Assessing Solution Quality in Stochastic Programs

Güzin Bayraksan · David P. Morton
guzinb@mail.utexas.edu · morton@mail.utexas.edu
Graduate Program in Operations Research and Industrial Engineering,
The University of Texas at Austin, Austin, TX 78712

January 26, 2005

Abstract

Determining whether a solution is of high quality (optimal or near optimal) is a fundamental question in optimization theory and algorithms. In this paper, we develop Monte Carlo sampling-based procedures for assessing solution quality in stochastic programs. Quality is defined via the optimality gap and our procedures' output is a confidence interval on this gap. We review a multiple-replications procedure that requires solution of, say, 30 optimization problems and then, we present a result that justifies a computationally simplified single-replication procedure that only requires solving one optimization problem. Even though the single replication procedure is computationally significantly less demanding, the resulting confidence interval might have low coverage probability for small sample sizes for some problems. We provide variants of this procedure that require two replications instead of one and that perform better empirically. We present computational results for a newsvendor problem and for two-stage stochastic linear programs from the literature. We also discuss when the procedures perform well and when they fail and provide preliminary guidelines for selecting a candidate solution.

1 Introduction

We consider a stochastic optimization problem of the form

$$z^* = \min_{x \in X} Ef(x, \tilde{\xi}), \tag{SP}$$

where f is a real-valued function that determines the cost of operating with decision x under a realization of the random vector $\tilde{\xi}$, whose distribution is assumed known. $X \subseteq \mathbb{R}^n$ denotes the set of constraints that the decision vector x must obey and E is the expectation operator. As simple as it is to state, (SP) represents a large class of problems that can be found in the statistics and operations research literature. For instance, classical maximum likelihood estimation can be cast as above where $-f$ is the log-likelihood function. Many problems in simulation can also be stated as (SP). For instance, one might be interested in minimizing the average work-in-process in a queueing network by allocating buffer capacity or servers. Our motivation comes from a special class of (SP) known as stochastic programs with recourse. The well-known two-stage stochastic linear program with recourse was introduced independently by [3, 7], in which

$$f(x,\tilde{\xi}) = cx \; + \; \min_{y \geq 0} \quad \tilde{q}y$$
$$\text{s.t} \quad \tilde{W}y = \tilde{r} - \tilde{T}x,$$

$X = \{x : Ax = b, x \geq 0\}$ and $\tilde{\xi} = (\tilde{q}, \tilde{W}, \tilde{r}, \tilde{T})$ is a random vector on $(\Xi, \mathcal{B}, P)$. This formulation can be extended to multiple stages, integer restrictions can be imposed in any of the stages and nonlinear constraints and objective function terms can be added. Stochastic programs with recourse have been successfully applied to problems from finance, energy, telecommunications, transportation, logistics and supply-chain management (e.g., [29]).

In this paper, we make the following assumptions with respect to (SP):

(A1) $f(\cdot, \tilde{\xi})$ is continuous on X, w.p.1,

(A2) $E \sup_{x \in X} f^2(x, \tilde{\xi}) < \infty$,

(A3) $X \neq \emptyset$ and is compact.

The first assumption is satisfied, for instance, by a two-stage stochastic linear program provided it has relatively complete recourse (i.e., for each feasible first stage decision, it is possible to find a feasible second stage decision for all scenarios). However, it eliminates consideration of two-stage stochastic integer programs when there are integrality constraints in the second stage. The second assumption guarantees existence of second moments and provides a needed uniform integrability condition. In some instances of (SP), X may naturally appear as an unbounded set. However, in most practical problems, a decision-maker would not be averse to specifying possibly large, but finite, simple bounds, $l \leq x \leq u$, making the feasible region bounded and hence compact, if also closed.

As the dimension of the random vector $\tilde{\xi}$ grows, (SP) gets harder and often impossible to solve exactly, unless the cost function f has simple structure, or the number of realizations is small. For continuous random vectors, minimization aside, taking the expectation might be very difficult if the dimension of the integral is high. For discrete random vectors, the problem size can grow exponentially in this dimension. In such cases, an intuitive approach is to resort to sampling and approximate the problem with

$$z_n^* = \min_{x \in X} \frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^i). \tag{SP$_n$}$$

$\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$ may be independent and identically distributed (i.i.d.) as $\tilde{\xi}$ or may be generated according to another sampling scheme. Let x^* denote an optimal solution to (SP) with optimal cost z^*. Similarly, let x_n^* and z_n^* denote an optimal solution and the optimal cost of (SP$_n$). Consistency and other asymptotic properties of estimators x_n^* and z_n^* have been studied extensively in the literature, see e.g., [2, 9, 17, 25].

In this paper, we discuss Monte Carlo sampling-based procedures for assessing solution quality in stochastic programs. Determining whether a solution is of high quality (optimal or near optimal) is a fundamental question in optimization theory and applications. Given a candidate solution $\hat{x}$, we define its quality by its optimality gap, $\mu_{\hat{x}} = Ef(\hat{x}, \tilde{\xi}) - z^*$. There are two problems associated with computing this quantity. First, z^* is not known and a lower bound (since we are dealing with a minimization problem) on z^* needs to be computed. In integer programming and nonlinear programming, for example, lower bounds are also useful for proving solution quality and are typically obtained through relaxed problems, where either the integrality constraints or some other complicating constraints are relaxed. An upper bound on z^* is readily available as the cost of the candidate solution. For stochastic programs, a second difficulty is that for a given $\hat{x} \in X$, it is not always possible to compute $Ef(\hat{x}, \tilde{\xi})$ exactly.

Monte Carlo simulation-based methods allow us to estimate an upper bound on the optimality gap for stochastic programs. In the next section, we briefly review how to construct confidence intervals (CIs) on the optimality gap using a multiple replications procedure [19]. Then, we show how to obtain a valid CI using only a single replication. In Section 4, we provide variants of this procedure that use two replications. In Section 5, we compare the empirical coverage results of the procedures for a newsvendor problem and for two-stage stochastic linear programs with recourse. In Section 6, we give more insight as to when the procedures perform well and discuss preliminary guidelines for selecting a candidate solution that can aid our procedures. Section 7 contains concluding remarks and a summary.

2 Multiple Replications Procedure

Let $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$ be i.i.d. from the distribution of $\tilde{\xi}$. Then, by interchanging minimization and expectation we obtain a statistical lower bound on z^*,

$$Ez_n^* = E\left[\min_{x \in X} \frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^i)\right] \leq \min_{x \in X} E\left[\frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^i)\right] = \min_{x \in X} Ef(x, \tilde{\xi}) = z^*. \quad (1)$$

This result establishes that z_n^* has a negative bias, $Ez_n^* - z^* \leq 0$. It can also be shown that $Ez_n^* \leq Ez_{n+1}^*$ for all n. This monotonicity result tells us that on average we obtain better estimates of the optimal value as the sample size increases.

Given a feasible decision $\hat{x} \in X$ and a sample size n for (SP$_n$), we bound the optimal value of (SP) using the above lower bound result, $Ez_n^* \leq z^* \leq Ef(\hat{x}, \tilde{\xi})$. The right inequality comes from suboptimality of $\hat{x}$. An upper bound on the optimality gap for $\hat{x}$ is then $Ef(\hat{x}, \tilde{\xi}) - Ez_n^*$. We estimate this quantity by

$$G_n(\hat{x}) = \frac{1}{n}\sum_{i=1}^{n} f(\hat{x}, \tilde{\xi}^i) - \min_{x\in X} \frac{1}{n}\sum_{i=1}^{n} f(x, \tilde{\xi}^i). \tag{2}$$

The first term on the right-hand side of (2) is an upper bound estimate and converges to $Ef(\hat{x}, \tilde{\xi})$, w.p.1, by the strong law of large numbers. The second quantity, z_n^*, is a lower bound estimate on z^*. In expectation, it provides a lower bound and under (A1)-(A3) converges to z^*, w.p.1 (see subsequent Proposition 1). When a common stream of random numbers, $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$, is used in calculating both terms in (2), $G_n(\hat{x}) \geq 0$, w.p.1. This approach also facilitates variance reduction.

Because of the minimization in (2), $G_n(\hat{x})$ (or, its scaled version $\sqrt{n}(G_n(\hat{x}) - \mu_{\hat{x}})$) is, in general, not normally distributed even as n grows large. Therefore, in [19] confidence intervals are constructed by employing replications, an approach frequently used in simulation for estimating the mean of a random variable with an unknown or non-normal distribution. We summarize below the multiple replications procedure (MRP) to construct a CI on the optimality gap. Let $t_{n,\alpha}$ be the $1-\alpha$ quantile of the Student's t distribution with n degrees of freedom.

MRP:

Input: Desired value of $0 < \alpha < 1$ (e.g., $\alpha = 0.10$), sample size n, replication size n_g and a candidate solution $\hat{x} \in X$.

Output: $(1-\alpha)$-level confidence interval on $\mu_{\hat{x}}$.

1. For $i = 1, 2, \ldots, n_g$,
   1.1. Sample i.i.d. observations $\tilde{\xi}^{i1}, \tilde{\xi}^{i2}, \ldots, \tilde{\xi}^{in}$ from the distribution of $\tilde{\xi}$,
   1.2. Solve (SP_n^i) using $\tilde{\xi}^{i1}, \tilde{\xi}^{i2}, \ldots, \tilde{\xi}^{in}$ to obtain x_n^{i*},
   1.3. Calculate $G_n^i(\hat{x}) = \frac{1}{n}\sum_{j=1}^{n}\left(f(\hat{x}, \tilde{\xi}^{ij}) - f(x_n^{i*}, \tilde{\xi}^{ij})\right)$.
2. Calculate gap estimate and sample variance by

$$\bar{G}(n_g) = \frac{1}{n_g}\sum_{i=1}^{n_g} G_n^i(\hat{x}) \quad \text{and} \quad s_G^2(n_g) = \frac{1}{n_g - 1}\sum_{i=1}^{n_g}\left(G_n^i(\hat{x}) - \bar{G}(n_g)\right)^2.$$

3. Output one-sided CI on $\mu_{\hat{x}}$,

$$\left[0, \bar{G}(n_g) + \frac{t_{n_g-1,\alpha} s_G(n_g)}{\sqrt{n_g}}\right]. \tag{3}$$

Even though $G_n(\hat{x})$ may not be normal, since $\bar{G}(n_g)$ is a sample mean of i.i.d. random variables, it is possible to use the standard central limit theorem (CLT) to construct an approximate $(1-\alpha)$-level CI for the optimality gap given in (3). Due to the negative bias of z_n^*, $E\bar{G}(n_g) \geq Ef(\hat{x}, \tilde{\xi}) - z^*$. Thus, for sufficiently large n_g, we can infer that

$$P\left(Ef(\hat{x}, \tilde{\xi}) - z^* \leq \bar{G}(n_g) + \frac{t_{n_g-1,\alpha} s_G(n_g)}{\sqrt{n_g}}\right) \approx 1 - \alpha \tag{4}$$

and hence that the CI formed by MRP will cover the optimality gap of $\hat{x}$ with the desired probability.

The lower bound given in (1) was independently introduced by Norkin et. al. [20] and used for global optimization of stochastic programs within a branch-and-bound methodology. Other algorithmic work that uses Monte Carlo simulation-based bounds and multiple replications includes [1, 18]. MRP has been applied to different kinds of problems in the literature including a bond portfolio model [4], a stochastic vehicle routing problem [16] and supply chain network design [24].

There is other related work on assessing solution quality in stochastic programs via Monte Carlo methods, some being in the context of specific algorithms. Higle and Sen [11] derive a bound on the optimality gap for two-stage stochastic linear programs that is motivated by the Karush-Kuhn-Tucker optimality conditions; see also, Shapiro and Homem-de-Mello [26]. Higle and Sen [12] have also proposed a statistical lower bound that is rooted in duality. Dantzig and Infanger [8] and Higle and Sen [14, 10] use Monte Carlo versions of lower bounds obtained in sampling-based adaptations of deterministic cutting-plane algorithms.

3 Single Replication Procedure

When applying the multiple replications procedure reviewed above, the replication size is typically taken to be $n_g \geq 30$ in an attempt to have a valid statistical inference. This constitutes a drawback as one needs to solve at least 30 optimization problems (in step 1.2) in order to determine whether a candidate solution is of high quality. In this section, we show how a single replication, $n_g = 1$, can be used to make a valid statistical inference on the quality of a candidate solution.

As before, we assume that the candidate solution $\hat{x} \in X$ is given, and we use the following additional notation. For a feasible solution, $x \in X$, let $\bar{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} f(x, \tilde{\xi}^i)$, $\sigma^2_{\hat{x}}(x) = \text{var}[f(\hat{x}, \tilde{\xi}) - f(x, \tilde{\xi})]$ and $s^2_n(x) = \frac{1}{n-1}\sum_{i=1}^{n}[(f(\hat{x}, \tilde{\xi}^i) - f(x, \tilde{\xi}^i)) - (\bar{f}_n(\hat{x}) - \bar{f}_n(x))]^2$. Note that $G_n(\hat{x})$ given in equation (2) can be written as $\bar{f}_n(\hat{x}) - z^*_n$, with the understanding that the same n observations $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$ are used in $\bar{f}_n(\hat{x})$ and z^*_n. We define z_α to satisfy $P(N(0,1) \leq z_\alpha) = 1 - \alpha$. Below we state the single replication procedure (SRP).

SRP:

Input: Desired value of $0 < \alpha < 1$, sample size n and a candidate solution $\hat{x} \in X$.

Output: $(1-\alpha)$-level confidence interval on $\mu_{\hat{x}}$.

1. Sample i.i.d. observations $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$ from the distribution of $\tilde{\xi}$.
2. Solve (SP_n) to obtain x^*_n.
3. Calculate $G_n(\hat{x})$ as given in (2) and

$$s^2_n(x^*_n) = \frac{1}{n-1}\sum_{i=1}^{n}\left[(f(\hat{x}, \tilde{\xi}^i) - f(x^*_n, \tilde{\xi}^i)) - (\bar{f}_n(\hat{x}) - \bar{f}_n(x^*_n))\right]^2.$$

4. Output one-sided CI on $\mu_{\hat{x}}$,

$$\left[0,\ G_n(\hat{x}) + \frac{z_\alpha s_n(x_n^*)}{\sqrt{n}}\right]. \tag{5}$$

The SRP differs from the MRP in that it uses a single replication and hence the sample variance is calculated differently. In the MRP, n_g i.i.d. observations of $G_n(\hat{x})$ are calculated and the sample variance of these gap estimates is used to form the CI. In contrast, only one value of $G_n(\hat{x})$ is calculated in SRP and the individual observations, $f(\hat{x}, \tilde{\xi}^i) - f(x_n^*, \tilde{\xi}^i)$ for $i = 1, \ldots, n$, are used to calculate the sample variance. In fact, $G_n(\hat{x})$ is the sample mean of these individual observations and $s_n^2(x_n^*)$ is the corresponding sample variance. Below, we show how solving a single replication yields enough information to make a valid statistical inference concerning the quality of a candidate solution *even though* $G_n(\hat{x})$ may not be asymptotically normal. Before stating the theorem, we give a proposition that establishes consistency of the estimators. Let X^* denote the set of optimal solutions to (SP) and let $x^*_{\min} \in \arg\min_{x \in X^*} \operatorname{var}[f(\hat{x}, \tilde{\xi}) - f(x, \tilde{\xi})]$ and $x^*_{\max} \in \arg\max_{x \in X^*} \operatorname{var}[f(\hat{x}, \tilde{\xi}) - f(x, \tilde{\xi})]$. In other words, $x^*_{\min}$ is an optimal solution with minimum variance of $f(\hat{x}, \tilde{\xi}) - f(x, \tilde{\xi})$, $\sigma^2_{\hat{x}}(x^*_{\min})$, among all the optimal solutions and likewise $x^*_{\max}$ is an optimal solution with maximum variance.

Proposition 1 *Assume (A1)-(A3), $\hat{x} \in X$, and that $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$ are i.i.d. as $\tilde{\xi}$. Then,*

(i) $z_n^* \to z^*$, w.p.1,

(ii) all limit points of $\{x_n^*\}$ lie in X^*, w.p.1,

(iii) $\sigma^2_{\hat{x}}(x^*_{\min}) \le \liminf_{n\to\infty} s_n^2(x_n^*) \le \limsup_{n\to\infty} s_n^2(x_n^*) \le \sigma^2_{\hat{x}}(x^*_{\max})$, w.p.1.

Proof. (A2) implies that $E \sup_{x \in X} f(x, \tilde{\xi}) < \infty$. Therefore, (i) follows immediately from Theorem A1 of [21, p.69]. (A1)-(A3) implies $\bar{f}_n(x)$ converges uniformly to $Ef(x, \tilde{\xi})$, w.p.1 on X. This coupled with (i) implies (ii). To prove (iii), we first show that the sequence of continuous functions $s_n^2(x)$ converges to $\sigma^2_{\hat{x}}(x)$ uniformly, w.p.1 on X. Let $g(x, \tilde{\xi}) = f(\hat{x}, \tilde{\xi}) - f(x, \tilde{\xi})$. Then, with $\bar{g}_n(x) = \frac{1}{n}\sum_{i=1}^n g(x, \tilde{\xi}^i)$ we have

$$s_n^2(x) = \frac{n}{n-1}\left\{\frac{1}{n}\sum_{i=1}^n \left(g(x, \tilde{\xi}^i) - Eg(x, \tilde{\xi})\right)^2 - \left(\bar{g}_n(x) - Eg(x, \tilde{\xi})\right)^2\right\}.$$

The first term in the curly brackets is a sample mean of i.i.d. random variables and by Lemma A1 of [21, p.67] converges uniformly, w.p.1, to $\sigma^2_{\hat{x}}(x) = \operatorname{var} g(x, \tilde{\xi})$. Also, by the same lemma, $\bar{g}_n(x)$ converges uniformly to $Eg(x, \tilde{\xi})$, w.p.1, i.e., $\sup_{x \in X} \left|\bar{g}_n(x) - Eg(x, \tilde{\xi})\right| \to 0$, w.p.1. This implies

$$\sup_{x \in X} \left(\bar{g}_n(x) - Eg(x, \tilde{\xi})\right)^2 = \left(\sup_{x \in X} \left|\bar{g}_n(x) - Eg(x, \tilde{\xi})\right|\right)^2 \to 0, \text{ w.p.1.}$$

The sum of these two terms, $a_n(x) = \frac{1}{n}\sum_{i=1}^n (g(x,\tilde{\xi}^i) - Eg(x,\tilde{\xi}))^2 - (\bar{g}_n(x) - Eg(x,\tilde{\xi}))^2$, then converges uniformly to $\sigma^2_{\hat{x}}(x)$, w.p.1. To show uniform convergence of $\frac{n}{n-1}a_n(x)$, consider the following inequality

$$\sup_{x\in X}\left|a_n(x) + \frac{a_n(x)}{n-1} - \sigma^2_{\hat{x}}(x)\right| \le \sup_{x\in X}\left|a_n(x) - \sigma^2_{\hat{x}}(x)\right| \quad + \\ \sup_{x\in X}\left|\frac{a_n(x) - \sigma^2_{\hat{x}}(x)}{n-1}\right| \quad + \quad \sup_{x\in X}\left|\frac{\sigma^2_{\hat{x}}(x)}{n-1}\right|.$$

By the above argument the first two terms on the right-hand side converge to 0, w.p.1. By (A2), $\sup_{x\in X}\sigma^2_{\hat{x}}(x) < \infty$. Thus, the last term also converges to 0, establishing uniform convergence.

Since X is compact, there exists a subsequence N along which $\{x^*_n\}_{n\in N}$ converges to a point in X, and by (ii) this point is in X^*, w.p.1. So, using the uniform convergence shown above,

$$\inf_{x\in X^*}\sigma^2_{\hat{x}}(x) \le \lim_{\substack{n\to\infty \\ n\in N}} s^2_n(x^*_n) \le \sup_{x\in X^*}\sigma^2_{\hat{x}}(x), \text{ w.p.1.}$$

The subsequence N is arbitrary and hence we obtain (iii). ■

When (SP) has multiple optimum solutions, we cannot expect $\{x^*_n\}$ to have a unique limit point. However, by part (ii) of Proposition 1, all its limit points belong the set of optimum solutions, X^*. Similarly, $\{s^2_n(x^*_n)\}$ may not have a unique limit. That is why "lim inf" and "lim sup" appear in part (iii) of Proposition 1 instead of a "lim." Note that by (A2), $\sigma^2_{\hat{x}}(x^*_{\max}) < \infty$. When X^* is a singleton, $x^*_n \to x^*$, w.p.1 and $\liminf_{n\to\infty} s^2_n(x^*_n) = \limsup_{n\to\infty} s^2_n(x^*_n) = \sigma^2_{\hat{x}}(x^*)$, w.p.1. We next present the main result regarding the validity of the SRP.

Theorem 2 *Assume (A1)-(A3), $\hat{x}\in X$, and that $\tilde{\xi}^1,\tilde{\xi}^2,\ldots,\tilde{\xi}^n$ are i.i.d. as $\tilde{\xi}$. Given $0<\alpha<1$, for the SRP,*

$$\liminf_{n\to\infty} P\left(\mu_{\hat{x}} \le G_n(\hat{x}) + \frac{z_\alpha s_n(x^*_n)}{\sqrt{n}}\right) \ge 1-\alpha. \tag{6}$$

Proof. When $\hat{x}\in X^*$, inequality (6) is trivial. Suppose $\hat{x}\notin X^*$, and recall that $z^*_n = \min_{x\in X}\bar{f}_n(x)$. Thus,

$$G_n(\hat{x}) = \bar{f}_n(\hat{x}) - z^*_n \ge \bar{f}_n(\hat{x}) - \bar{f}_n(x), \quad \forall x\in X.$$

Replacing x by $x^*_{\min}\in\arg\min_{x\in X^*}\sigma^2_{\hat{x}}(x)$ we obtain,

$$\begin{aligned} & P\left(G_n(\hat{x}) + \frac{z_\alpha s_n(x^*_n)}{\sqrt{n}} \ge \mu_{\hat{x}}\right) \\ \ge\ & P\left(\bar{f}_n(\hat{x}) - \bar{f}_n(x^*_{\min}) + \frac{z_\alpha s_n(x^*_n)}{\sqrt{n}} \ge \mu_{\hat{x}}\right) & (7) \\ =\ & P\left(\frac{(\bar{f}_n(\hat{x}) - \bar{f}_n(x^*_{\min})) - \mu_{\hat{x}}}{\sigma_{\hat{x}}(x^*_{\min})/\sqrt{n}} \ge -z_\alpha\frac{s_n(x^*_n)}{\sigma_{\hat{x}}(x^*_{\min})}\right), & (8) \end{aligned}$$

where in (8) we assume $\sigma^2_{\hat{x}}(x^*_{\min}) > 0$. Note that if $\sigma^2_{\hat{x}}(x^*_{\min}) = 0$ then $\mathrm{var}\left[\bar{f}_n(\hat{x}) - \bar{f}_n(x^*_{\min})\right] = \frac{1}{n}\sigma^2_{\hat{x}}(x^*_{\min}) = 0$ and it follows from (7) that (6) is again trivial. Let $D_n = \frac{(\bar{f}_n(\hat{x}) - \bar{f}_n(x^*_{\min})) - \mu_{\hat{x}}}{\sigma_{\hat{x}}(x^*_{\min})/\sqrt{n}}$, $a_n = \frac{s_n(x^*_n)}{\sigma_{\hat{x}}(x^*_{\min})}$ and $0 < \varepsilon < 1$, and for the moment assume $\alpha \le 1/2$ so that $z_\alpha \ge 0$. Then (8) can be rewritten as

$$\begin{aligned} & P\left(D_n \ge -z_\alpha a_n\right) \\ \ge\; & P\left(D_n \ge -(1-\varepsilon)z_\alpha,\ a_n \ge 1-\varepsilon\right) \\ =\; & P\left(D_n \ge -(1-\varepsilon)z_\alpha\right) + P\left(a_n \ge 1-\varepsilon\right) - P\left(\{D_n \ge -(1-\varepsilon)z_\alpha\} \cup \{a_n \ge 1-\varepsilon\}\right). \end{aligned} \tag{9}$$

Taking limits we obtain,

$$\liminf_{n\to\infty} P\left(\mu_{\hat{x}} \le G_n(\hat{x}) + \frac{z_\alpha s_n(x^*_n)}{\sqrt{n}}\right) \ge \Phi((1-\varepsilon)z_\alpha),$$

where Φ denotes the distribution function of the standard normal. By Proposition 1, the last two terms in (9) both converge to 1 and cancel out. Since $\bar{f}_n(\hat{x}) - \bar{f}_n(x^*_{\min})$ is a sample mean of i.i.d. random variables, by the CLT the first term in (9) converges to $\Phi((1-\varepsilon)z_\alpha)$. Letting ε shrink to zero gives the desired result, provided $\alpha \le 1/2$. When $\alpha > 1/2$ we replace $x^*_{\min}$ with $x^*_{\max} \in \arg\max_{x \in X^*} \sigma^2_{\hat{x}}(x)$ in (8) and then use a straightforward variation of the above argument. ■

Theorem 2 justifies construction of the approximate $(1-\alpha)$-level one-sided confidence interval for $\mu_{\hat{x}} = Ef(\hat{x}, \tilde{\xi}) - z^*$, given in (5) without requiring $G_n(\hat{x}) = \bar{f}_n(\hat{x}) - z^*_n$ to be asymptotically normal. The intuitive reason for this is that minimization of the sample mean in z^*_n, while making asymptotic analysis of this random variable more difficult, projects the normal distribution so that the resulting confidence interval is conservative. Because we estimate the sample variance in the SRP we recommend the more conservative Student t-quantiles, $t_{n-1,\alpha}$, when n is small.

We reviewed a procedure in which we use $n_g \ge 30$ replications and then introduced a procedure with just one replication, $n_g = 1$. Even though the single replication procedure is computationally significantly less demanding, solving a single minimization problem might also create some problems. For instance, in step 2 of the procedure, if the minimization problem used to calculate the gap estimate yields a solution x^*_n that is equal to $\hat{x}$, then both the gap estimate $G_n(\hat{x})$ and the variance estimate $s^2_n(x^*_n)$ are zero and consequently the CI on the optimality gap given in (5) has width zero. For small sample sizes, this can happen even though the candidate solution $\hat{x}$ is far from optimal. (Proposition 1 eliminates this possibility as the sample size grows large.) The following example illustrates this effect.

Example 1 *Consider the following problem,* $\{\min E[\tilde{\xi}x] : -1 \le x \le 1\}$, *where* $\tilde{\xi} \sim N(\mu, 1)$ *and* $\mu > 0$. *Note that (A1)-(A3) are satisfied. The optimal solution to this problem is* $x^* = -1$

and the candidate solution $\hat{x} = 1$ has the largest optimality gap of $\mu_{\hat{x}} = 2\mu$. Suppose we use the SRP with $\alpha = 0.10$ and $n = 50$ for the candidate solution $\hat{x} = 1$. When the random sample has $\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}^i < 0$, then $x_n^* = 1$ and $G_n(\hat{x}) = s_n(x_n^*) = 0$. Hence, for the problem instance with $\mu = 0.1$, the coverage probability $P\left(\mu_{\hat{x}} \leq G_n(\hat{x}) + z_\alpha s_n(x_n^*)/\sqrt{n}\right) \leq 1 - P(\bar{\xi} < 0) \simeq 0.760$ is below the desired level of 0.90 when a sample size of $n = 50$ is used.

This effect can be lessened by using a larger sample size or by performing more than one replication. The ideas used to show the validity of the single replication procedure can also be used to justify use of procedures with a small number of replications. In the next section, we focus on procedures with two replications.

4 Two-Replication Procedures

In this section we develop two procedures to assess solution quality in stochastic programs that use two replications. The first one, which we call the independent 2-replication procedure (I2RP), aims to eliminate the correlation between $G_n(\hat{x})$ and $s_n(x_n^*)$, by performing two independent replications, one to estimate the gap and the other to estimate $s_n(x_n^*)$.

I2RP:

Recall the definition of the SRP and replace step 3 by:

3′. Calculate $G_n^1(\hat{x})$ as given in (2) and to calculate the sample variance

3′.1. Sample i.i.d. observations $\tilde{\xi}^{n+1}, \tilde{\xi}^{n+2}, ..., \tilde{\xi}^{2n}$ from the distribution of $\tilde{\xi}$,

3′.2. Solve (SP$_n$) defined with respect to $\tilde{\xi}^{n+1}, \tilde{\xi}^{n+2}, ..., \tilde{\xi}^{2n}$ to obtain x_n^{2*},

3′.3. Calculate $s_n^2(x_n^{2*}) = \frac{1}{n-1}\sum_{i=1}^{n}[(f(\hat{x}, \tilde{\xi}^{n+i}) - f(x_n^{2*}, \tilde{\xi}^{n+i})) - (\bar{f}_n(\hat{x}) - \bar{f}_n(x_n^{2*}))]^2$, where the sample means in this sample variance computation are also with respect to the second sample.

The confidence interval on the optimality gap is formed exactly as in (5), where the gap point estimate, $G_n^1(\hat{x})$, comes from the first replication and the sample standard deviation, $s_n(x_n^{2*})$, comes from the second replication. Even though I2RP requires twice the computational effort compared to a single replication procedure, the correlation between these two estimates becomes zero. Following the ideas in the proof of Theorem 2, it can easily be shown that this procedure provides an asymptotically valid confidence interval. We formally state this in theorem below.

Theorem 3 *Assume (A1)-(A3), $\hat{x} \in X$, and that $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^{2n}$ are i.i.d. as $\tilde{\xi}$. Given $0 < \alpha < 1$, for the I2RP,*

$$\liminf_{n\to\infty} P\left(\mu_{\hat{x}} \leq G_n^1(\hat{x}) + \frac{z_\alpha s_n(x_n^{2*})}{\sqrt{n}}\right) \geq 1 - \alpha.$$

Proof. The proof of Theorem 2 remains the same when $s_n^2(x_n^*)$ is redefined as in step $3'$. ■

A natural extension of the I2RP is to use all the information available from the two replications. In other words, we have a single sample of size $2n$ and partition it (randomly) into two sets of size n. In each set we perform the SRP and average the two estimates. We call this the averaged two-replication procedure (A2RP).

A2RP:

Recall the definition of the MRP and fix $n_g = 2$. Replace steps 1.3, 2 and 3 by:

$1.3'$. Calculate $G_n^i(\hat{x})$ and $s_n^2(x_n^{i*})$.

$2'$. Calculate the estimates by taking the average,

$$G_n'(\hat{x}) = \frac{1}{2}\left(G_n^1(\hat{x}) + G_n^2(\hat{x})\right) \quad \text{and} \quad s_n^{2\,\prime} = \frac{1}{2}\left(s_n^2(x_n^{1*}) + s_n^2(x_n^{2*})\right). \tag{10}$$

$3'$. Output one-sided CI on $\mu_{\hat{x}}$,

$$\left[0,\ G_n'(\hat{x}) + \frac{z_\alpha s_n'}{\sqrt{2n}}\right].$$

Unlike the MRP, the sample variance, $s_n^2(x_n^{i*})$, for each sample $i = 1, 2$, is calculated as in the single replication procedure (in step $1.3'$) and these are averaged to obtain the variance estimator of the A2RP (in step $2'$). This variance estimator given in (10) is a pooled estimator, similar in spirit to that used in a two-sample t-test for testing the difference of means from populations with equal variance [6, p.396]. It is a consistent estimator, in the sense that $\liminf_{n\to\infty} s_n^{2\,\prime} \geq \sigma_{\hat{x}}^2(x_{\min}^*)$, w.p.1, by Proposition 1. A2RP provides an asymptotically valid CI on the optimality gap, as stated in the theorem below.

Theorem 4 *Assume (A1)-(A3), $\hat{x} \in X$, and that $\tilde{\xi}^{i1}, \tilde{\xi}^{i2}, \ldots, \tilde{\xi}^{in}$, $i = 1, 2$, are i.i.d. as $\tilde{\xi}$. Given $0 < \alpha < 1$, for the A2RP,*

$$\liminf_{n\to\infty} P\left(\mu_{\hat{x}} \leq G_n'(\hat{x}) + \frac{z_\alpha s_n'}{\sqrt{2n}}\right) \geq 1 - \alpha.$$

Proof. With an obvious extension of notation to index each sample, we have

$$\bar{f}_n^1(\hat{x}) - z_n^{1*} \geq \bar{f}_n^1(\hat{x}) - \bar{f}_n^1(x_{\min}^*) \qquad \text{and} \qquad \bar{f}_n^2(\hat{x}) - z_n^{2*} \geq \bar{f}_n^2(\hat{x}) - \bar{f}_n^2(x_{\min}^*). \tag{11}$$

Multiplying each of the inequalities in (11) by 1/2 and summing, we obtain

$$\begin{aligned} G_n'(\hat{x}) &\geq \frac{1}{2}\left([\bar{f}_n^1(\hat{x}) - \bar{f}_n^1(x_{\min}^*)] + [\bar{f}_n^2(\hat{x}) - \bar{f}_n^2(x_{\min}^*)]\right) \\ &= \bar{f}_{2n}(\hat{x}) - \bar{f}_{2n}(x_{\min}^*). \end{aligned}$$

Since $\bar{f}_{2n}(\hat{x}) - \bar{f}_{2n}(x_{\min}^*)$ is a sample mean of i.i.d. random variables, by the CLT, $\sqrt{2n}((\bar{f}_{2n}(\hat{x}) - \bar{f}_{2n}(x_{\min}^*)) - \mu_{\hat{x}})$ converges in distribution to a normal random variable with mean zero and variance

$\sigma^2_{\hat{x}}(x^*_{\min})$. Also, $\liminf_{n\to\infty} \frac{s'_n}{\sigma_{\hat{x}}(x^*_{\min})} \geq 1$, w.p.1, by Proposition 1. The rest of the proof for $\alpha \leq 1/2$ case is analogous to that of Theorem 2, and the proof for $\alpha > 1/2$ is again straightforward. ■

Note that the independent two-replication procedure uses $\sqrt{n}$ as the scaling factor whereas the averaged two-replication procedure uses $\sqrt{2n}$. Even though the two procedures use the same number of observations, the A2RP uses all of the information to form both estimators whereas I2RP uses half of the information for each estimator. However, I2RP eliminates the correlation between the gap and variance estimators. Now let us turn back to Example 1 to illustrate the two-replication procedures.

Example 2 *Consider the specific problem instance given in Example 1. Let* $\bar{\xi}_1 = \frac{1}{n}\sum_{i=1}^n \tilde{\xi}^i$ *be the sample mean of the first sample and likewise,* $\bar{\xi}_2$ *be the sample mean of the second sample. With* $\mu = 0.1$ *and* $n = 50$*, the probability of obtaining a CI of width 0 from I2RP or A2RP is* $P(\bar{\xi}_1 < 0)P(\bar{\xi}_2 < 0) = 0.057$*, from normal quantiles. Therefore, for the two-replication procedures that use a sample size of* $n = 50$ *for each replication, the coverage probabilities are bounded above by* 0.943*, compared to* 0.760 *for SRP in Example 1. For SRP that uses a sample size of* $2n = 100$*, the upper bound for the coverage probability is* 0.841*.*

5 Comparison of Empirical Coverage Results

In this section, we empirically analyze the small-sample behavior of the described procedures. We apply them to a newsvendor problem under uniform demand and to three small two-stage stochastic linear programs from the literature and compare empirical coverage probabilities.

5.1 Newsvendor Problem

The newsvendor problem is a classical example of a stochastic program with simple recourse and its properties are well known, e.g., [5, p.15]. We briefly review its formulation. Let r be the selling price of a newspaper, $0 < c < r$ be its cost to the vendor, and $\tilde{\xi}$ denote the nonnegative random demand. The vendor's problem is to find the number of papers to buy, x, so that the expected profit is maximized. So, the problem is formulated as $\max \left\{-cx + rE\min\{x,\ \tilde{\xi}\} : x \geq 0\right\}$ and its solution is given by x^* that solves $\inf_{x\geq 0} P(\tilde{\xi} \leq x) \geq (r-c)/r$, which is simply $\int_0^{x^*} dF(\xi) = (r-c)/r$, when the demand distribution is continuous with distribution function F. Note that the newsvendor problem is of the form (SP) with $f(x, \tilde{\xi}) = cx - r\min\{x,\ \tilde{\xi}\}$ and $X = \{x : x \geq 0\}$.

We assume $\tilde{\xi} \sim U(0, b)$, $b > 0$ and hence modify X to $\{x : 0 \leq x \leq b\}$. Note that (A1)-(A3) hold. To perform the tests, we set $\alpha = 0.10$. For the problem parameters, we use $c = 5$, $r = 15$ and $b = 10$. This problem has optimal solution $x^* = 6\frac{2}{3}$ with expected profit $z^* = 33\frac{1}{3}$. For the candidate solution $\hat{x}$, we pick a solution that has expected profit 10% from the optimum. We use

n	MRP	SRP	I2RP	A2RP	TRUE
50	0.9873 ± 0.0018	0.8756 ± 0.0017	0.9421 ± 0.0012	0.9273 ± 0.0012	0.9530 ± 0.0011
100	0.9741 ± 0.0026	0.8895 ± 0.0016	0.9299 ± 0.0013	0.9106 ± 0.0013	0.9360 ± 0.0013
200	0.9594 ± 0.0032	0.8898 ± 0.0016	0.9290 ± 0.0013	0.9124 ± 0.0013	0.9249 ± 0.0014
300	0.9483 ± 0.0036	0.8946 ± 0.0016	0.9257 ± 0.0014	0.9106 ± 0.0014	0.9188 ± 0.0014
400	0.9390 ± 0.0039	0.8944 ± 0.0016	0.9180 ± 0.0014	0.9061 ± 0.0014	0.9165 ± 0.0014
500	0.9359 ± 0.0040	0.8937 ± 0.0016	0.9192 ± 0.0014	0.9066 ± 0.0014	0.9140 ± 0.0015
600	0.9350 ± 0.0041	0.8962 ± 0.0016	0.9187 ± 0.0014	0.9079 ± 0.0014	0.9143 ± 0.0015
700	0.9299 ± 0.0042	0.8960 ± 0.0016	0.9153 ± 0.0014	0.9048 ± 0.0014	0.9124 ± 0.0015
800	0.9287 ± 0.0042	0.8959 ± 0.0016	0.9139 ± 0.0015	0.9058 ± 0.0015	0.9123 ± 0.0015
900	0.9317 ± 0.0041	0.8970 ± 0.0016	0.9146 ± 0.0015	0.9061 ± 0.0014	0.9118 ± 0.0015
1000	0.9267 ± 0.0043	0.8970 ± 0.0016	0.9143 ± 0.0015	0.9048 ± 0.0014	0.9105 ± 0.0015

Table 1: Empirical coverage results, $\hat{p}\pm1.645(\hat{p}(1-\hat{p})/k)^{1/2}$, for various values of n, where k =10,000 for MRP and 100,000 for SRP, I2RP, A2RP and TRUE. Confidence intervals for TRUE are calculated by using $G_n(\hat{x})$ from SRP and replacing $s_n(x_n^*)$ by $\sigma_{\hat{x}}(x^*)$ in (5).

$\hat{x} = 8.775$ with $Ef(\hat{x}, \tilde{\xi}) = 30$ and with an optimality gap of $\mu_{\hat{x}} = 3\frac{1}{3}$. This candidate solution has $\sigma^2_{\hat{x}}(x^*) = 140.79$. For the SRP, I2RP and A2RP we construct 100,000 confidence intervals and for the MRP, we take $n_g = 30$ and construct 10,000 intervals for each value of the sample size. We take sample sizes, n, between 50 and 1,000. For each sample size, n, we use the same observations as the SRP to form CIs for I2RP and A2RP. In other words, we compare SRP with sample size n with two-replication procedures that use the same n observations and a random partition of these observations into two samples of size $n/2$. To see how estimating $s_n(x_n^*)$ affects coverage, we form another CI by taking $G_n(\hat{x})$ from SRP and replacing $s_n(x_n^*)$ by $\sigma_{\hat{x}}(x^*)$ in (5). We denote this procedure as TRUE.

Table 1 summarizes the results. For each procedure, we report "coverage," i.e., the proportion, $\hat{p}$, of CIs containing the optimality gap and the half width, $1.645(\hat{p}(1-\hat{p})/k)^{1/2}$, of a 90% CI for the true coverage probability, where $k = 100,000$ for SRP and 10,000 for MRP. For example, when n =1,000 for the MRP the table indicates $\hat{p} = 0.9267$ so that we are confident at level 0.90 that the coverage probability, i.e., the left-hand side of (4), is in $[0.9224, 0.9310]$.

Figure 1 shows a plot of $\hat{p}$ versus n for each of the procedures. The coverage for the MRP exceeds the desired coverage of 90% but shrinks toward 90% as the sample size increases. The bias, $Ez_n^* - z^*$, constitutes a major part of the CI formed by MRP and thus this CI tends to overestimate the optimality gap. As indicated in Section 2, the bias shrinks as n increases and the coverage of MRP falls as n grows. The SRP, on the other hand, has slightly less than the desired coverage of 90%. Even though the bias is larger when the sample size is small, the number of times a single replication CI contains the optimality gap approaches 90% from below. With a more careful examination, we see a similar effect as illustrated in Example 1. For small sample sizes, $G_n(\hat{x})$ is

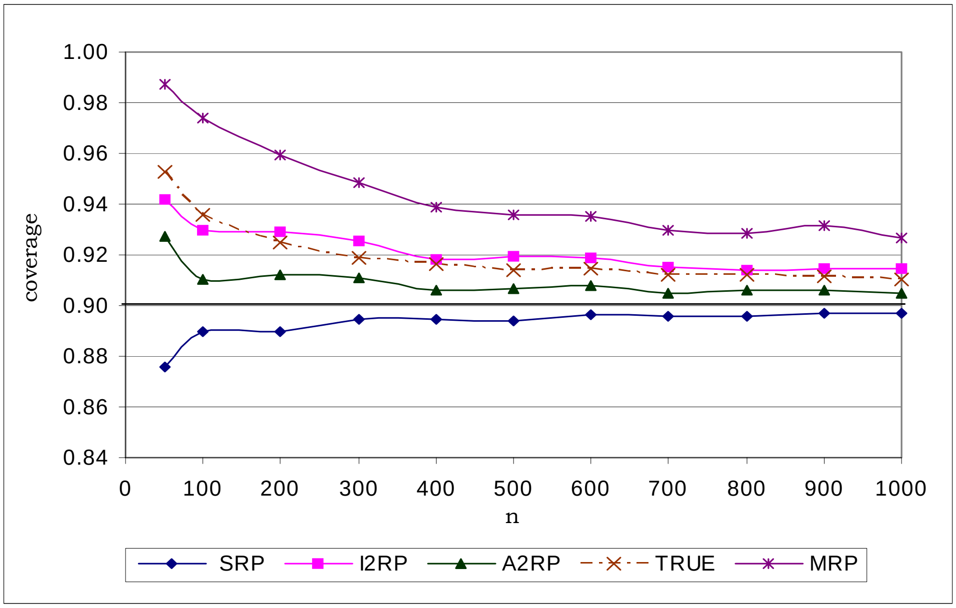

Figure 1: Empirical coverage probability ($\hat{p}$) versus sample size (n) for the newsvendor problem.

more variable and we have observed from the individual replications that when it is small, $s_n(x_n^*)$ also tends to be small, resulting in a narrow CI width. In particular, this happens when x_n^* is close to $\hat{x}$, even though $\hat{x}$ is not close to x^*. The two-replication procedures lessen this effect by using two samples and two estimates $x_{n/2}^*$. For this instance of the newsvendor problem, their coverage probabilities approach 90% from above.

5.2 Two-Stage Stochastic Linear Programs

In this section, we apply the procedures to three two-stage stochastic linear programs with recourse from the literature. The first one, denoted CEP1, is a capacity expansion planning problem with random demand. The dimension of the random vector $\tilde{\xi}$ for CEP1 is 3 and it has 216 total realizations. The second test problem, PGP2, is an electric power generation model, again with 3 stochastic parameters but with 576 realizations. Both CEP1 and PGP2 are described in [13, pp. 3-10]. The third test problem we use, denoted APL1P, can be found in [15]. It is a power expansion planning problem where $\tilde{\xi}$ has 5 independent elements and 1280 realizations. Since these test problems have small numbers of realizations, it is possible to calculate true optimality gaps and variances. Table 2 lists the candidate solutions we use for each problem.

To solve the sampling problems, we used the regularized decomposition algorithm of [22]. An accelerated implementation of this algorithm is in C++ [23] and we have modified this code to perform the tests. Again, we set $\alpha = 0.10$ and for each test problem under SRP, I2RP, A2RP and

TRUE, we construct 500 confidence intervals for various values of the sample size n. For MRP, we use $n_g = 30$ and construct 100 confidence intervals for different values of n. Tables 3, 4 and 5 list results for CEP1, PGP2 and APL1P, respectively. As before, we report "coverage," i.e., the proportion, $\hat{p}$, of CIs containing the optimality gap and the half width, $1.645(\hat{p}(1-\hat{p})/k)^{1/2}$, of a 90% CI for the true coverage probability, where $k = 100$ for MRP and $k = 500$ for the other four procedures. As PGP2 and APL1P have high variance relative to the optimality gap, the sampling error term in the CI for TRUE, i.e., $z_\alpha \sigma_{\hat{x}}(x^*)/\sqrt{n}$, dominates and results in mostly 100% coverage. (For the values of n we consider t versus normal quantiles make little difference.) Similarly, the MRP, while computationally more expensive than the single- and two-replication procedures is largely conservative with respect to its coverage results.

For CEP1 the optimal solution, x^*, is quite easy to find by a sampling problem. That is, the probability that x_n^* equals x^* is quite high even for small sample sizes. Therefore, CEP1 seems to have fairly good coverage for each of the procedures. In contrast, both PGP2 and APL1P yield different solutions, x_n^*, to sampling problems for values of n we consider. In fact, for PGP2 we have observed that the optimal solution x^* and the candidate solution given in Table 2 each appear as x_n^* almost 45% of the time when $n = 500$. Thus, due to the same effect illustrated in Example 1, the coverage results for this candidate solution are very low. Two-replication procedures have higher coverage compared to SRP but are still below the desired level of 90%. For APL1P, we have observed that the probability of obtaining x^* as x_n^* is even lower than PGP2. However, x_n^* takes a variety of different values for APL1P's sampling problems, compared to predominantly two distinct values for PGP2. Thus, the resulting coverage results are good for larger sample sizes for SRP and the two-replication procedures perform well even for small sample sizes.

6 Further Analysis and Preliminary Guidelines

It is well known in linear, integer and nonlinear programming that alternative equivalent formulations of a specific problem can differ dramatically in the ease with which certain algorithms can solve the problem. As a result, skilled modelers and analysts take such properties into account when formulating an optimization problem. With similar motivation, in this section we provide more insight and discuss preliminary guidelines as to the types of candidate solutions that are more amenable to our quality assessment procedures. As illustrated in Example 1 and the computational results of the previous section, it is harder to assess the quality of certain candidate solutions, $\hat{x}$, for some problems. This is particularly true when $\hat{x} \notin X^*$ is chosen to be a solution to an auxiliary sampling problem and this solution has a high probability of occurrence in the (SP_n) from the SRP. In such cases the coverage probability of the SRP can be low. That is, SRP can report a CI width which is too narrow, over-stating the quality of the candidate solution. Two-replication procedures are ways to reduce this effect, however they too may not be enough.

Example 3 *For the problem discussed in Examples 1 and 2, as $\mu \to 0$, the upper bound on the coverage probability, i.e., $1 - P($obtaining a CI of width 0$)$, for SRP approaches* 0.50 *and the same upper bound for I2RP and A2RP approaches* 0.75 *for all sample sizes. Note that for a fixed μ, we obtain $(1-\alpha)$-level coverage as $n \to \infty$. However, for a fixed n we obtain 0.50-level coverage for the SRP and 0.75-level coverage for the two-replication procedures as $\mu \to 0$.*

One alternative is to employ the more conservative multiple replications procedure, MRP. Another option is to average more than two replications, again at the expense of solving more optimization problems. For instance A3RP, the three-replication variant of A2RP, will increase the upper bound on the coverage probability from 0.75 to 0.875 as $\mu \to 0$ in Example 3. These methods assume the candidate solution is fixed and that we are trying to assess this particular solution's quality. Another option is to employ a single or two-replication procedure but improve the procedures' performance by modifying the way in which the candidate solution is selected. To this end, we examine PGP2 in more detail and restrict attention to the SRP, which appears to be our most "dangerous" procedure with respect to the possibility of undercoverage.

Table 6 lists the most frequent x_n^*'s to 10,000 sampling problems of size $n = 500$ for PGP2. We also report empirical coverage probabilities when taking each of these as the candidate solution under 500 repetitions of the SRP, again for a sample size of $n = 500$. The optimal solution, x^*, and the candidate solution used in the previous section, x_1, each appear approximately 45% of the time. Points x_1 and x_2 are quite close to each other (in terms discussed in more detail below) and they both result in very low coverage of the SRP. When $\hat{x} = x_1$ or x_2 and either of these points happens to solve the sampling problem in SRP, the resulting CI width is zero or nearly zero, lowering the coverage probability.

We say two points $x', x'' \in X$ coincide if $\text{var}[f(x', \tilde{\xi}) - f(x'', \tilde{\xi})] = 0$ and that they nearly coincide if this variance is small. This occurs if $x' = x''$ but it can also occur when x' and x'' are distinct. When a candidate solution $\hat{x}$ nearly coincides with a high probability $x_n^* \notin X^*$, the gap random variable is nearly degenerate and we can have undercoverage. So, even though x_2 from Table 6 has a relatively low probability of occurrence, it nearly coincides with the higher probability x_1, leading to low coverage.

Shapiro, Homem-de-Mello and Kim [28] define a condition number for convex, piecewise linear stochastic programs that have a unique, sharp optimum as

$$\kappa = \max_d \frac{\text{var}[f'(x^*, d)]}{E^2 f'(x^*, d)},$$

where $f'(x^*, d)$ denotes the directional derivative of $f(\cdot, \tilde{\xi})$ at x^* in the direction d and the maximization is over all feasible directions at x^*. Roughly speaking, the sample size required to achieve a desired probability that x_n^* equals x^* is proportional to κ. We note that all three stochastic linear programs we used in the previous section satisfy the assumptions in [28].

The condition number of CEP1 is estimated as 18.49 in [28] and our computational results suggest that $P(x_n^* = x^*)$ is very high even for small n, as also observed in [27]. We estimated $\kappa = 2.36 \times 10^5$ for PGP2 and from Table 6, an estimate of $P(x_{500}^* = x^*)$ is approximately 0.45. The other frequent x_n^* for PGP2 is x_1 and it is hard to assess the quality of candidate solutions that coincide with x_1. APL1P, on the other hand, has $\kappa = 7.73 \times 10^7$ [28] and our empirical results show that $P(x_{1000}^* = x^*)$ is around 0.27. The next two most frequent x_{1000}^* for APL1P show up approximately 15.57% and 10.57% of the time. Even though the condition number of APL1P is high, there are many different solutions to the sampling problems. This results in a smoother estimation of the optimality gap and the procedures we have proposed work fairly well in this situation despite the high condition number. For these three problems, a low condition number is associated with a problem whose candidate solutions are relatively easy to assess but a high condition number doesn't necessarily correspond to a problem whose candidate solutions are difficult to assess.

We consider an approach based on epsilon-optimal solutions to help avoid $\hat{x}$ and x_n^* coinciding. Suppose we generate $\hat{x}$ by solving ($\text{SP}_{n_{\hat{x}}}$), i.e., a sample-mean problem with sample size $n_{\hat{x}}$, and then assess its quality via SRP by solving a separate (SP_n). Here, $n_{\hat{x}}$ and n could be the same or differ (typically $n_{\hat{x}} \geq n$) and the same holds for the two "epsilons" used when approximately solving ($\text{SP}_{n_{\hat{x}}}$) and (SP_n). We believe solving ($\text{SP}_{n_{\hat{x}}}$) and (SP_n) approximately makes sense, particularly in light of the fact that our procedure's output is a confidence interval. There are clearly a number of possibilities but in our further computations we use $n_{\hat{x}} = n$ and solve (SP_n) with high precision but we obtain an epsilon-optimal solution to ($\text{SP}_{n_{\hat{x}}}$). When (SP) is a stochastic linear program this can be accomplished via an interior point method, which may also have the advantage of avoiding extreme points and may help avoid $\hat{x}$ coinciding with x_n^* and causing the kind of trouble we have illustrated.

We solved three ($\text{SP}_{n_{\hat{x}}=500}$)'s for PGP2 for specific samples that yield x_0, x_1 and x_2 as solutions using the standard primal-dual logarithmic barrier algorithm in CPLEX 8.1 (with no crossover to the simplex method) over a range of complementarity tolerances. The results are displayed in Figures 2, 3 and 4 for sampling problems ($\text{SP}_{n_{\hat{x}}}$) that yield $\hat{x} = x_0, x_1$ and x_2, respectively. We plot on the left-hand y-axis the empirical coverage probability of the SRP out of 500 repetitions for sample size $n = 500$, for the candidate solutions we obtain from solving this sampling problem with different levels of precision. This precision, labeled "suboptimality" on the x-axis represents the relative difference between primal and dual objective function values of the two-stage stochastic linear program upon termination of the barrier method. The right-hand y-axis plots the candidate solution's relative gap, $\mu_{\hat{x}}/z^*$, i.e., suboptimality in the true problem (SP). So, in Figure 2 a solution which is suboptimal by at most 5% in its (SP$_{500}$) is actually only 0.4% suboptimal in (SP). This is, in part, because the interior point method's suboptimality includes contributions from both first- and second-stage variables, and further because in a specific sample $\frac{1}{n}\sum_{i=1}^{n} f(x,\tilde{\xi}^i)$ may be more sensitive to small changes in x than $Ef(x,\tilde{\xi})$.

Not surprisingly, the results for the sampling problem with $\hat{x} = x_0 = x^*$ are very good over a wide range of levels of suboptimality. For the sampling problems with $x_n^* = x_1$ and x_2, the coverage probability reaches the desired level of 90% at about a barrier suboptimality tolerance around 3% and a relative gap in (SP) of roughly 0.3%. We note, however, that the method of obtaining a candidate solution via an epsilon-optimal solutions is not foolproof. In Figure 4 corresponding to the (SP$_{500}$) with optimal solution x_2, the barrier suboptimality level of 5% results in a candidate solution with poor coverage. This happens because the epsilon-optimal solution nearly coincides with x_1, the high frequency x_{500}^* with poor coverage (see Table 6). Before and after this tolerance level, the candidate solutions have good coverage. Overall, the preliminary computational results show that the range of tolerance values that give poor coverage is quite narrow and suggest that the use of epsilon-optimal solutions as a safeguarding technique when generating candidate solutions merits further investigation.

7 Conclusions

In this paper, we have developed Monte Carlo sampling-based procedures for assessing solution quality in stochastic programs. Compared to an earlier multiple replications procedure that requires solution of at least 30 optimization problems, the methods we have introduced require solution of one or two optimization problems. We illustrate through an example that even though the single replication procedure is computationally significantly less demanding, and even though its use is theoretically justified for sufficiently large samples, it can have low coverage probability for small sample sizes for some problems. Specifically, an illustrative example and computational results substantiate that when a solution $x_n^* \notin X^*$ to a sampling problem (SP$_n$) is chosen as the candidate solution $\hat{x}$ and this solution has a high probability of occurring as x_n^*, the coverage probability of the SRP can be quite low. So, we develop variants of this procedure that use two replications to lessen this effect. In some cases, the two-replication procedures might not be enough to reach a desired level of coverage. One alternative is to fix the candidate solution and employ more replications, and another is to generate a good candidate solution, in the sense that it is easier to assess its quality. To this end, we proposed using epsilon-optimal solutions. While there are a number of possibilities, we considered epsilon-optimal candidate solutions arising from solving an instance of (SP_n) via a barrier method for two-stage stochastic linear programs. At quite modest values for suboptimization, candidate solutions typically were less likely to coincide with an x_n^* and associated coverage results improved significantly.

8 Acknowledgements

The authors thank Andrzej Ruszczyński and Artur Świetanowski for access to their regularized decomposition code and Vaclav Dupač and Shane Henderson for valuable discussions. This research is partially supported by the National Science Foundation under Grant DMI-0217927.

References

[1] S. Ahmed and A. Shapiro. The sample average approximation method for stochastic programs with integer recourse. *Optimization Online*, 2002. http://www.optimization-online.org/.

[2] H. Attouch and R.J.-B. Wets. Approximation and convergence in nonlinear optimization. In O. Mangasarian, R. Meyer, and S. Robinson, editors, *Nonlinear Programming 4*, pages 367–394. Academic Press, New York, 1981.

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

[4] M. Bertocchi, J. Dupačová, and V. Moriggia. Sensitivity of bond portfolio's behavior with respect to random movements in yield curve: A simulation study. *Annals of Operations Research*, 99:267–286, 2000.

[5] J.R. Birge and F. Louveaux. *Introduction to Stochastic Programming.* Springer-Verlag, New York, 1997.

[6] G. Casella and R. L. Berger. *Statistical Inference.* Duxbury Press, Belmont, California, 1990.

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

[8] G.B. Dantzig and G. Infanger. A probabilistic lower bound for two-stage stochastic programs. Technical Report SOL 95-6, Department of Operations Research, Stanford University, November 1995.

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

[10] J.L. Higle and S. Sen. Statistical verification of optimality conditions for stochastic programs with recourse. *Annals of Operations Research*, 30:215–240, 1991.

[11] J.L. Higle and S. Sen. Stochastic decomposition: an algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research*, 16:650–669, 1991.

[12] J.L. Higle and S. Sen. Duality and statistical tests of optimality for two stage stochastic programs. *Mathematical Programming*, 75:257–275, 1996.

[13] J.L. Higle and S. Sen. *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*. Kluwer Academic Publishers, Dordrecht, 1996.

[14] J.L. Higle and S. Sen. Statistical approximations for stochastic linear programming problems. *Annals of Operations Research*, 85:173–192, 1999.

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

[16] A.S. Kenyon and D.P. Morton. Stochastic vehicle routing with random travel times. *Transportation Science*, 37:69–82, 2003.

[17] A.J. King and R.T. Rockafellar. Asymptotic theory for solutions in statistical estimation and stochastic programming. *Mathematics of Operations Research*, 18:148–162, 1993.

[18] A.J. Kleywegt, A. Shapiro, and T. Homem-de-Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(3):479–502, 2001.

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

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

[21] R.Y. Rubinstein and A. Shapiro. *Discrete Event Systems: Sensitivity and Stochastic Optimization by the Score Function Method*. John Wiley & Sons, Chichester, 1993.

[22] A. Ruszczyński. A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming*, 35:309–333, 1986.

[23] A. Ruszczyński and A. Świetanowski. Accelerating the regularized decomposition method for two stage stochastic linear problems. *European Journal of Operational Research*, 101:328–342, 1997.

[24] T. Santoso, S. Ahmed, M. Goetschalckx, and A. Shapiro. A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Reseach*, 2004. Forthcoming.

[25] A. Shapiro. Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30:169–186, 1991.

[26] A. Shapiro and T. Homem-de-Mello. A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming*, 81:301–325, 1998.

[27] A. Shapiro and T. Homem-de-Mello. On rate of convergence of monte carlo approximations of stochastic programs. *SIAM Journal on Optimization*, 11:70–86, 2001.

[28] A. Shapiro, T. Homem-de-Mello, and J. Kim. Conditioning of covex piecewise linear stochastic programs. *Mathematical Programming*, 94:1–19, 2002.

[29] S.W. Wallace and W.T. Ziemba, editors. *Applications of Stochastic Programming*. MPS-SIAM Series in Optimization, 2004. In progress.

Problem	$\hat{x}$	$Ef(\hat{x},\tilde{\xi})$	$\mu_{\hat{x}}$	$\sigma^2_{\hat{x}}(x^*)$
CEP1	$(0, 125, 875, 2500,$ $0, 625, 1375, 3000)$	$393,288.01$	$38,129.09$	$3,115,876,053.57$
PGP2	$(1.5, 5.5, 5, 4.5)$	448.46	1.14	$172,896.17$
APL1P	$(1111.11, 2300)$	$24,807.16$	164.84	$6,008,299.74$

Table 2: Candidate solutions used in tests.

n	MRP	SRP	I2RP	A2RP	TRUE
50	0.860 ± 0.057	0.912 ± 0.021	0.912 ± 0.021	0.920 ± 0.020	0.928 ± 0.019
100	0.940 ± 0.039	0.888 ± 0.023	0.898 ± 0.022	0.890 ± 0.023	0.912 ± 0.021
150	0.910 ± 0.047	0.912 ± 0.021	0.926 ± 0.019	0.906 ± 0.021	0.918 ± 0.020
200	0.920 ± 0.045	0.894 ± 0.023	0.906 ± 0.021	0.894 ± 0.023	0.906 ± 0.021

Table 3: Empirical coverage results for CEP1.

n	MRP	SRP	I2RP	A2RP	TRUE
50	1 ± 0	0.536 ± 0.037	0.708 ± 0.033	0.876 ± 0.024	1 ± 0
100	1 ± 0	0.572 ± 0.036	0.676 ± 0.034	0.766 ± 0.031	1 ± 0
200	1 ± 0	0.472 ± 0.037	0.792 ± 0.030	0.806 ± 0.029	1 ± 0
300	1 ± 0	0.662 ± 0.035	0.810 ± 0.029	0.906 ± 0.021	1 ± 0
400	1 ± 0	0.578 ± 0.036	0.712 ± 0.033	0.730 ± 0.033	1 ± 0
500	1 ± 0	0.504 ± 0.037	0.854 ± 0.026	0.864 ± 0.025	1 ± 0

Table 4: Empirical coverage results for PGP2.

n	MRP	SRP	I2RP	A2RP	TRUE
50	1 ± 0	0.782 ± 0.030	0.940 ± 0.017	0.932 ± 0.019	1 ± 0
100	1 ± 0	0.786 ± 0.030	0.910 ± 0.021	0.918 ± 0.020	1 ± 0
200	1 ± 0	0.828 ± 0.028	0.908 ± 0.021	0.902 ± 0.022	1 ± 0
300	1 ± 0	0.832 ± 0.028	0.918 ± 0.020	0.880 ± 0.024	1 ± 0
400	1 ± 0	0.850 ± 0.026	0.928 ± 0.019	0.886 ± 0.023	1 ± 0
500	1 ± 0	0.902 ± 0.022	0.940 ± 0.017	0.908 ± 0.021	1 ± 0
600	1 ± 0	0.894 ± 0.023	0.944 ± 0.017	0.910 ± 0.021	0.998 ± 0.003
700	1 ± 0	0.910 ± 0.021	0.964 ± 0.014	0.934 ± 0.018	0.990 ± 0.007
800	1 ± 0	0.910 ± 0.021	0.962 ± 0.014	0.934 ± 0.018	0.992 ± 0.007
900	1 ± 0	0.906 ± 0.021	0.965 ± 0.014	0.934 ± 0.018	0.984 ± 0.009
1000	1 ± 0	0.906 ± 0.021	0.956 ± 0.015	0.926 ± 0.019	0.990 ± 0.007

Table 5: Empirical coverage results for APL1P.

x_i		Frequency	$Ef(x_i, \tilde{\xi})$	μ_{x_i}	Coverage
$x^* = x_0$	(1.5, 5.5, 5, 5.5)	44.49%	447.324	0	1 ± 0
x_1	(1.5, 5.5, 5, 4.5)	43.90%	448.464	1.140	0.504 ± 0.037
x_2	(1.5, 5, 5, 5)	4.44%	448.511	1.186	0.504 ± 0.037
x_3	(1.5, 5.5, 5, 5)	3.54%	447.752	0.428	0.946 ± 0.017
x_4	(1.5, 5, 5, 6)	1.56%	447.376	0.051	0.970 ± 0.013

Table 6: Solutions to 10,000 (SP_{500}) for PGP2. We report coverage of SRP out of 500 repetitions for sample size of $n = 500$.

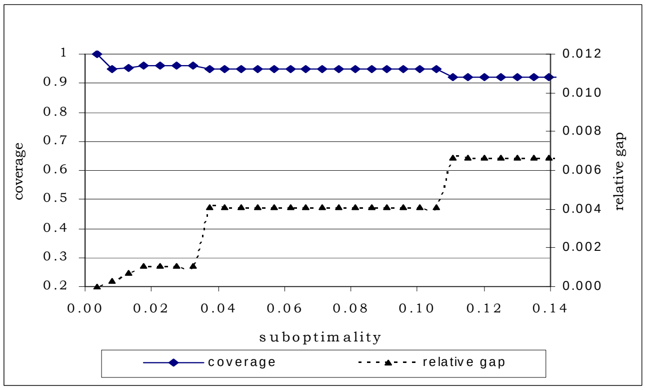

Figure 2: To generate a family of candidate solutions, $\hat{x}$, we obtain epsilon-optimal solutions to an (SP_{500}) of PGP2 using a primal-dual barrier method over a range of complementarity tolerances. The 500 scenarios for this (SP_{500}) yield $x_n^* = x^* = x_0$ when the problem is solved with sufficient precision. We plot the empirical coverage probability of the SRP out of 500 repetitions for sample size $n = 500$ and the candidate solution's relative gap, $\mu_{\hat{x}}/z^*$, versus the barrier method's duality gap.

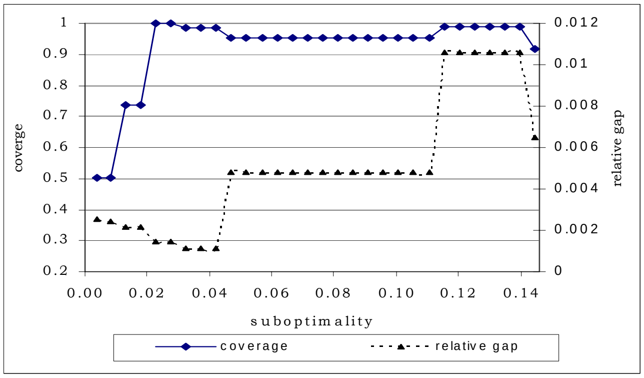

Figure 3: The figure is similar to Figure 2 except that the 500 scenarios for this (SP_{500}) yield $x_n^* = x_1$ when the problem is solved with sufficient precision.

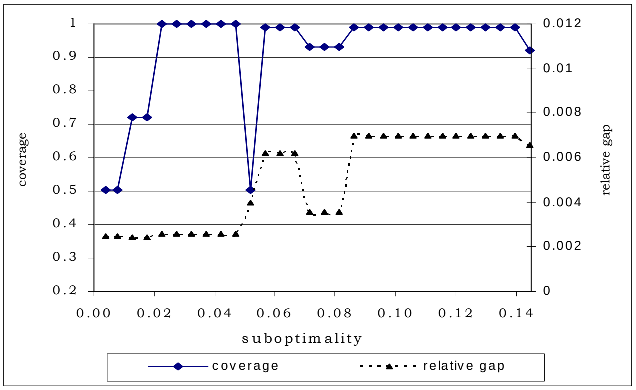

Figure 4: The figure is similar to Figures 2 and 3 except that the 500 scenarios for this (SP_{500}) yield $x_n^* = x_2$ when the problem is solved with sufficient precision.