

Overlapping Batches for the Assessment of Solution Quality in Stochastic Programs

David Love^a and Güzin Bayraksan^{b*}

^a Program in Applied Mathematics, University of Arizona, Tucson, AZ 85721, USA,

`dlove@math.arizona.edu`

^b Integrated Systems Engineering, The Ohio State University, Columbus, OH 43210, USA,

`bayraksan.1@osu.edu`

Abstract

Overlapping Batch Means (OBM) has long been used in simulation as a method of reusing data to generate variance estimators with asymptotically lower variance. In this paper, we extend the OBM method to stochastic programming by formulating a variant of the multiple replications procedure used for assessing solution quality. We show conditions under which the resulting optimality gap estimators are strongly consistent and have the same asymptotically lower variance as the classical results. Computational experiments for several test problems are presented, examining the small-sample behavior of the overlapping batches method in this context.

Keywords: Sample average approximation (SAA); Overlapping batch means; Stochastic programming; Optimality gap estimation

1 Introduction

Overlapping Batch Means (OBM) method, proposed by Meketon and Schmeiser [1984], is commonly used in simulation to obtain variance estimators with lower variances. While OBM is often used for variance estimators of means, it can also be used for variance estimators of non-means [Schmeiser et al., 1990] and more recently, it has been extended to estimators based on standardized time series (e.g., area, Cramér-von Mises) [Alexopoulos et al., 2007a,b]

*Corresponding author. Tel.: +1 614-292-1695. Fax: +1 614-292-7852

and their folded versions [Meterelliyoz et al., 2012]. To the best of our knowledge, OBM has never been studied in a stochastic optimization context before. In this paper we show that the same benefits can be obtained from overlapping batches when used in a stochastic optimization setting. A natural setting where batch means are used in stochastic optimization (esp. in stochastic programming) is for assessing solution quality. Assessing solution quality in stochastic programming typically involves generating point and interval estimators of the optimality gap of a given candidate solution using Monte Carlo simulation. Because the optimality gap estimators can be difficult to analyze statistically, a common approach is to use nonoverlapping batches of data to generate a number of optimality gap estimators. These individual estimators are then used to form point and interval estimators applying the classical nonoverlapping batch means method. This is referred to as the Multiple Replications Procedure (MRP) and was initially discussed in [Mak et al., 1999]. In this paper, we formulate an overlapping batches version of MRP, allowing for partial overlap, and provide conditions under which the same variance reduction results can be achieved by overlapping like in the classical setting. We also show that the resulting point estimators are strongly consistent and the interval estimators are asymptotically valid. We present computational results to examine small-sample behavior. Our experiments indicate that variance reduction is achieved with small sample sizes, while bias and coverage probability are unaffected.

We consider a stochastic optimization problem of the form

$$z^* = \min_{x \in X} \mathbb{E}f(x, \xi), \quad (\text{SP})$$

where x is a vector of decision variables, ξ is a vector of random variables, and X is the feasible set, which is assumed to be independent of ξ . Further, it is assumed that the distribution of ξ is known and that we can sample from it. Even though we will impose a more restrictive moment condition later, we assume that $\mathbb{E}f(x, \xi)$ is well defined and finite for all $x \in X$ and (SP) has a finite optimal solution achieved on X .

Unless f has a simple structure or the number of realizations of ξ is small, (SP) typically cannot be solved exactly. An approximation of (SP) can be generated by sampling from the distribution of ξ . While other sampling schemes are possible, let $\xi^1, \xi^2, \dots, \xi^m$ be an independent and identically distributed (i.i.d.) sample from the distribution of ξ . A sampling approximation of (SP), often called the Sample Average Approximation (SAA), is given by

$$z_m^* = \min_{x \in X} \frac{1}{m} \sum_{i=1}^m f(x, \xi^i). \quad (\text{SP}_m)$$

We denote an optimal solution to (SP) as x^* and an optimal solution to (SP_{*m*}) as x_m^* .

We wish to assess the quality of a candidate solution $\hat{x} \in X$. Assessing solution quality is important in practice because as (SP) typically cannot be solved exactly, one only has an approximate solution \hat{x} without verification of its quality. Assessing solution quality is also a critical component of stopping criteria in algorithms designed to (approximately) solve (SP). We define the quality of \hat{x} by its optimality gap, $\mathbb{E}f(\hat{x}, \xi) - z^*$. The lower the optimality gap, the higher the quality of the solution and a zero optimality gap implies that the solution is optimal. Note that the candidate solution $\hat{x} \in X$ may have been found in any way. We assume that this solution is given as input to our procedures, and hence fixed, throughout this paper.

The optimality gap $\mathbb{E}f(\hat{x}, \xi) - z^*$ often cannot be calculated exactly. First, for a given candidate solution $\hat{x} \in X$, evaluation of $\mathbb{E}f(\hat{x}, \xi)$ typically involves a difficult multidimensional integral. Second, we do not know z^* . In many fields of optimization, the second problem is alleviated by obtaining lower bounds on z^* through relaxations such as integrality, Lagrangian or semidefinite relations. As a result, it is common to evaluate an upper bound on the optimality gap. This motivates estimating the upper bound on the optimality gap in stochastic programming by using Monte Carlo simulation. Given a sample size m , an upper bound on the optimality gap can be obtained by

$$\mathbb{E}f(\hat{x}, \xi) - z^* \leq \mathbb{E}f(\hat{x}, \xi) - \mathbb{E}z_m^*$$

due to the inequality $\mathbb{E}z_m^* \leq z^*$ [Mak et al., 1999; Norkin et al., 1998]. This bound improves as the sample size increases, that is, $\mathbb{E}z_m^* \leq \mathbb{E}z_{m+1}^* \leq z^*$. A straightforward estimate of $\mathbb{E}f(\hat{x}, \xi)$ is the sample mean, $\frac{1}{m} \sum_{i=1}^m f(\hat{x}, \xi^i)$. Instead of $\mathbb{E}z_m^*$, simply z_m^* can be used. The resulting point estimator of optimality gap is then

$$\frac{1}{m} \sum_{i=1}^m f(\hat{x}, \xi^i) - z_m^*.$$

Here, we assume the same observations ξ^1, \dots, ξ^m are used in both terms.

Computing the above optimality gap estimate involves solving an optimization problem (SP_{*m*}) to obtain a lower bound estimator of z^* , z_m^* , which complicates the statistical analysis. As mentioned above, to enable statistical inference, the MRP of Mak et al. [1999] generates k independent estimators, each using sample size m (k nonoverlapping batches of size m) and averages them to obtain a point estimator. To form confidence intervals, the sample variance of these estimators is calculated (we further review MRP in §2.2.) An advantage of MRP is its applicability to a wide range of problems. With i.i.d. sampling, f can be linear or nonlinear, X can include integrality constraints or not. It is also easy to implement; thus, MRP has

been applied to a wide variety of problems, including finance, supply chain network design, etc.; see e.g., [Bertocchi et al., 2000; Janjarassuk and Linderoth, 2008; Santoso et al., 2005]. Recently, the approach of (nonoverlapping) batching has been used for assessing solution quality of stochastic programs with finitely many expected value [Wang and Ahmed, 2008] and stochastic dominance [Hu et al., 2012] constraints.

Our aim is to apply the idea of overlapping batches to MRP. Overlapping could help to obtain better variance estimators with simple data reuse, resulting in improved confidence intervals. Especially for problems where sampling error is dominant, this simple data reuse method could yield better estimators. We note several differences in this setting compared to the classical setting. In simulation output analysis, the point of interest is estimating the variance of the sample mean of a covariance stationary process. We are interested in estimating the variance of an optimality gap estimator. Notice that an optimality gap estimator not only has a sample mean ($\frac{1}{m} \sum_{i=1}^m f(\hat{x}, \xi^i)$) but also a minimized sample mean (z_m^*). Minimization changes the statistical properties of sample means. For instance, the central limit theorem may not hold for a minimized sample mean, even though it holds for each $x \in X$. Therefore, analysis of the asymptotic benefits of overlapping becomes more complicated. We overcome this technical difficulty by approximating the optimality gap estimators by their nonoptimized counterparts (see §4). The nonoptimized counterparts have the desired statistical properties and we establish convergence of the optimality gap estimators that use overlapping to their nonoptimized counterparts. Another difference is, once the data is generated through a simulation, it can be reused without much additional computational effort to obtain the OBM variance estimator. In this setting, due to the solution of a sampling problem (SP_m), the computational effort may increase with data reuse. This can be alleviated in two ways. First, near-optimal variance reduction can be obtained by partially overlapping the batches [Welch, 1987; Song and Schmeiser, 1993]. Partial overlap results in a fewer number of batches, hence fewer number of optimization problems need to be solved. In this paper, we consider variable amounts of overlap and examine this both theoretically and empirically. Second, in many solution methods, warm-starting can be used to quickly solve the sampling approximations with overlapping samples, considerably reducing solution time.

The rest of the paper is organized as follows. In the next section, we review relevant background information. We first discuss the overlapping batch means in §2.1 and then briefly go over MRP in §2.2. In §3, we present the Overlapping Multiple Replications Procedure (OMRP) with variable amounts of overlap. In §4, we show strong consistency of the OMRP point estimators, asymptotic validity of the OMRP confidence intervals, and the variance reduction results as in the classical case. In §5, we test the performance of OMRP on four

problems and end in §6 with a summary and conclusions.

2 Background

2.1 Overlapping Batch Means

Consider a covariance stationary stochastic process with mean μ and variance σ^2 . The task is to estimate μ given some realization of the stochastic process $\mathbf{y} = (y^1, y^2, \dots, y^n)$. Typically, this involves forming a confidence interval of the form $[L_\alpha(\mathbf{y}), U_\alpha(\mathbf{y})]$ for a given level of significance α such that $\mathbb{P}\{L_\alpha(\mathbf{y}) \leq \mu \leq U_\alpha(\mathbf{y})\} = 1 - \alpha$ or that this probability is approximately $1 - \alpha$ for large enough sample sizes. The usual estimator for μ is the sample mean $\bar{y} = \frac{1}{n} \sum_{i=1}^n y^i$, which is an unbiased estimator even in the presence of correlated data. However, the usual variance estimator $\frac{1}{n-1} \sum_{i=1}^n (y^i - \bar{y})^2$ can be inappropriate to use in the presence of correlated data, resulting in inaccurate interval estimators. To overcome this difficulty, various variance estimators have been proposed in the simulation literature; see, e.g., [Law, 2007]. We briefly review two of these variance estimators: nonoverlapping and overlapping batch means.

Nonoverlapping batch means takes n observations, y^1, \dots, y^n of the stochastic process and splits them into k batches of size m , where $k = \frac{n}{m}$ (for simplicity, assume, for now, $n = mk$). The sample mean of each batch is then computed as $\bar{y}_j = \frac{1}{m} \sum_{i=1}^m y^{m(j-1)+i}$, $j = 1, 2, \dots, k$, and the overall sample mean $\bar{y} = \frac{1}{k} \sum_{j=1}^k \bar{y}_j = \frac{1}{n} \sum_{i=1}^n y^i$ provides a point estimator of μ . The sample variance of the batch means is calculated as

$$\widehat{\text{Var}}(\bar{y}) = \frac{m}{n} \frac{1}{k-1} \sum_{j=1}^k (\bar{y}_j - \bar{y})^2, \quad (1)$$

and the $(1 - \alpha)$ -level approximate confidence interval (CI) is formed by $\bar{y} \pm t_{k-1, \alpha/2} \sqrt{\widehat{\text{Var}}(\bar{y})}$, where $t_{k-1, \alpha/2}$ denotes the $1 - \alpha/2$ quantile of the Student's t distribution with $k - 1$ degrees of freedom. Under appropriate conditions, $n\widehat{\text{Var}}(\bar{y})$ is a consistent estimator of σ^2 .

Overlapping batch means modifies this idea by taking batches given by $\bar{y}_j(m) = \frac{1}{m} \sum_{i=1}^m y^{(j-1)+i}$, for $j = 1, \dots, n - m + 1$. As an example, suppose each batch contains m observations and that there are a total of $n = 2m$ observations. In the nonoverlapping case, there are only two batches: the first batch contains observations y^1, y^2, \dots, y^m and the second batch contains observations $y^{m+1}, y^{m+2}, \dots, y^{2m}$. In the overlapping batches of Meketon and Schmeiser [1984], the first batch is as before, the second batch now consists of y^2, y^3, \dots, y^{m+1} , the third batch contains observations y^3, y^4, \dots, y^{m+2} and so on. This is what we call the

maximally-overlapping batch means method, because $m - 1$ observations are common to—and only one observation changes between—adjacent batches.

The sample variance estimator is updated to

$$\widetilde{\text{Var}}(\bar{y}) = \frac{1}{\left(\frac{n}{m} - 1\right)(n - m + 1)} \sum_{j=1}^{n-m+1} (\bar{y}_j(m) - \bar{y})^2. \quad (2)$$

Note that the corresponding batch means are no longer independent. However, by simply reusing the data in this fashion—and seemingly in a contradictory way that induces dependence—one obtains a better variance estimator. Meketon and Schmeiser [1984] show that: (i) the overlapping variance estimator has nearly the same bias as the standard nonoverlapping variance estimator, and (ii) the overlapping estimator has only two-thirds of the asymptotic variance of the nonoverlapping one. That is,

$$\frac{\text{Var}\left(\widetilde{\text{Var}}(\bar{y})\right)}{\text{Var}\left(\widehat{\text{Var}}(\bar{y})\right)} \rightarrow \frac{2}{3} \quad (3)$$

in the limit as batch size m and the number of batches (n/m) tend to infinity. Note that the degrees of freedom in (2) is slightly different than in [Meketon and Schmeiser, 1984]. The degrees of freedom in (2) makes $\widetilde{\text{Var}}(\bar{y})$ an unbiased estimator for i.i.d. data for all finite m and n with the same asymptotic benefits [Song and Schmeiser, 1993]. The $(1 - \alpha)$ -level approximate CI is similarly formed by $\bar{y} \pm t_{3(k-1)/2, \alpha/2} \sqrt{\widetilde{\text{Var}}(\bar{y})}$, with an adjustment to the degrees of freedom.

Welch [1987] established the relationship between OBM and spectral estimators and also considered variable amounts of overlap, see also [Song and Schmeiser, 1993]. We defer this discussion on variable overlap to §3.1 and continue with a brief review of MRP.

2.2 Multiple Replications Procedure

Taking n observations of the random variables ξ and splitting them into k nonoverlapping batches of size m , define

$$\begin{aligned} \bar{G}_j &= \frac{1}{m} \sum_{i=1}^m f(\hat{x}, \xi^{m(j-1)+i}) - \min_{x \in X} \frac{1}{m} \sum_{i=1}^m f(x, \xi^{m(j-1)+i}), \quad j = 1, 2, \dots, k \\ &= \frac{1}{m} \sum_{i=1}^m [f(\hat{x}, \xi^{m(j-1)+i}) - f(x_j^*, \xi^{m(j-1)+i})], \quad j = 1, 2, \dots, k, \end{aligned} \quad (4)$$

where x_j^* denotes an optimal solution to (SP_m) of batch j given on the right-hand side of (4). Since we are assessing the quality of a given solution $\hat{x} \in X$, we suppress it from the notation \bar{G}_j . The optimality gap estimator \bar{G}_j is similar to \bar{y}_j with individual observations $y^{m(j-1)+i} = f(\hat{x}, \xi^{m(j-1)+i}) - f(x_j^*, \xi^{m(j-1)+i})$, for $i = 1, \dots, m$ and $j = 1, \dots, k$. Notice that, in this case, y^i not only depends on observation i of batch j , $\xi^{m(j-1)+i}$, but all the observations in that batch $\xi^{m(j-1)+1}, \xi^{m(j-1)+2}, \dots, \xi^{m(j-1)+m}$ because x_j^* is the optimal solution to (SP_m) given in (4). As before, after k estimates are obtained, the overall mean of these estimates $\bar{\bar{G}} = \frac{1}{k} \sum_{j=1}^k \bar{G}_j$ provides a point estimator of the optimality gap. The sample variance is obtained as in (1) by $\widehat{\text{Var}}(\bar{\bar{G}}) = \frac{1}{k} \frac{1}{k-1} \sum_{j=1}^k (\bar{G}_j - \bar{\bar{G}})^2$, which results in a one-sided CI $\left[0, \bar{\bar{G}} + t_{k-1, \alpha} \sqrt{\widehat{\text{Var}}(\bar{\bar{G}})}\right]$ that has a level of significance, which is approximately α .

Notice that the minimization in the second term on the right-hand side of (4) gives rise to a biased gap estimator (recall the upper bound on the optimality gap). Thus, we can expect that the true probability of the optimality gap residing within the confidence interval to be greater than the $1 - \alpha$ suggested by the above calculation. This is shown empirically in [Bayraksan and Morton, 2006], which also investigates additional methods for using a smaller number of replications (e.g., 1 or 2) with an alternative variance estimator to compute a confidence interval. See also [Bayraksan and Morton, 2009; Partani et al., 2006; Partani, 2007] for variations of MRP aimed to reduce bias and variance.

3 Overlapping Multiple Replications Procedure

We begin our discussion with variably overlapping batches and then define the estimators of OMRP.

3.1 Variably Overlapping Batches

As before, let m denote the batch size, n the total sample size and $k = \lfloor \frac{n}{m} \rfloor$ be the number of nonoverlapping batches. In this paper, we use the *batch nonoverlap parameter* $1 \leq \gamma \leq m$ to denote how much neighboring batches do *not* overlap. For instance, $\gamma = m$ corresponds to the classical case of nonoverlapping batches and $\gamma = 1$ corresponds to the maximally overlapping case of Meketon and Schmeiser [1984]. The sample mean of each batch estimator is calculated similarly, $\bar{y}_j = \frac{1}{m} \sum_{i=1}^m y^{\gamma(j-1)+i}$, $j = 1, 2, \dots, \lfloor \frac{n-m}{\gamma} \rfloor + 1$, where $\lfloor \frac{n-m}{\gamma} \rfloor + 1$ is the number of batches used given n , m and γ . The sample variance estimator given in (2)

is changed to

$$\widetilde{\text{Var}}_\gamma(\bar{y}) = \frac{1}{\left(\frac{n}{m} - 1\right) \left(\left\lfloor \frac{n-m}{\gamma} \right\rfloor + 1\right)} \sum_{j=1}^{\left\lfloor \frac{n-m}{\gamma} \right\rfloor + 1} (\bar{y}_j - \bar{y})^2. \quad (5)$$

Note that with $\gamma = 1$, (5) reduces to (2) and with $\gamma = m$ and $n = mk$, (5) reduces to (1).

The amount of asymptotic variance reduction in this estimator depends on the asymptotic ratio of γ/m , which we denote by $\bar{\gamma}$. For example, when $\gamma = 1$ (or $\bar{\gamma} = 0$; the maximally overlapping case), the variance is reduced to two-thirds (66.67%) of the original nonoverlapping case, as given in (3). When 75% of observations overlap (i.e., when $\gamma = m/4$ or $\bar{\gamma} = 1/4$), the variance is 33/48th of the original (68.75%), which is near-optimal. When only half of the observations overlap ($\bar{\gamma} = 1/2$) the variance is 75% of original. In general, from the spectral analysis given in [Welch, 1987], the variance is reduced to

$$\bar{\gamma} \left(1 + 2 \sum_{j=1}^{\left\lfloor \frac{1}{\bar{\gamma}} \right\rfloor - 1} (1 - j\bar{\gamma})^2 \right) \quad (6)$$

of the nonoverlapping case. If $\frac{1}{\bar{\gamma}}$ is an integer, (6) simplifies to $\frac{2+\bar{\gamma}^2}{3}$.

A $(1 - \alpha)$ -level approximate CI on the mean can be formed using variably overlapping batches by $\bar{y} \pm t_{d_\gamma(k-1), \alpha/2} \sqrt{\widetilde{\text{Var}}_\gamma(\bar{y})}$, where the degrees of freedom increase d_γ is the multiplicative inverse of (6) [Welch, 1987]. If $\frac{1}{\bar{\gamma}}$ is an integer, $d_\gamma = \frac{3}{2+\bar{\gamma}^2}$. For maximally overlapping batches, the degrees of freedom is $d_\gamma = \frac{3}{2}$.

3.2 Definition of Estimators for OMRP

In order to apply overlapping batches to MRP, we need to keep track of solutions to sampling problems (SP_m) for each batch of size m . Toward this end, let $B(i)$ denote the set of batches $j \in \left\{1, 2, \dots, \left\lfloor \frac{n-m}{\gamma} \right\rfloor + 1\right\}$ observation ξ^i is used in, $i = 1, 2, \dots, n$. See Figure 1 for an example with $n = 18$, $m = 6$ and $\gamma = 2$. Here, the first batch uses observations $\xi^1, \xi^2, \xi^3, \xi^4, \xi^5, \xi^6$, the second batch uses $\xi^3, \xi^4, \xi^5, \xi^6, \xi^7, \xi^8$, and so on. So, $B(1) = \{1\}$, $B(3) = \{1, 2\}$, and $B(7) = \{2, 3, 4\}$. As before, we use x_j^* to denote an optimal solution to sampling problem (SP_m) formed using the j th batch; $x_j^* \in \operatorname{argmin}_{x \in X} \frac{1}{m} \sum_{i=1}^m f(x, \xi^{\gamma(j-1)+i})$, $j = 1, 2, \dots, \left\lfloor \frac{n-m}{\gamma} \right\rfloor + 1$.

The results on overlapping batch means occur in the limit as $n, m, k = n/m \rightarrow \infty$ [Damerdj, 1994, 1995; Meketon and Schmeiser, 1984; Song and Schmeiser, 1993; Welch, 1987]. Let n_l, m_l and k_l be sequences of numbers satisfying these requirements, then the

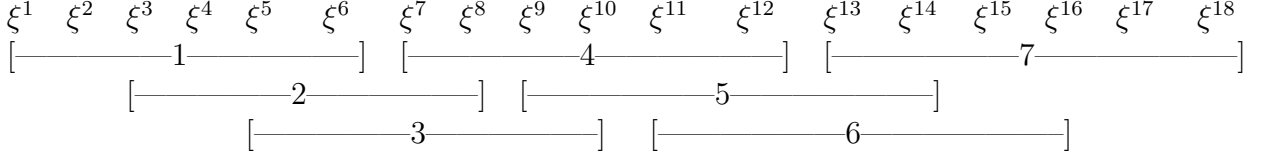


Figure 1: Visual representation of overlapping batches with $n = 18$, $m = 6$ and $\gamma = 2$. The brackets show which observations are used in each batch, and the numbers inside each bracket show the batch number j .

limits will be taken as $l \rightarrow \infty$. We assume a batching structure where $m_l = \lfloor n_l^r \rfloor$ for some $0 < r < 1$ as is typical in simulation, and n_l tends to infinity as $l \rightarrow \infty$. In addition, we may desire that the batch nonoverlap parameter change with l , to ensure that $\gamma_l/m_l = \bar{\gamma}$ converges to the constant $\bar{\gamma}$. Now we are ready to define the estimators for OMRP.

$$\bar{G}_{lj} = \frac{1}{m_l} \sum_{i=1}^{m_l} [f(\hat{x}, \xi^{\gamma(j-1)+i}) - f(x_j^*, \xi^{\gamma(j-1)+1})], \quad j = 1, 2, \dots, \left\lfloor \frac{n_l - m_l}{\gamma} \right\rfloor + 1, \quad (7)$$

$$\bar{G}_l = \frac{1}{n_l} \sum_{i=1}^{n_l} \frac{1}{|B(i)|} \sum_{j \in B(i)} [f(\hat{x}, \xi^i) - f(x_j^*, \xi^i)], \quad (8)$$

$$VG_l = \frac{1}{\left(\frac{n_l}{m_l} - 1\right) \left(\left\lfloor \frac{n_l - m_l}{\gamma} \right\rfloor + 1\right)} \sum_{j=1}^{\left\lfloor \frac{n_l - m_l}{\gamma} \right\rfloor + 1} (\bar{G}_{lj} - \bar{G}_l)^2. \quad (9)$$

The optimality gap estimator for each batch \bar{G}_{lj} is defined like (4) for general values of the nonoverlap parameter γ_l . We just removed the minimization in (4) and used directly an optimal solution of x_j^* of batch j . The overall mean, \bar{G}_l , is defined a little differently. Here, \bar{G}_l still uses each observation $i = 1, 2, \dots, n_l$ but also makes use of *all* the information collected throughout the batches. That is, if observation ξ^i is used in $|B(i)|$ number of batches, all the optimal solutions x_j^* corresponding to each batch $j \in B(i)$ are used for the lower bound estimator. Then, VG_l is defined in a similar fashion for the variably overlapping batches variance estimator (5).

4 Theoretical Results

In this section, we provide conditions under which several asymptotic properties hold. First, we show the consistency of the point estimators of optimality gap, \bar{G}_l , and variance, $n_l VG_l$. Then, we provide conditions under which the same benefits on variance reduction are observed in the stochastic programming setting by overlapping. Finally, we show the asymptotic validity of the OMRP interval estimator.

4.1 Assumptions

We make the following assumptions:

A1 Samples of the random variable ξ are i.i.d.

A2 $\sup_{x \in X} \mathbb{E}f(x, \xi)^4 < \infty$.

A3 (SP) has a unique optimal solution x^* and $\mathbb{P}\{x_m^* \neq x^*\} \rightarrow 0$ exponentially fast as $m \rightarrow \infty$. That is, there exists constants $M > 0$ and $c > 0$ such that $\log(\mathbb{P}\{x_m^* \neq x^*\}) \leq -cm$ for all $m > M$.

First, we restrict our attention to i.i.d. sampling by A1. The classical variance-reduction results for overlapping batches show that $\text{Var}(n_l V G_l)$ is reduced by overlapping, thus requiring at least finite fourth moments. Finally, our proof relies on an exponential rate of convergence of sampled solution to the true solution of the problem. Assumption A3 can be satisfied for different classes of problems with unique optimal solutions. Shapiro and Homem-de Mello [2000] establish exponential rates of convergence for a class of (SP), which include two-stage stochastic linear programs with recourse when the support of ξ is finite. Kleywegt et al. [2002] show that assumption A3 can be satisfied for discrete stochastic programs (i.e., when X is a finite set) under the condition that the moment generating function of $f(x^*, \xi) - f(x, \xi)$ for $x \in X \setminus \{x^*\}$ is finite valued on \mathbb{R} . This class of problems includes stochastic integer programs with recourse. We present computational results on both stochastic linear and integer programs with recourse in §5.

4.2 Nonoptimized Counterparts

The internal optimization in the batches in (7) makes a straightforward statistical analysis of the behavior of the estimators difficult. To overcome this problem, we introduce the following unbiased optimality gap estimators:

$$\bar{D}_{lj} = \frac{1}{m_l} \sum_{i=1}^{m_l} [f(\hat{x}, \xi^{\gamma^{(j-1)+i}}) - f(x^*, \xi^{\gamma^{(j-1)+i}})], \quad j = 1, 2, \dots, \left\lfloor \frac{n_l - m_l}{\gamma} \right\rfloor + 1, \quad (10)$$

$$\bar{D}_l = \frac{1}{n_l} \sum_{i=1}^{n_l} [f(\hat{x}, \xi^i) - f(x^*, \xi^i)], \quad (11)$$

$$V D_l = \frac{1}{\left(\frac{n_l}{m_l} - 1\right) \left(\left\lfloor \frac{n_l - m_l}{\gamma} \right\rfloor + 1\right)} \sum_{j=1}^{\left\lfloor \frac{n_l - m_l}{\gamma} \right\rfloor + 1} (\bar{D}_{lj} - \bar{D}_l)^2. \quad (12)$$

These are essentially the same as the variably overlapping batches estimators in §3.1 with $y^i = f(\hat{x}, \xi^i) - f(x^*, \xi^i)$. These are also defined identically to the original estimators (7)–(9) with the exception that x_j^* from (7) and (8) is replaced by an optimal solution x^* in (10) and (11). With \hat{x} and x^* fixed, estimators (10)–(12) have the same statistical properties as variably overlapping batches estimators in §3.1. Note that the optimal solution x^* is not known. However, the estimators (10)–(12) are used only to show convergence of (7)–(9); they are not necessary for carrying out the OMRP algorithm. Throughout the rest of this section, we assume that the optimality gap estimators (7)–(9) and their nonoptimized counterparts (10)–(12) use identical samples and values of n_l , m_l , and γ_l .

4.3 Consistency

We begin with two lemmas that establish exponential rates of convergence of certain intermediate quantities.

Lemma 1. *Under assumptions A1 and A3, $\mathbb{P} \left\{ \bar{\bar{D}}_l - \bar{\bar{G}}_l \neq 0 \right\} \rightarrow 0$ exponentially fast as $l \rightarrow \infty$.*

Proof. Let E^C denote the complement of event E . Note that

$$\begin{aligned} \mathbb{P} \left\{ \bar{\bar{D}}_l - \bar{\bar{G}}_l \neq 0 \right\} &\leq \mathbb{P} \left\{ \left(\left(\bigcap_{j=1}^{\lfloor \frac{n_l - m_l}{\gamma_l} \rfloor + 1} \{x_j^* = x^*\} \right)^C \right) \right\} \\ &= \mathbb{P} \left\{ \bigcup_{j=1}^{\lfloor \frac{n_l - m_l}{\gamma_l} \rfloor + 1} \{x_j^* \neq x^*\} \right\} \\ &\leq \left(\left\lfloor \frac{n_l - m_l}{\gamma_l} \right\rfloor + 1 \right) \mathbb{P} \{x_{m_l}^* \neq x^*\}. \end{aligned}$$

With a batching structure in which $m_l = \lfloor n_l^r \rfloor$ for some $0 < r < 1$, $\gamma_l \geq 1$, and by A3, the desired result follows. \square

Lemma 2. *Under assumptions A1–A3, both (i) $\mathbb{E}(\bar{D}_{lj} - \bar{G}_{lj})^4 \rightarrow 0$ and (ii) $\mathbb{E}(\bar{\bar{D}}_l - \bar{\bar{G}}_l)^4 \rightarrow 0$ exponentially fast as $l \rightarrow \infty$.*

Proof. For brevity, we drop the lj subscripts from \bar{D}_{lj} and \bar{G}_{lj} and l from $\bar{\bar{G}}_l$ and $\bar{\bar{D}}_l$. (i) We can decompose the expectation as

$$\begin{aligned} \mathbb{E}(\bar{D} - \bar{G})^4 &= \mathbb{E} \left[(\bar{D} - \bar{G})^4 \mid \bar{D} - \bar{G} \neq 0 \right] \mathbb{P} \{ \bar{D} - \bar{G} \neq 0 \} + 0 \\ &\leq \mathbb{E} \left[(\bar{D} - \bar{G})^4 \mid \bar{D} - \bar{G} \neq 0 \right] \mathbb{P} \{ x_{m_l}^* \neq x^* \}. \end{aligned}$$

The expectation term on the right-hand side is bounded by assumption A2 for i.i.d. samples. Assumption A3 states $\mathbb{P}\{x_{m_l}^* \neq x^*\} \rightarrow 0$ exponentially fast. Thus $\mathbb{E}(\bar{D} - \bar{G})^4 \rightarrow 0$ exponentially fast. (ii) This expectation term can be decomposed in the same manner

$$\mathbb{E}(\bar{D} - \bar{G})^4 = \mathbb{E} \left[(\bar{D} - \bar{G})^4 \mid \bar{D} - \bar{G} \neq 0 \right] \mathbb{P} \left\{ \bar{D} - \bar{G} \neq 0 \right\}.$$

With the expectation term on the right-hand side again bounded by A2, and the probability term converges to zero exponentially fast by Lemma 1. \square

We are now ready to show that the resulting point estimators are strongly consistent. We use the abbreviation a.s. to denote almost sure convergence. We also use $\mu_{\hat{x}} := \mathbb{E}[f(\hat{x}, \xi) - f(x^*, \xi)]$ to denote the optimality gap of \hat{x} and $\sigma_{\hat{x}}^2 := \text{Var}(f(\hat{x}, \xi) - f(x^*, \xi))$ to denote the associated variance term.

Theorem 1. *Under assumptions A1–A3, (i) $\bar{G}_l - \bar{D}_l \rightarrow 0$, a.s., and (ii) $n_l V G_l - n_l V D_l \rightarrow 0$, a.s. as $l \rightarrow \infty$. Therefore, \bar{G}_l is a consistent estimator of $\mu_{\hat{x}}$, and if $n_l V D_l$ is a consistent estimator of $\sigma_{\hat{x}}^2$, then $n_l V G_l$ is a consistent estimator of $\sigma_{\hat{x}}^2$.*

Proof. (i) By Lemma 1, $\mathbb{P}\{\bar{D}_l - \bar{G}_l \neq 0\} \rightarrow 0$ exponentially fast. Thus $\sum_l \mathbb{P}\{\bar{D}_l - \bar{G}_l \neq 0\} < \infty$. The desired result follows by the Borel-Cantelli Lemma. (ii) This proof is nearly identical to above. Note that

$$\mathbb{P}\{n_l V D_l - n_l V G_l \neq 0\} \leq \mathbb{P} \left\{ \left(\left(\bigcap_{j=1}^{\lfloor \frac{n_l - m_l}{\gamma_l} \rfloor + 1} \{x_j^* = x^*\} \right)^c \right) \right\}$$

and by the proof of Lemma 1, $\mathbb{P}\{n_l V D_l - n_l V G_l \neq 0\} \rightarrow 0$ exponentially fast, and thus $n_l V G_l$ is a consistent estimator of $\sigma_{\hat{x}}^2$, if $n_l V D_l$ is, by the Borel-Cantelli Lemma. \square

Part (i) of Theorem 1 provides conditions under which the point estimator \bar{G} of OMRP is a consistent estimator of the optimality gap, by establishing almost sure convergence of $\bar{G}_l - \bar{D}_l$ to 0 as $l \rightarrow \infty$. Notice that under assumption A1, $\bar{D}_l \rightarrow \mu_{\hat{x}}$, a.s. by the strong law of large numbers; as a result, $\bar{G} \rightarrow \mu_{\hat{x}}$, a.s. Similarly, part (ii) of Theorem 1 establishes almost sure convergence of $n_l V G_l - n_l V D_l$ to 0 as $l \rightarrow \infty$. As a result, if $n_l V D_l$ is a consistent estimator of $\sigma_{\hat{x}}^2$, so is $n_l V G_l$. Strong consistency of the variance estimator $n_l V D_l$ have been studied in [Damerdjji, 1994] for the cases of nonoverlapping and maximally overlapping batches. Under the additional moment assumption

A2' $\exists \epsilon > 0$ such that $\sup_{x \in X} \mathbb{E} f(x, \xi)^{4+\epsilon} < \infty$,

for a sampling scheme of $m_l = \lfloor n_l^r \rfloor$, Corollaries 3.4 and 4.2 of [Damerdji, 1994] show that $n_l V D_l$ is consistent if $\frac{2}{4+\epsilon} < r < 1$ for nonoverlapping batches and $\frac{2}{4+\epsilon} < r < \frac{1}{2}$ for maximally overlapping batches.

4.4 Variance Reduction

Meketon and Schmeiser [1984] show that $\frac{n_l}{m_l} \text{Var}(n_l V D_l) \rightarrow \alpha \sigma_{\hat{x}}^4$ where $\alpha = \frac{4}{3}$ for maximally overlapping batches and $\alpha = 2$ for nonoverlapping batches, resulting in the decrease of variance for the maximally overlapping batches estimator. Welch [1987] expands this result to variable amounts of overlap, showing that α is twice the value of (6). Damerdji [1995] shows that convergence of $\frac{n_l}{m_l} \text{Var}(n_l V D_l)$ to the appropriate $\alpha \sigma_{\hat{x}}^4$ —thus the mean-square consistency of $n_l V D_l$ —is guaranteed for $\frac{1}{2} + \frac{1}{4+\epsilon} < r < 1$ for both nonoverlapping and maximally overlapping batches under the batching structure $m_l = \lfloor n_l^r \rfloor$ and $4 + \epsilon$ finite moments assumption A2'. (With A2, the range of r that ensures mean-square consistency can be found by setting $\epsilon = 0$.)

To prove the variance reduction results in our setting, we wish to show that

$$\left(\frac{n_l}{m_l}\right) (\text{Var}(n_l V G_l) - \text{Var}(n_l V D_l)) \rightarrow 0,$$

for which it suffices to show $\sqrt{n_l/m_l} (n_l V G_l - n_l V D_l) \xrightarrow{L^2} 0$, where $\xrightarrow{L^2}$ denotes mean-square convergence.

Theorem 2. *Suppose assumptions A1–A3 hold, $\bar{\gamma} = \frac{\gamma_l}{m_l}$ is constant for all l , and the batching structure ensures $\frac{n_l}{m_l} \text{Var}(n_l V D_l) \rightarrow \alpha \sigma_{\hat{x}}^4$ where α is the twice the value of (6). Then, $\frac{n_l}{m_l} \text{Var}(n_l V G_l) \rightarrow \alpha \sigma_{\hat{x}}^4$ as $l \rightarrow \infty$ with the same α .*

Proof. For simplicity of exposition, we drop the subscript l and the $\frac{n_l}{m_l}$ for now and add them in later. Further, we use abbreviations $n_b = \lfloor \frac{n-m}{\gamma} \rfloor + 1$ to denote the number of batches and $df = \left(\frac{n}{m} - 1\right) \left(\lfloor \frac{n-m}{\gamma} \rfloor + 1\right)$ to denote the denominator term in $V G_l$ and $V D_l$. Note that

$$\begin{aligned} \mathbb{E}(nVG - nVD)^2 &= \left(\frac{n}{df}\right)^2 \mathbb{E}\left(\sum_{j=1}^{n_b} [(\bar{G}_j - \bar{G})^2 - (\bar{D}_j - \bar{D})^2]\right)^2 \\ &\leq \left(\frac{n}{df}\right)^2 \mathbb{E} n_b \sum_{j=1}^{n_b} [(\bar{G}_j - \bar{G})^2 - (\bar{D}_j - \bar{D})^2]^2 \end{aligned} \quad (13)$$

by the Cauchy-Schwartz inequality. By examining the batch structure in Figure 1 and the definition of \bar{G} in (8) one can see that the terms of the expectation are not identically

distributed. We will separate these into identically distributed “center” batches and “fringe” batches. For example, in Figure 1, only batch 4 is a center batch, while the others are fringe batches. Notice in Figure 1, we see $\lceil \bar{\gamma}^{-1} \rceil = 3$ sets of nonoverlapping batches $\{1, 4, 7\}$, $\{2, 5\}$, and $\{3, 6\}$, each of which has two fringe batches. In general, there are $\lceil \bar{\gamma}^{-1} \rceil$ sets of nonoverlapping batches, each of which will have two fringe batches, and $n_b - 2 \lceil \bar{\gamma}^{-1} \rceil$ center batches.

The fringe batches will come in identically distributed pairs. The first and last batch would be identically distributed, as would the second and second-to-last, etc. We know this because we could reverse the order of the i.i.d. sample and perform the batching again to get the exact same batch structure. Rewriting to take this into account, letting $A_j = (\bar{G}_j - \bar{G})^2 - (\bar{D}_j - \bar{D})^2$, and continuing from (13),

$$\begin{aligned}
& \left(\frac{n}{df} \right)^2 n_b \mathbb{E} \sum_1^{n_b} \left[(\bar{G}_j - \bar{G})^2 - (\bar{D}_j - \bar{D})^2 \right]^2 \\
&= \left(\frac{n}{df} \right)^2 n_b \mathbb{E} \left[\sum_{j=1}^{\lceil \bar{\gamma}^{-1} \rceil} [A_j]^2 + \sum_{j=\lceil \bar{\gamma}^{-1} \rceil+1}^{n_b - \lceil \bar{\gamma}^{-1} \rceil} [A_j]^2 + \sum_{j=n_b - \lceil \bar{\gamma}^{-1} \rceil + 1}^{n_b} [A_j]^2 \right] \\
&= \left(\frac{n}{df} \right)^2 n_b \sum_{j=\lceil \bar{\gamma}^{-1} \rceil + 1}^{n_b - \lceil \bar{\gamma}^{-1} \rceil} \mathbb{E} [A_j]^2 + \left(\frac{n}{df} \right)^2 (n_b) 2 \sum_{j=1}^{\lceil \bar{\gamma}^{-1} \rceil} \mathbb{E} [A_j]^2 \\
&= \left(\frac{n}{df} \right)^2 n_b (n_b - 2 \lceil \bar{\gamma}^{-1} \rceil) \mathbb{E} [A_j]^2 + \left(\frac{n}{df} \right)^2 (n_b) 2 \sum_{j=1}^{\lceil \bar{\gamma}^{-1} \rceil} \mathbb{E} [A_j]^2.
\end{aligned}$$

We have $\lceil \bar{\gamma}^{-1} \rceil + 1$ expectation terms, each of which has a coefficient that is less than $\left(\frac{n}{df} \right)^2 n_b^2 < (2m)^2$. So each expectation can be bounded by

$$\begin{aligned}
(2m)^2 \mathbb{E} [A_j]^2 &= (2m)^2 \mathbb{E} \left[(\bar{G}_j - \bar{G}) - (\bar{D}_j - \bar{D}) \right]^2 \left[(\bar{G}_j - \bar{G}) + (\bar{D}_j - \bar{D}) \right]^2 \\
&\leq (2m)^2 \sqrt{\mathbb{E} \left[(\bar{G}_j - \bar{G}) - (\bar{D}_j - \bar{D}) \right]^4} \sqrt{\mathbb{E} \left[(\bar{G}_j - \bar{G}) + (\bar{D}_j - \bar{D}) \right]^4} \quad (14)
\end{aligned}$$

$$\leq (2m)^2 \sqrt{\mathbb{E} \left[(\bar{G}_j - \bar{G}) - (\bar{D}_j - \bar{D}) \right]^4} \sqrt{8 \left(\mathbb{E} (\bar{G}_j - \bar{G})^4 + \mathbb{E} (\bar{D}_j - \bar{D})^4 \right)}, \quad (15)$$

where equation (14) follows from the Cauchy-Schwartz inequality, and equation (15) follows from Hölder’s inequality. Since we can bound both expectation terms in (15) by A2, we can demonstrate $\sqrt{\frac{n}{m}} (nVG - nVD) \xrightarrow{L^2} 0$ by focusing on only one of the terms. Focusing on the

first expectation term in (15) and returning the $\frac{n}{m}$, by Hölder's inequality,

$$\frac{n}{m}4m^2\mathbb{E}\left[(\bar{G}_j - \bar{G}) - (\bar{D}_j - \bar{D})\right]^4 \leq 32\frac{n}{m}m^2\left(\mathbb{E}(\bar{D}_j - \bar{G}_j)^4 + \mathbb{E}(\bar{D} - \bar{G})^4\right).$$

By Lemma 2, both expectation terms above converge to zero with exponential rates. With a batching structure $m_l = \lfloor n_l^r \rfloor$ for some $0 < r < 1$, we see that $\frac{n_l}{m_l}\mathbb{E}(n_lVG_l - n_lVD_l)^2 \rightarrow 0$. \square

Theorem 2 shows conditions under which the same asymptotic variance reduction results are achieved for the OMRP optimality gap variance estimator. We will empirically examine the variance reduction for small sample sizes in §5.

4.5 Asymptotic Validity of the Confidence Intervals

The overlapping variance estimator for stochastic programs displays the same desirable properties as in simulation, but we must still check that it results in a valid confidence interval. Let $I_l^G = \bar{G}_l + t_{d_l, \alpha}\sqrt{VG_l}$ be the width of the one-sided confidence interval generated by OMRP. The degrees of freedom d_l is the inverse of (6) using $\frac{n_l}{m_l}$ instead of $\bar{\gamma}$.

Theorem 3. *Suppose assumptions A1–A3 hold and the batching structure ensures n_lVD_l converges in probability to σ_x^2 . Then, for $\hat{x} \neq x^*$, $\lim_{l \rightarrow \infty} \mathbb{P}\{\mu_{\hat{x}} \leq I_l^G\} = 1 - \alpha$.*

Proof. Note that for $\hat{x} = x^*$, $\lim_{l \rightarrow \infty} \mathbb{P}\{\mu_{\hat{x}} \leq I_l^G\} = 1$. With $\hat{x} \neq x^*$, let $X_l = \frac{\bar{D}_l - \mu_{\hat{x}}}{\sigma_x^2/\sqrt{n_l}}$, $Y_l = \frac{\bar{G}_l - \mu_{\hat{x}}}{\sigma_x^2/\sqrt{n_l}}$, and $Z_l = \frac{\sigma_x^2}{\sqrt{n_lVD_l}} \frac{\sqrt{n_lVD_l}}{\sqrt{n_lVG_l}}$. It suffices to show $Y_lZ_l = \frac{\bar{G}_l - \mu_{\hat{x}}}{\sqrt{VG_l}} \Rightarrow N(0, 1)$, where \Rightarrow denotes convergence in distribution. First, we show that $|X_l - Y_l| \rightarrow 0$, a.s. Note that $|X_l - Y_l| > 0$ if and only if $|\bar{D}_l - \bar{G}_l| > 0$. In the proof of Theorem 1 we showed that $\{\bar{D}_l - \bar{G}_l \neq 0\}$ happens only finitely many times, so $|X_l - Y_l| \rightarrow 0$, a.s. by the Borel-Cantelli Lemma. Next, since $X_l \Rightarrow N(0, 1)$ by the central limit theorem and $|X_l - Y_l| \rightarrow 0$ in probability, $Y_l \Rightarrow N(0, 1)$. By hypothesis and Theorem 1, $Z_l \rightarrow 1$ in probability. Thus, by Slutsky's theorem, $Y_lZ_l \Rightarrow N(0, 1)$. \square

The asymptotic validity of the OMRP confidence interval requires n_lVD_l to converge in probability to σ_x^2 . This can be achieved either by almost sure or mean-square convergence, which gives a larger range of r in the batching structure.

5 Computational Results

In the previous section, we proved asymptotic results regarding the OMRP estimators. In this section, we apply OMRP to several test problems and examine small-sample behavior.

We begin our discussion by introducing the test problems used for numerically evaluating OMRP in §5.1, followed by our experimental setup in §5.2. Then, we present and discuss the numerical results in §5.3.

5.1 Test Problems

We empirically tested the effectiveness of OMRP on four test problems. All test problems are two-stage stochastic programs with recourse and were selected to study the behavior of OMRP under a variety of different situations—we will shortly discuss this below. The characteristics of these problems are given in Table 1. The first test problem is a newsvendor problem with demand distributed uniformly on $\{0, 0.0001, 0.0002, \dots, 10\}$; we denote it as NVD. The instance of the newsvendor problem can be found in [Bayraksan and Morton, 2006], except here we used a discrete uniform demand distribution. The second is a capacity expansion planning problem, denoted CEP1, with random demand vector ξ of dimension 3, with 216 total realizations [Higle and Sen, 1996]. The third is a stochastic knapsack problem denoted 10D, with 10 stochastic parameters and 10 binary decision variables [Kleywegt et al., 2002]. The final problem is a stochastic vehicle routing problem, denoted DB1, with 46 stochastic parameters and a total of 4.5×10^{25} realizations [Donohue and Birge, 1995].

Problem	# of 1st Stage Variables	# of 2nd Stage Variables	# of Stochastic Parameters	# of Scenarios
NVD	1	1	1	$10^5 + 1$
CEP1	8	15	3	216
10D	10	1	10	∞
DB1	5	102	46	4.5×10^{25}

Table 1: Test problem characteristics

We made several considerations in selecting the test problems. NVD can be solved very quickly, which allows for solving a large number of independent runs for examining the small-sample behavior of OMRP. CEP1 is also easy to solve; the optimal solution x^* can be obtained with high probability even for small sample sizes. This feature shows the behavior of VG when the gap CI width is determined mostly by sampling error, rather than bias. 10D demonstrates the behavior of OMRP when $|X|$ is finite. Furthermore, 10D has a high condition number [Kleywegt et al., 2002], making it unlikely that the optimal solution x^* is found when using a sampling solution, thus demonstrating the behavior of OMRP when the bias of the gap is significant. The optimal solutions of NVD, CEP1, and 10D are known, allowing for explicitly checking and verifying the results. In contrast, DB1 is a large-scale problem with an unknown optimal solution. This final test problem allows us to investigate how OMRP works on large-scale problems.

The candidate solutions used for NVD and CEP1 and their characteristics can be found in [Bayraksan and Morton, 2006]. We used the candidate solution $(0, 0, 1, 1, 1, 0, 1, 1, 0, 1)$ for 10D with $\mu_{\hat{x}} = 158.579$ and the candidate solution $(11, 14, 8, 11, 7)$ for DB1 with an estimated $\mu_{\hat{x}} = 1.08$. Because the optimal solution of DB1 is unknown, a Latin hypercube sample of size 50,000 was used to estimate $\mu_{\hat{x}}$, which was then used below to estimate the coverage probability of the OMRP CIs for this test problem.

5.2 Experimental Setup

Our experiments were conducted as follows. We used a total sample size of $n = 30m$, with several values of the batch size $m = 100, 200, \dots$. Then, at each value of m , we formed the OMRP estimators with varying amounts of overlap. We also set $\alpha = 0.10$, forming 90% confidence intervals on the optimality gap using the ORMP estimators. Results for NVD were compiled over 10,000 independent runs, while the CEP1 and 10D results were taken over 1,000 independent runs, and DB1 results were generated over 500 independent runs.

We solved the sampling approximations of the test problems as follows. While NVD can be written as a two-stage stochastic linear program with recourse, it can be solved quickly by sorting the demand values. We solved it using the quicksort algorithm in C++. The problems CEP1 and DB1 are two-stage stochastic linear programs with recourse and were solved with the regularized decomposition algorithm of Ruszczyński [1986], using the accelerated version of the algorithm implemented in C++ [Ruszczyński and Swietanowski, 1997]. 10D is a stochastic knapsack problem and was solved using dynamic programming. In all cases, we used the Mersenne Twister algorithm of Matsumoto and Nishimura [1998] to generate the random samples, using an implementation in C++ [Wagner and Berg, 2013].

5.3 Results of Experiments

Summary of results from these experiments are depicted in Figures 2 and 3. Figures 2(a) and 3 show the reduction of variance, with each term being normalized with respect to $\text{Var}(nVG)$ of the nonoverlapping case (MRP). The solid lines in these figures show the theoretical variance reduction from [Welch, 1987] given by (6). Empirical results agree with the theoretical values well, with more variability observed in CEP1, 10D, and DB1. These experiments suggest that similar variance reduction can be achieved in stochastic programming by overlapping the batches, even with small sample sizes. We note that the variance reduction in DB1 can be less than the asymptotic amount. Even with the relatively small sample sizes compared to the size of this problem, however, overlapping appears to achieve some variance reduction.

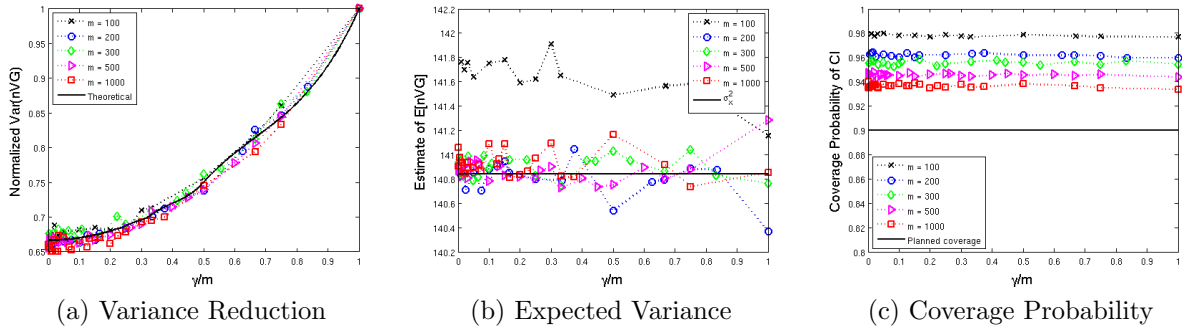


Figure 2: Summary of results for NVD. (a) Reduction in variance of nVG , (b) estimates of $E nVG$, and (c) coverage probability of the confidence intervals for various values of γ/m ($\gamma/m = 1$ denotes the nonoverlapping batches).

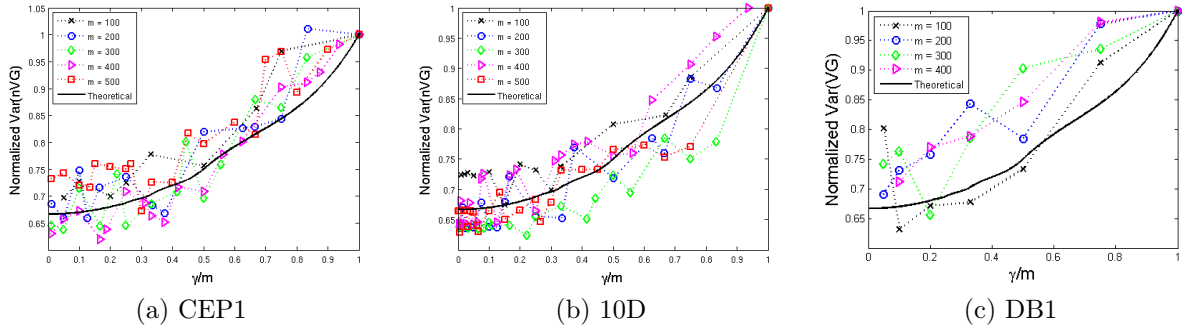


Figure 3: Variance Reduction for (a) CEP1, (b) 10D, and (c) DB1 for various values of γ/m .

Figure 2(b) shows estimates of $E nVG$ in the newsvendor problem for changing batch size and amount of overlap. We can see that the expectation of the variance estimator is not changed with increasing overlap. Although the estimate of $E nVG$ looks large for $m = 100$, the error from σ_x^2 is never more than 1%. Estimates for $E nVG$ showed a similar pattern for CEP1, 10D, and DB1. Those graphs are omitted for brevity.

Finally, Figure 2(c) shows the coverage probability of confidence intervals generated by OMRP for several values of m across varying values of γ/m for the newsvendor problem. The results of the classical overlapping batches estimators show that coverage probability does not change with the amount of overlap, which we can empirically see in this figure for OMRP for NVD. We observed the same for CEP1, 10D, and DB1. The coverage probabilities from applying the (nonoverlapping) MRP algorithm to these problems presented in [Bayraksan and Morton, 2006] agree with our results: for the newsvendor problem, coverage probability drops as m increases and the coverage probability of CEP1 (not shown) remains fairly constant around the desired value of 90%. Coverage probability for 10D and DB1 (not

shown) were very high, more than 0.99. The bias from solving (SP_m) for these problems seems to be much more significant than the variance.

Because variance reduction drops quickly as $\bar{\gamma}$ drops below 1, we recommend using an intermediate value of $\bar{\gamma}$, such as 1/3 or 1/4, to gain the majority of the variance reduction, while reducing the number of optimization problems to be solved. Warm-starting the algorithm to solve the sampling problems, when such a scheme is available, can significantly reduce the computation times.

6 Conclusions

We have extended previous work from simulation output analysis to assessing solution quality in stochastic programming by means of variably overlapping batches [Meketon and Schmeiser, 1984; Song and Schmeiser, 1993; Welch, 1987] for Monte Carlo simulation-based estimators of optimality gaps in stochastic programs [Mak et al., 1999]. We have provided conditions under which the resulting point estimator of the optimality gap and its associated sample variance are consistent, the same variance reduction can be achieved, and asymptotically valid confidence intervals can be formed. Empirical results with small sample sizes indicate that the asymptotic reductions in variance of the variance estimator show a similar decrease in OMRP, while bias and coverage probability remain unaffected.

Acknowledgments

The authors thank Andrzej Ruszczyński and Artur Świetanowski for access to their regularized decomposition code, which was used to solve CEP1 and DB1. An earlier version of this paper appeared in [Love and Bayraksan, 2011].

References

- C. Alexopoulos, N. T. Argon, D. Goldsman, N. M. Steiger, G. Tokol, and J. R. Wilson. Efficient computation of overlapping variance estimators for simulation. *INFORMS Journal on Computing*, 19(3):314–327, 2007a.
- C. Alexopoulos, N. T. Argon, D. Goldsman, G. Tokol, and J. R. Wilson. Overlapping variance estimators for simulation. *Operations Research*, 55(6):1090–1103, 2007b.
- G. Bayraksan and D. Morton. Assessing solution quality in stochastic programs via sampling.

- In *Tutorials in Operations Research*, volume 5, pages 102–122. INFORMS, Hanover, MD, 2009.
- G. Bayraksan and D. P. Morton. Assessing solution quality in stochastic programs. *Mathematical Programming*, 108(2):495–514, 2006.
- M. Bertocchi, J. Dupačová, and V. Moriggia. Sensitivity of bond portfolio’s behavior with respect to random movements in yield curve: A simulation study. *Annals of Operations Research*, 99:267–286, 2000.
- H. Damerджи. Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research*, 19(2):494–512, 1994.
- H. Damerджи. Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Operations Research*, 43(2):282–291, 1995.
- C. J. Donohue and J. R. Birge. An upper bound on the network recourse function. *Working Paper, Department of Industrial and Operations Engineering, University of Michigan*, 1995.
- J. Higle and S. Sen. *Stochastic Decomposition: A Statistical Method For Large Scale Stochastic Linear Programming*. Kluwer Academic Publishers, Dordrecht, 1996.
- J. Hu, T. Homem-de Mello, and S. Mehrotra. Sample average approximation of stochastic dominance constrained programs. *Mathematical Programming*, 133:171–201, 2012.
- U. Janjarassuk and J. T. Linderoth. Reformulation and sampling to solve a stochastic network interdiction problem. *Networks*, 52:120–132, 2008.
- A. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.
- A. Law. *Simulation Modeling and Analysis*. McGraw-Hill, 4 edition, 2007.
- D. Love and G. Bayraksan. Overlapping batches for the assessment of solution quality in stochastic programs. In S. Jain, R. Creasey, J. Himmelspach, K. White, and M. Fu, editors, *Proceedings of the 2011 Winter Simulation Conference*, pages 4184–4195, Piscataway, New Jersey, Dec. 2011. Institute of Electrical and Electronics Engineers, Inc.
- W. Mak, D. P. Morton, and R. K. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1-2):47 – 56, 1999.

- M. Matsumoto and T. Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, 1998.
- M. Meketon and B. Schmeiser. Overlapping batch means: Something for nothing? In S. Sheppard, U. Pooch, and D. Pegden, editors, *Proceedings of the 1984 Winter Simulation Conference*, pages 227–230, Piscataway, New Jersey, Dec. 1984. Institute of Electrical and Electronics Engineers, Inc.
- M. Meterelliyoz, C. Alexopoulos, and D. Goldsman. Folded overlapping variance estimators for simulation. *European Journal of Operational Research*, 220(1):135 – 146, 2012.
- V. Norikin, G. Pflug, and A. Ruszczyński. A branch and bound method for stochastic global optimization. *Mathematical Programming*, 83:425–450, 1998.
- A. Partani. *Adaptive Jackknife Estimators for Stochastic Programming*. Ph.D. dissertation, The University of Texas, Austin, Texas, 2007.
- A. Partani, D. Morton, and I. Popova. Jackknife estimators for reducing bias in asset allocation. In L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, editors, *Proceedings of the 2006 Winter Simulation Conference*, pages 783–791, Piscataway, New Jersey, Dec. 2006. Institute of Electrical and Electronics Engineers, Inc.
- A. Ruszczyński. A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming*, 35(3):309–333, 1986.
- A. Ruszczyński and A. Swietanowski. Accelerating the regularized decomposition method for two stage stochastic linear problems. *European Journal of Operational Research*, 101(2):328–342, 1997.
- T. Santoso, S. Ahmed, M. Goetschalckx, and A. Shapiro. A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*, 167:96–115, 2005.
- B. Schmeiser, T. N. Avramidis, and S. Hashem. Overlapping batch statistics. In O. Balci, R. P. Sadowski, and R. E. Nance, editors, *Proceedings of the 1990 Winter Simulation Conference*, pages 395–398, Piscataway, New Jersey, Dec. 1990. Institute of Electrical and Electronics Engineers, Inc.
- A. Shapiro and T. Homem-de Mello. On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. *SIAM Journal on Optimization*, 11(1):70–86, 2000.

- W. Song and B. Schmeiser. Variance of the sample mean: Properties and graphs of quadratic-form estimators. *Operations Research*, 41(3):501–517, 1993.
- R. Wagner and R. Berg. Mersenne Twister random number generator. <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/VERSIONS/C-LANG/MersenneTwister.h> [Last accessed 16-October-2013], 2013.
- W. Wang and S. Ahmed. Sample average approximation of expected value constrained stochastic programs. *Operations Research Letters*, 36:515–519, 2008.
- P. D. Welch. On the relationship between batch means, overlapping means and spectral estimation. In A. Thesen, H. Grant, and W. D. Kelton, editors, *Proceedings of the 1987 Winter Simulation Conference*, pages 320–323, Piscataway, New Jersey, Dec. 1987. Institute of Electrical and Electronics Engineers, Inc.