ESTIMATING THE ASYMPTOTIC VARIANCE WITH BATCH MEANS

by

Peter W. Glynn

Department of Operations Research
Stanford University
Stanford, CA 94305-4022

and

Ward Whitt

AT&T Bell Laboratories
Murray Hill, NJ 07974-2070

March 28, 1990

Revision: December 11, 1990

Abstract

We show that there is no batch-means estimation procedure for consistently estimating the asymptotic variance when the number of batches is held fixed as the run length increases. This result suggests that the number of batches should increase as the run length increases for sequential stopping rules based on batch means.

Keywords: simulation, estimation, batch means, consistency.

1. Introduction

The question investigated in the present paper arose in the study of sequential stopping rules for simulation estimators to achieve specified confidence intervals. The general goal is to have a procedure incorporated in the simulation program which enables the program to automatically terminate when the desired statistical precision has been attained. A desirable property of any candidate stopping rule is asymptotic validity as the prescribed width of the confidence interval approaches zero (and the resulting run length approaches infinity). By asymptotic validity, we mean that the probability of coverage (i.e., that the value to be estimated is contained in the confidence interval) converges to the designated value (e.g., 0.95) as the prescribed width of the confidence interval approaches zero. Of course, asymptotic validity is not sufficient for a procedure to be useful, but it seems to be a very desirable property.

In [5] we defined some sequential stopping rules and established general conditions for their asymptotic validity. There are two requirements: a functional central limit theorem for the estimation process and strong consistency (with-probability-one convergence) for the variance estimator. We then became interested in sequential stopping rules based on (non-overlapping) batch means. It seemed intuitively clear that asymptotic validity requires that the number of batches must increase without limit as the run length increases. In this paper, we establish a closely related result: We show that the number of batches must increase without limit as the run length increases in order to obtain a consistent (weak or strong) variance estimator. This result suggests that sequential stopping rules based on batch means should have the number of batches increase without limit as the run length increases, but we have only proved that the sufficient conditions for asymptotic validity in [5] are not satisfied with a fixed number of batches. It remains to show that asymptotic validity can not hold for batch means with a fixed number of batches.

Of course, it is well known that the standard variance estimator based on batch means is not consistent when the number of batches is held fixed (see Section 3). We want to establish a much stronger result, namely, that *there is no variance estimator based on a fixed number of batches that is consistent.* This result is not very difficult, but a proper formulation seems to require some care.

In Section 2 we formulate the problem and state the main theorem. In Section 3 we review the situation for the standard estimator. In Section 4 we prove the main theorem.

Our result also has relevance to fixed-run-length simulations. However, in that context, asymptotically valid confidence intervals are obtained anyway by cancellation methods, e.g., using the t distribution. For further discussion, see Schmeiser [9], Goldsman and Meketon [6], Sargent, Kang and Goldsman [8], Glynn and Iglehart [4] and Damerdji [2].

2. The Main Result

To be precise, we must first specify what we mean by an estimation procedure. To be interesting, an estimation procedure should apply to a large family of stochastic processes. Hence, let $X \equiv \{X(t) : t \geq 0\}$ be a measurable mapping from a measure space $(\Omega, \wedge\ )$ into $D \equiv D[0, \infty)$, the space of right-continuous real-valued functions on the interval $[0, \infty)$ with left limits, endowed with the usual Skorohod topology and associated Borel σ-field; e.g., see Ethier and Kurtz [3]. Of course, we want the underlying space $(\Omega, \wedge\ )$ to be sufficiently rich; it suffices to let $\Omega = D$ and $X(t)$ be the projection or coordinate map. We consider the set 3 of all probability measures P on $(\Omega, \wedge\ )$ such that X satisfies a FCLT with a Brownian motion limit, i.e., there exist finite constants $\mu \equiv \mu(P)$ and $\sigma \equiv \sigma(P)$ such that

$$n^{-1/2} \int_0^{nt} [X(s) - \mu] \, ds \Rightarrow \sigma B(t) \text{ as } n \to \infty \ , \tag{1}$$

where $\Rightarrow$ denotes weak convergence in D with respect to P and $B \equiv \{B(t) : t \geq 0\}$ is standard (zero drift, unit diffusion coefficient) Brownian motion. Our goal is to estimate σ^2, but we want our procedure to apply to all $P \in 3$. In other words, the procedure should apply to all stochastic processes X in D satisfying the functional central limit theorem (FCLT) (1). In fact, we only need convergence of all finite-dimensional distributions in (1), but the ordinary one-dimensional central limit theorem (CLT) obtained by setting $t = 1$ in (1) is not sufficient to analyze even the standard estimator (Section 3). Our formulation (1) assumes that X is a continuous-time stochastic process, but we could have $X(t) = X([t])$ for all t, where $[t]$ is the integer part of t, in which case X is effectively a discrete-time process and the normalized integral in (1) is asymptotically equivalent to the usual normalized sum.

To apply the method of batch means, we specify the number m of batches and the total run length T. We then construct our estimates from the m non-overlapping intervals of length T/m; i.e., let the i^{th} batch mean be

$$\overline{X}_i(T) = \frac{m}{T} \int_{(i-1)T/m}^{iT/m} X(s)\, ds \ , \ i = 1, \ldots, m \ . \tag{2}$$

We now want a procedure for combining the m observations $\overline{X}_1(T), \ldots, \overline{X}_m(T)$ in such a way that σ^2 is consistently estimated as $T \to \infty$. This ''combining transformation'' should not depend on the ''fine structure'' of the process X. In particular, it should not depend on μ and σ^2. Thus, in this context we say that an *estimation procedure* is a family of measurable mappings

$$g_T : R^m \to R \quad \text{for } T > 0 \ , \tag{3}$$

such that the estimate of σ^2 is $g_T(x_1, \ldots, x_m)$ when the total run length is T and $x_i = \overline{X}_i(T)$, $i = 1, \ldots, m$. Note that g_T can depend on T, but is independent of P.

We say that an estimation procedure is 3 -*consistent* if for each $P \in 3$

$$g_T(\overline{X}_1(T), \ldots, \overline{X}_m(T)) \Rightarrow \sigma^2(P) \ \text{as} \ T \to \infty \ . \tag{4}$$

Here $\Rightarrow$ denotes weak convergence with respect to P in R, which is equivalent to convergence in probability since $\sigma^2(P)$ is deterministic. Since we have a negative result, we focus on this weak consistency. We would have strong consistency if the convergence was w.p.1 with respect to P.

Here is our main result. It applies to any m.

Theorem 1. There does not exist a batch-means estimation procedure based on a fixed number of batches that is 3 *-consistent.*

In Section 3 we show what happens with the standard variance estimator. We see that we do not get consistency for σ^2 for any fixed m, but we can get as close as we wish by letting m be suitably large. In Section 4 we prove Theorem 1.

3. The Standard Estimator

The standard estimation procedure is specified by

$$g_T^*(x_1, \ldots, x_m) = \frac{T}{m(m-1)} \sum_{i=1}^{m} \left[x_i - \frac{1}{m} \sum_{k=1}^{m} x_k \right]^2 \tag{5}$$

for all $T > 0$, $m \geq 2$ and $(x_1, \ldots, x_m) \in R^m$. Let $\stackrel{d}{=}$ denote equality in distribution.

Theorem 2. Under (1),

$$\begin{aligned} g_T^*(\bar{X}_1(T), \ldots, \bar{X}_m(T)) &\Rightarrow \sigma^2 m^2 g_1^*([B(i/m) - B((i-1)/m)]\ ,\ 1 \leq i \leq m) \\ &\stackrel{d}{=} \frac{\sigma^2 \chi^2_{m-1}}{m-1} \ \text{ in } R\ , \end{aligned}$$

where χ^2_{m-1} is a chi-square random variable with $m-1$ degrees of freedom.

Proof. Note that

$$g_T^*(\overline{X}_1(T), \ldots, \overline{X}_m(T)) = \frac{m}{m-1} \sum_{i=1}^{m} \left[\sqrt{T} \frac{\int_{(i-1)T/m}^{iT/m} X(s)\,ds}{T} - \sqrt{T} \frac{\int_0^T X(s)\,ds}{mT} \right]^2$$

$$= \frac{m}{m-1} \sum_{i=1}^{m} \left[T^{-1/2} \int_{(i-1)T/m}^{iT/m} [X(s) - \mu]\,ds - m^{-1} T^{-1/2} \int_0^T [X(s) - \mu]\,ds \right]^2$$

$$\Rightarrow \sigma^2 \frac{m}{m-1} \sum_{i=1}^{m} \left[B(i/m) - B((i-1)/m) - \frac{B(1)}{m} \right]^2 \stackrel{d}{=} \frac{\sigma^2 \chi^2_{m-1}}{m-1} \quad \text{as } T \to \infty$$

by (1) and the continuous mapping theorem (Corollary 1.9 on p. 103 of [3]) using the function $h : D \to R$ defined for any $x \in D$ by

$$h(x) = \frac{m}{m-1} \sum_{i=1}^{m} [x(i/m) - x((i-1)/m) - x(1)]^2 . \quad \blacksquare$$

Note that $\sigma^2 \chi^2_{m-1}/(m-1)$ has mean σ^2 and variance $2\sigma^4/(m-1)$; e.g., see p. 168 of Johnson and Kotz [7]. Moreover, as m increases,

$$\frac{\sigma^2 \chi^2_m}{m} \Rightarrow \sigma^2 \quad \text{and} \quad \sqrt{m} \left[\frac{\sigma^2 \chi^2_m}{m} - \sigma^2 \right] \Rightarrow N(0, 2\sigma^4) , \tag{6}$$

where $N(a, b)$ denotes a normally distributed random variable with mean a and variance b. Hence, we can get as close as we want if we choose m suitably large. Moreover, we can obtain consistency under extra regularity conditions if $m \to \infty$ and $T \to \infty$ so that $T/m \to \infty$; see Goldsman and Meketon [6] and Damerdji [2]. In fact, Damerdji even proves strong consistency for a class of stochastic processes.

4. Proof of Theorem 1

To establish the negative result, it suffices to restrict attention to probability measures P such that $X(t) = X([t])$ for all $t \geq 0$, i.e., X is actually a discrete-time stochastic process, with $X(0) = 0$ and $\{X(k) : k \geq 1\}$ being a sequence of i.i.d. $N(0, \sigma^2)$ random variables. Then, for any m, when T is an integer multiple of m,

$$(\bar{X}_1(T), \ldots, \bar{X}_m(T)) \stackrel{d}{=} \left[\frac{m}{T} \sigma B(T/m), \ldots, \frac{m}{T} \sigma [B(T) - B(T(m-1)/m)] \right] , \qquad (7)$$

i.e., these batch means are distributed exactly as m i.i.d. $N(0, \sigma^2 m/T)$ random variables. Without loss of generality, we can remove the m/T factor by considering the transformed functions

$$\tilde{g}_T(x_1, \ldots, x_m) = g_T \left[\sqrt{\frac{m}{T}} x_1, \ldots, \sqrt{\frac{m}{T}} x_m \right] . \qquad (8)$$

Note that

$$g_T(\bar{X}_1(T), \ldots, \bar{X}_m(T)) \stackrel{d}{=} \tilde{g}_T(\sigma \mathbf{N}) \qquad (9)$$

for all T an integer multiple of m, where $\sigma \mathbf{N} \equiv (\sigma N_1, \ldots, \sigma N_m)$ and $\mathbf{N}$ is a fixed vector of i.i.d. $N(0, 1)$ random variables.

To have consistency, we must have

$$\tilde{g}_T(\sigma \mathbf{N}) \Rightarrow \sigma \text{ as } T \to \infty \qquad (10)$$

for all $\sigma > 0$, but this cannot happen for two or more different positive values of σ, say σ_1 and σ_2. To see this, first note that the convergence in probability for σ_1 in (10) implies that there is a deterministic subsequence $\{T_n : n \geq 1\}$ of $\{km : k \geq 1\}$ such that

$$\tilde{g}_{T_n}(\sigma_1 \mathbf{N}) \to \sigma_1 \quad \text{w.p.1 as } n \to \infty ; \qquad (11)$$

see Theorem 4.2.3 of Chung [1]. By (10), $\tilde{g}_{T_n}(\sigma_2 \mathbf{N}) \Rightarrow \sigma_2$ as $n \to \infty$. Hence, there is a deterministic subsequence $\{T'_n : n \geq 1\}$ of $\{T_n : n \geq 1\}$ such that

$$\tilde{g}_{T'_n}(\sigma_i \mathbf{N}) \to \sigma_i \quad \text{w.p.1 as } n \to \infty \qquad (12)$$

for both $i = 1$ and 2. Hence, for $i = 1$ and 2, $\tilde{g}_{T'_n}(x) \to \sigma_i$ for almost all x with respect to the law of $\sigma_i \mathbf{N}$, which implies that $\tilde{g}_{T_n'}(x) \rightarrow \sigma_i$ for almost all x with respect to Lebesgue measure on R^m, since $\sigma_i \mathbf{N}$ has a positive density with respect to Lebesgue measure. (See the appendix). However, it is not possible to have $\tilde{g}_{T_n'}(x)$ simultaneously converge almost everywhere with respect to Lebesgue measure to two different limits. (The set of convergence to one limit must be contained in the null set of non-convergence for the other limit.) ■

Acknowledgment. We thank Professors David Goldsman, Barry Nelson and James R. Wilson for helpful suggestions.

References

[1] Chung, K. L. (1974) *A Course in Probability Theory*, Second ed., Academic Press, New York.

[2] Damerdji, H. (1989) Strong consistency of the variance estimator in steady-state simulation output analysis. Department of Operations Research, Stanford University.

[3] Ethier, S. N. and T. G. Kurtz (1986) *Markov Processes, Characterization and Convergence*, Wiley, New York.

[4] Glynn, P. W. and D. L. Iglehart (1990) Simulation output analysis using standardized times series. *Math. Oper. Res.* 15, 1-16.

[5] Glynn, P. W. and W. Whitt (1991) The asymptotic validity of sequential stopping rules in stochastic simulation. *Ann. Appl. Prob.* 1, to appear.

[6] Goldsman, D. and M. Meketon (1986) A comparison of several variance estimators. School of Industrial and Systems Engineering, Georgia Institute of Technology.

[7] Johnson, N. L. and S. Kotz (1970) *Continuous Univariate Distributions-1, Distributions in Statistics*, Wiley, New York.

[8] Sargent, R. G., K. Kang and D. Goldsman (1989) An investigation of finite sample behavior of confidence interval estimation procedures. Department of Industrial Engineering, Syracuse University.

[9] Schmeiser, B. (1982) Batch size effects in the analysis of simulation output. *Oper. Res.* 30, 556-568.

Appendix

Here we give extra details showing that (12) implies that $\tilde{g}_{T_n'}(x) \to \sigma_1$ as $n \to \infty$ for almost all x with respect to Lebesgue measure on R^m. Let $A = \{x : \lim_{n \to \infty} \tilde{g}_{T_n'}(x) = \sigma_1\}$. Then

$$0 = P(\sigma_1 \mathbf{N} \in A^c) = \int_{A^c} f(x) \lambda(dx)$$

where λ is Lebesgue measure and f is the density of $\sigma_1 \mathbf{N}$, which is strictly positive almost everywhere. Let $B_n = \{x : n^{-1} \le f(x) < (n-1)^{-1}\}$ for $n \ge 2$ and $B_1 = \{x : 1 \le f(x)\}$. Then

$$0 = \int_{A^c} f(x) \lambda(dx) = \sum_{n=1}^{\infty} \int_{A^c \cap B_n} f(x) \lambda(dx) \ge \sum_{n=1}^{\infty} n^{-1} \lambda(A^c \cap B_n) \ge 0 ,$$

so that $\lambda(A^c \cap B_n) = 0$ for all n and

$$\lambda(A^c) = \sum_{n=1}^{\infty} \lambda(A^c \cap B_n) = 0 .$$