ISSN 1009-3095 *Journal of Zhejiang University SCIENCE* V.3, No.3, P.339-343, July-Aug., 2002
http://www.periodicals.com.cn; http://www.zju.edu.cn/English
http://www.zjupress.com; http://lib.zju.edu.cn/eindex.htm; jzu_s@mail.hz.zj.cn

B-splines smoothed rejection sampling method and its applications in quasi-Monte Carlo integration

LEI Gui-yuan(雷桂媛)
(*Department of Mathematics, Zhejiang University, Hangzhou 310028, China*)
E-mail: guiyuanlei@elong.com

Received Feb.2, 2001; revision accepted June 6, 2001

Abstract: The rejection sampling method is one of the most popular methods used in Monte Carlo methods. It turns out that the standard 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 proposed a B-splines smoothed rejection sampling method, which smoothed the characteristic function by B-splines smoothing technique without changing the integral quantity. Numerical experiments showed that the convergence rate of nearly $O(N^{-1})$ is regained by using the B-splines smoothed rejection method in importance sampling.

Keywords: Quasi-Monte Carlo, Monte Carlo, B-splines, Importance sampling, Numerical integration
Document code: A **CLC number:** O242.2 TB114.1

INTRODUCTION

Standard rejection sampling method is of importance in practical quasi-Monte Carlo methods such as decision process and sampling from a density function. However it is not as efficient as theoretically expected.

To estimate the integral

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

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

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

For Monte Carlo (MC) method, the points of P_N are independent identically distributed (i.i.d.) samples from the uniform distribution on I^s. For the quasi-Monte Carlo (QMC) method (Niederreiter, 1992), P_N is a low discrepancy point set and the error bound of the QMC estimate $E_N(f) = |I(f) - I_N|$ is the Koksma-Hlawka inequality

$$E_N(f) \leqslant D_N V(f) \qquad (3)$$

in which D_N is the discrepancy of the sequence and $V(f)$ is the variation of f in the Hardy-Krause sense. Rejection sampling method can be interpreted as the integration of a characteristic function. In general, characteristic functions have infnite variation, the exception being rectangles with sides parallel to the coordinate axes. So the Koksma-Hlawka inequality cannot be used to derive an upper bound and theoretical error bounds of size $O(N^{-1})$ are often not observed.

We smoothed the characteristic function by B-splines technique, and regained the error bounds of size $O(N^{-1})$. B-splines smoothed rejection sampling method is introduced in section 2. We also developed the importance sampling by using B-splines smoothed rejection sampling method in section 3. Numerical experiments will be given in section 4.

REJECTION METHOD

Standard Rejection Method, interpreted as integration of characteristic function

Let $p(\boldsymbol{x})$ be a probability density function defined on I^s. The algorithm of Standard Rejection can be described as follows:

1. Select $\gamma \geqslant \sup_{x \in I^s} p(\boldsymbol{x})$.

2. Repeat until N points have been accepted:

(a) Sample $(\boldsymbol{x}_t, y_t)$ from $U([0,1]^{s+1})$.

(b) If $y_t < \gamma^{-1} p(\boldsymbol{x}_t)$, accept trial point $\boldsymbol{x}_t$.

Otherwise, reject them.

By Bayes' formula, the density function of accepted points $p_{\text{accept}}(\boldsymbol{x})$ can be interpreted as a Monte Carlo evaluation of the following integral:

$$p_{\text{accept}}(\boldsymbol{x}) = \frac{\int_0^1 \chi(y < \gamma^{-1} p(\boldsymbol{x}))\mathrm{d}y}{\int_{I^s} [\int_0^1 \chi(y < \gamma^{-1} p(\boldsymbol{x}))\mathrm{d}y]\mathrm{d}\boldsymbol{x}} = \frac{p(\boldsymbol{x})/\gamma}{1/\gamma} = p(\boldsymbol{x}).$$

where $\chi(y < \gamma^{-1} p(\boldsymbol{x}))$ is characteristic function defined as:

$$\chi(y < \gamma^{-1} p(\boldsymbol{x})) = \begin{cases} 1, & \text{if } y < \gamma^{-1} p(\boldsymbol{x}), \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

So this algorithm produces an infinite sequence P of accepted points in s dimensions distributed according to $p(\boldsymbol{x})$.

But how well is the quality of the first N elements of the sequence P? We introduce the more general concept of discrepancy (Fang et al., 1994).

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

$$D_F(P_N) = \sup_{x \in I^s} |F_N(\boldsymbol{x}) - F(\boldsymbol{x})|. \quad (5)$$

The F-discrepancy is a measure for the quality of the representation of $F(\boldsymbol{x})$ by a point set P_N. As shown in the literature (Wang, 2000), The F-discrepancy is in fact the error of QMC integration of a characteristic function and the known theoretical bounds may only be $O(N^{-1/(s+1)})$, which is due to the discontinuity of the characteristic function. A smoothed rejection method was given (Wang, 2000; Moskowitz et al., 1996), which replaced the characteristic functions by continuous but non-differentiable functions. In section 2, we propose a B-splines smoothed rejection sampling method which introduced a differentiable weight function.

Smoothing characteristic function

We can smooth the discontinuous function by B-splines technique (Schumaker, 1981). It's well known that for any integrable function $f(x)$, $x \in R$, we call

$$f_h(x) = \frac{1}{h}\int_{x-\frac{h}{2}}^{x+\frac{h}{2}} f(t)\mathrm{d}t \quad (6)$$

its average function. Denote $D^{-1}f(x) = \int_a^x f(t)\mathrm{d}t$, we have

$$f_h(x) = h^{-1}\delta_h D^{-1} f(x)$$

where $\delta_h F(x) = F(x + \frac{h}{2}) - F(x - \frac{h}{2})$.

We apply the smoothing operator $h^{-1}\delta_h D^{-1}$ (h is smoothing width) to some simple and basic discontinuous functions. For example, if $f(x) = x_+$, then

$$f_h(x) = \frac{1}{2h}[(x + \frac{h}{2})_+^2 - (x - \frac{h}{2})_+^2] = \begin{cases} 0, & \text{if } x \leqslant -\frac{h}{2}, \\ (x + \frac{h}{2})^2/2h, & \text{if } -\frac{h}{2} < x \leqslant \frac{h}{2}, \\ x, & \text{if } x > \frac{h}{2}. \end{cases}$$

It is obvious that average function $f_h(x)$ is continuous function. When h is sufficiently small, $f_h(x)$ is the approximation function of $f(x)$. The difference between them is the function value over $[x - \frac{h}{2}, x + \frac{h}{2}]$.

Theorem 1 If $f_h(x)$ is the average function of $f(x)$ defined by Eq. (6), then

$$f_h(x) \approx f(x)$$

and $\lim_{h \to 0} f_h(x) = f(x)$.

We call the above described smoothing technique B-splines smoothing technique.

B-splines smoothed rejection sampling method, without changing the integral quantity

The characteristic function $\chi(y < \gamma^{-1} p$

$(\boldsymbol{x}))$ defined by Eq. (4) can also be smoothed by B-splines smoothing technique. We rewrite $\chi(y<\gamma^{-1}p(\boldsymbol{x}))$ as

$$W_0(\boldsymbol{x},y)=(\gamma^{-1}p(\boldsymbol{x})-y)_+^0, \quad 0\leqslant y\leqslant 1.$$

We apply the smoothing operator $(2h)^{-1}\delta_{2h}D^{-1}$ to $W_0(\boldsymbol{x},y)$, where smoothing width is $2h$.

$$W_\delta(\boldsymbol{x},y)=(2h)^{-1}\delta_{2h}D^{-1}W_0(\boldsymbol{x},y)=(2h)^{-1}[(\gamma^{-1}p(\boldsymbol{x})-y+h)_+-((\gamma^{-1}p(\boldsymbol{x})-y-h)_+]$$

Denote $f_1(y)=(\gamma^{-1}p(\boldsymbol{x})-y+h)_+$ and $f_2(y)=(\gamma^{-1}p(\boldsymbol{x})-y-h)_+$; applying the smoothing operator $h^{-1}\delta_hD^{-1}$ to $f_1(y)$ and $f_2(y)$ again, we get the differentiable weight function.

$$W_{\delta\delta}(\boldsymbol{x},\ y)=(2h)^{-1}[h^{-1}\delta_hD^{-1}f_1(y)-h^{-1}\delta_hD^{-1}f_2(y)]=\begin{cases}1, & \text{if } 0\leqslant y<\gamma^{-1}p(\boldsymbol{x})-\frac{3h}{2}\\ \dfrac{[(\gamma^{-1}p(\boldsymbol{x})-y+h)-\frac{(\gamma^{-1}p(\boldsymbol{x})-y-\frac{h}{2})^2}{2h}}{2h}, & \text{if } \gamma^{-1}p(\boldsymbol{x})-\frac{3h}{2}\leqslant y<\gamma^{-1}p(\boldsymbol{x})-\frac{h}{2},\\ \dfrac{(\gamma^{-1}p(\boldsymbol{x})-y+h)}{2h}, & \text{if } \gamma^{-1}p(\boldsymbol{x})-\frac{h}{2}\leqslant y<\gamma^{-1}p(\boldsymbol{x})+\frac{h}{2},\\ \dfrac{(\gamma^{-1}p(\boldsymbol{x})-y+\frac{3h}{2})^2}{4h^2}, & \text{if } \gamma^{-1}p(\boldsymbol{x})+\frac{h}{2}\leqslant y<\gamma^{-1}p(\boldsymbol{x})+\frac{3h}{2},\\ 0, & \text{if } \gamma^{-1}p(\boldsymbol{x})+\frac{3h}{2}\leqslant y\leqslant 1.\end{cases} \tag{7}$$

The function $W_{\delta\delta}(\boldsymbol{x},\ y)$ is our weight function used to replace characteristic function $\chi(y<\gamma^{-1}p(\boldsymbol{x}))$. The modified B-splines smoothed rejection sampling method is described as follows:

1. Select $\gamma\leqslant\sup_{x\in I^s}p(\boldsymbol{x})$.
2. Repeat until the sum of weight w_t is within one unit of N:
   (a) Sample $(\boldsymbol{x}_t,\ y_t)$ from $U([0,\ 1]^{s+1})$.
   (b) Set weight $w_t=W_{\delta\delta}(\boldsymbol{x}_t,\ y_t)$ to points $\boldsymbol{x}_t$, namely the accept probability of point $(\boldsymbol{x}_t,\ y_t)$ is w_t.

Theorem 2 $W_{\delta\delta}(\boldsymbol{x},\ y)$ is weight function defined by Eq. formula(7), $\chi(y<\gamma^{-1}p(\boldsymbol{x}))$ is characteristic function defined by Eq. (4). We have $\int_0^1 W_{\delta\delta}(\boldsymbol{x},y)\mathrm{d}y=\int_0^1\chi(y<\gamma^{-1}p(\boldsymbol{x}))\mathrm{d}y$. The sequence generated by B-splines smoothed rejection sampling method is distributed according to $p(\boldsymbol{x})$.

Proof Set $t=\gamma^{-1}p(\boldsymbol{x})-\frac{3h}{2}$, then

$$\int_0^1 W_{\delta\delta}(\boldsymbol{x},y)\mathrm{d}y=\int_0^t 1\mathrm{d}y+\int_t^{t+h}\frac{[(t-y+\frac{5h}{2})-\frac{(t-y+h)^2}{2h}]}{2h}\mathrm{d}y+\int_{t+h}^{t+2h}\frac{(t-y+\frac{5h}{2})}{2h}\mathrm{d}y+\int_{t+2h}^{t+3h}\frac{(t-y+3h)^2}{4h^2}\mathrm{d}y=t+\frac{11h}{12}+\frac{h}{2}+\frac{h}{12}=t+\frac{3h}{2}=\gamma^{-1}p(\boldsymbol{x})$$

So $\int_0^1 W_{\delta\delta}(\boldsymbol{x},y)\mathrm{d}y=\int_0^1\chi(y<\gamma^{-1}p(\boldsymbol{x}))\mathrm{d}y$. That is to say that the sequence generated by B-splines smoothed rejection sampling method is distributed according to $p(\boldsymbol{x})$.

We will apply the B-splines smoothed rejection sampling method to importance sampling in quasi-Monte Carlo integration in Section 3.

IMPORTANCE SAMPLING

Standard importance sampling

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

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

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 standard importance-sampled estimate is

$$I_N^{IS}=\frac{1}{N}\sum_{i=1}^{N}\frac{f(\boldsymbol{x}_i)}{p(\boldsymbol{x}_i)}, \qquad (8)$$

where $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ are i. i. d. samples from the density $p(\boldsymbol{x})$. Rejection sampling method is used robustly to sample points from $p(\boldsymbol{x})$. However, the improved performance for QMC method is not often observed, this degradation is due to the discontinuity of characteristic functions. We introduced the B-splines smoothed rejection sampling method into the importance sampling, and regained integration error of size $O(N^{-1})$.

Improving importance sampling with B-splines smoothed rejection sampling method

We rewrite the integral $I(f)$ as

$$\begin{aligned}I(f)&=\int_{I^s} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\\&=\gamma\int_{I^s}\frac{f(\boldsymbol{x})}{p(\boldsymbol{x})}\gamma^{-1}p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\\&=\gamma\int_{I^s}\frac{f(\boldsymbol{x})}{p(\boldsymbol{x})}\left[\int_0^1\chi(y<\gamma^{-1}p(\boldsymbol{x})\mathrm{d}y\right]\mathrm{d}\boldsymbol{x}\\&=\gamma\int_{I^s}\frac{f(\boldsymbol{x})}{p(\boldsymbol{x})}\left[\int_0^1 W_{\delta\delta}(\boldsymbol{x},y)\mathrm{d}y\right]\mathrm{d}\boldsymbol{x}\\&\approx\frac{\gamma}{N^*}\sum_{i=1}^{N^*}W_{\delta\delta}(\boldsymbol{x}_i,y_i)\frac{f(\boldsymbol{x}_i)}{p(\boldsymbol{x}_i)}\end{aligned}$$

The improved importance-sampled estimate of QMC integration is defined as

$$I_N^{(BIS)}=\frac{1}{N}\sum_{i=1}^{N^*}W_{\delta\delta}(\boldsymbol{x}_i,y_i)\frac{f(\boldsymbol{x}_i)}{p(\boldsymbol{x}_i)}, \qquad (9)$$

where $W_{\delta\delta}(\boldsymbol{x},y)$ is defined by Eq. (7) and N^* is chosen such that the sum of acceptance weights w_i is within one unit of N. It is obvious that

$$N\approx N^*/\gamma$$

Numerical experiments for improved importance sampling in MC and QMC integration will be given in the following section.

NUMERICAL EXPERIMENTS

In numerical experiments, the standard estimates, the standard rejection methods and B-splines smoothed rejection sampling methods for MC and QMC integration have been compared on several classical functions. The following estimates will be computed by Eqs. (2), (8) and (9):

Crude Monte Carlo:

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

Standard rejection method:

$$Y_N^{(2)}=(1/N)\sum_{i=1}^{N}f(\boldsymbol{x}_i)/p(\boldsymbol{x}_i),\ \boldsymbol{x}_i\sim p(\boldsymbol{x}_i),$$

accepted points;

B-spline smooth rejection:

$$Y_N^{(3)}=(1/N)\sum_{i=1}^{N^*}W_{\delta\delta}(\boldsymbol{x}_i,y_i)f(\boldsymbol{x}_i)/p(\boldsymbol{x}_i).$$

For a given N, take m samples of these estimates, denoted by $Y_N^{(j)}(k)$ for $1\leqslant k\leqslant m$ (using successive points from a single sequence). 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 the empirical standard deviation (sd error), defined as:

$$\hat{\sigma}_j=\frac{1}{(m-1)}\sum_{k=1}^{m}[Y_N^{(j)}(k)-\hat{I}^{(j)}]^2,\quad j=1,2,3. \qquad (10)$$

We use Halton sequences (Bratley et al., 1992) for quasirandom points and generate pseudorandom points using function ran2 (Press et al., 1992). Set $m=75$, that is to say that 75 runs for each estimate. Log-log plots are used so that slopes (which are presented parenthetically in the Figure keys) correspond to convergence rates.

One example is Monte Carlo and quasi-Monte Carlo integration over $I^7 = [0, 1]^7$ of the function

$$f(\boldsymbol{x}) = e^{1-(\sin^2(\frac{\pi}{2}x_1)+\sin^2(\frac{\pi}{2}x_2)+\sin^2(\frac{\pi}{2}x_3))}\arcsin \cdot (\sin(1) + \frac{x_1 + \cdots + x_7}{200})$$

Using the positive definite importance function:

$$p(\boldsymbol{x}) = \frac{1}{\eta}e^{1-(\sin^2(\frac{\pi}{2}x_1)+\sin^2(\frac{\pi}{2}x_2)+\sin^2(\frac{\pi}{2}x_3))}$$

Where $\eta = \int_{I^7} e^{1-(\sin^2(\frac{\pi}{2}x_1)+\sin^2(\frac{\pi}{2}x_2)+\sin^2(\frac{\pi}{2}x_3))}\mathrm{d}\boldsymbol{x} = e \cdot (\int_0^1 e^{-(\sin^2(\frac{\pi}{2}x)}\mathrm{d}x)^3$, which is easily approximated to high accuracy as a one-dimensional integral.

The resulting sd error using pseudorandom and quasirandom points are presented in Fig. 1 and Fig. 2.

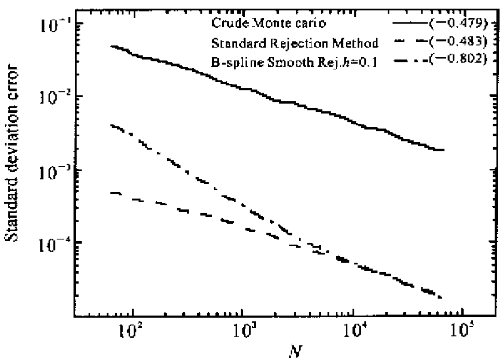

Fig. 1 The resulting standard deviation error using pseudorandom points

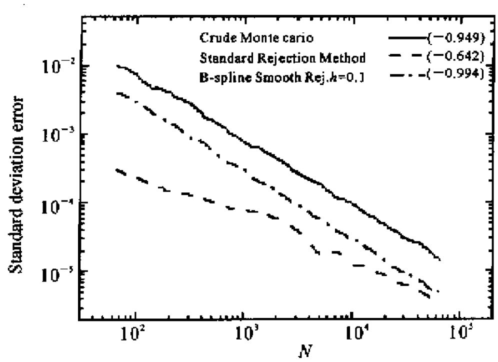

Fig. 2 The resulting standard deviation error using quasirandom points.

The computational examples show that:

1. QMC methods give much smaller errors than MC methods (with or without importance sampling) with the same sample size.

2. Both for QMC and MC, the importance sampling with B-splines smoothed rejection sampling is better than that with standard rejection sampling.

CONCLUSIONS

We can conclude that B-splines smoothed rejection sampling methods can improve the rejection method, and make the importance sampling more efficient in QMC methods. It can also be seen that use of modified differentiable weight functions in B-splines smoothed rejection sampling methods may improve MC methods. We expect that B-splines smoothing technique can be used more widely in MC and QMC methods, such as many methods involving decision process.

References

Bratley, P. , Fox, B. L. Niederreiter, H. , 1992. Implementation and tests of low-discrepancy sequences. *ACM Transactions on Modeling and Computer Simulation*. **2**(3): 195－213.

Fang, K. T. , Wang, Y. , 1994. Number-Theoretic Methods in Statistics. Chapman & Hall, London, p. 12－14.

Moskowitz, B. , Caflisch R. E. , 1996. Smoothness and dimension reduction in quasi-monte carlo methods. *Math. Comput. Modelling*. **23**(8－9): 37－54.

Niederreiter, H. , 1992. Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadel, p. 1－241.

Press, W. H. , Teukolsky, S. A. , Vetterling, W. T. , Flannery, B. P. , 1992. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press(2nd ed.), New York, p. 278－283.

Schumaker, L. , 1981. Spline Functions: Basic Theory. Wiley, New York, p. 118－134.

Wang Xiaoqun, 2000. Improving the rejection sampling method in quasi-Monte Carlo methods. *Journal of computational and applied Mathematics*. **114**: 231－246.