Overlapping Batches for Variance Reduction in Optimality Gap Estimation in Stochastic Programming

David Love[1] Güzin Bayraksan[2]

[1]Graduate Interdisciplinary Program in Applied Mathematics, University of Arizona

[2]Systems & Industrial Engineering, University of Arizona

INFORMS Annual Meeting 2009
October 13, 2009

The Basics

- Stochastic Process $\{X_t, t \geq 0\}$, discrete- or continuous-time or state space.
- (Long Term) Mean μ, Variance R_0.
- Want an estimate of μ given a realization $\{X^n\}$.

The Basics

- Stochastic Process $\{X_t, t \geq 0\}$, discrete- or continuous-time or state space.
- (Long Term) Mean μ, Variance R_0.
- Want an estimate of μ given a realization $\{X^n\}$.

- A family of point estimators, considered as $\bar{X} = \sum_{i=1}^{n} \alpha_i X_i$, where $\sum_{i=1}^{n} \alpha_i = 1$.
- Non-stationary: earlier observations given less weight. Stationary: $\alpha_i \equiv 1/n$.
- Evaluate performance of $\bar{X}$ by looking at its variance: s_n^2/n.

(Overlapping batches improves this measure of performance)

Nonoverlapping Batch Means

$$\begin{aligned} n &= 12 & \text{Total sample size} \\ m &= 4 & \text{Batch size} \\ k = \tfrac{n}{m} &= 3 & \text{Number of batches} \end{aligned}$$

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

Nonoverlapping Batch Means

$$n = 12 \quad \text{Total sample size}$$
$$m = 4 \quad \text{Batch size}$$
$$k = \frac{n}{m} = 3 \quad \text{Number of batches}$$

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

$$\bar{X}_1$$

Nonoverlapping Batch Means

$$\begin{aligned} n &= 12 && \text{Total sample size} \\ m &= 4 && \text{Batch size} \\ k &= \tfrac{n}{m} = 3 && \text{Number of batches} \end{aligned}$$

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

$$\underbrace{X^1 \; X^2 \; X^3 \; X^4}_{\bar{X}_1} \qquad \underbrace{X^5 \; X^6 \; X^7 \; X^8}_{\bar{X}_5}$$

Nonoverlapping Batch Means

$$n = 12 \quad \text{Total sample size}$$
$$m = 4 \quad \text{Batch size}$$
$$k = \frac{n}{m} = 3 \quad \text{Number of batches}$$

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

$$[\bar{X}_1] \qquad [\bar{X}_5] \qquad [\bar{X}_9]$$

Nonoverlapping Batch Means

$n = 12$	Total sample size
$m = 4$	Batch size
$k = \frac{n}{m} = 3$	Number of batches

$$\underbrace{X^1 \quad X^2 \quad X^3 \quad X^4}_{\bar{X}_1} \quad \underbrace{X^5 \quad X^6 \quad X^7 \quad X^8}_{\bar{X}_5} \quad \underbrace{X^9 \quad X^{10} \quad X^{11} \quad X^{12}}_{\bar{X}_9}$$

$$\widehat{\mathrm{Var}}_{k,m}(\bar{X}) = \frac{m}{n} \sum_{j=1}^{k} \frac{(\bar{X}_{m(j-1)}(m) - \bar{X})^2}{k-1}, \qquad \bar{X}_j(m) = \frac{1}{m} \sum_{i=0}^{m-1} X_{j+i}$$

Batch Means

Why not overlap the batches? (Meketon & Schmeiser, 1984)

Overlapping Batch Means

$$\begin{aligned} n = 12 &\quad \text{Total sample size} \\ m = 4 &\quad \text{Batch size} \\ k = \left\lfloor \frac{n}{m} \right\rfloor = 3 &\quad \text{Number of nonoverlapping batches} \end{aligned}$$

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

Overlapping Batch Means

$$\begin{aligned} n &= 12 && \text{Total sample size} \\ m &= 4 && \text{Batch size} \\ k = \left\lfloor \frac{n}{m} \right\rfloor &= 3 && \text{Number of nonoverlapping batches} \end{aligned}$$

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

$$[\text{———}\bar{X}_1\text{———}]$$

Overlapping Batch Means

$$
\begin{aligned}
n &= 12 && \text{Total sample size} \\
m &= 4 && \text{Batch size} \\
k = \left\lfloor \frac{n}{m} \right\rfloor &= 3 && \text{Number of nonoverlapping batches}
\end{aligned}
$$

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

$$[\text{———}\bar{X}_1\text{———}]$$

$$\langle\text{———}\bar{X}_2\text{———}\rangle$$

Overlapping Batch Means

$$
\begin{aligned}
n &= 12 && \text{Total sample size}\\
m &= 4 && \text{Batch size}\\
k = \left\lfloor \frac{n}{m} \right\rfloor &= 3 && \text{Number of nonoverlapping batches}
\end{aligned}
$$

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

$[\text{———}\bar{X}_1\text{———}]$

$\langle\text{———}\bar{X}_2\text{———}\rangle$

$\langle\text{———}\bar{X}_3\text{———}\rangle$

Overlapping Batch Means

$$n = 12 \quad \text{Total sample size}$$
$$m = 4 \quad \text{Batch size}$$
$$k = \left\lfloor \frac{n}{m} \right\rfloor = 3 \quad \text{Number of nonoverlapping batches}$$

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

$[$———$\bar{X}_1$———$]$

$\langle$———$\bar{X}_2$———$\rangle$

$\langle$———$\bar{X}_3$———$\rangle$

$\langle$———$\bar{X}_4$———$\rangle$

Overlapping Batch Means

$n = 12$	Total sample size
$m = 4$	Batch size
$k = \left\lfloor \frac{n}{m} \right\rfloor = 3$	Number of nonoverlapping batches

$$X^1 \quad X^2 \quad X^3 \quad X^4 \quad {\color{red}X^5} \quad {\color{red}X^6} \quad {\color{red}X^7} \quad {\color{red}X^8} \quad X^9 \quad X^{10} \quad X^{11} \quad X^{12}$$

$$[\text{———}\bar{X}_1\text{———}] \qquad {\color{red}[\text{———}\bar{X}_5\text{———}]}$$

$$\langle\text{———}\bar{X}_2\text{———}\rangle$$

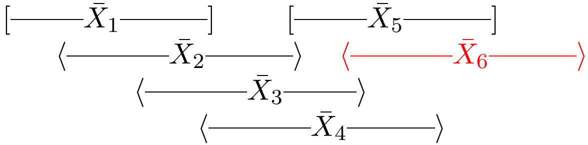
$$\langle\text{———}\bar{X}_3\text{———}\rangle$$

$$\langle\text{———}\bar{X}_4\text{———}\rangle$$

Overlapping Batch Means

$n = 12$	Total sample size
$m = 4$	Batch size
$k = \left\lfloor \frac{n}{m} \right\rfloor = 3$	Number of nonoverlapping batches

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



Overlapping Batch Means

$n = 12$	Total sample size
$m = 4$	Batch size
$k = \left\lfloor \frac{n}{m} \right\rfloor = 3$	Number of nonoverlapping batches

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

$[$———$\bar{X}_1$———$]$ $[$———$\bar{X}_5$———$]$

$\langle$———$\bar{X}_2$———$\rangle$ $\langle$———$\bar{X}_6$———$\rangle$

$\langle$———$\bar{X}_3$———$\rangle$ $\langle$———$\bar{X}_7$———$\rangle$

$\langle$———$\bar{X}_4$———$\rangle$

Overlapping Batch Means

$n = 12$	Total sample size
$m = 4$	Batch size
$k = \left\lfloor \frac{n}{m} \right\rfloor = 3$	Number of nonoverlapping batches

$$X^1 \quad X^2 \quad X^3 \quad X^4 \quad X^5 \quad X^6 \quad X^7 \quad \color{red}{X^8} \quad \color{red}{X^9} \quad \color{red}{X^{10}} \quad \color{red}{X^{11}} \quad X^{12}$$

$[\text{———}\bar{X}_1\text{———}]$ $[\text{———}\bar{X}_5\text{———}]$

$\langle\text{———}\bar{X}_2\text{———}\rangle$ $\langle\text{———}\bar{X}_6\text{———}\rangle$

$\langle\text{———}\bar{X}_3\text{———}\rangle$ $\langle\text{———}\bar{X}_7\text{———}\rangle$

$\langle\text{———}\bar{X}_4\text{———}\rangle$ $\color{red}{\langle\text{———}\bar{X}_8\text{———}\rangle}$

Overlapping Batch Means

$n = 12$	Total sample size
$m = 4$	Batch size
$k = \left\lfloor \frac{n}{m} \right\rfloor = 3$	Number of nonoverlapping batches

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

$$[\text{———}\bar{X}_1\text{———}] \qquad [\text{———}\bar{X}_5\text{———}] \qquad [\text{———}\bar{X}_9\text{———}]$$

$$\langle\text{———}\bar{X}_2\text{———}\rangle \qquad \langle\text{———}\bar{X}_6\text{———}\rangle$$

$$\langle\text{———}\bar{X}_3\text{———}\rangle \qquad \langle\text{———}\bar{X}_7\text{———}\rangle$$

$$\langle\text{———}\bar{X}_4\text{———}\rangle \qquad \langle\text{———}\bar{X}_8\text{———}\rangle$$

Overlapping Batch Means

$$
\begin{aligned}
n &= 12 \quad \text{Total sample size} \\
m &= 4 \quad \text{Batch size} \\
k = \left\lfloor \frac{n}{m} \right\rfloor &= 3 \quad \text{Number of nonoverlapping batches}
\end{aligned}
$$

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

$$[\text{———}\bar{X}_1\text{———}] \qquad [\text{———}\bar{X}_5\text{———}] \qquad [\text{———}\bar{X}_9\text{———}]$$

$$\langle\text{———}\bar{X}_2\text{———}\rangle \qquad \langle\text{———}\bar{X}_6\text{———}\rangle$$

$$\langle\text{———}\bar{X}_3\text{———}\rangle \qquad \langle\text{———}\bar{X}_7\text{———}\rangle$$

$$\langle\text{———}\bar{X}_4\text{———}\rangle \qquad \langle\text{———}\bar{X}_8\text{———}\rangle$$

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

Different Variance Estimators

Batch Mean $\bar{X}_j(m) = \frac{1}{m} \sum_{i=0}^{m-1} X_{j+i}$

Overall Mean $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

Nonoverlapping Variance $\widehat{\text{Var}}_{k,m}(\bar{X}) = \frac{m}{n} \sum_{j=1}^{k} \frac{(\bar{X}_{m(j-1)}(m) - \bar{X})^2}{k-1}$

Overlapping Variance $\widetilde{\text{Var}}_m\left(\bar{X}\right) = \frac{m}{n} \sum_{j=1}^{n-m+1} \frac{(\bar{X}_j(m) - \bar{X})^2}{n-2m+1}$

Important Results

- The variance estimator obtained by use of overlapping batches has essentially the same bias as the nonoverlapping estimator.
- Both estimators are unbiased in the limit as $n, m, k \to \infty$.
- Overlapping the batches reduces the variability of the variance estimator to 2/3 of the original.

$$\lim_{n,m,k\to\infty} \frac{\mathrm{Var}\left(\widetilde{\mathrm{Var}}_m\left(\bar{X}\right)\right)}{\mathrm{Var}\left(\widehat{\mathrm{Var}}^i_{k,m}(\bar{X})\right)} = \frac{2}{3}$$

Stochastic Optimization

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

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

where

f,	a real-valued function
$\mathbf{x}$,	a vector of decision variables
$\tilde{\xi}$,	a vector of random variables

Approximating Problem

We set up the approximating problem

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

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

where

$\tilde{\xi}^i$,	i.i.d. samples from the distribution of $\tilde{\xi}$, $i = 1, \ldots, n$
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 $\hat{\mathbf{x}} \in X$ (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.

Bounds on z^*

Upper Bound Since $\hat{\mathbf{x}}$ is probably not optimal, we can use $\mathbb{E} f(\hat{\mathbf{x}}, \tilde{\xi})$ as an upper bound on z^*.

Lower Bound $\mathbb{E} z_n^* = \mathbb{E}\left[\min_{\mathbf{x}\in X} \frac{1}{n}\sum_{i=1}^n f(\mathbf{x}, \tilde{\xi}^i)\right]$ is a lower bound on z^*.

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

(Mak & Morton (1999), Norkin, Pflug, Ruszczyński (1998))

Optimality Gap Estimation

The optimality gap $\mathbb{E}\left[f(\hat{\mathbf{x}}, \tilde{\xi})\right] - z^*$ is approximated by

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

- Since the approximated optimal solution is biased, the gap estimator is positively biased.
- Split our random sample into batches...

Multiple Replications Procedure

m, batch size; k, # of batches; $n = km$, total sample size

1. For $j = 1, \ldots, k$:
   1. Generate batch of samples $\tilde{\xi}^{(j-1)m+1}, \ldots, \tilde{\xi}^{jm}$.
   2. Find gap estimate
      $\bar{G}_m^j = \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} f(\hat{\mathbf{x}}, \tilde{\xi}^i) - \min_{\mathbf{x} \in X} \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} f(\mathbf{x}, \tilde{\xi}^i)$

Multiple Replications Procedure

m, batch size; k, # of batches; $n = km$, total sample size

1. For $j = 1, \ldots, k$:
   1. Generate batch of samples $\tilde{\xi}^{(j-1)m+1}, \ldots, \tilde{\xi}^{jm}$.
   2. Find gap estimate $\bar{G}_m^j = \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} f(\hat{\mathbf{x}}, \tilde{\xi}^i) - \min_{\mathbf{x} \in X} \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} f(\mathbf{x}, \tilde{\xi}^i)$
2. Calculate average gap estimate $\bar{\bar{G}} = \frac{1}{k} \sum_{j=1}^{k} \bar{G}_m^j$.

Multiple Replications Procedure

m, batch size; k, # of batches; $n = km$, total sample size

1. For $j = 1, \ldots, k$:
   1. Generate batch of samples $\tilde{\xi}^{(j-1)m+1}, \ldots, \tilde{\xi}^{jm}$.
   2. Find gap estimate
      $\bar{G}_m^j = \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} f(\hat{\mathbf{x}}, \tilde{\xi}^i) - \min_{\mathbf{x} \in X} \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} f(\mathbf{x}, \tilde{\xi}^i)$
2. Calculate average gap estimate $\bar{\bar{G}} = \frac{1}{k} \sum_{j=1}^{k} \bar{G}_m^j$.
3. Calculate sample variance $VG = \frac{1}{k(k-1)} \sum_{i=j}^{k} (\bar{G}_m^j - \bar{\bar{G}})^2$.

Multiple Replications Procedure

m, batch size; k, # of batches; $n = km$, total sample size

1. For $j = 1, \ldots, k$:
   1. Generate batch of samples $\tilde{\xi}^{(j-1)m+1}, \ldots, \tilde{\xi}^{jm}$.
   2. Find gap estimate $\bar{G}_m^j = \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} f(\hat{\mathbf{x}}, \tilde{\xi}^i) - \min_{\mathbf{x} \in X} \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} f(\mathbf{x}, \tilde{\xi}^i)$
2. Calculate average gap estimate $\bar{\bar{G}} = \frac{1}{k} \sum_{j=1}^{k} \bar{G}_m^j$.
3. Calculate sample variance $VG = \frac{1}{k(k-1)} \sum_{i=j}^{k} (\bar{G}_m^j - \bar{\bar{G}})^2$.
4. Find one-sided confidence interval $[0, \bar{\bar{G}} + t_{\alpha,k-1}\sqrt{VG}]$.

Overlapping Batches in MRP

- Now, we want to overlap the batches in the MRP.
- All of Meketon and Schmeiser's results occurred as $n, m, n/m \to \infty$. (Need $r \in (0,1)$, with $m \approx n^r$.)
- Choose sequences n_j, m_j to fit these properties as $j \to \infty$.

Assumptions

A1 Samples of the random variable $\tilde{\xi}$ are i.i.d.

A2 $z_n^* \to z^*$ a.s.

A3 $\exists \epsilon > 0$ such that $\mathbb{E}\left[(\sup_{x \in X} f(x, \tilde{\xi}))^{4+\epsilon}\right] < \infty$.

Partial Overlap

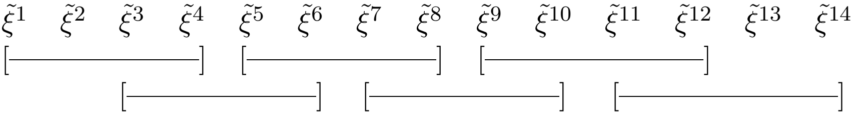


Figure: Graphical representation of $\gamma = 2$.

Let γ be the batch nonoverlap parameter.

- Number of new samples added to each new batch.
- $\gamma = 1$, maximally overlapping case discussed above.
- $\gamma = m$, classical nonoverlapping case.
- Could change along with n_j, m_j, use γ_j.

Definitions

A series of estimators

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

$$\bar{\bar{G}}_j = \frac{1}{|K(j)|} \sum_{i \in K(j)} \bar{G}^i_{m_j}$$

$$VG_j = \frac{m}{n} \frac{1}{df(j)} \sum_{i \in K(j)} (\bar{G}^i_{m_j} - \bar{\bar{G}}_j)^2$$

where

- $K(j)$, set of batches in the estimator (related to γ_j).
- $df(j)$, degrees of freedom of variance estimator.

Further Definitions

$$|K(j)| = \left\lfloor \frac{n_j - m_j}{\gamma_j} + 1 \right\rfloor$$

$$df(j) = \frac{n_j - k_j m_j}{\gamma_j} + (k_j - 2)\left\lceil \frac{m_j}{\gamma_j} \right\rceil + 1$$

Special cases:

Nonoverlapping	Maximally Overlapping
$\gamma_j = m_j$	$\gamma_j \equiv 1$
$\lvert K(j)\rvert = k_j,$	$\lvert K(j)\rvert = n_j - m_j + 1$
$df(j) = k_j - 1$	$df(j) = n_j - 2m_j + 1$

A Small Problem

Problem The internal minimization of

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

means we do not have the nice properties we want.

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

Beginnings of Proof

Unbiased Gap Estimator

$$\bar{D}^i_{m_j} = \frac{1}{m_j}\sum_{I=i}^{i+m_j-1} f(\hat{\mathbf{x}}, \tilde{\xi}^I) - \frac{1}{m_j}\sum_{I=i}^{i+m_j-1} f(\mathbf{x}^*, \tilde{\xi}^I)$$

$$\bar{\bar{D}}_j = \frac{1}{|K(j)|}\sum_{i\in K(j)} \bar{D}^i_{m_j}$$

$$VD_j = \frac{m}{n}\frac{1}{df(j)}\sum_{i\in K(j)} (\bar{D}^i_{m_j} - \bar{\bar{D}}_j)^2$$

- By (A2), $\bar{G}^i_m \to \bar{D}^i_m$ as $m \to \infty$.
- By Meketon, VD_j has decreased variance for overlapping batches.
- Show that $\mathrm{Var}(VG_j) \to \mathrm{Var}(VD_j)$. Equivalently, $VG_j \xrightarrow{L^2} VD_j$.

Sketch of Proof

1. Show that $VG_j \xrightarrow{P} VD_j$.
2. Show that VG_j^2 is uniformly integrable.
3. Then $VG_j \xrightarrow{L^2} VD_j$.

Convergence in Probability

Easier to show convergence in expectation

$$\mathbb{E}\left[|VG_j - VD_j|\right] = \frac{1}{df(j)} \sum_{i \in K(j)} \mathbb{E}\left[|(\bar{G}^i_{m_j} - \bar{\bar{G}}_j)^2 - (\bar{D}^i_{m_j} - \bar{\bar{D}}_j)^2\right]$$

$$\leq 2\mathbb{E}\left[|(\bar{G}^i_{m_j})^2 - (\bar{D}^i_{m_j})^2| \right.$$

$$\left. +2(\bar{G}^i_{m_j}\bar{\bar{G}}_j - \bar{D}^i_{m_j}\bar{\bar{D}}_j) + |\bar{\bar{G}}^2_j - \bar{\bar{D}}^2_j|\right].$$

All terms converge to zero pointwise (A2). Dominating function exists by (A3).

Uniform Integrability

Let

$$h(\tilde{\xi}^i) = \sup_{\hat{\mathbf{x}} \in X} f(\hat{\mathbf{x}}, \tilde{\xi}^i) - \min_{x \in X} f(x, \tilde{\xi}^i)$$

Then

$$\mathrm{Var}\left(VG_j^2\right) \le 4\mathbb{E}\left[(\bar{G}^1_{m_j})^4\right] \le 4\mathbb{E}\left[h^4(\tilde{\xi})\right]$$

and for uniform integrability

$$\begin{aligned}\mathbb{E}\left[(\bar{G}_1^m)^4 I_{(\bar{G}_1^m)^4 > t}\right] &\le \mathbb{E}\left[h^4(\tilde{\xi}^i) I_{(\bar{G}_1^m)^4 > t}\right] \\ &\le \sqrt[1+\epsilon/4]{\mathbb{E}\left[h^{4+\epsilon}(\tilde{\xi}^i)\right]} \sqrt[(4+\epsilon)/\epsilon]{\mathbb{P}\left\{(\bar{G}_1^m)^4 > t\right\}}\end{aligned}$$

And the $\mathbb{P}\left\{(\bar{G}_1^m)^4 > t\right\} \to 0$ as $t \to \infty$.

Result

Then $\text{Var}(VG_j) \to \text{Var}(VD_j)$, so we get the same variance reduction as we did in simulation!

Variance reduction

- $\gamma_j \equiv 1$, variance is reduced to $\dfrac{2}{3}$ of original.
- $\gamma_j = \dfrac{1}{3}m$, variance is reduced to $\dfrac{19}{27}$ of original.
- $\gamma_j = \dfrac{1}{2}m$, variance is reduced to $\dfrac{3}{4}$ of original.
- $\gamma_j = \dfrac{1}{N}m$, variance is reduced to $\dfrac{2N^2+1}{3N^2}$ of original.

Confidence Intervals

- Finally, given the above, we want to generate some one-sided confidence intervals on the true optimality gap, $\mu_{\hat{\mathbf{x}}}$.
- The interval estimators take the form

$$I_{G_j} = \bar{\bar{G}}_j + t_{\alpha,k-1}\sqrt{VG_j}$$
$$I_{D_j} = \bar{\bar{D}}_j + t_{\alpha,k-1}\sqrt{VD_j},$$

  for some $\alpha \in (0,1)$.
- We want to show that $\mathbb{P}\left\{\mu_{\hat{\mathbf{x}}} \leq I_{G_j}\right\}$ is a $100(1-\alpha)\%$ confidence interval.
- That is, $\mathbb{P}\left\{\mu_{\hat{\mathbf{x}}} \leq I_{G_j}\right\} \to 1-\alpha$.

Convergence of CI

We know that $I_{G_j} \xrightarrow{P} I_{D_j}$. The function $f(X) = 1_{X \in I}$ for some interval I is continuous a.s. for continuous random variables X. Then

$$\mathbb{P}\left\{ |1_{\mu_{\hat{\mathbf{x}}} \leq I_{G_j}} - 1_{\mu_{\hat{\mathbf{x}}} \leq I_{D_j}}| > \epsilon \right\} \rightarrow 0$$

and so we must have

$$\mathbb{P}\left\{ \mu_{\hat{\mathbf{x}}} \leq I_{G_j} \right\} \rightarrow \mathbb{P}\left\{ \mu_{\hat{\mathbf{x}}} \leq I_{D_j} \right\}$$

Result

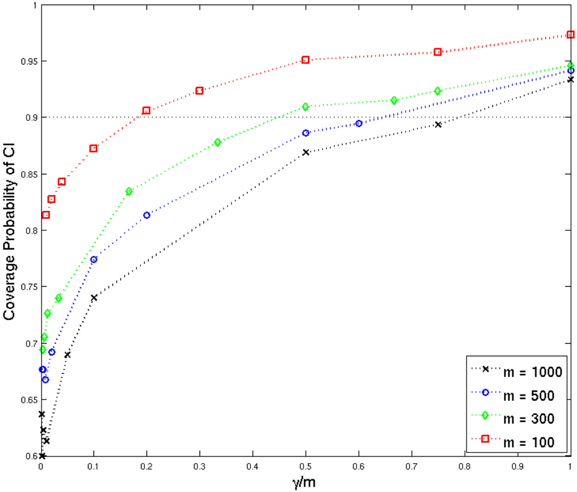
- We know from simulation theory the conditions that $\mathbb{P}\left\{\mu_{\hat{\mathbf{x}}} \leq I_{D_j}\right\} \to 1 - \alpha$.
- Now, we have an asymptotically valid confidence interval for stochastic optimization!

Newsvendor Problem

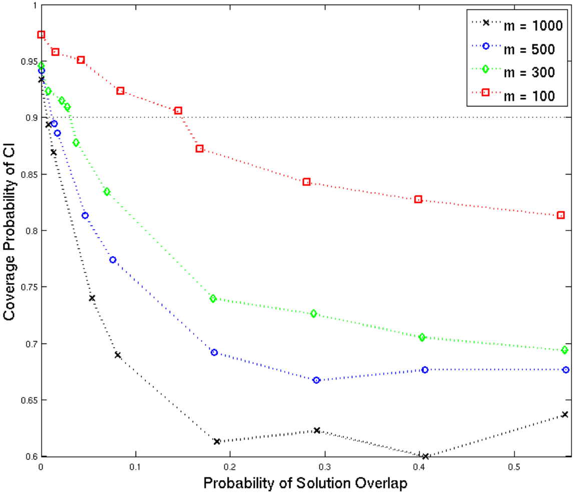
We tested the computational properties of our method on the newsvendor problem, with

- Selling price $r = 15$.
- Cost of paper $c = 5$.
- Distribution of $\tilde{d}$ is $U(0, 10)$.
- Optimal quantile $\frac{r-c}{r} = \frac{2}{3}$.

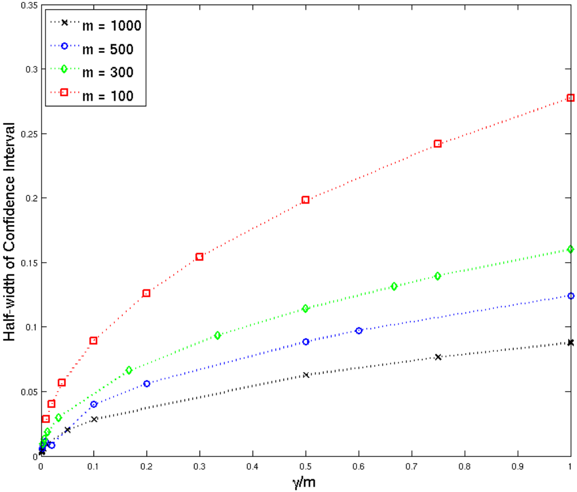
Coverage Probability with $k = 30$



Coverage Probability with $k = 30$



Width of CI with $k = 30$



Bibliography

1. C. Alexopoulos, N. T. Argon, D. Goldsman, G. Tokol, and J. R. Wilson. Overlapping Variance Estimators for Simulation. *Operations Research*, 55(6):1090-1103, 2007.
2. 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.
3. M. S. Meketon and B. Schmeiser. Overlapping batch means: something for nothing? In WSC *84: Proceedings of the 16th conference on Winter simulation*, pages 226-230, Piscataway, NJ, USA, 1984. IEEE Press.