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

ON STRONG CONSISTENCY OF THE VARIANCE ESTIMATOR

Halim Damerdji
Department of Industrial Engineering
University of Wisconsin-Madison
Madison, WI 53706

ABSTRACT

One way to construct a confidence interval for the mean constant of a stochastic process, is via consistent estimation of another parameter of the process, namely, the time-average variance constant. In this paper, we discuss strong consistency of the variance estimator for several methods of steady-state output analysis. These are; Batch Means (BM), Overlapping Batch Means (OBM), Spectral methods, and finally, Standardized Time Series (the area estimator of STS). A characterization of the spectral variance estimator is also presented; it is a generalization of OBM. Another estimator, which might be called Overlapping Area estimator, connects the area estimator with spectral methods.

1. INTRODUCTION

Suppose we observe the output sequence of a stochastic process $\{x_n : n \geq 1\}$, and that there exists an unknown parameter μ such that $\frac{1}{n}\sum_{i=1}^{n} x_i \Longrightarrow \mu$, as $n \to \infty$, where "$\Longrightarrow$" denotes weak convergence. The natural estimator of the parameter μ is evidently the sample mean, $\bar{x}(n) \equiv \frac{1}{n}\sum_{i=1}^{n} x_i$. To assess the accuracy of this estimator, typically, a confidence interval is constructed. For single long run methodology, there are two general approaches to this construction; i.) through consistent estimation of another parameter of the process, which is called the variance parameter, or ii.) through a cancellation method, such as Schruben's STS (1983).

Confidence intervals constructed via consistent estimation of the variance parameter, were shown in Glynn and Iglehart (1988), to have asymptotically shorter expected half-width, and smaller half-width variance. Furthermore, if the estimator is known to be strongly consistent (i.e. converging with probability one), then sequential procedures are asymptotically valid; see Chow and Robbins (1965).

We will discuss here strong consistency for several methods of steady-state output analysis. The general results were obtained in Damerdji (1987). In all the methods considered here, observations are grouped into batches; in applying one of these statistical procedures, one should decide, either beforehand or sequentially, the size of these batches relative to the total sample size. For simplicity, we will assume here that the batch size m_n is a fraction of the sample size, i.e. $m_n = n^{\alpha}$, where $0 < \alpha < 1$. The conditions we obtain in the theorems, provide theoretical insight into the relation between batch size and correlation of the process. A conclusion of our work is that batch size should be large when the process is heavily correlated. For the OBM and Spectral methods, the batch size cannot be too long though, as our conditions will indicate.

The assumption made on the process, is that it obeys a strong invariance principle; in Section 2, we discuss that it is a reasonable assumption in a simulation environment. The BM (resp. OBM, spectral, Area) method is taken up in Section 3 (resp. 4, 5, 6). Section 7 is the conclusion.

2. THE STRONG APPROXIMATION

Let S_n be the partial sum process, i.e. $S_n = \sum_{i=1}^{n} x_i$, and $S_0 = 0$. The strong invariance principle (or also strong approximation) states that the centered partial sum process $S_n - n\mu$, is close to a Brownian motion, and this with probability one. Using the tractability of the Brownian motion, one shows first the result sought for the Brownian motion, and then, by using the "closeness" from the strong approximation, infer the result for the original process.

From Philip and Stout (1975), strong approximation holds for a large class of processes; a brief discussion will be given below. Philip and Stout's conclusion is that for these processes, there exist constants σ and λ such that,

$$S_n - n\mu = \sigma B(n) + O(n^{1/2-\lambda}) \qquad w.p.1,$$

with $0 < \sigma < \infty$ and $0 < \lambda < 1/2$. The symbol "w.p.1" stands for "with probability one", while "O" denotes the classical big-Oh notation. The condition can be restated; there exist constants of the process σ and λ, such that for almost all sample paths ω of the pocess, there exists a constant $C(\omega)$ such that for all n,

$$|S_n(\omega) - n\mu - \sigma B(n,\omega)| < C(\omega)n^{1/2-\lambda}.$$

The constant λ will depend on the correlation of the process. The "nicer" the process is, in terms of correlation and moments, the closer it is to 1/2. If the correlation is high however, it is closer to 0. A lenghty discussion can be found in Damerdji (1987).

Philip and Stout (1975) showed that, under some more restrictions on the process, the strong approximation holds for example for i.) regenerative processes ($0 < \lambda < 1/4$), ii.) stationary φ-mixing ($0 < \lambda < 1/12$), and iii.) Strong mixing ($0 < \lambda < 1/264$). As mentioned in Glynn and Iglehart (1985), the regenerative property holds for a large class of processes, and so does strong mixing. See the latter reference for definitions. Note that these two assumptions do not include stationarity. In light of all this, the strong approximation assumption is then viable in a simulation context. We discussed above that the constant λ is ideally closer to 1/2 for low correlation processes. However, the bounds on λ are very small for these processes. This is due to mathematical difficulties, and we believe the true upper-bounds for λ are much closer to 1/2 than these. In all the remaining of the paper, we will assume that the strong approximation holds for the stochastic process under study.

3. BATCH MEANS

Batch Means, as discussed in Glynn and Iglehart (1988), is a cancellation method. The n observations are divided up into a fixed number k_n of batches, of size m_n each. The sample mean for each batch $j = 0, \ldots, k_n - 1$, is computed, i.e.

$$\bar{x}_j(m_n) = \frac{1}{m_n}\sum_{i=1}^{m_n} x_{jm_n+i}$$

If the process is well-behaved, then by some central limit theorem, $\bar{x}_j(m_n)$ will be asymptotically normal for a large batch size. Moreover, these values will also become asymptotically independent. Here, we will let the number of batches k_n grow to infinity as well, with the sample size. BM becomes then a consistent estimation method. Consider the sample variance of each batch, i.e.

$$\Gamma_{bm}(n) = \frac{1}{k_n - 1}\sum_{j=0}^{k_n-1}(\bar{x}_j(m_n) - \bar{x}(n))^2$$

This estimator is strongly consistent, as the following result indicates; its proof (for the general case) is in Damerdji (1987).

Theorem 1: If $\alpha > 1 - 2\lambda$, then

$$m_n\Gamma_{bm}(n) \to \sigma^2 \qquad w.p.1 \quad as\ n \to \infty.$$

Let us look at the condition of the theorem. For highly correlated processes, λ is closer to 0. For the theorem to be true then, α must be close to 1, and hence batches ought to be relatively large. On the other hand, for low correlation processes, λ is closer to 1/2, and hence α is not so restricted. Therefore, batch size is not crucial to get consistency in the low correlation case.

We discussed above that for regenerative processes, the constant λ for which the strong approximation holds, is smaller than 1/4. Hence, for $\lambda = 1/4$, the condition of the theorem reduces to $\alpha > 1/2$. The condition of the theorem suggests then, that in order to get strong consistency when applying BM for a "nice" regenerative stochastic process, one should take a number of batches very small compared to $\sqrt{n}$. We believe that one should take an even smaller number than that, as we discussed that the true λ should be closer to 1/2.

To give a flavor of how one uses the strong approximation assumption to carry out the proofs, we need to introduce some notation. Let,

$$\bar{A}_j(m_n) = \frac{1}{m_n}(B((j+1)m_n) - B(jm_n)),$$

$$\bar{A}(n) = \frac{1}{n}B(n),$$

and,

$$\widetilde{\Gamma}_{bm}(n) = \frac{1}{k_n - 1}\sum_{j=0}^{k_n-1}\left(\bar{A}_j(m_n) - \bar{A}(n)\right)^2.$$

The above quantities are the analogs of respectively, $\bar{x}_j(m_n)$, $\bar{x}(n)$, and $\Gamma_{bm}(n)$, but for the Brownian motion process. The proof goes in two steps.

i.) Show that $m_n\widetilde{\Gamma}_{bm}(n) \to 1$, $w.p.1$, as $n \to \infty$.

ii.) Using the strong approximation, show that

$$m_n\Gamma_{bm}(n) - \sigma^2 m_n\widetilde{\Gamma}_{bm}(n) \to 0 \quad w.p.1 \quad as \quad n \to \infty.$$

This of course implies that $m_n\Gamma_{bm}(n) \to \sigma^2$, with probability one as the sample size gets larger. These proofs can be found in Damerdji (1987).

4. OVERLAPPING BATCH MEANS

OBM, which was introduced by Meketon (1980), and Meketon and Schmeiser (1984), consists of overlapping the batches, that is, each observation starts a new batch. The resulting variance estimator is

$$\Gamma_{obm}(n) = \frac{m_n}{n - 2m_n + 1} \sum_{j=0}^{n-m_n} (\bar{x}_j(m_n) - \bar{x}(n))^2,$$

where $\bar{x}_j(m_n)$ is the sample mean over the j'th batch, i.e,

$$\bar{x}_j(m_n) = \frac{1}{m_n} \sum_{i=1}^{m_n} x_{j+i}.$$

In fact, from Meketon and Schmeiser (1984), the OBM estimator is closely related to the spectral estimator associated with the so-called modified Bartlett window. Let $\gamma_n(s)$ be the sample covariance of lag s, i.e. $\gamma_n(s) = \frac{1}{n}\sum_{t=1}^{n-s}(x_t - \bar{x}(n))(x_{t+s} - \bar{x}(n))$. We have the following lemma.

Lemma 2: (Meketon and Schmeiser, 1984)

$$\Gamma_{obm}(n) \approx \sum_{i=-(m_n-1)}^{m_n-1} (1 - \frac{|i|}{m_n})\gamma_n(i).$$

In the next section, we generalize the OBM estimator to include a large class of window kernels. Strong consistency of the OBM estimator will then follow as a special case.

5. SPECTRAL METHODS

To use a spectral estimator, one must choose a threshold value m_n, and a function $w_n(.)$ called window kernel. To be consistent with the other sections, we will call the threshold value (or also truncation point) m_n, the batch size. The spectral estimator, that we call $2\pi f_n(0)$, is given by,

$$2\pi f_n(0) = \sum_{i=-(m_n-1)}^{m_n-1} w_n(i)\gamma_n(i).$$

From Priestley (1981), in order to get consistency (in the mean square sense) of the spectral estimator, the truncation point must be such that $m_n \to \infty$, and also $n/m_n \to \infty$, as $n \to \infty$. The class of window kernels we consider here, will be large enough to include most of the windows available in the literature. Suppose that the window function is such that, i.) it is an even function between 0 and 1, ii.) $w_n(0) = 1$, and iii.) $w_n(s) = 0$ for $|s| \geq m_n$.

We now present a characterization of this estimator, but before that, we need to introduce some more notation. For $k = 1, \ldots, m_n$, let

$$\alpha_n(k) = m_n k^2 (w_n(k-1) - 2w_n(k) + w_n(k+1))$$

and consider the following variance estimator,

$$\Gamma_s(n) = \frac{1}{m_n n} \sum_{j=0}^{n-m_n} \sum_{k=1}^{m_n} \alpha_n(k) \left(\frac{S_{j+k} - S_j}{k} - \bar{x}(n) \right)^2.$$

Note that $\alpha_n(.)$ is a second difference.

Example 3: For the modified Bartlett window, i.e. $w_n(s) = 1 - |s|/m_n$, we have that $\alpha_n(k) = 0$ for $k = 1, \ldots, m_n - 1$, and $\alpha_n(m_n) = m_n^2$. Hence,

$$\Gamma_s(n) = \frac{m_n}{n} \sum_{j=0}^{n-m_n} (\bar{x}_j(m_n) - \bar{x}(n))^2 \approx \Gamma_{obm}(n).$$

We have the following proposition.

Proposition 4: For any window kernel from the class considered here,

$$\Gamma_s(n) \approx 2\pi f_n(0).$$

It is only approximate, due to some end-effects, as explained in the OBM case for example, in Meketon and Schmeiser (1984). Using this characterization, strong consistency of the general spectral estimator can be shown. See Damerdji (1987) for the proofs, and for the exact end-effects.

As an example, for the spectral estimator associated with the modified Bartlett window, the condition on the batch size is that $\alpha > 1 - 2\lambda$, and $\alpha < 1/2$. The first condition forces the batches to be large if the correlation is high, while the second condition tries to keep the batches relatively small, so that the end-effects remain asymptotically negligible. See the latter reference for the general case.

6. THE AREA ESTIMATOR

By modifying Schruben's area estimator (Schruben, 1983), that is by letting the number of batches grow to infinity with the sample size, one can show strong consistency of the resulting variance estimator. For $0 \leq j \leq k_n - 1$ and $1 \leq i \leq m_n - 1$, let

$$E_j(i) = (S_{jm_n+i} - S_{jm_n}) - \frac{i}{m_n}(S_{(j+1)m_n} - S_{jm_n})$$

and also

$$F_j(n) = \sum_{i=1}^{m_n-1} E_j(i).$$

From the latter reference, the area estimator is given by,

$$\Gamma_a(n) = \frac{12}{(m_n^3 - m_n)} \frac{1}{k_n} \sum_{j=0}^{k_n-1} F_j^2(n).$$

The conditions obtained for strong consistency will be identical to those of BM.

Theorem 5: If $m_n = n^\alpha$, and $\alpha > 1 - 2\lambda$, then

$$\Gamma_a(n) \rightarrow \sigma^2 \quad w.p.1 \text{ as } n \rightarrow \infty.$$

The same discussion about correlation and batch size, as in the BM case, follows. Finally, one can overlap the batches for the area estimator. For $j = 0, \ldots, n - m_n$, and $i = 1, \ldots, m_n - 1$, let

$$U_j(i) = (S_{j+i} - S_j) - \frac{i}{m_n}(S_{j+m_n} - S_j),$$

and,

$$V_j(n) = \sum_{i=1}^{m_n-1} U_j(i).$$

The variance estimator will be given by,

$$\Gamma_{oa}(n) = \frac{12}{(m_n - 1)(m_n^2 - 2m_n + 3)} \frac{1}{n} \sum_{j=0}^{n-m_n} V_j^2(n).$$

In fact, as in the OBM case, this estimator is almost identical to a spectral variance estimator. See Damerdji (1987) for the kernel window and the computations.

7. CONCLUSION

Strong consistency of the variance estimator for several methods of steady-state output analysis was discussed. The conditions obtained showed that batch size may be critical for consistency. When the process is heavily correlated, batches must be large enough. For low correlation however, batch size is not so crucial. For "nice" regenerative processes, we saw that for the BM and Area methods, the number of batches should be much smaller than $\sqrt{n}$. The characterization of the spectral variance estimator opens some interesting computational and numerical questions.

ACKNOWLEDGMENTS

The work described in this paper was part of a dissertation, undertaken under Professor Peter W. Glynn. The author is grateful to him. The author would also like to thank the Ministry of Higher Studies of Algeria, for financial support.

REFERENCES

Chow, Y.S., and Robbins, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Annals of Mathematical Statistics*, **36**, 457-462.

Damerdji, H. (1987). Strong consistency of the variance estimator in steady-state output analysis. In preparation

Glynn, P.W., and Iglehart, D.L., (1985). Large-sample theory for standardized time series: an overview. In: *Proceedings of the 1985 Winter Simulation Conference*, 128-134.

Glynn, P.W., and Iglehart, D.L., (1988). The theory of standardized time series. To appear in *Mathematics of Operations Research.*

Philip, W., and Stout, W., (1975). Almost sure invariance principles for partial sums of weakly dependent random variables. *Memoir of the American Mathematical Society*, **164**.

Meketon, M.S., (1980). The variance time-curve: theory, estimation, and application. Ph.D. Dissertation, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

Meketon, M.S., and Schmeiser, B., (1984). Overlapping Batch Means: something for nothing? In: *Proceedings of the 1984 Winter Simulation Conference*, 227-230.

Priestley, M.B., (1981). *Spectral Analysis and Time Series.* Academic Press, London.

Schruben, L.W., (1983). Confidence interval estimation using standardized time series. *Operations Research*, **31**, 1090-1108.

AUTHOR'S BIOGRAPHY

HALIM DAMERDJI is a Ph.D. student in the Department of Industrial Engineering at the University of Wisconsin-Madison. He received the equivalent of a B.S. degree in Mathematics in 1981, from the University of Sciences and Technology of Algiers, and a M.S. in Industrial Engineering in 1984, from the University of Wisconsin-Madison. His research interests are in simulation and related problems.

Halim Damerdji
Department of Industrial Engineering
University of Wisconsin-Madison
Madison, WI 53706, U.S.A.