Quasi-Monte Carlo, Discrepancies and Error Estimates[1]

Jiri Hoogland[2]

NIKHEF, Amsterdam, The Netherlands

Fred James[3]

CERN, Geneva, Switzerland

Ronald Kleiss[4]

University of Nijmegen, Nijmegen, The Netherlands

Abstract

We discuss the problem of defining an estimate for the error in Quasi-Monte Carlo integration. The key issue is the definition of an ensemble of quasi-random point sets that, on the one hand, includes a sufficiency of equivalent point sets, and on the other hand uses information on the degree of uniformity of the point set actually used, in the form of a discrepancy or diaphony. A few examples of such discrepancies are given. We derive the distribution of our error estimate in the limit of large number of points. In many cases, Gaussian central limits are obtained. We also present numerical results for the quadratic star-discrepancy for a number of quasi-random sequences.

1 The error problem

We discuss the problem of integration of a square integrable function over the s-dimensional unit hypercube $K = [0, 1)^s$, using a set of N points x_k,

$k = 1, 2, \ldots, N$. The actual integral is $J = \int_K dx \, f(x)$, and its numerical estimate is given by

$$S = \frac{1}{N} \sum_{k=1}^{N} f(x_k) \ . \tag{1}$$

Depending on the way in which the points x_k are chosen, we distinguish different integration methods: if the points come from some predetermined, deterministic scheme we have a quadrature rule, if they are considered to be iid uniform random numbers, we have Monte Carlo. An intermediate case is that of Quasi-Monte Carlo, where the points are considered to be part of a low-discrepancy sequence, but share the ergodic properties of a truly random sequence[5]. The integration error is defined as $\eta \equiv S - J$. Good integration methods are characterized by the fact that they typically lead to a small value of η, but, more importantly, allow one to obtain a good estimate of η. In the case of Monte Carlo, η is a stochastic variable, and hence has a probability density $P(\eta)$. For quasi-random point sets used in Quasi-Monte Carlo, we may (as we shall specify more precisely later on) also interpret η as having such a probability density: its form is the subject of this contribution.

In true Monte Carlo, the error distribution $P(\eta)$ is obtained by viewing the point set $\{x_1, x_2, \ldots, x_N\}$ as a typical member of an ensemble of random point sets, governed by the obvious Cartesian combined probability density

$$P_N(x_1, x_2, \ldots, x_N) = 1 \ , \tag{2}$$

so that the x_k are indeed iid uniform random numbers. The error η is then a random variable over this ensemble, with the following well-known results. In the first place, S is an unbiased estimator of J in the sense that $\langle \eta \rangle = 0$, where the average is over the ensemble of points sets. In the second place, for large N, $P(N)$ approaches a normal distribution according to the Central Limit Theorem. Finally, the variance of this distribution is given by $\langle \eta^2 \rangle = \mathrm{Var}(f)/N$, where Var denotes the variance. Note that, since we average over the integration points, the error distribution can depend only on the integrand itself.

The conceptual problem in the use of quasi-random rather than truly random point sets is the following: a quasi-random point set is *not* a 'typical' set of random points! Indeed, quasi-random point sets are very special, with carefully built-in correlations between the various points so that each new point tends to 'fill a gap' left by the previous ones. The usual Monte Carlo error estimate is therefore not really justified in Quasi-Monte Carlo. On the other hand, many different error bounds assure us that small errors are possible, and indeed likely when we apply low-discrepancy point sets. In the following, we shall discuss two approaches to a solution of this conundrum. Obviously, we can only summarize the main results here: technical details and pictures can be found in the references.

---

[5]We shall not discuss the case of point sets with fixed, predetermined N.

2 The Bayesian approach

The first way around the aforementioned conceptual problem is to interchange the rôles of integrand and point set: we view the integrand $f(x)$ as a typical member of some underlying class of functions and average over this class, so that the error depends only on a property of the point set. In practice, the choice of function class often entails a good deal of idealism or pot luck, as usual in a Bayesian approach to probability. We discuss several examples, in which we denote by $\langle\rangle_f$ an average over the probability measure governing the function class.

2.1 The Woźniakowski Lemma

Let the integrand $f(x)$ be chosen according to the Wiener (sheet) measure in s dimensions. This measure is Gaussian, with

$$\langle f(x)\rangle_f = 0 \quad , \quad \langle f(x), f(y)\rangle_f = \prod_{\mu=1}^{s} \min(x^\mu, y^\mu) \quad , \tag{3}$$

where the index μ labels the coordinates. We may then quote the Lemma from [1]:

$$\langle\eta\rangle_f = 0 \quad , \quad \left\langle\eta^2\right\rangle_f = D_2(x_1^*, x_2^*, \ldots, x_M^*) \quad , \tag{4}$$

where D_2 stands for the L_2 norm of the well-known star-discrepancy, and the x_k^* denotes the 'reflected' point, with $(x_k^*)^\mu = 1 - x^\mu$. In [2] it is shown, moreover, that the distribution $P(\eta)$ in this case is a Gaussian.

We have here the interesting general fact that the choice of a particular function class induces its own particular discrepancy. On the other hand, in many cases (such as in particle physics phenomenology) the Wiener measure is certainly not appropriate since it is dominated by integrands that are not locally smooth. In [3], *folded* Wiener sheet measures are studied with analogous results, but then again these describe functions that are much too smooth.

2.2 Induced discrepancies

In [2], we established the following general result. Let the measure on the function class be such that

$$\langle f(x_1)\rangle_f = 0 \quad , \quad \langle f(x_1)f(x_2)\rangle_f = \int dy\ h(x_1, y)h(x_2, y) \tag{5}$$

for all $x_{1,2}$ in K, for some $h(x, y)$. There is then an induced quadratic discrepancy, defined as follows:

$$\left\langle\eta^2\right\rangle_f = \int dy\ g(y)^2 \quad , \quad g(y) = \frac{1}{N}\sum_{k=1}^{N} h(x_k, y) - \int_K dx\ h(x, y) \quad . \tag{6}$$

3

Note that h is not necessarily in the same function class as the f, and indeed y may be defined in a space quite different from K. Note that, whenever the function class measure is Gaussian, then $P(\eta)$ will also be Gaussian. Generalizations to higher moments can be found in [2, 4].

2.3 Orthonormal function bases

As an example in $s = 1$, let $u_n(x)$ be an orthonormal set of functions on K, as follows:

$$u_0(x) = 1 \quad , \quad \int_K dx \, u_m(x) u_n(x) = \delta_{m,n} \quad , \tag{7}$$

with $m, n = 0, 1, 2, \ldots$. Let $f(x)$ admit of a decomposition

$$f(x) = \sum_{n \geq 0} v_n u_n(x) \quad , \tag{8}$$

and choose the measure such that the v_n are normally distributed around zero, with variance σ_n^2. The induced quadratic discrepancy D_2^{orth} is then defined[6] as

$$\left\langle \eta^2 \right\rangle_f = \frac{1}{N} D_2^{orth} \quad , \quad D_2^{orth} = \frac{1}{N} \sum_{n > 0} \sigma_n^2 \sum_{k,l=1}^{N} u_n(x_k) u_n(x_l) \quad . \tag{9}$$

A special case is that of the Fourier class, which has

$$u_{2n} = \sqrt{2} \cos(2\pi nx) \quad , \quad u_{2n-1} = \sqrt{2} \sin(2\pi nx) \quad , \quad n = 1, 2, 3, \ldots \tag{10}$$

The physically reasonable requirement that the *phase* of each mode n (made up from u_{2n-1} and u_{2n}) is uniformly distributed forces us to have the Gaussian measure, with in addition $\sigma_{2n-1} = \sigma_{2n}$, which property corresponds to translational invariance. We then have

$$D_2^{orth} = \frac{2}{N} \sum_{n>0} \sigma_{2n}^2 \left| \sum_{k=1}^{N} \exp(2i\pi nx_k) \right|^2 \quad . \tag{11}$$

Obviously, other orthonormal function bases are also possible, such as the system of Walsh functions; a further discussion, including the straightforward generalization to higher dimension, can be found in [4, 5, 6, 7]. Note that all such quadratic discrepancies are nonnegative by construction.

3 The discrepancy-based approach

Another way of establishing integration error estimates, which in our opinion does more justice to the spirit of Monte Carlo, is the following. Instead of considering all point sets of N truly random points, with the Cartesian probability density (2), we restrict ourselves to those point sets that have a given value of discrepancy, for some predefined type of discrepancy. In this way, information on the discrepancy performance of one's favorite quasi-random number sequence can be incorporated in the error estimate.

---

[6]This measure of discrepancy is also called *diaphony.*

3.1 Non-Cartesian distribution of points

We have then, instead of (2), a combined probability density P_N for the N points as follows. Let $D_N(x_1, \ldots, x_N)$ be *some* discrepancy defined on sets of N points in K, and suppose its value for the actual point set that is used in the integration be w. Then,

$$
\begin{aligned}
H(w) &= \int_K dx_1 \cdots dx_N \, \delta\left(D_N(x_1, \ldots, x_N) - w\right) \; , \\
P_N(w; x_1, \ldots, x_N) &= \frac{1}{H(w)} \delta\left(D_N(x_1, \ldots, x_N) - w\right) \\
&= 1 - \frac{1}{N} F_N(w; x_1, \ldots, x_N) \; , \quad (12)
\end{aligned}
$$

so that F_N measures the deviation from Cartesian (iid) uniformity. The quantity $H(w)$ is of course just the probability density for the discrepancy over the iid uniform random numbers, an object interesting in its own right. Let us also define marginal deviations as

$$
F_k(w; x_1, \ldots, x_k) = \int_K dx_{k+1} \cdots dx_N F_N(w; x_1, \ldots, x_N) \; . \quad (13)
$$

We can then simply establish, for instance, that, provided $F_1(w; x)$ vanishes for all x,

$$
\begin{aligned}
\left\langle \eta^2 \right\rangle &= \frac{1}{N}\left[\text{Var}(f) - \frac{N-1}{N} \int_K dx\,dy \; f(x) f(y) F_2(w; x, y)\right] \; , \\
\langle \eta \rangle &= 0 \; , \quad (14)
\end{aligned}
$$

where $\langle \rangle$ now denotes averaging with respect to $P_N(w; .)$. It is seen that we may expect a reduced error if F_2 is positive when x and y are 'close' together in some sense, *i.e.* if the points in the point set 'repel' each other. Note that only a small, $\mathcal{O}(1/N)$, deviation from uniformity is sufficient.

3.2 Error probability distribution

In many cases, it is actually possible to compute the F_2 mentioned above. In fact, especially in the case of discrepancies defined using orthonormal function bases, we can do much more. Using a Feynman-diagram technique described in detail in [6, 8], we can establish results for $P(\eta)$ as an asymptotic expansion in $1/N$. To leading order, we have

$$
\begin{aligned}
P(\eta) &= \frac{\sqrt{N/2\pi}}{2\pi i H(w)} \int_{-i\infty}^{+i\infty} dz \frac{1}{\sqrt{B(z)}} \exp\left[A(z) - \frac{\eta^2 N}{2B(z)}\right] \; , \\
A(z) &= -wz - \frac{1}{2} \sum_{n>0} \log(1 - 2z\sigma_n^2) \; , \\
B(z) &= \sum_{n>0} \frac{v_n^2}{1 - 2z\sigma_n^2} \; , \quad (15)
\end{aligned}
$$

where the z integral runs to the left of any singularities. This result holds, for N asymptotically large, for any discrepancy measure based on orthonormal functions as discussed above, and, moreover, for any reasonable f, even if it is not in the function class based on these orthonormal functions. The $1/N$ corrections are fully calculable, although we have not done so yet. Two corollaries follow immediately. In the first place,

$$\int\limits_0^\infty dw\, H(w)P(\eta) = \sqrt{N/2\pi V}\, e^{-\eta^2 N/2V} \quad , \quad V \equiv \sum_{n>0} v_n^2 = \mathrm{Var}(f) \quad , \qquad (16)$$

which recovers the Central Limit Theorem valid over the whole ensemble of N-point point sets with any w. In the second place, we obtain an integral representation for $H(w)$ by insisiting that $P(\eta)$ be normalized to unity:

$$H(w) = \frac{1}{2\pi i} \int\limits_{-i\infty}^{+i\infty} dz\, \exp\left[-zw - \frac{1}{2}\sum_{n>0} \log(1 - 2z\sigma_n^2)\right] \quad . \qquad (17)$$

Generalizations of these results only affect the sums over n.

3.3 Application 1: equal strengths

A simple model for a discrepancy is obtained by taking $\sigma_n = 1/2M$ for $n = 1, 2, \ldots, 2M$, and zero otherwise (with trivial extension to more dimensions). Let us then decompose the variance of the integrand as follows:

$$V = \mathrm{Var}(f) = \sum_{n>0} v_n^2 = V_1 + V_2 \quad , \quad V_1 = \sum_{n=1}^{2M} v_n^2 \quad , \quad V_2 = \sum_{n>2M} v_n^2 \quad , \qquad (18)$$

so that V_1 contains that part of the variance to which the discrepancy is sensitive (the 'covered' part) and V_2 the rest (the 'uncovered' part). We then have

$$\begin{aligned}
H(w) &= \frac{M^M}{\Gamma(M)} w^{M-1} e^{-Mw} \quad \sim \quad \exp\left(-\frac{M}{2}(w-1)^2\right) \quad , \\
P(\eta) &\sim \left(\frac{N}{2\pi(wV_1 + V_2)}\right)^{1/2} \exp\left(-\frac{\eta^2 N}{2(wV_1 + v_2)}\right) \quad ,
\end{aligned} \qquad (19)$$

where the approximations are valid for large M. We see that a new central limit theorem holds, where the variance of f has been modified so that its covered part is reduced by a factor w, according to intuition.

3.4 Application 2: harmonic model in one dimension

Let us concentrate on the case $s = 1$, and take $\sigma_{2n-1} = \sigma_{2n} = 1/n$, so that f is, on the average, square integrable, but its derivative is not. In that case we have,

$$H(w) = \sum_{m>0} (-1)^{m-1} m^2 e^{-wm^2/2} \quad , \qquad (20)$$

6

which is, apart from a trivial scaling, precisely the probability density of the Kolmogorov-Smirnov statistic. This is somewhat surprising since that statistic is based on the L_∞ norm of the standard star-discrepancy, a totally different object. In addition, we conjecture, that for values of w small compared to its expectation value $\langle w \rangle = \pi^2/3$, we shall have

$$P(\eta) \propto \exp\left[-\frac{\eta^2 N}{2C}\right] \quad , \quad C = \text{Var}(f)\frac{w}{\langle w \rangle} \quad ; \tag{21}$$

this would again indicate a reduction of the error estimate for small w. To date, we have not yet been able to prove this assertion.

3.5 The quadratic star-discrepancy

We have also obtained some results for the quadratic form of the standard star-discrepanc [4, 5]. Although this is not based on orthonormal functions and the analysis is hence more complicated, we have obtained the moment-generating function $G(z) = \langle e^{zw} \rangle$ for asymptotically large N, where w now stands for N times the quadratic star-discrepancy. More precisely, we have

$$
\begin{aligned}
G(z) &= \exp(\psi(z))/\sqrt{\chi(z)} \ , \\
\psi(z) &= -\frac{1}{2}\sum_{n>0} Q_s(2n-1)\log(1-2z_n) \ , \\
\chi(z) &= \frac{2^s}{2z}\sum_{n>0} Q_s(2n-1)\frac{z_n}{1-z_n} \ , \\
z_n &= (4/\pi^2)^s\frac{2z}{(2n-1)^2} \ .
\end{aligned}
\tag{22}
$$

Here, $Q_s(m)$ is the number of ways in which an odd positive integer m can be written as a product of s positive integers, including 1's. The function $H(w)$ can now be computed numerically for different s values. We have done so, and find that $H(w)$ very slowly approaches a Gaussian distribution as s increases. Indeed, the skewness of $H(w)$ is, for large s, approximately given by $(216/225)^{s/2}$ so that the approach to normality is indeed slow.

4 Numerical results for the quadratic star-discrepancy

Since we now have $H(w)$ for the quadratic star-discrepancy, we can reliably judge how well quasi-random number generators perform, since we can compare the discrepancy of their output with the behaviour of truly random points. Space does not permit us to show pictures, which can be found in [4, 5]. Here, we just describe the results. We have computed the quadratic star-discrepancy for a good pseudorandom generator (RANLUX), and for the Richtmeyer, Halton, Sobol', and Niederreiter sequences. We did this both

exactly, and by Monte Carlo. The latter method is actually faster for N larger than about 50,000 if we ask for 5% accuracy. We made runs of up to $N = 150,000$, and considered the lowest and highest discrepancy value in subsequent intervals of 1000. These we compared with the expected value $(2^{-s} - 3^{-s})/N$ for truly random points, and also plotted the standardized[7] form $\xi(w) = (w - \langle w \rangle)/\sqrt{\mathrm{Var}(w)}$. We considered dimensions from 1 up to 20. In all dimensions, RANLUX appears to mimic a truly random sequence quite well. The quasi-random generators perform very well in low dimensions, and generally the discrepancy falls further and further below than of a random sequence as N increases. There are exceptions, however: for instance, the Sobol' sequence for $s = 11$ degrades, and is actually worse than random for $N \sim 60,000$, again rapidly improving for larger N. Apart from taking this as a warning, we have not investigated the reason for such behaviour in detail. The biggest surprise was when we plotted the variable $\xi(w)$ (for instance, for $N = 150,000$) as a function of s. It appears that, as measured in this way, the performance of all quasi-random generators *improves* with increasing s! For s larger than 15 or so, all generators become rather bad, which is of course due to the fact that the onset of the asymptotic regime occurs for larger N as s increases. But what is more striking is the fact that the old-fashioned and simple Richtmyer generator performs as well as the modern, sophisticated Sobol' and Niederreiter sequences. We take this as an indication that the Richtmyer generator deserves more study, in particular since we have not attempted any optimization of its 'lattice point'.

---

[7]Again, here the $\langle \rangle$ stands for the expectation for truly random points.

References

[1] H. Woźniakoski, *Bull. AMS* **24**(1991)185.

[2] R. Kleiss, *Comp. Phys. Comm.* **71**(1992)39.

[3] S. Paskov, *J. Complexity* **9**(1993)291.

[4] J.K. Hoogland, Ph.D. thesis, University of Amsterdam, 1996.

[5] F. James, J. Hoogand and R. Kleiss, *Multidimensional sampling for simulation and integration: measures, discrepancies and quasi-random numbers*, to appear in *Comp. Phys. Comm.*

[6] J. Hoogland and R. Kleiss, *Discrepancy-based error estimates for Quasi-Monte Carlo. I: General formalism, Comp. Phys. Comm.* **98** (1996) 111.

[7] J. Hoogland and R. Kleiss, *Discrepancy-based error estimates for Quasi-Monte Carlo. II: Results for one dimension, Comp. Phys. Comm.* **98**(1996)128.

[8] J. Hoogland and R. Kleiss, *Discrepancy-based error estimates for Quasi-Monte Carlo. III: Error distributions and central limits*, in preparation.