Least Squares Importance Sampling for Libor Market Models

Luca Capriotti
Global Modelling and Analytics Group, Investment Banking Division, Credit Suisse Group One Cabot Square, London, E14 4QJ, United Kingdom, e-mail: luca.capriotti@credit-suisse.com

Abstract

A recently introduced Importance Sampling strategy based on a least squares optimization is applied to the Monte Carlo simulation of Libor Market Models. Such Least Squares Importance Sampling (LSIS) allows the automatic optimization of the sampling distribution within a trial class by means of a quick presimulation algorithm of straightforward implementation. With several numerical examples we show that LSIS can be extremely effective in reducing the variance of Monte Carlo estimators often resulting, especially when combined with stratified sampling, in computational speed-ups of orders of magnitude.

Keywords

Monte Carlo Simulations, Variance Reduction Techniques, Importance Sampling, Derivatives Pricing.

1. Introduction

The level of sophistication of the models employed by investment firms for pricing derivative securities is dramatically increasing in the continuous search for a possible edge against competitors. As a result, most of the models used in practice is too complex to be treated by analytic or deterministic numerical methods, and Monte Carlo simulation becomes more often than ever the only feasible means of pricing and hedging.

The main limitation of Monte Carlo simulations is their computational cost. In fact, being stochastic in nature, their outcome is always affected by a statistical error, that can be generally reduced to the desired level of accuracy by iterating the calculation for long enough time. This comes with a high computational cost as such statistical uncertainties, all things being equal, are inversely proportional to the square root of the number of statistically independent samples. Hence, in order to reduce the error by a factor of 10 one has to spend 100 times as much computer time. For this reason, to be used on a trading floor, Monte Carlo simulations often require to be run on large parallel computers with a high financial cost in terms of hardware, infrastructure, and software development.

Several approaches to speed-up Monte Carlo calculations, such as Antithetic Variables, Control Variates, and Importance Sampling, have been proposed over the last few years [6]. These techniques aim at reducing the variance per Monte Carlo observation so that a given level of accuracy can be obtained with a smaller number of iterations. In general, this can be done by exploiting some information known *a priori* on the structure of the problem at hand, like a symmetry property of the Brownian paths (Antithetic Variables), the value of a closely related security (Control Variates), or the form of the statistical distribution of the random samples (Importance Sampling). Antithetic Variables and Control Variates are the most commonly used variance reduction techniques, mainly because of the simplicity of their implementation, and the fact that they can be accommodated in an existing Monte Carlo calculator with a small effort. However, their effectiveness varies largely across applications, and is sometimes rather limited [6].

On the other hand, Importance Sampling techniques, although potentially more powerful, have not been employed much in professional contexts until recently. This is mainly because they generally involve a bigger implementation effort. Moreover, when used improperly, Importance Sampling can increase the variance of the Monte Carlo estimators, thus making its integration in an automated environment more delicate. Nonetheless, the potential efficiency gains at stake are so large that the interest in finding efficient Importance Sampling schemes is still very high.

The idea behind Importance Sampling is to reduce the statistical uncertainty of a Monte Carlo calculation by focusing on the most important sectors of the space from which the random samples are drawn. Such regions critically depend on both the random process simulated, and the structure of the security priced. For instance, for a deep out-of-the money Call option [13], the payoff sampled is zero for most of the iterations of a Monte Carlo simulation. Hence, simulating more samples with positive payoff reduces the variance. This can be done by changing the probability distribution from which the samples are drawn, and

reweighing the payout function by the appropriate likelihood-ratio (Radon-Nikodym derivative) in order to produce an unbiased result of the original problem [6].

Most of the work in Importance Sampling methods for security pricing has been done in a Gaussian setting [17,2,21,7,8,18,19,1,10] such the one arising from the simulation of a diffusion process. In this framework, Importance Sampling is achieved by modifying the drift term of the simulated process in order to drive the Brownian paths towards the regions that are the most important for the evaluation of the security. For instance, for the Call option above, this can be obtained by increasing the drift term up to a certain optimal level [17,2]. The different approaches proposed in the literature, essentially differ in the way in which such change of drift is found, and can be roughly divided into two families depending on the strategy adopted. The first strategy, common to the so-called adaptive Monte Carlo methods [21,18,19,1], aims to determine the optimal drift through stochastic optimization techniques that typically involve an iterative algorithm. On the other hand, the second strategy, proposed in a remarkable paper by Glasserman, Heidelberger, and Shahabuddin (GHS) [7], relies on a deterministic optimization procedure that can be applied for a specific class of payouts.

In a recent paper [5], we introduced the Least Squares Importance Sampling (LSIS) technique, as an alternative and flexible variance reduction strategy for Monte Carlo security pricing. This approach, originally proposed in Physics for the optimization of quantum mechanical wave functions of correlated electrons [20], was shown in Ref. [5] to provide an effective tool also for financial applications. In LSIS the determination of the optimal drift – or more in general of the most important regions of the sample space – is formulated in terms of a least squares minimization. This technique can be easily implemented and included in an existing Monte Carlo code, and simply relies on a standard least square algorithm for which several optimized libraries are available.

In this paper we apply the LSIS strategy to the simulation of a multi-factor Libor Market Model, and test its effectiveness on a variety of contracts. In addition, to further increase the computational efficiency we combine LSIS with stratified sampling [11]. The resulting variance reduction strategy is shown to be quite effective in a variety of cases, providing computational speed-ups of up to two order of magnitude.

In the following Section, we begin by discussing the simulation setting to which we apply the LSIS strategy. Then in Section 3 we review the main ideas behind Importance Sampling, and the principal approaches proposed in the financial literature. The rationale of LSIS is discussed in Section 4 together with the essential implementation details, and in Section 5 we illustrate how to combine LSIS with stratified sampling. Sections 6 and 7 discuss the Libor Market Model setting, and present the numerical results obtained with LSIS in this case. Finally, we draw our conclusions in Section 8.

2. The Setting

Although the variance reduction technique we discuss in this paper can be applied to a variety of financial problems, in the following we will focus on pricing applications that involve the simulation of multi-dimensional diffusions of the form

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dW_t. \tag{2.1}$$

Here the process $X(t)$ and the drift $\mu(X, t)$ are both L-dimensional real vectors, W_t is a N-dimensional standard Brownian motion, and the volatility, $\sigma(X, t)$, is a $L \times N$ real matrix. We will consider the problem of estimating the value at time $t = 0$, of contracts depending on the path followed by $X(t)$ within a certain interval $[0, T]$. This is given by the expectation value under the risk neutral probability measure, P [12] of the (discounted) payout functional $G[X(T)]$

$$V = E_P[G[X(T)]]. \tag{2.2}$$

Continuous time processes of the form (2.1) are typically simulated by sampling $X(t)$ on a discrete grid of points, $0 = t_0 < t_1 < \ldots < t_M = T$, by means, for instance, of a Euler scheme[a]

$$X_{i+1} = X_i + \mu(X_i, t)\Delta t_i + \sigma(X_i, t)\sqrt{\Delta t_i}\tilde{Z}_{i+1}, \tag{2.3}$$

where $X_i = X(t_i)$, $\Delta t_i = t_{i+1} - t_i$, and $\tilde{Z}_{i+1}$ is a N-dimensional vector of independent standard normal variates. In this representation, each discretized path for the vector process $X(t)$ can be put into a one to one correspondence with a set of $d = N \times M$ independent standard normal variables Z. As a result, the original problem of evaluating the expectation value of a functional of the realized path of the process $X(t)$ can be formulated as the calculation of expectation values of the form

$$V = E_P[G(Z)] = \int dZ\, G(Z)\, P(Z), \tag{2.4}$$

where $G(Z) = G(Z_1, \ldots, Z_d)$ is the scalar function obtained by discretizing the payout functional $G[X(T)]$ on a mesh of d sampling points, and the density is given by a d-dimensional standard normal distribution

$$P(Z) = N(0, I_d) \equiv (2\pi)^{-d/2} e^{-Z^2/2}, \tag{2.5}$$

where $Z^2 = Z \cdot Z$. For instance, for the familiar Call option in the Black-Scholes framework [13] one has $d = 1$, $P(Z) = (2\pi)^{-1/2} \exp(-Z^2/2)$ and

$$G(Z) = e^{-rT}\left(X_0 \exp\left[\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Z\right] - K\right)^+ \tag{2.6}$$

where r is the risk-free interest rate, σ is the volatility, X_0 and K are respectively the spot and strike price, and T the maturity of the option.

Whenever the dimension d of the state variable Z is large (say $d \gtrsim 5$) standard numerical quadrature approaches become highly inefficient, and Monte Carlo methods are the only feasible route for estimating expectation values of the form (2.4). To do so, one interprets Eq. (2.4) as a weighted average of the payout function $G(Z)$ over the possible configurations Z with weights given by the probability distribution $P(Z)$. This immediately leads to the simplest (and crudest) Monte Carlo estimator which is obtained by averaging the payout function over a sample of N_p *independent* values of the random variable Z generated according to the probability distribution $P(Z)$,

$$V \simeq \bar{V} = \frac{1}{N_p}\sum_{i=1}^{N_p} G(Z_i) \quad Z_i \sim P(Z). \tag{2.7}$$

In particular, the central limit theorem [14] ensures that, for big enough samples, the values of the estimator $\bar{V}$ are normally distributed around the true value, and converge for $N_p \to \infty$ towards V namely

$$V \simeq \frac{1}{N_p}\sum_{i=1}^{N_p} G(Z_i) \pm \frac{\Sigma}{\sqrt{N_p}}, \tag{2.8}$$

where $\Sigma^2 = E_P[G(x)^2] - E_P[G(x)]^2$ is the variance of the estimator and can be similarly approximated by

$$\Sigma^2 \simeq \frac{1}{N_p}\sum_{i=1}^{N_p} (G(Z_i) - \bar{V})^2. \tag{2.9}$$

Although Eq. (2.8) ensures the convergence of the Monte Carlo estimator to the expectation value (2.4), its practical utility depends on the magnitude of the variance, Σ^2. Indeed, the square root convergence in (2.8), implies that the number of replications N_p that are (asymptotically) necessary to achieve a given level of accuracy is proportional to the variance of the estimator[b]. Roughly speaking, such quantity is relatively small whenever the function $G(Z)$ is approximately constant over the region of values of Z that is represented the most among the random samples, i.e., the region that contains most of the probability mass of $P(Z)$. This is generally not the case for most of the pricing problems encountered in practice, and the calculation of accurate estimates of the expectation value (2.4) may require large sample sizes N_p, thus becoming computationally demanding.

3. Importance Sampling

The key observation underlying Importance Sampling is that the choice of extracting the random variable Z according to the probability distribution $P(Z)$ in order to sample stochastically Eq. (2.4), although natural, is by no means the only possible one. Indeed, the Monte Carlo integration can be performed by sampling an arbitrary probability distribution $\tilde{P}(Z)$ provided that the integral is suitably reweighed. In fact, using the identity

$$\int dZ\, G(Z)\, P(Z) = \int dZ \frac{G(Z)P(Z)}{\tilde{P}(Z)}\tilde{P}(Z), \tag{3.1}$$

an alternative estimator of the expectation value (2.4) is readily found as

$$V \simeq \tilde{V} = \frac{1}{N_p}\sum_{i=1}^{N_p} W(Z_i)G(Z_i) \quad Z_i \sim \tilde{P}(Z), \tag{3.2}$$

with the weight function given by $W(Z) = P(Z)/\tilde{P}(Z)$. The variance of the new Monte Carlo estimator reads

$$\tilde{\Sigma}^2 = \int dZ\, (W(Z)\, G(Z) - V)^2 \tilde{P}(Z) \tag{3.3}$$

and critically depends on the choice of the sampling probability distribution $\tilde{P}(Z)$. For non-negative functions $G(Z)$, the optimal choice of $\tilde{P}(Z)$ is the one for which $\tilde{\Sigma}$ vanishes, namely:

$$P_{\text{opt}}(Z) = \frac{1}{V}G(Z)P(Z). \tag{3.4}$$

In fact, the Monte Carlo estimator corresponding to such *optimal sampling distribution* reads

$$\tilde{V} \simeq \frac{1}{N_p}\sum_{i=1}^{N_p} W(Z_i)G(Z_i) = \frac{1}{N_p}\sum_{i=1}^{N_p} V, \tag{3.5}$$

leading to a constant value V on each Monte Carlo replication, and resulting therefore in zero variance[c]. Unfortunately, such a choice is not really viable as the normalization constant, V, is the expectation value (2.4) we want to calculate in the first place. Nevertheless, this observation provides the useful indication that the sampling density $\tilde{P}(Z)$, modulus a normalization, should be as close as possible to the product of the payout $G(Z)$ and the original multi-variate Gaussian distribution (2.5).

In this respect, Importance Sampling strategies generally choose a family of trial probability densities, $\tilde{P}_\theta(Z)$ *dash* depending on a set of N_θ real parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_{N_\theta})$ *dash* and aim at determining the one that minimize the variance of the estimator (3.3) within the class. In particular, Importance Sampling methods in security pricing generally try to guide the sampled paths towards the most important regions of the configuration space (i.e., where the contribution of the integrand is the largest), by means of a change of the drift terms of the process (2.1) or (2.3). The corresponding trial probability density reads

$$\tilde{P}_{\tilde{\mu}}(Z) = (2\pi)^{-d/2} e^{-(Z-\tilde{\mu})^2/2}, \tag{3.6}$$

where $\tilde{\mu}$ is a d-dimensional vector, and the weight function, as also expected from the Girsanov theorem [15], is

$$W_{\tilde{\mu}}(Z) = \exp[-\tilde{\mu}\cdot Z + \tilde{\mu}^2/2]. \tag{3.7}$$

A variety of approaches for the determination of the drift vector $\tilde{\mu}$ minimizing the variance of the estimator (3.3) has been recently proposed in the literature [21,7,8,18,19,1]. These can be roughly classified into two families depending on the strategy adopted.

The first strategy, common to the so-called adaptive Monte Carlo methods, is based on a stochastic minimization of the variance. Such minimization differs in details in the various methods but always involves an iterative procedure, to be performed in a preliminary Monte Carlo simulation.

In particular, Su and Fu [18,19], building upon previous work by Vazquez-Abad and Dufresne [21], used a gradient-based stochastic approximation, dubbed infinitesimal perturbation analysis, in order to estimate the optimal *uniform shift* of the drift for the diffusion (2.3), minimizing the variance of the estimator (3.3). In the notation of this Section, this translates in working with a trial density of the form (3.6) where the drift vector $\tilde{\mu}$ has components all equal to a single optimization parameter. The improvement of this method with respect to the one of Ref. [21], is that the minimization is carried out under the original probability measure, while in the latter the minimization was formulated under the trial probability measure. As a result, the stochastic minimization applies also for non differentiable payout, thus making

the approach more general. The application of this technique to partial average Asian options in a Black-Scholes market, and to Caplets under the Cox-Ingersoll-Ross model provides significative variance reductions [18,19].

Along similar ideas, Arouna [1] has recently proposed a different stochastic optimization methods for the determination of the optimal sampling density (3.6). Here, in contrast to the previous approach, all the components of the drift vector are independently optimized. The method relies on a truncated version of the Robbins-Monro algorithm that is shown to converge asymptotically to the optimal drift, and to provide an effective variance reduction in a variety of cases. However, as remarked by the same author, a critical aspect of the practical implementation of the Robbins-Monro algorithm is that it depends on the size of the iterative step. Hence, a particular care needs to be taken in order for the algorithm to be efficient[d].

On the other hand, the alternative strategy for the optimization of the trial density (3.6), proposed by Glasserman, Heidelberger, and Shahabuddin [7], relies on a saddle point approximation to minimize the variance of the estimator (3.3), or equivalently of its second moment (in the original measure)

$$m_2(\tilde{\mu}) = \int dZ\, W_{\tilde{\mu}}(Z) G(Z)^2 P(Z). \tag{3.8}$$

In fact, if the payout function $G(Z)$ is positive definite, by defining $F(Z) = \log G(Z)$ one can approximate Eq. (3.8) with the zero-order saddle point expansion

$$\begin{aligned}(2\pi)^{-d/2} &\int dZ \exp[2F(Z) - \tilde{\mu}\cdot Z + \tilde{\mu}^2/2 - Z^2/2] \\ &\simeq C \exp[\max_Z (2F(Z) - \tilde{\mu}\cdot Z + \tilde{\mu}^2/2 - Z^2/2)],\end{aligned}$$

where C is a constant. As a result, within this approximation, the problem of determining the optimal change of drift boils down to finding the vector μ such that

$$\max_Z (2F(Z) - \tilde{\mu}\cdot Z + \tilde{\mu}^2/2 - Z^2/2) \tag{3.9}$$

is minimum. It is easy to show that this is obtained by choosing $\tilde{\mu}^* = Z^*$ where Z^* is the point that solves the optimization problem

$$\max(F(Z) - Z^2/2), \tag{3.10}$$

or equivalently, for which the payout times the original distribution, $G(Z)P(Z)$, is maximum, i.e., Z^* corresponds to the maximum of the optimal sampling density, Eq. (3.4). The simplest interpretation of the saddle point approach is therefore that it approximates the zero variance distribution by means of a normal density with the same mode and variance.

This approach has been recently generalized to the continuous time in the Black-Scholes framework in a recent work by Guasoni and Robertson [10]. This formulation allows one to express the problem of the determination of the optimal drift in terms of a one-dimensional variational problem, and the solution of a Euler Lagrange equation.

The saddle point approach can be expected to be particularly effective in reducing the variance of the Monte Carlo estimator whenever the log payout function $F(Z)$ is close to be linear in the portion of the configuration space where most of the probability mass of $P(Z)$ lays. However, whenever the optimal sampling probability (3.4) cannot be accurately represented by a single Gaussian with the same mode and variance, the saddle point approximation is less beneficial. In particular, this approach turns out to be less effective whenever the structure of the payout function $G(Z)$ is such that the optimal sampling distribution (3.4) has a width which is very different from the one of the original distribution, or is multi-modal.

In the following Section we describe an alternative least squares strategy that is straightforward to implement and flexible enough to be applied in a generic Monte Carlo setting. Indeed, the Least Squares Importance Sampling (LSIS) is not limited to the determination of the optimal change of drift in a Gaussian model. Instead, it can be applied to any Monte Carlo simulation provided that a reasonable guess of the optimal sampling density is available. For this reason, in the next Section we will momentarily leave the Gaussian framework, and we will describe the rationale of LSIS in a more general setting.

4. Least Squares Importance Sampling

A practical approach to the search of an effective Importance Sampling distribution can be formulated in terms of a non-linear optimization problem. To this purpose, let us consider the family of trial probability densities, $\tilde{P}_\theta(Z)$. The variance of the estimator corresponding to $\tilde{P}_\theta(Z)$, Eq. (3.3), can be written in terms of the *original* probability distribution $P(Z)$ as

$$\tilde{\Sigma}^2_\theta = E_P[W_\theta(Z) G^2(Z)] - E_P[G(Z)]^2, \tag{4.1}$$

with $W_\theta(Z) = P(Z)/\tilde{P}_\theta(Z)$. Hence, the optimal Importance Sampling distribution within the family $\tilde{P}_\theta(Z)$ is the one for which the latter quantity, or equivalently the second moment (3.8) or

$$E_P[W_\theta(Z) G^2(Z)], \tag{4.2}$$

is minimum. The crucial observation is that the Monte Carlo estimator of this quantity,

$$m_2(\theta) \simeq \frac{1}{N'_p} \sum_{i=1}^{N'_p} (W_\theta(Z_i)^{1/2} G(Z_i))^2 \quad Z_i \sim P(Z), \tag{4.3}$$

can be interpreted as a non-linear least squares fit of a set of N'_p data points (x_i, y_i) with a function $y = f_\theta(x)$ parameterized by θ, with the correspondence $y_i \to 0$, $x_i \to Z_i$, and $f_\theta(x) \to W_\theta(Z)^{1/2} G(Z)$. The latter is a standard problem of statistical analysis that can be tackled with a variety of robust and easily accessible numerical algorithms, as the so-called Levenberg-Marquardt method [16].

Alternatively, to improve the numerical stability of the least-squares procedure, it is convenient in some situations to minimize, instead of (4.2), the pseudo-variance

$$\begin{aligned}S_2(\theta) &= E_P[(W_\theta(Z)^{1/2} G(Z) - V_T)^2] \\ &\simeq \frac{1}{N'_p} \sum_{i=1}^{N'_p} (W_\theta(Z_i)^{1/2} G(Z_i) - V_T)^2\end{aligned} \tag{4.4}$$

where the constant V_T is a guess of the option value. Indeed, the minimization of (4.4) is equivalent to the one of the real variance of the estimator (4.1) as

$$S_2(\theta) = \tilde{\Sigma}_\theta^2 + (E_P[G(Z)] - V_T)^2. \tag{4.5}$$

The algorithm for the determination of the optimal sampling distribution within a certain trial family can be therefore summarized as it follows:

(1) Generate a suitable number N_p' of replications of the state variables Z according to the *original* probability distribution $P(Z)$;
(2) Choose a trial probability distribution $\tilde{P}_\theta(Z)$, and an initial value of the vector of parameters θ;
(3) Set $x_i \to Z_i, f_\theta(x) \to W_\theta(Z)^{1/2}G(Z)$ and $y_i \to 0 (resp. y_i \to V_T)$ and call a least squares fitter, say LSQ $[x, y, f_\theta(X), \theta]$, providing the optimal $\theta = \theta^*$ by minimizing the second moment of the estimator $m_2(\theta)$, Eq. (4.3) [*resp.* $S_2(\theta)$, Eq. (4.4)].

Once the optimal parameters θ^* have been determined through the least squares algorithm, one can perform an ordinary Monte Carlo simulation by sampling the probability distribution $\tilde{P}_{\theta*}(Z)$, and calculating expectation values according to Eq. (3.2).

What makes LSIS a practical strategy is that just a relatively small number of replications $N_p' \ll N_p$ is usually required to determine the optimal parameters $\tilde{\theta}^*$. This is due to the fact that the configurations over which the optimization is performed are fixed. As a result of this form of *correlated sampling* [20], the difference in the $m_2(\theta)$'s for two sets of values of the parameters being optimized is much more accurately determined than the values of the $m_2(\theta)$'s themselves. This rather surprising feature is rooted in the fact that the minimization of Eq. (4.3) as a means to optimize the trial density, $\tilde{P}_\theta(Z)$, can be justified in terms of a genuine maximum likelihood criteria [4], and it is therefore independent on how accurately $m_2(\theta)$ approximates the quantity (4.2). As a result, the overhead associated with the optimization of the trial density is generally fairly limited, thus making LSIS a practical approach for variance reduction.

In a companion paper [5] we have demonstrated the effectiveness of LSIS by applying it to a variety of test cases. In particular, we have shown that LSIS provides variance reductions comparable or superior to those of the Importance Sampling methods most recently proposed in the financial literature [7,18,19,1]. As a simple example, for instance, below we briefly review the results obtained for standard Call and Put options in a Black-Scholes setting. In this case the payout function reads as in Eq. (2.6) (for the call), and the sampling density $P(Z)$ is a univariate standard normal distribution.

As discussed above, Importance Sampling techniques seek a sampling probability density $\tilde{P}_\theta(Z)$ as close as possible to the optimal sampling distribution, Eq. (3.4) (see Figure 1). The simplest choice for $\tilde{P}_\theta(Z)$, in this setting, is a Gaussian distribution of the form (3.6) (with $d = 1$), so that the only parameter θ to optimize is the drift $\tilde{\mu}$. We found that the least squares fitter was able to determine successfully the optimal $\tilde{\mu}$ with as little as $N_p' \simeq 50$ Monte Carlo replications.

In Tables 1 and 2 we compare the results obtained with LSIS with the ones obtained by means of the Robbins Monro (RM) adaptive Monte Carlo (as quoted in Ref. [1]), and the saddle point approach of GHS [7]. Here, as an indicator of the efficiency gains introduced by the different strategies of Importance Sampling, we have defined the variance ratio as

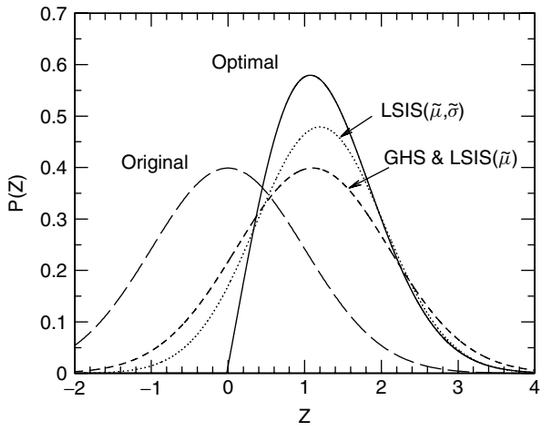

Fig. 1. Sampling probability density functions for a European Call option (2.6) with $T = 1$, $r = 0.05$, $\sigma = 0.3$, $X_0 = K = 50$ as obtained with LSIS [optimizing just the drift, LSIS($\tilde{\mu}$), and both the drift and the volatility, LSIS($\tilde{\mu}, \tilde{\sigma}$)], and the saddle point approximation of Ref. [7] (GHS). On this scale the results for LSIS($\tilde{\mu}$) and GHS are indistinguishable. The original (2.5) and the optimal (3.4) sampling densities are also shown for comparison.

$$\mathrm{VR} = \left(\frac{\sigma(\text{Crude MC})}{\sigma(\text{IS})} \right)^2 \tag{4.6}$$

where the numerator and denominator are the statistical errors (for the same number of Monte Carlo paths) of the Crude and the Importance Sampling estimators, respectively.

We found that the different methods produce a significative and comparable variance reduction. Intuitively, the change of drift is more effective for low volatility, and deep in and out of the money options (see also the discussion in the Introduction). In this case, the LSIS and GHS optimized trial distributions $\tilde{P}_{\tilde{\mu}}(Z)$ are very similar as shown Fig. 1. This could be expected as, in this case, the optimal Importance Sampling distribution (3.4) can be effectively approximated by a Gaussian with the same mode and variance, so that the GHS approach produces accurate results.

However, the LSIS method is not limited to Importance Sampling strategies based on a pure change of drift, and one can easily introduce additional optimization parameters in the trial density. For instance, in this example it makes sense to introduce the sampling volatility, $\tilde{\sigma}$,

$$\tilde{P}_{\tilde{\mu},\tilde{\sigma}}(Z) = (2\pi\tilde{\sigma}^2)^{-1/2} e^{-(Z-\tilde{\mu})^2/2\tilde{\sigma}^2}. \tag{4.7}$$

As illustrated in Fig. 1, by adjusting both $\tilde{\mu}$ and $\tilde{\mu}$, one obtains a trial density closer to the optimal one. This corresponds to an additional variance reduction up to over one order of magnitude, as shown in Tables 1 and 2.

5. Stratified Sampling

In a diffusive setting, LSIS can be naturally combined with stratified sampling [11] in order to achieve further variance reductions. In this Section

Table 1. Variance reductions (4.6) obtained with different Importance Sampling strategies. Comparison between LSIS, the adaptive Robbins-Monro (RM) algorithm (as quoted in Ref. [1]), and the saddle point approach of Ref. [7] (GSH): price of a European Call option on a lognormal asset (2.6) for different values of the volatility σ, and of the strike price K. The parameters used are $r = 0.05$, $X_0 = 50$, $T = 1.0$, and the number of simulated paths is 1,000,000 for Crude MC, LSIS and GHS, 50,000 for RM. Results for LSIS obtained by optimizing the drift only [LSIS $(\bar{\mu})$], and both the drift and the volatility [LSIS $(\bar{\mu}, \bar{\sigma})$] are reported. The uncertainties are reported in parentheses

σ	K	LSIS $(\bar{\mu})$	LSIS $(\bar{\mu}, \bar{\sigma})$	RM	GHS
0.1	30	104(1)	1700(100)	112(4)	100(1)
	50	7.8(1)	15(1)	7.8(4)	7.8(1)
	60	33.5(5)	84(5)	31(2)	33.5(5)
0.3	30	16.4(1)	51(1)	16.8(4)	14.8(2)
	50	9.9(5)	27(1)	11(2)	9.9(1)
	60	15.6(1)	35(1)	15.2(4)	14.2(1)

we illustrate how. We begin by reviewing the basic ideas underlying Stratification following Refs. [7,6].

Stratification is a technique that allows one to draw samples from a specified distribution in a more regular pattern thus reducing the variance. This is achieved by ensuring that the fraction of samples which falls in different subsets, or *strata*, of the domain of the random variable matches the theoretical probability of each subset. For example, in order to perform a stratified sampling of a single standard normal variable one can divide the real axis into M strata, such that the probability of the random variable to fall in any of them is $1/M$. This can be done easily by first dividing the unit interval (0, 1) into M segments of length $1/M$, and sampling uniformly from each of them. Then, each of the sampled uniform is mapped into a standard Gaussian by means of the inverse cumulative normal distribution. The resulting set of M variates will contain exactly one variable for each of the M strata of the real axis, and constitute therefore a stratified sample of the standard normal distribution. This simple algorithm can be therefore summarized as it follows:

(1) Draw M random variables, say $u^1, \ldots, u^M$, uniformly distributed in (0, 1).
(2) Define a new set of M random variables

$$v^{(i)} = \frac{i-1}{M} + \frac{u^{(i)}}{M},$$

with $i = 1, \ldots, M$, i.e., such that the i-th variable is uniformly distributed in the interval $(i - 1/M, i/M)$.
(3) Set

$$X^{(i)} = \Phi^{-1}(v^{(i)}),$$

where Φ is the standard normal cumulative density function. The variables $(X^{(1)}, \ldots, X^{(M)})$ constitute the sample of the standard normal distribution, stratified into M strata.

Although this procedure can be generalized to multi-dimensional normal variates, it becomes unpractical in high-dimension ($d \gtrsim 5$) for the same reason for which estimating the integral (2.4) by numerical quadrature becomes exponentially inefficient: if each dimension is divided into M strata, their total number scales as M^d. As a result, generating just one point on each stratum requires a sample size at least this large, thus becoming prohibitive for the values of $M \gtrsim 10$ that generally make Stratification effective in reducing the variance.

A feasible way of applying Stratification to the sampling of a multivariate normal distribution is to stratify only a specific one-dimensional projection of the random variable $Z \sim N(0, I_d)$. This is straightforward because, the projection of Z along a direction in $\mathbb{R}^d$ represented by a unit vector ξ, $\xi \cdot Z$, is a standard normal variable that can be stratified using

Table 2. Same as Table 1 for a European Put option

σ	K	LSIS $(\bar{\mu})$	LSIS $(\bar{\mu}, \bar{\sigma})$	RM	GHS
0.1	40	435(6)	571(9)	350(24)	435(6)
	50	8.8(1)	25(2)	9.6(4)	9.1(1)
	60	5.9(1)	17(1)	6.3(4)	5.9(1)
0.3	30	41(1)	69(2)	38(4)	40.8(5)
	50	5.8(1)	16.5(5)	6.2(4)	5.8(1)
	60	4.9(1)	13.9(2)	4.8(4)	4.4(1)

the one-dimensional algorithm described above. In addition, it is also easy to sample the vector Z conditional to a specific value of its projection $\xi \cdot Z$, as the conditional distribution $(Z|\xi \cdot Z = x)$ is itself normal and given by $N(x\xi, I_d - \xi\xi^t)$. The resulting algorithm leading to the stratification of Z along the direction ξ can be therefore summarized as it follows:

(1) Generate a stratified sample of $X^{(1)}, \ldots, X^{(M)}$ of the standard normal distribution as described above. Interpret $X^{(i)}$ as the i-th value of the one-dimensional projection $\xi \cdot Z$, of $Z \sim N(0, I_d)$.
(2) Draw M independent d-dimensional Gaussian variates $Y^{(i)}$ from $N(0, I_d)$.
(3) Set

$$Z^{(i)} = \xi X^{(i)} + (I_d - \xi\xi^t)Y^{(i)}.$$

The resulting set $(Z^{(1)}, \ldots, Z^{(M)})$ constitutes a sample from $N(0, I_d)$ stratified along the direction ξ into M strata.

Loosely speaking, the Stratification of a one-dimensional projection of a multi-dimensional normal variate has nearly the same effect of replacing the Monte Carlo integration with a numerical quadrature along the stratified direction ξ, while still using Monte Carlo for the remaining ones. Clearly, the choice of the direction ξ is critical for the Stratification to be effective in terms of variance reduction. This is likely to be the case if the output is strongly correlated to the value of the projection $\xi \cdot Z$.

As anticipated, the simplest possible strategy for Importance Sampling in a Gaussian framework, is to look for an optimal change of drift, i.e. to adopt the simple shifted Gaussian of Eq. (3.6) as trial probability density. In this setting, as suggested by Glasserman and collaborators [7], a natural choice for the direction of stratification is the optimal drift vector itself. This can be rigorously justified if the payout is a function of a linear combination of the Z_i's. However, in Refs. [7,8] and [5] it has been shown that this choice works in practice more in general, turning out to be highly effective in a variety of cases. In this paper, we also follow this strategy, and demonstrate its effectiveness for a variety of examples in the context of the Libor Market Model.

6. The Libor Market Model Setting

In the remainder of this paper we will apply the LSIS strategy, reviewed above, to the Libor Market Model of Brace, Gatarek and Musiela [3] for the arbitrage-free evolution of the forward Libor rates. In order to introduce this framework, we indicate with T_i, $i = 1, \ldots, M+1$, a set of $M+1$ bond maturities, with spacings $h = T_{i+1} - T_i$, assumed constant for simplicity. The Libor rate as seen at time t for the interval $[T_i, T_{i+1})$, $L_i(t)$, evolves according to the following stochastic differential equation

$$\frac{dL_i(t)}{L_i(t)} = \mu_i(L(t))dt + \sigma_i(t)^T dW_t, \quad 0 \le t \le T_i, \quad i = 1, \ldots, M, \qquad (6.1)$$

where W is a N-dimensional standard Brownian motion, $L(t)$ is the M-dimensional vector of Libor rates, and $\sigma_i(t)$ the N-dimensional vector of volatilities, both at time t. Here the drift term, as imposed by the arbitrage free conditions, reads

$$\mu_i(L(t)) = \sum_{j=\eta(t)}^{i} \frac{\sigma_i^T \sigma_j h L_j(t)}{1 + h L_j(t)}, \qquad (6.2)$$

where $\eta(t)$ denotes the index of the bond maturity immediately following time t, with $T_{\eta(t)-1} \le t < T_\eta(t)$.

Equation (6.1) can be simulated by applying a Euler discretization to the logarithms of the forward rates, and by dividing each interval $[T_i, T_{i+1})$ into n_e steps of equal width, $h_e = h/n_e$. This gives

$$\frac{L_i(n+1)}{L_i(n)} = \exp[(\mu_i(L(N)) - \sigma_i(n)||^2/2)h_e + \sigma_i^T(n)Z(n+1)\sqrt{h_e}], \quad (6.3)$$

for $i = \eta(nh_e), \ldots, \ldots, M$, and $L_i(n+1) = L_i(n)$ if $i < \eta(nh)$. Here Z is a N-dimensional vector of independent standard normal variables. Under the discretized model (6.3), the problem of evaluating the price of a contract written on a set of Libor rates is then formulated in the general form (2.4), and LSIS can be straight-forwardly applied.

In the following we will present results using a trial probability density involving displaced Gaussian multi-variate distributions of the form (3.6). This choice requires in principle the optimization of a number of parameters – the components of the drift vector $\tilde{\mu}$ – proportional to the number of Gaussian univariate Z_i necessary for the propagation of the Libor rates in the desired time horizon, namely $d = M \times N \times n_e$. As the number of time steps or the number of factors of the simulation increase, the complexity of the optimization problem increases as well. Nevertheless, as suggested in Ref. [8] and verified in the companion paper [5] for a variety of examples, one can significantly reduce the computation time associated with the optimization stage by approximating the drift vector with a continuous function parameterized by a small number of parameters. These are in turn tuned by the least square algorithm in order to determine an approximate optimal drift vector. We have found that a particularly effective realization of this approach is to approximate the drift vector by a piecewise linear function, parameterized by its values where it changes slope (the so-called *knot points*). In particular, in the simulation of the LMM we have found that by using a very limited number of knot points for each random factor (say for 1 to 5) one is able to achieve very effective variance reductions through LSIS and LSIS plus Stratification. Hence the simulation of the LMM required the optimization of a very small number of parameters (form 3 to 15, for $N = 3$) thus making the overhead associated with the presimulation stage rather limited. More precisely, we found that a few hundred Monte Carlo configurations and 10-20 iterations of the least squares fitter, were typically enough to determine the optimal drift vector. In addition, such vector generally changes continuously with the simulation parameters. As a result, an even faster convergence in the iterative procedure can be obtained by starting the pre-simulation from a drift vector optimized for a case with a similar set of parameters.

7. Numerical Results

The numerical results we present in this Section are based on the evolution of (6.3) in a three-factor ($N = 3$) model with $h = 1/4$ (a quarter of a year), and $n_e = 3$. Following Ref. [9], to keep things simple we take the volatilities to be functions of time to maturity

$$\sigma_i(t) = \sigma_{i-\eta(t)+1}(0), \tag{7.1}$$

with

$$\sigma_i^j(0) = \sigma_0(1+\alpha j)(1+\beta_i), \tag{7.2}$$

$j = 1, \ldots, 3, \alpha = \beta = 0.01,$ and $\sigma_0 = 0.2$. As initial Libor curve we take instead

$$L_i(0) = l_0(1+\beta_i), \tag{7.3}$$

with $l_0 = 5\%$.

As a first example we consider a Caplet for the interval $[T_m, T_{m+1})$ struck at K,

$$C_h(T_M) = \left(\Pi_{i=0}^m \frac{1}{1+hL_i(T_i)}\right) h(L_m(T_m) - K)^+. \tag{7.4}$$

Table 3 displays the estimated variance ratios obtained with LSIS, and the combination of LSIS and Stratification (LSIS+) introduced in Section 5 for a variety of maturities, and strike prices that range from in the money to out of the money. Here the results are all obtained using (3.6) as trial probability density, and by parameterizing the change of drift of each factor with a single parameter or knot point, corresponding to a rigid shift. We have verified that increasing the number of knots does not provide further sizable benefits in this case. As shown in Table 3, LSIS provides remarkable variance reductions, corresponding to a saving of roughly one order of magnitude in computational time, consistently across maturities. For fixed maturity, as expected, LSIS is more effective for out of the money strikes since in these cases the fraction of paths expiring worthless is more significant. These paths clearly provide little information, and tend to increase the variance of the sample. Changing the drift increases the fraction of paths which end up in the money thus making the sample more homogeneous. Conversely, as the maturity increases, the variance reduction provided by LSIS decreases as the outturn distributions of the Libor rates become more delocalized, and the change of drift strategy becomes less effective.

The combination of LSIS and Stratification provides for Caplets a tremendous variance reduction of up to two order of magnitude (see Table 3). However, the effectiveness of LSIS+ decreases sharply with maturity. Nevertheless, for the examples considered, it still gives around a factor of 40 in variance reduction for a 7 year maturity, thus resulting in extensive savings in computational time also for fairly long expiries.

Although important instruments for calibration, Caplets constitute an easy test ground for LSIS and LSIS+ as they are mostly sensitive to the single Libor rate determining the final payment. A more articulated example on which to assess the efficacy of LSIS are interest rate Caps. We consider contracts with first payment T_n and last payment T_M, and tenor h

$$Cap_h(T_n, T_M) = \sum_{l=n}^{M} C_h(T_l). \tag{7.5}$$

The results obtained for a variety of maturities and strike prices are shown in Table 4. In this case we have verified that $N_k = 3$ knot points provided the bulk of the variance reduction for the trial density function (3.6). The efficiency gains produced by LSIS, although slightly smaller

Table 3: Variance reductions (4.6) obtained with LSIS and LSIS plus Stratification (*LSIS+*) for Caplets, Eq. (7.4), in a three factor Libor Market Model, for different maturities T_M, and strike prices K. N_k is the number of knots per factor (see text). The number of simulated paths is 200,000. The uncertainties on the variance reductions are reported in parentheses

T_m (years)	K	N_k	LSIS	LSIS+
1.0	0.04	1	11.4(1)	1349(1)
1.0	0.055	1	13.3(2)	2300(2)
1.0	0.07	1	20.2(1)	4126(4)
2.5	0.04	1	14.0(1)	1189(1)
2.5	0.055	1	15.5(1)	897(1)
2.5	0.07	1	18.1(1)	1831(1)
5.0	0.040	1	12.7(1)	235.2(5)
5.0	0.060	1	12.5(1)	237.0(5)
5.0	0.080	1	14.5(1)	193.3(4)
7.0	0.04	1	7.9(3)	40.0(1)
7.0	0.055	1	8.5(4)	43.7(1)
7.0	0.07	1	8.5(4)	40(1)

Table 4: Variance reductions obtained with LSIS and LSIS plus Stratification (*LSIS+*) for Caps Eq. (7.5) in a three factor Libor Market Model, for $T_n = 0.25$ (years), different final maturities T_M, and strike prices K. N_k is the number of knots per factor (see text). The number of simulated paths is 200,000. The uncertainties on the variance reductions are reported in parentheses

T_m (years)	K	N_k	LSIS	LSIS+
1.0	0.04	3	10.6(5)	37.2(8)
1.0	0.055	3	9.7(3)	19.8(5)
1.0	0.07	3	13.6(5)	21.6(6)
2.5	0.04	3	16.2(5)	40.3(7)
2.5	0.055	3	12.0(4)	33.8(7)
2.5	0.07	3	15.7(5)	47.3(8)
5.0	0.04	3	14.9(5)	43.7(9)
5.0	0.055	3	14.5(6)	46.7(9)
5.0	0.07	3	15.6(6)	55(1)
7.0	0.04	3	13.0(6)	42.6(8)
7.0	0.055	3	12.2(5)	45.1(9)
7.0	0.07	3	12.6(4)	55(1)

than in the case of a single Caplet, are consistently around 10–15 for all the maturities considered. As expected, LSIS+ is not able to provide the massive variance reductions observed for Caps. Nonetheless, for the cases considered, it provides a further reduction of the variance with respect to LSIS of a sizable factor ranging from 2 to 4.

LSIS and LSIS+ result in remarkable computational savings also for Swaptions. Here we have considered contracts with expiry T_n to enter in a swap with payments dates $T_{n+1}, \ldots, T_{M+1}$, with the holder of the option paying a fixed rate K

$$V(T_n) = \sum_{i=n+1}^{M+1} B(T_n, T_i) h (S_n(T_n) - K)^+, \tag{7.6}$$

where $B(T_n, T_i)$ is the price at time T_n of a bond maturing at time T_i

$$B(T_n, T_i) = \Pi_{l=n}^{i-1} \frac{1}{1 + hL_l(T_l)}, \tag{7.7}$$

and the swap rate reads

$$S_n(T_n) = \frac{1 - B(T_n, T_{M+1})}{h \Sigma_{l=n+1}^{M=1} B(T_n, T_l)}. \tag{7.8}$$

The results are shown in 5 and indicate that LSIS provides variance reductions in the range $7 \div 20$ and LSIS+ further increases the computational efficiency by up to one order of magnitude.

As a final example – illustrating for a simple case the flexibility of LSIS – we consider the combination of a long Caplet and Flooret in a Straddle contract

$$St_h(T_m) = \left(\Pi_{i=0}^{m} \frac{1}{1 + hL_i(T_i)} \right) h |L_m(T_m) - K|. \tag{7.9}$$

In this case, the optimal sampling distribution (see Sec.3), proportional to the product of the payout and the Gaussian sampling distribution (2.5), has two well separated maxima because of the modulus in Eq. (7.9). As a result, a single mode trial probability density (3.6) provides limited variance reductions, especially for strikes at the money, where the relative importance of the two maxima is similar (see Table 6). However, the LSIS is not limited to a Gaussian trial density and one can use this flexibility to utilize a more accurate guess of the optimal sampling distribution. In particular, a better *ansatz* for the optimal density is represented by a bi-modal trial density of the form

$$\tilde{P}(Z) = (2\pi)^{-d/2} \left[w_a e^{-(Z-\mu_a)^2/2} + w_b e^{-(Z-\mu_b)^2/2} \right], \tag{7.10}$$

where $w_a + w_b = 1$ that can be optimized over μ_a, μ_b, and w_a. The simulation of a density of this form is straightforward as it simply implies choosing one of the two Gaussian components in (7.10) on each Monte Carlo step, and sample a configuration Z_i according to it. This can be done by extracting an auxiliary uniform random number $\xi \in [0, 1]$, and sampling Z_i according to the first Gaussian component if $\xi < w_a$, and according to the second otherwise.

Table 5. Variance reduction obtained with LSIS and LSIS plus Stratification (*LSIS+*) for Swaptions Eq. (7.6) in a three factor Libor Market Model. T_n is the option expiry and T_{M+1} is the final payment date of the underlying swap. K is the strike price. N_k is the number of knots per factor (see text). The number of simulated paths is 200,000. The uncertainties on the variance reductions are reported in parentheses.

T_n (years)	T_{M+1}	K	N_k	LSIS	LSIS+
0.5	1.5	0.04	3	6.8(3)	35.2(8)
0.5	1.5	0.055	3	10.5(4)	143(2)
0.5	1.5	0.07	3	21.2(6)	209(2)
0.5	2.5	0.04	3	7.0(3)	41.9(9)
0.5	2.5	0.055	3	9.8(3)	149(2)
0.5	2.5	0.07	3	18.6(5)	427(2)
0.5	5.5	0.04	3	6.8(3)	50(1)
0.5	5.5	0.055	3	8.5(3)	106(1)
0.5	5.5	0.07	3	12.0(4)	148(1)
1.0	6.0	0.04	3	8.0(4)	144(2)
1.0	6.0	0.055	3	8.6(3)	165(2)
1.0	6.0	0.07	3	12.7(4)	654(3)
2.0	7.0	0.04	3	9.2(3)	70(1)
2.0	7.0	0.055	3	9.7(3)	139(1)
2.0	7.0	0.09	3	13.9(4)	140(1)
5.0	10.0	0.04	5	7.3(4)	76(1)
5.0	10.0	0.055	5	7.4(3)	72(2)
5.0	10.0	0.09	5	7.5(4)	197(2)

Table 6: Variance reduction obtained with LSIS for a Straddle Eq. (7.9) in a three factor Libor Market Model, for different maturities T_m, and strike prices K. N_k is the number of knots per factor (see text). Results are shown using Eq. (3.6) [LSIS] and Eq. (7.10) [LSIS (MM)] as trial densities. The number of simulated paths is 200,000. The uncertainties on the variance reductions are reported in parentheses

T_m (years)	K	N_k	LSIS	LSIS(MM)
1.0	0.04	1	2.8(1)	5.8(1)
1.0	0.05	1	1.3(1)	5.3(1)
1.0	0.06	1	1.0(1)	3.9(1)
1.0	0.07	1	1.1(1)	3.4(1)
5.0	0.04	1	2.8(1)	8.7(1)
5.0	0.05	1	1.9(1)	6.5(1)
5.0	0.06	1	1.5(1)	4.9(1)
5.0	0.07	1	1.2(1)	4.0(1)

As shown in Table 6, using this trial density, LSIS improves significantly the computational efficiency also for Straddle contracts.

8. Conclusions

In this paper we have described the application of the recently introduced Least Squares Importance Sampling (LSIS) [5] to the simulation of Libor Market Models. Such variance reduction technique allows one to automatically optimize the sampling distribution within a chosen trial class by means of a presimulation algorithm of straightforward implementation.

What makes the approach practical in a financial context is that the overhead associated with the least squares optimization of the trial density is generally rather limited especially after reducing the dimensionality of the problem by means of a careful parametrization.

With several numerical examples we have shown that LSIS can be extremely effective in reducing the variance per sample of the simulation, thus resulting in remarkable speed-ups. Moreover, when used with Gaussian trial probability densities, LSIS can be naturally combined with Stratification thus providing further efficiency gains that can result in computational savings of orders of magnitude.

The efficacy of any Importance Sampling strategy is much dependent on how effectively the trial density function is able to reweigh the different regions of the sampled space in order to reduce the statistical fluctuations of the accumulated observables. These regions depends on both the model simulated, and the structure of the payout being priced. In this respect LSIS, when compared with previously methods, offers additional potential leeway as it is not limited to Gaussian trial densities. This becomes important when the structure of the optimal density is particularly complex e.g., with multi-modal features, or complicated correlation structures. In this paper we have illustrated this point with a simple multi-modal example. Further work is currently in progress in order to introduce more flexible probability distributions as trial densities.

Acknowledgments

It is a pleasure to acknowledge Gabriele Cipriani, David Shorthouse, and Mark Stedman for stimulating discussions, and Paul Glasserman for kind and useful correspondence. The opinion and views expressed in this paper are uniquely those of the author, and do not necessarily represent those of Credit Suisse Group.

FOOTNOTES & REFERENCES

[a] The use of other discretization schemes does not alter the present discussion.

[b] In particular, the Monte Carlo integration becomes unfeasible if the variance of the estimator diverges, giving rise to the so-called *sign-problem* instability. Although this problem is the crux of Monte Carlo simulations in several branches of the Physical Sciences, see, e.g., S. Sorella and L. Capriotti, Physical Review B **61**, 2599 (2000), this issue does not usually affect financial contexts.

[c] It is possible to show [16] that, when $G(Z)$ does not have a definite sign, the optimal sampling density has the similar form $P_{\rm opt} = |G(Z)|P(Z)/V$, although in this case the resulting variance is not zero.

[d] B. Arouna, *private communication*.

[1] B. Arouna. *Journal of Computational Finance*, 7:1245–1255, 2003.

[2] P. Boyle, M. Broadie, and P. Glasserman. *Journal of Economic Dynamics and Control*, 21:1257–1321, 1997.

[3] A. Brace, D. Gatarek, and M. Musiela. *Mathematical Finance*, 7:127–155, 1997.

[4] D. Bressanini, G. Morosi, and M. Mella. *Journal Chemical Physics*, 116:5345–5350, 2002.

[5] L. Capriotti. Least Squares Importance Sampling for Monte Carlo Security Pricing. *preprint*, 2006.

[6] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.

[7] P. Glasserman, P. Heidelberger, and P. Shahabuddin. *Mathematical Finance*, 9:117–152, 1999.

[8] P. Glasserman, P. Heidelberger, and P. Shahabuddin. *Journal of Derivatives*, 7:32–50, 1999.

[9] P. Glasserman and X. Zhao. *Journal of Computational Finance*, 3:5–39, 1999.

[10] P. Guasoni and S. Robertson. Optimal Importance Sampling with Explicit Formulas in Continuous Time. *preprint*, 2006.

[11] J.M. Hammersley and D.C. Handscomb. *Monte Carlo Methods*. Methuen, London, 1964.

[12] J. Harrison and D. Kreps. *Journal of Economic Theory*, 20:381–408, 1979.

[13] J. C. Hull. *Options, Futures and Other Derivatives*. Prentice Hall, New Jersey, 2002.

[14] O. Kallenberg. *Foundations of Modern Probability*. Springer, New York, 1997.

[15] M. Musiela and M. Rutkowski. *Martingale Methods in Financial Modelling*. Springer, New York, 2002.

[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C++ : the art of scientific computing*. Cambridge University Press, Cambridge, 2002.

[17] R. Reider. Working paper, 1993.

[18] Y. Su and M. C. Fu. In J.A. Joines, R.R Barton, K. Kang, and P.A. Fishwick, editors, *Proceedings of 2000 Winter Simulation Conference*, pages 587–596, Piscataway NJ, 2000. IEEE Press.

[19] Y. Su and M. C. Fu. *Journal of Computational Finance*, 5:27–50, 2002.

[20] C. J. Umrigar, K. G. Wilson, and W. Wilkins. *Physical Review Letters*, 60:1719–1722, 1988.

[21] F. Vazquez-Abad and D. Dufresne. In D.J. Medeiros, E.F. Watson, J.S. Carson, and M.S. Manivann, editors, *Proceedings of 1998 Winter Simulation Conference*, pages 1493–1500, Washington DC, 1998. IEEE Press.