Seoul Journal of Business
Volume 12, Number 1 (June 2006)

Quasi Random Sampling for Operations Management

Hongsuk Yang*
University of Utah
Utah, U.S.A.

Abstract

We look at the benefits of using a kind of quasi-random numbers to obtain more accurate results for a given number of simulation runs. We explore a sampling method with enhanced independence in multi-dimensional simulations by combining the ideas of stratified sampling and Latin Hypercube sampling. We test the new sampling method by comparing it with traditional stratified sampling and Latin Hypercube sampling applied to various operations management problems.

Keywords: quasi random sampling, stratified latin hypercube sampling, quasi monte carlo simulation

INTRODUCTION

If more than a few random factors such as demands, foreign exchange rate, or production yield, are involved, problems in operations management area often may not be solvable with an analytic approach. In those cases, we often conduct numerical analysis. Because of general applicability and ease of implementation, Monte Carlo simulation is often used to analyze complex operations problems. One can get more accurate results by letting the simulation run longer. There is the ubiquitous $1/\sqrt{n}$ law of statistical variation where n is the number of

* Assistant Professor, David Eccles School of Business, University of Utah (mgthy@business.utah.edu).

simulation runs (scenarios). That is, to reduce the standard error of an estimator by a factor of f, one needs to increase in the size of the experiment, or the number of simulation runs by f^2 (Law and Kelton 2000). Therefore, the amount of computer time required for a satisfactory accuracy may be embarrassingly long. There are a variety of "variance reduction" techniques available to increase the accuracy with a given amount of simulation. We discuss a general form of these techniques, quasi-random sampling. "Quasi" random sampling in a broad sense covers all variance reduction techniques that artificially manipulate the sampling procedure.

According to Judd (1998), Weyl found that infinite sequences of non-random points exist, which have a property similar to random sequences. Such sequences are said to be uniformly distributed (or equi-distributed) in the number theoretic sense. The points from these sequences are often called quasi-random in the narrow sense of the term. Several algorithms such as described by Haber, Niederreiter, Baker, Halton and Sobol are also available to generate quasi random numbers (Judd, 1998).

The ex ante uniformity of the samples makes the Monte Carlo estimate of an integral unbiased. We would like to develop ex post relatively uniform sequences even with the small number of scenarios because a small bias in sampling may result in huge distortion in optimal decision. Although there are well known equi-distributed sequences due to Wyle, Haber, Niederreiter, Baker and Sobol, they do not necessarily generate uniformly distributed samples if the number of scenarios is small and the dimensions are large. Thus we develop a quasi-random number generating procedure that is easy to understand and works well with small samples. This program generates uniform random real numbers between 0 and 1. We may generate any desired distributions by inversion method or other techniques (Ripley 1987).

In this paper, we look at the benefits of stratified Latin Hypercube sampling, the combination of two traditional variance reduction techniques: stratified sampling and Latin Hypercube sampling. First, we look at the various variance reduction sampling techniques in the next section. Then, we illustrate the benefit of stratified Latin Hypercube sampling in mean estimation compared with other variance techniques. Finally, we compare the performance of stratified Latin Hypercube with other sampling methods in several operations management problems.

METHODS FOR MONTE CARLO SIMULATION

Mathematically, Monte Carlo simulation is concerned with approximating the value of the integral of a function *g*( ) over a *d* dimensional unit hypercube (Robert and Casella 1999):

$$h = E(g(u)) = \int_{u_1=0}^{u_1=1} \Lambda \int_{u_d=0}^{u_d=1} g(u_1, \Lambda, u_d) du_d \quad \Lambda \quad du_1$$

Monte Carlo approximates *h* by

$$\hat{h} = \sum_{i=1}^{n} g(u_{i1}, \Lambda, u_{id}) / n$$

We restrict our attention to situations where *d* is known in advance. In terms of a physical process being simulated, it might be the behavior of a stock price over *d* periods. We might: 1) generate *d* independent uniform random variables in (0, 1); 2) these would be transformed into *d* lognormal random variables to represent relative stock price changes for each period; 3) if we are interested in the value of a “Asian option”, the *g*( ) function would be the amount by which the average price over the periods exceeds a strike price. The integral computes the expected value.

Variance reduction methods, including quasi-random numbers, are concerned with how to judiciously choose the s samples, $\{u_{i1}, \ldots, u_{id}\}$, so that the estimator has low variance.

We concentrate in particular on a method called Stratified Latin sampling and show that it tends to have low variance relative to 1) simple random sampling, 2) number theoretic quasi-random methods such as Weyl’s method, 3) antithetic variates, 4) traditional stratified sampling, 5) importance sampling, and 6) Latin Hypercube sampling. First we look at the each sampling method.

Simple random sampling

Random sampling selects points randomly in the d dimensional hypercube. Each point is independent of all others. Despite its inefficiency, simple random sampling without any variance reduction effort is widely used in research and commercial software.

Low discrepancy, Weyl sequence sampling

In number theory, the problem of distributing points as uniformly as possible over a d-dimensional unit cube has been studied. The uniformity of a point distribution can be measured by "discrepancy" (Niederreiter 1978). A formal definition of discrepancy is as follows:

For n points, x_1, K, x_n in the d-dimensional half-open unit cube $I^d = [0,1)^d$, $n \geq 1$, and subinterval J of I^d, we put $D(J;n) = \frac{A(J;n)}{n} - V(J)$, where $A(J;n)$ is the number of k, $1 \leq k \leq n$, with $x_k \in J$ and $V(J)$ is the volume of J. Then the "discrepancy" $\Delta(n)$ of the points x_1, K, x_n is defined by $D(n) = \sup_J |D(J,n)|$, where the supremum is extended over all half-open subintervals J.

The smaller discrepancy, the more uniform the distribution. Finite point sets in multi dimensional unit cube with a nearly uniform distribution are called low discrepancy sets. Low discrepancy points are of interest because of multi dimensional numerical integration. Thus, the construction of point set with as small discrepancy as possible is a central issue in quasi Monte Carlo simulation.

The simplest example of equi-distributed sequence is Weyl sequence (Judd 1998): $x_n = n\theta - [n\theta]$, where $n = 1, 2, \ldots$, for θ irrational.

Antithetic variates

We say U_1 and U_2 are antithetic if $U_2 = 1 - U_1$. Some authors (e.g., Bratley, Fox and Schrage 1989) use the less restrictive definition that U_1 and U_2 are antithetic if $Cov[U_1, U_2] < 0$. If g is monotone, then it generates negative correlation for $g(U)$ and $g(1 - U)$ and reduces variance in estimation. But it does not deal with multi dimensional aspects for the problems with multiple dimensions.

Stratified sampling

Classic examples of stratification arise in the design of sample surveys (Raghavarao 1971). Stratified sampling is applicable to multiple dimensions. By partitioning the d-dimensional unit hypercube into l equal-width intervals in each d dimension, we have l^d equal volume mini-hypercubes or strata, each having length $1/l$ in each dimension. Stratified sampling chooses one point randomly from each stratum or mini-hypercube.

Importance sampling

Importance sampling is related to weighted stratified sampling. Suppose we wish to estimate

$$h = E(g(u)) = \sum_{i=1}^{m} g(u \mid A_i)P(A_i),$$

where the events A_i are generated during a simulation and $g(u|A_i)$ is the expected value given that A_i occurs. In many systems there are some I for which $g(u|A_i)$ is large while simultaneously $P(A_i)$ is small. For example, for barrier options, it may be a rare event that the underlying stock price is exceptionally high or low and passes the barrier. Crossing a barrier, however, may have such a huge impact on payoff for the option holder that when it does occur it dominates all other payoffs. Similar examples can be found in inventory management with huge shortage penalties.

If we simulate in a standard way, the required time to obtain a reasonable estimate of h may be prohibitive. An alternative is to sample from a different distribution so that we can sample rare events more frequently, applying a correction factor to compensate. That is, suppose we simulate a model with known nonzero cdf V to u rather than correct uniform cdf U before inverse transform. Then if we let $\phi = g(\frac{U}{V})$,

$$\hat{h}_\phi = \frac{1}{n}\sum_{i=1}^{n}\phi(u_i)$$

is unbiased estimator of h.

This is weighted mean of $g(u)$ which weights inversely proportional to the "selection factor" U/V. Then

$$Var(\hat{h}_\phi) = \frac{1}{n}\int[\phi(u)-h]^2 dV(u) = \frac{1}{n}\int[\frac{gU}{V}-h]^2\, dV(u)$$

which can be small when $\frac{gU}{V}$ is nearly constant.

For many distributions, such as the normal distribution, much of the probability is concentrated around the mean, thus, many of the U_i's around the median will transform into essentially the same demand. A simple implementation of importance sampling, if we have 9 scenarios for a two-dimensional problem, is to partition the uniform unit hypercube with different length according to the size of variance. Since we expect higher variance in the tail parts of normal distribution, we have more samples in the extreme values in the uniform distribution before the inverse transformation.

Example: importance sampling with univariate normal distribution

If the output is a Normal distribution and we are allowed n samples, $n > 2$, in one dimension one of the fundamental questions on importance sampling may be how should the interval (0, 1) be partitioned so that the weighted estimator has minimum variance (and of course is unbiased). For simplicity, let a random variable x follow Normal distribution with mean zero and variance one.

Let $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$, $\Phi(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})dx$

and $\gamma(x) = \frac{1}{\sqrt{2\pi}}x\ \exp(-\frac{x^2}{2})$.

Sample size 3. Suppose we have partitioning points $t_1 \le t_2$ and three samples ($x_1 \le t_1 < x_2 \le t_2 < x_3$). Then the mean and variance for the first interval are

$$E(x_1) =: \mu_1 = \frac{1}{\Phi(t_1)} \int_{-\infty}^{t_1} x \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx$$

$$= \frac{1}{\Phi(t_1)} \left(-\frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) \bigg|_{-\infty}^{t_1} \right) = \frac{1}{\Phi(t_1)} \left(-\frac{1}{\sqrt{2\pi}} \exp(-\frac{t_1^2}{2}) \right)$$

$$= -\frac{\phi(t_1)}{\Phi(t_1)}$$

$$Var(x_1) =: \sigma_1^2 = \frac{1}{\Phi(t_1)} \int_{-\infty}^{t_1} x^2 \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx - m_1^2 .$$

$$= \frac{1}{\Phi(t_1)} \left[-\frac{1}{\sqrt{2\pi}} x \exp(-\frac{x^2}{2}) \bigg|_{-\infty}^{t_1} + \int_{-\infty}^{t_1} \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx \right]$$

$$- \left(-\frac{\phi(t_1)}{\Phi(t_1)} \right)^2$$

$$= \frac{1}{\Phi(t_1)} \left[-\frac{1}{\sqrt{2\pi}} t_1 \exp(-\frac{t_1^2}{2}) + \Phi(t_1) \right] - \left(-\frac{\phi(t_1)}{\Phi(t_1)} \right)^2$$

$$= 1 - \frac{\gamma(t_1)}{\Phi(t_1)} - \left(\frac{\phi(t_1)}{\Phi(t_1)} \right)^2$$

Similarly, the mean and variance for the second interval are

$$E(x_2) =: \mu_2 = -\frac{\phi(t_2) - \phi(t_1)}{\Phi(t_2) - \Phi(t_1)}$$

$$Var(x_2) =: \sigma_2^2 = 1 - \frac{\gamma(t_2) - \gamma(t_1)}{\Phi(t_2) - \Phi(t_1)} - \left(\frac{\phi(t_2) - \phi(t_1)}{\Phi(t_2) - \Phi(t_1)} \right)^2 .$$

Finally, the mean and variance for the third interval are

$$E(x_3) =: \mu_3 = -\frac{\phi(t_2)}{1 - \Phi(t_2)}$$

$$Var(x_3) =: \sigma_3^2 = 1 + \frac{\gamma(t_2)}{1-\Phi(t_2)} - \left(\frac{\phi(t_2)}{1-\Phi(t_2)}\right)^2 .$$

Define: P_i = probability that a random sample would fall in the k-th interval.

That is, $P_1 = \Phi(t_1)$, $P_2 = \Phi(t_2) - \Phi(t_1)$ and $P_3 = 1 - \Phi(t_2)$

Let our estimator of the sample mean be $\bar{x} = P_1x_1 + P_2x_2 + P_3x_3$. Then

$$E(\bar{x}) = \sum_{i=1}^{3} P_i m_i = \Phi(t_1)m_1 + (\Phi(t_2) - \Phi(t_1))m_2 + (1-\Phi(t_2))m_3 = 0$$

Therefore this sampling is unbiased (Casella and Berger 1990).

The variance of the above estimator is

$$\begin{aligned}
Var(\bar{x}) &= \sum_{i=1}^{3} P_i^2 \sigma_i^2 \\
&= P_1^2 + P_2^2 + P_3^2 - P_1\gamma(t_1) - P_2\{\gamma(t_2) - \gamma(t_1)\} + P_3\gamma(t_2) \\
&\quad - (\phi(t_1))^2 - (\phi(t_2) - \phi(t_1))^2 - (\phi(t_2))^2 \\
&= \Phi(t_1)^2 + \{\Phi(t_2) - \Phi(t_1)\}^2 + \{1-\Phi(t_2)\}^2 \\
&\quad - \Phi(t_1)\gamma(t_1) - \{\Phi(t_2) - \Phi(t_1)\}\{\gamma(t_2) - \gamma(t_1)\} \\
&\quad + [1-\Phi(t_2)\}\gamma(t_2) - (\phi(t_1))^2 - (\phi(t_2) - \phi(t_1))^2 - (\phi(t_2))^2 \\
&= \Phi(t_1)^2 + \{\Phi(t_2) - \Phi(t_1)\}^2 + \{1-\Phi(t_2)\}^2 \\
&\quad - 2\Phi(t_1)\gamma(t_1) - 2\Phi(t_2)\gamma(t_2) + \Phi(t_1)\gamma(t_2) + \Phi(t_2)\gamma(t_1) \\
&\quad + \gamma(t_2) - (\phi(t_1))^2 - (\phi(t_2) - \phi(t_1))^2 - (\phi(t_2))^2
\end{aligned}$$

Even though we do not have analytic solution for t_1 to minimize $Var(\bar{x})$, we may find a solution numerically using MATLAB. When $t_1 = -.5689$ (equivalently, $u_1 = .2847$ in (0,1) sampling interval), $Var(\bar{x}_{imp})$ is minimized with its value of .0609 with $Stdev(\bar{x}_{imp}) = .24678$.

Typical random sampling gives $Var(\bar{x}) = .3333$ and $Stdev(\bar{x}) = .57735$. If we apply stratified sampling with equal length partition (.3333 and .6667 in (0,1) sampling interval), the value of $Var(\bar{x}_{strat})$ is .0689 and $Stdev(\bar{x})$ = .26249.

Remark: $Var(\bar{x}_{start}) = Var(\bar{x}_{Random}) - \dfrac{1}{n^2}\sum_{i=1}^{n}(\mu_i - \mu)^2$

$$\sigma^2 = \int_{-\infty}^{\infty}(x-\mu_i+\mu_i-\mu)^2\,\phi(x)dx$$
$$= \int_{-\infty}^{\infty}(x-\mu_i)^2\,\phi(x)dx + \int_{-\infty}^{\infty}(\mu_i-\mu)^2\,\phi(x)dx$$
$$= \sum_{i=1}^{n}P_i\{\sigma_i^2 + (\mu_i-\mu)^2\}$$

Thus $Var(\bar{x}_{Random}) = \dfrac{1}{n}\left\{\sum_{i=1}^{n}P_i\sigma_i^2 + \sum_{i=1}^{n}P_i(\mu_i-\mu)^2\right\}$

If $P_i = \dfrac{1}{n}$, $Var(\bar{x}_{Random}) = \dfrac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 + \dfrac{1}{n^2}\sum_{i=1}^{n}(\mu_i-\mu)^2$ ■

Thus, if the output is Normal distributed, then importance sampling improves the precisions relative to simple equal partition stratified sampling by about 6% $(= \frac{.26249 - .24678}{.26249})$.

Sample size n > 4. When we extend the above approach to n sample case, we have similar results.

Let samples from each interval are $x_1 < x_2 < \Lambda < x_n$ and partition points are $t_1 < t_2 < \Lambda < t_{n-1}$.

The mean and variance expression of the first interval are the same as the three sample case.

The mean and variance of the *k*-th interval are

$$E(x_k) =: \mu_k = -\frac{\phi(t_k) - \phi(t_{k-1})}{\Phi(t_k) - \Phi(t_{k-1})}$$

$$Var(x_k) =: \sigma_k^2 = 1 - \frac{\gamma(t_k) - \gamma(t_{k-1})}{\Phi(t_k) - \Phi(t_{k-1})} - \left(\frac{\phi(t_k) - \phi(t_{k-1})}{\Phi(t_k) - \Phi(t_{k-1})}\right)^2.$$

The mean and variance of the last (*n*-th) interval are

$$E(x_n) =: \mu_n = -\frac{\phi(t_{n-1})}{1-\Phi(t_{n-1})}$$

$$Var(x_n) =: \sigma_n^2 = 1 + \frac{\gamma(t_{n-1})}{1-\Phi(t_{n-1})} - \left(\frac{\phi(t_{n-1})}{1-\Phi(t_{n-1})}\right)^2 .$$

Define P_i be probability that a random sample would fall in the k-th interval.

If our estimator of sample mean is $\bar{x} = \sum_{i=1}^{n} P_i x_i$, then

$$E(\bar{x}) = \sum_{i=1}^{n} P_i \mu_i = 0$$

Again, the estimator of the mean of n samples is unbiased.

Now the variance of the estimator is

$$\begin{aligned} Var(\bar{x}) &= \sum_{i=1}^{n} P_i^2 \sigma_i^2 \\ &= \Phi(t_1) + \{\Phi(t_2) - \Phi(t_1)\}^2 + \Lambda + \{1 - \Phi(t_{n-1})\}^2 \\ &\quad - \Phi(t_1)\gamma(t_1) - \{\Phi(t_2) - \Phi(t_1)\}\{\gamma(t_2) - \gamma(t_1)\} \\ &\quad - \Lambda + \{1 - \Phi(t_{n-1})\}\gamma(t_{n-1}) \\ &\quad - (\phi(t_1))^2 - (\phi(t_2) - \phi(t_1))^2 - \Lambda - (\phi(t_{n-1}))^2 . \end{aligned}$$

Although we may not have analytic solution for t_1, t_2, K, t_{n-1} to minimize $Var(\bar{x})$, we may find a good approximation by numerical search. The results for small number of samples are summarized in the table 1. It appears that as the partition increases, the precision increases more than proportionately.

The above analysis is for only one dimension. If our system is

Table 1. Partition Points for Importance Sampling

n	Partition points of (0,1): $\Phi(t_i)$	Var $(\bar{x}_{imp})$	$Stdev$ $(\bar{x}_{imp})$	Var $(\bar{x}_{strat})$	$Stdev$ $(\bar{x}_{strat})$	Var $(\bar{x}_{random})$	$Stdev$ $(\bar{x}_{random})$
3	.2847, .7153	.0609	.2468	.0689	.2625	.3333	.5773
4	.1838, .5, .8162	.0273	.1652	.0349	.1868	.2500	.5000
5	.1288, .3657, .6343, .8712	.0145	.1204	.0206	.1435	.2000	.4472

multidimensional, the above results still apply if the random variables in each dimension are independent Normal random variables and output measure is a linear function of the variables.

Latin hypercube sampling

Stratified sampling becomes less practical if the dimension d is large, because stratified sampling requires l^d samples. For example, if l is 2 (which is the smallest effective number), and d is 30, we need more than one billion samples. Latin Hypercube (LH) sampling can be used for any sample size specified in advance. LH is a one dimensional version of stratified sampling.

The LH method generates a sample by drawing independently between dimensions, but sampling without replacement from each of the l subintervals within a dimension. Roughly speaking, stratified sampling approximates the joint distribution of $g(\ \ )$ as closely as possible given the sample size, whereas LH sampling approximates the marginal distributions of $g(\ \ )$ as closely as possible.

According to Mckay, Beckman and Conover (1979), LH sampling is very good when the output is dominated by only a few dimensions. Stein (1985) tests the performance of Latin Hypercube with many random variables. To the best of our knowledge, LH sampling is the only variance reduction sampling method that has been in a commercial Monte Carlo software package for a number of years. The @RISK package from Palisade offers LH sampling as an option.

STRATIFIED LATIN HYPERCUBE SAMPLING

All the sampling methods in the previous section do not guarantee ex post uniformity if the sample size (simulation runs or the number of scenarios) is small. Stratified Latin Hypercube (SLH) sampling is best motivated by looking at the figure 1. Figure 1 illustrates the worst case of each sampling method with 4 samples in 2 dimensions. Typical random sampling could generate points that are concentrated together by chance. Compared with a typical random number generator, we can see

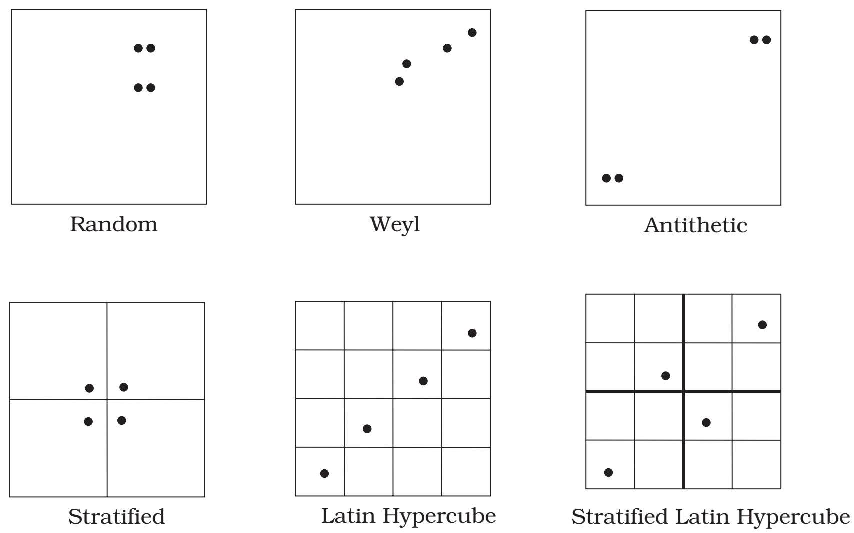

Figure 1. Worst case from each sampling (4 scenarios in 2 dimensions)

that points generated by LH sampling are more equally spaced even in the worst case. But samples may have strong positive correlation between two dimensions. Stratified sampling is better in terms of independence. For a given simulation, however, it need not generate samples that have equal marginal distributions.

We combine the idea of stratified sampling and Latin Hypercube sampling: keep marginal uniformity as in LH sampling and enhance the independence by sectioning as n-dimensional stratified sampling. SLH samples have more accurate marginal distributions than traditional stratified samples and have higher independence among dimensions than LH sampling. SLH sampling has been introduced as an option in the LINGO modeling system from LINDO systems (Schrage 1998a).

Variance of stratified latin hypercube

Let g_i be the result of the ith sample or scenario, for i = 1, Λ, n, that is,

$g_i = g(u_{i1}, \Lambda, u_{id})$.

The variance of our estimator of expected value will be $V = Var(\sum_{i=1}^{n} g_i / n)$. For all the methods we consider, except importance sampling, the g_i are identically (though not independently) distributed, so that $V = Var(g_1)/n + \sum_{i \neq j} Cov(g_i, g_j)/n^2 = Var(g_1)/n + 2\sum_{i>j} Cov(g_i, g_j)/n^2$.

For simple random sampling, $\sum_{i \neq j} Cov(g_i, g_j) = 0$. Each g_i is a draw from one of the l^d small cells of hypercubes in the d-dimension unit hypercube. For the quasi-random methods we want to argue that cells that are close together will not appear in the same sample. If we can argue that cells that are close are more positively correlated than two randomly selected cells, then we can argue that the summation over the remaining cells must result in $\sum_{i \neq j} Cov(g_i, g_j) < 0$.

For simple random sampling: $Var(\hat{h}_R) = \dfrac{1}{n} Var(h)$

$$Var(\hat{h}_S) = Var(\hat{h}_R) - \frac{1}{n^2} \sum_{i=1}^{n} (\mu_i - \mu)^2. \text{ } \textit{Thus } Var(\hat{h}_S) \le Var(\hat{h}_R).$$

From McKay, Beckman and Conover (1979),

$$Var(\hat{h}_{LH}) = Var(\hat{h}_R) + \frac{n-1}{n} \frac{1}{n^d (n-1)^d} \sum_{region} (\mu_i - \mu)(\mu_j - \mu)$$

Equivalently, $Var(\hat{h}_{LH}) = Var(\hat{h}_R) + \dfrac{n-1}{n} Cov(g_i, g_j)$.

They showed that for monotone functions since $\frac{1}{n^d (n-1)^d} \sum_{region} (\mu_i - \mu)(\mu_j - \mu) < 0$, $Var(\hat{h}_{LH}) \le Var(\hat{h}_R)$. Stein (1987) also showed for any function $Cov(g_i, g_j) \le 0$ asymptotically if the second moment of g is finite.

It is true that crude Monte Carlo has an advantage with regard to simple estimation of the variance of the estimator. With correlated schemes such SLH one has two options: either estimate the covariances (usually hard), or do several replications. The latter option, of course, implies extra computational effort. A third alternative may be to claim that the naive estimate of the variance of SLH based estimator is conservative. The argument might go something like this:

1) For a wide range of functions, e.g. continuous and monotonic we should be able to argue that the SLH estimator has lower variance than the simple random draws estimator (unfortunately, we do not know what it is).

2) Because the SLH samples are negatively correlated, the simple sample variance, disregarding the covariance terms (which we expect to be negative in SLH) will be higher than the sample variance from SLH sampling.

3) Putting (1) and (2) together, we get a conservative estimate of the variance of the SLH-based estimator.

We compare the performance of SLH with other sampling methods in several operations management problems in practical sense in the later section.

Stratified latin hypercube and importance sampling

We also introduce the combination of stratified Latin Hypercube sampling and importance sampling (lower right of figure 2). We split a unit interval (0, 1) to three unequal intervals (0, .25), (.25, .75) and (.75, 1). In comparison, we split a unit interval to equal size in stratified sampling. We test the

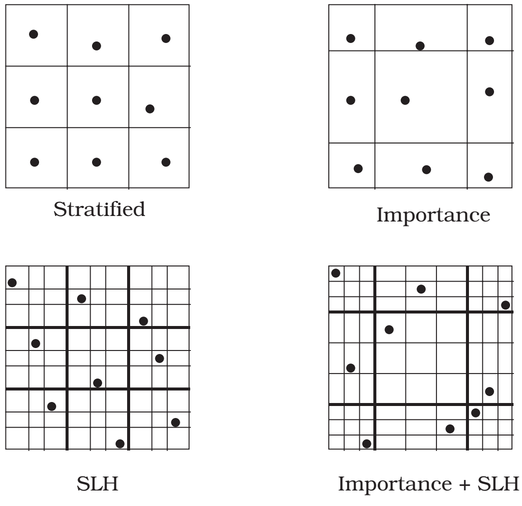

Figure 2. Stratified sampling vs. importance sampling

performance of the combination of stratified Latin Hypercube sampling and importance sampling with Petroleum Reserve Estimation problem in the later section. The result is summarized in the table 4. While the stratified Latin Hypercube sampling has the lowest variance of simulation for the petroleum reserve problem, the combination of importance sampling and stratified Latin Hypercube has the lowest bias in objective value.

APPLICATION TO OPERATIONS PROBLEMS

If an analytic solution is available, we can compare the performance of different sampling methods directly with the correct answer. If an analytic solution is not available, we can still compare the performance by: (1) standard error of simulations, given a number of scenarios or (2) convergence rate as we increase the number of scenarios.

Stochastic PERT

The PERT network example in this paper is based on Schrage (1998a).

We assume a triangular distribution for the time of each task. The minimum for each task is half of the mode and the maximum is double of the mode. The modes of triangular distribution for design, forecast, survey, price, schedule, costout and train are 10, 14, 3, 3, 7, 4, and 10, respectively.

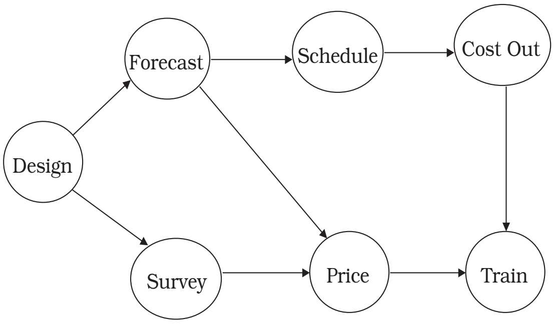

Figure 3. PERT network

Table 2. Simulation comparison for stochastic PERT (r = 1000, n = 64, d = 7)

	Random	Weyl	Antithetic	Stratified	LH	SLH
Mean	52.5296	52.2809	52.4972	52.4881	52.4998	52.5002
Std Error	0.5775	1.2354	0.0777	0.3455	0.0086	0.0080

This PERT network can be written as the following functional form:

Time = Design + max (max (Forecast, Survey) + Price, Forecast + Schedule + Costout) + Train.

Thus, the function to evaluate is a combination of sum and max functions. Since Latin Hypercube sampling and Stratified Latin Hypercube sampling perform well in sum and max functions, we may expect LH and SLH sampling to perform well for PERT. The result is summarized in the table 2. If expected activity times are used, the project length is 45. Thus, stochastic variation increased the expected project length by about 7 time units.

Petroleum reserve estimation

Murtha (2000) introduces the use of Monte Carlo simulation to the estimation of volumetric reserve, N = AhR, where A is area (in acre), h is height (in feet) and, R is STB per acre-feet. Each dimension follows a triangle distribution: A ~ triangle (1000, 2000, 4000), h ~ triangle (20, 50, 100), and R ~ triangle (80, 120, 200). The dimensions are assumed to be independent each other. Murtha conducted one simulation with 500 iterations (scenarios) using @RISK software with Latin Hypercube methods. We conducted 1,000 simulations with 8, 64 and 512 scenarios to compare sampling methods. The first three moments (mean, variance and skewness) of the distribution of the petroleum reserve may be of interest. Higher mean (expectation) and lower variance (risk) are preferred. Positive skewness may be preferred to avoid downside risk.

The comparison results with 64 scenarios are summarized in the table 3 where the numbers are scaled down by 10^6. SLH sampling outperforms the other sampling methods in the petroleum reserve estimation problem.

Table 3. Petroleum reserve estimation (r = 1000, n = 64, d = 3)

	Random	Weyl	Antithetic	Stratified	LH	SLH
Mean	17.6422	17.3686	17.6229	17.6367	17.6166	17.6304
Std Error	1.5282	1.2444	0.5900	0.6643	0.3904	0.2333
Variance	64.0433	60.0588	64.6739	65.1951	64.8850	65.5845
Std (Variance)	24.1684	18.1935	26.5705	15.1853	19.1346	13.2203
Skewness	0.8157	0.9481	0.7930	0.8816	0.8366	0.8960
Std (Skewness)	0.4915	0.5169	0.3861	0.4250	0.4682	0.4470

Table 4. Stratified sampling vs. importance sampling for mean estimation petroleum reserve estimation (r = 5000, n = 27, d = 3)

	Stratified	Importance	SLH	Imp+SLH
Mean	17.6169	17.6299	17.6244	17.6204
Std Error	0.6403	0.7339	0.2352	0.3965

Stratified sampling has better performance than Latin Hypercube sampling with 512 scenarios while Latin Hypercube sampling is better with 64 scenarios (or with just 8 scenarios). We may guess that independence is more important with a high number of scenarios.

Asian option pricing

A frequent use of simulation is pricing complex options (Glasserman, Heidelberger, and Shahabuddin 1999). Although option pricing is considered a finance application, evaluating real options is a fast growing area in operations management area. Among options with no closed form solution, the Asian option is one of the most popular. An Asian option gives the holder the right to buy (or sell) an underlying asset at a price equal to the average of the prices over the lifetime of the contract. This minimizes an option holder's exposure to erratic prices on a single day. Other popular options not having closed form solutions are barrier, spread, basket and lookback options such as max/min (Campbell, Lo and MacLinlay 1997).

Generally, there is no analytic solution for pricing path

Table 5. Simulation comparison for asian option pricing (*r* = 200, *n* = 1024, *d* = 10)

	Random	Weyl	Antithetic	Stratified	LH	SLH
Mean	1.8972	1.8807	1.8946	1.8896	1.8902	1.8897
Std Error	0.0811	0.1133	0.0578	0.0554	0.0373	0.0302

dependent options. Even though path-dependent options may be priced by the dynamic-hedging approach, the resulting partial differential equations are intractable. A discrete simulation approach is usually easier to implement.

For a test, we try to price an Asian option with
initial stock price = $100,
risk free interest rate = 5%,
and maturity = 10 days.

Stock prices are assumed to follow lognormal distribution with constant drift and volatility of 0 and 1, respectively. Stock prices are averaged daily and geometrically. We conducted 200 simulations with 1024 scenarios. The result shows SLH method has lower standard error for the estimation.

CONCLUSION AND DISCUSSION

Quasi Monte Carlo simulation can be used to analyze many complex problems in operations management as well as in other disciplines such as finance. Problems that are too complicated to have a closed form solution can be solved with a relatively small number of scenarios and a high accuracy.

Stratified Latin Hypercube sampling combines Latin square sampling with multidimensional stratified sampling. We compare this approach with other well-known variance reduction methods when applied to a variety of simulation problems in the operations management area, ranging from stochastic PERT networks, petroleum reserve evaluation, and option pricing. For all of the above operations problems, we illustrated that for practical purposes, Stratified Latin Hypercube sampling is never worse than other methods, and in many applications it is significantly better.

We set the number of scenarios to l^d for some l for the performance comparison in this paper. What if $n \neq l^d$? For example, if the number of scenarios is too small to cover all random variables in the model even with the lowest level of partitioning ($n < 2^d$), it is not clear how to design sampling to enhance independence and rate of convergence. Furthermore, if we have more than 2^d scenarios in the simulation, it is difficult to judge whether we had better disregard additional simulation capacity or not. Higher simulation capacity or more samples are not always desirable. On the contrary, sometimes more samples can produce worse simulation accuracy. For a simple example, when we analyze sum function with antithetic variates, if the number of samples is even, the simulation will generate an accurate result with zero standard error. But if we add one more sample, thus the number of samples becomes odd, then additional randomness could generates positive standard error in the simulation.

REFERENCES

Bratley, P. and B. Fox and L. Schrage (1989), *A Guide to Simulation*, 2nd ed. Springer-Verlag.

Campbell, J., A. Lo and A. C. MacLinlay (1997), *The Econometrics of Financial Markets*, Princeton University Press.

Casella, G and R. Berger (1990), *Statistical Inference*, Duxbury Press.

Glasserman, P., P. Heidelberger and P. Shahabuddin (1999), "Asymtotically Optimal Importance Sampling and Stratification for Pricing Path-Dependent Options, Mathematical Finance," 9(2), 117-152.

Judd, K. (1998), *Numerical methods in Economics*, Cambridge: MIT Press.

Kocis, L. and W. Whiten (1997), "Computational Investigations of Low-Discrepancy Sequences," *ACM Transactions on Mathematical Software*, 23, 266-294.

Law, A. M. and W. D. Kelton (2000), *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill.

Mckay, M., R. Beckman and W. Conover (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, 21(2), 239-245.

Murtha, J. (2000), *Decision Involving Uncertainty: An @RISK Tutorial for the Petroleum Industry*, Newfield: Palisade Press.

Niederreiter, H. (1978), "Quasi-Monte Carlo Methods and Pseudo

Random Numbers, *Bulletin of American Mathematical Society*," 84(6), 957-1041.

Raghavarao, D. (1971), *Constructions and Combinatorial Problems in Design of Experiments*, John Wiley & Sons, Inc.

Ripley, B. (1987), *Stochastic Simulation*, John Wiley & Sons, Inc.

Robert, C and G. Casella (1999), *Monte Carlo Statistical Methods*, Springer-Verlag.

Schrage, L. (1998a), *LINGO, the modeling language and optimizer*, Chicago: LINDO Systems Inc.

__________ (1998b), *Optimization Modeling with LINGO*, Chicago: LINDO Systems Inc.

Stein, M. (1985), *The Use of Latin Hypercube Sampling for Variance Reduction in Simulations with Many Random Parameters*, Yorktown Heights: Tech Report IBM Research Center.

__________ (1987), "Large Sample Properties of Simulations Using Latin Hypercube Sampling," *Technometrics*, 29(2), 143-151.