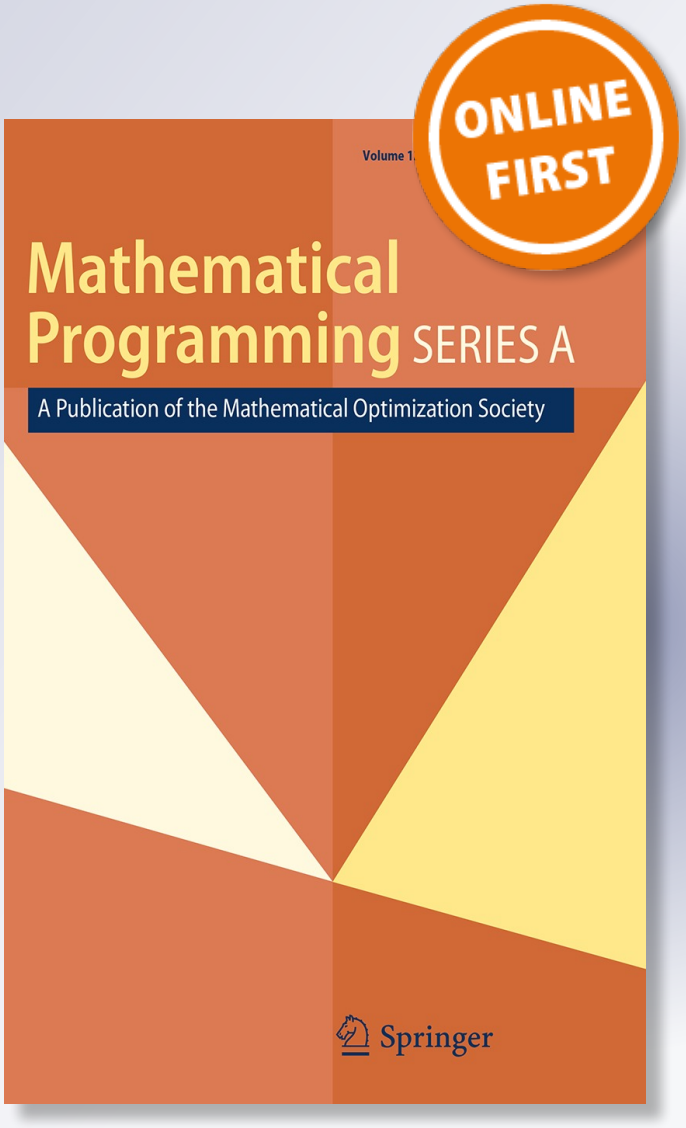
Simulation-based confidence bounds for two-stage stochastic programs

Peter W. Glynn & Gerd Infanger

Mathematical Programming
A Publication of the Mathematical Optimization Society

ISSN 0025-5610

Math. Program.
DOI 10.1007/s10107-012-0621-0



Springer

Your article is protected by copyright and all rights are held exclusively by Springer-Verlag Berlin Heidelberg and Mathematical Optimization Society. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Author's personal copy

Math. Program., Ser. A
DOI 10.1007/s10107-012-0621-0

FULL LENGTH PAPER

Simulation-based confidence bounds for two-stage stochastic programs

Peter W. Glynn · Gerd Infanger

Received: 14 April 2008 / Accepted: 25 November 2012
© Springer-Verlag Berlin Heidelberg and Mathematical Optimization Society 2013

Abstract This paper provides a rigorous asymptotic analysis and justification of upper and lower confidence bounds proposed by Dantzig and Infanger (A probabilistic lower bound for two-stage stochastic programs, Stanford University, CA, 1995) for an iterative sampling-based decomposition algorithm, introduced by Dantzig and Glynn (Ann. Oper. Res. 22:1–21, 1990) and Infanger (Ann. Oper. Res. 39:41–67, 1992), for solving two-stage stochastic programs. The paper provides confidence bounds in the presence of both independent sampling across iterations, and when common samples are used across different iterations. Confidence bounds for sample-average approximation then follow as a special case. Extensions of the theory to cover use of variance reduction and the dropping of cuts are also presented. An extensive empirical investigation of the performance of these bounds establishes that the bounds perform reasonably on realistic problems.

Mathematics Subject Classification 90C15 Stochastic programming · 90C05 Linear programming · 97K50 Probability theory · 62G15 Tolerance and confidence regions

Dedicated to the memory of George B. Dantzig.

P. W. Glynn · G. Infanger (✉)
Stanford, CA, USA
e-mail: infanger@stanford.edu

Published online: 16 February 2013

Author's personal copy

1 Introduction

Consider the two-stage stochastic linear program (with recourse) given by

$$\begin{array}{ll} \min & cx + z^*(x) \\ \text{s/t} & Ax = b \\ & x \geq 0, \end{array} \tag{1}$$

where $z^*(x) = Ez^*(x; \Lambda)$ and $z^*(x; \Lambda)$ is the minimum of the second-stage linear program

$$\begin{array}{ll} \min & qy \\ \text{s/t} & Wy = h + Tx \\ & y \geq 0, \end{array}$$

and $\Lambda \equiv (q, T, W, h)$ is random. When the sample space corresponding to Λ contains a large number of outcomes (or is countably infinite or continuous), use of sampling-based algorithms for solving (1) becomes attractive.

A key element in the reliable application of any numerical method is the ability to accurately assess its error. For a sampling-based procedure, such an assessment typically comes in the form of confidence statements. Such confidence statements have been previously proposed for the class of sampling-based procedures based on sample-average approximation (SAA). Mak et al. [30] show, for example, how upper and lower one-sided confidence intervals for the minimal objective value can be calculated (that are asymptotically valid as the sample size tends to infinity). Their approach actually applies to much more general stochastic programs than (1). In related work, Shapiro and Homem-de-Mello [39] propose a stopping criterion for SAA, as applied to (1), that is based on the asymptotic behavior of confidence regions related to the first-order optimality conditions for two-stage stochastic programs with recourse.

This paper's main contribution is the rigorous development of upper and lower one-sided confidence intervals for the minimum objective value of (1), when the underlying solution algorithm is the iterative sampling-based decomposition approach introduced by Dantzig and Glynn [10] and Infanger [20] that uses independent subproblems at each iteration. Our analysis also can be carried over to iterative procedures that use common random numbers so that previously generated subproblems are re-used at later iterations. Given that our analysis is largely motivated by such iterative implementations, we permit the sample size used in each iteration to increase from iteration to iteration (so as to permit the possibility of improving the accuracy over time). Because SAA is non-trivial to implement efficiently when one chooses to increase the sample size (especially when W is random), we view the existing fixed sample size SAA confidence interval theory as being complementary to the theory that we present here. However, since SAA is a special case of the iterative algorithm with common random number sampling with a fixed sample size, where the number of iterations is large enough to solve the SAA problem to optimality, we can present a confidence theory for SAA as well. The latter is a special case of our general theory.

Author's personal copy

Our confidence intervals provide a useful indicator of the quality of the final solution at algorithmic termination, and were first proposed in a technical report by Dantzig and Infanger [6] that lacked a complete theoretical justification. The current paper provides the theory necessary to make these confidence bounds rigorous and covers both discrete and continuous distributions on Λ. One important feature of our bounds, in comparison to the existing literature, is that our bounds require only that the sample size used at each iteration be large enough to justify the normal approximation (via the central limit theorem). In particular, there is no requirement that the number of iterations be so large that the currently computed minimizer can be assumed to be close to the true underlying minimizer. Thus, our theory is non-asymptotic in the iteration count and can be reliably used even when the number of iterations executed is small or when one is unsure that the algorithm has iterated to a solution close to the true minimizer. This stands in contrast to the prevailing literature that takes advantage of optimality conditions that are enforced asymptotically at the minimizer of (1); see, for example, previous work of Higle and Sen [16], Shapiro and Homem-de-Mello [39], and Shapiro et al. [41]. A further advantage of our confidence bounds for both the SAA and iterative implementations considered here is that they can be computed without the need to run the algorithm multiple independent times, as is required by the bounds proposed in Mak et al. [30] and Bayraksan and Morton [2], thereby saving on computational effort. An overview of the methods, competing with ours, for assessing the solution quality of stochastic programs can be found in Bayraksan et al. [3].

The stochastic linear program with recourse (1) was first introduced by Dantzig [8] and Beale [4], independently, and has been studied extensively in the years since. Dantzig and Madansky [7] and Van Slyke and Wets [44] proposed applying Dantzig-Wolfe and Benders decomposition, respectively, to solve (1) by sampling. Consistency results and convergence properties for SAA are discussed in Dupačová and Wets [11], King and Wets [24], King and Rockafellar [25], Robinson [33], Shapiro and Homem-de-Mello [39], and Shapiro et al. [41]. The general view is that SAA is appealing because once an approximating problem has been obtained via sampling, any deterministic method can be used to solve it, including decomposition, regularized decomposition (Ruszczyński [35], Ruszczyński and Svetanowski [36]), progressive hedging (Rockafellar and Wets [34]), and use of interior point methods (Lustig et al. [29]). Higle and Sen [14] present a stochastic decomposition method, closely related to our iterative decomposition algorithm, that involves generating one independent subproblem at each iteration and re-using previous random outcomes for estimation of the so-called Benders' cuts, and attempts to detect optimality via use of the (asymptotically valid) Karush-Kuhn-Tucker conditions for nonlinear programs. It should be noted that because stochastic decomposition requires collection and re-use of dual vertices in each iteration, the dual feasible region must be the same for all random outcomes and consequently the method requires fixed recourse (W deterministic).

This paper is organized as follows. In Sect. 2, we carefully describe the iterative sampling-based decomposition algorithm, where the samples used at each iteration are independent of those used at previous iterations. We then prove the asymptotic validity (as the number of samples per iteration goes to infinity) of our upper and

Author's personal copy

lower confidence bounds on the minimum of z^* of (1) (to be differentiated from the function $z^*(.)$ corresponding to the minimum of the second-stage linear program). Section 3 then extends the theory to cover the case in which subproblem instances generated previously are re-used by later iterations. (When the sample size is constant from iteration to iteration, this is just SAA implemented via a decomposition solution algorithm.) A key ingredient in our confidence bounds is the variance of the (random) subproblem minimum evaluated at the minimizer x^* of (1). Section 4 therefore discusses various approaches to estimating or bounding this variance. Section 5 is concerned with extending our results to modified versions of our basic algorithm (to cover use of variance reduction and to allow the dropping of cuts). We conclude in Sect. 6 with an extensive empirical investigation of the accuracy of our confidence bounds in the context of a suite of test problems.

2 Confidence bounds for iterative sampling-based decomposition: independent sampling

We start by describing the iterative sampling-based decomposition algorithm that we shall study. A key feature is its reliance on the dual linear program (LP), denoted $L^D(x;\Lambda)$, given by

$$\begin{aligned} \max\ & \pi(h + Tx) \\ \text{s/t}\ & \pi W \leq q \end{aligned}$$

that is dual to the (primal) subproblem $L^P(x;\Lambda)$ associated with (1), namely

$$\begin{aligned} \min\ & qy \\ \text{s/t}\ & Wy = h + Tx \\ & y \geq 0. \end{aligned}$$

We assume throughout this paper that:

A1. For each $x \in H \equiv \{x :\ Ax = b,\ x \geq 0\}$, $L^P(x;\Lambda)$ and $L^D(x;\Lambda)$ are almost surely (a.s.) feasible LP's.

The assumption that for each $x \in H$, $L^P(x;\Lambda)$ is a.s. feasible is called *relatively complete recourse* in the literature. Hence, if the problem has relatively complete recourse and if $L^D(x;\Lambda)$ is a.s. feasible for a single $x \in H$, A1 is satisfied.

A2. The problem (1) has at least one minimizer, with corresponding minimum value z^*.

Let $y^*(x;\Lambda)$ be a minimizer of $L^P(x;\Lambda)$ and let $\pi^*(x;\Lambda)$ be a maximizer of $L^D(x;\Lambda)$ (both assumed to be measurable selections in Λ; see Kall [22] for details of the existence of such a measurable selection), with corresponding common optimal value $z^*(x;\Lambda)$. An important parameter in the specification of the algorithm is the number n of primal subproblems that are independently sampled at every iteration of the decomposition algorithm. We denote the master solution associated with performing m iterations, each based on n subproblem samples, by $X_{m,n}$.

Author's personal copy

Algorithm 1 1. Set $X_{0,n} = x_0 \in H$ as our initial guess at some minimizer x^* of (1).
2. $m \leftarrow 1$.
3. Generate n independent realizations $\Lambda_{m,1} = (q_{m,1}, T_{m,1}, W_{m,1}, h_{m,1}), \ldots, \Lambda_{m,n} = (q_{m,n}, T_{m,n}, W_{m,n}, h_{m,n})$ from the distribution of $\Lambda = (q, T, W, h)$.
4. Solve the n LP's $L^D(X_{m-1,n}; \Lambda_{m,1}), \ldots, L^D(X_{m-1,n}; \Lambda_{m,n})$ for the corresponding maximizers $\pi^*(X_{m-1,n}; \Lambda_{m,1}), \ldots, \pi^*(X_{m-1,n}; \Lambda_{m,n})$.
5. Compute the sample means

$$G_{m,n} = \frac{1}{n}\sum_{j=1}^{n} \pi^*(X_{m-1,n}; \Lambda_{m,j}) T_{m,j},$$

$$g_{m,n} = \frac{1}{n}\sum_{j=1}^{n} \pi^*(X_{m-1,n}; \Lambda_{m,j}) h_{m,j}.$$

6. Solve the LP (master problem)

$$\begin{aligned} \min\, & cx + \theta \\ \text{s/t}\ \ & Ax = b \\ & -G_{i,n}x + \theta \geq g_{i,n},\ 1 \leq i \leq m \\ & x \geq 0, \end{aligned}$$

for a minimizer $X_{m,n}$. (Note that this LP is computing a minimizer of $cx + \max_{1\leq i\leq m}[G_{i,n}x + g_{i,n}]$ over $x \in H$.)
7. $m \leftarrow m + 1$ and return to 3.

The above algorithm produces a sequence of iterative approximations $X_{1,n} \ldots, X_{m,n}$ to the solution of (1). The associated sequence of approximating optimal objective values is given by $((cX_{j,n} + \max_{1\leq i\leq j}[G_{i,n}X_{j,n} + g_{i,n}]) : j \geq 1)$. Note that the family of n independent realizations of Λ that we generate at the m-th iteration are different from those generated at previous iterations. Step 3 of the algorithm is then requiring that $(\Lambda_{i,j} : i \geq 1, j \geq 1)$ be a collection of independent and identically distributed (iid) realizations drawn from the distribution of Λ. Let $\mathcal{F}_{m,n} = \sigma(\Lambda_{i,j} : 1 \leq i \leq m, 1 \leq j \leq n)$ (with $\mathcal{F}_{0,n} = \{\emptyset, \Omega\}$) be the σ-algebra generated by the $\Lambda_{i,j}$'s that are used by the first m iterations of the algorithm when sample size n is used at each iteration.

Our first result is an upper bound that is widely used in the setting of sampling-based optimization algorithms; see Asmussen and Glynn [1]. Let $z(x) = cx + E\, z^*(x; \Lambda)$ and note that because x^* is a minimizer of (1), it follows that $z(X_{m,n}) \geq z(x^*) \equiv z^*$. Hence $z(X_{m,n})$ is an upper bound on the minimum z^* of (1). Because $E\, z^*(.; \Lambda)$ is not typically available in closed form (for otherwise one would not be applying sampling to solve (1)), $z(X_{m,n})$ must be estimated by sampling. Suppose that we generate ℓ iid realizations of the distribution of Λ, independently of the previous realizations used to calculate $X_{m,n}$. More precisely, suppose that we generate $\Lambda_{m+1,1}, \ldots, \Lambda_{m+1,\ell}$; these random variates are clearly independent of $\mathcal{F}_{m,n}$. Then,

Author's personal copy

$$cX_{m,n} + \frac{1}{\ell}\sum_{j=1}^{\ell} z^*(X_{m,n};\, \Lambda_{m+1,j}) \tag{2}$$

is an estimator for $z(X_{m,n})$. If we set $\sigma^2(x) \equiv \mathrm{var} z^*(x;\, \Lambda)$ and

$$s_\ell^2(x) \equiv \frac{1}{\ell - 1}\sum_{i=1}^{\ell}\left(z^*(x;\, \Lambda_{m+1,i}) - \frac{1}{\ell}\sum_{j=1}^{\ell} z^*(x;\, \Lambda_{m+1,j})\right)^2,$$

we note that $\ell^{-1}s_\ell^2(X_{m,n})$ is an unbiased estimator for the conditional variance $\ell^{-1}\sigma^2(X_{m,n})$ of the estimator (2) (conditional on $\mathcal{F}_{m,n}$). The following upper (conditional) confidence bound on z^* thus follows easily from the central limit theorem (CLT). This upper bound has appeared in the literature many times before; see, e.g., Infanger [20]. We include it here as a complement to our lower bound contribution. (Throughout this paper, we use $N(0, 1)$ to denote a normal random variable with mean 0 and unit variance.)

Proposition 1 *Assume A1 and A2. Suppose that $\sigma^2(x) < \infty$ for all $x \in H$. If $\tilde{\eta}$ is chosen so that $P(N(0, 1) > \tilde{\eta}) = \alpha$, then*

$$\lim_{\ell\to\infty} P\left(z^* \le cX_{m,n} + \frac{1}{\ell}\sum_{j=1}^{\ell} z^*(X_{m,n};\, \Lambda_{m+1,j}) + \tilde{\eta}\sqrt{\frac{s_\ell^2(X_{m,n})}{\ell}} \mid \mathcal{F}_{m,n}\right) \ge 1-\alpha \quad a.s.$$

(with equality when $\sigma^2(X_{m,n}) > 0$).

The more subtle and difficult issue is that of a lower confidence bound on z^*. Our next result is the main theoretical contribution of this paper (along with Theorem 2).

Theorem 1 *Assume A1 and A2. Let x^* be any minimizer of (1) for which $\sigma^2(x^*) < \infty$. Then,*

$$\liminf_{n\to\infty} P\left(z^* \ge cX_{m,n} + \max_{1\le i\le m}(G_{i,n}X_{m,n} + g_{i,n}) - \eta\sqrt{\frac{\sigma^2(x^*)}{n}}\right) \ge 1 - \alpha,$$

provided that η is selected so that $P(N(0, 1) \le \eta) = (1 - \alpha)^{1/m}$.

Proof Because $X_{m,n}$ is a minimizer of

$$\min_{x\in H}\left[cx + \max_{1\le i\le m}(G_{i,n}x + g_{i,n})\right],$$

it follows that

$$cX_{m,n} + \max_{1\le i\le m}(G_{i,n}X_{m,n} + g_{i,n}) \le cx^* + \max_{1\le i\le m}(G_{i,n}x^* + g_{i,n}).$$

Author's personal copy

Hence,

$$\begin{aligned} & P\left(cX_{m,n} + \max_{1\le i\le m}(G_{i,n}X_{m,n} + g_{i,n}) \le cx^* + z^*(x^*) + \frac{\eta\sigma(x^*)}{\sqrt{n}}\right) \\ & \ge P\left(\max_{1\le i\le m}(G_{i,n}x^* + g_{i,n}) \le z^*(x^*) + \frac{\eta\sigma(x^*)}{\sqrt{n}}\right) \\ & = P\left(\max_{1\le i\le m}(G_{i,n}x^* + g_{i,n}) - z^*(x^*)\right) \le \frac{\eta\sigma(x^*)}{\sqrt{n}}\Bigg). \end{aligned} \tag{3}$$

Recall that $\pi^*(x;\Lambda_{i,j})$ is a maximizer of $L^D(x;\Lambda_{i,j})$, so

$$\pi^*(X_{i-1,n};\Lambda_{i,j})(h_{i,j} + T_{i,j}x) \le \pi^*(x;\Lambda_{i,j})(h_{i,j} + T_{i,j}x)$$

because $\pi^*(X_{i-1,n};\Lambda_{i,j})$ is feasible for $L^D(x;\Lambda_{i,j})$. By duality,

$$\pi^*(x;\Lambda_{i,j})(h_{i,j} + T_{i,j}x) = q_{i,j}y^*(x;\Lambda_{i,j}).$$

Hence,

$$\pi^*(X_{i-1,n};\Lambda_{i,j})(h_{i,j} + T_{i,j}x) \le q_{i,j}y^*(x;\Lambda_{i,j}) \text{ a.s.}$$

We conclude that

$$G_{i,n}x + g_{i,n} \le \frac{1}{n}\sum_{j=1}^{n} z^*(x;\Lambda_{i,j}) \text{ a.s.}$$

for $x \in H$, so that

$$G_{i,n}x^* + g_{i,n} - z^*(x^*) \le \frac{1}{n}\sum_{j=1}^{n} z^*(x^*;\Lambda_{i,j}) - z^*(x^*) \text{ a.s.}$$

The probability (3) may therefore be lower bounded by

$$P\left(\max_{1\le i\le m} n^{1/2}\left(\frac{1}{n}\sum_{j=1}^{n} z^*(x^*;\Lambda_{i,j}) - z^*(x^*)\right) \le \sigma(x^*)\eta\right). \tag{4}$$

Note that the above probability is 1 if $\sigma^2(x^*) = 0$. On the other hand, if $\sigma^2(x^*) > 0$, then $\{z^*(x^*;\Lambda_{i,j}) - z^*(x^*)) : 1 \le i \le m, 1 \le j \le n\}$ is a collection of iid mean zero random variables with finite variance $\sigma^2(x^*)$. Hence,

$$\left(\frac{n^{1/2}}{\sigma(x^*)}\left(\frac{1}{n}\sum_{j=1}^{n} z^*(x^*;\Lambda_{1,j}) - z^*(x^*)\right), \ldots,\right.$$

Author's personal copy

$$\left(\frac{n^{1/2}}{\sigma(x^*)}\left(\frac{1}{n}\sum_{j=1}^{n} z^*(x^*;\Lambda_{m,j}) - z^*(x^*)\right)\right.$$
$$\left.\Rightarrow (N_1(0,1),\ldots,N_m(0,1))\right.$$

as $n \to \infty$, where $N_1(0,1),\ldots,N_m(0,1)$ are m iid normal random variables with mean zero and unit variance. In view of (4), it follows that when $\sigma^2(x^*) > 0$,

$$\liminf_{n\to\infty} P\left(cX_{m,n} + \max_{1\le i\le m}(G_{i,n}X_{m,n} + g_{i,n}) \le cx^* + z^*(x^*) + \frac{\eta}{\sqrt{n}}\sigma(x^*)\right)$$
$$\ge P\left(\max_{1\le i\le m} N_i(0,1) \le \eta\right) = P(N(0,1)\le\eta)^m = 1-\alpha,$$

proving the theorem. □

Theorem 1 provides the key lower bound. Note that the lower bound involves one quantity, namely $\sigma^2(x^*)$, that is not computed during the course of our algorithm. Theorem 1 effectively reduces the problem of obtaining a lower confidence bound on z^* to estimating $\sigma^2(x^*)$ or (at least) to estimating (or computing) an upper bound on $\sigma^2(x^*)$. We discuss this question in greater detail in Sect. 4. It should be further noted that the quantile η satisfying $P(N(0,1)\le\eta) = (1-\alpha)^{1/m}$ increases rather slowly with m, specifically η is of the order of $\sqrt{2\log m}$ for m large; see Resnick [32]. As a consequence, one pays only a modest penalty when applying this confidence bound for m large.

In using an iterative sampling-based algorithm, it is natural to consider a modified version in which the sample size n_i describing the number of independent subproblems sampled at iteration i is permitted to depend on i. For example, as the number of iterations increases, we may wish to increase the sample size in order to improve the accuracy as the iterative scheme closes in on the optimizer. Let the sequence $(n_i : i \ge 1)$ be a (deterministic) sequence of integers corresponding to the successive sample sizes used at the different iterations. In this setting, the cut generated at iteration i has an associated sample size n_i, so we write it as (G_{i,n_i}, g_{i,n_i}). Assume that for $i \ge 1$, $n_i = c_i n(1+o(1))$, as $n\to\infty$, where $o(1)$ is a sequence that tends to zero as $n\to\infty$ and $c_i > 0$ for $i \ge 1$. Then, a slight adaptation of the proof in Theorem 1 shows that under the conditions stated there,

$$\liminf_{n\to\infty} P\left(z^* \ge cX_{m,n_m} + \max_{1\le i\le m}(G_{i,n_i}X_{m,n_m} + g_{i,n_i}) - \eta\sqrt{\frac{\sigma^2(x^*)}{n_1}}\right) \ge 1-\alpha,$$

where η is chosen so that $\prod_{i=1}^m P(N(0,1)\le\sqrt{\frac{c_i}{c_1}}\eta) = (1-\alpha)$. (Note that we have arbitarily chosen to express our confidence bound in terms n_1; we make no assumption here about the relative ordering of the c_i's. In particular, n_1 need not be the minimal sample size.) Such a value for η can easily be computed numerically, at least when m is not too large.

Author's personal copy

We conclude this section by noting that Proposition 1 provides an upper bound on z^* that involves a confidence statement that is conditional on $\mathcal{F}_{m,n}$, whereas Theorem 1 offers a lower bound on z^* relative to the unconditional distribution. A typical user of such confidence bounds prefers that either both statements be conditional or that both be unconditional. The argument underlying Theorem 1 fails badly, when one conditions on $\mathcal{F}_{m,n}$ (because, in that case, the $\Lambda_{i,j}$'s with $i \leq m$ and $j \leq n$ are no longer iid). As a consequence, we now offer a version of the upper bound (Proposition 1) that is unconditional.

Proposition 2 *In the presence of A1,*

$$\lim_{\ell \to \infty} P\left(z^* \leq cX_{m,n} + \frac{1}{\ell} \sum_{j=1}^{\ell} z^*(X_{m,n}; \Lambda_{m+1,j}) + \tilde{\eta} \sqrt{\frac{s_\ell^2(X_{m,n})}{\ell}} \right) \geq 1 - \alpha.$$

Proof According to Proposition 1, the limit of the conditional probability is at least equal to $1 - \alpha$. The result then follows from the bounded convergence theorem. □

3 Confidence bounds for iterative sampling-based decomposition: common random numbers

We now discuss a variant of Algorithm 1, in which we re-use subproblem instances generated at previous iterations in the current iteration. In the language of simulation, this is a use of *common random numbers*, as common samples are used across all iterations. Because the structure of the limit distribution (and associated confidence interval) differs substantially when one permits the sample size to be iteration-dependent (as opposed to using a constant sample size across all iterations), we will focus our discussion from the outset on the more general case of iteration-dependent sample sizes. As in Sect. 2, let n_i be the sample size associated with iteration i and assume that $n_i = c_i n(1 + o(1))$, as $n \to \infty$, where $c_i > 0$ for $i \geq 1$.

The following algorithm is a variant of Algorithm 1 that arises when one re-uses subproblem instances generated at previous iterations.

Algorithm 2
1. Set $X_0 = x_0 \in H$ as our initial guess at some minimizer x^* of (1).
2. $m \leftarrow 1, n_{\max} \leftarrow n_1$.
3. Generate n_1 independent realizations $\Lambda_1 = (q_1, T_1, W_1, h_1), \ldots, \Lambda_{n_1} = (q_{n_1}, T_{n_1}, W_{n_1}, h_{n_1})$ from the distribution of $\Lambda = (q, T, W, h)$.
4. Solve the n_m LP's $L^D(X_{m-1}; \Lambda_1), \ldots, L^D(X_{m-1}; \Lambda_{n_m})$ for the corresponding maximizers $\pi^*(X_{m-1}; \Lambda_1), \ldots, \pi^*(X_{m-1}; \Lambda_{n_m})$.
5. Compute the sample means

$$G_{m,n_m} = \frac{1}{n_m} \sum_{j=1}^{n_m} \pi^*(X_{m-1}; \Lambda_j) T_j,$$

$$g_{m,n_m} = \frac{1}{n_m} \sum_{j=1}^{n_m} \pi^*(X_{m-1}; \Lambda_j) h_j.$$

Author's personal copy

6. Solve the LP (master problem)

$$\begin{aligned} \min cx \quad &+ \theta \\ Ax \quad &= b \\ -G_{i,n_i}x \quad &+ \theta \ \geq g_{i,n_i}, \ 1 \leq i \leq m \\ x \quad &\geq 0, \end{aligned}$$

for a minimizer X_m.

7. $m \leftarrow m+1$
8. If $n_{\max} < n_m$, generate $n_m - n_{\max}$ additional independent realizations $\Lambda_{n_{\max}+1} = (q_{n_{\max}+1}, T_{n_{\max}+1}, W_{n_{\max}+1}, h_{n_{\max}+1}), \ldots, \Lambda_{n_m} = (q_{n_m}, T_{n_m}, W_{n_m}, h_{n_m})$ from the distribution of $\Lambda = (q, T, W, h)$.
9. $n_{\max} \leftarrow \max(n_{\max}, n_m)$ and return to 4.

This algorithm generates a sequence of approximations $X_1, X_2, \ldots$ to the minimizer of (1) that are constructed from a common sequence $\Lambda_n, n \geq 1$ of subproblem realizations that are used across all iterations. As in the independent setting of Sect. 2, the same upper bound confidence interval procedure can be used here, provided that (as in Sect. 2) the subproblem instances used to compute the upper bound are generated independently of the samples $\Lambda_1, \Lambda_2, \ldots$ used to calculate X_m.

As for the lower bound, the analog to Theorem 1 is presented next. (Recall that because the sample sizes determining the X_i's depend on the parameter n, the X_i's depend (implicitly) on n). To state the result, let $B = (B(t) : t \geq 0)$ be a standard Brownian motion with $B(0) = 0$.

Theorem 2 *Assume A1 and A2. Let x^* be any minimizer of (1) for which $\sigma^2(x^*) < \infty$. Then,*

$$\liminf_{n\to\infty} P\left(z^* \geq cX_m + \max_{1\leq i\leq m}(G_{i,n_i}X_m + g_{i,n_i}) - \eta'\sqrt{\frac{\sigma^2(x^*)}{n}}\right) \geq 1-\alpha, \quad (5)$$

where η' is selected so that $P\left(\max_{1\leq i\leq m}\frac{B(c_i)}{c_i} \leq \eta'\right) = 1-\alpha$.

Proof If one follows the same proof as in Theorem 1, the key is that (4) now becomes

$$P\left(\max_{1\leq i\leq m} n^{1/2}\left(\frac{1}{n_i}\sum_{j=1}^{n_i} z^*(x^*;\Lambda_j) - z^*(x^*)\right) \leq \sigma(x^*)\eta'\right). \quad (6)$$

But Donsker's theorem (see, for example, Billingsley [5, p. 68]) implies that

$$\begin{aligned} &n^{1/2}\left(\frac{1}{n_1}\sum_{j=1}^{n_1} z^*(x^*;\Lambda_j) - z^*(x^*), \ldots, \frac{1}{n_m}\sum_{j=1}^{n_m} z^*(x^*;\Lambda_j) - z^*(x^*)\right) \\ &\quad \Rightarrow \sigma(x^*)\left(\frac{B(c_1)}{c_1}, \ldots, \frac{B(c_m)}{c_m}\right) \end{aligned}$$

Author's personal copy

as $n \to \infty$. The continuous mapping principle then ensures that (6) converges to

$$P\left(\max_{1\le i\le m} \frac{B(c_i)}{c_i} \le \eta'\right),$$

which is $1-\alpha$. □

We note that the quantity η' will typically need to be computed numerically (for example, via a Monte Carlo root finding algorithm such as the Robbins-Monro method; see, for example, Kushner and Yin [26, p. 2]). An alternative, available when $c_i \ge 1$ for $i \ge 1$ (as would be typical), is to upper bound η' by η'', where η'' is defined as the root of $P(\sup_{t\ge 1} B(t)/t \le \eta'') = 1-\alpha$. To evaluate this latter distribution, recall that the process $(sB(1/s) : s > 0) \overset{\mathcal{D}}{=} (B(s) : s > 0)$ (where $\overset{\mathcal{D}}{=}$ denotes *equality in distribution*); see Karlin and Taylor [23, p. 351]. As a consequence,

$$\begin{aligned}\sup_{t\ge 1} B(t)/t &= \sup\{sB(1/s) : 0 < s \le 1\}\\ &\overset{\mathcal{D}}{=} \sup\{B(s) : 0 < s \le 1\}\\ &= \sup\{B(s) : 0 \le s \le 1\} \text{ a.s.}\end{aligned}$$

so $P(\sup_{t\ge 1} B(t)/t > z) = P(\sup_{0<s\le 1} B(s) > z) = 2P(N(0,1) > z)$ (by the reflection principle; Karlin and Taylor [23, p. 346]), so η'' satisfies $P(N(0,1) \le \eta'') = 1-\alpha/2$.

An important special case arises when one uses a common set of subproblem instances across all iterations, so that $n_1 = n_2 = \cdots = n_m = n$. This is effectively identical to using m iterations to solve a sample average approximation to (1), using sample size n. In this case, $\max_{1\le i\le m} B(c_i)/c_i = B(1)$, independent of m. Because m can be made arbitrarily large, it follows that X_∞ (i.e., the solution of the n-sample SAA solved to optimality) then satisfies the following lower (confidence) bound

$$\liminf_{n\to\infty} P\left(z^* \ge \left[cX_\infty + \frac{1}{n}\sum_{j=1}^{n} z^*(X_\infty;\Lambda_j)\right] - \eta'\sqrt{\frac{\sigma^2(x^*)}{n}}\right) \ge 1-\alpha, \quad (7)$$

where η' satisfies $P(B(1) \le \eta') = P(N(0,1) \le \eta') = 1-\alpha$. In view of the optimality bias of X_∞, (5) has, in this setting, a particularly natural form. A noteworthy feature of this confidence bound for SAA is that it can be computed internally from a single SAA optimization, and does not require running several independent replications of the SAA procedure in order to estimate the sampling error for the SAA estimator. The special case (7) (in which the n_i's are identical and $m = \infty$) was previously derived by Bayraksan and Morton [2]).

One common feature of the lower bounds presented here, as in Sect. 2, is that they require a knowledge of $\sigma^2(x^*)$ (i.e., the variance of $z^*(x^*;\Lambda)$ at a minimizer of (1)); this issue is addressed in the next section.

Author's personal copy

4 Bounding the variance $\sigma^2(x^*)$

One obvious means of estimating $\sigma^2(x^*)$ is to assume that the computed solution $X_{m,n}$ (in the setting of Algorithm 1) or X_m (in the setting of Algorithm 2) is close to x^*, so that $\sigma^2(X_{m,n})$ (or $\sigma^2(X_m)$) can be used to estimate $\sigma^2(x^*)$ in the confidence bound of Theorem 1 (or Theorem 2). There are many ways to formulate a rigorous result justifying such an estimator for $\sigma^2(x^*)$; the following is one (given in the context of Sect. 2).

Proposition 3 *Assume A1 - A2, $n_i = \lfloor c_i n \rfloor$, for $i \geq 1$ and $\inf\{c_i : i \geq 1\} > 0$. Suppose that $m = m_n \to \infty$ and $\ell = \ell_n \to \infty$ as $n \to \infty$ in such a way that $m_n/n^{1/2} \to 0$ as $n \to \infty$ and $X_{m,n} \to x^*$ as $n \to \infty$, where x^* is a minimizer of (1). If there exists ϵ for which $\sup\{|z^*(x;\Lambda)| : |x - x^*| \leq \epsilon\}$ is a bounded random variable then,*

$$\liminf_{n\to\infty} P\left(z^* \geq cX_{m,n} + \max_{1\leq i\leq m}(G_{i,n_i}X_{m,n} + g_{i,n_i}) - \eta\sqrt{\frac{s_\ell^2(X_{m,n})}{n_1}}\right) \geq 1-\alpha,$$

where $\eta = \eta_n$ is selected so that $\prod_{i=1}^m P(N(0,1) \leq \sqrt{\frac{c_i}{c_1}}\eta) = 1-\alpha$.

Proof We first establish that $s_\ell^2(X_{m,n})$ converges in probability to $\sigma^2(x^*)$ as $n \to \infty$. Under A1, it is known that $z^*(.;\Lambda)$ is continuous, see, e.g., Wets [45]. Hence, the bounded convergence theorem guarantees that $E\ z^*(.;\Lambda)^k$ is continuous in an ϵ-neighborhood of x^* for $k = 1, 2$, thereby ensuring that $\sigma^2(.)$ is a continuous function in the same ϵ-neighborhood of x^*. Write $s_\ell^2(X_{m,n}) - \sigma^2(x^*) = (s_\ell^2(X_{m,n}) - \sigma^2(X_{m,n}) + (\sigma^2(X_{m,n}) - \sigma^2(x^*))$, and note that $\sigma^2(X_{m,n}) \Rightarrow \sigma^2(x^*)$ as $n \to \infty$. On the other hand, Chebyshev's inequality (applied conditionally) implies that

$$\begin{aligned}
&P(|s_\ell^2(X_{m,n}) - \sigma^2(X_{m,n})| > \epsilon) \\
&\leq P(|s_\ell^2(X_{m,n}) - \sigma^2(X_{m,n})| > \epsilon, |X_{m,n} - x^*| \leq \epsilon) + P(|X_{m,n} - x^*| > \epsilon) \\
&\leq E\ \frac{\mathrm{var}(s_\ell^2(X_{m,n})|X_{m,n})}{\epsilon^2} I(|X_{m,n} - x^*| \leq \epsilon) + P(|X_{m,n} - x^*| > \epsilon).
\end{aligned}$$

But $\sup\{|z^*(x;\Lambda)| : |x - x^*| \leq \epsilon\}$ is a bounded random variable, so $\mathrm{var}(s_\ell^2(x)) \to 0$ as $\ell \to \infty$, uniformly over $|x - x^*| \leq \epsilon$. Hence, it follows that $s_\ell^2(X_{m,n}) - \sigma^2(X_{m,n})$ converges to zero in probability as $n \to \infty$, proving that $s_\ell^2(X_{m,n}) \Rightarrow \sigma^2(x^*)$ as $n \to \infty$.

The proposition is then an easy consequence of proving that

$$\begin{aligned}
\sup_w |P&\left(\max_{1\leq i\leq m} n^{1/2}\left(\frac{1}{n_i}\sum_{j=1}^{n_i} z^*(x^*;\Lambda_{i,j}) - z^*(x^*)\right) \leq w\right) \\
&- P\left(\max_{1\leq i\leq m} \sigma(x^*)\frac{1}{\sqrt{c_i}}N_i(0,1) \leq w\right)| \to 0
\end{aligned} \tag{8}$$

Author's personal copy

as $n \to \infty$. But the Berry–Esseen theorem (Feller [12, p. 542]) implies that the left-hand side of (8) is bounded by

$$\begin{aligned}
&\sup_w \left| \prod_{i=1}^m P\left(n^{1/2} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} z^*(x^*; \Lambda_{i,j}) - z^*(x^*) \right) \le w \right) \right. \\
&\quad \left. - \prod_{i=1}^m P\left(\sigma(x^*) \frac{N_i(0,1)}{\sqrt{c_i}} \le w \right) \right| \\
&\quad \le \sum_{i=1}^m \sup_w \left| P\left(n_i^{1/2} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} z^*(x^*; \Lambda_{i,j}) - z^*(x^*) \right) \le \sqrt{c_i} w \right) \right. \\
&\quad \left. - P(\sigma(x^*) N_i(0,1) \le \sqrt{c_i} w) \right| \\
&\quad \le \frac{3E\left| z^*\left(x^*; \Lambda_{i,j}\right) - z^*(x^*) \right|^3}{\sigma^3(x^*)} \sum_{i=1}^m \frac{1}{\sqrt{n_i}} \to 0
\end{aligned}$$

as $n \to \infty$, since $m/n^{1/2} \to 0$ and $\inf\{c_i : i \ge 1\} > 0$. □

However, as mentioned in the Introduction, our theory is intended to be non-asymptotic in the iteration count m. As a consequence, we do not wish to necessarily rely on the presumption that $X_{m,n}$ (or X_m) is close to x^*. In settings in which one wishes to avoid this asymptotic simplification, one then seeks upper bounds on $\sigma^2(x^*)$. While not the central focus of this paper, we suggest here three different ways of bounding $\sigma^2(x^*)$.

Method 1 We bound $\sigma^2(x^*)$ by $\sigma_*^2 \equiv \sup\{\sigma^2(x) : x \in H\}$ and estimate σ_*^2. To estimate σ_*^2, we sample r points $\chi_1, \ldots, \chi_r$ at random from within H, compute $s_\ell^2(\chi_i)$ at each of the r points and bound σ_*^2 via the sample-based quantity

$$\hat{\sigma}_*^2(r, \ell) = \max_{1 \le i \le r} s_\ell^2(\chi_i) \tag{9}$$

that uses a common set of Λ instances across all r points. An alternative is to generate independent Λ realizations of the second-stage subproblems at each of the r points, thereby calculating

$$\tilde{s}_\ell^2(\chi_i) = \frac{1}{\ell - 1} \sum_{j=1}^{\ell} \left(z^*(\chi_i; \Lambda_{m+i,j}) - \frac{1}{\ell} \sum_{k=1}^{\ell} z^*(\chi_i; \Lambda_{m+i,k}) \right)^2$$

for $1 \le i \le r$, and then compute

$$\tilde{\sigma}_*^2(r, \ell) = \max_{1 \le i \le r} \tilde{s}_\ell^2(\chi_i). \tag{10}$$

Author's personal copy

Note that because $s_\ell^2(\chi_i) \leq \max_{1\leq k\leq r} s_\ell^2(\chi_k)$, it follows that

$$\sigma^2(\chi_i) \leq E\,[\hat{\sigma}_*^2(r,\ell)|\chi_1,\ldots,\chi_r]$$

for $1 \leq i \leq r$. Furthermore if $\sigma(.)$ is continuous over H and if the sequence $\chi_1, \chi_2, \ldots$ is a.s. asymptotically dense in H, then $\max_{1\leq i\leq r} \sigma^2(\chi_i) \to \sigma_*^2$ a.s. as $r \to \infty$. It follows, under these conditions, that $\hat{\sigma}_*^2(r,\ell)$ is biased high for σ_*^2. In addition, $P(\hat{\sigma}_*^2(r,\ell) \leq \sigma_*^2 - \epsilon) \to 0$ as $r \to \infty$ for each $\epsilon > 0$, so that

$$\liminf_{n\to\infty, r\to\infty} P\left(z^* \geq cX_{m,n} + \max_{1\leq i\leq m}(G_{i,n}X_{m,n} + g_{i,n}) - \eta\sqrt{\frac{\hat{\sigma}_*^2(r,\ell)}{n}}\right) \geq 1-\alpha, \quad (11)$$

provided that η is selected so that $P(N(0,1) \leq \eta) = (1-\alpha)^{1/m}$. An identical conclusion follows for $\tilde{\sigma}_*^2(r,\ell)$. An implementation of this method requires specifying r, ℓ, and the algorithm used to generate the χ's.

Method 2 The second method uses the fact that $\sigma_*^2 \leq \sup\{E\ z^*(x;\Lambda)^2 : x \in H\}$. But

$$E\ z^*(x;\Lambda)^2 \leq E\ \sup_{x\in H} z^*(x;\Lambda)^2 \equiv E\ w(\Lambda)^2.$$

Assuming that $z^*(x;\Lambda)$ is non-negative a.s., the random variable $w(\Lambda)$ can be defined as the optimal objective of the (random) optimization problem

$$\begin{aligned} \max_{\pi,x}\ & \pi(h+Tx) \\ \text{s/t}\ & \pi W \leq q \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

It follows that

$$\frac{1}{\ell}\sum_{i=1}^{\ell} w(\Lambda_{m+1,i})^2 \to E\ w(\Lambda)^2 \geq \sigma_*^2 \ \text{a.s.}$$

as $\ell \to \infty$, so that $\ell^{-1}\sum_{i=1}^{\ell} w(\Lambda_{m+1,i})^2$ is an estimator of an upper bound on $\sigma^2(x^*)$, which can be used to implement a lower bound on z^*. However, the estimator requires solving a bilinear program ℓ times, which can be computationally challenging.

Method 3 Method 3 works with the variance $\sigma^2(x)$ directly, rather than the second moment bound $E\ z(x;\Lambda)^2$. We start by observing that

$$\sigma^2(x) = \frac{1}{2}E\ (z^*(x;\Lambda_{m+1,1}) - z^*(x;\Lambda_{m+1,2}))^2.$$

But $z^*(x;\Lambda_{m+1,i})$ is a maximizer of $L^D(x;\Lambda_{m+1,i}))$.

Author's personal copy

Assume that $W_{m+1,i} = D$ is deterministic. The feasible region of both dual linear programs are then contained in $F = \{\pi : \pi D \leq (q_{m+1,1} \vee q_{m+1,2})\}$, where $(q_{m+1,1} \vee q_{m+1,2})$ is the component-wise maximum of $q_{m+1,1}$ and $q_{m+1,2}$. Note that

$$\begin{aligned} |z^*(x; \Lambda_{m+1,1}) - z^*(x; \Lambda_{m+1,2})| &\leq \sup\{|\pi(h_{m+1,1} + T_{m+1,1}x) \\ &\quad -\pi(h_{m+1,2} + T_{m+1,2}x)| : \pi \in F\} \\ &= \sup\{|\pi(h_{m+1,1} - h_{m+1,2}) + \pi(T_{m+1,1} - T_{m+1,2})x)| : \pi \in F\}. \end{aligned}$$

So,

$$\begin{aligned} \sigma^2(x^*) \leq \frac{1}{2} E \sup\{(\pi(h_{m+1,1} - h_{m+1,2}) + \pi(T_{m+1,1} - T_{m+1,2})x)^2 : \\ \pi \in F, x \in H\} \equiv E\,\Psi, \end{aligned}$$

where Ψ is the (random) maximum of a quartic objective over a linear feasible set (with the feasible set having a "size" equal to that of a master plus a single subproblem). Let $\Psi_1, \ldots, \Psi_\ell$ be ℓ iid replicates of Ψ. Then,

$$\frac{1}{\ell} \sum_{i=1}^{\ell} |psi_i \to E\,\Psi \geq \sigma^2(x^*) \text{ a.s.},$$

yielding an upper bound on $\sigma^2(x^*)$. This method avoids sampling over H, as does Method 2 (but is again non-trivial to solve computationally).

5 Extensions

We present here several extensions to our key lower bound results, namely Theorems 1 and 2.

Extension 1: Confidence bounds that are non-asymptotic in the sample size n

Given our interest in this paper in lower bounds that are non-asymptotic in the iteration count m, it is worth noting that it is straightforward to also derive confidence bounds that are simultaneously non-asymptotic in the sample sizes used (modulo the knowledge of $\sigma^2(x^*)$). For example, in the setting of Algorithm 1 (with different sample sizes $n_1, n_2, \ldots$ with $n_i = \lceil c_i n \rceil$ for $i \geq 1$), Chebyshev's inequality implies that (4) can be bounded via

$$\begin{aligned} &\prod_{i=1}^{m} \left(1 - P(n_1^{1/2} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} z^*(x^*; \Lambda_{i,j}) - z^*(x^*)\right) > \sigma(x^*)\eta\right)\right) \\ &\quad \geq \prod_{i=1}^{m} \left(1 - \frac{c_1}{c_i} \frac{1}{\eta^2}\right). \end{aligned}$$

Author's personal copy

Thus, we are guaranteed that for every finite n and m,

$$P\left(z^* \geq cX_{m,n} + \max_{1\leq i\leq m}\left(G_{i,n}X_{m,n} + g_{i,n}\right) - \eta\sqrt{\frac{\sigma^2(x^*)}{n}}\right) \geq \prod_{i=1}^{m}\left(1 - \frac{c_1}{c_i}\frac{1}{\eta^2}\right).$$

Hence, if we choose η large enough that $\prod_{i=1}^{m}(1 - \frac{1}{c_i\eta^2}) > 1 - \alpha$, we obtain the required (non-asymptotic) lower bound. A similar use of Chebyshev's inequality can be implemented in the setting of Algorithm 2.

Of course, these error bounds involve the (unknown) quantity $\sigma^2(x^*)$. If a deterministic upper bound on $\sigma^2(x^*)$ is available (and used in place of $\sigma^2(x^*)$), one then obtains a potentially implementable confidence bound that is non-asymptotic in m and n. On the other hand, if sampling is used to upper bound $\sigma^2(x^*)$, one needs a (non-asymptotic) bound on the rate of convergence of $\hat{\sigma}_*^2(r, l)$ to σ_*^2; we leave this for future research.

Extension 2: Importance sampling

Suppose that we use a sample distribution $\tilde{P}$ (rather than the original P) from which to sample Λ, so as to concentrate more realizations of Λ in those regions that are most critical to determining the (expected) second-stage cost. We assume that $\tilde{P}$ satisfies

$$P(\Lambda \in dx) = l(x)\tilde{P}(\Lambda \in dx),$$

where P is the nominal distribution at Λ associated with the original probability distribution. The use of $\tilde{P}$ rather than P to generate outcomes of Λ is known as importance sampling. When Λ has densities p and $\tilde{p}$ under P and $\tilde{P}$, respectively, $l(.)$ can be expressed as $l(x) = p(x)/\tilde{p}(x)$ and is known as the likelihood ratio of P relative to $\tilde{P}$.

Because of the presence of importance sampling, we modify the i'th cut $(G_{i,n}, g_{i,n})$ to

$$\tilde{G}_{i,n} = \frac{1}{n}\sum_{j=1}^{n}\pi^*(X_{i-1,n}; \Lambda_{i,j})T_{i,j}l(\Lambda_{i,j}),$$

$$\tilde{g}_{i,n} = \frac{1}{n}\sum_{j=1}^{n}\pi^*(X_{i-1,n}; \Lambda_{i,j})h_{i,j}l(\Lambda_{i,j}).$$

and execute the algorithm with these modified cuts. The proof of Theorem 1 goes through with minor changes, namely $z^*(x; \Lambda_{i,j})$ must be replaced by $z^*(x; \Lambda_{i,j})l(\Lambda_{i,j})$, so that $\sigma^2(x^*)$ is replaced by $\tilde{\text{var}}z^*(x; \Lambda)l(\Lambda)$ (where $\tilde{\text{var}}(.)$ denotes the variance operator associated with distribution $\tilde{P}$); a similar adaptation can be made for Theorem 2.

When $\tilde{P}$ is chosen well, use of the sampling distribution $\tilde{P}$ rather than P can produce significant variance reductions, particularly in settings where there are second-stage

Author's personal copy

outcomes that occur rarely but have a big effect in the decisions to be made; see pp. 127–129 of Asmussen and Glynn [1] for further detail.

Extension 3: Control variates

In virtually all real-world applications of sampling-based methods for solving (1), a wide variety of expectations of the form $Ef(\Lambda)$ (for $f\mathcal{R}^{\ell}$-valued) can be computed either analytically in closed form or numerically to high precision (e.g., the mean of the right-hand side variables h and T). Let $C = f(\Lambda) - Ef(\Lambda)$ and note that $EC = 0$. Such a random variable C is called a control variate and, when intelligently used, can be used to induce a variance reduction in our sampling-based procedure. The idea is to replace the cut estimators $(G_{i,n}, g_{i,n})$ generated at iteration i by

$$G^c_{i,n} = \frac{1}{n}\sum_{j=1}^{n}(\pi^*(X_{i-1,n}; \Lambda_{i,j})T_{i,j} - \lambda_1 C_{i,j}),$$

$$g^c_{i,n} = \frac{1}{n}\sum_{j=1}^{n}(\pi^*(X_{i-1,n}; \Lambda_{i,j})h_{i,j} - \lambda_2 C_{i,j}).$$

where $C_{i,j} = f(\Lambda_{i,j}) - Ef(\Lambda)$ and $\lambda_1,\ \lambda_2$ are (deterministic) matrices (of row and column dimensions appropriate to make sense of the above expressions). One then executes the algorithm with these modified cuts.

As for importance sampling, the proof of Theorem 1 again goes through with minor changes. The key observation is to note that $z^*(x; \Lambda_{i,j})$ is replaced by

$$z^*(x, \Lambda_{i,j}) - \lambda_1 C_{i,j}x - \lambda_2 C_{i,j}$$

and $\sigma^2(x^*)$ is replaced by

$$\mathrm{var}(z^*(x^*, \Lambda_{i,j}) - \lambda_1 C_{i,j}x^* - \lambda_2 C_{i,j}).$$

When λ_1 and λ_2 are chosen so that $\mathrm{var}(z^*(x^*, \Lambda_{i,j}) - \lambda_1 C_{i,j}x^* - \lambda_2 C_{i,j})$ is smaller that $\sigma^2(x^*)$, the lower bound is tightened (in the sense that the length of the associated confidence interval for z^* is reduced). Note that a variance-minimizing choice of λ_1 and λ_2 (so as to minimize $\mathrm{var}(z^*(x^*, \Lambda_{i,j}) - \lambda_1 C_{i,j}x^* - \lambda_2 C_{i,j})$) requires knowledge of x^*, complicating the statistical estimation of the optimal control coefficient matrices λ_1^* and λ_2^* (in comparison with the estimation of the optimal control coefficients in the conventional control variate setting). Of course, if we approximate x^* by $X_{m,n}$ (as in estimating $\sigma^2(x^*)$ via $s^2_{\ell}(X_{m,n})$ in Sect. 4), the problem becomes much simpler and can be implemented via standard control variate regression-based methods; see Assmussen and Glynn [1, p. 138] . The challenge arises here when one is unwilling to make this simplifying approximation. This represents an interesting opportunity for future research.

Extension 4: Dropping cuts

In large-scale problems, it is often the case that the number of iterations required to accurately solve (1) is large. Our iterative algorithm can be slowed significantly

Author's personal copy

when one insists on keeping all m cuts. Instead, one may prefer on iteration m to (deterministically) retain the last j_m cuts (with $1 \leq j_m \leq m$), so that step 6 of our algorithm is implemented by solving

$$\begin{aligned} \min\, & cx + \theta \\ & Ax = b \\ -G_{i,n}x + \theta \; & \geq g_{i,n}, \; m - j_m + 1 \leq i \leq m \\ & x \geq 0, \end{aligned}$$

for a minimizer $X_{m,n}$. Our proof goes through without change to this setting. Note, however, that if the retained cuts are not deterministically specified (as would be the case if we were to choose to retain the "best" j_m cuts), our proof fails to be valid and a more sophisticated analysis would be required.

6 Numerical results

This paper provides confidence bound theory for a very rich class class of algorithms, given the number of methods proposed for upper bounding $\sigma^2(x^*)$ and the freedom of choice in determining the sample size as a function of iteration. As a consequence, we can not test all possible variants of what our theory covers. Instead, we limit ourselves to a computational investigation of our confidence interval methods in the simplest settings, namely that of the fixed sample size implementation associated with Algorithm 1 and that of the sample average approximation implementation when solving to optimality with a fixed sample size, as in the context of Algorithm 2. The results that we shall discuss in greater detail below suggest that our confidence intervals do indeed cover the minimum z^* to (1), precisely as our theory suggests should be the case. In addition, our experience with the test suite of problems below suggests that our proposed lower and upper confidence bounds generally are practically useful, in the sense that the spread of the bounds relative to z^* is sufficiently small in practice so as to provide useful information on the quality of the estimated solution. Such lower and upper confidence bounds on z^* serve to provide good estimates of the optimal decision. In particular, by generating an upper bound at the computed optimal solution (i.e., $X_{m,n}$) associated with the lower bound, we can determine how small is the "loss of performance" (as compared to use of x^*) due to use of $X_{m,n}$ as the decision.

We tested the confidence bounds methods described earlier in this paper on the following test problems: APL1P (Infanger [20]) and LandS (Louveaux and Smeers [28]) are electric power planning problems; GBD is an airline scheduling problem discussed in Dantzig [9]; STORM (Mulvey and Ruszczyński [31]), and STORMG2, a variant of STORM, are freight scheduling problems; 20Term is a vehicle location and freight operation problem devised by Gerd Infanger; and SNL is a telecomunications planning problem, communicated by Suvrajeet Sen as a small variant of the larger problem SSN (Sen et al. [37]). We used the problems LandS, GBD and Storm in the modified version of Linderoth et al. [27], with larger numbers of universe scenarios compared to the original versions of the test problems. All these test problems (with

Author's personal copy

Table 1 Test data: problem dimensions (rows, columns), number of (independent) random variables, number of scenarios, optimal objective z^* and standard deviation of the computed optimal second-stage cost $\sigma(x^*)$

Problem	Size master	Size $L^P(x;\Delta)$	RV	Scenarios	z^*	$\sigma(x^*)$
APL1P	3, 3	6, 9	5	1,280	24,642.3	4,808.8
LandS	3, 5	8, 12	3	10^6	225.6294	57.92
GBD	5, 18	6, 10	5	646,425	1,655.6278	665.95
STORMG2	186, 122	529, 1,259	1	1,000	15.580×10^6	1.3295×10^6
SNL	2, 8	15, 58	14	40,960,000	7.7727	9.9475
STORM	186, 122	529, 1,259	117	6×10^{81}	$15{,}498{,}739.41 \pm 19.11$	0.3319×10^6
20Term	4, 64	125, 764	40	1.1×10^{12}	$254{,}311.55 \pm 5.56$	8,679.18

the exception of SNL) have appeared previously in the literature (as just indicated) as computational examples in the stochastic programming domain.

Table 1 summarizes the size characteristics of each of the above problems. We use the term scenario to refer to each of the possible outcome values for the Λ's, so that the number of scenarios corresponds to the total number of subproblem outcomes in the sample space underlying problem (1). For the first four problems, the table includes their associated optimal objective values z^*, as computed by solving the problems exactly using decomposition without any sampling. For problems STORM and 20Term, we used the best objective and solution values (as provided by Linderoth et al. [27]) as sufficiently precise approximations to the exact values. Because our lower bound confidence interval procedure involves bounding the second-stage standard deviation $\sigma(x^*)$, we also report the exact value $\sigma(x^*)$, corresponding to the optimal solution x^* computed, for each problem (and accurate estimates for problems STORM and 20Term).

We slightly modified the above problems by adding box constraints (where necessary) to our first-stage feasible region, in particular lower and upper bounds to each first stage variable. As we shall see below, the box bounds both make the first stage region compact (which is convenient theoretically) and permit us to easily sample points within the first-stage region as required by Method 1 in Sect. 4. The box bounds were set reasonably far apart, in particular, for APL1P and LandS the lower/upper box bounds were 0/5000; for GBD, STORM and STORMG2 the bounds were 0/100; for SNL the lower/upper bounds were 0/10; and for 20Term the box lower/upper bounds were 0/1000.

In implementing our confidence bounds, we decided to use Method 1 (estimator (9) for the SAA case and estimator (10) for the independent sampling version) to upper bound the second-stage variance $\sigma^2(x^*)$. As just noted, these estimators require sampling a sequence of points within the first-stage feasible region. There are many such procedures that have been proposed for generating such a sequence; see, for example, Smith [43]. The most obvious such method is to sample points at random within a region containing H and to then reject points outside H. The difficulty with such a rejection algorithm is that H can be a "thin set" (e.g., H can be of low dimension, when A is rank deficient), causing such an algorithm to be very inefficient in such instances.

Author's personal copy

Our approach starts by sampling a point uniformly within the box. If the sampled point falls within the feasible region $H = \{x : Ax = b, x \geq 0\}$, we add it to the sample of feasible points. If the sampled point turns out infeasible, we project the point onto the feasible region H by solving the problem:

$$\min ||x - \hat{x}^k||^2, \text{ s/t } x \in H$$

to obtain as its optimal solution a feasible boundary point $\hat{x}_b^k \in H$. In order to obtain a feasible point in the interior of H, we compute a random convex combination of consecutive feasible boundary points as

$$\hat{x}_{feas}^k = \alpha^k \hat{x}_b^k + (1 - \alpha^k)\hat{x}_b^{k-1},$$

where each α^k is an independently distributed uniform random parameter between 0 and 1 and $\hat{x}_{feas}^k, k = 1, \ldots$ are feasible points with respect to H that are added to the sample of feasible points. We call this the *projection algorithm* and used this to generate $r = 30$ and $r = 50$ points within the feasible region. Of course, one obtains different lower bounds for these two different values of the parameter r. We used the projection algorithm rather than Smith [43] because the projection algorithm could be very easily implemented within the framework of the stochastic programming software (Infanger [19]) used for obtaining the numerical results of this paper (because solving the projection problem merely involves adding a quadratic term to the objective of the master problem).

6.1 Independent sampling

We first discuss our results in the setting of the fixed sample size implementation associated with Algorithm 1. Recall that in implementing our confidence bounds, we chose estimator (10) to upper bound the second-stage variance $\sigma^2(x^*)$.

The lower and upper bounds results are provided in Table 2, along with empirical coverages (based on running the sampling-based algorithm on each test problem 100 independent times) for the confidence intervals; the lower bounds are reported for three different estimators for bounding $\sigma(x^*)$ ($\tilde{\sigma}_*(r, \ell)$ for $r = 30$ and $r = 50$ and $s_\ell(X_{m,n})$, based on the "last" iterate $X_{m,n}$). As the table indicates, we use different values of sample size n across the various problems, because the number of scenarios varies so greatly across the problem suite. The lower and upper bounds are reported as a (signed) percentage of the true objective value, so that a lower bound of $-2\,\%$ means that the average computed lower confidence limit was $2\,\%$ lower than the true value, while an upper bound of $5\,\%$ means that the average upper confidence limit was $5\,\%$ above the true value.

We note that the upper and lower confidence bounds are practically useful for all our problems (in the sense that they are not so conservative as to be meaningless), and the coverage results for the estimated lower and upper bounds are also good. The coverage of the estimated lower bound is nearly always close to $100\,\%$, suggesting that the coverage guaranteed by Theorem 1 does indeed manifest itself at realistic

Author's personal copy

Table 2 Lower bound average value and coverage (based on estimators $\tilde{\sigma}_*(r, \ell)$ for sample sizes $r = 30$ and $r = 50$ and on estimator $s_\ell(X_{m,n})$) and upper bound average value and coverage

Problem	$\ell = n$	95 % Lower bound						95 % Upper bound	
		$r = 30$		$r = 50$		Last			
		Avg value (%)	Covg (%)	Avg value (%)	Covg (%)	Avg value (%)	Covg (%)	Avg value (%)	Covg (%)
	30	−7.88	100	−8.61	100	−4.96	98	7.37	95
APL1P	50	−5.85	100	−6.21	100	−4.10	97	5.48	99
$m = 20$	100	−3.83	100	−4.05	100	−2.87	100	4.18	96
	500	−1.67	100	−1.71	100	−1.47	100	1.81	97
	30	−9.21	100	−9.72	100	−6.94	100	9.82	98
LandS	50	−7.12	100	−7.61	100	−5.57	100	8.14	99
$m = 20$	100	−4.92	100	−4.87	100	−4.17	100	5.85	99
	500	−2.22	100	−2.26	100	−2.09	100	2.49	98
	30	−24.73	100	−25.72	100	−13.64	98	17.19	98
GBD	50	−19.02	100	−19.60	100	−12.06	100	14.36	98
$m = 20$	100	−13.21	100	−19.60	100	−8.69	100	10.09	100
	500	−5.93	100	−6.21	100	−4.08	100	4.54	97
	30	−2.54	100	−2.53	100	−2.33	100	3.02	97
STORMG2	50	−2.06	100	−2.02	100	−1.91	100	2.48	96
$m = 30$	100	−1.48	100	−1.41	100	−1.43	100	1.78	99
	500	−0.71	100	−0.71	100	−0.72	100	0.83	97
	30	−24.00	100	−24.90	100	−17.77	100	19.90	98
SNL	50	−18.61	100	−18.79	100	−14.57	100	17.98	100
$m = 20$	100	−13.68	100	−14.03	100	−11.39	100	12.75	100
	500	−6.96	100	−7.16	99	−6.20	100	6.71	100
	30	−0.83	100	−0.86	100	−0.71	100	1.0	95
STORM	50	−0.66	100	−0.65	100	−0.59	100	0.73	100
$m = 30$	100	−0.44	100	−0.46	100	−0.41	100	0.56	99
	500	−0.22	100	−0.22	100	−0.22	100	0.23	98
	30	−9.3	100	−9.92	100	−2.17	100	2.19	100
20Term	100	−4.92	100	−5.04	100	−1.37	100	1.80	100
$m = 1{,}200$	500	−2.22	100	−2.27	100	−0.71	100	1.03	100
	1,000	−1.63	100	−1.64	100	−0.58	100	0.83	100

values of the sample size per iteration. In all our test problems, the estimated standard deviation $s_l(X_{m,n})$ performs well as an estimator of $\sigma(x^*)$, even though it comes with no guarantees when the number of iterations is small. On the other hand, the maximum estimator $\tilde{\sigma}^*(r, l)$ (which does come with theoretical guarantees) typically inflates the standard deviation $\sigma(x^*)$ by only a relatively modest amount in all but one of the test problems, the exception being the 20Term problem. As expected, choosing $r = 50$ leads to a larger estimated upper bound for the standard deviation at the optimal solution and thus to a weaker (but statistically more reliable) estimated lower

Author's personal copy

Table 3 Average value (as a signed percentage of the true value) and coverage of the estimated standard deviation of second-stage cost: estimator $\tilde{\sigma}_*(r, \ell)$ for sample sizes $r = 30$, $r = 50$ and estimator $s_\ell(X_{m,n})$, based at last iteration

Problem	$\ell = n$	$r = 30$		$r = 50$		Last	
		Avg value (%)	Covg (%)	Avg value (%)	Covg (%)	Avg value (%)	Covg (%)
	30	27.61	100	33.13	100	−1.66	48
APL1P	50	21.14	100	24.39	100	0.99	40
$m = 20$	100	16.40	100	17.76	100	−1.15	42
	500	8.58	100	9.33	100	−0.20	45
	30	15.45	100	16.98	100	−1.83	41
LandS	50	12.40	100	14.19	100	−1.96	41
$m = 20$	100	9.16	100	9.86	100	−1.21	47
	500	4.77	94	5.24	98	−0.24	49
	30	50.47	100	53.78	100	−3.48	43
GBD	50	44.70	100	49.37	100	1.00	57
$m = 20$	100	40.14	100	49.37	100	0.04	50
	500	35.85	100	39.19	100	−0.74	49
	30	2.82	67	4.31	70	−1.83	43
STORMG2	50	2.90	67	3.31	70	−2.02	42
$m = 30$	100	1.95	68	2.18	78	−0.42	46
	500	0.28	52	0.71	60	−0.72	37
	30	28.39	100	33.56	100	−1.66	46
SNL	50	25.31	100	27.15	100	0.14	44
$m = 20$	100	20.12	100	23.16	100	−0.11	41
	500	16.83	100	18.26	100	1.94	64
	30	9.88	92	11.52	94	−1.35	54
STORM	50	7.70	84	7.81	89	0.00	49
$m = 30$	100	3.62	73	4.75	77	−0.64	44
	500	1.69	74	1.37	70	−0.24	47
	30	260.83	100	283.56	100	−0.11	45
20Term	100	240.50	100	248.30	100	3.57	65
$m = 1{,}200$	500	228.54	100	237.18	100	3.51	74
	1,000	225.99	100	229.05	100	4.81	83

bound than choosing $r = 30$. Table 3 gives, for various sample sizes ℓ and for sample sizes $r = 30$ and $r = 50$, the average percentage value (over 100 replications) of $(\tilde{\sigma}_*(r, \ell) - \sigma(x^*))$ relative to the standard deviation at the optimal solution $\sigma(x^*)$, as well as the coverage (the fraction of instances where $\tilde{\sigma}_*(r, \ell)$ was greater than $\sigma(x^*)$). As discussed in the paper, it may frequently be reasonable to assume that $X_{m,n}$ is close to a minimizer x^*, in which case $\sigma(X_{m,n})$ (and hence $s_l(X_{m,n})$) should be close to $\sigma(x^*)$. Consequently, we also report $s_l(X_{m,n})$ as a signed percentage of $\sigma(x^*)$.

Tables 4 and 5 present statistics for the estimated upper and lower bounds as a function of the number of iterations m. (The reported statistics are based on independent runs compared to the ones reported in Tables 2 and 3.) Because this set of experiments

Author's personal copy

Table 4 Model STORM, statistics for estimated lower bound and estimated upper bound as a function of iterations; sample size $n = 100$, $r = 30$, $\ell = n = 100$, 100 replications each

m	Lower bound			Upper bound			Lower bound	Upper bound
	95 %	Mean (%)	Median (%)	Median (%)	Mean (%)	95 %	Coverage (%)	Coverage (%)
5	−1.75	−1.48	−1.48	2.13	2.11	2.51	100	100
10	−0.94	−0.72	−0.73	0.68	0.69	1.26	100	99
15	−0.70	−0.56	−0.57	0.57	0.58	1.01	100	99
20	−0.63	−0.49	−0.49	0.56	0.54	1.03	99	100
25	−0.61	−0.46	−0.46	0.50	0.49	0.90	100	99
30	−0.58	−0.44	−0.44	0.50	0.52	0.92	100	99
35	−0.57	−0.44	−0.44	0.50	0.49	0.86	100	98
40	−0.55	−0.43	−0.44	0.46	0.49	0.85	100	100
45	−0.53	−0.42	−0.42	0.52	0.52	0.98	100	100
50	−0.52	−0.40	−0.41	0.46	0.46	0.85	100	97
55	−0.53	−0.40	−0.40	0.45	0.48	0.89	100	100
60	−0.49	−0.39	−0.39	0.50	0.51	0.88	100	99
65	−0.51	−0.38	−0.38	0.44	0.45	0.81	100	98
70	−0.47	−0.37	−0.37	0.44	0.45	0.77	100	97
75	−0.50	−0.38	−0.39	0.53	0.54	0.95	100	99
80	−0.50	−0.38	−0.39	0.53	0.54	0.95	100	99

Table 5 Model SNL, statistics for estimated lower bound and estimated upper bound as a function of iterations; sample size $n = 500$, $r = 30$, $\ell = n = 500$, 100 replications each

m	Lower bound			Upper bound			Lower bound	Upper bound
	95 %	Mean (%)	Median (%)	Median (%)	Mean (%)	95 %	Coverage (%)	Coverage (%)
5	−30.23	−27.07	−26.94	45.92	43.61	62.62	100	100
10	−16.86	−13.28	−13.18	8.71	8.87	14.95	100	100
15	−10.92	−9.26	−9.28	7.31	7.45	12.79	100	100
20	−8.23	−7.01	−7.05	5.84	6.07	10.12	100	100
25	−7.27	−6.11	−6.11	5.15	5.33	9.24	100	100
30	−6.52	−5.34	−5.36	5.22	5.15	8.57	100	100
35	−6.00	−5.06	−4.99	4.97	5.00	8.78	100	100
40	−5.80	−4.79	−4.83	5.20	5.17	9.10	100	99
45	−5.45	−4.56	−4.57	4.38	4.39	7.68	100	99
50	−5.46	−4.49	−4.51	4.45	4.51	7.86	100	100
55	−5.21	−4.41	−4.42	4.62	4.51	7.11	100	99
60	−5.22	−4.29	−4.29	4.46	4.40	8.10	100	99
65	−5.26	−4.19	−4.18	4.75	4.78	7.94	100	100
70	−4.96	−4.17	−4.22	4.71	4.52	7.32	100	99
75	−5.10	−4.10	−4.09	4.19	4.44	8.60	100	99
80	−5.01	−4.12	−4.13	4.51	4.69	7.76	100	100

Author's personal copy

Table 6 Lower bound average value and coverage (based on estimators $\hat{\sigma}_*(r, \ell)$ for sample sizes $r = 30$, $r = 50$ and on estimator $s_\ell(X_\infty)$) and upper bound average value and coverage

Problem	$\ell = n$	95 % Lower bound						95 % Upper bound	
		$r = 30$		$r = 50$		Last			
		Avg value (%)	Covg (%)	Avg value (%)	Covg (%)	Avg value (%)	Covg (%)	Avg value (%)	Covg (%)
APL1P	30	−7.48	99	−7.46	95	−6.33	92	6.25	95
	50	−5.84	98	−5.60	98	−5.00	96	4.91	91
	100	−3.87	99	−3.87	100	−3.55	99	3.32	97
	500	−1.50	97	−1.59	97	−1.49	96	1.57	95
LandS	30	−8.19	98	−8.76	98	−8.02	96	8.50	98
	50	−6.70	97	−6.71	97	−6.18	96	6.09	92
	100	−4.51	98	−4.67	98	−4.38	97	4.38	98
	500	−1.74	94	−1.85	95	−1.77	92	1.89	94
GBD	30	−16.56	100	−17.71	99	−12.83	96	12.63	93
	50	−11.74	100	−13.23	100	−9.58	94	8.74	93
	100	−8.67	100	−8.89	99	−6.48	93	6.61	96
	500	−3.86	97	−4.18	100	−3.12	92	3.15	94
STORMG2	30	−2.76	96	−2.65	97	−2.53	95	2.66	93
	50	−1.90	92	−2.29	96	−2.21	96	2.00	98
	100	−1.41	94	−1.36	94	−1.22	95	1.35	93
	500	−0.59	93	−0.57	93	−0.57	93	0.58	97
SNL	30	−15.81	98	−14.08	96	−11.32	91	13.72	99
	50	−11.62	97	−11.56	98	−9.68	95	10.15	95
	100	−8.06	98	−7.54	96	−6.30	93	7.16	97
	500	−3.47	99	−3.55	99	−3.02	97	2.93	97
STORM	30	−0.73	97	−0.74	96	−0.69	96	0.63	92
	50	−0.51	95	−0.52	95	−0.49	95	0.52	93
	100	−0.39	95	−0.37	95	−0.36	94	0.36	95
	500	−0.16	97	−0.17	95	−0.17	95	0.15	90
20Term	30	−3.66	100	−4.*l*1	100	−1.14	92	1.37	97
	100	−2.09	100	−2.08	100	−0.6	90	0.68	93
	500	−0.94	100	−0.94	100	−0.3	95	0.30	94
	1,000	−0.66	100	−0.67	100	−0.22	98	0.21	95

was intended to offer insight into the question of iterative convergence of the bounds, we selected STORM (due to its large size) and SNL (because of its large sample space) as candidate problems. As one might expect, the results demonstrate that the bounds have excellent coverage characteristics. They also make clear that if the number of iterations is large, eventually the noise associated with the fixed sample size n used at every iteration dominates the accuracy and no further iterative improvement can be gained.

For both problems the upper and lower bounds exhibit excellent coverage at every iteration, with the lower bound coverage almost always at 100 %. We note that choosing

Author's personal copy

Table 7 Average value and coverage of the bound on $\sigma(x^*)$ using $\hat{\sigma}_*(r, \ell)$ for $r = 30, r = 50$ and value $s_\ell(X_\infty)$

Problem	$\ell = n$	$r = 30$		$r = 50$		Last	
		Avg	Covg	Avg	Covg	Avg	Covg
APL1P	30	15.62	91	18.75	93	−0.67	49
	50	14.58	98	14.32	93	0.98	55
	100	10.87	95	10.07	93	0.20	48
	500	6.21	97	6.91	99	−0.20	47
LandS	30	10.87	88	8.34	87	−1.34	45
	50	7.76	88	7.92	89	−1.06	48
	100	6.06	92	6.31	92	−0.45	49
	500	3.33	87	4.16	94	−0.03	52
GBD	30	38.08	99	39.88	100	−0.58	47
	50	35.56	100	37.59	99	−1.50	44
	100	34.89	100	35.42	100	−1.12	44
	500	33.51	100	36.20	100	0.37	54
STORMG2	30	5.01	71	3.38	66	−1.35	41
	50	4.59	77	4.50	82	0.43	56
	100	3.07	75	2.63	75	0.10	51
	500	1.17	77	2.63	75	0.10	51
SNL	30	23.02	97	23.11	100	0.49	58
	50	21.09	99	19.78	100	−0.16	48
	100	17.54	100	19.31	100	0.74	53
	500	16.83	100	17.06	100	−0.58	48
STORM	30	8.36	83	7.48	78	0.16	48
	50	6.38	91	6.89	82	0.60	51
	100	4.36	78	4.11	79	−0.85	44
	500	1.18	65	1.35	72	−0.48	48
20Term	30	224.51	100	263.45	100	3.51	57
	100	243.83	100	241.75	100	5.71	73
	500	232.91	100	233.44	100	4.31	81
	1,000	230.15	100	230.96	100	4.72	92

a reasonable number of iterations is important. For example, a choice of 20 or 30 iterations is often not sufficient for obtaining the best solution possible.

6.2 Sample average approximation

We now discuss results in the setting of the sample average approximation associated with the special case of Algorithm 2, where $n_m = n$ for all m and m is large so that the problem is solved to optimality.

Note that in implementing our confidence bounds in the context of the sample average approximation, we chose estimator (9) to upper bound the second-stage variance

Author's personal copy

$\sigma^2(x^*)$. The upper bound and lower bound results (for the estimator (9), with $r = 30$ and $r = 50$ and using $s_\ell^2(X_\infty)$ as a bound on $\sigma^2(x^*)$) are provided in Table 6, along with empirical coverages (based on running the sampling-based algorithm on each test problem 100 independent times) for the confidence intervals. As for the case of independent sampling in Sect. 6.1, the average lower and upper bounds are reported as a (signed) percentage of the true objective value.

We note that also for the SAA, the upper and lower confidence bounds are reasonably tight for all problems, and the coverage results for the estimated lower and upper bounds are good. The coverage of the lower bound is always greater than or equal 95 %, indicating that also in the setting of the SAA the estimated bound on the standard deviation at the optimal solution is a statistically valid approach. Table 6 also gives the upper bound mean values and coverage results for various sample sizes n.

Table 7 gives, for various sample sizes ℓ and for sample sizes $r = 30$ and $r = 50$, the average percentage value (over 100 replications) of $(\hat{\sigma}_*(r, \ell) - \sigma(x^*))/\sigma(x^*)$ as percentage of the standard deviation at the optimal solution $\sigma(x^*)$, as well as the coverage (the fraction of instances where $\hat{\sigma}_*(r, \ell)$ was greater than $\sigma(x^*)$). The table also compares the results with the standard deviation estimator $s_\ell(X_\infty)$ obtained at the optimal solution X_∞ of the SAA problem. We note that using common random numbers results in smaller estimated values for $\hat{\sigma}_*(r, \ell)$ as compared to the estimator $\tilde{\sigma}_*(r, \ell)$ when using independent sampling.

In comparing the SAA bounds to those with independent sampling, we observe that the SAA intervals tend to be tighter. However, we do not view the empirical evidence presented here as a comprehensive computational comparison of the two approaches. Rather, we take the less ambitious view that our computations support only the conclusion that either method gives reasonable results.

Acknowledgments The authors wish to thank the referees and Associate Editor for their very helpful and insightful comments and suggestions, which have served to greatly improve both the content and exposition of this paper.

References

1. Asmussen, S., Glynn, P.W.: Stochastic Simulation: Algorithms and Analysis. Springer, New York (2007)
2. Bayraksan, G., Morton, D.P.: Assessing solution quality in stochastic programs. Math. Program. Series B **108**, 495–514 (2006)
3. Bayraksan, G., Morton, D.P., Partani, A.: Simulation-based optimality tests for stochastic programs. In: Infanger, G. (ed.) Stochastic Programming, pp. 37–55. Springer, New York (2011)
4. Beale, E.M.L.: On minimizing a convex function subject to linear inequalities. J. R. Stat. Soc. **17b**, 173–184 (1955)
5. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)
6. Dantzig, G.B., Infanger, G.: A probabilistic lower bound for two-stage stochastic programs, Report SOL 95–6. Stanford University, CA, Department of Operations Research (1995)
7. Dantzig, G.B., Madansky, M.: On the solution of two-staged linear programs under uncertainty. In: Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability. I, Neyman, J. (ed.), pp. 165–176 (1961)
8. Dantzig, G.B.: Linear programming under uncertainty. Manag. Sci. **1**, 197–206 (1955)
9. Dantzig, G.B.: Linear Programming and Extensions. Princeton University Press, Princeton (1963)
10. Dantzig, G.B., Glynn, P.W.: Parallel processors for planning under uncertainty. Ann. Oper. Res. **22**, 1–21 (1990)

Author's personal copy

11. Dupačová, J., Wets, R.J.-B.: Asymptotic behavior of statistical estimation and of optimal solutions of stochastic optimization problems. Ann. Stat. **16**, 1517–1549 (1988)
12. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 2. Wiley, New York (1971)
13. Glynn, P.W., Iglehart, D.L.: Importance sampling for stochastic simulations. Manag. Sci. **35**, 1367–1392 (1989)
14. Higle, J.L., Sen, S.: Stochastic decomposition: an algorithm for two-stage linear programs with recourse. Math. Oper. Res. **16**, 650–669 (1991)
15. Higle, J.L., Sen, S.: Finite master programs in stochastic decomposition. Math. Program. **67**, 143–168 (1994)
16. Higle, J.L., Sen, S.: Stochastic Decomposition: A statistical Method for Large-Scale Stochastic Linear Programming. Kluwer, Dordrecht (1996)
17. Higle, J.L., Sen, S.: Statistical approximations for stochastic linear programming problems. Ann. Oper. Res. **85**, 173–192 (1999)
18. Ho, J.K.: A successive linear optimization approach to the dynamic traffic assignment problem. Transp. Sci. **14**(4), 295–305 (1980)
19. Infanger, G.: DECIS User's Guide. Infanger Investment Technology, LLC., Mountain View, CA (1997)
20. Infanger, G.: Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs. Ann. Oper. Res. **39**, 41–67 (1992)
21. Infanger, G., Morton, D.: Cut sharing for multistage stochastic linear programs with interstage dependency. Math. Program. **75**, 241–256 (1996)
22. Kall, P.: Stochastic Linear Programming. Springer, Berlin (1976)
23. Karlin, S., Taylor, H.M.: Stochastic Processes. Academic Press, New York (1975)
24. King, A.J., Wets, R.J.-B.: Epi-consistency of convex stochastic programs. Stoch. Stoch. Reports **34**, 83–92 (1991)
25. King, A.J., Rockafellar, R.T.: Asymptotic theory for solutions in statistical estimation and stochastic programming. Math. Oper. Res. **18**, 148–162 (1993)
26. Kushner, H.J., Yin, G.: Stoch. Approx. Springer, New York (1997)
27. Linderoth, J., Shapiro, A., Wright, S.: The empirical behavior of sampling methods for stochastic programming. Ann. Oper. Res. **142**, 215–241 (2006)
28. Louveaux, F.V., Smeers, Y.: Optimal investment for electricity generation: a stochastic model and a test problem. In: Ermoliev, Y., Wets, R.J.-B. (eds.) Numerical Techniques for Stochastic Optimization. Springer, Berlin (1988)
29. Lustig, I.L., Mulvey, J.M., Carpenter, T.J.: Formulating two-stage stochastic programs for interior point methods. Oper. Res. **39**, 757–770 (1991)
30. 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)
31. Mulvey, J., Ruszczyński, A.: A new scenario decomposition method for large scale stochastic optimization. Oper. Res. **43**, 477–490 (1995)
32. Resnick, S.I.: Extreme Values, Regular Variation and Point Processes. Springer, New York (1987)
33. Robinson, S.M.: Analysis of sample path optimization. Math. Oper. Res. **21**, 513–528 (1996)
34. Rockafellar, R.T., Wets, R.J.-B.: Scenario and policy aggregation in optimization under uncertainty. Math. Oper. Res. **16**, 119–147 (1989)
35. Ruszczyński, A.: A regularized decomposition method for minimizing a sum of polyhedral functions. Math. Program. **35**, 309–333 (1986)
36. Ruszczyński, A., Svetanowski, A.: Accelerating the regularized decomposition method for two-stage stochastic linear programs. Eur. J. Oper. Res. **101**, 328–342 (1997)
37. Sen, S., Doverspike, R.D., Cosares, S.: Network planning with random demand. Telecommun. Syst. **3**, 11–30 (1994)
38. Sen, S., Zhou, Z., Huang, K.: Enhancements of two-stage stochastic decomposition. Comput. Oper. Res. **36**(8), 2434–2439 (2009)
39. Shapiro, A., Homem-de-Mello, T.: A simulation-based approach to two-stage stochastic programs with recourse. Math. Program. **81**, 301–325 (1998)
40. Shapiro, A., Homem-de-Mello, T.: On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. SIAM J. Optim. **11**, 70–86 (2000)
41. Shapiro, A., Homem-de-Mello, T., Kim, J.: Conditioning of convex piecewise linear stochastic programs. Math. Program. **94**, 1–19 (2002)

Author's personal copy

42. Shapiro, A.: Monte Carlo sampling methods. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programing. Elsevier, Amsterdam (2003)
43. Smith, R.L.: Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. Oper. Res. **32**, 1296–1308 (1984)
44. Van Slyke, R.M., Wets, R.: L-shaped linear programs with applications to optimal control and stochastic programming. SIAM J. Appl. Math. **17**, 638–663 (1969)
45. Wets, R.J.-B.: On the continuity of the value of a linear program and of related polyhedral-valued multifunctions. Math. Program. Study **14**, 14–29 (1985)