Proceedings of the 1987 Winter Simulation Conference
A. Thesen, H. Grant, W. David Kelton (eds.)

On The Relationship Between Batch Means, Overlapping Batch Means and Spectral Estimation

Peter D. Welch
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

Abstract

The conventional method of batch means and the overlapping batch means approach of Meketon and Schmeiser are related to spectral estimation via the time averaging of subsequence periodograms. It is shown that most of the reduction in the variance of the variance estimate can be achieved with modest amounts of overlapping. This may have practical implications because of the large number of batches required for the statistical test of lack of correlation. and the usual practice of rebatching the data after this test is passed.

1. INTRODUCTION

We let $\{X_j\}$ be a covariance stationary stochastic process with spectral density $p(f)$ and covariance function $\gamma(k)$. We let $\mu = E[X]$. We assume we have a sample $X_1, ..., X_N$ from $\{X_j\}$ and wish to generate a confidence interval for μ. This is one of the central problems of steady state simulation output analysis.

We consider the point estimator

$$\overline{X} = (1/N)\sum_{j=1}^{N} X_j$$

and form a confidence interval by generating an estimate $\hat{\sigma}^2(\overline{X})$, of the variance of $\overline{X}$. We form the confidence interval either assuming

$$(\overline{X} - \mu)/\hat{\sigma}(\overline{X}) \sim N(0, 1)$$

or by assuming

$$(\overline{X} - \mu)/\hat{\sigma}(\overline{X}) \sim t_n$$

where t_n indicates a t random variable with n degrees of freedom where n is chosen such that the coefficient of variation of a χ^2_n random variable is equal to the coefficient of variation of $\hat{\sigma}^2(\overline{X})$.

We will be discussing the method of batch means and will be providing a footnote to the work of Meketon [1980] and Meketon and Schmeiser [1984] which introduced the concept of overlapping batch means and pointed out its advantage over ordinary, non-overlapping batch means in the estimation of $\hat{\sigma}^2(\overline{X})$.

2. BACKGROUND

In the method of batch means the sample $\{X_j \; j = 1, ..., N\}$ is broken up into B contiguous subsamples of length M. The kth subsample is given by

$$X_{kj} = X_{(k-1)M+j} \qquad j = 1, ..., M.$$

We assume for simplicity that $BM = N$. Now generally in the application of the method the batch size is increased until, by some statistical test, it is decided that the batch means are approximately uncorrelated. Usually it is assumed that in order for such a test to be meaningful the number of batches should reasonably large, say about 40. After passing the test, because the consequences of correlated batches are more serious than the consequences of fewer batches, the batches are themselves batched and the batch size increased by some integral multiple. (Hence, typically, B and M will change after the statistical test.) Then with these 10 to 20 batches an estimate $\hat{\sigma}^2_1(\overline{X})$ is formed as

$$\hat{\sigma}^2_1(\overline{X}) = (1/B)(1/(B-1))\sum_{k=1}^{B}(\overline{X}_k - \overline{X})^2$$

where $\overline{X}_k$ is the kth batch mean, i.e.

$$\overline{X}_k = (1/M)\sum_{j=1}^{M} X_{(k-1)M+j} \qquad k = 1, ..., B.$$

Finally a confidence interval is formed assuming

$$(\overline{X} - \mu) / \hat{\sigma}_1(\overline{X}) \sim t_{B-1}.$$

Now, within this context, Meketon and Schmeiser showed that if we overlap the batches in forming the estimator of $\sigma^2(\overline{X})$ that is, if we form the estimator

$$\hat{\sigma}_2^2(\overline{X}) = (1/B)\,(1/(N-M+1)) \sum_{k=1}^{N-M+1} (XBAR_k - \overline{X})^2$$

where $XBAR_k = (1/M)\sum_{j=1}^{M} X_{k-1+j} \quad k = 1, \ldots, N-M+1.$

We obtain an estimate of $\sigma^2(\overline{X})$ whose variance is reduced by the factor 2/3. For the χ_n^2 distribution

$$n = 2\,E^2\,[\chi_n^2]\,/\,Var\,[\chi_n^2].$$

Hence, if we assume that $\hat{\sigma}_2^2(\overline{X})$ is distributed approximately as a multiple of a χ_n^2 random variable, the number of degrees of freedom should be multiplied by 3/2 and we can generate a confidence interval with less variability assuming

$$(\overline{X} - \mu)/\hat{\sigma}_2(\overline{X}) \sim \quad t_{3(B-1)/2}.$$

3. RELATIONSHIP TO SPECTRAL ESTIMATION

Viewed in terms of second order processes

$$\sigma^2(\overline{X}) = (1/N) \sum_{k=-(N-1)}^{N-1} ((N-k)/N)\,\gamma(k)$$

$$\simeq (1/N) \sum_{k=\infty}^{\infty} \gamma(k)$$

$$\simeq p(0)/N.$$

Hence estimating $\sigma^2(\overline{X})$ is tantamount to estimating the spectral density at zero frequency. This observation is by no means new, see e.g. Fishman [1978]. Now classical spectral estimators come in two forms. They are either frequency averages over the periodogram of all the observations or, less commonly, they are time averages over periodograms of subsequences of the data. These latter types of spectral estimates were first described by Bartlett [1950] and later generalized and discussed in Welch [1961] ; Godfrey, Bingham and Tukey [1967] and Welch [1967].

Bartlett described a method of spectral estimation which paralleled the method of batch means. Periodograms were taken over the non-overlapping contiguous segments $\{X_{kj} - \overline{X} : j = 1, \ldots, M\}$, $k = 1, \ldots, B$ defined earlier. These periodograms were of the form

$$I_k(f) = \left| \sum_{j=1}^{M} (X_{kj} - \overline{X})\, e^{-2\pi i\,(j-1)f/M} \right|^2 /M$$

and the spectral estimate was given by

$$\hat{p}_1(f) = (1/B) \sum_{k=1}^{B} I_k(f).$$

Hence

$$\hat{p}_1(0) = (1/B) \sum_{k=1}^{B} I_k(0)$$

$$= (1/BM) \left(\sum_{k=1}^{B} \left(\sum_{j=1}^{M} (X_{kj} - \overline{X})\right)^2\right)$$

$$= (1/BM) \sum_{k=1}^{B} (M\overline{X}_k - M\overline{X})^2$$

$$= (M/B) \sum_{k=1}^{B} (\overline{X}_k - \overline{X})^2$$

and hence

$$\hat{\sigma}_1^2(\overline{X}) = \hat{p}_1(0)\,/\,N = (1/B)^2 \sum_{k=1}^{B} (\overline{X}_k - \overline{X}^2)$$

$$\simeq (1/B)\,(1/(B-1)) \sum_{k-1}^{B} (\overline{X}_k - \overline{X})^2.$$

Thus the spectral estimate of $Var(\overline{X})$ obtained by averaging over periodograms of the batches is, approximately equal to the batch means estimate. To my knowledge this was not appreciated before the work of Meketon and Schmeiser. Bartlett however proposed this estimator only as a pedagogical step leading to a more accurate procedure involving the sample covariance function. This more accurate procedure used to estimate $p(0)$ as a step toward estimating $Var(\overline{X})$ is, in fact, the overlapping batch means approach of Meketon and Schmeiser.

Now the original "pedagogical" proposal of Bartlett for averaging periodograms of subseries has been generalized over the years by the addition of overlapping of the segments and the multiplication of the segments by a data window. It has been of interest for computational reasons. We assume segments again of length M each D units apart i.e. we let

$$X_k(j) = X_{D(k-1)+j} \qquad j = 1, \ldots, M$$

and assume for simplicity that the number of segments $C = 1 + ((N-M)/D)$ is an integer. Hence if $D = M$ we have the batches of non-overlapping batched means and if $D = 1$ we have the overlapping batches of Meketon and Schmeiser.

In the generalization (see Welch [1967]) we let

$$I_k(f) = \left| \sum_{j=1}^{M} W(j)\, X_k(j)\, e^{-2\pi i\,(j-1)f/M} \right|^2 /UM$$

where

$$U = (1/M) \sum_{j=1}^{M} W^2(j)$$

Here $W(j)$ is called a data window and $W(j) = 0$, $j < 1$ and $j > M$. The spectral estimate is the average of these periodograms, i.e.

$$\hat{p}_2(f) = (1/C) \sum_{k=1}^{C} I_k(f).$$

The spectral window is given by

$$H(f) = (1/MU) \mid \sum_{j=1}^{M} W(j)\, e^{-2\pi i f(j-1)} \mid^2.$$

Now, for a Gaussian process under fairly general conditions, when the passband of the estimator does not contain $f = 0$ or $1/2$ the variance of $\hat{p}_2(f)$ is given approximately by

$$Var[\hat{p}_2(f)] \simeq (p^2(f)/C) \left\{ \sum_{j=-\infty}^{\infty} \rho_D(j) \right\}$$

where $\rho_D(j) = [\sum_{k=1}^{M} W(k)\, W(k + jD)]^2 / [\sum_{k=1}^{M} W^2(k)]^2$ and at 0 and $1/2$ we have

$$Var[\hat{p}_2(0 \text{ or } 1/2)] \simeq (2p^2(0 \text{ or } 1/2)/C) \left(\sum_{j=-\infty}^{\infty} \rho_D(j) \right).$$

In the case of batch means

$$W(k) = 1 \qquad k = 1, \ldots, M$$
$$W(k) = 0 \qquad \text{otherwise}$$

Hence $\rho_D(j) = \sum_{|jD| \leq M} (1 - |jD|/M)^2$

and

$$Var[\hat{p}_2(0)] = (((2p^2(0))/((N - M + D)/D)) \sum_{j=-\infty}^{\infty} \rho_D(j)$$

Now if we consider the case $D = M$ corresponding to ordinary batch means, we obtain

$$Var[\hat{p}_2(0)] = 2p^2(0)/B.$$

and

$$Var[\hat{p}_2(0)] / E^2[\hat{p}_2(0)] = 2/B$$

which is correct for a χ^2_B random variable as we would expect.

Further if we consider the case corresponding to overlapping batch means we have $D = 1$ and

$$Var[\hat{p}_2(0)] = (2p^2(0)/(N - M + 1))(2M/3)$$

$$\simeq (2/3)(2p^2(0)/B).$$

The variance is reduced by a factor of $2/3$. This is the result of Meketon and Schmeiser .

4. EFFECT OF PARTIAL OVERLAPPING: A PRACTICAL SCHEME?

The formulas we have given above enable us to evaluate the change in the variance of the estimate $\hat{p}_2(0)$ for any choice of D. This could be of practical importance since as was pointed out earlier when the method of batch means is applied the number of batches required for testing for lack of correlation is greater than the number of required for a reasonable confidence interval. Hence the batches are rebatched for the confidence interval calculation. If they are rebatched by combining two batches into one, the method of overlapping batch means with $D = M/2$ can be easily applied. Here M is the batch size after rebatching. If they are rebatched by combining three batches into one, the method of overlapping batch means with $D = M/3$ can be easily applied; and so on. This should be computationally more efficient than calculating the sample correlation function (out to an uncertain lag M) as suggested in Meketon and Schmeiser. Also, as we will see, almost all the variance reduction can be achieved with these modest amounts of overlapping.

We first consider the case where $D = M/2$. We overlap by half the batch size. This would be easily achieved in practice if say one had forty batches for the test of independence and used them in an obvious fashion to realize thirty nine overlapped batches of twice the size. In this case

$$Var[\hat{p}_2(0)] = ((2p^2(0))/((N - (M/2))/(M/2)) \sum_{j=-\infty}^{\infty} \rho_{M/2}(j)$$

$$\simeq ((2p^2(0))/2B)(1 + (1/4) + (1/4))$$

$$\simeq (3/4)(2p^2(0)/B)$$

Thus if we overlap by one half the batch size we get a reduction in variance equal to .75 as against .67 for the most extreme overlapping. If we let $D = M/3$ we get

$$Var[\hat{p}_3(0)] \simeq (19/27)(2p^2(0)/B).$$

Hence in this case we get a reduction of .704 as against .67. Finally if $D = M/4$ we get 44/64 or .688. Below is a table giving the decrease in variance and the increase in degrees of freedom for these three cases and the extreme case.

D	Variance Decrease	Degrees of Freedom Increase
1	.667	1.5
M/4	.688	1.45
M/3	.704	1.42
M/2	.75	1.33

In the case where $D = 1$, an exact classical spectral estimator can be obtained by adding $M - 1$ zeros to the time series at either end and applying the extreme form of overlapping to the extended series. In this case the divisor is N not $N - M + 1$. The same trick would probably give marginal improvement when $D \neq 1$.

5. CONCLUSION

We have hopefully developed a useful additional to the overlapping batch means work of Meketon and Schmeiser. A large proportion of the variance reduction can be achieved with only modest overlapping. This modest overlapping is computationally consistent with the usual practice of rebatching the data to protect against residual correlation after a statistical test for lack of correlation has been passed.

6. REFERENCES

M. S. Meketon (1980), The Variance Time Curve: Theory, Estimation and Application, Ph.D. Dissertation, School of Operations Research and Industrial Engineering, Cornell University.

M. S. Meketon and B. Schmeiser (1984), "Overlapping Batch Means: Something for Nothing?", Proceedings of the Winter Simulation Conference. S. Sheppard, U. Pooch and D. Pegden (eds.), 227-230.

G. S. Fishman (1978), Principles of Discrete Event Simulation, Wiley-Interscience, New York.

M. S. Bartlett (1950), "Periodogram Analysis and Continuous Spectra", Biometrika, 37, 1-16.

P. D. Welch (1961), "A direct Digital Method of Power Spectrum Estimation", IBM J. Res. and Dev., 5, 141-156.

P. D. Welch (1967), "The Use of the Fast Fourier Transform for the Estimation of Spectra; A Method Based on Time Averaging Over Short, Modified Periodograms", IEEE Trans. Acoust. Electroacoust., AU - 15, 2, 70-73.

C. Bingham, M. D. Godfrey and J. W. Tukey, "Modern Techniques of Power Spectrum Estimation", IEEE Trans. Acoust. Electroacoust., AU-15, 2, 6-16.