Quasi Importance Sampling

Wolfgang Hörmann [a,b,1] and Josef Leydold [a,*]

[a] *University of Economics and Business Administration, Department of Statistics and Mathematics, Augasse 2-6, A-1090 Vienna, Austria*

[b] *IE Department, Boğaziçi University Istanbul, 80815 Bebek-Istanbul, Turkey*

Abstract

There arise two problems when the expectation of some function with respect to a nonuniform multivariate distribution has to be computed by (quasi-) Monte Carlo integration: the integrand can have singularities when the domain of the distribution is unbounded and it can be very expensive or even impossible to sample points from a general multivariate distribution. We show that importance sampling is a simple method to overcome both problems.

Key words: quasi-Monte Carlo method, nonuniform random variate generation, inversion method, importance sampling, Markov chain Monte Carlo
PACS: 02.70.-c, 02.70.Tt
1991 MSC: 65C05, 65C10, 65D32,

---

1 Introduction

The computation of expectations $\mathrm{E}_f(q)$ of functions $q(\mathbf{x})$ in $\mathbb{R}^d$ with respect to a given distribution F with probability density function $f(\mathbf{x})$ is one of the most important applications of *Monte Carlo* methods (*MC*):

$$\mathrm{E}_f(q) = \int_{\mathbb{R}^d} q(\mathbf{x})\, f(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int_{\mathbb{R}^d} q(\mathbf{x})\, \mathrm{d}F(\mathbf{x}) \,. \tag{1}$$

---

* Corresponding author. Tel +43 1 313 36–4695. FAX +43 1 313 36–738
Email addresses: `whoer@statistik.wu-wien.ac.at` (Wolfgang Hörmann), `leydold@statistik.wu-wien.ac.at` (Josef Leydold).

[1] This work was supported by the Austrian Science Foundation (FWF), project no. P16767-N12.

The naive (plain) Monte Carlo estimator is given by

$$\mathrm{E}_f(q) \approx \frac{1}{N} \sum_{i=1}^{N} q(\mathbf{X}_i) \ , \quad \mathbf{X}_i \sim F \tag{2}$$

where the $\mathbf{X}_i$ are drawn from distribution F. This approach is simple and works well in many situations, see, e.g., [1,2]. The convergence rate of this estimator is $\mathcal{O}(N^{-1/2})$, provided that the variance of the integrand is bounded. The sample of nonuniform random vectors is usually generated by transforming uniform (pseudo-) random numbers into random variates and vectors that follow the target distribution. For this task a lot of methods exist, see [3,4] for surveys. However, it must be noted that generating points $\mathbf{X} \sim F$ for (moderately) high dimensions is difficult or even prohibitively slow if F is a general multivariate distribution.

The convergence rate often can be increased when *highly uniform point sets* (*HUPS*, also called *low discrepancy sequences* or *quasi-random numbers*) are used instead of (pseudo-) random points. Such methods are called *quasi-Monte Carlo* methods (*QMC*). By the Koksma-Hlawka inequality the absolute integration error for computing $\int_{(0,1)^d} q(\mathbf{x})\,\mathrm{d}\mathbf{x}$ is bounded by the product of the total variation of q in the sense of Hardy and Krause and the star discrepancy of the used point set, i.e., $V(q) D^*_N(\mathbf{x}_1, \ldots, \mathbf{x}_N)$, see [5]. There exist HUPS where the star discrepancy (and thus the QMC estimator) converges with $\mathcal{O}(N^{-1} \log^d(N))$. When $\mathrm{E}_f(q)$ has to be evaluated with respect to some nonuniform distribution F with bounded domain, similar results exist [6].

The QMC approach requires point sets with low F-discrepancy, i.e., point sets $\{\mathbf{X}_i\}$ where $\{F(\mathbf{X}_i)\}$ has low discrepancy [7]. Such point sets are also created by applying appropriate transformation methods on low discrepancy sequences. However, for general multivariate distributions such transformations are often hard to find and/or numerically very expensive. Moreover, these may introduce singularities into our integration problem and thus convergence is not guaranteed by the Koksma-Hlawka inequality.

In this contribution we show that by means of *importance sampling* these two difficulties can be avoided. Such importance densities can be very simple.

2 Quasi-Monte Carlo and Nonuniform Distributions

We have to transform low discrepancy point sets into sets of points with low F-discrepancy when we want to compute the QMC estimator. However, the transformation methods that have been developed for nonuniform random variate generation often cannot be applied for QMC, since these destroy the structure of the underlying point set. Moreover, the theory of nonuniform random numbers does not directly apply when quasi-random numbers are used. (Halton [8] formulates a first theory of *quasi-probability.*) It is the authors' opinion that the problem of generating nonuniform *random* points and that of generating nonuniform *quasi*-random points should be seen as different problems. For the first one we have to transform uniform random numbers into random points. The correctness of the transformation is verified using probability theory. The structure of the used uniform *pseudo*-random point set is usually not taken into consideration. The latter problem of generating *quasi*-random points should be interpreted as transforming the integration problem with respect to F over $\mathbb{R}^d$ into an equivalent one over $(0,1)^d$ with respect to the Lebesgue measure, i.e., $\int_{\mathbb{R}^d} q(\mathbf{x})\, f(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int_{(0,1)^d} q^*(\mathbf{u})\, \mathrm{d}\mathbf{u}$ for some function q^*. This is required as HUPS are constructed to work for the integral over the unit cube. From this point of view it is somewhat surprising that most papers dealing with QMC methods for evaluating expectations $\mathrm{E}_f(q)$ do not consider the problem of appropriate transformations.

In dimension one ($d = 1$) the *inversion method* is usually the method of choice as it preserves the structure of the underlying point set. Quasi-/pseudo- random variates are then generated using $X = F^{-1}(U)$ where U is a $\mathcal{U}(0,1)$ quasi-/pseudo- random number. For our integration problem (1) this is equivalent to integrate

$$\int_{\mathbb{R}} q(x)\, f(x)\, \mathrm{d}x = \int_{(0,1)} q(F^{-1}(u))\; \mathrm{d}u \approx \frac{1}{N} \sum_{i=1}^{N} q(F^{-1}(U_i)) \;, \quad U_i \sim \mathcal{U}(0,1)\;.$$

There are some problems with this approach. First, the inverse CDF, F^{-1}, is often not given in closed form and thus numerical methods that only compute $F^{-1}(u)$ approximately have to be used. There exist fast methods for this task, see e.g. [9,10], but they either require the CDF or compute it by integrating the density function numerically. In the multivariate case the inversion method can be applied to the marginal distributions, if the components of the random vector $\mathbf{X}$ are stochastically independent. Otherwise, the conditional distribution method must be used which can be seen as the multivariate generalization of the inversion method. It requires the (inverse) CDF of conditional distributions of marginal distributions [4] which is practically never available in practice. Moreover, the F-discrepancy is increased when the components are not independent [11].

A more serious problem in the framework of QMC is the fact that the integrand $q(F^{-1}(u))$ is often unbounded and thus has unbounded variation when the support of the distribution is unbounded. This is for example the case when the m-th moment of the i-th variable has to be computed in Bayesian inference where $q(\mathbf{x}) = x_i^m$, or in derivative pricing in financial engineering when $q(\mathbf{x})$ behaves like $\exp(\sum_{i=1}^d x_i)$. In this case the Koksma-Hlawka inequality does not apply and convergence of the QMC estimator is not guaranteed. Nevertheless, it is well known that QMC often works for functions with singularities and other functions where the integrand has unbounded variation. Owen [12,13] has worked out theory to explain this observation for uniform distribution F when the integrand satisfies some growth condition and the low-discrepancy sequence avoids these singularities. He has shown that if an integrand q satiesfies $|\partial^u q(\mathbf{x})| \leq C||\mathbf{x} - \mathbf{z}||_p^{-A-|u|}$ for some $C < \infty$, $1 \leq p < \infty$ and some $0 < A < d$ near a singularity $\mathbf{z}$, then the estimator converges with $\mathcal{O}(N^{(-1+\varepsilon)((d-A)/d})$, ($\partial^u f$ denotes the partial derivative with respect to all coordinates in the index set u). Owen [12] gives examples where QMC integration of unbounded functions works very well and examples where the convergence is much worse. The results have been extended to the case where q is integrated with respect to a nonununiform distribution F with compact support [14,15].

An alternative method of transforming uniform point sets into those following the target distribution is due to Hlawka and Mück [11,16]. It implicitly avoids the points of singularities at the corners [14,15]. It can be seen as a rough approximation of the inversion method where the resolution of the points is $1/N$. It only works (well) for distributions with independent components and compact support.

The *acceptance/rejection* method [4] is the most efficient generation method for nonuniform random points in the framework of Monte Carlo integration. This principle can also be used for QMC integration. However, we need an additional dimension and the rejection step can be interpreted as integration of the indicator function of the set of acceptance. Such a function has unbounded variation. Computational experiences show that convergence can be as bad as for the naive Monte Carlo method. A possible solution is to replace the indicator function by some continuous function with the same integral (*smoothed rejection*, see [17,18]).

3 Importance Sampling

Both problems described above, namely difficulties of generating F- distributed points and integrands with singulatities, can be avoided by means of *importance sampling*. It is not astonishing that IS and QMC can be combined and this was also mentioned in some publications, see e.g. [19,20]. But it was not yet presented as a simple and very useful tool to avoid all problems with the generation of nonuniform quasi-/pseudo- random points. Otherwise, we think that techniques like smoothed rejection would not have been developed.

Importance sampling is based on the following identity. Let G and g be the distribution function and the density function of some distribution, called *importance distribution* in the sequel. Then

$$\mathrm{E}_f(q) = \int_{\mathbb{R}^d} q(\mathbf{x})\, f(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int_{\mathbb{R}^d} q(\mathbf{x}) \frac{f(\mathbf{x})}{g(\mathbf{x})}\, g(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int_{\mathbb{R}^d} q(\mathbf{x}) w(\mathbf{x})\, \mathrm{d}G(\mathbf{x})$$

where $w(\mathbf{x}) = f(\mathbf{x})/g(\mathbf{x})$. The importance distribution can be chosen such that it is easy to generate a sample of points that follow the importance density. In the case of QMC this will be the inversion method. In dimension one ($d = 1$) we then have the estimator

$$\begin{aligned} E_f(q) &= \int_{\mathbb{R}} q(x)\, f(x)\, \mathrm{d}x = \int_{(0,1)} q(G^{-1}(u))\, w(G^{-1}(u))\ \mathrm{d}u \\ &\approx \frac{1}{N} \sum_{i=1}^{N} q(G^{-1}(U_i)) w(G^{-1}(U_i))\ , \quad U_i \sim \mathcal{U}(0,1)\ . \end{aligned}$$

By a proper choice of the importance density g the integrand has bounded variation. It is enough that g has higher tails than the product of $q(x)f(x)$ to get rid of the singularity problem.

3.1 Distributions with Independent Components

For many important simulation applications (for example most discrete event simulations and many problems in finance) the random vectors of the input distribution F have independent components. Thus we can write our density $f(\mathbf{x}) = \prod_{i=1}^{d} f_i(x_i)$ with marginal densities f_i. In this case the generation of nonuniform (quasi-) random points is simple as the inversion method is applied to each component of the point from the underlying HUPS. But if the domain of F is unbounded the integrand can have singularities on the boundary of the unit cube, see Sect. 2.

As noted above we can eliminate these singularities by a proper choice of the importance density $g(\mathbf{x}) = \prod_{i=1}^{d} g_i(x_i)$. We use the estimator $E_f(q) \approx \frac{1}{N} \sum_{i=1}^{N} q(G^{-1}(\mathbf{U}_i)) w(G^{-1}(\mathbf{U}_i))$ where $\mathbf{U}_i \sim \mathcal{U}(0,1)^d$ and G^{-1} is computed componentwise. However, often f is only known up to an (unknown) multiple, see the Bayesian example below. Then we have to replace this estimate by the *ratio estimate* [21]

$$E_f(q) \approx \frac{\sum_{i=1}^{N} q(G^{-1}(\mathbf{U}_i)) w(G^{-1}(\mathbf{U}_i))}{\sum_{i=1}^{N} w(G^{-1}(\mathbf{U}_i))}\ .$$

It is possible that QMC is working very well in cases where the conditions for the Koksma-Hlawka inequality fail and where Owen's improved inequality predict a slower convergence. Moreover, these convergence results are of asymptotic nature and do not guarantee smaller errors for importance sampling when moderately large sample sizes are used. Thus we ran experiments to look at the error for typical sample sizes used in practice with N a power of two between 2^6 and 2^{22}. We experimented with f the normal distribution and integrands $q(x) = \sum x_i^r$, $q(x) = (\sum x_i)^r$ ($r = 2$ and 4) and $q(x) = \exp(\sigma \sum x_i)$ ($\sigma = 0.1$ and 0.5). We used a base-2 Niederreiter sequence [22], and a Sobol sequence [23] and repeated the experiments for fixed values of N with 1000 different randomly shifted point sets. We have compared our results with QMC without importance sampling and with MC (with and without importance sampling). The nonuniform random variates are generated by inversion. (We also have included the Box-Muller method with behaves similar.)

In all experiments we found a clear superiority of importance sampling over inversion, and clearly faster rates of convergence for dimensions one and two. The choice of the importance sampling density is not critical as long as it has tails that are heavier than those of f. In all cases importance sampling with QMC was clearly superior to importance sampling with MC. We took these results as a confirmation that our general considerations indicating that importance sampling is better than inversion for the use with QMC are correct. For larger dimensions the advantage is becoming smaller rapidly. Moreover, the results depend on q and f. As one example Figure 1 shows the root mean squared error (RMSE) for $q(x) = \exp(0.5 \sum x_i)$ and dimensions 2 and 8. In our experiments we encountered cases where inversion was a bit better than importance sampling for dimensions above five for all sample sizes we tried. We conjecture that this behavior changes for even larger sample sizes but we cannot prove this and must also omit that this question is of little practical importance as sample sizes larger than 10^6 are not common in practice.

We also ran experiments where f was the t-distribution with eight degrees of freedom and $q(x) = \sum x_i^4$. Then the integrand has a very "high" singularity at 0 and thus importance sampling should be much better than QMC with inversion. As expected the RMSE of QMC importance sampling is much (at least ten times) smaller than that of inversion and the faster rate of convergence is clearly visible for that example.

We also made two other observations in our experiments. First, the ratio estimate was better than the classical importance sampling estimate in some of the examples and never worse in the others. So we recommend to use the ratio estimate for importance sampling with QMC. The second experience was that finding a good importance sampling density for MC is a quite different objective than finding a good importance density for QMC. In particular, choosing g very close to f by some sophisticated method like the hat function of

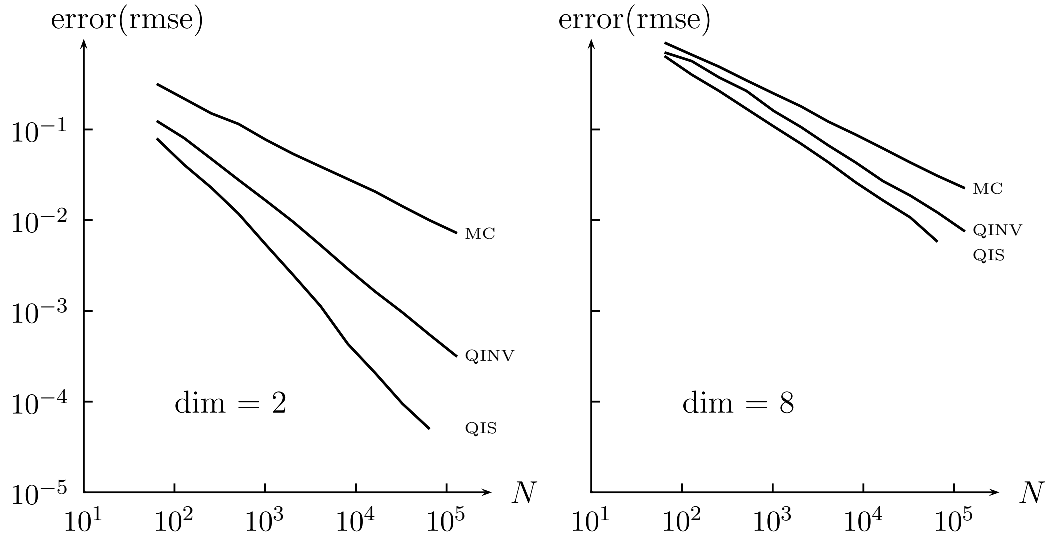

Fig. 1. Root mean squared error for f independent normal and $q(x) = \exp(0.5 \sum x_i)$. MC ... MC with inversion, QINV ... QMC with inversion, QIS ... QMC with importance sampling.

transformed density rejection (see [4, §4]) was inferior to "simple" methods like using the double exponential distribution with a rough estimate for optimal scale parameters in the QMC setting (whereas the sophisticated method was superior for MC).

3.2 General Multivariate Distributions

If the multivariate distribution F is not the product of independent univariate distributions the generation of random vectors from the distribution F is a major problem. Here importance sampling allows to select a simple distribution G not too different from F. The selection of G for general multivariate and multimodal distributions is by no means straight forward. For integrals that occur for Bayesian parameter estimation problems f is often unimodal. It is then general practice (see e.g. [24]) to search for the mode of f and use the Hessian in the mode for a linear transformation that transforms F into a multivariate distribution with approximately independent components. Using double-exponential importance sampling densities for these independent components leads to an importance sampling algorithm that is well suited for QMC methods as well. Of course this approach can again be interpreted as a change of variable in the original integral (1).

We demonstrate the application of QMC for such problems by means of a random effect Poisson model for pump-failure data (see [25]). It is possible to

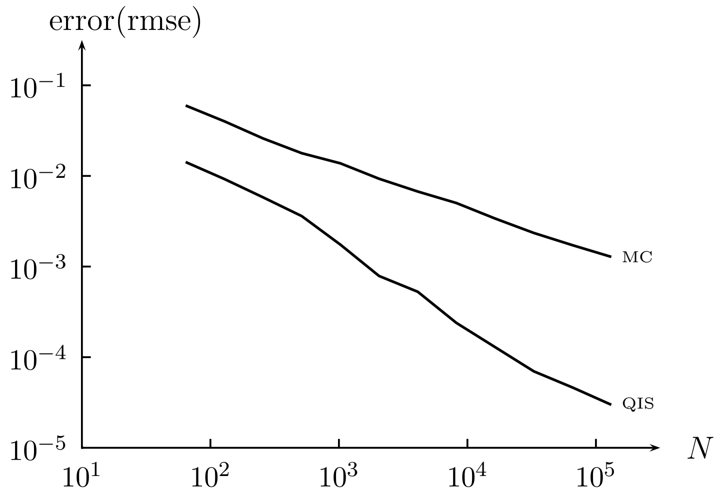

Fig. 2. Root mean squared error for estimating the first parameter of the pump model.

write down the posterior of the two hyper-parameters of that model in closed form. To obtain the Bayesian parameter estimates we have to calculate the expectations of the two marginal distributions of the posterior; this means that we calculate our standard integral with f the posterior density and $q(x) = x_i$. Note that the integral of f is not known in this application and we can therefore only use the ratio estimate.

Figure 2 shows the error of MC and QMC integration for different sample sizes. The advantage of QMC is very obvious: For QMC with $n = 1000$ we can expect the same error as for MC with $n = 10^5$. This is of considerable practical importance as the evaluation of the density f (i.e., the posterior distribution of the Bayesian model) for this example like for most others is very expensive. Up to our knowledge the application of quasi importance sampling procedures to Bayesian problems is not suggested in the literature. A possible reason is that in recent years Markov Chain Monte Carlo (MCMC) methods are more frequently used than importance sampling. There are some attempts to use the QMC approach for MCMC (quasi-Markov chain Monte Carlo method). However, the concept is certainly far from clear to interpret and theoretical considerations assume strong conditions on the underlying point sets [26]. Up to now there are no empirical results clearly indicating the practical value of the combination of QMC with MCMC. We are therefore convinced that our results of Figure 2 clearly suggest that importance sampling is the more natural way to speed up Bayesian integration by applying QMC.

In [27] simple and easy to use importance densities for multivariate integration are proposed.

4 Conclusions

We have demonstrated that the combination of QMC and importance sampling leads to a simple and very useful method for calculating expectations with respect to multivariate distributions. This method can be used to circumvent the definition of nonuniform quasi-random variates. Interpreted as a parameter transformation method it can be shown that it allows to get rid of singularities of the integrand which increases the speed of convergence of QMC. In the case of complicated multivariate distributions the application of QMC techniques is much easier for importance sampling than for Markov chain Monte Carlo methods.

References

[1] G. S. Fishman, Monte Carlo. Concepts, algorithms, and applications, Springer Series in Operations Research, Springer, New York, 1996.

[2] M. Evans, T. Swartz, Approximating Integrals via Monte Carlo and Deterministic Methods, Oxford University Press, 2000.

[3] L. Devroye, Non-Uniform Random Variate Generation, Springer-Verlag, New-York, 1986.

[4] W. Hörmann, J. Leydold, G. Derflinger, Automatic Nonuniform Random Variate Generation, Springer-Verlag, Berlin Heidelberg, 2004.

[5] H. Niederreiter, Random number generation and quasi-Monte Carlo methods, Vol. 63 of SIAM CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1992.

[6] B. Tuffin, Simulation accélérée par les méthodes de Monte Carlo et quasi-Monte Carlo: théorie et applications, Ph.D. thesis, Université de Rennes 1 (1997).

[7] K.-T. Fang, Y. Wang, Number-theoretic Methods in Statistics, Vol. 51 of Monographs on Statistics and Applied Probability, Chapman and Hall, London, 1994.

[8] J. H. Halton, An outline of quasi-probability. Why quasi-Monte-Carlo methods are statistically valid and how their errors can be estimated statistically, Monte Carlo Methods and Applications 10 (3–4) (2004) 183–196.

[9] W. Hörmann, J. Leydold, Continuous random variate generation by fast numerical inversion, ACM Trans. Model. Comput. Simul. 13 (4) (2003) 347–362.

[10] J. H. Ahrens, K. D. Kohrt, Computer methods for efficient sampling from largely arbitrary statistical distributions, Computing 26 (1981) 19–31.

[11] E. Hlawka, R. Mück, Über eine Transformation von gleichverteilten Folgen II, Computing 9 (1972) 127–138.

[12] A. B. Owen, Halton sequences avoid the origin, SIAM Reviews 48, to appear.

[13] A. B. Owen, Randomized QMC and point singularities, in: H. Niederreiter, D. Talay (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2004, Springer-Verlag, Berlin, 2005, to appear.

[14] J. Hartinger, R. F. Kainhofer, R. F. Tichy, Quasi-Monte Carlo algorithms for unbounded, weighted integration problems, Journal of Complexity 20 (5) (2004) 654–668.

[15] J. Hartinger, R. F. Kainhofer, Non-uniform low-discrepancy sequence generation and integration of singular integrands, in: H. Niederreiter, D. Talay (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2004, Springer-Verlag, Berlin, 2005, to appear.

[16] E. Hlawka, Gleichverteilung und Simulation, Sitzungsber., Abt. II, Österr. Akad. Wiss., Math.-Naturwiss. Kl. 206 (1997) 183–216.

[17] B. Moskowitz, R. E. Caflisch, Smoothness and dimension reduction in quasi-Monte Carlo methods, Math. Comput. Modelling 23 (8–9) (1996) 37–54.

[18] J. Leydold, W. Hörmann, Smoothed transformed density rejection, Monte Carlo Methods and Applications 10 (3–4) (2004) 393–402.

[19] R. Schürer, Adapive quasi-Monte Carlo integration based on MISER and VEGAS, in: H. Niederreiter (Ed.), Monte Carlo and Quasi-Monte Carlo Methods 2002, Springer, Berlin, Heidelberg, 2004, pp. 393–406.

[20] G. Ökten, Error reduction techniques in quasi-Monte Carlo integration, Mathematical and Computer Modelling 30 (7–8) (1999) 61–69.

[21] T. Hesterberg, Weighted average importance sampling and defensive mixture distributions, Technometrics 37 (2) (1995) 185–194.

[22] P. Bratley, B. L. Fox, H. Niederreiter, Implementation and tests of low-discrepancy sequences, ACM Trans. Model. Comput. Simul. 2 (3) (1992) 195–213.

[23] P. Bratley, B. L. Fox, Algorithm 659: implementing Sobol's quasirandom sequence generator, ACM Trans. Math. Software 14 (1) (1988) 88–100.

[24] J. Geweke, Bayesian inference in econometric models using monte carlo integration, Econometrica 57 (6) (1989) 1317–1339.

[25] E. George, U. Makov, A. Smith, Conjugate likelihood distributions, Scandinavian Journal of Statistics 20 (1993) 147–156.

[26] A. B. Owen, S. D. Tribble, A quasi-Monte Carlo Metropolis algorithm, technical report, available at `http://www-stat.stanford.edu/~owen/reports/` (2005).

[27] W. Hörmann, New importance sampling densities, technical report, available at `http://statistik.wu-wien.ac.at/papers/2005-05.wh.pdf` (2005).