Proceedings of the 1990 Winter Simulation Conference
Osman Balci, Randall P. Sadowski, Richard E. Nance (eds.)

OVERLAPPING BATCH STATISTICS

Bruce W. Schmeiser
Thanos N. Avramidis
Sherif Hashem

School of Industrial Engineering
Purdue University
West Lafayette, Indiana 47907

ABSTRACT

Batch-means algorithms have long been used to estimate the standard error of sample means from stationary simulation output. We discuss the extension of batching algorithms from sample means to more-general estimators. We provide assumptions sufficient for unbiasedness and convergence and provide computationally efficient algorithms for variances and quantiles. Although the definitions, discussion, and examples generalize to general batching estimators, we consider only the completely overlapping version.

1. INTRODUCTION

One purpose of stochastic simulation experimentation is to estimate the (possibly multivariate) performance measure, θ, of the model of interest. The simulation experiment produces an output sequence $\{Y_i\}_{i=1}^n = Y_1, Y_2, ..., Y_n$, from which the point estimator, $\hat{\theta}$, is computed. Output analysis is the process of estimating the sampling error of $\hat{\theta}$. Often the sampling error is measured by the variance of $\hat{\theta}$, $\text{var}(\hat{\theta})$, or by its square root, the standard error. The estimate of the standard error has a variety of uses, including confidence intervals, tolerance intervals, methods for comparing systems, or as a stand-alone indication of sampling error.

The extensive literature for estimating the standard error of $\hat{\theta}$ is reviewed in most graduate-level simulation textbooks. Most of the attention is focused on $\hat{\theta} = \bar{Y}$, the sample mean of the output process. In the simplest case, independent identically distributed (iid) output processes, the square root of the ratio of the sample variance and the sample size, $[\sum_{i=1}^n (Y_i - \bar{Y})^2/(n(n-1))]^{1/2}$, is well accepted. A variety of methods have been proposed for output analysis for the sample mean of stationary processes. Most do not extend directly to the more-general setting where $\hat{\theta}$ is not the sample mean.

In Section 2 we discuss the straightforward extension of overlapping batch means [Meketon and Schmeiser 1984] to overlapping batch statistics (obs), estimation of standard errors for estimates of performance measures that are not sample means. Schmeiser [1990] contains a general discussion of obs estimators. We specialize the discussion to sample variances in Section 3 and to sample quantiles in Section 4.

2. STATISTICAL BEHAVIOR OF OVERLAPPING BATCH STATISTICS

The variance of $\hat{\theta}$ can be estimated by the (obs) estimator

$$\hat{V}(m) = \left[\frac{m}{n-m}\right] \frac{\sum_{j=1}^{n-m+1} (\hat{\theta}_j - \hat{\theta})^2}{(n-m+1)},$$

where $\hat{\theta}_j$ is defined analogously to $\hat{\theta}$ but is a function of only $Y_j, Y_{j+1}, ..., Y_{j+m-1}$, the data in the j^{th} batch of size m. In the special case of overlapping batch means (obm), $\hat{\theta}_j = m^{-1}\sum_{i=j}^{j+m-1} Y_i$.

That $\hat{V}(m)$ is a reasonable family of estimators of the variance of the sample mean is well established. Computation is $O(n)$, as discussed in Meketon and Schmeiser [1984] and Schmeiser and Song [1987]. Statistical properties depend upon the batch size m, with bias decreasing and variance increasing with m, but asymptotically the bias of obm is that of the nonoverlapping-batch-means estimator (nbm) and has two-thirds the variance of nbm.

$\hat{V}(m)$ is also a reasonable family of estimators of the variance of nonmeans. Nonmeans include all point estimators that are not the sample mean of the process Y. Typical examples include variances, standard deviations, higher-order moments especially skewness and kurtosis, and quantiles. Probabilities are means. In estimating nonmeans, we usually use an estimator commonly used for iid data. Such point estimators of marginal-distribution properties remain applicable with dependent data, although often with the burden of asymptotically negligible bias.

Unlike for the point estimator, large sample size does not alleviate all problems with estimating the standard error, where bias and variance are affected fundamentally by dependence. In Section 2.1 we study bias of $\hat{V}(m)$, in Section 2.2 we study the variance of $\hat{V}(m)$, and in Section 2.3 we discuss our assumptions on $\hat{\theta}$ and the data process.

2.1 Bias of OBS Estimators

Ideally, $\hat{V}(m)$ should be unbiased; that is, $E\hat{V}(m) = \text{var}(\hat{\theta})$. We provide here three assumptions that are sufficient for unbiasedness. Often these assumptions hold only in the limit as batch size m and/or sample size n grow large, as discussed in Section 2.3.

Assumption 1. $E\hat{\theta} = E\hat{\theta}_1 = E\hat{\theta}_2 = \cdots = E\hat{\theta}_{n-m+1} = \theta$.
Assumption 2. For some positive constant c,
$\text{var}(\hat{\theta}_1) = \text{var}(\hat{\theta}_2) = \cdots = \text{var}(\hat{\theta}_{n-m+1}) = c/m$ and
$\text{var}(\hat{\theta}) = c/n$.
Assumption 3. $\text{cov}(\hat{\theta}_j, \hat{\theta}) = \text{var}(\hat{\theta})$.

From these three assumptions, we obtain two intermediate results. First, from Assumption 2 we have

$$\frac{\text{var}(\hat{\theta})}{\text{var}(\hat{\theta}_j)} = \frac{c/n}{c/m} = \frac{m}{n}.$$

Second, for an arbitrary batch j the expected squared deviation is

$$\begin{aligned} E(\hat{\theta}_j - \hat{\theta})^2 &= E((\hat{\theta}_j - \theta) - (\hat{\theta} - \theta))^2 \\ &= E(\hat{\theta}_j - \theta)^2 - 2E(\hat{\theta}_j - \theta)(\hat{\theta} - \theta) + E(\hat{\theta} - \theta)^2 \\ &= \text{var}(\hat{\theta}_j) - 2\text{cov}(\hat{\theta}_j, \hat{\theta}) + \text{var}(\hat{\theta}) \\ &= \left(\frac{n-m}{n}\right)\text{var}(\hat{\theta}_j), \end{aligned}$$

where the third equality follows from Assumption 1 and the fourth equality follows from Assumptions 2 and 3.

Combining these two intermediate results yields

$$\begin{aligned} E\hat{V}(m) &= \left[\frac{m}{n-m}\right] \frac{\sum_{j=1}^{n-m+1} E(\hat{\theta}_j - \hat{\theta})^2}{(n-m+1)} \\ &= \left[\frac{m}{n}\right]\left[\frac{n}{n-m}\right] E(\hat{\theta}_j - \hat{\theta})^2 \\ &= \left[\frac{\text{var}(\hat{\theta})}{\text{var}(\hat{\theta}_j)}\right]\left[\frac{n}{n-m}\right]\left[\left[\frac{n-m}{n}\right]\text{var}(\hat{\theta}_j)\right] \\ &= \text{var}(\hat{\theta}) . \end{aligned}$$

2.2 Variance of OBS Estimators

Having established assumptions under which the obs estimators are unbiased, we now add one additional assumption sufficient for convergence. The discussion here parallels the usual discussion for obm estimators.

By Assumption 2, $\text{var}(\hat{\theta})$ goes to zero as n increases, so we hope for a stronger result than $\text{var}(\hat{V}(m))$ going to zero with n. We focus on $n\hat{V}(m)$ to estimate $n\text{var}(\hat{\theta})$, which by Assumption 2 is a constant. We show here that under weak conditions, $\text{var}(n\hat{V}(m))$ goes to zero.

For any process and any point estimator

$$n\,\text{var}(n\hat{V}(m)) = n^3\,\text{var}\left[\left[\frac{m}{n-m}\right]\frac{\sum_{j=1}^{n-m+1}(\hat{\theta}_j-\hat{\theta})^2}{n-m+1}\right]$$

$$= \left[\frac{n^3m^2}{(n-m+1)(n-m)^2}\right]\left[\frac{\sum_{j=1}^{n-m+1}\sum_{k=1}^{n-m+1}\text{cov}((\hat{\theta}_j-\hat{\theta})^2,(\hat{\theta}_k-\hat{\theta})^2)}{n-m+1}\right].$$

We now make

> Assumption 4. The sequence of squared differences of the batch statistics and the point estimator is a covariance stationary process having a finite sum of autocorrelations.

Let R_h denote $\text{cov}((\hat{\theta}_j-\hat{\theta})^2, (\hat{\theta}_{j+h}-\hat{\theta})^2)$, the lag-$h$ covariance of the squared-differences process. Then for a fixed batch size m and fixed sample size n we have $n\,\text{var}(n\hat{V}(m))$ equal to

$$\left[\frac{n^3m^2}{(n-m+1)(n-m)^2}\right]\left[\text{var}((\hat{\theta}_j-\hat{\theta})^2) + 2\sum_{h=1}^{n-m+1}(1-\frac{h}{n-m+1})R_h\right]$$

and for large sample sizes

$$\lim_{n\to\infty} n\text{var}(n\hat{V}(m)) = m^2\sum_{h=-\infty}^{\infty} R_h ,$$

assuming that the batch size m is controlled so that $\lim_{n\to\infty} n/(n-m) = 1$. Therefore $\text{var}(n\hat{V}(m))$ goes to zero.

2.3 Discussion of the Four Assumptions

The four assumptions of Sections 2.1 and 2.2 are sufficient to provide obs estimates that are unbiased and converge in a meaningful way to the variance of the point estimator. Here we restate the assumptions in words and briefly discuss their implications and interpretations.

Assumption 1 states that the point estimator and each batch statistic are unbiased estimators of θ.

Assumption 2 states that the variances of the point estimator $\hat{\theta}$ and of the batch statistics $\hat{\theta}_j$ decrease proportionally with sample size. This assumption is the usual limiting behavior $\lim_{n\to\infty} n\text{var}(\hat{\theta}) = c$, which also holds for the batch statistics since they are defined analogously to $\hat{\theta}$.

Assumption 3 states that the covariance between each batch statistic and the point estimator equals the variance of the point estimator. When Assumption 2 holds, Assumption 3 is equivalent to stating that $\text{corr}(\hat{\theta}_j, \hat{\theta}) = (m/n)^{1/2}$; that is, the squared correlation is the ratio of the number of observations in a batch to the number of observations in the entire sample.. For intuition about this assumption, consider the regression model $\hat{\theta} = \beta_0 + \beta_1\hat{\theta}_j + \varepsilon$ for an arbitrary batch j. Then the squared correlation is the fraction of the variance of $\hat{\theta}$ explained by $\hat{\theta}_j$. Since $\hat{\theta}_j$ is a function of m of the n observations, it is reasonable that $\hat{\theta}_j$ explains m/n of the variance of $\hat{\theta}$. When autocorrelation is present, the m observations in the batch contain information about the other $n-m$ observations, causing some error. For very early and very late batches, this error is about half of that incurred by the middle batches.

Assumptions 1, 2 and 3 seldom hold for finite sample sizes. However, for all the classical estimators these assumptions hold in the limit as batch size grows large.

Assumption 4 requires distant observations to be weakly correlated. This ensures that additional observations contain new information.

In the special case of $\hat{\theta}$ being the sample mean of iid observations, the obs (overlapping batch means, obm) estimator is unbiased and converges with run length, since the four assumptions hold even for finite batch and sample sizes.

Note that although the bias of $\hat{V}(m)$ decreases with the batch size m, the variance of $\hat{V}(m)$ decreases with sample size n for any batch size m.

3. OVERLAPPING BATCH VARIANCES

In this section we specialize the obs estimator to estimate the variance of the sample variance. The problem is to estimate the marginal-distribution variance $\sigma^2 = E(Y_i-\mu)^2$, where $\mu = EY_i$, and to obtain an estimate of the variance of the associated estimator.

The point estimator is the sample variance,

$$S^2 = \frac{\sum_{j=1}^{n}(Y_j-\bar{Y})^2}{n-1} .$$

David [1985] notes that the expected value of S^2 is $n(\sigma^2 - \text{var}(\bar{Y}))/(n-1)$ under any dependence structure. In the special case of stationary data, the expected value is $\sigma^2(1 - (2/n)\sum_{h=1}^{n-1}((n-h)/(n-1))\text{corr}(Y_1, Y_h))$. Therefore, if the sum of correlations is finite, S^2 is an asymptotically unbiased estimator for the marginal variance σ^2.

Our problem is to estimate the variance of S^2. To this end, we specialize the obs estimator to variances to obtain the overlapping batch variances (obv) estimator

$$\hat{V}_{\text{obv}}(m) = \frac{m}{n-m}\left[\frac{1}{n-m+1}\sum_{h=1}^{n-m+1}(S^2_{h,m}-S^2)^2\right], \qquad (3\text{-}1)$$

where m (satisfying $2 \le m \le n-1$) is the batch size and $S^2_{h,m}$ is the sample variance of $\{Y_h, \ldots, Y_{h+m-1}\}$. We will simply write $\hat{V}_{\text{obv}}$ when the batch size is implied or of no importance.

Schmeiser and Song [1987] give a Fortran subroutine for computing the overlapping batch means estimate $\hat{V}_{\text{obm}}$ of the variance of the sample mean that requires $O(n)$ time. In fact the computation is performed in one pass through the data. The same approach

can be used for $\hat{V}_{obv}$. After expanding the squares in Equation (3-1), we obtain

$$\hat{V}_{obv}(m) = \frac{m\left[\sum_{j=1}^{n-m+1}(S_{j,m}^2)^2 - S^2[2\sum_{j=1}^{n-m+1}S_{j,m}^2 - (n-m+1)S^2]\right]}{(n-m)(n-m+1)} \quad (3\text{-}2)$$

Appendix A contains a Fortran subroutine that, given a batch size m, computes $\hat{V}_{obv}(m)$ in $O(n)$ time, based on Equation (3-2).

4. OVERLAPPING BATCH QUANTILES

We now specialize obs estimators to quantiles, obq. The qth quantile of the marginal-distribution of Y is the smallest constant y satisfying $P(Y \le y) = q$. We consider the point estimator $\hat{\theta} = Y_{(\lceil qn \rceil)}$, where $Y_{(j)}$ is the j^{th} order statistic and $\lceil . \rceil$ denotes the ceiling, the smallest integer less than or equal to the argument. Other quantile point estimators, such as linear combinations of order statistics could be used, in which case the algorithms of this section would undergo straightforward modifications.

The objective of this section is to provide a computationally efficient method for estimating quantiles and at the same time obtain an estimate of the resulting standard error. Two obq algorithms are introduced to deal with this problem and a comparison between these algorithms regarding both their time and space complexities is also provided.

Algorithm (A)

This algorithm is fairly simple and straight forward. The given set of n observations is sorted, then the $\lceil qn \rceil^{th}$ smallest element is used to estimate the q^{th} quantile. The same process is repeated for each batch and the $\lceil qm \rceil^{th}$ smallest element is used as an estimate of the q^{th} quantile within the batch.

The initial sorting of the n data points requires $O(n\log n)$ time, while locating the qn^{th} smallest element can be done in $O(1)<$ time. Thus obtaining the grand estimate requires $O(n\log n)$ time. Similarly, for each batch $O(m\log m)$ time is needed. Hence, Algorithm (A) requires $O(n\log n + nm\log m)$ time.

It can be easily shown that Algorithm (A) requires $O_s(n)$ space.

Algorithm (B)

To obtain a faster run time, both phases of Algorithm (A) will be attacked. First, the initial *sorting* of the n observations to find the k^{th} order statistic will be replaced by the *selection* procedure SELECT with parameters $k = \lceil qn \rceil$ and S the set of observations.

```
Procedure SELECT (k,S);
if n < n_c then
   sort S;
   return the k^th-smallest element in S;
else
   partition S into ⌊n/5⌋ sequences of 5 elements each with
      up to four left-over elements;
   sort each of the 5-element sequences;
   let M be the set of the medians of the 5-element sets;
   d ← SELECT (⌈ |M|/2 ⌉, M);
   let S_1, S_2, and S_3 be the sequences of elements in S
     less than, equal to, and greater than d, respectively;
   if |S_1| ≥ k, then
      return SELECT (k,S_1)
   elseif (|S_1| + |S_2| ≥ k), then
      return d
   else
      return SELECT (k-|S_1|-|S_2|, S_3)
   endif
endif
```

Here n_c is a machine-dependent constant, possibly about 50, that determines whether n is so small that a complete sort is appropriate.

This procedure takes $O(n)$ run time and hence it is more efficient than the initial sorting used in the first phase of Algorithm (A), which takes $O(n\log n)$ time. A complete analysis of the above procedure is given in Aho et al. [1974].

The second enhancement over Algorithm (A) is to replace the repeated sorting used in estimating the q^{th} quantile for each batch by the following procedure:

i. Construct a balanced 2–3 tree [Aho et al. 1983] with the m observations of the first batch. On each vertex that is not a leaf we store the number of leaves in every subtree rooted at a son of this vertex.

ii. Find the k^{th}-smallest element in the 2–3 tree for the first batch.

iii. Update the 2-3 tree by deleting the first observation in the first batch and inserting the extra observation found in the second batch. Now the 2-3 tree holds the data points of the second batch, hence obtain the k^{th}-smallest element of the second batch.

iv. Repeat step *iii* for the remaining batches and hence obtain the required estimate of the standard error.

The time complexity of Algorithm (B) is of $O(n\log m)$ and that its space complexity is $O_s(n)$. Algorithm (B) is better than Algorithm (A) when comparing their relative time complexities for any choice of m. Since $m = O(n^{1/3})$ is the optimal batch size for obm estimators, of particular interest is to choose m to be $O(n^c)$, where $0 \le c \le 1$. Then the time complexity becomes $O(n^{1+c}\log n)$ for Algorithm (A) and $O(n\log n)$ for Algorithm (B).

5. DISCUSSION

The idea of batch statistics can be applied to completely overlapping (as here), partially overlapping, adjacent nonoverlapping (classical batching), and spaced batches. (Song and Schmeiser [1989] discuss the general family of batch-means estimators having two parameters: batch size and distance between the the first observations of each batch. Welch [1987] studies the statistical costs of partial overlapping.) Although we discuss only completely overlapped estimators here, generalization to general batching estimators is direct by substituting the general $\hat{\theta}$ for $\bar{Y}$. No other family of estimators enjoys such direct extension beyond sample means.

An open issue is the choice of batch size m to balance bias and variance. Schmeiser [1982] discusses the trade-off in the context of confidence intervals for nonoverlapping batch means. Song and Schmeiser [1988a,b] consider estimator variances and covariances of variance estimators for the standard error of sample means. Limiting behavior for sample means is discussed by Goldsman and Meketon [1986] and Schmeiser and Song [1990]. But we have little information about choosing batch size for nonmeans. A convenient result would be that the optimal batch size for sample means are nearly optimal for general statistics, since this would allow the known results for means to be used for other statistics.

APPENDIX A. SUBPROGRAM FOR OVERLAPPING BATCH VARIANCES

```
      subroutine obv (x,n,m,xvar,vxvar)
c.....thanos avramidis
c     purdue university
c     july 1989.
c.....overlapping-batch-variances estimator
c     of the variance of S^2, where S^2 is
c     the estimator of the process variance.
c.....parameter definitions
c    ..input
c       n:      number of observations
c       x:      vector of observations
c    ..output
c       vxbar: estimated variance of the
c         sample mean
c       vxvar: estimated variance of the
c         sample variance
      real x(n)
c.... process the first m observations
      sumx  = 0.0
      sum2x = 0.0
      do 10 i=1,m
         sumx  = sumx  + x(i)
10       sum2x = sum2x + x(i) * x(i)
      sbvar  = (sum2x - sumx*sumx/m) / (m-1)
      s2bvar = sbvar * sbvar
c.....process observations m+1 through n
      sumd  = 0.0
      sum2d = 0.0
      do 20 i=m+1,n
         sumd  = sumd  + x(i-m)
         sum2d = sum2d + x(i-m) * x(i-m)
         sumx  = sumx  + x(i)
         sum2x = sum2x + x(i) * x(i)
         bsum  = sumx  - sumd
         bsum2 = sum2x - sum2d
         bvar  = (bsum2 - bsum*bsum/m)/(m-1)
         sbvar = sbvar  + bvar
20       s2bvar = s2bvar + bvar * bvar
c.....calculate grand variance
      xvar = (sum2x - sumx*sumx/n)/(n-1)
c.....calculate the obv estimator
      vxvar = (s2bvar -
     +        xvar*(2.*sbvar-(n-m+1)*xvar)) /
     +        ((n-m+1.)*(n-m)/m)
      return
      end
```

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under Grant No. DMS-871779 to Purdue University and David Ross Grant No. 6901627 from the Purdue Research Foundation. We acknowledge helpful discussions with W. David Kelton, Barry L. Nelson, Wheyming Tina Song, Michael R. Taaffe, and James R. Wilson.

REFERENCES

Aho, A.V., J.E. Hopcroft and J.D. Ullman (1974), *The Design and Analysis of Computer Algorithms,* Addison Wesley, Reading, MA.

Aho, A.V., J.E. Hopcroft and J.D. Ullman (1983), *Data Structures and Algorithms,* Addison Wesley, Reading, MA.

David, H.A. (1985), "Bias of S^2 under Dependence", *American Statistician 39,* 201.

Goldsman, D. and M.S. Meketon (1986), "A Comparison of Several Variance Estimators," Technical Report J-85-12, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA.

Meketon, M.S. and B.W. Schmeiser (1984), "Overlapping batch means: Something for nothing?" In *Proceedings of the Winter Simulation Conference,* S. Sheppard, U.W. Pooch, and C.D. Pegden, Eds. 227-230.

Schmeiser, B.W. (1982), "Batch Size Effects in the Analysis of Simulation Output," *Operations Research 30,* 3, 556-568.

Schmeiser, B.W. (1990), "Simulation Experimentation," In *Handbook of Operations Research and Management Science, Volume 2: Stochastic Models,* Dan Heyman and Matthew Sobel, Eds. North Holland, Amsterdam, The Netherlands, 295-330.

Schmeiser, B.W. and W.-M.T. Song (1987), "Correlation among Estimators of the Variance of the Sample Mean," In *Proceedings of the Winter Simulation Conference,* A. Thesen, H. Grant, and W.D. Kelton, Eds. 309-317.

Schmeiser, B. and Song, W.-M.T. (1990), "Optimal Mean-Squared-Error Batch Sizes," Revision, Technical Report SMS-89-3, School of Industrial Engineering, Purdue University, West Lafayette, IN.

Song, W.-M.T. and B.W. Schmeiser (1988a), "On the Dispersion Matrix of Estimators of the Variance of the Sample Mean in the Analysis of Simulation Output," *Operations Research Letters 7,* 5, 259-266.

Song, W-M.T. and B.W. Schmeiser (1988b), "Minimal-Mse Linear Combinations of Variance Estimators of the Sample Mean," In *Proceedings of the Winter Simulation Conference,* M.A. Abrams, P.L. Haigh, and J.C. Comfort, Eds. 414-421.

Song, W.-M.T. and B.W. Schmeiser (1988c), "Estimating Standard Errors: Empirical Behavior of Asymptotic Mse-Optimal Batch Sizes," In *Computing Science and Statistics: Proceedings of the* 20^{th} *Symposium on the Interface,* E.J. Wegman, D.T. Gantz, and J.J. Miller, Eds. American Statistical Association, Alexandria, VA, 575-580.

Song, W.-M.T. and B.W. Schmeiser (1989), "Variance of the Sample Mean: Properties and Graphs of Quadratic-Form Estimators," Technical Report SMS-89-27, School of Industrial Engineering, Purdue University, West Lafayette, IN.

Welch, P.D. (1987), "On the Relationship between Batch Means, Overlapping Batch Means, and Spectral Estimation," In *Proceedings of the Winter Simulation Conference,* A. Thesen, H. Grant, and W.D. Kelton, Eds. 320-323.