On the Warnock-Halton quasi-standard error

Art B. Owen
Stanford University

October 2005

Abstract

This paper investigates an error estimate proposed by Warnock and studied by Halton (2005). That error estimate is simply the sample standard error applied to certain non-randomized quasi-Monte Carlo points. This quasi-standard error (QSE) closely tracks the actual error in an example, and looks to be at least as accurate as a standard error based on random replication. We also show that the quasi-standard error is not unreasonably large in its intended use. But there are QMC constructions for which the QSE severely underestimates the true error. Moreover, discrepancy considerations do not separate these counter-examples from other cases where the method might be reliable. We conclude that the QSE is not yet ready to be trusted in applications.

1 Introduction

Quasi-Monte Carlo integration provides a deterministic alternative to Monte Carlo integration. When the integrand is square integrable then MC has a root mean squared error of $O(n^{-1/2})$ based on n function evaluations. Under the stronger condition that the integrand has finite variation in the sense of Hardy and Krause, the RMSE of QMC is $O(n^{-1+\epsilon})$ for any $\epsilon > 0$.

While QMC has a smaller error, estimating that error from data has been problematic. To this end, various schemes for randomizing QMC points have been proposed. See Owen (1999) and L'Ecuyer & Lemieux (2002) for surveys.

This paper investigates some unrandomized error estimates for quasi-Monte Carlo integration proposed in an unpublished technical report of Tony Warnock, widely discussed in the QMC community, and studied recently by Halton (2005). These methods amount to a deterministic replication treated as if it were a random replication.

The outline of this paper is as follows. Section 2 introduces the problem and our notation. Section 3 presents a numerical example based on the Richtmyer sequence in which the quasi-standard error performs well. Section 4 shows that the QSE can fail. Examples with Sobol' sequences and certain lattice rules give a QSE of zero for any f. Nobody would use the QSE with such sequences. The reason to mention them is that they have very good discrepancy properties. Any proof that the QSE is reliable in some setting, has then to contend with the fact that the QSE can fail, even on points with good discrepancy. The analysis will have to go deeper than discrepancy. Section 5 combines a result from Halton (2005) with the Koksma-Hlawka inequality, to give a sense in which the QSE is at least not unreasonably large. Section 6 has conclusions.

Halton (2005) makes the case for QMC and its error estimation without any recourse to randomness, replacing probability by quasi-probability. The paper is lengthy but contains a helpful synopsis in section 25 at the end. Part D of the synopsis refers to sections 22 and 24 of the article, covering a central limit theorem for QMC and bounds on what is here called the quasi-standard error. Halton's acknowledgements give Warnock much credit for the ideas. Unfortunately no published work of Warnock is available on the topic.

2 Notation

In quasi-Monte Carlo integration we estimate the integral

$$I = \int_{[0,1]^d} f(x)dx$$

by the average

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

for carefully chosen points $x_i \in [0,1]^d$.

As an example of such carefully chosen points consider $x_i = (x_{i1}, \dots, x_{id})$ where $x_{ij} = \{i\sqrt{p_j}\}$. Here the notation $\{z\}$ means $z - \lfloor z \rfloor$ where $\lfloor z \rfloor$ is the greatest integer less than or equal to z. These x_i are the Richtmyer points (Richtmyer 1951). Typically p_j is taken to be the j'th prime number.

A simple form of randomization, due to Cranley & Patterson (1976) can be used to get a data-based estimate of the quasi-Monte Carlo error. For

$r = 1, \dots, R \geq 2$ let U_r be independent random vectors with the uniform distribution on $[0,1]^d$. Then let $z_{ir} = \{x_i + U_r\}$ and put

$$\hat{I}_r = \frac{1}{n}\sum_{i=1}^{n} f(z_{ir}), \quad \hat{I} \ = \frac{1}{R}\sum_{r=1}^{R} \hat{I}_r, \quad \text{and,} \quad \widehat{V}(\hat{I}) = \frac{1}{R(R-1)}\sum_{r=1}^{R}(\hat{I}_r - \hat{I})^2. \tag{1}$$

Then $E(\hat{I}) = I$ and $\widehat{V}(\hat{I})$ is an unbiased estimate of the variance of $\hat{I}$. In practice one typically uses pseudo-random U_r instead of random ones.

Warnock (2001) has proposed a form of quasi-replication in which the points $z_{i1}, \dots, z_{iR}$ are taken from a dR dimensional quasi-Monte Carlo rule instead of from a randomization. That technical report is no longer available from Los Alamos. As of October 2005 it was possible to obtain a PDF file (Warnock 2002) of slides of a talk by Warnock that described the results. I have used the description in those slides. Let $p_1, \dots, p_{dR}$ be the first dR prime numbers. One version of Warnock's proposal takes

$$z_{ir} = \Bigl(\{i\sqrt{p_{(r-1)d+1}}\}, \dots, \{i\sqrt{p_{rd}}\}\Bigr) \tag{2}$$

for $i = 1, \dots, n$ and $r = 1, \dots, R$ and then defines $\hat{I}_r$, $\hat{I}$, and $\widehat{V}(\hat{I})$ using the same formulas (1) as used in randomization. In this case we call $\sqrt{\widehat{V}(\hat{I})}$ the quasi-standard error of $\hat{I}$ and wonder when it is a reliable guide to the size of the true error $|\hat{I} - I|$.

3 Empirical investigation

Warnock considers several examples including one with $d = 1$ and $f(x) = 3x^4 + \sin^2(6\pi x)$ for which of course $I = \int_0^1 f(x)dx = 1.1$.

Warnock (2002) finds a very close match between the quasi-standard error and the actual error. Specifically $4.604\sqrt{\widehat{V}(\hat{I})}$ and $|\hat{I} - I|$ track each other very closely on his plot over several orders of magnitude in n. The value 4.604 comes from the threshold for a 99% confidence interval based on a t test with 4 degrees of freedom.

I have been able to substantially reproduce the figure on page 18 of Warnock (2002), in the middle panel of Figure 1. For my results, I found that the QSE tracked the actual error much more closely than did 4.6 times the QSE. Multiplying in a factor of 4.6 raises the QSE noticeably above the actual errors, so I have not done so. The figure in Warnock (2002) used a

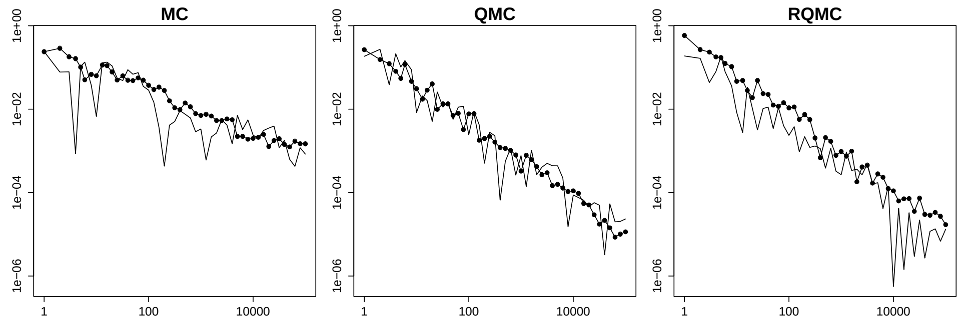

Figure 1: Each panel plots the quasi-standard error and the actual error $|\hat{I} - I|$ versus sample size n running from 1 to 10^5. There are 10 points per power of 10 in n. The standard error from 5 replicates is plotted with solid dots. The actual error is plotted with a connected line that is more jagged than the one for the QSE. From left to right the points were: simple Monte Carlo, Richtmyer points, and randomized QMC points.

much larger range for n, and this would have diminished the visual effect of a factor of 4.6. The leftmost panel in Figure 1 shows comparable results for MC points. The rightmost panel shows comparable results for randomized QMC points: the Richtmyer points are given a Cranley-Patterson rotation. It appears that the QSE works at least as well, possibly better than, sampling based SEs do in this example.

When preparing plots like Figure 1 it is important to use only a sparse subsequence of n values. Otherwise the rare sample sizes n with error near zero will make the plot very hard to read. Such points commonly arise. The error may well change sign at some point as n increases and so it is not surprising that it might pass near zero.

4 Reliability

For the quasi-standard error to be reliable, it should never be too much smaller than the actual error. We show here that certain QMC constructions make the quasi-standard error zero, for any f.

For $i = 1, \dots, n$, let $z_i \in [0, 1]^{dR}$. The point z_i is made by concatenating the vectors $z_{i1}, \dots, z_{iR} \in [0, 1]^d$. In applications of the Warnock-Halton quasi-standard error, the points z_i are chosen to be of low discrepancy with respect to the $U[0, 1]^{dR}$ distribution.

Halton (2005) points out that a consequence of the $U[0, 1]^{dR}$ distribution for z_i is statistical independence of the components $z_{i1}, \ldots, z_{iR}$. So if the z_i have low discrepancy with respect to $U[0,1]^{dR}$ then the components $z_{i1}, \ldots, z_{iR}$ and hence also the function values $f(z_{i1}), \ldots, f(z_{iR})$ are nearly independent of each other. For deterministic z_i, "nearly independent" means that the distribution of $(f(z_{j1}), \ldots, f(z_{jR}))$ where the random j is uniformly distributed in $\{1, \ldots, n\}$ is close to a product distribution on $\mathbb{R}^R$.

Even if f_{i1} is independent of f_{i2} we may not have $\hat{I}_1$ independent of $\hat{I}_2$. The reason is that f_{i1} may be dependent on f_{j2} for some $j \neq i$. There are $O(n^2)$ such pairs to consider and so even a small dependence among them may overwhelm the near independence of pairs with a common index i.

To take an extreme illustration, consider $d = 1$ and $n = 2^m$ and let z_i be the i'th point of the R dimensional Sobol' sequence. Then, as is well known, for each $r = 1, \ldots, R$, the n values z_{ir} for $i = 1, \ldots, n$ are a permutation of the van der Corput sequence. It follows that for this case, all of the $\hat{I}_r$ are the same and the quasi-standard error is 0 (for any f). A similar phenomenon happens when the z_i are taken from a lattice rule for which each component takes the values $0, 1/n, 2/n, \ldots, (n-1)/n$. A QSE of zero would surely raise an alarm in practice. But in applications with $d > 1$ and a nearly additive integrand, a more troublesome problem arises. Then the QSE could be far too small, but not so obviously small as to raise the alarm.

Sobol' points and lattice rules form a partial counter-example to the quasi-standard error. They show that neither low dR dimensional discrepancy of the $z_1, \ldots, z_n$, nor near independence of $(f(z_{j1}), \ldots, f(z_{jR}))$ suffice to make the quasi-standard error reliable. Similar difficulties arise for $d > 1$ when Sobol' or lattice points are used and f is additive or nearly so in it's d inputs.

To determine when the QSE does not underestimate the error, the joint behavior of points x_{ir} and x_{js} for $i \neq j$ has to be taken into account. Notice that the discrepancy of $z_1, \ldots, z_n$ can change completely if $z_{1r}, \ldots, z_{nr}$ are replaced by a permutation $z_{\pi_r(1)r}, \ldots, z_{\pi_r(n)r}$ but that neither $\hat{I}$ nor $\widehat{V}(\hat{I})$ changes under such permutation.

If the method really does work for the Richtmyer or any other points, then it must be due to deeper properties than discrepancy of z_i. Also, simply changing from one kind of d-dimensional discrepancy to another would not be enough to identify when the method works. When using the QSE, we need to avoid constructions in which the points of one replication can, after a permutation, be brought very close to the points of another replication.

5 Conservatism

There is a sense in which the QSE is at least not too large, in its intended use. We can of course make the QSE far too large by taking nR points with low discrepancy in $[0,1]^d$ and splitting them into R groups of n points where each group has high discrepancy. For example the points in group r might all have their first component in the interval $[(r-1)/R, r/R)$. But such a decomposition is unfair because it is not the way that the QSE is intended to be used. Instead we consider the case where each of the R groups has low discrepancy.

Let $z_{r1},\dots,z_{rn}$ have d-dimensional star discrepancy $D^*_{r,n}$. Next let $\overline{D}^*_n = \max_{1\le r\le R} D^*_{r,n}$, and suppose that $\overline{D}^*_n \le C_d(\log n)^d/n$ for some $C_d<\infty$. The power of $\log(n)$ could also be taken to be $d-1$ for some QMC constructions, but the power d is appropriate for extensible sequences.

Halton (2005) showed that

$$\sqrt{\frac{1}{R}\sum_{r=1}^R(\hat I_r-\hat I)^2} \le \frac{1}{2}\Bigl[\max_r \hat I_r - \min_r \hat I_r\Bigr] \le \max_r |\hat I_r - I|,$$

from which we find, by the Koksma-Hlawka inequality, that

$$\sqrt{\widehat V(\hat I)} \le \frac{1}{\sqrt{R-1}}\max_r|\hat I_r - I| \le \frac{\overline{D}^*_n V_{\mathrm{HK}}(f)}{\sqrt{R-1}} \le \frac{C_d(\log n)^d V_{\mathrm{HK}}(f)}{n\sqrt{R-1}}, \qquad (3)$$

where $V_{\mathrm{HK}}(f)$ is the total variation of f in the sense of Hardy and Krause. The Koksma-Hlawka bound for the pooled estimator is

$$|\hat I - I| \le D^*_n(z_{11}, z_{21},\dots,z_{n1},\dots,z_{nR})V_{\mathrm{HK}}(f).$$

No known extensible QMC construction in d dimensions had discrepancy $o(\log(n)^d/n)$. So the Koksma-Hlawka bound for $\hat I - I$ on any QMC points used in practice is at least

$$c_d V_{\mathrm{HK}}(f)(\log(nR))^d/(nR)$$

for some $c_d>0$. Accordingly $\widehat V(\hat I)$ does not overestimate the Koksma-Hlawka bound on $|\hat I - I|$ by a factor of more than

$$\frac{C_d}{c_d}\frac{R}{\sqrt{R-1}}\Bigl(\frac{\log n}{\log nR}\Bigr)^d.$$

It is possible that a lucky cancellation among the individual errors $\hat I_r - I$ could give rise to an error $\hat I - I$ of zero, and then any positive quasi-standard error might seem too big. But this can also happen with randomization based standard errors, so it is not particularly a disadvantage of the QSE.

6 Conclusions

It is very interesting that the QSE can perform so well in numerical examples, and this performance remains mysterious. The quasi-standard error can grossly underestimate the actual error for some QMC sequences. Whether it does so depends on properties of the sample points not captured by the discrepancy of the point sequence $z_1, \dots, z_n$. Until an analysis has been carried out showing that the QSE for the Richtmyer or other QMC construction does not severely underestimate error, it is inadvisable to use the QSE. Genuine randomizations are preferable. They have theoretical support, and there exist well tested pseudo-random number generators to perform them.

The possibility of getting better than random error estimates for better than random quadrature is intriguing and seems to merit more study. We conclude by mentioning some results in this direction. Better than random error estimates would seem to require that the points from one replicate tend to avoid the points from another replicate. This is what we expect when the replicates are drawn from non-overlapping consecutive indices of a single d dimensional QMC point set. Berblinger & Schlier (1991) proposed splitting the points of the Halton sequence into a number of consecutive runs and treating them as replicates. Owen (1997) suggests splitting (randomized) digital nets into internal replicates, each with the digital net property. Both of these methods tend to be conservative because while within-group error decreases as in QMC, the between-group error is taken to decrease at only the MC rate. Snyder (2000) estimates and corrects for this conservativism by varying the size of the replicated group. The technique uses an assumption that the error decay as a power of n over the range studied. Snyder obtains good empirical results on the test functions of Genz (1984).

Acknowledgments

I thank Tony Warnock and John Halton for discussions at the MCM 2005 conference. I thank Christoph Schlier and Christiane Lemieux for helpful comments. This work was supported by the U.S. NSF under grant DMS–0306612.

References

Berblinger, M. & Schlier, Ch. (1991), 'Monte Carlo intergration with quasi-random numbers: some experience', *Computer Physics Communications* **66**, 157–166.

Cranley, R. & Patterson, T. (1976), 'Randomization of number theoretic methods for multiple integration', *SIAM Journal of Numerical Analysis* **13**, 904–914.

Genz, A. (1984), Testing multidimensional integration routines, *in* 'Tools, Methods and Languages for Scientific and Engineering Computation', North-Holland, Amsterdam, pp. 81–94.

Halton, J. (2005), 'Quasi-probability', *Monte Carlo methods and applications* **11**(3), 203–350.

L'Ecuyer, P. & Lemieux, C. (2002), A survey of randomized quasi-Monte Carlo methods, *in* M. Dror, P. L'Ecuyer & F. Szidarovszki, eds, 'Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications', Kluwer Academic Publishers, pp. 419–474.

Owen, A. B. (1997), 'Monte Carlo variance of scrambled equidistribution quadrature', *SIAM Journal of Numerical Analysis* **34**(5), 1884–1910.

Owen, A. B. (1999), Monte Carlo, quasi-Monte Carlo, and randomized quasi-Monte Carlo, *in* H. Niederreiter & J. Spanier, eds, 'Monte Carlo and quasi-Monte Carlo Methods 1998', Springer-Verlag, Berlin, pp. 86–97.

Richtmyer, R. (1951), The evaluation of definite integrals and quasi-Monte Carlo method based on the properties of algebraic numbers, Technical report, Los Alamos Scientific Laboratory, Los Alamos, NM.

Snyder, W. C. (2000), 'Accuracy estimation for quasi-Monte Carlo simulations', *Mathematics and Computers in Simulation* **54**, 131–143.

Warnock, T. (2001), Effective error estimates for quasi-Monte Carlo computations, Technical Report LA-UR-01-1950, Los Alamos National Labs.

Warnock, T. (2002), 'Effective error estimates for quasi-Monte Carlo computations'. http://lib-www.lanl.gov/la-pubs/00367143.pdf.