h+x\|^2_{\mathbb{H}_T}$ is convex and finite, hence norm continuous. Thus, it is also weakly lower semi-continuous. Since F is weakly continuous, then the function $x\mapsto 2F(x)-M\|x+h\|^2_{\mathbb{H}_T}+\|h\|^2_{\mathbb{H}_T}$ is weakly upper semi-continuous. Assumption 3.4 implies that

$$\begin{aligned}2F(x)-M\|x+h\|^2_{\mathbb{H}_T}+\|h\|^2_{\mathbb{H}_T}&\le 2K_1+2K_2\|x\|^\alpha_\infty-M\|x+h\|^2_{\mathbb{H}_T}+\|h\|^2_{\mathbb{H}_T}\\&\le 2K_1+2K_2T^{\alpha/2}\|x\|^\alpha_{\mathbb{H}_T}-M\|x+h\|^2_{\mathbb{H}_T}+\|h\|^2_{\mathbb{H}_T}\end{aligned}$$

Since $\alpha<2$, the coercivity property follows:

$$\lim_{\|x\|_{\mathbb{H}_T}\to\infty}\left(2F(x)-M\|h+x\|^2_{\mathbb{H}_T}+\|h\|^2_{\mathbb{H}_T}\right)=-\infty$$

and the existence of a maximizer follows by upper semi-continuity.

The remaining part of the Proof of Theorem 3.6 requires some preliminaries on the theory of Large Deviations. Here are summarized some basic definitions, mainly with the purpose of introducing notation, while the reader is referred to the monographs of Deuschel and Stroock (1989), Dupuis and Ellis (1997) and Dembo and Zeitouni (1998) for an extensive treatment of topic.

Definition A.2 Let $(X, \mathcal{B})$ be a metric space with its Borel σ-algebra, and $I : X \mapsto [0, +\infty]$ a lower semi-continuous function. A family of measures $(\mu_\varepsilon)_{\varepsilon \in (0,\delta)}$ satisfies a *large deviation principle* with *good rate function I* if:

i) $\{x \in X : I(x) \leq \alpha\}$ is compact for all $\alpha \in \mathbb{R}$;

ii) For all sets $A \in \mathcal{B}$:

$$-\inf_{x \in A^\circ} I(x) \leq \liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(A) \leq \limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(A) \leq -\inf_{x \in \bar{A}} I(x)$$

On the Wiener space the following result holds. It is originally due to Schilder (1966), while modern proofs can be found in Dembo and Zeitouni (1998, Theorem 5.2.3) and Deuschel and Stroock (1989, Theorem 1.3.27)

Theorem A.3 (Schilder) *Let $X = \mathbb{W}_T$ and μ_ε be the probability on $\mathbb{W}_T$ induced by the process $\sqrt{\varepsilon}W$, where W is a standard Brownian Motion. Then $(\mu_\varepsilon)_{\varepsilon \in (0,\delta)}$ satisfies a large deviations principle (Definition A.2) with good rate function:*

$$I(x) = \begin{cases} \frac{1}{2}\int_0^T \dot{x}_t^2 dt & \text{if } x \in \mathbb{H}_T \\ +\infty & \text{if } x \in \mathbb{W}_T \setminus \mathbb{H}_T \end{cases} \tag{A.2}$$

The next result, known as Varadhan's Lemma, is the extension of the Laplace approximation for integrals to a general (infinite-dimensional) setting. A proof can be found in Dembo and Zeitouni (1998, Theorem 4.3.1).

Lemma A.4 (Varadhan) *Let $(Z_\varepsilon)_{\varepsilon \in (0,\delta)}$ be a family of X-valued random variables, whose laws $\mu_\varepsilon = Z_\varepsilon(P)$ satisfy a large deviations principle (Definition A.2) with rate function I. If $H : X \mapsto \mathbb{R}$ is a continuous function which satisfies:*

$$\limsup_{\varepsilon \to 0} \varepsilon \log E\left[\exp\left(\frac{\gamma}{\varepsilon} H(Z_\varepsilon)\right)\right] < \infty \tag{A.3}$$

for some $\gamma > 1$, then:

$$\lim_{\varepsilon \to 0} \varepsilon \log E\left[\exp\left(\frac{1}{\varepsilon} H(Z_\varepsilon)\right)\right] = \sup_{x \in X}\left(H(x) - I(x)\right) \tag{A.4}$$

The present setting requires a slight generalization of this result, as to allow $H : X \mapsto [-\infty, \infty)$ rather than $H : X \mapsto \mathbb{R}$. The following Lemma (cf. Glasserman et al. (1999, Lemma 2.1)) provides the necessary extension.

Lemma A.5 *Let $H : X \mapsto [-\infty, \infty)$. Under the assumptions of Lemma A.4, the following holds for any $A \in \mathcal{B}$:*

$$\begin{aligned}\sup_{x \in A^\circ}\left(H(x) - I(x)\right) &\leq \liminf_{\varepsilon \to 0} \varepsilon \log\left(\int_{A^\circ} \exp\left(\frac{1}{\varepsilon} H(Z_\varepsilon)\right) d\mu_\varepsilon\right) \\ &\leq \limsup_{\varepsilon \to 0} \varepsilon \log\left(\int_{\bar{A}} \exp\left(\frac{1}{\varepsilon} H(Z_\varepsilon)\right) d\mu_\varepsilon\right) \leq \sup_{x \in \bar{A}}(H(x) - I(x))\end{aligned}$$

Proof The second inequality is trivial, while the first one follows directly from Dembo and Zeitouni (1998, Lemma 4.3.4), fixing $x \in A^\circ$ instead of $x \in X$. For the third inequality, note that if $F \equiv -\infty$ the result holds trivially. Assuming F is not identically $-\infty$, let C be a closed subset of X. For $M > 0$, consider the set $C_M = C \bigcap \{F(x) \geq -M\}$ which is closed by the continuity of F. Thus, one has that:

$$\begin{aligned}\int_C \exp\left(\frac{1}{\varepsilon}F\left(Z_\varepsilon\right)\right)d\mu_\varepsilon &= \int_{C_M} \exp\left(\frac{1}{\varepsilon}F\left(Z_\varepsilon\right)\right)d\mu_\varepsilon + \int_{C\setminus C_M} \exp\left(\frac{1}{\varepsilon}F\left(Z_\varepsilon\right)\right)d\mu_\varepsilon \\ &\leq \int_{C_M} \exp\left(\frac{1}{\varepsilon}F(Z_\varepsilon)\right)d\mu_\varepsilon + \exp\left(-\frac{M}{\varepsilon}\right)\mu_\varepsilon\left(C\setminus C_M\right)\end{aligned}$$

Since $(\mu_\varepsilon)_{\varepsilon\in(0,\delta)}$ satisfy the LDP with good rate function I:

$$\limsup_{\varepsilon\to 0} \varepsilon \log\left(\exp\left(-\frac{M}{\varepsilon}\right)\mu_\varepsilon\left(C\setminus C_M\right)\right) \leq -M - \inf_{x\in\overline{C\setminus C_M}} I\left(x\right)$$

Using Varadhan's Lemma on $F\mathbf{1}_{C_M}$ (Dembo and Zeitouni 1998, Exercise 4.3.11)

$$\limsup_{\varepsilon\to 0} \varepsilon \log\left(\int_{C_M} \exp\left(\frac{1}{\varepsilon}F(Z_\varepsilon)\right)d\mu_\varepsilon\right) \leq \sup_{x\in C_M}\left(F\left(x\right) - I\left(x\right)\right)$$

and hence (cfr. Dembo and Zeitouni (1998, Lemma 1.2.15)):

$$\begin{aligned}&\limsup_{\varepsilon\to 0} \varepsilon \log\left(\int_{C_M} \exp\left(\frac{1}{\varepsilon}F(Z_\varepsilon)\right)d\mu_\varepsilon + \left(\exp\left(-\frac{M}{\varepsilon}\right)\mu_\varepsilon\left(C\setminus C_M\right)\right)\right) \\ &\leq \max\left(\sup_{x\in C_M}\left(F\left(x\right) - I\left(x\right)\right), -M - \inf_{x\in\overline{C\setminus C_M}} I\left(x\right)\right) \\ &\leq \max\left(\sup_{x\in C}\left(F\left(x\right) - I\left(x\right)\right), -M\right)\end{aligned}$$

The claim follows as $M \to \infty$.

Lemma A.6 *Let F satisfy Assumption 3.4, and define $F_h : \mathbb{W} \mapsto \mathbb{R}$ as:*

$$F_h(x) = 2F\left(x\right) - \int_0^T \dot{h}_t dx_t + \frac{1}{2}\int_0^T \dot{h}_t^2 dt$$

Then F_h is well-defined, norm continuous and satisfies (A.3) *for any $h \in \tilde{\mathbb{H}}_T$ and $\gamma > 1$.*

Proof Since F is continuous, the continuity of F_h will follow from the continuity of $x \mapsto \int_0^T \dot{h} dx_t$. Since $\dot{h}$ has finite variation on $[0, T]$ for each $h \in \tilde{\mathbb{H}}_T$, the integral $\int_0^T \dot{h} dx_t$ is defined path-wise in the Stieltjes sense. For any $f \in \mathbb{W}_T$, integration by parts and the continuity of f imply that:

$$\left|\int_0^T \dot{h} df_t\right| = \left|\dot{h}(T)f(T) - \int_0^T f_t d\dot{h}_t\right| \leq \|f\|_{\mathbb{W}_T}\| \operatorname{Var}(\dot{h})$$

where $\mathrm{Var}(\dot{h})$ denotes the total variation of $\dot{h}$. Thus continuity follows by the finite variation assumption. To check the integrability condition (A.3), apply the Cauchy-Schwarz inequality to see that:

$$\varepsilon \log E_P \left[\exp\left(\frac{\gamma}{\varepsilon} \left(2F\left(\sqrt{\varepsilon} W\right) - \int_0^T \dot{h}_t d\left(\sqrt{\varepsilon} W\right)_t + \frac{1}{2}\int_0^T \dot{h}_t^2 dt \right)\right)\right]$$
$$\leq \frac{\gamma}{2}\int_0^T \dot{h}_t^2 dt + \frac{\varepsilon}{2}\log E_P\left[\exp\left(-\frac{2\gamma}{\sqrt{\varepsilon}}\int_0^T \dot{h}_t dW_t\right)\right] + \frac{\varepsilon}{2}\log E_P\left[\exp\left(\frac{4\gamma}{\varepsilon}F\left(\sqrt{\varepsilon}W\right)\right)\right] \tag{A.5}$$

The first term is finite. For the second, observe that $\int_0^T \dot{h}_t dW_t \sim N\left(0, \int_0^T \dot{h}_t^2 dt\right)$, whence:

$$\limsup_{\varepsilon\to 0} \frac{\varepsilon}{2}\log E_P\left[\exp\left(-\frac{2\gamma}{\sqrt{\varepsilon}}\int_0^T \dot{h}_t dW_t\right)\right] = \gamma^2 \int_0^T \dot{h}_t^2 dt < \infty$$

It remains to consider the last term in (A.5). Assumption 3.4 implies that:

$$\frac{\varepsilon}{2}\log E_P\left[\exp\left(\frac{4\gamma}{\varepsilon}F\left(\sqrt{\varepsilon}x\right)\right)\right] \leq 2\gamma K_1 + \frac{\varepsilon}{2}\log E\left[\exp\left(\frac{4\gamma K_2}{\varepsilon^{1-\alpha/2}}\left(\sup_{0\leq t\leq T}|W(t)|\right)^{\alpha}\right)\right]$$

and one has to check that the last term is finite. To see this, observe that:

$$E_P\left[\exp\left(\frac{4\gamma K_2}{\varepsilon^{1-\alpha/2}}\left(\sup_{0\leq t\leq T}|W(t)|\right)^{\alpha}\right)\right] \leq 2E_P\left[\exp\left(\frac{4\gamma K_2}{\varepsilon^{1-\alpha/2}}\left(\sup_{0\leq t\leq T}x(t)\right)^{\alpha}\right)\right]$$
$$\leq 4\sqrt{\frac{2}{\pi T}}\int_0^\infty \exp\left(\frac{4\gamma K_2}{\varepsilon^{1-\alpha/2}}b^\alpha - \frac{1}{2T}b^2\right) db$$

where the first inequality follows from the formula $E_P[X] = \int_0^\infty P(X \geq b)\, db$, combined with the elementary estimate:

$$P\left(\sup_{0\leq t\leq T}|W_t| \geq b\right) \leq 2P\left(\sup_{0\leq t\leq T} W_t \geq b\right)$$

The second inequality follows from the classical distribution:

$$P\left(\sup_{0\leq t\leq T} W_t \in db\right) = \sqrt{\frac{2}{\pi T}}\exp\left(-\frac{b^2}{2T}\right).$$

Applying Lemma A.7 below, for $A = \frac{4\gamma K_2}{\varepsilon^{1-\alpha/2}}$, $B = \frac{1}{2T}$ yields:

$$\int_0^\infty \exp\left(\frac{4\gamma}{\varepsilon^{1-\alpha/2}}K_2 b^\alpha - \frac{1}{2T}b^2\right) db$$
$$\leq \exp\left(\frac{4\gamma K_2}{\varepsilon^{1-\alpha/2}}\left(4\gamma K_2 \alpha T\right)^{\frac{\alpha}{2-\alpha}}\varepsilon^{-\alpha/2} + \frac{1}{2T}\left(4\gamma K_2\alpha T\right)^{\frac{2}{2-\alpha}}\varepsilon^{-1}\right)$$
$$\times\left(\left(4\gamma K_2\alpha T\right)^{\frac{1}{2-\alpha}}\varepsilon^{-1/2} + \sqrt{\frac{2\pi}{\min\left(\frac{1}{T}, \frac{1}{T}(2-\alpha)\right)}}\right)$$

which, after letting $N = 4\gamma K_2 \alpha T$ and $M = \min\left(\frac{1}{T}, \frac{1}{T}(2-\alpha)\right)$ reduces to

$$\exp\left(\frac{1}{\varepsilon}\frac{1}{T}N^{\frac{2}{2-\alpha}}\left(\frac{1}{\alpha}-\frac{1}{2}\right)\right)\left(N^{\frac{1}{2-\alpha}}\varepsilon^{-1/2}+\sqrt{\frac{2\pi}{M}}\right)$$

Thus:

$$\limsup_{\varepsilon\to 0}\varepsilon\log\left(2\sqrt{\frac{2}{\pi T}}\int_0^\infty \exp\left(\frac{4\gamma}{\varepsilon^{1-\alpha/2}}K_2 b^\alpha - \frac{1}{2T}b^2\right)db\right)^{1/2}$$
$$\leq \frac{1}{2}\frac{1}{T}N^{\frac{2}{2-\alpha}}\left(\frac{1}{\alpha}-\frac{1}{2}\right) < \infty$$

which proves the claim.

Lemma A.7 *Let $A, B > 0$, $\alpha \in (0,2)$ and set $\underline{b} = \left(\frac{\alpha A}{2B}\right)^{\frac{1}{2-\alpha}}$. Then the function $f(b) = Ab^\alpha - Bb^2$ satisfies the estimate*

$$\int_0^\infty \exp(f(b))\,db \leq \exp\left(A\underline{b}^\alpha - B\underline{b}^2\right)\left(\underline{b}+\sqrt{\frac{2\pi}{\min(2B, 2B(2-\alpha))}}\right) \quad \text{(A.6)}$$

Proof Note that:

$$f'(b) = \alpha A b^{\alpha-1} - 2Bb$$
$$f''(b) = \alpha(\alpha-1)Ab^{\alpha-2} - 2B$$
$$f'''(b) = \alpha(\alpha-1)(\alpha-2)Ab^{\alpha-3}$$

Let $\underline{b}$ be as given in the statement of the lemma and note that $f'(\underline{b}) = 0$ and for $b < \underline{b}, f'(b) > 0$ and for $b > \underline{b}, f'(b) < 0$. Thus, $\underline{b}$ is the unique global maximum of $f(b)$. Upon inspecting the derivatives of f, it follows that $f''(b) < -2B < 0$ for $\alpha \leq 1$, and $f'''(b) < 0$ for $1 < \alpha < 2$. This implies thatfor $b > \underline{b}$

$$f''(b) < f''(\underline{b}) = -2B(2-\alpha)$$

and taking the Taylor expansion of f around $\underline{b}$:

$$f(b) = A\underline{b}^\alpha - B\underline{b}^2 + \frac{1}{2}(b-\underline{b})^2 f''(\xi(b))$$

for some $\xi(b) \in [b, \underline{b}]$ if $b < \underline{b}$ and $\xi(b) \in [\underline{b}, b]$ if $b > \underline{b}$. Note that for $b > \underline{b}, f''(\xi(b)) < \max(-2B, -2B(2-\alpha))$. Thus,

$$\int_0^\infty \exp\left(Ab^\alpha - Bb^2\right)db = \int_0^{\underline{b}} \exp\left(Ab^\alpha - Bb^2\right)db + \int_{\underline{b}}^\infty \exp\left(Ab^\alpha - Bb^2\right)db$$
$$\leq \exp\left(A\underline{b}^\alpha - B\underline{b}^2\right)\left(\underline{b}+\int_{\underline{b}}^\infty \exp\left(-\frac{1}{2}(b-\underline{b})^2\min(2B, 2B(2-\alpha))\right)db\right)$$
$$\leq \exp\left(A\underline{b}^\alpha - B\underline{b}^2\right)\left(\underline{b}+\int_{-\infty}^\infty \exp\left(-\frac{(b-\underline{b})^2}{2(1/\min(2B, 2B(2-\alpha)))}\right)db\right)$$
$$= \exp\left(A\underline{b}^\alpha - B\underline{b}^2\right)\left(\underline{b}+\sqrt{\frac{2\pi}{\min(2B, 2B(2-\alpha))}}\right)$$

Proof of Theorem 3.6

By Lemma A.6, Lemma A.5 can be applied to the set $A = \mathbb{W}_T$, which implies i). To prove ii), set $M = \frac{1}{2}$ in Lemma A.1 to prove the existence of a maximizer for (3.5). Analogously, $h \equiv 0$, $M = 1$ yields a maximizer for (3.6).

It remains to prove iii). In view of i), and since $\int_0^T (\dot{h}_t - \dot{x}_t)^2 dt \geq 0$, for any $h \in \mathbb{H}_T$ it follows that:

$$L(h) = \sup_{x \in \mathbb{H}_T} \left(2F(x) + \frac{1}{2} \int_0^T (\dot{x}_t - \dot{h}_t)^2 dt - \int_0^T \dot{x}_t^2 dt \right) \geq \sup_{x \in \mathbb{H}_T} \left(2F(x) - \int_0^T \dot{x}_t^2 dt \right) \tag{A.7}$$

which implies the inequality:

$$\inf_{h \in \mathbb{H}_T} L(h) \geq 2F(\hat{h}) - \int_0^T \dot{\hat{h}}_t^2 dt \tag{A.8}$$

and hence $\hat{h}$ is asymptotically optimal if (3.7) is satisfied. For the uniqueness part, consider two distinct solutions h, g to (3.6). Strict convexity implies that:

$$L(h) \geq 2F(g) + \frac{1}{2} \int_0^T (\dot{g}_t - \dot{h}_t)^2 dt - \int_0^T \dot{g}_t^2 dt > 2F(g) - \int_0^T \dot{g}_t^2 dt = 2F(h) - \int_0^T \dot{h}_t^2 dt$$

which contradicts the optimality of h, and uniqueness follows.

References

Dembo, A. and Zeitouni, O. (1998), *Large deviations techniques and applications*, Vol. 38 of *Applications of Mathematics (New York)*, second edn, Springer-Verlag, New York.

Deuschel, J.-D. and Stroock, D. W. (1989), *Large deviations*, Vol. 137 of *Pure and Applied Mathematics*, Academic Press Inc., Boston, MA.

Dufresne, D. (2001), 'The integral of geometric Brownian motion', *Adv. in Appl. Probab.* **33**(1), 223–241.

Dupuis, P. and Ellis, R. S. (1997), *A weak convergence approach to the theory of large deviations*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York. , A Wiley-Interscience Publication.

Dupuis, P. and Wang, H. (2004), 'Importance sampling, large deviations, and differential games', *Stoch. Stoch. Rep.* **76**(6), 481–508.

Dupuis, P. and Wang, H. (2005), 'Dynamic importance sampling for uniformly recurrent Markov chains', *Ann. Appl. Probab.* **15**(1A), 1–38.

Geman, H. and Yor, M. (1992), 'Quelques relations entre processus de Bessel, options asiatiques et fonctions confluentes hypergéométriques', *C. R. Acad. Sci. Paris Sér. I Math.* **314**(6), 471–474.

Glasserman, P., Heidelberger, P. and Shahabuddin, P. (1999), 'Asymptotically optimal importance sampling and stratification for pricing path-dependent options', *Math. Finance* **9**(2), 117–152.

Glasserman, P. and Wang, Y. (1997), 'Counterexamples in importance sampling for large deviations probabilities', *Ann. Appl. Probab.* **7**(3), 731–746.

Kemna, A. and Vorst, A. (1990), 'A pricing method for options based on average values', *Journal of Banking and Finance* **14**, 113–129.

Levy, E. (1992), 'Pricing European average rate currency options', *Journal of International Money and Finance* **11**(5), 474–491.

Lyasoff, A. (2006), 'The integral of geometric brownian motion revisited', *Preprint* .

Schilder, M. (1966), 'Some asymptotic formulas for Wiener integrals', *Trans. Amer. Math. Soc.* **125**, 63–85.

Siegmund, D. (1976), 'Importance sampling in the Monte Carlo study of sequential tests', *Ann. Statist.* **4**(4), 673–684.

Stroock, D. W. (1993), *Probability theory, an analytic view*, Cambridge University Press, Cambridge.

Turnbull, S. and Wakeman, L. (1991), 'A Quick Algorithm for Pricing European Average Options', *The Journal of Financial and Quantitative Analysis* **26**(3), 377–389.

Varadhan, S. R. S. (1966), 'Asymptotic probabilities and differential equations', *Comm. Pure Appl. Math.* **19**, 261–286.