POLYTECHNIC UNIVERSITY
Department of Computer and Information Science

QUASI-MONTE CARLO METHOD USING LOW-DISCREPANCY NUMBERS

K. Ming Leung

Abstract: We discuss algorithms to generate some of the popular quasi-random sequences, and how these numbers can be used in quasi-Monte Carlo simulations.

Directory

- Table of Contents
- Begin Article

Last Revision Date: March 23, 2004 Version 1.0

Table of Contents

1. Introduction

As we have seen, the use of N points uniformly random points in Monte Carlo simulations leads to an error term that decreases as $N^{-1/2}$. Each new point sampled adds linearly to an accumulated sum that will become the function average. But each of these points also adds linearly to an accumulated sum that will become the variance. The estimated error comes from the square root of this variance, hence the power $N^{-1/2}$. This convergence rate is often not high enough to produce sufficiently accurate results.

By changing the way how these points are sampled, we can avoid this inverse square root convergence. As an example, we can choose sample points that lie on a rectangular grid, and to sample each point exactly once in a certain order. The Monte Carlo simulation then follows a deterministic scheme and the fractional error then decreases much faster, at least as fast as N^{-1}.

The problem with a grid is that one has to decide in advance how fine it should be. It is inconvenient to ”sample until” some convergence or termination condition is met. Is there an intermediate scheme for picking sample points ”at random”, yet spread out in some self-avoiding way, lowering the chance that clustering may occurs with uniformly random points.

The answer to the above question is ”yes”. Sequences of n-tuples that are generated in a deterministic fashion and fill up space more uniformly than uncorrelated random points are called quasi-random sequences. The sample points in the sequence are designed to ”maximally avoid” each other. These sequences are also known as low discrepancy sequences. There are variations in the definition of discrepancy. They represent ways to quantify the notion of deviation from uniform distribution, or in other words, for the irregularity in the distribution.

2. Van der Corput Sequence

The simplest example is van der Corput's sequence.[1] We start by choosing an integer $b \geq 2$ as the base. Then the jth number H_j in the sequence is obtained by the following steps:

1. Express j as a number in base b.

2. Reverse the digits and put a radix point (i.e., a decimal point in base b) in front of the digits. The result is H_j in base b.
3. Then convert H_j back to the decimal system.

These numbers $H_j, j = 0, 1, \cdots$ gradually fill up the interval from 0 to 1.

As an example, suppose we choose a base $b = 3$ and want to obtain the $j = 17$ number in the van der Corput sequence. First we note that 17 can be expressed as $1 \times 3^2 + 2 \times 3^1 + 2 \times 3^0$, and so 17 is given by 122 in base 3. Reversing the digits and putting a radix point in front gives $H_j = 0.221$ in base 3. Converting the number back to decimal number gives $2 \times 3^{-1} + 2 \times 3^{-2} + 1 \times 3^{-3} = \frac{25}{27}$.

The following table shows the first 18 van der Corput numbers in base 3. Column 1 is the representation of j in the decimal system. Column 2 is it's representation in base 3. The j-th van der Corput number in base 3 is given in column 3. Finally in column 4 is the j-th member in the van der Corput sequence.

j (decimal)	j (base 3)	H_j (base 3)	H_j (decimal)
0	0	0.0	0
1	1	0.1	1/3
2	2	0.2	2/3
3	10	0.01	1/9
4	11	0.11	4/9
5	12	0.21	7/9
6	20	0.02	2/9
7	21	0.12	5/9
8	22	0.22	8/9
9	100	0.001	1/27
10	101	0.101	10/27
11	102	0.201	19/27
12	110	0.011	4/27
13	111	0.111	13/27
14	112	0.211	22/27
15	120	0.021	7/27
16	121	0.121	16/27
17	122	0.221	25/27

It is not hard to see how the van der Corput sequence works: Every time the number of digits representing j increases by one place, j's digit-reversed fraction becomes a factor of b finer-meshed. Thus the process is one of filling in all the points from 0 to 1 on a sequence of finer and finer Cartesian grids – and in a kind of maximally spread-out order on each grid (since for example the most rapidly changing digit in j controls the most significant digit of the fraction).

3. More Detailed Description

Let $b \geq 2$ be an integer. Any integer $n \geq 0$ can be expressed in base b notation as

$$n = d_j d_{j-1} \cdots d_1 d_0 (\text{base b}) = d_j b^j + \cdots + d_1 b + d_0$$

where d_i is an integer ≥ 0 and $\leq b-1$, for $i = 0, 1, \ldots, j$.

Define the radical inverse function

$$\phi_b(n) = 0.d_0 d_1 \cdots d_j (\text{base b}) = \frac{d_0}{b} + \frac{d_1}{b^2} + \cdots + \frac{d_j}{b^{j+1}},$$

which provides a 1-1 map between the set of natural numbers and the set of rational numbers in $[0, 1)$. The van der Corput sequence in base b is then the sequence: $\phi_b(n)$ for $n = 0, 1, \ldots$.

To obtain $\phi_b(n+1)$ from $\phi_b(n)$ we need to add $\frac{1}{b} = 0.1$(base b) to $\phi_b(n)$ with a rightward carry, *i.e.* a carry in the opposite direction than that in the usual addition.

The s-dimensional Halton sequence is defined as

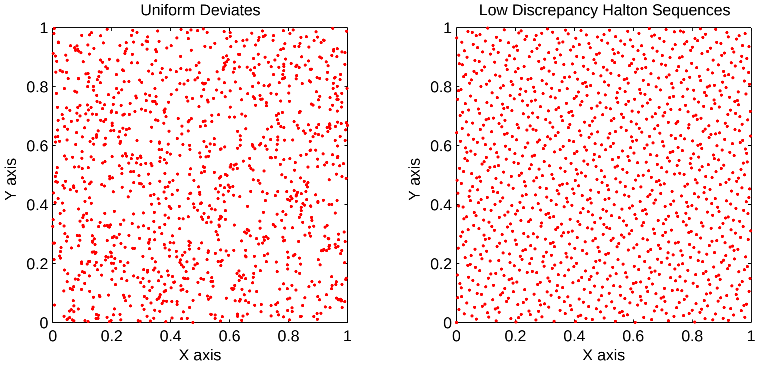
$$\mathbf{x}_n = (\phi_{b_1}(n), \phi_{b_2}(n), \cdots, \phi_{b_s}(n)), \quad n = 0, 1, \cdots,$$

where $b_1, b_2, \cdots, b_s$ are pairwise prime numbers greater than 1. They are often chosen to be the first s prime numbers starting from 2.

For example in 2 dimension, we use $s_1 = 2$ and $s_2 = 3$ to generate the Halton sequence. The first 10 points of the sequences are

$$(0,0), (\frac{1}{2}, \frac{1}{3}), (\frac{1}{4}, \frac{2}{3}), (\frac{3}{4}, \frac{1}{9}), (\frac{1}{8}, \frac{4}{9}),$$
$$(\frac{5}{8}, \frac{7}{9}), (\frac{3}{8}, \frac{2}{9}), (\frac{7}{8}, \frac{5}{9}), (\frac{1}{16}, \frac{8}{9}), (\frac{9}{16}, \frac{1}{27})$$

We now generate 1000 quasi-random points in the 2-dimensional Halton sequence and plot them in the following diagram. For comparison reasons, points generated by pseudo-random number generators are also shown. We see that the usual pseudo-random numbers have a much stronger clustering effect than numbers from the Halton's sequence. We also check and find that the numbers from the Halton sequence have a much smaller variance than from the pseudo-random numbers. Recall that the theoretical value for the variance should be $1/12 \approx 0.8333333$.


Uniform Deviates

Low Discrepancy Halton Sequences

4. The N-element Hammersley Point Set

The N-element Hammersley point set in the bases $b_1, \cdots, b_{s-1}$, each $b_i \geq 2$, is defined for $n = 0, 1, \cdots, N-1$ as

$$\mathbf{x}_n = (\frac{n}{N}, \phi_{b_1}(n), \phi_{b_2}(n), \cdots, \phi_{b_{s-1}}(n)).$$

For example, the 2-dimensional 10-element Hammersley point set contains

$$(0,0), (\frac{1}{10}, \frac{1}{2}), (\frac{1}{5}, \frac{1}{4}), (\frac{3}{10}, \frac{3}{4}), (\frac{2}{5}, \frac{1}{8}),$$
$$(\frac{1}{2}, \frac{5}{8}), (\frac{3}{5}, \frac{3}{8}), (\frac{7}{10}, \frac{7}{8}), (\frac{4}{5}, \frac{1}{16}), (\frac{9}{10}, \frac{9}{16})$$

5. More General Halton Sequences

Since sequences used in quasi-Monte Carlo are deterministic, it is not possible to use the Central Limit Theorem for error estimation as is done in Monte Carlo. It is desirable to develop techniques which combine the potential high accuracy of quasi-Monte Carlo with the practical error estimation ability of Monte Carlo. By introducing some randomness in low discrepancy sequences, one can use statistical methods for error analysis. Confidence interval can be obtained using the Central Limit Theorem.

Any number $x \in [0, 1]$ can be written as

$$x = \sum_{k=0}^{\infty} \frac{u_k}{b^{k+1}},$$

where u_k is an integer in $0, 1, \cdots, b - 1$. Define the rightward carry addition $\oplus$ so that

$$x \oplus \frac{1}{b} = \frac{u_m + 1}{b^{m+1}} + \sum_{k>m} \frac{u_k}{b^{k+1}},$$

where m is the first k in which $u_k \neq b - 1$.

For example,

$$(0.11101101)_2 \oplus \frac{1}{2} = (0.00011101)_2$$

$$(0.2210012)_3 \oplus \frac{1}{3} = (0.0020012)_3.$$

Next, the Von Neumann-Kakutani transformation of x is defined to be

$$T_b(x) = x \oplus \frac{1}{b}.$$

The operator T_b is called the Von Neumann-Kakutani transformation. It is a piecewise linear mapping from the interval $[0,1]$ to itself.

This transformation is then used to define a sequence $x_n, n = 0, 1, \cdots$ recursively:

$$x_{n+1} = T_b(x_n),$$

with a given starting point $x_0 \in [0,1]$. Let us denote the n-th iteration of T_b by T_b^n:

$$T_b^n(x) = \begin{cases} x, & n = 0 \\ T_b(T_b^{n-1}(x)), & n \geq 1. \end{cases}$$

Note that if $x_0 = 0$, the sequence is precisely the van der Corput sequence. However if $x_0 = 0.d_0 d_1 \cdots d_j$(base b), and so if we denote the integer $m_0 = d_j \cdots d_1 d_0$ in base b, then we have $x_0 = \phi_b(m_0)$ and $T_b^n(x_0) = \phi_b(m_0 + n)$ for $n = 1, 2, \cdots$. The sequence $T_b^n(x_0)$ for $n = 0, 1, \cdots$ is the original van der Corput sequence skipping the first

m_0 terms.

6. Generalization to Higher-Dimensions

The above discussion can be generalized straightforwardly to higher dimensions. If $b_1, b_2, \cdots, b_s$ are relative prime numbers greater than 1, then we can let $\mathbf{x}_0 = (x_{0,1}, x_{0,2}, \cdots, x_{0,s})$ define the sequence $\mathbf{x}_n = (x_{n,1}, x_{n,2}, \cdots, x_{n,s})$ for $n = 1, 2, \cdots$ by

$$x_{n+1,i} = T_{b_i}(x_{n,i}), \qquad i = 1, 2, \cdots, s.$$

In vector notation we write $\mathbf{x}_{n+1} = \mathbf{T}(x_n)$. The n-th iteration of $\mathbf{T}$ is denoted by $\mathbf{T}^n = (T^n_{b_1}, T^n_{b_2}, \cdots, T^n_{b_s})$

$\mathbf{x}_0$ is the start point of the sequence. The original Halton sequence results if $\mathbf{x}_0 = 0$. If $\mathbf{x}_0 = (x_{0,1}, x_{0,2}, \cdots, x_{0,s})$ with $x_{0,i} = 0.d_{0,i}d_{1,i} \cdots d_{k_i,i}$ (base b), then letting integer $m_{0,i} = d_{k_i,i} \cdots d_{1,i}d_{0,i}$ (base b), the sequence

$$\mathbf{T}^n(\mathbf{x}_0) = (\phi_{b_1}(m_{0,i} + n), \cdots, \phi_{b_s}(m_{0,s} + n)), \qquad n = 0, 1, \cdots,$$

is then the skipped Halton's sequence.

For any start point $\mathbf{x}_0$, the sequence $\mathbf{T}^n(\mathbf{x}_0)$, for $n = 0, 1, \cdots$, is a deterministic low discrepancy sequence. In actual computation with finite precision, each such sequence is actually a skipped Halton sequence

7. Random-Start Halton Sequence

Let $\mathbf{x}_0 = (x_{0,1}, x_{0,2}, \cdots, x_{0,s})$ be a random vector with uniform distribution on $[0,1]^s$. The sequence $\mathbf{x}_n \in [0,1]^s$ defined by

$$\mathbf{x}_n = \mathbf{T}^n((x_0) = (T_{b_1}^n(x_{0,1}), \cdots, T_{b_s}^n(x_{0,s})), \quad n = 0, 1, \cdots,$$

is called a random-start Halton sequence. Each random-start Halton sequence is a low discrepancy sequence. Such randomization preserves the uniformity of the Halton sequence. Multiple sequences with randomized start point will be used for error estimation.

8. Further Remarks

1. Very often the uniformity of the various low discrepancy sequences can be reduced further by permuting the digits in the digits expansion.
2. There are other low discrepancy sequences, such as those of Sobol and Faure.
3. More recent ones are based on the so called$(t, m, s)-$nets and (t, s)-sequences and the digital method for their construction.[3]

References

[1] J. G. van der Corput, *Verteilungsfunktionen* I, II, Nederl. Akad. Wetensch. Proc. Ser. B, 38, p. 813-821, 1058-1066, 1935. 4

[2] J. H. Halton, *Numerische Mathematik*, vol. 2, p. 84-90, 1960.

[3] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, (SIAM, Philadelphia, 1992). 15

[4] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery *Numerical Recipes*, Second Edition, Sec. 7.7, p. 299 (Cambridge University Press, 1992).

[5] Ilya M. Sobol, *A Primer for the Monte Carlo Method*, (CRC Press, 1994).

[6] D. E. Knuth, *The Art of Computer Programming: Vol. 2: Seminumerical Algorithms*, Second Edition, (Addison Wesley, 1981).

[7] J. Banks, J. S. Carson II, B. L. Nelson, and D. M. Nicol, *Discrete-Event System Simulation*, Third Edition, Ch. 8 (Prentice Hall, 2000).

[8] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, (McGraw-Hill, New York, 2000).

[9] P. Bratley, B. L. Fox and E. L. Schrage, *A Guide to Simulation*, (Springer-Verlag, New York, 1983).

[10] L. Devroye, *Non-Uniform Random Variate Generation*, (Springer-Verlag, New York, 1986).