Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

IMPORTANCE SAMPLING FOR MULTIMODAL FUNCTIONS AND APPLICATION TO PRICING EXOTIC OPTIONS

Athanassios N. Avramidis

Département d' Informatique et de Recherche Opérationnelle
Université de Montréal, C.P. 6128, Succ. Centre-Ville
Montréal, H3C 3J7, CANADA

ABSTRACT

We consider importance sampling (IS) to increase the efficiency of Monte Carlo integration, especially for pricing exotic options where the random input is multivariate Normal. When the importance function (the product of integrand and original density) is multimodal, determining a good IS density is a difficult task. We propose an Automated Importance Sampling DEnsity selection procedure (AISDE). AISDE selects an IS density as a mixture of multivariate Normal densities with modes at certain local maxima of the importance function. When the simulation input is multivariate Normal, we use principal component analysis to obtain a reduced-dimension, approximate importance function, which allows efficient identification of a good IS density via AISDE in original problem dimensions over 100. We present Monte Carlo experimental results on randomly generated option-pricing problems (including path-dependent options), demonstrating large and consistent efficiency improvement.

1 INTRODUCTION

Consider the problem of estimating the integral

$$\nu = \mathrm{E}_f[h(Z)] = \int h(z) f(z) dz < \infty, \qquad (1)$$

via Monte Carlo sampling for a *response function* $h : \mathcal{R}^d \to [0, \infty)$, where Z is a d-dimensional random vector with known density function f.

Importance sampling (IS) is known as a very effective method for reducing the variance (more generally, increasing the efficiency) of the Monte Carlo estimate of ν. Let g be any d-dimensional density that is positive on the support of f, i.e., $f(z) > 0 \Rightarrow g(z) > 0$. We write

$$\nu = \int h(z) \frac{f(z)}{g(z)}\, g(z) dz = \mathrm{E}_g\left[h(Z)\frac{f(Z)}{g(Z)}\right],$$

where E_g denotes expectation under the new density g. Monte Carlo estimation with importance sampling proceeds as follows:

1. Generate $\{Z_i\}_{i=1}^n$ as independent, identically distributed (i.i.d.) under g.
2. Calculate

$$\widehat{\nu}(g) = n^{-1} \sum_{i=1}^{n} \frac{(h \cdot f)(Z_i)}{g(Z_i)}.$$

We will refer to $\widehat{\nu}(g)$ as an *importance sampling estimator* of ν. The density g is called the *importance sampling density*. We call the product of integrand and original density, $(h \cdot f)(z) \equiv h(z) f(z)$, the *importance function*. Sampling from g may be more (or less) costly than sampling from f, which affects the *estimation efficiency*, defined as the inverse product of an estimator's variance times the associated computing cost.

The class of integration applications that motivated this work is pricing high-dimensional exotic options with option-pricing models where the stochastic factors are multivariate Normal. In exotic option pricing, when cast as an integration problem as in (1) with f being the multivariate standard normal density, the importance function may be multimodal and possibly have modal regions far from each other. In this setting, selecting a good IS density is a nontrivial problem. We give background on option pricing and review existing IS methods in this context in Section 2.

In this paper, we are interested in *automated* and *robust* methods for identifying an IS density. By "automated", we mean that no analytical manipulation of the integral is performed, except for the trivial rewriting of the integral to account for the choice of sampling density. By "robust", we loosely mean that efficiency improvement should be obtained over a wide class of response functions–in particular, including the case where the importance function is multimodal. IS density selection methods that fit this loose definition of robustness have been proposed in the

past, most notably in the context of Bayesian integration by Oh and Berger (1993) and West (1993). We discuss the difficulties that arise in implementing these procedures in the next section.

Our main contribution is an automated, general-purpose, and robust algorithmic framework for IS density selection. A specialized implementation within this framework is presented as a procedure for Automated Importance Sampling DEnsity selection (AISDE). AISDE delivers an IS density with multivariate Normal components centered at certain modes (local maxima) of the importance function $h \cdot f$. To identify the modes, AISDE performs repeated maximizations of $h \cdot f$ invoking a generic unconstrained optimization routine. Each maximization is initialized at a good point, determined on the basis of a random sample from a sampling density. When the random input is multivariate Normal, we use principal components analysis to obtain a reduced-dimension, approximate importance function, which allows efficient identification of a good IS density via AISDE in original problem dimensions over 100.

This paper is organized as follows. In Section 3 we review methods for identifying a good IS density, with focus on robust methods. In Section 4 we motivate and develop Procedure AISDE. Section 5 develops the dimension-reduction technique and the corresponding Monte Carlo estimation with importance sampling. In Section 6 we report results of a Monte Carlo study demonstrating the effectiveness of AISDE in the application of pricing high-dimensional exotic options. We summarize our findings and suggest some extensions in Section 7.

2 IMPORTANCE SAMPLING FOR PRICING EXOTIC OPTIONS

Let S_t^j, $j = 1, \ldots, k$ denote the time-t value of the k stochastic factors underlying the option. These factors may correspond directly to the price of a tradeable assets such a stocks, or, they may be pricing-model parameters such as an interest rate, forward rate, or stochastic volatility. The values of factors are monitored in discrete time over the set of monitoring times $t_i = iT/m, i = 0, \ldots, m$, equally spaced between time 0 and time T, where T is the calendar option expiration time. At time 0, the factor vector has known value $\mathbf{S}_0 = (s_0^1 s_0^2 \ldots s_0^k)$.

Let $\mathbf{S}_{t_i} = (S_{t_i}^1 \; S_{t_i}^2 \; \ldots S_{t_i}^k)$ denote the vector of all stochastic factors at time $t_i, i = 1, \ldots, m$. The *option payoff* is some nonnegative function $p(\cdot)$ applied to the set of all factor values $\mathbf{S}_{t_i}, i = 1, \ldots, m$. From arbitrage-pricing theory, the arbitrage-free price of the option is the expectation of $p(\cdot)$ with respect to a so-called *risk-neutral measure*. For a rigorous treatment of arbitrage pricing theory, see Duffie (1996) and Harrison and Pliska (1981); for an excellent and mathematically lighter treatment, see Baxter and Rennie (1996).

The prevailing class of pricing models postulates that the vector of stochastic factors follows k-dimensional Geometric Brownian motion. In these models, under the risk-neutral measure, we have

$$\mathbf{R}_{t_i} \equiv \ln(\mathbf{S}_{t_i} ./ \mathbf{S}_{t_{i-1}}) \sim N_k\left(\mu(t_i - t_{i-1}),\, \Sigma\sqrt{t_i - t_{i-1}}\right),$$

$$\mathbf{R}_{t_i}, i = 1, \ldots, m \text{ are independent,} \tag{2}$$

where "./" denotes element-wise division; "$\sim$" means "is distributed as"; $N_k(\mu, \Sigma)$ denotes the k-variate Normal distribution with mean μ and covariance matrix Σ; μ is the risk-neutral drift vector; Σ is the covariance matrix of factor log-returns over one time unit. (For brevity, we skip the details of determination of the drift vector μ under the risk-neutral measure.) We defer the remaining details of casting option pricing as in (1) to Section 6.

In exotic option pricing, the importance function may be multimodal and possibly have modal regions far from each other, making the determination of a good IS density a difficult task. This is typically the case for a call option on the maximum of $k > 1$ factors and for an outperformance option, which is a call option on the difference between two factors.

Notably, the IS density proposed in Glasserman et al. (1999) typically fails for such options, often substantially increasing the variance. Another IS density selection procedure that in our experience proved ineffective (the variance was roughly unchanged) is the algorithm by Lepage (1978) that appears in the classic *Numerical Recipes in C* by Press et al. (1992). Closer inspection of the reason of failure of these methods reveals that they are designed for relatively narrow classes of response functions and/or original densities f. In the case of Glasserman et al. (1999), the effectiveness of the density is shown under the assumption (roughly) that the logarithm of the importance function is a concave function on the support of the response function (the option payoff), which may be violated by exotic-option payoffs. The Lepage procedure is designed to identify the variance-minimizing *separable* IS density, i.e., a density that is the product of univariate densities.

3 IMPORTANCE SAMPLING WITH MULTIMODAL IMPORTANCE FUNCTIONS

We assume throughout this paper that $h(z) \geq 0$ and $\nu > 0$. Our entire development extends easily to arbitrary h by writing $\nu = \int h^+(z) f(z) dz - \int h^-(z) f(z) dz$, where h^+ and h^- are the positive and negative part of the integrand h, respectively, and then estimating separately the two integrals of nonnegative functions.

Importantly for simulation efficiency, the *variance-minimizing IS density* is

$$g^*(z) = \frac{1}{\nu} h(z) f(z) \tag{3}$$

since it would lead to a zero-variance estimator if it were possible to both sample from it and evaluate it in closed form. However, actually evaluating $g^*(Z)$ is clearly infeasible, as ν is the unknown quantity to be estimated. The message from (3) is that, to reduce variance, an IS density should approximate–as much as possible–the importance function.

Many of the IS density selection methods that have been developed are designed for a unimodal importance function. When the importance function is multimodal, there can be serious difficulties in finding a good density (van Dijk and Kloek 1980). We focus our review on two approaches for the multimodal case; both are motivated by integration (and more generally, inference) with Bayesian posterior distributions.

West (1993) proposes a kernel density estimation technique. Based on a sample from an appropriate density, the candidate IS density is a mixture of kernels (densities) centered at each of the sampled points. Kernel density estimation is extremely intensive computationally, as it involves by definition a number of density components equal to the sample size. To make the IS density practical to use, West proposes a heuristic procedure for iteratively collapsing pairs of the mixture components to a single component until the total number of components in the mixture is as small as deemed appropriate by the analyst.

Oh and Berger (1993) use as importance sampling density a mixture of multivariate t density functions in dimension d. Mixtures of t's have many attractive properties:

(a) They can represent very irregular forms of functions (van Dijk and Kloek 1980).
(b) They allow easy and fast random variate generation.
(c) They allow flexibility in controlling the tail behavior (thickness of tails) of the density.

The authors assume a capability to identify the *important modes* of the importance function. [They do not define precisely this notion. Loosely speaking, a mode is important if the function is large at the mode (or the integral is large at a region appropriately linked to the mode) relative to the other modes.] They choose the degrees of freedom for each t component based on application-specific considerations. Their procedure performs constrained continuous minimization of a Monte Carlo estimate of the squared coefficient of variation, where: (a) the components are initially centered at the known modes; and (b) the decision variables are the mixture weights, the mean vectors, and covariance matrices of all the components.

In implementing the Oh-Berger (OB) procedure, there is a key difficulty. Quoting the authors, "Note that we thus assume a capability to identify the modes (or at least the important modes) of the integrand. This can, of course, be a difficult task". (In our terminology, the integrand of Oh and Berger is the importance function, $h \cdot f$, the product of response function h times the original density f.) Beyond this difficulty, there is another important issue that must be addressed with respect to efficiency: if $h \cdot f$ has many modes, then even if it were feasible and computationally viable to identify all modes, the efficiency of a mixture IS density with too many components would suffer from the high cost of evaluating the IS density.

4 FRAMEWORK FOR AUTOMATED IMPORTANCE SAMPLING DENSITY SELECTION (AISDE)

In this paper, the candidate IS densities considered belong to the family

$$g(\cdot) = \sum_{i=1}^{m} \alpha_i \phi(\cdot; \theta_i) \tag{4}$$

where m is a positive integer; $\alpha_i, i = 1, \ldots, m$ are positive mixing weights such that $\sum_{i=1}^{m} \alpha_i = 1$; $\theta_i \in \mathcal{R}^d, i = 1, \ldots, m$; and $\phi_d(\cdot; \theta)$ is the density of the d-variate Normal distribution with mean θ and identity covariance matrix. This family is flexible in terms of location and weighing of the component densities, while being constrained to have unit covariance on each component. The choice of unit covariance is made to simplify our subsequent exposition, but is not restricting on our development. In view of the covariance restriction, unless otherwise stated, we will assume that the original density f is the product of univariate densities with unit variance.

Our approach to density selection is logically positioned before the OB procedure in the density selection process. We do not require a priori knowledge of any of the modes of the importance function $h \cdot f$ and focus on efficiently identifying modes that are important in reducing the variance.

To begin our development, we define the variance and the second moment under importance sampling, respectively, as functions of the IS density:

$$\sigma^2(g) = \nu_2(g) - \nu^2$$

where

$$\nu_2(g) \equiv \mathrm{E}_g\left[\left(h(Z)\frac{f(Z)}{g(Z)}\right)^2\right].$$

A mixture IS density g with many components is typically substantially costlier to evaluate than f. To model the efficiency of candidate IS densities, we use the follow-

ing simple model that captures the essential Monte Carlo computing cost components. Define the constants c_f, c_h, and c_ϕ as the expected per-replication computing cost of random-variate generation (i.e., sampling from f), evaluation of the response function h, and evaluation of ϕ, respectively. Writing $g(M)$ to explicitly denote the number of components of g, the efficiency of g relative to f is

$$\mathrm{Eff}(g(M)) = \frac{\sigma^2(f)}{\sigma^2(g(M))} \frac{c_f + c_h}{c_f + c_h + (M+1)c_\phi}. \quad (5)$$

In practice, good estimates of the computing-cost constants may be either a priori known or estimated dynamically during the density estimation itself. In the remainder, we assume these as known constants.

Let $\mathcal{M} = \{z_1, z_2, \ldots, z_{|\mathcal{M}|}\}$ be the set of all modes (local maxima) of $h \cdot f$. We will select an IS density by attempting to obtain a good solution to the optimization problem

$$\begin{aligned} \max \quad & \mathrm{Eff}(g(\mathcal{N})) && (6)\\ \text{s.t.} \quad & \mathcal{N} \subseteq \mathcal{M} && (7)\\ & g(\mathcal{N}) = \textstyle\sum_{j\in\mathcal{N}} \alpha_j \phi(z_j) && (8)\\ & z_j \in \mathcal{M}, \quad \text{each } j && (9)\\ & \alpha_j = \tfrac{(h\cdot f)(z_j)}{\sum_{\ell\in\mathcal{N}}(h\cdot f)(z_\ell)}. && (10) \end{aligned}$$

That is, the selection problem restricts attention to densities in the class (4), further restricted as follows:

- Constraint (8) says that each component ϕ of g is centered at (has mean) a mode of the importance function $h \cdot f$.
- Constraint (10) says that the mixture components are weighed in proportion to the value of the importance function at the corresponding mode.

Briefly, our approach to obtaining a good solution to (6) is as follows. In view of the constraints (8) and (9), it is necessary to identify some or all of the modes of $h \cdot f$. For this task, we simply use a standard off-the-shelf unconstrained optimization routine, say MAXIMIZE. The remaining work is to find a good $\mathcal{N}$.

There are two main considerations in the search for a good $\mathcal{N}$. The first one has to do with the impact of $\mathcal{N}$ to efficiency, as opposed to simply variance. In general, for $\mathcal{N}_1 \subseteq \mathcal{N}_2$, we expect $\mathrm{Var}(g(\mathcal{N}_2) \le \mathrm{Var}(g(\mathcal{N}_1)$, by arguing that a larger set of modes allows g more flexibility to approximate $h \cdot f$. However, the computing-cost component of efficiency decreases with $|\mathcal{N}|$, so we may have $\mathrm{Eff}(g(\mathcal{N}_2) \le \mathrm{Eff}(g(\mathcal{N}_1)$.

The second, more important consideration is the practical issue of controlling the maximization computing effort. We focus our discussion on the effects of problem dimension d. Actual integration via Monte Carlo requires $O(d)$ work, i.e., it is linear in problem dimension. In contrast, the work of MAXIMIZE is typically $O(d^3)$ or $O(d^4)$, when derivatives are user-provided or approximated via finite differences within MAXIMIZE, respectively. Thus, as the dimension d increases, invoking MAXIMIZE will be not practical. However, for moderate problem dimension (say ≤ 30), the work per call to MAXIMIZE may be quite small relative to the Monte Carlo total budget. With this case in mind, and considering that $h \cdot f$ may have many modes, the main consideration is to identify modes efficiently, namely:

- Identify a new mode with each call to MAXIMIZE.
- Identify earlier (rather than later) the modes with higher impact on reducing variance.

With these considerations, we propose Procedure AISDE (Automatic Importance Sampling Density Estimation). In the following paragraph, we summarize the key steps of AISDE, accompanied with motivating comments and discussion. A commented pseudocode of AISDE with full details is given in Figure 1.

1. Generate a sample $Z_i, i = 1, \ldots, n$, independent and identically distributed (i.i.d.) from a *sampling density*. For simplicity, use the sampling density f.
2. At iteration 0, initialize the candidate IS density g_0 to f. Note that for any density g constructed independent of the sample, an unbiased estimate of the second moment $\nu_2(g)$ is

$$\begin{aligned} \widehat{\nu_2(g)} &= n^{-1} \sum_{i=1}^{n} \left(\frac{h(Z_i) f(Z_i)}{g(Z_i)} \right)^2 \frac{g(Z_i)}{f(Z_i)} \\ &= n^{-1} \sum_{i=1}^{n} \frac{h^2(Z_i) f(Z_i)}{g(Z_i)}. \end{aligned}$$

3. (At the M-th update iteration, use the following notation: $\mathcal{N}_M$ is the set of all known modes of $h \cdot f$; and g_M is the candidate IS density.) Until a certain termination criterion is met, do:
   (a) Identify the sample point that contributes most to the estimate of the variance $\sigma^2(g_M)$,

$$i^* = \mathrm{argmax}_{1\le i\le n} \frac{h^2(Z_i) f(Z_i)}{g_M(Z_i)}.$$

   We can expect to most improve g_M by increasing it near Z_{i^*}. Thus the most promising region in which to look for undiscovered modes of $h \cdot f$ is near Z_{i^*}.

(b) Invoke MAXIMIZE to maximize $h \cdot f$, starting at the point Z_{i^*}:

$$z = \text{MAXIMIZE}(h \cdot f; \text{start at } Z_{i^*}).$$

If z is not in $\mathcal{N}_M$, then the M-th update is

i. $M = M + 1$.

ii. Add the mode z to the set of known modes $\mathcal{N}_M$;

iii. Update the IS density g_M as in (8)-(10) where $\mathcal{N} = \mathcal{N}_M$.

iv. Update estimates of the variance and efficiency of g_M.

4. Output density g_A as the one with maximum estimated efficiency over all update iterations.

4.1 Convergence and Statistical Properties of AISDE

We briefly discuss the asymptotic behavior of AISDE with respect to the sample size n and the number of maximizations n_O. Of course in practice AISDE will likely be most useful with modest n and small n_O. We need the notion of attraction set of a mode z_j, defined as the set $\mathcal{A}_j = \{z \in \mathcal{R}^d : \text{MAXIMIZE}(h \cdot f; \text{start at } z) \text{ returns } z_j\}$.

The maximum necessary value of n_O is n (in this case, each sampled point serves as a starting point for one maximization.) First consider a fixed sample size n. As n_O increases to n, the set of modes of the importance function identified by AISDE, $\mathcal{N}$, increases to a (possibly strict) subset of $\mathcal{M}$ (the strict case occurs if there exists a mode j of $h \cdot f$ such that none of the sampled points $\{Z_i\}_{i=1}^n$ belongs to $\mathcal{A}_j$. Now letting $n = n_O \to \infty$, under the mild assumption that the attraction set of each point in $\mathcal{M}$ has positive probability under f, $\mathcal{N}$ converges to $\mathcal{M}$, and the associated weights $\alpha_i, i = 1, \ldots, \mathcal{N}$ converge. In summary, g_A converges to a density whose components are in one-to-one correspondence with the modes of $h \cdot f$ as the computer budget allocated to AISDE grows appropriately.

Given that g_M was constructed explicitly to reduce the sample-based variance estimate at the previous M-iteration, S^2_{M-1}, we expect the variance estimate S^2_M to be biased low, i.e., underestimate the true variance $\sigma^2(g_M)$. Recall, however, that the reason for obtaining these estimates is to *compare* the successive candidate density variances and efficiencies, so the quantity we are implicitly estimating is the *difference (or ratio) of variances* across update iterations, which is expected to be less biased than the individual variance estimates. Our computational experience confirmed this negative bias. Ultimately, however, the really relevant quantity is the efficiency of g_A, evaluated empirically in Section 6.

Procedure AISDE (Automatic Importance Sampling DEnsity Selection)

$M = 0; \quad \mathcal{S} = \emptyset; \quad \mathcal{N}_0 = \emptyset$

1. Generate an i.i.d. sample of size n from density f
   $\{Z_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} f$
2. Calculate unbiased estimates of ν, $\sigma^2(f)$, and ν^2:
   $\widehat{\nu} = n^{-1} \sum_{i=1}^n h(Z_i)$
   $S^2 = (n-1)^{-1} \sum_{i=1}^n h^2(Z_i) - n\widehat{\nu}^2$
   $\widehat{\nu^2} = \widehat{\nu}^2 - n^{-1} S^2$.
3. Initialize the first candidate IS density
   $g_0(\cdot) = f(\cdot)$
4. Each sample point gives an estimate of $\nu_2(g_0)$
   $Y_i^{(0)} = \frac{h^2(Z_i) f(Z_i)}{g_M(Z_i)}, \quad i = 1, 2, \ldots, n$

While (Termination Condition)

5. Find the sample point that contributes most to $\widehat{\sigma^2(g_M)}$
   $i^* = \text{argmax}_i \{Y_i^{(M)} : i \notin \mathcal{S}\}$
   $\mathcal{S} = \mathcal{S} \cup \{i^*\}$
6. Maximize $h \cdot f$, starting at Z_{i^*}
   $z = \text{MAXIMIZE}(h \cdot f; \text{start at } Z_{i^*})$
7. Check if the maximizer is valid and new
   If (MAXIMIZE Converged and $z \notin \mathcal{N}_M$)
8. Update mode information
   $M = M + 1; \quad z_M = z; \quad \mathcal{N}_M = \mathcal{N}_{M-1} \cup \{z_M\}$
9. Set IS density mixture weights
   $\alpha_M = \frac{(h \cdot f)(z_M)}{\sum_{j=1}^M (h \cdot f)(z_j)}$
10. Update the IS density (evaluated at the sample points)
    $g_M(Z_i) = \alpha_M \phi(Z_i; z_M) + (1 - \alpha_M) g_{M-1}(Z_i), \quad i = 1, \ldots, n$
11. Update the individual-point estimates of $\nu_2(g_M)$
    $Y_i^{(M)} = \frac{h^2(Z_i) f(Z_i)}{g_M(Z_i)}, \quad i = 1, 2, \ldots, n$
12. Estimate the efficiency of g_M relative to f
    $S^2_M = n^{-1} \sum_{i=1}^n Y_i^{(M)} - \widehat{\nu^2}$
    $\widehat{\text{Eff}(g_M)} = (S^2 / S^2_M) \frac{c_f + c_h}{c_f + c_h + (M+1)c_\phi}$

    End If

End (Termination Condition)

13. Output density with maximum estimated efficiency
    $M^* = \text{argmax}_i \{\widehat{\text{Eff}(g_M)}, i = 1, \ldots, M\}$
    $g_A(\cdot) = \sum_{j=1}^{M^*} \alpha_j \phi(\cdot; z_j)$

Figure 1: AISDE Pseudocode

5 DIMENSION REDUCTION FOR IMPORTANCE SAMPLING VIA PRINCIPAL COMPONENT ANALYSIS

We restrict attention to the special case with simulation input $X \sim N_d(\mu_X, \Sigma_X)$, where "$\sim$" means "is distributed as"; N_d denotes the d-variate Normal distribution; μ_X is the mean vector; Σ_X is a nonsingular covariance matrix. The problem is to estimate $\nu = \mathrm{E}[h_1(X)]$ for a given function $h_1(\cdot)$. In this case, the importance function can often be well-approximated by lower-dimensional functions, thus enabling effective use of procedures such as AISDE requiring maximization in high dimension.

From standard linear algebra, there exists a $d \times d$ orthogonal matrix U such that

$$U'\Sigma_X U = \Lambda, \quad \Lambda = \mathrm{diag}(\lambda_1^2, \lambda_2^2, \ldots, \lambda_d^2)$$

where the "diag" notation means that Λ is a diagonal matrix whose diagonal elements are the argument of diag.

$$\lambda_1^2 \geq \lambda_2^2 \geq \ldots \geq \lambda_d^2 > 0.$$

The multivariate Normal distribution has the special property that any linear transformation of a multivariate Normal vector is also multivariate Normal. Thus X can be represented as

$$X = \mu_X + UY, \quad Y \sim N_d(0_d, \Lambda). \tag{11}$$

where 0_d is a d-vector of zeros. Since Y consists of d independent Normal random variables with decreasing variances, the parsimonious approach to approximating X is to restrict the transformation implied by U to the first $d_\mathrm{R} < d$ elements in Y. The reduced dimension d_R may be either selected directly or determined by selecting the "proportion of total variance to keep" via a parameter $0 < \delta \leq 1$. In the latter case, we take

$$d_\mathrm{R} = \min\left\{k : \sum_{j=1}^{k} \lambda_j^2 \geq \delta \sum_{j=1}^{d} \lambda_j^2\right\}$$

Partition $Y = (Y_\mathrm{K} \;\; Y_\mathrm{D})$, where Y_K and Y_D are the fist d_R and the remaining elements of Y, respectively. Let $\Lambda_\mathrm{K} = \mathrm{diag}(\lambda_1^2, \lambda_2^2, \ldots, \lambda_{d_\mathrm{R}}^2)$; $\Lambda_\mathrm{D} = \mathrm{diag}(\lambda_{d_\mathrm{R}+1}^2, \lambda_{d_\mathrm{R}+2}^2, \ldots, \lambda_d^2)$. Let U_K be the $d \times d_\mathrm{R}$ matrix formed by the first d_R columns of U; and let U_D be the $d \times (d - d_\mathrm{R})$ matrix formed by the last $d - d_\mathrm{R}$ columns of U.

The proposed approximation to X is

$$\tilde{X} = \mu_X + U_\mathrm{K}\tilde{Y_\mathrm{K}}, \quad \tilde{Y_\mathrm{K}} \sim N_{d_\mathrm{R}}(0_{d_\mathrm{R}}, \tilde{\Lambda}_\mathrm{K}), \tag{12}$$

where

$$\tilde{\Lambda}_\mathrm{K} = \rho\,\mathrm{diag}(\lambda_1^2, \lambda_2^2, \ldots, \lambda_{d_\mathrm{R}}^2), \quad \rho = \frac{\sum_{j=1}^{d} \lambda_j^2}{\sum_{j=1}^{d_\mathrm{R}} \lambda_j^2}.$$

The vector $\tilde{Y_\mathrm{K}}$ is conceptually equivalent to Y_K, except that the sum of variances of its elements is adjusted via the inflation factor $\rho \geq 1$ to equal the sum of variances of the full-dimensional input Y in (11). Note that $\tilde{X}$ has the original dimension d, but is generated as a linear transformation of the d_R-dimensional $\tilde{Y_\mathrm{K}}$.

Consider a hypothetical simulation (approximating the original simulation) where the random imput is $\tilde{Y_\mathrm{K}}$, of dimension d_R, and the output whose expectation is to be estimated is $h_1(\tilde{X})$. Corresponding to the approximating simulation is the *approximate importance function*

$$r(z) = h_1(\mu_X + U_\mathrm{K}\tilde{\Lambda}_\mathrm{K} z)\phi_{d_\mathrm{R}}(z). \quad z \in \mathcal{R}^{d_\mathrm{R}} \tag{13}$$

where ϕ_{d_R} is the d_R-dimensional standard Normal density. It is important that there is flexibility in choosing d_R, with the obvious tradeoff that as d_R is reduced, the accuracy of (12) as an approximation to (11) will deteriorate.

Based on this flexible development of an approximate, lower-dimensional importance function, we propose that Monte Carlo estimation of ν via importance sampling can proceed in two steps. Step 1 is to determine an IS density for the reduced-dimension, approximate importance function in (13). Procedure AISDE may be used in this step to obtain an IS density g for sampling Y_K. This reduction of dimension is crucial, in view of the very fast (cubic or quartic) growth of the MAXIMIZE work with problem dimension. Step 2 is estimation of ν via Monte Carlo with importance sampling as follows. X is sampled according to the exact representation (11), where the elements of Y_K are sampled via a new density g; the elements of Y_D are sampled via the original density ϕ_{d-d_R}. The exact procedure for step 2 is listed in Figure 2. We name this procedure *Importance Sampling on Selected Principal Components* (ISSPC).

Proposition 1. *For any density g strictly positive on $(-\infty, \infty)^{d_\mathrm{R}}$, (15) is an unbiased estimate of* $\mathrm{E}[h_1(X)]$.

Proof. The representation (11) is equivalent to the representation

$$X = \mu_X + U\Lambda Z, \quad Z \sim N_d(0_d, I_d),$$

where I_d is the identity matrix of dimension d. Under the original Z-density ϕ_d, $Z_\mathrm{K} \sim \phi_{d_\mathrm{R}}$, $Z_\mathrm{D} \sim \phi_{d-d_\mathrm{R}}$, and $Z_\mathrm{K}, Z_\mathrm{D}$ are independent. In view of the sampling of Z_K and Z_D in (14), the likelihood ratio equals

$$\frac{\phi_{d_\mathrm{R}}(Z_\mathrm{K})\phi_{d-d_\mathrm{R}}(Z_\mathrm{D})}{g(Z_\mathrm{K})\phi_{d-d_\mathrm{R}}(Z_\mathrm{D})},$$

Importance Sampling on Selected Principal Components

INPUTS:

- $X \sim N_d(\mu_X, \Sigma_X)$; μ_X is the mean vector Σ_X is a nonsingular covariance matrix;
- IS density g (d_{R}-dimensional)
- $\phi_{d-d_{\mathrm{R}}}$ is the $(d - d_{\mathrm{R}})$-variate standard Normal density

GOAL: Estimate $\nu = \mathrm{E}[h_1(X)]$.

1. Generate

$$Z_{\mathrm{K}} \sim g, \quad Z_{\mathrm{D}} \sim \phi_{d-d_{\mathrm{R}}}; \quad Z_{\mathrm{K}}, Z_{\mathrm{D}} \text{ independent} \tag{14}$$

$$Y_{\mathrm{K}} = \Lambda_{\mathrm{K}} Z_{\mathrm{K}}; \quad Y_{\mathrm{D}} = \Lambda_{\mathrm{D}} Z_{\mathrm{D}}$$

2. Set

$$X = \mu_X + U_{\mathrm{K}} Y_{\mathrm{K}} + U_{\mathrm{D}} Y_{\mathrm{D}}$$

3. Evaluate

$$h_1(X) \frac{\phi_{d_{\mathrm{R}}}(Z_{\mathrm{K}})}{g(Z_{\mathrm{K}})} \tag{15}$$

Figure 2: Importance Sampling on Selected Principal Components

completing the proof. ∎

6 APPLICATION TO EXOTIC OPTION PRICING AND EXPERIMENTAL EVALUATION

We focus on the application of the techniques of Sections 4 and 5 to pricing exotic options. Recalling (2), and defining the vector of all log-returns $X = (\mathbf{R}_{t_1}\mathbf{R}_{t_2} \ldots \mathbf{R}_{t_m})$, we have that X is d-variate Normal, where $d = km$, with known mean and covariance matrix. Observe that the vector of factor prices is recovered from X as

$$\mathbf{S}_{t_i} = \mathbf{S}_0 . * \exp\left(\sum_{j=1}^{i} \mathbf{R}_{t_j}\right).$$

where ".*" denotes element-wise multiplication.

We report experimental results for two exotic option payoff functions. The payoff of a call option on the maximum of the path-wise arithmetic average of factors is

$$p_1(X) = \left(\max_k \left\{ m^{-1} \sum_{i=1}^{m} S_{t_i}^k \right\} - K\right)^+$$

where $x^+ \equiv \max(x, 0)$; K is the *strike price*. The payoff of a call option on the maximum of factors at expiration with a down-and-out barrier on the minimum of factors is

$$p_2(X) = (\max_k S_T^k - K)^+ 1\left\{\min_k S_{t_i}^k \geq b, i = 1, \ldots, m\right\}$$

where b is the *barrier* value. A special feature is that the corresponding importance functions are multimodal and thus require techniques such as the ones developed in this paper. Recalling the discussion of arbitrage pricing in the second paragraph of Section 2, the option price is $c_i = \mathrm{E}[p_i(X)]$, $i = 1, 2$.

We tested the robustness of AISDE against problem instances that were generated randomly as follows. The parameters μ and Σ in (2) were: Σ is a diagonal matrix, with diagonal equal to $(0.1e_k + 0.7U)^2$, where e_k is a k-vector of ones; U is a k-vector uniformly distributed on $(0, 1)^k$; the squaring of the vector in parentheses is element-wise; $\mu = -0.05e_k - \frac{1}{2}\mathrm{diag}(\Sigma)$, where $\mathrm{diag}(\Sigma)$ is the k-vector of diagonal elements of Σ. The factor vector at time 0 was $\mathbf{S}_0 = 60e_k + 30V$, where V is a k-vector uniformly distributed on $(0, 1)^k$. For the barrier, we took $b = 30$. The strike price K was set subsequently by increasing K in small amounts until the coefficient of variation (CV) of $p_i(X)$ exceeded 5. The large target CV value of 5 aims to set up problem instances such that importance sampling is most needed. The option expiration time was $T = 1$ year. The number of monitoring times was $m = 10$.

AISDE was implemented as follows. We took $n = 10000$. The termination condition in the While statement was $\widehat{\mathrm{Eff}(g_M)} < \widehat{\mathrm{Eff}(g_{M-1})}$. The dimension reduction technique was implemented by setting $\delta = 0.9$, i.e., we used as many principal components as necessary to "cover" 90% of the total problem variance.

Performance measures were estimated as follows. The variance ratio, $\mathrm{VR} = \sigma^2(f)/\sigma^2(g_{\mathrm{A}})$, was estimated as the ratio of sample variances based on 10 independent macroreplications of the standard and IS estimate, where each of the latter estimates was the average over 32000 independent replications. The efficiency ratio, ER, was estimated as in (5), where the variance ratio was estimated as we just discussed, and the constants c_f, c_h, and c_ϕ were estimated to sufficient accuracy for the computing platform MATLAB on which all experiments were performed.

Importance sampling estimation of the option price based on a first-stage density estimation via AISDE has two computing cost components: (1) The one-time cost of IS density estimation, measured here by T_{A}, the CPU time

consumed by AISDE; (2) Just as with standard Monte Carlo, the cost per simulation replication. Let γ denote the target confidence level of a Monte-Carlo based confidence interval

of the option price. To put the cost (1) in perspective, define the *break-even relative accuracy*

$$\mathrm{BRA}(T_\mathrm{A}, \gamma) = x_{1-\gamma/2}\mathrm{CV}\sqrt{\frac{\mathrm{ER}}{\mathrm{ER}-1}\frac{c_f + c_h}{T_\mathrm{A}}}$$

where $x_{1-\gamma/2}$ is the $(1-\gamma/2)$-quantile of the standard Normal distribution. It is easy to check that BRA has the following property for any given confidence level γ: BRA is the minimum ratio (CI half-width/option price) such that standard Monte Carlo requires less CPU time than AISDE-based importance sampling.

Tables 1 and 2 contain results for the option payoffs p_1 and p_2, respectively. In each table, there are three panels corresponding to different values of the number of factors k. In each panel, each of the first 5 rows corresponds to one randomly generated problem instance and is correspondingly numbered under the column labeled "Probl #"; the 6th row labeled "AVG" gives the geometric average of the variance ratio and efficiency ratio over the 5 problem instances. For each problem instance, we report: the reduced dimension d_R; the number of modes M^* of the selected IS density g_A; the estimated variance ratio VR; the estimated efficiency ratio, ER; and the break-even relative accuracy $\mathrm{BRA}(T_\mathrm{A}, 95\%)$.

Clearly density g_A yields large and consistent efficiency improvement. In addition, the large values of the break-even relative accuracy, BRA, indicate that the CPU cost of AISDE, T_A, is justified by the efficiency improvement. For example, in Table 1, panel 1, Problem 1, unless the user is satisfied with a ratio (CI half-width/option price) > 0.173, AISDE-based importance sampling is preferred to standard Monte Carlo.

In a larger set of experiments than the one reported here, we tested AISDE on a wider selection of payoff functions and in original problem dimension up to 140 (7 factors, 20 monitoring times). This experimental evaluation over random problem instances demonstrated that AISDE is powerful and robust–it yielded large efficiency improvement in a very large percentage of randomly generated problems.

Table 1: Performance of AISDE-Based Importance Sampling for Option Payoff p_1; $k = 3, 5, 7$; $m = 10$

k	Probl #	d_R	M^*	VR	ER	BRA
	1	5	3	35.7	30.9	0.173
	2	4	3	46.0	40.1	0.182
3	3	5	3	37.2	32.2	0.134
	4	4	2	48.0	43.2	0.240
	5	4	2	76.5	68.9	0.262
	AVG			46.8	41.2	
	1	7	3	47.8	43.6	0.169
	2	8	5	32.0	27.9	0.146
5	3	7	5	37.5	32.8	0.151
	4	7	5	40.6	35.5	0.144
	5	7	5	33.8	29.5	0.135
	AVG			37.9	33.4	
	1	10	6	36.7	30.6	0.117
	2	10	7	56.3	45.9	0.170
7	3	10	7	58.0	47.3	0.252
	4	10	2	297.4	274.2	0.658
	5	10	6	98.1	81.9	0.347
	AVG			81.0	68.4	

Table 2: Performance of AISDE-Based Importance Sampling for Option Payoff p_2; $k = 3, 5, 7$; $m = 10$

k	Probl #	d_R	M^*	VR	ER	BRA
	1	4	3	14.4	13.1	0.235
	2	5	3	15.0	13.5	0.171
3	3	4	3	19.7	17.9	0.193
	4	4	2	15.4	14.3	0.111
	5	4	3	15.6	14.2	0.158
	AVG			15.9	14.5	
	1	7	5	10.4	9.3	0.196
	2	7	5	12.7	11.3	0.217
5	3	7	5	11.5	10.2	0.120
	4	7	5	12.1	10.8	0.282
	5	7	5	11.0	9.8	0.139
	AVG			11.5	10.3	
	1	10	7	5.0	4.3	0.174
	2	10	7	7.5	6.5	0.178
7	3	10	7	7.8	6.8	0.163
	4	10	7	5.5	4.8	0.150
	5	10	7	8.6	7.5	0.140
	AVG			6.7	5.9	

7 CONCLUSION AND SUGGESTIONS FOR FUTURE WORK

Both the Oh-Berger (OB) and our new Procedure AISDE appear to be robust, i.e., yield efficiency improvement with a consistency that is hard to "break". The power of these procedures stems from the flexibility of mixtures of normal or t densities in approximating the many types of importance functions that may be encountered in applications. Our contribution is towards automating and making computationally efficient the task of locating the mixture components at modes of the importance function, a central task that does not appear to have been previously addressed. There is a natural combination of our Procedure AISDE and procedures such as OB, namely: first apply AISDE to obtain a candidate density g_A and then apply OB to improve g_A

via optimization over the component weights, modes, and covariance matrices.

The idea of dimension reduction via principal component analysis significantly increases the range of problem dimensions that can be addressed effectively via AISDE. Unfortunately, this development leverages special properties of the Multivariate Normal distribution and does not immediately extend to other distributions.

The impressive performance of AISDE in our experiments in option pricing does not of course guarantee efficiency improvement in a given new integration application. It would be interesting to study experimentally the properties of integration problems that may "break" the observed robustness of AISDE. Such properties include the heaviness of tails of the original density and perhaps more pathological response functions than the ones we have encountered.

REFERENCES

Baxter, M., and A. Rennie. 1996. *Financial calculus*. Cambridge: Cambridge University Press.

Duffie, D. 1996. *Dynamic asset pricing theory*. 2d ed. Princeton, New Jersey: Princeton University Press.

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

Harrison, M., and S. Pliska. 1981. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications* 11:215–260.

Lepage, G. P. 1978. A new algorithm for adaptive multidimensional integration. *Journal of Computational Physics* 27 (2):192–203.

Oh, M. S., and J. O. Berger. 1993. Integration of multimodal functions by monte carlo importance sampling. *Journal of the American Statistical Association* 88:450–456.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. *Numerical recipes in c: The art of scientific computing*. 3d ed. Cambridge: Cambridge University Press.

van Dijk, H. J., and T. Kloek. 1980. Further experience in bayesian analysis using monte carlo integration. *Journal of Econometrics* 14:307–328.

West, M. 1993. Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society B* 55 (2):409–422.

AUTHOR BIOGRAPHY

ATHANASSIOS (THANOS) N. AVRAMIDIS is an Invited Researcher at the "Département d' Informatique et de Recherche Opérationnelle" at the University of Montreal, Canada. He has been on the faculty at Cornell University and a consultant with SABRE Decision Technologies. His recent research interests are in Monte Carlo simulation methods with applications to derivative valuation, hedging, and risk management. He holds M.S. and Ph.D. degrees from the School of Industrial Engineering at Purdue University and a diploma in Mechanical Engineering from the University of Thessaloniki in Greece. He can be reached via e-mail at <avramidis99@yahoo.com>.