ERROR ESTIMATES IN MONTE CARLO AND QUASI-MONTE CARLO INTEGRATION* **

Achilleas Lazopoulos

Theoretical Physics Department, Radbout University of Nijmegen
6500 HC Nijmegen, The Netherlands
e-mail: lazopoul@sci.kun.nl

(Received October 11, 2004)

While the Monte Carlo approach to integration dominates any numerical calculation in particle physics phenomenology, the Quasi-Monte Carlo method, which promises better performance, is restrained to relatively minor applications. One of the reasons is the difficulty in estimating reliably the error when using Quasi-Monte Carlo point sequences. The classical Monte Carlo estimator, that consistently overestimates the error, has been used up to now. We review the situation on the error estimators for classical Monte Carlo and present a new estimator for Quasi-Monte Carlo.

PACS numbers: 02.70.Uu

1. Monte Carlo and Quasi-Monte Carlo

1.1. Introduction

In numerical integration, the main problem is not to obtain a numerical answer for the integral, but rather, on the one hand, to ensure that the inherent numerical error is as small as possible, and, on the other hand, to estimate this error as precisely as possible.

In this paper, we shall be concerned with the integration errors arising in Monte Carlo and Quasi-Monte Carlo integration. In these methods, the integration nodes are distributed in a (more or less) stochastic manner, and the integration errors are, therefore, of an essentially probabilistic nature. The difference between Monte Carlo and Quasi-Monte Carlo is that in the

* Presented at the final meeting of the European Network "Physics at Colliders", Montpellier, France, September 26–27, 2004.

** Research supported by the E.U. contract no. HPMD-CT-2001-00105.

former, the integration points are iid[1] uniform in the integration region[2], while in the latter the integration points are not chosen independently, but rather with an explicit interdependence so that their overall distribution is "smoother", in a sense discussed below.

In stochastic integration methods of the Monte Carlo or Quasi-Monte Carlo types, the integration error is itself an estimate, which contains its own error. That this is not an academic point becomes clear when we realize that the error estimate is routinely used to provide *confidence levels* for the integral estimate (be it based either or Chebyshev or Central-Limit-Theorem, Gaussian rules); and a mis-estimate of the integration error can lead to a serious under- or overestimate of the confidence level. As an example, suppose that the Central Limit Theorem is applicable, so that the integration result is drawn from a Gaussian distribution centered around the true integral value. One standard deviation, as estimated by Monte Carlo, corresponds to a two-sided confidence level of 68%. If the error estimate is off by 50% (admittedly a large value), the actual confidence level may then be anything between 38% and 87%.

From this consideration we are, therefore, led to a hierarchy of error estimates: the *first-order* error is that on the integral estimate, while the *second-order* error is the error on the error estimate. This in turn has, of course, its own *third-order* error, and so on. Higher orders than the second one, however, appear to be too academic for practical relevance, but we should like to argue that, in any serious integration problem, the second-order error ought to be included. In what follows we shall discuss the first- and second-order error estimates.

1.2. Monte Carlo estimators

In this section we briefly review the probabilistic theory underlying Monte Carlo integration. This is of course well known, but we include it here so that the significant difference with the Quasi-Monte Carlo can become clear.

Throughout this paper we shall consider integration problems over the d-dimensional unit hypercube $C = [0,1)^d$. The integrand is a function $f(\vec{x})$, which we shall assume real and non-negative, and, of course, integrable over C. We shall define:

$$J_m = \int_C f(\vec{x})^m \, d^d\vec{x} \qquad m = 1, 2, 3, \dots , \tag{1}$$

[1] iid stands for "independent, identically distributed".

[2] This ignores the possible interpretation of stratified and importance sampling methods of variance reduction. These can, at any rate, always be formulated in terms of methods using iid uniform integration points.

so that J_1 is the required integral. Note that J_m is not necessarily finite for $m \geq 2$. In Monte Carlo we assume N integration points, to be chosen iid from the uniform probability distribution over C. This means that the *point set* $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N\}$ on which the integration is based is *assumed* to be a typical member of an ensemble of such point sets, in such a way that the combined probability distribution of the N points over this ensemble is the uniform iid one

$$P_N(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) = 1\,. \tag{2}$$

We shall take the averages over this ensemble.

Let us assume that a point set X has been generated, and the values of the integrand $f(\vec{x})$ at all these points have been computed. These we shall denote by $f_j \equiv f(\vec{x}_j)$, $j = 1, 2, \ldots, N$. From these we can compute the discrete analogues of the integrals J_m, which are computable in linear time (that is, time proportional to N):

$$S_m = \sum_{j=1}^{N} (f_j)^m\,. \tag{3}$$

The Monte Carlo estimate of the integral is then

$$E_1 = \frac{1}{N}\, S_1\,. \tag{4}$$

The expected value of E_1 over the above ensemble of point sets is then given by

$$\langle E_1 \rangle = \frac{1}{N} \sum_i \langle f_i \rangle = \int\limits_C f(\vec{x})\; d^d\vec{x} = J_1\,, \tag{5}$$

which is indeed the required integral: this is the basis for the Monte Carlo method. Its usefulness appears if we compute the variance of E_1

$$\sigma\,(E_1)^2 = \left\langle E_1^2 \right\rangle - \langle E_1 \rangle^2 = \frac{1}{N}\left(J_2 - J_1^2\right)\,. \tag{6}$$

Since this decreases as N^{-1}, the Monte Carlo method actually converges for large N. Note that the leading, $\mathcal{O}(N^0)$, terms of $\langle E_1^2 \rangle$ and $\langle E_1 \rangle^2$ cancel against each other: this is a regular phenomenon in variance estimates of this kind[3]. The variance $\sigma\,(E_1)^2$ is estimated by the first-order error estimator

$$E_2 = \frac{1}{N^2} S_2 - \frac{1}{N^3} S_1^2\,, \tag{7}$$

[3] It should be pointed out that what we estimate is the average of the squared error, rather than the error itself, and squaring and averaging do *not* commute. In fact, this is another reason why the second-order estimate is relevant.

for which we have

$$\langle E_2 \rangle = \sigma\left(E_1\right)^2 + \mathcal{O}(N^{-2})\,. \tag{8}$$

Since N is usually quite large, at least 10,000 or so, we feel justified in working only to leading order in N. The squared error of E_2 is computed to be, to leading order in N,

$$\sigma\left(E_2\right)^2 = \frac{1}{N^3}\left(J_4 - 4J_3J_1 - J_2^2 + 8J_2J_1^2 - 4J_1^4\right)\,, \tag{9}$$

for which the estimator is

$$E_4 \;=\; \frac{1}{N^7}\left(N^3S_4 - 4N^2S_3S_1 - N^2S_2^2 + 8NS_2S_1^2 - 4S_1^4\right)\,. \tag{10}$$

which can also be computed in linear time; we have

$$\langle E_4 \rangle = \sigma\left(E_2\right)^2 + \mathcal{O}(N^{-4})\,. \tag{11}$$

Some details on the computation of leading-order expectation values of this type, as well as (for purposes of illustration) the form of the third- and fourth-order error estimators, will be given in [1].

1.3. Quasi-Monte Carlo estimators

In contrast to the case of regular Monte Carlo, the technique of Quasi-Monte Carlo relies on point sets in which the points are *not* chosen iid from the uniform distribution, but rather interdependently. To make this more specific, let us consider a point set X of N points. For such a point set, we may define a *measure of non-uniformity*, called a *discrepancy* or, as in this paper, a *diaphony*. Its precise definition is presented below: for now, suffice it to demand that there exist a function $D(X)$ of the point set, which increases with its non-uniformity: $D(X) = 0$ if the point set is perfectly uniform in all possible respects, an ideal situation that can never be obtained for any finite point set. The Quasi-Monte Carlo method consists of using point sets X for which $D(X)$ has some value s which is (very much) smaller than $\langle s \rangle$, the value that may be expected for truly iid uniform ones.

Given that such "quasi-random" point sets can be obtained, how does one use them in numerical integration? The obvious issue here is to determine of what ensemble the quasi-random point set X can be considered to be a "typical" member. In this paper, we should like to advocate the viewpoint that, since the main additional property of the quasi-random point set that distinguishes it from truly random point sets is its "anomalously small" discrepancy D, the ensemble ought to consist of those point sets that are

iid uniformly, with the additional constraint that the discrepancy D has the particular value $D(X) = s$ for the actually used point set[4].

The multi-point distribution of such a point set P_N is now no longer simply unity, since that would imply independence of the points in the point set. Let us, therefore, write the *multipont distribution* as

$$P_N(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) = 1 - \frac{1}{N} F_N(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N)\,, \tag{12}$$

where we have anticipated a factor $1/N$ before the *multipoint correlation* F_N. Since an obvious requirement on the correlation function is that it must be independent of the ordering of the points, $F_N(s; \ldots)$ must be totally symmetric; moreover, we must have

$$F_k(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_K) = \int\limits_C F_{k+1}(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k, \vec{x}_{k+1})\; d^d\vec{x}_{k+1}\,. \tag{13}$$

Finally, since the minimum requirement is that the Quasi-Monte Carlo integral be unbiased, we must have

$$P_1(s; \vec{x}_1) = 1\,, \tag{14}$$

so that

$$\int\limits_C F_2(s; \vec{x}_1, \vec{x}_2)\; d^d\vec{x}_2 = 0\,. \tag{15}$$

The above remains, of course, to be proven for any prescription used for the construction of the correlation function, and we shall do so in the next section, for our particular prescription. This establishes the properties of our ensemble of point sets X on which, to our view, the Quasi-Monte Carlo estimates ought to be based. We shall indicate the "Quasi-Monte Carlo" nature of the estimators by the superscript (q). The first estimator is that of the integral

$$E_1^{(q)} = \frac{1}{N} \sum f_j\,. \tag{16}$$

Here, and in the rest of this section, the sums will run from 1 to N. Denoting by the subscript (q) averages with respect to the "quasi-random" ensemble discussed above, we then have

$$\left\langle E_1^{(q)} \right\rangle_{(q)} = \int\limits_C f(\vec{x})\; P_1(s; \vec{x})\; d^d\vec{x} = J_1\,, \tag{17}$$

[4] We do not examine the possible alternative that the point sets in the ensemble must have discrepancy *in the neighborhood* of the observed value s; this amounts to the distinction between the micro-canonical and the canonical ensemble in statistical mechanics.

as before: owing to the fact that the one-point distribution is uniform, the Quasi-Monte Carlo estimate is indeed unbiased just as the Monte Carlo one. The distinction between the two methods appears in the first-order error estimate. Let us define

$$\alpha(\vec{x}_i, \vec{x}_j) = 1 + F_2(s; \vec{x}_i, \vec{x}_j)\,, \tag{18}$$

then we have

$$\sigma\left(E_1^{(q)}\right)^2_{(q)} = \frac{1}{N}\left(J_2 - \int f_1 f_2 \alpha_{12}\right) + \mathcal{O}\left(\frac{1}{N^2}\right), \tag{19}$$

where we have adopted the straightforward convention for integrals

$$\int f_1 f_2 \alpha_{12} = \int\limits_C f(\vec{x}_1)\, f(\vec{x}_2)\, \alpha(\vec{x}_1, \vec{x}_2)\, d^d\vec{x}_1\, d^d\vec{x}_2\,, \tag{20}$$

etcetera. As before, we shall happily neglect terms that are subleading in $1/N$. The advantage of the Quasi-Monte Carlo method is now clear: if we can ensure that $\alpha_{12} > 1$ "where it counts", that is, generally, when $\vec{x}_1$ and $\vec{x}_2$ are "close" in some sense, then the Quasi-Monte Carlo error will be smaller than the Monte Carlo one. A good Quasi-Monte Carlo point set, therefore, is one in which the points "repel" each other to some extent.

The first-order error estimate is now simply

$$E_2^{(q)} = \frac{1}{N^2}\sum f_i^2 - \frac{1}{N^3}\sum f_i f_j \alpha_{ij}\,. \tag{21}$$

It is simple to show that, indeed

$$\left\langle E_2^{(q)} \right\rangle_{(q)} = \sigma\left(E_1^{(q)}\right)^2_{(q)} + \mathcal{O}(N^{-2})\,, \tag{22}$$

however, *evaluating* $E_2^{(q)}$ in linear time is less trivial since a simple expression for $F_2(s;\ldots)$ has to be derived. We shall discuss this later. The variance of the estimator $E_2^{(q)}$ can be evaluated to

$$\begin{aligned}\sigma\left(E_2^{(q)}\right)^2 \;=\; \frac{1}{N^3}\Bigg(&\int f_i^4 - 4\int f_i^3 f_j \alpha_{ij} - \int f_i^2 f_j^2 \alpha_{ij} \\ &+ 4\int f_i^2 f_k f_l \alpha_{ik}\alpha_{kl} + 4\int f_i^2 f_k f_l \alpha_{ik}\alpha_{il} \\ &- 4\int f_i f_j f_k f_l \alpha_{ij}\alpha_{jk}\alpha_{kl}\Bigg) + \mathcal{O}(N^{-4})\,,\end{aligned} \tag{23}$$

for which the corresponding estimator (to leading order) is

$$\begin{aligned} E_4^{(q)} &= \frac{1}{N^7}\Big(N^3\sum f_i^4 - 4N^2\sum f_i^3 f_j\alpha_{ij} - N^2\sum f_i^2 f_j^2\alpha_{ij} \\ &\quad + 4N\sum f_i^2 f_k f_l\alpha_{ik}\alpha_{kl} + 4N\sum f_i^2 f_k f_l\alpha_{ik}\alpha_{il} \\ &\quad - 4\sum f_i f_j f_k f_l\alpha_{ij}\alpha_{jk}\alpha_{kl}\Big)\,. \end{aligned} \tag{24}$$

A technique with Feynman graphs has been deviced, that yields the higher order error estimators, where the role of $\hbar$ is played by $1/N$. The details of this technique, which allowed us to also find E_8 will be discussed in [1]. It goes without saying that the substitution $\alpha_{ij} \to 1$ will reduce all the Quasi-Monte Carlo results to the regular Monte Carlo ones.

2. Multipoint distributions

2.1. The multi-point distribution in general

Under the premise that the Quasi-Monte Carlo point sequence we use is a typical member of the ensemble of point sets with the particular value of diaphony $D(X) = s$, the Quasi-Monte Carlo analogue of Eq. (2) would then be the assumption

$$P_N(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_N) = \frac{1}{H(s)}\delta(D(X)-s)\,, \tag{25}$$

where s is, again, the observed value of the discrepancy of X, on which P_N must now of course depend; and $H(s)$ is the probability density to happen upon a point sets X with this discrepancy in the regular Monte Carlo ensemble:

$$H(s) = \int\limits_C \delta(D(X)-s)\; d^d\vec{x}_1\; d^d\vec{x}_2\; d^d\vec{x}_N\,. \tag{26}$$

2.2. Diaphony

We consider a point set X with N elements, given in C. The non-uniformity of the point set X can be described by its *diaphony*

$$D(X) = \frac{1}{N}\sum_{j,k=1}^{N}\beta(\vec{x}_j,\vec{x}_k)\,, \tag{27}$$

with

$$\begin{aligned} \beta(\vec{x}_j,\vec{x}_k) &= \hat{\sum_{\vec{n}}}\,\sigma_{\vec{n}}^2\; e_{\vec{n}}(\vec{x}_j)\bar{e}_{\vec{n}}(\vec{x}_k)\,, \\ e_{\vec{n}}(\vec{x}) &= \exp(2i\pi\;\vec{n}\cdot\vec{x})\,. \end{aligned} \tag{28}$$

Here, the vectors $\vec{n} = (n_1, n_2, \ldots, n_d)$ form the integer lattice, and the hat denotes the sum over all $\vec{n}$ except $\vec{n} = \vec{0}$. We may also write

$$D(X) = \frac{1}{N} \hat{\sum_{\vec{n}}} \sigma_{\vec{n}}^2 \left| \sum_{j=1}^{N} e_{\vec{n}}(\vec{x}_j) \right|^2 , \tag{29}$$

so that we recognize the diaphony as a measure of how well the various Fourier modes are integrated by the point set X[5]. The diaphony is, therefore, seen to be related to the "spectral test", well-known in the field of random-number generator testing. For the *mode strengths* $\sigma_{\vec{n}}^2$ we have

$$\sigma_{\vec{n}}^2 \leq 0 , \qquad \hat{\sum_{\vec{n}}} \sigma_{\vec{n}}^2 = 1 . \tag{30}$$

The latter convention simply establishes the overall normalization of D. The advantage of this diaphony over, say, the usual (star) discrepancy is the fact that it is translation-invariant:

$$\beta(\vec{x}_j, \vec{x}_k) = \beta(\vec{x}_j - \vec{x}_k) , \tag{31}$$

so that point sets X and X' that differ only by a translation (modulo 1) have the same nonuniformity: the diaphony is actually defined on the hypertorus rather than on the hypercube. Also, the diaphony is *tadpole-free*

$$\int_C \beta(\vec{x}) \, d^d\vec{x} = 0 . \tag{32}$$

Moreover, we shall use $\sigma_{\vec{n}}^2$ such that $\sigma_{\vec{n}}^2 = \sigma_{\vec{n}'}^2$ if the two lattice vectors $\vec{n}$ and $\vec{n}'$ differ only by a permutation of their components. Thus, X and X' will also have the same nonuniformity if they differ by a global permutation of the coordinates of the points.

2.3. Multipoint distribution by Laplace transform

As discussed above, let $H(s)$ be the probability that the point set X has diaphony equal to s, that is, $D(X) = s$. The underlying ensemble of point sets is that of sets of N iid uniformly distributed points, *i.e.* the same ensemble underlying the usual Monte Carlo error estimates.

[5] It should be noted, however, that this specific choice for $e_{\vec{n}}(x)$ is by no means mandatory. The only necessary condition here and in all that follows is that the functions $e_{\vec{n}}(x)$ form a complete orthonormal set in $[0,1]^D$.

Let us define

$$G_p(z;\vec{x}_1,\dots\vec{x}_p) = \int d^d\vec{x}_{k+1}\dots\vec{x}_N e^{zD(x)}\,. \tag{33}$$

Then, we have

$$\begin{aligned} H(s) &= \int\limits_C d^d\vec{x}_1\; d^d\vec{x}_2\dots d^d\vec{x}_N\;\delta(D(X)-s) \\ &= \frac{1}{2i\pi}\int\limits_{-i\infty}^{+i\infty} e^{-zs}\,G_0(z)\,dz\,, \end{aligned} \tag{34}$$

where the integration contour runs to the left of all the singularities of $G_0(z)$; and the multipoint distribution for p points averaged over all point sets X with diaphony s, is given by

$$\begin{aligned} P_p(s;\vec{x}_1,\vec{x}_2,\dots,\vec{x}_p) &= \frac{1}{H(s)}R_p(s;\vec{x}_1,\vec{x}_2,\dots,\vec{x}_p)\,, \\ R_p(s;\vec{x}_1,\vec{x}_2,\dots,\vec{x}_p) &= \frac{1}{2i\pi}\int\limits_{-i\infty}^{+i\infty} e^{-zs}\,G_p(z)\,dz\,. \end{aligned} \tag{35}$$

It has been shown [2] that

$$G_0(z) = \exp\Big(-\frac{1}{2}\sum_{\hat{\vec{n}}}\ln\big(1-2z\sigma^2_{\vec{n}}\big)\Big) + O\left(\frac{1}{N^2}\right) \tag{36}$$

and

$$G_2(z;x,y) = G_0(z)\Big(1+\frac{1}{N}\sum_{\hat{\vec{n}}}\frac{2z\sigma^2_{\vec{n}}}{1-2z\sigma^2_{\vec{n}}}\,e_{\vec{n}}(x)\,e^*_{\vec{n}}\Big) + O\left(\frac{1}{N^2}\right)\,. \tag{37}$$

Upon insertion of Eqs. (36), (37) in the formula for the two-point correlation function we get

$$F_2(s;x,y) = \sum_{\hat{\vec{n}}}\omega_{\vec{n}}\,e_{\vec{n}}(x)\,e^*_{\vec{n}}(y)\,, \tag{38}$$

with

$$\omega_{\vec{n}} = \frac{\int\limits_{-i\infty}^{i\infty} dz e^{-zs}G_0(z)\frac{-2z\sigma^2_{\vec{n}}}{1-2z\sigma^2_{\vec{n}}}}{\int\limits_{-i\infty}^{i\infty} dz\,e^{-zs}G_0(z)}\,. \tag{39}$$

Except in the very simplest cases, a complete evaluation of Eq. (38) is nontrivial. A simplification arises if s is much smaller than its expectation value 1 (which is anyway the aim in quasi-Monte Carlo), or if the Gaussian limit is applicable, namely when the number of modes with non-negligible $\sigma^2_{\vec{n}}$ becomes large in such a way that no single mode dominates. In practice, this happens when the dimensionality of C becomes large. Fortunately, these are precisely the situations of interest. The position of the saddle point for $H(s)$, $\hat{z}$, is given by

$$\hat{\sum_{\vec{n}}} \frac{\sigma^2_{\vec{n}}}{1 - 2z\sigma^2_{\vec{n}}} = s\,. \tag{40}$$

For $s \ll 1$, therefore, $\hat{z}$ is large and negative. Since to first order the same saddle point may be used for R_2, we find the attractive result

$$F_2(s; \vec{x}_1, \vec{x}_2) \approx \hat{\sum_{\vec{n}}} \frac{-2\hat{z}\sigma^2_{\vec{n}}}{1 - 2\hat{z}\sigma^2_{\vec{n}}} e_{\vec{n}}(\vec{x}_1)\bar{e}_{\vec{n}}(\vec{x}_2)\,. \tag{41}$$

The formulae (40) and (41) suffice, in our approximation, to compute all the multipoint correlations.

3. The estimator

3.1. The estimator analyzed

Inserting Eq. (41) in the equation for our estimator (Eq. 21) we arrive at the following estimator for the Quasi-Monte Carlo error

$$E_2^{(q)} = \frac{1}{N^2}\sum f_i^2 - \frac{1}{N^3}\left(\sum f_i\right)^2 - \frac{1}{N^3}\sum_{\hat{\vec{n}}} \omega_{\vec{n}} \Big| \sum_i f_i e_{\vec{n}}(x_i)\Big|^2\,, \tag{42}$$

with

$$\omega_{\vec{n}} = \frac{-2\hat{z}\sigma^2_{\vec{n}}}{1 - 2\hat{z}\sigma^2_{\vec{n}}}\,. \tag{43}$$

We are still free to choose the exact form of the weights $\sigma^2_{\vec{n}}$ at will, under the constraints of Eq. (30). Our choice is the so called Jacobi weights

$$\sigma^2_{\vec{n}} = K e^{-\lambda \vec{n}^2} \tag{44}$$

with

$$K^{-1} = \sum_{\hat{\vec{n}}} e^{-\lambda \vec{n}^2}\,. \tag{45}$$

The parameter λ is regulating the sensitivity of the diaphony: as $\lambda \to 0 \Rightarrow \sigma_{\vec{n}} \to 1$ for every mode while as $\lambda \to \infty \Rightarrow \sigma_{\vec{n}} \to 0$. The first case

corresponds to a super-sensitive diaphony, useless for practical purposes, whereas the second case corresponds to a non-sensitive diaphony that would value equally all point-sets ($D(X) = 0$ always). In effect λ defines the number of "active" modes in the diaphony. We choose $\lambda = 0.1$.

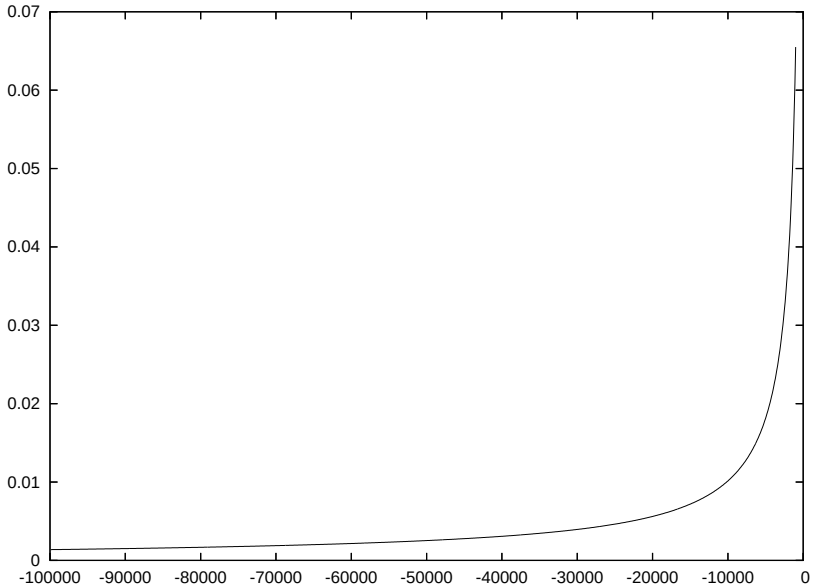

Fig. 1. The value of the diaphony as a function of the value of the saddle point through $s = \sum_n (\sigma_n^2/(1 - 2z\sigma_n^2))$ for $\lambda = 0.1$ and $d = 2$.

Since the estimator (Eq. (42)) contains an infinite sum over modes, we have to give a prescription as to which modes one is entitled to disregard. Fig. 1 shows the dependence of the saddle point $\hat{z}$ to the value of the diaphony s. As the value of s becomes small the saddle point becomes quickly large and negative $\hat{z} \ll 0$. Then $-2\hat{z}\sigma_{\vec{n}}^2 \to \infty$ for low modes and $-2\hat{z}\sigma_{\vec{n}}^2 \to 0$ for higher modes, when $\sigma_{\vec{n}}^2/|\hat{z}| \to 0$. We can thus safely neglect these higher modes in the estimator. As long as the value of the diaphony is small, which is in any case the goal in Quasi-Monte Carlo the profile of $\omega_{\vec{n}}$ depends only on the choice of λ, which regulates the sensitivity of the diaphony. We see, therefore, that the estimator inherits the sensitivity of the diaphony in a direct way.

It is easy to see that the estimator averages (to leading order in N) in a positive definite quantity. This leaves open the possibility for a negative error estimate, in which case the error on the above estimator becomes particularly relevant. In this case the higher order terms ($1/N^2$ supressed) that we have neglected are actually important, which indicates that as N becomes large the negative error effect should disappear.

3.2. The estimator plotted

In the following we present a number of plots that show how both the classical and the quasi error estimates behave as a function of the number of points N. We use a RANLUX [3] generator for pseudo-random point-sequences and a `Van de Corput` [4] generator with base $2, 3, 5, 7, 11, \ldots$ for quasi-random sequences. All integrations are performed in the unit hypercube $[0,1]^D$.

3.2.1. Test functions

The test functions we use consist of a subset of the test functions used by Schlier in [5], along with two Gaussian functions with constant and dimensionally dependent width. We have

$$\mathrm{TF13} : f(\vec{x}) = \prod_{k=1}^{D} \frac{|4x_k - 2| + k}{1 + k}\,, \tag{46}$$

which averages to $J_1 = 1$. This test function is especially taylored for a Van de Corput sequence, since in $D = 1$ it is perfectly integrated by such a sequence with base 2.

$$\mathrm{TF2} : f(\vec{x}) = \prod_{k=1}^{D} k\cos(kx_k)\,, \tag{47}$$

which averages to $J_1 = \prod_k \sin(k)$. This function should be difficult to integrate in high dimensions.

$$\mathrm{TF4} : f(\vec{x}) = \sum_{k=1}^{D} \prod_{j=1}^{k} x_j\,, \tag{48}$$

which averages to $J_1 = 1 - (1/2^D)$. It is chosen as a simple example of a function that is not a product of one dimensional functions.

$$\mathrm{TF5} : f(\vec{x}) = Ke^{-(\vec{x}-\vec{x}_0)^2/2\sigma^2} \qquad \sigma^2 = 0.01\,, \tag{49}$$

which averages to $J_1 = 1$ (K is the normalization factor). This is the standard Gaussian peak with fixed width. The fixed width results in a rapid decrease, in higher dimensions of the subset of the integration volume where the function is non-zero, making the integration cumbersome (the higher the dimension, the more points are needed). To fix this side effect we also use

$$\mathrm{TF6} : f(\vec{x}) = Ke^{-(\vec{x}-\vec{x}_0)^2/2\sigma^2}\,, \qquad \sigma^2 = \frac{1}{4\pi(1+a)^{2/D}}\ a = 10\,, \tag{50}$$

which averages to $J_1 = 1$ (K is the normalization factor). The width in this function increases in such a way as to keep the ratio of the useful integration volume to the total fixed.

3.2.2. Plots

In the following some plots of the error and its estimates as functions of the number of points N are shown in a logarithmic scale. For a more detailed presentation the reader is deferred to [1]. Here the classical error estimate, based on the iid assumption is presented, along with three versions of quasi error estimators, $E_2^{q5}, E_2^{q10}, E_2^{q15}$. The superscript next to q denotes the squared length of the higher modes included in the sum of Eq. (42). Thus E_2^{q10} includes modes with $\vec{n}^2 \leq 10$. The real error made is included for comparison. Moreover, in the first plot the performance of a pseudo-random point set produced by `RANLUX` is presented for comparison. The observation, there, that the pseudo-random pointsets perform much worse than the quasi-random ones persists in all other cases. The error under the pseudo-random set is not depicted in the rest of the plots for reasons of clarity. The reader is deferred to [1] for a more elaborate treatment.

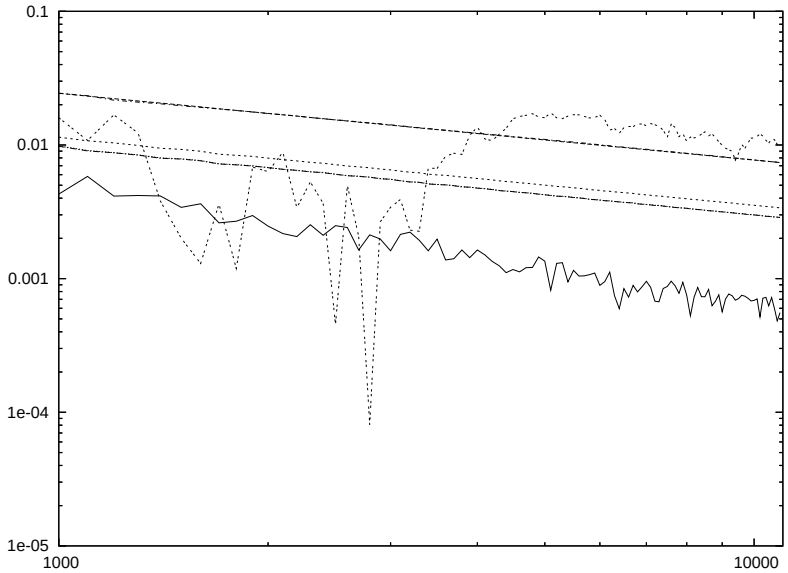

Fig. 2. TF2, $d = 2$ log-plot of the real error (continuous line), and then from top to bottom the classical estimate (slashed), E_2^{q5} (small-slashed), E_2^{q10} (dotted) and E_2^{q15} (slashed–dotted). The fluctuating dotted line in this plot is the error with the `RANLUX` pseudorandom point set. The comparison with the real Quasi-Monte Carlo error shows that the quasi point set integrates much better. Further more, the classical error estimator is far off the real error whereas the quasi estimators are approaching the real error as more modes are added to the sum.

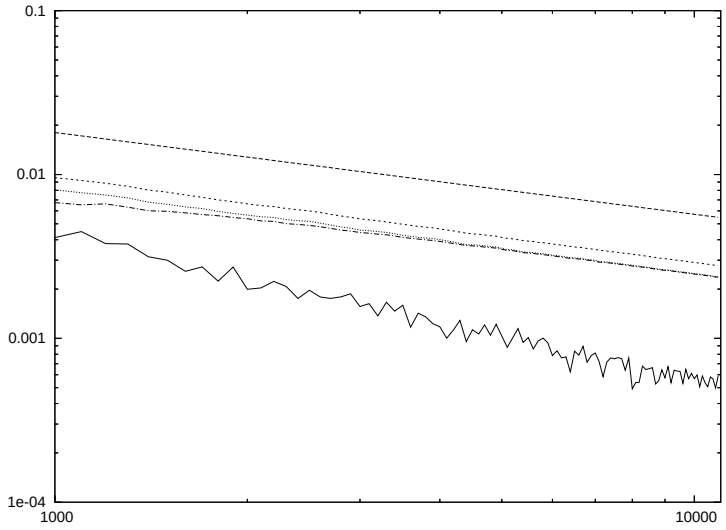

Fig. 3. TF4, $d = 3$ log-plot of the real error (continuous line), and then from top to bottom the classical estimate (slashed), E_2^{q5} (small-slashed), E_2^{q10} (dotted) and E_2^{q15} (slashed–dotted). Here TF4 is shown in 3 dimensions, the need for more modes becomes apparent though the improvement in the error estimate using the quasi estimator is still significant.

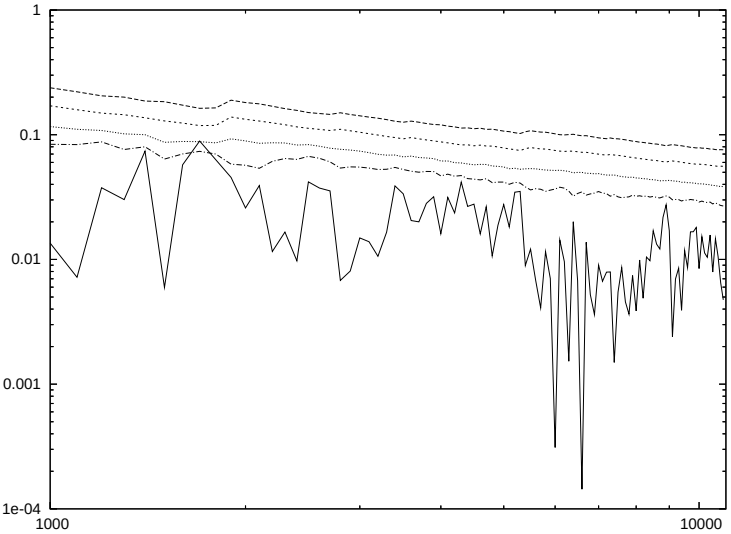

Fig. 4. TF5, $d = 3$ log-plot of the real error (continuous line), and then from top to bottom the classical estimate (slashed), E_2^{q5} (small-slashed), E_2^{q10} (dotted) and E_2^{q15} (slashed–dotted). The Gaussian peak test function in 3 dimensions behaves much better. The estimator is not only improved but also approaching very well the real error made by the use of the Van de Corput sequence.

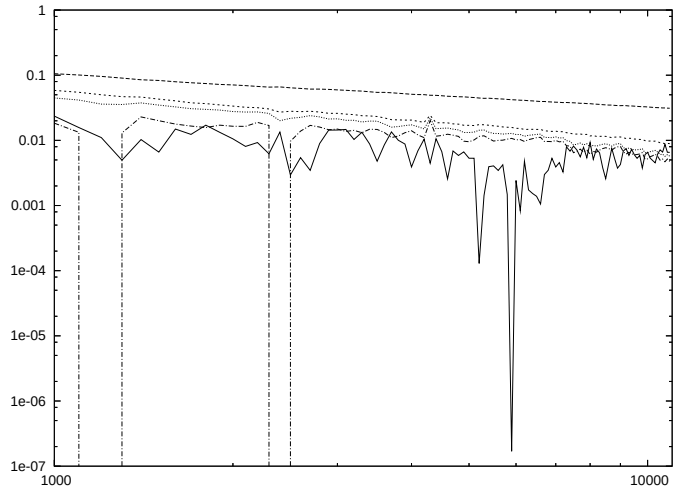

Fig. 5. TF6, $d = 4$ log-plot of the real error (continuous line), and then from top to bottom the classical estimate (slashed), E_2^{q5} (small-slashed), E_2^{q10} (dotted) and E_2^{q15} (slashed–dotted). The quasi estimators approximate well the error, but now the negative error effect appears. The fact that for $N \to 10^4$ the estimator returns to positive values suggests that the negative values effect is indeed a statistical fluctuation and not a systematic error. This is the case where the error on the error estimator becomes relevant.

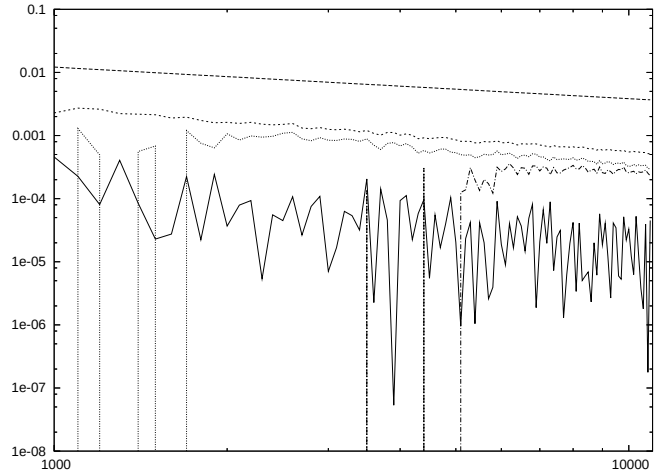

Fig. 6. TF13, $d = 3$ log-plot of the real error (continuous line), and then from top to bottom the classical estimate (slashed), E_2^{q5} (small-slashed), E_2^{q10} (dotted) and E_2^{q15} (slashed–dotted). Here the error is particularly small and the error on the error makes the estimator with modes of squared length up to 10 and up to 15 negative for small N. The conclusion should be once more that the error on the error is important in such cases. As this second order error decreases with N the error itself becomes positive approaching its strictly positive mean value.

4. Outlook

We have seen that the error estimator suggested in this paper performs always better than the classical estimator when one uses Quasi-Monte Carlo point sequences. The price to pay is the raise in the complexity of the computation of the estimator from linear to linear times the number of modes involved. When the dimensionality of the integral is high the number of modes that are close to zero and, therefore, most relevant for every reasonable diaphony definition, becomes overwhelmingly large. The solution out of this deadlog could be reached through sampling over the mode sum in an efficient way. This we defer to further research.

Moreover, one could decide to consider the point set in question as a typical member of the ensemble of pointsets with a value s of the diaphony which is not precisely $D(X)$ but narrowly distributed around $D(X)$. The statistical ensemble would then be different and there are indications that this approach would be more convenient for progress in the analytic part of the calculation.

Finally, further experience with more realistic test functions and the implementation of more advanced Quasi-Monte Carlo sequences (like for example the Niederreiter sequences [6]) would be in order for further improving the present results.

REFERENCES

[1] R. Kleiss, A. Lazopoulos, to be published

[2] J. Hoogland, R. Kleiss, `hep-ph/9601270`.

[3] M. Luescher, *Comput. Phys. Commun.* **79**, 100 (1994); F. James, *Comput. Phys. Commun.* **79**, 110 (1994).

[4] A. Halton, *J. Numer. Math.* **2**, 84 (1960).

[5] Ch. Schlier, *Comput. Phys. Commun.* **159**, 93 (2004).

[6] H. Niederreiter, *J. Number Theory* **30**, 51 (1988).