Proceedings of the 1986 Winter Simulation Conference
J. Wilson, J. Henriksen, S. Roberts (eds.)

LARGE AND SMALL SAMPLE COMPARISONS OF VARIOUS VARIANCE ESTIMATORS

David Goldsman
School of ISyE
Georgia Institute of Technology
Atlanta, GA 30332

Keebom Kang
Department of I.E.
University of Miami
Coral Gables, FL 33124

Robert G. Sargent
Department of IE & OR
Syracuse University
Syracuse, NY 13244

ABSTRACT

The objective of this report is to investigate the behavior of various variance estimators arising in computer simulation experiments. We give large and small sample results pertaining to the bias, variance, and confidence interval performance of nonoverlapping batched means, overlapping batched means, and standardized time series estimators.

1. INTRODUCTION

A primary concern in simulation output analysis is confidence interval estimation for the mean μ of a stationary stochastic process, $X_1,\ldots,X_n$. There are a number of difficulties involved in estimating confidence intervals for μ. One problem arises from the presence of serial correlation in the simulation output; that is, the X_i's are not independent. Serial correlation can result in the violation of the conditions required for proper confidence interval estimation. For example, if the X_i's are positively correlated, then confidence interval estimators (c.i.e.'s) for μ based on the sample mean $\bar{X}_n \equiv \Sigma X_i/n$ and the variance estimator $S^2 = \Sigma(X_i - \bar{X}_n)^2/(n-1)$ may have coverage probability lower than the nominal coverage:

$$\Pr\{ \mu \varepsilon \bar{X}_n \pm t_{n-1,1-½\alpha} S/\sqrt{n} \} < 1 - \alpha ,$$

where $t_{n-1,1-½\alpha}$ is the upper $1-½\alpha$ quantile of a t-distribution with n-1 degrees of freedom.

Another difficulty is related to the fact that the underlying stochastic process may not produce stationary output, especially during the initial portion of the simulation run. For instance, customers from the early portion of a queueing system simulation initialized empty and idle have lower expected waiting times than do later "steady state" customers. In this article, we shall restrict ourselves to the analysis of stationary simulation output.

Over the last two decades, a number of confidence interval estimation methodologies have been proposed and studied: nonoverlapping batched means, independent replications, ARMA time series modeling, spectral representation, and regeneration [cf. Fishman (1978) and Bratley, Fox, and Schrage (1983) for reviews of these five techniques.] Two recently developed methodologies are standardized time series [Schruben (1983)] and overlapping batched means [Meketon and Schmeiser (1984)].

The main difference among the above methodologies concerns the estimation of $\mathrm{Var}(\bar{X}_n)$. This paper studies a number of variance estimators to better understand their large and small sample behavior in confidence interval procedures. In the next section, we introduce several estimators of interest. Section 3 discusses large and small sample properties of these variance estimators. Section 4 gives a summary and conclusions.

2. SEVERAL VARIANCE ESTIMATORS

In this section, we will review some variance estimators and provide the necessary notation.

2.1 Nonoverlapping Batched Means (NOBM)

Suppose that we divide the stationary process $X_1,\ldots,X_n$ into $b>1$ contiguous, nonoverlapping batches each of which has m observations ($n = bm$). Batch i consists of $X_{(i-1)m+1}$, $X_{(i-1)m+2}$, ..., X_{im}, $i=1,\ldots,b$. Define

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

as the i-th *batched mean*, $i=1,\ldots,b$.

The method of NOBM assumes that the batched means are approximately i.i.d. normal random variables with unknown mean μ and variance σ^2/m. Then a $100(1-\alpha)\%$ c.i.e. for μ is given by

$$\Pr\{\ \mu \in \bar{\bar{X}}_n \pm t_{b-1,1-\frac{1}{2}\alpha}(\hat{V}_N/n)^{\frac{1}{2}}\ \} \doteq 1-\alpha \quad (1)$$

where

$$\hat{V}_N \equiv m\sum_{j=1}^{b}[\ \bar{X}_{j,m} - \bar{\bar{X}}_n\ ]^2/(b-1)$$

is the NOBM estimator for $\sigma^2 = \lim_{n\to\infty} n\mathrm{Var}(\bar{\bar{X}}_n)$.

2.2 Overlapping Batched Means (OBM)

The method of OBM has recently been developed for use in simulations by Meketon (1980) and Meketon and Schmeiser (1984). We define the i-th *overlapping batched mean*, $i=1,\ldots,n-m+1$, as

$$\bar{X}(i,m) \equiv \frac{1}{m}\sum_{j=0}^{m-1} X_{i+j}\ .$$

The OBM estimator for σ^2 is given by

$$\hat{V}_O \equiv m\sum_{i=1}^{n-m+1}[\bar{X}(i,m) - \bar{\bar{X}}_n]^2/[(n-m+1)(1-\tfrac{m}{n})]\ .$$

Based on results from Meketon (1980), Meketon and Schmeiser propose that the degrees of freedom from $\hat{V}_O$ be equal to 1.5 times the degrees of freedom from the NOBM estimator $\hat{V}_N$; however, Monte Carlo results show that when n/m is small, the proposed degrees of freedom may be too small. Schmeiser (1986) has suggested a slightly revised formula for the OBM degrees of freedom.

2.3 Standardized Time Series (STS)

STS uses a functional central limit theorem to transform $X_1,\ldots,X_n$ into a process which is asymptotically distributed as a Brownian bridge. Properties of Brownian bridges are then used to estimate confidence intervals for μ.

Suppose that the stationary process $X_1,\ldots,X_n$ [satisfying other mild assumptions from Schruben (1983)] is divided into b batches of size m. For $i=1,\ldots,b$ and $j=1,\ldots,m$, denote the j-th cumulative mean from batch i as:

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

For $0\le t\le 1$, and all i and j, let

$$S_{i,j} \equiv \bar{X}_{i,m} - \bar{X}_{i,j} \quad \text{and}$$

$$T_{i,m}(t) \equiv \frac{\|mt\|S_{i,\|mt\|}}{\sigma\sqrt{m}}$$

where $\|\cdot\|$ is the "floor" function and $T_{i,m}(t)$ is the *standardized time series* from the i-th batch. Schruben shows that as $m\to\infty$, $T_{i,m}(t)$ converges in distribution to a standard Brownian bridge (i.e., conditioned Brownian motion).

Finally, define for all batches,

$$\hat{A}_i \equiv \sum_{j=1}^{m} jS_{i,j},$$

$$\hat{K}_i \equiv \underset{j}{\operatorname{argmax}}\ \{jS_{i,j}\}, \quad \text{and}$$

$$\hat{S}_i \equiv \hat{K}_i S_{i,\hat{K}_i}.$$

The following are estimators for σ^2:

Area estimator:

$$\hat{V}_A \equiv \frac{12}{(m^3-m)b}\sum_{i=1}^{b}\hat{A}_i^2 \approx \frac{\sigma^2\chi^2(b)}{b}$$

Maximum estimator:

$$\hat{V}_M \equiv \sum_{i=1}^{b}\frac{m\hat{S}_i^2}{3b\hat{K}_i(m-\hat{K}_i)} \approx \frac{\sigma^2\chi^2(3b)}{3b}$$

(where "$\stackrel{\cdot}{\sim}$" reads "is approximately distributed as") Schruben shows that $\hat{V}_A$ and $\hat{V}_M$ are asymptotically independent of $\hat{V}_N$. Hence, additional estimators for σ^2 are:

Combined NOBM-Area estimator:

$$\hat{V}_{NA} \equiv [(b-1)\hat{V}_N + b\hat{V}_A]/(2b-1) \stackrel{\cdot}{\sim} \frac{\sigma^2\chi^2(2b-1)}{2b-1}$$

Combined NOBM-Maximum estimator:

$$\hat{V}_{NM} \equiv [(b-1)\hat{V}_N + 3b\hat{V}_M]/(4b-1) \stackrel{\cdot}{\sim} \frac{\sigma^2\chi^2(4b-1)}{4b-1}$$

To construct confidence intervals based on the above estimators, we substitute the appropriate degrees of freedom and variance expressions into equation (1).

3. PROPERTIES OF THE VARIANCE ESTIMATORS

We survey results concerning the bias and variance of the NOBM, OBM, and STS variance estimators, and the performance of each estimator in confidence interval estimation procedures for μ.

3.1 Bias of the Estimators

The bias of an estimator is the difference between the expected value of the estimator and the parameter to be estimated. If we denote a generic variance estimator by $\hat{V}$, then

$$\text{Bias} = \sigma^2 - E[\hat{V}].$$

All of the variance estimators under consideration here are asymptotically unbiased as the batch size m becomes large.

We first state results which are analogous to the asymptotic continuous time results from Goldsman and Meketon (1986) (G-M): For large m and b,

$$\text{Bias}(\hat{V}_N) \doteq \frac{c}{m} - \frac{\sigma^2}{b},$$

$$\text{Bias}(\hat{V}_O) \doteq \frac{c}{m} - \frac{\sigma^2}{b},$$

$$\text{Bias}(\hat{V}_A) \doteq \frac{3c}{m}, \text{ and}$$

$$\text{Bias}(\hat{V}_{NA}) \doteq \frac{2c}{m} - \frac{\sigma^2}{2b},$$

where c is a constant and σ^2 is defined in Section 2.1. (These bias results do not include small order terms.)

We see that, for large m and b, the NOBM and OBM methods yield approximately the same bias. Since the bias expression for the area estimator does not include the σ^2/b term, the area estimator is not directly comparable to the NOBM and OBM estimators; however, if b is very large or if σ^2 is small, then we would expect the area estimator's bias to be approximately three times that of the NOBM and OBM estimators. Analogous results are not available for the maximum and combined NOBM-maximum estimators.

It is possible to calculate the exact bias of the variance estimators for specific stochastic processes. As an example, we compare the exact bias of the NOBM and area estimators for the familiar AR(1) process: $X_i = \varphi X_{i-1} + \varepsilon_i$, $i = 1,2,\ldots$, where the ε_i's are i.i.d. $\text{Nor}(0,1-\varphi^2)$ and $-1 < \varphi < 1$.

Result 1: $$E[\hat{V}_N] = \frac{1+\varphi}{1-\varphi} - \frac{2\varphi}{(1-\varphi)^2 m} + \frac{2\varphi^{m+1}}{(1-\varphi)^2 m}.$$

Result 2: $$E[\hat{V}_A] = \frac{1+\varphi}{1-\varphi} + \frac{12\varphi\left[\frac{-m^2+1-\varphi^m(m+1)^2}{2} + \frac{2\varphi[1-(m+1-m\varphi)\varphi^m]}{(1-\varphi)^2}\right]}{(m^3-m)(1-\varphi)^2}$$

$$= \frac{1+\varphi}{1-\varphi} - \frac{6\varphi}{(1-\varphi)^2 m} + \text{small order terms}.$$

Results 1 and 2 follow from (A.5-7) and (A.5-15), respectively, of Goldsman (1984).

Since $\sigma^2 = (1+\varphi)/(1-\varphi)$ for the AR(1) process, the bias of the area estimator is asymptotically three times that of the NOBM (as expected). It is also possible to show that the bias of the OBM estimator is approximately the same as that of the NOBM estimator. It is clear that the bias decreases to 0 as the batch size $m \to \infty$.

The following remarks illustrate why the bias of a variance estimator $\hat{V}$ is an important consideration in confidence interval estimation. Let $T \equiv (\bar{\bar{X}}_n-\mu)/(\hat{V}/n)^{1/2}$. Then T is asymptotically (as $b \to \infty$) normally distributed with mean 0 and variance $\sigma^2/E[\hat{V}]$ [instead of Nor(0,1)]. If $E[\hat{V}] < \sigma^2$ (i.e., if bias > 0), then making the false assumption that T is Nor(0,1) will result in smaller than desired coverage when confidence intervals for μ are calculated; this coverage asymptotically converges to

$$2\Phi\left\{ z_{1-\frac{1}{2}\alpha} \left(E[\hat{V}]/\sigma^2\right)^{1/2} \right\} - 1 , \qquad (2)$$

where $\Phi(\cdot)$ is the standard Normal c.d.f. and $z_{1-\frac{1}{2}\alpha}$ is the upper $1-\frac{1}{2}\alpha$ quantile of the Nor(0,1) [see e.g., Schruben (1980)]. Table 1 uses equation (2) and Results 1 and 2 to find the limiting coverages for the AR(1) model.

Table 1 - Limiting Coverages (as $b\to\infty$) for the Mean of the AR(1) Process $X_i = 0.9X_{i-1} + \varepsilon_i$ $(1-\alpha = 0.90)$

m	NOBM	AREA
2	0.397	0.095
4	0.521	0.166
8	0.652	0.293
16	0.763	0.476
32	0.836	0.667
64	0.871	0.795
128	0.887	0.854
256	0.894	0.879
512	0.897	0.890
1024	0.898	0.895
∞	0.900	0.900

For this AR(1) case, the NOBM coverage reaches the nominal value $1-\alpha = 0.9$ more quickly than the area estimator.

3.2 Variance of the Estimators

G-M report that as m and b become large,

$$\mathrm{Var}(\hat{V}_N) \doteq 2\sigma^4/b ,$$

$$\mathrm{Var}(\hat{V}_O) \doteq \tfrac{4}{3}\sigma^4/b ,$$

$$\mathrm{Var}(\hat{V}_A) \doteq 2\sigma^4/b ,$$

$$\mathrm{Var}(\hat{V}_{NA}) \doteq \sigma^4/b ,$$

$$\mathrm{Var}(\hat{V}_M) \doteq \tfrac{2}{3}\sigma^4/b , \text{ and}$$

$$\mathrm{Var}(\hat{V}_{NM}) \doteq \tfrac{1}{2}\sigma^4/b .$$

Note that the NOBM and area estimators have approximately the same asymptotic variance. G-M also calculate the mean squared error of the variance estimators.

3.3 Large Sample Confidence Interval Estimation

The purpose of a confidence interval estimator (c.i.e.) is to indicate the accuracy and the variability of a point estimator of μ. To evaluate the performance of a c.i.e., the following measures are commonly used: (a) the achieved coverage of μ, (b) the expected half-length, E[H], of the confidence interval, and (c) the variance of the half-length, Var(H), where we denote the random variable corresponding to the half-length of a generic c.i.e. by H. Among c.i.e.'s which achieve the desired coverage ($1-\alpha$, say), that c.i.e. which yields the smallest E[H] is preferred. It is also preferred to have small Var(H), this meaning that the c.i.e. is "stable."

We first survey asymptotic properties of the c.i.e.'s. As m and b become large, all of the c.i.e.'s under study achieve nominal coverage. Following Schmeiser (1982) and Goldsman and Schruben (1984), we can derive:

$$H \approx \frac{\sigma t_{d,1-\frac{1}{2}\alpha}}{(nd)^{1/2}} \chi(d) ,$$

$$E[H] \doteq \frac{\sigma}{\sqrt{n}} t_{d,1-\frac{1}{2}\alpha}(2/d)^{1/2} \frac{\Gamma(\frac{1}{2}(d+1))}{\Gamma(\frac{1}{2}d)} , \text{ and}$$

$$\mathrm{Var}(H) \doteq \frac{\sigma^2}{n} t^2_{d,1-\frac{1}{2}\alpha} \left\{ 1 - \frac{2}{d} \left[\frac{\Gamma(\frac{1}{2}(d+1))}{\Gamma(\frac{1}{2}d)} \right]^2 \right\},$$

where $\chi(d)$ denotes the chi distribution with d degrees of freedom (d being the appropriate degrees of freedom associated with a particular variance estimator) and $\Gamma(\cdot)$ is the gamma function.

The above papers show that as m becomes large, the STS area, combined NOBM-area, maximum, and combined NOBM-maximum estimators all yield smaller E[H] and Var(H) than the NOBM and OBM methods; as b becomes large, all of the estimators have approximately the same E[H].

3.4 Small Sample Confidence Interval Estimation

Small sample analysis of c.i.e.'s is difficult, and we must resort to Monte Carlo methods. In one such study, we generated observations from a steady state AR(1) process with $\varphi = 0.9$. [The steady state process can be directly generated by taking the initial observation of each run from a Nor(0,1) distribution.] We considered batch sizes $m = 2^k$ (k=0,1,..,10) and numbers of batches $b = n/m = 2$ and 16. A confidence interval was calculated for each combination of m and n/m for the different estimators (NOBM, OBM, area, and combined NOBM-area); we shall not report here on the maximum and combined NOBM-maximum estimators since it turned out that they performed poorly in this small sample environment. We replicated 1000 independent runs, and Figures 1 and 2 show the achieved coverages (the proportion of the 1000 confidence intervals which contain μ) and average half-lengths (EHL) produced by the c.i.e.'s of the study.

We first discuss Figure 1, where the number of batches is "large" (b = 16). The behavior of the NOBM and OBM coverages is about the same when b = 16, mainly because the NOBM and OBM variance estimators have the same asymptotic bias; see Figure 1(a). Since the bias of the area estimator is asymptotically three times that of NOBM, the area estimator's achieved coverage approaches the desired nominal value more slowly than the NOBM's. The EHL of the NOBM c.i.e. is the largest; the EHL of the OBM's c.i.e. is nearly the same. However, as $m \to \infty$, all four estimators have the same EHL, as explained in the previous subsection; see Figure 1(b).

We now refer to Figure 2, in which case b is "small" (b = 2). Here, the OBM method yields approximately the same coverage and EHL as the area and combined NOBM-area c.i.e.'s. NOBM approaches the desired coverage fastest. But as m becomes large, the OBM and STS estimators outperform the NOBM c.i.e., since the OBM and STS estimators achieve the desired coverage while simultaneously obtaining smaller EHL; see Figures 2(a) and 2(b).

4. SUMMARY

This paper investigated the behavior of a number of estimators for $\sigma^2 = \lim_{n\to\infty} n\mathrm{Var}(\bar{X}_n)$. In particular, we studied the popular batched means estimator as well as estimators arising from overlapping batched means and standardized time series. We reported asymptotic and small sample results pertaining to the bias and variance of the estimators, and to their performance in confidence interval estimation.

In order to form a conclusive picture of the small sample c.i.e. performance characteristics of the NOBM, OBM, and STS c.i.e.'s, the authors are currently conducting a large-scale Monte Carlo study involving a number of stochastic processes. Our tentative results from these experiments show that the NOBM estimator reaches the desired coverage more quickly than the other estimators. Yet OBM and STS perform better when sufficient observations are available.

ACKNOWLEDGEMENTS:

We thank Mr. M.J. Rao for his computer assistance. The first author was partially supported by the Mathematical Sciences Institute at Cornell University.

REFERENCES

Bratley, P., B.L. Fox, and L.E. Schrage (1983). *A Guide to Simulation*. Springer-Verlag, New York.

Fishman, G.S. (1978). *Principles of Discrete Event Simulation*. John Wiley & Sons, New York.

Goldsman, D. (1984). *On Using Standardized Time Series to Analyze Stochastic Processes*. Ph.D. Dissertation, School of O.R. & I.E., Cornell Univ., Ithaca, NY.

Goldsman, D. and M.S. Meketon (1986). "A Comparison of Several Variance Esimators." School of ISyE Tech. Report J-85-12, Georgia Institute of Technology, Atlanta.

Goldsman, D. and L. Schruben (1984). "Asymptotic Properties of Some Confidence Interval Estimators for Simulation Output." *Management Science*, 30, 1217-1225.

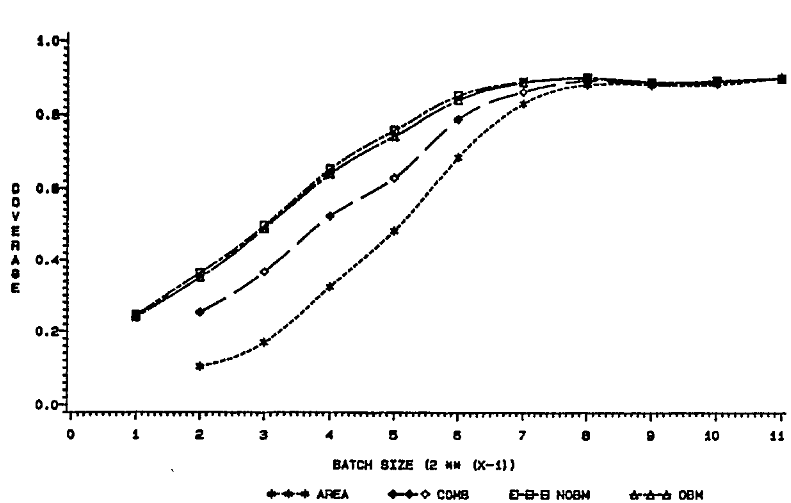

Figure 1(a). Coverage Plot for AR(1)
(No of Batches = 16)

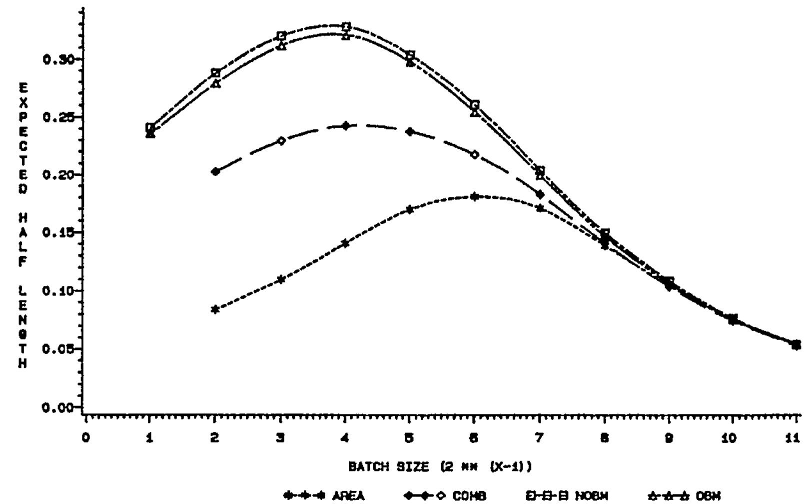

Figure 1(b). EHL Plot for AR(1)
(No of Batches = 16)

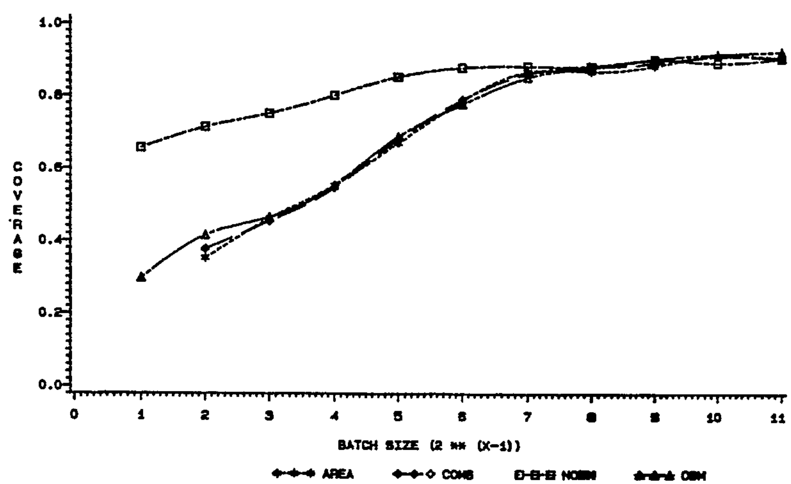

Figure 2(a). Coverage Plot for AR(1)
(No of Batches = 2)

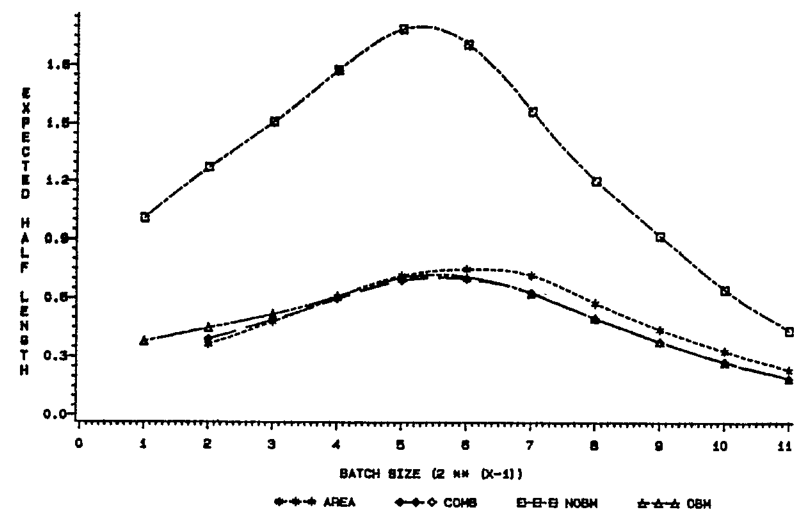

Figure 2(b). EHL Plot for AR(1)
(No of Batches = 2)

Meketon, M.S. (1980). *The Variance Time Curve: Theory, Estimation and Application.* Ph.D. Dissertation, School of O.R. & I.E., Cornell Univ., Ithaca, NY.

Meketon, M.S. and B.W. Schmeiser (1984). "Overlapping Batch Means: Something for Nothing?" *Proc. Winter Simul. Conf.*, 227-230.

Schmeiser, B.W. (1982). "Batch Size Effects in the Analysis of Simulation Output." *Operations Research*, 30, 556-568.

Schmeiser, B.W. (1986). Personal Communication.

Schruben, L. (1980). "A Coverage Function for Interval Estimators of Simulation Response." *Management Science*, 26, 18-27.

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

AUTHORS' BIOGRAPHIES

DAVE GOLDSMAN is an Assistant Professor of Industrial and Systems Engineering at Georgia Tech. He holds a Ph.D. from Cornell, where he spent last summer as a Visiting Scientist. His research interests include simulation output analysis, and ranking and selection.

Dave Goldsman
School of ISyE
Georgia Tech
Atlanta, GA 30332
(404) 894-2365

KEEBOM KANG is an Assistant Professor of Industrial Engineering at the University of Miami. He received his Ph.D. from the School of Industrial Engineering, Purdue University. During the summer of 1986, he visited Syracuse University to conduct research in simulation modeling and analysis.

Keebom Kang
Department of Industrial Engineering
University of Miami
Coral Gables, FL 33124
(305) 284-2370

ROBERT G. SARGENT is a Professor of Industrial Engineering and Operations Research and a member of the Computer and Information Sciences Faculty at Syracuse University. Dr. Sargent has served the Winter Simulation Conferences in several capacities, including being a member of the Board of Directors for ten years, Board Chairman for two years, General Chairman of the 1977 Conference, and Co-editor of the 1976 and 1977 Conferences. Professor Sargent was Department Editor of Simulation Modeling and Statistical Computing for the *Communications of the ACM* for five years, has served as Chairman of the TIMS College on Simulation and Gaming, and has received service awards from ACM, IIE, and the Winter Simulation Conference Board of Directors. He currently is an ACM National Lecturer, a member of the Executive Committee of the IEEE Computer Society Technical Committee on Simulation, and a Director-at-large of the Society for Computer Simulation. Dr. Sargent received his education at the University of Michigan. His current research interests include model validation, simulation methodology, simulation application performance evaluation, and applied operations research. Professor Sargent is a member of AAAM, the New York Academy of Sciences, Sigma Xi, ACM, IIE, ORSA, SCS, and TIMS, and is listed in Who's Who in America.

Robert G. Sargent
Department of IE & OR
Syracuse University
Syracuse, New York 13244-1240
(315)423-4348