

# On Figures of Merit for Randomly-Shifted Lattice Rules

Pierre L'Ecuyer and David Munger

**Abstract** Randomized quasi-Monte Carlo (RQMC) can be seen as a variance reduction method that provides an unbiased estimator of the integral of a function  $f$  over the  $s$ -dimensional unit hypercube, with smaller variance than standard Monte Carlo (MC) under certain conditions on  $f$  and on the RQMC point set. When  $f$  is smooth enough, the variance converges faster, asymptotically, as a function of the number of points  $n$ , than the MC rate of  $\mathcal{O}(1/n)$ . The RQMC point sets are typically constructed to minimize a given parameterized measure of discrepancy between their empirical distribution and the uniform distribution. These parameters can give different weights to the different subsets of coordinates (or lower-dimensional projections) of the points, for example. The ideal parameter values (to minimize the variance) depend very much on the integrand  $f$  and their choice (or estimation) is far from obvious in practice. In this paper, we survey this question for randomly-shifted lattice rules, an important class of RQMC point sets, and we explore the practical issues that arise when we want to use the theory to construct lattices for applications. We discuss various ways of selecting figures of merit and for estimating their ideal parameters (including the weights), we examine how they can be implemented in practice, and we compare their performance on examples inspired from real-life problems. In particular, we look at how much improvement (variance reduction) can be obtained, on some examples, by constructing the points based on function-specific figures of merit compared with more traditional general-purpose lattice-rule constructions.

---

Pierre L'Ecuyer and David Munger  
Département d'Informatique et de Recherche Opérationnelle, Université de Montréal,  
C.P. 6128, Succ. Centre-Ville, Montréal, H3C 3J7, Canada, [lecuyer@iro.umontreal.ca](mailto:lecuyer@iro.umontreal.ca),  
[mungerd@gmail.com](mailto:mungerd@gmail.com)

## 1 Introduction: Monte Carlo and Randomized Quasi-Monte Carlo

We are concerned with the problem of estimating the integral of a function  $f : [0, 1]^s \rightarrow \mathbb{R}$  over the  $s$ -dimensional unit hypercube  $[0, 1]^s = \{\mathbf{u} = (u_1, \dots, u_s) : 0 \leq u_j < 1 \text{ for all } j\}$ , by evaluating  $f$  at  $n$  points in this hypercube and taking the average. The integral can be written as

$$\mu = \mu(f) = \int_{[0,1]^s} f(\mathbf{u}) \, d\mathbf{u} = \mathbb{E}[f(\mathbf{U})] \quad (1)$$

where  $\mathbb{E}$  denotes the mathematical expectation and  $\mathbf{U} = (U_1, \dots, U_s) \sim \mathcal{U}(0, 1)^s$  (a random vector uniformly distributed over  $(0, 1)^s$ ). A large class of applications fit this framework [14, 18, 19].

The standard *Monte Carlo* (MC) method generates  $n$  independent realizations of  $\mathbf{U}$ , say  $\mathbf{U}_0, \dots, \mathbf{U}_{n-1}$ , and estimates  $\mu$  by

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=0}^{n-1} f(\mathbf{U}_i).$$

This estimator is unbiased and its variance converges as  $\mathcal{O}(n^{-1})$  when  $n \rightarrow \infty$ .

*Randomized quasi-Monte Carlo* (RQMC) employs an estimator of the same form,

$$\hat{\mu}_{n,\text{rqmc}} = \frac{1}{n} \sum_{i=0}^{n-1} f(\mathbf{U}_i) \quad (2)$$

where  $\mathbf{U}_i \sim U[0, 1]^s$  for each  $i$ , so  $\mathbb{E}[\hat{\mu}_{n,\text{rqmc}}] = \mu$  (the estimator is unbiased), but the  $\mathbf{U}_i$ 's are no longer independent. The aim is to have  $\text{Var}[\hat{\mu}_{n,\text{rqmc}}] < \text{Var}[\hat{\mu}_n]$ . For this, the randomized points are constructed so that  $P_n = \{\mathbf{U}_0, \dots, \mathbf{U}_{n-1}\} \subset [0, 1]^s$  covers  $[0, 1]^s$  more evenly than typical independent random points, in the sense that some selected (expected) measure of discrepancy between the empirical distribution of  $P_n$  and the uniform distribution over  $[0, 1]^s$  is smaller. Two popular classes of RQMC point sets are randomly-shifted lattices and digitally-shifted nets. For background on RQMC methods, including lattice rules, see [14, 16, 18, 19, 20, 21] and the references given there.

In this paper, we focus on randomly-shifted lattice rules, where  $P_n$  is the intersection of a randomly-shifted lattice with  $[0, 1]^s$  [21]. For any given square-integrable  $f$  (that is, for which  $\text{Var}[f(\mathbf{U}_i)] < \infty$ ),  $\text{Var}[\hat{\mu}_{n,\text{rqmc}}]$  can be written explicitly in terms of the square Fourier coefficients of  $f$  and on the lattice. Conceptually, one could compute the optimal lattice for  $f$  by solving an optimization problem that minimizes this variance expression with respect to the lattice parameters. However, this is impractical because these Fourier coefficients are usually unknown, and there are infinitely many, so we have to

rely on suboptimal strategies. The variance expression is usually replaced by a figure of merit with fewer parameters, and those parameters are selected by heuristic methods that take into account the class of functions  $f$  to be considered.

The aim of this paper is two-fold. First, we give a partial overview of current knowledge on randomly-shifted lattice rules from a practical viewpoint. Then, we examine the issues that arise when we want to exploit this theoretical knowledge in applications. In particular, we explore the impact of the choice of figure of merit, the choice of weights given to the different subsets of coordinates in discrepancy measures, we compare empirical performances of these choices in terms of the RQMC variance, we compare the convergence rate for the variance that are typically observed empirically (for reasonable values of  $n$ ) to the theoretical asymptotic rates (when  $n$  goes to infinity) which are based on bounds that are usually not tight, and see what can be observed in the common situation where the integrand is discontinuous or unbounded. We always assume that  $s$  is fixed. We do not consider complexity and tractability issues.

The remainder is organized as follows. In Section 2, we recall basic definitions and known results on randomly-shifted lattice rules and the corresponding explicit variance expressions. In Section 3, we discuss how we could conceivably select a lattice adaptively to reduce the variance expression if the Fourier coefficients of  $f$  were known, or could be estimated easily when needed. The main purpose is to show the difficulty of doing this. We describe and implement a selection method that starts with a large set of lattices and eliminates them one by one, by visiting a sequence of important terms in the variance expression and by keeping, at each step, only the lattices that eliminate those large variance terms. The procedure is very effective on the small examples on which we try it, where the Fourier coefficients are known. But for typical real-life problems, the Fourier coefficients are unknown and estimating them would be too time-consuming, so we need other heuristics. In Section 4, we examine previously-proposed figures of merit defined as discrepancies that assign a weight to each subset of coordinates (or projection), using a functional ANOVA decomposition of  $f$ , and we suggest ways of specifying the weights as functions of the ANOVA variance components, for Sobolev classes of integrands with square integrable partial derivatives up a given order. When  $s$  is large, having a different weight for each projection may give a model with too many parameters. Parameterized choices of weights with fewer parameters are discussed in Section 5. They include order-dependent weights and product weights, in particular, and we examine ways of setting those weights. In Section 6, we discuss figures of merit defined in terms of the lengths of shortest nonzero vectors in dual lattices. In Section 7, we briefly recall the algorithms we have used to search for good lattices with respect to the selected figures of merit. Then, in the following sections, we perform empirical experiments with some examples. Our goal is to estimate the convergence rate of the RQMC variance as a function of  $n$  and

the variance reduction compared with standard MC, in the practical range of values of  $n$ , and to assess the impact of the choice of figure of merit (and weights) on this variance, at least for our selected examples. Motivated by the fact that discontinuous integrands are very frequent in applications, we start in Section 9 with simple indicator functions. We give examples where lattice rules are still effective, but where a standard measure of discrepancy can be (sometimes) a very bad predictor of the performance. This illustrates the difficulty of defining good and robust figures of merit in general. In one case, we make the integrand continuous by taking a conditional expectation with respect to one random variable (after generating the other ones) and we examine the effect of this. In Section 10, we consider a stochastic activity network example inspired from a real-life application, where the integrand is also an indicator function, and we extend the study made in Section 9 to this slightly more complicated setting. The examples of Section 9 were constructed as simplifications of that of Section 10, to try to better understand the behavior of randomly-shifted lattice rules in those situations. Finally, in Section 11, we consider the pricing of Asian-style options, with and without a barrier. Our examples have been inspired from real-life problems, and as it turns out, none of them satisfies the smoothness conditions that guarantee a fast convergence of the variance (such as  $\mathcal{O}(n^{-4})$ ) with the best lattice constructions. This seems to correspond to many typical real-life problems. An online appendix provides detailed results of our experiments.

The good news is that in the great majority of cases, most reasonable choices of figures of merit and weights provide lattices that perform well, for those examples, provided that none of the relevant weights is zero and the irrelevant weights do not dominate too much. This means that there is no need to work hard to estimate the ANOVA variances accurately. Faced with an important application, one may want to spend a small fraction of the available computing budget at the beginning to estimate ANOVA components very roughly, or to just explore a few choices of weights and compare the variances in pilot runs. Also, the convergence rate of the variance observed empirically for reasonable values of  $n$  (up to a few millions) is slower than the asymptotic rates proved theoretically for smooth functions. On the other hand, this observed rate is always better than  $\mathcal{O}(1/n)$ , even for discontinuous and unbounded integrands, in our examples.

## 2 Randomly-Shifted Lattice Rules

An *integration lattice* is a discrete vector space defined by

$$L_s = \left\{ \mathbf{v} = \sum_{j=1}^s z_j \mathbf{v}_j \text{ such that each } z_j \in \mathbb{Z} \right\},$$

where  $\mathbf{v}_1, \dots, \mathbf{v}_s \in \mathbb{R}^s$  are linearly independent over  $\mathbb{R}$  and  $L_s$  contains  $\mathbb{Z}^s$ , the integer vectors. A *lattice rule* approximates  $\mu$  by the average of  $f(\mathbf{u}_0), \dots, f(\mathbf{u}_{n-1})$ , where  $P_n^0 = \{\mathbf{u}_0, \dots, \mathbf{u}_{n-1}\} = L_s \cap [0, 1)^s$ . Almost all lattice rules used in practice have *rank 1*, which means that the points of  $P_n^0$  can be enumerated as  $\mathbf{u}_i = i\mathbf{v}_1 \bmod 1$  for  $i = 0, \dots, n-1$ , where  $n\mathbf{v}_1 = \mathbf{a}_1 = (a_1, \dots, a_s) \in \{0, 1, \dots, n-1\}$  is the *generating vector*. We have a *Korobov rule* if  $\mathbf{a}_1$  has the form  $\mathbf{a}_1 = (1, a, a^2 \bmod n, \dots)$  for some integer  $a \in \mathbb{Z}_n$ . For more details on lattice rules, see [4, 11, 19, 21]. For any subset of coordinates  $\mathbf{u} \subseteq \{1, \dots, s\}$ , the *projection*  $L_s(\mathbf{u})$  of  $L_s$  over the subspace determined by  $\mathbf{u}$  is also a lattice, in  $|\mathbf{u}|$  dimensions. In this paper, we assume that all lattices are of rank 1 and that the coordinates  $a_1, \dots, a_s$  of the generating vector are *all* relatively prime to  $n$  (when  $n$  is prime, this is automatic). When the latter holds for the first coordinate, the lattice rule is called a rank-1 simple rule [4]. Here we are assuming more: our assumption implies that the projection of  $P_n^0$  over the subspace determined by any nonempty subset of coordinates contains exactly  $n$  points and this projection is always  $\{0, 1/n, \dots, (n-1)/n\}$  in the case of a single coordinate. Therefore, there is no need to measure the uniformity of these one-dimensional projections.

The point set  $P_n^0$  can be turned into an RQMC point set  $P_n$  by a *random shift modulo 1*, defined as follows [5, 21]: Generate a single random point  $\mathbf{U}$  uniformly over  $(0, 1)^s$  and add it to each point of  $P_n^0$ , modulo 1, coordinate-wise:  $\mathbf{U}_i = (\mathbf{u}_i + \mathbf{U}) \bmod 1$ . Then, each  $\mathbf{U}_i$  is uniformly distributed over  $[0, 1)^s$  and  $\hat{\mu}_{n,\text{rqmc}}$  is an unbiased estimator of  $\mu$ , while  $P_n$  inherits the lattice structure of  $P_n^0$ .

A key issue is whether (and when)  $\hat{\mu}_{n,\text{rqmc}}$  has smaller variance than the MC estimator  $\hat{\mu}_n$ . An exact expression for the variance can be obtained in terms of the Fourier coefficients of the integrand  $f$ , as follows. If  $f$  has Fourier expansion

$$f(\mathbf{u}) = \sum_{\mathbf{h} \in \mathbb{Z}^s} \hat{f}(\mathbf{h}) e^{2\pi i \mathbf{h}^t \mathbf{u}},$$

then (see [15])

$$\text{Var}[\hat{\mu}_{n,\text{rqmc}}] = \sum_{\mathbf{0} \neq \mathbf{h} \in L_s^*} |\hat{f}(\mathbf{h})|^2, \tag{3}$$

where  $L_s^* = \{\mathbf{h} \in \mathbb{R}^s : \mathbf{h}^t \mathbf{v} \in \mathbb{Z} \text{ for all } \mathbf{v} \in L_s\} \subseteq \mathbb{Z}^s$  is the *dual lattice* to  $L_s$ . This variance depends on  $f$  and on  $L_s$ . For any given  $f$ , an optimal lattice  $L_s$  from the viewpoint of variance reduction would minimize  $D^2(P_n^0) = \text{Var}[\hat{\mu}_{n,\text{rqmc}}]$ . This suggests a figure of merit of the general form

$$\mathcal{M}_w(P_n^0) = \sum_{\mathbf{0} \neq \mathbf{h} \in L_s^*} w(\mathbf{h}), \tag{4}$$

with weights  $w(\mathbf{h})$  that mimic the anticipated behavior of the  $|\hat{f}(\mathbf{h})|^2$ . It may be tempting to refer to (4) as a measure of discrepancy. However it does not

necessarily measure a departure from the uniform distribution. For certain functions  $f$ , the best lattice does not necessarily cover the space very evenly.

It is known [21] that if  $f$  has square-integrable mixed partial derivatives up to order  $\alpha$  (an integer), and the periodic continuations of its derivatives up to order  $\alpha - 2$  are *continuous* across the unit cube boundaries, then

$$|\hat{f}(\mathbf{h})|^2 = \mathcal{O}((\max(1, h_1), \dots, \max(1, h_s))^{-2\alpha}). \quad (5)$$

Moreover, there is a rank-1 lattice with  $\mathbf{v}_1 = \mathbf{v}_1(n)$  such that

$$\mathcal{P}_{2\alpha} = \sum_{\mathbf{0} \neq \mathbf{h} \in L_s^*} (\max(1, h_1), \dots, \max(1, h_s))^{-2\alpha} = \mathcal{O}(n^{-2\alpha+\delta}) \quad (6)$$

for any  $\delta > 0$ . Note that  $\mathcal{P}_{2\alpha}$  in (6) is the RQMC variance for a *worst-case*  $f$  having

$$|\hat{f}(\mathbf{h})|^2 = (\max(1, h_1), \dots, \max(1, h_s))^{-2\alpha},$$

so the convergence order in (6) applies when (5) holds. This worst-case  $f$  is not necessarily representative of functions encountered in applications, and therefore,  $\mathcal{P}_{2\alpha}$  is not necessarily the most appropriate figure of merit.

For the preceding bound to hold with  $\alpha \geq 2$ , the periodic continuation of  $f$  must be continuous. When it is not, which is often the case,  $f$  can be transformed into a function  $\tilde{f}$  having the same integral and a continuous periodic continuation, by compressing the graph of  $f$  horizontally along each coordinate and then making a mirror copy with respect to  $1/2$ . This gives  $\tilde{f}(u_1, \dots, u_s) = f(v_1, \dots, v_s)$  where  $v_j = 2u_j$  for  $u_j < 1/2$  and  $v_j = 2(1 - u_j)$  for  $u_j \geq 1/2$ . In practice, instead of changing  $f$ , we would stretch the (randomized) points by a factor of two along each coordinate, and fold them back. This is equivalent. That is, each coordinate  $U_{i,j}$  of  $\mathbf{U}_i$  is replaced by  $2U_{i,j}$  if  $U_{i,j} < 1/2$  and by  $2(1 - U_{i,j})$  otherwise. This is the *baker's transformation*. When  $f$  is sufficiently smooth, this can reduce the RQMC variance from  $\mathcal{O}(n^{-2+\delta})$  to  $\mathcal{O}(n^{-4+\delta})$  [12].

### 3 Adaptive Search for Lattices that Avoid the Large Fourier Coefficients

Searching for a lattice that minimizes the variance expression (3) for each  $f$  that we want to integrate is certainly impractical, because the Fourier coefficients are usually unknown and there are infinitely many. If we estimate them, we would have to do it for each  $f$  and this is likely to take more time than applying RQMC to estimate  $\mu$ . We nevertheless explore empirically, in this section, what we could do if we knew (or could estimate on-demand, at low cost) those Fourier coefficients and how much we could gain by exploiting this knowledge (or by finding the optimal lattice for the problem at hand in

any other way). In situations where the gain can be significant, it may be worthwhile to investigate ways of identifying the most important Fourier coefficients.

We start with a simple function for which we know the Fourier expansion. But even in that case, the figure of merit (the variance) in (3) involves an infinite number of terms. Heuristic ways of handling this could be to truncate the sum to a finite subset  $B \subset \mathbb{Z}^s$ ,

$$\sum_{\mathbf{0} \neq \mathbf{h} \in L_s^* \cap B} |\hat{f}(\mathbf{h})|^2,$$

or to the largest  $q$  square coefficients  $|\hat{f}(\mathbf{h})|^2$ . But this is hard to implement. The following heuristic truncates the sum adaptively by exploring the dual space. It makes sense in the situation where the  $|\hat{f}(\mathbf{h})|^2$  tend to decrease with each  $|h_j|$ . It starts with a large set  $\mathcal{L}$  of lattices (or a large set of generating vectors  $\mathbf{v}_1$ , for a given  $n$ ). Then the method searches for vectors  $\mathbf{h}$  with large weights  $w(\mathbf{h}) = |\hat{f}(\mathbf{h})|^2$ , via a neighborhood search starting at  $\mathbf{h} = \mathbf{0}$ , keeping a sorted list (as in Dijkstra's shortest path algorithm), and eliminates successively from  $\mathcal{L}$  the lattices whose dual contains  $\mathbf{h}$  for the next largest  $w(\mathbf{h})$  found so far, until a single lattice remains. It is stated as Algorithm 1 (the scope of the **while** and **for** statements are specified by the indentation). The ordered set  $\mathcal{Q}$  can be implemented as a priority queue. This algorithm requires a definition of neighborhood in the space  $\mathbb{Z}^s$  of vectors  $\mathbf{h}$ . For example, one can define the neighborhood of  $\mathbf{h}$ ,  $\mathcal{N}(\mathbf{h})$ , as the set of vectors that differ from  $\mathbf{h}$  by only one coordinate, and by one unit only. When the  $|\hat{f}(\mathbf{h})|^2$  are unknown, we may think of estimating them whenever they are needed in the algorithm, dynamically.

---

**Algorithm 1** : Dual-Space Exploration

---

**Require:** a set of lattices  $\mathcal{L}$  and a weight function  $w$

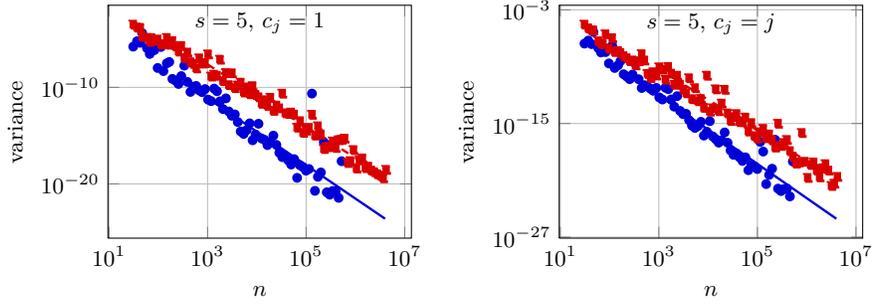
```

 $\mathcal{Q} \leftarrow \mathcal{N}(\mathbf{0})$  // vectors  $\mathbf{h}$  to be visited, sorted by decreasing weight  $w(\mathbf{h})$ 
 $\mathcal{M} \leftarrow \mathcal{N}(\mathbf{0})$  // vectors  $\mathbf{h}$  that have already entered  $\mathcal{Q}$ 
while  $|\mathcal{L}| > 1$  do
     $\mathbf{h} \leftarrow$  remove first vector from  $\mathcal{Q}$ 
    for all lattices  $L_s \in \mathcal{L}$  such that  $\mathbf{h} \in L_s^*$  do
        remove  $L_s$  from  $\mathcal{L}$ 
        if  $|\mathcal{L}| = 1$  then
            return the single lattice  $L_s \in \mathcal{L}$  and exit
        for all  $\mathbf{h}' \in \mathcal{N}(\mathbf{h}) \setminus \mathcal{M}$  do
            add  $\mathbf{h}'$  to  $\mathcal{M}$  and to  $\mathcal{Q}$  with priority (weight)  $w(\mathbf{h}')$ 

```

---

One can also define a *component-by-component* (CBC) version of this construction algorithm, as follows. For each coordinate  $j$ ,  $j = 1, \dots, s$ , we apply the algorithm for a set  $\mathcal{L}$  of  $j$ -dimensional lattices with common (fixed)  $j - 1$  first coordinates, determined in the previous steps, and we select the  $j$ th coor-



**Fig. 1** Estimated variance vs.  $n$ , for  $s = 5$ , in log-log scale, with  $c_j = 1$  (left) and  $c_j = j$  (right), using lattices constructed with the dual-space exploration algorithm ( $\bullet$ ), and the CBC algorithm with the  $\mathcal{P}_{\gamma,2}$  criterion ( $\blacksquare$ ).

minate by visiting all  $j$ -dimensional vectors  $\mathbf{h}$  having nonzero  $j$ th coordinate, as in Algorithm 1.

*Example 1.* To experiment with this algorithm, we consider the product V-shaped function

$$f(\mathbf{u}) = \prod_{j=1}^s \frac{|4u_j - 2| + c_j}{1 + c_j},$$

for which

$$\hat{f}(\mathbf{h}) = \prod_{\{j: h_j \text{ is odd}\}} \frac{4}{(1 + c_j)\pi^2 h_j^2}.$$

We take  $s = 5$  dimensions, first with  $c_j = 1$  and then with  $c_j = j$ . We applied the CBC version of the dual-space exploration algorithm for prime values of  $n$  ranging from  $2^5 - 1 = 31$  to  $2^{19} - 1 = 524,287$ , to construct a 5-dimensional lattice for each  $n$ , then we estimated the RQMC variance for this lattice by the empirical variance with 100 independent random shifts.

Figure 1 shows the empirical variance as a function of  $n$ , in the lower (dark) line. The upper line represents the RQMC variance with lattices obtained by a CBC construction using the criteria  $\mathcal{P}_{\gamma,2}$  defined in (9), with weights selected based on estimated ANOVA variance components as explained in Section 4. This is arguably the best available construction method for general applications among those that we have tried in our experiments. The figure shows that for this small example, our dual-space exploration method does much better. The reason is that by constructing the lattice in terms of a figure of merit that takes into account the individual Fourier coefficients, we can be more accurate in selecting the vectors  $\mathbf{h}$  that we want to eliminate from the dual lattice, and thus kick out more of the important terms from the variance expression (3), than if we use a criterion such as  $\mathcal{P}_{\gamma,2}$  that just put weights on subsets of coordinates.

For the dual-space exploration, with  $n = 2^{16} + 1$ , the variance was reduced by a factor of  $1.7 \times 10^{14}$  for  $c_j = 1$  and  $3.0 \times 10^{16}$  for  $c_j = j$ , compared with MC. Empirically, the variance decreases approximately as  $\mathcal{O}(n^{-3.46})$  for  $c_j = 1$  and  $\mathcal{O}(n^{-3.61})$  for  $c_j = j$ . (There is one outlying value, for  $n = 2^{17} - 1 = 131,071$ , where the algorithm did poorly for  $c_j = 1$ , as can be seen in Figure 1.) For the lattice constructions based on  $\mathcal{P}_{\gamma,2}$ , on the other hand, the variance was reduced by only  $1.8 \times 10^{12}$  in the best case, and decreased empirically (approximately) as  $\mathcal{O}(n^{-3.24})$ .

We also tried with the  $\mathcal{M}'_{\gamma,2}$  criterion defined in (14), and the results were worse than with  $\mathcal{P}_{\gamma,2}$ .

The dual-space exploration algorithm performs much better, for this small example, than the other methods discussed in forthcoming sections. However, in typical situations, the Fourier coefficients are unknown, not always monotonously decreasing with the components of  $\mathbf{h}$ , have to be estimated during the exploration, and the dimension can be much larger than 5. Then, this search approach is unlikely to remain practical and effective. We will discuss alternatives in the following.

## 4 ANOVA Decomposition and Projection-Dependent Weights

Given that the Fourier expansion and the sum (3) have too many terms to be convenient figures of merit for selecting the lattice parameters, one could seek decompositions of  $f$  into a smaller number of terms than in (3), and define measures that take into account the relative importance of those terms. A popular one is the functional ANOVA decomposition [8, 18, 20], where  $f(\mathbf{u}) = f(u_1, \dots, u_s)$  is written as

$$f(\mathbf{u}) = \sum_{\mathbf{u} \subseteq \{1, \dots, s\}} f_{\mathbf{u}}(\mathbf{u}) = \mu + \sum_{i=1}^s f_{\{i\}}(u_i) + \sum_{i,j=1:j \neq i}^s f_{\{i,j\}}(u_i, u_j) + \dots$$

where

$$f_{\mathbf{u}}(\mathbf{u}) = \int_{[0,1]^{|\bar{\mathbf{u}}|}} f(\mathbf{u}) \, d\mathbf{u}_{\bar{\mathbf{u}}} - \sum_{\mathbf{v} \subset \mathbf{u}} f_{\mathbf{v}}(\mathbf{u}_{\mathbf{v}})$$

$\bar{\mathbf{u}}$  is the complement of  $\mathbf{u}$ , and  $\mathbf{u}_{\mathbf{v}}$  refers to the projection of  $\mathbf{u}$  on the subspace determined by  $\mathbf{v}$ . The *MC variance* then decomposes as

$$\sigma^2 = \sum_{\mathbf{u} \subseteq \{1, \dots, s\}} \sigma_{\mathbf{u}}^2, \quad \text{where } \sigma_{\mathbf{u}}^2 = \text{Var}[f_{\mathbf{u}}(\mathbf{U})].$$

The variance components  $\sigma_{\mathbf{u}}^2$  can be estimated by the algorithm given in [25], using either MC or (preferably) RQMC to estimate the integrals.

For any  $\mathbf{h} \in \mathbb{Z}^s$ , let

$$\mathbf{u}(\mathbf{h}) = \mathbf{u}(h_1, \dots, h_s) = \{j \in \{1, \dots, s\} : h_j \neq 0\}.$$

The RQMC variance with a randomly-shifted lattice rule decomposes as:

$$\text{Var}[\hat{\mu}_{n,\text{rqmc}}] = \sum_{\emptyset \neq \mathbf{u} \subseteq \{1, \dots, s\}} \sum_{\mathbf{h} \in L_s^* : \mathbf{u}(\mathbf{h}) = \mathbf{u}} |\hat{f}(\mathbf{h})|^2 = \sum_{\mathbf{u} \subseteq \{1, \dots, s\}} \text{Var}[\hat{\mu}_{n,\text{rqmc}}(f_{\mathbf{u}})]. \quad (7)$$

The idea here is to adopt a criterion as in (4), but with weights  $w(\mathbf{h})$  that depend on a smaller number of parameters, namely one parameter per projection  $\mathbf{u}$ . For this, following [7] and others, we take

$$w(\mathbf{h}) = \gamma_{\mathbf{u}(\mathbf{h})} \prod_{j \in \mathbf{u}} h_j^{-2\alpha} \quad (8)$$

for all  $\mathbf{h} \in \mathbb{Z}^s$ , where  $\alpha$  is a positive integer to be selected, and the  $\gamma_{\mathbf{u}}$  are arbitrary positive real numbers which we call *projection-dependent weights*. Some authors call them *general weights* [6, 7], although their form is much less general than the arbitrary weights  $w(\mathbf{h})$  in (4). With the weights (8), the figure of merit (4) becomes the *weighted*  $\mathcal{P}_{2\alpha}$  criterion [7]:

$$\begin{aligned} \mathcal{P}_{\gamma, 2\alpha}(P_n^0) &= \sum_{\mathbf{0} \neq \mathbf{h} \in L_s^*} \gamma_{\mathbf{u}(\mathbf{h})} (\max(1, h_1), \dots, \max(1, h_s))^{-2\alpha} \\ &= \sum_{\emptyset \neq \mathbf{u} \subseteq \{1, \dots, s\}} \frac{1}{n} \sum_{i=0}^{n-1} \gamma_{\mathbf{u}} \left[ \frac{-(-4\pi^2)^\alpha}{(2\alpha)!} \right]^{|\mathbf{u}|} \prod_{j \in \mathbf{u}} B_{2\alpha}(u_{i,j}), \end{aligned} \quad (9)$$

where  $\mathbf{u}_i = (u_{i,1}, \dots, u_{i,s}) = i\mathbf{v}_1 \bmod 1$  is the  $i$ th lattice point before the shift,  $|\mathbf{u}|$  is the cardinality of  $\mathbf{u}$ , and  $B_{2\alpha}$  is the Bernoulli polynomial of order  $2\alpha$ .

This criterion comes naturally in the following setting. Let  $\mathcal{F}_\alpha$  be the class of functions  $f : [0, 1]^s \rightarrow \mathbb{R}$  for which for each  $\mathbf{u} \subseteq \{1, \dots, s\}$ , the partial derivative of order  $\alpha$  with respect to  $\mathbf{u}$  is square integrable, and (if  $\alpha \geq 2$ ) the partial derivatives of orders 0 to  $\alpha - 2$  of the periodic continuation of  $f$  over  $\mathbb{R}^s$  are continuous. Define the square variation of  $f \in \mathcal{F}_\alpha$  by

$$V_\gamma^2(f) = \sum_{\mathbf{u} \subseteq \{1, \dots, s\}} V_\gamma^2(f_{\mathbf{u}}) = \sum_{\emptyset \neq \mathbf{u} \subseteq \{1, \dots, s\}} \frac{1}{\gamma_{\mathbf{u}} (4\pi^2)^{\alpha|\mathbf{u}|}} \int_{[0,1]^{|\mathbf{u}|}} \left| \frac{\partial^{|\mathbf{u}|}}{\partial \mathbf{u}^\alpha} f_{\mathbf{u}}(\mathbf{u}) \right|^2 d\mathbf{u} \quad (10)$$

(which depends on the  $\gamma_{\mathbf{u}}$ 's). Then, for any constant  $K > 0$ , the largest RQMC variance over the class of functions  $f \in \mathcal{F}_\alpha$  for which  $V_\gamma^2(f) \leq K$  is equal to  $K\mathcal{P}_{\gamma, 2\alpha}(P_n^0)$ , and the maximum is reached for a worst-case function whose square Fourier coefficients are

$$|\hat{f}(\mathbf{h})|^2 = K\gamma_{\mathbf{u}(\mathbf{h})}(\max(1, h_1), \dots, \max(1, h_s))^{-2\alpha}.$$

See [6, 14] for the details. The constant  $K$  is just a scale factor that can be incorporated in the weights  $\gamma_{\mathbf{u}}$ , so we can assume that  $K = 1$ . The worst-case function can then be written as

$$f_{\alpha}^*(\mathbf{u}) = \sum_{\mathbf{u} \subseteq \{1, \dots, s\}} \sqrt{\gamma_{\mathbf{u}}} \prod_{j \in \mathbf{u}} \frac{(2\pi)^{\alpha}}{\alpha!} B_{\alpha}(u_j).$$

The ANOVA variance components for this function are

$$\sigma_{\mathbf{u}}^2 = \gamma_{\mathbf{u}} \left[ \text{Var}[B_{\alpha}(U)] \frac{(4\pi^2)^{\alpha}}{(\alpha!)^2} \right]^{|\mathbf{u}|} = \gamma_{\mathbf{u}} \left[ |B_{2\alpha}(0)| \frac{(4\pi^2)^{\alpha}}{(2\alpha)!} \right]^{|\mathbf{u}|} \stackrel{\text{def}}{=} \gamma_{\mathbf{u}}(\kappa(\alpha))^{-|\mathbf{u}|} \quad (11)$$

where  $\kappa(\alpha)$  is a constant that depends on  $\alpha$ . In particular, we have

$$\kappa(1) = \frac{3}{\pi^2} \approx 0.30396, \quad \kappa(2) = \frac{45}{\pi^4} \approx 0.46197, \quad \kappa(3) \approx 0.49148,$$

and  $\kappa(\alpha)$  increases with  $\alpha$  for  $\alpha \geq 1$  and converges to  $1/2$  when  $\alpha \rightarrow \infty$ .

To be consistent with our choice of  $\mathcal{P}_{\gamma, 2\alpha}$  as a criterion, we can select the weights  $\gamma_{\mathbf{u}}$  as if the function  $f$  that we want to integrate has the same form as  $f_{\alpha}^*$ . That is, we take the weights given by the formula

$$\gamma_{\mathbf{u}} = \sigma_{\mathbf{u}}^2(\kappa(\alpha))^{|\mathbf{u}|},$$

in which the variance components  $\sigma_{\mathbf{u}}^2$  are replaced by estimates. These estimates can be obtained with the algorithm of [25], for example. This formula can be generalized slightly to

$$\gamma_{\mathbf{u}} = \sigma_{\mathbf{u}}^2 \rho^{|\mathbf{u}|}, \quad (12)$$

where  $0 < \rho \leq 1$  is a constant to be selected. In view of the behavior of  $\kappa(\alpha)$ , it makes sense to take  $\rho \leq 0.5$ , and smaller when we think that  $f$  is less smooth.

It is known that for any  $\alpha > 1$ , any  $\delta > 0$ , and any choices of weights  $\gamma_{\mathbf{u}}$ , there are rank-1 lattices for which  $\mathcal{P}_{\gamma, 2\alpha}(P_n^0) = \mathcal{O}(n^{-2\alpha+\delta})$ , and the corresponding vectors  $\mathbf{v}_1$  can be constructed explicitly one coordinate at a time, by a component-by-component construction method [6].

## 5 Further Heuristics for Choosing the Weights

In (9), there are  $2^s - 1$  parameters  $\gamma_{\mathbf{u}}$  to specify, which is too many when  $s$  is large. It is also hard to estimate these  $\sigma_{\mathbf{u}}^2$  with reasonable relative error when they are small compared with  $\sigma^2$  and this typically occurs for most

subsets  $\mathbf{u}$  when  $|\mathbf{u}|$  increases. This motivates the introduction of more parsimonious models for the weights, with fewer parameters. As mentioned in the first paragraph of Section 2, the one-dimensional projections are all the same under our assumptions, so the weights of the one-dimensional subsets  $|\mathbf{u}|$  are irrelevant and we can set them to zero in the selection criterion; that is, restrict the sum in (9) to the subsets  $\mathbf{u}$  of cardinality  $|\mathbf{u}| \geq 2$ . We always do that in our experiments when searching for good lattices. Note that multiplying all weights by the same constant has no impact on the selection of  $\mathbf{v}_1$ , since it does not change the relative importance of the projections, so we can fix one of them (the largest one, for example) to 1. But there still remains  $2^s - s - 2$  projections weights to specify.

One way to reduce the number of parameters in (9) (and the likelihood of overfitting) is to bundle (partition) the subsets  $\mathbf{u}$  in subgroups, and force the same  $\gamma_{\mathbf{u}}$  within each subgroup. A well-known example of this is to take *order-dependent weights*, where  $\gamma_{\mathbf{u}}$  depends only on the cardinality of  $\mathbf{u}$ , say  $\gamma_{\mathbf{u}} = \gamma_r$  when  $|\mathbf{u}| = r$ , for  $r = 2, \dots, s$ . To specify those  $\gamma_r$ , we can estimate each  $\sigma_r^2 = \sum_{\{\mathbf{u}:|\mathbf{u}|=r\}} \sigma_{\mathbf{u}}^2$ , which represents the total variance captured by the  $\binom{s}{r}$  projections of order  $r$ , and plug it in the formula

$$\gamma_r = C \rho^r \sigma_r^2 \binom{s}{r}^{-1},$$

where  $C > 0$  is an arbitrary scaling constant. This gives  $s - 1$  parameters to estimate.

In one special case, we can simply assume that  $\gamma_r = \gamma^{r-2}$  for all  $r \geq 2$ , for some constant  $\gamma$ , and estimate  $\gamma$  by least-squares fitting of the linear regression model

$$\ln C + r \ln \rho + 2 \ln \sigma_r - \ln \binom{s}{r} = (r - 2) \ln \gamma + \varepsilon_r$$

(for example), by finding  $\ln C$  and  $\ln \gamma$  that minimize  $\sum_{r=2}^{\infty} \varepsilon_r^2$ . The resulting weights are *geometric order-dependent weights*. With *constant order-truncated weights*, one simply takes  $\gamma_{\mathbf{u}} = 1$  for  $|\mathbf{u}| \leq d$  and  $\gamma_{\mathbf{u}} = 0$  otherwise, for a given integer  $d \geq 2$ . Wang [26] suggests this with  $d = 2$ .

A different type of parameterization, used in [10, 11, 24], assigns a weight  $\gamma_j$  to each coordinate  $j$  and uses the *product weights*  $\gamma_{\mathbf{u}} = \prod_{j \in \mathbf{u}} \gamma_j$ . Again, we can estimate the parameters  $\gamma_j$  by matching the ANOVA variances, ignoring the one-dimensional projections. One way of doing this is to fit the weights (12) where the variance components are estimated, over all *two-dimensional* projections, via a least-squares procedure. Then we rescale all the weights by a constant factor to match the ratio of average estimated weights (12) over the *three-dimensional* projections to that over the two-dimensional projections.

More specifically, we first minimize

$$R = \sum_{k=1}^s \sum_{j=1}^{k-1} \left( \tau_j \tau_k - \rho^2 \sigma_{\{j,k\}}^2 \right)^2$$

with respect to  $\tau_1, \dots, \tau_s$ , where  $\tau_j$  can be viewed as the unscaled weight for projection  $j$ , and where the variance components  $\sigma_{\mathbf{u}}^2$  for  $|\mathbf{u}| = 2$  are replaced by their estimates. Differentiating this expression with respect to  $\tau_j$  and equating to 0, we obtain, for each  $j$ ,

$$\tau_j \sum_{k=1, k \neq j}^s \tau_k^2 = \sum_{k=1, k \neq j}^s \tau_k \rho^2 \sigma_{\{j,k\}}^2.$$

We solve this (heuristically) by an iterative fixed-point algorithm:

$$\tau_j^{(0)} = \max_{k,l=1,\dots,s} \rho \sigma_{\{k,l\}}, \quad \tau_j^{(i+1)} = \frac{\sum_{k=1, k \neq j}^s \tau_k^{(i)} \rho^2 \sigma_{\{j,k\}}^2}{\sum_{k=1, k \neq j}^s \left( \tau_k^{(i)} \right)^2},$$

for  $i = 1, 2, \dots$ . We then rescale the weights via  $\gamma_j = c\tau_j$  where the constant  $c$  satisfies

$$\frac{\sum_{k=1}^s \sum_{j=1}^{k-1} \tau_j \tau_k}{\sum_{k=1}^s \sum_{j=1}^{k-1} \sum_{l=1}^{j-1} \tau_j \tau_k \tau_l} = c \frac{\sum_{k=1}^s \sum_{j=1}^{k-1} \rho^2 \sigma_{\{j,k\}}^2}{\sum_{k=1}^s \sum_{j=1}^{k-1} \sum_{l=1}^{j-1} \rho^3 \sigma_{\{j,k,l\}}^2}$$

in which the sum of weights of order 3 is again replaced by an estimate.

## 6 Figures of Merit Based on the Spectral Test

In view of the variance expression (3) and its decomposition (7), and because we normally expect the square Fourier coefficients  $|\hat{f}(\mathbf{h})|^2$  to decrease with the size of  $\mathbf{h}$  (when  $f$  is smooth, we know from (5) that these coefficients must converge at the given rate), it seems to make sense to define a criterion that penalizes the short non-zero vectors  $\mathbf{h}$  in the dual lattice  $L_s^*$ , as well as in the dual lattices  $(L_s(\mathbf{u}))^*$  to the projections  $L_s(\mathbf{u})$ . Note that  $(L_s(\mathbf{u}))^*$  is the projection over  $\mathbf{u}$  of  $\{\mathbf{h} \in L_s^* : \mathbf{u}(\mathbf{h}) \subseteq \mathbf{u}\}$ , but not the projection of  $L_s^*$  over  $\mathbf{u}$ .

For each  $\mathbf{u}$ , one can compute the Euclidean length  $\ell_{\mathbf{u}}$  of a shortest nonzero vector in  $(L_s(\mathbf{u}))^*$ . There is a known tight theoretical upper bound  $\ell_r^*(n)$  on the length of a shortest nonzero vector in a lattice with  $n$  points per unit of volume in  $r$  dimensions [3, 15], and we can divide  $\ell_{\mathbf{u}}$  by  $\ell_{|\mathbf{u}|}^*(n)$  to obtain a standardized measure between 0 and 1, and raise it to some power  $\beta > 0$ , for each  $\mathbf{u}$ , or take the reciprocal to obtain a measure of non-uniformity larger or equal to 1. To give more weight to more important projections, this

measure can in turn be multiplied by some weight  $\gamma_{\mathbf{u}}$ , for each  $\mathbf{u}$ . Then we can take either the sum or the minimum (worst case) over a selected class  $\mathcal{J}$  of nonempty subsets  $\mathbf{u}$  of  $\{1, \dots, s\}$ . The role of  $\beta$  is to amplify or reduce the relative importance of bad projections (those having a small value of  $\ell_{\mathbf{u}}/\ell_{|\mathbf{u}|}^*(n)$ ) in the criterion. This gives the following figures of merit

$$\mathcal{M}_{\gamma,\beta}(P_n^0) = \sum_{\mathbf{u} \in \mathcal{J}} \gamma_{\mathbf{u}} \left( \frac{\ell_{\mathbf{u}}}{\ell_{|\mathbf{u}|}^*(n)} \right)^{\beta}, \quad (13)$$

$$\mathcal{M}'_{\gamma,\beta}(P_n^0) = \sum_{\mathbf{u} \in \mathcal{J}} \gamma_{\mathbf{u}} \left( \frac{\ell_{|\mathbf{u}|}^*(n)}{\ell_{\mathbf{u}}} \right)^{\beta}, \quad (14)$$

$$\widetilde{\mathcal{M}}_{\gamma,\beta}(P_n^0) = \min_{\mathbf{u} \in \mathcal{J}} \gamma_{\mathbf{u}} \left( \frac{\ell_{\mathbf{u}}}{\ell_{|\mathbf{u}|}^*(n)} \right)^{\beta}, \quad \text{and} \quad (15)$$

$$\widetilde{\mathcal{M}}'_{\gamma,\beta}(P_n^0) = \max_{\mathbf{u} \in \mathcal{J}} \gamma_{\mathbf{u}} \left( \frac{\ell_{|\mathbf{u}|}^*(n)}{\ell_{\mathbf{u}}} \right)^{\beta}. \quad (16)$$

The criteria (13) and (15) are to be maximized, whereas (14) and (16) are to be minimized. In (15) and (16), only the quality of the worst-case projections matters, and we do not care about the quality of the other ones, whereas in (13) and (14), the quality of all the projections contributes to the sum, so these criteria encourage improvements on all projections, not only the worst ones. The two variants (15) and (16) are equivalent in terms of which lattice maximizes or minimizes them, if we invert the weights (although we do not invert the weights in our examples). On the other hand, (13) and (14) are really different. For a fixed  $\beta$  and fixed weights, in (13) the bad projections have a small importance in the sum (they only “fail to score high”) whereas in (14) they have more importance because they bring a large penalty.

The computing time of  $\ell_{\mathbf{u}}$  increases only very slowly with  $n$  (roughly at a logarithmic rate), in contrast to that of  $\mathcal{P}_{\gamma,2\alpha}$ , but on the other hand it is exponential in  $s$  in the worst case. In practice, it can be computed reasonably quickly for  $s$  up to 30 or so, and  $n$  as large as we want. A computational advantage of the criteria (15) and (16) is that poor lattices can be eliminated quickly (on average) without having to compute all the  $\ell_{\mathbf{u}}$ 's. For example, in (15), the lattice can be eliminated as soon as we have a small enough upper bound on  $\gamma_{\mathbf{u}}\ell_{\mathbf{u}}/\ell_{|\mathbf{u}|}^*(n)$  for some  $\mathbf{u}$  (for this, we do not even need to know  $\ell_{\mathbf{u}}$  exactly). For all the criteria based on a sum to be minimized, we can also stop and eliminate the lattice whenever the sum exceeds a given value (e.g., if it exceeds the best value found so far).

A special case of (15) was used in [15], with  $\beta = 1$  and

$$\begin{aligned} \mathcal{J} &= \mathcal{J}(t_1, \dots, t_d) \\ &= \{\mathbf{u} = \{1, \dots, r\} \text{ for } 2 \leq r \leq t_1\} \\ &\quad \cup \{\mathbf{u} = \{j_1, \dots, j_r\} \text{ such that } 1 = j_1 < \dots < j_r \leq t_r \text{ and } 2 \leq r \leq d\}. \end{aligned}$$

This was inspired by criteria used for random number generators having a lattice structure [13]. The main drawback of this criterion is that many projections are not considered at all; they can be very bad and this is not reflected by the figure of merit.

All these criteria can also be defined based on the lengths of the shortest vectors in the *primal lattices*  $L_s(\mathbf{u})$ , instead of their dual lattices  $(L_s(\mathbf{u}))^*$ , and permuting minimization for maximization. This makes little difference in terms of the uniformity of retained lattices. The length of the shortest vector in  $L_s(\mathbf{u})$  represents the minimal distance between any two lattice points, and we want this distance to be as large as possible.

## 7 Searching for Lattice Parameters

Once we have selected a discrepancy measure (or figure of merit) and specified the weights, the next step is to search for lattices that minimize this measure, for a given  $n$ . In our experiments, we will use (and compare) the following strategies.

In the case of Korobov lattices, there is a single parameter that can take at most  $n - 1$  values, so we will simply make an exhaustive search for the best vector  $\mathbf{a}_1 = (1, a, a^2, \dots, \dots)$  over all admissible integers  $a$ .

For general rank-1 lattices, under our assumptions, there could be up to  $(n - 1)^{s-1}$  combinations and an exhaustive search is usually out of the question (for example, this happens as soon as  $s$  exceeds a few units if  $n$  is around a million, which is not unusual). A standard construction method in this context is the *component by component (CBC) construction algorithm*, which works as follows [22, 23]:

Let  $a_1 = 1$ ;  
 For  $j = 2, \dots, s$ , find  $a_j \in \{1, \dots, n - 1\}$ ,  $\gcd(a_j, n) = 1$ , such that  
 $(a_1, \dots, a_{j-1}, a_j)$  minimizes the selected figure of merit for the  
 first  $j$  dimensions.

We will also use the following streamlined search method, which replaces the exhaustive search over  $a_j$  at each step by a search over a small number of different random candidates  $a_j$  (the number  $r$  of candidates can be from 20 to 100, for example, depending on the computing budget that we are willing to devote to this).

Let  $a_1 = 1$ ;

For  $j = 2, \dots, s$ , try  $r$  random  $a_j \in \{1, \dots, n-1\}$ ,  $\gcd(a_j, n) = 1$ , and retain the one for which  $(a_1, \dots, a_{j-1}, a_j)$  minimizes the selected figure of merit for the first  $j$  dimensions.

## 8 Experimental methodology

We summarize our experimental setting for the empirical results reported in the following sections. For each example where this is relevant, we first estimate the ANOVA variance components of the integrand by the method of [25], using RQMC with  $2^{20} - 3 = 1,048,573$  lattice points and 1000 independent replications (random shifts). The lattice used for this (for all examples) was constructed by a randomized CBC search with  $r = 50$  using geometric order-dependent weights with  $\gamma = 0.5$ . Next, we select the criteria among (9) or (13)–(16) and the types of weights that we want to consider. The weights are selected as functions of the estimated ANOVA variances, using the strategies described in Sections 4 and 5. Occasionally, the ANOVA variance estimators are zero or take a smaller value than their precision. Then, we give these projections a weight equal to  $1/100$  of the smallest nonzero computed weight. For each selected criterion and type of weight, we construct lattices using random CBC searches with  $r = \min(50, n-1)$ , for 86 different prime values of  $n$  ranging from  $2^5 - 1 = 31$  to  $2^{22} - 3 = 4,194,301$ . Then, for each retained lattice, we estimate the RQMC variance with 100 independent replications.

When constructing lattices with the  $\mathcal{P}_{\gamma, 2\alpha}$  criterion for use with the baker's transformation, we set  $\alpha = 2$ ; otherwise, we set  $\alpha = 1$ . The weights for the criteria based on the spectral test are taken simply as  $\gamma_{\mathbf{u}} = \sigma_{\mathbf{u}}^2$ , where the latter is estimated, and we take  $\mathcal{J} = \{\mathbf{u} : \emptyset \neq \mathbf{u} \subseteq \{1, \dots, s\}\}$ , unless indicated otherwise.

In most cases, the variance behaves approximately linearly in logarithmic scale for  $n \geq 10^3$ . Then we fit a linear model of the form

$$\ln \text{Var}[\mu_{n, \text{rqmc}}] = \ln a_0 - \nu \ln n + \varepsilon \quad (17)$$

for positive constants  $a_0$  and  $\nu$ , where  $\varepsilon$  represents a noise term. We do this by computing the values  $\hat{a}_0$  and  $\hat{\nu}$  of  $a_0$  and  $\nu$  that minimize the sum of squares of the values of  $\varepsilon$  for the 61 (out of 86) values of  $n$  that are greater than  $10^3$ . Our estimated (or empirical) convergence rate is then  $\mathcal{O}(n^{-\hat{\nu}})$ . We report the precision on our estimates of  $\hat{\nu}$  via 95% confidence intervals, assuming that  $\varepsilon$  is normally distributed with mean 0 and variance  $S_\varepsilon^2$  (we have checked empirically that this is indeed a reasonable assumption).

In (17), the parameters  $a_0$  and  $\nu$  tell us how the log-variance decreases “on average” as a function of  $n$ , for a given lattice construction procedure and a given example. They are the primary quality indicators for the procedure. The

parameter  $\varepsilon$  represents the departure of the log-variance from the linear model for the particular lattice selected at a given  $n$ , together with the estimation error in the RQMC log-variance because it is based on a finite number of replications. The latter error can be made arbitrarily small by making more independent replicates of the RQMC estimator. The departure of the true log-variance from the linear model typically has a larger contribution to  $\varepsilon$  in our examples. This departure depends on the lattice parameters that are retained by the selection algorithm for the given  $n$ ; it is intrinsic to the lattice construction procedure and it generally depends on the criterion and type of weights. A small standard deviation  $S_\varepsilon$  means that the linear model is a better predictor of the performance for a given  $n$ , and that the returned lattices tend to be more robust and reliable in terms of RQMC variance. When the linear models for two or more criteria predict similar RQMC variances, we should prefer the one with the smallest  $S_\varepsilon$ .

We define the variance reduction factor (VRF) for a specific  $n$ -point randomly-shifted lattice rule as the variance  $\sigma^2$  of the MC estimator divided by  $n$  times the variance of the RQMC estimator. We estimate  $\sigma^2$  by the empirical variance  $S_n^2$ . In some cases, we replace the RQMC variance of the specific lattice at a given value of  $n$  by the variance  $\hat{a}_0 n^{-\hat{\nu}}$  interpolated from our linear model in log scale, and we report the corresponding interpolated VRF,  $\widehat{\text{VRF}}(n) = n^{\hat{\nu}-1} S_n^2 / \hat{a}_0$ , usually with  $n = 2^{20}$ . This interpolation is more stable than the actual variance at a given  $n$ .

Detailed results of our experiments are given in the online appendix. In the following sections, we only summarize these results.

## 9 An Indicator Function

In our first set of experiments, we consider a simple discontinuous integrand defined as the indicator that a sum of  $s$  independent random variables exceeds a given threshold. We assume that  $Y_1, \dots, Y_s$  are independent random variables, and that  $Y_j$  is exponential with rate  $\lambda_j$ , for each  $j$ . We estimate the probability  $\mu = \mathbb{P}[Y_1 + \dots + Y_s > x]$  by MC or RQMC, for some constant  $x$ . The *basic estimator* is  $X = \mathbb{I}[Y_1 + \dots + Y_s > x]$ , where  $\mathbb{I}$  denotes the indicator function. It corresponds to the discontinuous  $s$ -dimensional integrand

$$f(u_1, \dots, u_s) = \mathbb{I}[F_1^{-1}(u_1) + \dots + F_s^{-1}(u_s) > x],$$

where  $F_j^{-1}(u_j) = -\ln(1 - u_j)/\lambda_j$  is the inverse cdf of  $Y_j$  evaluated at  $u_j$ .

We also consider the *conditional MC* (CMC) estimator

$$\begin{aligned} X_{\text{cmc}} &= \mathbb{P}[Y_1 + \dots + Y_s > x \mid Y_1 + \dots + Y_{s-1}] \\ &= \exp[-\lambda_s(x - Y_1 - \dots - Y_{s-1})] \cdot \mathbb{I}[x - Y_1 - \dots - Y_{s-1} \geq 0]. \end{aligned}$$

The associated integrand,

$$f(u_1, \dots, u_{s-1}) = 1 - F_s(x - F_1^{-1}(u_1) - \dots - F_{s-1}^{-1}(u_{s-1})),$$

has dimension  $s-1$  and is continuous, but has a discontinuity in its first-order derivatives, because the cdf of  $Y_s$ ,  $F_s(y) = [1 - \exp(-\lambda_s y)] \cdot \mathbb{I}[y > 0]$ , has a discontinuous derivative at  $y = 0$ .

For the one-dimensional case, it is known (see [17]) that the basic RQMC estimator can take only two values and its variance converges as  $\mathcal{O}(n^{-2})$  regardless of the choice of lattice. Using the dual-space exploration algorithm here does not work well because the Fourier coefficients do not decrease monotonously with  $\|\mathbf{h}\|$ .

We simulated these estimators for  $s = 2, \dots, 6$ , for the following four cases:  $\lambda_j = 1$ ,  $\lambda_j = j^{-1}$ , and  $\lambda_j = j^{-2}$ , with  $x$  chosen so that the probability  $\mu$  to be estimated is close to 0.5, and  $\lambda_j = j^{-1}$  with  $x$  chosen so that  $\mu$  is near 0.1.

To select the lattice parameters, we tried the criterion (9) with  $\alpha = 1$  for both the basic and CMC estimators, with  $\alpha = 2$  for the CMC estimator with the baker transformation, and the criteria (13)–(16) with  $\beta = 1$  and 2, with projection-dependent, product, order-dependent and geometric order-dependent weights in all cases. In general, the observed convergence rates (reported in the online appendix) do not vary too much when only  $\lambda_j$  or  $x$  changes. Although none of the integrands here meets the smoothness requirements that justify using the  $\mathcal{P}_{\gamma, 2\alpha}$  criterion, in the sense that we have no guaranteed convergence rate for the variance of the corresponding RQMC estimators, lattices constructed with that criterion and projection-dependent weights gave slightly higher values of  $\widehat{\text{VRF}}(2^{20})$  and  $\hat{\nu}$  together with smaller values of  $\hat{S}_\varepsilon$  on average, compared to those obtained with criteria based on the spectral test. They give empirical convergence rates exponents of approximately  $\hat{\nu} \approx (s+1)/s$  for the basic estimator. For the CMC estimator, the best convergence rates, of  $\mathcal{O}(n^{-2})$  without the baker's transformation and of  $\mathcal{O}(n^{-4})$  with the baker's transformation, are obtained at  $s = 2$  and degrade as a function of  $s$  down to around  $\mathcal{O}(n^{-1.5})$  and  $\mathcal{O}(n^{-1.6})$ , respectively, at  $s = 6$ . The improvement on the convergence rate due to the baker's transformation is clear at  $s = 2$  or 3 but seems marginal for  $s \geq 4$ . The observed convergence rates for the CMC estimator for  $s = 2$  were expected, because in that case the integrand is one-dimensional, is continuous in  $(0, 1)$  but its periodic continuation is not, and these are the known convergence rates for such an integrand [12].

Regarding the choices of weights, our estimations of the ANOVA variances suggested that product and order-dependent weights were not justified, yet we found no clearly observable best choice of weights for the basic estimator. For the CMC estimator, however, projection-dependent weights, when used with the  $\mathcal{P}_{\gamma, 2\alpha}$  criterion, consistently offer good performance.

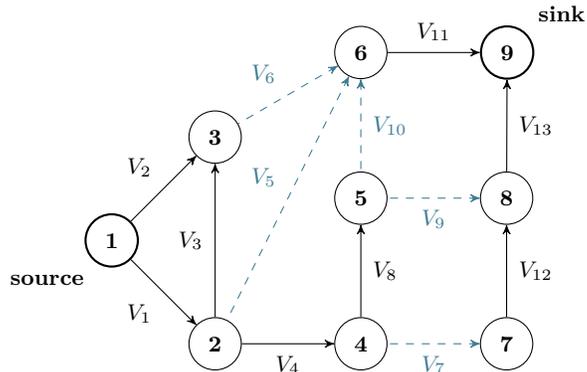
We also examined a simpler case with  $s = 2$  where  $Y_1 \sim U[0, m)$  for some  $m \in [0, 1)$  and  $Y_2 \sim U[0, 1)$ . Our experiments with  $m = 0.375$  and  $x = 0.8$  revealed that some lattices with excellent values of standard figures of merit, such as  $\mathcal{P}_{2\alpha}$  and those based on the spectral test, are not among the best in terms of variance reduction. These criteria, introduced earlier, are not really appropriate in this situation, because they do not take into account the alignment between the lattice points and the discontinuity, which turns out to be a key factor here. On the other hand, even for this specific artificial example, by examining all lattices for a given  $n$ , we found a clear positive correlation between the RQMC variance and  $\mathcal{P}_{2\alpha}$ . Here, the choice of weights is not an issue, because there is a single projection in more than one dimension.

## 10 Example: a Stochastic Activity Network

We consider the stochastic activity network example taken from [9] and represented in Figure 2, where the lengths  $V_1, \dots, V_{13}$  of edges  $1, \dots, 13$  are independent random variables with distribution functions  $F_1, \dots, F_{13}$ , respectively. We take the same cdf's  $F_j$  as in [2, Section 4.1]. For the activities  $j = 1, 2, 4, 11, 12$ , we have  $V_j = \max(0, \tilde{V}_j)$  where  $\tilde{V}_j$  is a normally distributed random variable with mean  $\theta_j$  and variance  $\theta_j^2/16$ . The other  $V_j$ 's are exponential with mean  $\theta_j$ . The values of  $\theta_1, \dots, \theta_{13}$  are 13.0, 5.5, 7.0, 5.2, 16.5, 14.7, 10.3, 6.0, 4.0, 20.0, 3.2, 3.2, 16.5, respectively. See [1, 2, 15] for a complete description of the problem. We are interested in estimating the probability that the longest path from source (node 1) to sink (node 9) has a length  $T$  larger than some constant  $x$  with the estimator  $X = \mathbb{I}[T > x]$ . We also consider the CMC estimator obtained by simulating the  $V_j$ 's only for the edges that are *not* in the cut set  $\mathcal{L} = \{5, 6, 7, 9, 10\}$ , and taking the probability that  $T > x$  conditional on those  $V_j$ 's, as in [1]. This CMC estimator can be written as

$$X_{\text{CMC}} = \mathbb{P}[T > x \mid \{V_j : j \notin \mathcal{L}\}] = 1 - \prod_{j \in \mathcal{L}} \mathbb{P}[V_j \leq x - P_j] \quad (18)$$

where  $P_j$  is the length of the longest path that goes through edge  $j$  when we put  $V_j = 0$  (i.e., we exclude edge  $j$ ). The main motivation for considering this estimator is that it is continuous as a function of the  $V_j$ 's that are generated (and therefore as a function of the underlying uniform random numbers), in contrast to the original estimator  $X$ , and it is also easy to compute. This example generalizes the problems and the CMC estimators considered in the previous section. This integrand has dimension  $s = 13$  with the basic estimator  $X$ , and dimension  $s = 8$  with the CMC estimator  $X_{\text{CMC}}$ . We also



**Fig. 2** Graph of the stochastic activity network with 9 nodes and 13 links. The dashed links are not simulated in the CMC variant.

estimated  $\mathbb{E}[T]$  by simulation; the corresponding integrand has dimension  $s = 13$ .

For all types of estimators, we have estimated the ANOVA variances and observed that they vary a lot across projections of a given order, so we do not expect order-dependent or geometric weights to work well. In our experiments, we found that the  $\mathcal{P}_{\gamma,2\alpha}$  criterion (with  $\alpha = 1$  for the standard estimator and  $\alpha = 2$  for the CMC estimator) performed well in all cases, with relatively high values of  $\widehat{\text{VRF}}(2^{20})$  and  $\hat{\nu}$ , together with low values of  $\hat{S}_\varepsilon$ , with slightly better performance for projection-dependent and product weights. We also found that using the inappropriate order-dependent or geometric weights does not guarantee poor performance—in some cases the VRF's were even slightly higher than with the more appropriate projection-dependent and product weights—but it makes it more unpredictable, with VRFs as low as half of the best ones in some cases. The criteria based on the spectral test did not perform as well as  $\mathcal{P}_{\gamma,2\alpha}$ , at least for projection-dependent and product weights. The standard and CMC estimators had similar qualitative behavior, but the observed VRFs were much larger for the CMC estimator. For example, the best VRF for  $x = 60$  interpolated at  $n = 2^{20}$  is 27 with an empirical convergence rate of 1.20 for the standard estimator, obtained with the  $\mathcal{P}_{\gamma,2\alpha}$  criterion with projection-dependent weights and  $\alpha = 1$ . For the same case but with CMC estimator and  $\alpha = 2$ , we observed a fitted VRF of  $4.4 \times 10^3$  with an empirical convergence rate of 1.51. The baker's transformation offered very little improvement on the CMC estimator. All these results are in the online appendix.

## 11 Example: Asian Call Option

We consider an Asian call option based on the price  $S(t)$  of single asset at times  $t_0 = 0, t_1, \dots, t_s$ , with payoff:

$$Y = e^{-\tilde{r}t_s} \max \left[ 0, \frac{1}{s} \sum_{j=1}^s S(t_j) - K \right],$$

where  $\tilde{r}$  is the risk-free interest rate and  $K$  is the strike price. The asset price is a geometric Brownian motion:

$$S(t) = S(0) \exp[(\tilde{r} - \sigma^2/2)t + \sigma B(t)],$$

where  $\{B(t) : t \geq 0\}$  is a standard Brownian motion, and  $\sigma$  is the volatility. We also consider a down-and-in variant of the Asian option with payoff

$$Y' = Y \cdot \mathbb{I} \left[ \min_{j=1, \dots, s} S(t_j) \leq K' \right],$$

where  $K'$  is a barrier. We estimate  $\mathbb{E}[Y]$  and  $\mathbb{E}[Y']$  with MC and RQMC. For our experiments, we set  $S(0) = 100$ ,  $K = 100$ ,  $K' = 80$ ,  $\tilde{r} = 0.05$ ,  $\sigma = 0.5$ ,  $t_j = j/s$  for  $j = 0, \dots, s$ , and  $s = 6$ . To simulate  $S(t_1), \dots, S(t_s)$ ,  $Y$  and  $Y'$ , we sample a standard normal vector  $\mathbf{Z} = (Z_1, \dots, Z_s)$  with  $Z_j = \Phi^{-1}(U_j)$ , where  $\Phi$  is the standard normal distribution function. Then we generate  $\mathbf{B} = (B(t_1), \dots, B(t_s)) = \mathbf{A}\mathbf{Z}$ , where  $\mathbf{C} = \mathbf{A}\mathbf{A}^t$  is the covariance matrix of  $\mathbf{B}$  with elements  $c_{j,k} = \sigma^2 \min(t_j, t_k)$ . We consider two standard choices for the decomposition  $\mathbf{A}\mathbf{A}^t$ . The first is the Cholesky factorization where  $\mathbf{A}$  is a lower triangular matrix. The second, based on principal component analysis (PCA), is  $\mathbf{A} = \mathbf{P}\mathbf{D}^{1/2}$ , where  $\mathbf{P}$  is the matrix of right eigenvectors of  $\mathbf{C}$  and  $\mathbf{D}$  is a diagonal matrix that contains the eigenvalues of  $\mathbf{C}$  sorted by increasing order so that the components of  $\mathbf{B}$  depend more on the first components of  $\mathbf{Z}$  than on the others.

For the Asian option with PCA, our estimations of the ANOVA variances showed that projection  $\{1\}$  itself accounts for nearly 99% of the total variance for the Asian option, whereas with Cholesky all projections of order 1 together account for only 73% of the total variance. For the down-and-in option, the largest part of the variance is contributed by projections of order 2 and more, and PCA barely improves the situation with respect to Cholesky by raising from 9% to 14% the percentage of variance contributed by projections of order 1. Note that there is only one projection of order 6 and it accounts for 9.4% and 13% of the total variance for Cholesky and PCA, respectively.

The Asian option payoff function is continuous with respect to the uniforms, but the down-and-in variant is not, so we use the baker's transformation for the former but not for the latter. For the PCA case, we show in Table 1 the fitted VRF's and empirical convergence rates for various types

| Asian option (PCA), $s = 6$ |              |            |                   |                                |                   |                       |
|-----------------------------|--------------|------------|-------------------|--------------------------------|-------------------|-----------------------|
| criterion                   | construction | $r$        | weight type       | $\widehat{\text{VRF}}(2^{20})$ | $\hat{\nu}$       | $\hat{S}_\varepsilon$ |
| $\mathcal{P}_{\gamma,4}$    | CBC          | 50         | proj.-dep.        | $3.1 \times 10^5$              | $1.846 \pm 0.004$ | 0.325                 |
|                             |              |            | product           | $3.1 \times 10^5$              | $1.840 \pm 0.005$ | 0.335                 |
|                             |              |            | order-dep.        | $1.6 \times 10^5$              | $1.707 \pm 0.008$ | 0.632                 |
|                             |              |            | geometric         | $2.4 \times 10^5$              | $1.784 \pm 0.005$ | 0.399                 |
|                             |              |            | order 2           | $1.4 \times 10^5$              | $1.710 \pm 0.010$ | 0.852                 |
|                             | –            | proj.-dep. | $3.5 \times 10^9$ | $1.870 \pm 0.020$              | 0.317             |                       |
| Korobov                     | 50           | proj.-dep. | $2.6 \times 10^9$ | $1.825 \pm 0.005$              | 0.354             |                       |
|                             | –            | proj.-dep. | $3.0 \times 10^9$ | $1.850 \pm 0.010$              | 0.333             |                       |
| $\mathcal{M}_{\gamma,2}$    | CBC          | 50         | proj.-dep.        | $1.7 \times 10^5$              | $1.751 \pm 0.007$ | 0.545                 |
| $\mathcal{M}'_{\gamma,2}$   | CBC          | 50         | proj.-dep.        | $2.2 \times 10^5$              | $1.807 \pm 0.007$ | 0.492                 |

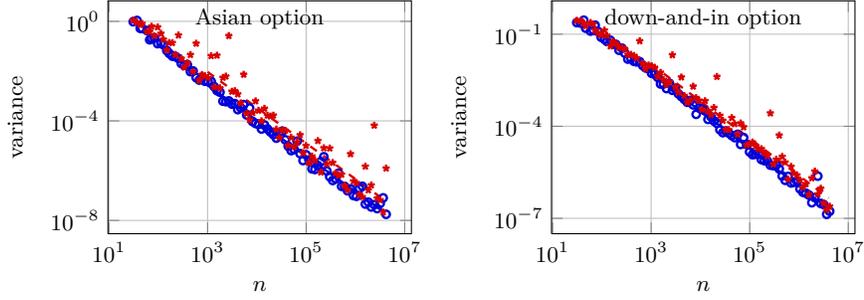
  

| down-and-in option (PCA), $s = 6$ |              |            |             |                                |                   |                       |
|-----------------------------------|--------------|------------|-------------|--------------------------------|-------------------|-----------------------|
| criterion                         | construction | $r$        | weight type | $\widehat{\text{VRF}}(2^{20})$ | $\hat{\nu}$       | $\hat{S}_\varepsilon$ |
| $\mathcal{P}_{\gamma,4}$          | CBC          | 50         | geometric   | 7.8                            | $1.180 \pm 0.003$ | 0.238                 |
|                                   |              |            | product     | 7.5                            | $1.212 \pm 0.004$ | 0.332                 |
|                                   |              |            | proj.-dep.  | 7.5                            | $1.169 \pm 0.004$ | 0.267                 |
|                                   |              |            | order-dep.  | 7.1                            | $1.149 \pm 0.005$ | 0.372                 |
|                                   |              |            | order 2     | 4.0                            | $1.160 \pm 0.010$ | 0.793                 |
|                                   | –            | proj.-dep. | 9.0         | $1.193 \pm 0.009$              | 0.227             |                       |
| Korobov                           | 50           | proj.-dep. | 7.1         | $1.195 \pm 0.005$              | 0.341             |                       |
|                                   | –            | proj.-dep. | 7.6         | $1.181 \pm 0.008$              | 0.217             |                       |
| $\mathcal{M}_{\gamma,2}$          | CBC          | 50         | proj.-dep.  | 6.0                            | $1.160 \pm 0.004$ | 0.313                 |
| $\mathcal{M}'_{\gamma,2}$         | CBC          | 50         | proj.-dep.  | 6.2                            | $1.183 \pm 0.007$ | 0.500                 |

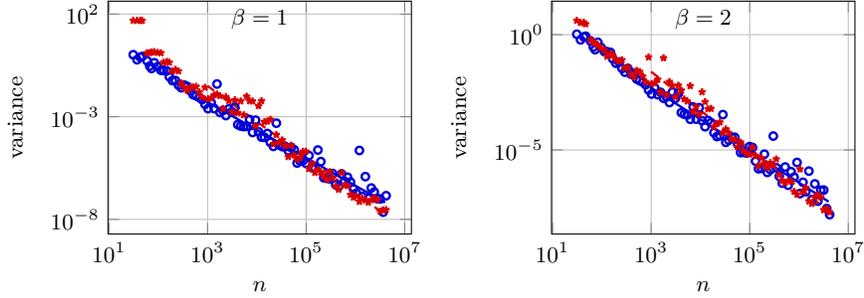
**Table 1** Fitted variance reduction factors at  $n = 2^{20}$  and empirical convergence rates for the Asian and down-and-in options. The baker's transformation was applied for the Asian option, but not for the down-and-in option. When CBC is followed by a value of  $r$ , it refers to random CBC, otherwise it refers to exhaustive CBC, and similarly for Korobov. Order-dependent of order 2 is abbreviated as *order 2*.

of weights with the  $\mathcal{P}_{\gamma,2\alpha}$  criterion using random CBC construction, and for projection-dependent weights for the  $\mathcal{M}_{\gamma,2}$  and  $\mathcal{M}'_{\gamma,2}$  criteria. The error on  $\ln \hat{a}_0$  (not shown in the table) is in general of the order of one tenth of the value of  $\hat{S}_\varepsilon$  or less. Besides constant order-truncated weights at order 2, which yield poor performance as confirmed in Figure 3, the other types of weights all seem to offer comparable performance. With PCA, compared to Cholesky, the VRF's are much higher, the convergence with  $n$  is faster, and there is less noise in the observed variances (see the appendix).

We compared the relative performance of the criteria  $\mathcal{M}_{\gamma,\beta}$ ,  $\mathcal{M}'_{\gamma,\beta}$ ,  $\widetilde{\mathcal{M}}_{\gamma,\beta}$  and  $\widetilde{\mathcal{M}}'_{\gamma,\beta}$ , for  $\beta = 1$  and 2, and for projection-dependent, product, order-dependent and geometric order-dependent weights. With Cholesky factorization,  $\widetilde{\mathcal{M}}_{\gamma,\beta}$  and  $\widetilde{\mathcal{M}}'_{\gamma,\beta}$ , based on the worst projection, generally yield faster convergence and larger VRF than their counterparts  $\mathcal{M}_{\gamma,\beta}$  and  $\mathcal{M}'_{\gamma,\beta}$  based on a weighted sum over all projections. Besides this, it is hard to discriminate between criteria and weight types. We illustrate part of these observations in Figure 4, where we compare (14) to (16) for  $\beta = 1$  and 2 and product weights.



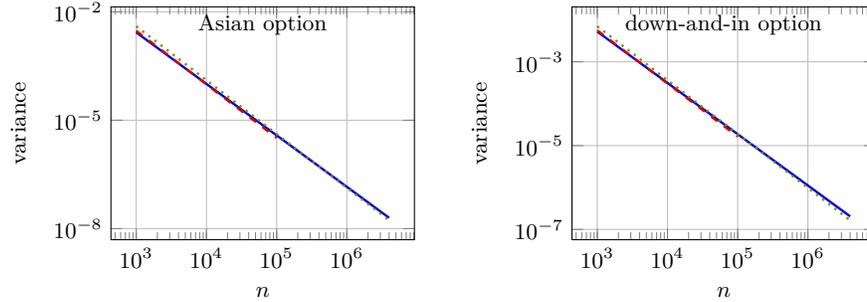
**Fig. 3** Estimated and fitted variance of the RQMC estimator, using lattices constructed with the  $\mathcal{P}_{\gamma,2\alpha}$  criterion, for the Asian option with  $\alpha = 2$  and the baker's transformation (left) and for the down-and-in option with  $\alpha = 1$  without the baker's transformation (right), using Cholesky factorization, with projection-dependent weights ( $\circ$ ) and with constant order-dependent weights truncated at order 2 ( $\star$ ).



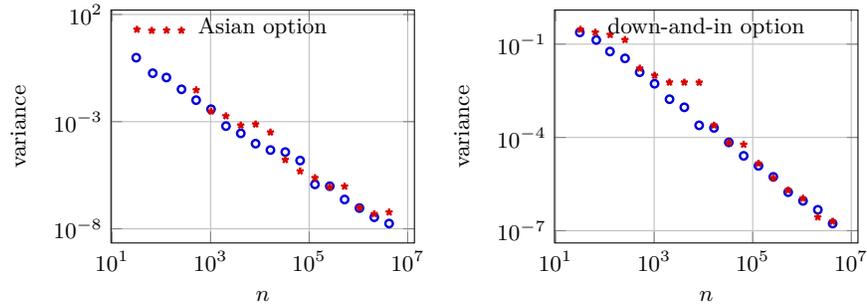
**Fig. 4** Estimated and fitted variance of the RQMC estimator for the Asian option (Cholesky) with the baker transformation, using lattices constructed with the  $\mathcal{M}'_{\gamma,\beta}$  ( $\circ$ ) and  $\widetilde{\mathcal{M}}'_{\gamma,\beta}$  ( $\star$ ) criteria, with  $\beta = 1$  (left) and  $\beta = 2$  (right) and with product weights.

The observed variances are more noisy on average when using (16), but the convergence seems faster. When using PCA, on the other hand, we did not observe any significant difference in the results across different criteria. The easy explanation is that for integrands where only a small part of the variance lies in projections of order two or more, all criteria and weight types under consideration here are practically equivalent in terms of the variance of the RQMC estimator.

In Table 1, we also give some results for exhaustive CBC and Korobov constructions for projection-dependent weights, for the Asian option. Random Korobov means that we tried  $r$  random values of  $a$ . The exhaustive CBC construction generally provides a slightly better variance reduction than random CBC, and the Korobov construction is slightly worse than CBC, but the difference is thin, as can be seen in Figure 5. Note that because the cost of exhaustive CBC increases with  $n$  (there are  $(s-1)(n-1)$  vectors to examine) we have results only for  $n \leq 10^5$  in this case.

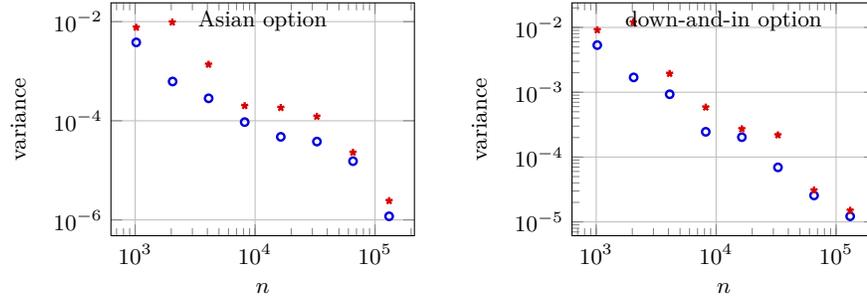


**Fig. 5** Estimated and fitted variance of the RQMC estimator, using lattices constructed with the  $\mathcal{P}_{\gamma,2\alpha}$  criterion and projection-dependent weights, for the Asian option with  $\alpha = 2$  and the baker's transformation (left) and for the down-and-in option with  $\alpha = 1$  without the baker's transformation (right), using Cholesky factorization, with random CBC construction with  $r = 50$  (—), exhaustive CBC construction (---) or exhaustive Korobov construction (⋯⋯).



**Fig. 6** Estimated and fitted variance of the RQMC estimator for lattices constructed with the  $\mathcal{P}_{\gamma,2\alpha}$  criterion with projection-dependent weights, for the Asian option with the baker's transformation and  $\alpha = 2$  (left) and the down-and-in option without the baker's transformation with  $\alpha = 1$  (right), for with prime values of  $n$  ( $\circ$ ) and for values of  $n$  that are powers of two ( $\star$ ).

We also constructed lattices using values of  $n$  that are powers of two. In some cases, they clearly produced larger RQMC variances than lattices with prime  $n$ , as illustrated in Figure 6. But in most cases, the variances for  $n$  prime or a power-of-two are comparable. For instance, this occurs for the example of Figure 6, but with product weights instead of projection-dependent weights. Note that in order to have each  $a_j$  relatively prime with  $n$  for  $j = 2, \dots, s$  when  $n$  is a power of two,  $a_j$  has to be an odd number, which means that for each component of the generating vector  $\mathbf{a}$  except the first which is fixed at  $a_1 = 1$ , there is only half the number of possible values to consider. In other words, the space of lattice parameters is  $2^{s-1}$  times smaller for values of  $n$  that are powers of two than for prime values of  $n$ . This could be part of the explanation.

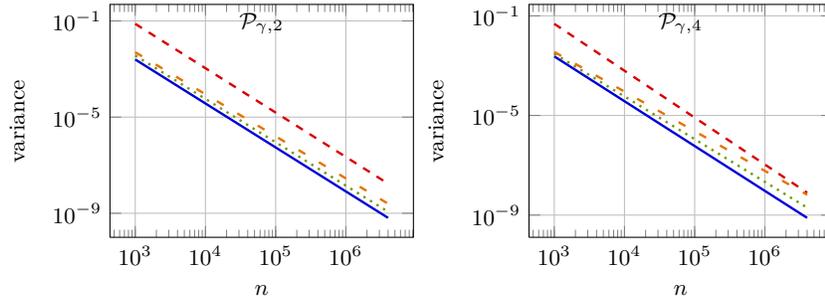


**Fig. 7** Fitted variance of the RQMC estimator for the Asian option with the baker's transformation (left) and the down-and-in option without the baker's transformation (right) with Cholesky factorization, for lattices constructed using the  $\mathcal{P}_{\gamma,2\alpha}$  criterion with random CBC with  $r = 50$  ( $\circ$ ) and the  $\widetilde{\mathcal{M}}_{\gamma,1}$  criterion as in (15) with  $\mathcal{J} = \mathcal{J}(32, 24, 16, 12)$  ( $\star$ ) and Korobov construction.

We also did a few experiments with the  $\widetilde{\mathcal{M}}_{\gamma,1}$  criterion as in (15), with  $\mathcal{J} = \mathcal{J}(32, 24, 16, 12)$ , as proposed in [15]. As shown in Figure 7, this criterion does not perform well. It does not appear appropriate for the problems at hand, because too many coordinates (up to 32) are considered by the criterion whereas projections of order 5 and 6 are ignored.

Finally, to show a situation where projection-dependent weights perform clearly better than other types of weights, we give some results for an (artificial) example where we have two independent Asian options, each with  $s = 3$  and the same parameters, and the payoff is the sum of the payoffs of the two options. Of course, we could estimate the expected payoff of each of the two options by RQMC separately and add up, but here, for the purpose of the illustration, we simulate the first option using the first three coordinates of the six-dimensional point set and the second option, using the last three coordinates. Then, the ANOVA variances are non-zero only for projections  $\mathbf{u}$  such that  $\emptyset \neq \mathbf{u} \subseteq \{1, 2, 3\}$  or  $\emptyset \neq \mathbf{u} \subseteq \{4, 5, 6\}$ . There are thus only 14 out of 63 total projections that are relevant to the problem. This way, order-dependent weights are unlikely to perform well, because they give significant weight to the 9 irrelevant projections of order 2 and to the 18 irrelevant projections of order 3, rather than concentrate the weights over the important projections. We expect product weights to do even worse, because they waste their weights to these and on the 22 other irrelevant projections of order 4, 5, and 6. This is confirmed in Figure 8 and Table 2. Interestingly, the lattices obtained using  $\alpha = 1$  appear more robust than those with  $\alpha = 2$ , even if the baker's transformation was used in both cases.

In summary, other choices of weights frequently perform almost as well as (more general) projection-dependent weights even when they are not really justified, which is good news, but there are situations where the projection-dependent weights really perform much better.



**Fig. 8** Fitted variance of the RQMC estimator for the sum of two Asian payoffs, with Cholesky factorization, using the baker's transformation and criterion  $\mathcal{P}_{\gamma,2\alpha}$  with  $\alpha = 1$  (left) and  $\alpha = 2$  (right), using projection-dependent weights (—), product weights (---), order-dependent weights (⋯⋯), and geometric weights (-.-).

| weight type | $\mathcal{P}_{\gamma,2}$       |                   |                       | $\mathcal{P}_{\gamma,4}$       |                   |                       |
|-------------|--------------------------------|-------------------|-----------------------|--------------------------------|-------------------|-----------------------|
|             | $\widehat{\text{VRf}}(2^{20})$ | $\hat{\nu}$       | $\hat{S}_\varepsilon$ | $\widehat{\text{VRf}}(2^{20})$ | $\hat{\nu}$       | $\hat{S}_\varepsilon$ |
| proj.-dep.  | $1.9 \times 10^5$              | $1.829 \pm 0.005$ | 0.351                 | $1.7 \times 10^5$              | $1.800 \pm 0.004$ | 0.328                 |
| product     | $7.2 \times 10^3$              | $1.85 \pm 0.03$   | 1.88                  | $1.5 \times 10^4$              | $1.88 \pm 0.02$   | 1.35                  |
| order-dep.  | $1.1 \times 10^5$              | $1.80 \pm 0.01$   | 0.669                 | $7.2 \times 10^4$              | $1.72 \pm 0.01$   | 0.738                 |
| geometric   | $5.6 \times 10^4$              | $1.75 \pm 0.01$   | 1.10                  | $2.6 \times 10^4$              | $1.59 \pm 0.01$   | 1.00                  |

**Table 2** Estimated  $\widehat{\text{VRf}}$ ,  $\hat{\nu}$  and  $\hat{S}_\varepsilon$  for the RQMC estimator of the sum of two Asian options for the criterion  $\mathcal{P}_{\gamma,2\alpha}$  with  $\alpha = 1$  and 2 with the baker's transformation in both cases.

## 12 Conclusion

The optimal lattice, which minimizes the variance when estimating an integral by a randomly shifted lattice rule, depends on the integrand  $f$ , and optimizing this lattice is harder in general than computing the integral itself. The idea of constructing efficient adaptive algorithms by estimating the Fourier coefficients or the variance components, for general applications, is attractive at first sight, but estimating those quantities with reasonable accuracy usually too costly. Fortunately, crude estimates of the variance components are generally sufficient to identify the subsets of coordinates on which to put more weight when constructing the lattice, and doing this with a weighted  $\mathcal{P}_{\gamma,2\alpha}$  figure of merit with projection-dependent weights is a robust approach that gives very good results in most examples that we have tried. In fact, lattices constructed based on a weighted  $\mathcal{P}_{\gamma,2\alpha}$  with reasonable choices of weights, such as order-dependent weights that decrease geometrically (but not too fast) with the cardinality of coordinate subsets, perform well enough in most cases. Such lattices could be provided in general-purpose RQMC software. On the other hand, lattices constructed with lousy choices of weights, that give too little weight to some important projections (for example, giving weight

only to the projections of order 2), or too much weight to several irrelevant projections, often perform poorly. We also saw counterexamples (indicator functions in two dimensions) where a lattice having the best  $\mathcal{P}_{\gamma,2\alpha}$  performs very poorly, not because of a poor choice of weights, but because  $\mathcal{P}_{2\alpha}$  is not always a relevant measure in these examples. Thus, all the practical methods that we can propose to define a figure of merit for general applications are heuristic and none is foolproof. However, these counterexamples were constructed on purpose and such cases are rarely encountered in applications.

The theoretical asymptotic convergence rate of  $\mathcal{O}(n^{-2\alpha+\delta})$  for  $\mathcal{P}_{\gamma,2\alpha}$  and for the RQMC variance for certain classes of smooth functions is rarely observed in the practical range of values of  $n$ , say up to a few millions. The rates we have observed empirically, with the best lattices we found, are typically somewhere between  $\mathcal{O}(n^{-2})$  and  $\mathcal{O}(n^{-1})$ . Interestingly, this applies not only to smooth functions  $f$ , but also to non-smooth integrands, and even to discontinuous and unbounded ones.

An ongoing project related to this study is to build integrated software tools that can construct lattices based on a variety of parameterized figures of merit, with flexibility for the choices of weights (or parameters), and feed them to simulation software for arbitrary RQMC applications. This will include lattices extensible in both the dimension  $s$  and the number of points  $n$ . Hopefully, this will put these RQMC methods closer to the hands of practitioners and promote their utilization in a large variety of applications.

The online appendix to this paper can be found at <http://www.iro.umontreal.ca/~lecuyer/myftp/papers/mcqmc-plenary-app.pdf>

**Acknowledgements** This research has been supported by NSERC-Canada grant No. ODGP0110050 and a Canada Research Chair to the first author. Computations were performed using the facilities of the Réseau québécois de calcul haute performance (RQCHP).

## References

1. Avramidis, A.N., Wilson, J.R.: Integrated variance reduction strategies for simulation. *Operations Research* **44**, 327–346 (1996)
2. Avramidis, A.N., Wilson, J.R.: Correlation-induction techniques for estimating quantiles in simulation experiments. *Operations Research* **46**(4), 574–591 (1998)
3. Conway, J.H., Sloane, N.J.A.: *Sphere Packings, Lattices and Groups*, 3rd edn. *Grundlehren der Mathematischen Wissenschaften* 290. Springer-Verlag, New York (1999)
4. Cools, R., Nuyens, D.: A Belgian view on lattice rules. In: A. Keller, S. Heinrich, H. Niederreiter (eds.) *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pp. 3–21. Springer-Verlag, Berlin (2008)
5. Cranley, R., Patterson, T.N.L.: Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis* **13**(6), 904–914 (1976)
6. Dick, J., Sloan, I.H., Wang, X., Wozniakowski, H.: Liberating the weights. *Journal of Complexity* **20**(5), 593–623 (2004)

7. Dick, J., Sloan, I.H., Wang, X., Wozniakowski, H.: Good lattice rules in weighted Korobov spaces with general weights. *Numerische Mathematik* **103**, 63–97 (2006)
8. Efron, B., Stein, C.: The jackknife estimator of variance. *Annals of Statistics* **9**, 586–596 (1981)
9. Elmaghraby, S.: *Activity Networks*. Wiley, New York (1977)
10. Hickernell, F.J.: A generalized discrepancy and quadrature error bound. *Mathematics of Computation* **67**(221), 299–322 (1998)
11. Hickernell, F.J.: Lattice rules: How well do they measure up? In: P. Hellekalek, G. Larcher (eds.) *Random and Quasi-Random Point Sets, Lecture Notes in Statistics*, vol. 138, pp. 109–166. Springer-Verlag, New York (1998)
12. Hickernell, F.J.: Obtaining  $O(N^{-2+\epsilon})$  convergence for lattice quadrature rules. In: K.T. Fang, F.J. Hickernell, H. Niederreiter (eds.) *Monte Carlo and Quasi-Monte Carlo Methods 2000*, pp. 274–289. Springer-Verlag, Berlin (2002)
13. L'Ecuyer, P.: Good parameters and implementations for combined multiple recursive random number generators. *Operations Research* **47**(1), 159–164 (1999)
14. L'Ecuyer, P.: Quasi-Monte Carlo methods with applications in finance. *Finance and Stochastics* **13**(3), 307–349 (2009)
15. L'Ecuyer, P., Lemieux, C.: Variance reduction via lattice rules. *Management Science* **46**(9), 1214–1235 (2000)
16. L'Ecuyer, P., Lemieux, C.: Recent advances in randomized quasi-Monte Carlo methods. In: M. Dror, P. L'Ecuyer, F. Szidarovszky (eds.) *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*, pp. 419–474. Kluwer Academic, Boston (2002)
17. L'Ecuyer, P., Munger, D., Tuffin, B.: On the distribution of integration error by randomly-shifted lattice rules. *Electronic Journal of Statistics* **4**, 950–993 (2010)
18. Lemieux, C.: *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer-Verlag, New York, NY (2009)
19. Niederreiter, H.: *Random Number Generation and Quasi-Monte Carlo Methods, SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*, vol. 63. SIAM, Philadelphia, PA (1992)
20. Owen, A.B.: Latin supercube sampling for very high-dimensional simulations. *ACM Transactions on Modeling and Computer Simulation* **8**(1), 71–102 (1998)
21. Sloan, I.H., Joe, S.: *Lattice Methods for Multiple Integration*. Clarendon Press, Oxford (1994)
22. Sloan, I.H., Kuo, F.Y., Joe, S.: On the step-by-step construction of quasi-Monte Carlo rules that achieve strong tractability error bounds in weighted Sobolev spaces. *Mathematics of Computation* **71**, 1609–1640 (2002)
23. Sloan, I.H., Reztov, A.: Component-by-component construction of good lattice rules. *Mathematics of Computation* **71**, 262–273 (2002)
24. Sloan, I.H., Woźniakowski, H.: When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals. *Journal of Complexity* **14**, 1–33 (1998)
25. Sobol', I.M., Myshetskaya, E.E.: Monte Carlo estimators for small sensitivity indices. *Monte Carlo Methods and Applications* **13**(5–6), 455–465 (2007)
26. Wang, X.: Constructing robust good lattice rules for computational finance. *SIAM Journal on Scientific Computing* **29**(2), 598–621 (2007)