Online Supplement to
"Efficient Computation of Overlapping Variance Estimators for Simulation"
INFORMS Journal on Computing

Christos Alexopoulos
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA, christos@isye.gatech.edu

Nilay Tanık Argon
Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, Wisconsin 53706-1572, USA, nilay@engr.wisc.edu

David Goldsman
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA, sman@gatech.edu

Natalie M. Steiger
Maine Business School, University of Maine, Orono, Maine 04469-5723, USA, nsteiger@maine.edu

Gamze Tokol
Decision Analytics, Atlanta, Georgia 30306, USA, gamze@mindspring.com

James R. Wilson
Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Campus Box 7906, Raleigh, North Carolina 27695-7906, USA, jwilson@ncsu.edu

The purpose of this online supplement is to present the following: (i) linear time algorithms to obtain overlapping area and Cramér–von Mises estimators; (ii) experimental results on the bias, variance, mean-squared error, and efficiency of the NBM and overlapping variance estimators; and (iii) analytical results on the bias and variance of the overlapping estimators when they are applied to i.i.d. normal data.

---

1. Algorithms

We start with the algorithms for computing overlapping area and CvM estimators with several weighting functions in $O(n)$ time. Figures 1 and 2 present the algorithms for obtaining the overlapping area estimator $\mathcal{A}^{\mathrm{O}}(f; b, m)$ for weighting functions $f_r(t) = \sum_{p=0}^{r} c_p t^p$ and $f_{\cos,j}(t) = \sqrt{8}\pi j \cos(2\pi j t)$, respectively. Figure 3 presents an algorithm to obtain the overlapping CvM estimator $\mathcal{C}^{\mathrm{O}}(g; b, m)$ with weighting function $g(t) = \sum_{p=0}^{r} c_p t^p$.

Figure 1: Algorithm for the Overlapping Area Estimator with Polynomial Weighting Functions

Step 1: Initialization

$V \leftarrow 0$, $a_0 \leftarrow 0$, $a_1 \leftarrow 0$, $S^{(p)} \leftarrow 0$ for $p = 0, \ldots, r$, and $k \leftarrow 1$

Step 2: Calculate $S^{(p)}_{0,m}$ for $p = 0, \ldots, r$; $a_0(f_r)$; $a_1(f_r)$; and $V_0(f_r)$

Repeat:

$S^{(p)} \leftarrow S^{(p)} + k^p Y_k$ for $p = 0, \ldots, r$; $V \leftarrow V - f_r(k/m)S^{(0)}$

$a_0 \leftarrow a_0 + f_r(k/m)$, $a_1 \leftarrow a_1 + k f_r(k/m)$

$k \leftarrow k + 1$

Until $k = m + 1$

$V \leftarrow V + S^{(0)} a_1/m$

Step 3: Calculate $W_i(f_r)$, $V_i(f_r)$ for $i = 1, \ldots, n - m$; accumulate $\mathcal{A}^{\text{O}}(f_r; b, m)$ for all overlapping batches

$\mathcal{A}^{\text{O}} \leftarrow V^2/m^3$ and $i \leftarrow 1$

Repeat:

Calculate $S^{(p)}_{i,m}$ backwards in p to save storage space

$p \leftarrow r$

Repeat:

$S^{(p)} \leftarrow \sum_{x=0}^{p} \binom{p}{x}(-1)^{p-x} S^{(x)} + m^p Y_{i+m} - Y_i I_{\{p=0\}}$, $p \leftarrow p - 1$

Until $p = -1$

$W \leftarrow \sum_{p=0}^{r} c_p S^{(p)}/m^p$, $V \leftarrow V + (Y_{i+m} - Y_i)a_1/m - W + Y_i a_0$

$\mathcal{A}^{\text{O}} \leftarrow \mathcal{A}^{\text{O}} + V^2/m^3$, $i \leftarrow i + 1$

Until $i = n - m + 1$

Return $\mathcal{A}^{\text{O}} \leftarrow \mathcal{A}^{\text{O}} \big/ (n - m + 1)$

2. Efficiency of the Variance Estimators

In Alexopoulos et al. (2006), we summarize the results of an experimental performance evaluation of the new overlapping variance estimators; and we find that they perform efficaciously compared with their nonoverlapping counterparts—at least with respect to the measures of bias and variance. In particular, there is little difference between corresponding nonoverlapping and overlapping estimators when it comes to bias; yet overlapping variance estimators often achieve substantial savings compared with nonoverlapping variance estimators when it comes to variance. We have also shown that all of the overlapping estimators studied herein require computation on the same order of magnitude as that of NBM, the simplest estimator in terms of computation.

A natural question to ask is: just how equal are the $O(n)$ computational efforts required by the various variance estimators? We might reasonably expect that among these $O(n)$ estimators, NBM requires less work than any of the batched STS estimators, which in turn require less work than the corresponding overlapping STS estimators. The good news is that, for all practical purposes, the differences in computational effort

Figure 2: Algorithm for the Overlapping Area Estimator with Trigonometric Weighting Functions

Step 1: Initialization

$j \leftarrow$ user-selected positive integer, $V \leftarrow 0$, $a_0 \leftarrow 0$, $a_1 \leftarrow 0$, $Q \leftarrow 0$, $Q' \leftarrow 0$, $S^{(0)} \leftarrow 0$, and $k \leftarrow 1$

Step 2: Calculate $S^{(0)}_{0,m}$, Q_m, Q'_m, $a_0(f_{\cos,j})$, $a_1(f_{\cos,j})$, and $V_0(f_{\cos,j})$

Repeat:

$$S^{(0)} \leftarrow S^{(0)} + Y_k, \quad V \leftarrow V - f_{\cos,j}(k/m)S^{(0)}$$
$$Q \leftarrow Q + \cos(\alpha_j k)Y_k, \quad Q' \leftarrow Q' + \sin(\alpha_j k)Y_k$$
$$a_0 \leftarrow a_0 + f_{\cos,j}(k/m), \quad a_1 \leftarrow a_1 + k f_{\cos,j}(k/m), \quad k \leftarrow k+1$$

Until $k = m+1$

$V \leftarrow V + S^{(0)} a_1/m$

Step 3: Calculate Q_i, Q'_i, $W_i(f_{\cos,j})$, $V_i(f_{\cos,j})$ for $i = 1, \ldots, n-m$; accumulate $\mathcal{A}^{\mathrm{O}}(f_{\cos,j}; b, m)$ for all overlapping batches

$\mathcal{A}^{\mathrm{O}} \leftarrow V^2/m^3$ and $i \leftarrow 1$

Repeat:

$$Q \leftarrow Q + \cos(\alpha_j(i+m))Y_{i+m} - \cos(\alpha_j i)Y_i$$
$$Q' \leftarrow Q' + \sin(\alpha_j(i+m))Y_{i+m} - \sin(\alpha_j i)Y_i$$
$$W \leftarrow \sqrt{8}\pi j(\cos(\alpha_j i)Q + \sin(\alpha_j i)Q')$$
$$V \leftarrow V + (Y_{i+m} - Y_i)a_1/m - W + Y_i a_0$$
$$\mathcal{A}^{\mathrm{O}} \leftarrow \mathcal{A}^{\mathrm{O}} + V^2/m^3, \quad i \leftarrow i+1$$

Until $i = n-m+1$

Return $\mathcal{A}^{\mathrm{O}} \leftarrow \mathcal{A}^{\mathrm{O}}/(n-m+1)$

are meaningless, especially when one regards the observations as "expensive" compared to the subsequent computational costs once the observations are in hand.

Nevertheless, we undertook a great deal of Monte Carlo work to evaluate the actual effort necessary to compute the various variance estimators. In one of our simplest experiments, we generated $n = 160{,}000$ observations of the waiting time process for the $M/M/1$ queue with 80% traffic intensity; and we organized the resulting time series into batches of size $m = 4{,}000$, 8,000, and 16,000 so that we could obtain precise estimates of the total computing time $\mathbb{C}$, bias $\mathbb{B}$, variance $\mathbb{V}$, and mean-squared error $\mathbb{M}$ of the corresponding variance estimators $\mathcal{N}(b,m)$, $\mathcal{A}^{\mathrm{O}}(f_0; b, m)$, $\mathcal{A}^{\mathrm{O}}(f_2; b, m)$, $\mathcal{C}^{\mathrm{O}}(g_0; b, m)$, and $\mathcal{C}^{\mathrm{O}}(g_2^\star; b, m)$ for $b = 40$, 20, and 10. Table 1 reports the results based on 16,000 i.i.d. replications of each of the variance estimators under study.

Each total computing-time entry $\mathbb{C}$ in Table 1 is the sum of two components:

1. On an IBM T43p laptop computer with an Intel Pentium M processor having a speed of 2.13 GHz and

Figure 3: Algorithm for the Overlapping CvM Estimator with Polynomial Weighting Functions

Step 1: Initialization

$S \leftarrow 0$, $M \leftarrow 0$, $J \leftarrow 0$, $a_p \leftarrow 0$ for $p \in \{0, 1, 2\}$, $S^{(p)} \leftarrow 0$ for $p = 0, \ldots, r+1$

$Z^{(p)} \leftarrow 0$ for $p = 0, \ldots, r$, $\ell \leftarrow 1$, and $k \leftarrow 1$

Step 2: Calculate $\boldsymbol{S_{0,m}^{(p)}}$ for $p = 0, \ldots, r+1$, $\boldsymbol{U_\ell}$ for $\boldsymbol{\ell = 2, \ldots, m+1}$, $\boldsymbol{M(1)}$, and $\boldsymbol{a_p(g)}$ for $\boldsymbol{p \in \{0, 1, 2\}}$; initialize the second term of (4) from Alexopoulos et al. (2006)

Repeat:

  $S^{(p)} \leftarrow S^{(p)} + k^p Y_k$ for $p = 0, \ldots, r+1$, $\quad J \leftarrow J + g(k/m)kS^{(0)}/m$

  if $\ell = 1$

    $U_2 \leftarrow g(1/m)Y_1$

  else

    Calculate $\boldsymbol{Z_{\ell,1,\ell-1}^{(p)}}$ backwards in $\boldsymbol{p}$ to save storage space

    $p \leftarrow r$

    Repeat:

      $Z^{(p)} \leftarrow \sum_{x=0}^{p} \binom{p}{x} Z^{(x)} + Y_{\ell-1}, \quad p \leftarrow p-1$

    Until $p = -1$

    $U_{\ell+1} \leftarrow U_\ell + Y_\ell a_0 - \sum_{p=0}^{r} c_p Z^{(p)}/m^p + g(\ell/m)S^{(0)}$

  endif

  $a_p \leftarrow a_p + k^p g(k/m)$ for $p \in \{0, 1, 2\}$, $\quad M \leftarrow M + g(k/m)S^{(0)}, \quad \ell \leftarrow \ell + 1, \quad k \leftarrow \ell$

Until $\ell = m + 1$

Step 3: Calculate $\boldsymbol{W_i(g_{r+1})}$ for $\boldsymbol{i = 1, \ldots, n-m}$, $\boldsymbol{U_\ell}$ for $\boldsymbol{\ell = m+2, \ldots n}$, $\boldsymbol{M(\ell)}$ for $\boldsymbol{\ell = 2, \ldots, n-m+1}$; accumulate the third term of (4) from Alexopoulos et al. (2006) and $\boldsymbol{\mathcal{C}^{\mathrm{O}}(g; b, m)}$

$K \leftarrow M$, $G \leftarrow a_0$, $R \leftarrow Y_1 M$, $\mathcal{C}^{\mathrm{O}} \leftarrow a_2(S^{(0)}/m)^2 - 2S^{(0)}J$, $i \leftarrow 1$

Repeat:

  Calculate $\boldsymbol{S_{i,m}^{(p)}}$ backwards in $\boldsymbol{p}$ to save storage space

  $p \leftarrow r + 1$

  Repeat:

    $S^{(p)} \leftarrow \sum_{x=0}^{p} \binom{p}{x} (-1)^{p-x} S^{(x)} + m^p Y_{i+m} - Y_i I_{\{p=0\}}, \quad p \leftarrow p-1$

  Until $p = -1$

  $W \leftarrow \sum_{p=0}^{r} c_p S^{(p+1)}/m^{p+1}, \quad J \leftarrow J + W - Y_i a_1/m$

  $K \leftarrow K + \sum_{p=0}^{r} c_p S^{(p)}/m^p - a_0 Y_{\ell-m}, \quad M \leftarrow M + K - U_{\ell-m+1}, \quad R \leftarrow R + Y_{\ell-m+1}M$

  if $\ell \leq n - m + 1$

    Calculate $\boldsymbol{Z_{\ell,1,m}^{(p)}}$ backwards in $\boldsymbol{p}$ to save storage space

    $p \leftarrow r$

    Repeat:

      $Z^{(p)} \leftarrow \sum_{x=0}^{p} \binom{p}{x} Z^{(x)} + Y_{\ell-1} - (m+1)^p Y_{\ell-m-1} I_{\{\ell \geq m+2\}}, \quad p \leftarrow p-1$

    Until $p = -1$

    $U_{\ell+1} \leftarrow U_\ell + Y_\ell a_0 - \sum_{p=0}^{r} c_p Z^{(p)}/m^p$

  else

    $G \leftarrow G - g((\ell - n + m - 2)/m), \quad S \leftarrow S + Y_\ell, \quad p \leftarrow r$

    Calculate $\boldsymbol{Z_{\ell,\ell-1-(n-m),m}^{(p)}}$ backwards in $\boldsymbol{p}$ to save storage space

    Repeat:

      $Z^{(p)} \leftarrow \sum_{x=0}^{p} \binom{p}{x} Z^{(x)} + Y_{\ell-1} I_{\{\ell = n-m+2\}} - (m+1)^p Y_{\ell-(m+1)}, \quad p \leftarrow p-1$

    Until $p = -1$

    $U_{\ell+1} \leftarrow U_\ell + Y_\ell G - \sum_{p=0}^{r} c_p Z^{(p)}/m^p - g((\ell - 1 - n + m)/m)S$

  endif

  $\mathcal{C}^{\mathrm{O}} \leftarrow \mathcal{C}^{\mathrm{O}} + a_2(S^{(0)}/m)^2 - 2S^{(0)}J, \quad \ell \leftarrow \ell + 1, \quad i \leftarrow i + 1$

Until $\ell = n + 1$

$\ell \leftarrow n - m + 2$

Step 4: Calculate $\boldsymbol{M(\ell)}$ for $\boldsymbol{\ell = n-m+2, \ldots, n}$; accumulate the third term of (4) and $\boldsymbol{\mathcal{C}^{\mathrm{O}}(g; b, m)}$

Repeat:

  $M \leftarrow M - U_\ell, \quad R \leftarrow R + Y_\ell M, \quad \ell \leftarrow \ell + 1$

Until $\ell = n + 1$

Return $\mathcal{C}^{\mathrm{O}} \leftarrow (\mathcal{C}^{\mathrm{O}} + R)/(m^2(n - m + 1))$

Table 1: Efficiency Analysis of the Variance Estimators

Performance	Variance Estimator				
Measure	$\mathcal{N}(b,m)$	$\mathcal{A}^{\mathrm{O}}(f_0;b,m)$	$\mathcal{A}^{\mathrm{O}}(f_2;b,m)$	$\mathcal{C}^{\mathrm{O}}(g_0;b,m)$	$\mathcal{C}^{\mathrm{O}}(g_2^\star;b,m)$
Results for $m=4{,}000$ and $b=40$					
$\mathbb{C}_{\text{v.est}}$ (sec)	0.005396	0.05525	0.06967	0.05276	0.05702
$\mathbb{C}$ (sec)	4.225	4.275	4.290	4.273	4.277
$\mathbb{B}$	–28.25	–74.20	–12.83	–120.1	–15.92
$\mathbb{V}$	430,984	287,271	332,457	237,308	326,708
$\mathbb{M}$	431,782	292,777	332,621	251,741	326,961
$\mathbb{Q}$	1	0.6744	0.7831	0.5568	0.7673
$\mathbb{Q}'$	1	0.6861	0.7821	0.5896	0.7665
Results for $m=8{,}000$ and $b=20$					
$\mathbb{C}_{\text{v.est}}$ (sec)	0.005344	0.05397	0.06926	0.05106	0.05989
$\mathbb{C}$ (sec)	4.225	4.274	4.289	4.271	4.280
$\mathbb{B}$	–23.89	–37.14	–3.06	–60.32	–6.825
$\mathbb{V}$	637,247	379,785	426,471	313,398	417,670
$\mathbb{M}$	637,817	381,164	426,481	317,036	417,717
$\mathbb{Q}$	1	0.6028	0.6794	0.4971	0.6639
$\mathbb{Q}'$	1	0.6044	0.6788	0.5024	0.6634
Results for $m=16{,}000$ and $b=10$					
$\mathbb{C}_{\text{v.est}}$ (sec)	0.005359	0.05138	0.06845	0.06373	0.08658
$\mathbb{C}$ (sec)	4.225	4.271	4.288	4.284	4.307
$\mathbb{B}$	–14.15	–22.79	–4.44	–31.57	–11.90
$\mathbb{V}$	1,084,522	549,573	618,916	430,888	600,620
$\mathbb{M}$	1,084,722	550,092	618,936	431,885	600,762
$\mathbb{Q}$	1	0.5123	0.5792	0.4028	0.5645
$\mathbb{Q}'$	1	0.5127	0.5791	0.4037	0.5645

1.00 GB of RAM, we computed the average execution time $\mathbb{C}_{\text{sim}} = 4.22$ seconds per run for an Arena simulation model (Kelton, Sadowski, and Sturrock 2004) of the $M/M/1$ queue in which 160,000 customer waiting times are merely computed and tabulated in an Arena Record module before the associated entity is disposed.

2. On the same laptop computer, we computed the average execution time $\mathbb{C}_{\text{v.est}}$ (in seconds) to compute the variance estimators $\mathcal{N}(b, m)$, $\mathcal{A}^{\text{O}}(f_0; b, m)$, $\mathcal{A}^{\text{O}}(f_2; b, m)$, $\mathcal{C}^{\text{O}}(g_0; b, m)$, and $\mathcal{C}^{\text{O}}(g_2^\star; b, m)$ from a set of $n = 160{,}000$ queue waiting times for the $M/M/1$ queue *that were already resident in computer memory*. To compute $\mathbb{C}_{\text{v.est}}$ for, say, the NBM variance estimator $\mathcal{N}(b, m)$, we used the PORTLIB function DTIME of Fortran PowerStation (Microsoft Corp. 1995) to obtain the total CPU time required to compute 16,000 replications of $\mathcal{N}(b, m)$; and then a similar experiment was performed for the overlapping variance estimators $\mathcal{A}^{\text{O}}(f_0; b, m)$, $\mathcal{A}^{\text{O}}(f_2; b, m)$, $\mathcal{C}^{\text{O}}(g_0; b, m)$, and $\mathcal{C}^{\text{O}}(g_2^\star; b, m)$ separately.

Thus for each variance estimator in Table 1, we report $\mathbb{C}_{\text{sim}}$, $\mathbb{C}_{\text{v.est}}$, and $\mathbb{C} = \mathbb{C}_{\text{sim}} + \mathbb{C}_{\text{v.est}}$. We take the NBM variance estimator $\mathcal{N}(b, m)$ as the baseline with performance measures $\mathbb{C}_0$, $\mathbb{B}_0$, $\mathbb{V}_0$, and $\mathbb{M}_0$; and for any other variance estimator with corresponding performance measures $\mathbb{C}_1$, $\mathbb{B}_1$, $\mathbb{V}_1$, and $\mathbb{M}_1$, we see that the relative efficiency statistics are given by

$$\mathbb{Q} = \mathbb{C}_1\mathbb{V}_1 \big/ \big(\mathbb{C}_0\mathbb{V}_0\big) \quad \text{and} \quad \mathbb{Q}' = \mathbb{C}_1\mathbb{M}_1 \big/ \big(\mathbb{C}_0\mathbb{M}_0\big). \tag{1}$$

On any reasonably sized practical applications, the estimator computation times are insignificant compared to the times required to obtain the observations, thus rendering NBM noncompetitive in terms of the relative efficiency measures $\mathbb{Q}$ and $\mathbb{Q}'$.

3. Exact Results for i.i.d. Normal Case

The purpose here is simply to see whether the new estimators perform as advertised on an example for which we are able to calculate the expected value and variance of the variance estimators exactly. Suppose that we sample from an i.i.d. standard normal process, $\{Y_i : i = 1, \ldots, n\}$. After some tedious algebra that is detailed below, we obtain the results displayed in Table 2 describing the bias and variance properties of some of the proposed overlapping variance estimators. In Table 2 all the bias results are exact; on the other hand, in the tabulated expressions for the variance of the overlapping estimators, we omitted terms of the form

$O(1/m)$ (for fixed b). We see that the entries in Table 2 indeed match up nicely with those in Table 1 of Alexopoulos et al. (2006) with $\gamma = 0$. Of course, in this very special case, the natural estimator for $\sigma^2 = 1$ would have been the usual sample variance S^2, which is unbiased and has variance $2/(n-1)$. What follows are the derivations of the entries appearing in Table 2.

Table 2: Bias and Variance for Variance Estimators in I.i.d. Normal Case

Estimator	Bias	Variance
$\mathcal{A}^{\mathrm{O}}(f_0; b, m)$	$-\frac{1}{m^2}$	$\frac{24b-31}{35(b-1)^2}$
$\mathcal{A}^{\mathrm{O}}(f_2; b, m)$	$\frac{7}{2m^2} + \frac{63}{2m^4} - \frac{36}{m^6}$	$\frac{3514b-4359}{4290(b-1)^2}$
$\mathcal{C}^{\mathrm{O}}(g_0; b, m)$	$-\frac{1}{m^2}$	$\frac{88b-115}{210(b-1)^2}$
$\mathcal{C}^{\mathrm{O}}(g_2^\star; b, m)$	$\frac{4}{m^2} - \frac{5}{m^4}$	$\frac{10768b-13605}{13860(b-1)^2}$
$\mathcal{O}(b, m)$	0	$\frac{4b^3-11b^2+4b+6}{3(b-1)^4}$

Overlapping Area Estimator

We first obtain exact results for the overlapping area estimator. Using the arguments found in, e.g., §4 of Foley and Goldsman (1999), one can apply a little algebraic elbow grease to obtain

$$A_i^{\mathrm{O}}(f; m) \;=\; \frac{1}{m^3}\left[\sum_{j=1}^{m} h(f;\, j, m) Y_{i+j-1}\right]^2,$$

$i = 1, \ldots, n-m+1$, where

$$h(f;\, j, m) \;\equiv\; \sum_{\ell=1}^{m} \tfrac{\ell}{m} f(\tfrac{\ell}{m}) - \sum_{\ell=j}^{m} f(\tfrac{\ell}{m}),$$

and where we notice that the $h(f;\, j, m)$'s need only be calculated once beforehand.

We now calculate the mean and variance of the overlapping area estimator for the current example. First, we give a general expression for the mean. Since the Y_i's are i.i.d. with unit variance, we have

$$\begin{aligned} \mathrm{E}[\mathcal{A}^{\mathrm{O}}(f;b,m)] &= \mathrm{E}[A_1^{\mathrm{O}}(f;m)] \\ &= \frac{1}{m^3}\mathrm{E}\left[\left(\sum_{j=1}^{m} h(f;j,m)Y_j\right)^2\right] \\ &= \frac{1}{m^3}\mathrm{Var}\left(\sum_{j=1}^{m} h(f;j,m)Y_j\right) \\ &= \frac{1}{m^3}\sum_{j=1}^{m} h^2(f;j,m). \end{aligned} \tag{2}$$

Before moving ahead to the variance of the overlapping area estimator, we state a useful lemma.

Lemma 1 (Patel and Read 1996) If (Z_1, Z_2) is bivariate normal with marginal means equal to zero, then $\mathrm{Cov}(Z_1^2, Z_2^2) = 2\mathrm{Cov}^2(Z_1, Z_2)$.

In addition, define for $k = 0, 1, 2, \ldots$,

$$\begin{aligned} R_k^A &\equiv \mathrm{Cov}(A_1^{\mathrm{O}}(f;m), A_{1+k}^{\mathrm{O}}(f;m)) \\ &= \mathrm{Cov}\left(\frac{1}{m^3}\left(\sum_{i=1}^{m} h(f;i,m)Y_i\right)^2, \frac{1}{m^3}\left(\sum_{j=1}^{m} h(f;j,m)Y_{k+j}\right)^2\right) \\ &= \frac{2}{m^6}\mathrm{Cov}^2\left(\sum_{i=1}^{m} h(f;i,m)Y_i, \sum_{j=1}^{m} h(f;j,m)Y_{k+j}\right) \quad \text{(by Lemma 1)} \\ &= \frac{2}{m^6}\left[\sum_{i=1}^{m}\sum_{j=1}^{m} h(f;i,m)h(f;j,m)\mathrm{Cov}(Y_i, Y_{k+j})\right]^2 \\ &= \frac{2}{m^6}\left[\sum_{i=k+1}^{m} h(f;i,m)h(f;i-k,m)\right]^2, \end{aligned} \tag{3}$$

since the Y_i's are i.i.d. Then the definition of $\mathcal{A}^{\mathrm{O}}(f;b,m)$ and stationarity imply that

$$\begin{aligned} &(n-m+1)\,\mathrm{Var}(\mathcal{A}^{\mathrm{O}}(f;b,m)) \\ &\quad= \frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\sum_{j=1}^{n-m+1}\mathrm{Cov}(A_i^{\mathrm{O}}(f;m), A_j^{\mathrm{O}}(f;m)) \\ &\quad= R_0^A + 2\sum_{k=1}^{n-m}\left(1-\frac{k}{n-m+1}\right)R_k^A \\ &\quad= R_0^A + 2\sum_{k=1}^{m-1}\left(1-\frac{k}{n-m+1}\right)R_k^A, \end{aligned} \tag{4}$$

again since the underlying observations are i.i.d.

Now we see what happens to the area and variance of the overlapping area estimator when we apply specific weighting functions to the i.i.d. standard normal process.

Example 1 For the overlapping area estimator with constant weighting function $f_0(t) = \sqrt{12}$, it is easy to show that

$$h(f_0; j, m) \;=\; \sqrt{12}\Big(j - \frac{m+1}{2}\Big).$$

Plugging this into Equations (2)–(4) leads to

$$\mathrm{E}[\mathcal{A}^{\mathrm{O}}(f_0; b, m)] \;=\; 1 - \frac{1}{m^2},$$

which matches up with the appropriate entry in Table 1 of Alexopoulos et al. (2006) (taking into account that $\gamma = 0$), and

$$\mathrm{Var}(\mathcal{A}^{\mathrm{O}}(f_0; b, m)) \;=\; \frac{(m^2-1)\big[m^5(24b-31) + 24m^4 + 66m^3 b + o(m^3 b)\big]}{35m^5\big[m(b-1)+1\big]^2} \;\doteq\; \frac{24b-31}{35(b-1)^2},$$

for large m and b, which is in synch with the appropriate variance entry from Table 1 of Alexopoulos et al. (2006). ◁

Example 2 Consider the first-order unbiased overlapping area estimator with quadratic weights $f_2(t) = \sqrt{840}(3t^2 - 3t + 1/2)$. After some algebra, we have

$$h(f_2; i, m) \;=\; \frac{\sqrt{210}\,\big[4i^3 - 6i^2(m+1) + 2i(m^2+3m+1) - m(m+1)\big]}{2m^2}.$$

Again using (2)–(4), we eventually obtain

$$\mathrm{E}[\mathcal{A}^{\mathrm{O}}(f_2; b, m)] \;=\; \frac{2m^6 + 7m^4 + 63m^2 - 72}{2m^6},$$

and

$$\mathrm{Var}(\mathcal{A}^{\mathrm{O}}(f_2; b, m)) \;=\; \frac{3514m + O(1/m)}{4290(bm-m+1)} - \frac{26m^2 + O(1)}{132(bm-m+1)^2} \;\doteq\; \frac{3514b - 4359}{4290(b-1)^2},$$

for large m and b, in agreement with the results found in Table 1 of Alexopoulos et al. (2006). ◁

Remark 1 Some additional algebra—not reported here—allows us to verify the overlapping area estimator results for the AR(1) process analyzed via Monte Carlo in §4.1 of Alexopoulos et al. (2006). ◁

Overlapping CvM Estimator

We next look at the overlapping CvM estimator for an i.i.d. standard normal process. As for the overlapping area estimator, we begin with a general expression for the expected value of the variance estimator. Since the Y_i's are i.i.d. with unit variance, we have

$$\begin{aligned}
\mathrm{E}[\mathcal{C}^{\mathrm{O}}(g;b,m)] &= \mathrm{E}[C_1^{\mathrm{O}}(g;m)] \;=\; \frac{1}{m^2}\sum_{k=1}^{m} g(\tfrac{k}{m})k^2\mathrm{E}[(\bar{Y}_{1,m}^{\mathrm{O}}-\bar{Y}_{1,k}^{\mathrm{O}})^2] \\
&= \frac{1}{m^2}\sum_{k=1}^{m} g(\tfrac{k}{m})k^2\,\mathrm{Var}(\bar{Y}_{1,m}-\bar{Y}_{1,k}) \\
&= \frac{1}{m^2}\sum_{k=1}^{m} g(\tfrac{k}{m})k^2\,[\mathrm{Var}(\bar{Y}_{1,m})+\mathrm{Var}(\bar{Y}_{1,k})-2\mathrm{Cov}(\bar{Y}_{1,m},\bar{Y}_{1,k})] \\
&= \frac{1}{m^2}\sum_{k=1}^{m} g(\tfrac{k}{m})k^2\left(\frac{1}{m}+\frac{1}{k}-\frac{2}{m}\right) \\
&= \frac{1}{m^2}\sum_{k=1}^{m} g(\tfrac{k}{m})k^2\left(\frac{1}{k}-\frac{1}{m}\right). \qquad (5)
\end{aligned}$$

Before undertaking our calculation of the overlapping CvM estimator's variance, we start off with a preliminary calculation: For $i, j = 1, \ldots, m-1$ and $k = 0, 1, 2, \ldots$,

$$\begin{aligned}
&\mathrm{Cov}\left(\bar{Y}_{1,m}^{\mathrm{O}}-\bar{Y}_{1,i}^{\mathrm{O}},\, \bar{Y}_{1+k,m}^{\mathrm{O}}-\bar{Y}_{1+k,j}^{\mathrm{O}}\right) \\
&= \frac{1}{m^2}\sum_{\alpha=1}^{m}\sum_{\beta=1}^{m}\mathrm{Cov}(Y_\alpha,Y_{k+\beta}) - \frac{1}{jm}\sum_{\alpha=1}^{m}\sum_{\beta=1}^{j}\mathrm{Cov}(Y_\alpha,Y_{k+\beta}) \\
&\quad -\frac{1}{im}\sum_{\alpha=1}^{i}\sum_{\beta=1}^{m}\mathrm{Cov}(Y_\alpha,Y_{k+\beta}) + \frac{1}{ij}\sum_{\alpha=1}^{i}\sum_{\beta=1}^{j}\mathrm{Cov}(Y_\alpha,Y_{k+\beta}) \\
&= \frac{1}{m^2}\max(m-k,0) - \frac{1}{jm}\min(j,\max(m-k,0)) - \frac{1}{im}\max(i-k,0) + \frac{1}{ij}\min(j,\max(i-k,0)) \\
&\quad \text{(since the } Y_i\text{'s are i.i.d. with unit variance)} \\
&= \begin{cases}
\left(1+\frac{k}{m}\right)\left(\frac{1}{i}-\frac{1}{m}\right), & \text{if } k<i \text{ and } j\le i-k \\
-\frac{k}{m^2}+\left(1-\frac{k}{i}\right)\left(\frac{1}{j}-\frac{1}{m}\right), & \text{if } k<i \text{ and } i-k<j\le m-k \\
-k\left(\frac{1}{i}-\frac{1}{m}\right)\left(\frac{1}{j}-\frac{1}{m}\right), & \text{if } k<i \text{ and } m-k<j \\
-\frac{k}{m^2}, & \text{if } i\le k<m \text{ and } j\le m-k \\
-\left(1-\frac{k}{m}\right)\left(\frac{1}{j}-\frac{1}{m}\right), & \text{if } i\le k<m \text{ and } m-k<j \\
0, & \text{if } m\le k
\end{cases}. \qquad (6)
\end{aligned}$$

For $k = 0, 1, 2, \ldots$, let

$$\begin{aligned} R_k^C &\equiv \operatorname{Cov}(C_1^{\mathrm{O}}(g; m), C_{1+k}^{\mathrm{O}}(g; m)) \\ &= \operatorname{Cov}\left(\frac{1}{m^2}\sum_{i=1}^{m} g(\tfrac{i}{m}) i^2 \big(\bar{Y}_{1,m}^{\mathrm{O}} - \bar{Y}_{1,i}^{\mathrm{O}}\big)^2, \frac{1}{m^2}\sum_{j=1}^{m} g(\tfrac{j}{m}) j^2 \big(\bar{Y}_{1+k,m}^{\mathrm{O}} - \bar{Y}_{1+k,j}^{\mathrm{O}}\big)^2\right) \\ &= \frac{2}{m^4}\sum_{i=1}^{m-1}\sum_{j=1}^{m-1} g(\tfrac{i}{m}) g(\tfrac{j}{m}) i^2 j^2 \operatorname{Cov}^2\Big(\bar{Y}_{1,m}^{\mathrm{O}} - \bar{Y}_{1,i}^{\mathrm{O}}, \bar{Y}_{1+k,m}^{\mathrm{O}} - \bar{Y}_{1+k,j}^{\mathrm{O}}\Big), \end{aligned}$$

the last step following from Lemma 1. Evocation of Equation (6) then reveals that

$$\begin{aligned} R_k^C = \frac{2}{m^4}\Bigg\{ &\frac{k^2}{m^4}\sum_{i=1}^{k}\sum_{j=1}^{m-k} g(\tfrac{i}{m}) g(\tfrac{j}{m}) i^2 j^2 + \big(1 - \tfrac{k}{m}\big)^2 \sum_{i=1}^{k}\sum_{j=m-k+1}^{m-1} g(\tfrac{i}{m}) g(\tfrac{j}{m}) i^2 \left(1 - \tfrac{j}{m}\right)^2 \\ &+ \big(1 + \tfrac{k}{m}\big)^2 \sum_{i=k+1}^{m-1}\sum_{j=1}^{i-k} g(\tfrac{i}{m}) g(\tfrac{j}{m}) j^2 \left(1 - \tfrac{i}{m}\right)^2 \\ &+ \sum_{i=k+1}^{m-1}\sum_{j=i-k+1}^{m-k} g(\tfrac{i}{m}) g(\tfrac{j}{m}) i^2 j^2 \left(-\tfrac{k}{m^2} + \left(1 - \tfrac{k}{i}\right)\left(\tfrac{1}{j} - \tfrac{1}{m}\right)\right)^2 \\ &+ k^2 \sum_{i=k+1}^{m-1}\sum_{j=m-k+1}^{m-1} g(\tfrac{i}{m}) g(\tfrac{j}{m}) \left(1 - \tfrac{i}{m}\right)^2 \left(1 - \tfrac{j}{m}\right)^2 \Bigg\}, \end{aligned} \tag{7}$$

for $k = 0, 1, \ldots, m-1$, and $R_k^C = 0$ for $k \geq m$. Finally, as in the derivation of Equation (4), we have

$$(n - m + 1)\operatorname{Var}(\mathcal{C}^{\mathrm{O}}(g; b, m)) = R_0^C + 2\sum_{k=1}^{m-1}\left(1 - \frac{k}{n - m + 1}\right) R_k^C. \tag{8}$$

With Equations (5), (7), and (8) in hand, we can calculate the mean and variance of the overlapping CvM estimator when specific weighting functions are applied to the i.i.d. normal process.

Example 3 For the overlapping CvM estimator with constant weighting function $g_0(t) = 6$, we obtain

$$\mathrm{E}[\mathcal{C}^{\mathrm{O}}(g_0; b, m)] = 1 - \frac{1}{m^2},$$

which is consistent with Table 1 of Alexopoulos et al. (2006) (with $\gamma = 0$), and

$$\operatorname{Var}(\mathcal{C}^{\mathrm{O}}(g_0; b, m)) = \frac{(88b - 115)m^7 + 88m^6 + m^5(588b - 756) + O(m^4 b)}{210m^5(m(b-1)+1)^2} \doteq \frac{88b - 115}{210(b-1)^2},$$

for large m and b, which is again consistent with Table 1 of Alexopoulos et al. (2006). $\triangleleft$

Example 4 For the overlapping CvM estimator with quadratic weighting function $g_2^{\star}(t) = -24 + 150t - 150t^2$, we have

$$\mathrm{E}[\mathcal{C}^{\mathrm{O}}(g_2^{\star}; b, m)] = 1 + \frac{4}{m^2} - \frac{5}{m^4},$$

which agrees with Table 1 of Alexopoulos et al. (2006) (with $\gamma = 0$); and

$$\begin{aligned} \mathrm{Var}(\mathcal{C}^{\mathrm{O}}(g_2^\star; b, m)) &= \frac{3876480m^{10}(bm+1) - 4897800m^{11} + O(m^9 b)}{4989600m^9(m(b-1)+1)^2} \\ &\doteq \frac{10768b - 13605}{13860(b-1)^2}, \end{aligned}$$

for large m and b, which agrees with Table 1 of Alexopoulos et al. (2006). $\triangleleft$

Overlapping Batch Means Estimator

Finally, we examine the OBM estimator for an i.i.d. standard normal process. The expected value is given by

$$\begin{aligned} \frac{(n-m+1)(n-m)}{nm}\,\mathrm{E}[\mathcal{O}(b,m)] &= \sum_{i=1}^{n-m+1} \mathrm{E}[(\bar{Y}^{\mathrm{O}}_{i,m} - \bar{Y}_n)^2] \\ &= \sum_{i=1}^{n-m+1} \mathrm{Var}(\bar{Y}^{\mathrm{O}}_{i,m} - \bar{Y}_n) \\ &= \sum_{i=1}^{n-m+1} \Big[\mathrm{Var}(\bar{Y}^{\mathrm{O}}_{i,m}) - 2\mathrm{Cov}(\bar{Y}^{\mathrm{O}}_{i,m},\ \bar{Y}_n) + \mathrm{Var}(\bar{Y}_n)\Big] \\ &= (n-m+1)\left[\frac{1}{m} - \frac{2}{n} + \frac{1}{n}\right], \end{aligned}$$

so that $\mathrm{E}[\mathcal{O}(b,m)] = 1$, in agreement with Table 1 of Alexopoulos et al. (2006) (with $\gamma = 0$).

As for the variance of the OBM estimator, we have

$$\begin{aligned} &\frac{(n-m+1)^2(n-m)^2}{2n^2m^2}\,\mathrm{Var}(\mathcal{O}(b,m)) \\ &= \frac{1}{2}\sum_{i=1}^{n-m+1}\sum_{j=1}^{n-m+1} \mathrm{Cov}\Big((\bar{Y}^{\mathrm{O}}_{i,m} - \bar{Y}_n)^2,\ (\bar{Y}^{\mathrm{O}}_{j,m} - \bar{Y}_n)^2\Big) \\ &= \sum_{i=1}^{n-m+1}\sum_{j=1}^{n-m+1} \mathrm{Cov}^2(\bar{Y}^{\mathrm{O}}_{i,m} - \bar{Y}_n,\ \bar{Y}^{\mathrm{O}}_{j,m} - \bar{Y}_n) \quad \text{(by Lemma 1)} \\ &= \sum_{i=1}^{n-m+1}\sum_{j=1}^{n-m+1} \Big[\mathrm{Cov}(\bar{Y}^{\mathrm{O}}_{i,m}, \bar{Y}^{\mathrm{O}}_{j,m}) - \mathrm{Cov}(\bar{Y}^{\mathrm{O}}_{i,m}, \bar{Y}_n) - \mathrm{Cov}(\bar{Y}^{\mathrm{O}}_{j,m}, \bar{Y}_n) + \mathrm{Var}(\bar{Y}_n)\Big]^2 \\ &= \sum_{i=1}^{n-m+1}\sum_{j=1}^{n-m+1} \left[\mathrm{Cov}(\bar{Y}^{\mathrm{O}}_{i,m}, \bar{Y}^{\mathrm{O}}_{j,m}) - \frac{1}{n}\right]^2 \\ &= (n-m+1)\left[\mathrm{Var}(\bar{Y}^{\mathrm{O}}_{1,m}) - \frac{1}{n}\right]^2 + 2\sum_{j=1}^{n-m+1}(n-m+1-j)\left[\mathrm{Cov}(\bar{Y}^{\mathrm{O}}_{1,m}, \bar{Y}^{\mathrm{O}}_{1+j,m}) - \frac{1}{n}\right]^2 \\ &= (n-m+1)\left[\frac{1}{m} - \frac{1}{n}\right]^2 + 2\sum_{j=1}^{m-1}(n-m+1-j)\left[\frac{m-j}{m^2} - \frac{1}{n}\right]^2 + \frac{2}{n^2}\sum_{j=m}^{n-m+1}(n-m+1-j), \end{aligned}$$

so that, after carrying out some algebra,

$$
\begin{aligned}
\mathrm{Var}(\mathcal{O}(b,m)) &= \frac{4b^3 - 11b^2 + 4b + 6}{3(b-1)^4} + \frac{2b^2}{3m(b-1)^2} - \frac{2b(b^4 - 3b^3 + 5b^2 - 10b + 10)}{3(n-m+1)(b-1)^4} \\
&\quad + \frac{b^2(b^2 - 6b + 8)}{3(n-m+1)^2(b-1)^4} \\
&= \frac{4b^3 - 11b^2 + 4b + 6}{3(b-1)^4} + O(1/m),
\end{aligned}
$$

which is in agreement with Table 1 of Alexopoulos et al. (2006). □

References

Alexopoulos, C., N. T. Argon, D. Goldsman, N. M. Steiger, G. Tokol, and J. R. Wilson. 2006. Efficient computation of overlapping variance estimators for simulation. Technical report. School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA.

Foley, R. D., and D. Goldsman. 1999. Confidence intervals using orthonormally weighted standardized times series. *ACM Transactions on Modeling and Computer Simulation* **9** 297–325.

Kelton, W. D., R. P. Sadowski, and D. T. Sturrock. 2004. *Simulation with Arena*, 3rd ed. McGraw-Hill, New York.

Microsoft Corp. 1995. *Fortran PowerStation Reference*. Microsoft Corp., Redmond, WA.

Patel, J. K., and C. B. Read. 1996. *Handbook of the Normal Distribution*, 2nd ed. Marcel Dekker, New York.