Quasi Monte Carlo

Lecture #10: Thursday, 30 October 1997
Lecturer: Eric Veach
Scribe: Matt Pharr
Reviewer: David Hoffman

1 Overview

For Monte Carlo integration to be effective, we want to use sample points that are uniformly distributed across the domain of integration. The last lecture introduced Latin hypercube sampling, which gives samples that are well-stratfied with respect to any one dimension, and orthogonal array sampling, which gives samples that are well-distributed with respect to any combination of two, three, or more dimensions.

Quasi Monte Carlo (QMC) is based on the approach of removing randomness from the generation of sampling sequences; instead, the idea is to look for fixed sequences of points that perform *better* than random sequences, at least for certain classes of functions. Techniques have been developed to generate both finite and infinite sequences of well distributed quasi-random points; in general, slightly better distributions are possible if the number of sample points needed is known in advance.

2 Discrepancy

The first question we have to answer is, given a sequence of samples, how well distributed are the samples (and thus, how effective will they be in estimating the values of the integrals we want to compute.) Discrepancy is a measure of irregularity of distribution which is used to evaluate the quality of QMC sampling sequences. The basic idea is to look at various regions of the domain and compare the volume of these regions to the number of sample points inside them. In general, one fourth of the volume should correspond to one fourth of the sample points, etc.

2.1 Definition

Assume we're evaluating a set of points defined over $[0,1]^s$. To compute the discrepancy of the points, we first pick a family of shapes $\mathcal{B}$ which are subsets of $[0,1]^s$. For example, boxes with one corner at the origin are often used. This corresponds to:

$$\mathcal{B} \;=\; \{B = [0,v_1] \times [0,v_2] \times \cdots \times [0,v_s] \mid 0 \le v_i \le 1 \text{ for all } i\} \;\; .$$

Given a sequence of sample points $P = x_1, \ldots, x_N$, the discrepancy of P with respect to $\mathcal{B}$ is

$$D_N(B, P) = \sup_{B \in \mathcal{B}} \left| \frac{\#\{x_i \in B\}}{N} - \lambda(B) \right|,$$

where $\lambda(B)$ is the volume of B. In other words, we're finding the maximum difference between the fraction of points inside one of the shapes and the volume of the shape. When the set of shapes $\mathcal{B}$ is the set of boxes with a corner at the origin (described above), this is called the *star discrepancy* $D_N^*(P)$.

Other popular sets of shapes to use to compute discrepancy include arbitrary axis aligned boxes, hyperplanes that cut the domain into two pieces, etc.

2.2 Some Examples

Consider the set of points in one dimension

$$x_i = i/N.$$

We can see that the star discrepancy of x_i is

$$D_N^*(x_1, \ldots, x_n) = \frac{1}{N}.$$

For example, take the interval $B = [0, \frac{1}{N})$. Then $\lambda(B) \approx \frac{1}{N}$, but $\#\{x_i \in B\} = 0$. This is the interval (along with the intervals $[0, \frac{2}{N})$, etc.) with the largest difference between volume and the fraction of points contained.

We can improve on the star discrepancy of this sequence by modifying it slightly:

$$x_i = \frac{i - \frac{1}{2}}{N}.$$

Then

$$D_N^*(x_i) = \frac{1}{2N}.$$

A theorem due to H. Niederreiter (one of the main pioneers of QMC) provides some bounds for the star discrepancy of a sequence of points in 1D:

$$D_N^*(x_i) = \frac{1}{2N} + \max_{1 \le i \le N} \left| x_i - \frac{2i-1}{2N} \right|.$$

And thus, the second example's sequence has the lowest possible discrepancy for a sequence in 1D. In general, it is much easier to analyze and compute bounds for the discrepancy of sequences in 1D than in higher dimensions.

Another often used type of discrepancy is the *extreme discrepancy*, which uses arbitrary axis-aligned boxes for the set of shapes. The extreme discrepancies of the 1D examples given above are both $\frac{1}{N}$.

3 Theoretical Performance of QMC

The Koksma-Hlawka theorem gives an upper bound for the error in QMC evaluation of integrals, as a product of two factors: one factor depends on the discrepancy of the point set used, while the other depends on the function being integrated. Given an integral

$$I = \int_{[0,1]^s} f(x_1, \ldots, x_s)\, dx_1 \cdots dx_s,$$

and a set of sample points $P = (p_1, \ldots, p_N)$, consider an estimate of the form

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} f(p_i).$$

The Koksma-Hlawka theorem says that

$$|I - \hat{I}| \leq V(f)\, D_N^*(P). \tag{1}$$

Thus, the error is split into a component $V(f)$ that depends only on the function being integrated and a component $D_N^*(P)$ that depends only on the point sequence. Therefore, so long as $V(f)$ is bounded (and it isn't always bounded), the lower we can make the discrepancy of the points, the lower the maximum error will be.

In s dimensions, it is possible to get sequences such that

$$D_N^*(P) = O\left(\frac{(\log N)^{s-1}}{N}\right).$$

In particular, note that for $s = 1$,

$$D_N^*(P) = O\left(\frac{1}{N}\right).$$

As the number of dimensions increases, we can't do as well as we can in 1D, but it's nearly as good. Note that this convergence rate is much better than the $O(N^{-\frac{1}{2}})$ that standard Monte Carlo gives. However, in higher dimensions the $(\log N)^{s-1}$ factor can be quite large. For example if $s = 6$, then $(\log N)^{s-1}/N > N^{-1/2}$ for all $N < 10^15$. So, this error bound is not always very useful in practice.

$V(f)$ is called the *variation of f in the sense of Hardy and Krause*. It's easy to define in one dimension:

$$V(f) = \int_0^1 |f'(x)|\, dx,$$

as long as the derivative $f'(x)$ is continuous. Basically, it's the sum of the heights of all the monotonic segments of f (see Figure 1).

In two or more dimensions, if f is discontinuous, the variation is infinite and the bounds 1 are meaningless. In three or more dimensions, if f' is discontinuous, the variation is infinite. In general, in s dimensions, the first $s - 2$ derivatives of f must be

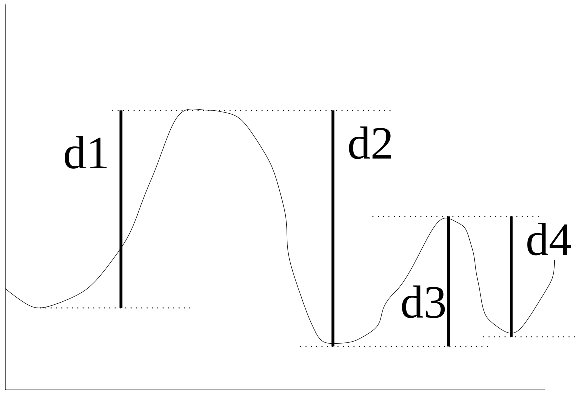

Figure 1: $V(f) \approx d1 + d2 + \ldots$

continuous for $V(f)$ to be bounded. In spite of the lack of theoretical bounds when $V(f)$ is unbounded, however, QMC can still do better than standard MC in practice.

4 Constructing Low Discrepancy Sequences

4.1 1D: Radical inverse sequences

One of the simplest low discrepancy sequences to construct is the *radical inverse sequence*. The idea is to write down an integer in a given base b, and then reflect the digits around the decimal point. When the base used is $b = 2$, this is called the *van der Corput sequence*:

1	.1	1/2
10	.01	1/4
11	.11	3/4
100	.001	3/8
101	.101	5/8
⋮	⋮	⋮

It recursively splits the intervals of the 1D line in half. The discrepancy of this sequence is

$$D_N^*(P) = O\left(\frac{\log N}{N}\right).$$

This matches the best discrepancy that has been attained for infinite sequences in s dimensions:

$$D_N^*(P) = O\left(\frac{(\log N)^s}{N}\right).$$

For finite sequences, better discrepancy is possible; the bound is

$$D_N^*(P) = O\left(\frac{(\log N)^{s-1}}{N}\right).$$

In terms of theoretical bounds on discrepancy, it is believed that these attained bounds are as good as it gets, but the best proven lower bound for sequences (in an arbitrary dimension s) is

$$D_N^*(P) = \Omega\left(\frac{(\log N)^{\frac{s}{2}}}{N}\right)$$

while for a finite point set it is

$$D_N^*(P) = \Omega\left(\frac{(\log N)^{\frac{s-1}{2}}}{N}\right).$$

Thus, most of the continuing work in constructing low-discrepancy sequences centers around reducing the constant factor k (in one dimension):

$$D_N^*(P) = O\left(\frac{\log N}{N}\right) \leq k\frac{\log N}{N}$$

for $N \geq 2$. The best k that has been attained so far is 0.224; it has been proven that $k = 0.06$ is unattainable.

4.2 Two and more dimensions

4.2.1 Halton Sequences

Halton sequences build on the radical inverse used in Van Der Corput sequences and introduce $\phi_b(i)$, the radical inverse of i in base b. To generate an n dimensional sequence, we use the radical inverse base b, with a different base for each dimension:

$$x_i = (\phi_2(i), \phi_3(i), \phi_5(i), \ldots, \phi_{p_s}(i)).$$

The first s prime numbers, $p_1, \ldots, p_s$ are used as the bases for the s dimensions. The discrepancy of these sequences is good:

$$D_N^*(x_i) = O\left(\frac{(\log N)^s}{N}\right).$$

Note that in each dimension, the Halton sequence fills in the sample points from left to right as the unit digit increases from 0 to $b - 1$. Permutations of the sequence can improve their performance in practice.

4.2.2 Hammersley Points

If you know the number of samples to be taken in advance, discrepancy can be improved slightly. Hammersley point sets are defined by:

$$x_i = (\frac{i}{N}, \phi_2(i), \phi_3(i), \ldots, \phi_{s-1}(i)),$$

where N is the total number of samples to be taken. See "Ray Tracing and Irregularities of Distribution", by Don Mitchell, in the proceedings of the Third Eurographics Rendering Workshop for more on the uses of Hammersley points in graphics (more precisely, the Hammersley-Zaremba points, which use permutations as mentioned above).

4.2.3 (t, s)-sequences and (t, m, s)-nets

A family of low discrepancy sequences called (t, s)-sequences and (t, m, s)-nets has been constructed based on looking at the distributions of points with respect to b-ary boxes: these are axis aligned boxes, coincident with the lines of $(\frac{1}{b})^i$. They are defined by:

$$E = \prod_{i=1}^{s} [a_i b^{-d_i}, (a_i + 1) b^{-d_i}],$$

where $d_i \geq 0$, $0 \leq a_i < b^{d_i}$, and $\prod$ denotes a product of intervals over all dimensions.

For example, with $b = 5$ and $s = 3$, valid boxes include those of size $1 \times \frac{1}{5} \times \frac{1}{5}$, $\frac{1}{25} \times \frac{1}{125} \times 1$, etc., where the box is aligned on each axis to an integer multiple of its dimension on that axis.

(t, s)-sequences are infinite sequences with low discrepancy with respect to b-ary boxes, and (t, m, s)-nets are finite sequences with similarly good discrepancy; details of the construction of these sequences was beyond the scope of the lecture.

(t, m, s)-nets have some particularly nice properties. By definition, s is the number of dimensions we are integrating over, $0 \leq t \leq m$, and the total number of points $N = b^m$. (t, m, s)-nets are constructed so that if E is a b-ary box with volume $\lambda(E) = b^{t-m}$, then

$$\#\{x_i \in E\} = b^t.$$

The best case of this is when $t = 0$; then any b-ary box of size b^{-m} will have exactly one point in it–exactly what we'd want!

5 Summary

QMC can be a big win when doing numerical integration; keep in mind, though, that discontinuities in the integrand prevent it from being as powerful as one might expect from theoretical bounds in the presence of smooth integrands. Another complication is that classic estimates of variance can't be computed when using QMC, since computing an estimate again will always give the same result.