Proceedings of the 1995 Winter Simulation Conference
ed. C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman

A MAXIMUM-LIKELIHOOD INTERPRETATION OF BATCH MEANS ESTIMATORS

Kevin J. Healy

School of Industrial Engineering
Purdue University
West Lafayette, IN 47907-1287, U.S.A.

ABSTRACT

We show how the classical batch means estimator of the variance parameter of a strictly stationary dependent stochastic process can be viewed as a maximum likelihood estimator based on asymptotic properties of the standardized time series of observations from the process.

1 INTRODUCTION

There exist several methods for estimating steady state parameters of a strictly stationary dependent simulation output process from a single simulation run. Let $\mathbf{X} = \{X_1, X_2, ..., X_n\}$ represent the output of such a process with mean $\mu = E(X_i)$, and process variance parameter

$$\begin{aligned}\sigma^2 &= Var(X_i) + 2\sum_{i=1}^{\infty}\gamma_X(i) \\ &= \lim_{n\to\infty} n\bar{X}_n\end{aligned}$$

where

$$\gamma_X(i) = Cov(X_1, X_{1+i})$$

and

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Usually, the mean value of the process, μ, is the parameter of interest but estimating σ^2 is also of interest both as a direct measure of the variability of the process as well as in constructing confidence intervals to assess the precision of $\bar{X}_n$. Several estimation procedures have been proposed including batch means, spectral, regenerative, autoregressive, and standardized time series (STS) methods. In a study of the STS method, Glynn and Iglehart (1990) identify a general class of transformations of the original process that lead to STS interval estimation procedures and one mapping in particular that yields the batch means method. In this paper we point out another interesting relationship that exists between the methods in which the batch means estimator of the process variance parameter can be viewed in the context of STS as arising from the straightforward application of the method of maximum likelihood.

In Section 2 we review briefly the batch means and STS methods. In Section 3 we derive the corresponding estimator of the process variance parameter from first principles using the method of maximum likelihood. In Section 4, we conclude by suggesting some potentially useful applications of the result.

2 BATCH MEANS AND STS

In the method of batch means, the random sample $X_1, X_2, ..., X_n$ is divided into b adjacent, nonoverlapping, batches each of size m. As a matter of convenience, we assume b and m are positive integers and the sample size is such that $n = mb$. The i^{th} batch consists of observations $X_{1+(i-1)m}, X_{2+(i-1)m}, \ldots, X_{im}$. For $i = 1, 2, ...b$,

$$\bar{X}_{im} = \frac{1}{j}\sum_{k=1}^{m} X_{k+(i-1)m}$$

is the sample mean of the observations from the ith batch. The sample variance of the process $\{\bar{X}_{1m}, \bar{X}_{2m}, \ldots, \bar{X}_{bm}\}$ is denoted by

$$S^2 = \frac{1}{b-1}\sum_{i=1}^{b}(\bar{X}_{im} - \bar{X}_n)^2$$

For fixed $b > 1$, it can be shown (Schmeiser 1982) that

$$\frac{m(b-1)S^2}{\sigma^2} \longrightarrow \chi^2_{b-1} \quad as\ m \to \infty$$

where χ^2_{b-1} denotes a Chi-squared random variable with $b-1$ degrees of freedom which leads to the following method of moments estimator for σ^2

$$\hat{\sigma}^2 = mS^2. \tag{1}$$

An interval estimation procedure for μ is obtained from the related result

$$\frac{\sqrt{b}(\bar{X}_n - \mu)}{\hat{\sigma}_{BM}} \longrightarrow T_{b-1} \text{ as } m \rightarrow \infty$$

where T_{b-1} denotes a Student's t random variable with $b-1$ degrees of freedom.

The basis of the STS method (Schruben 1983) is the existence of a functional limit theorem for the sample mean of the simulated process. Let

$$Y_n(t) = \frac{[nt](\bar{X}_n([nt]) - \mu t)}{\sqrt{n}}, \quad 0 \leq t \leq 1$$

where $\bar{X}_n(j)$ denotes the cumulative average of the first j observations of the original process and [ ] denotes the greatest integer function. To apply the STS method, it is necessary to assume that

$$\mathbf{Y}_n \Longrightarrow \sigma \mathbf{W} \quad as\ n \rightarrow \infty$$

where $W(t)$ is the standard Weiner (Brownian motion) process. It follows that the so-called standardized series

$$T_n(t) = Y_n(1) - Y_n(t), \quad 0 \leq t \leq 1, \tag{2}$$

converges weekly to $\sigma B(t)$ where B is the Brownian bridge process defined as $W(t)$ conditioned such that $W(0) = W(1) = 0$. From this can be derived several estimators for σ^2 as well as their asymptotic distributions including one based on the signed area under the standardized process

$$A_n = \int_0^1 T_n(t)dt$$

and an another based on the maximum of the standardized process

$$M_n = T_n(\acute{t}_n)$$

where

$$\acute{t}_n = \inf_{0 \leq t,s \leq 1} \{t : T_n(t) \geq T_n(s)\}$$

is the random time index where the standardized series achieves a maximum.

In their characterization of the STS method, Glynn and Iglehart (1990) identify a general class of real-valued transformations of the cumulative average process, $\bar{\mathbf{X}}_n$, that obey a scalar central limit theorem and lead to interval estimation procedures for μ. Specifically, they establish properties of the transformation $g(\cdot)$ sufficient to insure

$$\frac{\bar{X}_n(n) - \mu}{g(\bar{\mathbf{X}}_n)} \longrightarrow \frac{W(1)}{g(\mathbf{W})} \quad as\ n \rightarrow \infty.$$

The defining characteristic of the mapping is that the asymptotic distribution of the resultant pivot is, in each case, independent of the unknown scale parameter σ due to the effect of an algebraic cancellation that occurs much like it does in the construction of a t-statistic. In an application of this result, they demonstrate that

$$g_b(x) = \left(\frac{b}{b-1} \sum_{i=1}^{b} (\Delta_m x(i/b) - x(1)/b)^2 \right)^{1/2}$$

where $\Delta_b x(t) = x(t) - x(t - 1/b)$ yields the batch means-based pivot defined in (1). In particular, the quantity

$$\frac{1}{b-1} \sum_{i=1}^{b} (\Delta_b T_n(i/b) - T_n(1)/b)^2 \tag{3}$$

can be associated with the batch means-based estimator for σ^2 given by (1). In the following section, we show how this estimator can be derived in an alternate way by straightforward application of the method of maximum likelihood in the context of the limiting distribution of the process $T_n(t)$.

3 A Maximum Likelihood Interpretation

By definition, the functional central limit theorem that applies to (2) implies that all finite dimensional distributions of the standardized process $T_n(t)$ are also convergent. In particular, for any fixed $b \geq 2$ the vector of $b-1$ equally spaced points in time on the standardized process,

$$(\ T_n(1/b),\ T_n(2/b), \ldots,\ T_n((b-1)/b)\ ),$$

converges weakly to the random vector

$$\sigma(\ B(1/b),\ B(2/b), \ldots,\ B((b-1)/b)\ )$$

as $n \rightarrow \infty$. For $0 \leq s \leq t \leq 1$, the covariance of the zero mean, normal random variables $B(s)$ and $B(t)$ is

$$Cov(B(s), B(t)) = s(1-t) \tag{4}$$

so that the limiting joint density

$$f(\mathbf{x}, \sigma^2) = \frac{1}{(2\pi)^{\frac{b-1}{2}} \sqrt{|\Sigma|}} \exp\left\{ -\frac{1}{2} \mathbf{x} \Sigma^{-1} \mathbf{x}' \right\}$$

is multivariate normal with zero mean vector and covariance matrix Σ with entries

$$\Sigma_{(i,j)} = \sigma^2 (\tfrac{min(i,j)}{b})(1 - \tfrac{max(i,j)}{b}), \quad 1 \leq i,j \leq b-1.$$

It is also not difficult to show that for $1 \leq i, j \leq b-1$ the entries of Σ^{-1} take the following form

$$\Sigma^{-1}_{(i,j)} = \begin{cases} \frac{2b}{\sigma^2} & \text{if } i = j \\ -\frac{b}{\sigma^2} & \text{if } |i-j| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

More generally, $f(\mathbf{x}, \sigma^2)$ is a one-parameter exponential family member in σ^2 with canonical representation

$$f(\mathbf{x}, \theta) = \exp\{c(\theta)T(\mathbf{x}) + d(\theta) + S(\mathbf{x})\}\, I_A(\mathbf{x})$$

where $\theta = \sigma^2$,

$$C(\theta) = -\frac{1}{\sigma^2}, \quad d(\theta) = -\ln\left\{(2\pi)^{\frac{b-1}{2}}\sqrt{|\Sigma|}\right\},$$

$$T(\mathbf{x}) = b\sum_{i=1}^{b-1}(x_i^2 - x_i x_{i+1}), \quad S(\mathbf{x}) = 0,$$

and $A = \Re^{b-1}$. General theory (see, for instance, Bickel and Doksum 1977) prescribes that the the maximum likelihood estimator of θ is the unique solution to the equation

$$E_\theta[T(\mathbf{X})] \quad = \quad T(\mathbf{x})$$

which in this case yields

$$E_{\sigma^2}\left[b\sum_{i=1}^{b-1}(x_i^2 - x_i x_{i+1})\right] \quad = \quad b\sum_{i=1}^{b-1}(x_i^2 - x_i x_{i+1}).$$

with $x_i = T_n(i/b)$ and $x_b = T_n(1) = 0$. After some simple algebraic manipulations involving the repeated application of (4), one can easily verify that the resultant maximum likelihood estimator is the same as (3).

4 CONCLUSIONS

The maximum likelihood interpretation of the batch means process variance parameter has several potentially interesting applications. For example, variants of the batch means method, including overlapping and spaced batched means methods lend themselves to similar interpretations. There has also been a significant amount of work employing various methods to show that these procedures for estimating the process variance parameter are strongly consistent as well as consistent in the mean square sense (see for instance, Damerdji 1995 and Chien, Goldsman, and Melamed 1994). It is possible that large sample theory applicable to maximum likelihood estimators may provide a useful framework for establishing such results.

REFERENCES

Bickel, P. J., K. J. Doksum. 1977. *Mathematical Statistics : Basic Ideas and Selected Topics* Oakland: Holden-Day, Inc.

Chien, C., D. Goldsman, and B. Melamed. 1994. A large-sample result for batch means. Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta.

Damerdji, H. 1995. Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Operations Research* 43:282–291.

Glynn, P. W., D. L. Iglehart. 1990. Simulation output analysis using standardized time series. *Mathematics of Operations Research* 15:1–16.

Schmeiser, B. 1982. Batch size effects in the analysis of simulation output. *Operations Research* 30:556–558.

Schruben, L. 1983. Confidence interval estimation using standardized time series. *Operations Research* 31:1090–1108.

AUTHOR BIOGRAPHY

KEVIN J. HEALY is an assistant professor in the School of Industrial Engineering at Purdue University. He earned B.S. and M.S. degrees in Industrial Engineering from The Pennsylvania State University and a Ph.D. in Operations Research from Cornell University. His research interests include stochastic optimization, simulation output analysis, and simulation modeling formalisms.