Wolfgang Hörmann and Josef Leydold

Importance Sampling to Accelerate the Convergence of Quasi-Monte Carlo

Paper

Importance Sampling to Accelerate the Convergence of Quasi-Monte Carlo

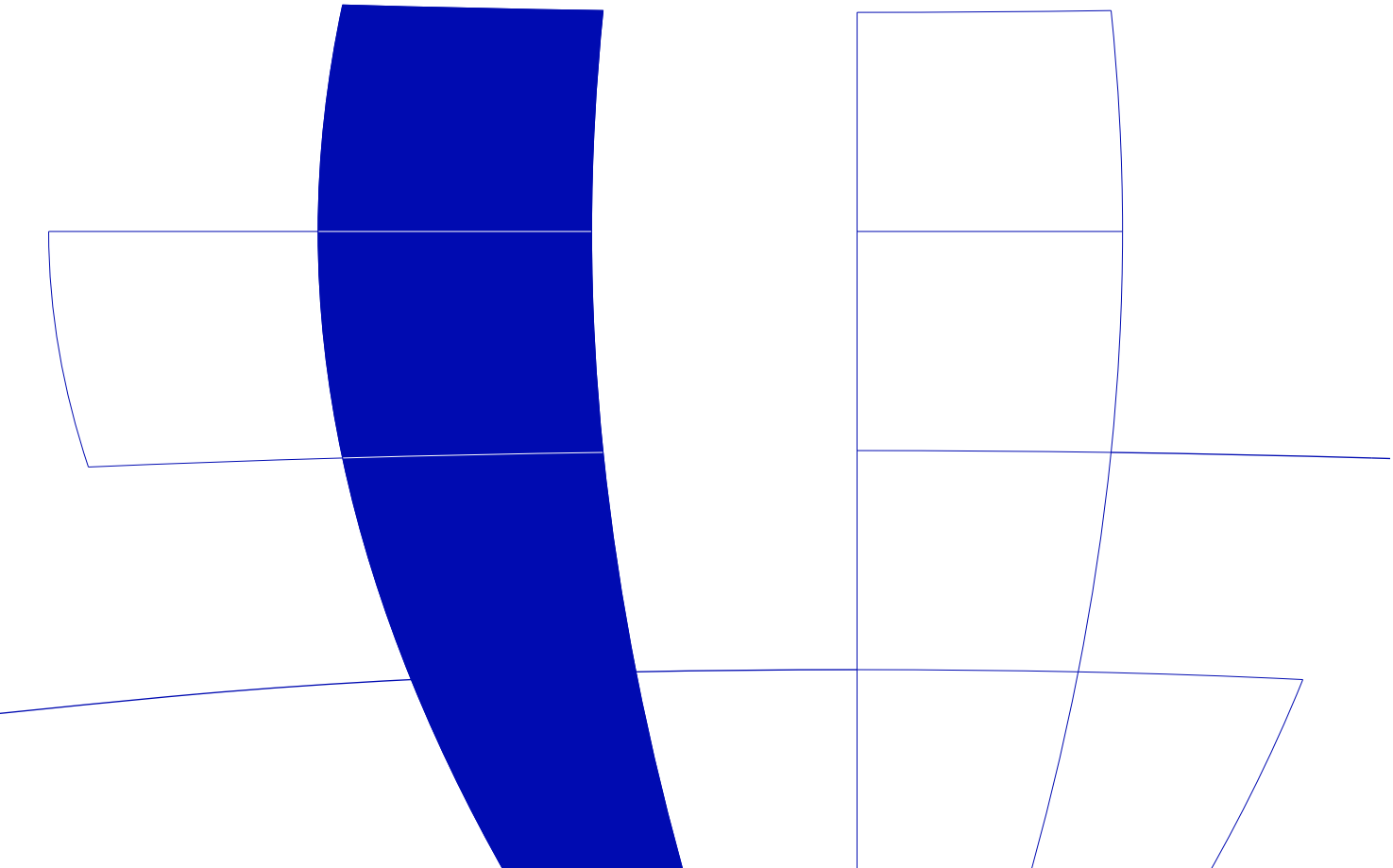
Wolfgang Hörmann, Josef Leydold

Department of Statistics and Mathematics
Wirtschaftsuniversität Wien

Research Report Series

Report 49
February 2007

http://statmath.wu-wien.ac.at/



Importance Sampling to Accelerate the Convergence of Quasi-Monte Carlo

WOLFGANG HÖRMANN AND JOSEF LEYDOLD
Department of Statistics and Mathematics
Vienna University of Economics and Business Administration

Importance sampling is a well known variance reduction technique for Monte Carlo simulation. For quasi-Monte Carlo integration with low-discrepancy sequences it was neglected in the literature although it is easy to see that it can reduce the variation of the integrand for many important integration problems. For lattice rules importance sampling is of highest importance as it can be used to obtain a smooth periodic integrand. Thus the convergence of the integration procedure is accelerated. This can clearly speed up QMC algorithms for integration problems up to dimensions 10 to 12.

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]:

General Terms: Algorithms

Additional Key Words and Phrases:

1. INTRODUCTION

A standard problem in scientific computing is the evaluation of the expectation of a function $q(x)$ with respect to a multivariate density $f(x)$. The integral can be written as

$$\mathrm{E}_f(q(x)) = \int_{\mathbb{R}^d} q(x) f(x)\, dx\ ,$$

where $x = (x_1, x_2, \ldots, x_d)$ denotes a vector in $\mathbb{R}^d$.

The direct way to solve our integral with Monte Carlo integration is to generate a sequence of random vectors X_i with density proportional to $f(x)$ and use the sample mean of $q(X_i)$ as an estimate of the value of the integral. If the generation of the required sequence X_i is very difficult or impossible we can draw a sample Y_i from a similar density $g(x)$ (called *importance sampling density*) and calculate a weighted average of $q(Y_i)$ which is again an unbiased estimate of the integral. The principle of *importance sampling* (IS) can be written as:

$$\mathrm{E}_f(q(x)) = \mathrm{E}_g(q(x)\, w(x)) = \int_{\mathbb{R}^d} q(x) w(x) g(x)\, dx \quad \text{with} \quad w(x) = \frac{f(x)}{g(x)}\ .$$

Besides its application in situations where direct sampling of the target distribution f should be avoided, importance sampling is also a well-known variance reduction technique in stochastic simulations especially in the situation of rare events.

Author's address: Vienna University of Economics and Business Administration, Department of Statistics and Mathematics, Augasse 2-6, A-1090 Vienna, Austria
email: leydold@statistik.wu-wien.ac.at
This work was supported by the Austrian Science Foundation (FWF), project no. P16767-N12.

February 13, 2007

The topic of this paper is the combination of importance sampling with quasi-Monte Carlo (QMC) techniques where (pseudo-) random points are replaced by *low-discrepancy sequences* (also called *highly uniform point sets*, HUPS). There is hardly any literature on this approach; a discussion about fundamental properties can be found in [Hörmann and Leydold 2005]. We therefore try to give a first answer to the question of how to combine IS with QMC for solving integration problems. Which are useful IS densities and how should we select their parameters? Which low-discrepancy sequences are best suited for the combination with IS? Is IS able to speed up QMC in general?

In this paper we are concentrating on the simple but important special case where the random vector X has independent and identically distributed (iid.) components. In this case the multivariate density $f(x)$ is the n-fold product of the same univariate density, $f(x) = \prod_{i=1}^{d} f_1(x_i)$. Hence it is very natural to select the IS density $g(x)$ as a n-fold product of the same univariate density as well, $g(x) = \prod_{i=1}^{d} g_1(x_i)$. Thus the choice of the IS density is mainly a one-dimensional procedure but we will see below that the dimension of the problem still has some impact. Notice that we also can deal with distributions with dependent components of X within this framework by means of some transformation of the random vector inside the function $q(x)$.

The paper is organized as following: In Sect. 2.1 we shortly describe the choice of IS densities for Monte Carlo integration as presented in the literature, whereas Sect. 2.2 discusses qualitative considerations for combining IS and low-discrepancy sequences. Section 2.3 demonstrates how IS can speed up the numerical integration using lattice rules and demonstrates the practical importance of these considerations on three examples each in dimensions 4, 8 and 12. Section 3 compares the performance of numerical integration based on IS when using pseudo-random numbers (MC), low-discrepancy sequences and lattice rules for these examples.

2. PRINCIPLES FOR SELECTING THE IMPORTANCE SAMPLING DENSITY

2.1 Monte Carlo Integration

The question of proper choices of IS densities for Monte Carlo integration was studied in the literature, see e.g. [Geweke 1989] and [Hesterberg 1995]). The objectives to select an IS density are clear: Evaluation of the density must not be expensive and it allows for easy generation of random vectors. Moreover, the variance of the MC estimator should be as small as possible.

As a first answer to this problem it is well known that the optimal IS density is proportional to $|q(x)|f(x)$. However, this result is not of practical relevance as the integral below this function is not known. A more important observation is that the variance of the IS estimator is only bounded if the tails of the IS density are at least as high as the tails of f. Thus it is usually suggested to consider a distribution family with sufficiently high tails and try to find a "good" parameter value. In some cases it is possible to calculate the variance of the IS estimator as a function of the parameter of the IS density or at least we can obtain an upper bound. Then the optimal parameter can be found by solving an optimization problem. If there is no bound for the variance available it is still possible to use an empirical approach and compare the variances of pilot runs with different parameter values. But this approach may be problematic as the estimate of the variance can be very

unstable, especially if the tails of the IS density are not high enough. If a parameter selection procedure is too time consuming or if different functions $q(x)$ are going to be integrated it is only possible to try to stick to the general rule: "Select an IS density with a shape similar to f but with higher tails."

2.2 Quasi-Monte Carlo Integration

We were not able to find much suggestions for the selection of IS densities for QMC in the literature. For example, the natural question whether it is still sensible to minimize the variance of the experiment was apparently neglected. As a matter of fact in many QMC papers the authors do not even consider the possibility of importance sampling. They just apply the inversion method for computing $\mathrm{E}_f(q(X))$ in the framework of QMC. As low-discrepancy sequences are always defined for the unit hypercube it makes sense to see how this approach can be rewritten in integral notation. Denoting the inverse of the CDF of the univariate density f_1 by $F^{-1}(u_i)$ and using the abbreviation $F^{-1}(u) = (F^{-1}(u_1), F^{-1}(u_2), \ldots, F^{-1}(u_d))$ we have:

$$\mathrm{E}_f(q(x)) = \int_{(0,1)^d} q(F^{-1}(u))\, du\ . \tag{1}$$

It is easy to see that for a density with unbounded support and unbounded $q(x)$ this change of variables leads to an integrand that may be unbounded for $u_i = 1$ or $u_i = 0$. Thus the integrand has unbounded variation and the Koskma-Hlawka inequality, which is considered the main theoretical argument for using low-discrepancy sequences, is no longer applicable. For densities with low tails this problem is of little practical relevance (Owen [2006] gives a mathematical argument for this fact). But it is very easy to get entirely rid of this problem by using importance sampling. Analogously we denote the inverse of the CDF of the univariate density g_1 by $G^{-1}(u_i)$ and use the abbreviation $G^{-1}(u) = (G^{-1}(u_1), G^{-1}(u_2), \ldots, G^{-1}(u_d))$. Then we can write the integral (1) with respect to u as

$$\mathrm{E}_f(q(x)) = \mathrm{E}_g(q(x)\, w(x)) = \int_{(0,1)^d} q(G^{-1}(u)) w(G^{-1}(u))\, du\ .$$

It is enough that the tails of the IS density are higher than the tails of f and that $q(x)$ has no pole to guarantee that the IS integrand is bounded for both, $u_i = 0$ and $u_i = 1$. This is a qualitative argument why importance sampling for distributions with unbounded support is expected to work well with low-discrepancy sequences; at least better than using the inversion method with the target distribution f. Consider the example where f is the standard normal distribution and $q(x) = (x+1)^2 I_{(-1,\infty)}$ ($I_{(a,b)}$ denotes the indicator function of the interval (a, b)). Figure 1 shows the integrand with respect to u without IS (l.h.s.) and for IS where the IS density $g(x)$ is the density of the normal distribution with $\sigma = 2$. Clearly the integrand has a pole at 1 for the naive algorithm but no pole when IS is used.

The bounded variation argument is a qualitative argument why IS should work well with QMC but it does not supply us with any guidance how to find an optimal IS density for QMC; to get rid of the pole for the integrand it is enough that the IS density has tails not lower than the original density. Thus for low-discrepancy sequences we have to use empirical evidence to decide if the application of IS can really reduce the integration error. We tried all the integration problems of the next

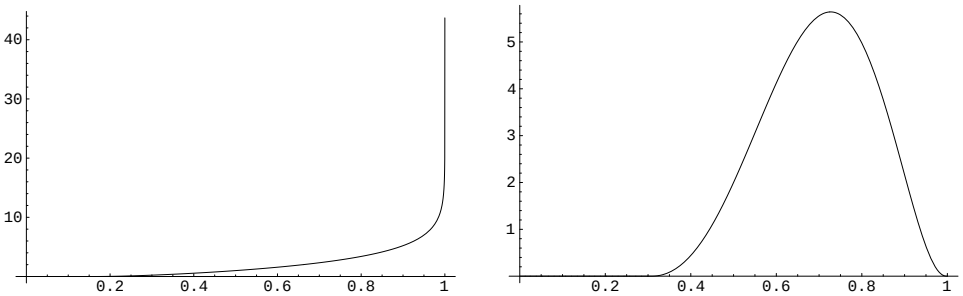


Fig. 1. The integrand with respect to u of $\mathrm{E}_f(q(x))$ for $q(x) = (x+1)^2 I_{(-1,\infty)}$ and $f(x)$ the standard normal density: naive algorithm (l.h.s.) and IS with $g(x)$ normal distribution with $\sigma = 2$ (r.h.s.).

section and the IS densities used there for the Sobol and the Niederreiter sequence as implemented in the GNU Scientific Library [GSL]. The results showed that, especially for intermediate and large dimensions (say larger than or equal to 8), the optimal IS density for those two low-discrepancy sequences is very similar to the optimal IS density for MC. The limited space does not allow to present the detailed results but the overall performance of low-discrepancy sequences in comparison to MC and lattice rules are presented in Sect. 3.

2.3 Lattice Rules

We started with very simple one-dimensional integrals to get an idea how to select IS densities that are useful for QMC. We made experiments to see for which types of functions on $(0,1)$ the integration error is small. Of course a constant function is always integrated exactly, but these experiments reminded us of the fact that just a few equally spaced points can be enough to integrate quite complicated smooth periodic functions (that is for a function which is periodically continued on $\mathbb{R}$). To obtain a similar result for higher dimensions we have to use lattice rules (see [Sloan and Joe 1994]), which allow for a convergence faster than $O(N^{-1+\varepsilon})$ when integrating smooth periodic functions. What is the connection to IS? We can try to select the IS density such that the resulting integrand with respect to u is smooth and periodic as we then have reasons to expect that the integration error is reduced or the convergence is even accelerated. For lattice rules the integration error tends to zero with rate $O(N^{-\alpha}(\log N)^{\alpha\, d})$ for periodic functions (where roughly spoken $\alpha = 1$ for continuous and $\alpha = k{+}1$ for k-times continuously differentiable functions). This rate can be much faster than the well known worst case rate of $O(N^{-1}\log^d N)$ for low-discrepancy sequences.

Figure 1 is a first indication that IS is able to transform a non-periodic integration problem into a smooth and periodic problem as the density and its derivative are tending to 0 for u tending to 0 or 1. As an IS density with higher tails leads to weights that converge to 0 faster it is easy to see that higher tails of the IS density also imply that the integrand and several derivatives are 0 for $u = 0$ and $u = 1$. This means that the integrand is smoother periodic if the tails of the IS density are higher. This fact is illustrated by Figure 2 that shows the same integration problem as Figure 1 but where σ has respective values 1.1, 3, and 5. Figure 2 also

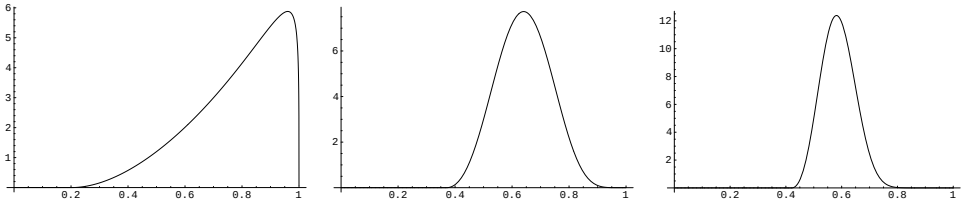


Fig. 2. The integrand of Figure 1 for $g(x)$ normal with: $\sigma = 1.2$ (l.h.s.), $\sigma = 3$ (center) and $\sigma = 5$ (r.h.s.).

demonstrates the trade-off between smoothing and periodizing the integrand on the one side and increasing the variance not too much on the other. Using larger values for σ leads to a smoother periodic integrand whereas the variance of the Monte Carlo estimator is increased.

We expect that the function $q(x)$ and especially its smoothness properties play an important role for the question whether IS with lattice rules can lead to a considerable error reduction for QMC integration. Also the selection of the IS density should have an impact on the integration error. For f standard normal the most natural candidate for the IS density is a normal density with $\sigma > 1$; it has the computational advantage that the evaluation of the weights is numerically very simple and due to the larger variance this IS density leads to an integrand which is 0 on the boundary of the unit cube and thus periodic. Other natural candidates would be the logistic distribution or the t-distribution which both have higher tails than the normal distribution and thus lead to an even smoother periodic integrand.

Sidi [1993] suggested a general approach for obtaining smooth periodic integrands (see also [Sloan and Joe 1994, p. 36]). It makes use of the "sine transform"

$$\phi(t) = t - \frac{1}{2\,\pi} \sin 2\pi t \;, \qquad \phi'(t) = 1 - \cos 2\pi t \;.$$

It can be used to obtain generally applicable IS densities that should be especially well suited for lattice rules as ϕ' is a periodic function that is integrated without error by lattice rules. For a density $f(x)$ with CDF $F(x)$ this transformation leads to the IS density

$$g(x) = \frac{f(x)}{\phi'(F(x))} \text{ with } w(x) = \phi'(F^{-1}(u)) \text{ and } G^{-1}(u) = \phi(F^{-1}(u)) \,.$$

If $F^{-1}(u)$ is available the algorithm using this IS density is very simple. Fast algorithms that only require evaluation of the CDF F are available for this task, see [Hörmann and Leydold 2003].

As a first example that demonstrates nicely the potential of combining IS with lattice rules we try an integral with respect to a smooth function $q(x)$.

2.3.1 *Example: Calculating the covariance of a multinormal vector.* To calculate the covariance of two variables (e.g. variables 2 and 3) of a multinormal vector Y with mean 0 we have to evaluate the expectation $\mathrm{E}_f(y_2\, y_3)$ where f is the density of the respective multinormal distribution. Note that similar calculations are common in Bayesian statistics as the posterior distribution is often close to a multinormal

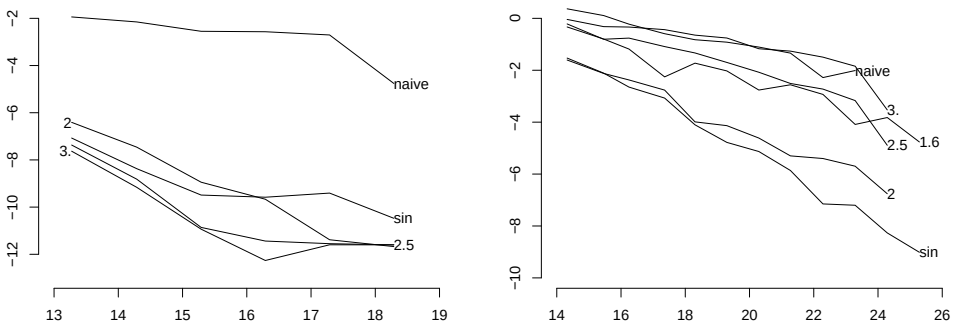


Fig. 3. Covariance example: the common logarithm of the relative RMSE for calculating the covariance is plotted against the base-2 logarithm of the sample size; The numbers in the plot show the parameter σ used for the normal IS density; dimension 4 (l.h.s.) and dimension 12 (r.h.s)

distribution and the correlation of two variables is of interest as it is the correlation between two parameter estimates. It is well known that a multinormal vector is easily generated by transforming a standard normal vector with the Cholesky factor of the variance-covariance matrix (see, e.g., [Hörmann et al. 2004]). We can thus apply importance sampling directly to the standard normal vector X; the function $q(x)$ is then a quadratic form in the x_i's and thus smooth. As variance-covariance matrix we decided to take unit variances and all covariances equal to a fixed value of ρ. To assess the integration error obtained for different IS densities and sample sizes we repeated each integration 100 times for randomly selected values of ρ between 0.8 and 0.9 and calculated the relative root mean squared error of these 100 experiments.

In our experiments we used embedded lattice rules as suggested by Sloan and Joe [1994, Chapt. 10] and also utilized the parameters and the error estimate introduced there. We compared the "naive" algorithm (i.e. direct generation of the normal variates without IS), the normal IS density with several values for $\sigma > 1$ and also the sine transform. Figure 3 shows the resulting root mean squared error (RMSE) for several values of σ (which resulted in small RMSE) for the normal IS density together with the results for the naive algorithm and for the sine transform.

The results indicate that, contrary to the situation for MC and for other low-discrepancy sequences (see Sect. 3), importance sampling really helps to reduce the integration error when used together with lattice rules and a smooth function $q(x)$; comparing the results of "naive" with those of the best IS density the improvement is really impressing and we can also clearly see the improved rate of convergence. As the integrand of the naive experiment is not smooth when periodically continued on $\mathbb{R}$ (in opposition to the other integrands) this nicely supports the conjecture that lattice rules are superior only for smooth periodic functions. The results also indicate (as expected) that the optimal IS density changes with the sample size and especially with the dimension. The sine transform has worse performance than the other IS densities for dimension 4 and better for dimension 12.

2.3.2 *Expectations with respect to the exponential distribution.* As second example we consider the function

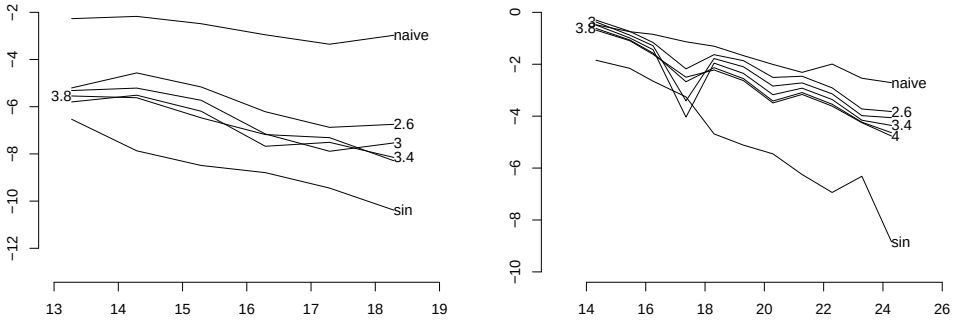


Fig. 4. Exponential distribution example: axes like in Figure 3. The numbers in the plot show the parameter b used for the Gamma(1/2) IS density; dimension 4 (l.h.s.) and dimension 12 (r.h.s)

$$q(x) = \left(\left(\sum_{i=1}^{d} x_i - c \right)^2 \right) + \sum_{i=1}^{d} (a_i - x_i\, b_i)^2$$

and calculate its expectation with respect to a vector of iid. exponential variates.

For integrals with respect to a vector of iid. exponential random variates the selection of the IS density $g(x)$ is less obvious. To obtain an integrand that vanishes for $u_i = 0$ it is necessary that $w(x) = 0$ for $x = 0$. To reach that aim we need an IS density with pole at 0. The Gamma(1/2,b) distribution, for example, is a sensible choice. Note that nobody would suggest such an IS density for a MC algorithm. In our experiments we chose the parameters a_i and c randomly with uniform distribution over $(0.5, 2)$ and the parameters b_i uniformly over $(0.2, 0.4)$. Again we repeated the experiment with 100 randomly selected parameter sets and report the relative RMSE in Figure 4.

The results show similar to the experiment above that importance sampling improves the lattice rule results by an impressive margin. The selection of the "strange" Gamma(1/2) IS density for the exponential distribution is useful. However, in this experiment the sine transform leads to the best results.

2.3.3 *Example: Option pricing for Asian Options.* As a third example we present the results of using importance sampling and lattice rules to evaluate the option price of an Asian option using the geometric average, which allows for a closed-form solution of the integral. (For option pricing by simulation see e.g. [Glasserman 2004] or [Charnes 2000].) This example is quite different to the two above as the pay-off function of an Asian option is continuous but not differentiable in each point.

The number of control points for an Asian option is equal to the dimension of the integral; we select the control intervals equally spaced, the time to maturity 0.25 years, the risk free rate equal to 0.05; value at present is 100 and the exercise price is randomly chosen from a uniform distribution over $(98, 102)$, the volatility uniformly over $(0.2, 0.4)$. Figure 5 reports the RMSE calculated for 100 randomly chosen parameter values.

Comparing with "naive" we can again see that IS is necessary for lattice rules to reach good results. The choice of the best IS density is not clear here. It changes

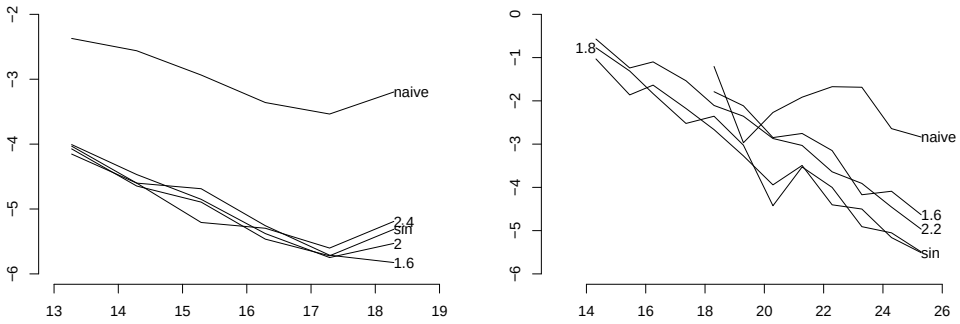


Fig. 5. Asian option pricing example; axes and labels like in Figure 3; dimension 4 (l.h.s.) and dimension 12 (r.h.s)

with the sample size and with the dimension. The sine transform is again a good choice but not better than using the normal distribution with a well chosen σ.

2.4 Quality of the Heuristic Error Bound

As pointed out by Sloan and Joe [1994] one important advantage of the method of "embedded lattice rules" is that they allow the computation of a simple heuristic error estimate. Thus we also checked the quality of this error estimate in our experiments and noticed that, not astonishingly, the error estimate works better for smooth integrands.

We calculated about 9000 integrals for each of our three examples. Analysis of the error statistics of all these integrals for dimensions 4, 8 and 12 for the three IS densities that lead to the smallest error and for all sample sizes we used showed that:

—For the covariance example the percentage of errors larger than the true error is about 1.3%.

—For the exponential distribution example the percentage of errors larger than the true error is about 12%.

—For the Asian option pricing example the percentage of errors larger than the true error is about 14%.

The observed error rate is small to moderate. So we may say that the error estimate is useful but we should not trust it blindly.

3. COMPARING THE PERFORMANCE

We have seen above that the combination of IS with lattice rules works good for small to moderate dimensions. We can even expect a faster rate of convergence if the function $q(x)$ is smooth and the dimension not too high. But for the practitioner it is of more importance to see which method works best for moderate sample sizes up to say 2^{20} or 2^{23} he is normally using for his real-world problems. In this section we therefore compare the best IS lattice method with the best IS Monte-Carlo and the best IS low-discrepancy sequence method. As a first comment it is interesting that in all our examples IS is very important for the lattice methods as the integrand is only smooth as periodically continued function when using IS. In contrast to the

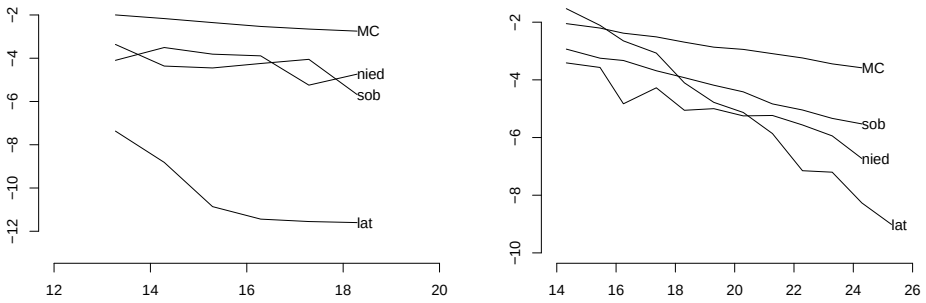


Fig. 6. Covariance example: dimension 4 (l.h.s.) and dimension 12 (r.h.s)

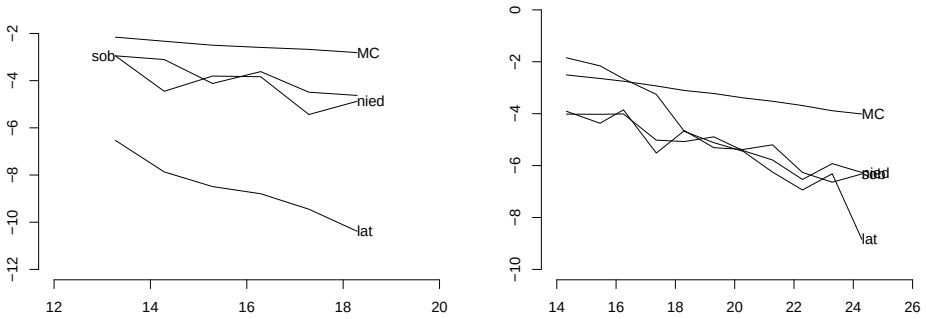


Fig. 7. Exponential distribution example: dimension 4 (l.h.s.) and dimension 12 (r.h.s)

situation with lattice rules for the integration examples we tried (including the three examples demonstrated here and four others) the difference between IS and naive simulation is small for Monte Carlo and also for other low-discrepancy sequences.

Looking at Figures 6 to 8 we can see that for dimension 4 the situation is very clear: IS lattice is superior to the other methods for the two smooth integrands as it has a faster rate of convergence. The situation remains very similar for dimension 8 so we are omitting the corresponding figures here. For dimension 12 the situation is different. For the covariance example the faster convergence is still clearly visible, probably because the problem is in its nature only a 2-dimensional problem; but even here the break-even point is for a sample-size above 2^{20}. For the exponential distribution example and the Asian option pricing example the faster convergence of the IS lattice method is no longer clearly visible for dimension 12. Nevertheless IS lattice is the best among the methods we have tried.

Another finding from Figures 6 to 8 is that it does pay to use quasi-Monte Carlo methods for well-behaved integrands if precise results are required. Even in dimension 12 they are clearly superior to Monte Carlo.

4. CONCLUSIONS

To calculate the expectation of smooth functions with respect to a random vector with iid. components the combination of lattice rules with importance sampling can speed up the convergence of the numerical integration. The importance sampling

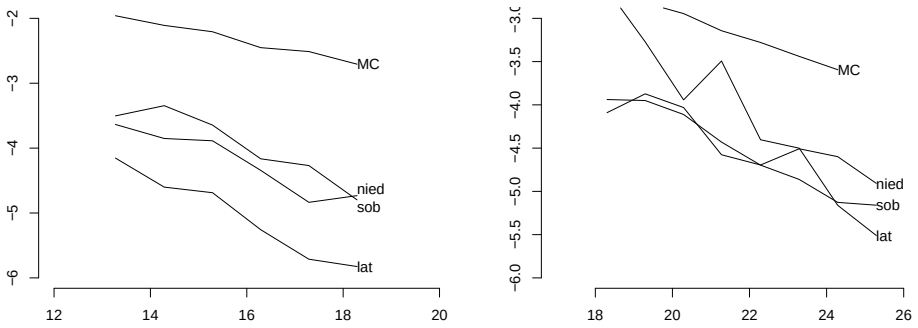


Fig. 8. Asian option pricing example; dimension 4 (l.h.s.) and dimension 12 (r.h.s)

density must be selected such that the integrand is transformed into a periodic function. For small dimensions the advantage of this approach is very clear and for many smooth functions this combination remains superior to all (quasi-) Monte Carlo methods up to dimension 12. This is also true for functions that are not differentiable everywhere.

REFERENCES

Charnes, J. M. 2000. Using simulation for option pricing. In *Proceedings of the 2000 Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, Eds. 151–157.

Geweke, J. 1989. Bayesian inference in econometric models using monte carlo integration. *Econometrica 57,* 6, 1317–1339.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.

GSL. *GSL – GNU Scientific Library, version 1.4*. `http://www.gnu.org/software/gsl/`.

Hesterberg, T. 1995. Weighted average importance sampling and defensive mixture distributions. *Technometrics 37,* 2, 185–194.

Hörmann, W. and Leydold, J. 2003. Continuous random variate generation by fast numerical inversion. *ACM Trans. Model. Comput. Simul. 13,* 4, 347–362.

Hörmann, W. and Leydold, J. 2005. Quasi importance sampling. Preprint Series 57, Department of Applied Statistics and Data Processing, Wirtschaftsuniversität Wien, Augasse 2–6, A-1090 Wien, Austria. `http://epub.wu-wien.ac.at/english/`, document ID: `oai:epub.wu-wien.ac.at:epub-wu-01_9fe`.

Hörmann, W., Leydold, J., and Derflinger, G. 2004. *Automatic Nonuniform Random Variate Generation*. Springer-Verlag, Berlin Heidelberg.

Owen, A. B. 2006. Halton sequences avoid the origin. *SIAM Reviews 48*, 487–503.

Sidi, A. 1993. A new variable transformation for numerical integration. In *Numerical Integration IV*, H. Brass and G. Hämmerlin, Eds. Birkhäuser, Basel, 359–373.

Sloan, I. H. and Joe, S. 1994. *Lattice Methods for Multiple Integration*. Clarendon Press, Oxford.