Proceedings of the 1984
Winter Simulation Conference
S. Sheppard, U. Pooch, D. Pegden (eds.)

OVERLAPPING BATCH MEANS: SOMETHING FOR NOTHING?

Marc S. Meketon
Operations Research Department
AT&T Bell Laboratories
Holmdel, NJ 07733

Bruce Schmeiser*
School of Industrial Engineering
Purdue University
West Lafayette, IN 47907

ABSTRACT

Nonoverlapping batch means (NOLBM) is a well-known approach for estimating the variance of the sample mean. In this paper we consider an overlapping batch means (OLBM) estimator that, based on the same assumptions and batch size as NOLBM, has essentially the same mean and only 2/3 the asymptotic variance of NOLBM. Confidence interval procedures for the mean based on NOLBM and OLBM are discussed. Both estimators are compared to the classical estimator of the variance of the mean based on sums of covariances.

INTRODUCTION

Consider a covariance stationary stochastic process $\{X\}$ having mean μ and variance R_0, defined over a discrete or continuous time parameter and having a discrete or continuous state space. Estimation of μ from a realization $\{x\}$ is a common problem, especially in computer simulation of stochastic systems.

The family of point estimators usually considered is $\Sigma_{i=1}^{n} \alpha_i x_i$, where $\Sigma_{i=1}^{n} \alpha_i = 1$. In the nonstationary case early observations are often weighted less, but in the stationary case $\bar{x} = n^{-1} \Sigma_{i=1}^{n} x_i$ is used almost exclusively. We study only $\bar{X}$ here, but note that $\bar{X}$ is not the minimum variance linear estimator, which would place greater weight on both early and late observations when autocovariances are positive.

Confidence interval procedures for μ based on dependent data have been widely studied. Recent textbooks on stochastic simulation provide good discussions: Bratley, Fox and Schrage [1], Fishman [2] and Law and Kelton [3]. The problem is to find functions of the data $u_\alpha(\{x\})$ and $v_\alpha(\{x\})$ such that $P\{U_\alpha(\{X\}) \leq \mu \leq V_\alpha(\{X\})\} = 1-\alpha$ (in which case the procedure is called *valid*) while obtaining reasonable interval width and stability. Typically $U_\alpha = \bar{X} - H_\alpha$ and $V_\alpha = \bar{X} + H_\alpha$, where the half width $H_\alpha = q_{\alpha/2}\sqrt{\hat{V}(\bar{X})}$ and $q_{\alpha/2}$ is a constant reflecting the joint distribution of $\bar{X}$ and the estimator of the variance of the sample mean, $\hat{V}(\bar{X})$.

The method of batch means is of interest in this paper. The usual approach, based on nonoverlapping batch means (NOLBM), is $\bar{X} \pm H_{\alpha,k}$, where

$H_{\alpha,k} = t_{\alpha/2,k-1}\, S_k/\sqrt{k}$ and $S_k^2 = (k-1)^{-1}\, \Sigma_{j=1}^{k} (\bar{X}_j - \bar{X})^2$. Typically, one of three types of batch means point estimators are used:

$$\bar{X}_j = m^{-1}\, \Sigma_{i=(j-1)m+i}^{jm} X_i \quad \text{and} \quad \bar{X} = n^{-1}\, \Sigma_{i=1}^{n} X_i$$

$$\bar{X}_j = t_0^{-1} \int_{(i-1)t_0}^{it_0} X(t)dt \quad \text{and} \quad \bar{X} = t_*^{-1} \int_0^{t_*} X(t)dt$$

or

$$\bar{X}_j = t_0^{-1}\, (N(it_0)-N((i-1)t_0) \quad \text{and} \quad \bar{X} = t_*^{-1}\, N(t_*)$$

where $N(t)$ is a counting process with $N(0)=0$. We treat the first case here, but analogous results hold for the second and third cases by replacing summations with integrations, m with t_0, and n with t_*.

The second section discusses NOLBM, the third section discusses OLBM, and the fourth section relates both NOLBM and OLBM to the classical estimator based on the sum of covariances.

NONOVERLAPPING BATCH MEANS

Performance of the NOLBM procedure depends on the joint distribution of $\bar{X}$ and S_k^2. A valid procedure results when

(a) $\bar{X}$ is normally distributed,
(b) $\bar{X}$ and S_k^2 are independent,

and

(c) $(k-1)S_k^2/\sigma_k^2$ has a chi-square distribution with $k-1$ degrees of freedom,

where σ_k^2 , variance of each of k batches, equals $n/mV(\bar{X})$

In NOLBM the batch size m (or equivalently the number of batches k when the sample size n is fixed) is chosen with regard to the last two criteria, since the first is unaffected by batch size. Since these two criteria are difficult to measure in an application, two other criteria are typically substituted:

(d) the batch means are independent

and

(e) the batch means are normally distributed.

(See, e.g., Fishman [4], Law and Carson [5], Mechanic and McKay [6], and Schriber and Andrews [7].) Criteria (d) and (e), which are sufficient to ensure (b) and (c), are satisfied in the limit as $n \to \infty$ and $m \to \infty$. Therefore, at least asymptotically, increasing the batch size m (or equivalently for a fixed sample size n decreasing the number of batches k) moves the procedure toward validity.

Balancing the quest for validity is the need for short and stable confidence intervals, as usually measured by the mean and coefficient of variation of the half width, respectively. Schmeiser [8] quantifies and discusses the effect of m and k on these properties.

*The second author received time and resource support from AT&T Bell Laboratories, where he spent the summer of 1984 visiting the Operations Research Department.

OVERLAPPING BATCH MEANS

Because of the role of batch independence in NOLBM as discussed in the last section, the idea of using overlapping batch means (OLBM) can appear, at best, unnatural, since the common observations in the overlapping batches causes substantial positive correlation. However, the idea in many ways is a good one, as discussed in this section, because batch size rather than batch independence is the crucial element.

We consider the OLBM estimator of $V(\bar{X})$

$$\tilde{V}_m(\bar{X}) = (m/n)\ \Sigma_{j=1}^{n-m+1}\ (\bar{X}_j(m) - \bar{X})^2\ /\ (n-2m+1)$$

where $\bar{X}_j(m) = m^{-1}\ \Sigma_{i=0}^{m-1}\ X_{j+i}$ is the batch mean of size m beginning with observation X_j. As you progress through this section, the denominator (n-2m+1) will begin to seem like the obvious choice (if it doesn't now). As a beginning, note that for the extreme case of m=1 this estimator reduces to S_n^2/n, which is the usual estimator for the independent case. Also note the coefficient (m/n) is used rather than 1/k, which is consistent with using $\bar{X}$ rather the average of the batch means, which ignores the last fraction of a batch.

The remainder of this section studies properties of $\tilde{V}_m$. We show that this estimator and a NOLBM estimator have essentially the same bias, but that the OLBM estimator has only 2/3 the asymptotic variance. We also show the covariances with $\bar{X}$ are essentially equal, which implies the correlation between $\tilde{V}_m$ and $\bar{X}$ is greater than that between the NOLBM estimator and $\bar{X}$. Finally we show the computational effort is not prohibitive.

The results of this section follow almost directly from Proposition 1, which relates the OLBM estimator $\tilde{V}_m(\bar{X})$ to the NOLBM estimators

$$\hat{V}^i_{k,m}(\bar{X}) = (m/n)\ \Sigma_{j=1}^{k}\ (\bar{X}_{m(j-1)+i} - \bar{X})^2\ /\ (k-1)$$

which is the estimator for $V(\bar{X})$ arising from k NOLBM beginning with observation i. The subscript k indicating the number of batches is superfluous, since $k = \lfloor (n-i+1)/m \rfloor$, but we carry k explicitly to make the argument more clear. On the other hand, the sample size n is not carried explicitly. The proofs of the propositions indicate the main argument, often with algebraic detail, but are often not rigorously stated here.

Proposition 1 states that the OLBM estimator $\tilde{V}_m(\bar{X})$ is a weighted average of the NOLBM estimators.

Proposition 1.

$$\tilde{V}_m(\bar{X}) = \frac{(k-1)\Sigma_{i=1}^{k'}\hat{V}^i_{k,m}(\bar{X}) + (k-2)\Sigma_{i=k'+1}^{m}\hat{V}^i_{k-1,m}(\bar{X})}{n-2m+1}$$

where $k' = n - km + 1$.

Proof: Substitute the definitions of $\tilde{V}_m(\bar{X})$ and $V^i_{k,m}(\bar{X})$ and verify that $(k-1)k' + (k-2)(m-k') = n - 2m + 1$. ||

Proposition 2 says that the bias of $\tilde{V}_m(\bar{X})$ is essentially the same as that for $\hat{V}_{k,m}(\bar{X})$.

Proposition 2.

$$E(\tilde{V}_m(\bar{X})) \simeq V(\bar{X}) - (2\,m\ \Sigma_{j=1}^{k-1}\ \bar{R}_j)\ /\ (n - 2m + 1)$$

where $\bar{R}_j = Cov(\bar{X}_h(m), \bar{X}_{h+jm}(m))$ for all h.

Proof: Recall that for the NOLBM estimators

$$E(\hat{V}^i_{k,m}(\bar{X})) = V(\bar{X}) - 2(k-1)^{-1}\ \Sigma_{j=1}^{k-1}\ (1 - \tfrac{j}{k})\ \bar{R}_j$$

Substituting this result into the expression obtained by passing the expected value operator through the summations in the definition of the OLBM estimator and simplifying yields

$$E(\tilde{V}_m(\bar{X})) = V(\bar{X}) - \frac{2m\ \Sigma_{j=1}^{k-1}\ (1 - \frac{j}{k})\ \bar{R}_j}{n - 2m + 1} + \frac{2(m-k')\ [\Sigma_{j=1}^{k-2}\ \bar{R}_j + k^{-1}\bar{R}_{k-1}]}{k(k-1)(n - 2m + 1)}$$

Since the last term is negligible, the result is obtained. ||

The bias, of course, is the weighted average of the biases for k and k-1 NOLBM estimators. In the limit as batch size grows, $\bar{R}_j$ decreases and all the estimators are unbiased.

As might be expected since $\tilde{V}_m(\bar{X})$ is a weighted average of the NOLBM estimators, the OLBM estimator has a smaller variance.

Proposition 3.

$$\lim_{\substack{n \to \infty \\ m \to \infty \\ n/m \to \infty}} V(\tilde{V}_m(\bar{X}))\ /\ V(\hat{V}^i_{k,m}(\bar{X})) = 2/3$$

Proof: The limit of the denominator is $(m/n)^2\ 2\sigma_k^4/(k-1)$, which follows directly from $(k-1)S_k^2/\sigma_k^2$ having variance equal to $2(k-1)$ since it has a chi-square distribution. The limit of the numerator is identical, except the coefficient is 4/3 rather than 2, as shown in Meketon [9]. ||

Propositions 2 and 3 suggest that we should consider confidence interval procedures based on OLBM. The three obvious (extreme) possibilities are

$$\bar{X} \pm t_{\alpha/2,\lfloor n/m \rfloor -1}\ \sqrt{\tilde{V}_m(\bar{X})}$$

$$\bar{X} \pm t_{\alpha/2,\lfloor n/m \rfloor -1}\ \sqrt{\tilde{V}_{(3/2)m}(\bar{X})}$$

and

$$\bar{X} \pm t_{\alpha/2,(3/2)\lfloor n/m \rfloor -1}\ \sqrt{\tilde{V}_m(\bar{X})}$$

The first is the direct substitution of $\tilde{V}_m(\bar{X})$ for $\hat{V}_{k,m}(\bar{X})$ with no change in batch size m or the constant t. The second increases the batch size by 50%. The third increases the degrees of freedom of the constant by 50%. The third in some ways seems the most natural, and Fishman [10, p.284] suggests this is the customary modification in a similar situation. We have not studied these three possibilities empirically, but the next proposition suggests the larger batch size of the second procedure is appealing.

Proposition 4.

$$Corr(\tilde{V}_m(\bar{X}),\bar{X}) > Corr(\hat{V}^i_{k,m}(\bar{X}),\bar{X})$$

Proof: Let $c = \mathrm{Cov}(\hat{V}^1_{k,m}(\bar{X}),\bar{X})$. Then to a very slight error $c \simeq \mathrm{Cov}(\hat{V}^i_{k,m}(\bar{X}),\bar{X})$ for $i = 2,3,\ldots,k'$. To a slightly larger error we also have $c \simeq \mathrm{Cov}(\hat{V}^i_{k-1,m}(\bar{X}),\bar{X})$ for $i = k'+1,\ldots,m$. Since the OLBM estimator $\tilde{V}_m(\bar{X})$ is a weighted average of the NOLBM estimators, all of which have essentially covariance with $\bar{X}$ of c, then $\mathrm{Cov}(\tilde{V}_m(\bar{X}),\bar{X}) \simeq \mathrm{Cov}(\hat{V}^i_{k,m}(\bar{X}),\bar{X})$.

However, from Proposition 2 we know the variance is larger for NOLBM, which makes the correlation for NOLBM smaller. ||

The good news for OLBM is that the correlation is zero when NOLBM is zero. The bad news is that OLBM are less robust. However, the second confidence interval procedure from the last page uses batch sizes 50% larger than the NOLBM batch size, making the variances in the comparison essentially equal. This procedure then has less bias than NOLBM, similar variance and similar correlation with $\bar{X}$.

We leave the issue of confidence interval procedures now in favor of considering computational issues and relationship to other estimators.

The computational effort required for $\tilde{V}_m(\bar{X})$ at first appears quite large. A little thought, however, quickly yields the following algorithm for any given m and n:

$a \leftarrow m\, \Sigma^n_{i=1} x_i / n$

$b \leftarrow \Sigma^m_{i=1} x_i$

$s \leftarrow (b-a)^2$

$j \leftarrow 0$

While $j < n-m$ do

  $j \leftarrow j + 1$

  $b \leftarrow b + x_{j+m} - x_j$

  $s \leftarrow s + (b-a)^2$

End

$s \leftarrow m^{-2}s$

$\tilde{V}_m(\bar{X}) \leftarrow (m/n)(n - 2m + 1)^{-1}\, s$

While reasonably efficient, this algorithm must be repeated, except for the first step, for each value of m considered. Relationships developed in the next section yield a more efficient algorithm when many values of m are to be considered.

OLBM, NOLBM AND THE CLASSICAL ESTIMATOR

Since $V(\bar{X}) = n^{-1}[R_0 + 2\, \Sigma^{n-1}_{j=1} (1-\frac{j}{n})\, R_j]$, a reasonable estimator of $V(\bar{X})$ is to substitute estimators of the autocovariances into the equation to obtain what we will call the classical estimator. Proposition 5, the main result of this section, states that the OLBM estimator can be viewed as a classical estimator.

Proposition 5.

$$\hat{V}_m(\bar{X}) \simeq (n - 2m + 1)^{-1}[\hat{R}_0 + 2\sum^{m-1}_{j=1} (1 - \tfrac{j}{m})\, \hat{R}_j]$$

where $\hat{R}_j = n^{-1} \sum^{n-j}_{i=1} (X_j - \bar{X})(X_{j+i} - \bar{X})$ for $j = 0,1,\ldots,m-1$.

Proof: $V_m(X) = (m/n)\sum^{n-m+1}_{j=1} (\bar{X}_j(m)-\bar{X})^2 / (n-2m+1)$

$$= (m/n)\sum^{n-m+1}_{j=1} [\sum^{m-1}_{i=0} \frac{X_{j+i} - \bar{X}}{m}]^2 / (n-2m+1)$$

$$= (mn)^{-1}\sum^{n-m+1}_{j=1}\sum^{m-1}_{i=0}\sum^{m-1}_{\ell=0} (X_{j+i}-\bar{X})(X_{j+\ell}-\bar{X})/(n-2m+1)$$

$$\simeq \frac{m\sum^{n}_{i=1}(X_i-\bar{X})^2 + 2\sum^{m-1}_{i=1}(m-i)\sum^{n-i}_{j=1}(X_j-\bar{X})(X_{j+i}-\bar{X})}{mn\,(n - 2m + 1)}$$

$$= \frac{\sum^{n}_{i=1}(X_i-\bar{X})^2 + 2\sum^{m-1}_{i=1} (1-\frac{i}{m}) \sum^{n-i}_{j=1}(X_j-\bar{X})(X_{j+i}-\bar{X})}{n\,(n - 2m + 1)}$$

$$= (n - 2m + 1)^{-1}\, [\hat{R}_0 + 2\sum^{m-1}_{i=1} (1 - \tfrac{i}{m})\, \hat{R}_i]$$

The approximate equality arises because of the end effects; that is, some early and some late cross-products would appear more often if they had occurred in the center of the data. The reduction of the triple sum to a double sum in the fourth step of the proof may be substantiated algebraically or by organizing the terms graphically, as shown in Figure 1 for n=8 and m=3.

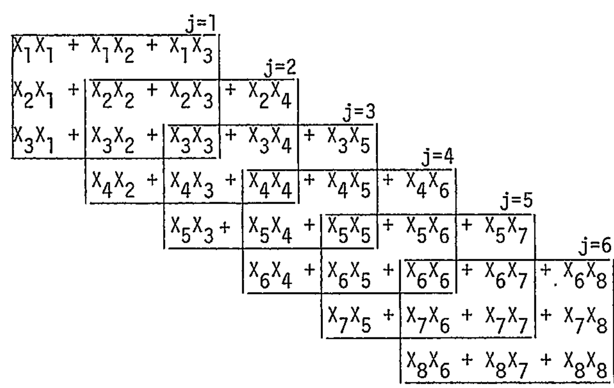

Figure 1. Graphical illustration of the terms summed in the OLBM estimator.

Each cross-product occurs once for each rectangle in which it is included. The doubling term arises from symmetry. The main diagonal corresponds to estimating R_0, the first off-diagonal corresponds to R_1, and so forth. ||

Proposition 5 says that the OLBM estimator $\tilde{V}_m(\bar{X})$ is algebraically equivalent, other than for minor end effects, to the classical estimator that uses the same number of autocorrelations as the OLBM uses in each batch.

Fishman [10] discusses the classical estimator in the context of spectral estimation. The specific estimator considered there is

$$m_3/n = (n-p)^{-1}\, [\hat{R}_0 + 2\sum^{p-1}_{j=1} (1 - \tfrac{j}{p})\, \hat{R}_j]$$

which differs only in the coefficient. Thus the asymptotic aspects of Propositions 2 and 3 can be shown via Proposition 5 and the known asymptotic properties of m_3.

The relationship of Proposition 5 is useful for calculating the OLBM estimator for various values of the batch size m, since only the autocovariance estimates are needed and they can be collected cumulatively.

As suggested by Fishman [10] and Meketon [9], plotting $\tilde{V}_m(\bar{X})$ as a function of m can be useful for determining an appropriate batch size m. In fact such plotting suggests another estimator for $V(\bar{X})$ at a still higher level:

$$\tilde{V}_{m_1,m_2}(\bar{X}) = \sum_{m=m_1}^{m_2} \beta_m \tilde{V}_m(\bar{X}) \text{ where } \sum_{m=m_1}^{m_2} \beta_m = 1$$

We now briefly comment on the relationship of the NOLBM estimator $\hat{V}_{k,m}(\bar{X})$ and the OLBM estimator $\tilde{V}_m(\bar{X})$. Recall Figure 1 with the overlapping rectangles corresponding to the overlapping batch means. Not surprisingly, $\hat{V}^1_{k,m}(\bar{X})$ corresponds to including only rectangles 1, m+1, 2m+1, ..., which are adjacent but do not overlap, resulting in fewer terms being used in the estimator. In particular, the estimators for $\hat{R}_j$ corresponding to the terms in the j th off-diagonal are missing terms that are as useful as the terms included. Similarly, $\hat{V}^i_{k,m}(\bar{X})$ corresponds to rectangles i, m+i, 2m+i, ...

SUMMARY

We have studied the relationship between nonoverlapping batch means, overlapping batch means, and classical estimators for $V(\bar{X})$. The overlapping batch means estimator, $\tilde{V}_m(\bar{X})$, has been shown to be algebraically equivalent, other than for end effects, to the classical estimators using m-1 covariances.

A potential reason for the unpopularity of the classical estimators arises in Proposition 4, where the covariance of the overlapping estimator with the point estimator for the mean was seen to be greater than the same quantity for nonoverlapping batch means, except in the limit when both are zero. However, given the popularity of nonoverlapping batch means and the near domination of nonoverlapping batch means by overlapping batch means, we think the use of overlapping batches deserves further study.

REFERENCES

1. P. Bratley, B.L. Fox and L.E. Schrage, *A Guide to Simulation*, Springer-Verlag, New York, 1983, 383 pp.

2. G.S. Fishman, *Principles of Discrete Event Simulation*, Wiley-Interscience, New York, 1978, 514 pp.

3. A.M. Law and W.D. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, New York, 400 pp.

4. G.S. Fishman, "Grouping Observations in Digital Simulation," *Management Science*, 24 (1978), 510-521.

5. A.M. Law and J.S. Carson II, "A Sequential Procedure for Determining the Length of a Steady State Simulation," *Operations Research*, 27 (1979), 1011-1025.

6. H. Mechanic and W. McKay, "Confidence Intervals for Averages of Dependent Data in Simulations II," Technical Report ASDD 17-202, IBM Corporation, Yorktown Heights, New York.

7. T.J. Schriber and R.W. Andrews, "Interactive Analysis of Simulation Output by the Method of Batch Means," *Proceedings of the Winter Simulation Conference*, 1979, 513-525.

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

9. M.S. Meketon, *The Variance Time Curve: Theory, Estimation and Application*, Ph.D. Disseration, School of Operations Research and Industrial Engineering, Cornell University.

10. G.S. Fishman, *Concepts and Methods in Discrete Event Digital Simulation*, Wiley-Interscience, New York, 1973.