

Overlapping Batches for the Assessment of Solution Quality in Stochastic Programs

David Love¹ Güzin Bayraksan²

¹Graduate Interdisciplinary Program in Applied Mathematics, University of Arizona

²Systems & Industrial Engineering, University of Arizona

12th International Conference on Stochastic Programming
August 20, 2010

- 1 Introduction
 - Problem Description
 - Previous Results

- 2 Overlapping Multiple Replications Procedure (OMRP)
 - Defining Estimators
 - Theoretical Issues
 - Extension

- 3 Computational Results

Problem Description

$$\min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}f(\mathbf{x}, \tilde{\xi}) \quad (\text{SP})$$

$$\mathbf{x}^* \in \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} \mathbb{E}f(\mathbf{x}, \tilde{\xi})$$

- \mathbf{x} Vector of decision variables.
- $\tilde{\xi}$ Vector of problem parameters.
- f Objective function, may depend on $\tilde{\xi}$.
 - Lower semicontinuous on x a.s.
 - $\mathbb{E}f(\mathbf{x}, \tilde{\xi})$ defined and finite for all x .
- \mathcal{X} Feasible region.
 - Nonempty, closed, compact, independent of $\tilde{\xi}$.

Project Goal

- We will combine two previously published results
 - One from stochastic programming
 - One from simulation
- To develop a new variation on answering the question:

Given a candidate solution, $\hat{\mathbf{x}}$, to the stochastic problem, how can we estimate how good it is, compared to the optimal solution?

Approximating Problem

Following Mak, Morton and Wood (1999), we set up the approximating problem

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

$$\mathbf{x}_n^* \in \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}, \tilde{\xi}^i)$$

where

- $\tilde{\xi}^i$, i.i.d. samples from the distribution of $\tilde{\xi}$, $i = 1, \dots, n$
- z_n^* , optimal objective value of the approximating problem
- \mathbf{x}_n^* , optimal solution of the approximating problem

Defining a “Good” Candidate Solution

Suppose we have a candidate solution $\hat{\mathbf{x}}$ (maybe we solved (SP_n)). How good is it?

- We can look at its “optimality gap”,

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

- Clearly is positive.
- Want a statistical estimate on size.

Optimality Gap Estimation

The optimality gap $\mathbb{E}f(\hat{\mathbf{x}}, \tilde{\xi}) - \min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}f(\mathbf{x}, \tilde{\xi})$ is approximated by

$$\frac{1}{M} \sum_{i=1}^M f(\hat{\mathbf{x}}, \tilde{\xi}^i) - \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{M} \sum_{i=1}^M f(\mathbf{x}, \tilde{\xi}^i)$$

- Since the approximated optimal solution is biased, the gap estimator is positively biased.
- Bias monotonically decreases as batch size M increases.
- To do statistics, we repeat the above multiple times...

Multiple Replications Procedure

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \frac{N}{M} = 3$ Number of batches

$\tilde{\xi}^1$ $\tilde{\xi}^2$ $\tilde{\xi}^3$ $\tilde{\xi}^4$ $\tilde{\xi}^5$ $\tilde{\xi}^6$ $\tilde{\xi}^7$ $\tilde{\xi}^8$ $\tilde{\xi}^9$ $\tilde{\xi}^{10}$ $\tilde{\xi}^{11}$ $\tilde{\xi}^{12}$

Multiple Replications Procedure

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \frac{N}{M} = 3$ Number of batches

$$\begin{array}{cccccccccccccc}
 \tilde{\xi}^1 & \tilde{\xi}^2 & \tilde{\xi}^3 & \tilde{\xi}^4 & \tilde{\xi}^5 & \tilde{\xi}^6 & \tilde{\xi}^7 & \tilde{\xi}^8 & \tilde{\xi}^9 & \tilde{\xi}^{10} & \tilde{\xi}^{11} & \tilde{\xi}^{12} \\
 \left[\text{-----} \right] & & & & & & & & & & & \\
 \bar{G}_1 & & & & & & & & & & &
 \end{array}$$

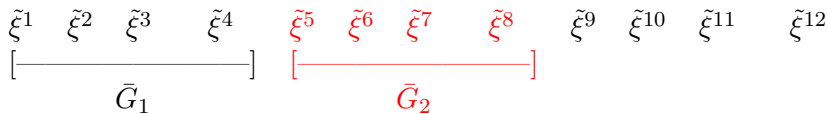
$$\bar{G}_1 = \frac{1}{M} \sum_{i=1}^M f(\hat{\mathbf{x}}, \tilde{\xi}^{M(1-1)+i}) - \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{M} \sum_{i=1}^M f(\mathbf{x}, \tilde{\xi}^{M(1-1)+i})$$

Multiple Replications Procedure

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \frac{N}{M} = 3$ Number of batches



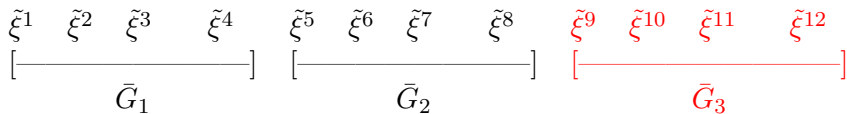
$$\bar{G}_2 = \frac{1}{M} \sum_{i=1}^M f(\hat{\mathbf{x}}, \tilde{\xi}^{M(2-1)+i}) - \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{M} \sum_{i=1}^M f(\mathbf{x}, \tilde{\xi}^{M(2-1)+i})$$

Multiple Replications Procedure

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \frac{N}{M} = 3$ Number of batches



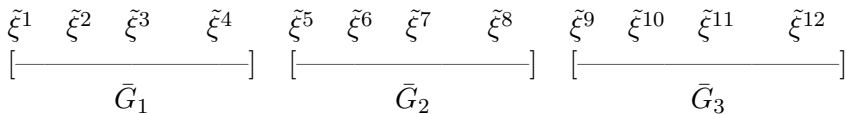
$$\bar{G}_3 = \frac{1}{M} \sum_{i=1}^M f(\hat{\mathbf{x}}, \tilde{\xi}^{M(3-1)+i}) - \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{M} \sum_{i=1}^M f(\mathbf{x}, \tilde{\xi}^{M(3-1)+i})$$

Multiple Replications Procedure

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \frac{N}{M} = 3$ Number of batches



$$\bar{G}_j = \frac{1}{M} \sum_{i=1}^M f(\hat{\mathbf{x}}, \xi^{M(j-1)+i}) - \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{M} \sum_{i=1}^M f(\mathbf{x}, \xi^{M(j-1)+i})$$

$$VG = \frac{1}{(k-1)k} \sum_{j=1}^k (\bar{G}_j - \bar{\bar{G}})^2 \quad \bar{\bar{G}} = \frac{1}{k} \sum_{j=1}^k \bar{G}_j$$

Batch Means

- This situation is much like that of simulation:
 - Partition simulation output into batches.
 - Do statistics on the batches.
- Meketon & Schmeiser (1984) proposed that batches could be overlapped.

Overlapping Batches

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \lfloor \frac{N}{M} \rfloor = 3$ Number of nonoverlapping batches

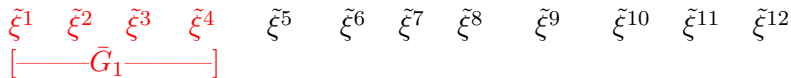
$\tilde{\xi}^1$ $\tilde{\xi}^2$ $\tilde{\xi}^3$ $\tilde{\xi}^4$ $\tilde{\xi}^5$ $\tilde{\xi}^6$ $\tilde{\xi}^7$ $\tilde{\xi}^8$ $\tilde{\xi}^9$ $\tilde{\xi}^{10}$ $\tilde{\xi}^{11}$ $\tilde{\xi}^{12}$

Overlapping Batches

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \lfloor \frac{N}{M} \rfloor = 3$ Number of nonoverlapping batches

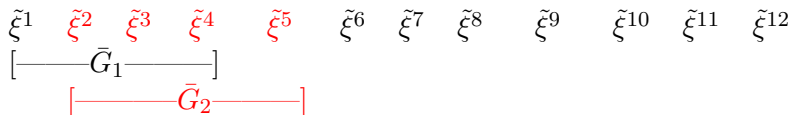


Overlapping Batches

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \lfloor \frac{N}{M} \rfloor = 3$ Number of nonoverlapping batches

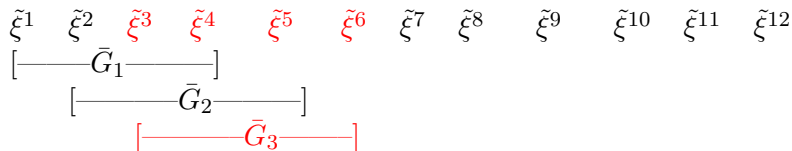


Overlapping Batches

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \lfloor \frac{N}{M} \rfloor = 3$ Number of nonoverlapping batches

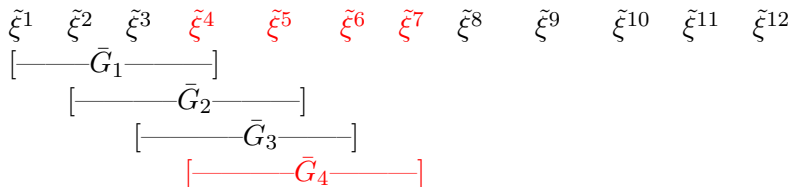


Overlapping Batches

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \lfloor \frac{N}{M} \rfloor = 3$ Number of nonoverlapping batches

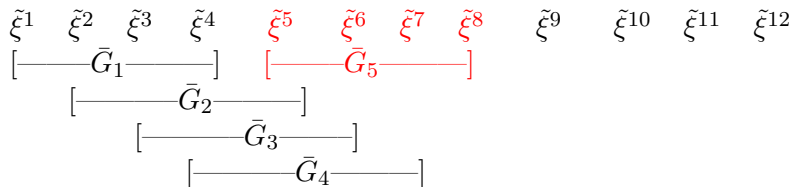


Overlapping Batches

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \lfloor \frac{N}{M} \rfloor = 3$ Number of nonoverlapping batches

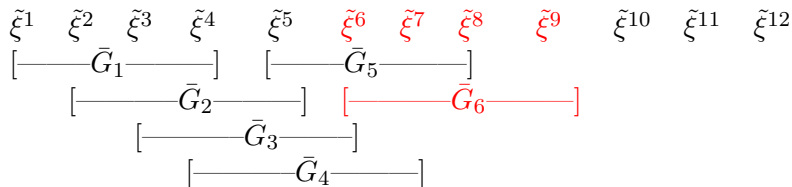


Overlapping Batches

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \lfloor \frac{N}{M} \rfloor = 3$ Number of nonoverlapping batches

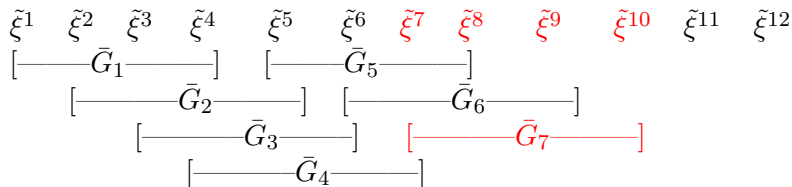


Overlapping Batches

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \lfloor \frac{N}{M} \rfloor = 3$ Number of nonoverlapping batches

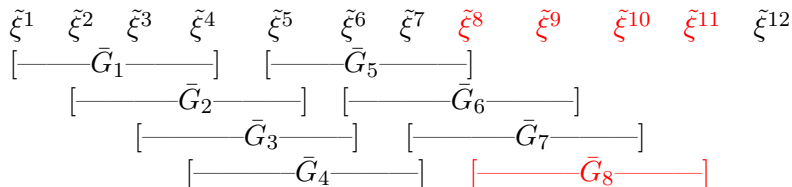


Overlapping Batches

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \lfloor \frac{N}{M} \rfloor = 3$ Number of nonoverlapping batches

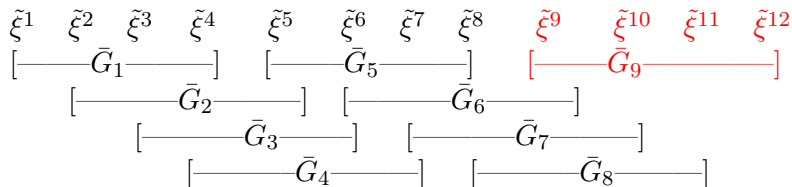


Overlapping Batches

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \lfloor \frac{N}{M} \rfloor = 3$ Number of nonoverlapping batches

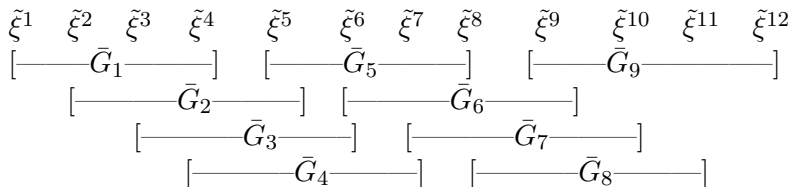


Overlapping Batches

$N = 12$ Total sample size

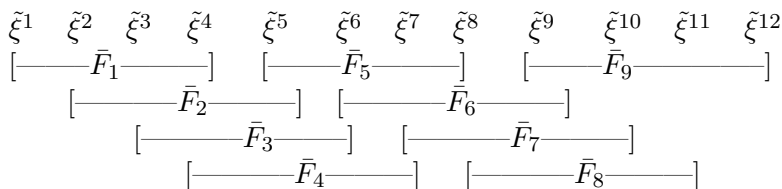
$M = 4$ Batch size

$k = \lfloor \frac{N}{M} \rfloor = 3$ Number of nonoverlapping batches



$$VG = \frac{1}{(k-1)(N-M+1)} \sum_{j=1}^{N-M+1} (\bar{G}_j - \bar{G})^2$$

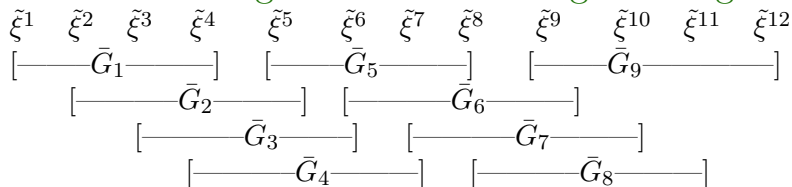
What About the Grand Average?



- A small excursion to simulation, and suppose we wish to estimate $\mathbb{E}f(\hat{\mathbf{x}}, \tilde{\xi})$ (rather than $\mathbb{E}f(\hat{\mathbf{x}}, \tilde{\xi}) - z^*$).
- Meketon and Schmeiser simply average all observations:

$$\bar{F}_j = \frac{1}{M} \sum_{i=1}^M f(\hat{\mathbf{x}}, \tilde{\xi}^{M(j-1)+i}) \quad \bar{F} = \frac{1}{N} \sum_{i=1}^N f(\hat{\mathbf{x}}, \tilde{\xi}^i)$$

Extending to Stochastic Programming



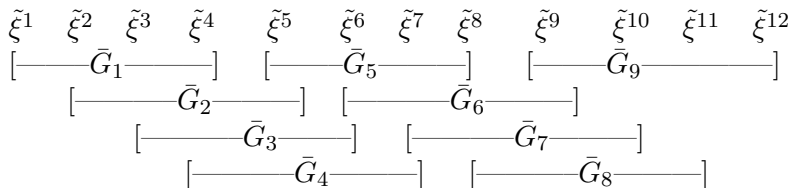
- Our batch average

$$\bar{G}_j = \frac{1}{M} \sum_{i=1}^M \left[f(\hat{\mathbf{x}}, \tilde{\xi}^{M(j-1)+i}) - f(\mathbf{x}_{M,j}^*, \tilde{\xi}^{M(j-1)+i}) \right]$$

- But what would the grand average be?

$$\bar{\bar{G}} = \frac{1}{N} \sum_{i=1}^N \left[f(\hat{\mathbf{x}}, \tilde{\xi}^i) - f(??, \tilde{\xi}^i) \right]$$

A Small Problem



Problem The internal minimization of

$$\bar{G}_j = \frac{1}{M} \sum_{i=1}^M f(\hat{\mathbf{x}}, \tilde{\xi}^{M(j-1)+i}) - \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{M} \sum_{i=1}^M f(\mathbf{x}, \tilde{\xi}^{M(j-1)+i})$$

means the batches have individual properties, making it difficult to take the grand average.

Solution We must average in two directions.

Different Variance Estimators

Batch Mean $\bar{G}_j =$

$$\frac{1}{M} \sum_{i=1}^M \left[f(\hat{\mathbf{x}}, \tilde{\xi}^{M(j-1)+i}) - f(\mathbf{x}_{M,j}^*, \tilde{\xi}^{M(j-1)+i}) \right]$$

Overall Mean

$$\bar{\bar{G}} = \frac{1}{N} \sum_{i=1}^N \frac{1}{|N(i)|} \sum_{j \in N(i)} \left[f(\hat{\mathbf{x}}, \tilde{\xi}^i) - f(\mathbf{x}_{M,j}^*, \tilde{\xi}^i) \right]$$

Nonoverlapping Variance $VG(\text{NOL}) = \frac{1}{(k-1)k} \sum_{j=1}^k (\bar{G}_j - \bar{\bar{G}})^2$

Overlapping Variance

$$VG(\text{OL}) = \frac{1}{(k-1)(N-M+1)} \sum_{j=1}^{N-M+1} (\bar{G}_j - \bar{\bar{G}})^2$$

Assumptions

- A1** Samples of the random variable $\tilde{\xi}$ are i.i.d.
- A2** $z_n^* \rightarrow z^*$ a.s.
- A3** $\exists \epsilon > 0$ such that $\mathbb{E} \left[\left(\sup_{x \in X} f(x, \tilde{\xi}) \right)^{4+\epsilon} \right] < \infty$.

Important Result

- Bias in the variance estimator is essentially unaffected by the decision to overlap.
- Both overlapping and nonoverlapping estimators are unbiased in the limit as $N, M, k \rightarrow \infty$.
- Overlapping batches lowers the variance of $VG(\text{OL})$, the variance estimator of the optimality gap.

$$\lim_{N, M, k \rightarrow \infty} \frac{\text{Var}(VG(\text{OL}))}{\text{Var}(VG(\text{NOL}))} = \frac{2}{3}$$

Another Small Problem

Problem The internal minimization of

$$\bar{G}_j = \frac{1}{M} \sum_{i=1}^M f(\hat{\mathbf{x}}, \tilde{\xi}^{M(j-1)+i}) - \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{M} \sum_{i=1}^M f(\mathbf{x}, \tilde{\xi}^{M(j-1)+i})$$

means we do not have the nice properties we want.

Solution Introduce a sequence of “non-minimized” gap estimators, \bar{D}_j . Show convergence between these two series.

Estimators for Proof

Unbiased Gap Estimator

- $\bar{D}_j = \frac{1}{M} \sum_{i=1}^M \left[f(\hat{\mathbf{x}}, \tilde{\xi}^{M(j-1)+i}) - f(\mathbf{x}^*, \tilde{\xi}^{M(j-1)+i}) \right]$
- $\bar{\bar{D}} = \frac{1}{N} \sum_{i=1}^N \left[f(\hat{\mathbf{x}}, \tilde{\xi}^i) - f(\mathbf{x}^*, \tilde{\xi}^i) \right]$
- $VD_j(\text{OL}) = \frac{1}{(k-1)(N-M+1)} \sum_{j=1}^{N-M+1} (\bar{D}_j - \bar{\bar{D}})^2$

Proposed Gap Estimator

- $\bar{G}_j = \frac{1}{M} \sum_{i=1}^M \left[f(\hat{\mathbf{x}}, \tilde{\xi}^{M(j-1)+i}) - f(\mathbf{x}_{M,j}^*, \tilde{\xi}^{M(j-1)+i}) \right]$
- $\bar{\bar{G}} = \frac{1}{N} \sum_{i=1}^N \frac{1}{|N(i)|} \sum_{j \in N(i)} \left[f(\hat{\mathbf{x}}, \tilde{\xi}^i) - f(\mathbf{x}_{M,j}^*, \tilde{\xi}^i) \right]$
- $VG(\text{OL}) = \frac{1}{(k-1)(N-M+1)} \sum_{j=1}^{N-M+1} (\bar{G}_j - \bar{\bar{G}})^2$

Partial Overlap

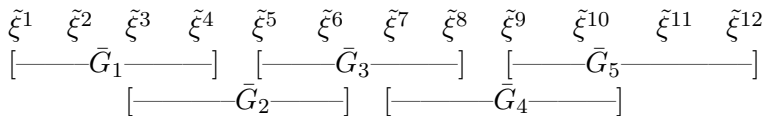


Figure: Graphical representation of $\gamma = 2$.

Let γ be the batch nonoverlap parameter.

- γ gives the number of new samples added to each new batch.
- $\gamma = 1$, maximally overlapping case discussed above.
- $\gamma = M$, classical nonoverlapping case.
- More useful: $\frac{\gamma}{M} \in (0, 1]$.

Result

Variance reduction of Sample Variance

- $\gamma \equiv 1$, variance is reduced to $\frac{2}{3}$ of original.
- $\gamma = \frac{1}{3}M$, variance is reduced to $\frac{19}{27}$ of original.
- $\gamma = \frac{1}{2}M$, variance is reduced to $\frac{3}{4}$ of original.
- $\gamma = \frac{1}{L}M$, variance is reduced to $\frac{2L^2 + 1}{3L^2}$ of original.

Newsvendor Problem

- You purchase newspapers from a supplier at cost c , and sell them to the public at price r .
- Number of customers $\tilde{d} \geq 0$ is a random variable.
- Purchase x papers every day.
- Maximize expected profit $\mathbb{E} \left[-cx + r \min \{x, \tilde{d}\} \right]$.

Optimal solution occurs when $\mathbb{P} \left[\tilde{d} \leq x \right] = \frac{r-c}{r}$.

News vendor Problem

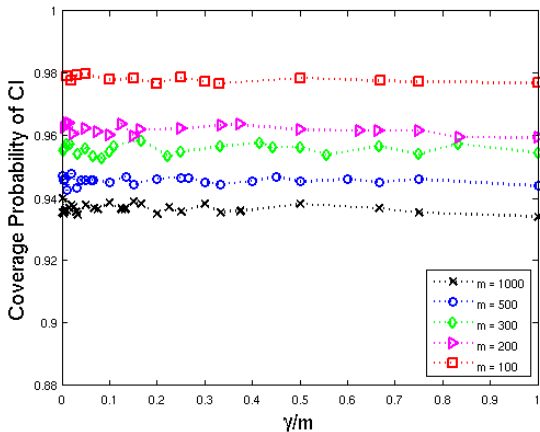
We tested the computational properties of our method on the news vendor problem, with

- Selling price $r = 15$.
- Cost of paper $c = 5$.
- Optimal quantile $\frac{r-c}{r} = \frac{2}{3}$.
- Daily demand of newspapers is $\tilde{d} \sim U(0, 10)$.
- $\hat{\mathbf{x}} = 8.7749$.

Computational Method

- Take a variety of batch sizes M .
- For each, choose sample size $N = 30M$.
- Compute confidence intervals for γ/M
 - Near 0, for fully overlapping
 - $\gamma/M = 1$ for nonoverlapping
- Repeat each computation 10,000 times for estimates.
- Construct a 90% confidence interval.

Coverage Probability



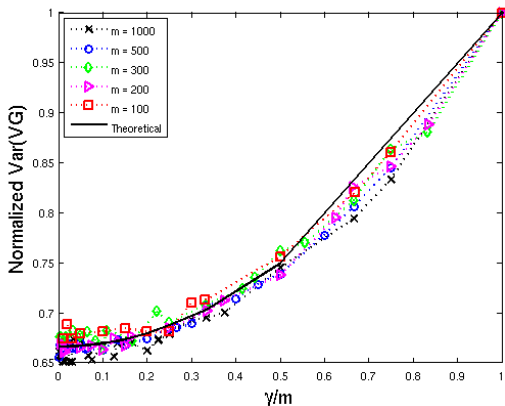
Variance Reduction

- The main result:

$$\lim_{N, M, k \rightarrow \infty} \frac{\text{Var}(VG(\frac{\gamma}{M}))}{\text{Var}(VG(\text{NOL}))}$$

is reduced by overlapping.

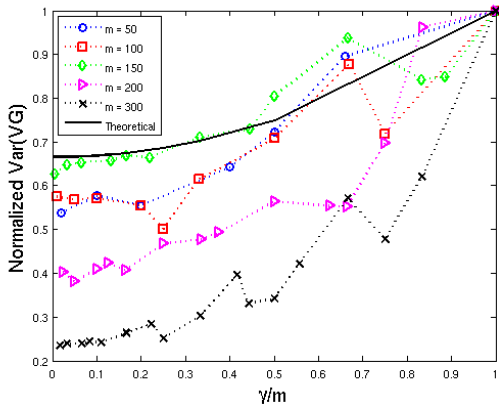
Variance Reduction of Variance Estimator



SLP-2 Test Problems

- Also tested OMRP on several Two-Stage Stochastic Linear Programs with Recourse.
- Solved with the regularized decomposition algorithm of Ruszczyński and Świetanowski.
- Results for APL1P presented here.
 - Power expansion planning problem with 5 independent stochastic parameters and 1280 total realizations.
- Similar results on confidence interval coverage probability.

Variance Reduction in APL1P



Conclusions

- Began with old results:
 - Variance reduction by overlapping batches.
 - Solution quality assessment by MRP.
- I have shown that these results can be combined in stochastic programming.
- Computational results show small-sample behavior is in close agreement with asymptotic results.

References

- ① 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.
- ② M. S. Meketon and B. Schmeiser. Overlapping batch means: something for nothing? In *WSC Proceedings of the 1984 conference on Winter simulation*, pages 226-230, Piscataway, NJ, USA, 1984. IEEE Press.
- ③ P. D. Welch. On the relationship between batch means, overlapping means and spectral estimation. *Proceedings of the 1987 Winter Simulation Conference*, pages 320-323, New York, NY, USA, 1987. IEEE Press.

Thanks for Listening!