An introduction to quasi-random numbers

By George Levy, Numerical Algorithms Group Ltd.

Introduction

Monte-Carlo simulation and random number generation are techniques that are widely used in financial engineering as a means of assessing the level of exposure to risk. Typical applications include the pricing of financial derivatives and senario generation in portfolio management. In fact many of the financial applications that use Monte-Carlo simulation involve the evaluation of various stochastic integrals which are related to the probabilities of particular events occurring.

A case in point is the pricing of a simple European option, where the value of a call option is $V_c = e^{-rT}\ \bar{E}[\max((S^T - X)),0)]$ while the value of a put is $V_p = e^{-rT}\ \bar{E}[\max((X - S^T),0)]$. Here X is the strike price, T is the maturity of the option, r is the risk free interest rate, S^T is the market value of the asset at maturity and $\bar{E}[]$ denotes the expectation operator.

The value of a European put is therefore, $V_p = e^{-rT} \int_{-\infty}^{\infty} p(S^T)\ \max((X - S^T),0)\ dS^T$ where $p(S^T)$ is the probability that the asset will have market value S^T at maturity.

If it is assumed that the value of the asset follows geometric Brownian motion and $p(S^T)$ is the lognormal distribution then the Black-Scholes formula [1] can be used to price the options as follows:

$$V_c = S^0\ N(d_1) - e^{-rT} X\ N(d_2),$$
$$V_p = -S^0\ N(-d_1) + e^{-rT} X\ N(-d_2)),$$

where

$$d_1 = (\log\ (S^0 / X) + (r - s^2/2)T)/(s\sqrt{T}),\ d_2 = d_1 - s\ \sqrt{T}$$

and $N(x) = \dfrac{1}{\sqrt{2p}} \int_{-\infty}^{x} e^{-x^2/2} dx$

where S^0 is the current value of the asset, s is the volatility of the asset, and $N(x)$ is the cumulative standard normal distribution.
In many cases however, the assumptions of constant volatility and a lognormal distribution for S^T are quite restrictive. Real financial applications may require a variety of extensions to the standard Black-Scholes model. Common requirements are for: non-lognormal distributions, time varying volatilities, caps, floors, barriers etc. In these circumstances it is often the case that there is no closed form solution to the problem. Monte-Carlo simulation can then provide a very useful means of evaluating the required integrals.

Monte-Carlo Integration

When we evaluate the integral of a function, $f(x)$, in the s-dimensional unit cube, I^S, by the Monte-Carlo method we are in fact calculating the average of the function at a set of randomly sampled points. This means that each point adds linearly to the accumulated sum that will become the integral and also linearly to the accumulated sum of squares that will become the variance of the integral.

When there are N sample points the integral is:

$$n = \frac{1}{N}\sum_{i=1}^{N} f(x^i)$$

where n is used to denote the approximation to the integral and $x^1, x^2, \ldots, x^N$ are the N, s-dimensional, sample points.
If a pseudo-random number generator is used the points x^i will be (*should be*) independently and identically distributed. From standard statistical results [2] we can then estimate the expected error of the integral as follows:

If we set $c^i = f(x^i)$ then since x^i is independently and identically distributed c^i is also independently and identical distributed. The mean of c^i is n and the variance is $Var(c^i) = \Delta^2$. It is a well known statistical property that the variance of n is given by $Var(n) = \frac{\Delta^2}{N}$. We can therefore conclude that the estimated integral n has a standard error of $\Delta N^{-1/2}$. This means that the estimated error of the integral will decrease at the rate of $N^{-1/2}$.

Is it possible to achieve a better convergence than this? If sample points are chosen that lie on a Cartesian grid and we sample each grid point exactly once then the Monte-Carlo method effectively becomes a deterministic quadrature scheme, whose fractional error decreases at the rate of N^{-1} or faster. The trouble with the grid approach is that it is necessary to decide in advance how fine it should be, and all the grid points need to be used. It is therefore not possible to sample until some convergence criterion has been met.

Quasi-random number sequences seek to bridge the gap between the flexibility of pseudo-random number generators and the advantages of a regular grid. They are designed to have a high level of uniformity in multidimensional space, but unlike pseudo-random numbers they are not statistically independent.

Quasi-random sequences

Quasi-random numbers are also called low discrepancy sequences. The discrepancy of a sequence is a measure of its uniformity and is defined as follows:

Given a set of points $x^1, x^2, \cdots, x^N \in I^S$ and a subset $G \subset I^S$, define the counting function $S_N(G)$ as the number of points $x^i \in G$. For each $x = (x_1, x_2, \ldots, x_s) \in I^S$, let G_x be the rectangular s-dimensional region

$G_x = [0, x_1) \times [0, x_2) \times \cdots \times [0, x_S)$ with volume $x_1 x_2, \ldots, x_N$. Then the discrepancy of the points $x^1, x^2, \ldots, x^N$ is given by:

$$D_N^*(x^1, x^2, \ldots, x^N) = \sup_{x \in I^S} | S_N(G_x) - N x_1 x_2, \cdots, x_S | .$$

The discrepancy is therefore computed by comparing the actual number of sample points in a given volume of multidimensional space with the number of sample points that *should be* there assuming a uniform distribution.

It can be shown that the discrepancy of the first N terms of quasi-random sequence has the form:

$$D_N^*(x^1, x^2, \ldots, x^N) \le C_S (\log N)^S + O((\log N)^{S-1}) \text{, for all } N \ge 2$$

The principal aim in the construction of low-discrepancy sequences is thus to find sequences in which the constant C_S is as small as possible. Various sequences have been constructed to achieve this goal. Here we consider the following quasi-random sequences:

- Niederreiter [3]
- Sobol [4]
- Faure [5]

The results of using NAG random number generator software [6] with GenStat graphics [7] is shown below. Figures 1-3 illustrate the visual uniformity of the sequences. They were created by generating 1000, 16-dimensional sample points, and then plotting the 4th dimensional component of each point against its 5th dimensional component.

In Figure 1, it can be seen that the pseudo-random sequence exhibits clustering of points, and there are regions with no points at all.

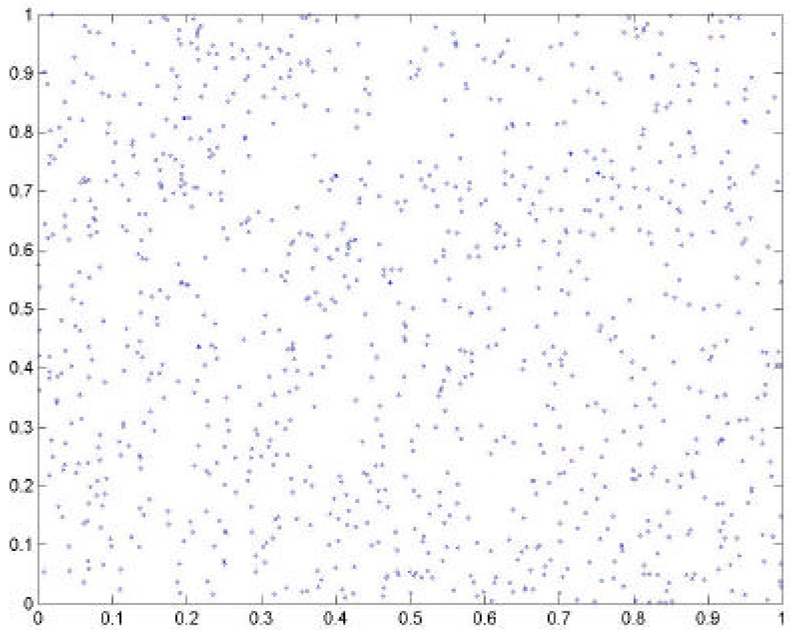

Figure 1: Pseudo-random sequence points.

Visual inspection of Figure 2 and Figure 3 show that both the Sobol and Niederreiter quasi-random sequences appear to cover the area more uniformly.

It is interesting to note that the Sobol sequence appears to be a structured lattice which still has some gaps. The Niederreiter sequence on the other hand appears to be more irregular and covers the area better. However, we can't automatically conclude from this that the Niederreiter sequence is the best. This is because we haven't considered all the other possible pairs of dimensions.

Perhaps the easiest way to evaluate the random number sequences is to use them to calculate an integral.

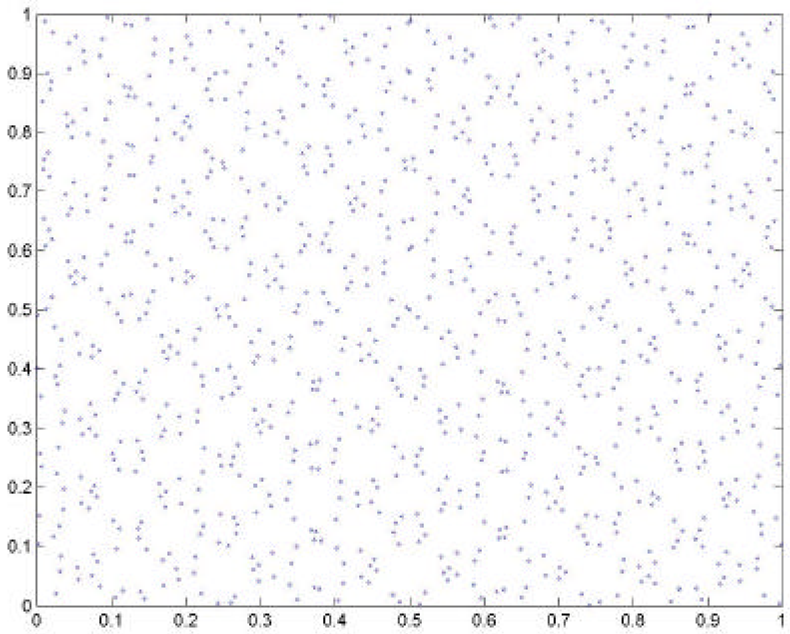

Figure 2: Sobol sequence points.

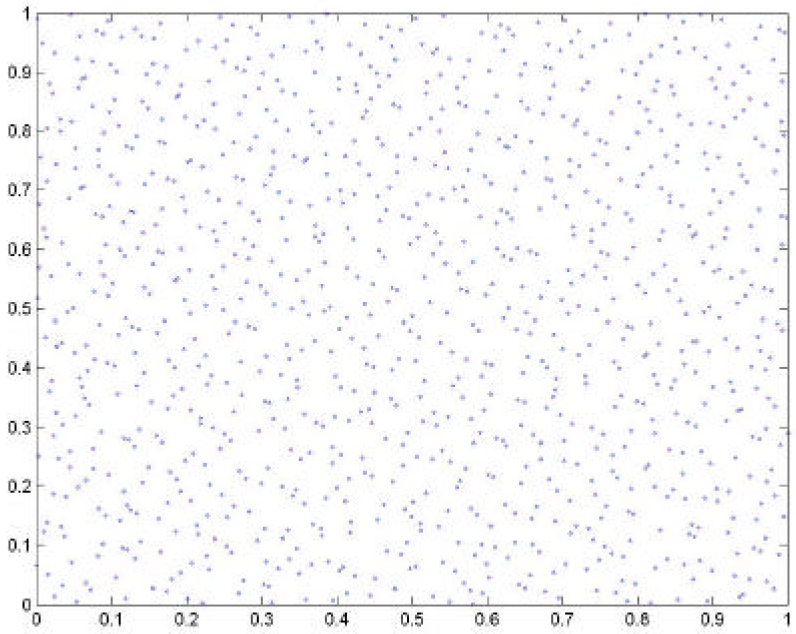

Figure 3: Niederreiter sequence points.

In Figure 4 Monte-Carlo results are presented for the calculation of the six dimensional integral:

$$I = \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \prod_{i=1}^{6} (i \ \cos(ix_i))\, dx_1\ dx_2\ dx_3\ dx_4\ dx_5\ dx_6$$

The exact value of this integral is:

$I = \prod_{i=1}^{6} \sin(i)$, which for $i = 6$, gives $I = -0.0219$

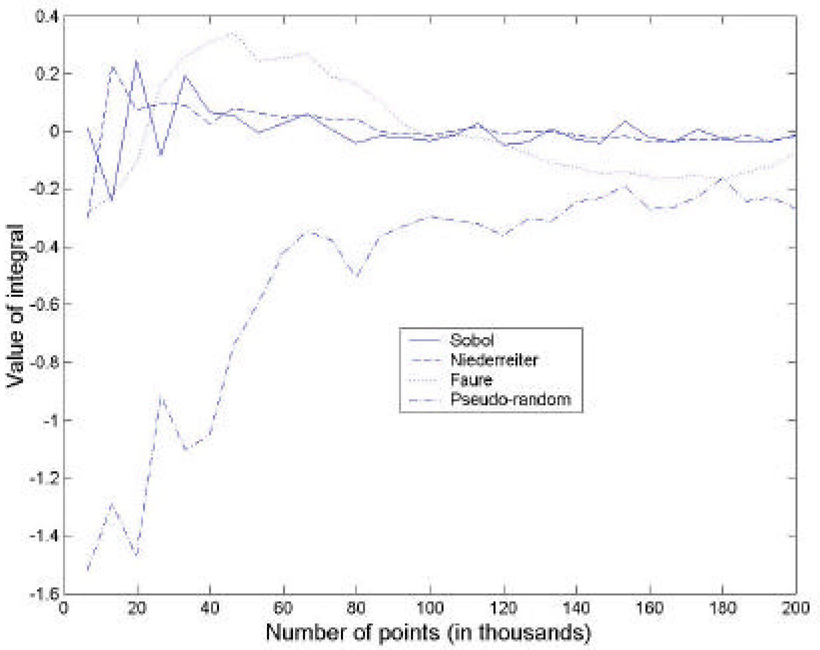

Figure 4: Monte Carlo integration using random numbers.

It can be seen that the pseudo-random sequence gives the worst performance. But as the number of points increases its approximation to the integral improves. Of the quasi-random sequences it can be seen that the Faure sequence has the worst performance, whilst both the Sobol and Neiderreiter sequences give rapid convergence to the solution.

To conclude it has been shown that quasi-random sequences can evaluate integrals more efficiently than pseudo-random sequences. They thus provide financial engineers with a very useful technique for risk assessment.

George Levy works at NAG Ltd UK, he can be contacted at george@nag.co.uk.

References

[1] Hull., J. C. *Options, Futures and other Derivatives*, Prentice Hall International Inc, 3rd Edition 1997.

[2] Goldberger, A. S. *A course in Econometrics*, Havard University Press, 1997.

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

[4] Sobol, I. M. The distribution of points in a cube and the approximate evaluation of integrals, *USSR Comput. Math. Math. Phys.* 7, 4, 86-112, 1967.

[5] Faure, H. Discrepance de suites associees a un systeme de numeration (en dimension s). *Acta Arith.* 41, 337-351 1982.

[6] NAG Ltd, *The Fortran 77 Library Mark 20*, NAG Ltd, Oxford, UK, 2001.

[7] *GenStat*, VSN International Ltd, Oxford, UK.