ELSEVIER

Journal of Computational and Applied Mathematics 114 (2000) 231–246

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Improving the rejection sampling method in quasi-Monte Carlo methods

Xiaoqun Wang[1]

Department of Applied Mathematics, Tsinghua University, Beijing 100084, People's Republic of China

Received 14 May 1997; received in revised form 5 April 1999

Abstract

The rejection sampling method is one of the most popular methods used in Monte Carlo methods. In this paper, we investigate and improve the performance of using a deterministic version of rejection method in quasi-Monte Carlo methods. It turns out that the "quality" of the point set generated by deterministic rejection method is closely related to the problem of quasi-Monte Carlo integration of characteristic functions, whose accuracy may be lost due to the discontinuity of the characteristic functions. We propose a method of smoothing characteristic functions in a rather general case. We replace the characteristic functions by continuous ones, without changing the value of the integrals. Using this smoothing technique, we modify the rejection method. An extended smoothed rejection method is described. Numerical experiments show that the extended smoothed rejection method is much more efficient than the standard quasi-Monte Carlo and the unsmoothed rejection method when used with low discrepancy sequences.

MSC: 65C05; 65D30

Keywords: Quasi-Monte Carlo methods; Low discrepancy sequences; Monte Carlo methods; Rejection sampling; Numerical integration

1. Introduction

1.1. Quasi-Monte Carlo methods

Evaluation of complicated multidimensional integrals is a common computational problem occurring in many areas of science such as computational physics, statistics, computer graphics and mathematical finance [2,4,8,9,14].

[1] This work was supported by the State Education Commission of China.
E-mail address: xwang@math.tsinghua.edu.cn (X. Wang)

0377-0427/00/$ - see front matter
PII: S0377-0427(99)00194-6

To estimate the integral

$$I(f) = \int_{I^s} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \tag{1}$$

where $I^s = [0,1]^s$ is the s-dimensional unit cube, and $f : I^s \to \mathbb{R}$, one chooses a point set $P_N = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ in I^s and compute the estimate

$$I_N = \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{x}_i). \tag{2}$$

For Monte Carlo (MC) method, the points of P are independent identically distributed (i.i.d.) samples from the uniform distribution on I^s. For the Quasi-Monte Carlo (QMC) method, P_N is a *low discrepancy point set* [13] or *lattice rules* [15]. The error bound of the QMC estimate is the form (see [7])

$$\left| \int_{I^s} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{x}_i) \right| \leqslant V(f)D(P_N), \tag{3}$$

where $D(P_N)$ is the *discrepancy* of P_N, which measures the regularity of P_N, and $V(f)$ is the *variation* of f over I^s, which measures the variability of the integrand f. To make the above statement more precise, the *star-discrepancy* $D^*(P_N)$ (see the next section) and the variation in the sense of Hardy and Crause are defined; and the inequality (3) reduces to the classical Koksma–Hlawka inequality [13]. Infinite low discrepancy sequences with star-discrepancy $\mathrm{O}(N^{-1}(\log N)^s)$ have been constructed [5,6,13,16]. So QMC integration for integrands of bounded variation (in the sense of Hardy and Crause) has theoretical error bound $\mathrm{O}(N^{-1}(\log N)^s)$, which is much faster than the probabilistic error bound $\mathrm{O}(N^{-1/2})$ of the MC methods.

When the integrand has unbounded variation, the Koksma–Hlawka inequality cannot be used to derive an upper bound. The computational experiments (see [1,11]) show that for some integrands of unbounded variation, the observed convergence rate is only between $\mathrm{O}(N^{-1/2})$ and $\mathrm{O}(N^{-2/3})$. The limitation of QMC on smoothness of integrand is significant, because many applications involve decisions of some sort, which usually corresponds to characteristic functions.

1.2. The importance sampling and the rejection sampling

Importance sampling is probably the most widely used variance reduction technique in MC methods. Rewrite the integral $I(f)$ as

$$I(f) = \int_{I^s} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{I^s} \frac{f(\boldsymbol{x})}{p(\boldsymbol{x})}\,p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \tag{4}$$

where $p(\boldsymbol{x})$ is an importance function, which is chosen such that it mimics the behavior of $f(\boldsymbol{x})$ over I^s. The importance-sampled estimate is

$$I_N^{(\mathrm{IS})} = \frac{1}{N}\sum_{i=1}^{N} \frac{f(\boldsymbol{x}_i)}{p(\boldsymbol{x}_i)}, \tag{5}$$

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ are i.i.d. samples from the density $p(\boldsymbol{x})$.

A crucial part in this method is that sampling from the density $p(\boldsymbol{x})$ is required. If the density $p(\boldsymbol{x})$ is simple, the transformation method or the conditional distribution method can be used [3]. For more general density, some other generation procedures must be used, such as the *rejection sampling* method.

The importance sampling and the rejection sampling have obvious deterministic analogies in QMC methods. However, it was found [12] that the rejection sampling technique lose its effectiveness in QMC setting (means QMC performs no better than standard MC). The reason is that the rejection-sampling-based estimate (5) can be interpreted as the integration of a discontinuous integrand, which involves a characteristic function coming from the decision of acceptance or rejection.

Attempts have been made to modify the rejection sampling. The *smoothed* version of the standard rejection method is described in [12].

In this paper, we investigate and improve the performance of deterministic rejection method in QMC methods. We study the "quality" of the point sets generated by standard deterministic rejection method in Section 2. It turns out that the quality is closely related to the problem of QMC integration of characteristic functions, whose accuracy may be lost due to the discontinuity of the characteristic functions. In Section 3, in a rather general case, a method of smoothing characteristic functions is proposed. The characteristic functions are replaced by continuous ones, which have the same value of integral. This smoothing method is used to improve the rejection method in Section 4. An extended smoothed version is described, which is a generalization of the method in [12]. Numerical experiments are given in Section 5.

2. Deterministic rejection in quasi-Monte Carlo methods — how well does it behave?

2.1. Discrepancy and F-discrepancy

Definition 2.1. The *star-discrepancy* of a point set $P_N = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ in I^s is defined by

$$D^*(P_N) = \sup_{J \in \mathscr{J}^*} \left| \frac{1}{N} \sum_{i=1}^{N} c_J(\boldsymbol{x}_i) - m(J) \right|,$$

where $\mathscr{J}^*$ is the family of all subintervals of I^s of the form $\prod_{i=1}^{s} [0, u_i)$, $c_J(\boldsymbol{x})$ is the characteristic function of J, and $m(J)$ is the volume of J with respect to the Lebesque measure.

Star-discrepancy can be seen as maximum integration error for the characteristic functions of the set in the family $\mathscr{J}^*$.

Definition 2.2. An infinite sequence $\{\boldsymbol{x}_i\}$ in I^s is called *uniformly distributed* (or *equidistributed*) if for all Riemann integrable function $f(\boldsymbol{x})$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{x}_i) = \int_{I^s} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}.$$

It should be observed that an infinite sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ of points on I^s is uniformly distributed if and only if $\lim_{N \to \infty} D^*(P_N) = 0$ (see [13]).

Now let $\boldsymbol{x} = (X_1, X_2, \ldots, X_s)$ be a random vector with probability density function $p(\boldsymbol{x})$ and cumulative distribution function (c.d.f.) $F(\boldsymbol{x})$ defined on I^s. We introduce the more general concept of discrepancy [4].

Definition 2.3. Assume that $P_N = \{\boldsymbol{x}_i,\ i = 1, \ldots, N\}$ is a set of points in I^s, and $F_N(\boldsymbol{x})$ is the empirical distribution of P_N, i.e., $F_N(\boldsymbol{x}) = (1/N)\sum_{i=1}^{N} c_{\{\boldsymbol{x}_i \leqslant \boldsymbol{x}\}}$. The *F-discrepancy* of P_N with respect to $F(\boldsymbol{x})$ is defined by

$$D_F(P_N) = \sup_{\boldsymbol{x} \in I^s} |F_N(\boldsymbol{x}) - F(\boldsymbol{x})|.$$

The F-discrepancy is a measure for the quality of the representation of $F(\boldsymbol{x})$ by a point set P_N. It is in fact the Smirnov–Kolmogorov statistic for goodness-of-fit. When $F(\boldsymbol{x})$ is the uniform distribution on I^s, the F-discrepancy reduces to the star-discrepancy. Another reasonable measure of representation is the so-called *quasi-F-discrepancy* (see [4]).

Points that aim for minimizing F-discrepancy or quasi-F-discrepancy are called *representative points* of $F(\boldsymbol{x})$. The generation of such points is a central part in many applications (such as integration, optimization, experimental design, etc.).

2.2. Standard deterministic rejection method

Let $p(\boldsymbol{x})$ be a probability density function defined on I^s, and suppose that $0 < p(\boldsymbol{x}) < M$. The algorithm of standard rejection can be described as follows:

1. Generate $(\boldsymbol{x}, y)$ from the uniform distribution on I^{s+1} ($\boldsymbol{x} \in I^s,\ y \in [0,1]$).
2. If $y < M^{-1}p(\boldsymbol{x})$, accept the point $\boldsymbol{x}$; otherwise, reject the point $\boldsymbol{x}$.
3. Repeat Steps 1 and 2, until the necessary number of points have been accepted.

This random rejection sampling algorithm has the following deterministic version in QMC methods ("deterministic rejection algorithm"):

1. Generate a low discrepancy sequence P of points in I^{s+1}:

$$P = \{(\boldsymbol{x}_i, y_i) \in I^{s+1}\colon \boldsymbol{x}_i \in I^s,\ y_i \in [0,1],\ i = 1, 2, \ldots\}.$$

The point set consisting of the first N elements of P is denoted by P_N.

2. For $i = 1, 2, \ldots$, if $y_i < M^{-1}p(\boldsymbol{x}_i)$, accept the point $\boldsymbol{x}_i$; otherwise, reject it.

This algorithm produces a sequence Q of accepted points in I^s. Let $Q_{K(N)}$ denote the set of accepted points in the first N "acceptance–rejection decisions". $K(N)$ is the number of points in the set $Q_{K(N)}$. Now we investigate the goodness of representation of the sequence Q for $F(\boldsymbol{x})$. For any point $\boldsymbol{x}^* \in I^s$, define the set

$$E(\boldsymbol{x}^*) = \{(\boldsymbol{x}, y) \in I^{s+1}\colon \boldsymbol{x} < \boldsymbol{x}^*,\ y < M^{-1}p(\boldsymbol{x}),\ \boldsymbol{x} \in I^s,\ y \in [0,1]\}. \tag{6}$$

Let $A([\boldsymbol{0}, \boldsymbol{x}^*), Q_{K(N)})$ denote the number of points of $Q_{K(N)}$ falling in the subinterval $[\boldsymbol{0}, \boldsymbol{x}^*)$, and $A(E(\boldsymbol{x}^*), P_N)$ denote the number of points of P_N falling in the set $E(\boldsymbol{x}^*)$. Then for any point $\boldsymbol{x}^* \in I^s$, we have $A([\boldsymbol{0}, \boldsymbol{x}^*), Q_{K(N)}) = A(E(\boldsymbol{x}^*), P_N)$.

The absolute difference between the empirical distribution $F_{K(N)}(\boldsymbol{x})$ of the accepted point set $Q_{K(N)}$ and the distribution $F(\boldsymbol{x})$ can be estimated as follows:

$$|F_{K(N)}(\boldsymbol{x}^*) - F(\boldsymbol{x}^*)| = \left|\frac{A([\boldsymbol{0}, \boldsymbol{x}^*), Q_{K(N)})}{K(N)} - F(\boldsymbol{x}^*)\right|$$

$$\leqslant \frac{N}{K(N)}\left\{\left|\frac{A(E(\boldsymbol{x}^*), P_N)}{N} - \frac{F(\boldsymbol{x}^*)}{M}\right| + F(\boldsymbol{x}^*)\left|\frac{K(N)}{N} - \frac{1}{M}\right|\right\}. \tag{7}$$

Note that the volume of the set $E(\boldsymbol{x}^*)$ is

$$\mathrm{Vol}(E(\boldsymbol{x}^*)) = \int_{[\boldsymbol{0}, \boldsymbol{x}^*]}\left(\int_0^{M^{-1}p(\boldsymbol{x})} \mathrm{d}y\right)\mathrm{d}\boldsymbol{x} = \frac{F(\boldsymbol{x}^*)}{M}.$$

If the sequence P is uniformly distributed over I^{s+1}, then for any point $\boldsymbol{x}^* \in I^s$,

$$\left|\frac{A(E(\boldsymbol{x}^*), P_N)}{N} - \frac{F(\boldsymbol{x}^*)}{M}\right| \to 0 \quad \text{as } N \to \infty.$$

Specifically, put $\boldsymbol{x}^* = (1, \ldots, 1)$, then $F(\boldsymbol{x}^*) = 1$ and $A(E(\boldsymbol{x}^*), P_N) = K(N)$. So

$$\left|\frac{K(N)}{N} - \frac{1}{M}\right| \to 0 \quad \text{as } N \to \infty,$$

and for sufficiently large N, $N/K(N) \leqslant M + 1$. Therefore, from (7), we have

$$|F_{K(N)}(\boldsymbol{x}^*) - F(\boldsymbol{x}^*)| \to 0 \quad \text{as } N \to \infty.$$

Theorem 2.4. *Let P be a uniformly distributed sequence of points in I^{s+1}, and Q the sequence obtained by the deterministic rejection algorithm. Then for any $\boldsymbol{x}^* \in I^s$,*

$$|F_{K(N)}(\boldsymbol{x}^*) - F(\boldsymbol{x}^*)| \to 0 \quad \textit{as } N \to \infty. \tag{8}$$

where $F_{K(N)}(\boldsymbol{x}^*)$ is the empirical distribution of the "accepted" point set $Q_{K(N)}$.

Suppose that P_N is a low discrepancy point set with discrepancy $\mathrm{O}(N^{-1}(\log N)^{s+1})$. What is the convergence rate of (8)? As we see that the "quality" of the sequence Q is determined by the term

$$\left|\frac{A(E(\boldsymbol{x}^*), P_N)}{N} - \mathrm{Vol}(E(\boldsymbol{x}^*))\right|. \tag{9}$$

This term is in fact the error of QMC integration of the characteristic function of the set $E(\boldsymbol{x}^*)$ over I^{s+1}, that is

$$\left|\frac{A(E(\boldsymbol{x}^*), P_N)}{N} - \mathrm{Vol}(E(\boldsymbol{x}^*))\right| = \left|\frac{1}{N}\sum_{i=1}^{N} c_{E(\boldsymbol{x}^*)}(\boldsymbol{x}_i, y_i) - \int_{I^s}\int_0^1 c_{E(\boldsymbol{x}^*)}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}y\right|, \tag{10}$$

where $c_{E(\boldsymbol{x}^*)}$ is the characteristic function of $E(\boldsymbol{x}^*)$. In general, the Koksma–Hlawka inequality cannot be used here to derive an upper bound for the right-hand side of (10), because the variation of $c_{E(\boldsymbol{x}^*)}$ can be infinite (the variation is finite, only if the set $E(\boldsymbol{x}^*)$ is rectangles with sides parallel to the coordinate axes).

The known theoretical bounds for (9) (see [13]) may only be $O(N^{-1/(s+1)})$, which is too pessimistic. Now we provide an informal argument similarly as in [1], which leads to the convergence rate of the integration error (10) to be of the order $O(N^{-(s+2)/2(s+1)})$. Consider the case of $\boldsymbol{x}^* = (1,\ldots,1)$. In this case the set defined in (6) is

$$E^* = \{(\boldsymbol{x}, y) \in I^{s+1}\colon\ y < M^{-1}p(\boldsymbol{x}),\ \boldsymbol{x} \in I^s,\ y \in [0,1]\}.$$

Choose $A(\boldsymbol{x})$ and $B(\boldsymbol{x})$ to be step functions, such that $0 \leqslant A(\boldsymbol{x}) \leqslant M^{-1}p(\boldsymbol{x}) \leqslant B(\boldsymbol{x}) \leqslant 1$, for all $\boldsymbol{x} \in [0,1]$, and such that $B(\boldsymbol{x}) - A(\boldsymbol{x})$ is of the order $O(N^{-1/(s+1)})$ (this is the size of the mean distance of the neighbouring points of P_N). The set

$$E_a^b = \{(\boldsymbol{x}, y) \in I^{s+1}\colon A(\boldsymbol{x}) \leqslant y \leqslant B(\boldsymbol{x}),\ \boldsymbol{x} \in I^s\}$$

is called the *domain of discontinuity*. Decompose the characteristic function c_{E^*} into two functions, $c_{E^*} = f_c + f_d$, where f_c equals to c_{E^*} outside E_a^b, and equals to zero inside E_a^b; f_d equals to c_{E^*} inside E_a^b, and zero outside E_a^b. So f_c has finite variation and f_d inherits the discontinuity of c_{E^*}. The volume of the domain of discontinuity $\mathrm{Vol}(E_a^b)$ is of the order $O(N^{-1/(s+1)})$, and the number of points falling in E_a^b, $A(E_a^b, P_N)$, is of the order $O(N^{s/(s+1)})$. We can suppose that in the domain of discontinuity E_a^b low discrepancy points behave like pseudo-random ones [1,11], then

$$\left| \int_{I^s} \int_0^1 f_d \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}y - \frac{1}{N}\sum_{i=1}^{N} f_d(\boldsymbol{x}_i, y_i) \right| = O(\mathrm{Vol}(E_a^b)(A(E_a^b, P_N))^{-1/2}) = O(N^{-(s+2)/2(s+1)}).$$

Taking into account that f_c has finite variation, one can conclude that the integration error (10) is of the order $O(N^{-(s+2)/2(s+1)})$.

The experiments in [1,11] show that the integration errors of characteristic functions (i.e., the error term (9)) can be approximated by $cN^{-\alpha}$ for a real constant c and $\frac{1}{2} \leqslant \alpha \leqslant 1$.

Now suppose that for all $\boldsymbol{x}^* \in I^s$

$$\left| \frac{A(E(\boldsymbol{x}^*), P_N)}{N} - \mathrm{Vol}(E(\boldsymbol{x}^*)) \right| \leqslant c N^{-\alpha} \quad (1/2 \leqslant \alpha < 1). \tag{11}$$

Then

$$\left| \frac{A(E(\boldsymbol{x}^*), P_N)}{N} - \frac{F(\boldsymbol{x}^*)}{M} \right| \leqslant cN^{-\alpha},$$

and specifically

$$\left| \frac{K_N}{N} - \frac{1}{M} \right| \leqslant cN^{-\alpha}.$$

Therefore, from (7), for some constant C, we have for all $\boldsymbol{x}^* \in I^s$

$$|F_{K(N)}(\boldsymbol{x}^*) - F(\boldsymbol{x}^*)| \leqslant CN^{-\alpha}.$$

Theorem 2.5. *Let P be a low discrepancy sequence of points in I^{s+1}. Under the assumption of* (11), *the sequence Q obtained by the deterministic rejection algorithm has F-discrepancy of the order* $O(N^{-\alpha})$ (*where* $\frac{1}{2} \leqslant \alpha < 1$).

We have showed that the quality of the point sets generated by deterministic rejection algorithm is closely related to the errors of QMC integration of characteristic functions. An informal argument and practical experiences show that even for characteristic functions, QMC methods can be better than MC, but their advantages seem to be less than predicted by the theory for functions of finite variation. Consequently, the F-discrepancy of the point set generated by deterministic rejection algorithm is smaller than that of the point set generated by random rejection algorithm, but we cannot expect that it is of the order of nearly $O(N^{-1})$. Recall that the F-discrepancy of the point set generated by the transformation method or the conditional distribution method may in the order of nearly $O(N^{-1})$ (see [17]).

3. Smoothing of characteristic functions

Let B be a Jordan-measurable subset (i.e., subset for which the characteristic function is Riemann integrable) of I^s, and $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_N\}$ be a low discrepancy set of points in I^s. As was suggested in [4,10,13], the integral

$$\int_{I^s} c_B(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},$$

(that is the volume of B) can be approximated by standard QMC estimate

$$\frac{1}{N}\sum_{i=1}^{N} c_B(\boldsymbol{x}_i),$$

where $c_B(\boldsymbol{x})$ is the characteristic function of B. But this approximation often leads to large errors. The loss of accuracy is due to the discontinuity or lack of smoothness in the integrand. How to improve the performance of QMC integration of characteristic functions? Our attempt is to replace the characteristic function by a continuous one, without changing the value of the integral.

The following case deserves special attention. Let $h(\boldsymbol{x})$ be a continuous function defined on I^s. Without loss of generality, assume that $0 < h(\boldsymbol{x}) < 1$. Define the set

$$E = \{(\boldsymbol{x}, y) \in I^{s+1}\colon\ y < h(\boldsymbol{x}),\ \boldsymbol{x} \in I^s,\ y \in [0,1]\}. \tag{12}$$

For convenience of notation, here E is a subset of $(s+1)$-dimensional unit cube I^{s+1}. Note that the integration of the characteristic function of the set E over I^{s+1} is related to the integration of $h(\boldsymbol{x})$ over I^s:

$$\int_{I^s} h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{I^s}\int_0^1 c_E(\boldsymbol{x}, y)\,\mathrm{d}y\,\mathrm{d}\boldsymbol{x}. \tag{13}$$

We will smooth the characteristic function of E, $c_E(\boldsymbol{x}, y)$, *without changing the value* of the integral. Choose functions $a(\boldsymbol{x})$ and $b(\boldsymbol{x})$, such that

$$0 \leqslant a(\boldsymbol{x}) < h(\boldsymbol{x}) < b(\boldsymbol{x}) \leqslant 1 \quad \text{for all } \boldsymbol{x} \in I^s,$$

and such that $a(\boldsymbol{x})$ and $b(\boldsymbol{x})$ are easy to evaluate. Define the *smoothed function* $w(\boldsymbol{x}, y)$ as follows:

$$w(\boldsymbol{x}, y) = \begin{cases} 1 & \text{if } y \in [0, a(\boldsymbol{x})), \\ 1 + \dfrac{h(\boldsymbol{x}) - b(\boldsymbol{x})}{(b(\boldsymbol{x}) - a(\boldsymbol{x}))(h(\boldsymbol{x}) - a(\boldsymbol{x}))}(y - a(\boldsymbol{x})) & \text{if } y \in [a(\boldsymbol{x}), h(\boldsymbol{x})), \\ \dfrac{h(\boldsymbol{x}) - a(\boldsymbol{x})}{(b(\boldsymbol{x}) - a(\boldsymbol{x}))(h(\boldsymbol{x}) - b(\boldsymbol{x}))}(y - b(\boldsymbol{x})) & \text{if } y \in [h(\boldsymbol{x}), b(\boldsymbol{x})), \\ 0 & \text{if } y \in [b(\boldsymbol{x}), 1]. \end{cases} \tag{14}$$

Theorem 3.1. *Let the function $w(\boldsymbol{x}, y)$ be defined by* (14). *Then*

(i) *The function $w(\boldsymbol{x}, y)$ is continuous with respect to y over* $[0, 1]$.
(ii) *The integration of $w(\boldsymbol{x}, y)$ over I^{s+1} is the same as that of the characteristic function $c_E(\boldsymbol{x}, y)$*:

$$\int_{I^s} \int_0^1 w(\boldsymbol{x}, y)\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x} = \int_{I^s} \int_0^1 c_E(\boldsymbol{x}, y)\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}. \tag{15}$$

Proof. The continuity of $w(\boldsymbol{x}, y)$ with respect to y is obvious. We now prove (15). Since

$$\int_0^1 w(\boldsymbol{x}, y)\, \mathrm{d}y = \int_0^{a(\boldsymbol{x})} 1 \cdot \mathrm{d}y + \int_{a(\boldsymbol{x})}^{h(\boldsymbol{x})} w(\boldsymbol{x}, y)\, \mathrm{d}y + \int_{h(\boldsymbol{x})}^{b(\boldsymbol{x})} w(\boldsymbol{x}, y)\, \mathrm{d}y.$$

According to the construction of $w(\boldsymbol{x}, y)$, by direct computation we obtain

$$\int_{a(\boldsymbol{x})}^{h(\boldsymbol{x})} w(\boldsymbol{x}, y)\, \mathrm{d}y = h(\boldsymbol{x}) - a(\boldsymbol{x}) + \frac{(h(\boldsymbol{x}) - a(\boldsymbol{x}))(h(\boldsymbol{x}) - b(\boldsymbol{x}))}{2(b(\boldsymbol{x}) - a(\boldsymbol{x}))}$$

and

$$\int_{h(\boldsymbol{x})}^{b(\boldsymbol{x})} w(\boldsymbol{x}, y)\, \mathrm{d}y = -\frac{(h(\boldsymbol{x}) - a(\boldsymbol{x}))(h(\boldsymbol{x}) - b(\boldsymbol{x}))}{2(b(\boldsymbol{x}) - a(\boldsymbol{x}))}.$$

So we have

$$\int_0^1 w(\boldsymbol{x}, y)\, \mathrm{d}y = h(\boldsymbol{x}). \tag{16}$$

Therefore, integrating both sides of (16) with respect to $\boldsymbol{x}$ over I^s, we have

$$\int_{I^s} \int_0^1 w(\boldsymbol{x}, y)\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x} = \int_{I^s} h(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}.$$

The relation (15) follows immediately from (13). □

The function $w(\boldsymbol{x}, y)$ is continuous, but may not be differentiable at $y = a(\boldsymbol{x})$, $y = h(\boldsymbol{x})$, or $y = b(\boldsymbol{x})$. Many authors point out that the continuity of the integrand is important for QMC integration, but the differentiability does not seem to have any advantage [1]. This fact gives us confidence to replace the characteristic function $c_E(\boldsymbol{x}, y)$ by the continuous one $w(\boldsymbol{x}, y)$. The choice of functions $a(\boldsymbol{x})$ and $b(\boldsymbol{x})$ may influence on the effect of smoothing. In general, the differences $h(\boldsymbol{x}) - a(\boldsymbol{x})$ and $b(\boldsymbol{x}) - h(\boldsymbol{x})$ should not be made too small. Otherwise, we will make the integrand only "formally" continuous (i.e., with sharp edges), and cannot expect too much gain in accuracy. In practice, if

$h(\boldsymbol{x})$ is relatively easy to evaluate, we can choose $a(\boldsymbol{x}) = h(\boldsymbol{x}) - \delta$, and $b(\boldsymbol{x}) = h(\boldsymbol{x}) + \delta$ ($\delta \geqslant 0$). We can also choose $a(\boldsymbol{x})$ and $b(\boldsymbol{x})$ to be piecewise constant functions. Another simple choice is $a(\boldsymbol{x}) = 0$, and $b(\boldsymbol{x}) = 1$.

The smoothing method described in this section will be demonstrated on the rejection sampling method, which involves a characteristic function in the integrand, coming from the decision to accept or reject.

4. Improved rejection sampling procedure in quasi-Monte Carlo methods

Now suppose that our aim is to estimate the integral $I(f)$ defined by (1). To speed up the standard MC and QMC estimates (2), the importance sampling technique is often used. Suppose the importance function $p(\boldsymbol{x})$ has been chosen, such that it models the behavior of the integrand $f(\boldsymbol{x})$. Rewrite the integral $I(f)$ as in (4). The corresponding estimate of the integral $I(f)$ is $I_N^{(\mathrm{IS})}$ defined in (5). For MC, the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ are i.i.d. samples from density $p(\boldsymbol{x})$. For QMC, these points are deterministic, which aim for minimizing the F-discrepancy or quasi-F-discrepancy.

To use importance sampling, it is required to sample from the distribution $p(\boldsymbol{x})$. "Simple" density can be sampled by the transformation or the conditional distribution methods. For more general density, the rejection sampling method is widely used. The advantage of rejection sampling is that it does not require inverse of the cumulative distribution function.

However, the rejection sampling method in its standard form is not well suited for low discrepancy sequences [12].

The estimate (5), in which the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ are generated from $p(\boldsymbol{x})$ by rejection sampling method, is in fact a slight modification of the standard MC estimates for the following integral:

$$I(f) = M \int_{I^s} \int_0^1 \frac{f(\boldsymbol{x})}{p(\boldsymbol{x})} \cdot c_{E^*}(\boldsymbol{x}, y)\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}, \tag{17}$$

where

$$E^* = \left\{(\boldsymbol{x}, y) \in I^{s+1} \colon y < M^{-1} p(\boldsymbol{x}),\ \boldsymbol{x} \in I^s,\ y \in [0,1]\right\}. \tag{18}$$

Direct use of low discrepancy sequences (instead of pseudo-random numbers) to the integrand in (17) may lead to large errors, because the integrand in (17) contains a characteristic function, corresponding to the decision of accept or reject.

To effectively use importance sampling in QMC setting for estimating integral (4), the rejection method is modified by replacing the accept/rejection decision with the assignment of an appropriate "acceptance" weight to each sample point.

Choose functions $A(\boldsymbol{x})$ and $B(\boldsymbol{x})$, such that

$$0 \leqslant A(\boldsymbol{x}) < p(\boldsymbol{x}) < B(\boldsymbol{x}) \leqslant M \quad \text{for all } \boldsymbol{x} \in I^s,$$

and such that $A(\boldsymbol{x})$ and $B(\boldsymbol{x})$ are cheaper to evaluate than $p(\boldsymbol{x})$.

Note that the set E^* defined by (18) has the form of (12) (corresponding to $h(\boldsymbol{x}) = M^{-1}p(\boldsymbol{x})$). According to (14), the smoothed function of $c_{E^*}(\boldsymbol{x}, y)$ has the next form

$$W(\boldsymbol{x}, y) = \begin{cases} 1 & \text{if } My \in [0, A(\boldsymbol{x})), \\ 0 & \text{if } My \in [B(\boldsymbol{x}), M], \\ 1 + \dfrac{p(\boldsymbol{x}) - B(\boldsymbol{x})}{(B(\boldsymbol{x}) - A(\boldsymbol{x}))(p(\boldsymbol{x}) - A(\boldsymbol{x}))}(My - A(\boldsymbol{x})) & \text{if } My \in [A(\boldsymbol{x}), p(\boldsymbol{x})), \\ \dfrac{p(\boldsymbol{x}) - A(\boldsymbol{x})}{(B(\boldsymbol{x}) - A(\boldsymbol{x}))(p(\boldsymbol{x}) - B(\boldsymbol{x}))}(My - B(\boldsymbol{x})) & \text{if } My \in [p(\boldsymbol{x}), B(\boldsymbol{x})). \end{cases} \tag{19}$$

This is obtained by putting $h(\boldsymbol{x}) = M^{-1}p(\boldsymbol{x})$, $a(\boldsymbol{x}) = M^{-1}A(\boldsymbol{x})$, and $b(\boldsymbol{x}) = M^{-1}B(\boldsymbol{x})$ in (14).

Based on Theorem 3.1 and the relation (16), the function $W(\boldsymbol{x}, y)$ is continuous with respect to y, and satisfies

$$\int_0^1 c_{E^*}(\boldsymbol{x}, y)\,\mathrm{d}y = \int_0^1 W(\boldsymbol{x}, y)\,\mathrm{d}y = M^{-1}p(\boldsymbol{x}). \tag{20}$$

Therefore, the integral $I(f)$ can be rewritten as

$$I(f) = M\int_{I^s}\int_0^1 \frac{f(\boldsymbol{x})}{p(\boldsymbol{x})} \cdot c_{E^*}(\boldsymbol{x}, y)\,\mathrm{d}y\,\mathrm{d}\boldsymbol{x} = M\int_{I^s}\int_0^1 \frac{f(\boldsymbol{x})}{p(\boldsymbol{x})} \cdot W(\boldsymbol{x}, y)\,\mathrm{d}y\,\mathrm{d}\boldsymbol{x}. \tag{21}$$

The standard MC approximation of the integral on the right-hand side of (21) corresponds to the following *extended smoothed rejection procedure*.

Step 1. Choose functions $A(\boldsymbol{x})$ and $B(\boldsymbol{x})$ as indicated above.
Step 2. Generate $(\boldsymbol{x}_i, y_i)$ from the uniform distribution on I^{s+1}.
Step 3. Assign each sample point an acceptance weight $W(\boldsymbol{x}_i, y_i)$, defined by (19).
Step 4. Repeat Steps 2 and 3, until

$$\sum_{i=1}^{N^*} W(\boldsymbol{x}_i, y_i) \approx N, \tag{22}$$

where N^* is the total number of trial points.

The theory behind this procedure is based on the following theorem, which is a generalization of the result in [12].

Theorem 4.1. *The density function of accepted points* $\boldsymbol{x}$ *using the extended smoothed rejection procedure is* $p(\boldsymbol{x})$.

Proof. Because an "a priori" distribution of $\boldsymbol{x}$ is uniform distribution on I^s (Step 2), and its acceptance probability (given $\boldsymbol{x}$) is

$$\int_0^1 W(\boldsymbol{x}, y)\,\mathrm{d}y = M^{-1}p(\boldsymbol{x})$$

(Step 3, and see (20)). So by Bayes' formula, the density function of accepted points $\boldsymbol{x}$ is given by

$$\frac{1 \cdot \int_0^1 W(\boldsymbol{x}, y)\,\mathrm{d}y}{\int_{I^s}\{1 \cdot \int_0^1 W(\boldsymbol{u}, y)\,\mathrm{d}y\}\,\mathrm{d}\boldsymbol{u}} = \frac{M^{-1}p(\boldsymbol{x})}{\int_{I^s} M^{-1}p(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}} = \frac{M^{-1}p(\boldsymbol{x})}{M^{-1}} = p(\boldsymbol{x}). \qquad \square$$

By comparison with the original smoothed rejection method [12], the extended smoothed rejection method gives more freedom to construct functions $A(\boldsymbol{x})$ and $B(\boldsymbol{x})$ which model the importance function from below and from above. Under the special choices of $A(\boldsymbol{x}) = p(\boldsymbol{x}) - \frac{1}{2}M\delta$, and $B(\boldsymbol{x}) = p(\boldsymbol{x}) + \frac{1}{2}M\delta$, $(\delta \geqslant 0)$, our procedure reduces to the original smoothed rejection method in [12].

Another advantage of the extended smoothed rejection method is that it economizes some computation of the density function. Because in Step 3 of our procedure, if $My \geqslant B(\boldsymbol{x})$, then according to (19), $W(\boldsymbol{x}, y) = 0$, there is no need to calculate the density $p(\boldsymbol{x})$. This will become more clear in the following estimate (23). This is a great economy, especially when $p(\boldsymbol{x})$ is expensive to calculate (this often occurs in practice). In the original technique [12], it is always required to calculate the density $p(\boldsymbol{x})$. Therefore, the extended smoothed rejection method is a more flexible and more economic procedure than the original one.

The extended smoothed rejection method can be used both in MC and QMC setting. In QMC setting, the points $(\boldsymbol{x}_i, y_i)$ in Step 2 are the points of certain low discrepancy sequence on I^{s+1}.

The estimate for the integral $I(f)$ using the extended smoothed rejection is

$$\frac{1}{N}\sum_{i=1}^{N^*} W(\boldsymbol{x}_i, y_i)\frac{f(\boldsymbol{x}_i)}{p(\boldsymbol{x}_i)}, \tag{23}$$

where $W(\boldsymbol{x}_i, y_i)$ is defined by (19), and the N^* is chosen such that the sum of the acceptance weights is approximately equal to N (see (22)).

The estimate (23) is a slight modification of the next standard MC or QMC approximations

$$I(f) = M\int_{I^s}\int_0^1 \frac{f(\boldsymbol{x})}{p(\boldsymbol{x})} \cdot W(\boldsymbol{x}, y)\,\mathrm{d}y\,\mathrm{d}\boldsymbol{x} \approx \frac{M}{N^*}\sum_{i=1}^{N^*} \frac{f(\boldsymbol{x}_i)}{p(\boldsymbol{x}_i)} \cdot W(\boldsymbol{x}_i, y_i).$$

When one uses the estimate (23), there is no need to determine which sample points are accepted or rejected, this decision is replaced by allowing each sample point to have appropriate acceptance weight, which has corresponding contribution in the sum of the estimate (23).

Similarly as in Section 3, the differences $B(\boldsymbol{x}) - p(\boldsymbol{x})$ and $p(\boldsymbol{x}) - A(\boldsymbol{x})$ should not be made too small. Moreover, they should not be made too large, because they determine the number of necessary sample points in order to reach a total acceptance weight of size N and so determine the necessary evaluations of $p(\boldsymbol{x})$ and $f(\boldsymbol{x})$.

5. Computational experiments

In this section, the standard MC and QMC estimates, the standard rejection method and the extended smoothed methods will be compared on several examples. The following estimates will be

computed:

$$\text{Standard}: Y_N^{(1)} = (1/N)\sum_{i=1}^{N} f(\boldsymbol{x}_i), \quad \boldsymbol{x}_i \sim \mathrm{U}([0,1]^s);$$

$$\text{Rej. Meth.}: Y_N^{(2)} = (1/N)\sum_{i=1}^{N} f(\boldsymbol{x}_i)/p(\boldsymbol{x}_i), \quad \boldsymbol{x}_i \sim p(\boldsymbol{x}), \text{ accepted points};$$

$$\text{Ex. Sm. Rej.}: Y_N^{(3)} = (1/N)\sum_{i=1}^{N^*} W(\boldsymbol{x}_i, y_i) f(\boldsymbol{x}_i)/p(\boldsymbol{x}_i).$$

All three of these methods are performed both in MC setting and in QMC setting. For a given value of N, take m samples of these estimates, denoted by $Y_N^{(j)}(k)$ for $1 \leqslant k \leqslant m$. In QMC setting, we use multiple random-start Halton sequences, which have the additional advantage of practical error estimation [18].

The final approximation of integral $I(f)$ is given by $\hat{I}^{(j)} = (1/m)\sum_{k=1}^{m} Y_N^{(j)}(k)$. In all cases the errors can be estimated by sample variance:

$$\hat{\sigma}_j^2 = \frac{1}{m(m-1)}\sum_{k=1}^{m} [Y_N^{(j)}(k) - \hat{I}^{(j)}]^2, \quad j = 1,2,3. \tag{24}$$

All these estimates (both in MC and QMC setting) are compared with the standard MC estimate (with the same sample size) by comparing their relative efficiencies, defined by

$$\mathrm{REff}(\hat{I}^{(j)}) = \frac{\sigma^2(\hat{I}^{(\mathrm{mc})})}{\sigma^2(\hat{I}^{(j)})}, \quad j = 1,2,3.$$

Here $\hat{I}^{(\mathrm{mc})}$ is the standard MC estimate (it is the same as $\hat{I}^{(1)}$ in MC setting).

Example 1. Consider the evaluation of the following integral, which is a modification of the test integral introduced in [12]:

$$\int_{I^5} \exp\left(\sum_{i=1}^{5} a_i x_i^2 \left(1 + \frac{1}{2}\sin\left(\sum_{j=1, j\neq i}^{5} x_j\right)\right)\right) \mathrm{d}\boldsymbol{x}.$$

Three different choices of the parameters $a_1, \ldots, a_5$ will be considered:

(i) $a_1 = 1,\ a_2 = \frac{1}{2},\ a_3 = a_4 = a_5 = \frac{1}{5}$ (this is the case considered in [12]);
(ii) $a_k = 1/k,\ 1 \leqslant k \leqslant 5$;
(iii) $a_k = 1/k^2,\ 1 \leqslant k \leqslant 5$.

Choose the importance distribution

$$p(\boldsymbol{x}) = \frac{1}{C}\exp\left(\sum_{i=1}^{5} a_i x_i^2\right),$$

where $C = \int_{I^5} \exp(\sum_{i=1}^{5} a_i x_i^2)\,\mathrm{d}\boldsymbol{x}$ is the normalized constant, which can be easily computed as the product of 5 one-dimensional integrals. We can choose functions $A(\boldsymbol{x})$ and $B(\boldsymbol{x})$ as follows:

$$A(\boldsymbol{x}) = \frac{1}{C}\left(1 + \sum_{i=1}^{5} a_i x_i^2\right), \quad B(\boldsymbol{x}) = \frac{1}{C}\exp\left(\sum_{i=1}^{5} a_i\right).$$

Table 1
Comparison of the estimated standard derivation $\hat{\sigma}_j$ and the relative efficiency (in parentheses) for Example 1, case (i), with 64 runs

	Monte Carlo			Quasi-Monte Carlo		
N	St. MC	Rej. Meth.	Ex. Sm. Rej.	St. QMC	Rej. Meth.	Ex. Sm. Rej.
256	9.68e-3	3.44e-3	2.95e-3	1.27e-3	1.24e-3	7.99e-4
	(1.00)	(7.90)	(10.77)	(58.33)	(61.25)	(146.82)
1024	4.95e-3	1.71e-3	1.31e-3	3.57e-4	5.05e-4	2.63e-4
	(1.00)	(8.38)	(14.38)	(192.36)	(96.07)	(352.12)
4096	2.91e-3	9.04e-4	6.88e-4	8.72e-5	2.14e-4	7.03e-5
	(1.00)	(10.35)	(17.91)	(1113.73)	(184.26	(1713.40)
16384	1.65e-3	4.39e-4	3.52e-4	2.23e-5	9.01e-5	2.16e-5
	(1.00)	(14.05)	(21.84)	(5049.30)	(333.73)	(5833.77)

Table 2
Comparison of the estimated standard derivation $\hat{\sigma}_j$ and the relative efficiency (in parentheses) for Example 1 with $a_k = 1/k$ and 64 runs

	Monte Carlo			Quasi-Monte Carlo		
N	St. MC	Rej. Meth.	Ex. Sm. Rej.	St. QMC	Rej. Meth.	Ex. Sm. Rej.
256	1.08e-2	4.20e-3	3.72e-3	1.43e-3	1.42e-3	9.54e-4
	(1.00)	(6.70)	(8.52)	(57.23)	(58.86)	(130.19)
1024	5.50e-3	1.86e-3	1.42e-3	3.99e-4	5.27e-4	3.01e-4
	(1.00)	(8.75)	(14.95)	(189.99)	(108.87)	(333.89)
4096	3.22e-3	1.05e-3	8.86e-4	1.02e-4	2.07e-4	8.83e-5
	(1.00)	(9.38)	(13.21)	(996.57)	(242.65)	(1331.48)
16384	1.75e-3	4.92e-4	3.86e-4	2.64e-5	9.05e-5	2.35e-5
	(1.00)	(12.69)	(20.64)	(4429.01)	(375.73)	(5545.94)

The estimated standard derivation $\hat{\sigma}_j$ and the relative efficiencies (in parentheses) are given in Tables 1–3.

Example 2. The second example is the evaluation of the following integral [12]:

$$\int_{I^7} \exp\left(1 - \left(\sin^2\left(\frac{\pi}{2}x_1\right) + \sin^2\left(\frac{\pi}{2}x_2\right) + \sin^2\left(\frac{\pi}{2}x_3\right)\right)\right) \arcsin\left(\sin(1) + \frac{x_1 + \cdots + x_7}{200}\right) \mathrm{d}\boldsymbol{x}.$$

Choose the importance function

$$p^*(\boldsymbol{x}) = \frac{1}{C^*}\exp\left(1 - \left(\sin^2\left(\frac{\pi}{2}x_1\right) + \sin^2\left(\frac{\pi}{2}x_2\right) + \sin^2\left(\frac{\pi}{2}x_3\right)\right)\right),$$

Table 3
Comparison of the estimated standard derivation $\hat{\sigma}_j$ and the relative efficiency (in parentheses) for Example 1 with $a_k = 1/k^2$ and 64 runs

	Monte Carlo			Quasi-Monte Carlo		
N	St. MC	Rej. Meth.	Ex. Sm. Rej.	St. QMC	Rej. Meth.	Ex. Sm. Rej.
256	6.87e-3	2.71e-3	2.55e-3	6.88e-4	7.87e-4	5.22e-4
	(1.00)	(6.44)	(7.25)	(99.81)	(76.29)	(173.04)
1024	3.81e-3	1.32e-3	1.11e-3	1.90e-4	2.81e-4	1.43e-4
	(1.00)	(8.34)	(11.62)	(402.73)	(184.66)	(710.92)
4096	2.05e-3	6.11e-4	4.93e-4	4.14e-5	1.04e-4	4.31e-5
	(1.00)	(11.32)	(17.38)	(2459.01)	(392.27)	(2269.63)
16384	1.21e-3	2.60e-4	2.38e-4	1.29e-5	4.36e-5	1.00e-5
	(1.00)	(21.72)	(25.91)	(8892.91)	(774.19)	(14752.43)

Table 4
Comparison of the estimated standard derivation $\hat{\sigma}_j$ and the relative efficiency (in parentheses) for Example 2 with 64 runs

	Monte Carlo			Quasi-Monte Carlo		
N	St. MC	Rej. Meth.	Ex. Sm. Rej.	St. QMC	Rej. Meth.	Ex. Sm. Rej.
256	3.93e-3	5.80e-4	4.61e-4	3.46e-4	1.82e-4	1.22e-4
	(1.00)	(45.77)	(72.37)	(128.59)	(463.29)	(1024.14)
1024	1.99e-3	3.36e-4	2.63e-4	7.48e-5	6.60e-5	3.91e-5
	(1.00)	(35.19)	(57.40)	(711.32)	(911.47)	(2601.38)
4096	9.77e-4	1.34e-4	1.12e-4	2.50e-5	2.69e-5	1.04e-5
	(1.00)	(53.16)	(74.98)	(1525.33)	(1317.08)	(8908.90)
16384	5.39e-4	6.82e-5	5.22e-5	4.92e-6	1.18e-5	2.60e-6
	(1.00)	(62.46)	(106.86)	(11997.50)	(2078.36)	(43174.52)

where

$$C^* = \int_{I^7} \exp\left(1 - \left(\sin^2\left(\frac{\pi}{2}x_1\right) + \sin^2\left(\frac{\pi}{2}x_2\right) + \sin^2\left(\frac{\pi}{2}x_3\right)\right)\right) \mathrm{d}\boldsymbol{x}$$

$$= \mathrm{e}\left(\int_0^1 \exp\left(-\sin^2\left(\frac{\pi}{2}x\right)\right) \mathrm{d}x\right)^3,$$

which can be easily computed with high accuracy as a one-dimensional integral. Choose functions $A(\boldsymbol{x})$ and $B(\boldsymbol{x})$ as follows:

$$A(\boldsymbol{x}) = \frac{1}{C^*}\mathrm{e}^{-2}, \qquad B(\boldsymbol{x}) = \frac{1}{C^*}\mathrm{e}.$$

The computational results are given in Table 4.

The computational results show the following.

- QMC methods give much smaller errors than MC (with or without importance sampling) with the same sample size.
- For MC, smoothing has a little effect on errors (the smoothing method is always a little, but not significantly, better than the unsmoothed method).
- For QMC, importance sampling with the unsmoothed rejection may not improve upon standard QMC (the former may even worse than the later).
- For QMC, importance sampling combining with the extended smoothed rejection is always the best one in our experiments. It behaves much better than the standard QMC and than the unsmoothed rejection method.

The computational results seem quite encouraging. Smoothing can improve the performance of rejection sampling method, and make the importance sampling more efficient in QMC methods. There is strong reason to use extended smoothed rejection procedures in QMC integration.

Acknowledgements

The author would like to thank Kai-Tai Fang and Fred Hickernell and the referee for their valuable comments and suggestions.

References

[1] M. Berblinger, Ch. Schlier, T. Weiss, Monte Carlo integration with quasi-random numbers: experience with discontinuous integrands, Comput. Phys. Commun. 99 (1997) 151–162.
[2] R.E. Caflish, W. Morokoff, A. Owen, Valuation of Mortgage backed securities using Brownian bridges to reduce effective dimension, J. Comp. Finance 1 (1997) 27–46.
[3] L.P. Devroye, Non-Uniform Random Variate Generation, Springer, New York, 1986.
[4] K.T. Fang, Y. Wang, Number-Theoretic Methods in Statistics, Chapman & Hall, London, 1994.
[5] H. Faure, Discrépance de suites associées à un système de numération (en dimension s), Acta Arith. 41 (1982) 337–351.
[6] J.H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, Numer. Math. 2 (1960) 84–90.
[7] F.J. Hickernell, A generalized discrepancy and quadrature error bound, Math. Comput. 67 (1998) 299–322.
[8] F. James, J. Hoogland, R. Kleiss, Multidimensional sampling for simulation and integration: measures, discrepancies, and quasi-random numbers, Comput. Phys. Commun. 99 (1997) 180–220.
[9] A. Keller, Quasi-Monte Carlo Methods in Computer Graphics: the Global Illumination Problem, Lectures in Applied Mathematics, vol. 32, 1996, pp. 455–469.
[10] C. Lecot, Error bounds for quasi-Monte Carlo integration with nets, Math. Comp. 65 (1996) 179–187.
[11] W. Morokoff, R.E. Caflish, Quasi-Monte Carlo integration, J. Comp. Phys. 122 (1995) 218–230.
[12] B. Moskowitz, R.E. Caflish, Smoothness and dimension reduction in quasi-Monte Carlo methods, Math. Comput. Modelling 23 (1996) 37–54.
[13] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, SIAM, Philadelphia, 1992.
[14] S.H. Paskov, J.F. Traub, Faster valuation of financial derivatives, J. Portfolio Management 22 (1995) 113–120.
[15] I.H. Sloan, S. Joe, Lattice Methods for Multiple Integration, Oxford University Press, Oxford, 1994.
[16] I.M. Sobol', On the distribution of points in a cube and the approximate evaluation of integrals, Zh. Vychisli. Mat. Mat. Fiz. 7 (1967) 784–802 (in Russian).

[17] X.Q. Wang, Quasi-Monte Carlo methods for estimating a multivariate regression function, Chinese J. Appl. Probab. Statist. 14 (1998) 351–358.
[18] X.Q. Wang, F.J. Hickernell, Randomized Halton sequences, Working paper, Hong Kong Baptist University, 1999.