Math. Program., Ser. B 108, 495–514 (2006)

Digital Object Identifier (DOI) 10.1007/s10107-006-0720-x

Güzin Bayraksan · David P. Morton

Assessing solution quality in stochastic programs

Received: August 31, 2004 / Accepted: November 11, 2005
Published online: April 25, 2006 – © Springer-Verlag 2006

Abstract. Determining whether a solution is of high quality (optimal or near optimal) is fundamental 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 we propose using ε-optimal solutions to strengthen the performance of our procedures.

1. Introduction

We consider a stochastic optimization problem of the form

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

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. $X \subseteq \mathbb{R}^d$ 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 in the statistics and operations research literature including our motivation, stochastic programs with recourse. The well-known two-stage stochastic linear program with recourse was introduced independently by [4, 9], in which

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

G. Bayraksan: Systems and Industrial Engineering, University of Arizona,
P.O. Box 210020, Tucson, AZ 85721, USA.
e-mail: `guzinb@sie.arizona.edu`

D. P. Morton: Graduate Program in Operations Research,
The University of Texas at Austin, Austin, Texas 78712, USA.
e-mail: `morton@mail.utexas.edu`

Mathematics Subject Classification (1991): 90C15

$X = \{x : Ax = b, x \geq 0\}$ and $\tilde{\xi} = \left(\tilde{q}, \tilde{W}, \tilde{r}, \tilde{T}\right)$ 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., [32]).

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

(A1) $f\left(\cdot, \tilde{\xi}\right)$ is continuous on X, w.p.1,
(A2) $E \sup_{x \in X} f^2\left(x, \tilde{\xi}\right) < \infty$,
(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. 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 most practical problems, a decision-maker would not be averse to specifying possibly large, but finite, simple bounds, $l \leq x \leq u$, making X bounded and hence compact, if also closed.

As the dimension of $\tilde{\xi}$ grows, (SP) gets harder and often impossible to solve exactly, unless f has simple structure, or the number of realizations is small. In cases where it is not possible to solve (SP) exactly, an intuitive approach is to resort to sampling and approximate the problem with

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

$\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$ 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, 11, 19, 28].

This paper describes 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 fundamental in optimization. Given a candidate solution $\hat{x}$, we define its quality by its optimality gap, $\mu_{\hat{x}} = Ef\left(\hat{x}, \tilde{\xi}\right) - z^*$. There are two difficulties associated with computing this quantity. First, z^* is not known and a lower bound (since we have 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 relaxations. 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 $Ef\left(\hat{x}, \tilde{\xi}\right)$ 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 [22]. Then, we show how to obtain a valid CI using only a single replication.

An earlier version of this result appeared in [3]. In Section 4, we provide variants of this procedure that use two replications. In Section 5, we compare the empirical coverage results of our procedures for a newsvendor problem and for two-stage stochastic linear programs with recourse. In Section 6, we give more insight on the procedures' performance. And, we propose using ε-optimal solutions to strengthen that performance. Finally, 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^*,

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

This result establishes that z_n^* has a negative bias, $Ez_n^* - z^* \le 0$. It can also be shown that $Ez_n^* \le Ez_{n+1}^*$ 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, $Ez_n^* \le z^* \le Ef\left(\hat{x}, \tilde{\xi}\right)$. The right inequality comes from suboptimality of $\hat{x}$. An upper bound on the optimality gap for $\hat{x}$ is then $Ef\left(\hat{x}, \tilde{\xi}\right) - Ez_n^*$. We estimate this quantity by

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

The first term on the right-hand side of (2) is an upper bound estimate and converges to $Ef\left(\hat{x}, \tilde{\xi}\right)$, 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), $G_n\left(\hat{x}\right) \ge 0$, 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 [22] 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 nonnormal 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, and, for later use, let z_α be that of the standard normal.

MRP:

Input: Value $\alpha \in (0,1)$ (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 $G_n^i(\hat{x}) = \frac{1}{n}\sum_{j=1}^{n}\left(f\left(\hat{x}, \tilde{\xi}^{ij}\right) - f\left(x_n^{i*}, \tilde{\xi}^{ij}\right)\right)$.
2. Calculate gap estimate and sample variance by

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

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, as given in (3). Due to the negative bias of z_n^*, $E\bar{G}(n_g) \geq Ef\left(\hat{x}, \tilde{\xi}\right) - z^*$. 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. [23] and used for global optimization of stochastic programs within a branch-and-bound method. Other algorithmic work that uses Monte Carlo simulation-based bounds and multiple replications includes [1, 20]. MRP has been applied to different kinds of problems in the literature including a bond portfolio model [5], a stochastic vehicle routing problem [18] and supply chain network design [27].

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 [13] 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 [29]. Higle and Sen [14] have also proposed a statistical lower bound that is rooted in duality. Dantzig and Infanger [10] and Higle and Sen [12, 16] 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 $\bar{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} f\big(x, \tilde{\xi}^i\big)$, $\sigma_{\hat{x}}^2(x) = \mathrm{var}\big[f\big(\hat{x}, \tilde{\xi}\big) - f\big(x, \tilde{\xi}\big)\big]$ and $s_n^2(x) = \frac{1}{n-1}\sum_{i=1}^{n}\big[\big(f(\hat{x}, \tilde{\xi}^i) - f\big(x, \tilde{\xi}^i\big)\big) - (\bar{f}_n(\hat{x}) - \bar{f}_n(x))\big]^2$. 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^*. Below we state the single replication procedure (SRP).

SRP:

Input: Value $\alpha \in (0, 1)$, 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

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

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

$$\left[0, G_n(\hat{x}) + \frac{t_{n-1,\alpha} s_n\left(x_n^*\right)}{\sqrt{n}}\right]. \tag{5}$$

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, $f\left(\hat{x}, \tilde{\xi}^i\right) - f\left(x_n^*, \tilde{\xi}^i\right)$ 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 $s_n^2\left(x_n^*\right)$ is the corresponding sample variance. Below, we show how carrying out 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 the following proposition, which establishes consistency of the estimators. The proposition's hypothesis defines $x_{\min}^*$ and $x_{\max}^*$. In words, these are the optimal solutions to (SP) with minimum and maximum variance, respectively, of $f\left(\hat{x}, \tilde{\xi}\right) - f\left(x, \tilde{\xi}\right)$.

Proposition 1. *Let* $X^* = \arg\min_{x\in X} Ef(x,\tilde{\xi})$, $x^*_{\min} \in \arg\min_{x\in X^*} \operatorname{var}[f(\hat{x},\tilde{\xi}) - f(x,\tilde{\xi})]$, *and* $x^*_{\max} \in \arg\max_{x\in X^*} \operatorname{var}[f(\hat{x},\tilde{\xi}) - f(x,\tilde{\xi})]$. *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) $\sigma^2_{\hat{x}}(x^*_{\min}) \le \liminf_{n\to\infty} s^2_n(x^*_n) \le \limsup_{n\to\infty} s^2_n(x^*_n) \le \sigma^2_{\hat{x}}(x^*_{\max})$, *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 [24, p.69]. (A1)–(A3) imply $\bar{f}_n(x)$ converges uniformly to $Ef(x,\tilde{\xi})$, w.p.1 on X. This coupled with (i) implies (ii). To prove (iii), we first show the sequence of continuous functions $s^2_n(x)$ 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

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

The first term in the curly brackets is a sample mean of i.i.d. random variables and by Lemma A1 of [24, 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., $\sup_{x\in X}|\bar{g}_n(x) - Eg(x,\tilde{\xi})| \to 0$, w.p.1. This implies

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

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

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

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 to the set of optimum solutions, X^*. Similarly, $\left\{s_n^2\left(x_n^*\right)\right\}$ 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), $\sigma_{\hat{x}}^2\left(x_{\max}^*\right) < \infty$. When X^* is a singleton, $x_n^* \to x^*$, w.p.1 and $\liminf_{n\to\infty} s_n^2\left(x_n^*\right) = \limsup_{n\to\infty} s_n^2\left(x_n^*\right) = \sigma_{\hat{x}}^2(x^*)$, w.p.1. We next present the main result regarding the validity of the SRP.

Theorem 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 $\tilde{\xi}$. Given $0 < \alpha < 1$, for the SRP,*

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

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,

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

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

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

where in (8) we assume $\sigma_{\hat{x}}^2\left(x_{\min}^*\right) > 0$. Note that if $\sigma_{\hat{x}}^2\left(x_{\min}^*\right) = 0$ then $\operatorname{var}\left[\bar{f}_n\left(\hat{x}\right) - \bar{f}_n\left(x_{\min}^*\right)\right] = \frac{1}{n}\sigma_{\hat{x}}^2\left(x_{\min}^*\right) = 0$ and it follows from (7) that (6) is again trivial. Let $D_n = \frac{\left(\bar{f}_n\left(\hat{x}\right) - \bar{f}_n\left(x_{\min}^*\right)\right) - \mu_{\hat{x}}}{\sigma_{\hat{x}}\left(x_{\min}^*\right)/\sqrt{n}}$, $a_n = \frac{s_n\left(x_n^*\right)}{\sigma_{\hat{x}}\left(x_{\min}^*\right)}$ 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

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

Taking limits we obtain,

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

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 $\bar{f}_n(\hat{x}) - \bar{f}_n\left(x_{\min}^*\right)$

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 1 justifies construction of the approximate $(1-\alpha)$-level one-sided confidence interval for $\mu_{\hat{x}} = Ef\big(\hat{x}, \tilde{\xi}\big) - z^*$, given in (5) without requiring $G_n\left(\hat{x}\right) = \bar{f}_n\left(\hat{x}\right) - z^*_n$ 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. This notion of projection is formalized in [24, Theorem 6.4.2] which states that the scaled errors of z^*_n converge in distribution to the minimum of a collection of normal random variables. More specifically, $\sqrt{n}\left(z^*_n - z^*\right) \Rightarrow \inf_{x \in X^*} Z(x)$, where each $Z(x)$ is a mean-zero normal random variable defined via $\sqrt{n}\left(\bar{f}_n(x) - z^*\right) \Rightarrow Z(x)$, $x \in X^*$. (Here, "$\Rightarrow$" denotes convergence in distribution.) If X^* is a singleton then $\inf_{x \in X^*} Z(x)$ is normally distributed. Otherwise, the precise nature of this random element is dictated by the dependency among $Z(x)$ for $x \in X^*$. We also note that because we estimate the sample variance in the SRP, the statement of the procedure uses the more conservative Student's t quantile, $t_{n-1,\alpha}$.

Before proceeding, we recall some terminology associated with confidence interval estimation (see, e.g., [8, §9.1]). An *interval estimator*, $\tilde{I}$, of a real-valued parameter of interest (e.g., $\mu_{\hat{x}}$) is a random set. The MRP and SRP form respective interval estimators (3) and (5). The *coverage probability*, or simply the *coverage*, of an interval estimator is the probability that the random interval $\tilde{I}$ contains the parameter of interest, e.g., $P\big(\mu_{\hat{x}} \in \tilde{I}\big)$. Theorem 1 characterizes the coverage probability of the SRP's interval estimator $\tilde{I} = \tilde{I}_n$, as the sample size n grows large. Similarly, as indicated in Section 2, the standard CLT gives a result for the coverage probability of the MRP's interval estimator $\tilde{I} = \tilde{I}_{n_g}$, as the number of replications n_g grows large. Such theoretical results are typically used to justify employing the associated procedure, provided the sample size n or replication size n_g is sufficiently large. Practical interpretation of what is meant by "sufficiently large" is usually guided by extensive empirical testing of the procedures. With this aim, preliminary empirical coverage results are described in Sections 5 and 6. First, however, we examine undesirable effects that can arise with small sample sizes, and we provide procedures to mitigate these effects.

MRP is a procedure in which we use $n_g \ge 30$ replications and SRP is 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 difficulties. 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 $G_n\left(\hat{x}\right)$ and the variance estimate $s^2_n\left(x^*_n\right)$ 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, $\left\{\min E[\tilde{\xi}x] : -1 \le x \le 1\right\}$, 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 $G_n\left(\hat{x}\right) = s_n\left(x_n^*\right) = 0$. Hence, for the problem instance with $\mu = 0.1$, the coverage probability $P\left(\mu_{\hat{x}} \le G_n\left(\hat{x}\right) + t_{n-1,\alpha}s_n\left(x_n^*\right)/\sqrt{n}\right) \le 1 - P\left(\bar{\xi} < 0\right) \simeq 0.760$ 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 $G_n\left(\hat{x}\right)$ and $s_n\left(x_n^*\right)$, by performing two independent replications, one to estimate the gap and the other to estimate $s_n\left(x_n^*\right)$.

I2RP:

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

3′. Calculate $G_n^1\left(\hat{x}\right)$ in the same way as given in (2), and carryout the following sub-steps to calculate the sample variance:
  3′.1. Sample i.i.d. observations $\tilde{\xi}^{n+1}, \tilde{\xi}^{n+2}, \ldots, \tilde{\xi}^{2n}$ from the distribution of $\tilde{\xi}$,
  3′.2. Solve (SP_n) defined with respect to $\tilde{\xi}^{n+1}, \tilde{\xi}^{n+2}, \ldots, \tilde{\xi}^{2n}$ to obtain x_n^{2*},
  3′.3. Form $s_n^2\left(x_n^{2*}\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(f\left(\hat{x}, \tilde{\xi}^{n+i}\right) - f\left(x_n^{2*}, \tilde{\xi}^{n+i}\right)\right) - \left(\bar{f}_n(\hat{x}) - \bar{f}_n\left(x_n^{2*}\right)\right)\right]^2$, 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, $G_n^1\left(\hat{x}\right)$, comes from the first replication and the sample standard deviation, $s_n\left(x_n^{2*}\right)$, 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 1, it can easily be shown that this procedure provides an asymptotically valid confidence interval. We formally state this in the next theorem.

Theorem 2. *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,*

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

Proof. The proof of Theorem 1 remains the same when $s_n^2\left(x_n^*\right)$ 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 $G_n^i\left(\hat{x}\right)$ and $s_n^2\left(x_n^{i*}\right)$.
2′. Calculate the estimates by taking the average,

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

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

$$\left[0, G_n'\left(\hat{x}\right) + \frac{t_{2n-1,\alpha} s_n'}{\sqrt{2n}}\right].$$

Unlike the MRP, the sample variance, $s_n^2\left(x_n^{i*}\right)$, 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, e.g., [8, p.396]. It is a consistent estimator, in the sense that $\sigma_{\hat{x}}^2\left(x_{\min}^*\right) \le \liminf_{n\to\infty} s_n^{2\,\prime} \le \limsup_{n\to\infty} s_n^{2\,\prime} \le \sigma_{\hat{x}}^2\left(x_{\max}^*\right)$, w.p.1, by Proposition 1. A2RP provides an asymptotically valid CI on the optimality gap, as stated in the theorem below.

Theorem 3. *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,*

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

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

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

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

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

Since $\bar{f}_{2n}(\hat{x}) - \bar{f}_{2n}(x^*_{\min})$ is a sample mean of i.i.d. random variables, by the CLT, $\sqrt{2n}\big(\big(\bar{f}_{2n}(\hat{x}) - \bar{f}_{2n}(x^*_{\min})\big) - \mu_{\hat{x}}\big)$ 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 the $\alpha \leq 1/2$ case is analogous to that of Theorem 1, 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 problem instance of 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

Theorems 1, 2 and 3 show that the confidence intervals formed using the three procedures, SRP, I2RP and A2RP are asymptotically valid. In other words, these theorems establish that the CIs have the desired coverage probability as the sample size grows large ($n \to \infty$). Example 1 suggests that these results might not hold for SRP for small values of n. To investigate how the procedures behave for small sample sizes, we apply them to a newsvendor problem under uniform demand and to three small two-stage stochastic linear programs from the literature, and we report empirical coverage probabilities.

When using a specific procedure on one of these problems, we perform the following experiment: We set $\alpha = 0.10$ and repeat the procedure k times for varying values of sample sizes, n. For a given value of n, we form $\hat{p}$, the fraction of the k repetitions in which the CI contains the true gap. Quantity $\hat{p}$ is an estimator of the true coverage probability p. Ideally, we would like to have $p \geq 1 - \alpha = 0.90$. Note that $\hat{p}$ is a (scaled) binomial random variable, and for sufficiently large values of k, we can use the CLT to form a 90% confidence interval on the true coverage probability via $\hat{p} \pm 1.645\left(\hat{p}\left(1-\hat{p}\right)/k\right)^{1/2}$. So, the output of one of our procedures is a confidence interval on the optimality gap. In this section, our experiments yield yet another confidence interval, namely, a CI on the coverage probability associated with the procedure's output for various values of n. We note that this is a standard way to assess coverage results in the simulation literature, e.g., [21, pp. 508–509].

In order to lessen the effect of variation in the samples, for each sample size n, we use the same observations from 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 $n/2$. For MRP, we set the number of replications to $n_g = 30$ and typically take k smaller than the single or two-replication procedures as the computational requirement is higher. To understand how the estimator $s_n\left(x_n^*\right)$ affects coverage, we form another CI by taking $G_n\left(\hat{x}\right)$ from SRP and replacing $s_n\left(x_n^*\right)$ by $\sigma_{\hat{x}}\left(x^*\right)$ in (5). We denote this procedure as TRUE. We now turn to the computational results for the test problems.

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., [6, 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. The problem is formulated as $\max\{-cx + rE\min\{x, \tilde{\xi}\} : x \geq 0\}$, 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. 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 $\hat{x} = 8.775$ with $Ef(\hat{x}, \tilde{\xi}) = 30$ and with an optimality gap of $\mu_{\hat{x}} = 3\frac{1}{3}$. This candidate solution has $\sigma_{\hat{x}}^2(x^*) = 140.79$. For the SRP, I2RP, A2RP and TRUE we construct k =100,000 confidence intervals and for the MRP, we construct k =10,000 intervals for each value of the sample size. We take sample sizes, n, from 50 to 1000. Table 1 summarizes the results. For example, when n =1000, for the MRP, the table indicates $\hat{p} = 0.9267$ so that we are confident at level 0.90 that the true coverage probability, i.e., the left-hand side of (4), is in [0.9224, 0.9310].

Table 1. Empirical coverage results, $\hat{p} \pm 1.645\left(\hat{p}\left(1-\hat{p}\right)/k\right)^{1/2}$, 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 $G_n\left(\hat{x}\right)$ from SRP and replacing $s_n\left(x_n^*\right)$ by $\sigma_{\hat{x}}(x^*)$ in (5)

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

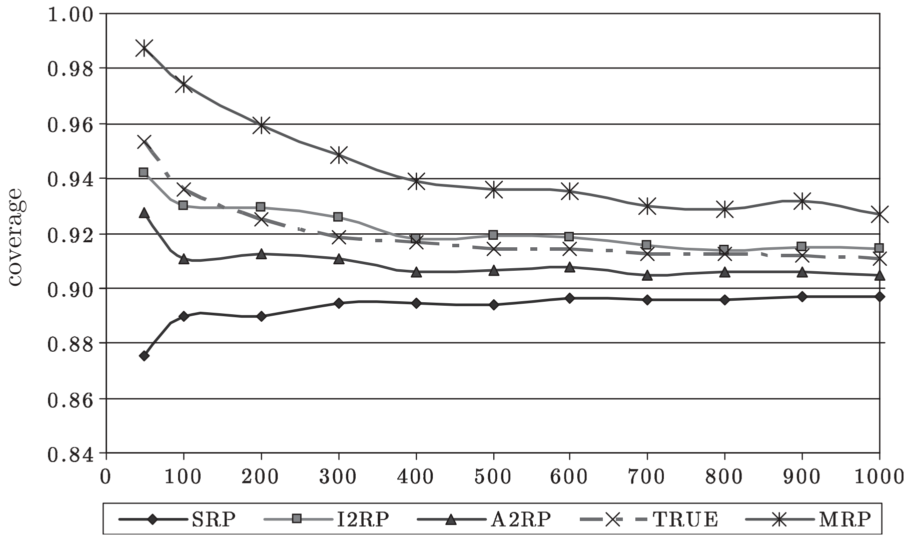

Fig. 1. Empirical coverage probability $(\hat{p})$ versus sample size (n) for the newsvendor problem

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, $G_n\left(\hat{x}\right)$ is more variable and we have observed from the individual replications that when it is small, $s_n\left(x_n^*\right)$ 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 our 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 [15, pp. 3–10]. The third test problem we use, denoted APL1P, can be found in [17]. 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. In particular, Table 2 lists the candidate solutions we use for each problem in the "$\hat{x}$" column. For example, the dimension of the candidate solution $\hat{x}$ for PGP2 is 4, and this candidate solution is approximately 0.25% ($= 100 \times \mu_{\hat{x}}/z^*$) from the optimal. The final column gives the values of $\sigma_{\hat{x}}(x^*) = \text{var}^{1/2}[f(\hat{x}, \tilde{\xi}) - f(x^*, \tilde{\xi})]$. (The optimum solution to each of the three problems is unique. We also note that for CEP1 our x corresponds, in terms of the notation of [15, p.5], to the vector $(x_1, \ldots, x_4, z_1, \ldots, z_4)$.)

To solve the sampling problems, we used the regularized decomposition (RD) algorithm of [25]. An accelerated implementation of this algorithm is in C++ [26], and we have modified this code to perform the tests. For each test problem under SRP, I2RP, A2RP and TRUE, we construct $k = 500$ confidence intervals for various values of the sample size n. For MRP, we use $n_g = 30$ and construct $k = 100$ confidence intervals for the same values of n. Tables 3, 4 and 5 list results for CEP1, PGP2 and APL1P, respectively. 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 often results in 100% coverage. 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 the values of

Table 2. Candidate solutions used in tests

Problem	$\hat{x}$	$Ef(\hat{x}, \tilde{\xi})$	$\mu_{\hat{x}}$	$\sigma_{\hat{x}}(x^*)$
CEP1	(0, 125, 875, 2500, 0, 625, 1375, 3000)	393,288.01	38,129.09	55,690.34
PGP2	(1.5, 5.5, 5, 4.5)	448.46	1.14	82.69
APL1P	(1111.11, 2300)	24,807.16	164.84	1,893.03

Table 3. Empirical coverage results for CEP1

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 4. Empirical coverage results for PGP2

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 5. Empirical coverage results for APL1P

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	0.992 ± 0.007
500	1 ± 0	0.902 ± 0.022	0.940 ± 0.017	0.908 ± 0.021	0.966 ± 0.013
600	1 ± 0	0.894 ± 0.023	0.944 ± 0.017	0.910 ± 0.021	0.968 ± 0.013
700	1 ± 0	0.910 ± 0.021	0.964 ± 0.014	0.934 ± 0.018	0.966 ± 0.013
800	1 ± 0	0.910 ± 0.021	0.962 ± 0.014	0.934 ± 0.018	0.962 ± 0.014
900	1 ± 0	0.906 ± 0.021	0.965 ± 0.014	0.934 ± 0.018	0.948 ± 0.016
1000	1 ± 0	0.906 ± 0.021	0.956 ± 0.015	0.926 ± 0.019	0.956 ± 0.015

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

As illustrated in Example 1 and the computational results of the previous section, in some problems the nature of the candidate solution $\hat{x}$ and the solution(s) x_n^* can lead to inferior performance of our procedures. The procedures can work poorly when $\hat{x} \notin X^*$ is chosen to be a candidate solution from an auxiliary sampling problem, and this solution has a high probability of occurrence as an x_n^* solution to an (SP$_n$) used in our procedures. In such cases, especially the SRP can report a CI width which is too narrow, over-stating the quality of the candidate solution. (The cause is actually a bit more subtle, as we describe below in the context of PGP2.) Two-replication procedures reduce this effect, but they may not be enough. Let us return to the problem discussed in Examples 1 and 2 and show that for a fixed value of n, however large, the two-replication procedures can have low coverage.

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 $\mu \to 0$. □

We say two points $x', x'' \in X$ coincide if $\sigma^2_{x'}(x'') \equiv \mathrm{var}[f(x', \tilde{\xi}) - f(x'', \tilde{\xi})] = 0$ and that they nearly coincide if this variance is small. When a candidate solution $\hat{x}$ nearly

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

x_i	Frequency	$Ef(x_i, \tilde{\xi})$	μ_{x_i}	Coverage	$\sigma_{x_i}(x_0)$	$\sigma_{x_i}(x_1)$
$x_0 = (1.5, 5.5, 5, 5.5)$	44.49%	447.324	0	1 ± 0	0.00	82.69
$x_1 = (1.5, 5.5, 5, 4.5)$	43.90%	448.464	1.140	0.504 ± 0.037	82.69	0.00
$x_2 = (1.5, 5, 5, 5)$	4.44%	448.511	1.186	0.504 ± 0.037	82.83	6.65
$x_3 = (1.5, 5.5, 5, 5)$	3.54%	447.752	0.428	0.946 ± 0.017	41.10	43.95
$x_4 = (1.5, 5, 5, 6)$	1.56%	447.376	0.051	0.970 ± 0.013	6.64	83.64

coincides with a high probability $x_n^* \notin X^*$, the gap random variable is nearly degenerate, and we can have undercoverage. Of course, this occurs if $\hat{x} = x_n^*$, but it can also occur when these two points are distinct as we illustrate next for PGP2.

Table 6 enumerates as $x_0, x_1, \ldots, x_4$, the most frequent x_n^* solutions 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 using $k = 500$ repetitions of the SRP, again for a sample size of $n = 500$. The optimal solution, $x^* = x_0$, and the candidate solution used in the previous section, x_1, each appear roughly 45% of the time. For each solution, x_i, $i = 0, \ldots, 4$, the table also lists $\sigma_{x_i}(x)$ for $x = x_0$ and $x = x_1$. So, x_1 is suboptimal to (SP) but is a frequent solution to (SP_n) and hence, has poor coverage. Point x_2 is a relatively infrequent solution to (SP_n), but still has poor coverage because it nearly coincides with x_1 (i.e., $\sigma_{x_2}(x_1)$ is relatively small). Finally, we note that even though x_4 nearly coincides with x_0, x_4 has good coverage since coinciding with an optimal solution does not degrade the coverage probability.

One way to try to avoid the difficulties illustrated above 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. In the next subsection, we propose another approach in which we solve (SP_n) suboptimally when carrying out, e.g., the SRP or the A2RP.

6.1. ε-Optimal solutions

In this section, we consider an approach based on ε-optimal solutions to help avoid $\hat{x}$ and x_n^* coinciding. While $\hat{x}$ can be generated by any method, typically, $\hat{x}$ may be obtained by solving a sample-mean problem with sample size $n_{\hat{x}}$, $\left(SP_{n_{\hat{x}}}\right)$. Then, we 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 $\left(SP_{n_{\hat{x}}}\right)$ and (SP_n). There are clearly a number of possibilities, but since our focus is on assessing solution quality, we use ε-optimal solutions when solving (SP_n) for the SRP. We believe solving (SP_n) approximately makes sense, particularly in light of the fact that our procedure's output is a confidence interval. We similarly use ε-optimal solutions for the two sampling problems solved in I2RP and A2RP.

For our computational results, as before, we used the RD algorithm of [25, 26]. This is similar to the multicut version of the L-Shaped method [7], except that the master problem has a regularizing proximal term. The quadratic proximal term can result in considerable computational savings. Because of the regularizing term, the objective function value of the master problem in RD does not provide a lower bound on the problem's optimal value. So, we first run RD with the proximal term and then remove the proximal term in order to obtain a lower bound on z_n^*. When this procedure terminates, we have $\underline{z}_n^* \leq z_n^* \leq \overline{z}_n^*$. If we solve the problem precisely, $\underline{z}_n^* = z_n^* = \overline{z}_n^*$. However, in this section we solve the sampling problem (SP_n) with varying degrees of suboptimality, and use the lower bound, $\underline{z}_n^*$, for constructing the gap and the sampling variance estimates.

We applied this methodology to the two candidate solutions, x_1 and x_2 of PGP2, which have poor coverage results, using SRP, I2RP and A2RP. Figures 2 and 3 show the results of our computations. On the y-axis, we plot the procedures' empirical coverage probability out of $k = 500$ repetitions for a sample size of $n = 500$. On the x-axis, we plot the level of suboptimality with which we solved the sampling problems (SP_n) to obtain the confidence interval estimators. Suboptimality is measured as $\left(\overline{z}_n^* - \underline{z}_n^*\right) / \min\left\{\left|\overline{z}_n^*\right|, \left|\underline{z}_n^*\right|\right\}$. The vertical dashed line is the ratio of the optimality gap of the candidate solution to the optimal value of PGP2, $\mu_{\hat{x}}/z^*$. This value is 2.5×10^{-3} for $\hat{x} = x_1$. Even though this is measured with respect to z^* and suboptimality is measured with respect to z_n^*, we expect to have almost 100% coverage after this point. However, all of the procedures reach the desired level of coverage earlier than this, with a suboptimality level of around 1.5×10^{-3}. The two-replication procedures reach the desired coverage of 0.90 at a suboptimality level of 1×10^{-3}.

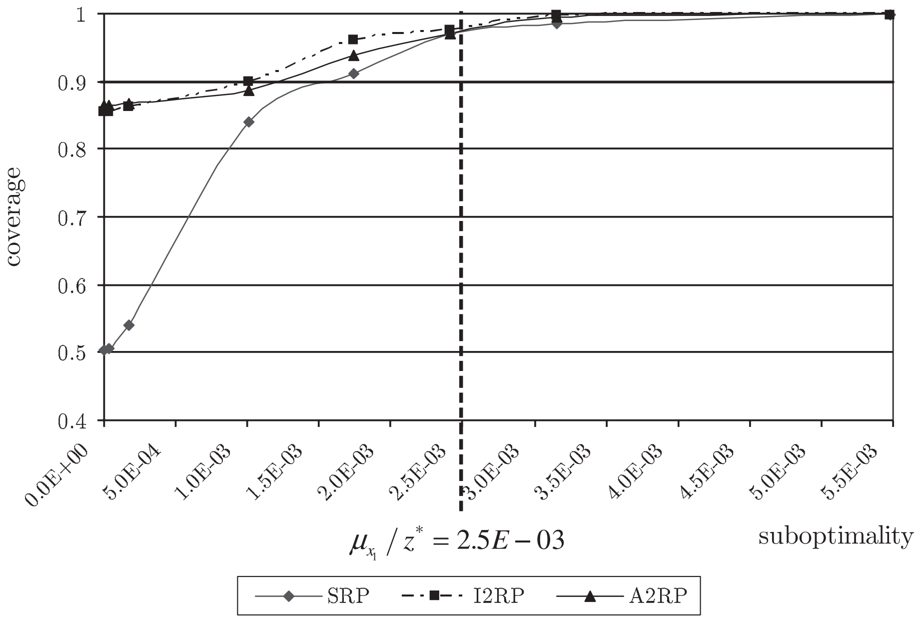

Fig. 2. We plot the empirical coverage probability out of $k = 500$ repetitions with sample size $n = 500$ for the candidate solution $\hat{x} = x_1$ of PGP2. We solved the sampling problems (SP_n) in the estimation of SRP, I2RP and A2RP with varying levels of suboptimality. The vertical dashed line represents the ratio of the optimality gap of x_1 to the optimal value of PGP2

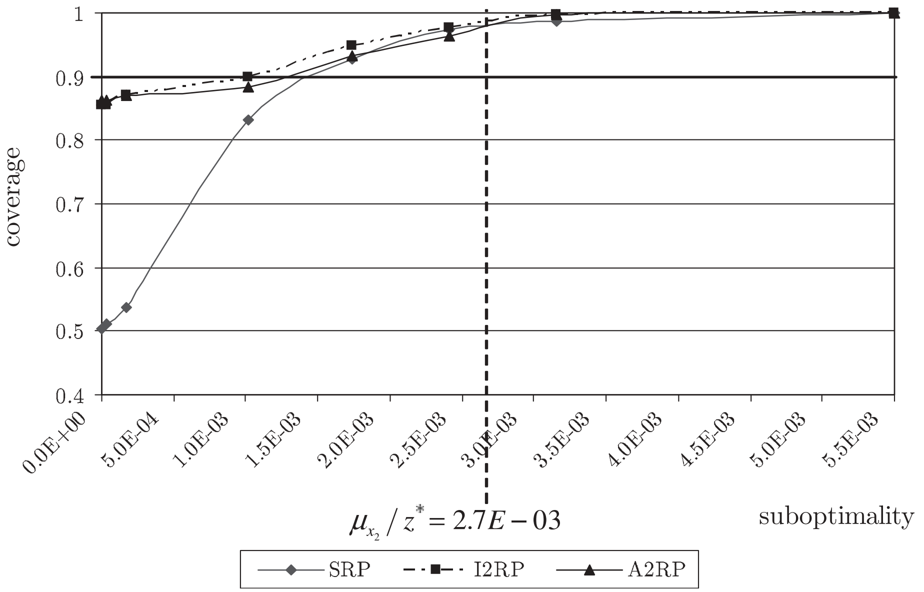

Fig. 3. This figure is similar to Figure 2 except that $\hat{x} = x_2$

6.2. *Preliminary guidelines*

The goal of this paper is to present computationally attractive alternatives to the MRP for assessing solution quality in stochastic programs. That said, there are at least three situations in which we recommend the use of MRP instead of SRP or one of the two-replication procedures. First, if computation is quite cheap, then we recommend the more conservative MRP. Second, if the desire for conservative coverage results is paramount we again recommend MRP. Third, if sampling error dominates the point estimate of the optimality gap then we recommend MRP, because it is computationally cheaper to reduce sampling error by increasing the number of replications in the MRP rather than increasing the sample size in the SRP or one of the two-replication procedures. Otherwise, based on our computational experience to date, we recommend using A2RP with ε-optimal solutions. Relative to the SRP, the two-replication procedures help to avoid $\hat{x}$ and x_n^* coinciding and hence improve coverage results. The use of ε-optimal solutions further assist in improving coverage results for two reasons. First, like the two-replication procedures, the likelihood of coinciding solutions is reduced. Second, our use of lower bounds $\underline{z}_n^*$ in place of z_n^* when solving (SP$_n$) suboptimally has the effect of inflating the confidence interval width. In our preliminary computations, small suboptimality levels (e.g., 0.10–0.20%) were enough to reach the desired level of coverage. Finally, we note that relative to I2RP, the A2RP uses all of the observations in both the point estimate of the gap and the sampling-error estimate. Empirically, this advantage appears to come without detrimental effects on coverage.

7. Conclusions

In this paper, we develop 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 introduce 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 $\hat{x} \notin X^*$ to an auxiliary sampling problem (SP_n) is chosen as the candidate solution, and this solution has a high probability of occurring as the x_n^* used in the SRP, the coverage probability can be quite low. So, we develop variants of this procedure that use two replications to lessen this effect. And, with the same motivation we propose use of ε-optimal solutions when solving the sampling problem (SP_n) used in our confidence interval estimation procedures. Our preliminary computational results seem to indicate that at quite modest values of suboptimization, this approach works well.

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

References

1. Ahmed, S., Shapiro, A.: The sample average approximation method for stochastic programs with integer recourse. Optimization Online. http://www.optimization-online.org/ (2002)
2. Attouch, H., Wets, R.J.-B.: Approximation and convergence in nonlinear optimization. In: Mangasarian O., Meyer R., Robinson S. (eds.), Nonlinear Programming, **4**, Academic Press, New York, 1981, pp. 367–394
3. Bayraksan, G., Morton, D.P.: Testing solution quality in stochastic programming: a single replication procedure. In: Proceedings of the 16th Symposium of IASC on Computational Statistics. Physica-Verlag/Springer, Prague, Czech Republic, 2004
4. Beale, E.M.L.: On minimizing a convex function subject to linear inequalities. J. Roy. Stat. Soc. **17B**, 173–184 (1955)
5. Bertocchi, M., Dupačová, J., Moriggia, V.: Sensitivity of bond portfolio's behavior with respect to random movements in yield curve: a simulation study. Ann. Oper. Res. **99**, 267–286 (2000)
6. Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer-Verlag, New York, 1997
7. Birge, J.R., Louveaux, F.V.: A multicut algorithm for two-stage stochastic linear programs. European J. Oper. Res. **34**, 384–392 (1988)
8. Casella, G., Berger, R.L.: Statistical Inference. Duxbury Press, Belmont, California, 1990
9. Dantzig, G.B.: Linear programming under uncertainty. Manage. Sci. **1**, 197–206 (1955)
10. Dantzig, G.B., Infanger, G.: A probabilistic lower bound for two-stage stochastic programs. Technical Report SOL 95-6, Department of Operations Research, Stanford University, November 1995
11. Dupačová, J., Wets, R.J.-B.: Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. Ann. Stat. **16**, 1517–1549 (1988)
12. Higle, J.L., Sen, S.: Statistical verification of optimality conditions for stochastic programs with recourse. Ann. Oper. Res. **30**, 215–240 (1991)
13. Higle, J.L., Sen, S.: Stochastic decomposition: an algorithm for two-stage linear programs with recourse. Math. Oper. Res. **16**, 650–669 (1991)
14. Higle, J.L., Sen, S.: Duality and statistical tests of optimality for two stage stochastic programs. Math. Program. **75**, 257–275 (1996)
15. Higle, J.L., Sen, S.: Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming. Kluwer Academic Publishers, Dordrecht, 1996

16. Higle, J.L., Sen, S.: Statistical approximations for stochastic linear programming problems. Ann. Oper. Res. **85**, 173–192 (1999)
17. Infanger, G.: Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs. Ann. Oper. Res. **39**, 69–95 (1992)
18. Kenyon, A.S., Morton, D.P.: Stochastic vehicle routing with random travel times. Transport. Sci. **37**, 69–82 (2003)
19. King, A.J., Rockafellar, R.T.: Asymptotic theory for solutions in statistical estimation and stochastic programming. Math. Oper. Res. **18**, 148–162 (1993)
20. Kleywegt, A.J., Shapiro, A., Homem-de-Mello, T.: The sample average approximation method for stochastic discrete optimization. SIAM J. Optim. **12**, 479–502 (2001)
21. Law, A.M., Kelton, W.D.: Simulation Modeling and Analysis, 3rd Edition. McGraw-Hill, Boston, 2000
22. Mak, W.K., Morton, D.P., Wood, R.K.: Monte Carlo bounding techniques for determining solution quality in stochastic programs. Oper. Res. Lett. **24**, 47–56 (1999)
23. Norkin, V.I., Pflug, G.Ch., Ruszczyński, A.: A branch and bound method for stochastic global optimization. Math. Program. **83**, 425–450 (1998)
24. Rubinstein, R.Y., Shapiro, A.: Discrete Event Systems: Sensitivity and Stochastic Optimization by the Score Function Method. John Wiley & Sons, Chichester, 1993
25. Ruszczyński, A.: A regularized decomposition method for minimizing a sum of polyhedral functions. Math. Program. **35**, 309–333 (1986)
26. Ruszczyński, A., Świetanowski, A.: Accelerating the regularized decomposition method for two stage stochastic linear problems. European J. Oper. Res. **101**, 328–342 (1997)
27. Santoso, T., Ahmed, S., Goetschalckx, M., Shapiro, A.: A stochastic programming approach for supply chain network design under uncertainty. European J. Oper. Res. **167**, 96–115 (2005)
28. Shapiro, A.: Asymptotic analysis of stochastic programs. Ann. Oper. Res. **30**, 169–186 (1991)
29. Shapiro, A., Homem-de-Mello, T.: A simulation-based approach to two-stage stochastic programming with recourse. Math. Program. **81**, 301–325 (1998)
30. Shapiro, A., Homem-de-Mello, T.: On rate of convergence of Monte Carlo approximations of stochastic programs. SIAM J. Optim. **11**, 70–86 (2001)
31. Shapiro, A., Homem-de-Mello, T., Kim, J.: Conditioning of convex piecewise linear stochastic programs. Math. Program. **94**, 1–19 (2002)
32. Wallace, S.W., Ziemba, W.T., (eds.): Applications of Stochastic Programming. MPS-SIAM Series in Optimization, Philadelphia, 2005