

OVERLAPPING BATCHES FOR THE ASSESSMENT OF SOLUTION QUALITY IN STOCHASTIC PROGRAMS

David Love

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

Güzin Bayraksan

Department of Systems and Industrial Engineering
University of Arizona
Tucson, AZ 85721, USA

ABSTRACT

We investigate the use of overlapping batches for assessing solution quality in stochastic programs. Motivated by the original use of overlapping batches in simulation, we present a variant of the multiple replications procedure that reuses data via variably overlapping batches to obtain alternative variance estimators. These estimators have lower variances, where the degree of variance reduction depends on the amount of overlap. We provide several asymptotic properties and present computational results to examine small-sample behavior.

1 INTRODUCTION

We consider a stochastic program 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 $X \neq \emptyset$ and compact, 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 $f(\cdot, \xi)$ is lower semicontinuous on X , almost surely (a.s.). This implies that (SP) has a finite optimal solution achieved on X .

Let $\xi^1, \xi^2, \dots, \xi^n$ be an independent and identically distributed (iid) sample from the distribution of ξ . A sampling approximation of (SP) is given by

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

We denote an optimal solution to (SP) as x^* and an optimal solution to (SP_n) as x_n^* .

We are interested in assessing the quality of a candidate solution $\hat{x} \in X$, which may have been found in any way, defined by its optimality gap $\mathbb{E}f(\hat{x}, \xi) - z^*$. This is important in practice since (SP) typically cannot be solved exactly, so one only has an approximate solution \hat{x} without verification of its quality. Assessing solution quality is also an indispensable part of stopping criteria in algorithms. In many fields of optimization, it is common to evaluate an upper bound on the optimality gap through relaxations such as integrality or Lagrangian relations. In stochastic programming, given a sample size n , an upper bound on the optimality gap can be obtained by $\mathbb{E}f(\hat{x}, \xi) - z^* \leq \mathbb{E}f(\hat{x}, \xi) - \mathbb{E}z_n^*$ due to the inequality $\mathbb{E}z_n^* \leq z^*$. This bound improves as the sample size increases, that is, $\mathbb{E}z_n^* \leq \mathbb{E}z_{n+1}^* \leq z^*$ (Mak, Morton, and Wood 1999, Norkin, Pflug, and Ruszczyński 1998). For stochastic programs, one way to estimate an upper bound on optimality gaps, $\mathbb{E}f(\hat{x}, \xi^i) - \mathbb{E}z_n^*$, is by Monte Carlo sampling. A straightforward estimate of $\mathbb{E}f(\hat{x}, \xi)$ is the sample mean, $\frac{1}{n} \sum_{i=1}^n f(\hat{x}, \xi^i)$. Instead of $\mathbb{E}z_n^*$, simply z_n^* can be used. The resulting point estimator is

$G_n(\hat{x}) = \frac{1}{n} \sum_{i=1}^n f(\hat{x}, \xi^i) - z_n^*$. Here, we assume the same observations ξ^1, \dots, ξ^n are used in both terms of $G_n(\hat{x})$, which results in variance reduction.

Computing $G_n(\hat{x})$ involves solving an optimization problem (SP_n) to obtain a lower bound estimator z_n^* , which complicates the statistical analysis. To enable statistical inference, the *multiple replications procedure* (MRP) of Mak, Morton, and Wood (1999) generates k independent estimators of $G_m(\hat{x})$, each using sample size m ($mk = n$) and averages them to obtain a point estimator. To form confidence intervals, the sample variance of these estimators is calculated. This is essentially a nonoverlapping batch means estimator (k batches of size m) commonly used in simulation; see, e.g., Law (2007). We further review MRP estimators in §2.2.

In this paper, we extend MRP by overlapping the batches. Overlapping batch means (OBM) was introduced by Meketon and Schmeiser (1984). Suppose each batch contains m observations and that there are a total of $2m$ observations. In the nonoverlapping case, there are only two batches: the first batch contains observations $\xi^1, \xi^2, \dots, \xi^m$ and the second batch contains observations $\xi^{m+1}, \xi^{m+2}, \dots, \xi^{2m}$. Meketon and Schmeiser (1984) overlap the batches so that the first batch is as before, the second batch now consists of $\xi^2, \xi^3, \dots, \xi^{m+1}$, the third batch consists of $\xi^3, \xi^4, \dots, \xi^{m+2}$ and so on. As a result, the corresponding batch means are no longer independent. However, in reusing the data in this fashion, one obtains a better variance estimator: a variance estimator that asymptotically has two-thirds of the variance of the usual batch means variance estimator but the bias of the two estimators is approximately the same. The idea of overlapping batches can be used for variance estimators of non-means (Schmeiser, Avramidis, and Hashem 1990) and recently, has been extended to estimators based on standardized time series (e.g., area, Cramér-von Mises) for steady-state simulation output analysis (Alexopoulos et al. 2007, Alexopoulos et al. 2007). In this paper, we explore its use for assessing solution quality in stochastic programs. Similar to Welch (1987) and Song and Schmeiser (1993), we consider partial overlap between batches. We provide computational results on two-stage stochastic linear programs with recourse and examine small sample behavior. Our computational experiments indicate that the asymptotic variance reduction is achieved with small sample sizes, while bias and coverage probability are unaffected. We also provide conditions under which the point estimators are consistent.

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 the multiple replications procedure in §2.2. In §3, we present the overlapping multiple replications procedure (OMRP) with variable overlap. In §4, we prove several asymptotic properties of the OMRP estimators and in §5, we test the performance of OMRP on two problems. We end in §6 with a summary and conclusions.

2 BACKGROUND

2.1 Overlapping Batch Means

Consider a covariance stationary stochastic process which has 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 $\mathbb{P}[L_\alpha(\mathbf{y}) \leq \mu \leq U_\alpha(\mathbf{y})] = 1 - \alpha$. 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 severely biased 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). As our aim is to investigate the use of overlapping batches for assessing solution quality in stochastic programs, in what follows, we briefly review two of these variance estimators, namely the 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 that $n = mk$). The sample mean of

each batch is then computed as

$$\bar{y}_j = \frac{1}{m} \sum_{i=1}^m y^{m(j-1)+i}, j = 1, 2, \dots, k \tag{1}$$

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

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

and the $(1 - \alpha)$ -level approximate confidence interval (CI) is then 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.

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 $j = 1, \dots, n - m + 1$. 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. Given this, the sample variance estimator is updated to

$$\widetilde{\text{Var}}(\bar{y}) = \frac{1}{\left(\frac{n}{m} - 1\right)(n - m + 1)} \sum_{j=1}^{n-m+1} (\bar{y}_j(m) - \bar{y})^2. \tag{3}$$

Meketon and Schmeiser (1984) show several attractive properties of the overlapping batches variance estimator, the two most important of which are: (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,

$$\frac{\text{Var}\left(\widetilde{\text{Var}}(\bar{y})\right)}{\text{Var}\left(\widehat{\text{Var}}(\bar{y})\right)} \rightarrow \frac{2}{3} \tag{4}$$

in the limit as batch size m and then the number of batches (n/m) tend to infinity. Note that the degrees of freedom in (3) is slightly different than in Meketon and Schmeiser (1984). The degrees of freedom in (3) makes $\widetilde{\text{Var}}(\bar{y})$ an unbiased estimator for iid 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 $\bar{y} \pm t_{3(k-1)/2, \alpha/2} \sqrt{\widetilde{\text{Var}}(\bar{y})}$, where (maximally) overlapping results in a 3/2 increase in the degrees of freedom (Welch 1987).

Welch (1987) established the relationship between OBM and spectral estimators and also considered partial overlap, see also Song and Schmeiser (1993). In our application to assessment of solution quality, we also consider varying the amount of overlap between neighboring batches. We defer this discussion on variable overlap to §3.1 and continue with a brief review of MRP.

2.2 Multiple Replications Procedure

Given a candidate solution $\hat{x} \in X$ to a stochastic program (SP), the task is to estimate its optimality gap $\mathbb{E}f(\hat{x}, \xi) - z^*$. Recall that MRP uses the upper bound on the optimality gap, $\mathbb{E}f(\hat{x}, \xi) - \mathbb{E}z_m^*$ (for a given sample size m), to construct a point estimator and a CI using the nonoverlapping batches method described above. Again, taking n observations of the random variables and splitting them into k batches of batch size m , define (for a candidate solution \hat{x})

$$\bar{G}_j = \frac{1}{m} \sum_{i=1}^m f(\hat{x}, \xi^{m(j-1)+i}) - \min_{x \in X} \frac{1}{m} \sum_{i=1}^m f(x, \xi^{m(j-1)+i}), \quad j = 1, 2, \dots, k. \tag{5}$$

Since we are assessing the quality of a given solution $\hat{x} \in X$, we suppress it from the notation. \bar{G}_j is similar to \bar{y}_j given in (1) except that the second term on the right-hand side of (5) minimizes a sample mean over a feasible set X . As before, after k batch estimators are obtained, the overall mean of these estimators, $\bar{\bar{G}} = \frac{1}{k} \sum_{j=1}^k \bar{G}_j$, provides a point estimator of the optimality gap. The sample variance is obtained as in (2) 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 approximately a level of significance of α .

Notice that the use of z_m^* in the \bar{G}_j (the second term on the right-hand side of (5)) 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 suggests 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, Morton, and Popova (2006), and Partani (2007) for variations of MRP aimed to reduce bias and variance.

An advantage of MRP is its applicability to a wide range of problems. With iid sampling, (SP) can be linear or nonlinear, X can include integrality constraints or not. It is also easy to implement, thus, has been applied to a variety of problems (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, de Mello, and Mehrotra 2011) constraints.

3 OVERLAPPING MULTIPLE REPLICATIONS PROCEDURE

Our aim is to apply the idea of overlapping batches to MRP. We note several differences in this setting compared to the simulation setting. In simulation, 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^*). First, 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 $x \in X$. We overcome this 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 to their nonoptimized counterparts. Second, once the data is generated through a simulation, it can be reused without much additional computational effort to obtain the OBM variance estimator. In our setting, due to the solution of a sampling problem (SP_m), the computational effort can increase with data reuse. Fortunately, near-optimal variance reduction can be obtained by partially overlapping the batches. Partial overlap results in a fewer number of batches; hence, fewer number of optimization problems need to be solved. Moreover, in many solution methods, warm-starting can be used to solve sampling approximations with overlapping samples, considerably reducing solution time.

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, $\bar{y}_j(m, \gamma) = \frac{1}{m} \sum_{i=1}^m y^{\gamma(j-1)+i}$, $j = 1, 2, \dots, \lfloor \frac{n-m}{\gamma} \rfloor + 1$, 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

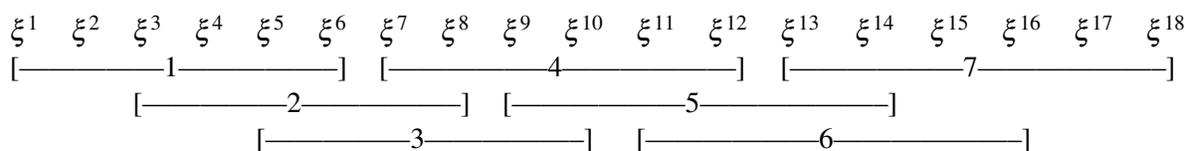


Figure 1: Visual representation of overlapping batches with $n = 18$, $m = 6$ and $\gamma = 2$. The brackets show which observations are used in each batch, and the numbers inside each bracket show the batch number j .

(3) is changed to

$$\widetilde{\text{Var}}_{\gamma}(\bar{y}) = \frac{1}{\binom{n}{m} \left(\lfloor \frac{n-m}{\gamma} \rfloor + 1 \right)} \sum_{j=1}^{\lfloor \frac{n-m}{\gamma} \rfloor + 1} (\bar{y}_j(m, \gamma) - \bar{y})^2. \tag{6}$$

Note that with $\gamma = 1$, (6) reduces to (3) and with $\gamma = m$ and $n = mk$, (6) reduces to (2).

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

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 $\frac{3N^2}{2N^2+1}$ when $\tilde{\gamma} = 1/N$ for integer N , and 3/2 in the maximally overlapping case (Welch 1987).

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 $N(i)$ denote the set of batches $j \in \{1, 2, \dots, \lfloor \frac{n-m}{\gamma} \rfloor + 1\}$ observation ξ^i is used in, $i = 1, 2, \dots, n$. $|N(i)|$ then gives the number of batches observation ξ^i is used in. 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, $N(1) = \{1\}$, $N(2) = \{1\}$, $N(3) = \{1, 2\}$ and $N(7) = \{2, 3, 4\}$. We use x_j^* to denote an optimal solution to sampling problem (SP_m) formed using the j th batch; $x_j^* \in \text{argmin}_{x \in X} \frac{1}{m} \sum_{i=1}^m f(x, \xi^{\gamma(j-1)+i})$, $j = 1, 2, \dots, \lfloor \frac{n-m}{\gamma} \rfloor + 1$.

The results on overlapping batch means occur in the limit as $n, m, k = n/m \rightarrow \infty$ (Damerджи 1994, Damerджи 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 limits will be taken as $l \rightarrow \infty$. For instance, this can be achieved with $m_l = n_l^r$ for some $0 < r < 1$, where n_l tends to infinity as $l \rightarrow \infty$. In particular, the results of Damerджи (1995) indicate that the batch size need to grow at a rate $r \in (1/2, 1)$ to have a consistent variance estimator in the mean-square sense. In addition, we may desire that the batch nonoverlap parameter change with l , to ensure that $\gamma_l/m_l = \tilde{\gamma}_l$ converges to the constant $\tilde{\gamma}$. Now we are ready to define the estimators for OMRP.

$$\bar{G}_j(m_l, \gamma_l) = \frac{1}{m_l} \sum_{i=1}^{m_l} f(\hat{x}, \xi^{\gamma_l(j-1)+i}) - \frac{1}{m_l} \sum_{i=1}^{m_l} f(x_j^*, \xi^{\gamma_l(j-1)+1}), \quad j = 1, 2, \dots, \lfloor \frac{n_l - m_l}{\gamma_l} \rfloor + 1, \quad (7)$$

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

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

The optimality gap estimator for each batch, $\bar{G}_j(m_l, \gamma_l)$, is defined like (5) for general values of the nonoverlap parameter γ_l . We just removed the minimization in (5) and used directly the optimal solution of x_j^* of batch j . The overall mean, $\bar{\bar{G}}_l$, is defined a little differently. Here, $\bar{\bar{G}}_l$ 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 $|N(i)|$ batches, then all the optimal solutions x_j^* corresponding to each batch $j \in N(i)$ are used for the lower bound estimator. Then, VG_l is defined in a similar fashion for the variable overlapping batches variance estimator given in (6).

4 THEORETICAL RESULTS

We begin this section by listing several additional assumptions on (SP) and the sampling scheme. Next, we introduce the nonoptimized counterparts of (7)–(9). We then use these nonoptimized estimators to establish certain asymptotic properties of OMRP estimators.

4.1 Assumptions

We make the following assumptions:

- A1** Samples of the random vector ξ are iid.
- A2** $z_n^* \rightarrow z^*$, almost surely, as $n \rightarrow \infty$.
- A3** $\exists \varepsilon > 0$ such that $\mathbb{E} [|\sup_{x \in X} f(x, \xi)|^{4+\varepsilon}] < \infty$.

First, we restrict our attention to iid sampling by A1. Then, we require optimal objective function values from sampling problems z_n^* converge to z^* , almost surely via A2. Assumption A2 is not uncommon in the literature and a significant amount of work has gone into documenting the conditions under which this assumption is true (see, e.g., Attouch and Wets (1981), Dupačová and Wets (1988), King and Rockafellar (1993), Shapiro (1991), and the survey of Shapiro (2003)). We are interested in variance of the variance estimator; therefore, we need (at least) the fourth moment of the objective function. Assumption A3 warrants this and is used as a sufficient condition for uniform integrability.

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

$$\bar{D}_j(m_l, \gamma_l) = \frac{1}{m_l} \sum_{i=1}^{m_l} f(\hat{x}, \xi^{\gamma_l(j-1)+i}) - \frac{1}{m_l} \sum_{i=1}^{m_l} f(x^*, \xi^{\gamma_l(j-1)+i}), \quad j = 1, 2, \dots, \lfloor \frac{n_l - m_l}{\gamma_l} \rfloor + 1, \quad (10)$$

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

$$VD_l = \frac{1}{(\frac{n_l}{m_l} - 1)(\lfloor \frac{n_l - m_l}{\gamma_l} \rfloor + 1)} \sum_{j=1}^{\lfloor \frac{n_l - m_l}{\gamma_l} \rfloor + 1} (\bar{D}_j(m_l, \gamma_l) - \bar{D}_l)^2. \quad (12)$$

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 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.

4.3 Consistency

In this section, we show the consistency of \bar{G} in the almost sure sense and the consistency of VG in the mean square sense. Throughout this section, we will denote convergence in L^p by $\xrightarrow{L^p}$, convergence in probability by \xrightarrow{P} , and almost sure convergence by $\xrightarrow{a.s.}$. For brevity, and to highlight the focus on the stochastic programming estimators becoming similar to the classical estimators, we will abbreviate convergence statements such as $VG_l - VD_l \xrightarrow{L^2} 0$ by $VG_l \xrightarrow{L^2} VD_l$ as $l \rightarrow \infty$.

The first result provides conditions under which the point estimator \bar{G} of OMRP is a strongly 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, $\bar{D}_l \xrightarrow{a.s.} \mathbb{E}f(\hat{x}, \xi) - z^*$ as $l \rightarrow \infty$ by the strong law of large numbers; as a result, $\bar{G}_l \xrightarrow{a.s.} \mathbb{E}f(\hat{x}, \xi) - z^*$.

Theorem 1 Suppose that assumptions A1 and A2 hold and $\bar{\gamma} = \gamma_l/m_l = 1/N$ for some integer N for all l . Then, $\bar{G}_l \xrightarrow{a.s.} \bar{D}_l$ as $l \rightarrow \infty$.

Proof. We use the shorthand notation $f_{ij} = f(\hat{x}, \xi^i) - f(x_j^*, \xi^i)$, for $1 \leq i \leq n_l$ and $1 \leq j \leq \lfloor \frac{n_l - m_l}{\gamma_l} \rfloor + 1$. Since not all combinations of i and j are represented, we set $f_{ij} = 0$ if it is not otherwise defined. For brevity, we suppress the subscript l in the below proof. Then, the total gap estimator is

$$\bar{G} = \frac{1}{n} \sum_{i=1}^n \frac{1}{|N(i)|} \sum_{j \in N(i)} f_{ij}.$$

We can ignore the “ends” of the sample asymptotically, and concentrate on the center $n - 2(m - 1)$ points, i.e., $m \leq i \leq n - m + 1$. This is justified because $\frac{n - 2(m - 1)}{n} \rightarrow 1$. Then

$$\bar{G} \approx \frac{1}{n} \sum_{i=m}^{n-m+1} \frac{1}{|N(i)|} \sum_{j \in N(i)} f_{ij}.$$

For these values of i , with the assumption that m/γ is integer, $|N(i)| = m/\gamma = N$. Partition the batches into sets $J_\lambda = \{j = 1, \dots, \lfloor \frac{n-m}{\gamma} \rfloor + 1 : j \bmod N = \lambda\}$ for $\lambda = 0, 1, \dots, N - 1$. (See Figure 1, and note that

$J_0 = \{3, 6\}$, $J_1 = \{1, 4, 7\}$ and $J_2 = \{2, 5\}$. Then, we can exchange the order of summation

$$\bar{G} \approx \frac{1}{N} \sum_{\lambda=0}^{N-1} \frac{1}{n} \sum_{j \in J_\lambda, i=m}^{n-m+1} f_{ij}.$$

Now, each interior sum is effectively the total gap estimator \bar{G} in the nonoverlapping case $\gamma \equiv m$ with $\frac{n}{m} - 2$ or $\frac{n}{m} - 3$ number of batches (we can again asymptotically ignore the ends within each λ). Thus, under assumptions A1 and A2, each interior sum converges to \bar{D} . \square

We now turn our attention to VG and show mean square convergence of $VG_l - VD_l$ to 0 as $l \rightarrow \infty$. We begin with two lemmas.

Lemma 1 Suppose that assumption A3 holds. Then, $(\bar{G}_j(m_l, \gamma_l))^4$ is uniformly integrable.

Proof. We again drop l from the notation for simplicity. Define

$$g(\hat{x}, \xi^i) = f(\hat{x}, \xi^i) - \inf_{x \in X} f(x, \xi^i) \quad h(\xi^i) = \sup_{x \in X} g(x, \xi^i)$$

so that $\bar{G}_j(m, \gamma) \leq \frac{1}{m} \sup_{x \in X} \sum g(x, \xi^i) \leq \frac{1}{m} \sum h(\xi^i)$. We can show that $(\bar{G}_j(m, \gamma))^2$ and $(\bar{G}_j(m, \gamma))^4$ are uniformly integrable, beginning with the inequality

$$\begin{aligned} \mathbb{E}(\bar{G}_1(m, \gamma))^2 &\leq \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \mathbb{E}h(\xi^i)h(\xi^j) \\ &\leq \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \sqrt{\mathbb{E}h^2(\xi^i)} \sqrt{\mathbb{E}h^2(\xi^j)} \\ &= \mathbb{E}h^2(\xi^i) \end{aligned}$$

and with a similar calculation showing that $\mathbb{E}(\bar{G}_1(m, \gamma))^4 \leq \mathbb{E}h^4(\xi^i)$. Then, to show uniform integrability, we will show that $\mathbb{E}(\bar{G}_1(m, \gamma))^4 I_{(\bar{G}_1(m, \gamma))^4 > t} \rightarrow 0$ as $t \rightarrow \infty$ uniformly in m . By above, we have

$$\mathbb{E}(\bar{G}_1(m, \gamma))^4 I_{(\bar{G}_1(m, \gamma))^4 > t} \leq \mathbb{E}h^4(\xi^i) I_{(\bar{G}_1(m, \gamma))^4 > t}.$$

We may then use Hölder's inequality,

$$\begin{aligned} \mathbb{E}(\bar{G}_1(m, \gamma))^4 I_{(\bar{G}_1(m, \gamma))^4 > t} &\leq \mathbb{E}h^4(\xi^i) I_{(\bar{G}_1(m, \gamma))^4 > t} \\ &\leq^{(4+\varepsilon)/4} \sqrt{\mathbb{E}h^2(\xi^i)^{4+\varepsilon}} \sqrt{\mathbb{P}[(\bar{G}_1(m, \gamma))^4 > t]} \end{aligned}$$

for some $\varepsilon > 0$. By assumption A3 and Markov's inequality, $\mathbb{P}[(\bar{G}_1(m, \gamma))^4 > t] \leq \mathbb{E}(\bar{G}_1(m, \gamma))^4 / t \rightarrow 0$ as $t \rightarrow \infty$, so we have uniform integrability. \square

Lemma 2 Suppose that assumptions A1-A3 hold and $\bar{\gamma} = \gamma_l / m_l = 1/N$ for some integer N for all l . Then, $VG_l \xrightarrow{P} VD_l$ as $l \rightarrow \infty$.

Proof. We can show that $VG_l \xrightarrow{P} VD_l$ by showing that $VG_l \xrightarrow{L^1} VD_l$ as $l \rightarrow \infty$:

$$\begin{aligned} \mathbb{E}|VG_l - VD_l| &= \frac{1}{\left(\frac{n_l}{m_l} - 1\right) \left(\lfloor \frac{n_l - m_l}{\gamma_l} \rfloor + 1\right)} \sum_{j=1}^{\lfloor \frac{n_l - m_l}{\gamma_l} \rfloor + 1} \mathbb{E} \left| (\bar{G}_j(m_l, \gamma_l) - \bar{G}_l)^2 - (\bar{D}_l(m_l, \gamma_l) - \bar{D}_l)^2 \right| \\ &\leq 2\mathbb{E} \left[|(\bar{G}_j(m_l, \gamma_l))^2 - (\bar{D}_j(m_l, \gamma_l))^2| + 2(\bar{G}_j(m_l, \gamma_l)\bar{G}_l - \bar{D}_j(m_l, \gamma_l)\bar{D}_l) + |\bar{G}_l - \bar{D}_l| \right]. \end{aligned}$$

Each term in parentheses above converges to zero in probability by assumption A2 and Theorem 1, and each term is uniformly integrable, so we have convergence in expectation. \square

Theorem 2 Suppose that assumptions A1-A3 hold and $\bar{\gamma} = \gamma_l/m_l = 1/N$ for some integer N for all l . Then, $VG_l \xrightarrow{L^2} VD_l$ as $l \rightarrow \infty$.

Proof. Sufficient conditions for mean-square convergence are given in Theorem 6.6.2 of Resnick (1999): if $VG_l \xrightarrow{P} VD_l$ and VG_l^2 is uniformly integrable, then $VG_l \xrightarrow{L^2} VD_l$ as $l \rightarrow \infty$. Lemmas 1 and 2 establish the assumptions of the quoted theorem. \square

From the above analysis, we can infer that the bias of VG_l is essentially the same across values of $\bar{\gamma}$ because of convergence in mean, $VG_l \xrightarrow{L^1} VD_l$ as $l \rightarrow \infty$. We see this in our experiments in §5.

5 COMPUTATIONAL RESULTS

We empirically tested the effectiveness of OMRP on two problems. The first test problem is a newsvendor problem with $U(0,10)$ distributed demand and the second is a capacity expansion planning problem, denoted CEP1, with random demand vector ξ of dimension 3, with 216 total realizations. Both problems can be written as two-stage stochastic linear programs with recourse and have unique optimal solutions. A description of CEP1 can be found in Hagle and Sen (1996). The instance of the newsvendor problem as well as the candidate solution used for this problem can be found in Bayraksan and Morton (2006). For CEP1, we used candidate solution (650, 650, 650, 650, 150, 150, 150, 150), which has optimality gap $1.28215\text{E}+06$ (z^* for CEP1 is 355,158.92). CEP1 was solved with the regularized decomposition algorithm of Ruszczyński (1986), using the accelerated version of the algorithm, implemented in C++ by Ruszczyński and Swietanowski (1997). We modified this code to use the Mersenne Twister algorithm to generate random samples (Wagner, Matsumoto, and Nishimura 2009). The newsvendor problem was solved using the quicksort algorithm in C++.

For both test problems, we used the sampling scheme $n = 30m$, with several values of the batch size m . We selected the number of batches in the nonoverlapping case as 30 because this value is used commonly in the literature for MRP estimators. Then, at each value of m , we formed OMRP estimators, with varying amounts of overlap. Recall that $\gamma/m = 1$ denotes the case of nonoverlapping batches and the amount of overlap increases for smaller values of γ/m ; for instance, $\gamma/m = 1/2$ denotes the case where half of the observations between batches overlap, etc. We also formed 90% confidence intervals on the optimality gap using ORMP estimators. Results for the newsvendor problem were compiled over 10,000 independent runs, while the CEP1 results were taken over 1,000 independent runs. Summary of results from these experiments are depicted in Figures 2 and 3.

Figures 2(a) and 3(a) show the reduction of variance, with each term being normalized with respect to $\text{Var}(VG)$ of the nonoverlapping case (MRP). The solid lines in these figures show the theoretical variance reduction from Welch (1987), computed at the points $\gamma/m = 1, 1/2, 1/3, \dots$. Empirical results agree with the theoretical values well, with more variability observed in CEP1. These computational results suggest that similar variance reduction can be achieved in stochastic programming by overlapping the batches.

Figures 2(b) and 3(b) show estimates of $\mathbb{E}VG$ for changing batch size and amount of overlap. As claimed in §4.3, we can see that the expectation of the variance estimator is not changed with varying amount of overlap and is essentially the same as the nonoverlapping case. (The decrease in $\mathbb{E}VG$ as m increases is due to the fact that VG provides an estimator of $\text{Var}(\bar{G})$, which shrinks as $n = 30m$ increases.)

Finally, Figures 2(c) and 3(c) show the coverage probability of confidence intervals generated by OMRP for several values of m across varying values of γ/m . 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 these figures for OMRP. The coverage probabilities from applying the (nonoverlapping)

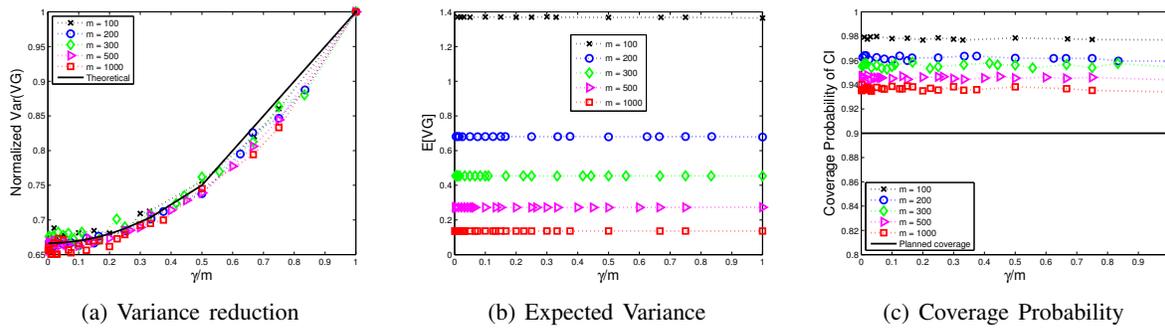


Figure 2: Summary of results for the newsvendor problem: (a) reduction in variance of VG , (b) estimates of $\mathbb{E}VG$, and (c) coverage probability of the confidence intervals for various values of γ/m . ($\gamma/m = 1$ denotes the nonoverlapping batches.)

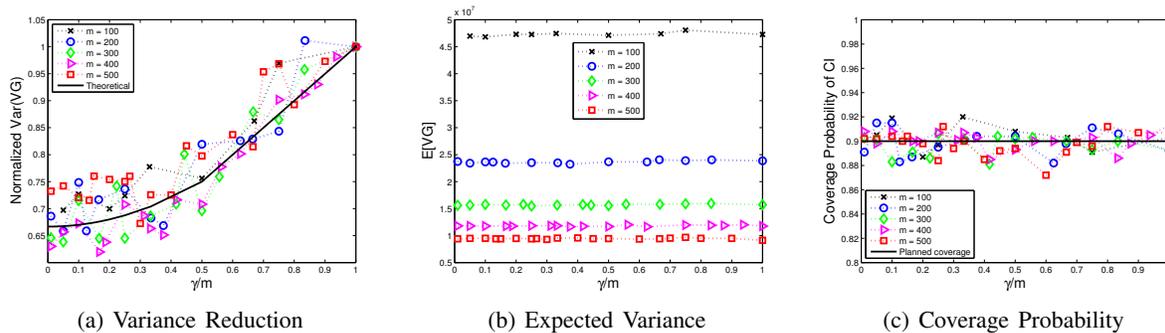


Figure 3: Summary of results for CEP1: (a) reduction in variance of VG , (b) estimates of $\mathbb{E}VG$, and (c) coverage probability of the confidence intervals for various values of γ/m . ($\gamma/m = 1$ denotes the nonoverlapping batches.)

MRP algorithm to these problems presented in Bayraksan and Morton (2006) agree with the results presented here. As in Bayraksan and Morton (2006), the coverage probability of the newsvendor problem drops as m increases. The coverage probability of CEP1, on the other hand, remains fairly constant around the desired value of 90%, like the suboptimal solution used in Bayraksan and Morton (2006).

Because variance reduction drops quickly as $\tilde{\gamma}$ drops below 1, the authors recommend using an intermediate value of $\tilde{\gamma}$, such as $1/3$, 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 as scheme is available, can reduce the solution times.

6 CONCLUSION

We have extended previous work from simulation analysis that use variably overlapping batches to form variance estimators (Meketon and Schmeiser 1984, Song and Schmeiser 1993, Welch 1987) to Monte Carlo sampling-based estimators of optimality gaps in stochastic programming (Mak, Morton, and Wood 1999). The point estimator in this case is also slightly changed. We have provided conditions under which this point estimator of the optimality gap is strongly consistent and that the resulting variance estimator converges in mean-square to a nonoptimized counterpart. 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. This research is supported in part by the National Science Foundation under Grants DMS-0602173 and EFRI-0835930.

REFERENCES

- Alexopoulos, C., N. T. Argon, D. Goldsman, N. M. Steiger, G. Tokol, and J. R. Wilson. 2007. “Efficient computation of overlapping variance estimators for simulation”. *INFORMS Journal On Computing* 19 (3): 314–327.
- Alexopoulos, C., N. T. Argon, D. Goldsman, G. Tokol, and J. R. Wilson. 2007. “Overlapping Variance Estimators for Simulation”. *Operations Research* 55 (6): 1090–1103.
- Attouch, H., and R. Wets. 1981. “Approximation and convergence in nonlinear optimization”. In *Nonlinear Programming 4*, edited by O. Mangasarian, R. Meyer, and S. Robinson, 367–394. Academic Press, New York.
- Bayraksan, G., and D. Morton. 2009. “Assessing Solution Quality in Stochastic Programs via Sampling”. In *Tutorials in Operations Research*, Volume 5, 102–122. Hanover, MD: INFORMS.
- Bayraksan, G., and D. P. Morton. 2006. “Assessing solution quality in stochastic programs”. *Mathematical Programming* 108 (2): 495–514.
- Bertocchi, M., J. Dupačová, and V. Moriggia. 2000. “Sensitivity of Bond Portfolio’s Behavior with Respect to Random Movements in Yield Curve: A Simulation Study”. *Annals of Operations Research* 99:267–286.
- Damerdji, H. 1994. “Strong consistency of the variance estimator in steady-state simulation output analysis”. *Mathematics of Operations Research* 19 (2): 494–512.
- Damerdji, H. 1995. “Mean-square consistency of the variance estimator in steady-state simulation output analysis”. *Operations Research* 43 (2): 282–291.
- Dupačová, J., and R. Wets. 1988. “Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems”. *The Annals of Statistics* 16:1517–1549.
- Higle, J., and S. Sen. 1996. *Stochastic decomposition: a statistical method for large scale stochastic linear programming*. Dordrecht: Kluwer Academic Publishers.
- Hu, J., T. H. de Mello, and S. Mehrotra. 2011. “Sample Average Approximation for Stochastic Dominance Constrained Programs”. Technical report, Northwestern University. Available via <http://www.optimization-online.org> [accessed October 22, 2011].
- Janjarassuk, U., and J. T. Linderoth. 2008. “Reformulation and Sampling to Solve a Stochastic Network Interdiction Problem”. *Networks* 52:120–132.
- King, A. J., and R. T. Rockafellar. 1993. “Asymptotic theory for solutions in statistical estimation and stochastic programming”. *Mathematics of Operations Research* 18:148–162.
- Law, A. 2007. *Simulation Modeling and Analysis*. 4 ed. McGraw-Hill.
- Mak, W., D. P. Morton, and R. K. Wood. 1999. “Monte Carlo bounding techniques for determining solution quality in stochastic programs”. *Operations Research Letters* 24 (1-2): 47 – 56.
- Meketon, M., and B. Schmeiser. 1984, December. “Overlapping batch means: Something for nothing?”. In *Proceedings of the 1984 Winter Simulation Conference*, edited by S. Sheppard, U. Pooch, and D. Pegden, 227–230. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Norkin, V., G. Pflug, and A. Ruszczyński. 1998. “A branch and bound method for stochastic global optimization”. *Mathematical Programming* 83:425–450.
- Partani, A. 2007. *Adaptive Jackknife Estimators for Stochastic Programming*. Ph.D. dissertation, The University of Texas, Austin, Texas.

- Partani, A., D. Morton, and I. Popova. 2006, December. "Jackknife estimators for reducing bias in asset allocation". In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 783–791. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Resnick, S. 1999, October. *A Probability Path*. Birkhäuser Boston.
- Ruszczynski, A. 1986. "A regularized decomposition method for minimizing a sum of polyhedral functions". *Mathematical programming* 35 (3): 309–333.
- Ruszczynski, A., and A. Swietanowski. 1997. "Accelerating the regularized decomposition method for two stage stochastic linear problems". *European Journal of Operational Research* 101 (2): 328–342.
- Santoso, T., S. Ahmed, M. Goetschalckx, and A. Shapiro. 2005. "A Stochastic Programming Approach for Supply Chain Network Design under Uncertainty". *European Journal of Operational Research* 167:96–115.
- Schmeiser, B., T. N. Avramidis, and S. Hashem. 1990, December. "Overlapping Batch Statistics". In *Proceedings of the 1990 Winter Simulation Conference*, edited by O. Balci, R. P. Sadowski, and R. E. Nance, 395–398. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Shapiro, A. 1991. "Asymptotic analysis of stochastic programs". *Annals of Operations Research* 30:169–186.
- Shapiro, A. 2003. "Monte Carlo sampling methods". In *Handbooks in Operations Research and Management Science, Volume 10: Stochastic Programming*, edited by A. Ruszczyński and A. Shapiro, 353–425. Elsevier.
- Song, W., and B. Schmeiser. 1993. "Variance of the sample mean: Properties and graphs of quadratic-form estimators". *Operations Research* 41 (3): 501–517.
- Wagner, R.J. and M. Matsumoto and T. Nishimura 2009, September. "Mersenne Twister Random Number Generator". <http://www-personal.umich.edu/~wagnerr/MersenneTwister.html>.
- Wang, W., and S. Ahmed. 2008. "Sample average approximation of expected value constrained stochastic programs". *Operations Research Letters* 36:515–519.
- Welch, P. D. 1987, December. "On the relationship between batch means, overlapping means and spectral estimation". In *Proceedings of the 1987 Winter Simulation Conference*, edited by A. Thesen, H. Grant, and W. D. Kelton, 320–323. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

DAVID LOVE is a graduate student in the Interdisciplinary Program in Applied Mathematics at the University of Arizona. His research focuses on Monte Carlo based quality assessment in stochastic programming. His email address is dlove@email.arizona.edu.

GÜZİN BAYRAKSAN is an Assistant Professor in the Department of Systems and Industrial Engineering at the University of Arizona. Her research interests include stochastic optimization, particularly Monte Carlo sampling-based methods for stochastic programming, with applications to water resources management. Her email address is guzinb@sie.arizona.edu and her web page is www.sie.arizona.edu/faculty/guzinb.