Proceedings of the 1994 Winter Simulation Conference
ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila

ON THE BATCH MEANS AND AREA VARIANCE ESTIMATORS

Halim Damerdji

Department of Industrial Engineering
North Carolina State University
Raleigh, NC 27695–7906, U.S.A.

ABSTRACT

In steady-state simulation output analysis, construction of a consistent estimator of the the variance parameter of the process may be desirable in a number of instances. For example, if the process obeys a central limit theorem and an estimator of the variance is available, one may, then, construct an asymptotically valid confidence interval for the process-mean parameter. Centered moments (e.g., bias, variance, skewness, etc.) of an estimator are familiar measures of goodness of that estimator. Also, a central limit theorem involving the estimator provides its asymptotic rate of convergence.

We consider here the batch means and the standardized time series area variance estimators, in their nonclassical setting, and provide asympotic expressions for their centered moments as well as central limit theorems. As a by-product, consistency in the mean-square sense of these estimators is obtained. Our assumption on the process does not include stationarity nor covariance stationarity (although we are in the steady-state context).

1 INTRODUCTION

Often, the goal of a steady-state simulation is to estimate a mean system performance (only the univariate case is considered here). Let us assume that μ, the mean system performance of interest, can be estimated via a discrete-time stochastic process $\{X_n : n \geq 0\}$. The most natural estimator is, of course, the grand sample mean $\overline{\overline{X}}(n) \equiv (1/n) \sum_{i=1}^{n} X_i$, where n is the total number of observations of the process. Assuming ergodicity, the grand sample mean is close to μ, but is still variable nonetheless. The traditional approach to assess the precision of this estimator is to construct a confidence interval around it, and there are two general classes of methods for such a construction: the class of *cancellation* methods and the class of *consistent estimation* methods. The limit theorems these methods are based on involve the t-distribution for cancellation methods (typically) and the normal distribution for consistent estimation methods.

Consistent estimation methods are based around the construction of an estimator of a process parameter called the time-average variance constant of the process (often denoted σ^2). We will simply call σ^2 the variance of the process. When the process is covariance stationary (see Anderson (1971) for a precise definition), the process variance is simply the sum of the covariances at all lags. In the general ergodic case (i.e., the process is ergodic, yet nonnecessarily stationary nor covariance stationary), this process parameter is the constant which appears in the central limit theorem

$$\sqrt{n}\left(\overline{\overline{X}}(n) - \mu\right) \implies \sigma N(0,1)$$

as $n \to \infty$, assuming it holds. The symbol "$\implies$" denotes convergence in distribution and $N(a, b)$ stands for the normal distribution with mean a and variance b. The process variance σ^2 is also sometimes defined as the limit, as the sample size n grows, of $n\mathrm{Var}(\overline{\overline{X}}(n))$. If a consistent estimator $\Gamma(n)$ of σ^2 is available, an asymptotically valid confidence interval ensues from the above central limit theorem. However, there are also other advantages to consistent estimation of the variance parameter. On the basis of mean and variance of the confidence interval half-width, Glynn and Iglehart (1990) show that consistent-estimation-based confidence intervals are better behaved asymptotically than cancellation-based ones. (As they point out, however, it is not known whether one class dominates the other or not in the small-sample case.) Another advantage of consistent estimation of the process variance is that it provides additional information about the system. For example, when comparing two systems with comparable mean performances, one may decide to choose the system that has lower variability. The process variance is a measure of variability, and an estimator of the variance would provide an estimate of this variability. Another advantage of consistent estimation of the process variance is the following. From the above central limit theorem, one gets (assuming an extra condition) that $n\text{Var}(\overline{\overline{X}}(n)) \approx \sigma^2$ for n large, and, therefore, an estimate of the variance also provides an estimate of the variability of the sample mean, variability measured in terms of variance.

An advantage of *strong* consistency, or also consistency with probability one, is that it is one of the sufficient conditions of Glynn and Whitt (1992) in order to construct a fixed-width confidence interval for the process mean parameter. Strong consistency of the batch means and area variance estimators is considered in Damerdji (1994). *Mean-square* consistency of the variance estimator allows one to consider the bias and variance of this estimator, which are going down to 0; bias and variance of an estimator are familiar measures of "goodness" of that estimator. Mean-square consistency of the variance estimator also allows one to measure the quality, in terms of mean-square error, in the estimation of μ; see Goldsman and Meketon (1993). Also, research on *batch size* selection is often based on the assumption of mean-square consistency of the variance estimator, and, hence, the variance estimators must be a priori consistent in the mean-square sense; see Goldsman and Meketon (1993) and Schmeiser and Song (1993). (In a number of steady-state output analysis methods, a batch size parameter must be preset; this choice may dramatically affect the performance of the procedure.)

In this paper, we will consider the batch means and the standardized time series area variance estimators, and provide mean-square consistency. Asymptotic expressions for the centered moments (up to any order) of these variance estimators are given, as well as central limit theorems. We do obtain the expected result that the batch means and area variance estimators converge to the process variance at rate one over the square root of the number of batches.

We assume that the process obeys a strong approximation, or also, a strong invariance principle; this is discussed in the next section. Let us now define the two estimators. Let $\Gamma_{\text{bm}}(n)$ and $\Gamma_{\text{a}}(n)$ be the variance estimators associated with, respectively, the batch means and the standardized time series area methods. In both methods, the n observations $X_1,\ldots,X_n$ are divided up into a number k of adjacent batches, each of length m. This parameter m is called the batch size. In the *classical* setting, the number of batches is fixed; there, the batch means and area methods are *cancellation* methods, and no estimation of the process variance is attempted. If we let the number of batches (and the batch size) grow with the sample size, the batch means and area methods are *consistent* estimation methods: in order to construct an asymptotically valid confidence interval for the mean, the limiting theorem will be based on the normal distribution (and not on the t-distribution, which is typically the case for cancellation methods). Some notation is needed. For $j = 0, \ldots, k-1$, let $\overline{X}_j(m) \equiv m^{-1}\sum_{i=1}^{m} X_{jm+i}$. We assume that $km = n$ for simplicity. The batch means variance estimator is defined by

$$\Gamma_{\text{bm}}(n) \equiv \frac{m}{k-1}\sum_{j=0}^{k-1}\left(\overline{X}_j(m) - \overline{\overline{X}}(n)\right)^2,$$

and the area variance estimator by

$$\Gamma_{\text{a}}(n) \equiv \frac{12}{m(m^2-1)k}\sum_{j=0}^{k-1} F_j^2,$$

where $F_j = (1/2)\sum_{i=1}^{m}(m+1-2i)X_{jm+i}$. The batch means method is well known. The standardized time series approach was introduced by Schruben (1983).

2 THE ASSUMPTION ON THE PROCESS

The results presented in this paper will be under the assumption that the process statisfies a strong invariance principle, or also, a strong approximation. Consider the partial-sum process $\{S_n \equiv \sum_{i=1}^{n} X_i : n \geq 1\}$. We assume that there exists a process-constant $\delta \in (0, 1/2]$ and a random variable C with finite second moment such that, for almost all sample trajectories ω of the process, there exists an integer $n_0 \equiv n_0(\omega)$ such that we have

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

for all $\lambda \in (0, \delta)$ and $n \geq n_0$. The process $\{B(n) : n \geq 0\}$ is the standard Brownian motion process.

From the assumption, the parameter μ is the process mean parameter, while σ^2 is the process variance. This assumption also implies a strong law of large numbers and a central limit theorem. See Damerdji (1994) for an elaboration. The process constant δ depends on the autocorrelation and moment structure of the process; it is closer to $1/2$ for processes having little autocorrelation and admitting high moments, while it is closer to 0 for processes with high autocorrelation and/or low-only moments. See Damerdji (1994) for a lengthy discussion.

A wide class of ergodic processes do satisfy the assumption of strong approximation. A process is said to be *weakly dependent* if events far apart in time are almost independent (in a certain sense). Ergodic Markov chains are one example. Processes with the φ-mixing or the strong mixing property are other examples. Philipp and Stout (1975) show that a number of weakly dependent processes do satisfy a strong invariance principle.

A number of stochastic processes considered in a steady-state simulation are of the regenerative type, that is, there exists an increasing sequence $\{T_r : r \geq 0\}$ of finite (with probability one) random times such that the sequence $\{(X_{T_r}, \ldots, X_{T_{r+1}-1}, \tau_r) : r \geq 0\}$ of random vectors is independent and identically distributed, where $\tau_r \equiv T_{r+1} - T_r$. (The definition is not complete.) For future reference, let $U_r \equiv \sum_{j=T_r}^{T_{r+1}-1} |X_j|$ and $V_r \equiv \sum_{j=T_r}^{T_{r+1}-1} |X_j|^4$. From Damerdji (1993), a regenerative process will satisfy the assumption of strong approximation if $\mathrm{E}V_0 < \infty$, $\mathrm{E}\tau_0^2 < \infty$, and $\mathrm{E}U_0^2 < \infty$; we have $\delta = 1/4$ and $C = U_0 + 2$. Regenerative processes that satisfy the above conditions will be called "nice regenerative processes" in this paper.

In any event, we do not see the assumption of strong approximation as being stringent in the steady-state setting. We stress that the upper bound δ is related to the autocorrelation and moment structure of the process.

3 MEAN SQUARE CONSISTENCY, BIAS, AND VARIANCE

As previously mentioned, an estimator's centered moments such as bias, variance, skewness, and kurtosis are measures of goodness of that estimator. A central limit theorem is also informative about the quality of an estimator as it provides a rate of convergence. In this section, asymptotic expressions for the bias and variance of the batch means and area variance estimators are given. Consistency in the mean-square sense of these estimators is obtained as a by-product. (An estimator $Y(n)$ of a parameter θ is said to be consistent in the mean-square sense if $\lim_{n\to\infty} E[(Y(n) - \theta)^2] = 0$.) We place ourselves in the *nonclassical* setting by allowing the *number of batches* in the two procedures, as well as the batch sizes, to grow with the sample size.

We now make some assumptions on the batch size in the procedures. Consider batch sizes of the type $m = \gamma n^\alpha$, where $0 < \alpha < 1$ and $\gamma > 0$. (The theory for general batches is developped in Damerdji (1993).) We also assume that α is such that

$$1 - \delta < \alpha < 1.$$

Note that α is close to 1 for processes with high correlation, since δ is close to 0; therefore, $m = \gamma n^\alpha$ will be relatively large. If the process has low correlation, however, δ is closer to 1/2, and so α is closer to 1/2 also. Ideally, we would have liked α to be closer to 0 in this case, but 1/2 is the best lower bound we could achieve using our bounding techniques. For a nice regenerative process, we get that $\delta = 1/4$, and so, we consider $\alpha \in (3/4, 1)$.

In the remainder of the paper, we will be using the classical Big-Oh notation, i.e., $a(n) = O(b(n))$ if $|a(n)/b(n)|$ is bounded by a constant for n large enough. In the next proposition, the batch means and area variance estimators are generically denoted by $\Gamma(n)$.

Proposition 1 *We have that*

a)

$$\mathrm{Bias}\Big(\Gamma(n)\Big) = O\Big(m^{-1/2}n^{1/2-\lambda}(\log n)^{1/2}\Big)$$

b)

$$\frac{n}{m}\mathrm{Var}\Big(\Gamma(n)\Big) = 2\sigma^4 + O\Big(m^{-1}n^{1-\lambda}(\log n)^{1/2}\Big).$$

The rate of convergence we obtain on the bias is far from being as sharp as that obtained by Goldsman and Meketon (1993); computing the moments directly, as they did in their paper, gives sharper rates. We are able, however, to give a sharp bound on the variance.

As a consequence of the above proposition, it follows that the batch means and variance estimators are consistent in the mean-square sense. For the batch means method, this result was previously obtained by Carlstein (1986) and Goldsman and Melamed (1992) (the process was assumed to be strongly mixing and with stationary increments in the respective works). Here, we used a different assumption on the process, namely the strong approximation.

4 HIGHER CENTERED MOMENTS OF THE VARIANCE ESTIMATORS

We use here the power of the strong approximation and present asymptotic expressions for any order centered moment of the batch means and area variance estimators. Quantities, such as skewness and kurtosis, based on higher moments of an estimator are informative about the quality of that estimator.

Define the skewness and the kurtosis of $\Gamma_{\mathrm{bm}}(n)$ by $\mathrm{Skew}\Big(\Gamma_{\mathrm{bm}}(n)\Big) \equiv \mathrm{E}\Big[\Big(\Gamma(n) - \sigma^2\Big)^3\Big]$ and

$\text{Kurt}\left(\Gamma_{\text{bm}}(n)\right) \equiv \text{E}\left[\left(\Gamma(n) - \sigma^2\right)^4\right]$, respectively. These are nonstandard definitions, used for simplicity. Recall that we are not assuming that the process is stationary, but that it only satisfies the strong approximation. We get that

$$(k-1)^{3/2}\text{Skew}\left(\Gamma_{\text{bm}}(n)\right) = 8\sigma^6(k-1)^{-1/2} + O\left((k-1)^{-3/2}\right) + O\left(m^{-1}n^{1-\lambda}(\log n)^{1/2}\right)$$

and

$$(k-1)^2\text{Kurt}\left(\Gamma_{\text{bm}}(n)\right) = 12\sigma^8 + O\left((k-1)^{-1}\right) + O\left(m^{-1}n^{1-\lambda}(\log n)^{1/2}\right).$$

Similar results can be obtained for the area variance estimator, but with $(k-1)$ replaced by n/m in the above two expressions.

If the batch size m grows sufficiently fast, we are able to get the sharper bound

$$(k-1)^2\text{Skew}\left(\Gamma_{\text{bm}}(n)\right) = 8\sigma^6 + O\left((k-1)^{-1}\right) + O\left(m^{-3/2}n^{3/2-\lambda}(\log n)^{1/2}\right).$$

If the batch size grows sufficiently fast, we have, then, that $\text{Skew}\left(\Gamma_{\text{bm}}(n)\right) \approx 8\sigma^6/k^2$. For a nice regenerative process (and so $\alpha = 3/4$ from a previous discussion), we get that $\text{Skew}\left(\Gamma_{\text{bm}}(n)\right) \approx (8\sigma^6/\gamma^2)n^{-1/2}$. See also Chien (1993) for an asymptotic expression of the batch means variance estimator up to the third moment (via its cumulants).

In order to give expansions for the pth centered moment, we need to assume that $\text{E}[C^{2p}] < \infty$. We get the following proposition, where $\{K_q : q \geq 1\}$ is a sequence of integers, related to the χ^2-distribution (see Kendall, Stuart, and Ord (1987, p. 507) for a description).

Proposition 2 *If $\text{E}[C^{2p}] < \infty$, we have that*

$$(k-1)^{p/2}\text{E}\left[\left(\Gamma_{\text{bm}}(n) - \sigma^2\right)^p\right] = K_p\sigma^{2p} + O\left((k-1)^{-1}\right) + O\left(m^{-1}n^{1-\lambda}(\log n)^{1/2}\right)$$

for p even, and, for p odd, that

$$(k-1)^{p/2}\text{E}\left[\left(\Gamma_{\text{bm}}(n) - \sigma^2\right)^p\right] = K_p\sigma^{2p}(k-1)^{-1/2} + O\left((k-1)^{-3/2}\right) + O\left(m^{-1}n^{1-\lambda}(\log n)^{1/2}\right).$$

The results presented here are large-sample ones.

5 CENTRAL LIMIT THEOREMS

Central limit theorems can also be obtained. These give rates of convergence. One may also be able to construct an asymptotically valid confidence interval for the process variance, if one of the goals of the simulation is to estimate that parameter.

Proposition 3 *For the batch means and the area methods, we have that*

$$\sqrt{k}\left(\Gamma(n) - \sigma^2\right) \Longrightarrow N(0, 2\sigma^4)$$

as $n \to \infty$.

For the batch means case, this theorem can also be found in Carlstein (1986). It is new for the area estimator. An asymptotically $(1-\eta)100\%$-valid confidence interval for σ^2 is given by

$$\left[\frac{\Gamma(n)}{1 + z_{\eta/2}\sqrt{2/k}}, \frac{\Gamma(n)}{1 - z_{\eta/2}\sqrt{2/k}}\right],$$

where $z_{\eta/2}$ is the $(1-\eta/2)$-quantile of the standard normal distribution.

6 CONCLUSION

We have provided here asymptotic expressions for the moments of the batch means and the area variance estimators. As a by-product, we showed consistency in the mean-square sense of these two estimators. Central limit theorems were also provided. These results extend to the continuous-time case.

Variance estimators other than the batch means and area estimators, such as the weighted area and the Cramér-von Mises area variance estimators have been investigated in Damerdji and Goldsman (1993).

REFERENCES

Anderson, T. W. (1971). *Statistical Analysis of Time Series*, Wiley, New York.

Carlstein, E. (1986). "The Use of Subseries for Estimating the Variance of a Statistic from a Stationary Sequence," *Annals of Statistics*, **14**, 1171–1179.

Chien, C. (1993). "Cumulants of Batch Means for Stationary Processes," Technical Report.

Damerdji, H. (1993). "Mean-Square Consistency of the Variance Estimator in Steady-State Simulation Output Analysis," Tech. Rep. 93–7, Department of Industrial Engineering, North Carolina State University.

Damerdji, H. (1994). "Strong Consistency of the Variance Estimator in Steady-State Simulation

Output Analysis," *Math. of Operations Research*, **19**, 494–512.

Damerdji, H. and D. Goldsman (1993). "Consistency of Several Variants of the Standardized Time Series Variance Estimator," Tech. Rep. 93-14, Department of Industrial Engineering, North Carolina State University.

Glynn, P. W. and D. L. Iglehart (1990). "Simulation Output Analysis Using Standardized Time Series," *Math. of Operations Research*, **15**, 1–16.

Glynn, P.W. and W. Whitt (1992). "The Asymptotic Validity of Sequential Stopping Rules for Stochastic Simulations," *Annals of Applied Probability*, **2**, 180–198.

Goldsman, D. and B. Melamed (1992). "A Large Sample Result for batch Means," Tech. Rep. School of ISyE, Georgia Tech.

Goldsman, D. and M. S. Meketon (1993). "A Comparison of Several Variance Estimators for Stationary Increments Stochastic Processes," Tech. Rep. No. J–85–12, School of ISyE, Georgia Tech.

Kendall, M., A. Stuart, and J. K. Ord (1987). *Kendall's Advanced Theory of Statistics, Volume 1*, 5th edn., Oxford University Press, New York.

Philipp, W., and W. Stout (1975). "Almost Sure Invariance Principles for Partial Sums of Weakly Dependent Random Variables," *Mem. Amer. Math. Soc.*, **161**.

Schruben, L. W. (1983). "Confidence Interval Estimation Using Standardized Time Series," *Operations Research*, **31**, 1090–1108.

Schmeiser, B. W. and W.-M. T. Song (1993). "Optimal Mean-Squared-Error Batch Sizes," to appear in *Management Science.*

HALIM DAMERDJI is an Assistant Professor in the Department of Industrial Engineering at North Carolina State University. His research interests are in simulation and applied stochastic processes. He is a member of ORSA.