

ADVANCED OUTPUT ANALYSIS FOR SIMULATION

Andrew F. Seila

Department of Management Sciences and Information Technology
University of Georgia
Athens, GA 30602, U.S.A.

ABSTRACT

The purpose of this paper is to review some advanced aspects of methods for analyzing data produced by simulations. The review focuses on methods for estimating parameters of stationary output processes. The techniques include some variations of the batch means method, sequential methods, standardized time series estimators, methods based upon Hoeffding's inequality, quantile estimation, and multivariate estimation methods.

1 INTRODUCTION

The purpose of most simulations is to develop an understanding of system behavior, with the goal of using that understanding to make decisions involving the system. The process of developing and running the simulation involves some combination of the following activities: preliminary analysis of system operations; model design and coding; verification; validation; experimental design; performing simulation runs to produce output data; and statistical analysis of output data to estimate parameters. Each of these activities contributes to understanding of system behavior, but often the bottom line is knowing the values of one or more system parameters that measure performance. This paper focuses on the last step: advanced methods for analysis of output data to estimate one or more parameters that describe system performance.

There are many statistical topics that are related to output analysis: variance reduction, experimental design, comparison methods, optimization and response surface methodology, and initialization techniques. While these are important topics and are intimately related to

output analysis, they will not be discussed in this tutorial. Most of these topics are discussed in other papers in this volume and a number of references at the end of this paper relate to these topics.

Simulations can be classified as terminating or non-terminating (Law 1980, 1983). Terminating simulations run only until some stopping criterion is met and data from terminating simulations have the characteristic that they are dependent upon the initial state of the system. Normally, analysis of such data involves running independent replications and applying standard statistical techniques to compute a confidence interval for the parameter of interest (Law 1980, 1983; Seila 1991). This paper does not deal with methods for analyzing data from terminating simulations.

Non-terminating simulations can conceptually run indefinitely. We will assume that the system is stationary, and therefore the output process is stationary. In brief, an output process is stationary if any finite collection of observations has the same distribution as the same collection shifted by an arbitrary amount.

A weaker form of stationarity is wide-sense or covariance stationarity. A process is said to be wide-sense stationary if $E(X_t) = E(X_{t+h})$, $Var(X_t) = Var(X_{t+h})$ and $Cov(X_t, X_{t+h})$ depends only upon h for all t and all h . The analysis methods considered in this paper apply to strictly stationary and covariance stationary processes.

This tutorial presents some advanced topics in the analysis of output data when the output process is stationary. Some of it is a republication of the output analysis tutorial presented at the 1991 Winter Simulation Conference (Seila 1991). A number of excellent

texts are available to provide the specific algorithms and calculations for many of the methods discussed (Bratley, Fox and Schrage 1983; Fishman 1987b; Kleijnen 1974, 1975, 1982 Law and Kelton 1991; Lewis and Orav 1989; Rubinstein 1981; Welch 1983). Sections 2 through 6 involve methods for estimating the mean of the output process. Section 2 presents a general discussion of issues concerned with estimating the stationary mean. Topics relating to the batch means method are presented in Section 3. Section 4 discusses methods for sequential estimation. Sections 5 and 6 present recently developed techniques using standardized time series and Hoeffding's inequality. Finally, quantile estimation and multivariate methods for estimating multiple parameters simultaneously are the topics of Sections 7 and 8, respectively.

2 GENERAL CONSIDERATIONS IN ESTIMATING THE MEAN

Since any simulation run must be started in a specific initial state, the data produced will depend upon the initial conditions. However, as contrasted to terminating simulations, steady-state simulations eventually produce data that does not depend upon the initial state of the system (or at least, the dependence is sufficiently weak that it can be ignored). Performance measures of interest are defined in terms of the steady-state behavior of the system. A great deal of effort has gone into developing estimators for the mean of a stationary output process. For example, one might be interested in estimating the steady-state mean waiting time for a customer in a queue or item in a production system. The stationary mean can be conceptualized in two ways. Let X_1, X_2, \dots, X_n be the waiting times observed from n customers while the system is operating in steady-state. The mean waiting time is

$$\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i.$$

This is the average waiting time for a very large number of customers. Another way to interpret the steady-state mean waiting time is to consider an arbitrary customer who enters the system without knowledge of the system state. His waiting time is a random variable that has some (unknown) distribution. The stationary mean waiting time is the expected value of this random variable.

Suppose that observations are binary random variables, having the value 1 if a specified event occurs and 0 otherwise. Then, the mean of these observations is the probability that the event occurs. For example, let $X_i = 1$ if the i^{th} customer in a queueing system waits more than 2 minutes, and $X_i = 0$ otherwise. Then, the expected value of X_i is the probability that a randomly selected customer must wait more than 2 minutes, or the proportion of customers in a very long run who must wait more than 2 minutes. Since proportions, or probabilities, are means of appropriately defined binary observations, methods for estimating means can also be applied to estimate probabilities.

The point estimator for the mean is the overall sample mean:

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

Let $\sigma_{\bar{X}}^2$ denote the variance of the sample mean \bar{X}_n and suppose that $s_{\bar{X}}^2$ is an estimator for $\sigma_{\bar{X}}^2$. Then the $100(1 - \alpha)$ -percent confidence interval for μ is $\bar{X}_n \pm t_{d, 1 - \alpha/2} s_{\bar{X}}$, where $t_{d, \alpha}$ is the 100α percentage point for the Student's t -distribution with d degrees of freedom. Various methods for computing the confidence interval differ primarily in the estimator $s_{\bar{X}}$ and the number of degrees of freedom, d .

When analyzing data generated by a stationary simulation, one must deal with two problems: (1) the presence of initial data that constitutes a transient portion of the output process (the startup problem), and (2) autocorrelation among the stationary observations. The method for dealing with the initial transient portion is normally to delete an initial segment of observations from the data prior to applying estimation methods. The problem one faces, however, is to determine where one should truncate the data to remove the initial transient portion. This is a very difficult problem for which several solution methods have been proposed, but also one for which none has proved uniformly satisfactory (Gafarian, Ancker and Morisaku 1978; Fishman 1972; Kelton 1987; Schruben 1982; Schruben, Singh and Tierney 1983; Kelton and Law 1983; Welch 1981; Wilson and Pritsker 1978a, b). Other research has been concerned with how the initial state of the system should be chosen so that the transient portion will be minimized (Kelton 1989 and references therein).

The effect of having autocorrelation among

the data is to increase the difficulty of estimating the variance of the sample mean. Frequently, the observations are positively autocorrelated. The effect of positive autocorrelation is to cause the usual sample standard error to be a biased estimator of the standard error of the mean. If σ^2 is the variance of each X_i and ρ_j is the autocorrelation between X_i and X_{i+j} , i.e., $\rho_j = \text{Corr}(X_i, X_{i+j})$, the variance of \bar{X} is given approximately by:

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \left(1 + 2 \sum_{j=1}^{\infty} \rho_j \right)$$

If X_1, X_2, \dots are independent, so that $\rho_j = 0$ for $j \neq 0$, then the sample variance of the mean s^2/n is an unbiased estimator for $\sigma_{\bar{X}}^2$. However, if autocorrelation is present, the term $\sum_{j=1}^{\infty} \rho_j$ introduces a bias to the sample variance of the mean s^2/n , as an estimator $\sigma_{\bar{X}}^2$. A confidence interval computed without accounting for autocorrelation will be too short and have a true confidence coefficient which is smaller than the nominal value used to compute the interval. That is, a confidence interval which is computed to have a confidence coefficient of 95-percent may have a true confidence coefficient much smaller than .95.

3 THE BATCH MEANS METHOD

In the batch means method (Conway 1963; Mechanic and McKay 1966; Law 1977; Fishman 1978a; Schriber and Andrews 1979; Schmeiser 1982), data from the stationary portion of the run are grouped into batches and the sample mean is computed for each batch. Under some mild technical conditions, it is known that if the batch size is large enough, the batch means will be approximately uncorrelated (Law and Carson 1979). If this is the case, the batch means can be treated as a sequence of independent, identically distributed observations of the sample mean and the usual method can be used to compute a confidence interval for the mean.

The difficulty with the batch means method is that the user must decide how large the batches should be, and consequently, if the number of observations in the run is fixed, the number of batches to use. If the batch size is too small, the batch means will not be uncorrelated and the

confidence interval will often be too narrow and have an erroneously low coverage probability. If the batch size is too large, the confidence interval will be unnecessarily large because the number of batches will be too few. See Schmeiser (1982) for a discussion of batch size effects. A number of methods to determine the batch size have been proposed (Fishman 1978a, b; Law and Carson 1979).

Fishman's method is simple but has proven to work successfully. In this method, one starts with a sample size which is a power of 2 and batches of size 1. Then, the method iterates through a procedure which performs an hypothesis test for lack of autocorrelation between adjacent batch means and successively doubles the batch size and halves the number of batches if significant autocorrelation is found. The procedure stops when either the autocorrelation between adjacent batch means is not significantly different from zero, or the number of batches is less than 8. In the former case, the final batch size is used; in the latter, the entire sample is declared too small and additional observations must be generated. Schmeiser (1982) has noted that if the autocorrelation function is positive and decreasing, then one should almost never need more than approximately 30 batches, so in this case Fishman's method can be modified to start with 32 batches, say, instead of n batches. As a by-product of Fishman's method, one can plot the p-value of the hypothesis test for zero autocorrelation between adjacent batch means against the log of the batch size. In normal circumstances, this is an increasing function. Therefore, an appropriate batch size can be selected from this plot by noting the first batch size where the p-value exceeds the desired significance level.

The advantage of the batch means method is that it seems to use the data more efficiently than many alternatives. By deleting the initial transient only once, the amount of data that is discarded in the run is minimized. If the length of the initial transient is misjudged too large, this error is not multiplied by the number of replications as in the independent replications method, and if it is misjudged too small, this error may cause the first batch mean to be biased but is not likely to affect other batch means. A study comparing several methods (Law and Kelton 1984) has shown that the batch means method is competitive with all of the other methods that have been proposed and is generally superior to them in terms of producing the smallest, most accurate confidence intervals.

Meketon and Schmeiser (1984) have suggested that the batch means method could be improved by forming batches that overlap. In this scheme, which they called overlapping batch means, the first batch consists of observations 1 through m , where m is the batch size. Then, the second batch consists of observations 2 through $m+1$, the third batch has observations 3 through $m+2$, and so on. The estimator of $\sigma_{\bar{X}}^2$ is the usual sample variance of the batch means, except that in this case there are $n-m+1$ batches and their sample means are clearly not uncorrelated. Meketon and Schmeiser showed that, for large batches, overlapping batch means produce a more stable estimator of $\sigma_{\bar{X}}^2$ because the variance of $s_{\bar{X}}^2$ for the batch means formed from overlapping batch means is one third smaller than the variance of $s_{\bar{X}}^2$ formed using the usual non-overlapping batch means. Welch (1987) later showed that both traditional batch means and overlapping batch means are specific cases of spectral estimation at frequency 0, and moreover, that most of the variance reduction from using overlapping batch means can be obtained by forming three or four sub-batches from each batch and using these sub-batches to compute overlapping batch means. For example, if a run consists of 40 batches of 32 observations each for a total of 1280 observations, one can divide each batch into 4 sub-batches of 8 observations each. Then, the first (overlapping) batch consists of observations 1 through 32; the second, of observations 9 through 40; the third, of observations 17 through 48, and so on. The batch mean is computed for these batches, and the sequence of batch means is then used to compute a variance estimate for use in the confidence interval calculation. This procedure will result in a reduction of approximately 32 percent in the variance of the variance estimate.

4 SEQUENTIAL METHODS

It is frequently the case that the analyst wishes to estimate the parameter with a specified precision. For example, in a manufacturing simulation, one might wish to estimate the mean processing time for a product with precision plus-or-minus 1 hour, or estimate the mean utilization of a particular machine with precision plus-or-minus .05. In this case, the desired precision determines the amount of data needed to compute the confidence interval. However, some measure of variance is required in order to compute the appropriate sample size, and this quantity is

unknown. Sequential methods (Kabak 1968; Fishman 1977; Law and Kelton 1982) solve this problem by sequentially collecting data and testing to determine if enough observations have been collected to assure the desired precision of the confidence interval. Many of the sequential methods are based upon the results of Chow and Robbins (1965) and Nadas (1969). The advantage of sequential methods is that, if the technical assumptions of the method apply, the precision of the confidence interval is guaranteed; however, the number of observations generated, and therefore, the amount of computer time required for the simulation is not predictable in advance. This opens the possibility that the simulation run could require more time than is available. A second difficulty with sequential methods is that they must be built into the simulation; i.e., one cannot just run the simulation, store the output data on a file, and analyze the data separately from the simulation run. Instead, the test for the end of the run must be made as the simulation is running. If these two considerations do not pose a problem, sequential methods are a preferable approach to computing confidence intervals.

5 STANDARDIZED TIME SERIES

Another estimator, called an *area estimator*, is based upon the theory of standardized time series and has been proposed as an alternative to these techniques for estimating the standard error of the mean (Schruben 1983). This estimator assumes that the process has the ϕ -mixing property, which, informally, states that if the process runs for a sufficiently long time, observations in the distant past are approximately independent of those in the present. This is a property which is easy to assume but for many models is difficult to prove. If the process has the ϕ -mixing property, a suitably standardized version of the sample mean can be shown to converge to a Brownian Bridge, and this property is used to develop an estimator for the variance of the sample mean based upon the area under the standardized sample mean. The area estimator has been shown to be less efficient in many cases than the batch means and spectral estimators. A recent paper (Glynn and Iglehart 1990) shows the relationships between batch means and area estimators, and compares their efficiencies for large sample sizes. However, recent work on weighted versions of the area estimator appears promising (Goldsmann, Meketon and Schruben 1990).

6 SPECIAL METHODS FOR PROPORTIONS

Empirical studies (Law 1983; Law and Kelton 1982, 1984) have shown that the actual confidence coefficient for many of the methods is somewhat less than the theoretical value. Thus, in practice confidence intervals are not as reliable as one would like. This is a result of the fact that the methods for computing confidence intervals are based upon the asymptotic distribution of the sample statistics. Thus, the confidence coefficient is accurate only in the limit as the sample size approaches infinity. Recently, Hoeffding's inequality, which applies to the sample mean for observations that are bounded, has been used to develop conservative confidence intervals for proportions (or probabilities) (Fishman 1986). These confidence intervals, which apply to independent observations, require that an expression involving the sample proportion be solved numerically. The widths of the confidence intervals are approximately 30-percent larger than those using asymptotic normal theory, but they have the desirable property that the confidence coefficient is guaranteed to be at least as large as the nominal value used to compute the confidence interval.

7 METHODS FOR ESTIMATING QUANTILES

A p -quantile or $100p^{\text{th}}$ percentile is a value such that a specified proportion, p , of the observations are less than the quantile. For example, the .9-quantile of waiting time for customers in a stationary queueing system is a value, θ , such that the probability that an arbitrary customer must wait less than θ is .90. Quantiles are useful parameters if the objective of the simulation study is capacity planning. Unfortunately, however, quantiles, especially extreme quantiles (p close to 0 or 1), are much more difficult to estimate than means.

Several methods have been proposed for estimating quantiles when the data is generated by a stationary simulation (Iglehart 1976; Moore 1980; Seila 1982a, b; Heidelberger and Lewis 1984). Most of these methods use the sample quantile as the point estimate. Heidelberger and Lewis suggest using the maximum transform, in which the sample is summarized by being grouped into groups of k observations, and the maxima of the k observations in each group selected to represent the sample. Then, the r -quantile is estimated, where

$r = p^k$. When the observations are mutually independent, theory is available to justify the maximum transformation; however, if the observations are dependent, there is no theoretical justification. The sample p -quantile, P_n , is computed by selecting the observation X_l such that 100 p percent of the observations are less than X_l and 100(1- p)-percent are greater. The methods differ in the way that the variance of X_l is computed.

8 MULTIVARIATE ESTIMATION

Frequently, one wishes to use the same simulation run to estimate two or more parameters simultaneously. For example, in a manufacturing system, one may wish to estimate the mean processing time for items along with the utilization of a particular critical machine. Normally, the mean processing time, which one desires to minimize, increases with machine utilization, which one desires to maximize. Thus, there is a trade-off between these two parameters.

Some special techniques (Seila 1984, Charnes 1991) have been developed for multivariate estimation in certain special cases. More generally, however, Bonferroni's inequality can be used to compute a conservative confidence coefficient for a set of simultaneous confidence intervals. If k confidence intervals are computed with confidence coefficients $1 - \alpha_1, 1 - \alpha_2, \dots, 1 - \alpha_k$, then the probability that all k confidence intervals simultaneously include their parameters is at least $1 - \sum_{i=1}^k \alpha_i$. Therefore, if one wants two confidence intervals to have simultaneous confidence coefficient .95, each can be computed to have an individual confidence coefficient of .975. If each of five confidence intervals has confidence coefficient .98, the simultaneous confidence coefficient for all five is no less than .90. This is a very general technique and has been shown to be rather accurate (Schruben 1981).

The batch means method has been adapted to multivariate output that is synchronized in the sense that the observations on all parameters can be formed into a sequence of vectors (Chen and Seila 1987). This method uses Scheffe confidence intervals and was shown to produce reliable confidence regions. Like Fishman's method, this method chooses the batch size by testing for first-order serial autocorrelation among the (vector) batch means. Charnes found that a good test statistic for this test is the F-approximation to the

Wilks likelihood ratio test; see Charnes (1990). Charnes (1989) has also examined the use of multivariate autoregressive models to estimate the mean. While these approaches have seen some success, this area is the subject of ongoing research.

8 OTHER METHODS

Other methods have also been proposed for estimating the mean of a stationary output process. The spectral method (Duket and Pritsker 1978; Fishman 1978; Heidelberger and Welch 1981a, b) estimates $\sigma^2_{\bar{X}}$ using established methods for estimating the spectrum of a time series. The concepts and calculations involved in applying this method are more complex than those presented so far, but it has been shown to perform well in a number of cases, and software is available to efficiently compute the spectrum. Other time series methods (Schriber and Andrews 1984) attempt to fit the output data to an empirical model and use the estimated parameters of the model to estimate the quantity (2).

For systems where the output process is regenerative (Ross 1989), special methods to estimate the mean and quantiles have been proposed. A regenerative process has the property that the sequence of observations can be grouped into a sequence of independent, identically distributed cycles of observations. It is known for regenerative processes that the mean of the process is equal to the ratio of the mean of the sum of the observations over a cycle divided by the mean number of observations in the cycle. This fact has been used to develop a confidence interval for the mean (Fishman 1973, 1974; Crane and Iglehart 1974, 1975; Crane and Lemoine 1977; Iglehart 1978). It has been observed (Law 1984) that in many cases the batch means method performed as well as or better than the regenerative method.

REFERENCES

- Bratley, P., B. L. Fox and L. E. Schrage. 1983. *A guide to simulation*, Springer-Verlag, New York, NY.
- Charnes, J. M. 1989. Statistical analysis of multivariate discrete-event simulation output. Ph.D. Thesis, Department of Operations and Management Science, University of Minnesota, Minneapolis, Minnesota.
- Charnes, J. M. 1990. Power comparisons for the multivariate batch-means method, In *Proceedings of the 1990 Winter Simulation Conference*, eds. O. Balci, R. P. Sadowski, R. E. Nance, 281-287. Institute of Electrical and Electronics Engineers, San Diego, California.
- Charnes, J. M. 1991. Multivariate simulation output analysis, In *Proceedings of the 1991 Winter Simulation Conference*, eds. B. L. Nelson, W. D. Kelton and G. M. Clark, 187-193. Institute of Electrical and Electronics Engineers, Phoenix, Arizona.
- Chen, D. R. and A. F. Seila. 1987. Multivariate inference in stationary simulation using batch means, In *Proceedings of the 1987 Winter Simulation Conference*, eds. A. Thesen, H. Grant, W. D. Kelton, 302-304. Institute of Electrical and Electronics Engineers, Atlanta, Georgia.
- Chow, Y. S. and H. Robbins. 1965. On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Annals of Mathematical Statistics* **36**: 457-462.
- Conway, R. W. 1963. Some tactical problems in digital simulation. *Management Science* **10**: 47-61.
- Crane, M. A. and D. L. Iglehart. 1974. Simulating stable stochastic systems: I. general multiserver queues. *Journal of the ACM* **21**: 103-113.
- Crane, M. A. and D. L. Iglehart. 1975. Simulating stable stochastic systems: III. regenerative processes and discrete-event simulations, *Operations Research* **23**: 33-45.
- Crane, M. A. and A. J. Lemoine. 1977. *An introduction to the regenerative method for simulation analysis*, Springer-Verlag, New York, NY.
- Duket, S. D. and A. A. B. Pritsker. 1978. Examination of simulation output using spectral methods. *Math. Comput. Simulation* **20**: 53-60.
- Fishman, G. S. 1972. Bias considerations in simulation experiments. *Operations Research* **20**: 785-790.
- Fishman, G. S. 1973. Statistical analysis for queueing simulations. *Management Science* **20**: 363-369.
- Fishman, G. S. 1974. Estimation in multiserver queueing simulations. *Operations Research* **22**: 72-78.
- Fishman, G. S. 1977. Achieving specific accuracy in simulation output analysis. *Communications of the Association for Computing Machinery* **20**: 310-315.
- Fishman, G. S. 1978a. Grouping observations in digital simulation. *Management Science* **24**:

- 510-521.
- Fishman, G. S. 1978b. *Principles of discrete event simulation*, John Wiley, New York, NY.
- Fishman, G. S. 1986. Confidence intervals for the mean and a proportion in the bounded case. Technical Report No. UNC/OR/TR-86/19, Curriculum in Operations Research and Systems Analysis, University of North Carolina, Chapel Hill, NC.
- Gafarian, A. V., C. J. Ancker Jr. and F. Morisaku. 1978. Evaluation of commonly used rules for detecting steady-state" in computer simulation. *Naval Research Logistics Quarterly* **25**: 511-529.
- Glynn, P. W. and D. L. Iglehart. 1990. Simulation analysis using standardized time series. *Mathematics of Operations Research* **15**: 1-16.
- Goldsmann, D., M. Meketon and L. Schruben. 1990. Properties of standardized time series weighted area variance estimators. *Management Science* **36**: 602-612.
- Heidelberger, P. and P. A. W. Lewis. 1984. Quantile estimation in dependent sequences. *Operations Research* **32**: 185-209.
- Heidelberger, P. and P. D. Welch. 1981. A spectral method for confidence interval generation and run length control in simulations. *Communications of the Association for Computing Machinery* **24**: 233-245.
- Heidelberger, P. and P. D. Welch. 1981. Adaptive spectral methods for simulation output analysis. *IBM Journal of Research and Development* **25**: 860-876.
- Iglehart, D. L. 1976. Simulating stable stochastic systems: VI. quantile estimation. *Journal of the Association for Computing Machinery* **23**: 347-360.
- Iglehart, D. L. 1978. The regenerative method for simulation analysis. In *Current Trends in Programming Methodology, Vol. III. Software Engineering*, K. M. Chandy and R. T. Yeh, Eds. Prentice-Hall, Englewood Cliffs, NJ, 52-71.
- Kabak, I. W. 1968. Stopping rules for queuing simulations. *Operations Research* **16**: 431-437.
- Kelton, W. D. 1987. Stochastic simulation: initial transient techniques. *Systems and Control Encyclopedia; Theory, Technology, Applications*, M. G. Singh, Ed. Pergamon Press, Oxford, 4651-4654.
- Kelton, W. D. and A. M. Law. 1983. A new approach for dealing with the startup problem in discrete event simulation. *Naval Research Logistics Quarterly* **30**: 641-658.
- Kelton, W. D. 1989. Random initialization methods in simulation. *IEEE Transactions* **21**: 355-367.
- Kleijnen, J. P. C. 1974. *Statistical techniques in simulation, part I*, Marcel Dekker, New York, NY.
- Kleijnen, J. P. C. 1975. *Statistical techniques in simulation, part II*, Marcel Dekker, New York, NY.
- Kleijnen, J. P. C. 1982. Statistical aspects of simulation: an updated survey. *Statistica Neerlandica* **36**: 165-186.
- Law, A. M. 1977. Confidence intervals in discrete event simulation: a comparison of replication and batch means. *Naval Research Logistics Quarterly* **24**: 667-678.
- Law, A. M. 1980. Statistical analysis of the output data from terminating simulations. *Naval Research Logistics Quarterly* **27**: 131-143.
- Law, A. M. 1983. Statistical analysis of simulation output data. *Operations Research* **31**: 983-1029.
- Law, A. M. and W. D. Kelton. 1991. *Simulation modeling and analysis*, McGraw-Hill, New York, NY.
- Law, A. M. and W. D. Kelton. 1982b. Confidence intervals for steady-state simulations, II: a survey of sequential procedures. *Management Science* **28**: 550-562.
- Law, A. M. and W. D. Kelton. 1984. Confidence intervals for steady-state simulations: I. a survey of fixed sample size procedures. *Operations Research* **32**: 1221-1239.
- Law, A. M. and J. S. Carson. 1979. A sequential procedure for determining the length of a steady-state simulation. *Operations Research* **27**: 1011-1025.
- Lewis, P. A. W. and E. J. Orav. 1989. *Simulation methodology for statisticians, operations analysts, and engineers, volume I*, Wadsworth, Inc., Belmont, CA.
- Mechanic, H. and W. McKay. 1966. Confidence intervals for averages of dependent data in simulations II. Technical Report ASDD 17-202, IBM Corporation, Yorktown Heights, NY.
- Meketon, M. S. and B. Schmeiser. 1984. Overlapping batch means: something for nothing?, In *Proceedings of the 1984 Winter Simulation Conference*, eds. S. Sheppard, U. Pooch and D. Pegden, 227-230. Institute of Electrical and Electronics Engineers, Dallas, Texas.
- Moore, L. W. 1980. Quantile estimation in regenerative processes. Ph.D. Thesis, Curriculum in Operations Research and Systems Analysis, University of North Carolina, Chapel Hill, NC.

- Nadas, A. 1969. An extension of the theorem of Chow and Robbins on sequential confidence intervals for the mean. *Annals of Mathematical Statistics* **40**: 667-671.
- Rubinstein, R. Y. 1981. *Simulation and the monte carlo method*, John Wiley, New York, NY.
- Schmeiser, B. W. and K. Kang. 1981. Properties of batch means from stationary ARMA(1,1) time series. Technical Report 81-3, School of Industrial Engineering, Purdue University, West Lafayette, IN.
- Schmeiser, B. W. 1982. Batch size effects in the analysis of simulation output. *Operations Research* **30**: 556-568.
- Schriber, T. J. and R. W. Andrews. 1979. Interactive analysis of simulation output by the method of batch means. *Proceedings of the 1979 Winter Simulation Conference*, ACM, New York, NY.
- Schriber, T. J. and R. W. Andrews. 1984. ARMA-based confidence intervals for simulation output analysis. *American Journal of Mathematical and Management Sciences* **4**: 345-374.
- Schruben, L. W. 1981. Control of initialization bias in multivariate simulation response. *Communications of the Association for Computing Machinery* **24**: 246-252.
- Schruben, L. W. 1982. Detecting initialization bias in simulation output. *Operations Research* **30**: 569-590.
- Schruben, L. W. 1983. Confidence interval estimation using standardized time series. *Operations Research* **31**: 1090-1108.
- Schruben, L., H. Singh and L. Tierney. 1983. Optimal tests for initialization bias in simulation output, *Operations Research* **31**: 1167-1178.
- Seila, A. F. 1982a. A batching approach to quantile estimation in regenerative simulations. *Management Science* **28**: 573-581.
- Seila, A. F. 1982b. Percentile estimation in discrete event simulation. *Simulation* **39**: 193-200.
- Seila, A. F. 1984. Multivariate simulation output analysis. *American Journal of Mathematical and Management Sciences* **4**: 313-334.
- Seila, A. F. 1991. Output analysis for simulation. In *Proceedings of the 1990 Winter Simulation Conference*, eds. O. Balci, R. P. Sadowski and R. E. Nance, 49-54. The Society for Computer Simulation, International, San Diego, CA.
- Welch, P. D. 1981. On the problem of the initial transient in steady state simulation. Technical Report, IBM Watson Research Center, Yorktown Heights, NY.
- Welch, P. D. 1983. The statistical analysis of simulation results. *The Computer Performance Modeling Handbook*, S. S. Lavenberg, Ed., Academic Press, New York, 268-328.
- Welch, P. D. 1987. On the relationship between batch means, overlapping batch means and spectral estimation, In *Proceedings of the 1987 Winter Simulation Conference*, eds. A. Thesen, H. Grant, W. D. Kelton, 320-323. Institute of Electrical and Electronics Engineers, Atlanta, Georgia.
- Wilson, J. R. and A. A. B. Pritsker, A. A. B. 1978. A survey of research on the simulation startup problem. *Simulation* **31**: 55-58.
- Wilson, J. R. and A. A. B. Pritsker. 1978. Evaluation of startup policies in simulation experiments. *Simulation* **31**: 79-89.

AUTHOR BIOGRAPHY

ANDREW F. SEILA is an Associate Professor of Management Sciences at The University of Georgia. He received the Ph.D. degree in operations research in 1976, from the University of North Carolina at Chapel Hill. His research interests include simulation model development and validation, simulation program development, and output analysis. Dr. Seila is a member of TIMS, ORSA ASA and the TIMS College on Simulation, and is Program Chairman for the 1994 Winter Simulation Conference.