OUTPUT ANALYSIS FOR SIMULATION

Andrew F. Seila

Department of Management Sciences and Information Technology
University of Georgia
Athens, GA 30602

ABSTRACT

The purpose of this paper is to review methods for analyzing data produced by simulations. The review discusses the objectives of output analysis and examines the need to apply statistical methods. Methods for estimating parameters from terminating and steady-state simulations are presented. For steady-state simulations, methods for computing estimates of the mean, as well as proportions and percentiles, are presented. Multivariate estimation techniques are discussed, and some practical advice is provided concerning output analysis in practice.

1 OBJECTIVES OF DATA ANALYSIS AND THE NEED TO COMPUTE CONFIDENCE INTERVALS

Simulations are run in order to gain an understanding of the behavior of the system under study. This understanding of the behavior is an important input to decisionmaking concerning the system. Frequently, the system's behavior is summarized by the values of one or more parameters such as mean waiting time, mean utilization of a resource, mean cost per unit time, proportion of items that are delayed longer than a specified time, or a specified percentile of the distribution of a system variable. Consider the following examples: In an analysis of the design of a planned manufacturing facility, the designers are interested in selecting the combination of machines, buffer capacities and plant layout that will minimize the mean time from the start of production until the finished product is shipped (mean waiting time). In a hospital, the administration is interested in maximizing the proportion of time that a very expensive CAT scanner is used (utilization of a

resource). The management of a transportation system which incurs costs related to personnel, vehicle usage, downtime and repair, and ridership wishes to determine the routes, schedules and maintenance plan that will minimize the mean cost of operation. In a transportation/distribution model for perishable produce from South America, the managers are concerned about the proportion of produce that cannot be delivered to market within 6 days of shipment. In a warehouse inventory model for major appliances, managers wish to determine the warehouse capacity such that 98-percent of the time, the warehouse capacity is sufficient to handle the inventory (98th percentile of inventory level). Since these quantities cannot be measured directly, data generated by the simulation must be used to infer their values. The objective of output analysis is to estimate the value(s) of one or more unknown parameters by applying appropriate statistical techniques to the data collected from the simulation.

Unfortunately, many simulations are run without applying statistical analysis to the output. That is, after many months and tens of thousands of dollars or more of development cost, the simulation is run once for each scenario to be analyzed, a single value such as the average cost is computed, and this number is treated as if it is the correct parameter value. For example, average tardiness for jobs in a manufacturing system is to be evaluated under two dispatching rules. After completing 1000 jobs under each scenario, average tardiness is computed to be 1.2 days under rule 1 and 2.3 days under rule 2. If all of the "inputs" to the simulation (job arrival and processing times, etc.) were deterministic, the outputs would be deterministic and one could say with absolute certainty that rule 1 produces smaller average tardiness than rule 2. However, virtually all simulations are run with at least some random "inputs," so the "outputs" are random. Under this

28

circumstance, one can say that for the particular set of jobs presented to the two systems, rule 1 provided superior performance, but a more general statement cannot be made without additional data analysis. This additional data analysis is needed to assess the *variability* of the sample results. Variability is usually expressed through a *confidence interval*, such as "The mean tardiness for jobs under rule 1 is 1.2 ± 0.6 days," or through a probability statement, such as "The probability that mean tardiness under rule 1 is less than mean tardiness under rule 2 is 0.994." Without a confidence interval or a probability satement, one cannot know whether the sample results are representative or not.

This tutorial is a republication of the same tutorial that was presented at the 1990 Winter Simulation Conference (Seila, 1990). Section 2 discusses the types of data that can be produced by simulations. Techniques for analyzing data from terminating simulations are presented in Section 3. Section 4 presents several methods for estimating the stationary mean from simulation data. Sections 5 and 5 discuss problems with estimating proportions and percentiles. Sequential methods for computing a fixed-width confidence interval for the stationary mean are discussed in section 8, and section 9 presents some practical ideas concerning output analysis in practice. Finally, section 10 introduces variance reduction techniques. There are several excellent textbooks (Kleijnen 1974, 1975; Fishman 1978b; Law and Kelton 1982a; Rubinstein 1981; Bratley, Fox and Schrage 1983; Lewis and Orav 1989) and surveys (Kleijnen 1982; Law 1983; Welch 1983) of output analysis methods for simulation. There is not enough room to present all of the methods in detail in this paper. The reader is referred to these references for the details of each method.

2 TYPES OF DATA PRODUCED BY SIMULATIONS

Simulations can be classified into two categories: terminating and nonterminating (Law 1980, 1983). Terminating simulations are run only until some stopping criterion is met. The stopping criterion depends upon the nature of the system and the purpose of the analysis. Some systems, such as a bank lobby or a retail store, start each day in the same state (usually empty and idle), operate for a specified period of time (10 hours, say), and terminate. Other systems, such as a manufacturing system, can operate indefinitely, but one might be interested only in the behavior of the system up to the time when the 100th item is produced or when a particular parts buffer becomes 75-percent full. When this is the case, the simulation is considered to terminate when the special event occurs, and the parameter to be estimated depends only upon the data collected up to that event, such as the mean time until the event occurs, or the mean cost per unit time prior to the terminating event.

Nonterminating simulations can conceptually run indefinitely. These systems are "balanced" so that they ultimately reach a steady-state, or stationary, pattern of behavior. The term "steady-state" does not mean that the system settles into some sort of equilibrium or deterministic behavior, but rather that the intital conditions that existed at the start of the simulation no longer influence the system's behavior. For example, in a manufacturing system, after the system has operated for several days, it will behave the same whether it initially had no work-in-process or it had 20 items being produced initially. Moreover, when a system is stationary, the behavior of observations does not depend upon when they are collected. Stated more precisely, if $X_1, X_2, ..., X_k$ is a sequence of observations from the simulation, it has the same distribution as $X_{h+1}, X_{h+2}, ..., X_{h+k}$ for any h = 1, 2, ... Whether and when this steady-state behavior is reached depends upon the characteristics of the system and the length of time the simulation has been running.

The simulation can produce either discrete or continuous observations. Discrete observations are recorded as a sequence of values $X_1, X_2, ..., X_k$. These can be produced by recording a measurement on each entity as it encounters a particular event in the simulation. For example, one might record the waiting time in the queue for each customer in a queueing simulation when the customer encounters the event that starts service, or one might record the delay from entry of an order until the order is shipped for each order in a manufacturing system. Discrete observations can also be produced by sampling a variable in the system. For example, one might sample the number of parts in a parts buffer every 10 minutes or every time a new part arrives. Continuous observations are produced by recording the level, or value, of a system variable over time. For example, $Q(t)$ might represent the number of customers in a queue at time t, or $P(t)$ might represent the number of parts in a production system at time t. In discrete event simulations the values of system variables change only when an event occurs, so the value of a continuously observed process such as $Q(t)$ or $P(t)$ is constant except for jumps at times

when relevant events occur. This makes recording a process easy since one only needs to record, for each event that occurs that affects the process, the time of the event and the new value of the process. This also makes computing time averages easy because the time average can be expressed as:

$$\overline{Q} = \frac{1}{T} \int_0^T Q(t)\, dt$$

$$= \frac{1}{T} \left[\sum_{j=1}^{n(T)} Q_j (t_j - t_{j-1}) + Q_{N(t)} (T - t_{N(t)}) \right],$$

where $n(T)$ is the number of events up to time T.

3   DATA ANALYSIS FOR TERMINATING SIMULATIONS

The parameters that are estimated from terminating simulations explicitly depend upon the initial state of the system. For example, in a simulation of a bank lobby, one might wish to estimate the teller utilization, which is the proportion of time a particular teller is busy. This proportion would depend upon the initial number of customers in the bank. If the simulation started with 10 customers waiting for service, the teller utilization would be higher than if the simulation started with no customers waiting. Since the value of the parameter depends upon the initial conditions, data must be generated by independently replicating the simulation run using the same initial conditions to start each replication. If the run is replicated n times, with each replication producing a single observation, X_i, the data will consist of n observations $X_1, X_2, \ldots, X_n$. If the runs are made using independent random number seeds, the observations will be independent and identically distributed, and the techniques that are normally applied to i.i.d. data can be applied here. A $100(1 - \alpha)$ percent confidence interval for the mean is given by

$$\overline{X} \pm t_{1 - \alpha/2, n-1}\, s_{\overline{X}}, \qquad (1)$$

where $s_{\overline{X}} = \frac{s}{\sqrt{n}}$, s^2 is the sample variance of $X_1, X_2, \ldots, X_n$:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n \left(X_i - \overline{X} \right)^2,$$

and $t_{1 - \alpha/2, n-1}$ is the $100(1 - \alpha/2)^{\text{th}}$ percentile of the Student's t-distribution with $n - 1$ degrees of freedom.

4   ESTIMATING THE MEAN FOR STEADY-STATE SIMULATIONS

As opposed to terminating simulations, steady-state simulations eventually produce data that does not depend upon the initial state of the system. Performance measures that one is interested in 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 interpretation of this can be thought of in two ways. Let $W_1, W_2, \ldots, W_n$ be the waiting times observed from n customers while the system is operating in steady-state. The mean waiting time is

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n W_i.$$

This is the average waiting time for a very large number of customers. Another interpretation of 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 steady-state mean waiting time is the mean of this distribution.

If observations are defined to have the value 1 if a given event is observed and 0 otherwise, the mean of these observations is the probability that the event is observed. For example, let $X_i = 1$ if the ith 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 observations, methods for estimating means can also be applied to estimate probabilities.

When analyzing data generated by a stationary simulation, one must deal with two problems: (1) the presence of an initial transient

portion (the startup problem), and (2) autocorrelation in the observations. The initial transient portion is usually deleted from the data prior to the application of 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 (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 initial transient portion will be minimized (Kelton 1989 and references therein).

The effect of having autocorrelation among the data is to make it difficult to estimate the variation in the sample mean. In many cases 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 observations separated by j lags, i.e., $\rho_j = \text{Corr}(X_i, X_{i+j})$, the variance of $\overline{X}$ is given approximately by:

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

Thus, the term $\sum_{j=1}^{\infty} \rho_j$ causes the sample standard error, $s/\sqrt{n}$, which is an estimator of $\sigma/\sqrt{n}$, to be biased downward. 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 of only 50-percent.

4.1 Independent Replications Method with Initial Transient Deletion

The simplest way to deal with autocorrelation among the data is to independently replicate the simulation run. On each replication, the initial transient portion is deleted and the sample mean of the stationary observations is computed. Let m represent the number of replications and $\overline{X}_k$ be the sample mean computed from the kth replication. Then $\overline{X}_1, \overline{X}_2, ..., \overline{X}_m$ is a

sample of i.i.d. observations, and the confidence interval for the mean can be computed from this data using the usual methods for independent observations (1).

There are several potential difficulties with using independent replications with initial transient deletion. First, it is critical that the number of observations to be deleted be chosen correctly. If too few are deleted, every replication will contain transient observations which will bias every sample mean. If too many are deleted, data will be wasted and the confidence interval will be wider than necessary. Secondly, this method uses data inefficiently. For example, if every replication consists of 1000 transient observations and 1000 stationary observations, this method will discard half of the data generated. Finally, one must intervene in the simulation by stopping the run, collecting the data and reinitializing the system at the end of each replication. These seem like substantial problems, and they can be burdensome, but the method of independent replications has the advantages that the sample means computed are "guaranteed" to be mutually independent, and one is not confronted with decisions associated with the other methods that will be discussed below. The independent replications method is recommended for those situations where one is reasonably sure that the system reaches steady-state quickly and run lengths may be limited.

4.2 Batch Means Method

It seems that once the system reaches steady-state and data are being collected, one should just continue collecting data from the steady-state and somehow use data from one long run, rather than a number of short independent runs, to estimate the mean. This is the essence of the batch means method (Conway 1963; Mechanic and McKay 1966; Law 1977; Fishman 1978; 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. 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 (1).

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 erroneous 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. A number of methods to determine the batch size have been proposed (Fishman 1978; Law and Carson 1979).

The advantage of the batch means method is that it uses 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 competetive 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.

4.3 Regenerative Method

Some systems have the property that at random points in time the system "regenerates" itself in the sense that future behavior is independent of the past and the patterns of development after these points are identical (Ross 1985). An example of such a system is a simple queueing system where interarrival times are identically distributed, service times are identically distributed, and all interarrival and service times are mutually independent. The regeneration points are those times when a customer arrives to find the system empty and idle. When this occurs the arrivals and departures of customers in the future will be independent of arrivals and departures in the past (since interarrival times and service times are mutually independent), and they will be an identical replica of those after any other empty and idle state, in a probabilistic sense (since interarrival times are identically distributed and service times are identically distributed). The empty and idle state for this system is a regeneration state. Indeed, to determine if a state is a regeneration state, one can ask whether, when the system enters the state, it can proceed from this point without any knowledge of the history of the system prior to

entering the state. If the answer is "yes," the state is a regeneration state.

If a system is regenerative and is started in a regeneration state, its run can be characterized by a sequence of "cycles" of data. Each cycle consists of observations collected between entrances to a regeneration state. These cycles of data will consist of random numbers of observations, but statistics computed from distinct cycles are independent since observations from distinct cycles are independent. An initial transient period does not have to be considered since the run consists of a sequence of independent cycles of data.

It is known for regenerative processes that the mean of the data 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). The confidence interval is only marginally more difficult to compute than the procedures in the batch means and independent replications methods. One would expect that since the regenerative method makes use of the special properties of the data that it would produce more accurate confidence intervals, but this has generally not been observed to be the case (Law 1984). In many cases the batch means method performed as well as or better than the regenerative method.

One difficulty with the regenerative method is that many systems either do not possess regeneration states or it is very difficult to identify states that are regeneration states. The process of showing that a state is regenerative can be very tedious mathematically and therefore beyond the abilities of many simulation practitioners. On the other hand, some systems have many regeneration states and the user is faced with the decision about which state to use.

4.4 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 the quantity (2) using known 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).

Recently, a new estimator, called an *area estimator*, 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, the distant past of the process is approximately independent of 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 process can be shown to converge to a Brownian Bridge, and this property is used to develop the area estimator for the variance of the 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.

5 SPECIAL METHODS FOR PROPORTIONS

Empirical studies (Law 1983, 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.

6 METHODS FOR ESTIMATING PERCENTILES

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

Several methods have been proposed for estimating percentiles when the data is generated by a stationary simulation (Iglehart 1976; Moore 1980; Seila 1982a, b; Heidelberger and Lewis 1984). These methods all use the sample percentile as the point estimate. The sample p-percentile, P_n, is computed by sorting the data into ascending order and selecting the observation such that p-percent of the observations is less than P_n and $(100 - p)$-percent is greater than P_n. The methods differ in the way that the variance of P_n is computed.

7 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) 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 $1 - \alpha_1$, $1 - \alpha_2$, ..., $1 - \alpha_k$, 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).

8   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 precision determines the amount of data needed to compute the confidence interval. However, some measure of variance is needed in order to compute the required sample size, and this quantity is unknown. Sequential methods (Kabak 1968; Fishman 1977; Law and Kelton 1982b) 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.

9   OUTPUT ANALYSIS IN PRACTICE

Many simulation languages and special-purpose simulation software packages have built-in report generators. These generators produce averages and other statistics ("variances," percentiles and other quantities) for resource levels, queue lengths, and other measurements generated by the simulation. Generally, the reports produced in this way should be viewed with great skepticism. If standard statistical formulas are used to compute the statistics, without dealing with the startup problem or the autocorrelation among observations, the statistics computed will be inaccurate and misleading. Before using these report generators, one is well advised to read carefully the documentation provided by the software vendor and be sure that accepted methodology is implemented in the data analysis module. If the documentation does not describe fully or accurately exactly how each value given on the report is calculated, the best decision is to not use the report generator, but instead, store the data generated by the simulation on a separate file, and analyze this data using methods that are known and accepted. Frequently, this can be done using available statistical packages such as SAS, SPSS, MINITAB and others without having to write data analysis programs. Another alternative is to use an electronic spreadsheet such as LOTUS 123, Quattro PRO and others, to do the data analysis calculation. The current generation of spreadsheets has powerful capabilities for analyzing data as well as graphing and presenting results.

10 VARIANCE REDUCTION TECHNIQUES

This paper has introduced the reader to methods for analyzing output data from simulations, especially discrete event simulations. Data analysis is a small but important part of an overall simulation analysis design. Techniques are available for increasing the efficiency of simulations by allowing parameter estimates to have smaller variance or, equivalently, computing parameter estimates with the same variance using fewer observations. These techniques, which are called variance reduction techniques, operate generally by controlling the random number streams used to drive the simulation (common random numbers, anththetic variates), exploiting the relationship between the output statistics of interest and other statistics in the model (control variates), or modifying the way the model is sampled (importance sampling); See the survey by Wilson (1984). Proper simulation output analysis cannot be performed by itself, but must be planned in conjunction with the overall design of the simulation experiment in order to produce the most precise estimates possible.

REFERENCES

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

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 Tatical Problems in Digital Simulation. *Management Science* **10**: 47-61.

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. *Concepts and Methods in Discrete Event Digital Simulation*, John Wiley, New York, NY.

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.

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. *IIE 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. 1982a. *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 .

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.

Ross, S. M. 1985. *Introduction to Probability Models*, Third Edition, Academic Press, Orlando, FL.

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. 1990. 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.

Wilson, J. R. 1984. Variance Reduction Techniques for Digital Simulation. *American Journal of Mathematical and Management Sciences* **4**: 277-312.

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.

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.