Combining Latin hypercube sampling with other variance reduction techniques

N. Packham[*†]

October 21, 2014

Abstract

We consider the problem of reducing the variance of Monte Carlo estimators of high-dimensional estimation problems by combining the variance reduction techniques Latin hypercube sampling with dependence (LHSD), control variates and importance sampling. Under some standard conditions, the resulting estimators are consistent and asymptotically unbiased, and a central limit theorem holds. The effectiveness of the combined variance reduction methods is investigated by pricing an Asian basket call option. When comparing the effectiveness with existing combined variance reduction techniques, it turns out that techniques highly tailored to the specific problem are more effective, but among the methods that make no use of specific information, LHSD performs best. Since LHSD is easy to apply, our results indicate that it is worthwhile to test the performance pickup of integrating LHSD into arbitrary simulation problems even when a variance reduction technique is already in place.

JEL Classification: C15, C63, G13

Keywords: Monte Carlo simulation, variance reduction, Latin hypercube sampling (with dependence), control variates, importance sampling, derivative pricing

1 Introduction

The variance of a Monte Carlo estimator is a key figure for assessing the quality of an estimation, as a lower variance indicates a higher precision of the estimation. In particular for multivariate estimations – for example, when a financial derivative to be valued depends on several underlyings or values of the underlying at different time points – variance reduction techniques are indispensable to obtain a feasible result within an acceptable computing time. Monte Carlo simulation techniques are ubiquitous in many areas such as the valuation of financial derivatives, and different techniques to reduce the variance of an estimator have been developed, see e.g. Glasserman (2004).

We consider the problem of reducing the variance of a Monte Carlo estimator targeted at a random vector of dependent random variables by combining several variance reduction techniques. More specifically, we consider Latin hypercube sampling with dependence (LHSD) as developed in

[*]Department of Finance, Frankfurt School of Finance & Management, Sonnemannstr. 9-11, 60314 Frankfurt, Germany , Phone: +49 69 154008 723, Fax: +49 69 154008 4723, Email: n.packham@fs.de

[†]I am grateful to Dr. Mohammad Majidi, d-fine GmbH, for helpful comments.

Packham and Schmidt (2010), combined with control variates and importance sampling. LHSD – a generalization of Latin hypercube sampling (LHS) introduced by McKay, Beckman, and Conover (1979) – is a general technique that can be applied to any Monte Carlo estimator, while both control variates and importance sampling typically exploit problem-specific properties. Both LHS and LHSD can be viewed as multivariate adaptations of stratified sampling. Loosely speaking, these methods ensure that a sample of random numbers is well spread across its domain. They are simple to use and often achieve an effective variance reduction. The general procedure for combined variance reduction is to introduce an LHS / LHSD step just after drawing random numbers and then proceed with the other variance reduction technique in the usual way.

In this paper, we first derive consistency, asymptotic unbiasedness and a central limit theorem for the resulting combined estimators, under some standard conditions.

Second, we investigate the effectiveness of the resulting estimators by valuing an Asian basket call option. The payoff of this option depends on the stock prices of a basket of stocks monitored at several time points, and as such, is a high-dimensional problem for which no analytic solution is known. We find that the combination of LHSD and control variates is effective for options with low strikes, whereas LHSD combined with importance sampling is effective for options with a high strike. The latter observation can be explained by the fact that a high proportion of simulation outcomes of such an option is zero. Importance sampling decreases the likelihood of sampling from the out-of-the-money region, and LHSD then operates mainly on the in-the-money region. The variance reduction ratios achieved in importance sampling depend largely on the problem-at-hand and on the actual measure change chosen. General methods of importance sampling, such as kernel density estimators, as in Zhang (1996), are computationally very expensive, and thus applicable only when the individual simulations are very time-intensive.

In a second example we are concerned with the valuation of a (one-asset) Asian option. The example is taken from Glasserman, Heidelberger, and Shahabuddin (1999), who analyse the effectiveness of combinations of importance sampling and stratification. Both, the application of importance sampling and stratification are highly tailored to the problem, and as such outperform LHSD. In Lemieux and La (2005), the same example is used to compare importance sampling combined with Quasi-Monte Carlo (QMC) and Randomized Quasi-Monte Carlo (RQMC) methods. In QMC simulation, well-chosen deterministic numbers are used in place of random number in order to increase the accuracy of the estimation problem, see e.g. Niederreiter (1992). Similar to LHSD, (R)QMC methods do not make use of problem-specific properties. In our examples, LHSD outperforms the (R)QMC methods most of the time.

Our results show that LHSD with importance sampling is two to three times more effective (in terms of standard deviation) than standard LHS with importance sampling. Whilst it is impossible to say anything general about the performance pickup from incorporating LHS / LHSD in simulation problems, a particular appeal of these methods lies in their generality and simplicity, so that one can easily test their effectiveness on a case-by-case basis.

The paper is structured as follows: in Section 2 we explain the general principle of variance reduction and we introduce the variance reduction techniques that are used in the paper, namely Latin hypercube sampling (LHS, LHSD), control variates and importance sampling. Section 3 contains the theoretical analysis of the resulting combined estimators (consistency and central limit theorem). Section 4 demonstrates and analyses the effectiveness of the combined estimators. Finally, we conclude in Section 5.

2 Variance reduction techniques under consideration

Reducing the variance of a Monte Carlo estimator allows for increasing the efficiency of a simulation by either saving computation time or achieving a more accurate result. This is especially important for multivariate problems, as the computation time of a simulation grows with the number of dimensions and quickly becomes intractable for high dimensions.

More specifically, on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, let $(X^1, \ldots, X^d)$ be a random vector with distribution function F. Suppose the goal is to estimate $\mathbb{E}h(X^1, \ldots, X^d)$ with $h : \mathbb{R}^d \to \mathbb{R}$ Borel-measurable and F-integrable. The usual Monte Carlo estimator based on n independent samples $(X_i^1, \ldots, X_i^d)$, $i = 1, \ldots, n$ is $1/n \sum_{i=1}^n h(X_i^1, \ldots, X_i^d)$. The estimator is strongly consistent, that is,

$$\frac{1}{n} \sum_{i=1}^{n} h(X_i^1, \ldots, X_i^d) \stackrel{\mathbf{P}\text{–a.s.}}{\to} \mathbb{E}h(X^1, \ldots, X^d), \quad \text{as } n \to \infty.$$

The usual central limit theorem states that the scaled estimator converges to a normal distribution, that is,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(h(X_i^1, \ldots, X_i^d) - \mathbb{E}h(X^1, \ldots, X^d) \Big) \stackrel{\mathcal{L}}{\to} \mathrm{N}(0, \sigma^2),$$

with $\sigma^2 = \mathrm{Var}(h(X^1, \ldots, X^d))$. The central limit theorem serves as an indicator of the speed of convergence in the sense that the estimator is approximately normally distributed, that is,

$$\frac{1}{n} \sum_{i=1}^{n} \Big(h(X_i^1, \ldots, X_i^d) - \mathbb{E}h(X^1, \ldots, X^d) \Big) \approx Z, \tag{1}$$

where $Z \sim \mathrm{N}(0, \sigma^2/n)$. The goal of variance reduction is to find consistent estimators with convergence properties similar to Equation (1), but whose variance is smaller than σ^2 above.

2.1 Latin hypercube sampling with dependence (LHSD)

Latin hypercube sampling (LHS) is a well-studied variance reduction technique for random vectors with independent components, see McKay et al. (1979), Stein (1987), Owen (1992), (Glasserman, 2004, Section 4.4). The concept of LHS is generalised to dependent random vectors, called LHSD, by Packham and Schmidt (2010). The method is also briefly considered by Stein (1987).

The principal idea of LHS and LHSD is to ensure a uniform placement of (uniform) random numbers on the d-dimensional unit cube. This is achieved by partitioning the $[0, 1]$ interval into so-called *strata* of equal probability and by transforming the random numbers in such a way that in each dimension, exactly one sample is placed into each stratum. In LHS, the strata in each dimension are chosen randomly, whereas in LHSD they are chosen according to the samples' rank statistics.

More specifically, let $U^1, \ldots, U^d$ be uniform random variables on $[0, 1]$, and assume that the goal is to estimate $\mathbb{E}g(U^1, \ldots, U^d)$ with $g : [0, 1]^d \to \mathbb{R}$ Borel-measurable and C-integrable, where C denotes the copula of $(U^1, \ldots, U^d)$.[1] Suppose further given n independent samples $(U_i^1, \ldots, U_i^d)$, $i = 1, \ldots, n$, with the same distribution as $(U^1, \ldots, U^d)$.

[1] A copula is the distribution function of a random vector with uniform marginals, see e.g. Joe (1997) and Nelsen (1999).

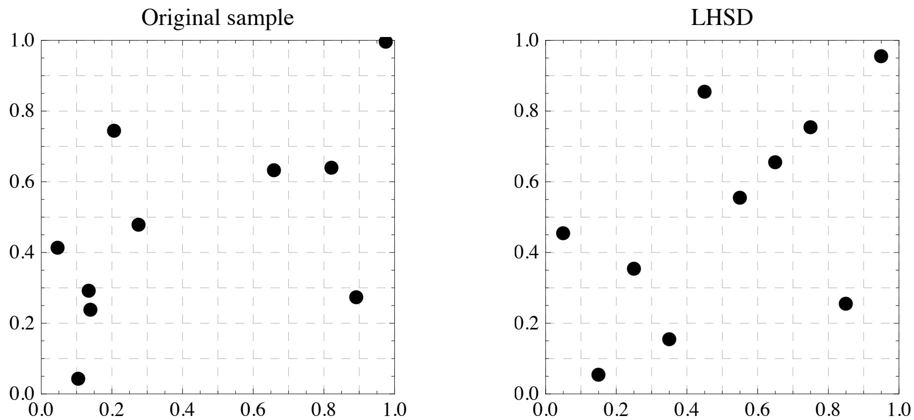

Figure 1: Left: Original sample $(U_1^1, U_1^2), \ldots, (U_{10}^1, U_{10}^2)$ linked with a Gaussian copula with correlation $\rho = 1/2$. Right: Corresponding LHSD sample $(V_{1,10}^1, V_{1,10}^2), \ldots, (V_{10,10}^1, V_{10,10}^2)$.

If C is the independence copula (i.e., the components of $(U^1, \ldots, U^d)$ are independent), then LHS transforms the samples according to $\dfrac{\pi_i^j - 1 + U_i^j}{n}$, $i = 1, \ldots, n$, $j = 1, \ldots, d$, where $\pi^j = (\pi_1^j, \ldots, \pi_n^j)$ is a random permutation of $\{1, \ldots, n\}$ drawn from the distribution that assigns equal probability to all permutations.

Now assume that $(U^1, \ldots, U^d)$ are not independent. The principal idea of LHSD is to preserve the dependence as much as possible by choosing a particular permutation for the transformation. This requires notion of a rank statistic. Fixing a dimension j, the i-th rank statistic $r_{i,n}^j$ denotes the position of the i-th sample when ordering $(U_1^j, \ldots, U_n^j)$ in ascending order. Formally,

$$r_{i,n}^j := \sum_{k=1}^n \mathbf{1}_{\{U_k^j \leq U_i^j\}}.$$

The existence $\mathbf{P}$–a.s. of such an ordering follows from the continuity of the uniform distribution. Intuitively, for large n, we expect the proportion of samples smaller than U_i^j to correspond approximately to $r_{i,n}^j/n$. Furthermore, $r_{i,n}^j/n$ is just the empirical distribution function of $U_1^j, \ldots, U_n^j$ evaluated at U_i^j, which is well-known to converge $\mathbf{P}$–a.s. to U_i^j, as $n \to \infty$, by the Glivenko-Cantelli Theorem.

A Latin hypercube sample (with dependence) can then be defined as

$$V_{i,n}^j := \frac{r_{i,n}^j - 1 + 1/2}{n}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, d. \tag{2}$$

This particular variant places each sample in the middle of its stratum; for other variants we refer to Packham and Schmidt (2010). In each dimension, exactly one sample is placed in each stratum, which enforces that the marginals are uniformly spread on the $[0, 1]$ interval. At the same time, since each sample converges to its original value as $n \to \infty$, the dependence structure of the samples is protected for sufficiently large n. An example of a Latin hypercube sample is given in Figure 1. The LHSD estimator is given by

$$\frac{1}{n}\sum_{i=1}^{n} g(V_{i,n}^1,\dots,V_{i,n}^d).$$

In Packham and Schmidt (2010), it is shown that for g bounded and continuous C-a.e., the LHSD estimator is strongly consistent and asymptotically unbiased, and, under some further restrictions, a central limit theorem is proved. It is further demonstrated that LHSD achieves a variance reduction in practical applications.

Since LHSD is easy to apply and does not depend on the specific problem at hand, it is natural to ask, if a further pickup in variance reduction is achieved when combining LHSD with other variance reduction techniques. Combining LHSD with other variance reduction techniques is achieved by using an LHSD sample $(V_{i,n}^1,\dots,V_{i,n}^d)$, $i = 1,\dots,n$, instead of the random numbers $(U_i^1,\dots,U_i^d)$, $i = 1,\dots,n$, as inputs.

When dealing with dependent random vectors, one can also apply LHS by first applying the LHS transform to independent random numbers and then introducing the dependence. However, this typically affects the stratification of the samples, losing much of the appeal of LHS. Our results indicate that it is generally better to use LHSD, which stratifies after introducing the dependence.

2.1.1 LHSD and random vectors with arbitrary marginals

The analysis so far is based on random vectors with uniform marginals. Let us now provide the link between random vectors with uniform marginals and vectors with non-uniform marginals.

Consider a random vector $X = (X^1,\dots,X^d)$ with marginal distribution functions $F_1,\dots,F_d$ and joint distribution function F. Define the right-inverse of F_i as the function $F_i^{(-1)} : [0,1] \to \mathbb{R}\cup\{\pm\infty\}$ with $F_i^{(-1)}(u) := \inf\{x : F(x) > u\}$, for $u \in [0,1]$.

Then, for a Borel-measurable function $h : \mathbb{R}^d \to \mathbb{R}$, there exists a Borel-measurable function $g : [0,1]^d \to \mathbb{R}$ such that

$$h(X^1,\dots,X^d) \stackrel{\mathcal{L}}{=} h(F_1^{(-1)}(U^1),\dots,F_d^{(-1)}(U^d)) =: g(U^1,\dots,U^d), \tag{3}$$

where the joint distribution of $(U^1,\dots,U^d)$ is determined by the copula corresponding to F and $F_1,\dots,F_d$. Equation (3) is essentially a consequence of Sklar's Theorem, Sklar (1959), which establishes the relationship between copulas and distribution functions.

For the results in Section 3 we note that whenever $F_1,\dots,F_d$ are continuous and h is F-a.e. continuous, then g is C-a.e. continuous, and that whenever h is right-continuous and of bounded variation (in the sense of Hardy-Krause), then g is right-continuous and of bounded variation (see Section 3 for the definitions of bounded variation and right-continuity for multivariate functions).

2.2 Control variates

The principal idea of the method of control variates is to use information about the errors in estimates of known quantities to reduce the error in the estimate of an unknown quantity. For a detailed treatment see e.g. (Glasserman, 2004, Section 4.1).

Let $X_1,\dots,X_n$ be the output of a Monte Carlo simulation, where each X_i follows the distribution of X. Assume further that another output Y_i with the same distribution as Y is calculated along with X_i and that the expectation $\mathbb{E}Y$ of Y is known. Then the sample based on the control variates $Y_1,\dots,Y_n$ is given by

$$X_i(b) = X_i - b(Y_i - \mathbb{E}Y), \quad i = 1,\dots,n,$$

where $b \in \mathbb{R}$. The control variates estimator is given by

$$\bar{X}(b) = \bar{X} - b(\bar{Y} - \mathbb{E}Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - b(Y_i - \mathbb{E}Y)). \tag{4}$$

It is easily observed that the control variates estimator is unbiased and consistent. The estimator's variance is $\sigma^2(b)/n$, with

$$\sigma^2(b) := \mathrm{Var}(X_i(b)) = \sigma_X^2 - 2b\sigma_Y\sigma_X\rho_{XY} + b^2\sigma_Y^2, \tag{5}$$

where $\sigma_X^2 = \mathrm{Var}(X)$, $\sigma_Y^2 = \mathrm{Var}(Y)$ and ρ_{XY} denotes the correlation between X and Y. The optimal coefficient b^* that minimises Equation (5) is given by $b^* = \rho_{XY}\sigma_X/\sigma_Y$, yielding a variance of the estimator Equation (4) of

$$\mathrm{Var}(\bar{X}(b)) = \frac{\sigma_X^2}{n}(1 - \rho_{XY}^2),$$

which is smaller than the variance σ_X^2/n of the usual Monte Carlo estimator $\bar{X} = 1/n\sum_{i=1}^{n} X_i$.

In practice, ρ_{XY} and σ_X, and hence the coefficient b^*, are often unknown, in which case one can use the following estimate of the coefficient based on the sample estimates of ρ_{XY} and σ_X:

$$\hat{b}_n = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}. \tag{6}$$

To estimate $\mathbb{E}g(U^1, \ldots, U^d)$ as introduced in Section 2.1 and with $f(U^1, \ldots, U^d)$ as the control variate, that is, setting $X = g(U^1, \ldots, U^d)$ and $Y = f(U^1, \ldots, U^d)$, the combined estimator LHSD with control variates is given by

$$\frac{1}{n}\sum_{i=1}^{n}\left(g(V_{i,n}^1, \ldots, V_{i,n}^d) - b(f(V_{i,n}^1, \ldots, V_{i,n}^d) - \mathbb{E}f(U^1, \ldots, U^d))\right). \tag{7}$$

2.3 Importance sampling

The key idea of importance sampling is to reduce variance by changing the probability measure from which random elements are generated by assigning more weight to "important" outcomes. For a detailed treatment see e.g. (Glasserman, 2004, Section 4.6).

Suppose for simplicity that under $\mathbf{P}$, $X = (X^1, \ldots, X^d)$ has a density $f^{\mathbf{P}}$ and let $f^{\mathbf{Q}}$ be the density for X under another measure $\mathbf{Q}$ such that $f^{\mathbf{P}}(x) > 0$ implies $f^{\mathbf{Q}}(x) > 0$, for $x \in \mathbb{R}^d$.

Then,

$$\mathbb{E}^{\mathbf{P}}[h(X)] = \int h(x)\,\frac{f^{\mathbf{P}}(x)}{f^{\mathbf{Q}}(x)}\,f^{\mathbf{Q}}(x)\,\mathrm{d}x = \mathbb{E}^{\mathbf{Q}}\left[h(X)\,\frac{f^{\mathbf{P}}(X)}{f^{\mathbf{Q}}(X)}\right],$$

where $\mathbb{E}^{\mathbf{P}}$ and $\mathbb{E}^{\mathbf{Q}}$ express that expectation is taken with respect to the measures $\mathbf{P}$ and $\mathbf{Q}$, respectively. The ratio $\dfrac{f^{\mathbf{P}}(X)}{f^{\mathbf{Q}}(X)}$ is the so-called *likelihood ratio*. The corresponding importance sampling estimator based on iid draws $X_i = (X_i^1, \ldots, X_i^d)$ of X (from $\mathbf{Q}$), is given by

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i)\frac{f^{\mathbf{P}}(X_i)}{f^{\mathbf{Q}}(X_i)}. \tag{8}$$

It is easily seen that the estimator is unbiased and strongly consistent. The estimator, when combining LHSD and importance sampling, is

$$\frac{1}{n}\sum_{i=1}^{n} g(V_{i,n}^1,\dots,V_{i,n}^d)\,\frac{f^{\mathbf{P}}(F_{1,\mathbf{Q}}^{(-1)}(V_{i,n}^1),\dots,F_{d,\mathbf{Q}}^{(-1)}(V_{i,n}^d))}{f^{\mathbf{Q}}(F_{1,\mathbf{Q}}^{(-1)}(V_{i,n}^1),\dots,F_{d,\mathbf{Q}}^{(-1)}(V_{i,n}^d))}, \tag{9}$$

with $g(U^1,\dots,U^d) := h(F_{1,\mathbf{Q}}^{(-1)}(U^1),\dots,F_{d,\mathbf{Q}}^{(-1)}(U^d))$ as in Equation (3), and where $F_{j,\mathbf{P}}^{(-1)}$ denotes the right-inverse of the distribution function $F_{j,\mathbf{P}}$ of X^j under $\mathbf{P}$ and likewise for $\mathbf{Q}$.

3 Consistency and Central Limit Theorem of combined estimators

We consider the combination of LHSD with control variates, and the combination of LHSD with importance sampling. First, we establish consistency of the combined estimators, provided g is bounded and fulfills some continuity conditions. It follows directly that the estimator is asymptotically unbiased. Second, under some further conditions, we derive a central limit theorem that allows us to give an interpretation of the sample variance derived from the combined estimators. Although the technical conditions in the statements below may appear restrictive (e.g. excluding unbounded payoffs of financial derivatives), this does not pose serious limitations when applying Monte Carlo simulation. In fact, these conditions are commonly found in statements involving Quasi-Monte Carlo methods.

Proposition 1. *Let $g:[0,1]^d\to\mathbb{R}$ be bounded and continuous C-a.e.*

(i) Let $f:[0,1]^d\to\mathbb{R}$ be bounded and continuous C-a.e. Then, the combined LHSD and control variates estimator, Equation (7), is strongly consistent, that is,

$$\frac{1}{n}\sum_{i=1}^{n}\left(g(V_{i,n}) - b(f(V_{i,n}) - \mathbb{E}f(U))\right) \overset{\mathbf{P}\text{–a.s.}}{\longrightarrow} \mathbb{E}g(U), \qquad \text{as } n\to\infty.$$

(ii) Let $\mathbf{Q}$ be another probability measure and suppose that $\mathbf{P}$ is absolutely continuous with respect to $\mathbf{Q}$, $\mathbf{P}\ll\mathbf{Q}$, and suppose further that $f^{\mathbf{P}}$ and $f^{\mathbf{Q}}$ are densitities for X under measures $\mathbf{P}$ and $\mathbf{Q}$, respectively, and that the likelihood ratio $\frac{f^{\mathbf{P}}}{f^{\mathbf{Q}}}$ is bounded and continuous $\mathbf{P}$–a.s.. Then, the combined LHSD and importance sampling estimator, Equation (9), is strongly consistent (with respect to $\mathbf{Q}$), that is,

$$\frac{1}{n}\sum_{i=1}^{n} g(V_{i,n})\,\frac{f^{\mathbf{P}}(F_{1,\mathbf{P}}^{(-1)}(V_{i,n}^1),\dots,F_{d,\mathbf{Q}}^{(-1)}(V_{i,n}^d))}{f^{\mathbf{Q}}(F_{1,\mathbf{Q}}^{(-1)}(V_{i,n}^1),\dots,F_{d,\mathbf{Q}}^{(-1)}(V_{i,n}^d))} \overset{\mathbf{Q}\text{–a.s.}}{\longrightarrow} \mathbb{E}^{\mathbf{Q}}\left[g(U)\,\frac{f^{\mathbf{P}}(F_{1,\mathbf{P}}^{(-1)}(U^1),\dots,F_{d,\mathbf{P}}^{(-1)}(U^d))}{f^{\mathbf{Q}}(F_{1,\mathbf{Q}}^{(-1)}(U^1),\dots,F_{d,\mathbf{Q}}^{(-1)}(U^d))}\right] = \mathbb{E}^{\mathbf{P}}[g(U)],$$

as $n\to\infty$. Since $\mathbf{P}\ll\mathbf{Q}$, the claim also holds $\mathbf{P}$–a.s.

Proof. The proof follows directly from Proposition 4.1 of Packham and Schmidt (2010), which establishes strong consistency for the LHSD estimator when $g:[0,1]^d\to\mathbb{R}$ is bounded and C-a.e. continuous. □

Corollary 2. *The estimators in Proposition 1 are asymptotically unbiased, that is, their expectation converges to* $\mathbb{E}g(U^1, \dots, U^d)$*, as* $n \to \infty$.

Proof. This is an application of the Dominated Convergence Theorem. $\square$

The Central Limit Theorem below is stated for the bivariate case. This captures the essential features of the convergence property while avoiding notational clutter. A multivariate version of the Central Limit Theorem for LHSD is found in Aistleitner, Hofer, and Tichy (2012). First we need some definitions.

Definition 1. A function $g : [0,1]^2 \to \mathbb{R}$ is of *bounded variation (in the sense of Hardy-Krause)*, if there exists a constant K such that

(i) for every bounded rectangle $[a,b] \times [c,d] \subseteq [0,1]^2$, for all m, n and points $a = x_0 < x_1 < \cdots x_m = b$, $c = y_0 < y_1 < \cdots < y_n = d$,

$$\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |g(x_i, y_j) + g(x_{i+1}, y_{j+1}) - g(x_i, y_{j+1}) - g(x_{i+1}, y_j)| \leq K,$$

(ii) for every $u \in [0,1]$, $v \mapsto g(u,v)$ is a function whose variation is bounded by K,

(iii) for every $v \in [0,1]$, $u \mapsto g(u,v)$ is a function whose variation is bounded by K.

Definition 2. A function $g : [0,1]^2 \to \mathbb{R}$ is *right-continuous* if for any sequence $(u_n^1, u_n^2)_{n \geq 1}$, with $u_n^j \downarrow u^j$, $j = 1, 2$, $\lim_{n\to\infty} g(u_n^1, u_n^2) = g(u^1, u^2)$.

Theorem 3 (Central Limit Theorem)**.** *Let the copula* C *of* (U^1, U^2) *have continuous partial derivatives, and let* $g : [0,1]^2 \to \mathbb{R}$ *be of bounded variation (in the sense of Hardy-Krause) and right-continuous.*

(i) Let $f : [0,1]^2 \to \mathbb{R}$ *be of bounded variation and right-continuous. Then, for the combined LHSD and control variates estimator, Equation (7), as* $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(g(V_{i,n}^1, V_{i,n}^2) - b(f(V_{i,n}^1, V_{i,n}^2) - \mathbb{E}f(U^1, U^2)) - \mathbb{E}g(U^1, U^2) \right) \xrightarrow{\mathcal{L}} \mathrm{N}(0, \sigma^2_{LHSD+CV}).$$

(ii) Let $\mathbf{Q}$ *be another probability measure, let the distribution function* $F^{\mathbf{Q}}(F_{1,\mathbf{Q}}^{(-1)}(U^1), F_{2,\mathbf{Q}}^{(-1)}(U^2))$ *of* (U^1, U^2) *have continuous partial derivatives and suppose that* $\mathbf{P} \ll \mathbf{Q}$. *Suppose further that the Radon-Nikodym derivative* $\frac{f^{\mathbf{P}}}{f^{\mathbf{Q}}}$ *is of bounded variation and right-continuous. Then, as* $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(g(V_{i,n}^1, V_{i,n}^2) \frac{f^{\mathbf{P}}(F_{1,\mathbf{P}}^{(-1)}(V_{i,n}^1), F_{2,\mathbf{P}}^{(-1)}(V_{i,n}^2))}{f^{\mathbf{Q}}(F_{1,\mathbf{Q}}^{(-1)}(V_{i,n}^1), F_{2,\mathbf{Q}}^{(-1)}(V_{i,n}^2))} - \mathbb{E}g(U^1, U^2) \right) \xrightarrow{\mathcal{L}} \mathrm{N}(0, \sigma^2_{LHSD+IS}).$$

Proof. The proof follows directly from Theorem 5.4 of Packham and Schmidt (2010), which establishes a central limit theorem for the standalone LHSD estimator when C has continuous partial derivatives and when g is of bounded variation and right-continuous. $\square$

Explicit expressions for $\sigma^2_{\text{LHSD+CV}}$ and $\sigma^2_{\text{LHSD+IS}}$ can be derived from Theorem 5.4 in Packham and Schmidt (2010). In practice, one typically simulates the estimator by running several independent iterations of the simulation, and then takes the sample variance as an estimate of $\sigma^2_{\text{LHSD+CV}}$, resp. $\sigma^2_{\text{LHSD+IS}}$.

4 Examples

We investigate the effectiveness of combining LHSD with other variance reduction techniques by studying two examples. In the first example we price an Asian basket call option, whose payoff depends on the sum of several underlying assets monitored at several points in time. This is a path-dependent option in a high-dimensional setting, and simulation is a standard valuation approach. The example below is taken from Imai and Tan (2007).

The second example is on pricing an Asian option (i.e., the one asset case). It is taken from Glasserman et al. (1999) (see also (Glasserman, 2004, Section 4.6.2)), who analyse the combination of stratification and importance sampling. Using the same example, Lemieux and La (2005) study combinations of control variates and importance sampling based on RQMC methods. These results allows us to compare combinations of LHSD with other stratification methods, resp. quasi-Monte Carlo methods, and importance sampling.

Pricing options at varying strikes allows us to analyse the effectiveness for situations where sampling is done efficiently over the sample space (case of in-the-money options) compared to situations where the area of interest is hit only by few samples (out-of-the-money options case). In particular, we would expect combinations of stratification techniques with importance sampling to be effective when the region of interest is small, as importance sampling will ensure that samples are drawn predominantly from the region of interest, while stratification ensures that the samples are evenly spaced in this region.

The importance sampling methods applied below are tailored to the applications. Nonparametric generic methods, such as importance sampling with kernel density estimators (Zhang, 1996; Swiler and West, 2010), have proven to be very slow, so that they are applicable only when individual simulations are CPU-intensive.

4.1 Valuation of Asian basket call option

Assume a basket of m assets, with S^i_t the price of the i-th asset at time t, $i = 1, \ldots, m$. Fixing a maturity T, the strike K, a set of n monitoring time points $0 < t_1 < t_2 < \cdots < t_n = T$ and weights w^{ij}, with $\sum_{i=1}^m \sum_{j=1}^n w^{ij} = 1$, the payoff of the Asian basket call option on the m-asset basket is given by

$$\max\left(\sum_{i=1}^m \sum_{j=1}^n w^{ij} S^i_{t_j} - K, 0\right). \tag{10}$$

We assume that asset prices follow a geometric Brownian motion; more specifically, $S^1, \ldots, S^m$ is the solution of the stochastic differential equation

$$\mathrm{d}S^i_t = rS^i_t \,\mathrm{d}t + \sigma^i S^i_t \,\mathrm{d}W^i_t, \quad i = 1, \ldots, m,$$

where r is the risk-free interest rate, σ^i is the volatility of the i-th asset and $(W^1, \ldots, W^m)$ is an m-dimensional Brownian motion, whose components W^i and W^k are correlated with ρ^{ik}, $1 \le i, k \le m$.

The solution of the SDE is given by

$$S_t^i = S_0^i \exp\left(\left(r - \frac{(\sigma^i)^2}{2}\right)t + \sigma^i W_t^i\right), \quad i = 1, \dots, m. \tag{11}$$

Pricing the option requires simulating the paths of each asset at the monitoring time points. Assume that the points $t_1, \dots, t_n$ are equidistant and let $\Delta t = T/n$, so that $t_j = j\Delta t$. Let Σ be the $m \times m$ covariance matrix with entries $(\rho^{ik}\sigma^i\sigma^k\Delta t)_{i,k=1,\dots,m}$. Let $\tilde{\Sigma}$ be the $nm \times nm$ matrix generated from Σ via

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma & \Sigma & \cdots & \Sigma \\ \Sigma & 2\Sigma & \cdots & 2\Sigma \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma & 2\Sigma & \cdots & n\Sigma \end{pmatrix}.$$

The asset prices are simulated according to Equation (11), with the random vector $\tilde{W} = (\sigma^1 W_{t_1}^1, \dots, \sigma^m W_{t_1}^m, \sigma^1 W_{t_2}^1, \dots, \sigma^m W_{t_2}^m, \dots, \sigma^1 W_{t_n}^1, \dots, \sigma^m W_{t_n}^m)'$ derived via

$$\tilde{W} = \tilde{C}Z,$$

where $\tilde{C}$ is such that $\tilde{C}\tilde{C}' = \tilde{\Sigma}$ (e.g. the Cholesky decomposition) and Z is a vector of nm independent standard normal random variables. The payoff at time T of the Asian basket option can then be written as

$$\max(g(\tilde{W}) - K, 0),$$

with

$$g(\tilde{W}) = \sum_{k=1}^{nm} \exp(\nu^k + \tilde{W}_k) \tag{12}$$

$$\nu^k = \ln(w^{k_1\ k_2} S_0^{k_1}) + (r - (\sigma^{k_1})^2/2)t_{k_2}, \tag{13}$$

where

$$k_1 = (k-1) \bmod m + 1$$
$$k_2 = \lfloor (k-1)/m \rfloor + 1, \quad k = 1, \dots, mn.$$

In this approach, simulation of option payoffs involves the computation of products of high-dimensional matrices.

4.1.1 Application of LHS and LHSD

To apply LHS, Z is simulated, by drawing uniform random numbers, introducing an LHS step and then applying the inverse normal distribution cdf $\mathrm{N}^{(-1)}$.

To apply LHSD, $\tilde{W}$ is first simulated using the Cholesky decomposition $\tilde{C}$ of $\tilde{\Sigma}$, and then an LHSD step is introduced in each dimension over all simulations. In other words, each dimension is stratified by applying the transformation (2) to generate a LHSD sample from random numbers.

4.1.2 Application of Control Variates

To apply the control variates method, the underlying assets are used as control variates. More precisely, the variables X_i, $i = 1, \dots, n$ in the control variates estimator, Equation (4), are the sampled option payoffs, and the control variates Y_i, $i = 1, \dots, n$ are given by

$$Y_i = \sum_{k=1}^{m} \sum_{j=1}^{n} \frac{S_{t_j}^k}{nm}.$$

Since the optimal coefficient b^* is not known, the parameter estimate $\hat{b}_n$ of Equation (6) is used. As b^* depends essentially on the correlation σ_{XY}, which measures the linear relationship between X and Y, the control variates estimator typically performs better for options with smaller strikes (in-the-money options) and performs worse for options with greater strikes (out-of-the-money options).

4.1.3 Application of Importance Sampling

Independent random variables Importance sampling requires that we choose a measure according to which random numbers are sampled. Inspection of Equation (8) indicates that we should sample proportional to $h(X)\, f^{\mathbf{P}}(X)$. For a Geometric Brownian motion, a change of measure with equivalent measures $\mathbf{P}$ and $\mathbf{Q}$ corresponds to a change of the drift; in our setup this implies that the mean of Z is changed. We follow Glasserman et al. (1999) (see also Section 4.6.2 of Glasserman (2004)).

Using that $\tilde{W}_k = \sigma^{k_1} W_{t_{k_2}}^{k_1} = \sum_{j=1}^{nm} \tilde{c}_{kj} Z_j = \tilde{c}_k Z$, where $\tilde{c}_{..}$ refers to the elements of $\tilde{C}$ and where $\tilde{c}_k$ denotes the k-th row of $\tilde{C}$, write the option payoff using Equation (12) in terms of $Z \sim \mathrm{N}(0, I)$:

$$G(Z) = \max \left(\sum_{k=1}^{nm} \exp\left(\nu^k + \tilde{c}_k Z\right) - K, 0 \right),$$

with ν^k, $k = 1, \dots, nm$, given by Equation (13). Pricing the option amounts to evaluating $\mathbf{e}^{-rT} \mathbb{E} G(Z)$.

The change in measure is applied by changing the mean of Z from 0 to some vector μ. The likelihood ratio when changing measure from $\mathrm{N}(0, I)$ to $\mathrm{N}(\mu, I)$ is given by

$$\prod_{i=1}^{mn} \frac{f_0(Z_i)}{f_{\mu_i}(Z_i)} = \exp\left(-\mu^T Z + \frac{1}{2} \mu^T \mu \right), \tag{14}$$

where f_μ denotes the normal density with mean μ and variance 1. It follows that

$$\mathbb{E} G(Z) = \mathbb{E}_\mu \left[G(Z) \mathbf{e}^{-\mu^T Z + \frac{1}{2} \mu^T \mu} \right],$$

where $\mathbb{E}_\mu$ denotes the expectation when $Z \sim \mathrm{N}(\mu, I)$. The estimator is unbiased, and we would like to choose a vector μ that produces a low-variance estimator.

Since G takes only non-negative values, we may write $G(z) = \exp(F(z))$. Furthermore, taking expectation of Z under the measure $\mathbf{P}_\mu$ is equivalent to taking expectation of $\mu + Z$ under the original measure. Thus,

$$\begin{aligned} \mathbb{E} G(Z) = \mathbb{E}\left[\mathbf{e}^{F(Z)} \right] &= \mathbb{E}_\mu \left[\mathbf{e}^{F(Z)}\, \mathbf{e}^{-\mu^T Z + \frac{1}{2}\mu^T \mu} \right] \\ &= \mathbb{E}\left[\mathbf{e}^{F(\mu+Z)}\, \mathbf{e}^{-\mu^T(\mu+Z) + \frac{1}{2}\mu^T \mu} \right] = \mathbb{E}\left[\mathbf{e}^{F(\mu+Z)}\, \mathbf{e}^{-\mu^T Z - \frac{1}{2}\mu^T \mu} \right]. \end{aligned}$$

A particular choice of μ is obtained by a first-order approximation of F,

$$\mathbf{e}^{F(\mu+Z)}\,\mathbf{e}^{-\mu^T Z-\frac{1}{2}\mu^T\mu} \approx \mathbf{e}^{F(\mu)+\nabla F(\mu)Z}\,\mathbf{e}^{-\mu^T Z-\frac{1}{2}\mu^T\mu}, \tag{15}$$

with $\nabla F(\mu)$ the gradient of F at μ. If μ is chosen to satisfy $\nabla F(\mu) = \mu^T$, then the right-hand side of Equation (15) is a constant with no dependence on Z. Hence, if Equation (15) held exactly, then this would produce a zero-variance estimator, and the hope is that this produces a low-variance estimator if Equation (15) holds only approximately. Observe that $\nabla F(z) - z^T = 0$ is equivalent to $\nabla G(z) - z^T G(z) = 0$.

Returning to the Asian basket call option, observe first that

$$\frac{\partial G(z)}{\partial z_l} = \mathbf{1}_{\{\text{in-the-money}\}} \sum_{k=1}^{nm} \exp\left(\nu^k + \tilde{c}_k z\right)\,\tilde{c}_{kl}.$$

Combining the above, we obtain the conditions

$$\frac{\partial G(z)}{\partial z_l} - z_l G(z) = \mathbf{1}_{\{\text{in-the-money}\}} \left[\sum_{k=1}^{nm} \exp\left(\nu^k + \tilde{c}_k z\right)\,\tilde{c}_{kl} - z_l \left(\sum_{k=1}^{nm} \exp\left(\nu^k + \tilde{c}_k z\right) - K\right)\right] = 0,$$

for $l = 1, \ldots, nm$. Restricting to the case where the option is in-the-money, solving for z_l yields

$$z_l = \frac{\sum_{k=1}^{nm} \exp\left(\nu^k + \tilde{c}_k z\right)\,\tilde{c}_{kl}}{G(z)}. \tag{16}$$

The values of $z = (z_1, \ldots, z_{nm})^T$ can be found using e.g. multivariate Newton's method for root finding. Fix-point iteration has proven to be unstable as this may converge to a fix point outside the range where the option is in-the-money. The computation time of z depends heavily on the initial values plugged into the root finding algorithm and may easily annihilate the pick-up in variance reduction. However, when running the simulation on a regular (e.g. daily) basis, then one can expect the computation to be efficient by choosing the current value of z as initial value.

Dependent random variables Applying importance sampling to a random vector whose components are not independent, as is the case when combining LHSD and importance sampling, requires adjusting the likelihood ratio accordingly. Recall that the likelihood ratio based on independent $Z = (Z_1, \ldots, Z_{nm})^T$ is given by Equation (14). A measure change from $\mathrm{N}(0, I)$ to $\mathrm{N}(\mu, I)$ when Z are the inputs can be translated into a corresponding measure change when $\tilde{W}$ are the inputs as follows: under the original measure we have $\tilde{W}_k = \tilde{c}_k Z$ where $\tilde{c}_k$ denotes the k-th row of $\tilde{C}$, whereas under the new measure we have $\tilde{W}_k = \tilde{c}_k(Z + \mu)$, with $Z \sim (0, I)$ as before. When $\tilde{W}_k$ are the inputs, the measure is thus changed from $\mathrm{N}(0, \tilde{\Sigma})$ to $\mathrm{N}(\tilde{C}\mu, \tilde{\Sigma})$. The corresponding likelihood ratio is

$$\frac{f_{0,\tilde{\Sigma}}(\tilde{W})}{f_{\tilde{C}\mu,\tilde{\Sigma}}(\tilde{W})} = \exp\left(-\frac{1}{2}\left(\tilde{W}^T \tilde{\Sigma}^{-1} \tilde{W} - (\tilde{W} - \tilde{C}\mu)^T \Sigma^{-1} (\tilde{W} - \tilde{C}\mu)\right)\right),$$

where $f_{\mu,\Sigma}$ denotes the multivariate normal density with mean vector μ and covariance matrix Σ, and with $\tilde{W} \sim \mathrm{N}(\tilde{C}\mu, \tilde{\Sigma})$.

Table 1: Parameters of Asian basket option

Parameter	Value
Number of assets	$m = 10$
Initial asset value	$S_0^j = 100,\ j = 1, \ldots, m$
Asset volatility	$\sigma^j = 0.1 + (j-1)/(m-1) \cdot 0.4,\ j = 1, \ldots, m$
Correlation	$\rho^{ij} = 0.4,\ 1 \leq i < j \leq m$
Interest rate	$r = 0.04$
Maturity	$T = 1$ (years)
Strike	$K = 90, 100, 110$
Number of time steps	$n = 250$
Weights	$w^{ij} = 1/(nm),\ i = 1, \ldots, n,\ j = 1, \ldots, m$

4.2 Results

4.2.1 Asian basket call option

Based on the data set of Imai and Tan (2007), we calculate prices of an Asian basket option by simulating $\tilde{W}$ using a Cholesky decomposition using various combinations of variance reduction techniques. The parameters of the example are given in Table 1. As in Imai and Tan (2007), we compute 10 iterations of 4,096 simulations. We calculate options prices using plain Monte Carlo simulation, LHSD and LHS, and combined with control variate and importance sampling, yielding nine different setups. LHS was implemented by applying LHS to independent random numbers, which were then correlated. For control variates we used the parameter $\hat{b}_n$ from Equation (6).

The results are given in Table 2. Aside from prices and standard deviations (of the estimator), the table shows the ratios of the estimators' standard deviations relative to the standard deviation of the plain Monte Carlo estimator, the ratios of CPU-time consumed relative to the CPU-time consumed by the plain Monte Carlo estimator, and an indicator of efficiency, given by the product of variance and expected computing time (relative to the plain Monte Carlo estimator). An estimator with a smaller efficiency indicator is preferred over an estimator with a higher indicator. Computation times for importance sampling exclude the calculation of the measure (cf. Equation (16)), which is costly when no proper initial guess is present. Furthermore, it should be noted that the efficiency indicator will be in favour of estimators with smaller variance if the calculation of the payoff – which affects all methods – is costly.

For an easy comparison, the ratios of the estimators' standard deviations relative to the standard deviations of the plain Monte Carlo estimator are shown in Figure 2.

First, observe that the variance reduction with LHS/LHSD depends strongly on the strike price of the option and decreases with increasing strike. The greater the strike, the more often the option expires out-of-the money, which diminishes the effect of stratification. A similar observation is made in an example from (Glasserman, 2004, pp. 242–243), where an Asian call option is priced using LHS.

Next, there is a pick-up in variance reduction when applying control variates combined with the standalone methods other than LHSD, especially in the case where the strike price of the option is low (this is evident, as the correlation with the underlying basket of stocks, which serves as the control variate, is greater the lower the strike price). However, combining LHSD with control variates performs equally well to LHSD standalone.

Table 2: Simulated prices of an Asian basket option (parameters in Table 1) for strikes $K \in \{90, 100, 110\}$. The results are based on 10 iterations with 4096 simulations each. The numbers in parentheses denote the sample standard deviation based on the 10 iterations. The columns labelled "ratio" denote the ratio of the standard deviations of plain Monte Carlo simulation with the respective method.

	$K = 90$			$K = 100$			$K = 110$		
	price	std. dev.	ratio	price	std. dev.	ratio	price	std. dev.	ratio
MC	12.353	(0.113)	1.000	5.658	(0.130)	1.000	2.016	(0.071)	1.000
MC + CV	12.337	(0.018)	6.250	5.659	(0.046)	2.818	2.017	(0.058)	1.210
MC + IS	12.332	(0.049)	2.307	5.650	(0.045)	2.881	2.040	(0.022)	3.244
LHS	12.334	(0.042)	2.679	5.694	(0.095)	1.370	2.035	(0.061)	1.153
LHS + CV	12.330	(0.028)	4.090	5.689	(0.072)	1.802	2.035	(0.053)	1.338
LHS + IS	12.327	(0.050)	2.244	5.629	(0.045)	2.871	2.030	(0.020)	3.531
LHSD	12.333	(0.013)	8.543	5.661	(0.031)	4.128	2.015	(0.043)	1.658
LHSD + CV	12.334	(0.013)	8.545	5.662	(0.031)	4.128	2.016	(0.043)	1.658
LHSD + IS	12.331	(0.046)	2.462	5.649	(0.034)	3.790	2.041	(0.018)	4.019

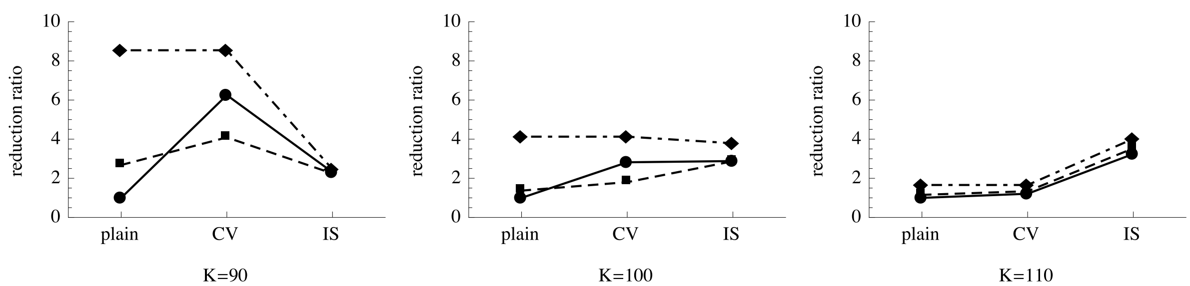

Figure 2: Ratios of reduction in standard deviation relative to plain Monte Carlo simulation (column "ratio" in Table 2). Circles denote Monte Carlo simulation, squares denote LHS and diamonds denote LHSD.

However, the picture is mixed when combining importance sampling with LHS or LHSD: for low strikes (in-the-money case), the variance is actually increased, whereas for large strikes (out-of-the-money case), the variance is reduced. This shows that combinations of stratification and importance sampling (as implemented here) complement each other well when the region of interest is small. This is intuitive as stratification alone fails to achieve satisfactory results when the option's strike is high, as many outcomes will be zero regardless of stratification. Importance sampling ensures that more stratified samples are placed in the region where the option is in-the-money. On the other hand, combining LHS/LHSD in combination with importance sampling performs poor when the region of interest corresponds largely to the sampling region. We shall see in the following examples that this behaviour also occurs with other stratification techniques.

In all cases, LHSD performs better than the respective plain Monte Carlo or LHS counterparts, which shows that performing an LHSD step is generally worthwhile doing. In the following example, it will therefore also be interesting to see how LHSD performs compared to other stratification techniques.

Finally, assessing the performance pickup requires taking into account the CPU time required for each simulation pass. However, since CPU times are hardware-, software- and implementation-specific, we merely give some indications on the implementation at hand (Mathematica 9.0 on Mac OS X). A key figure for assessing the overall efficiency is the product of variance (of the estimator) and expected CPU time per estimation (cf. Section 1.1.3 of Glasserman (2004)). In these terms, both control variates and importance sampling perform well relative to the naive Monte Carlo estimator, while LHS and LHS combined with control variates perform worse. LHS combined with importance sampling and LHSD-combined estimators perform better than naive Monte Carlo simulation, and only slightly worse than the uncombined control variates and importance sampling estimators. Excluded from these observations is the initial CPU time required for determining the measure change in importance sampling, (16), which depends strongly on the starting values. More generally, it should be taken into account that the the payoff's complexity plays a significant role: the greater the CPU time required to calculate the payoff, the smaller the significance of the CPU time of the procedure for generating random numbers.

4.2.2 Asian call option

We now turn to the example in Glasserman et al. (1999) (see also Section 4.6.2 of Glasserman (2004)) of valuing an Asian call option (i.e., the one asset case). Here, the combination of stratification and importance sampling is analysed. Using the same example, Lemieux and La (2005) study combinations of control variates and importance sampling based on QMC and RQMC methods.

The stratification of Glasserman et al. (1999) works as follows: Suppose that $Z_1, \ldots, Z_l$ are l iid n-dimensional realisations of $Z \sim \mathrm{N}(0, I)$. Let $(-\infty, k_1], (k_1, k_2], \ldots, (k_{l-1}, \infty)$ be a partition of the real line into l strata of equal probability, that is, for $X \sim \mathrm{N}(0,1)$, we have $\mathbf{P}(X \in (k_j, k_{j+1}]) = 1/l$ for all $j \in \{1, \ldots, l\}$. The principal idea is to choose a linear projection $v^T Z$, with $v \in \mathbb{R}^n$ and $\|v\| = 1$, where the last condition ensures that $v^T Z \sim \mathrm{N}(0,1)$, and generate random numbers such that $(v^T Z_1, \ldots, v^T Z_l)$ are stratified, that is, where the k-th sample is placed in the k-th stratum. The algorithm for generating $(Z_1, \ldots, Z_l)$ fulfilling this property is given in Section 4.3.2 of Glasserman (2004). It remains to choose $v \in \mathbb{R}^n$. Glasserman et al. (1999) provide two choices of $v \in \mathbb{R}^n$, the first one of which is to set $v = \mu$, with μ the drift change from importance sampling, see Section 4.1.3, based on the idea that μ is an "important" path. The second method derives v from a second-order approximation of the payoff. The resulting variance reduction ratios of both methods are roughly of the same order, so that we consider only the first method.

The parameters of the example are $S_0 = 50$, $T = 1$, $r = 0.05$, $K \in \{45, 50, 55\}$. The asset volatility is $\sigma \in \{0.1, 0.3\}$ and the number of monitoring time points is $n \in \{16, 64\}$. The results, given in Table 3, are from 100 iterations of 4096 simulations. A first observation is that stratification has by far the greatest impact on reducing the variance. Second, the combined estimators achieve a greater variance reduction when the option is out-of-the money than in the standalone cases. As in the earlier example, for options far in-the-money, the effect may be reversed, that is, importance sampling may actually worsen the variance reduction. This effect is also observed in Lemieux and La (2005) for combinations of Quasi-Monte Carlo (QMC) methods and importance sampling. As in the previous example, we observe that importance sampling and the other methods complement each other well for out-of-the money options (or more generally, when sampling predominantly in "unimportant" regions without importance sampling).

When comparing LHS and LHSD we see that, in general, LHSD performs better than LHS, but the variance reduction of LHSD comes nowhere near the ratios achieved with stratification.

Table 3: Simulated prices of an Asian option. The parameters are $S_0 = 50$, $T = 1$, $r = 0.05$ with the number of monitoring time points n, volatility σ and strike K of the option indicated below. The results are based on 100 iterations with 4096 simulations each. The columns labelled "ratio" denote the ratio of the standard deviations of plain Monte Carlo simulation with the respective method.

			$K = 45$			$K = 50$			$K = 55$		
	n	σ	price	std.dev.	ratio	price	std.dev.	ratio	price	std.dev.	ratio
MC	16	0.1	6.057	(0.048)	1.000	1.917	(0.038)	1.000	0.202	(0.012)	1.000
IS			6.054	(0.013)	3.796	1.920	(0.013)	2.914	0.203	(0.003)	4.333
ST			6.055	(0.001)	87.049	1.920	(0.000)	84.398	0.202	(0.000)	34.308
IS+ST			6.055	(0.001)	80.512	1.920	(0.000)	111.331	0.202	(0.000)	163.458
LHS			6.055	(0.003)	14.625	1.919	(0.013)	2.895	0.203	(0.009)	1.339
LHS+IS			6.055	(0.010)	4.609	1.920	(0.011)	3.569	0.202	(0.002)	5.326
LHSD			6.055	(0.001)	54.102	1.919	(0.004)	10.599	0.202	(0.003)	3.898
LHSD+IS			6.055	(0.004)	12.357	1.920	(0.005)	8.145	0.203	(0.001)	7.917
MC	64	0.3	7.049	(0.127)	1.000	4.026	(0.092)	1.000	2.088	(0.081)	1.000
IS			7.027	(0.042)	2.994	4.027	(0.032)	2.857	2.082	(0.019)	4.159
ST			7.020	(0.003)	40.557	4.022	(0.003)	33.033	2.080	(0.003)	32.167
IS+ST			7.021	(0.003)	37.611	4.023	(0.002)	39.614	2.080	(0.002)	50.844
LHS			7.026	(0.043)	2.930	4.021	(0.050)	1.846	2.086	(0.048)	1.686
LHS+IS			7.018	(0.036)	3.574	4.022	(0.027)	3.453	2.077	(0.018)	4.512
LHSD			7.021	(0.010)	12.954	4.022	(0.013)	7.334	2.079	(0.013)	6.314
LHSD+IS			7.020	(0.017)	7.477	4.023	(0.014)	6.551	2.080	(0.010)	8.251

This certainly has to do with the fact that stratification and importance sampling use highly problem-specific information in an optimised way: stratifying a linear projection requires choosing a direction $v \in \mathbb{R}^n$ and importance sampling requires choosing a target probability measure. In the case of the Asian option, both methods use the linear drift vector μ, but for more complex options or models it may not be straightforward to choose these quantities in a meaningful way, and it may be time-consuming to compute them, potentially reducing or even annihilating the benefit from variance reduction. LHS and LHSD, on the other hand, are applied without problem-specific knowledge, that is, they may be applied to any valuation problem in the way they are described in this paper. Regarding computational overhead, LHS requires drawing random permutations, which can be efficiently done with the *Fisher-Yates shuffle* (Knuth, 1998, Section 3.4.2). This algorithm generates a random permutation of n numbers in $O(n)$ computing time. Computing the ranks for LHSD essentially requires sorting the samples, which can be done in $O(n \log n)$ computing time.

Finally, comparing the variance reduction achieved by LHSD with the error reduction achieved by the QMC-methods in Lemieux and La (2005), we find that LHSD dominates QMC most of time, with three simulation results where QMC outperforms LHSD. The differences are substantial only for in-the-money options where the standard error by LHSD is up to 2.5 times smaller than for QMC.

5 Conclusion

Efficient simulation techniques are essential for fast pricing of contingent claims via Monte Carlo simulation. This applies especially to high-dimensional valuation problems, such as options depending on several underlyings, as the computational effort typically increases non-linearly with the dimension. The aim of variance reduction techniques is to produce estimators with a smaller variance than the usual Monte Carlo estimator, allowing to save computation time or increasing accuracy.

In this paper, we study combinations of Latin hypercube sampling specifically targeted at dependent random vectors with control variates and importance sampling. LHS techniques are general transforms of random numbers, while control variates and importance sampling make use of problem-specific information. Because of this, combining these types of variance reduction techniques may prove effective. Under some mild conditions, the resulting estimators are consistent and asymptotically unbiased, and we derive a central limit theorem.

We investigate the effectiveness of applying combined variance reduction methods by pricing an Asian basket call option with 10 assets monitored at 250 time points, yielding a 2500-dimensional estimation problem. It turns out that combining LHSD with control variates yields no significant pick-up compared to applying LHSD standalone. LHSD with importance sampling performs best (factor 4 in terms of standard deviation) for options with a high strikes (i.e., out-of-the-money options). For options with a low strike, the combination of LHSD with importance sampling actually increases the variance compared to standalone LHSD. This effect is also observed in other studies of combinations of stratification and importance sampling techniques (e.g. Lemieux and La, 2005).

In a second example, we further compare the effectiveness with other combinations of variance reduction techniques, and find that problem-specific stratification methods outperform LHSD. This is not surprising as a problem-specific method is tailored to optimise the variance reduction of a particular problem, whereas LHSD is a variance reduction technique that requires no specific knowledge about the problem. Among the problem-independent stratification-like methods, LHSD performs best.

LHSD is not restricted to the types of examples studied above, but can be integrated into any kind of simulation problem, such as models beyond the Black-Scholes model and exotics beyond Asian options. Since LHSD is easy to apply, our results indicate that it is worthwhile to test the performance pickup of integrating LHSD into arbitrary simulation problems even when a variance reduction technique is already in place.

References

Christoph Aistleitner, Markus Hofer, and Robert Tichy. A central limit theorem for Latin hypercube sampling with dependence and application to exotic basket option pricing. *International Journal of Theoretical & Applied Finance*, 15(7), 2012.

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

Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2004.

Junichi Imai and Ken Seng Tan. A general dimension reduction technique for derivative pricing. *Journal of Computational Finance*, 10(2):129–155, 2007.

Harry Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC, 1997.

G. Knuth. The Art of Computer Programming, vol. 2: Seminumerical Algorithms. *Reading, Massachusetts*, 1998.

C. Lemieux and J. La. A study of variance reduction techniques for American option pricing. In *Proceedings of the 2005 Winter simulation conference*, pages 1884–1891. IEEE Press, 2005.

M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21: 239–245, 1979.

Roger B. Nelsen. *An Introduction to Copulas*. Springer, 1999.

H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63. SIAM, 1992.

A. B. Owen. A central limit theorem for Latin hypercube sampling. *Journal of the Royal Statistical Society*, Series B 54, 13:541–551, 1992.

N. Packham and W. M. Schmidt. Latin hypercube sampling with dependence and applications in finance. *Journal of Computational Finance*, 13(3):81–111, 2010.

A. Sklar. Fonctions de répartition a n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Universite de Paris*, 8:229–231, 1959.

Michael Stein. Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2):143–151, 1987. ISSN 0040-1706. Correction in 32:367.

Laura P. Swiler and Nicholas J. West. Importance sampling: Promises and limitations. In *Proceedings of the 51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, April 2010.

P. Zhang. Nonparametric importance sampling. *Journal of the American Statistical Association*, 91(435):1245–1253, 1996.