The Impact of Sampling Methods on Bias and Variance in Stochastic Linear Programs

Michael B. Freimer
SignalDemand, Inc., 301 Howard Street, Suite 1950, San Francisco, CA 94105, USA,
michael.freimer@SignalDemand.com

Jeffrey T. Linderoth
Industrial and Systems Engineering, Lehigh University, 200 W. Packer Ave., Bethlehem, PA 18015, USA,
jtl3@lehigh.edu

Douglas J. Thomas
Smeal College of Business, The Pennsylvania State University, 463 Business Building, State College, Pennsylvania 16802-3603, USA, dthomas@psu.edu

Stochastic linear programs can be solved approximately by drawing a subset of all possible random scenarios and solving the problem based on this subset, an approach known as sample path optimization. The value of the optimal solution to the sampled problem provides an estimate of the true objective function value. This estimator is known to be optimistically biased; the expected optimal objective function value for the sampled problem is lower (for minimization problems) than the optimal objective function value for the true problem. We investigate how two alternative sampling methods, antithetic variates and Latin Hypercube sampling, affect both the bias and variance, and thus the mean squared error (MSE), of this estimator. For a simple example, we analytically express the reductions in bias and variance obtained by these two alternative sampling methods. For eight test problems from the literature, we computationally investigate the impact of these sampling methods on bias and variance. We find that both sampling methods are effective at reducing mean squared error, with Latin Hypercube sampling outperforming antithetic variates. Whether the bias reduction or variance reduction plays a larger role in MSE reduction is problem and parameter specific.

Key words: stochastic programming; sample path optimization; antithetic variates; Latin Hypercube sampling
History:

1. Introduction

Two-stage stochastic linear programs arise in a variety settings. At the first stage, values are chosen for a set of design variables; for example, the design variables may represent a set of production line capacities, while the second stage decisions may be production quantities. The objective function of the first-stage problem requires us to evaluate the expected value of the solution to a second-stage linear program (e.g., a production scheduling problem), some of whose parameters (e.g., demand) are stochastic. Furthermore, the design variables from the first stage appear in the constraints of the second-stage linear program (LP). Early formulations of this problem were given by Dantzig (1955) and Beale (1955).

The motivation for this paper is the efficient solution of such two-stage design problems. Throughout the paper we adopt a modified version of the notation for two-stage stochastic programs presented by Kleywegt and Shapiro (2001):

$$\text{MP}: \quad z^*_{\text{MP}} \stackrel{\text{def}}{=} \min_x \mathbb{E}_\omega \left[Q(x,\omega)\right] + g(x), \text{ s.t. } Ax = b, \ x \geq 0,$$

where $g(x)$ is a deterministic function of x, and $Q(x,\omega)$ represents the optimal objective function value of the second-stage problem:

$$\text{P}: \quad Q(x,\omega) \stackrel{\text{def}}{=} \min_y q(\omega)^T y, \text{ s.t } T(\omega)x + W(\omega)y = h(\omega), \ y \geq 0.$$

Here $q(\omega) \in \mathbb{R}^n, T(\omega) \in \mathbb{R}^{\ell \times m}, W(\omega) \in \mathbb{R}^{\ell \times n}$, and $h(\omega) \in \mathbb{R}^\ell$ may be random (functions of the realization ω). When $g(x) = c^T x$ and $W(\omega)$ is deterministic, we have a two-stage stochastic linear program with fixed recourse.

Standard solution methods, such as the L-shaped method, suppose that the random components of the problem have finite support. If there are K possible realizations (scenarios), each with probability p_k, an equivalent extensive form of MP can be written:

$$\text{EF}: \min_{x,y_1,\ldots,y_K} \sum_{i=1}^{K} p_i q(\omega_i)^T y_i + g(x), \text{ s.t.}$$

$$Ax = b, \ x \geq 0, \ T(\omega_i)x + W(\omega_i)y_i = h(\omega_i), \ y_i \geq 0, \ i = 1, 2, \ldots, K.$$

The L-shaped method takes advantage of this problem's structure by performing a Dantzig-Wolfe decomposition of the dual or a Benders decomposition of the primal (Birge and Louveaux, 1997).

In many practical problems the number of possible scenarios K is prohibitively large, so a Monte Carlo approximation of MP is used. We will refer to this approach as sample path optimization. The idea is to draw N realizations (sample paths) and optimize over this representative sample. More specifically, let $MP_N(\omega_1, \ldots, \omega_N)$ denote a realization of the N-sample path problem. That is,

$$\text{MP}_N: \quad z^*_{\text{MP}_{N(\omega_1,\ldots,\omega_N)}} \stackrel{\text{def}}{=} \min_x N^{-1} \sum_{i=1}^{N} Q(x, \omega_i) + g(x). \text{ s.t. } Ax = b, \; x \geq 0.$$

The solution to MP_N is used to approximate the solution to the original problem MP.

In this paper we examine three different sampling procedures for generating the approximating problem MP_N: independent sampling (IS), antithetic variates (AV), and Latin Hypercube sampling (LH). Suppose X is a random element of the data $\{q, T, W, h\}$ having invertible cdf F, so $Q_i(x, \omega_i)$ is a function of $X(\omega_i)$. Under independent sampling we generate N independent values $\{U_1, \ldots, U_N\}$ uniformly distributed on [0,1], and we use $X(\omega_i) = F^{-1}(U_i)$ to compute $Q_i(x, \omega_i)$. Under AV, rather than drawing N independent numbers, we draw $N/2$ antithetic pairs $\{(U_i, 1-U_i), i = 1, 2, \ldots, N/2\}$ to obtain $\{U_1, \ldots, U_N\}$. Under LH, the interval [0,1] is divided into N segments, $[(i-1)/N, i/N]$, $i = 1 \ldots, N$, and a sample is generated uniformly from each segment. These samples are shuffled to obtain $\{U_1, \ldots, U_N\}$.

We are concerned with the impact of the sampling procedure on the use of $z^*_{\text{MP}_N}$ as an estimator for z^*_{MP}. As with any statistical estimation problem, two important measures of performance are the estimator's variance and bias, which are combined as the mean squared error (MSE). (The MSE is equal to the bias squared plus the variance.) The relative contribution of variance and bias to MSE is problem- and parameter-specific. In Section 2 we develop an example based on the newsvendor problem in which the fraction of MSE due to variance changes dramatically with the cost parameters of the problem. In Section 3 we investigate a series of computational examples from the literature which tell a similar story. In some problems the bulk of MSE is due to variance, and in others the bias predominates.

The AV, and LH sampling procedures are usually prescribed for reducing variance (Law and Kelton, 2000), and indeed Higle (1998) investigates the use of AV and other techniques to reduce the variance of

$$N^{-1} \sum_{i=1}^{N} Q_i(x, \omega_i) + g(x),$$

which is an unbiased estimate of $\mathbb{E}_\omega [Q(x, \omega)] + g(x)$ for an arbitrarily chosen value of x. Here we are concerned with estimating z^*_{MP}, which is $\mathbb{E}_\omega [Q(x, \omega)] + g(x)$ evaluated at x^*_{MP}, the unknown optimal solution to MP. In Section 2 we derive analytic expressions for the variance of the estimator $z^*_{\mathrm{MP}_N}$ in the context of the newsvendor problem. We show this variance is reduced under LH, but the effect of AV depends on the problem parameters. We use this example to motivate the computational results of Section 3, where the variance reduction benefits of LH and AV are shown to be highly problem dependent.

In addition to variance reduction, we show the AV and LH sampling procedures may also reduce the bias of $z^*_{\mathrm{MP}_N}$. Under fairly general conditions, the solution to MP_N approaches that of MP with probability 1 as the number of realizations N increases (Dupačová and Wets, 1988). However, the solution to MP_N is biased in the sense that the expectation of the optimal objective function value of MP_N is less than that of MP. Mak, Morton, and Wood (1999) show that:

$$\mathbb{E}_{(\omega_1,\ldots,\omega_N)}\left[z^*_{\mathrm{MP}_{N(\omega_1,\ldots,\omega_N)}}\right] \leq \mathbb{E}_{(\omega_1,\ldots,\omega_{N+1})}\left[z^*_{\mathrm{MP}_{N+1(\omega_1,\ldots,\omega_{N+1})}}\right] \leq z^*_{\mathrm{MP}}\ \forall N. \tag{1.1}$$

A related issue is that the optimal solution $x^*_N(\omega_1,\ldots,\omega_N)$ of MP_N may be suboptimal with respect to the objective function $\mathbb{E}_\omega\left[Q(x,\omega)\right]+g(x)$ of MP. We refer to:

$$\mathbb{E}_\omega\left[Q(x^*_{\mathrm{MP}_{N(\omega_1,\ldots,\omega_N)}},\omega)\right] + g\left(x^*_{\mathrm{MP}_{N(\omega_1,\ldots,\omega_N)}}\right) \tag{1.2}$$

as the *actual* cost of the sample path problem and $z^*_{\mathrm{MP}_N}$ as the *perceived* cost.

In Section 2 we show that in the context of the newsvendor problem, AV and LH bring the both perceived and actual costs closer to z^*_{MP}. (The effect on the perceived cost is equivalent to reducing the bias of $z^*_{\mathrm{MP}_N}$.) This example is again used to motivate results in Section 3, where we examine bias reduction in a number of computational examples. This computational work extends a related paper by Linderoth et al. (2006), which examines the impact of LH on the bias of $z^*_{\mathrm{MP}_N}$ and on an upper bound for z^*_{MP} with a set of empirical examples.

2. The Newsvendor Problem

In this section we develop an example based on the newsvendor problem in order to illustrate the effects of different sampling procedures. The newsvendor problem can be expressed as a two-stage stochastic program as follows. In the first stage we choose an order quantity x. After demand D has been realized, we decide how much of the available stock y to sell. Assume demand is uniformly distributed on the interval $[0,1]$, and there is a shortage cost

$\alpha \in (0,1)$ and an overage cost $1-\alpha$. The second stage problem is

$$\text{P}: \quad Q(x,D) \stackrel{\text{def}}{=} \min_y \{(1-\alpha)(x-y) + \alpha(D-y) \mid y \le x, y \le D\}.$$

The solution to P is $\min(x,D)$. Let $TC(x)$ be the expected total cost associated with order quantity x:

$$\begin{aligned} TC(x) &\stackrel{\text{def}}{=} \mathbb{E}[Q(x,D)] \\ &= \mathbb{E}\left[\min_y \{(1-\alpha)(x-y) + \alpha(D-y) \mid y \le x, y \le D\}\right], \end{aligned} \tag{2.1}$$

so MP is $\min_x TC(x)$. Furthermore:

$$\begin{aligned} TC(x) &= (1-\alpha)\mathbb{E}(x-D)^+ + \alpha\mathbb{E}(D-x)^+ \\ &= (1-\alpha)\int_0^x (x-z)dz + \alpha \int_x^1 (z-x)dz \\ &= (1-\alpha)\frac{x^2}{2} + \alpha\frac{(1-x)^2}{2}. \end{aligned}$$

The cost-minimizing solution is therefore $x^* = \alpha$, and the optimal expected total cost is:

$$TC^* \stackrel{\text{def}}{=} TC(\alpha) = (1-\alpha)\frac{\alpha^2}{2} + \alpha\frac{(1-\alpha)^2}{2} = \frac{\alpha(1-\alpha)}{2}. \tag{2.2}$$

The N-sample path version of this problem is:

$$z^*_{\text{MP}_{N(D_1,\ldots,D_N)}} \stackrel{\text{def}}{=} \min_x N^{-1} \sum_{i=1}^N \left[(1-\alpha)(x-D_i)^+ + \alpha(D_i - x)^+\right]. \tag{2.3}$$

The optimal solution $\hat{x}$ to (2.3) is the $\lceil \alpha N \rceil^{\text{th}}$order statistic of the demands $\{D_1,\ldots,D_N\}$. The k^{th} of N order statistics uniformly distributed on $(0,1)$ has a Beta distribution with parameters $k, N-k$ (see Hogg and Craig, 1978, p. 159); therefore, under independent sampling, $\hat{x}$ has a Beta distribution with parameters $\lceil \alpha N \rceil$ and $(N - \lceil \alpha N \rceil + 1)$.

We next compute $Var[z^*_{\text{MP}_N}]$, the variance with respect to demands $D_1,\ldots,D_N$. The analysis is based on the following expression for variance involving random variables X and Y:

$$Var(X) = Var_Y[E(X|Y)] + E_Y[Var(X|Y)].$$

(See, for example, Law and Kelton (2000).) The following equations hold under any sampling procedure:

$$
\begin{aligned}
Var[z^*_{\text{MP}_{N(D_1,\ldots,D_N)}}] =& Var\left[\frac{1}{N}\sum_{i=1}^{N}(1-\alpha)(\hat{x}-D_i)^+ + \alpha(D_i-\hat{x})^+\right] \\
=& Var_{\hat{x}}\left[\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}(1-\alpha)(\hat{x}-D_i)^+ + \alpha(D_i-\hat{x})^+ \middle| \hat{x}\right]\right] \\
&+ \mathbb{E}_{\hat{x}}\left[Var\left[\frac{1}{N}\sum_{i=1}^{N}(1-\alpha)(\hat{x}-D_i)^+ + \alpha(D_i-\hat{x})^+ \middle| \hat{x}\right]\right]. \qquad (2.4)
\end{aligned}
$$

Under independent sampling, $\hat{x}$ has a Beta distribution with parameters $\lceil \alpha N \rceil$ and $(N - \lceil \alpha N \rceil + 1)$, so:

$$
\begin{aligned}
\mathbb{E}[\hat{x}] &= \frac{\lceil \alpha N \rceil}{N+1} &(2.5)\\
\mathbb{E}[\hat{x}^2] &= \frac{(\lceil \alpha N \rceil)(\lceil \alpha N \rceil + 1)}{(N+1)(N+2)} &(2.6)\\
Var[\hat{x}] &= \frac{(\lceil \alpha N \rceil)(N - \lceil \alpha N \rceil + 1)}{(N+1)^2(N+2)}. &(2.7)
\end{aligned}
$$

Since $\hat{x}$ is the $\lceil \alpha N \rceil^{th}$ order statistic, $\lceil \alpha N \rceil - 1$ of the demand values are uniformly distributed below $\hat{x}$, and the remaining $N - \lceil \alpha N \rceil$ are uniformly distributed above. The first term on the right side of (2.4) becomes:

$$
\begin{aligned}
&Var_{\hat{x}}\left[\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}(1-\alpha)(\hat{x}-D_i)^+ + \alpha(D_i-\hat{x})^+ \middle| \hat{x}\right]\right] \\
&= Var_{\hat{x}}\left[(1-\alpha)\left(\frac{\lceil \alpha N \rceil - 1}{N}\right)\frac{1}{\hat{x}}\int_0^{\hat{x}}(\hat{x}-z)dz + \alpha\left(\frac{N-\lceil \alpha N \rceil}{N}\right)\frac{1}{1-\hat{x}}\int_{\hat{x}}^{1}(z-\hat{x})dx\right] \\
&= \left(\frac{\lceil \alpha N \rceil - 1 + \alpha - \alpha N}{2N}\right)^2 Var(\hat{x}). \qquad (2.8)
\end{aligned}
$$

The second term on the right side of (2.4) becomes:

$$
\begin{aligned}
&\mathbb{E}_{\hat{x}}\left[Var\left[\frac{1}{N}\sum_{i=1}^{N}(1-\alpha)(\hat{x}-D_i)^+ + \alpha(D_i-\hat{x})^+ \middle| \hat{x}\right]\right] \\
&= \mathbb{E}_{\hat{x}}\left[\frac{1}{N^2}\left[(1-\alpha)^2(\lceil \alpha N \rceil - 1)\frac{\hat{x}^2}{12} + \alpha^2(N - \lceil \alpha N \rceil)\frac{(1-\hat{x})^2}{12}\right]\right]. \qquad (2.9)
\end{aligned}
$$

Combining (2.4) through (2.9) gives $Var[z^*_{\text{MP}_N}]$ under independent sampling.

We next examine the expected performance of $\hat{x}$ with respect to the original objective function $TC(\cdot)$. The expected *actual* total cost using the sample path optimization solution is:

$$\begin{aligned} \mathbb{E}_{\hat{x}}\left[TC(\hat{x})\right] &= \int_0^1 TC(z) f_{\hat{x}}(z)dz = \frac{1-\alpha}{2}\mathbb{E}\left[\hat{x}^2\right] + \frac{\alpha}{2}\left(1 - 2\mathbb{E}\left[\hat{x}\right] + \mathbb{E}\left[\hat{x}^2\right]\right) \\ &= \frac{1}{2}\mathbb{E}\left[\hat{x}^2\right] - \alpha\mathbb{E}\left[\hat{x}\right] + \frac{\alpha}{2}. \end{aligned} \tag{2.10}$$

The expected *perceived* cost of the sample path optimization solution (i.e., $\mathbb{E}[z^*_{\text{MP}_N}]$) is:

$$\begin{aligned} &\mathbb{E}_{D_1,\ldots,D_N}\left[N^{-1}\sum_{i=1}^{N}(1-\alpha)(\hat{x}-D_i)^+ + \alpha(D_i - \hat{x})^+\right] \\ &= \int_0^1 \left[(1-\alpha)\left(\frac{\lceil \alpha N\rceil - 1}{N}\right)\frac{1}{u}\int_0^u (u-z)dz + \alpha\left(\frac{N-\lceil\alpha N\rceil}{N}\right)\frac{1}{1-u}\int_u^1 (z-u)dz\right] f_{\hat{x}}(u)du \\ &= \frac{(1-\alpha)}{2}\left(\frac{\lceil\alpha N\rceil - 1}{N}\right)\mathbb{E}\left[\hat{x}\right] + \frac{\alpha}{2}\left(\frac{N-\lceil\alpha N\rceil}{N}\right)\left(1-\mathbb{E}\left[\hat{x}\right]\right). \end{aligned} \tag{2.11}$$

We derive the second line by conditioning on the value of $\hat{x}$, the $\lceil\alpha N\rceil^{\text{th}}$order statistic, in which case $\lceil\alpha N\rceil - 1$ of the demand values are uniformly distributed below $\hat{x}$, and the remaining $N - \lceil\alpha N\rceil$ are distributed above. The bias of $z^*_{\text{MP}_N}$ under independent sampling is computed by substituting (2.5) into (2.11) and subtracting (2.2).

Not surprisingly, both the expected actual and expected perceived costs approach the optimal cost (2.2) as N increases. If αN is integer, the expected actual cost of the sample path solution (SPS) computed by substituting (2.5) and (2.6) into (2.10) is:

$$\frac{\alpha}{2}\left[1 - \frac{\alpha N^2 + 4\alpha N - N}{(N+1)(N+2)}\right],$$

and the expected perceived cost of the sample path solution computed by substituting (2.5) into (2.11) is:

$$\frac{\alpha(1-\alpha)}{2}\left(\frac{N}{N+1}\right).$$

Therefore using (2.2):

$$\frac{\mathbb{E}_{\hat{x}}\left[TC(\hat{x})\right]}{TC^*} = \left(\frac{1}{1-\alpha}\right)\left(1 - \frac{\alpha N^2 + 4\alpha N - N}{(N+1)(N+2)}\right), \text{ and} \tag{2.12}$$

$$\frac{\mathbb{E}\left[z^*_{\text{MP}_{N(D_1,\ldots,D_N)}}\right]}{TC^*} = \frac{N}{N+1}. \tag{2.13}$$

Expressions (2.12) and (2.13) approach 1 (from above and below respectively) as N increases.

We can also use expressions (2.4) through (2.9), which provide $Var[z^*_{\text{MP}_N}]$, and expressions (2.5), (2.11) and (2.2), which provide $Bias[z^*_{\text{MP}_N}]$, to compute the MSE. It is worth

N \ α	0.6	0.7	0.8	0.9
2	7.4	2.0	79.7	85.3
5	27.6	41.0	25.9	51.7
10	19.4	19.3	19.0	18.0
100	2.9	2.9	2.9	2.9

Table 1: Percent of MSE due to bias in the newsvendor problem

pointing out that the relative importance of variance and bias to MSE in the newsvendor problem depends greatly on the choice of α and N. Table 1 presents the percent of MSE due to bias for various combinations of parameter values. Notice that when N is small and α is close to 1, bias becomes a much more significant issue than variance.

2.1 Antithetic Variates

We next examine the effect of antithetic variates on variance and bias for the newsvendor problem. Under AV we draw $N/2$ antithetic pairs $\{(D_i, 1 - D_i), i = 1, 2, \ldots, N/2\}$, rather than N independent values $\{D_1, \ldots, D_N\}$. These correlated values are used in the sample path problem (2.3).

In the subsequent analysis we suppose $\alpha > 1/2$, although similar computations can be performed for lower values. The solution to the sample path problem is still the $\lceil \alpha N \rceil^{\text{th}}$order statistic. With antithetic variates, we know that $N/2$ of the points will lie below $1/2$, so the solution to the sampled problem in this case, $\hat{x}_{AV}$, is the $\lceil \alpha N - N/2 \rceil^{\text{th}}$ order statistic of $N/2$ random variables uniformly distributed on $[1/2, 1]$. Hence $\hat{x}_{AV} = 1/2 + X/2$, where X has a Beta distribution with parameters $\lceil \alpha N - N/2 \rceil$ and $N/2 - \lceil \alpha N - N/2 \rceil + 1$. Therefore:

$$\mathbb{E}[\hat{x}_{AV}] = \frac{\lceil \alpha N \rceil + 1}{N + 2}, \tag{2.14}$$

and

$$Var[\hat{x}_{AV}] = \frac{(\lceil \alpha N \rceil - N/2)\,(N - \lceil \alpha N \rceil + 1)}{4\,(N/2 + 1)^2\,(N/2 + 2)}. \tag{2.15}$$

If αN and $(\alpha N - N/2)$ are integers, then $\mathbb{E}[\hat{x}] < \mathbb{E}[\hat{x}_{AV}] < \alpha$, so the expectation of the AV solution is closer to the optimal solution of the original problem. Furthermore $\text{Var}\,[\hat{x}_{AV}] < \text{Var}[\hat{x}]$.

To compute $Var[z^*_{\text{MP}_N}]$ under AV, we again use (2.4) and condition on the value of $\hat{x}_{AV}$. The first term on the right side of (2.4) becomes (2.16). We develop this expression by considering the $N/2$ points that lie between $1/2$ and 1 and their corresponding antithetic pairs. The first term inside the brackets corresponds to $\hat{x}_{AV}$ and its antithetic partner. The second term corresponds to the $\lceil \alpha N \rceil - N/2 - 1$ points lying below $\hat{x}_{AV}$ but above $1/2$ and their antithetic partners lying between $1 - \hat{x}_{AV}$ and $1/2$. The third term corresponds to the $N - \lceil \alpha N \rceil$ points above $\hat{x}_{AV}$ and their partners, lying between 0 and $1 - \hat{x}_{AV}$. The points other than $\hat{x}_{AV}$ and $1 - \hat{x}_{AV}$ are uniformly distributed in the given range conditioned on the value of $\hat{x}_{AV}$. This yields the following expression for the variance:

$$
\begin{aligned}
&Var_{\hat{x}_{AV}} \left[\mathbb{E} \left[\frac{1}{N} \sum_{i=1}^{N} (1-\alpha)(\hat{x}_{AV} - D_i)^+ + \alpha(D_i - \hat{x}_{AV})^+ \middle| \hat{x}_{AV} \right] \right] \\
=&Var_{\hat{x}_{AV}} \left[\frac{(1-\alpha)(\hat{x}_{AV} - (1 - \hat{x}_{AV}))}{N} \right. \\
&\quad + (1-\alpha) \left(\frac{\lceil \alpha N \rceil - N/2 - 1}{N} \right) \frac{1}{\hat{x}_{AV} - 1/2} \int_{1/2}^{\hat{x}_{AV}} ((\hat{x}_{AV} - z) + (\hat{x}_{AV} - 1 + z)) dz \\
&\quad \left. + \left(\frac{N - \lceil \alpha N \rceil}{N} \right) \frac{1}{1 - \hat{x}_{AV}} \int_{\hat{x}_{AV}}^{1} \left[\alpha(z - \hat{x}_{AV}) + (1-\alpha)(\hat{x}_{AV} - 1 + z) \right] dz \right] \qquad (2.16) \\
=&Var_{\hat{x}_{AV}} \left[\frac{(-2\alpha N + N + \lceil \alpha N \rceil)\hat{x}_{AV} - \lceil \alpha N \rceil + \alpha N}{2N} \right] \qquad (2.17)
\end{aligned}
$$

The second term on the right side of (2.4) becomes:

$$
\begin{aligned}
&\mathbb{E}_{\hat{x}_{AV}} \left[Var \left[\frac{1}{N} \sum_{i=1}^{N} (1-\alpha)(\hat{x}_{AV} - D_i)^+ + \alpha(D_i - \hat{x}_{AV})^+ \middle| \hat{x}_{AV} \right] \right] \\
=&E_{\hat{x}_{AV}} \left[(N - \lceil \alpha N \rceil) \frac{(1 - \hat{x}_{AV})^2}{12N^2} \right] \\
=&\frac{N - \lceil \alpha N \rceil}{12N^2} \left(Var[\hat{x}_{AV}] + E[\hat{x}_{AV}]^2 - 2E[\hat{x}_{AV}] + 1 \right) \qquad (2.18)
\end{aligned}
$$

Combining (2.4) with (2.14), (2.15), (2.17), and (2.18) gives $Var[z^*_{\text{MP}_N}]$.

To derive the expected perceived cost under AV, we again condition on the value of $\hat{x}_{AV}$. As above, the three terms in the integrand correspond to the antithetic partner of $\hat{x}_{AV}$, those demand values lying below $\hat{x}_{AV}$, and those demand values lying above $\hat{x}_{AV}$ (but whose antithetic partners lie below):

$$\int_{1/2}^{1} [N^{-1}(1-\alpha)(u-(1-u))$$
$$+ N^{-1}(1-\alpha)(\lceil \alpha N \rceil - N/2 - 1)\left(\frac{1}{u-1/2}\right)\int_{1/2}^{u}((u-z)+(u-1+z))dz$$
$$+ N^{-1}(N-\lceil \alpha N \rceil)\frac{1}{1-u}\int_{u}^{1}(\alpha(z-u)+(1-\alpha)(u-1+z))dz\ ]\ f_{\hat{x}_{AV}}(u)du$$
$$= \frac{(-2N\alpha + N + \lceil \alpha N \rceil)\mathbb{E}[\hat{x}_{AV}] - \lceil \alpha N \rceil + \alpha N}{2N}.$$

The expected actual cost under AV is computed from (2.10) using the distribution of $\hat{x}_{AV}$. Figure 1 plots the expected actual and perceived costs as percentages of the optimal cost with both independent sample paths and antithetic pairs of sample paths, using a cost ratio of $\alpha = 0.8$. (To facilitate comparison across sampling methods, this figure includes lines for Latin Hypercube sampling which will be discussed in the next section.) Note that the use of antithetic pairs reduces the gaps between the expected cost of the optimal solution and both the actual and perceived costs of the sample path solution. In particular, the bias of $z^*_{\mathrm{MP}_N}$ is reduced.

If αN is integer, the expected actual cost of the sample path solution under AV simplifies to:

$$\frac{\alpha(1-\alpha)N^2 + (8\alpha-1)(1-\alpha)N + 2}{2(N+2)(N+4)},$$

and the expected perceived cost under AV is:

$$\frac{1-\alpha}{2}\frac{\alpha N + 1}{N+2}.$$

Therefore using (2.2), under AV we have:

$$\frac{\mathbb{E}_{\hat{x}}\left[TC(\hat{x})\right]}{TC^*} = \frac{N^2}{(N+2)(N+4)} + \frac{(8\alpha-1)(1-\alpha)N+2}{\alpha(1-\alpha)(N+2)(N+4)}; \tag{2.19}$$

$$\frac{\mathbb{E}\left[z^*_{\mathrm{MP}_N(D_1,\ldots,D_N)}\right]}{TC^*} = \frac{\alpha N + 1}{\alpha(N+2)}. \tag{2.20}$$

Expressions (2.19) and (2.20) approach 1 (from above and below respectively) as N increases.

Interestingly, while the use of AV decreases the bias for all values of α in the range (1/2, 1), it increases the variance for some values of α and N. The combination of the two effects is reflected in the MSE. Figure 2 shows the change in MSE obtained by using antithetic pairs.

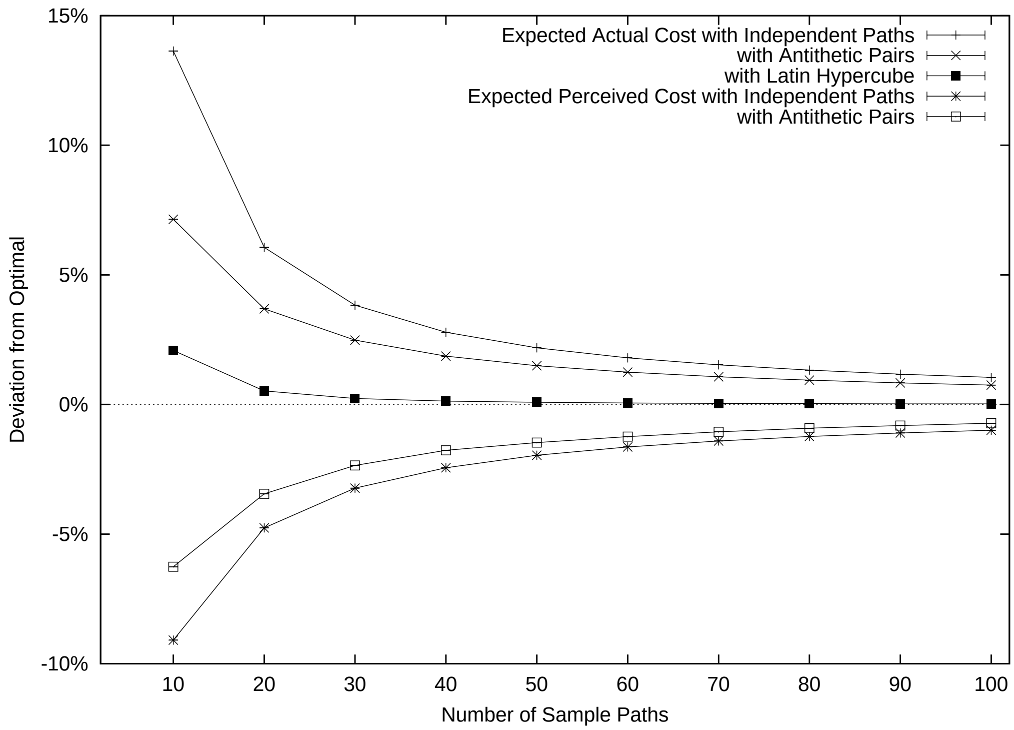

Figure 1: Expected performance of sample path solution under IS, AV, and LH for the newsvendor problem, as a function of the number of sample paths

Note that for $\alpha = 0.6$ and $\alpha = 0.7$, use of antithetic pairs increases MSE (the "reduction" is negative). As we will see in Section 3, it is possible, although apparently rare, for AV and LH to increase MSE.

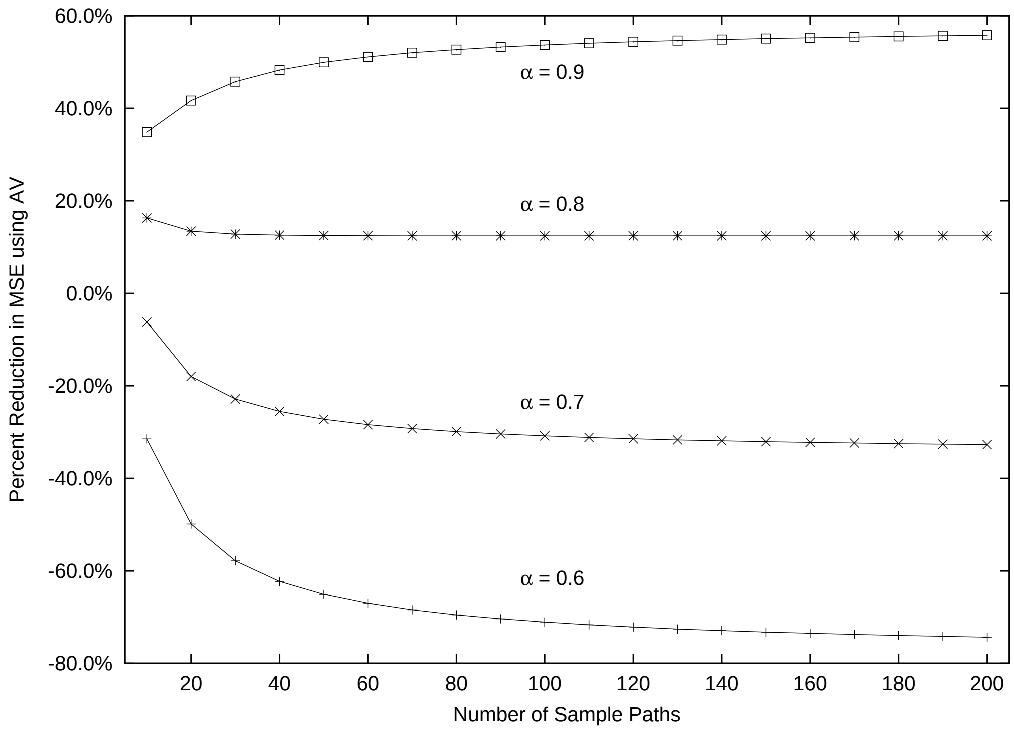

Figure 2: Change in MSE obtained by using antithetic variates instead of independent samples for the newsvendor problem, as a function of the number of sample paths

2.2 Latin Hypercube Sampling

We can also attack the variance and bias of $z^*_{\mathrm{MP}_N}$ using Latin Hypercube sampling, another technique usually prescribed for variance reduction (McKay et al., 1979). In this one-dimensional problem, we divide the interval $[0,1]$ into N equal segments; the i^{th} demand value D_i is drawn uniformly from the i^{th} segment. The solution to the sample path problem under Latin Hypercube sampling, $\hat{x}_{LH}$, is the demand value drawn from the $\lceil \alpha N \rceil^{\mathrm{th}}$ segment, which is uniformly distributed on $[(\lceil \alpha N \rceil - 1)/N, \lceil \alpha N \rceil / N]$. Under AV we have $\mathbb{E}[\hat{x}] < \mathbb{E}[\hat{x}_{AV}] < \alpha$ when αN is integer. Similarly we now have $\mathbb{E}[\hat{x}_{LH}] < \alpha$; however the relationships between $\mathbb{E}[\hat{x}_{LH}]$ and the values $\mathbb{E}[\hat{x}_{AV}]$ and $\mathbb{E}[\hat{x}]$ depend on the choice of α and N.

Since under LH the i^{th} demand value D_i is uniformly distributed on $[(i-1)/N, i/N]$, we compute $Var[z^*_{\text{MP}_N(D_1,\ldots,D_N)}]$ directly:

$$
\begin{aligned}
&Var[z^*_{\text{MP}_N(D_1,\ldots,D_N)}] \\
&\quad = Var\left[\frac{1}{N}\sum_{i=1}^{N}(1-\alpha)(\hat{x}_{LH}-D_i)^+ + \alpha(D_i-\hat{x}_{LH})^+\right] \\
&\quad = \frac{1}{N^2}Var\left[\sum_{i=1}^{\lceil\alpha N\rceil-1}(1-\alpha)(D_{\lceil\alpha N\rceil}-D_i) + \sum_{i=\lceil\alpha N\rceil+1}^{N}\alpha(D_i-D_{\lceil\alpha N\rceil})\right] \\
&\quad = \frac{1}{N^2}\left[(\lceil\alpha N\rceil-1+\alpha-\alpha N)^2 Var[D_{\lceil\alpha N\rceil}] + (1-\alpha)^2\sum_{i=1}^{\lceil\alpha N\rceil-1}Var[D_i] + \alpha^2\sum_{i=\lceil\alpha N\rceil+1}^{N}Var[D_i]\right] \\
&\quad = \frac{\lceil\alpha N\rceil(\lceil\alpha N\rceil-2\alpha N-1)+\alpha N(\alpha N-\alpha+2)}{12N^4}.
\end{aligned}
$$

The derivation of the expected perceived cost under LH is also straightforward:

$$
\begin{aligned}
&\mathbb{E}_{D_1,\ldots,D_N}\left[N^{-1}\sum_{i=1}^{N}(1-\alpha)(\hat{x}_{LH}-D_i)^+ + \alpha(D_i-\hat{x}_{LH})^+\right] \\
&= N^{-1}\mathbb{E}_{D_1,\ldots,D_N}\left[\sum_{i=1}^{\lceil\alpha N\rceil-1}(1-\alpha)(D_{\lceil\alpha N\rceil}-D_i) + \sum_{i=\lceil\alpha N\rceil+1}^{N}\alpha(D_i-D_{\lceil\alpha N\rceil})\right] \\
&= N^{-1}\left[\sum_{i=1}^{\lceil\alpha N\rceil-1}N^{-1}(1-\alpha)(\lceil\alpha N\rceil-i) + \sum_{i=\lceil\alpha N\rceil+1}^{N}N^{-1}\alpha(i-\lceil\alpha N\rceil)\right] \\
&= \frac{\lceil\alpha N\rceil(\lceil\alpha N\rceil-2\alpha N-1)+\alpha N(N+1)}{2N^2}.
\end{aligned}
$$

When αN is integer this expression reduces to $\alpha(1-\alpha)/2$; comparing this to (2.2) we see the perceived cost estimate is unbiased. The expected actual cost under LH is computed from (2.10):

$$
\begin{aligned}
&\mathbb{E}_{D_1,\ldots,D_N}\left[\hat{x}_{LH}^2\right]/2 - \alpha\mathbb{E}_{D_1,\ldots,D_N}\left[\hat{x}_{LH}\right] + \alpha/2 \\
&= \frac{1}{2}\left[\frac{1}{12N^2} + \left(\frac{\lceil\alpha N\rceil}{N}-\frac{1}{2N}\right)^2\right] - \alpha\left(\frac{\lceil\alpha N\rceil}{N}-\frac{1}{2N}\right) + \frac{\alpha}{2} \\
&= \frac{1}{8N^2}\left[\frac{1}{3}+(2\lceil\alpha N\rceil-1)^2\right] - \frac{\alpha}{2N}(2\lceil\alpha N\rceil-1) + \frac{\alpha}{2}.
\end{aligned}
$$

When αN is integer this expression reduces to $\alpha(1-\alpha)/2 + 1/(6N^2)$. Figure 1 plots the expected actual cost under Latin Hypercube sampling as a percentage of the optimal cost. We see that LH reduces the gap between the expected actual cost of the sample path solution and the optimal cost more effectively than AV.

As with AV, the use of LH also can also reduce the variance of the optimal objective function estimator (i.e., the variance of the perceived cost) for the stochastic LP. This is true for any value of α when N is greater than 3. Again, the combination of the two effects is reflected in the MSE, although as we noted above, the bias for this particular newsvendor case is equal to zero, so the MSE and the variance are equal. Figure 3 shows the dramatic reduction obtained in MSE by using Latin Hypercube sampling.

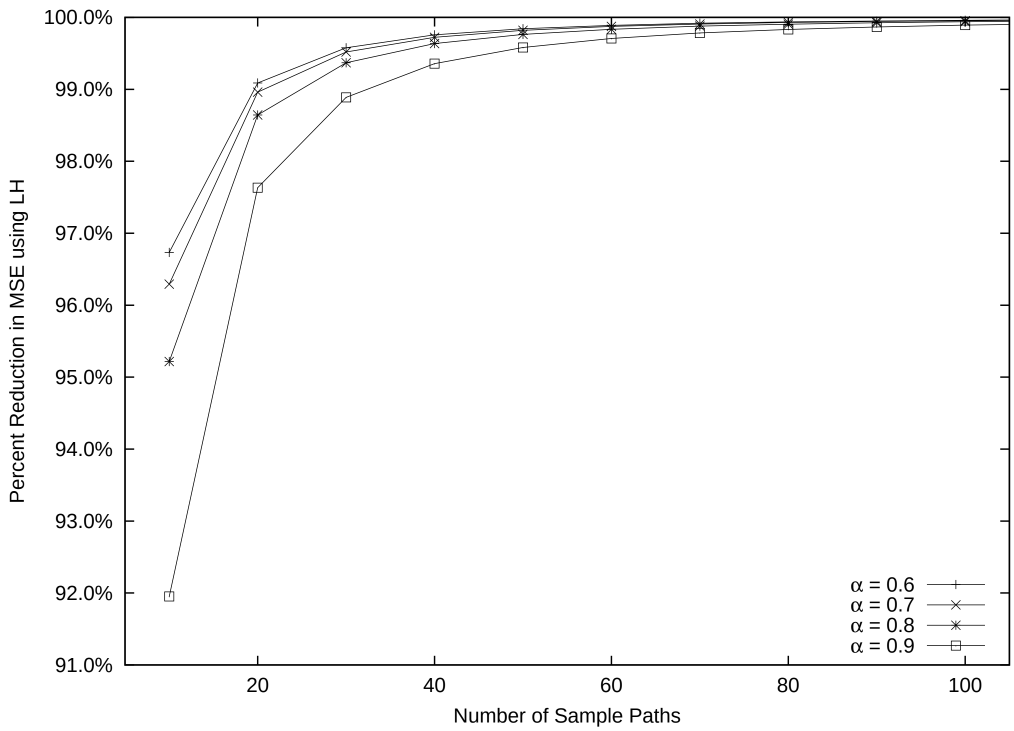

Figure 3: Change in MSE obtained by using Latin Hypercube sampling for the newsvendor problem, as a function of the number of sample paths

3. Computational Examples

In Sections 2.1 and 2.2 we analytically characterized the effects of sampling methods on the bias and variance of the solution to a simple sample path problem. Here, we present

Name	Application	Source	Scenarios
20term	Vehicle Positioning	Mak et al. (1999)	1.1×10^{12}
apl-1p	Power Expansion Planning Problem	Infanger (1992)	1284
fleet	Fleet Planning	Powell and Topaloglu (2005)	8.5×10^{113}
gbd	Aircraft Allocation	Dantzig (1963)	6.5×10^{5}
LandS	Electrical Investment Planning	Louveaux and Smeers (1988)	10^{6}
snip	Stochastic Network Interdiction	Janjarassuk and Linderoth (2006)	3.7×10^{19}
ssn	Telecommunication Network Design	Sen et al. (1994)	10^{70}
storm	Flight Scheduling	Mulvey and Ruszczyński (1995)	6×10^{81}

Table 2: Description of test instances

empirical results of applying these sampling methods to a set of more complicated test problems. The section contains a brief description of our test problems, a description of our approach for obtaining statistical estimates for perceived and actual cost, a description of our computational platform, and the results of the experiments.

3.1 Test Problems

The test problems are two-stage stochastic linear programs with recourse that were obtained from the literature. Table 2 contains details about each of the problems. The problem *fleet* is a fleet management problem available from the page `http://www.orie.cornell.edu/~huseyin/research/research.html#Fleet_20_3`. The problem *snip* is a (linear relaxation) of a stochastic network interdiction problem available at the page `http://coral.ie.lehigh.edu/sp-instances/`. The remaining problems are described in Linderoth et al. (2006) and available from the companion web site `http://www.cs.wisc.edu/~swright/stochastic/sampling/`.

3.2 Methodology

Perceived Cost Estimates As indicated by the inequalities in (1.1), and previously shown by Norkin et al. (1998) and Mak et al. (1999), the expected perceived cost:

$$\mathbb{E}_{(\omega_1,\ldots,\omega_N)}\left[z^*_{\mathrm{MP}_{N(\omega_1,\ldots,\omega_N)}}\right]$$

is a biased estimate of z^*_{MP}, the value of the optimal solution. First, we generate M independent (and identically distributed) samples of size N: $(\omega_1^1,\ldots,\omega_N^1),\ldots,(\omega_1^M,\ldots,\omega_N^M)$. We define $\ell_j, j = 1, 2, \ldots, M$, to be the solution value of the jth sample path problem:

$$\ell_j \stackrel{\text{def}}{=} z^*_{\mathrm{MP}_{N(\omega_1^j,\ldots,\omega_N^j)}}$$

and compute the value:

$$\mathcal{L}_{N,M} \stackrel{\text{def}}{=} \frac{1}{M} \sum_{j=1}^{M} \ell_j.$$

The statistic $\mathcal{L}_{N,M}$ provides an unbiased estimate of $\mathbb{E}_{(\omega_1,\ldots,\omega_N)}\left[z^*_{\mathrm{MP}_{N(\omega_1,\ldots,\omega_N)}}\right]$. Since the M samples are i.i.d, we can construct an approximate $(1-\alpha)$ confidence interval for $\mathbb{E}_{(\omega_1,\ldots,\omega_N)}\left[z^*_{\mathrm{MP}_{N(\omega_1,\ldots,\omega_N)}}\right]$:

$$\left[\mathcal{L}_{N,M} - \frac{z_{\alpha/2} s_{\mathcal{L}}(M)}{\sqrt{M}}, \mathcal{L}_{N,M} + \frac{z_{\alpha/2} s_{\mathcal{L}}(M)}{\sqrt{M}}\right], \tag{3.1}$$

where

$$s_{\mathcal{L}}(M) \stackrel{\text{def}}{=} \sqrt{\frac{1}{M-1} \sum_{j=1}^{M} \left(\ell^j - \mathcal{L}_{N,M}\right)^2}. \tag{3.2}$$

For small values of M, one can use $t_{\alpha/2,M-1}$ critical values instead of $z_{\alpha/2}$, which will produce slightly bigger confidence intervals.

Actual Cost Estimates Since a solution to the sample path problem MP_N may be suboptimal with respect to the true objective function, we estimate the expected actual cost of $x^*_N(\omega_1, \ldots, \omega_N)$, an optimal solution to MP_N. We estimate the expected actual cost of a sample path problem of size N in the following manner. First, we generate M samples of size N: $(\omega_1^1, \ldots, \omega_N^1), \ldots, (\omega_1^M, \ldots, \omega_N^M)$ and solve the sample-path problem MP_N for each sample yielding:

$$x^*_j \in \arg \min_{Ax=b, x\geq 0} N^{-1} \sum_{i=1}^{N} Q_i(x, \omega_i^j) + g(x), \;\; j = 1, 2, \ldots, M.$$

Note that this is the same calculation necessary to compute a lower bound on the optimal objective value, and the computational effort required is to solve M sample path problems, each containing N scenarios. Next, for each candidate solution x^*_j, we take a new, Latin Hypercube sample of size N', $(\omega_1^j, \ldots, \omega_{N'}^j)$ and compute the quantity:

$$a_j = \sum_{i=1}^{N'} Q(x^*_j, \omega_i^j) + g(x^*_j). \tag{3.3}$$

Latin Hypercube sampling appears to be superior to the other two methods for variance reduction; thus, we use this technique to estimate expected actual cost no matter what sampling method was used to obtain x^*_j. Since x^*_j is fixed, this computation required the solution of N' independent linear programs. The quantity:

$$\mathcal{A}_{N,M} \stackrel{\text{def}}{=} \frac{1}{M} \sum_{j=1}^{M} a_j$$

is an unbiased estimate of the expected actual cost:

$$\mathbb{E}_{(\omega_1,\ldots,\omega_N)} \left[\mathbb{E}_\omega \left[Q(x^*_{\text{MP}_{N(\omega_1,\ldots,\omega_N)}}, \omega) \right] + g(x^*_{\text{MP}_{N(\omega_1,\ldots,\omega_N)}}) \right].$$

Since the random quantities a^j are i.i.d., we can similarly construct an approximate $(1-\alpha)$ confidence interval for this quantity.

3.3 Computational Platform

The computational experiments presented here were performed on a non-dedicated, distributed computing platform known as a *computational grid* (Foster and Kesselman, 1999). The computational platform was created with the aid of the Condor software toolkit (Livny et al., 1997), which can be configured to allow for the idle cycles of machines to be donated to a "Condor pool".

In order to create the sampled problems, we use the SUTIL software toolkit (Czyzyk et al., 2005). Specifically for this work, SUTIL was equipped with the ability to sample two-stage stochastic programs using an antithetic variates sampling technique. An important feature of SUTIL, necessary when running in a distributed and heterogeneous computing environment, is its ability to obtain the *same* value for a random vector ω^j on different processors and at different points of the solution algorithm (say different iterations of the LShaped method). This is a nontrivial implementation issue and is accomplished in SUTIL by employing an architecture- and operating-system-independent random number stream, storing and passing appropriate random seed information to the participating processors, and performing some recalculation of random vectors in the case that the vectors in a sample are correlated.

In order to solve the sampled problems, we use the code **atr** of Linderoth and Wright (2003). The algorithm is a variation of the well-known LShaped algorithm (Van Slyke and Wets, 1969) that has been enhanced with mechanisms for reducing the synchronization requirements of the algorithm (useful for the distributed computing environment), and also with a $\|\cdot\|_\infty$-norm trust region to help stabilization of the master problem. The initial iterate of the algorithm was taken to be the solution of a sampled instance of intermediate size.

3.4 Computational Results

Our computational experiments were designed to examine the impact of different sampling methods on the bias and variance of the perceived cost of 2-stage stochastic linear programs solved via sample path optimization. Recall that bias and variance reduction combine to improve the mean squared error of the solution to the sample path problem. In the results presented here, the optimal solution value to the full problem, z^*_{MP}, is unknown, so we cannot calculate the bias. Since we know that for minimization problems, the expected value of the sample path solution, $\mathbb{E}\left[z^*_{\text{MP}_N}\right]$, is less than the true optimal solution, z^*_{MP}, we can test whether or not one sampling method reduces bias as compared to another by testing whether or not the expected value of the sample path solution, $\mathbb{E}\left[z^*_{\text{MP}_N}\right]$, is significantly larger and therefore closer to the true optimal solution, z^*_{MP}.

Using samples drawn in an independent fashion, samples drawn using antithetic variates, and samples drawn using Latin Hypercube sampling, the following experiment was performed. For all instances described in Table 2 except **apl-1p**, confidence intervals for both expected perceived cost and expected actual cost (as defined in Section 3.2) were computed for $M = 1000$ for $N \in \{10, 25, 50, 75, 100\}$, $M = 50$ for $N \in \{500, 1000\}$, and $M = 10$ for $N \in \{5000, 10000, 50000\}$. The value N' used in the calculation of a_j (3.3) was $N' = 500$ for $M \in \{10, 25, 50, 75, 100\}$ and $N' = 20,000$ for $N \in \{500, 1000, 5000, 10000, 50000\}$. In all tables and figures, we use $z_{0.975} \approx 1.96$ when $M = 50, 1000$ ($N \in \{10, 25, 50, 75, 100, 500, 1000\}$) and $t_{0.025,M-1} \approx 2.685$ when $M = 10$ ($N \in \{5000, 10000, 50000\}$). Since the full problem **apl-1p** has only 1284 realizations, we only perform the smaller experiments with $M \in \{10, 25, 50, 75, 100\}$. The complete experiment required the solution of more than one billion linear programs, so the ability to run in the powerful distributed setting of the computational grid was of paramount importance to this work.

Confidence intervals for expected perceived cost and expected actual cost are contained in the appendix. Table 3 shows the results of t-tests for bias reduction and F-tests for variance reduction for the expected perceived cost. (Since each trial was independently generated, and variances are significantly different in some cases, we use one-sided, unpaired student t-tests, assuming unequal variance.) We use the symbol $\succ$ to indicate when a test assumes one method is preferred to another. Note that for this set of problems, statistically significant bias reduction occurs with AV occasionally and with LH frequently. Both AV and LH sampling methods are effective in reducing variance, with LH reducing variance as compared to IS in almost all cases. Table 4 shows the percentage reduction in bias and variance for cases where the reduction is statistically significant. As discussed above, both bias and variance combine to determine the quality of the lower bound estimate. Table 5 shows the percentage reduction in mean squared error for cases where there is a significant reduction bias or variance.

Reductions in bias and variance combine to reduce MSE. Whether bias or variance improvements are responsible for the biggest improvement in MSE is problem specific. We consider two of the eight problems here. Results from the full set of experiments are availabile in the appendix. Figures 4 and 5 show estimates of expected perceived and expected actual cost with confidence intervals for problem **LandS**. A horizontal reference line is shown on the perceived and actual cost figures for each problem. For all problems except **apl-1p**, we do not know the optimal objective value, so we use the expected actual cost for $N = 50,000$ with LH sampling as the reference value. In Figure 5, one can easily see that both AV and LH dramatically reduce variance as compared to IS. The statistical tests reveal reduction in bias for some values of N, but the improvement in the estimate of $\mathbb{E}[z^*_{\mathrm{MP}_N}]$ for this problem comes mostly from variance reduction. Consider the $N = 100$ case specifically. With independent sampling and $N = 100$, the bias is approximately 0.38 while the variance is approximately 34.4. With LH sampling, the bias and variance estimates are 0.02 and 0.05. So, while bias is substantially reduced, it is the variance reduction that drives a large improvement in MSE.

Figures 6 and 7 tell a different story for problem **ssn**. For this problem, bias reduction plays a far more substantial role in improving MSE. Again, consider the $N = 100$ case. Under IS, bias is approximately 3.03 and variance is approximately 2.95. Under LH, bias and variance are approximately 1.03 and 1.4. For the $N = 10$ case with **ssn**, LH and AV decrease bias but actually *increase* variance. These were the only instances in our experiments where variance was increased in a statistically significant way. This result appears to be driven by the fact that the random variables in this problem represent demand demand flows on a capacitated network. With independent sampling and $N = 10$, the optimal objective value for the sampled problems is nearly always zero; all demand is met. The recurrence of this zero value gives low variance (albeit biased) estimate of the optimal objective value. LH and AV reduce the bias, but in doing so reduce the frequency of zero values and thus increase the variance. This phenomenon disappears for larger values of N.

		Bias Reduction (t-test)			Variance Reduction (F-test)		
Instance	N	AV $\succ$ IS	LH $\succ$ IS	LH $\succ$ AV	AV $\succ$ IS	LH $\succ$ IS	LH $\succ$ AV
20	10	0.0298	0.0041	0.0326	0.0000	0.0000	0.0001
	25	0.0173	0.0041		0.0000	0.0000	0.0000
	50			0.0049	0.0000	0.0000	0.0000
	75				0.0000	0.0000	0.0000
	100				0.0000	0.0000	0.0000
	500				0.0000	0.0000	
	1000				0.0000	0.0000	
apl-1p	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	25	0.0000	0.0000	0.0342	0.0000	0.0000	0.0000
	50		0.0022	0.0004	0.0000	0.0000	0.0000
	75	0.0277	0.0010		0.0000	0.0000	0.0000
	100		0.0011	0.0006	0.0000	0.0000	0.0000
fleet	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	25	0.0000	0.0000	0.0011	0.0000	0.0000	0.0000
	50	0.0204	0.0000	0.0014	0.0000	0.0000	0.0000
	75		0.0001	0.0084	0.0000	0.0000	0.0000
	100		0.0042		0.0000	0.0000	0.0000
	500				0.0003	0.0000	0.0263
	1000				0.0135	0.0000	0.0001
gbd	10		0.0005	0.0244	0.0000	0.0000	0.0000
	25				0.0000	0.0000	0.0000
	50		0.0102		0.0000	0.0000	0.0000
	75				0.0000	0.0000	0.0000
	100			0.0090	0.0000	0.0000	0.0000
	500					0.0000	0.0000
	1000					0.0000	0.0000
LandS	10				0.0000	0.0000	
	25		0.0259		0.0000	0.0000	0.0000
	50			0.0142	0.0000	0.0000	0.0000
	75	0.0104	0.0068		0.0000	0.0000	0.0000
	100	0.0279	0.0260		0.0000	0.0000	0.0000
	500				0.0000	0.0000	0.0000
	1000				0.0000	0.0000	0.0095
snip	10		0.0000	0.0000	0.0000	0.0000	0.0000
	25		0.0000	0.0000	0.0000	0.0000	0.0000
	50		0.0000	0.0000	0.0000	0.0000	0.0000
	75		0.0000	0.0000	0.0000	0.0000	0.0000
	100	0.0328	0.0000	0.0000	0.0000	0.0000	0.0000
	500				0.0006	0.0000	0.0000
	1000					0.0000	0.0000
ssn	10		0.0006	0.0028			
	25	0.0006	0.0000	0.0000		0.0000	0.0000
	50	0.0417	0.0000	0.0000		0.0000	0.0000
	75		0.0000	0.0000	0.0391	0.0000	0.0000
	100		0.0000	0.0000		0.0000	0.0000
	500		0.0000	0.0000		0.0010	0.0004
	1000		0.0010	0.0000		0.0302	0.0213
storm	10			0.0004	0.0000	0.0000	0.0000
	25				0.0000	0.0000	0.0000
	50			0.0024	0.0000	0.0000	0.0000
	75				0.0000	0.0000	0.0000
	100				0.0000	0.0000	0.0000
	500				0.0000	0.0000	0.0000
	1000				0.0000	0.0000	0.0000

Table 3: p-values (from t and F test) for cases where LH or AV sampling methods result in a statistically significant (at 5% or better) reduction of bias or variance

Instance	N	Bias Reduction (t-test)			Variance Reduction (F-test)		
		AV ≻ IS	LH ≻ IS	LH ≻ AV	AV ≻ IS	LH ≻ IS	LH ≻ AV
20	10	44.76 %	62.23 %	31.63 %	90.12 %	92.36 %	22.66 %
	25	55.30 %	68.12 %		84.17 %	88.01 %	24.22 %
	50			54.82 %	89.27 %	92.96 %	34.44 %
	75				88.58 %	92.49 %	34.24 %
	100				87.71 %	92.42 %	38.34 %
	500				91.52 %	92.72 %	
	1000				89.81 %	93.33 %	
apl-1p	10	58.68 %	88.01 %	70.98 %	67.19 %	97.41 %	92.10 %
	25	74.32 %	90.30 %	62.24 %	63.07 %	95.61 %	88.12 %
	50		75.96 %	67.69 %	69.06 %	98.56 %	95.33 %
	75	52.99 %	75.69 %		69.54 %	97.73 %	92.55 %
	100		73.31 %	62.14 %	69.82 %	98.54 %	95.16 %
fleet	10	45.19 %	64.26 %	34.79 %	50.81 %	74.26 %	47.66 %
	25	50.71 %	69.83 %	38.79 %	52.19 %	80.26 %	58.71 %
	50	35.13 %	71.08 %	55.43 %	42.83 %	79.92 %	64.87 %
	75		63.51 %	46.97 %	48.14 %	80.09 %	61.61 %
	100		50.32 %		42.72 %	78.79 %	62.97 %
	500				65.42 %	81.83 %	47.46 %
	1000				51.17 %	84.55 %	68.37 %
gbd	10		93.15 %	85.55 %	47.90 %	98.65 %	97.42 %
	25				43.03 %	95.97 %	92.93 %
	50		101.94 %		38.79 %	99.89 %	99.83 %
	75				44.25 %	98.57 %	97.44 %
	100			100.04 %	41.96 %	100.00 %	100.00 %
	500					100.00 %	100.00 %
	1000					100.00 %	100.00 %
LandS	10				99.38 %	98.90 %	
	25		90.47 %		95.32 %	99.68 %	93.14 %
	50			60.01 %	99.30 %	99.81 %	73.07 %
	75	88.69 %	93.77 %		97.87 %	99.82 %	91.72 %
	100	92.81 %	94.08 %		99.35 %	99.85 %	77.03 %
	500				98.83 %	99.85 %	87.01 %
	1000				99.64 %	99.83 %	52.91 %
snip	10		51.26 %	49.41 %	32.34 %	83.54 %	75.67 %
	25		77.12 %	74.33 %	28.30 %	90.57 %	86.85 %
	50		82.32 %	79.11 %	30.41 %	91.82 %	88.25 %
	75		75.01 %	74.58 %	33.51 %	93.27 %	89.87 %
	100	33.91 %	85.98 %	78.78 %	29.97 %	93.93 %	91.34 %
	500				63.53 %	92.28 %	78.83 %
	1000					94.86 %	93.50 %
ssn	10		1.39 %	1.24 %			
	25	4.35 %	53.83 %	51.73 %		24.60 %	31.43 %
	50	3.27 %	52.38 %	50.77 %		51.25 %	54.47 %
	75		53.43 %	53.16 %	12.25 %	47.70 %	40.40 %
	100		65.81 %	64.55 %		52.84 %	46.63 %
	500		102.27 %	102.27 %		61.80 %	64.56 %
	1000		66.89 %	78.15 %		46.61 %	48.70 %
storm	10			52.31 %	99.76 %	99.90 %	59.62 %
	25				95.89 %	99.92 %	98.10 %
	50			72.38 %	99.75 %	99.91 %	65.65 %
	75				98.22 %	99.91 %	95.08 %
	100				99.73 %	99.91 %	65.53 %
	500				98.40 %	99.87 %	91.91 %
	1000				99.65 %	99.92 %	77.83 %

Table 4: Estimates of percentage reduction in bias and variance for cases where bias and variance reduction are statistically significant (at 5% or better)

Instance	N	MSE Reduction AV $\succ$ IS	LH $\succ$ IS	LH $\succ$ AV
20	10	89.73%	92.23%	24.40%
	25	84.10%	88.04%	24.75%
	50	88.89%	92.89%	36.04%
	75	88.45%	92.39%	34.17%
	100	87.70%	92.32%	37.59%
	500	91.31%	92.76%	
	1000	87.99%	92.89%	
apl1p	10	68.81%	97.57%	92.20%
	25	64.54%	95.80%	88.16%
	50	68.89%	98.56%	95.38%
	75	69.74%	97.75%	92.57%
	100	69.73%	98.55%	95.22%
fleet	10	55.65%	77.54%	49.34%
	25	55.68%	81.84%	59.02%
	50	43.59%	80.51%	65.45%
	75	48.31%	80.34%	61.96%
	100	43.04%	78.67%	62.56%
	500	64.17%	82.24%	50.42%
	1000	54.20%	84.95%	67.13%
gbd	10	48.27%	98.66%	97.42%
	25	43.05%	95.98%	92.94%
	50	39.00%	99.89%	99.83%
	75	44.28%	98.57%	97.44%
	100	41.68%	100.00%	100.00%
	500		100.00%	100.00%
	1000		100.00%	100.00%
LandS	10	99.36%	98.89%	
	25	95.32%	99.68%	93.09%
	50	99.29%	99.81%	73.26%
	75	97.88%	99.82%	91.63%
	100	99.35%	99.85%	76.87%
	500	98.81%	99.85%	87.19%
	1000	99.62%	99.83%	55.61%
snip	10	34.00%	85.41%	77.90%
	25	29.56%	91.54%	87.99%
	50	31.52%	92.53%	89.09%
	75	35.64%	94.49%	91.44%
	100	29.61%	94.66%	92.41%
	500	67.14%	87.65%	62.43%
	1000		94.20%	93.17%
ssn	10		1.84%	1.68%
	25	7.06%	74.42%	72.48%
	50	4.45%	73.50%	72.26%
	75	3.58%	71.65%	70.60%
	100		79.66%	77.87%
	500		83.09%	83.66%
	1000		64.10%	76.96%
storm	10	99.76%	99.90%	59.62%
	25	95.89%	99.92%	98.10%
	50	99.75%	99.91%	65.65%
	75	98.22%	99.91%	95.08%
	100	99.73%	99.91%	65.53%
	500	98.40%	99.87%	91.91%
	1000	99.65%	99.92%	77.83%

Table 5: Estimates of percentage reduction in mean squared error for cases where there is a statistically significant (at 5% or better) reduction in bias or variance.

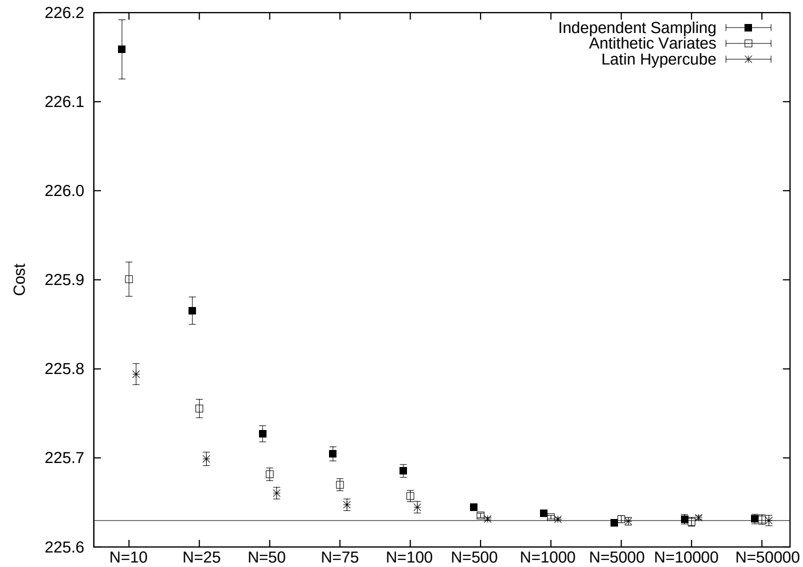

Figure 4: Expected Actual Cost Estimates for **LandS**

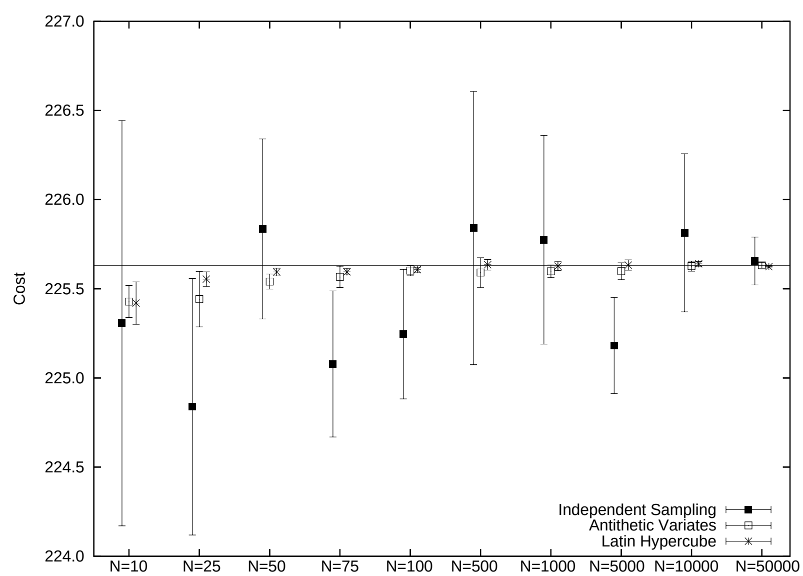

Figure 5: Expected Perceived Cost Estimates for **LandS**

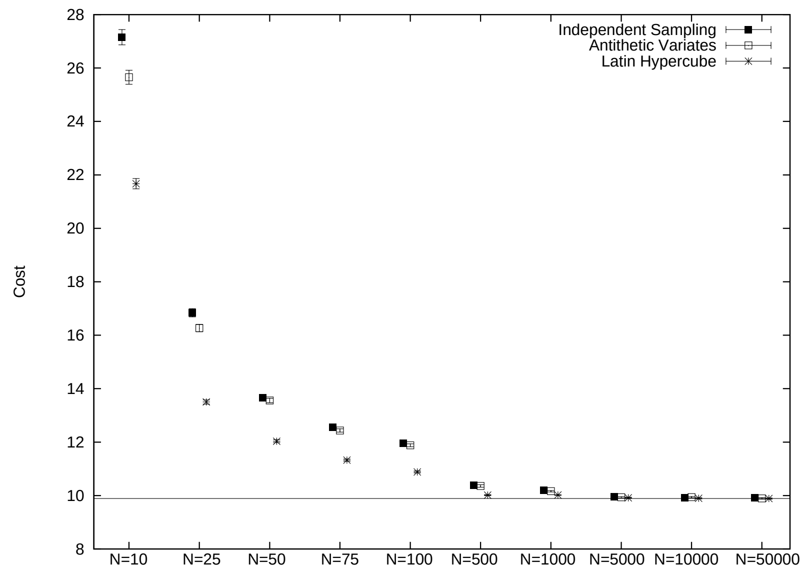

Figure 6: Expected Actual Cost Estimates Estimates for **ssn**

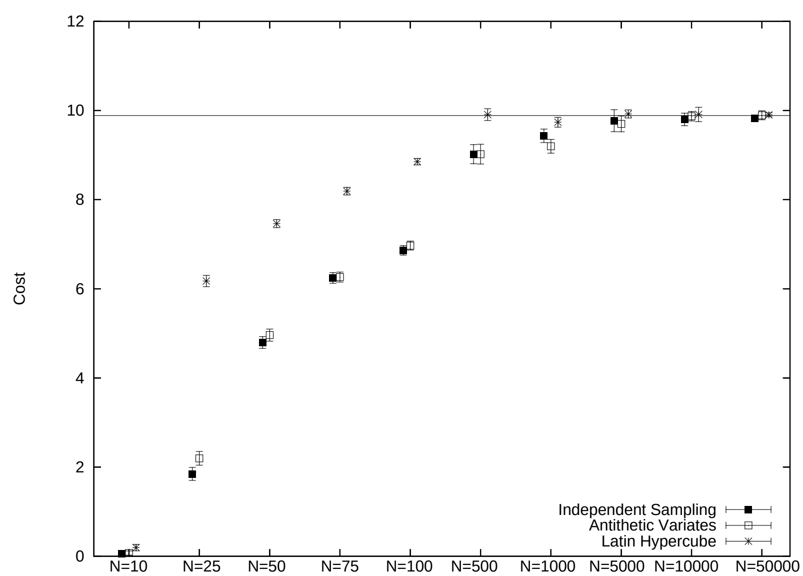

Figure 7: Expected Perceived Cost Estimates for **ssn**

4. Conclusion

Sample path optimization is a convenient method for solving stochastic programs; however a gap is introduced between the optimal solution and both the expected actual and expected perceived cost of the sample path solution. We have investigated two variations of sample path optimization where samples are drawn in antithetic pairs or using Latin Hypercube sampling. For a version of the simple newsvendor problem, we show that both the antithetic samples approach and the Latin Hypercube approach, techniques commonly used for variance reduction, reduce the solution *bias* as compared to sample path optimization with independent samples. For the newsvendor problem, the Latin Hypercube approach reduces variance of the sample path solution, while antithetic variates may increase or decrease the variance, depending on the cost parameters.

Using a computational grid, we perform extensive computational experiments investigating these same sampling methods on large-scale, two-stage, stochastic programs from the literature. We find that both sampling techniques are frequently effective at reducing variance, and Latin Hypercube sampling is often effective at reducing bias. The relative importance of bias and variance reduction is problem specific.

Acknowledgments

We would like to thank Shane Henderson and Tito Homen-de-Mello for their helpful comments on earlier versions of this paper. The work of author Linderoth is supported in part by the US National Science Foundation (NSF) under grant OCI-0330607 and by the Mathematical, Information and Computational Sciences subprogram of the Office of Science, US Department of Energy, under grant DE-FG02-05ER25694. Computational resources are provided in part by equipment purchased by the NSF through the IGERT Grant DGE-9972780. We also acknowledge the computing resources and support of the Graduate Education and Research Services Group at Penn State.

References

Beale, E. M. 1955. On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society* **17B** 173–184.

Birge, J. R., R. Louveaux. 1997. *Introduction to Stochastic Programming*. Springer, New York.

Czyzyk, J., J. Linderoth, J. Shen. 2005. SUTIL: A utility library for handling stochastic programs. user's manual. Available at `http://coral.ie.lehigh.edu/sutil`.

Dantzig, G. B. 1955. Linear programming under uncertainty. *Management Science* **1** 197–206.

Dantzig, G. B. 1963. *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ.

Dupačová, J., R. J.-B. Wets. 1988. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *The Annals of Statistics* **16** 1517–1549.

Foster, I., C. Kesselman. 1999. Computational grids. I. Foster, C. Kesselman, eds., *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann. Chapter 2.

Higle, J. L. 1998. Variance reduction and objective function evaluation in stochastic linear programs. *INFORMS Journal on Computing* **10** 236–247.

Hogg, R., A. Craig. 1978. *Introduction to mathematical statistics*. 4th ed. Macmillan.

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

Janjarassuk, U., J. Linderoth. 2006. Reformulation and sampling to solve a stochastic network interdiction problem. Technical Report 06T-001, Lehigh University Department of Industrial and Systems Engineering.

Kleywegt, A. J., A. Shapiro. 2001. Stochastic optimization. *Handbook of Industrial Engineering*, 3rd ed. John Wiley & Sons, New York, 2625–2649.

Law, A., D. Kelton. 2000. *Simulation Modeling and Analysis*. 3rd ed. McGraw-Hill, Boston.

Linderoth, J., A. Shapiro, S. Wright. 2006. The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research* **142** 215–241.

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

Livny, M., J. Basney, R. Raman, T. Tannenbaum. 1997. Mechanisms for high throughput computing. *SPEEDUP* **11**.

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

Mak, W. K., D. P. Morton, R. K. Wood. 1999. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters* **24** 47–56.

McKay, M. D., R. J. Beckman, W. J. Conover. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21** 239–245.

Mulvey, J. M., A. Ruszczyński. 1995. A new scenario decomposition method for large scale stochastic optimization. *Operations Research* **43** 477–490.

Norkin, V., G. Pflug, A. Ruszczyński. 1998. A branch and bound method for stochastic global optimization. *Mathematical Programming* **83** 425–450.

Powell, W., H. Topaloglu. 2005. Fleet management. S. Wallance, W. Ziemba, eds., *Applications of Stochastic Programming*. MPS-SIAM Series on Optimization, Philadelphia.

Sen, S., R. D. Doverspike, S. Cosares. 1994. Network planning with random demand. *Telecommunications Systems* **3** 11–30.

Van Slyke, R., R.J-B. Wets. 1969. L-shaped linear programs with applications to control and stochastic programming. *SIAM Journal on Applied Mathematics* **17** 638–663.