Math. Program., Ser. A (2013) 142:107–131
DOI 10.1007/s10107-012-0563-6

FULL LENGTH PAPER

A probability metrics approach for reducing the bias of optimality gap estimators in two-stage stochastic linear programming

Rebecca Stockbridge · Güzin Bayraksan

Received: 21 June 2011 / Accepted: 24 May 2012 / Published online: 21 June 2012

Abstract Monte Carlo sampling-based estimators of optimality gaps for stochastic programs are known to be biased. When bias is a prominent factor, estimates of optimality gaps tend to be large on average even for high-quality solutions. This diminishes our ability to recognize high-quality solutions. In this paper, we present a method for reducing the bias of the optimality gap estimators for two-stage stochastic linear programs with recourse via a probability metrics approach, motivated by stability results in stochastic programming. We apply this method to the Averaged Two-Replication Procedure (A2RP) by partitioning the observations in an effort to reduce bias, which can be done in polynomial time in sample size. We call the resulting procedure the Averaged Two-Replication Procedure with Bias Reduction (A2RP-B). We provide conditions under which A2RP-B produces strongly consistent point estimators and an asymptotically valid confidence interval. We illustrate the effectiveness of our approach analytically on a newsvendor problem and test the small-sample behavior of A2RP-B on a number of two-stage stochastic linear programs from the literature. Our computational results indicate that the procedure effectively reduces bias. We also observe variance reduction in certain circumstances.

Electronic supplementary material The online version of this article (doi:10.1007/s10107-012-0563-6) contains supplementary material, which is available to authorized users.

R. Stockbridge
Graduate Interdisciplinary Program in Applied Mathematics,
University of Arizona, Tucson, AZ 85721, USA

G. Bayraksan (✉)
Department of Systems and Industrial Engineering,
University of Arizona, Tucson, AZ 85721, USA
e-mail: guzinb@sie.arizona.edu

Keywords Stochastic programming · Monte Carlo simulation · Stability analysis · Bias reduction techniques · Confidence intervals · Probability metrics

Mathematics Subject Classification (2000) 90C15

1 Introduction

In this paper, we combine a Monte Carlo sampling-based approach to optimality gap estimation with stability results for a class of two-stage stochastic linear programs, with the intention of reducing the bias of Monte Carlo sampling-based optimality gap estimators. Stability results use probability metrics to provide continuity properties of optimal values and optimal solution sets with respect to perturbations of the original probability distribution of the random vector; see, e.g., the survey by Römisch [34]. Stability results have been successfully used for scenario reduction in stochastic programs; see, e.g., [9,13,14]. In this paper, we present a way to use stability results to motivate bias reduction.

We consider a stochastic optimization problem of the form

$$\min_{x\in X} \mathbb{E} f(x,\tilde{\xi}) = \min_{x\in X} \int_{\Xi} f(x,\xi)P(d\xi), \tag{SP}$$

where $X \subseteq \mathbb{R}^{d_x}$ represents the set of constraints the decision vector x of dimension d_x must satisfy and $\tilde{\xi}$ is a random vector of dimension d_ξ on $(\Xi, \mathscr{B}, P)$ with support $\Xi \subseteq \mathbb{R}^{d_\xi}$ and distribution P that does not depend on x. The function $f : X \times \Xi \to \mathbb{R}$ is assumed to be a Borel measurable, real-valued function, with inputs being the decision vector x and a realization ξ of the random vector $\tilde{\xi}$. Throughout the paper, we use $\tilde{\xi}$ to denote the random vector and ξ to denote its realization. The expectation operator $\mathbb{E}$ is taken with respect to P. We use z^* to denote the optimal objective function value of (SP) and x^* to denote an optimal solution to (SP). The set of optimal solutions is given by $X^* = \arg\min_{x\in X} \mathbb{E} f(x,\tilde{\xi})$.

We are interested in assessing the quality of a solution to (SP). That is, given a candidate (feasible) solution $\hat{x} \in X$ to (SP), we would like to determine whether it is optimal or nearly optimal. Assessing solution quality is critically important since many real-world problems cast as (SP)—such as two-stage stochastic linear programs with recourse, which are the focus of this paper—cannot be solved exactly and one is often left with an approximate solution $\hat{x} \in X$ without verification of its quality. This is also fundamental in optimization algorithms, as these algorithms use quality assessment iteratively, e.g., every time a new candidate solution is generated, they need to identify an optimal or nearly optimal solution to stop.

We define the quality of a solution $\hat{x} \in X$ to be its optimality gap, denoted $\mathscr{G}$, where $\mathscr{G} = \mathbb{E} f(\hat{x},\tilde{\xi}) - z^*$. (Throughout this paper, we are concerned with a fixed $\hat{x} \in X$, so we simply use $\mathscr{G}$ and suppress the dependence on $\hat{x}$.) The smaller the optimality gap, the higher the quality of the candidate solution, and a zero optimality gap implies $\hat{x}$ is optimal. The optimality gap $\mathscr{G}$ cannot be evaluated explicitly, particularly as the optimal value z^* is not known. Furthermore, exact evaluation of $\mathbb{E} f(\hat{x},\tilde{\xi})$ may not be possible. Monte Carlo sampling-based methods bypass these difficulties by allowing us to form statistical estimators of optimality gaps [2,27,29]. These take as input a candidate solution $\hat{x} \in X$ and a sample size n, and form point and interval estimators on $\mathscr{G}$. They are easy to implement, provided a sampling approximation of (SP) with moderate sample sizes can be solved, and they can be used in conjunction with any specialized solution procedure to solve the sampling approximations of the underlying problem. These methods have been successfully applied to problems in finance [5], stochastic vehicle routing [22,41], and supply chain network design [37].

It is well known that Monte Carlo statistical estimators of optimality gaps are biased [27,29]. That is, on average, they tend to over-estimate $\mathscr{G}$ for a finite sample size n. When a statistical estimate of $\mathscr{G}$ turns out to be large, this can be due to either the bias or the variance of the estimator, or it can simply be due the fact that $\mathscr{G}$ is large. When bias is the dominant factor, estimates of $\mathscr{G}$ tend to be large even if we have a high-quality solution. This significantly diminishes our ability to identify an optimal or nearly optimal solution.

Bias reduction in statistics and simulation is a well-established topic and resampling methods such as jackknife and bootstrap are commonly used for this purpose [10,38]. In stochastic programming, while there has been a lot of focus on variance reduction techniques [1,6,8,15,20,23,26], bias reduction has received relatively little attention. Only a few studies exist for this purpose. Freimer et al. [12] study the effect on bias of different sampling schemes mainly used for variance reduction, such as antithetic variates and Latin Hypercube sampling on bias. These schemes can successfully reduce the bias of the estimator of z^* with minimal computational effort; however, the entire optimality gap estimators are not considered. Partani [30] and Partani et al. [31], on the other hand, develop a generalized jackknife technique for bias reduction in MRP optimality gap estimators.

In this paper, bias reduction is motivated by the stability results in stochastic programming rather than adaptation of well-established sampling or bias reduction techniques. We specifically apply the bias reduction approach to the Averaged Two-Replication Procedure (A2RP) of [2] and use a particular stability result from [34] involving the Kantorovich metric. Utilizing the Kantorovich metric to calculate distances between probability measures results in a significant computational advantage (see Sect. 4.1). The specific stability result we use, however, restricts (SP) to a class of two-stage stochastic linear programs with recourse (see Sect. 2). The bias reduction approach presented in the paper does not require resampling—like the bootstrap and jackknife methods commonly used in statistics—and thus no additional sampling approximation problems need to be solved. The cost of bias reduction, however, comes from solving a minimum-weight matching problem, which is used to partition a random sample so as to reduce bias by minimizing the Kantorovich metric. Minimum-weight matching is a well-known combinatorial optimization problem for which efficient algorithms exist. It can be solved in polynomial time in sample size n and the computational burden is likely to be minimal compared to solving (approximations of) real-world stochastic programs with hundreds of stochastic parameters.

Partitioning a random sample in an effort to reduce bias as we do in this paper results in observations that are no longer independent nor identically distributed. We show that the resulting distributions on the partitioned subsets converge weakly to P, the original distribution of $\tilde{\xi}$, almost surely (a.s.). We also provide conditions under which the point estimators are strongly consistent (i.e., converge to the true value a.s. as opposed to in probability; simply referred to as consistent from now on) and the interval estimator is asymptotically valid. Our computational experiments indicate that variance may also be reduced, especially when $\hat{x} = x^*$. Thus, the method presented in this paper has the potential to produce more reliable estimates of the optimality gap and increase our ability to recognize optimal or nearly optimal solutions.

The rest of the paper is organized as follows. In the next section, we formally define our problem setup and list necessary assumptions. In Sect. 3, we give an overview of optimality gap estimation in stochastic programs and review A2RP from [2]. We also present the stability result from [34]. We then introduce our bias reduction technique in Sect. 4 and illustrate the technique on an instance of a newsvendor problem in Sect. 5. Asymptotic properties of the resulting estimators are provided in Sect. 6. In Sect. 7, we present our computational experiments on a number of test problems. Finally, in Sect. 8, we provide a summary and outline future work.

2 Framework

While (SP) encompasses many classes of problems, in this paper, we focus on a particular class dictated by the specific stability result we use to motivate the proposed bias reduction technique (see Sect. 3.2). We consider two-stage stochastic linear programs with recourse, where $f(x, \tilde{\xi}) = cx + h(x, \tilde{\xi})$, $X = \{x : Ax = b, x \geq 0\}$, and $h(x, \xi)$ is the optimal value of the minimization problem

$$\min_y \{qy : Wy = R(\xi) - T(\xi)x, \ \ y \geq 0\}.$$

The above problem has fixed recourse (W is non-random) and stochasticity only on the right-hand side ($R(\xi)$ and $T(\xi)$ are random). We assume that X and Ξ are convex polyhedral. We also assume that T and R depend affine linearly on ξ, which allows for modeling first-order dependencies between them, such as those that arise in commonly-used linear factor models. Furthermore, we assume that our model has relatively complete recourse, i.e., for each $(x, \xi) \in X \times \Xi$, there exists $y \geq 0$ such that $Wy = R(\xi) - T(\xi)x$, and dual feasibility, i.e., $\{\pi : \pi W \leq q\} \neq \emptyset$. These assumptions are needed to ensure the stability result presented in Sect. 3.2. We make the following additional assumptions:

(A1) $X \neq \emptyset$ and is compact,

(A2) Ξ is compact.

Assumption (A1) requires that the problem be feasible and the set of feasible solutions be closed and bounded. Let $\mathscr{P}(\Xi)$ be the set of probability measures on Ξ with finite first order moments, i.e., $\mathscr{P}(\Xi) = \left\{Q : \int_\Xi \|\xi\| Q(d\xi) < \infty\right\}$. It follows immediately from assumption (A2) that $P \in \mathscr{P}(\Xi)$, a condition required by our theoretical results.

For the class of problems we consider, $f(\cdot, \xi)$ is convex in x for all fixed $\xi \in \Xi$, and $f(x, \xi)$ satisfies the following Lipschitz continuity condition for all $x, x' \in X$ and $\xi \in \Xi$, for some $L > 0$:

$$|f(x, \xi) - f(x', \xi)| \leq L \max\{1, \|\xi\|\} \|x - x'\|,$$

where $\|\cdot\|$ is some norm (see Proposition 22 in [34]). This result leads directly to the continuity of $f(\cdot,\xi)$ in x for fixed ξ. We also note that $f(x,\cdot)$ is Lipschitz continuous and thus continuous in ξ; see e.g., Corollary 25 in [34]. Assumptions (A1) and (A2) along with continuity in both variables imply that $f(x,\xi)$ is uniformly bounded, i.e., $\exists\, C<\infty$ such that $|f(x,\xi)|<C$ for each $x\in X, \xi\in\Xi$, a condition necessary for establishing consistency of our point estimators (see Sect. 6.2). Uniform boundedness also ensures that $f(x,\xi)$ is a real-valued function and enforces the relatively complete recourse and dual feasibility assumptions. In addition, convexity and continuity in x implies that $\mathbb{E} f(x,\tilde{\xi})<\infty$ is convex and continuous in x. Hence, (SP) has a finite optimal solution on X, and so X^* is non-empty.

The results presented in this paper require precise probabilistic modeling of the Monte Carlo sampling performed. In particular, the expectations and the almost sure statements are made with respect to the underlying product measure. An overview of this framework is as follows. Let $(\Omega,\mathscr{A},\hat{P})$ be the space formed by the product of a countable sequence of identical probability spaces $(\Xi_i,\mathscr{B}_i,P_i)$, where $\Xi_i=\Xi$, $\mathscr{B}_i=\mathscr{B}$, and $P_i=P$, for $i=1,2,\ldots$, and let ξ^i denote an outcome in the sample space Ξ_i. An outcome $\omega\in\Omega$ then has the form $\omega=(\xi^1,\xi^2,\ldots)$. Now, define the countable sequence of projection random variables $\{\tilde{\xi}^i:\Omega\to\Xi, i=1,2,\ldots\}$ by $\tilde{\xi}^i(\omega)=\xi^i$. Then, the collection $\{\tilde{\xi}^1,\ldots,\tilde{\xi}^n\}$ is a random sample from $(\Xi,\mathscr{B},P)$, and $\tilde{\xi}:=\xi^1$ is a random variable with distribution P.

3 Background

In this section, we give an overview of the Monte Carlo sampling-based techniques for assessing the quality of a candidate solution by estimating its optimality gap. In particular, we review the A2RP from [2] to which we will apply the bias reduction technique presented in Sect. 4. We also provide a stability result from [34] that is fundamental to the bias reduction technique.

3.1 Assessing solution quality

3.1.1 Optimality gap estimation

Given a candidate solution $\hat{x}\in X$, we would like to determine whether it is optimal or nearly optimal. This can be achieved through investigating $\mathscr{G}$, the optimality gap of $\hat{x}$, where $\mathscr{G}=\mathbb{E} f(\hat{x},\tilde{\xi})-z^*$. Since $\mathscr{G}$ usually cannot be evaluated explicitly, we use Monte Carlo sampling to provide an approximation of (SP) and exploit the properties of this approximation to estimate the optimality gap. We first approximate P, using the observations from a random sample $\{\tilde{\xi}^1,\ldots,\tilde{\xi}^n\}$ described in Sect. 2, by the empirical distribution $P_n(\cdot)=\sum_{i=1}^n \frac{1}{n}\delta_{\{\tilde{\xi}^i\}}(\cdot)$. The use of $(\cdot)$ indicates that P_n is a probability measure on Ξ. We then approximate (SP) by

$$\min_{x\in X}\frac{1}{n}\sum_{i=1}^n f(x,\tilde{\xi}^i)=\min_{x\in X}\int_{\Xi} f(x,\xi)P_n(d\xi). \qquad (\mathrm{SP}_n)$$

Let x_n^* denote an optimal solution to (SP$_n$) and z_n^* denote the corresponding optimal value. As mentioned above, it is most convenient throughout the paper to interpret expectations and almost sure statements relating to z_n^* with respect to the underlying probability measure $\hat{P}$. For instance, $\mathbb{E}z_n^* = \int_\Omega z_n^*(\omega)\hat{P}(d\omega)$.

By interchanging minimization and expectation, we have

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

In other words, $\mathbb{E}z_n^*$ provides us with a lower bound on z^*. An upper bound on $\mathscr{G}$, $\mathbb{E}f(\hat{x},\tilde{\xi}) - z^*$, is then given by $\mathbb{E}f(\hat{x},\tilde{\xi}) - \mathbb{E}z_n^*$. We estimate $\mathbb{E}f(\hat{x},\tilde{\xi}) - \mathbb{E}z_n^*$ by

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

With fixed $\hat{x}\in X$, $\frac{1}{n}\sum_{i=1}^{n} f(\hat{x},\tilde{\xi})$ is an unbiased estimator of $\mathbb{E}f(\hat{x},\tilde{\xi})$ due to i.i.d. sampling. However, since $\mathbb{E}z_n^* - z^* \le 0$,

$$\mathbb{E}G_n \ge \mathbb{E}f(\hat{x},\tilde{\xi}) - z^*,$$

and hence G_n is a biased estimator of the optimality gap. We assume the same observations are used in both terms on the right-hand side in (1), so $G_n \ge 0$. This results in variance reduction through the use of common random variates. Consequently, compared to z_n^*, which has the same bias, bias can be a more prominent factor in G_n.

It is well-known that the bias decreases as the size of the random sample increases [27,29]. That is, $\mathbb{E}z_n^* \le \mathbb{E}z_{n+1}^*$ for all n. However, the rate the bias shrinks to zero can be slow, e.g., of order $\mathrm{O}(n^{-1/2})$; see for instance Example 4 in [3]. One way to reduce bias is to simply increase the sample size. However, significant increases in sample sizes are required to obtain a modest reduction in bias, and increasing the sample size is not computationally desirable since obtaining statistical estimators of optimality gaps requires solving a sampling approximation problem.

We note that there are other approaches to assessing solution quality. Some of these approaches are motivated by the Karush–Kuhn–Tucker conditions, see, e.g., [17,40]. There is also work on assessing solution quality with respect to a particular sampling-based algorithm, typically utilizing the bounds obtained through the course of the algorithm, see, e.g., [7,16,19,25]. We apply the bias reduction technique on A2RP, originally introduced in [2]. Given a candidate solution $\hat{x}\in X$, A2RP produces point and interval estimators on $\mathscr{G}$ and can be used as a standalone procedure or within a sequential framework [4].

3.1.2 The Averaged Two-Replication Procedure

Let n be even and let $\{\tilde{\xi}^1,\ldots,\tilde{\xi}^n\}$ be a random sample from P. Now, let $\mathscr{I}^1$ be a uniformly distributed random variable independent of $\{\tilde{\xi}^1,\ldots,\tilde{\xi}^n\}$ taking values in the set of all subsets of $\{1, \ldots, n\}$ of size $n/2$, and let I_1 be an instance of $\mathscr{I}^1$. Let $I_2 = (I_1)^C$; that is, I_2 contains all $n/2$ elements of $\{1, \ldots, n\}$ that are not in I_1. This is essentially equivalent to generating two independent random samples of size $n/2$. However, we prefer to use the notation I_1 and I_2 to emphasize the random partitioning. Later, the proposed bias reduction technique will alter this partitioning mechanism.

Let $P_{I_k}(\cdot) = \sum_{i \in I_k} \frac{2}{n} \delta_{\{\tilde{\xi}^i\}}(\cdot)$ be the empirical probability measure formed on the kth set of observations, $k = 1, 2$. Similar to the definition of (SP$_n$), let (SP$_{I_k}$) denote the problem

$$\min_{x \in X} \frac{1}{n} \sum_{i \in I_k} f(x, \tilde{\xi}^i) = \min_{x \in X} \int_{\Xi} f(x, \xi) P_{I_k}(d\xi), \qquad (\mathrm{SP}_{I_k})$$

$x^*_{I_k}$ denote an optimal solution to (SP$_{I_k}$), and let $z^*_{I_k}$ be the optimal value, for $k = 1, 2$. Let z_α be the $1 - \alpha$ quantile of the standard normal distribution. A2RP is as follows:

A2RP
Input: Desired value of $\alpha \in (0, 1)$, even sample size n, and a candidate solution $\hat{x} \in X$.
Output: $(1 - \alpha)$-level confidence interval (CI) on $\mathscr{G}$.

1. Sample i.i.d. observations $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$ from P.
2. Generate a random partition of $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$ via I_1 and I_2, and produce P_{I_1} and P_{I_2}.
3. For $k = 1, 2$:
   3.1. Solve (SP$_{I_k}$) to obtain $x^*_{I_k}$ and $z^*_{I_k}$.
   3.2. Calculate:

$$G_{I_k} = \frac{2}{n} \sum_{i \in I_k} f(\hat{x}, \tilde{\xi}^i) - z^*_{I_k} \quad \text{and}$$
$$s^2_{I_k} = \frac{1}{n/2 - 1} \sum_{i \in I_k} \left[\left(f(\hat{x}, \tilde{\xi}^i) - f(x^*_{I_k}, \tilde{\xi}^i) \right) - G_{I_k} \right]^2.$$

4. Calculate the optimality gap and sample variance estimators by taking the average; $G_I = \frac{1}{2}\left(G_{I_1} + G_{I_2}\right)$ and $s^2_I = \frac{1}{2}\left(s^2_{I_1} + s^2_{I_2}\right)$.
5. Output one-sided confidence interval on $\mathscr{G}$:

$$\left[0, G_I + \frac{z_\alpha s_I}{\sqrt{n}}\right]. \qquad (2)$$

For notational simplicity, we suppress the dependence on the sample size n and candidate solution $\hat{x}$ when defining the estimators. A set of conditions under which the point estimator G_I of A2RP is consistent, and the interval estimator of A2RP in (2) is asymptotically valid are given in [2]. To understand the behavior of the interval estimator for small sample sizes, we consider the probability that it contains the optimality gap, referred to as its *coverage probability*, or simply *coverage*, and compare it to the desired coverage of $1-\alpha$.

A2RP is a modification of the Single Replication Procedure (SRP), also presented in [2]. Rather than dividing the n observations into two subsets, SRP produces a single optimality gap estimator and sample variance estimator using all the n observations. The bias of the optimality gap estimator is thus decreased compared to A2RP. However, for small sample sizes it can happen that x_n^* coincides with the candidate solution $\hat{x}$. The optimality gap and sample variance estimators are then zero (e.g., when $\hat{x} = x_n^*$, G_n in (1) is zero). This also results in a confidence interval estimator of width zero, even though the candidate solution may be significantly suboptimal. For a detailed discussion on coinciding solutions, we refer the readers to Section 6 of [2]. We repeat Examples 1 and 2 from [2] as an illustration:

Example 1 Consider the problem $\{\min \mathbb{E}[\tilde{\xi}x] : -1 \leq x \leq 1\}$, where $\tilde{\xi} \sim N(\mu, 1)$ and $\mu > 0$. The optimal pair is $(x^*, z^*) = (-1, -\mu)$. We examine the candidate solution $\hat{x} = 1$, which has the largest optimality gap of 2μ. If the random sample satisfies $\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\xi}^i < 0$, then $x_n^* = 1$ coincides with $\hat{x}$, and so the point and interval estimators of SRP are zero. Setting $\mu = 0.1$, $\alpha = 0.10$ and $n = 50$, and using normal quantiles, we obtain an upper bound on the coverage of SRP as $1 - P(\bar{\xi} < 0) \approx 0.760$, which is considerably below the desired coverage of 0.90. Now consider A2RP that uses two samples of size $n/2 = 25$ each. Let $\bar{\xi}_1 = \frac{2}{n}\sum_{i \in I_1} \tilde{\xi}^i$ be the sample mean of the first subset of observations, and similarly let $\bar{\xi}_2$ be the sample mean of the second subset. In this case, the probability of obtaining a confidence interval estimator of non-zero width is given by $1 - P(\bar{\xi}_1 < 0)P(\bar{\xi}_2 < 0) \approx 1 - (0.308)^2 \approx 0.905$.

Due to the difficulties that can arise when using SRP, we focus on A2RP and aim to reduce the bias of this optimality gap estimator. We will return to the above example in Sect. 4.2 to see how bias reduction affects coinciding solutions. Now, we turn our attention to a stability result which forms the basis for the proposed reduction technique.

3.2 A stability result

Stability results in stochastic programming quantify the behavior of (SP) when P, the original distribution of $\tilde{\xi}$, is perturbed. In this paper, we are particularly interested in changes in the optimal value z^* under perturbations of P; however, stability results also examine the changes in the solution sets X^*. In this section, we will use $z^*(P)$ to denote the optimal value of (SP) when the distribution of $\tilde{\xi}$ is P. Similarly, $z^*(Q)$ denotes the optimal value under the distribution Q, a perturbation of P.

Probability metrics, which calculate distances between probability measures, can provide upper bounds on $|z^*(P) - z^*(Q)|$, the change in the optimal value. One such probability metric relevant for the class of problems we consider is $\hat{\mu}_d(P, Q)$, the Kantorovich metric with cost function $d(\xi^1, \xi^2) = ||\xi^1 - \xi^2||$, where $||\cdot||$ is a norm.

The following result—a restatement of Corollary 25 to Theorem 23 in [34], written to match our paper's notation—provides continuity properties of optimal values of (SP) with respect to perturbations of P.

Theorem 1 *Let only $T(\xi)$ and $R(\xi)$ be random, and assume that relatively complete recourse and dual feasibility hold. Let $P \in \mathscr{P}(\Xi)$, and X^* be non-empty. Then, there exist constants $L > 0, \delta > 0$ such that*

$$\left|z^*(P) - z^*(Q)\right| \le L\hat{\mu}_d(P, Q)$$

whenever $Q \in \mathscr{P}(\Xi)$ and $\hat{\mu}_d(P, Q) < \delta$.

All conditions necessary to apply Theorem 1 for the class of problems we consider are satisfied, as specified in Sect. 2; see [34] for details. The above stability result implies that if P and Q are sufficiently close with respect to the Kantorovich metric, then the optimal value of (SP) behaves Lipschitz continuously with respect to changes in the probability distribution.

Suppose that P is a discrete probability measure placing masses $p_1, \dots, p_{N_P}$ on the points $\{\xi^1, \dots, \xi^{N_P}\}$ in Ξ, respectively, and Q is a discrete measure with masses $q_1, \dots, q_{N_Q}$ on the points $\{\nu^1, \dots, \nu^{N_Q}\}$ in Ξ, respectively. Then the Kantorovich metric can be written in the form of the Monge–Kantorovich transportation problem, which formulates the transfer of mass from P to Q:

$$\hat{\mu}_d(P, Q) = \min_{\eta} \left\{ \sum_{i=1}^{N_P} \sum_{j=1}^{N_Q} \|\xi^i - \nu^j\| \eta_{ij} : \sum_{i=1}^{N_P} \eta_{ij} = q_j, \forall j;\ \sum_{j=1}^{N_Q} \eta_{ij} = p_i, \forall i;\ \eta_{ij} \ge 0, \forall i, j \right\}. \tag{MKP}$$

This is the well-known *transportation problem*, where P can be viewed to have N_P supply nodes, each with supply $p_i, i = 1, \dots, N_P$; similarly, Q has N_Q demand nodes, each with demand $q_j, j = 1, \dots, N_Q$; and total supply and demand match, i.e., $\sum_{i=1}^{N_P} p_i = \sum_{j=1}^{N_Q} q_j = 1$. Thus, $\hat{\mu}_d(P, Q)$ is the minimum cost of transferring mass from P to Q. Representing the Kantorovich metric as the optimal value of a well-known, efficiently solvable optimization problem is vital in allowing us to implement the bias reduction technique described in the next section.

4 Bias reduction via probability metrics

In this section, we present a technique to reduce the bias in sampling-based estimates of z^* in stochastic programs and apply it to the A2RP optimality gap estimators. We begin by discussing the motivation behind the technique and explaining the connection with Theorem 1. We then formally state the resulting procedure to obtain variants of the A2RP optimality gap estimators after bias reduction.

4.1 Motivation for bias reduction technique

Consider a partition of n observations $\{\tilde{\xi}^1, \dots, \tilde{\xi}^n\}$ given by index sets S_1 and S_2, where

(i) $S_1, S_2 \subset \{1, \ldots, n\}$ and $S_2 = (S_1)^C$,
(ii) $|S_1| = |S_2| = n/2$, and
(iii) each $\tilde{\xi}^i$, $i \in S_1 \cup S_2$, receives probability mass $2/n$.

Note that S_1 and S_2 are functions of the random sample $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$. This is a generalization of the partitioning performed via I_1 and I_2, where we now allow dependencies between S_1 and S_2. For any given $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$, we have

$$\frac{1}{2}(z^*_{S_1} + z^*_{S_2}) = \frac{1}{2}\left(\min_{x\in X} \frac{2}{n}\sum_{i\in S_1} f(x, \tilde{\xi}^i) + \min_{x\in X} \frac{2}{n}\sum_{i\in S_2} f(x, \tilde{\xi}^i)\right)$$
$$\leq \min_{x\in X} \frac{1}{n}\left(\sum_{i\in S_1} f(x, \tilde{\xi}^i) + \sum_{i\in S_2} f(x, \tilde{\xi}^i)\right) = \min_{x\in X} \frac{1}{n}\sum_{i=1}^{n} f(x, \tilde{\xi}^i) = z^*_n.$$

Therefore, by the monotonicity of expectation, the following inequality holds:

$$\frac{1}{2}(\mathbb{E}z^*_{S_1} + \mathbb{E}z^*_{S_2}) \leq \mathbb{E}z^*_n \leq z^*. \tag{3}$$

Inequality (3) indicates that when n observations are divided in two, the expected gap between $\frac{1}{2}(z^*_{S_1} + z^*_{S_2})$ and z^* grows. This motivates us to partition the observations via index sets S_1 and S_2 that maximize $\frac{1}{2}(\mathbb{E}z^*_{S_1} + \mathbb{E}z^*_{S_2})$. This approach will help to alleviate the increase in bias that results from using two subsets of $n/2$ observations rather than one set of n observations. Since $\frac{1}{2}(\mathbb{E}z^*_{S_1} + \mathbb{E}z^*_{S_2})$ is always bounded above by $\mathbb{E}z^*_n$, we equivalently aim to minimize $\mathbb{E}z^*_n - \frac{1}{2}(\mathbb{E}z^*_{S_1} + \mathbb{E}z^*_{S_2})$. This problem can be hard to solve, but an approximation is obtained by:

$$\mathbb{E}z^*_n - \frac{1}{2}(\mathbb{E}z^*_{S_1} + \mathbb{E}z^*_{S_2}) \leq \frac{1}{2}\mathbb{E}\left[\left|z^*_n - z^*_{S_1}\right| + \left|z^*_n - z^*_{S_2}\right|\right].$$

We would thus like to minimize $\mathbb{E}[|z^*_n - z^*_{S_1}| + |z^*_n - z^*_{S_2}|]$, but again this is a hard problem. In an effort to achieve this, we focus on $|z^*_n - z^*_{S_1}| + |z^*_n - z^*_{S_2}|$ for a given sample of size n. By viewing the empirical measure P_n of the random sample $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$ as the original measure and the measures $P_{S_k}(\cdot) = \sum_{i\in S_k} \frac{2}{n}\delta_{\{\tilde{\xi}^i\}}(\cdot)$, $k = 1, 2$, as perturbations of P_n, we appeal to Theorem 1 to obtain an upper bound containing probability metrics:

$$|z^*_n - z^*_{S_1}| + |z^*_n - z^*_{S_2}| \leq L\hat{\mu}_d(P_n, P_{S_1}) + L\hat{\mu}_d(P_n, P_{S_2}). \tag{4}$$

As a result, we aim to reduce the bias of the optimality gap estimator by partitioning the observations according to sets S_1 and S_2 that minimize $\hat{\mu}_d(P_n, P_{S_1}) + \hat{\mu}_d(P_n, P_{S_2})$. By minimizing these metrics, we would like P_{S_1} and P_{S_2} to mimic P_n as much as possible. This way, we may expect the resulting optimal values of the partitions to be closer to z^*_n, reducing the bias induced by partitioning.

We note that several approximations were used above and (4) is valid only when P_n and P_{S_k} are sufficiently close in terms of the Kantorovich metric, for $k = 1, 2$. However, it is natural to think of P_{S_k} as a perturbation of P_n even though Theorem 1 does not specify how close they should be. The advantage of using these approximations is that it results in an easily solvable optimization problem (see the discussion below). Even though it is approximate, we present strong evidence that the proposed bias technique can be successful via analytical results on a newsvendor problem in Sect. 5 and numerical results for several stochastic programs from the literature in Sect. 7.

Let us now examine $\hat{\mu}_d(P_n, P_{S_k})$, $k = 1, 2$ when we have a *known* partition via S_1 and S_2. Suppose given a realization of the random sample $\{\xi^1, \ldots, \xi^n\}$ and the corresponding empirical measure P_n, we have identified S_1 and S_2 that satisfy (i)–(iii) above. Because $d(\xi^i, \xi^j) = ||\xi^i - \xi^j|| = 0$ whenever $\xi^i = \xi^j$, for $i \in \{1, \ldots, n\}$ and $j \in S_k$, the Monge–Kantorovich problem (MKP) in this setting turns into an *assignment problem* (for a given index set S_k):

$$\hat{\mu}_d(P_n, P_{S_k}) = \min_{\eta} \left\{ \sum_{i \in S_2} \sum_{j \in S_1} \|\xi^i - \xi^j\| \eta_{ij} : \sum_{i \in S_2} \eta_{ij} = \frac{1}{n}, \forall j; \right.$$
$$\left. \sum_{j \in S_1} \eta_{ij} = \frac{1}{n}, \forall i; \ \eta_{ij} \geq 0, \forall i, j \right\}.$$

Furthermore, since the cost function $d(\xi^i, \xi^j) = ||\xi^i - \xi^j||$ is symmetric, $\hat{\mu}_d(P_n, P_{S_1})$ = $\hat{\mu}_d(P_n, P_{S_2})$. It follows that if we minimize $\hat{\mu}_d(P_n, P_{S_1})$, we automatically minimize $\hat{\mu}_d(P_n, P_{S_2})$. Therefore, identifying sets S_1 and S_2 that minimize the sum of the Kantorovich metrics is equivalent to finding an S_1 that minimizes $\hat{\mu}_d(P_n, P_{S_1})$. Thus, to attempt to reduce the bias of the optimality gap estimator, we wish to find an index set of size $n/2$ that solves the problem:

$$\min \left\{ \hat{\mu}_d \left(P_n, P_{S_1} \right) : S_1 \subset \{1, \ldots, n\}, \ |S_1| = n/2. \right\} \tag{PM}$$

Note that this is the well-known *minimum-weight perfect matching problem*. Given a graph with n nodes and m edges, it can be solved in polynomial time of $O(mn \log n)$ [28]. The running time for our problem is $O(n^3 \log n)$ since we have a fully connected graph. A special case of (PM) when ξ is univariate is solvable in $O(n \log n)$ via a sorting algorithm, as the optimal solution is to place the odd order statistics in one subset and the even order statistics in the other. For large-scale stochastic programs, solving instances of (SP_n) can be expected to be the computational bottleneck compared to solving (PM).

4.2 The Averaged Two-Replication Procedure with Bias Reduction

In this section, we present the Averaged Two-Replication Procedure with Bias Reduction (A2RP-B) that results from adapting A2RP to include the bias reduction technique described in Sect. 4.1. To distinguish from the uniformly chosen subsets I_1 and I_2 defined Sect. 3.1.2, we denote an optimal solution to (PM) by J_1 and let $J_2 = (J_1)^C$. The resulting probability measures are denoted P_{J_k}, $k = 1, 2$, where $P_{J_k} = \sum_{i \in J_k} \frac{2}{n} \delta_{\tilde{\xi}^i}$.

A2RP-B
Input: Desired value of $\alpha \in (0, 1)$, even sample size n, and a candidate solution $\hat{x} \in X$.
Output: $(1 - \alpha)$-level confidence interval on $\mathscr{G}$.

1. Sample i.i.d. observations $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$ from P.
2. Generate J_1 and J_2 by solving (PM), and produce P_{J_1} and P_{J_2}.
3. For $k = 1, 2$:
   3.1. Solve (SP_{J_k}) to obtain $x^*_{J_k}$ and $z^*_{J_k}$.
   3.2. Calculate:

$$G_{J_k} = \frac{2}{n} \sum_{i \in J_k} f(\hat{x}, \tilde{\xi}^i) - z^*_{J_k} \quad \text{and}$$

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

4. Calculate the optimality gap and sample variance estimators by taking the average; $G_J = \frac{1}{2}\left(G_{J_1} + G_{J_2}\right)$ and $s^2_J = \frac{1}{2}\left(s^2_{J_1} + s^2_{J_2}\right)$.
5. Output one-sided confidence interval on $\mathscr{G}$:

$$\left[0, G_J + \frac{z_\alpha s_J}{\sqrt{n}}\right]. \tag{5}$$

A2RP-B differs from A2RP in Step 2. Here, a minimum-weight perfect matching problem (PM) is solved to obtain an optimal partition of the observations via the index sets J_1 and J_2. Note that the elements in J_1 and J_2 depend on the observations $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$, and so J_1 and J_2 are random variables acting on Ω. Hence, J_1 and J_2 are *not* independent of $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$, distinguishing P_{J_1} and P_{J_2} from P_{I_1} and P_{I_2}. The random partitioning mechanism of I_1 and I_2 results in i.i.d. observations in P_{I_1} and P_{I_2}. Unfortunately, this property is lost in P_{J_1} and P_{J_2}. Nevertheless, we prove in Sect. 6 that the point estimators G_J and s^2_J are consistent and the interval estimator given by (5) is asymptotically valid.

We conclude this section by updating Example 1 to include the effects of bias reduction. We will illustrate A2RP-B in more detail on an instance of a newsvendor problem in the next section.

Example 2 Consider the problem described in Example 1. As before, $\alpha = 0.10$ and $n = 50$. Let $\bar{\xi}_{J_1} = \frac{2}{n} \sum_{i \in J_1} \tilde{\xi}^i$ be the sample mean of the first subset of 25 observations (all odd order statistics), and similarly let $\bar{\xi}_{J_2}$ be the sample mean of the second subset (all even order statistics) after solving (PM). To estimate an upper bound on the coverage of A2RP-B, we ran 1,000,000 independent runs in MATLAB and computed the proportions of runs $\bar{\xi}_{J_1}$ and $\bar{\xi}_{J_2}$ were negative. This resulted in the estimate $1 - P(\bar{\xi}_{J_1} < 0)P(\bar{\xi}_{J_2} < 0) \approx 1 - (0.363)(0.145) \approx 0.947$. Compared to A2RP with $P(\bar{\xi}_k < 0) \approx 0.308$ for each subset $k = 1, 2$, after solving (PM), the sample mean of the first subset shifted slightly downward, increasing this probability, whereas the sample mean of the second subset shifted slightly upward, decreasing this probability. As a result, the probability of obtaining coinciding solutions is decreased. Hence the upper bound on the coverage of A2RP-B is greater than A2RP for this problem.

In general, A2RP-B may be viewed as in between SRP and A2RP. Like A2RP, it can lower the occurrence of coinciding solutions while at the same time having a lower bias like SRP. Computational results in Sect. 7 seem to support this hypothesis.

5 Illustration: newsvendor problem

Before presenting theoretical results, we illustrate the above bias reduction technique on an instance of a newsvendor problem. For this problem, we are able to derive analytical results, and therefore can compare the optimality gap estimators produced by A2RP, A2RP-B, and SRP to examine the efficacy of the bias reduction technique.

The specific newsvendor problem we consider is as follows: a newsvendor would like to determine the number of newspapers to order daily, x, in order to maximize expected daily profit. Each copy sells at a price r and costs the newsvendor c, where $0 < c < r$. The daily demand, $\tilde{\xi}$, is assumed to be random with a $U(0, b)$ distribution. The problem can be expressed as

$$\min_{x \geq 0} \mathbb{E}\left[cx - r\min\{x, \tilde{\xi}\}\right]. \tag{6}$$

The optimal solution is $x^* = b(r-c)/r$ and the optimal value is $z^* = -b(r-c)^2/(2r)$. Note that (6) can be rewritten as a two-stage stochastic linear program in the form presented in Sect. 2.

Prior to finding expressions for the biases of the optimality gap estimators, we note two results that are used in the subsequent derivations. First, let $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$ be a random sample of size n from a $U(0, b)$ distribution, and let $\{\tilde{\xi}^{(1)}, \ldots, \tilde{\xi}^{(n)}\}$ denote the ordering of the random sample, i.e., $\tilde{\xi}^{(1)} \leq \tilde{\xi}^{(2)} \leq \cdots \leq \tilde{\xi}^{(n)}$. The optimal solution to the approximated problem (SP_n) using this random sample is $x_n^* = \tilde{\xi}^{(l^*)}$, where $l^* = \lceil (r-c)n/r \rceil$. The optimal value of ($\text{SP}_n$) is thus

$$z_n^* = cx_n^* - \frac{r}{n}\sum_{i=1}^{n} \min\{x_n^*, \tilde{\xi}^i\} = c\tilde{\xi}^{(l^*)} - \frac{r}{n}\sum_{i=1}^{l^*-1} \tilde{\xi}^{(i)} - \frac{r}{n}\left(n - l^* + 1\right)\tilde{\xi}^{(l^*)}.$$

Second, recall that the ith order statistic from a $U(0, b)$ random sample of size n satisfies $\tilde{\xi}^{(i)}/b \sim \beta(i, n+1-i)$, where $\beta(\alpha_1, \alpha_2)$ denotes a random variable having a Beta distribution with parameters α_1 and α_2.

We now determine the bias of G_I, the optimality gap estimator produced by A2RP. In this case, the n observations are randomly partitioned into two subsets of size $n/2$, generating the corresponding sampled problems (SP$_{I_k}$), $k = 1, 2$. Relabel the observations $\tilde{\xi}^i$, $i \in I_1$, as $\tilde{\xi}^i_{I_1}$, and similarly for I_2. The optimal solution to (SP$_{I_k}$) is $x^*_{I_k} = \tilde{\xi}^{(i^*)}_{I_k}$, where $i^* = \lceil (r-c)n/2r \rceil = (r-c)n/2r + \kappa$, for some $\kappa \in [0, 1)$. The ith order statistic of each subset satisfies $\tilde{\xi}^{(i)}_{I_k}/b \sim \beta\left(i, \frac{n}{2} + 1 - i\right)$. After some algebra, the bias of G_I is

$$z^* - \frac{1}{2}(\mathbb{E}z^*_{I_1} + \mathbb{E}z^*_{I_2}) = -\frac{b}{n(n+2)r}\left[2\kappa(\kappa-1)r^2 - cnr + c^2 n\right]. \tag{7}$$

The analysis changes somewhat under A2RP-B. The newsvendor problem is univariate in ξ, and so (PM) places the odd order statistics in one subset and the even order statistics in the other. Since the order statistics are computed from the original sample size of n, the ith order statistic follows a $\beta(i, n+1-i)$ distribution. Note that after solving (PM), the observations in each subset are no longer i.i.d., since order statistics are neither identically distributed nor independent. Solving the sampling problem using the first subset of observations leads to the optimal solution $x^*_{J_1} = \tilde{\xi}^{(2i^*-1)}$ and using the second set of observations produces the optimal solution $x^*_{J_2} = \tilde{\xi}^{(2i^*)}$. Following the same steps, we calculate the bias of G_J as

$$z^* - \frac{1}{2}(\mathbb{E}z^*_{J_1} + \mathbb{E}z^*_{J_2}) = -\frac{b}{2n(n+1)r}\left[4\kappa(\kappa-1)r^2 - cnr + c^2 n\right]. \tag{8}$$

We now consider the limiting behavior of the percentage reduction in the bias of the optimality gap estimator going from A2RP to A2RP-B, which is given by subtracting expression (8) from expression (7) and normalizing by (7). We get

$$\%\ \text{Red. in Bias} = 1 - \frac{-\frac{b}{2n(n+1)r}\left[4\kappa(\kappa-1)r^2 - cnr + c^2 n\right]}{-\frac{b}{n(n+2)r}\left[2\kappa(\kappa-1)r^2 - cnr + c^2 n\right]} \to 1 - \frac{1}{2} \text{ as } n \to \infty.$$

Therefore, the percentage reduction in the bias converges to 50 % as $n \to \infty$. So, simply partitioning the random sample into odd and even order statistics [the result of solving (PM)] gives an optimality gap estimator with asymptotically half the bias compared to using a random partitioning. This result holds regardless of the values of the parameters r, c, and b for this specific newsvendor problem, so parameter choices that change the bias of the optimality gap estimator will not alter the large-sample behavior of the bias reduction technique. For small sample size behavior of this newsvendor problem, see Sect. 7.3. Our numerical results indicate that convergence of the percentage reduction in bias is achieved very quickly, e.g., around a sample size of $n = 100$.

Finally, we compare A2RP-B and A2RP to SRP. Observe that replacing n with $2n$ in (7) gives the bias of the optimality gap estimator produced by SRP. Consequently, the ratio of the bias of the A2RP optimality gap estimator to the bias of the SRP estimator converges to 2 as $n \to \infty$, indicating that partitioning the observations into two random subsets doubles the bias for larger sample sizes. In contrast, the ratio of the biases of the A2RP-B and SRP optimality gap estimators converges to 1 as $n \to \infty$. In essence, the bias reduction technique performs "anti-partitioning" for this problem by eliminating the additional bias introduced from the partitioning.

6 Theoretical properties

We now prove that the estimators G_J and s_J^2 of A2RP-B are strongly consistent and that A2RP-B provides an asymptotically valid confidence interval on the optimality gap. This is important because applying a bias reduction technique can sometimes result in overcorrection of the bias and lead to undesirable behavior. In this section, we show that asymptotically such unwanted behavior does not happen for our method. The technical difficulty in the consistency proofs for the A2RP-B estimators comes from the fact that the proposed bias reduction technique destroys the i.i.d. nature of the observations in the partitioned subsets of observations. Recall that in A2RP, the uniform partitioning of the observations preserves the i.i.d. property, but this is not the case for A2RP-B; see Sect. 5 for an illustration from the newsvendor problem. Hence, it is necessary to generalize the consistency proofs in [2] to cover the non-i.i.d. case arising from solving (PM).

6.1 Weak convergence of empirical measures

We first establish the weak convergence of the empirical probability measures P_{J_1} and P_{J_2} to P, the original distribution of $\tilde{\xi}$, a.s. This provides the structure necessary to obtain consistent estimators.

Theorem 2 *Assume that $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$ is an i.i.d. sample from distribution P and (A2) holds. Then the probability measures on the partitioned sets obtained by solving (PM), P_{J_1} and P_{J_2}, converge weakly to P, the original distribution of $\tilde{\xi}$, a.s.*

Proof Since $\hat{\mu}_d$ is a metric, by the triangle inequality we have that

$$\hat{\mu}_d(P, P_{J_1}) \leq \hat{\mu}_d(P, P_n) + \hat{\mu}_d(P_n, P_{J_1}).$$

Also, $\hat{\mu}_d(P_n, P_{J_1}) \leq \hat{\mu}_d(P_n, P_{I_1})$, since the partitioning of the observations via J_1 minimizes the Kantorovich metric; hence, the random partition provides an upper bound. Therefore,

$$\hat{\mu}_d(P, P_{J_1}) \leq \hat{\mu}_d(P, P_n) + \hat{\mu}_d(P_n, P_{I_1}),$$

and by applying the triangle inequality again, we obtain

$$\hat{\mu}_d(P, P_{J_1}) \leq \hat{\mu}_d(P, P_n) + \hat{\mu}_d(P, P_n) + \hat{\mu}_d(P, P_{I_1}) = 2\hat{\mu}_d(P, P_n) + \hat{\mu}_d(P, P_{I_1}).$$

We would like to show that $\hat{\mu}_d(P, P_{J_1}) \to 0$ as $n \to \infty$, a.s. First, applying the Strong Law of Large Numbers (SLLN) for all bounded, continuous functions on Ξ

gives that the random empirical measure P_n converges weakly to the non-random measure P, a.s. This combined with (A2) yields $\int_\Xi \|\xi\| P_n(d\xi) \to \int_\Xi \|\xi\| P(d\xi)$, a.s. Hence, applying Theorem 6.3.1 of [32], we obtain $\hat{\mu}_d(P, P_n) \to 0$ as $n \to \infty$, a.s. and similarly, $\hat{\mu}_d(P, P_{I_1}) \to 0$ as $n \to \infty$, a.s. The second statement follows from the fact that P_{I_1} is essentially the same as $P_{n/2}$. Combining these, we obtain that $2\hat{\mu}_d(P, P_n) + \hat{\mu}_d(P, P_{I_1}) \to 0$, a.s. Therefore, $\hat{\mu}_d(P, P_{J_1}) \to 0$, a.s., and another application of Theorem 6.3.1 of [32] implies that P_{J_1} converges weakly to P, a.s. The same argument holds for P_{J_2}. □

Even though we lose the i.i.d. property of the observations in the partitioned subsets after minimizing the Kantorovich metrics, Theorem 2 shows the weak convergence of the resulting probability measures to the original measure.

6.2 Consistency of point estimators

We now show the consistency of the estimators G_J and s_J^2 in the almost sure sense. For a fixed $\hat{x} \in X$, define $\sigma_{\hat{x}}^2(x) = \mathrm{var}\big(f(\hat{x}, \tilde{\xi}) - f(x, \tilde{\xi})\big)$, and denote the optimal solutions that minimize and maximize $\sigma_{\hat{x}}^2(x)$ by $x_{\min}^* \in \arg\min_{x \in X^*} \sigma_{\hat{x}}^2(x)$ and $x_{\max}^* \in \arg\max_{x \in X^*} \sigma_{\hat{x}}^2(x)$, respectively. Note that since $f(x, \tilde{\xi})$ is continuous in x, $\mathbb{E} f(x, \tilde{\xi})$ is continuous, and hence X^* is closed (and therefore compact). In addition, $\sigma_{\hat{x}}^2(x)$ is continuous, and thus $\arg\min_{x \in X^*}$ and $\arg\max_{x \in X^*}$ are nonempty.

Theorem 3 *Assume $\hat{x} \in X$, $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$ is an i.i.d. sample from distribution P, and (A1) and (A2) hold. Fix $0 < \alpha < 1$. Let n be even and consider A2RP-B. Then,*

(i) *all limit points of $x_{J_k}^*$ lie in X^*, a.s., for $k = 1, 2$;*
(ii) *$z_{J_k}^* \to z^*$, a.s., as $n \to \infty$, for $k = 1, 2$;*
(iii) *$G_J \to \mathscr{G}$, a.s., as $n \to \infty$;*
(iv) *$\sigma_{\hat{x}}^2(x_{\min}^*) \le \liminf_{n\to\infty} s_J^2 \le \limsup_{n\to\infty} s_J^2 \le \sigma_{\hat{x}}^2(x_{\max}^*)$, a.s.*

Proof (i) First, note from Theorem 2 that the probability measures on the partitioned subsets converge weakly to the original distribution of $\tilde{\xi}$ as $n \to \infty$, a.s. As a result, for $k = 1, 2$, $\int_\Xi f(x, \xi) P_{J_k}(d\xi)$ epi-converges to $\int_\Xi f(x, \xi) P(d\xi)$ as $n \to \infty$, a.s., by Theorem 3.9 of [43]. Thus by Theorem 3.9 of [43], all limit points of $x_{J_k}^*$ lie in X^*, a.s., for $k = 1, 2$.

(ii) Using epi-convergence, Theorem 7.33 of [33] along with assumptions (A1) and (A2) give that $z_{J_k}^*$ converges to z^*, a.s., as $n \to \infty$.

(iii) By definition, $G_J = \frac{1}{2}\left[G_{J_1} + G_{J_2}\right]$ where $G_{J_k} = \frac{2}{n}\sum_{i \in J_k} f(\hat{x}, \tilde{\xi}^i) - z_{J_k}^*$, for $k = 1, 2$. For a feasible $x \in X$, define $\bar{f}_n(x) = \frac{1}{n}\sum_{i=1}^n f(x, \tilde{\xi}^i)$. Then $G_J = \bar{f}_n(\hat{x}) - \frac{1}{2}(z_{J_1}^* + z_{J_2}^*)$. Since the original sample is formed using n i.i.d. observations, $\bar{f}_n(\hat{x})$ converges to $\mathbb{E} f(\hat{x}, \tilde{\xi})$, a.s., by the SLLN. Furthermore, by part (ii), $\frac{1}{2}(z_{J_1}^* + z_{J_2}^*)$ converges to z^*, a.s., as $n \to \infty$. We conclude that $G_J \to \mathbb{E} f(\hat{x}, \tilde{\xi}) - z^*$, a.s., as $n \to \infty$.

(iv) Let $w(x,\xi) = 2C - (f(\hat{x},\xi) - f(x,\xi))$ and $\bar{w}_{J_k} = \frac{2}{n}\sum_{i\in J_k} w(x,\tilde{\xi}^i)$, for a given $x \in X$, $k = 1, 2$. Recall that the constant C gives a uniform bound on $f(x,\xi)$; see Sect. 2. We define $w(x,\xi)$ in this fashion to enforce non-negativity. Altering our notation slightly and fixing $\hat{x} \in X$, we define $s^2_{J_k}(x) = \frac{1}{n/2-1}\sum_{i\in J_k}\left(w(x,\tilde{\xi}^i) - \bar{w}_{J_k}(x)\right)^2$. Note that $s^2_{J_k}(x^*_{J_k})$ is equivalent to $s^2_{J_k}$ defined in Sect. 4.2. Rewriting, we obtain

$$s^2_{J_k}(x) = \frac{n/2}{n/2-1}\left[\left(\frac{2}{n}\sum_{i\in J_k} w^2(x,\tilde{\xi}^i)\right) - \left(\bar{w}_{J_k}(x)\right)^2\right]. \qquad (9)$$

We show that the sequence of functions $\{s^2_{J_k}(x)\}$ converges uniformly to $\sigma^2_{\hat{x}}(x)$, a.s., as $n \to \infty$, for $k = 1, 2$. To this end, we first examine the two terms inside the brackets in (9).

By the uniform boundedness of $f(x,\xi)$, $|f(\hat{x},\xi) - f(x,\xi)| \le 2C$; hence, $w(x,\xi) \ge 0$ for all $x \in X, \xi \in \Xi$. It also immediately follows that $w(x,\cdot)$ is bounded in ξ since for all $x \in X$, $|w(x,\xi)| \le 4C$, and $w(x,\cdot)$ is continuous in ξ for the class of problems we consider. Therefore, for each $x \in X$, by the definition of weak convergence and using Theorem 2, we have $\bar{w}_{J_k}(x) \to \mathbb{E}w(x,\tilde{\xi})$, a.s., as $n \to \infty$, i.e., the SLLN holds pointwise, a.s. Since $f(\cdot,\xi)$ is convex in x, $w(\cdot,\xi)$ is convex in x (note that $\hat{x}$ is fixed). Hence, we apply Corollary 3 from [39] to obtain $\sup_{x\in X}\left|\bar{w}_{J_k}(x) - \mathbb{E}w(x,\tilde{\xi})\right| \to 0$, a.s., as $n \to \infty$, i.e., $\bar{w}_{J_k}(x)$ converges uniformly to $\mathbb{E}w(x,\tilde{\xi})$, a.s., as $n \to \infty$. Note that $w^2(x,\cdot)$ is bounded and continuous in ξ, and because $w(\cdot,\xi) \ge 0$, $w^2(\cdot,\xi)$ is also convex in x. Hence, following the same steps as above, we conclude that $\frac{2}{n}\sum_{i\in J_k} w^2(x,\tilde{\xi}^i)$ converges uniformly to $\mathbb{E}w^2(x,\tilde{\xi})$, a.s., as $n \to \infty$. Combining these, it follows that $a_{J_k}(x) := \left[\left(\frac{2}{n}\sum_{i\in J_k} w^2(x,\tilde{\xi}^i)\right) - \left(\bar{w}_{J_k}(x)\right)^2\right]$ converges uniformly, a.s., as $n \to \infty$ to $\mathbb{E}w^2(x,\tilde{\xi}) - \left(\mathbb{E}w(x,\tilde{\xi})\right)^2 = \mathrm{var}(w(x,\tilde{\xi})) = \sigma^2_{\hat{x}}(x)$.

The remainder of the proof follows as a slightly modified version of the proof of Proposition 1 in [2]. Specifically, the uniform convergence of $s^2_{J_k}(x) = \frac{n/2}{n/2-1}a_{J_k}(x)$ to $\sigma^2_{\hat{x}}(x)$ and the subsequent bounds on $\liminf_{n\to\infty} s^2_{J_k}(x^*_{J_k})$ and $\limsup_{n\to\infty} s^2_{J_k}(x^*_{J_k})$ can be shown in a similar way for $k = 1, 2$. Averaging produces the final result. □

Parts (i) and (ii) of Theorem 3 establish the consistency of $x^*_{J_k}$ and $z^*_{J_k}$, an optimal solution and the optimal value of (SP$_{J_k}$). Similarly, parts (iii) and (iv) establish the consistency of G_J and s^2_J, the point estimators produced by A2RP-B. Note that if (SP) has a unique optimal solution; that is, $X^* = \{x^*\}$, then part (i) implies that $x^*_{J_k} \to x^*$, for $k = 1, 2$, and part (iv) implies that $\lim_{n\to\infty} s^2_J = \sigma^2_{\hat{x}}(x^*)$, a.s., as $n \to \infty$.

6.3 Asymptotic validity of the interval estimator

In our final result, we show the asymptotic validity of the confidence interval estimator produced by A2RP-B, given in (5). This justifies the construction of an approximate confidence interval after bias reduction.

Theorem 4 *Assume $\hat{x} \in X$, $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$ is an i.i.d. sample from distribution P, and (A1) and (A2) hold. Fix $0 < \alpha < 1$. Let n be even and consider A2RP-B. Then,*

$$\liminf_{n\to\infty} P\left(\mathscr{G} \le G_J + \frac{z_\alpha s_J}{\sqrt{n}}\right) \ge 1 - \alpha.$$

Proof First, note that if $\hat{x} \in X^*$, then the inequality is satisfied automatically. Suppose now that $\hat{x} \notin X^*$. As in the proof of part (iii) of Theorem 3, we express G_J as $G_J = \bar{f}_n(\hat{x}) - \frac{1}{2}\left(z^*_{J_1} + z^*_{J_2}\right)$, where $\bar{f}_n(x) = \frac{1}{n}\sum_{i=1}^n f(x, \tilde{\xi}^i)$. Since $z^*_{J_k} = \min_{x\in X} \sum_{i\in J_k} f(x, \tilde{\xi}^i)$ for $k = 1, 2$, $G_J \ge \bar{f}_n(\hat{x}) - \bar{f}_n(x)$, for all $x \in X$. Noting that $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$ is an i.i.d. sample, the rest of the proof proceeds as in the proof of Theorem 1 in [2]. □

7 Computational experiments

In Sect. 6, we proved asymptotic results regarding the consistency and validity of estimators produced using A2RP-B. In this section, we apply A2RP-B to several test problems in order to examine its small-sample behavior. We begin our discussion by introducing the test problems used for evaluating the bias reduction technique, followed by the experimental setup in Sects. 7.1 and 7.2. Then, in Sect. 7.3, we present the results of our experiments and discuss computational effort. We end our discussion by providing insights gained from our experiments in Sect. 7.4. Additional details and discussion are presented in the online resource.

7.1 Test problems

To fully evaluate the efficacy of the proposed bias reduction technique, we consider four test problems from the literature; namely the newsvendor problem (denoted NV), APL1P, PGP2, and GBD. All four problems are two-stage stochastic linear programs with fixed recourse and stochasticity on the right-hand side, and can be solved exactly, allowing us to compute exact optimality gaps. Characteristics of these problems are summarized in Table 1. NV is defined as in Sect. 5 and can be solved analytically. We set the cost of one newspaper, c, to be 5, and its selling price, r, to 15. The demand $\tilde{\xi}$ is assumed to have a $U(0, 10)$ distribution. The electric power generation model PGP2 of [18] has 3 stochastic parameters and 576 scenarios. APL1P is a power expansion problem with 5 independent stochastic parameters and 1,280 scenarios [21]. GBD, described in [11], is an aircraft allocation model. The version we use has 646,425 scenarios generated by 5 independent stochastic parameters.

The standard formulations of these three problems differ slightly from the formulation presented in Sect. 2, in that $\xi := (R, T)$ rather than R and T being functions of ξ. This discrepancy can be easily remedied by defining the functions $R(\xi)$ and $T(\xi)$ in our formulation to be the coordinate projections of ξ, so with a slight abuse of notation, $R(\xi) = R$ and $T(\xi) = T$. Then $R(\xi)$ and $T(\xi)$ satisfy the affine linearity

Table 1 Test problem characteristics

Problem	No. of 1st stage variables	No. of 2nd stage variables	No. of stochastic parameters	No. of scenarios
NV	1	1	1	∞
PGP2	4	16	3	576
APL1P	2	9	5	1,280
GBD	17	10	5	646,425

Table 2 Optimal and suboptimal candidate solutions

Problem	x^*	Suboptimal $\hat{x}$	z^*	$\mathscr{G}$
NV	$6\frac{2}{3}$	8.775	$33\frac{1}{3}$	$3\frac{1}{3}$
PGP2	(1.5, 5.5, 5, 5.5)	(1.5, 5.5, 5, 4.5)	447.32	1.14
APL1P	(1800, 1571.43)	(1111.11, 2300)	24,642.32	164.84
GBD	(10, 0, 0, 0, 0, 12.48, 1.19, 5.33, 0, 4.24, 0, 20.76, 7.81, 0, 7.20, 0, 0)	(10, 0, 0, 0, 0, 12.43, 1.22, 5.33, 0, 4.32, 0, 20.68, 8.05, 0, 6.95, 0, 0)	1,655.63	1.15

assumption in Sect. 2 and we can express the problems in the form assumed in this paper. All test problems satisfy the required assumptions and can be solved exactly.

We selected two candidate solutions, $\hat{x}$, for each test problem listed in Table 1. The first candidate solution is the optimal solution, i.e., $\hat{x} = x^*$. Note that all these problems have a unique optimal solution. (We provide computations on newsvendor instances with multiple optimal solutions in the online resource.) The second candidate solution is a suboptimal solution. For NV, APL1P, and PGP2, the suboptimal solution is the solution used in the computational experiments in [2]. We selected a suboptimal solution for GBD by solving an independent sampling problem and setting its solution as the candidate solution. Table 2 summarizes the optimal and suboptimal solutions used in our computational experiments, along with the optimal value and the optimality gap of the suboptimal candidate solution.

7.2 Experimental setup

The primary objective of our computational experiments is to determine the reduction in the bias of the point estimator G_J of A2RP-B compared to the estimator G_I of A2RP for finite sample sizes n. It is well-known that bias reduction techniques in statistics can increase the variance of an estimator; therefore, we use the mean-squared error (MSE) to capture both effects. Recall that if $\hat{\theta}$ is an estimator of θ, the MSE of $\hat{\theta}$ is given by $\mathbb{E}(\hat{\theta} - \theta)^2 = (\mathbb{E}\hat{\theta} - \theta)^2 + \mathbb{E}(\hat{\theta} - \mathbb{E}\hat{\theta})^2$, where the first term is the square of the bias and the second term is the variance of $\hat{\theta}$.

Our experiments were conducted as follows. First, for each test problem, we fixed a candidate solution (optimal or suboptimal) and set $\alpha = 0.10$. Then, we applied A2RP and A2RP-B for a variety of sample sizes ($n = 50, 100, 200, \dots, 1000$) to test the small-sample behavior and to observe any potential trends as n increases. For each independent run, we used the same random number stream for both A2RP and A2RP-B to enable a direct comparison of the two procedures. We used a batch size of m to estimate the biases of G_I and G_J by averaging across m independent runs. We also obtained single estimates of $\text{var}(G_I)$, $\text{var}(G_J)$, $\text{MSE}(G_I)$, and $\text{MSE}(G_J)$ using the m runs. In order to obtain better estimates, we repeated this procedure M times, resulting in a total of $m \times M$ independent runs. The means of the M estimates of the bias, variance, and MSE and the $m \times M$ confidence interval widths were used to compute percentage reductions.

Since the stochastic parameters of APL1P take values that vary by several orders of magnitude, we used a weighted Euclidean norm to better calculate the distance between scenarios when defining (PM). We used the standard Euclidean norm for the other test problems. For NV, we used the quicksort algorithm (in C++) to solve the sampling approximations (SP_{I_k}) and (SP_{J_k}), $k = 1, 2$, as the optimal solution is a sample quantile of demand. We also used the quicksort algorithm to perform the minimum-weight perfect matching. For all other test problems, we used the regularized decomposition (RD) code (in C++) by Świetanowski and Ruszczyński [35,36] to solve the sampling approximations. We modified this code to use the Mersenne Twister algorithm to generate random samples [42]. To solve (PM), we used the Blossom V code of Kolmogorov [24]. We note that there are multiple ways to partition the observations given a solution to (PM); we simply chose our partition based on the output from Blossom V. Given that NV and its corresponding matching problem can be solved efficiently, we set $m = 1{,}000$ and $M = 1{,}000$ for a total of 1,000,000 independent runs for each sample size n for this problem. For the other problems, we used $m = 10$ and $M = 1{,}000$ for a total of 10,000 independent runs for each n. For PGP2, APL1P, and GBD, we used the high performance computing center at the University of Arizona and for NV, we used the MORE Institute facilities.

Finally, we know from Theorem 4 that the confidence intervals will attain the desired coverage of 0.90 for large sample sizes. However, given that bias reduction may reduce the width of the confidence interval estimator, it is important to consider the change in coverage for small sample sizes when applying bias reduction. We estimated the coverage for each algorithm and sample size. This was done by computing $\hat{p}$, the proportion of the $m \times M$ independent runs in which the confidence interval contained the optimality gap. Note that when the candidate solution is optimal, the optimality gap is 0, and so the coverage is always trivially 1. The estimator $\hat{p}$ is a scaled binomial random variable, and hence for the suboptimal candidate solution we formed a 90 % confidence interval on the coverage via $\hat{p} \pm 1.645\sqrt{\hat{p}(1-\hat{p})/10^6}$ for NV and $\hat{p} \pm 1.645\sqrt{\hat{p}(1-\hat{p})/(5 \times 10^4)}$ for the other test problems.

We now present the computational results for each candidate solution, beginning with the optimal candidate solution.

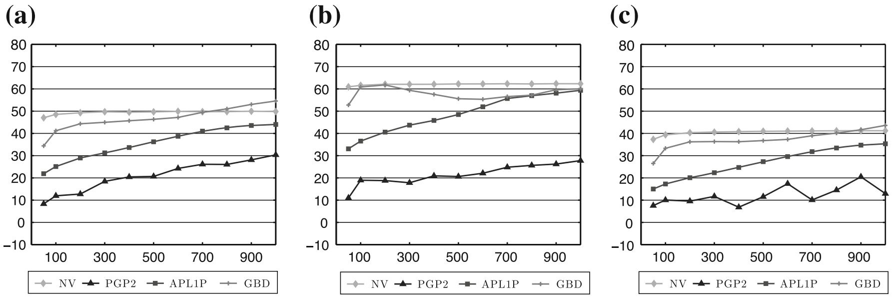

Fig. 1 Percentage reductions between A2RP and A2RP-B in **a** bias and **b** MSE of optimality gap estimator, and **c** CI width with respect to sample size n for optimal candidate solutions

7.3 Results of experiments

7.3.1 Optimal candidate solution

Figure 1 depicts a summary for all sample sizes in terms of the percentage reduction in the bias and MSE of the optimality gap estimator and the width of the confidence interval estimator between A2RP and A2RP-B when $\hat{x}$ is fixed to the optimal solution.

In particular, Fig. 1a shows the percentage reduction between the biases of G_I and G_J. The results for NV match the theory presented in Sect. 5, and we note the very fast convergence of the percentage reduction in the bias to 50 %. For the other test problems, we observe a monotonic increase in the percentage reduction in the bias with sample size. The summary of results on the MSE of the optimality gap estimator is depicted in Fig. 1b. We observe that the proposed bias reduction technique not only reduces the bias but also the variance of the optimality gap estimator. Like the bias, we observe increases in variance reduction as n increases. As a result, the percentage reduction in the MSE is notable for all test problems, and is roughly monotonically increasing. Finally, Fig. 1c shows the percentage reduction in the CI width at an optimal candidate solution. Because the optimality gap of an optimal solution is zero, reduction in the interval widths in this case is desirable. We again observe an increasing trend with sample size. Detailed results are provided in the online resource.

7.3.2 Suboptimal candidate solution

We now consider the suboptimal candidate solutions. Figure 2 shows plots of the percentage reductions in the bias and the MSE of the optimality gap estimator and in the CI width. Although the bias of the optimality gap estimator is independent of the candidate solution, its variance, and hence MSE, depends on the candidate solution. The MSE of the optimality gap estimator is reduced mainly at smaller sample sizes. One exception is PGP2, which exhibits MSE reduction across all values of n. The percentage reduction in the width of the confidence interval estimator is fairly small, with the exception of GBD for small sample sizes. Table 3 provides confidence intervals on the coverage for $n = 200$. The coverages are lowered under A2RP-B in every case; however, they remain close to 90 %, with the exception of PGP2. Note that PGP2

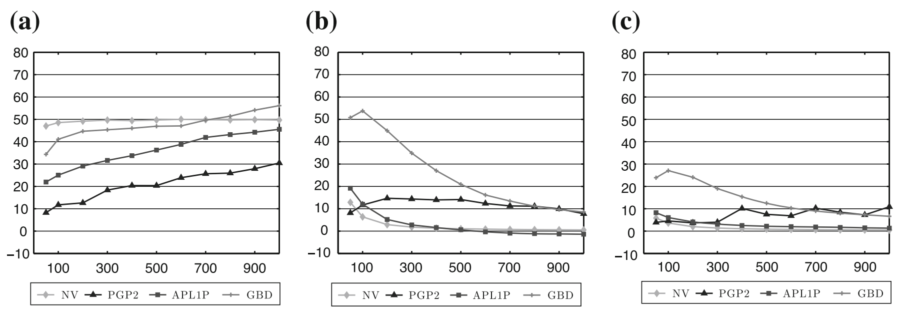

Fig. 2 Percentage reductions between A2RP and A2RP-B in **a** bias and **b** MSE of optimality gap estimator, and **c** CI width with respect to sample size n for suboptimal candidate solutions

Table 3 Confidence interval estimator for suboptimal candidate solutions ($n = 200$)

Problem	A2RP	A2RP-B
NV	0.912 ± 0.000	0.894 ± 0.001
PGP2	0.821 ± 0.006	0.792 ± 0.007
APL1P	0.899 ± 0.005	0.867 ± 0.006
GBD	0.979 ± 0.002	0.939 ± 0.004

is known to have low coverage when A2RP is used [2]. A2RP-B reduces coverage for PGP2 but is still much higher than SRP, which yields coverage probabilities of 0.50–0.60 at the same candidate solution [2].

7.4 Discussion

In this section, we summarize insights gained from our computational experiments and discuss our findings.

- The percentage reduction in the bias of the optimality gap estimator tends to increase as n increases. We hypothesize that this is due to the stability result that motivates the bias reduction technique. Recall that Theorem 1 requires P and Q to be to sufficiently close, and at larger sample sizes, we expect P_n and $P_{J_k}, k = 1, 2,$ to be closer.
- The bias reduction technique works well when an optimal candidate solution is used. In this case, both the bias and the variance are reduced, resulting in a significant reduction in the MSE of the optimality gap point estimator.
- At a suboptimal candidate solution, bias reduction is not affected, but the bias reduction technique reduces the variance, and hence the MSE, at smaller sample sizes. However, it can sometimes increase the variance at higher sample sizes, weakening the MSE reduction. The coverage is slightly reduced.

Suppose we solve an independent sampling problem (or use any other method) to obtain a candidate solution. Fixing this solution, we apply A2RP to obtain an estimate of its optimality gap. If this estimate is large, then we do not know if it is a good solution or not. Note that even when an optimal solution is obtained, this estimate can be large due to bias or variance. Alternatively, the candidate solution itself may have a large optimality gap. Suppose the candidate solution obtained is indeed an optimal solution. Then, the use of A2RP-B can significantly increase our ability to detect that this is an optimal solution. Our results indicate that A2RP-B reduces the bias, variance, and MSE of the optimality gap point estimator, and the width of the interval estimator, at an optimal solution. The risk in doing so is a decrease in the coverage at suboptimal solutions. The reduction in bias remains same but the variance and MSE are mainly reduced at smaller sample sizes at suboptimal solutions, indicating that A2RP-B provides a more reliable point estimator at suboptimal solutions at smaller sample sizes.

If identifying optimal solutions is of primary importance, then, we recommend A2RP-B. If, on the other hand, conservative coverage is the primary concern, then we recommend the Multiple Replications Procedure (MRP) of Mak et al. [27], which is known to be more conservative; see the computational results and also the preliminary guidelines in [2].

8 Summary and future work

In this paper, we present a bias reduction technique for a class of stochastic programs that is rooted in a stability result. The proposed technique partitions the observations by minimizing the Kantorovich metrics between the empirical measure of the original sample and the probability measures on the resulting partitioned observations. This amounts to solving a minimum-weight perfect matching problem, which is polynomially solvable in the sample size. The bias reduction technique is applied to the A2RP optimality gap estimators for a given candidate solution. Analytical results on an instance of a newsvendor problem and computations indicate that bias reduction technique can reduce the bias introduced by partitioning while maintaining appropriate coverage. We show that the optimality gap and sample variance estimators of A2RP-B are consistent and the confidence interval estimator is asymptotically valid. Preliminary computational results suggest that the technique works well for optimal candidate solutions, decreasing both the bias and the variance of the optimality gap estimator, and hence the MSE. For suboptimal solutions, bias reduction is unaffected but variance and MSE reduction are weakened. Coverage is slightly lowered after bias reduction.

Future work includes the application of the bias reduction technique to other optimality gap estimators such as the MRP estimator, which is formed by averaging $k \geq 30$ independent estimates of G_I. By further partitioning, ideas discussed in this paper can be used to form a variant of MRP. This also raises some interesting questions. For instance, is there an optimal number of partitions that maximizes relative bias reduction? What happens when different partitioning schemes are used (e.g., one partition gets 1/3 of the observations, the other 2/3)? These merit further investigations. Another area of future research is studying the use of the A2RP-B estimators in a sequential setting [4].

For other classes of stochastic programs, stability results use different probability metrics [34]. Minimizing these metrics could potentially lead to more complicated problems than the minimum-weight perfect matching problem that arose in the current context; see for instance the scenario reduction techniques for different classes of stochastic programs [13,14]. However, quick solution methods could still provide significant reduction in bias. This is another area of future research. Finally, future work includes comparison of this approach to other bias reduction methods such as the jackknife estimators of [30,31] and other sampling techniques.

Acknowledgments The authors are grateful to David Morton for valuable conversations and suggestions, the associate editor and the referees whose suggestions significantly improved the paper, Andrzej Ruszczyński and Artur Świetanowski for access to their regularized decomposition code, Péguy Pierre-Louis for assistance with the computational experiments, and Roman Garnett for helpful discussions. This work has been partially supported by the National Science Foundation through grants DMS-0602173, EFRI-0835930, and CMMI-1151226.

References

1. Bailey, T., Jensen, P., Morton, D.: Response surface analysis of two-stage stochastic linear programming with recourse. Nav. Res. Logist. **46**, 753–778 (1999)
2. Bayraksan, G., Morton, D.: Assessing solution quality in stochastic programs. Math. Program. **108**, 495–514 (2006)
3. Bayraksan G., Morton D.: Assessing solution quality in stochastic programs via sampling. In: Oskoorouchi, M.R. (ed.) INFORMS TutORials in Operations Research, vol. 6, pp. 102–122. INFORMS, Hanover, MD (2009)
4. Bayraksan, G., Morton, D.: A sequential sampling procedure for stochastic programming. Oper. Res. **59**, 898–913 (2011)
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.: Quasi-Monte Carlo approaches to option pricing. Technical report 94-19, Department of Industrial and Operations Engineering, University of Michigan (1994)
7. Dantzig, G., Infanger, G.: A probabilistic lower bound for two-stage stochastic programs. Technical report SOL 95-6, Department of Operations Research, Stanford University (1995)
8. Drew, S., de Mello, T.H.: Quasi-Monte Carlo strategies for stochastic optimization. In: Proceedings of the 2006 Winter Simulation Conference, pp. 774–782 (2006)
9. Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: an approach using probability metrics. Math. Program. **95**, 493–511 (2003)
10. Efron, B., Tibshirani, R.: An Introduction to the Bootstrap. Chapman & Hall, New York (1993)
11. Ferguson, A., Dantzig, G.: The allocation of aircraft to routes: an example of linear programming under uncertain demands. Manag. Sci. **3**, 45–73 (1956)
12. Freimer, M.B., Thomas, D.J., Linderoth, J.T.: The impact of sampling methods on bias and variance in stochastic linear programs. Comput. Optim. Appl. **51**(1), 51–75 (2012)
13. Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming. Comput. Optim. Appl. **24**, 187–206 (2003)
14. Henrion, R., Küchler, C., Römisch, W.: Scenario reduction in stochastic programming with respect to discrepancy distances. Comput. Optim. Appl. **43**, 67–93 (2009)
15. Higle, J.: Variance reduction and objective function evaluation in stochastic linear programs. INFORMS J. Comput. **10**, 236–247 (1998)
16. Higle, J., Sen, S.: Statistical verification of optimality conditions for stochastic programs with recourse. Ann. Oper. Res. **30**, 215–240 (1991)
17. Higle, J., Sen, S.: Stochastic decomposition: an algorithm for two-stage linear programs with recourse. Math. Oper. Res. **16**, 650–669 (1991)
18. Higle J., Sen, S. (1996) Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming. Kluwer, Dordrecht
19. Higle, J., Sen, S.: Statistical approximations for stochastic linear programming problems. Ann. Oper. Res. **85**, 173–192 (1999)

20. Homem-de-Mello, T.: On rates of convergence for stochastic optimization problems under non-i.i.d. sampling. SIAM J. Optim. **19**, 524–551 (2008)
21. Infanger, G.: Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs. Ann. Oper. Res. **39**, 69–95 (1992)
22. Kenyon, A., Morton, D.: Stochastic vehicle routing with random travel times. Transp. Sci. **37**, 69–82 (2003)
23. Koivu, M.: Variance reduction in sample approximations of stochastic programs. Math. Program. **103**, 463–485 (2005)
24. Kolmogorov, V.: Blossom V: a new implementation of a minimum cost perfect matching algorithm. Math. Program. Comput. **1**, 43–67 (2009)
25. Lan, G., Nemirovski, A., Shapiro, A.: Validation analysis of robust stochastic approximation method. Math. Program. (2011). doi:10.1007/s10107-011-0442-6
26. Linderoth, J., Shapiro, A., Wright, S.: The empirical behavior of sampling methods for stochastic programming. Ann. Oper. Res. **142**, 219–245 (2006)
27. Mak, W., Morton, D., Wood, R.: Monte Carlo bounding techniques for determining solution quality in stochastic programs. Oper. Res. Lett. **24**, 47–56 (1999)
28. Mehlhorn, K., Schäfer, G.: Implementation of $O(nm \log n)$ weighted matchings in general graphs: the power of data structures. J. Exp. Algorithm. **7** (2002). doi:10.1145/944618.944622
29. Norkin, V., Pflug, G., Ruszczyński, A.: A branch and bound method for stochastic global optimization. Math. Program. **83**, 425–450 (1998)
30. Partani, A.: Adaptive jacknife estimators for stochastic programming. Ph.D. thesis, The University of Texas at Austin (2007)
31. Partani, A., Morton, D., Popova, I.: Jackknife estimators for reducing bias in asset allocation. In: Proceedings of the 2006 Winter Simulation Conference, pp. 783–791 (2006)
32. Rachev, S.T.: Probability Metrics and the Stability of Stochastic Models. Wiley, New York (1991)
33. Rockafellar, R., Wets, R.B.: Variational Analysis. Springer, Berlin (1998)
34. Römisch, W. : Stability of stochastic programming problems. In: Ruszczyński, A., Shapiro, A. (eds.) Handbooks in Operations Research and Management Science, vol. 10: Stochastic Programming, pp. 483–554. Elsevier, Amsterdam (2003)
35. Ruszczyński, A.: A regularized decomposition method for minimizing a sum of polyhedral functions. Math. Program. **35**, 309–333 (1986)
36. Ruszczyński, A., Świetanowski, A.: Accelerating the regularized decomposition method for two stage stochastic linear problems. Eur. J. Oper. Res. **101**, 328–342 (1997)
37. Santoso, T., Ahmed, S., Goetschalckx, M., Shapiro, A.: A stochastic programming approach for supply chain network design under uncertainty. Eur. J. Oper. Res. **167**, 96–115 (2005)
38. Shao, J., Tu, D.: The Jackknife and Bootstrap. Springer, New York (1995)
39. Shapiro, A.: Monte Carlo sampling methods. In: Ruszczyński, A., Shapiro, A. (eds.) Handbooks in Operations Research and Management Science, vol. 10: Stochastic Programming, pp. 353–425. Elsevier, Amsterdam (2003)
40. Shapiro, A., Homem-de-Mello, T.: A simulation-based approach to two-stage stochastic programming with recourse. Math. Program. **81**, 301–325 (1998)
41. Verweij, B., Ahmed, S., Kleywegt, A., Nemhauser, G., Shapiro, A.: The sample average approximation method applied to stochastic vehicle routing problems: a computational study. Comput. Optim. Appl. **24**, 289–333 (2003)
42. Wagner, R.: Mersenne twister random number generator. http://svn.mi.fu-berlin.de/seqan/releases/seqan_1.3/seqan/random/ext_MersenneTwister.h (2009). Last accessed June 7 2012
43. Wets, R. : Stochastic programming: solution techniques and approximation schemes. In: Bachem, A., Grötschel, M., Korte, B. (eds.) Mathematical Programming: The State of the Art (Bonn 1982), pp. 560–603. Springer, Berlin (1983)