Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D.M. Nicol, and R. M. Fujimoto, eds.

QUASI-MONTE CARLO STRATEGIES FOR STOCHASTIC OPTIMIZATION

Shane S. Drew
Tito Homem-de-Mello

Department of Industrial Engineering
and Management Sciences
Northwestern University
Evanston, IL 60208-3119, U.S.A.

ABSTRACT

In this paper we discuss the issue of solving stochastic optimization problems using sampling methods. Numerical results have shown that using variance reduction techniques from statistics can result in significant improvements over Monte Carlo sampling in terms of the number of samples needed for convergence of the optimal objective value and optimal solution to a stochastic optimization problem. Among these techniques are stratified sampling and Quasi-Monte Carlo sampling. However, for problems in high dimension, it may be computationally inefficient to calculate Quasi-Monte Carlo point sets in the full dimension. Rather, we wish to identify which dimensions are most important to the convergence and implement a Quasi-Monte Carlo sampling scheme with padding, where the important dimensions are sampled via Quasi-Monte Carlo sampling and the remaining dimensions with Monte Carlo sampling. We then incorporate this sampling scheme into an external sampling algorithm (ES-QMCP) to solve stochastic optimization problems.

1 INTRODUCTION

In this paper, we consider two-stage stochastic programs with fixed recourse of the form:

$$\min_{x \in X}\{g(x) := c^T x + \mathbb{E}[Q(x,\xi)]\} \qquad (1)$$

where X is a subset of $\mathbb{R}^n$, ξ is a random vector in $\mathbb{R}^s$, and the second stage stochastic program $Q(x,\xi)$ is defined as

$$Q(x,\xi) := \inf\{q^T y : Wy \geq h(\xi) - T(\xi)x,\ y \geq 0\}. \quad (2)$$

Thus the random vector ξ is made up of components of vector h and matrix T. We also assume that all of the components of ξ are independent of each other and that this problem has a finite solution. We shall refer to (1) as the *true optimization problem.*

Often though, $\mathbb{E}[Q(x,\xi)]$ cannot be written in closed form or it cannot be easily calculated due to a large sample space for ξ. In these cases, we can approximate the expectation with a sample average:

$$\hat{Q}_N(x) := \frac{1}{N}\sum_{j=1}^{n} Q(x,\xi^j), \qquad (3)$$

where the ξ^j are random samples from the distribution of ξ.

From our family of estimators $\{\hat{Q}_N(\cdot)\}$, we can construct another stochastic program

$$\min_{x \in X}\{\hat{g}_N(x) := c^T x + \hat{Q}_N(x)\} \qquad (4)$$

which we shall refer to as the *sampled optimization problem.* Such an algorithm is referred to as an *external sampling algorithm.* These methods are also known as Sample Average Approximation (SAA) or Sample Path Optimization as $\hat{g}(x_N)$ is obtained by average the results of N sample paths $\xi^1,\ldots,\xi^N$. If we let $\hat{v}_N$ denote the optimal objective value to the sampled optimization problem and $\hat{x}_N$ an optimal solution, then $\hat{v}_N$ and $\hat{x}_N$ are approximations to the true optimal objective value v^* and some true optimal solution x^*.

When the samples $\xi^1,\ldots,\xi^N$ are independent and identically distributed (i.i.d.), $\hat{g}_N(x)$ is referred to as a Monte Carlo estimator of $g(x)$ and the approach of solving the sampled optimization problem is usually referred to as the sample average approximation method or sample path optimization. The external sampling approach using Monte Carlo methods has been well-studied in the context of stochastic optimization (see Shapiro 2003 for a compilation of results). If x^* is the unique optimal solution to the true optimization problem, then $\hat{x}_N \to x^*$ and $\hat{v}_N \to v^*$ under some general conditions. Also, the sequence of optimal objective values

1-4244-0501-7/06/$20.00 ©2006 IEEE

$\{\hat{v}_N\}$ satisfies a central limit theorem. Namely,

$$\sqrt{N}(\hat{v}_N - v^*) \Rightarrow \text{Normal}(0, \sigma^*)$$

where " $\Rightarrow$ " denotes convergence in distribution and $\sigma^* := Var[G(x^*)]$. Thus the convergence of optimal objective values is of order $N^{-1/2}$.

Often though, either the sample size N required to guarantee a small error is extremely large or it is computationally expensive to evaluate the function $Q(x, \xi)$ at too many fixed values of ξ. A natural next step is to consider variance reduction techniques from the simulation and statistics literature and apply them to the stochastic optimization problem. Among the variance reduction techniques that have been used in the context of stochastic optimization are antithetic variates, control variates, and importance sampling.

In this paper we focus on Quasi-Monte Carlo (QMC) techniques. In QMC, we deterministically choose a point set to sample. The goal is for the point set to closely replicate a random sample from a uniform distribution. Some of these methods have been shown to have errors on the order of $\frac{(\log N)^s}{N}$ when computing the sample average. QMC sampling does have its drawbacks though. For one, it can be computationally expensive to generate QMC point sets for high dimensions. Also, while the $\frac{(\log N)^s}{N}$ rate of convergence is asymptotically superior to the Monte Carlo rate of $\frac{1}{\sqrt{N}}$, it is dependent on the dimension and often does not become superior to Monte Carlo until N is extremely large. This is impractical for a sampling algorithm. However, QMC often outperforms its theoretical rate of convergence. This suggests that the sample spaces of these problems are really in some lower dimension– i.e., that some random variables are considerably more important to the problem than others. We could then construct a Quasi-Monte Carlo sampling scheme with padding where the most important variables are sampled using Quasi-Monte Carlo and the remaining variables are sampled using a computationally less expensive scheme such as Monte Carlo or Latin Hypercube.

A few papers have looked specifically at the use of QMC methods for external sampling schemes. Kalagnanam and Diwekar (1997) show empirical results using Hammersley sequences. Koivu (2005) shows that, under mild assumptions, when Quasi-Monte Carlo methods are used to solve (1) the estimator function $\hat{g}_N$ epiconverges to the true function g, which guarantees that the optimal values and optimal solutions converge with probability one. Homem-de-Mello (2006) shows that, under proper conditions, the optimal value $\{\hat{v}_N\}$ of (4) converges to the optimal value v^* of (1) at the same rate as that at which $\hat{g}_N(x)$ converges to $g(x)$ for a fixed x. This ensures that rates of convergence for pointwise estimators under QMC carry over to the optimization case. None of those papers, however, addresses the question of how to select a subset of the random variables on which QMC sampling will be applied.

In section 2 of this paper, we describe QMC sampling in detail. In section 3, we discuss a strategy for determining the important subset of random variables for two-stage stochastic programs with fixed recourse and we propose two heuristics for estimating importance. In section 4, we develop an external sampling algorithm using Quasi-Monte Carlo sampling with padding (ES-QMCP) to solve two-stage stochastic programs. In section 5, we test our algorithm against a comparable algorithm which uses Monte-Carlo sampling on four small two-stage stochastic programs. Finally, in section 6, we present our conclusions.

2 QUASI-MONTE CARLO SAMPLING

Suppose that X is a random variable in $\mathbb{R}^s$ with all of its arguments mutually independent, $g : \mathbb{R}^s \mapsto \mathbb{R}$ is a measurable function, and we wish to estimate $I = \mathbb{E}[g(X)]$. One way to do this is via numerical integration. We choose a point set $\{\xi^1, \ldots, \xi^N\}$ from the sample space and then calculate the sample average $\hat{I} = \frac{1}{N}\sum_{j=1}^N g(\xi^j)$ from that point set. When the elements of the point set are independent and identically distributed (i.e., each point is sampled randomly from the entire sample space), the numerical integration method is called a Monte Carlo method. It is well-known from statistics that if the function $g(X)$ has a finite second moment then the error of the sample average approximation $\hat{I}_{MC}$ is of order $\frac{1}{\sqrt{N}}$.

In Quasi-Monte Carlo sampling, we wish to select points $\xi^1, \ldots, \xi^N$ that are approximately uniformly spaced avoiding large gaps or clusters. The difference between the empirical distribution of the QMC point set and the uniform distribution is quantified by the *discrepancy*. As a result, most of the research on QMC methods has focused on finding low-discrepancy point sets. Comprehensive reviews of QMC methods can be found in (Niederreiter 1992) and (L'Ecuyer and Lemieux 2002).

One type of QMC point set that yields low discrepancy is the so-called (t, m, s)*-net.* For completeness, we include a definition below.

Definition 1 *An* elementary interval *of base b in dimension s is a subinterval E of the form*

$$E = \prod_{j=1}^{s} \left[\frac{a_j}{b^{d_j}}, \frac{a_j+1}{b^{d_j}}\right)$$

for nonnegative integers $\{a_j\}$ and $\{d_j\}$ with $a_j < b^{d_j}$ for all j. E then has volume $b^{-\sum_j d_j}$.

Let m be a nonnegative integer. A finite sequence of $N = b^m$ points is called a (t, m, s)*-net in base b* if every elementary interval in base b of volume $\frac{1}{b^{m-t}}$ contains exactly b^t points of the sequence. Note, that any (t, m, s)-net is also a

(u,m,s)-net for any integer $u \in [t,m]$. Thus smaller values of t are more desirable. The discrepancy (and integration error) of a (t,m,s)-net in base b is $O\left(\frac{(\log(N))^{s-1}}{N}\right)$.

However, as a QMC point set is entirely deterministic, the estimator $\hat{I} = \frac{1}{N}\sum_{j=1}^{N} g(\xi^j)$ has zero variance. This makes it difficult to calculate the integration error $|\hat{I}_{QMC} - I|$. In fact, all we have is an upper bound for the integration error from the Koksma-Hlawka inequality, which states that the integration error of a function using a QMC point set is bounded above by the product of the total variation of the function and the discrepancy of the point set. One way to rectify this problem is by using randomized QMC methods. In a randomized QMC method, each individual sample point is now uniformly distributed over the sample space, but the point set as a whole still preserves the low-discrepancy property. Randomized QMC methods are discussed in detail in (Fox 2000) and (Owen 2000).

We focus on scrambled (t,m,s)-nets proposed by Owen (1995). In this method, the digits of the base b expansion of each element of the net are randomly scrambled in a particular way. The resulting scrambled (t,m,s)-net is still a (t,m,s)-net with probability one. Thus all of the properties of (t,m,s)-nets can also be applied to the scrambled nets. Another important property is that the point set from a scrambled net is uniformly distributed over the unit hypercube. This is a consequence of the scrambling and is independent of whether or not the pre-scrambled point set was a (t,m,s)-net. As a result, the integration error can be calculated by taking multiple independent replications of the scrambled net (from the same original net).

2.1 Effective Dimension

While (t,m,s)-nets and scrambled (t,m,s)-nets compare favorably to Monte Carlo sampling methods, there is still one major disadvantage. Although the Quasi-Monte Carlo integration error of $\frac{(\log(N))^{s-1}}{N}$ is asymptotically superior to the Monte Carlo error bound of $\frac{1}{\sqrt{N}}$, it is not advantageous until N is very large, unless the dimension s is small. Even for $s = 5$, we must have $N \geq 2.1 \times 10^{12}$ for QMC to be theoretically advantageous. Yet for many problems, numerical results show that QMC methods have a much lower integration error, even for reasonable sample sizes. This suggests that either the upper bound from the Koksma-Hlawka inequality is not tight or that the problem is actually in some lower dimension. That in turn has led to the notion of effective dimension (Caflisch, Morokoff, and Owen 1997), which we describe below.

If the function $g(\cdot)$ is square-integrable, then $g(\cdot)$ can be written

$$g(\xi) = \sum_{A \in \{1,\dots,s\}} g_A(\xi) \tag{5}$$

where $g_A(\xi)$ depends only on the components ξ_j with $j \in A$. Also, $g_A(\xi)$ satisfies the properties:

1. $\int_0^1 g_A(\xi)d\xi_j = \begin{cases} 0, & \text{if } j \in A \\ g_A(\xi), & \text{if } j \notin A. \end{cases}$
2. $\int_{[0,1)^s} g_u(\xi)g_v(\xi)d\xi = 0, \quad \text{if } u \neq v.$

$g_A(\xi)$ is calculated by looking at the portion of g that is not determined by subsets of A and then averaging over all of the components not in A, i.e.,

$$g_A(x) = \int_{[0,1)^{A^c}} \left(g(\xi) - \sum_{v \subset A} g_v(\xi) \right) d\xi^{A^c}. \tag{6}$$

The goal of any sampling method is to estimate $\bar{g} := \int_{[0,1)^s} g(\xi)d\xi$ (equivalent to letting A be the empty set in (6)). From the above properties, the variance of g, defined as $\sigma^2 := \int (g(\xi) - \bar{g})d\xi$, satisfies

$$\sigma^2 = \sum_{A \in \{1,\dots,s\}} \sigma_A^2 \tag{7}$$

where $\sigma_A^2 := \int g_A(\xi)^2 d\xi$. Hence, this is called the ANOVA (analysis of variance) decomposition of g. One can then define the effective dimension of the problem based on the terms of the ANOVA decomposition that contribute most toward the overall variance.

The effective dimension of g in the truncation sense is the smallest integer s_T such that

$$\sum_{A \subset \{1,2,\dots,s_T\}} \sigma_A^2 \geq (1-\varepsilon)\sigma^2. \tag{8}$$

Typically, ε is chosen to be 0.01. Note that the value of s_T is dependent on the order in which the input variables are indexed. This definition implies that only a small number of the input variables are important to the problem, thus there is little need for variance reduction techniques on the remaining variables. This leads to the concept of *padding*, where the important variables are integrated using a QMC or randomized QMC point set and the remaining variables are integrated using something computationally less expensive such as the midpoint of the interval, a Monte Carlo sample (Spanier 1995), or a Latin Hypercube sample (Owen 1998).

3 DETERMINING THE IMPORTANT SUBSET OF RANDOM VARIABLES

We now apply the concept of QMC sampling with padding to the two-stage stochastic program. By calculating the effective dimension of the function in the truncation sense, we can identify which random variables have the largest contribution of variance to the function. We then can construct a smaller QMC point set on just these components and pad the sample using some other method such as Monte Carlo sampling on the remaining components. However, the ANOVA decomposition of an s-dimensional function requires computing 2^s integrals. This can be quite costly at each stage of an optimization algorithm. We instead look to a different method to estimate the importance of each random variable.

Since Quasi-Monte Carlo sampling is a variance reduction tool, it seems natural to label the important variables as the ones that contribute the most to the overall variance of the function. Let V_k be the contribution of random variable k to the overall variance of the function. Our goal is to estimate the V_k via heuristic methods.

For the two-stage stochastic program, the function of interest is the second stage optimal value $Q(x, \xi)$. Since the random variables are elements of the right-hand side $h - Tx$ of the second stage problem, which are not in the objective of the primal problem, it is often more convenient to look at the dual of the second stage stochastic program:

$$\sup\{\pi^T(h(\xi) - T(\xi)x) : \pi^T W \leq q^T, \pi \geq 0\} \qquad (9)$$

Thus the total variance of the optimal value is

$$\text{Var}\left(\sum_k \pi_k^*(h_k - \sum_j T_{kj}x_j)\right). \qquad (10)$$

where π_k^* is the optimal dual multiplier for constraint k. By expanding the sums, we can rewrite the optimal value as $\sum_i Z_i$ where each Z_i is the product of some constraint's (call it constraint k) optimal dual multiplier π_k^* and a term that (a) involves h_k and/or elements from the k^{th} row of T and (b) contains *at most* one random component of ξ. (In the case where each constraint in the primal problem contains exactly one random variable, then $Z_i = \pi_i^*(h_i - T_i x)$. Otherwise, we split out the h_i and the elements of row T_i accordingly). While we assume that the random variables themselves are mutually independent, the terms in the objective function are usually dependent due to the interactions of the dual multipliers. The variance of the optimal value is now

$$\text{Var}(\sum_i Z_i) = \sum_i \sum_j \text{Cov}(Z_i, Z_j). \qquad (11)$$

By definition, the variance of the optimal value can also be written as $\sum_{k=1}^s V_k$. Our goal then is to estimate the individual covariance terms in (11) and assign each to one of the V_k.

The Z_i can be partitioned into two sets

1. A := the set of all Z_i containing exactly one random component of ξ (contains s elements).
2. B := the set of all Z_i containing zero random components of ξ (contains $m - s$ elements).

and reordered so that $Z_1, \ldots, Z_s$ correspond to the same random variables as $V_1, \ldots, V_s$. We now propose two heuristics to estimate the V_k.

Our first heuristic is to estimate V_k by

Heuristic 1 :
$\hat{V}_k = |\sum_{j\in A} Cov(Z_k, Z_j) + 2\sum_{j\in B} Cov(Z_j, Z_k)|, \quad k = 1, \ldots s.$ □

Here the contribution of each random variable to the overall variance includes the variance of its own Z term plus the covariance with the Z terms of the other random components of ξ plus *both* covariance terms with the Z terms that do not have a random component of ξ (as the second covariance term would otherwise be unassigned). We ignore the terms Cov(Z_i, Z_j) where $i, j \in B$ as they should theoretically be small since there are no random components present. We take the absolute value to account for terms that greatly affect the variance in either direction.

This method involves estimating at most $s(m-s)$ terms in the covariance matrix (due to the symmetry of the matrix) though this number can be reduced by ignoring Z terms that are identically zero. While this is most likely less than the 2^s terms of the ANOVA decomposition, since m is at least as large as the number of constraints in the primal problem, these estimates can still be quite cumbersome for very large problems.

Our second heuristic is to estimate V_k by

Heuristic 2 :
$\hat{V}_k = Var(Z_k), \quad k = 1, \ldots s.$ □

This is a very crude approximation which completely ignores any dependence between the dual multipliers. However, since it only involves estimating at most s terms of the covariance matrix rather than the $s(m-s)$ terms of the first heuristic, it may be preferable for large problems.

Once we have calculated the $\hat{V}_k$ via one of the heuristics, we are now ready to select the important subset. We propose the following algorithm:

Algorithm 1: Selecting the Important Set I

1. Let total variance $V = \sum_k \hat{V}_k$, important variance $V^I = 0$, important set $I = \emptyset$, and unimportant set $M = \{1, \ldots, S\}$.
2. Rank the $\hat{V}_k$ in descending order.
3. Select the largest remaining $\hat{V}_k$. Let $V^I \leftarrow (V^I + \hat{V}_k)$. Let $I \leftarrow I \bigcup \{k\}$ and $M \leftarrow M - \{k\}$.
4. If $\frac{V^I}{V} \geq 0.9$ or $|I| = 10$ then terminate the algorithm with important set I. Otherwise, return to step 3.

We are now ready to implement this method into an external sampling algorithm to solve two-stage stochastic programs.

4 AN EXTERNAL SAMPLING ALGORITHM USING QMC SAMPLING WITH PADDING (ES-QMCP)

In order to estimate elements of the covariance matrix from the second stage stochastic program (2), we need a first stage solution x. Thus the algorithm we propose is an iterative algorithm where the sampled stochastic program is solved to obtain a new x at each iteration using a QMC point set with padding and then that x is used to determine the set of important variables for the next iteration. The initial x value is determined by solving the sampled optimization problem using a Monte Carlo point set (as we have not yet identified the important random variables by that time). To estimate each random variable's contribution V_k to the variance and determine the set of important random variables, we estimate $\mathbb{E}[Q(x,\xi)]$ using a *Monte Carlo* point set. Then from the samples we can estimate the covariance matrix of the Z_k terms (as defined in section 3). In order to gauge each random variable's full effect on the overall variance, it is essential that we employ a sampling scheme with no variance reduction while estimating the covariance terms. The full External Sampling Algorithm using Quasi-Monte Carlo Sampling with Padding (ES-QMCP) is detailed below.

Algorithm ES-QMCP

Below, $N^{(0)}$ and α are pre-specified parameters or the algorithm.

1. *Initialization:*
   (a) Rewrite the second stage dual objective as $\sum_k Z_k$ as described in section 3.
   (b) Set iteration count $i \leftarrow 0$, important set $I^{(0)} = \emptyset$, and sample size $N^{(0)}$.
   (c) Using a Monte Carlo point set of size $N^{(0)}$, solve the sampled optimization program to determine first stage solution $x^{*(0)}$ and optimal value $\psi^{(0)}$.
2. *Increment:* Let $i \leftarrow i+1$ and $N^{(i)} = \alpha N^{(i-1)}$.
3. Using $N^{(i)}$ *Monte Carlo* samples to solve the *second stage* stochastic program at $x = x^{*(i-1)}$, calculate the estimates $\hat{V}_k$, $k = 1,\ldots,s$ using one of the two heuristics described in section 3.
4. Using Algorithm 1, select the important set $I^{(i)}$ for iteration i.
5. Using a padded point set — a Quasi-Monte Carlo point set for random variables $k \in I^{(i)}$ and a Monte Carlo point set on random variables $k \notin I^{(i)}$ — solve the sampled optimization program to determine first stage solution $x^{*(i)}$ and optimal value $\psi^{(i)}$.
6. If a stopping criterion is satisfied, then terminate the algorithm with optimal first stage solution $x^{*(i)}$ and optimal value $\psi^{(i)}$. Otherwise return to step 2. □

In our implementation we used $N^{(0)} = 8$, $\alpha = 2$ and the algorithm was stopped when $\frac{|\psi^{(i)} - \psi^{(i-1)}|}{\psi^{(i-1)}} \leq \epsilon$, i.e., when the relative change in optimal value between iterations was less than some tolerance ϵ. We used $\epsilon = 0.01$. One could use more sophisticated criteria, but we used this simple one for a preliminary implementation. We will compare this algorithm to the corresponding algorithm ES-MC where all sampling is done with crude Monte Carlo, thus making it unnecessary to estimate the importance of the random variables.

Algorithm ES-MC

1. *Initialization:*
   (a) Set iteration count $i \leftarrow 0$, and sample size $N^{(0)}$.
   (b) Using a Monte Carlo point set of size $N^{(0)}$, solve the sampled optimization program to determine first stage solution $x^{*(0)}$ and optimal value $\psi^{(0)}$.
2. *Increment:* Let $i \leftarrow i+1$ and $N^{(i)} = \alpha N^{(i-1)}$.
3. Using a Monte Carlo point set, solve the sampled optimization program to determine first stage solution $x^{*(i)}$ and optimal value $\psi^{(i)}$.
4. If a stopping criterion is satisfied, then terminate the algorithm with optimal first stage solution $x^{*(i)}$ and optimal value $\psi^{(i)}$. Otherwise return to step 2. □

5 NUMERICAL RESULTS

We now look at four test problems:

- gbd – the airline fleet assignment problem first proposed by (Dantzig 1963)
- LandS1, LandS2 – an electrical investment planning problem that originally appeared in (Louveaux and Smeers 1988). LandS2 is just a version of LandS1 with the objective coefficients shuffled.
- APL1P – a model for electric power capacity expansion from (Infanger 1992).

Since our test problems are all relatively small (3 to 5 random variables) we can explicitly write out the dual for the second stage problems. Given a problem with s random variables, we hope to identify the subset I of variables that are most important to the optimal value of the problem. In our QMC sampling scheme with padding, the random variables in I are generated via a scrambled $(t,m,|I|)$-sequence in base 2 and the remaining $s - |I|$ variables by Monte Carlo sampling. For an s-dimensional problem, there are 2^s possible sampling schemes (technically, the order of the QMC variables matters so there would be $\sum_{k=0}^{s}\binom{s}{k}k!$ sampling schemes, but for simplicity we will assume that the order is irrelevant). Our samples are generated using the publicly available random sampling routines from (Friedel and Keller 2002) which we built into the SUTIL library (Czyzyk, Linderoth, and Shen 2005).

We are interested in testing two results:

- How well our two heuristics for estimating the V_k explain the importance of the random variables
- The performance of the algorithm ES-QMCP using each of the two heuristics

For the first item, we quantify a random variable's importance by how much variance reduction is gained by including it the set of important variables. Since the test problems are all small, we can estimate the expected objective value of the second stage stochastic programs using *all* 2^s sampling schemes (for a given x value). For each sampling scheme, we solve 10 realizations of the second stage program with a sample size of $N = 1024$ using the ATR solver (Linderoth and Wright 2003). We then estimate the overall variance of that estimator for each sampling scheme by the variance across the replications for the sampling scheme. Let the variance estimate for sampling scheme i, $i = 1, \ldots, 2^s$ be Y_i.

From the 2^s variance estimates we perform a linear regression. We assume that the variance for sampling scheme i can also be estimated as a linear function

$$\hat{Y}_i = \hat{\beta}_0 + \sum_{j=1}^{s} \hat{\beta}_j X_{ij} \tag{12}$$

where $\hat{\beta}_j$ is the effect on the variance from random variable j (usually negative to indicate variance reduction), $\hat{\beta}_0$ is the constant term, and X_{ij} is one if random variable j is included in the important set and zero otherwise. So each $\hat{\beta}_j$ is a rough measure of the importance of random variable j. From statistics (e.g., Tamhane and Dunlop 2000), $\hat{\beta}_j$ can be calculated as

$$\hat{\beta}_j = \frac{\sum_{i=1}^{2^s} \lambda_{ij} \bar{Y}_i}{2^{s-1}} \tag{13}$$

where $\bar{Y}_i = \frac{\sum_{i=1}^{2^s} Y_i}{2^s}$ and λ_{ij} is $+1$ if random variable j is in the important set for sampling scheme i and -1 if not (for $j = 0$, $\lambda_{ij} = +1$ for all i). We also can assess the quality of the regression model by calculating

$$R^2 = \frac{\sum_{i=1}^{2^s}(Y_i - \hat{Y}_i)}{\sum_{i=1}^{2^s}(Y_i - \bar{Y}_i)}.$$

Table 1: Assessing Heuristics to Identify Important Variables - gbd

	Random		$\hat{V}_j$	
j	Variable	$\hat{\beta}_j$	Heuristic 1	Heuristic 2
0	Constant	372.92	n/a	n/a
1	d_1	-87.25	$127,701.88$	$133,094.52$
2	d_2	-25.43	$5,690.23$	$4,966.30$
3	d_3	-34.53	$13,656.90$	$11,517.01$
4	d_4	-240.68	$285,821.12$	$295,241.44$
5	d_5	-27.35	142.38	31.45
Correlation with $\hat{\beta}$:			-0.9866	-0.9856

Regression $R^2 = 0.7617$

In our problem descriptions, we also assess the quality of our two heuristics for estimating the importance by calculating the correlation between the $\hat{V}_j$ and the $\hat{\beta}_j$. As mentioned earlier, those values depend on the first-stage solution x; the numbers we report below were obtained with $x = x^*$, the optimal solution (which is known as those problems have been solved in the literature). This is done only with the purpose of evaluating the heuristics; the ES-QMCP and ES-MC algorithms, of course, do not require knowledge of the optimal solution.

After the describing the test problems, we look at the performance of the ES-QMCP algorithm using the two heuristics on those problems versus the corresponding ES-MC algorithm.

5.1 gbd

We begin with the fleet assignment problem first described by (Dantzig 1963). Four types of aircrafts are to be assigned to five different routes (with three of the aircraft/route combinations not valid). The random variables are the customer demands for each route (denoted d_j). This is a relatively simple problem where the dual variables are independent and thus the importance of each random variable can be quantified in closed form.

Using the random demand scenarios from (Linderoth, Shapiro, and Wright 2005), we calculate our $\hat{V}_j$ estimates using both heuristics and compare to our regression estimates $\hat{\beta}_j$. Since all of the terms in the objective are independent, we expect our covariance estimates between different terms to be approximately zero and for the two heuristics to give similar results for the values of the $\hat{V}_j$. Not surprisingly, this is the case and both are highly correlated with the $\hat{\beta}_j$ estimates from the regression (see Table 1) with random variables d_1 and d_4 accounting for nearly all of the variance.

Table 2: Assessing Heuristics to Identify Important Variables - LandS1

j	Random Variable	$\hat{\beta}_j$	$\hat{V}_j$ Heuristic 1	$\hat{V}_j$ Heuristic 2
0	Constant	1.34	n/a	n/a
1	d_1	-1.16	$2,066.60$	$3,206.57$
2	d_2	-0.08	918.19	$1,251.99$
3	d_3	0.48	66.73	41.61
Correlation with $\hat{\beta}$:			-0.9954	-0.9989

Regression $R^2 = 0.6645$

5.2 LandS

Our next example is an electrical investment planning problem that originally appeared in (Louveaux and Smeers 1988). The first stage decision involves allocating capacity to four new technologies subject to capacity minimums and budget restrictions. The second stage decision is regarding the production of three different modes of electricity for each technology to meet customer demand subject to the capacities from the first stage. In this problem the customer demand for each mode of electricity is random (again denoted d_j). We use the modifications to the demand scenarios from (Linderoth, Shapiro, and Wright 2005) where the random variables have identical independent distributions with 100 equally likely realizations. Here however, the dual variables are not independent and thus the terms in the dual objective function are correlated.

We consider two forms of this problem. The first is the original problem which we will refer to as LandS1. Here the cost coefficients c_{ij} in the primal objective function (where i is the technology and j the mode of electricity) take the form $c_{i2} = 6c_{i3}$ and $c_{i1} = 10c_{i3}$. Since the demand random variables are identically distributed, we expect the relationship between the c_{ij} to cause d_1 to be the most important random variable, followed by d_2 then d_3. It turns out that this is indeed the case. In fact the difference in importance is so drastic that there is little room for improvement by including the covariance terms in our estimates of the $\hat{V}_j$ (see Table 2).

To alleviate the effects of the structure of the c_{ij}, we consider a second form LandS2 of the problem where we shuffle the 12 c_{ij} coefficients. Here in Table 3, we see that the correlations between the $\hat{V}_j$ and $\hat{\beta}_j$ are lower than the other problems and that surprisingly the inclusion of the covariance terms into the heuristics actually gives a worse correlation. It is worthwhile to note however that the R^2 for the regression where we estimate the $\hat{\beta}_j$ is much lower than that of the other problem, so it is possible that the $\hat{\beta}_j$ are not good estimates.

Table 3: Assessing Heuristics to Identify Important Variables - LandS2

j	Random Variable	$\hat{\beta}_j$	$\hat{V}_j$ Heuristic 1	$\hat{V}_j$ Heuristic 2
0	Constant	0.23	n/a	n/a
1	d_1	-0.20	305.88	668.54
2	d_2	0.06	9.80	13.40
3	d_3	0.05	207.85	379.20
Correlation with $\hat{\beta}$:			-0.7897	-0.8602

Regression $R^2 = 0.5055$

Table 4: Assessing Heuristics to Identify Important Variables - apl1p

j	Random Variable	$\hat{\beta}_j$	$\hat{V}_j$ Heuristic 1	$\hat{V}_j$ Heuristic 2
0	Constant	$20,458$	n/a	n/a
1	a_1	-514	$2,714,095$	$11,267,140$
2	a_2	-436	$1,635,107$	$7,257,204$
3	d_1	-204	$2,266,447$	$1,966,338$
4	d_2	$-10,821$	$6,565,340$	$3,942,965$
5	d_3	$-10,308$	$9,656,419$	$6,888,416$
Correlation with $\hat{\beta}$:			-0.9308	0.2093

Regression $R^2 = 0.8364$

5.3 apl1p

Our final problem is a model for electric power capacity expansion first described in (Infanger 1992). The problem consists of two electrical generators that can operate at three levels. The first stage problem is to determine capacity for the two generators. The second stage decision is to determine how much to produce at each of the three operation levels on the two generators subject to the the availability of the two generators (random variables denoted a_i) and customer demand for each of the three operation levels (random variables denoted d_j).

Again, the terms in the dual objective function are correlated due to the structure of the dual constraints. We see in Table 4 that the inclusion of the covariance terms into the heuristic for $\hat{V}_j$ significantly improves the correlation with the $\hat{\beta}_j$. In fact, Heuristic 2 indicates that a_1 and a_2 are the two most important random variables, while Heuristic 1 and the regression indicate that they are significantly less important than d_2 or d_3.

Table 5: A Comparison of the ES-QMCP and ES-MC Algorithms

gbd	Optimal Value		Iterations		Total Samples	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
MC	1,671.94	24.95	7.0	4.3	8,248	9,326
QMCP	1,648.46	24.17	4.0	1.1	590	446
	True Optimal Value = 1,655.62					

LandS1	Optimal Value		Iterations		Total Samples	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
MC	227.89	6.60	4.8	2.1	670	550
QMCP	225.03	1.04	2.0	0.0	104	0
	True Optimal Value = 225.62					

LandS2	Optimal Value		Iterations		Total Samples	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
MC	126.54	2.52	3.4	2.4	254	491
QMCP	127.72	1.15	2.0	0.0	104	0
	True Optimal Value = 128.20					

apl1p	Optimal Value		Iterations		Total Samples	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
MC	24,375.16	685.52	4.8	2.5	670	992
QMCP-1	24,647.62	204.30	3.6	1.0	411	197
QMCP-2	24,686.32	29.13	3.4	1.0	360	208
	True Optimal Value = 24,642.32					

5.4 Numerical Results of the Algorithm

We now test our algorithm ES-QMCP against the Monte Carlo analogue ES-MC for each of our test problems (including both versions of LandS). Five replications of each algorithm were run for each problem. For gbd, LandS1, and LandS2, since the choice of heuristic does not seem to effect the choice of the important subset we only run algorithm ES-QMCP using Heuristic 1. For apl1p, where the importance of the random variables can vary depending on the heuristic we run ES-QMCP using both Heuristic 1 and Heuristic 2 (denoted ES-QMCP-1 and ES-QMCP-2 respectively).

We use three measures to assess the performance of the algorithms: (1) the optimal value at the end of the algorithm, (2) the number of iterations of the algorithm until convergence, and (3) the total number of samples used in the algorithm. Note that in ES-QMCP since we need to estimate the important subset, each iteration uses twice as many samples as an iteration of ES-MC. In Table 5, we report the mean and the half-width of the 95% confidence interval about the means for each performance measure.

From our results, we see that ES-QMCP performs better than ES-MC in all areas. The average optimal values from the algorithm are closer to the actual results; it requires fewer iterations; and on average uses fewer samples even though each iteration itself is more expensive. Also, the variances of these measures are lower under ES-QMCP suggesting that we can more accurately predict when the algorithm will stop.

For apl1p, where we run the algorithm using two different heuristics for estimating the important variables, the two heuristics seem indistinguishable at 5 replications though it is indeed possible that one heuristic would prove superior if the number of replications were increased.

6 CONCLUSIONS

Our empirical data suggests that external sampling algorithms that incorporate Quasi-Monte Carlo methods are more efficient than their Monte Carlo counterparts in terms of the number of samples needed. Still QMC point sets are more time-intensive to generate than Monte Carlo samples. Additionally computing time must be devoted to determining the important subset of random variables. Ultimately, there is a tradeoff between the additional time necessary to come up with good estimates of the important subset and the number of fewer samples required by the algorithm. Research is ongoing to quantify this tradeoff.

REFERENCES

Caflisch, R. E., W. J. Morokoff, and A. B. Owen. 1997. Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *Journal of Computational Finance* 1:27–46.

Czyzyk, J., J. Linderoth, and J. Shen. 2005. SUTIL: A utility library for handling stochastic programs. User's Manual. Software available at `http://coral.ie.lehigh.edu/sutil`.

Dantzig, G. 1963. *Linear programming and extensions*. Princeton, New Jersey: Princeton University Press.

Fox, B. L. 2000. *Strategies for quasi-Monte Carlo*. Norwell, MA: Kluwer Academic Publishers.

Friedel, I., and A. Keller. 2002. Fast generation of randomized low-discrepancy point sets. In *Monte Carlo and quasi-Monte Carlo methods, 2000 (Hong Kong)*, 257–273. Berlin: Springer. Software available at `http://www.multires.caltech.edu/software/libseq/`.

Homem-de-Mello, T. 2006. On rates of convergence for stochastic optimization problems under non-i.i.d. sampling. Manuscript, available on *Optimization Online* (`www.optimization-online.org`).

Infanger, G. 1992. Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs. *Annals of Operations Research* 39:69–95.

Kalagnanam, J., and U. Diwekar. 1997. An efficient sampling technique for off-line quality control. *Technometrics* 39 (3): 308–319.

Koivu, M. 2005. Variance reduction in sample approximations of stochastic programs. *Mathematical Programming*:463–485.

L'Ecuyer, P., and C. Lemieux. 2002. Recent advances in randomized quasi-Monte Carlo methods. In *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*, ed. M. Dror, P. L'Ecuyer, and F. Szidarovszky, 419–474. Boston: Kluwer Academic Publishers.

Linderoth, J. T., A. Shapiro, and S. J. Wright. 2005. The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*. forthcoming.

Linderoth, J. T., and S. J. Wright. 2003. Implementing a decomposition algorithm for stochastic programming on a computational grid. *Computational Optimization and Applications* 24:207–250. Special Issue on Stochastic Programming.

Louveaux, F., and Y. Smeers. 1988. Optimal investments for electricity generation: A stochastic model and a test problem. In *Numerical techniques for stochastic optimization problems*, ed. Y. Ermoliev and R. J.-B. Wets, 445–452. Berlin: Springer-Verlag.

Niederreiter, H. 1992. *Random number generation and quasi-Monte Carlo methods*. Philadelphia, PA: SIAM.

Owen, A. B. 1995. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte Carlo and quasi-Monte Carlo methods in scientific computing (Las Vegas, NV, 1994)*, Volume 106 of *Lecture Notes in Statist.*, 299–317. New York: Springer.

Owen, A. B. 1998. Latin supercube sampling for very high-dimensional simulations. *ACM Transactions on Modeling and Computer Simulation* 8:71–102.

Owen, A. B. 2000. Monte Carlo, quasi-Monte Carlo, and randomized quasi-Monte Carlo. In *Monte Carlo and quasi-Monte Carlo methods 1998 (Claremont, CA)*, 86–97. Berlin: Springer.

Shapiro, A. 2003. Monte Carlo sampling methods. In *Handbook of Stochastic Optimization*, ed. A. Ruszczynski and A. Shapiro. Amsterdam, Netherlands: Elsevier Science Publishers B.V.

Spanier, J. 1995. Quasi-monte carlo methods for particle transport problems. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, ed. H. Niederreiter and P. Shiue, 121–148. New York: Springer-Verlag.

Tamhane, A., and D. Dunlop. 2000. *Statistics and data analysis from elementary to intermediate*. Upper Saddle River, New Jersey: Prentice-Hall.

AUTHOR BIOGRAPHIES

SHANE S. DREW is a graduate student in the Department of Industrial Engineering and Management Sciences at Northwestern University. His e-mail address is <s-drew@northwestern.edu>.

TITO HOMEM-DE-MELLO is an Associate Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. He received his PhD degree in Operations Research from the Georgia Institute of Technology in 1998. His research interests include stochastic optimization, simulation analysis methodology, and stochastic models for revenue management. His e-mail and web addresses are <tito@northwestern.edu> and <http://users.iems.northwestern.edu/˜tito/>.