Steady-State Simulation of Queueing Processes:a Survey of Problems and Solutions

KRZYSZTOF PAWLIKOWSKI

Department of Computer Science, University of Canterbury, Christchurch, New Zealand

For years computer-based stochastic simulation has been a commonly used tool in the performance evaluation of various systems. Unfortunately, the results of simulation studies quite often have little credibility, since they are presented without regard to their random nature and the need for proper statistical analysis of simulation output data.

This paper discusses the main factors that can affect the accuracy of stochastic simulations designed to give insight into the steady-state behaviour of queueing processes. The problems of correctly starting and stopping such simulation experiments, to obtain the required statistical accuracy of the results are addressed. In this survey of possible solutions, the emphasis is put on their possible applications in the sequential analysis of output data, which adaptively decides about continuing a simulation experiment until the required accuracy of results is reached. A suitable solution for deciding upon the starting point of a steady-state analysis and two techniques for obtaining the final simulation results to a required level of accuracy are presented, together with pseudo-cod implementations.

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]: Statistical Computing, Statistical Software; G.4 [**Mathematical Software**]: Efficiency, Reliability and Robustness; G.m [**Miscellaneous**]: Queueing Theory; I.6.4 [**Simulation and Modeling**]: Model Validation and Analysis

General Terms: Algorithms, Performance, Theory

Additional Key Words and Phrases: Automation of simulation experiments, initial transient period, precision of simulation results, sequential analysis of confidence intervals, statistical analysis of simulation output data, stopping rules for steady-state simulation

INTRODUCTION

Computer-based stochastic simulation, traditionally regarded as a last resort tool (if analytical methods fail), has become a valid and commonly used method of performance evaluation. This popularity is due to the continuing development of more powerful and less expensive computers, as well as significant achievements in software engineering. One can observe a trend towards integrating simulation methodology with concepts and methods of artificial intelligence [Artificial Intelligence 1988]. Various user-friendly simulation packages offer visual interactive capabilities; traditional discrete-event simulation modeling is more and more frequently supported by object and logic-oriented programming and various concepts of artificial intelligence [Bell and O;Kneefe 1987;

Gates et al. 1988; Jackman and Medeiros 1988; Kerckhoffs and Vansteenkiste 1986; Knapp 1986; Oren and Zeigler 1987; Ruiz-Mier and Talavage 1987; Stairmand and Kreutzer 1988; Zeigler 1987]. All these developments offer users increasingly powerful and versatile techniques for performance evaluation, leading towards automatic, knowledge-based simulation packages. Simulation programming techniques and languages are discussed in numerous publications, including textbooks by Bulgren [1982], Kreutzer [1986], Law and Kelton [1982], and Payne [1982].

Applying simulation to the modeling and performance analysis of complex systemscan be compared to the surgical scalpel [Shannon 1981], whereby "in the right hand [it] can accomplish tremendous good, but it must be used with great care and by someone who knows

what they are doing". One of the applications where simulation has become increasingly popular is the class of dynamic systems with random input and output processes, represented for example by computer communication networks. In such cases, regardless of how advanced the programming methodology applied to simulation modeling is, since simulated events are controlled by random numbers, the results produced are nothing more than statistical samples. Therefore, various simulation studies, frequently reported in technical literature, can be regarded as programming exercises only. The authors of such studies, after putting much intellectual effort and time into building simulation models and then writing and running programs, have very little or no interest in a proper analysis of the simulation results. It is true that "the purpose of modeling is insight, not numbers" [Hamming 1962], but proper insight can only be obtained from correctly analysed numbers. Other modes of presenting results, for example animation, can be very attractive and useful when the model is validated, but nothing can substitute the need for statistical analysis of simulation output data in studies aimed at performance analysis; see also Schruben [1987].

In the stochastic simulation of, for example, queueing systems "computer runs yield a mass of data but this mass may turn into a mess." If the random nature of the results is ignored, "instead of an expensive simulation model, a toss of the coin had better be used" [Kleijnen, 1979]. Statistical inference is an absolute necessity in any situation when the same (correct) program produces different (but correct) output data from each run. Any sequence $x_1, x_2, \cdots, x_n$ of such output data simply consists of realisations of random variables $X_1, X_2, \cdots, X_n$. Examples illustrating this fact may be found in Kelton [1986], Law [1982], Law and Kelton [1982a], and Welch [1983, Sec.6.1] .

The simplest objective of simulation studies is the estimation of the mean μ_x of an analysed process from the sequence of collected observations $x_1, x_2, \cdots, x_n$, by calculating the average as follows:

$$\overline{X}(n) = \sum_{i=1}^{n} \frac{x_i}{n}. \tag{1}$$

Such an average assumes a random value, depends on the sequence of observations. The accuracy with which it estimates an unknown parameter μ_x can be assessed by the probability

$$P(|\overline{X}(n) - \mu_x| < \Delta_x) = 1 - \alpha \tag{2a}$$

or

$$P(\overline{X}(n) - \Delta_x \leq \mu_x \leq \overline{X}(n) + \Delta_x) = 1 - \alpha, \tag{2b}$$
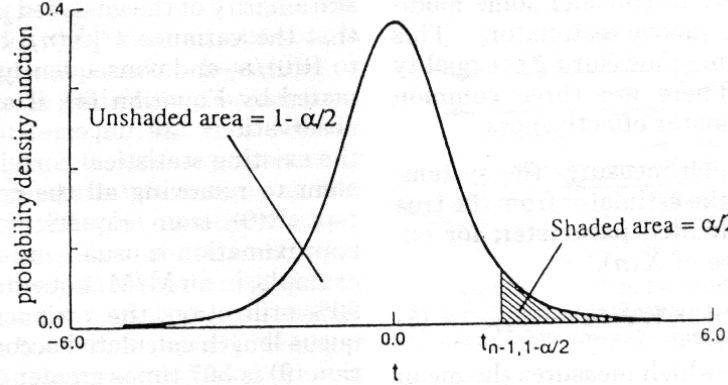
Fig. 1: $P[T_{n-1} \le t_{n-1}, 1 - \frac{\alpha}{2}] = 1 - \frac{\alpha}{2}$

where Δ_x is the half-width of the confidence interval for the estimator and $(1 - \alpha)$ is the confidence level, $0 < \alpha < 1$. Thus, if the width $2\Delta_x$ of the confidence interval is found for an assumed confidence level of $(1-\alpha)$ and the simulation experiment were repeated a number of times, the interval $(\overline{X}(n) - \Delta_x, \overline{X}(n) + \Delta_x)$ would contain an unknown average μ_x in $100(1 - \alpha)\%$ of cases and would not in $100\alpha\%$ of cases. It is well known that if observations $x_1, x_2, \cdots, x_n$ can be regarded as realizations of independent and normally distributed random variables $X_1, X_2, \cdots, X_n$, then

$$\Delta_x = t_{n-1,1-\alpha/2}\hat{\sigma}[\overline{X}(n)], \qquad (3)$$

where

$$\hat{\sigma}^2[\overline{X}(n)] = \sum_{i=1}^{n} \frac{\{x_i - \overline{X}(n)\}^2}{n(n-1)} \qquad (4)$$

is the (unbiased) estimator[1] of the variance of $\overline{X}(n)$, and $t_{n-1,1-\alpha/2}$ is the upper $(1 - \alpha/2)$ critical point obtained from the t-distribution

with $(n-1)$ degrees of freedom. In other words, for given $1 - \alpha/2$, assuming the t-distribution for the random variable $T_{n-1} = (\overline{X}(n) - \mu_x)/\hat{sigma}[\overline{X}(n)]$, we get that $P[T_{n-1} \le t] = 1 - \alpha/2$ for $t = t_{n-1,1-\alpha/2}$; see Figure 1 and Appendix A. For this reason, $t_{n-1,1-\alpha/2}$ is also called the $(1 - \alpha/2)$ quantile, or percentile, of the t-distribution with $(n - 1)$ degrees of freedom. For $n > 30$, the t-distribution can be replaced by the standard normal distribution. In that case $t_{n-1,1-\alpha/2}$ in Equation (3) should be replaced by $z_{1-\alpha/2}$, which is the upper $(1 - \alpha/2)$ critical point obtained from the standard normal distribution or, equivalently, the $(1 - \alpha/2)$ quantile of the standard normal distribution; see Appendix A. Commonly used values of $t_{n-1,1-\alpha/2}$ and $z_{1-\alpha/2}$ have been tabularized and can be found in many textbooks; see, for example, Trivedi [1982, Appendix 3, and p.489]. (Warning: The definitions used for obtaining tabularized values should always be checked. For example, $t_{n-1,1-\alpha/2}$ and $z_{1-\alpha/2}$ are sometimes denoted as $t_{n-1,\alpha/2}$ and $z_{\alpha/2}$, respectively.)

Equation (3) can also be applied if the observations $x_1, x_2, \cdots, x_n$ represent random variables which are not normally distributed. That is, if the observations are realizations of inde-

---

[1] Following standard notation, $\hat{a}$ means an estimator of the parameter a.

pendent and identically distributed (i.i.d.) random variables $X_1, X_2, \cdots, X_n$, then according to the central limit theorem (see Appendix A), the distribution of the variable $\overline{X}(n)$ tends to the normal distribution as the number of collected observations n increases. In practice, Equation (3) gives a good approximation for $n > 100$. Results obtained from Equations (1) and (3) are called *point* and *interval estimates*, respectively. Both of them are important: The former characterizes the system analysed, and the latter states the accuracy of the obtained characteristics.

If observations $x_1, x_2, \cdots, x_n$ cannot be regarded as realizations of i.i.d. random variables, we have to consider some modifications to the above estimators. This raises the problem of measuring the quality of estimators. There are three common measures of estimator effectiveness :

1. The *bias*, which measures the systematic deviation of the estimator from the true value of the estimated parameter; for example, in the case of $\overline{X}(n)$,

$$Bias[\overline{X}(n)] = E[\overline{X}(n) - \mu_x]. \quad (5)$$

2. The *variance*, which measures the mean (squared) deviation of the estimator from its mean value; that is,

$$\sigma^2[\overline{X}(n)] = E[\{\overline{X}(n) - E[\overline{X}(n)]\}^2]. \quad (6)$$

3. The *mean square error* (MSE) of the estimator, defined as

$$MSE[\overline{X}(n)] = E\{[\overline{X}(n) - \mu_x]^2\}. \quad (7)$$

Note that from these definitions,

$$MSE[\overline{X}(n)]$$
$$= \{Bias[X(n)]\}^2 + \sigma^2[X(n)]. \quad (8)$$

The main analytical problem encountered in the analysis of simulation results is that they are usually highly correlated and thus do not satisfy the precondition of statistical independence. If observations $x_1, x_2, \cdots, x_n$ represent an autocorrelated and stationary sequence of random variables $X_1, X_2, \cdots, X_n$, then the variance of $\overline{X}(n)$ is given by the formula

$$\sigma^2[\overline{X}(n)] = \frac{[R(0) + 2\sum_{k=1}^{n-1}(1 - \frac{k}{n})R(k)]}{n}, \quad (9)$$

where

$$R(k) = E[(X_i - \mu_x)(X_{i-k} - \mu_x)],$$
$$0 \le k \le n - 1 \quad (10)$$

is the autocovariance of order k (the lag k component of the autocorrelation function $R(k)$) of the sequence. The autocovariances defined in Equation (10) are independent of the index i due to the assumed stationarity of the analysed processes. Note that the variance $\sigma^2[\overline{X}(n)]$ can be reduced to $R(0)/n$, and consequently could be estimated by Equation (4), if and only if the observations are uncorrelated. Neglecting the existing statistical correlation is equivalent to removing all the components except $R(0)$ from Equation (9). Such an approximation is usually unacceptable. For example, in an M/M/1 queueing system with 90% utilization, the variance of the mean queue length calculated according to Equation (9) is 367 times greater than that from Equation (4), [Blomqvist 1967]; see Law and Kelton [1982a, p.146] for another example. Any variance analysis disregarding correlations among the observations would lead either to an excessively optimistic confidence interval for μ_x, in the case of positively correlated observations, or to an excessively pessimistic confidence interval for μ_x , in the case of negatively correlated observations; see Equation (3). A positive correlation between observations is typical in simple queueing systems without feedback connections and it is stronger

for a higher system utilization; see, for example, Daley [1968] for correlation analysis of the M/M/1 queue.

Generally, the analysis of variance of correlated processes, and the analysis of their autocorrelation functions in particular, is a complex statistical problem and therefore creates a major difficulty in the statistical analysis of simulation output data. In *terminating* (or *finite-horizon*) *simulation* used for studying the behaviour of systems during specified intervals of time, the above problem can be overcome by making a number of independent replications of the simulation experiment. In that case the means of individual observations collected during different simulation runs can be regarded as a sequence of independent (secondary) output data, and Equation (4) can be applied. Exhaustive discussions on the statistical analysis of output data from terminating simulation can be found for example, in Kleijnen [1979, 1987], Law [1980], and Law and Kelton [1982a, Sec. 8.5].

In this paper, we discuss *steady state (infinite horizon) simulation*, aimed to give insight into the behaviour of queueing processes after a long period of time. The methodology for this kind of simulation study is much more complicated. After launching, a queueing process is initially in a nonstationary phase (warm-up period). Then, if the process is stable, it moves asymptotically towards a steady state (statistical equilibrium), although different parameters usually tend to the steady state with different rates. Since observations gathered during the initial transient periods do not characterize the steady state, a natural idea is to discard all such observations before further analysis. This requires an estimation of the effective length of the initial transient period. Ignoring the existence of this period can lead to a significant bias of the final results. On the other hand, the removal of any observations increases the variance of estimates, which in turn can increase

the value of the mean-square error [Donnelly and Shannon 1981; Fishman 1972; Turnquist and Sussman 1977; Wilson and Pritsker 1978a]. Thus, a decision whether to delete or not to delete initial observations depends on the assumed criterion of goodness of the estimators. This also affects methods used to collect observations, which are discussed in the next Section 1. These and other aspects of the problem of initialization are presented broadly in Section 2.

Several methods of data collection and analysis has been proposed to overcome the theoretical problems that arise from the correlated nature of observations collected during steady-state simulation. We survey these methods in Section 1. They are distinguished by the way they estimate the variance of observed processes; the estimate is needed for determining the width of the confidence intervals. Usually, the method impose special requirements on how the output data from simulation experiments should be collected and preprocessed, depending on whether they attempt to weaken or even remove statistical dependencies among observations or take the actual correlations among observations into consideration. Among them one can distinguish:

- the method of independent replications

- the method of batch means

- the method of overlapping batch means

- the method of uncorrelated sampling

- the method of regenerative cycles

- the method based on spectral analysis

- the method based on autoregressive representation

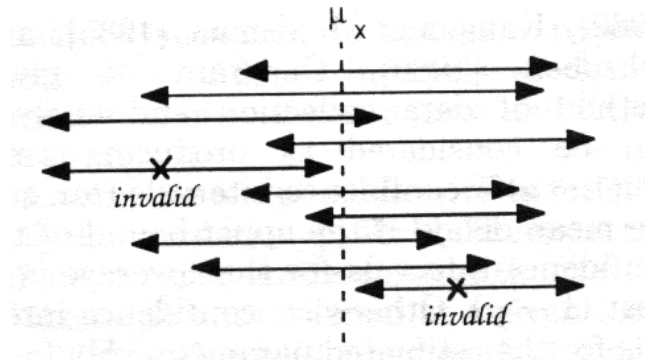- the method based on standardized time series

Fig. 2: Valid and invalid confidence intervals for μ_x in coverage analysis (iin this example the coverage equals 80%)

Various approximations assumed during statistical analysis of simulation output data may bias both point and interval estimates. For example, the final confidence interval should theoretically contain the true value of the estimated parameter with probability $(1 - \alpha)$, or equivalently, if an experiment is repeated many times, in $(1-\alpha)100\%$ of cases; but various difficulties in satisfying theoretical assumptions can cause the real rate of the confidence intervals containing the true parameter to differ significantly from $(1 - \alpha)$. The robustness of the above methods of data collection and analysis is usually measured by the *coverage of confidence intervals*, defined as the frequency with which the intervals $(\overline{X}(n) - \Delta_x, \overline{X}(n) + \Delta_x)$ contain the true parameter μ_x, at a given confidence level $(1 - \alpha)$, $0 < \alpha < 1$; see Figure 2. Thus, the coverage analysis can be applied only to systems with theoretically well-known behaviour, since the value of μ_x has to be known. Any analysed method must be applied in a statistically significant number of repeated simulation experiments (usually 200 or more replications) to determine the fraction of experiments producing the final confidence intervals covering the true mean value of the estimated parameter.

The coverage error and its main sources were theoretically analysed by Glynn [1982]. Kang and Goldsman [1985], and Schruben [1980]. Generally, a given method of data collection and analysis can be considered as producing valid $100(1 - \alpha)\%$ confidence intervals (for, say, the mean delay) if the upper bound of the confidence intervals for the coverage is at least $(1 - \alpha)$. Otherwise, confidence intervals for the estimated parameter should be regarded as invalid and the method as inaccurate. A few additional measures for the effectiveness of methods used for data collection and analysis were proposed in Schruben [1981]. The weakest point of such approach is that there is no theoretical basis for extrapolating results found for simple, analytically tractable systems to more complex systems, which are the real subjects of simulation studies; see Fox [1978].

Some conclusions on the quality of the methods can be also obtained from the analysis of the asymptotic properties of the variance estimators $\hat{\sigma}^2[\overline{X}(n)]$ used by particular methods for determining the width of confidence intervals. Namely, the quality of various variance estimators can be compared by comparing the limit values of their bias,

$$Bias\{\hat{\sigma}^2[\overline{X}(n)]\}$$
$$= E\{\hat{\sigma}^2[\overline{X}(n)] - \sigma^2[\overline{X}(n)]\} \qquad (11)$$

and variance $Var\{\hat{\sigma}^2[\overline{X}(n)]\}$, as the number of observations tends to infinity; that is, as $n \to \infty$; see Goldsman and Meketon [1985] and Goldsman et al. [1986]. Alternatively, Schmeiser [1982] proposed studying the asymptotic properties of the expected values and variances of the halfwidth of confidence intervals Δ_x generated by a method; see also Glynn and Iglehart [1985] and Goldsman and Schruben [1984]. Following such criteria, one can say that the method using the variance estimator with the smallest bias and smallest variance, or using the estimator of the width of confidence interval having the smallest expected value and smallest variance, is (asymptotically) superior to others. Unfortunately, these criteria are not universal since a small bias can be accompanied by large variance or vice versa. In terms of confidence intervals, it can mean wide and very variable confidence intervals giving good coverage, or, conversely, stable and narrow confidence intervals giving poor coverage.

Even if we were able to collect independent and identically distributed output data from simulation runs, we cannot be fully protected from erroneous conclusions because of inherent variations of simulation output data caused by the pseudorandom nature of input data; see Pidd [1984, Sec.8.4.2] for a more detailed discussion. Certainly, one of the most important issues of any simulation experiment is to use proper input data. In the case of queueing processes this usually means selecting a good generator of uniformly distributed (pseudo)random numbers. The state of art in this area is summarized in Park amd Miller[1988].

The experimental accuracy of simulation can be further improved by using *variance reduction techniques* (VRT) developed for reducing the variance of recorded results without affecting their mean values. Surveys of variance reduction techniques may be found in Bratley et al. [1983, Sec.2], Cheng [1986], Frost et al. [1988], Kleijnen [1974, Ch.3], Law and Kelton [1982, Ch.11], Nelson and Schmeiser [1986], Wilson [1983, 1984]. Although all of these techniques can be applied to the method of independent replications, only some of them can be fitted into other methods of data collection and aanlysis. Two VRTs, known as the method of control variables (or control variates) and importance sampling, seem to be the most frequently advocated; see, for example Anonuevo and Nelson [1986], Frost et al. [1988], Izydorczyk et al. [1984], Lavenberg and Welch [1981], Lavenberg et al. [1982], and Venkartraman and Wilson [1985], for discussions of these methods and their applications. Unfortunately, despite the fact that various VRTs have been extensively studied theoretically since the very beginning of digital simulation [Harling 1958], most of them have found only limited prac-

tical application. The reson is that they can be difficult to implement in simulation studies of even moderately complex systems, since they are strongly model dependent and/or require a substantial amount of computing resources. The simplest VRT in the context of queueing processes [Law 1975], consists of a direct estimation of the mean time in queue (without service time), since this estimator usually has a smaller variance than (direct) estimators of the mean time in the system or the mean queue length. Next, the estimates of the latter group of parameters can be obtained indirectly by applying, for example, Little's formula. The efficiency of this approach has been proved for G/G/c queueing systems [Carson and Law 1980]. A variance reduction as high as 90% has been reported, although generally the reduction tends to 0 as the system utilization tends to 100%. Indirect estimation has been found justifiable even for more complex queueing systems, provided that the interarrival and waiting times are negatively correlated, [Glynn and Whitt 1989]. Another VRT, for achieving more stable confidence intervals by reducing the variance $Var\{\hat{\sigma}^2[\overline{X}(n)]\}$, has been proposed by Meketon and Schmeiser [1984]. They showed that the variance reduction can be achieved by calculating estimators from overlapping subsequences of output data; see the method of overlapping batch means in Section 1.

The methods for data collection and analysis can be used either in their fixed-sample-size versions or in their sequential versions. In the former case, statistical analysis is performed once at the end of the simulation experiment when a predetermined number of observations, assumed to be sufficient to get results of a required accuracy, has been collected. The present survey of methods for data collection and analysis,

used in steady-state simulation, focusses on their sequential versions, in which the length of simulation is increased sequentially from one checkpoint to the next until a prespecified accuracy of the point estimators is obtained. Such procedures, which automatically control the length of simulation experiments, are very desirable in user-friendly simulation packages. Sequential statistical analysis is also more efficient, since it is usually difficult to determine a priori the length of simulation needed by a fixed-size procedure that would be sufficient to obtain a required width of confidence intervals at the assumed level of confidence Law and Kelton [1982b, 1984].

Among the few possible criteria for stopping the simulation, probably the most useful one is based on the relative (half) width of the confidence interval at a given confidence level $(1 - \alpha)$ defined as the ratio

$$\epsilon = \frac{\Delta_x}{\overline{X}(n)} \qquad 0 < \epsilon < 1; \qquad (12)$$

c.f. Equation (2).

The above ratio is also called the *relative precision of the confidence interval*. The simulation experiment is stopped at the first checkpoint for which $\epsilon \leq \epsilon_{max}$, where ϵ_{max} is the required limit relative precision of the results at the $100(1 - \alpha)\%$ confidence level, $0 < \epsilon_{max} < 1$. Note that if

$1 - \alpha$

$$\leq P[|\overline{X}(n) - \mu_x| \leq \epsilon|\overline{X}(n)|], \qquad (13)$$

then, for $\mu_x \neq 0$,

$P[|\overline{X}(n) - \mu_x| \leq \epsilon|\overline{X}(n)|]$

$= P[|\overline{X}(n) - \mu_x| \leq \epsilon|\overline{X}(n) - \mu_x + \mu_x|]$

$\leq P[|\overline{X}(n) - \mu_x| \leq \epsilon|\overline{X}(n) - \mu_x| + \epsilon|\mu_x|].$

and finally

$1 - \alpha \leq P[|\overline{X}(n) - \mu_x| \leq \epsilon|\overline{X}(n)|]$

$$\leq P\left[\frac{|\overline{X}(n) - \mu_x|}{|\mu_x|} \leq \frac{\epsilon}{1 - \epsilon}\right], \qquad (14)$$

where

$$\frac{|\overline{X}(n) - \mu_x|}{|\mu_x|} \qquad (15)$$

is called the relative error of the confidence interval.

Sequential reasoning about the statistical accuracy of results is particularly advisable if higher precision is required. There is a danger, however, since when precision requirements are increased, the resulting confidence intervals have a greater chance of not containing the true value of parameters (*the narrower the confidence interval, the worse the coverage effect*). One can also expect that lower coverage is more probable in the case of negatively correlated observations for which it is more likely to get an underestimated value of their variance. Higher accuracy requirements can also unacceptably lengthen simulation runs controlled by a sequential procedure. In this context, any variance reduction technique can be regarded as a technique for speeding up the simulation, since any decrease in the value of the variance decreases the width of the resulting confidence intervals, and a specified accuracy can be met more quickly.

The sequential approach is regarded by some as the only alternative to steady-state

simulation, see Bratley [1983, p.101]. Some statisticians believe that it is possible to devise a procedure that would fully automatically conduct data collection and analysis using a sequential rule for assessing the accuracy of estimates; however, a fully acceptable solution has not yet been invented. Relatively few commercially available simulation packages offer some degree of automation of statistical analysis; c.f. Catalog of Simulation Software [1987]. For example, sequential procedures, automated to some extent and based on independent replications, regenerative, and spectral methods of data collection and analysis are implemented in Research Queueing Package (REQS), [Mac-Nair 1985; Sauer et al. 1984], and in its network oriented extension Performance Evaluation Tool(PET) [Bharath-Kumar and Kermani 1984]. A method of batch means is incorporated in SIMSCRIPT II.5 [Mills 1987] and its specialized variations such as Network II.5 and COMNET II.5. Partial automationm of data analysis is also offered in Queueing Network Analysis Package Version 2 (QNAP2) [Potier 1984, 1986].

The effectiveness of proposed methods generally depends on the level of a priori knowledge of the system's behaviour. Successful fully automated implementations have been reported only for some restricted classes of queueing processes. A fully automated procedure that could be used in stochastic simulation studies of a broader class of systems by users having little knowledge or interest in the statistical analysis of the output data is a matter for future investigation. The quality of presently known methods depends on the selection of values for the various parameters involved, which requires some knowledge of the analysed systems' dynamics. In Section 3 we present details of two sequential procedures for data

collection and analysis which were implemented in simulation studies of data communication protocols reported in Asgarkhani and Pawlikowski [1989] and Pawlikowski and Asgarkhani [1988].

This report is not addressed to statisticians. We try to avoid the strict mathematical formulation of the problems and use basic statistical terminology. Readers are referred to the references for more details.

1 METHODS OF DATA COLLECTION AND ANALYSIS

During the last 25 years of discussion on the methodology of statistical analysis of output data from steady-state simulation, initiated by Conway's [1963] "Some Tactical Problems in Digital Simulation," a variety of methods for data collection and analysis has been proposed to circumvent the nonstationarity of simulated queueing processes (especially the initial nonstationarity caused by the existence of the initial transient period) and the autocorrelation of events (correlations among collected observations). As has been mentioned, these methods either try to weaken (or remove) autocorrelations among observations or to exploit the correlated nature of observations in analysis of variance needed for determining confidence intervals for the estimated parameters.

In **the method of replications**, adopted from terminating simulation, the problem with autocorrelated nature of the original output data is overcome in a conceptually simple way: The simulation is repeated a number of times, each time using a different, independent sequence of random numbers, and the mean value of observations collected during each run is computed.

These means are used in further statistical analysis as secondary, evidently independent and identically distributed, output data. By the central limit theorem (see Appendix A), such data also become approximately normally distributed. Following these motivations, if m observations are collected during each replication, the sequence of n primary observations from $k_b = n/m$ replications $(x_{11}, x_{12}, \cdots, x_{1m})$, $(x_{21}, x_{22}, \cdots, x_{2m})$, $\cdots$, $(x_{kb1}, x_{kb2}, \cdots, x_{kbm})$ is replaced by the sequence of their means $\overline{X}_1(m), \overline{X}_2(m), \cdots, \overline{X}_{kb}(m)$, where

$$\overline{X}_i(m) = \frac{1}{m} \sum_{j=1}^{m} x_{ij}, \qquad (16)$$

which are used to obtain the point and interval estimates of the process. Namely, adopting Equations (1)-(4), we get the estimator of the mean μ_x as

$$\overline{\overline{X}}(k_b, m) = \frac{1}{k_b} \sum_{i=1}^{k_b} \overline{X}_i(m), \qquad (17)$$

which, for $n = k_b m$ is numerically equivalent to $\overline{X}(n)$. We also set the $100(1 - \alpha)\%$ confidence interval of μ_x as

$$\overline{\overline{X}}(k_b, m) \pm t_{k_b-1,1-\alpha/2} \hat{\sigma}[\overline{\overline{X}}(k_b, m)], \qquad (18)$$

where

$$\hat{\sigma}^2[\overline{\overline{X}}(k_b, m)]$$

$$= \sum_{i=1}^{k_b} \frac{\{\overline{X}_i(m) - \overline{\overline{X}}(k_b, m)\}^2}{k_b(k_b - 1)} \qquad (19)$$

is the estimator of the variance of $\overline{\overline{X}}(k_b, m)$, and $t_{k_b-1,1-\alpha/2}$, for $0 < \alpha < 1$, is the upper $(1 - \alpha/2)$ critical point from the t-distribution with $k_b - 1$ degrees of freedom.

There are different opinions on the effectiveness of this method as compared to other methods of data collection and analysis, all of which are based on a single (longer) run of the simulation experiment. Arguments defending the method of replications are provided by the results of Kelton and Law [1984], Lavenberg [1981, p. 114], and Turnquist and Sussman [1977], which reveal that better accuracy of the point estimator measured by its MSE [see Equation (7)] can be achieved if the simulation is run a few times rather than if it is run only one time . But Cheng [1976] argues that such a policy cannot always be correct; see also Madansky [1976]. On the other hand, the method of replications appears to be much more sensitive to the nonstationarity of observations collected during the initial transient period than methods based on single simulation runs, since any new replication begins with a new warm-up period. If the bias of the estimator $\overline{\overline{X}}(k_b, m)$ is our main concern, then data collected during the initial transient period should be discarded, (see Section 2), and in Equation (17) $\overline{\overline{X}}_i(m)$ should be replaced by

$$\overline{X}_i(m - n_{oi}) = \frac{1}{m - n_{oi}} \sum_{j=n_{oi}+1}^{m} x_{ij}, \qquad (20)$$

where n_{oi} is the number of observations discarded from the $i_t h$ replication. Thus, the total number of initial observations discarded from k_b replications would be about $k_b - 1$ times larger than in corresponding single run methods. In the sequential version of the method, new replications are generated until the required accuracy is reached. It was found that proper estimation of the length of the initial transient period can significantly improve the final coverage of confidence intervals obtained by the method of

replications. There is a trade-off between the number of replications and their length for achieving a required accuracy of estimators. Fishman [1978, p. 122] suggests using at least 100 observations in each replication (i.e., $m - n_{oi} \geq 100$) to secure normality of the replication means. Moreover, results of Law [1977] and Kelton [1984] show that it is better to keep replications longer than to make more replications, since that will usually improve the final coverage too.

All other methods of data collection and analysis have been developed for obtaining steady-state estimators from single simulation runs rather than from multiple replications. In the **method of batch means**, first mentioned by Blackman and Tuckey in [1958] and Conway et al. [1959], the recorded sequence of n original observations $x_1, x_2, \cdots, x_n$ is divided into a series of nonoverlapping batches $(x_{11}, x_{12}, \cdots, x_{1m})$, $(x_{21}, x_{22}, \cdots, x_{2m})$, $\cdots$, of size m, and batch means $\overline{X}_1(m), \overline{X}_2(m), \cdots, \overline{X}_{k_b}(m)$, corresponding to the means over replications from Equation (16), are next used as (secondary) output data in statistical analysis of the simulation results. The mean μ_x is estimated by $\overline{X}(n) = \overline{\overline{X}}(k_b, m)$ [see Equation (17)], and the confidence interval is given by Equations (18)-(19), and with k_b meaning now the number of batches and m meaning the batch size, $k_b = n/m$. This approach is based on the assumption that observations more separated in time are less correlated. Thus, for sufficiently long batches of observations, batch means should be (almost) uncorrelated; see Brillinger [1973] for a formal justification. By the central limit theorem (see Appendix A), batch means can also be regarded as approximately normally distributed, which justifies the application of Equation (18). If the bias of the estimator

$\overline{\overline{X}}(k_b, m)$ is our main concern, then again the effective length of the initial transient period should be determined (see Section 2), and the first n_o observations collected during this period should be deleted. Thus, the division of observations into k_b batches of size m should begin with setting $x_{11} = x_{n_o+1}$.

Selection of a batch size that ensures uncorrelated batch means appears to be the main problem associated with this method. Another problem is selecting a suitable length of the initial transient period. A natural solution is to estimate correlation between batch means starting from an initial batch size m_1, and if the correlation cannot be ignored, increase the batch size and repeat the test. At this stage, the method in its sequential version requires two procedures: the first sequentially testing for an acceptable batch size and the second sequentially testing the accuracy of estimators. Correlation between the means of batches of size m can be measured by estimators of the autocorrelation coefficients

$$\hat{r}(k, m) = \frac{\hat{R}(k, m)}{\hat{R}(0, m)} \qquad (21)$$

where

$$\hat{R}(k, m) = \frac{1}{k_b - k} \sum_{i=k+1}^{k_b} [\overline{X}_i(m)$$

$$- \overline{X}(n)][\overline{X}_{i-k}(m) - \overline{X}(n)] \qquad (22)$$

is the estimator of autocovariance of lag, $k = 0, 1, 2, \cdots$, in the sequence of batch means $\overline{X}_1(m), \overline{X}_2(m), \cdots, \overline{X}_{k_b}(m)$. The sequence of batch means can be regarded as non-autocorrelated when all $\hat{r}(k, m), k = 1, 2, \cdots$, assume small magnitudes, say, if they are less than 0.05.

One can also determine the threshold for neglecting the autocorrelations in a statis-

tical way, by testing their values at an assumed level of significance; see Adamn [1983] and Welch [1983, p. 306]. The main analytical problem is caused by the fact that $\hat{r}(k, m)$'s of higher order are less reliable since they are calculated from fewer data points[2]. The higher the lag of an autocovariance, the fewer the observations available to estimate this autocovariance within a batch. Usually it is suggested to consider autocovariances of the lag not greater than 25% of the sample size [Box and Jenkins 1970, p. 33] or even of 8 -10% (c.f., Geisler [1964]). Law and Carson [1978] have proposed a procedure for selecting the batch size for processes with autocovariances monotonically decreasing with the value of the lag; see also Law and Kelton [1982a]. In such a case, only the lag 1 autocorrelation has to be taken into account. In this procedure three types of behaviour of $\hat{r}(1, m)$ as a function of m are distinguished. In the same class of processes Fishman [1978b] has proposed testing batch means against autocorrelation using von Neumann's [1941] statistic. One version of Fishman's procedure can be applied to processes with positive values of $\hat{r}(1, m)$, which decrease monotonically with m, whereas another includes cases when $\hat{r}(1, m)$ is a function oscillating in a damped harmonic fashion, assuming both positive and negative values [Fishman 1978, p. 240]. A sequential procedure using the former version together with the control variates variance reduction technique is presented in Anonuevo and Nelson [1986].

In this procedure observations are batched not by count but by time, that is, over equal time intervals, whose length is specially selected, giving uncorrelated sequence of time means over the intervals.

Procedures proposed for selecting the batch size m^* use various statistical techniques and various criteria, hence they usually lead to very different batch sizes. The statistical tests involved typically require many more batch means to be tested against autocorrelation than is needed for getting results with a required precision. Consequently, it has been reported that some of these procedures can lead to interval estimates with very poor coverage, caused, among other reasons, by accepting batch sizes that are too small. For example, the above-mentioned Fishman's procedures can select batches of as few as eight observations. Law [1983] refers to simulation studies of M/M/1 queues in which the method of batch means with the procedure proposed in Law and Carson [1979] was used. Using $k_b = 10$ batches of size $m = 32$, for system utilisation $\rho = 0.9$, and 500 repeated simulation experiments, the achieved coverage of the nominal 90% confidence intervals was only 63%. For these reasons, Kleijnen et al. [1982] suggest the use of a modified Fishman's procedure accepting batches at least 100 observations long, whereas Welch [1983, p. 307] recommends constructing batches at least 5 times larger than the size m^* given by a test against autocorrelation, provided that at least 10 such batches can be recorded.

Schmeiser [1982] theoretically analysed the trade-off between the number of batches, the batch size, and the coverage of confidence intervals. The results show that usually the number of batches used in the analysis of confidence intervals should be not less

---

[2] The variance of the estimator $\hat{R}(k, m)$ is reduced if the factor $1/(k_b - k)$ in Equation (22) is replaced by $1/k_b$. But this variance reduction is followed by an increase of the bias of the estimator; see Parzen [1961]

than 10, and need not be greater than 30, if the simulation run is long enough to secure an adequate degree of normality and independency of batch means. This means that having determined a batch size that gives neglegibly correlated and approximately normal batch means (which can sometimes require even a few hundred batches to be tested), there is no need to use more than $k_b = 30$ batches to obtain confidence intervals with a good coverage. Thus confidence intervals can be analysed constructing a smaller number of longer batches. Such a transformation improves the normality and independence of batch means and, as such, usually yields better coverage of the confidence intervals. Although the problem of selecting a suitable batch size has not yet been satisfactorily solved, the above method of batch means generally behaves better than the method of replications [Law 1977] and is quite often applied in practice. An example of its sequential version, offering automated analysis of simulation output data, is presented in Section 3.

As mentioned in Introduction, to generate short and stable confidence intervals an estimator of variance $\sigma^2[\overline{X}(n)]$ should have a small variance itself. The theory of statistics says that this requirement is equivalent to using variance estimators with high degrees of freedom, since the number of degrees of freedom is inversely proportional to the variance of such estimators. Meketon and Schmeiser have shown that the variance of the variance estimator can be reduced by introducing overlapping batches of observations, [Meketon 1980; Meketon and Schmeiser 1984]. This solution is applied in the **method of overlapping batch means**, a modification of the previous method, which in this context is known as the *method of nonoverlapping*

batch means.

Following Meketon and Schmeiser, the variance of $\overline{X}(n)$ is estimated as

$$\hat{\sigma}_{ob}^2[\overline{X}(n)] = (\tfrac{n}{m} - 1)^{-1}$$

$$\sum_{j=1}^{n-m+1} \frac{\{\overline{X}_j(m) - \overline{X}(n)\}^2}{n-m+1}. \qquad (23)$$

where

$$\overline{X}_j(m) = \frac{1}{m} \sum_{i=0}^{m-1} x_{j+i} \qquad (24)$$

is the batch mean of size m beginning with the observation x_j, and $\overline{X}(n)$ is the overall mean, averaged over all observations, [see Equation (1)]. Then the confidence interval can be approximated as

$$\overline{X}(n) \pm t_{,1-\alpha/2}\hat{\sigma}_{ob}[\overline{X}(n)], \qquad (25)$$

where $t_{,1-\alpha/2}$ is the upper $(1 - \alpha/2)$ critical point of the t-distribution with $k = 1.5(\lfloor n/m \rfloor - 1)$ degrees of freedom; $\lfloor a \rfloor$ denotes the "floor" function of a, the greatest integer not greater than a. Thus the number of degrees of freedom is 1.5 times greater here than in the method of nonoverlapping batch means [c.f. Equation(18)], with $k_b = n/m$ nonoverlapping batches. It can be proved that if each new observation starts the next batch of observations, then, assuming the same batch size m, the method of overlapping batch means gives asymptotically (as $n to \infty$) more stable confidence intervals, since the variance of $\hat{\sigma}_{ob}^2[\overline{X}(n)]$ equals 2/3 of the variance of $\hat{\sigma}^2[\overline{\overline{X}}(k_b, m)]$ for nonoverlapping batch means, [Meketon and Schmeiser 1984]. The bias of both variance estimators remains practically the same [Goldsman et al. 1986]. It has been shown that even overlapping batches only by half

of their size gives an asymptotic variance of $\hat{\sigma}^2_{ob}[\overline{X}(n)]$ equal to 75% of the variance of the estimator calculated from nonoverlapping batches with the same batch size [Welch 1987]. An important feature of this technique is that it makes it possible to increase the size of batches within a given length of simulation run without decreasing the number of batches. All of this makes the estimator given by Equation (23) quite attractive, despite it being computationally more complex, although it remains computationally tractable. The optimal batch size for this estimator (and for others), in terms of its asymptotic mean square error, was analysed by Goldsman and Meketon [1989] and Song [1988].

Since the method of batch means was originally invented to obtain less correlated data, discarding some observations between consecutive batches should be a natural and effective way to obtain additional decrease in the correlation between batch means. An extreme solution is applied in the **method of uncorrelated sampling** in which only single observations, each of ν observations apart, are retained and all other observations are discarded; see Schmidt and Ho [1988] and Solomon [1983, p. 200]. The distance between consecutive retained observations should be selected large enough to make the correlation between them negligible. When this is done and the initial n_o observations from the transient period are discarded, the sequence of K retained observations $x_{n_o+1}, x_{n_o+\nu+1}, \cdots, x_{n_o+(K-1)\nu+1}$ contains realizations of (almost) independent and identically distributed random variables. Thus, the mean μ_x can be estimated by

$$\overline{X}_{us}(K) = \frac{1}{K} \sum_{i=0}^{K-1} x_{n_o+i\nu+1}. \qquad (26)$$

Its confidence interval is

$$\overline{X}_{us}(K) \pm t_{K-1,1-\alpha/2}\hat{\sigma}_{us}[\overline{X}_{us}(K)], \qquad (27)$$

where

$$\hat{\sigma}^2_{us}[\overline{X}_{us}(K)] = \sum_{i=0}^{K-1} \frac{\{x_{n_o+i\nu+1} - \overline{X}_{us}(K)\}^2}{K(K-1)}$$

$$(28)$$

and $t_{K-1,1-\alpha/2}$ is the upper $(1-\alpha/2)$ critical point obtained from the t-distribution with $(K-1)$ degrees of freedom. The size ν of separating intervals can be selected sequentially, applying the same tests as for determining the batch size in the method of (nonoverlapping) batch means, although one can expect that the size of intervals used for removing correlations between individual observations will usually be smaller than the batch size required for making batch means uncorrelated. In the example considered in Solomon [1983], the separating intervals of length $\nu = 25$ were selected by applying the Spearman rank correlation test. No results on effectiveness of this method are available, but some consider that it wastes too many observations. In fact, as shown by Conway [1963], the benefit of introducing the separating intervals is doubtful since it increases the variance of estimates. Let us also note that one of the reasons for batching observations is to make them more normally distributed. Thus, Equation (27) can give quite a poor approximation to the confidence interval if the analysed process is not a normal one.

In the **regenerative method** observations are also grouped into batches, but the batches are of random length, determined by successive instants of time at which the simulated process starts afresh (in the probabilistic sense), that is, at which its future

state transitions do not depend on the past. In the theory of regenerative processes (see, e.g., Shedler [1987]), which gives the theoretical support for this method, such instants of time are called *regeneration points*, and the states of the processes at these points are called *regeneration states*. The special nature of the process behaviour after each regeneration point—its fresh rebirth—causes batches of observations collected during different *regenerative cycles* (i.e., within periods of time bounded by consecutive regeneration points) to be statistically independent and identically distributed. So are the means of these batches. For example, the regeneration points in the behaviour of simple singleserver queueing systems are clearly the time instants at which newly arriving customers find the system empty and idle. From any such moment on, no event from the past influences the future evolution of the system. More examples are given in Welch [1983, p. 317] and Shedler [1987, Sec. 2.1]. Note that usually a few, or even infinitely many, different sequences of regeneration points (for different types of regeneration states) can be distinguished in the behaviour of a system.

As a consequence of the identical distributions of output data collected within consecutive regenerative cycles, the problem of initialization vanishes if a simulation experiment commences from a selected regeneration point. The regenerative method was first suggested by Cox and Smith [1961, p. 136], then independently developed by Fishman [1973b, 1974], and by Crane and Iglehart [1974, 1975a]. Because of the random length of batches, these methods require special estimators, usually in the form of a ratio of two variables. In particular, if observations $x_1, x_2, \cdots, x_n$ are collected during N consecutive regenerative cycles, then the mean μ_x of the observed process is estimated by

$$\overline{X}_{re}(N) = \frac{\overline{Y}(N)}{\overline{T}(N)}, \qquad (29)$$

where

$$\overline{Y}(N) = \sum_{i=1}^{N} \frac{Y_i}{N}, \qquad (30)$$

$$\overline{T}(N) = \sum_{i=1}^{N} \frac{T_i}{N}, \qquad (31)$$

In the above equations,

$$T_i = n_{i+1} - n_i \qquad (32)$$

is the length of the $i_t h$ regenerative cycle or, equivalently, the number of observations collected during the cycle i, n_i is the serial number of an observation collected at the $i_t h$ regeneration point, and

$$Y_i = \sum_{j=n_i}^{n_{i+1}-1} x_j. \qquad (33)$$

Thus Y_i is the sum of observations collected during the $i_t h$ regenerative cycle. If sufficiently many regenerative cycles is recorded, then the $100(1 - \alpha)\%$ confidence interval of unknown parameter μ_x is bounded by

$$\overline{X}_{re}(N) \pm z_{1-\alpha/2}\hat{\sigma}_{re}[\overline{X}_{re}(N)], \qquad (34)$$

where $z_{1-\alpha/2}$ is the upper $(1 - \alpha/2)$ critical point from the standard normal distribution, and

$$\hat{\sigma}_{re}^2[\overline{X}_{re}(N)]$$
$$= \frac{(s_Y^2 - 2\overline{X}_{re}(N)s_{YT} + [\overline{X}_{re}(N)]^2 s_T^2)}{\overline{T}(N)\sqrt{N}},$$

$$(35)$$

$$s_Y^2 = \sum_{i=1}^{N} \frac{\{Y_i - \overline{Y}(N)\}^2}{(N-1)}, \qquad (36)$$

$$s_{YT} = \sum_{i=1}^{N} \frac{\{Y_i - \overline{Y}(N)\}\{T_i - \overline{T}(N)\}}{(N-1)},$$

$$(37)$$

$$s_T^2 = \sum_{i=1}^{N} \frac{\{T_i - \overline{T}(N)\}^2}{(N-1)}. \qquad (38)$$

It can be shown that $\overline{X}_{re}(N)$ given by Equation (29) is a biased estimator of μ_x (the mean value of the ratio of two variables is approximated by the ratio of their mean values, which generally is not correct), although it is a consistent estimator, which means that $\overline{X}_{re}(N)$ tends to μ_x with probability 1, as $N to \infty$. Additionally, the asymptotic normality of the ratio estimator $\overline{X}_{re}(N)$, on which the formula given by Equation (34) is based, is questionable even for relatively large N. Thus this method eliminates the bias of initialization but introduces new sources of systematic errors caused by special forms of estimators. Some efforts have been made to obtain less biased estimators than those of Equations (29) and (34). Less biased estimators of μ_x have been proposed in Fishman [1977] (Tin's estimator), [Iglehart 1975] (the "jackknife" estimator) and Minh [1987]. Comparative studies reported in Gunther and Wolff [1980], Iglehart [1975, 1978], and Law and Kelton [1982b] show that using the jackknife approach for the mean and variance estimation can significantly improve the accuracy of the estimates, although some question the generality of these results [Bratley 1983, p. 92]. In some reported cases, especially if a small number of regenerative cycles is recorded,

the performance of the regenerative method appears to be poor indeed, worse than that of the method of (nonoverlapping) batch means (see Law and Kelton [1982b, 1984]).

An effective sequential, regenerative procedure for output data analysis has been proposed by Fishman [1977]. Because of reservations about the appropriateness of the assumption of the approximate normality of $\overline{X}_{re}(N)$, the procedure is equipped with a statistical test for normality of the collected data (the Shapiro-Wilk test; see Shapiro and Wilk [1965] or Bratley et al. [1983, App. A]). This normality test requires grouping output data (means over observations collected during consecutive regenerative cycles) into fixed size batches. Fishman [1977] proposed using batches containing data collected during at least 100 cycles and increasing the size of batches if the normality test fails. Results presented in Law and Kelton [1982b] show that this method, although rather more complicated numerically because of testing for normality, produces more accurate results in comparison with both a sequential "plain" regenerative method proposed by Lavenberg and Sauer [1977] and a sequential method of (nonoverlapping) batch means proposed by Law and Carson [1979]. A sophisticated modification of the regenerative method was also proposed by Heidelberger and Lewis [1981], who suggest interactive intervention by users in the process of data collection and analysis for achieving better accuracy. The regenerative method of data collection and analysis requires a regeneration state to be well chosen so as to ensure that a sufficient number of output data can be collected for statistical analysis. To satisfy this requirement, a few approximations to the method have been proposed; see Crane and Iglehart [1975b], Crane and Lemoine

[1977], Gunther and Wolff [1980], Heidel-
berger [1979]. Gunter and Wolff [1980] pro-
posed replacing single regeneration states by
sets of states and defining (almost) regener-
ative cycles bounded by entries of the sim-
ulated process to such sets of states rather
than to a single regeneration state as in the
original method. Such modification can lead
to even better accuracy of results than that
obtained by the original (accurate) regener-
ative method, at least in the cases reported
in Gunther and Wolff [1980]. But users
must still select a proper set of (almost) re-
generative states, which can sometimes in-
volve substantial preparatory work. This ap-
proach certainly deserves to be more thor-
oughly compared with others. On the other
hand, selecting the most frequently occur-
ring regeneration state does not guarantee
the best quality of the estimator. For exam-
ple, Calvin [1988] has shown that such a se-
lection may even result in the estimator with
the largest variance. Thus, a general crite-
rion for selecting regeneration states still re-
mains an open question. The random length
of regenerative cycles makes the control of
the accuracy of results more difficult, since
stopping the simulation at a nonregenerative
point can cause a substantial additional bias,
[Meketon and Hiedelberger 1982]. Any vari-
ant of the regenerative method offers very at-
tractive solution to the main "tactical" prob-
lems of stochastic simulation, but it requires
a deeper a priori knowledge of the simulated
processes.

As has been said, one can also try to
take into account the correlated nature of
observations when the variance, needed for
the analysis of confidence intervals, is esti-
mated. The simplest, but usually heavily
biased, estimator of the variance $\hat{\sigma}^2[\overline{X}(n)]$
can be obtained directly from Equation (9).
Namely,

$$\hat{\sigma}^2[\overline{X}(n)]$$

$$= \frac{1}{n}\left[\hat{R}(0) + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\hat{R}(k)\right], \quad (39)$$

where

$$\hat{R}(k) = \frac{1}{n-k}$$

$$\cdot \sum_{i=k+1}^{n}[x_i - \overline{X}(n)][x_{i-k} - \overline{X}(n)], \quad (40)$$

for $0 \le k \le n-1$. This estimator can be
improved by discarding $\hat{R}(k)$'s of higher or-
der since, as has been mentioned discussing
the problem of selecting the batch size in
the method of batch means, they are less
reliable because they are calculated from
fewer data points. Since the above formulas
assume that the analysed observations are
taken from a stationary process, the estima-
tion should be forwarded by detecting the ef-
fective length of the initial transient period
to discard initial nonstationary data.

The serial correlation of observations col-
lected during simulation experiments is more
effectively exploited in the **method based
on spectral analysis**, first proposed by
Fishman [1967] and Fishman and Kiviat
[1967]. The method assumes that obser-
vations represent the stationary and au-
tocorrelated sequence $X_1, X_2, \cdots, X_n$, and
shifts their analysis into the frequency do-
main by applying a Fourier transformation
to the autocorrelation function $\{R(k)\}, k =
0, 1, 2, \cdots$, yielding the *spectral density func-
tion*

$$p_x(f) = R(0) + 2\sum_{j=1}^{\infty}R(j)cos(2\Pi fj), \quad (41)$$

for $-\infty \le f \le +\infty$ [Brilliinger 1981; Jenk-
ins and watts 1968].

Note that because of the randomness of the collected observations, the spectral density function is a random function too. Comparing Equations (41) and (9) one can see that, for sufficiently large n,

$$\sigma^2[\overline{X}(n)] \cong \frac{p_x(0)}{n}. \qquad (42)$$

Thus, the estimator of $\sigma^2[\overline{X}(n)]$ can be obtained from an estimator $p_x(f)$ at $f = 0$. Several techniques have been proposed for obtaining good estimators of the spectral density function $p_x(f)$. Most of them follow classical techniques of spectral estimation based on the concept of *spectral windows* (special weighting functions introduced for lowering the final bias of the estimators) [Fishman 1973a, 1987a; Jenkins and Watts 1968; Marks 1981]. It can been shown that applying a modification of the so-called Bartlett window gives an estimator of the variance from Equation (42) equivalent to that from the overlapping batch means, [Damerdji 1987; Welch 1987]. The best results in the sense of coverage were reported by applying the so-called Tukey–Hanning window [Jenkins 1968; Law and Kelton 1984]. Using this approach, one can determine the confidence interval of μ_x

$$\overline{X}(n) \pm t_{k,1-\alpha/2}\hat{\sigma}_{sp}[\overline{X}(n)], \qquad (43)$$

assuming

$$\hat{\sigma}^2_{sp}[\overline{X}(n)] = \frac{\hat{p}_x(0)}{n}, \qquad (44)$$

and $t_{k,1-\alpha/2}$ as the upper $(1-\alpha/2)$ critical point of the t-distribution with k degrees of freedom. The value of f_k depends here on the ratio of n/k_{max}, where k_{max} is the value of the upper lag considered in the autocorrelation function $R(k)$[Bratley et al. 1983,p. 97;

Fishman 1973a]. This approach can sometimes produce quite accurate final results (see Law and Kelton[1984]), but it cannot be regarded as a good candidate for a more user friendly implementation because of its rather sophisticated nature. In particular there is no definitive method for choosing the parameter $_k$ [Bratley 1983, p.97; Fishman 1978a, p. 265; Law and Kelton 1984].

The usefulness of spectral windows in reducing the bias of the estimate $\hat{p}_x(0)$ has been questioned in Duket and Pritsker [1978], Hiedelberger an dWelch [1981a, 1981b], Wahba [1980]. The last three papers propose estimating $p_x(0)$ from the *periodogram* $\{\pi(j/n)\}, j = 0, 1, \cdots$, of the sequence of observations $x_1, x_2, \cdots, x_n$. This periodogram is a function of the discrete Fourier transforms $\{A_x(j)\}$ of the observations $x_1, x_2, \cdots, x_n$, namely,

$$\Pi(\frac{j}{n}) = \frac{|A_x(j)|^2}{n} \qquad (45)$$

and

$$A_x(j) = \sum_{s=1}^{n} x_s exp\left[-\frac{2\pi i(s-1)j}{n}\right], \quad (46)$$

where[3] $i = \sqrt{-1}$. It can be shown that for $0 < j < n/2$

$$p_x\left(\frac{j}{n}\right) \approx E\left[\Pi\left(\frac{j}{n}\right)\right]. \qquad (47)$$

To find an unbiased estimate of $p_x(0)$ the periodogram is transformed into a smoother function, namely, into the logarithm of the averaged periodogram

$$L(f_j)$$

$$= log\left\{\frac{1}{2}\left[\Pi\left(\frac{2j-1}{n}\right) + \Pi\left(\frac{2j}{n}\right)\right]\right\} \quad (48)$$

---

[3] The symbol i has this special meaning only in Equation 46

for $f_j = (4j - 1)/n$. Next, this smoother function (but still not the smoothest one) is approximated by a polynomial to get its value at zero. The whole approach is discussed in detail in Hiedelberger and Welch [1981a] together with a method for calculating $_k$, the number of degrees of freedom needed in Equation (43); see also the Appendix B. Despite a number of approximations involved, the method produces quite accurate results, in particular in terms of coverage.

Heidelberger and Welch [1981a] proposed a sequential version of their method that uses a limited number of (aggregated) output data points instead of a growing number of individual observations, since, as they show, both individual observations and their batch means (of arbitrary size) can be used in the variance analysis. Namely, if n observations are grouped into b batches[4] of m observations each, then for $n = bm$

$$\frac{p_x(0)}{n} = \frac{p_{\overline{X}(m)}(0)}{b} \qquad (49)$$

where, for $-\infty \leq f \leq +\infty$,

$$p_{\overline{X}(m)}(f) = R(0, m)$$

$$+2 \sum_{j=1}^{\infty} R(j, m) cos(2\Pi j f) \qquad (50)$$

is the spectral density function of the autocorrelation function $R(k, m)(k = 0, 1, 2, \cdots)$ of the batch means; see Equation (22). This insensitivity of the method to batching the observations allows the batch size to be increased dynamically (starting from $m = 1$), keeping in memory only a limited number of the batch means. A special batching/rebatching procedure is presented in Heidelberger and Welch [1981a, 81b]. It appears to be an efficient way of limiting the required memory space. A modified version of this method is presented in Section 3. The method has demonstrated a good coverage of confidence intervals in various applications, even if data are collected asynchronously in time [Asgarkhani and Pawlikowski 1989], despite claims based on the basic assumption of discrete Fourier transformation, that it should be applied only in simulation experiments in which observations are collected at equally spaced time intervals [Bratley et al. 1983, p. 96].

Another approach for estimating the variance of correlated observations collected during a single simulation run is applied in the **method based on autoregressive representation** developed by Fishman [1971, 1973a, 1978a]. Again it is assumed that after having decided about observations gathered during the initial transient period, the analysed sequence of observations $x_1, x_2, \cdots, x_n$ represents a stationary process. The main assumption of the method is that such a sequence of originally correlated observations can be transformed into a sequence of independent and identically distributed random variables, called its *autoregressive representation*. The autoregressive representation $y_{q+1}, y_{q+2}, \cdots, y_n$ of order q is defined by the transformation

$$y_i = \sum_{k=0}^{q} c_k (x_{i-k} - \mu_x), \qquad (51)$$

for $i = q + 1, q + 2, \cdots, n$. It can be shown, [Fishman 1978], that

$$E[y_i] = 0, \qquad (52)$$

---

[4] The symbol b, instead of k_b, is used here to emphasize the insensetivity of the method of the number of batches.

$E[\overline{Y}(n-q)]$

$$= \sum_{i=q+1}^{n} \frac{y_i}{(n-q)} \approx C[\overline{X}(n) - \mu_x], \qquad (53)$$

and

$$\sigma^2[\overline{Y}(n-q)] \approx C^2 \sigma_{ar}^2[\overline{X}(n)], \qquad (54)$$

where $C = c_0 + c_1 + \ldots + c_q, c_0 = 1$. $\sigma^2[\overline{Y}(n-q)]$, as the variance of the mean of i.i.d. random variables, can be estimated using Equation (4), provided the coefficients $q, c_1, c_2, \cdots, c_q$ are known. The correct autoregressive order q can be determined by examining the convergance of the distribution of a test statistic to an F distribution (also known as the Fisher distribution, or the Snedecor distribution, or the variance-ratio distribution)[Bratley 1983; Hannan 1970, p. 336] or to a χ^2 distribution [Fishman 1978a, p. 251; Hannan 1970, p. 336]. Having selected q, the estimates of the coefficients of $c_1, c_2, \cdots, c_q$ can be found from a set of q linear equations of the form

$$\sum_{i=1}^{q} \hat{c}_i \hat{R}(k-i) = -\hat{R}(k), \qquad (55)$$

for $k = 1, 2, \cdots, q$, where $\hat{R}(k)$ is the estimate of the lag k autocovariance from the sequence of original observations $x_1, x_2, \cdots, x_n$ [Fishman 1978, p. 249]. Next, having determined $\hat{\sigma}^2[\overline{Y}(n-q)]$, one can easily find the estimate of the variance of $\overline{X}(n)$, since from Equation (54),

$$\hat{\sigma}_{ar}^2[\overline{X}(n)] = \frac{\hat{\sigma}^2[\overline{Y}(n-q)]}{C^2}. \qquad (56)$$

Finally, assuming that the variable $\sqrt{n}(\overline{X}(n) - \mu_x)/\hat{\sigma}_{ar}[\overline{X}(n)]$ is governed by the t-distribution with

$$k = n\frac{C}{2}\left[\sum_{j=0}^{q}(q-2j)\hat{c}_j\right]^{-1} \qquad (57)$$

degrees of freedom (see arguments given in Fishman [1978a, p. 252]), the resulting confidence interval of μ_x is determined as

$$\overline{X}(n) \pm t_{k,1-\alpha/2}\hat{\sigma}_{ar}[\overline{X}(n)]. \qquad (58)$$

The main restriction of the last method seems to be the required existence of an autoregressive representation of the simulated process. Results of empirical studies of the method's efficiency published in Fishman [1971] were not very encouraging, since frequently below 80%. These results were, however, achieved in short simulation runs. Andrews and Schriber, in their studies of the autoregressive method reported in Andrews and Schriber [1978] and Schriber and Andrews [1979, 1981], observed a significant variability in the average widths of confidence intervals in simulation experiments. Law and Kelton [1984], after comparative studies of different fixed-size methods of data analysis, also that the autoregressive approach does not offer better results than other, computationally simpler methods of data analysis. And, in contrast to both the method of batch means and the method of spectral analysis, the improvement of the final coverage when increasing the number of collected observations was very slow. Continuous execution of the test for determining the autoregressive order q and solving the sets of equations for determining the coefficients $c_1, c_2, \cdots, c_q$ can be time consuming in a sequential version, especially if longer sequences of observations have to be collected.

The **method of standardized time series**, originally proposed by Schruben [1983, 1985], relies on the convergence of standardized random processes to a Wiener

random process with independent incre-
ments, also known as a Brownian bridge
process. It is an application of the the-
ory of dependent random processes [Billings-
ley 1968, Chaps. 20 and 21] and its func-
tional central limit theorem, which is a gen-
eralization of the (scalar) central limit theo-
rem presented in Appendix A. According to
this approach, an analysed sequence of ob-
servations is first divided into subsequences
(batches) of observations, and each of them
is transformed into its standard form re-
quired by the functional central limit theo-
rem. Next, various functions of the trans-
formed sequence can be analysed to con-
struct the confidence interval of $\overline{X}(n)$. The
method requires that the analysed process
be stationary; thus, initial observations rep-
resenting its nonstationary warm-up period
should be discarded before the sequence of
n remaining observations is divided into
k_b nonoverlapping batches, each of size
m. The ith batch, containing the observa-
tions $x_{(i-1)m+1}, x_{(i-1)m+2}, \cdots, x_{(i-1)m+m}, i =
1, 2, \cdots, k_b$, is transformed into the stan-
dardized process $T_i(t), \leq t \leq 1$, where

$$T_i(t) = \frac{\lfloor mt \rfloor [\overline{X}_i(m) - \overline{X}_i(\lfloor mt \rfloor)]}{\hat{\sigma}[\overline{X}_i(m)]\sqrt{m}} \quad (59)$$

for $0 < t \leq 1$ ($\lfloor a \rfloor$ denotes the greatest in-
teger not greater than a), and $T_i(0) = 0$. In
this formula,

$$\overline{X}_i(k) = \frac{1}{k} \sum_{j=1}^{k} x_{(i-1)m+j} \quad (60)$$

is the cumulative average of the first k ob-
servations in the ith batch, and $\hat{\sigma}^2[\overline{X}_i(m)]$
is the variance estimator of the ith batch
mean. The functional central limit theo-
rem says that in the limit, as $m \to \infty$, any
standardized process $T_i(t), 0 \leq t \leq 1$, for

$i = 1, 2, cdots$, becomes the Brownian bridge
process (a mathematical model of Brown-
ian motion on the $[0, 1]$ interval [Billingsley
1968]). This fact has been used by Schruben
[1983], who proposed to use two functions of
$T_i(t)$ to estimate the variance of $\overline{X}(n)$: (i)
the maximum of $T_i(t), 0 \leq t \leq 1$, and (ii)
the sum of $T_i(k/m)$, from $k = 1 tom$. The
former function is used in the *maximum es-
timator* of $\sigma^2[\overline{X}(n)]$,
$\hat{\sigma}^2_{max}[\overline{X}(n)]$

$$= \sum_{j=1}^{k_b} \frac{k_{i,max}}{3k_b^2(m - k_{i,max})} \quad (61)$$

$[\overline{X}_i(m) - \overline{X}_i(k_{i,max})]^2,$

where

$$k_{i,max} = min\{k : k[\overline{X}_i(m) - \overline{X}_i(k)]$$

$$\geq j[\overline{X}_i(m) - \overline{X}_i(j)]; \quad (62)$$

$$j = 1, 2, \cdots, m\}$$

is the location of the (first if ties occur) max-
imum of $T_i(k/m)$ as a function of k, Golds-
man and Schruben [1984], and gives the con-
fidence interval of μ_x,

$$\overline{X}(n) \pm t_{3k_b, 1-\alpha/2}\hat{\sigma}_{m}ax[\overline{X}(n)], \quad (63)$$

where the constant $t_{3k_b, 1-\alpha/2}$ is from the t-
distribution with $3k_b$ degrees of freedom.
The latter function is used in the *area es-
timator*

$$\hat{\sigma}^2_A[\overline{X}(n)] = \frac{12}{(m^2 - 1)n^2} \sum_{i=1}^{k_b} A_i^2, \quad (64)$$

where

$$A_i = \sum_{j=1}^{m} (j - 0.5(m + 1))x_{(i-1)m+j}; \quad (65)$$

[Song 1988]. It gives the confidence interval

$$\overline{X}(n) \pm t_{k_b, 1-\alpha/2} \hat{\sigma}_A [\overline{X}(n)]. \qquad (66)$$

Goldsman and Schruben [1984] showed that the former estimator is asymptotically superior to the later one in the sense that as $m \to \infty$, it produces narrower and more stable confidence intervals. On the other hand, it can perform poorly when batches are short. They also showed that if m is large, the method of standardized time series is superior to the method of nonoverlapping batch means. Another positive feature of this method is that despite the sophisticated statistical techniques involved, the estimators have simple numerical forms. For a further improvement of their quality, Damerdji [1987] proposed to use overlapping batches of observations as is done in the method of overlapping batch means, and Foley and Goldsman [1988] have modified the Equation (59), introducing the so-called orthonormally weighted standardized time series, which gives asymptotically narrower and more stable confidence intervals than the original one. Unfortunately, no simple rule for selecting the batch size is available, and only a few practical implementations of the method of standardized time series have been reported. Theoretical analysis of asymptotic cases shows that the method usually requires longer batches than the method of nonoverlapping or overlapping batch means, [Song 1988].

The estimators of variance that have been presented in this section require various, more or less complex ways of data collection. They also have their own statistical strengths and weaknesses. This has led to the idea of finding a robust method of simulation output data analysis by combining different estimators of variance $\sigma^2[\overline{X}(n)]$

into a composite estimator, since, by theoretical arguments, combinations of independent variance estimators should have a smaller variance and, in consequence, give better coverage of confidence intervals than its components. Such an effect has been achieved by Schruben [1983], with a linear combination of the area estimator (64), or the maximum estimator (61), and the estimator calculated from nonoverlapping batch means; see results of the quality analysis in [Goldsman and Schruben 1984; Goldsman et al. 1986; Song 1988]. The first of these two combined estimators has been applied in a sequential procedure for the analysis of simulation output data described in Duersch and Schruben [1986]. More sophisticated linear combinations of estimators are discussed in Schmeiser and Song [1987] and Song and Schmeiser [1988].

Research in the area of variance estimation continues. The diversity of existing methods of data collection and analysis, and the variance estimators which they use, requires a more thorough comparison of their quality. Some results of comparative studies can be found in Glynn and Iglehart [1985], Goldsman and Meketon [1985, 1989], Goldsman and Schruben [1984], Goldsman et al. [1986], Kang and Goldsman [1985], and Song [1988]. Readers interested in current developments in this area are encouraged to browse through recent publications in, for example, "Operations Research," "Management Science," and the annual *Proceedings of the Winter Simulation Conference*.

2   THE PROBLEM OF INITIALIZATION

It is well known that just after initialization any queueing process with nondeterministic, random streams of arrivals and/or random service times is in a transient phase,

during which its (stochastic) characteristics vary with time. This is caused by the fact that, like any (stochastic) dynamic system, such queueing systems or networks initially "move" along nonstationary trajectories. After a period of time, the system approaches its statistical equilibrium on a stationary trajectory if the system is stable, or remains permanently on a nonstationary trajectory if the system is unstable[5]. Note that in practice only queuing systems with infinite populations of customers and unlimited queue capacities can never enter a stationary trajectory and this happens if the average request for service is equal to or greater than the average supply of service; that is, if

$$\lambda \geq c\mu_s, \qquad (67)$$

where λ is the mean arrival rate, $1/\mu_s$ is the mean service time, and c is the number of service facilities. In such a case, the queues will eventually increase in length with time and the system becomes permanently congested. On the other hand, queueing systems with limited queue capacities always reach an (inner) statistical equilibrium, even if the system's load expressed by the traffic intensity $\rho = \lambda/c\mu_s$ is much greater than 1. In such a case, internally stationary queuing systems are in the nonstationary environment of streams of rejected customers. Of course, output data collected during transient periods do not characterise steady state behaviour of simulated systems, and so they can cause quite significant deviation of the final "steady state" results from their true values. Although it seems quite natural that the deletion of untypical initial observations should result in better steady state estimators, the problem to delete or not to delete is a perennial dilemma of stochastic simula-

tion practice. Each of these two alternatives has its advocates. The answer depends on the assumed measure of goodness and the resource limitations of simulation experiments (the maximum possible number of recorded observations). The influence that the initial transient data can have on the final results is a function of the strength of the autocorrelation of collected observations. With no restrictions imposed on the length of the simulation run, this influence can be arbitrarily weakened by running the simulating program sufficiently longer. But in most practical situations simulation experiments are more or less restricted in time, and that time can be more or less effectively used to calculate estimators. If all initial output data are retained, the bias of the point estimator $\overline{X}(n)$ is greater than if they were deleted.

Contrary opinions on the usefulness of deletion are caused by the fact that it increases the variance of the point estimator [Fishman 1972, 1973a, Sec. 10.3; Turnquist and Sussman 1977] and, in effect, can increase its MSE, see Equation (8). Let us note that an increase of the variance can be compensated for by applying one of the variance reduction techniques. Deletion of initial observations seems to be justified if the variance of the estimator is smaller than the squared bias and/or if observations are strongly correlated (the initial conditions have a longer effect on the evolution of the system in time). On the other hand, Blomqvist [1970] showed that for long run simulations of GI/G/1 queuing systems the minimum MSE of the mean delay usually occurs for the truncation point $n_o = 0$, which supports the thesis that no initial observations need to be deleted. Results of experiments conducted by Turnquist [1977],

---

[5] We could also mention the periodic limiting behavior of D/D/c queuing systems and their networks.

and Wilson and Pritsker [1978b] provide the same argument.

The usefulness/uselessness of data deletion also depends on methods used for data collection and analysis. Independent replications give much more "contaminated" data than methods of data collection based on single runs, since each replication begins with a new initial transient period. Consequently data deletion seems to be more crucial for plurality of transient periods than for just one transient period in one long run. In an example discussed in Kelton and Law [1983] the estimator of mean delay in an M/M/1 queue obtained from replications of 500 observations, without initial deletions, was biased -43.2%, for $\rho = 0.95$. In the case of higher accuracy requirements, a significant bias of estimators will normally increase their chances of being outside the theoretical confidence intervals, thus it will decrease the coverage of confidence intervals. Law [1983] Law and Kelton [1984] analysed the influence of initial data deletion on the coverage of the final results in the case of the method of independent replications and stated a clear improvement of the actual coverage (experimental values of confidence levels) to levels near nominal theoretical values $(1 - \alpha)$, without unduly widening confidence intervals, especially if replications were not too long and/or not too many observations were deleted. In methods of data analysis based on single runs and an assumption that the observed process is stationary, deletion of data from the initial nonstationary period improves approximate stationarity of the remaining process.

The nature of the convergence of simulated processes to steady state depends on many factors; the initial conditions of simulation are one of them. Conway [1963] advised a careful selection of starting states (typical ones for steady state of the simulated process) to shorten the duration of the initial transient phase. Since then many trials have been undertaken to determine the optimal initial conditions in the sense that they would cause the weakest influence of the transient phase on the steady-state results, but ambiguous conclusions have been reached. Madansky [1976] proved that the MSE of the mean queue length in simulation studies of M/M/1 queuing systems (without data deletion) can reach its minimum value if they are initialized as empty and idle, that is, in their modal states. Wilson and Pritsker [1978b], having examined a slightly broader class of queuing processes, concluded that the optimal (in the MSE sense) initial state is the most likely state in statistical equilibrium (the mode of the steady state distribution) if it differs from the empty and idle state. Moreover they found that a judicious selection of initial conditions can be more effective than the deletion of initial data. Similar conclusions were reached by Donnelly and Shannon [1981] after a more methodical investigation. Following this line of reasoning, Murray and Kelton [1988] proposed to use the method of independent replications with randomly selected initial states taken from an initial state distribution. Such a distribution could be constructed during the first k_o replications (the pilot phase of the simulation experiment) and used during the remaining $k_b - k_0$ replications (the productive phase). Results reported in Murray and Kelton [1988] show that such an approach can be effective both for reducing the bias of point estimates and for increasing the coverage of confidence intervals, without unduly increasing the variance or mean square error of estimates.

On the other hand, Kelton and Law showed experimentally by investigating

queuing systems with exponential and Er-
langian distributions of interarrival and ser-
vice times that the shortest transient periods
occur if the simulated processes start from
states slightly larger than their steady-state
means [Kelton 1985; Kelton and Law 1985].
For example, in the M/M/1 queue the mean
delay reaches its steady state in the shortest
time (with accuracy $\pm 0.01, \rho = 0.9$) if the
initial queue length is 15, while the steady-
state mean queue length equals 9. These
results were theoretically justified by Abate
and Whitt [1987b], who indicated that the
optimal initial state when the mean value is
estimated is about one-and-a-half times the
steady-state mean. It was also shown that
starting from a state much larger than the
mean can result in a very long transient pe-
riod. Thus, because in real situations the
steady-state mean is unknown, it is much
safer to initialise systems as empty and idle,
particularly if the bias of an estimator con-
cerns us more than its MSE.

Having decided to discard data collected
during transient periods, we face the next
problem: how long such periods last. In
simulation practice we can encounter both
very short initial transient effects and tran-
sient effects that are spread over tens of
thousands of observations [Heidelberger and
Welch 1983].

2.1   The Duration Of The Initial Transient Period

The problem of determining the duration
of the initial transient period in simulation
runs appears to be complicated, even if we
restrict ourselves to estimators of means val-
ues only. The first rules of thumb were pro-
posed by Conway [1963] and Tocher [1963].
Conway suggested the following rule:

*R1.   In a time series of observations
$x_1, x_2, cdots, x_n$ the initial transient lasts un-
til the first of the series is neither the maxi-
mum nor minimum of the rest.*

This rule of thumb, associating the be-
ginning of steady state with the occurrence
of the first "typical" observation, appears to
give a poor approximation of the duration
of the initial transient. As was shown in Ga-
farian et al. [1978], using this rule we can
significantly overestimate the length of the
initial transient for small ρ and underesti-
mate it for high ρ; see [Wilson and Pritsker
1978b].

The performance of a system can be re-
garded as a cyclic evolution of system's basic
operations. For this reason, Tocher [1963, p.
176] suggested this rule:

*R2. The initial transient period is over if the
longest cycle distinguished in the behaviour
of the simulated system has been executed at
least three or four times.*

No results concerning the effectiveness of
this rule are available. The duration of the
initial transient period is also analysed in
the queueing theory. It has been shown that
the rate at which the mean queue lengths or
the mean delays tend to their steady state
is, after some period of time, dominated
by a term of the form $\exp(-t/\tau)$, where τ
is called the *relaxation time* of the queue.
Thus, the constant τ may be used for speci-
fying an upper bound on the length of time
after which the influence of the initial state
is negligible. For example, one can conclude
this rule:

R3. The initial transient period is over after the time $t_\beta = -\tau \ln \beta$, where β is the permissible relative residue of the initial state, $0 < \beta < 1$.

Thus, assuming $\beta \leq 0.02$, at $t = 4\tau$ we find that the queue characteristics are within 2% of their steady-state values; or in other words, output data collected from that point of time should be biased by initial states by less than 2%. The analysis of relaxation times was initiated by Morse [1955], who considered the correlation function of the M/M/1 queue length. Cohen [1982] analysed transient distributions of queue lengths and determined the relaxation time for GI/G/1 queueing systems. These appear to be from 9 to 2 times greater than Morse's results for M/M/1 systems, as ρ changes from 0.1 to 1.0. This diversity of results has stimulated search for approximate formulas for relaxation times, such as Newell's result for GI/G/1 under heavy traffic queues [Newell 1971] and results for Markovian queueing systems obtained by Odoni and Roth [1983]. The latter, having studied various Markovian systems, proposed to approximate the relaxation times by

$$\tau = \frac{C_a^2 + C_s^2}{2.8\mu_s(1 - \sqrt{\rho})^2}, \qquad (68)$$

where C_a^2 and C_s^2 are the coefficients of variation for the interarrival and service times, respectively, and $1/\mu_s$ is the mean service time. The usefulness of the last formula in simulation has been studied in the case of M/M/1 and M/E_k/1 queueing systems, [Roth 1985; Roth and Rutan 1985]. More detailed analysis of the transient behavior of some stochastic processes has been reported by Abate and Whitt [1987a, 1987b, 1987c,

1988], who analysed the relaxation times in the M/M/1 queues and Brownian motion processses (the latter are used to approximate queueing processes in a heavy traffic scenario [Kleinrock 1976]). All these results show that more heavily loaded systems tend more slowly to their statistical equilibrium. Abate and Whitt have also proved that higher moments of queue parameters have longer relaxations times than corresponding lower moments [Abate and Whitt 1987a]. Thus, mean values tend to a steady state faster than, for example, variances.

Relaxation times have also been analysed theoretically for some simple queueing networks; for example, Blanc [1985a] analysed the relaxation time in an open network of K service centers with a Poisson arrival stream, an unlimited number of servers at each center, general distribution of service times, and a homogenous transition matrix. He showed that the relaxation time in such a network has an upper limit, namely,

$$\tau \leq \frac{K}{\mu_s}. \qquad (69)$$

Equality occurs for tandem connections of queueing systems. The case of two queueing systems in series has been analysed in more detail in Blanc [1985b].

In addition, a conjectural relaxation time for Jacksonian queueing networks with K single server centers has been proposed. For more complex queueing networks the relaxation times have not yet been theoretically determined. But the usefulness of even known formulas for relaxation times can be questioned in simulation studies. They can be used only as first approximations of the duration of simulated initial transients, since it has been shown that estimators of the mean values from simulation tend to their steady state more slowly than exponentially;

for example, Anderson [1985] showed that in queueing systems with limited queue capacities the rate at which the estimator of mean queue length tends to its steady state eventually becomes inversely proportional to time. It has also been shown that the standard deviation of estimators converges even more slowly, namely, in inverse proportion to the square root of time [Anderson 1985; Fishman 1967]. Both of these facts have found application in various heuristic rules proposed for determining the duration of the initial transient period.

Studying the convergence of a moving average of output data to determine a possible end of the initial transient period is attributable to Gordon [1969, p. 285] and Emshoff and Sisson [1970, Sec. 8.2]; see Solomon [1983, p. 195]. The simplest way would be to find an instant of time at which the running mean $\overline{X}(n)$ [Equation (1)] approaches a constant level with a given accuracy $\delta, \delta \leq 0$. Thus, we can assume the following rules:

R4. in a time series of observations $x1, x2, \cdots, xi, \cdots,$ *the initial transient period is over after* n_o *observations if k consecutive values of the running mean $\overline{X}(i)$ recorded after the observation n_o differ less then $100\delta\%$ from $\overline{X}(n-o+k)$, that is, for all* $i, n_o < i \leq n_o + k$,

$$\frac{|\overline{X}(n_o + k) - \overline{X}(i)|}{|\overline{X}(n_o + k)|} < \delta. \qquad (70)$$

The stabilization of $\overline{X}(n)$ should be tested over a sufficiently long sequence of observations, so the parameter k should be large (in the statistical sense); that is, $k \leq 30$. The above rule has two weaknesses. First, as has been indicated by Conway [1963], accumulative statistics such as running means usually stabilize very slowly with time, and so usually give overestimated values of n_o. Second, fluctuations of the running mean $\overline{X}(n)$, calculated over data collected during a single simulation run, can continue for a long time. For these reasons the above rule and its various modifications are usually used with the method of replications, and the inequality (70) is applied to the running mean after having additionally smoothed it by averaging over replications. Despite this, the resulting length of the initial transient period is still usually overestimated [Gafarian et al. 1978; Roth 1985; Roth and Rutan 1985; Wilson and Pritsker 1978a]. Welch proposed a special technique for smoothing running mean that uses the concept of a moving window within which mean values over replications are additionally averaged, producing a smoother (but still highly correlated) process [Welch 1983, p. 293]. No results on the effectiveness of this technique have been published.

Another rule of thumb can be based on the supposition that in steady state typical observations are evenly dispersed around the mean value. For example, Fishman [1973b, p. 275] proposed the following:

R5. The initial transient period is over after n_o *observations if the time series* $x_1, x_2, \cdots, x_{n_o}$ *crosses the mean $\overline{X}(n_o)k$ times.*

This rule is sensitive to the value of k [Gafarian et al. 1978]. Too large a value will usually lead to an overestimated value of n_o regardless of system's utilization, whereas too small a value can result in an underestimated n_o in more heavily loaded systems. In Gafarian et al. [1978], $k = 25$ was recommended for M/M/1/∞ queueing systems, whereas in Wilson and Pritsker [1978b] $k = 7$ was chosen for the M/M/1/15 system. The

system-dependent selection of the parameter k in the rule R5 seems to be too arduous for potential users.

Yet another approach, which Solomon [1983] attributes to Emshoff and Sisson [1970], is based on the χ^2 goodness-of-fit test applied for selecting a time from which the numbers of observations below and above the running mean are equal (in the statistical sense). According to this test, the sequence of observations should be partitioned into batches of at least $m_o = 10$ observations each (Solomon selected $m_o = 30$).Then one can conclude the following rule:

R6. *In a time series of observations* $x_1, x_2, \cdots , x_n$ *the initial transient is over after* n_o *observations if the* χ^2 *goodness-of-fit test confirms that in the batch of observations* $x_{n_o+1}, x_{n_o+2}, \cdots , x_{n_o+m_o}$ *following the observation* n_o *the numbers of observations above and below the running mean* $\overline{X}(n_o)$ *are about the same.*

Rule R6 seems to be quite simple and independent of any system-related parameter. No results are available on its effectiveness and relationship to other criteria. For reducing the fluctuations of analysed sequences and saving memory space if long transient periods are expected, batches of individual observations can be replaced by their mean values. Note that such batches must be introduced before the procedures for selecting the size of uncorrelated batch means discussed in Section 1 can be applied. There are no established rules in this case for selecting the batch size. If a statistical test is used to help decide about the length of the initial transient, we should follow the requirements of the test or use statistically large batches, which usually means taking $m_o > 30$. Otherwise the only recommendation is to select a batch size which gives the desired data re-

duction while retaining the stabilizing trend of the original sequence. After the batch size m_o is selected, the sequence of batch means $\overline{X}_1(m_o), \overline{X}_2(m_o), \cdots$, can be tested in a similar way to the sequence of the original observations. For example, Wilson and Pritsker [1978a] formulated the following rule, which they attributed to Schriber [1974] (Solomon [1983, p. 195] attributed it to Emshoff and Sisson [1970]):

R7. *In a time series of batch means* $\overline{X}_1(m_o), \overline{X}_2(m_o), \cdots$, *the initial transient is over after* b_o *batches, that is, after* $n_o = b_o m_o$ *observations, if the* k *most recent batch means all fall within an interval of width* δ_1 *, that is, if*

$$|\overline{X}_{b_o-i}(m_o) - \overline{X}_{b_o-j}(m_o)| \le \delta_1, \qquad (71)$$

for $0 \; i < k - 1, 0 < j < k - 1$.

This rule, like rule R4, is sensitive to the value of the parameter k, which should depend on the variability of the observed process. A small value for k, for example $k = 2$, as was assumed in Wilson and Pritsker [1978a] and Solomon [1983, p. 196], can lead to an underestimation of n_o, since the difference between averages, having dropped only k times below δ_1, can easily rise again to an unacceptable level, as in an example considered in Solomon [1983, p. 197].

For further data reduction and additional smoothing of the tested sequence, Kelton and Law [1983] proposed applying such a batching technique in connection with the method of independent replications; see Section 1. Namely, they batched the sequence of mean observations (means over replications) and then analysed the sequence of the means of these batches, assuming rule R8:

R8. *In a time series* $\overline{X}_1(m_o), \overline{X}_2(m_o), \cdots$,

the initial transient period is over after the batch b_o, that is, after $n_o = m_o b_o$ observations, if the sequence of the batch means after the batch b_o can be approximated by a straight line with zero slope.

Rule R8 can be applied only in the case of monotonic convergence to steady state but, as was proved by Kiefer and Wolfowitz [1955], in any stable, initially empty-and-idle GI/G/c queueing system the mean delay-in-queue grows monotonically in time. Kelton and Law [1983] proposed testing the slope of the regression line backwards after collecting an assumed number of observations. The test for zero slope is over a fixed number of batch means (if zero slope is confirmed, the test is repeated over an earlier sequence of batch means to find whether the initial transient period had expired earlier). If the test fails at the beginning, a new check point is chosen after gathering further output data from the simulation. Note that this requires that the process of collecting new observations in all previously stopped replications be continued again. Because correlations between batch means can still be significant, they are approximated by a straight line using a generalized least-squares procedure proposed by Amemiya [1973], which allows for autocorrelation of the analysed data. For additional saving of memory space, the number of batches could be kept constant by allowing the size of batches to grow when simulation runs are continued. The procedure implementing rule R8 appears to be quite effective, especially in lowering the MSE of estimators [Roth 1985; and Roth and Rutan 1985].

Rules R4 - R8 are based on the convergence of the mean of observations to its steady-state value. Other criteria of convergence are also possible. For example, because the variance of the mean of observations taken from a stationary process is approximately inversely proportional to the number of observations [Fishman 1973a, p. 281; Gafarian et al. 1978], that is,

$$\hat{\sigma}^2[\overline{X}(n)] = \frac{C_1}{n} + o(\frac{1}{n}), \qquad (72)$$

[cf. Equation (9)] where C_1 is a positive constant and n is the number of observations, Gordon [1969] proposed rule R9:

R9. In a time series $x_1, x_2, \cdots, x_n$ the initial transient is over after the observation n_o if the graph $(\log n, \log \hat{sigma}[\overline{X}(n)])$, becomes approximately linear with slope -0.5 from this observation on.

To smooth variations of the analysed curve, Gordon [1969] proposed analysing the variance of the mean of observations averaged over a number of replications. This rule was analysed in Gafarian et al. [1978] and Wilson and Pritsker [1978a], using Equation (4) to calculate $\hat{\sigma}^2[\overline{X}(n)]$, thus rejecting existing correlations between observations. In this case the rule R9 can give an overestimated value of n_o. No results have been published on the effectiveness of this rule when more accurate estimators of $\sigma^2[\overline{X}(n)]$ are applied.

Fishman [1971, p. 29] proposed equating the variance of the mean of autocorrelated observations with the variance of the mean of a hypothetical sequence of independent observations to find the number of collected (autocorrelated) observations equivalent, in the above sense, to one independent (hypothetical) observation. After some simplification we get the following rule:

R10.   In a time series of observations $x_1, x_2, \cdots, x_n, \cdots$, the initial transient is

over after

$$n_o = 2 \sum_{k=1}^{n-1} (1 - \frac{k}{n}) \frac{\hat{R}(k)}{\hat{R}(0)} \qquad (73)$$

observations, where $\hat{R}(k)$ is the estimator of the autocorellation of the lag $k, 0 \le k \le n-1$; see Equation (40)

The sequence of observations collected after the observation n_o should be (approximately) independent of the initial conditions. The autocovariance estimators $\hat{R}(k)$ should be analysed with caution; see the comments after Equation (22). Comparing this with the results given for example in Roth and Rutan [1985], one can state that Rule R10 usually gives underestimated values of n_o; no exhaustive comparisons of this rule with other rules are available.

As an example of more sophisticated truncation rules let us mention the rule proposed in Beall [1982] based on the analysis of the convergence of the distribution functions to their steady-state forms. The rule has been analysed in the context of autoregressive moving average (ARMA) processes [Brockwell and Davis 1987, Chap. 3]. All the above-mentioned rules proposed for determining the length of the initial transient periods are either quite elaborate and, as such, do not ensure an accurate control of the initialization bias, or can determine precisely the length of the initial transient period but only for restricted classes of simulated processes and/or by applying sophisticated techniques to collect and analyze the output data. This can unnecessarily lengthen the time of simulation experiments, especially if the required accuracy is tested sequentially; cf. the rule R8. Usually the effectiveness of these rules also strongly depends on the specific parameters they use, but little or no guidance is available on how to select val-

ues for these parameters. Some of the rules have been implemented as built-in options offered in simulation packages such as GPSS, SLAM and SIMSCRIPT II.5 [Law and Kelton 1982a]. Thus, potential users should be aware of their limitations.

A promising approach for detecting the expiration of the initial transient period is offered by statistical stationarity tests based on the theory of dependent stochastic processes developed by Billingsley [1968]. According to this approach, we have the following rule: R11:

R11. The initial transient data have been removed from a given sequence of observations if the (standardized) sequence determined over the remaining observations behaves in a way consistent with a standard (stationary) stochastic process.

Such tests were first invented by Schruben [1982] and Schruben et al. [1983] as a special application of the same theory which is applied in the method of standardized time series for generating confidence intervals; see Section 1. These tests are based on the high sensitivity of the sequence of partial sums

$$S_k = \overline{X}(n) - \overline{X}(k), \qquad (74)$$

$k = 0, 1, 2, cdots$, and $S_0 = S_n = 0$, to the presence of initialization bias in $\overline{X}(n)$; $\overline{X}(n)$ and $\overline{X}(k)$ are means over n and k first observations, respectively; see Equation (1). Following this phenomenon, tests proposed in Schruben [1982] and Schruben et al. [1983] analyse the convergence of a standardized sequence $\{T(t)\}, 0 \le t \le 1$, to the Brownian bridge process with zero mean and variance equal 1. The sequence $\{T(t)\}$ is the standardized sequence of the partial sums S_k,

namely,

$$T(t) = \frac{\lfloor nt \rfloor S_{\lfloor nt \rfloor}}{\hat{\sigma}[\overline{X}(n)]\sqrt{n}}, \qquad (75)$$

for $0 < t \leq 1$ (where $\lfloor a \rfloor$ denotes the greatest integer not greater than a), and $T(0) = 0$.

Heidelberger and Welch [1983] listed a few other standardized sequences that can be used to find statistics for the above rule. Rejection or acceptance of the hypothesis that a given subsequence of observations is stationary or, equivalently, that the initial transient period is not included in the observations, depends on the probability characterizing the scalar value calculated from the considered sequence. Despite the sophisticated theory hidden behind these tests, they appear to be quite simple numerically and can be applied to a wide class of simulated processes. A sequential version of one of the tests proposed by Schruben et al. [1983] is presented in Section 3. The main practical problem with their implementations is that they require a priori knowledge of the variance estimator $\hat{\sigma}^2[\overline{X}(n)]$ of the simulated process in steady state. To estimate this variance, one can use a sequence of observations collected at some distance from an assumed truncation point, assuming that the process is then at least closer to steady state. Schruben [1982] also describes a test that does not require the variance $\sigma^2[\overline{X}(n)]$, suggesting that it should perform not worse than others. No results of its evaluation are available. Note that searching for the beginning of the stationary phase by means of any statistical test can be shorten if it is preceded by one of the rules of thumb, for example, the rule R1 or R6, to find a rough approximation of the length of the initial transient period.

Recently Vassilacopoulos [1989] proposed a new, simple statistical test, which does not require the calculation of the variance of the simulated process. But as its author stated, the test should be evaluated extensively before its adopted in simulation practice.

3 SEQUENTIAL PROCEDURES FOR STEADY STATE SIMULATION: EXAMPLES

This section presents in detail two sequential procedures for stopping the simulation experiment when the required relative precision of confidence intervals is achieved. The first procedure is based on the method of spectral analysis; the second applies the method of the nonoverlapping batch means. Since both these procedures require the analysed sequence of observations to be stationary, each of them has to be preceded by a sequential procedure for detecting the length of the initial transient (nonstationary) period. Thus, when these procedures are applied, simulation experiments comprise two stages: Stage 1 for determining the length of the initial transient period and Stage 2 during which the steady-state behaviour is simulated and analysed, as illustrated in Figure 3.

3.1 Detecting the Length of the Initial Transient Period

The sequential procedure presented here is based on a stationarity test proposed by Schruben et al. [1983]. It is used to test the hypothesis that a sufficient number of initial transient data has been (or has not been) discarded. As in any statistical test, the value of a chosen statistic calculated from the tested sequence of observations

is compared with the corresponding value from a standard sequence, and the decision about rejection or acceptance of the hypothesis is taken at an assumed significance level $\alpha_t, 0 < \alpha_t < 1$. The significance level can be regarded as the probability of erroneously rejecting the hypothesis that the tested process is stationary. To get a first approximation for the truncation point,n_o^* , we can use one of the heuristic rules R1-R10 presented in Section 2. For example, in simulation studies of data communication protocols, [Pawlikowski and Asgarkhani 1988] rule R5 was applied (assuming $k = 25$). A flow chart of the procedure is given in Figure 4.

The problem encountered during testing a sequence of n_t observations for stationarity is that the steady-state estimator for the variance $\sigma^2[\overline{X}(n)]$, and the number κ of degrees of freedom for its χ^2 distribution, has to be known earlier than we know that the process has entered the stationary region. To get a robust estimate of that variance, the estimation should be done using only a subsequence of the last n_v observations from the sequence of n_t observations tested for stationarity; that is, taking $n_t \geq \gamma_v n_v$, where $\gamma_v(\gamma_v \geq 2)$ is the "safety" coefficient for the variance estimator to represent the steady state. Assuming larger value of γ_v, the last n_v of n_t observations are more likely to be from the stationary region, even if the truncation point of the initial transient period has been initially underestimated. Both the value of n_v and n_t should be selected after having taken into account the minimum sample sizes required by the method of variance analysis and the stationarity test. Heidelberger and Welch [1983] assumed $n_v \geq 100, \gamma_v = 2$. Having assumed such value of n_v we have at least $n_t = 200$ observations stored in a buffer for testing

against nonstationarity, which is the size of the sample assumed by Schruben [1982]. Since the number of observations tested for stationarity should be larger if longer transient

periods are suspected, one can assume that

$$n_t = max(\gamma_v n_v, \gamma n_o^*), \qquad (76)$$

where γn_o^* is the smallest length of one step in sequential testing for stationarity for a given $n_o^*, \gamma > 0$.

Thus, after having discarded n_o^* observations, the next n_t observations are collected and the last n_v observations are used to find $\sigma^2[\overline{X}(n)]$ and κ. The variance and the degrees of freedom of its χ^2 distribution can be estimated using a few different methods, presented, for example, in Fishman [1973a, p. 289], Heidelberger and Welch [1981a], and Schruben [1982]. The method desribed by Heidelberger and Welch [1981a], based on the spectral analysis of observations, was used in simulation studies of data communication protocols [Asgarkhani and Pawlikowski 1989; Pawlikowski and Asgarkhani 1988] and is summarized in Appendix B.
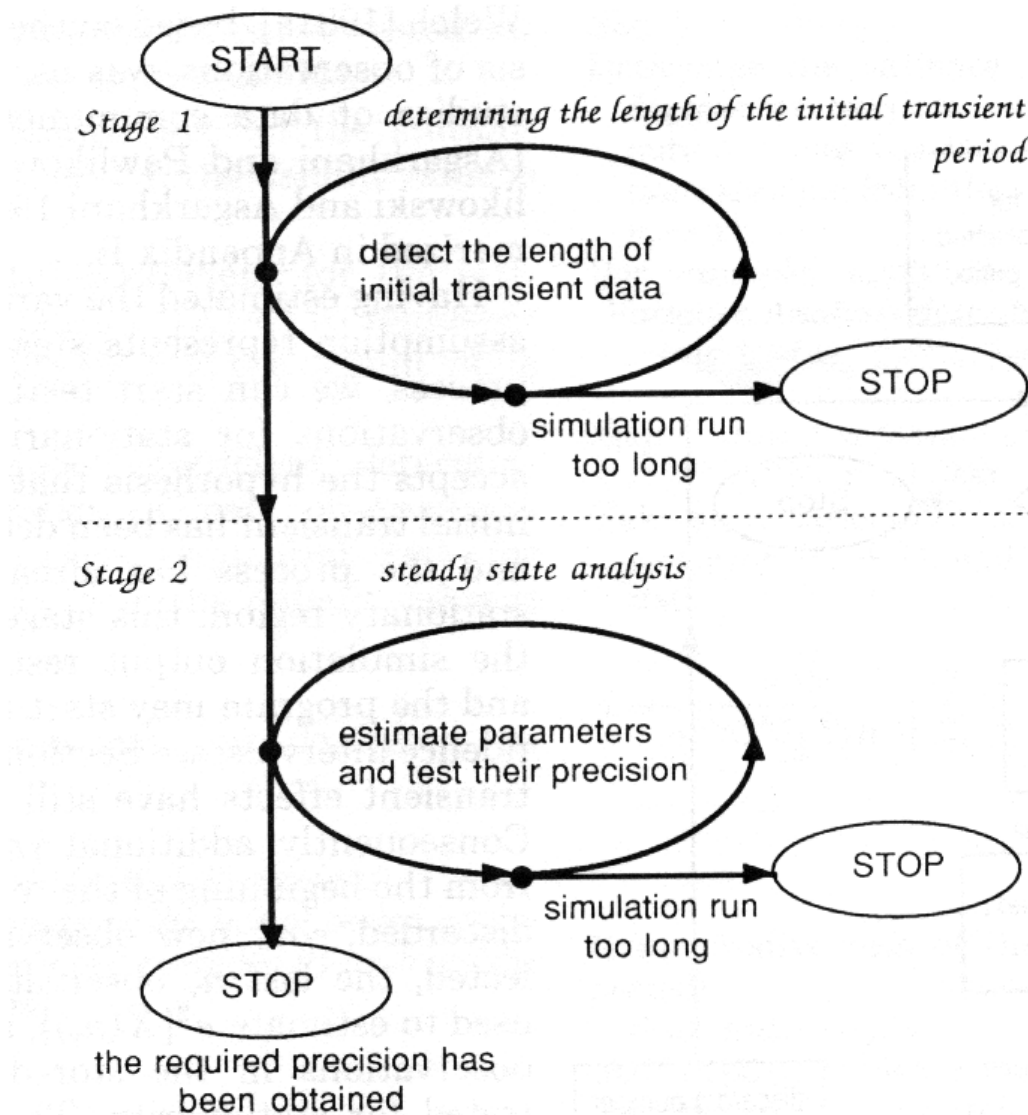
Fig. 3: Two-stage sequential methods for data collection and analysis in steady-state simulation.
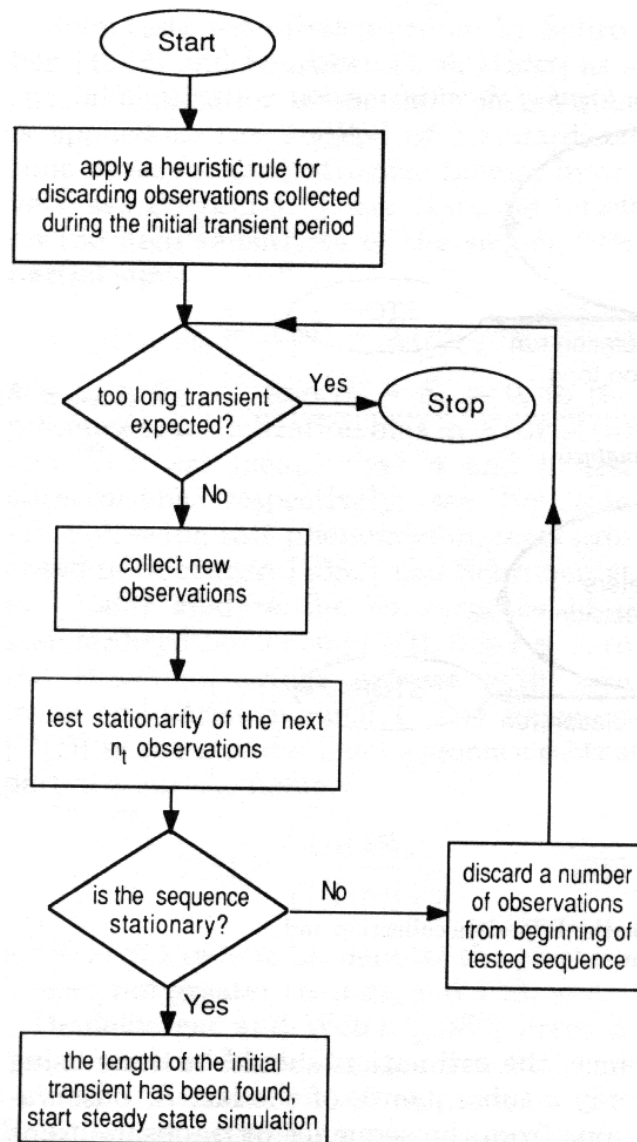
Fig. 4: Sequential analysis of the length of the initial transient period.

Having estimated the variance, which by assumption represents steady state of the process, we can start testing the first n_t observations for stationarity. If the test accepts the hypothesis that the end of the initial transient has been detected correctly and the process has already entered its stationary region, this stage of analysis of the simulation output results is finished and the program may start to analyze confidence intervals; see Section 3.2. Otherwise transient effects have still been detected. Consequently additional γn_o^* observations from the beginning of the test sequence are discarded, γn_o^* new observations are collected, the last n_v observations are again used to estimate $\sigma^2[\overline{X}(n_v)]$, and the first n_t observations in the stored sequence are tested for stationarity. This procedure is continued until the stationarity of the sequence of observations is confirmed or $n_{o,max}$ (the upper limit of the number of tested observations) is reached. As the longest acceptable length of the initial transient period, one can assume $n_{o,max} = 0.5n_{max}$, where n_{max} is the maximum length of the simulation run. If the transient phase extends beyond $n_{o,max}$ observations, either the allowed maximum length of the simulation run is too short or the system is unstable.

Summing, up, the procedure requires the following parameters:

n_max The maximum allowed length of the simulation run measured in the number of recorded observations (to be decided in advance. In simulation studies reported in Asgarkhani and Pawlikowski [1989] and Pawlikowski and Asgarkhani [1988] n_{max} was not less than $100,000$);

$n_{o,max}$ The maximum allowed length of the initial transient period (the default value is $n_{o,max} = 0.5n_{max}$);

n_v The length of the sequence used for estimating the steady-state variance $\sigma^2[\overline{X}(n)]$ (the default value is $n_v = 100$);

n_t The length of the sequence tested for stationarity; see Equation (76).

α_t The significance level of the stationarity test ($0 < \alpha_t < 1$; the default value is $\alpha_t = 0.05$).

γ_v The "safety" coefficient for the estimator of variance $\sigma^2[\overline{X}(n)]$ to represent the steady state ($\gamma_v \leq 2$; the default value is $\gamma_v = 2$).

γ The "exchange" coefficient, determining the number of new observations included in each sequential test for stationarity ($\gamma > 0$; the default valueiis $\gamma = 0.5$).

The porcedure can be described as follows:

procedure *DetectInitialTransient*;
{*determine the length of the initial period applying the Schruben's test preceded by an heuristic rule of truncation*}

Step 1
Start the simulation run from the empty-and-idle state;
apply one of the heuristic deletion rules R1-R10 {*see Section 2*} to determine n_o^*;
{n_o^* *is the first approximation of the number of observations to be deleted*}
if (the initial transient period embraces more than $n_{o,max}$ observations) **then**
     goto *Step 6*
else $n_o := n_o^*$; discard first n_o observations

endif;
if $\gamma_v n_v \geq \gamma n_o^*$ **then** $n_t := \gamma_v n_v$ else $n_t := \gamma n_o^*$

    endif;

$\Delta n := n_t$;

{n_t *observations will be tested for stationarity, Δn old observations will be replaced by new ones*}

$\alpha_{t1} := \alpha_t$;

{*the initial value of significance level of the test for stationarity; see Step 4*}

Step 2
if $n_o + n_t \leq n_{o,max}$ **then**

    append Δ_n observations to the tested sequence;

{*some of these observations may have been already collected when the heuristic deletion rule was applied*}

    goto *Step 3*
else   {*if $n_o + n_t > n_{o,max}$*}

        goto *Step 6*
endif;

Step 3
Determine the variance $\hat{\sigma}^2[\overline{X}(n_v)]$ and κ, the degrees of freedom of the variance distribution, using the last n_v collected observations starting from the observation $(n_o + n_t - n_v + 1)$;

{*for example, apply the procedure SpectralVariance Analysis described in Appendix B, assuming $x_s = x_{n_o+n_t-n_v+s}$, for $s = 1, 2, cdots, n_v$*}

Step 4
{*the test for stationarity, [Schruben et al. 1983, p. 1173]*}

Take all n_t observations, starting from theobservation $(n_o + 1)$, and calculate the test statistic

$$T = \frac{\sqrt{45}}{n_t^{1.5} n_v^{0.5} \hat{\sigma}[\overline{X}(n_v)]} \sum_{k=1}^{n_t} k(1 - \frac{k}{n_t})[\overline{X}(n_t) - \overline{X}(k)],$$

$$\text{where } \overline{X}(i) = \sum_{j=n_o+1}^{n_o+i} \frac{xj}{i};$$

if (a negative bias of the mean $\overline{X}(n_t)$ is suspected)

    then goto *Step 5*
elsif (a positive bias is suspected) **then** $T := -T$

{*the reason for considering bias of an assumed sign is that one-sided tests are usually more powerful than their two-sided correspondents*}

else
{*if a sign of the initial bias is difficult to predict, then prepare data for a two-sided test*}

    $T := |T|$; $\alpha_{t1} := \alpha_t/2$
endif;

Step 5
if $T \leq t_{k,1-\alpha_{t1}}$ **then**
{$t_{k,1-\alpha_{t1}}$ *is the upper $(1 - \alpha_{t1})$ critical point from the t-distribution with κ degrees of freedom*}

    write ('the initial transient period is not longer than n_o observations');

    start sequential analysis of confidence intervals

{*call one of procedures presented in Section 3.2*}

    else   {*if $T > t_{k,1-\alpha_{t1}}$*}

    discard first γn_o^* observations from the tested sequence;

    $n_o := n_o + \gamma n_o^*$; $\gamma n := \gamma n_o^*$;   **goto** *Step 2*
endif

Step 6
{*if the initial transient period embraces more than $n_{o,max}$ observations*}

stop the simulation run;
write ('the initial transient period embraces more than $n_{o,max}$ observations, or the simulated process is unstable')
end *DetectInitialTransient*.

The effectiveness of this procedure strongly depends on the effectiveness of the variance estimator $\hat{\sigma}^2[\overline{X}(n)]$.

3.2 Sequential Testing for a Required Accuracy of Results

Both sequential procedures for stopping a simulation experiment which are presented here require the analysed sequence of observations to be representative of steady state, so n_o observations representing the initial transient period have to have been discarded beforehand.

The first procedure is a modified version of the spectral method of analysis proposed by Heidelberger and Welch [1981a, 1981b, 1983]. Its simplified flow chart is given in Figure 5. As mentioned in Section 1, the method allows the reduction of the number of individual data items stored in memory during a simulation experiment by batching individual observations into batches of size $2m, m = 1, 2, \cdots$, and replacing them with the batch means. These means are stored in the buffer AnalysedSequence of size $2M$. Whenever $2M$ batch means over batches of m observations are recorded, they are consolidated into M means over batches of size $2m$. Subsequent observations are lumped into M successive batches of size $2m$, and if more than M such new batches are needed, the rebatching procedure is repeated. Thus, the buffer AnalysedSequence can be implemented simply as a one-dimensional array of size $2M$. The accuracy of estimators is measured by the relative precision ϵ of confidence intervals, defined in Equation (12), and the simulation is stopped if $\epsilon \leq \epsilon_{max}$, where

$\epsilon_m ax$ is the acceptable maximum value of the relative precision of the final results. The current values of ϵ are evaluated at consecutive checkpoints ω_k ($k = 1, 2, \cdots ; \omega_k \leq n_{max}$), that is, each time when ($\omega_{k+1} - \omega_k$) new observations have been collected. To limit the number of possible checkpoints, one can assume that they are geometrically distributed;that is, for a given ω_1,

$$\omega_{k+1} = min\{\lfloor \gamma_a(\omega_k - n_o) \rfloor + n_o, n_{max}\}, \quad (77)$$

where $k = 1, 2, \cdots ; \lfloor x \rfloor$ denotes the greatest integer not greater than x and $\gamma_a > 1$. To avoid too large a distance between consecutive checkpoints, one can assume $\omega_1 = max(2M, 2n_o)$ and a constant value for the step between checkpoints above a thershold point. Following a different reasoning, selection of the value of ω_1 satisfying the inequality $\lfloor 0.1(n_{max} - n_o) \rfloor$

$$+n_o \leq \omega_1 \leq \lfloor 0.2(n_{max} - n_o) \rfloor + n_o \quad (78)$$

was suggested in Heidelberger and Welch [1983]. The estimator of variance $\sigma^2[\overline{X}(n)]$, and the number κ of its degrees of freedom needed in analysis of confidence intervals, are calculated here using the procedure SpectralVarianceAnalysis. It requires at least $n_v \geq 100$ batch means to be available (see Appendix B). This approach is described below by a pseudocode procedure that uses the following parameters:

n_o The number of discarded initial observations (determined by the procedure DetectInitialTransient).

n_{max} The maximum allowed length of the simulation run, measured in the number of recorded observations ($n_{max} \geq max(3n_o, n_o + 2M)$; to be decided in advance).
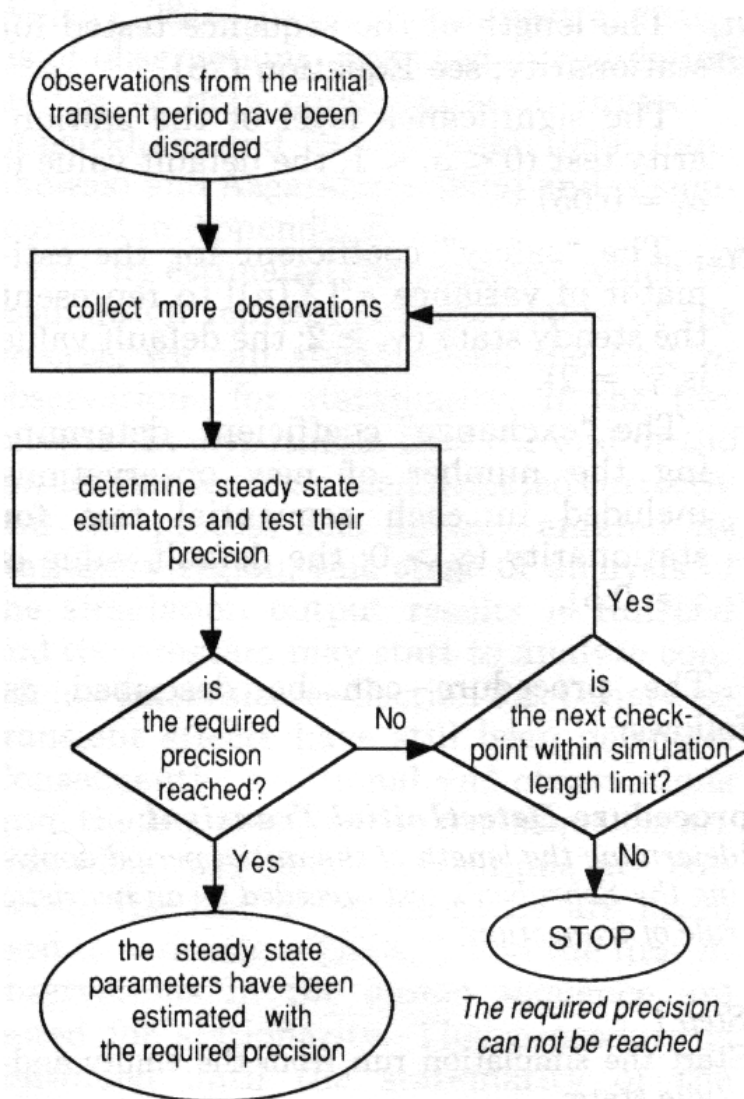
Fig. 5: Flow chart for the spectral method.

n_v The length of the sequence used for estimating the variance $\sigma^2[\overline{X}(n)](n_v \geq 100)$; the default value is $n_v = 100$).

γ_a The checkpoint incremental coefficient for sequential testing for accuracy ($\gamma_a > 1$; the default value is $\gamma_a = 1.5$).

$(1 - \alpha)$ The assumed confidence level of the final results ($0 < \alpha < 1$; the default value is $alpha = 0.05$).

ϵ_{max} The maximum acceptable value of the relative precision of confidence intervals ($0 < \epsilon_{max} < 1$; the default value is $\epsilon_{max} = 0.05$).

procedure *SpectralAnalysis*;

{*sequential analysis of simulation output data based on spectral analysis of the series of collected observations*}

const $M = 100$;

{*the default value of the minimum number of data points for the analysis of confidence intervals; $M \geq n_v \geq 100$; see Section 3.1*}

procedure *Batching*;

{*preparation of secondary output data; transformation of individual observations into the sequence of n_o more than $2M$ batch means of sequentially increasing batch size*}

begin

{*calculate the batch mean $\overline{X}_j(m)$, and store it as the jth data item in the buffer AnalysedSequence*}

    $\overline{X}_j(m) := sum/m$;

    if $j = 2M$ **then**

{*consolidate $2M$ means of batches of size m into M means of batches of size $2m$*}

       for   s:=1 **to** M **do**

         $\overline{X}_s(2m) := 0.5(\overline{X}_{2s-1}(m)+\overline{X}2s(m))$

       enddo

       $m := 2m$; $j := M$

    endif

    $j := j + 1$; $sum := 0$

{*start to calculate the next batch mean*}

end *Batching*;

procedure *Estimation*;

{*sequentially calculate estimates and test their precision until the required precision is reached*}

begin

    find the estimator $\hat{\sigma}^2[\overline{X}(n_v)]$ for the sequence $\overline{X}_{j-n_v+1(m)}, \cdots, \overline{X}_j(m)$, the last n_v batch means stored in the buffer AnalysedSequence, and determine κ, the degrees of freedom of the distribution of $\hat{\sigma}^2[\overline{X}(n_v)]$;

{*apply the procedure SpectralVarianceAnalysis;see Appendix B*}

calculate the relative half width of the confidence interval at the confidence level $(1 - \alpha)$ for the current checkpoint ω_k:

$$\epsilon = \frac{t_{k,1-\alpha/2}\hat{\sigma}[\overline{X}(n_v)]}{\overline{X}(jm)},$$

where $\overline{X}(jm) = \frac{\sum_{s=1}^{j}\overline{X}_s(m)}{j}$

is the current value of the estimated mean after jm observations, and $t_{k,1-\alpha/2}$ is the upper $(1 - \alpha/2)$ critical point of the t-distribution with κ degrees of freedom;

{*test conditions of stopping the simulation run*}

    if   ($\epsilon \leq \epsilon_{max}$) **then**

{*print the final results and stop the simulation*}

       StopSimulation := **true**

    else

{*the required precision has not been reached yet; determine the next, $(k + 1)st$, checkpoint*}

       $k := k + 1$;

       $\omega_k := min(\lfloor\gamma_a(\omega_{k-1} - n_o)\rfloor + n_o, n_{max})$

    endif

end *Estimation*;

begin   {*main procedure*}

    $m := 1$;

{*the initial batch size*}

    $k := 1$;

{*the initial checkpoint is after $\omega_k = \omega_1$ observations*}

    $\omega_o := n_o$; $i := 1$;

{*having discarded n_o observations, collect next observations starting from the observation $(\omega_o + i)$*}

    $\omega_1 := max(2M, 2n_o)$;

{*the default location of the first checkpoint*}

    sum := 0; $j := 1$;

{*start calculating the 1st batch mean*}

    StopSimulation := **false** ;

{*a condition of stopping the simulation has not been met yet*}

while (**not** StopSimulation) **do**
{*collect and process new* ($\omega_k - \omega_{k-1}$)
observations}
    get the observation
    x_{ω_o+i};
sum := sum + x_{ω_o+i};
    if ($i \bmod m = 0$) **then** *Batching*
endif;
    if ($i = \omega_k$) **then** *Estimation*
endif;
    if (**not** StopSimulation) **then**
        $i := i + 1$;
        if ($i > n_{max} - n_o$) **then**
            write ('the required accuracy
can not be reached; either restart the sim-
ulation using a new seed (seeds) of random
number generator, or increase n_{max}, or in-
crease α, or increase ϵ_{max}');
{*for obtaining a new sequence of pseudoran-
dom numbers, independent from the previous
one, assume the last number from the previ-
ous sequence as the first number of the new
sequence*}
            StopSimulation := **true**
        endif
    endif
enddo;
    write ('the required precision of re-
sults has been obtained having collected',
$n_o + jm$, 'observations; the final relative pre-
cision:', 100ϵ%, 'the final $(1 - \alpha)100$% confi-
dence interval:', $\overline{X}(jm)[1 \pm \epsilon]$)
end *SpectralAnalysis*.

The next sequential procedure for stop-
ping simulation experiments when the re-
quired precision of results is reached is based
on the method of nonoverlapping batch
means. Its simplified flow chart is given
in Figure6 (only the main loops of com-
putations are depicted). As in the case of
the procedure SpectralAnalysis, the initial
nonstationary sequence of n_o observations

should have been discarded earlier, apply-
ing, for example, the procedure DetectIni-
tialTransient. For weakening serial corre-
lations of analysed output data, individual
observations are replaced here by the less
correlated means of their batches. Thus,
the problem of direct analysis of confidence
intervals from correlated observations is re-
placed by the problem of determining the
batch size m^*, such that batch means are
(almost) uncorrelated at a given level of sig-
nificance. Generally, in a sequence of corre-
lated data the autocorrelation coefficients of
lag $k, k = 1, 2, \cdots$, [see Equations (21) and
(22)] are not necessarily decreasing as the
lag increases, although all autocorrelation
coefficients are zero if the sequence contains
uncorrelated observations. For this reason
we follow here the test proposed by Adam
[1983]: a given batch size can be accepted
as the batch size for approximately uncor-
related batch means if all L autocorrelation
coefficients of lag $k(k = 1, 2, \cdots, L)$ are sta-
tistically negligible at a given significance
level $\beta_k, 0 < \beta_k < 1$. The analytical prob-
lems encountered during estimation of the
autocorrelation coefficients suggest that the
number of considered lags should be limited
to $L = 0.1k_{b0}$, where k_{b0} is the number of
batch means tested for autocorrelation; see
comments in Section 1. The autocorrelation
coefficients can be better estimated by the
so-called jackknife estimators [Miller 1974],
which are usually less biased than the ordi-
nary estimators defined by Equations (21)
and (22). A jackknife estimator of autocor-
relation coefficient of lag k for a sequence of
batch means of size m is calculated from the
following formula:

$\hat{\hat{r}}(k, m)$

$$= 2\hat{r}(k, m) - \frac{\hat{r}'(k, m) + \hat{r}''(k, m)}{2} \qquad (79)$$

where the estimators on the right-hand-side are calculated like ordinary estimators of autocorrelation coefficients [see Equations (21) and (22)], except that $\hat{r}(k, m)$ is the estimator over all k_{b0} batch means, whereas $\hat{r}'(k, m)$ and $\hat{r}''(k, m)$ are estimators over the first and the second half of the analysed sequence of k_{b0} batch means, respectively.

Let us note that:

1. to get acceptable estimators of the autocorrelation coefficients, at least 50 batch means should be available, [Box and Jenkins 1970, p. 33]; thus in the case of jackknife estimators one should assume $k_{b0} \geq 100$.

2. to ensure approximate normality of batch means, the size of considered batches should not be less than 50 [Adams 1983].

3. to get an acceptable overall significance level β when testing the value of L autocorrelation coefficients of lag $k(k = 1, 2, \cdots, L)$, each at the significance level β_k, we have to assume

$$\beta < \sum_{k=1}^{L} \beta_k, \qquad (80)$$

hence in practice L should not be too large. This restriction is irrelevant if the autocorrelation coefficients decrease monotonically with the value of the lag, since then only $\hat{\hat{r}}(1, m)$ has to be considered.

To avoid wastefully collecting an excessive number of observations, especially when testing batch sizes, the procedure Batch-MeansAnalysis uses two buffers for storing batch means: a buffer called ReferenceSequence is used to store the batch means $\overline{X}_1(m_o), \overline{X}_2(m_o), \cdots$ of a batch size m_o, and a buffer AnalysedSequence is used for storing an assumed number k_{b0} of batch means over batches of size $m_s = s m_o (s = 1, 2, \cdots)$, formed from the batch means kept in the ReferenceSequence. Thus, since the number of data items collected in the ReferenceSequence grows in time during a simulation run, a linked list of batch means seems to be a proper data structure for this buffer. The number of data items in the AnalysedSequence is limited to k_{b0}, so it can be implemented as an ordinary one-dimensional array. By selecting m_o properly, we can secure a sequential increase of tested batch sizes slower than in the batching schemes proposed in Adams [1983], Fishman [1973a], Law and Carson [1979] to reduce the resultant simulation run length. For neutralizing the observed randomness of the estimators of correlation coefficients, $m^* = m_s$ is selected as the final batch size of weakly correlated means iff the hypothesis of all zero autocorrelation coefficients is accepted in two successive tests, both for the batch size m_{s-1} and m_s.

Having selected the batch size m^*, one can sequentially analyse the accuracy of results by calculating confidence intervals from Equations (18) and (19), which are valid for independent and identically distributed batch means. The sequence of batch means for batch size m^* kept in the buffer ReferenceSequence can be sequentially appended by new batch means if more observations are needed to improve the accuracy of results.
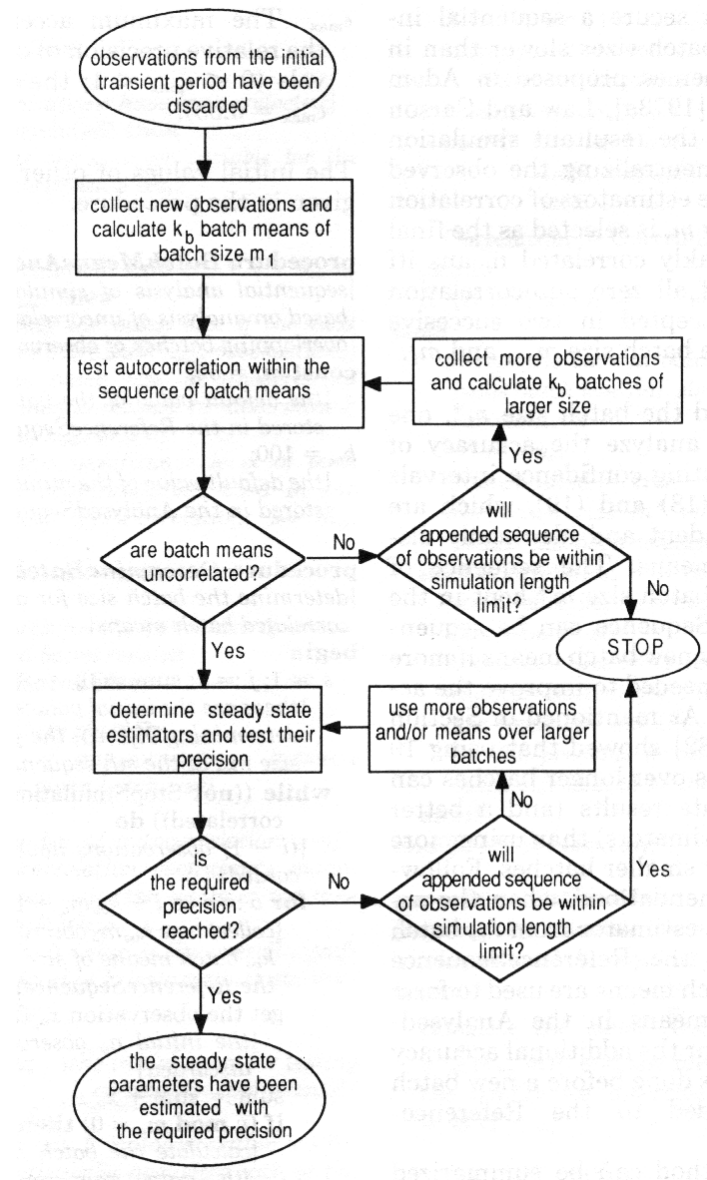
Fig. 6: Flow chart for the method of batch means.

As mentioned in Section 1, Schmeiser [1982] showed that using 10 to 30 batch means over longer batches can give more accurate results (and a better coverage of the estimators) than using more batch means but over smaller batches. Following these recommendations, when the accuracy test of the estimator from k_{be} batch means stored in the ReferenceSequence fails, then these k_{be} batch means are used to form $k_{bo} = 30$ batch means in the AnalysedSequence buffer for the additional accuracy test. Such test is done before a new batch mean is appended to the ReferenceSequence.

The whole method can be summarized by the following pseudocode procedure, which requires the following parameters:

n_o The number of discarded initial observations (determined by the procedure DetectInitialTransient).

n_{max} The maximum allowed length of the simulation run, measured as the number of recorded observations ($n_{max} \geq n_o + m_o k_{bo}$; to be determined in advance).

$1 - \alpha$ The assumed confidence level of the final results ($0 < \alpha < 1$; the default value is $\alpha = 0.1$).

ϵ_{max} The maximum acceptable value of the relative precision of confidence intervals ($0 < \epsilon_{max} < 1$; the default value is $\epsilon_{max} = 0.1$).

The initial values of other parameters are given in the procedure.

procedure *BatchMeansAnalysis*;
{ *sequential analysis of simulation output data based on analysis of uncorrelated means of nonoverlapping batches of observations*}

const $m_o = 50$;
{ *the default value of the batch size for means stored in the ReferenceSequence*}
$k_{bo} = 100$;
{ *the default value of the number of batch means stored in the AnalysedSequence*}

procedure *DetermineBatchSize*;
{ *determine the batch size for approximately uncorrelated batch means*}
begin
    s := 1; j := 1; sum := 0;
{ *these are the initial values of parameters for determining $\overline{X}_j(m_o)$, the jth batch mean of size m_o, at the sth sequential step*}
    while ((**not** StopSimulation)
    and (**not** Uncorrelated)) **do**
{ *$(i - 1)$ observations have been already recorded*}
        for v := i **to** $i + k_{bo}m_o - 1$ **do**
{ *collect new $k_{bo}m_o$ observations for the next k_{bo} batch means of size m_o to be store in the ReferenceSequence*}
            get the observation $x_{n_\alpha} + v$;
{ *the initial n_o observations have been discarded*}
            sum := sum $+ x_{n_\alpha} + v$;
            if (v **mod** $m_o = 0$) **then**
{ *calculate the batch mean $\overline{X}_j(m_o)$; the jth data item in the ReferenceSequence*}
                $\overline{X}_j(m_o)$ =sum/m_o;
                sum := 0; j := j + 1
            endif
{ *the next batch mean has been determined*}
        enddo
{ *the next k_{bo} batch means have been determined*}
        for v := 1 **to** k_{bo} **do**
{ *consolidate sk_{bo} batch means $\overline{X}_1(m_o)$, $\overline{X}_2(m_o), \cdots$,from the ReferenceSequence into k_{bo} batch means $\overline{Y}_1(sm_o), \overline{Y}_2(sm_o), \cdots$,in the AnalysedSequence, which will be tested for autocorrelation*}

$$\overline{Y}_v(sm_o) := \sum_{r=1}^{s} \overline{X}_{(v-1)s+r}(m_o)/s$$

   enddo
   TestCorrelation;
   $i := i + k_{bo}m_o$;
{n_{o+i-1} *observations have been collected*}
   if (**not** Uncorrelated) **then**
{*re-initialize the counter variable for the next autocorrelation test*}
    s:= s+1
   endif
   if (**not** Uncorrelated) **and** (n_o+ i +$k_{bo}m_o - 1 > n_{max}$) **then**
{*stop selecting the batch size if the next test requires more than n_{max} observations*}
    write ('the batch size of uncorrelated batch means can not be determined; increase n_{max} or β');
{β *is the significance level of tests against autocorrelation used in the procedure Test-Correlation; see Eq. (80)*}
    StopSimulation := **true**
   endif
  enddo
{ *of search for the batch size of uncorrelated batch means*}
end **DetermineBatchSize**;

procedure **TestCorrelation**;
{ *test significance of autocorrelations between batch means for a given batch size*}
const $L = 10$;
{*the default number of autocorrelation coefficients; if autocorrelation coefficients monotonically decrease with the value of lag then $L = 1$, else $L:=0.1 \ k_{bo}$*}
$\beta = 0.1$;
{*the default value of the overall significance level of L tests against autocorrelation*}
begin
  Correlation :=0;
{*the initial value for testing L correlation coefficients*}
  for $k := 1$ **to** L **do**
{ *test whether all L autocorrelation coefficients are statistically negligible each at the β_k significance level*}
   calculate the jacknife estimator $\hat{\hat{r}}(k, sm_o)$ of the autocorrelation coefficient of lag k for the sequence $\overline{Y}_1(sm_o)$, $\overline{Y}_2(sm_o)$, $\cdots$, $\overline{Y}_{k_{bo}}(sm_o)$ stored in the AnalyzedSequence; { *apply Eq. (79) for $m = sm_o$*}
   $\beta_k := \beta/L$;
   $\hat{\sigma}^2[\hat{\hat{r}}(k, sm_o)]:=$
    if (k=1) **then** (1/ k_{bo})
     else ($[1+2\sum_{u=1}^{k-1}\hat{\hat{r}}^2(u, sm_o)]/k_{bo}$)
   endif
{ *the variance of the autocorrelation coefficient of lag k [Bartlett 1946]*}
   if ($|\hat{\hat{r}}(k, sm_o)| < z_{1-\beta_k/2}\hat{\sigma}[\hat{\hat{r}}(k, sm_o)]$)
{ $z_{1-\beta_k/2}$ *is the upper $(1-\beta_k/2)$ critical point of the standard normal distribution*}
    then $\hat{\hat{r}}(k, sm_o):=0$
   endif
{ *the lag k autocorrelation is statistically negligible at the confidence level $1 - \beta_k$*}
   Correlation := Correlation+$|\hat{\hat{r}}(k, sm_o)|$;
  enddo
  if ((Correlation=0) **and** AcceptableSize)
then
{ *accept the current batch size since the previous batch size has already given negligible correlations*}
   $m^* := sm_o$;
   Uncorrelated := **true**
{ *the batch size m^* of uncorrelated batch means has been selected*}
  elsif (Correlation=0) **then**
{ *start collecting next observations for testing larger batch size*}
   AcceptableSize := **true**
{ *autocorrelations for the current batch size*}

are negligible but they were not negligible for the previous batch size, thus the next batch size should be considered}
  endif
end ***TestCorrelation***;

procedure ***Estimation***;
{ sequentially calculate estimates and test their precision until the required precision is reached}
begin
calculate the mean $\overline{\overline{X}}(k_{be}, m)$, its variance $\hat{\sigma}^2[\overline{\overline{X}}(k_{be}, m)]$, and the relative half width ϵ of the confidence intervals at the confidence level $(1 - \alpha)$:

$$\epsilon := t_{k_{be}-1, 1-\alpha/2}\hat{\sigma}[\overline{\overline{X}}(k_{be}, m)]/\overline{\overline{X}}(k_{be}, m),$$

using the whole sequence of k_{be} data items stored in the ReferenceSequence;
{ apply Eqs. (17)-(19), for $k_b := k_{be}$, $m := m^*$ and $\overline{X}_i(m) := \overline{X}_i(m)$}
  if $((\epsilon > \epsilon_{max})$ **and** $(kbe \bmod 30 = 0))$ **then**
{ the additional test of accuracy of estimators after consolidation of k_{be} batch means from the ReferenceSequence into $k_{bo} = 30$ means of longer batches stored in the AnalysedSequence; see the recommendations in Schmeiser [1982]}
   $m := k_{be}/30$;
    for $j := 1$ **to** 30 **do**
    $\overline{Y}_j(mm^*) := \sum_{r=1}^{m} \overline{X}_{(j-1)m+r}(m^*)/m$;

calculate the mean $\overline{\overline{X}}(k_{bo}, mm^*)$, its variance $\hat{\sigma}^2[\overline{\overline{X}}(k_{bo}, mm^*)]$, and the relative half width ϵ of the confidence interval at the confidence level $(1 - \alpha)$:

$$\epsilon :=$$
$$t_{k_{bo}-1, 1-\alpha/2}\hat{\sigma}[\overline{\overline{X}}(k_{bo}, mm^*)]/\overline{\overline{X}}(k_{bo}, mm^*),$$
from the sequence of 30 data items stored in the AnalyzedSequence;
{ apply Eqs. (17)-(19), for $k_{bo}:= 30$, $m:= mm^*$, $\overline{X}_i(m) := \overline{Y}_i(mm^*)$}
   enddo
  endif
   if $(\epsilon \le \epsilon_{max})$
{ the required precision has been reached}
   then write('the required precision of results has been obtained after', $n_o + k_{be}m^*$, 'observations recorded');
   StopSimulation := **true**
  elsif $(n_o + (k_{be} + 1)m^* \le n_{max})$ **then**
{ k_{be} batch means of size m^* have been used to estimate the mean μ_x, but the required precision has not been reached yet; collect the next batch of m^* observations and store their mean in the ReferenceSequence}
   sum := 0;
   for $j := 1$ **to** m^* **do**
    get the observation $x_{k_{be}m^*+j}$;
    sum := sum + $x_{k_{be}m^*+j}$
   enddo
   $k_{be} := k_{be} + 1$;
   $\overline{X}_{k_{be}}(m^*) := sum/m^*$
  else
{ the required precision has not been reached; too short the simulation run assumed}
   write ('the required accuracy can not be reached; either restart the simulation using a new seed (seeds) of random number generator, or increase n_{max}, or increase α, or increase ϵ_{max}');
{ for obtaining a new sequence of pseudorandom numbers, independent from the previous one, assume the last number from the previous sequence as the first number of the new sequence}
   StopSimulation := **true**
  endif
end ***Estimation***;

begin *{ main procedure}*
  StopSimulation := **false** ;

{ *a condition of stopping the simulation has not been met yet* }

    AcceptableSize := **false** ;

{ *a batch size for uncorrelated batch means has not been found yet* }

    Uncorrelated := **false** ;

{ *the batch size for uncorrelated batch means has not been determined yet* }

    $\epsilon := 1 - \epsilon_{max}$;

{ *the initial precision $\epsilon > \epsilon_{max}$* }

    $i := 1$;

{ *having discarded n_o observations, collect the next observations starting from the observation $(n_o + 1)st$* }

DetermineBatchSize ;

{ *if the batch size m^* has been selected then the ReferenceSequence contains sk_{bo} batch means of size m^*/s and the AnalysedSequence contains k_{bo} batch means of size m^** }

    $k_{be} := k_{bo}$;

{ *the size of the ReferenceSequence at the beginning of estimation* }

    for $j := 1$ **to** k_{be} **do**

{ *prepare data for estimation; consolidate sk_{bo} batch means of size $m_o = m^*/s$ in the ReferenceSequence into $k_{be} = k_{bo}$ batch means of size m^** }

$$\overline{X}_j(m^*) := \sum_{r=1}^{s} \overline{X}_{(j-1)s+r}(m_o)/s$$

    enddo

    while (Uncorrelated **and**

        (**not** StopSimulation))

do

{ *sequentially calculate and test the precision of the estimators until the required precision or the maximum length of simulation run is reached* }

Estimation ;

    enddo

    write ('the final relative precision:', $100\epsilon\%$, 'the final $100(1 - \alpha)\%$ confidence interval:', $\overline{\overline{X}}(k_{be}, m^*)[1 \pm \epsilon]$) ;

end ***BatchMeansAnalysis***.

In practice, the last procedure often uses a sequential search for the batch size of uncorrelated batch means rather than the sequential stopping rule for the simulation run (sequential testing the precision of estimates). This happens because usually many more batch means have to be tested during the first stage (testing against autocorrelation) than during the second stage (testing the precision of the results). The number of observations recorded, $k_{bo}m^*$, when the mean and the width of confidence intervals are to be calculated for the first time is usually much larger than it is required for obtaining the required level of accuracy [Schmeiser 1982]. From this point of view, the spectral method of analysis is more thrifty. On the other hand, the method of batch means uses a very simple estimator of the variance when the width of confidence intervals is analysed. Exhaustive comparative studies of both procedures have not been performed yet, but the reported results indicate that the spectral method is usually more efficient, both in the sense of the coverage recorded in reference experiments and in the sense of the final simulation run lengths for obtaining the required accuracy of the results, at least in investigated classes of processes; [Pawlikowski and Asgarkhani 1988].

4   SUMMARY AND GENERAL COMMENTS

We have discussed in detail methods for dealing with the main phenomena encountered in steady-state simulation of queueing processes: the inherent initial nonstationarity and the permanent autocorrelation of collected observations. The emphasis is on methods for the sequential anal-

ysis of simulation output data, bearing in mind their possible implementation in user-friendly simulation packages, which would produce results automatically. In such a context, methods of analysis based on single simulation runs seem to be more attractive than the methods of independent replications.

The survey has been limited to analysis of a single measure of performance, namely, the point and interval estimates of the sample mean. Nevertheless, this methodology can easily be modified to estimate other performance measures, as long as their estimators are based on cumulative statistics. For example, sequential analysis of simulation output data can easily be applied to estimate the variance, or the probability that the analysed variate lies in some fixed interval [Welch 1983, Sec. 6.3]. It cannot be applied directly to estimate quantiles, since their estimators require that sequences of collected observations be pre-sorted and as such are not amenable to cumulative statistics calculated while the simulation is in progress. Estimation of quantiles by using the method of spectral analysis is discussed in Heidelberger and Lewis [1984]; the same application of the regenerative method is considered in Iglehart [1976] and Seila [1982a]. Computational problems associated with quantile estimation are discussed in Jain and Chlamtac [1985] and Raatikainen [1987, 1988].

The methodology of simultaneous analysis of more than one measure of performance during one simulation experiment is much less advanced than the methodology of univariate estimation discussed in this paper. The main theoretical problems of multivariate estimation and the methods which can be used for determining the confidence regions in such multiresponse simulation experiments are discussed in Charnes and Kelton [1988], Chen and Seila [1987], Friedman [1984], Law and Kelton [1982a, p. 308], and Seila [1982b, 1983]. The initialization bias in multiresponse simulation can be controlled by applying a statistical test proposed by Schruben [1981]. Various variance reduction techniques that are applicable in this kind of simulation experiments are discussed in Bauer et al. [1987], Rubinstein and Marcus [1985], Venkartraman and Wilson [1986], and Yang and Nelson [1988].

Specific statistical problems accompanying comparative simulation studies of alternative systems are surveyed in Law and Kelton [1982a, Chap. 9]; see also Balmer [1987], Clark [1986], Friedman and Friedman [1986], and Goldsman [1983, 1986]. (Note that in these applications the bias of estimators is not important, as long as the estimators are equally biased.) Stochastic simulation is also used for sensitivity analysis of performance measures and optimization of various systems and processes under uncertainty. Special methods for estimating the gradient of a performance measure with respect to a selected input parameter (e.g., the changes of the mean delay of messages in a communication network with respect to the mean message length) during a single simulation run are discussed [Ho 1987; Ho and Cao 1983; Ho et al. 1983; Reiman and Weiss 1986; Rubinstein 1989; Suri and Zazanis 1988].

This paper is restricted to the statistical analysis of simulation output data, but practitioners are aware that that is not the only problem that must be overcome in obtaining useful results. Before observations are collected, the processes for which performance is to be investigated have to be properly modeled, and each model should be validated and verified to make the simulation experiments credible. Various aspects of the validation and verification of simu-

lation models are discussed in [Banks and Carson 1984; Bulgren 1982; Law and Kelton 1982a; Sargent 1986; Shannon 1981; Velayas and Levary 1987].

Having presented methods of data collection and analysis which are used in stochastic steady-state simulation, one can conclude that no definite conclusions can be made about their applicability. The need for more exhaustive comparative studies expressed by Schriber and Andrews [1981] still to be a live issue.

The length of simulation runs remains a critical issue, especially in the case of simulation studies of complex systems, which often can be performed only if a technique for speeding up the process of data collection is applied. In this context an important role could be played by the variance reduction techniques, which, by reducing the variance, narrow confidence intervals and consequently reduce the number of steps needed by sequential procedures for reaching the required accuracy of results. But, as was mentioned in Introduction, practical implementations of variance reduction techniques have been reported infrequently. Among the recently published ones are those in Izydorczyk et al. [1984] and Walrand [1988]. Another area where the duration of simulation can be very critical is the analysis of performance measures that depend on events occuring infrequently. Some techniques proposed for speeding up rare events simulation are discussed in Cottrell [1983], Parekh and Walrand [1989], and Shahabuddin et al. [1988].

Performance evaluation studies based on simulation experiments can be accellerated by the decomposition of analysed systems into subsystems, which are modelled separately, applying both simulation and analytical models, mutually interacting if neces-

sary. The efficiency of such an approach was studied extensively in Blum et al. [1985], in the context of queuing network models of computer systems. A survey of hybrid techniques can be found in Shanthikumar [1983]; see also Frost et al. [1988] and Kurose and Mouftah [1988]. The time needed for simulation studies can also be significantly reduced in multiprocessor systems. In such an environment individual processors can be used for running independent replications of one simulation experiment, or, in a more sophisticated solution, logical processes occuring during one simulation run can be executed in parallel by different processors. The main problem encountered in the latter case is the synchronization among processes run on different processors. There exists a danger that the contribution of large synchronization overheads will actually slow the simulation experiment. Various techniques of distributed and parallel simulation are discussed in Chandy and Misra [1981], Fujimoto [1988], Misra [1986], Nicol [1988a, 1988b], Reynolds [1988], Unger [1988], and Wagner and Lazowsa[1989]; see also Baik and Zeigler [1985], DeCegama [1987], Krishnamurthy et al. [1985], Kumar [1986], and Kurose and Mouftah [1988]. Specific problems relating to the statistical analysis of output data which accompany distributed simulation are discussed in Heidelberger [1986].

The diversity of methodological approaches and the complexity of some statistical techniques used in simulation output analysis has motivated attempts to develop expert systems that advise on selecting the best simulation methodology for accomplishing the required statistical computations; see, for example, Haddock [1987], Hahn [1985], Hand [1985], Nielsen [1986], and O'keefe [1986]. One of the first sta-

tistical expert systems specifically oriented towards steady-state simulation is discussed in Mellichamp and Park [1989]. It is hoped that further developments in the area of expert systems and applied statistics will make possible to design fully automated, knowledge-based simulation packages.

APPENDIX A. ELEMENTARY STATISTICAL CONCEPTS

Appendix A contains elementary statistical concepts needed to understand the problems discussed in the paper. A deeper and more detailed discussion of the subject can be found in the numerous textbooks on applied probability theory and statistics.

Stochastic simulation involves experiments which mimic various events by means of numbers generated by (pseudo)-random numbers generators. Therefore, it is natural that the analysis of the output data should be based on the methods of statistical inference. These methods can help to estimate the characteristics and/or parameters of simulated processes. Accepting the random nature of simulation output data, one should consider any sequence $x_1, x_2, \cdots, x_n$ of observations collected during such experiments as the sequence of realizations of random variables $X_1, X_2, \cdots, X_n$, sampled from a stochastic process $X(t)$ at instants of time $t = t_i, i = 1, 2, \cdots, n$. If the process $X(t)$ is stationary, what means that its stochastic characteristics are time invariant, then the random variables $X_1, X_2, \cdots, X_n$ have the same (but unknown) probability distribution. In the context of this paper, the process is stationary if it enters its steady-state region, and steady state simulation is concerned with modeling and analysis of such processes only.

Any function $f(X_1, X_2, \cdots, X_n)$ of random variables $X_1, X_2, \cdots, X_n$ is called a *statistic*. A statistic $\hat{\theta}(X_1, X_2, \cdots, X_n)$ used for estimating a parameter θ of the distribution of the analysed process $X(t)$ is called an estimator of θ. The number of independent random variables used in the definition of a statistic is called its *number of degrees of freedom* or, simply, its *degrees of freedom*. The value $\hat{\theta}(x_1, x_2, \cdots, x_n)$, that is, the value of function $\hat{\theta}(X_1, X_2, \cdots, X_n)$ for a given sequence of observations $x_1, x_2, \cdots, x_n$ is known as an *estimate* of θ. It means that estimators, as functions of random variables, are random variables themselves, and estimates are simply realisations of these random variables.

A good estimator should be *unbiased*; see [Trivedi 1982] for other desired properties of estimators. If the estimator $\hat{\theta}^* = \hat{\theta}^*(X_1, X2, \cdots, X_n)$ of the parameter θ is unbiased, then its statistical average

$$\mathrm{E}[\hat{\theta}^*(X_1, X_2, \cdots, X_n)] = \theta \qquad (A1)$$

For example, it can be shown that the estimator $\overline{X}(n)$ given in Equation (1) is an unbiased estimator of the average μ_x of a stationary process $X(t)$, since $\mathrm{E}[\overline{X}(n)] = \mu_x$. Similarly, one can show that if random variables $X_1, X_2, \cdots, X_n$ are independent and identically distributed, then the unbiased estimator of the variance $\sigma^2[X]$ of the stationary process $X(t)$ is

$$\sigma^2[X] = \sum_{i=1}^{n} \frac{\{X_i - \overline{X}(n)\}^2}{(n-1)}, \qquad (A2)$$

while the unbiased estimator of the variance of the estimator $\overline{X}(n)$ is

$$\sigma^2[\overline{X}(n)] = \frac{\sigma^2[X]}{n}, \qquad (A3)$$

see Equation (4).

Theoretical justification for using $\overline{X}(n)$ as the estimator of μ_x is provided by the *law of large numbers*, which states that if observations $x_1, x_2, \cdots, x_n$ are realizations of independent random variables, taken from a stationary process $X(t)$, then

$$\lim_{n\to 0} P[|\overline{X}(n) - \mu_x| \leq \epsilon] = 1, \qquad (A4)$$

for any $\epsilon > 0$.

To construct confidence intervals for $\overline{X}(n)$ we have to assume a probability distribution for $\overline{X}(n)$. If $\overline{X}(n)$ is the weighted sum of independent and normally distributed random variables $X_i, i = 1, 2, \cdots, n$, with unknown variance $\sigma^2[X]$, then it can be proved that the statistic

$$T_{n-1} = \frac{\overline{X}(n) - \mu_x}{\sigma[\overline{X}(n)]}, \qquad (A5)$$

has the standard (Student) t-distribution with (n-1) degrees of freedom. Using this distribution, one can easily find the value $t = t_{n-1,1-\alpha/2}$ for it, given $\alpha, 0 < \alpha < 1$,

$$P(-t_{n-1,1-\alpha/2} \leq T_{n-1} \leq t_{n-1,1-\alpha/2})$$
$$= 1 - \alpha. \qquad (A6)$$

Consequently, the confidence interval of μ_x at the $(1 - \alpha)$ confidence level is given by

$$P(\overline{X}(n) - t_{n-1,1-\alpha/2}\hat{\sigma}[\overline{X}(n)] \leq \mu_x$$
$$\leq \overline{X}(n) + t_{n-1,1-\alpha/2}\hat{\sigma}[\overline{X}(n)]) = 1 - \alpha, \quad (A7)$$

see Figure 1. The limiting t-distribution, for $n \to \infty$ is the standard normal distribution described by the density function

$$p_z(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2). \qquad (A8)$$

In practice, the approximation of a t-distribution by the normal distribution is already acceptable for samples of $n > 30$ observations. In such cases, one can easily find the value $z = z_{1-\alpha/2}$ from the standard normal distribution for which

$$P(\overline{X}(n) - z_{1-\alpha/2}\hat{\sigma}[\overline{X}(n)] \leq \mu_x$$
$$\leq \overline{X}(n) + z_{1-\alpha/2}\hat{\sigma}[\overline{X}(n)]) = 1 - \alpha, (A9)$$

which is a good approximation for the confidence interval of μ_x at the $(1-\alpha)$ confidence level if $n > 30$. In the case of multivariate estimation, confidence intervals are replaced by confidence regions (rectangles, etc.)

A justification of the assumption that the estimator $\overline{X}(n)$ is a normal random variable even if observations $x_1, x_2, \cdots, x_n$ are not is given by the *central limit theorem*. This theorem says that the sums of independent and identically distributed random variables tend to be normally distributed, even though the components are not, if $n \to \infty$. From this theorem, one can show that the limit distribution (for $n \to \infty$) of the random variable

$$Z = \frac{\overline{X}(n) - \mu_x}{\sigma[X]/\sqrt{n}} \qquad (A10)$$

is given by Equation (A8). Note that it requires the variance $\sigma^2[X]$ of the random variables $X_1, X_2, \cdots, X_n$ to be known. Since this is not the case in practice, if $X_1, X_2, \cdots, X_n$ are not normally distributed, the limiting distribution of the statistics T_{n-1} can be only approximately normal (even if the variables are statistically independent). The quality of this approximation depends on the accuracy of the variance estimator.

APPENDIX B. SPECTRAL ANALYSIS OF VARIANCE.

As was mentioned in Section 1, Heidelberger and Welch [1981a, 1981b] proposed estimating the variance of the estimator of mean in an autocorrelated sequence of observations from the smoothed averaged periodogram of this sequence; see Equations (45)-(50). The

series of approximations they propose can be summarised as follows.

Procedure *Spectral Variance Analysis*;
{Preconditions:

$x_1, x_2, \cdots, x_{n_v}$ A sequence of n_v observations taken from a stationary process ($n_v \geq 100$, the default value is $n_v = 100$).

n_{ap} The number of points of the averaged periodogram used to fit it to a polynomial by applying the least squares procedure ($n_{ap} \leq n_v/4$, the default value is $n_{ap} = 25$).

δ The degree of the polynomial fitted to the logarithm of the averaged periodogram (the default value is $\delta = 2$).

C_σ A normalizing constant, chosen to make $\hat{p}_x(0)$ approximately unbiased (for the default values of n_{ap} and δ, take $C_\sigma = 0.882$ [Heidelberger and Welch, Table 1], where these constants are denoted as K, d, and $C_1(K, d)$, respectively)}

Step 1
Calculate $2n_{ap}$ values of the periodogram of the sequence $x_1, x_2, \cdots, x_{n_v}$:

$\Pi(\frac{j}{n_v})$

$$= |\sum_{s=1}^{n_v} x_s \exp[-\frac{2\pi i(s-1)j}{n_v}]|^2 / n_v, \qquad (B1)$$

{*cf. Eqs. (45) and (46)*}
for $j = 1, 2, \cdots, 2n_{ap}$, and $i = \sqrt{-1}$;

Step 2
Calculate n_{ap} values of the function $\{L(fj)\}$, for $j = 1, 2, \cdots, n_{ap}$; where $fj = (4j - 1)/2n_v$, and

$L(fj)$

$$= \log\{[\Pi((2j-1)/n_v) + \Pi(2j/n_v)]/2\}; (B2)$$

Step 3 Apply the least squares extrapolation procedure {*see, for example, Press et al. [1986, p. 509]*} *for determining the coefficient a_o in the polynomial*

$$g(f) = \sum_{s=0}^{\delta} a_s f^s, \qquad (B3)$$

fitted to the function $\{L(fj) + 0.270\}$, $j = 1, 2, \cdots, n_{ap}$;
{*the value a_o is an unbiased estimate of log $p_x(0)$*}

Step 4
Calculate

$$\hat{p}_x(0) = C_\sigma e^{a_o}; \qquad (B4)$$

$$\hat{\sigma}_{sp}^2[\overline{X}(n_v)] = \hat{p}_x(0)/n_v; \qquad (B5)$$

determine κ;
{*the degrees of freedom κ for the χ^2 distribution of $\hat{\sigma}_{sp}^2[\overline{X}(n_v)]$, for given n_{ap} and δ, are given in Heidelberger and Welch [1981, Table 1] and denoted there by $C_2(K, d)$; for $n_v = 100, n_{ap} = 25$ and $\delta = 2 : \kappa = 7$*}
end *Spectral Variance Analysis*.

If an analysed sequence of output data consists of n_v batch means $\overline{X}_1(m)$, $\overline{X}_2(m)$, $\cdots$, $\overline{X}_{n_v}(m)$, each over m observations, instead of n_v individual observations x_1, x_2, $\cdots$, x_{n_v} as above, the batch means should replace individual observations in Equation (B1). Then the periodogram $\{\Pi(j/n)\}$ and the logarithm of the averaged periodogram $\{L(fj)\}$ are functions of the sequence $\overline{X}_1(m)$, $\overline{X}_2(m)$, $\cdots$, $\overline{X}_{n_v}(m)$, and Equation (B4) gives us the estimate $\hat{p}_{\overline{X}(m)}(0)$, the spectral density function of this sequence of batch means at $f = 0$. Now, with n_v meaning the number of batch means, each over m observations, we obtain

$$\hat{p}_x(0) = m\hat{p}_{\overline{X}(m)}(0); \qquad (B6)$$

see Equation (49). Thus, the estimator of the variance of the mean $\overline{X}(n_v m)$, over $n_v m$ individual observations, is

$$\hat{\sigma}^2_{sp}[\overline{X}(n_v m)] = \frac{\hat{p}_{\overline{X}(m)}(0)}{n_v}. \qquad (B7)$$

The degrees of freedom can be determined as in the previous case; see the comment in Step 4 of the procedure.

ACKNOWLEDGEMENTS

References

[1] J. Abate and W.Whitt. Transient behavior of regulated brownian motion, i: Starting at the origin. In *Adv. Appl. Probab.*, volume 19, pages 560–598, 1987a.

[2] J. Abate and W.Whitt. Transient behavior of regulated brownian motion, ii: Non-zero initial conditions. In *Adv. App. Probab.*, volume 19, pages 599–631, 1987b.

[3] J. Abate and W.Whitt. Transient behavior of the m/m/1 queue: Starting at the origin. In *Queue. Sys.*, volume 2, pages 41–65, 1987c.

[4] J. Abate and W.Whitt. Transient behavior of the m/m/1 queue via laplace transform. In *Adv. Appl. Probab.*, volume 20, pages 145–178, 1988.

[5] N.R. Adam. Achieving a confidence interval for parameters estimated by simulation. In *Management Sci.*, volume 29, pages 856–866, 1983.

[6] T. Amemiya. Generalized least squares with an estimated autocovariance matrix. In *Econometria*, volume 41, pages 723–732, 1973.

[7] K. Anderson. Some steady-state problems in simulation. In *Proceedings of Annual Simulation Symposium*, pages 79–93. IEEE, 1985.

[8] R.W. Andrews and T.J.Schriber. Interactive analysis of output from gpss-based simulations. In *Proceedingsof the 1978 Winter Simulation Conference*, pages 267–278. ACM/IEEE, 1978.

[9] R. Anonuevo and B.L. Nelson. Automated estimation and variance reduction for steady state simulation. In *Proc of the 1986 Winter Simulation Conference*, pages 871–875. ACM/IEEE, 1986.

[10] Artificial Intelligence and Simulation:. The diversity of applications. In *Proceedings of the SCS Multiconference*. Society for Computer Simulation, 1988.

[11] M. Asgarkhani and K.Pawlikowski. Simulation studies of mixed traffic on satellite channels using tdma-reservation protocol. In *Proceedings of the 8th Phoenix International Conference on Computers and Communications*, pages 195–200, 1989.

[12] D. Baik and B.P. Zeigler. Performance evaluation of hierarchical distributed

simulators. In *Proceedings of the 1985 Winter Simulation Conference*, pages 421–427. IEEE, 1985.

[13] D.W. Balmer. Polishing the analysis of the statistical output of comparative simulation experiments. *Simulation*, pages 123–126, 1987.

[14] J. Banks and J.S. Carson. In *Discrete Event System Simulation*, 1984.

[15] M.S. Bartlett. On the theoretical specification of the sampling properties of autocorrelated time series. In *J. Roy. Statist. Soc. Ser. B,*, pages 27–41, 1946.

[16] K.W. S. Venkatraman Bauer and J.R.Wilson. Estimation procedures based on contorl variates with known covariance matrix. In *Proc. 1987 Winter Simulation Conf.*, pages 334–341. IEEE, 1987.

[17] C.W. Beall. A regression technique or determining steady state conditions in time series simulations. In *Proc. 1982 Winter Simulation Conf.*, pages 439–448. IEEE, 1982.

[18] P.C. Bell and R.M. O'Kneefe. A regression technique or determining steady state conditions in time series simulations. In *Proc. 1982 Winter Simulation Conf.*, pages 439–448. IEEE, 1982.

[19] K. Bharath-Kumar and P. Kermani. Performance evaluation tool (pet): An analysis tool for computer communication networks. In *J. Select . Areas Commun.*, pages 229–225. IEEE, 1984.

[20] P. Billingsley. Convergence of probability measures. 1968.

[21] R.B. Blackman and J.W. Tuckey. The measurement of power spectra from the point of view of communications engineering (part 1. and part 2.). In *Bell Systems Tech. J.*, pages 185–282 and 485–569, 1958.

[22] J.P.C. Blanc. The transient behaviour of networks with infinite servers nodes. In *Performance '84, ed. E. Gelenbe*, pages 159–174, 1984.

[23] J.P.C. Blanc. The relaxation time of two queueing systems in series. In *Commun. Statist. Stochastic Models*, volume 1, pages 1–16, 1985.

[24] N. Blomqvist. The covariance function of the m/g/1 queueing system. In *Skandinavisk Aktuarietidskrift*, pages 157–174, 1967.

[25] N. Blomqvist. On the transient behavior of the gi/g/1 waiting-times. In *Skandinavisk Aktuarietidskrift*, pages 118–129, 1970.

[26] et al. Blum, A. Experiments with decomposition of extended queueing network models. In *Modelling Techniques and Tools for Performance Analysis*, pages 623–638, 1985.

[27] G.E.P. Box and G.M. Jenkins. In *Time Series Analysis, Forecasting and Control*, 1970.

[28] P. B.L. Fox Bratley and L.E. Schrage. A guide to simulation. *Springer-Verlag*, 1983.

[29] D.R. Brillinger. Estimation of the mean of a stationary time series by sampling. In 1973, editor, *J. of Applied Probability*, 419–431.

[30] D.R. Brillinger. Time series: Data analysis and theory. *Holden-Day*, 1981.

[31] P.J. Brockwell and R.A. Davis. Time series: Theory and methods. *Springer-Verlag*, 1987.

[32] W.G. Bulgren. Discrete system simulation. *Prentice-Hall*, 1982.

[33] J.M. Calvin. Covariance of regenrative mean and variance estimators for the markov chains. In *Proc. 1988 Winter Simulation Conf.*, pages 473–475. IEEE, 1988.

[34] J.S. Carson and A.M. Law. Conservation equations and variance reduction in queueing simulations. In *Operations Res.*, pages 535–546, 1980.

[35] Catalog. of simulation software. *Simulation*, pages 165–181, 1987.

[36] K.M. Chandy and J.Misra. Asynchronous distributed simulation via a sequence of parallel computations. In *Comm. of the ACM*, pages 198–205, 1981.

[37] J.M. Charnes and W.D. Kelton. A comparison of confidence region estimators for multivariate simulation output. In *Proc. 1988 Winter Simulation Conf.*, pages 458–465. IEEE, 1988.

[38] R.C.H. Cheng. A note on the effect of initial conditions on a simulation run. In *Operational Res. Quart.*, pages 467–470, 1976.

[39] R.C.H. Cheng. Variance reduction methods. In *Proc. 1986 Winter Simulation Conf.*, pages 60–68. IEEE, 1986.

[40] D. R. Chen and A.F. Seila. Multivariate inference in stationary simulation using batch means. In *Proc. 1987 Winter Simulation Conf.*, pages 302–304. IEEE, 1987.

[41] G.M. Clark. A bonferroni selection procedure when using common numbers with unknown variances. In *Proc. 1987 Winter Simulation Conf.*, pages 313–315. IEEE, 1986.

[42] J.W. Cohen. The single server queue. 1982.

[43] R.W. Conway. Some tactical problems in digital simulation. In *Management Sci.*, pages 47–61, 1963.

[44] M. J.C.Fort Cottrell and G.Malgouyres. Largedeviations and rare events in the study of stochastic algorithms. In *IEEE Trans. Automatic Control*, pages 907–918, 1983.

[45] D.R. Cox and S.W.L. Smith. Queues. *Wiley, N.Y.*, 1961.

[46] M. Crane and D.L. Iglehart. Simulating stable stochastic systems: I. general multiserver queues. In *J. of the ACM*, pages 103–113, 1974.

[47] M.A. Crane and D.L. Iglehart. Simulating stable stochastic systems: Iii. regenerative processes and discrete-event simulations. In *Operations Res.*, pages 33–45, 1975.

[48] M.A. Crane and D.L. Iglehart. Simulating stable stochastic systems. iv: Approximation techniques. In *Management Sci.*, pages 1215–1224, 1975.

[49] M.A. Crane and A.J. Lemoine. An introduction to the regenerative method for simulation analysis. In *Lecture Notes in Control and Information Science, No. 4, A.W. Balakrishnan and M. Thoma (eds), Springer-Verlag*, 1977.

[50] H. Damerdji. On strong consistency of the variance estimator. In *Proc. 1987 Winter Simulation Conf.*, pages 305–308. IEEE, 1987.

[51] D.J. Daley. The serial correlation coefficients of waiting times in a stationary simple server queue. In *J. Australian Math. Society*, pages 683–699, 1968.

[52] A. DeCegama. Parallel processing simulation of large computer networks. In *Proc. 1987 Symposium on the Simulation of Computer Networks*, pages 51–62, Colorado, 1987. IEEE.

[53] J.N. Donnelly and R.E. Shannon. Minimum mean-squared error estimates for simulation experiments. In *Comm. of the ACM*, pages 253–259, 1981.

[54] R.R. Duersch and L.W. Schruben. An interactive run length control for simulations on pc's. In *Proc. 1986 Winter Simulation Conf.*, pages 866–870. IEEE, 1986.

[55] S.D. Duket and A.A.B. Pritsker. Examination of simulation output using spectral methods. In *Math. Comput. Simul.*, pages 53–60, 1978.

[56] J.R. Emshoff and R.L. Sisson. Design and use of computer simulation models. 1970.

[57] G.S. Fishman and P.J. Kiviat. The analysis of simulation-generated time series. In *Management Sci.*, pages 525–557, 1967.

[58] G.S. Fishman. Problems in the statistical analysis of simulation experiments: The comparison of means and the length of sample records. In *Comm. of the ACM*, pages 94–99, 1967.

[59] G.S. Fishman. Estimating sample size in computing simulation experiments. In *Management Sci.*, pages 21–38, 1971.

[60] G.S. Fishman. Bias considerations in simulation experiments. In *Operations Res.*, pages 685–790, 1972.

[61] G.S. Fishman. Concepts and methods in discrete event digital simulation. 1973.

[62] G.S Fishman. Statistical analysis for queueing simulation. In *Management Sci.*, pages 363–369, 1973.

[63] G.S. Fishman. Estimation in multiserver queueing simulations. In *Operations Res.*, pages 72–78, 1974.

[64] G.S. Fishman. Achieving specific accuracy in simulation output analysis. In *Comm. of the ACM*, pages 310–315, 1977.

[65] G.S. Fishman. Principles of discrete event simulation. *Wiley, N. Y.*, pages 510–521, 1978.

[66] G.S. Fishman. Grouping observations in digital simulation. In *Management Sci.*, pages 510–521, 1978.

[67] R.D. Foley and D. Goldsman. Confidence intervals and orthonormally weighted standardized time series. In *Proc. 1988 Winter Simulation Conf.*, pages 422–424, N.Y., 1988. IEEE.

[68] B. Fox. Estimation and simulation. In *Management Sci.*, pages 860–861, 1978.

[69] L.W. Friedman and H.H. Friedman. Comparing simulated alternatives using a distributionfree statistic with blocking by random number stream. In *Simulation*, pages 68–70, 1986.

[70] L.W. Friedman. Multivariate simulation output analysis: Past, present, and future. In *Proc. 1984 Winter Simulation Conf.*, pages 277–281. IEEE, 1984.

[71] W.W. Laure Frost, V.S. and K.S. Shanmugan. Efficient techniques for simulation of computer communication networks. In *IEEE J. on Selected Areas in Comm.*, pages 146–157, 1988.

[72] R.M. Fujimoto. Performance measurements of distributed simulation strategies. In *Proc. of Distributed Simulation 1988, San Diego, USA, Society for Computer Simulation*, pages 14–20, 1988.

[73] et al. Gafarian, A.V. Evaluation of commonly used rules for detecting 'steady state' in computer simulation. In *Naval Research Logistics Quart.*, pages 511–529, 1978.

[74] B. S.Cammarata Gates and J.Rothenberg. A demon facility for object-oriented simulation

language. In *Proc. 1988 Summer Simulation Conf., Society for Computer Simulation*, pages 667–673, 1988.

[75] M.A. Geisler. The size of simulation samples required to compute certain inventory characteristics with stated precision and confidence. In *Management Sci.*, pages 261–271, 1964.

[76] P.W. Glynn. Coverage error for confidence intervals arising in simulation output analysis. In *Proc. 1982 Winter Simulation Conf.*, pages 369–375. IEEE, 1982.

[77] P.W. Glynn and D.L. Iglehart. Large-sample theory for standardized time series: An overview. In *Proc. 1985 Winter Simulaton Conf.*, pages 129–134. IEEE, 1985.

[78] P.W. Glynn and W. Whitt. Indirect estimation via l=λw. In *Management Sci.*, pages 82–103, 1989.

[79] D. Goldsman. Ranking and selection in simulation. In *Proc. 1983 Winter Simulation Conf.*, pages 387–393. IEEE, 1983.

[80] D. Goldsman and L.Schruben. Asymptotic properties of some confidence interval estimators for simulation output. In *Management Sci.*, pages 1217–1225, 1984.

[81] D. Goldsman and M.S. Meketon. A comparison between standardized time series and overlapping batched means. In *Proc. 1985 Winter Simulation Conf.*, pages 208–210. IEEE, 1985.

[82] K. Kang Goldsman, D. and R.G. Sargent. Large and small sample comparisons of various variance estimators. In *Proc. 1986 Winter Simulation Conf.*, pages 278–284. IEEE, 1986.

[83] D. Goldsman. Tutorial on indifference -zone normal means ranking and selection procedures. In *Proc. 1986 Winter Simulation Conf.*, pages 370–375. IEEE, 1986.

[84] D. Goldsman and M.S. Meketon. A comparison of several variance estimators. In *Operations Res.*, 1989.

[85] G. Gordon. System simulation. 1969.

[86] Gross D. and C.M. Harris. Fundamentals of queueing theory. 1974.

[87] F.L. Gunther and R.W. Wolff. The almost regenerative method for stochastic system simulations. In *Operations Res.*, pages 387–393, 1980.

[88] J. Haddock. An expert system framework based on a simulation generator. In *Simulation*, pages 45–53, 1987.

[89] G.J. Hahn. More intelligent statistical software and statistical expert systems: Future directions. In *The American Statistician*, pages 1–16, 1985.

[90] R.W. Hamming. Numerical methods for scientists and engineers. 1962.

[91] D.J. Hand. Statistical expert systems: Necessary attributes. In *J. of Applied Statistics*, pages 19–27, 1985.

[92] E.J. Hannan. Multiple time series. 1970.

[93] J. Harling. Simulation techniques in operations research - a review. In *Operations Res.*, pages 307–319, 1958.

[94] P. Heidelberger. A variance reduction technique that increases regeneration frequency. In *Current Issues in Computer Simulation*, pages 257–269, 1979.

[95] P. Heidelberger and P.D. Welch. A spectral method for confidence interval generation and run length control in simulations. In *Comm. of the ACM*, pages 233–45, 1981.

[96] P. Heidelberger and P.D. Welch. Adaptive spectral methods for simulation output analysis. In *IBM Research Develop.*, pages 860–876, 1981.

[97] P. Heidelberger and P.A.W. Lewis. Regression - adjusted estimates for regenerative simulations, with graphics. In *Comm. of the ACM*, pages 260–273, 1981.

[98] P. Heidelberger and P.D. Welch. Simulation run length control in the presence of an initial transient. In 1983, editor, *Operations Res.*, 1109–1144.

[99] P. Heidelberger and P.A.W. Lewis. Quantile estimation in dependent sequences. In *Operations Res.*, pages 185–209, 1984.

[100] P. Heidelberger. Statistical analysis of parallel simulation. In *Proc. 1986 Winter Simulation Conf.*, pages 290–295. IEEE, 1986.

[101] Y. C. Ho. Performance evaluation and perturbation analysis of discrete event dynamic systems. In *IEEE Trans.*

on *Automatic Control*, pages 563–572, 1987.

[102] Y.C. Ho and X.Cao. Perturbation analysis and optimization of queueing networks. In *J. Optimization Theory Applications*, pages 559–582, 1983.

[103] X.Cao Ho, Y.C. and C.Cassandras. Infinitesimal and finite perturbation analysis for queueing networks. In *Automatica*, pages 439–445, 1983.

[104] D.L. Iglehart. Simulating stable stochastic systems. v: Comparison of ratio estimators. In *Naval Research Logistics Quart.*, pages 553–565, 1975.

[105] D.L. Iglehart. Simulating stable stochastic systems vi: Quantile estimation. In *J. of the ACM*, pages 347–360, 1976.

[106] D.L. Iglehart. The regenerative method for simulation analysis. In *Current Trends in Programming Methodology*, pages 52–71, 1978.

[107] J. Mierzwa Izydorczyk, J. and A. Wolisz. A control variable method for increasing the efficiency of multiple access protocol simulation. In *Performance of Computer Communication Systems*, pages 95–108, 1984.

[108] J. Jackman and D.J.Medeiros. A graphical methodology for simulating communication networks. In *IEEE Trans. on Comm.*, pages 459–464, 1988.

[109] R. Jain and I.Chlamtac. The p2 algorithm for dynamic calculation of quantiles and histograms without storing observations. In *Comm. of the ACM*, pages 1076–1085, 1985.

[110] G.M. Jenkins and D.G. Watts. Spectral analysis and its applications. 1968.

[111] K. Kang and D.Goldsman. The correlation between mean and variance estimators. In *Proc.1985 Winter Simulation Conf.*, pages 211–216, 1985.

[112] W.D. Kelton and A.M. Law. A new approach for dealing with the startup problem in discrete event simulation. In *Naval Research Logistics Quart.*, pages 641–658, 1983.

[113] W.D. Kelton and A.M. Law. An analytical evaluation of alternative strategies in steady-state simulation. In *Operations Res.*, pages 169–184, 1984.

[114] W.D. Kelton and A.W. Law. The transient behaviour of the m/m/s queue, with implications for steady-state simulation. In *Operations Res.*, pages 378–396, 1985.

[115] W.D. Kelton. Transient exponential-erlang queues and steady-state simulation. In *Comm. of the ACM*, pages 741–749, 1985.

[116] W.D. Kelton. Statistical design and analysis. In *Proc. 1986 Winter Simulation Conf.*, pages 45–51. IEEE, 1986.

[117] E.J.H. Kerckhoffs and G.C.Vansteenkiste. The impact of advanced information processing on simulation - an illustrative review. In *Simulation*, pages 17–26, 1986.

[118] J. Kiefer and J. Wolfowitz. On the theory of queues with many servers. In *Trans. of the American Math. Society*, pages 1–18, 1955.

[119] J.P.C. Kleijnen. Statistical techniques in simulation. *Marcel Dekker*, 1, 1974.

[120] J.P.C. Kleijnen. The role of statistical methodology in simulation. In *Methodology in Systems Modelling and Simulation*, 1979.

[121] R. Van der Ven Kleijnen, J.P.C. and B. Saunder. Testing independence of simulation subruns: A note on the power of the von neumann test. In *Eur. J. Operational Res.*, pages 92–93, 1982.

[122] J.P.C. Kleijnen. Statistical tools for simulation practitioners. *Marcel Dekker*, 1987.

[123] L. Kleinrock. Queueing systems, vol.2: Computer applications. *Wiley*, 2, 1976.

[124] V. Knapp. The smalltalk simulation environment. In *Proc. 1986 Winter Simulation Conf.*, pages 125–128. IEEE, 1986.

[125] W. Kreutzer. System simulation. programming styles and languages. *Addison Wesley*, 1986.

[126] U. Chandra sekaran Krishnamurthy, M. and S.Sheppard. Two approaches to the implementation of a distributed simulation system. In *Proc. 1985 Winter Simulation Conf.*, pages 435–443. IEEE, 1985.

[127] D. Kumar. A novel approach to sequential simulation. *IEEE Software*, pages 25–33, 1986.

[128] J.F. Kurose and H.T. Mouftah. Computer-aided modeling, analysis, and design of communication networks. In *IEEE J. Select. Areas Commun.*, pages 130–145, 1988.

[129] S.S. Lavenberg and C.H. Sauer. Sequential stopping rules for the regenerative method of simulation. In *IBM J. Research Develop.*, pages 545–558, 1977.

[130] et. al. Lavenberg, S.S. The initial transient in steady state simulation (panel discussion). In *Proc. 1981 Winter Simulation Conf.*, pages 113–120. IEEE, 1981.

[131] S.S. Lavenberg and P.D. Welch. A perspective on the use of control variables to increase the efficiency of monte carlo simulations. In *Management Sci.*, pages 322–335, 1981.

[132] T.L. Moeller Lavenberg, S.S. and P.D. Welch. Statistical results on control variables with application to queueing network simulation. In *Operations Res.*, pages 182–202, 1982.

[133] A.M. Law. Eficient estimators for simulated queueing systems. In *Management Sci.*, pages 30–41, 1975.

[134] A.M. Law. A confidence interval in discrete event simulation: A comparison of replication and batch means. In *Naval Research Logistics Quart.*, pages 667–678, 1977.

[135] A.M. Law. Statistical analysis of the output data from terminating simulations. In *Naval Research Logistics Quart.*, pages 131–143, 1980.

[136] A.M. Law. Statistical analysis of simulation output data. In *Operations Res.*, pages 983–1029, 1983.

[137] A.M. Law and J.C. Carson. A sequential procedure for determining the lenght of a steady state simulation. In *Operations Res.*, pages 1011–1025, 1979.

[138] A.M. Law and D.W.Kelton . Simulation, modeling and analysis. In *McGraw-Hill*, 1982.

[139] A.M. Law and D.W.Kelton. Confidence intervals for steady state simulations. ii: A survey of sequential procedures. In *Management Sci.*, pages 550–562, 1982.

[140] A.M. Law and W.D. Kelton. Confidence intervals for steady state simulations. i: A survey of fixed sample size procedures. In *Operations Res.*, pages 1221–1239, 1984.

[141] E.A. MacNair. An introduction to the research queueing package. In *Proc. 1985 Winter Simulation Conf.*, pages 257–262. IEEE, 1985.

[142] A. Madansky. Optimal initial conditions for a simulation problem. In *Operations Res.*, pages 172–577, 1976.

[143] N.B. Marks. Further investigation into spectral analysis for confidence intervals in steady state simulation. In *Proc. 1981 Winter Simulation Conf.*, pages 461–464. IEEE, 1981.

[144] M.S. Meketon. The variance time-curve: Theory, estimation, and application. In *Ph.D. Dissertation, School of Operations Research and Industrial Engineering*, Cornell University, Ithaca, N.Y., 1980.

[145] M.S. Meketon and P. Heidelberger. A renewal theoretic approach to bias reduction on regenerative simulations. In *Management Sci.*, pages 173–181, 1982.

[146] M.S. Meketon and B.Schmeiser. Overlapping batch means: Something for nothing ? In *Proc.1984 Winter Simulation Conf.*, pages 227–230. IEEE, 1984.

[147] R.G. Miller. The jackknife - a review. *Biometrika*, pages 1–15, 1974.

[148] J.M. Mellichamp and Y.H.Park. A statistical expert system for simulation analysis. In *Simulation*, pages 134–139, 1989.

[149] R. Mills. Statistical analysis of steady state simulation output data with simscript ii.5. In *CACI*, Los Angeles, 1987.

[150] D.L. Minh. Simulating gi/g/k queues in heavy traffic. In *Management Sci.*, pages 1192–1199, 1987.

[151] J. Misra. Distributed discrete event simulation. In *ACM Computing Surveys*, pages 39–66, 1986.

[152] I. Mitrani. Simulation techniques for discrete event systems. In *Cambridge Academic Press*, 1982.

[153] P.M. Morse. Stochastic processes of waiting lines. In *Operations Res.*, pages 255–261, 1955.

[154] J.R. Murray and W.D.Kelton. Initializing for bias reduction: Some analytical considerations. In *Proc.1988 Winter Simulation Conf.*, pages 546–548. IEEE, 1988.

[155] B.L. Nelson and B.W. Schmeiser. Decomposition of some well-known variance reduction techniques. In *J.Statist.Comput.Simul.*, pages 183–209, 1986.

[156] G.F. Newell. Applications of queueing theory. In *Chapman & Hall*, London, 1971.

[157] D.M. Nicol. Parallel discrete event simulation of fcfs stochastic queueing networks. In *Parallel Programming: Experience with Applications, Languages and Systems*, pages 124–137. ACM SIGPLAN, 1988.

[158] D.M. Nicol. High performance parallelized discrete-event simulation of stochastic queueing networks. In *Proc. 1988 Winter Simulation Conf.*, pages 237–241. IEEE, 1988.

[159] N.R. Nielsen. Expert systems and simulation. In *Computer Networks and Simulation III*, pages 41–52, S. Schoemaker (ed.), North-Holland, 1986.

[160] A.R. Odoni and E. Roth. An empirical investigation of the transient behaviour of stationary queueing systems. In *Operations Res.*, pages 306–314, 1983.

[161] R.M. O'Keefe. Simulation and expert systems - a taxonomy and some examples. In *Simulation*, pages 10–16, 1986.

[162] T.I. O ren and B.P. Zeigler. Artificial intelligence in modelling and simulation: Directions to explore. In *Simulation*, pages 131–134, 1987.

[163] S. Parekh and J.Walrand. A quick simulation of excessive backlogs in networks of queues. In *IEEE Trans. Automatic Control*, pages 54–66, 1989.

[164] S.K. Park and K.W. Miller. Random number generators: Good ones are hard to find. In *Comm. of the ACM*, pages 1992–1201, 1988.

[165] E. Parzen. Mathematical consideration in the estimation of spectra. In *Technometrics*, pages 167–190, 1961.

[166] K. Pawlikowski and M. Asgarkhani. Sequential procedures in simulation studies of satellite protocols, studies in telecommunication. In *Proc. ITC'12, Torino, Italy, 1988, Teletraffic Science for New Cost-Effective Systems, Systems and Services*, volume 6, 12, pages 4.3B.3.1–4.3B.3.7, North Holland, 1988, 1989.

[167] J.A. Payne. Introduction to simulation. *McGraw-Hill, N.Y.*, 1982.

[168] M. Pidd. Computer simulation in managment science. *Wiley , Chichester*, 1984.

[169] D. Potier. New user's introduction to qnap2. In *Rapport Technique INRIA*, number 40, 1984.

[170] D. Potier. The modeling package qnap2 and applications to computer networks simulation. In *Computer Networks and Simulation III*, pages 235–265, North Holland, 1986.

[171] et al. Press, W.H. Numerical recipes. the art of scientific computing. In *Cambridge University Press*, 1986.

[172] K.E.E. Raatikainen. Simultaneous estimation of several percentiles. In *Simulation*, pages 159–164, 1987.

[173] K.E.E. Raatikainen. Validating percentiles of response times in queueing network models. In *Proc. of 1988 Summer Simulation Conf.*, pages 136–141. Society for Computer Simulation, 1988.

[174] R. Reddy. Epistemology of knowledge based simulation. In *Simulation*, pages 162–166, 1987.

[175] M.I. Reiman and A.Weiss. Sensitivity analysis via likelihood ratios. In *Proc. 1986 Winter Simulation Conf.*, pages 285–289. IEEE, 1986.

[176] P.F. Reynolds. A spectrum of options for parallel simulation. In *Proc. 1988 Winter Simulation Conf.*, pages 325–332. IEEE, 1988.

[177] E. Roth and A.H. Rutan. A relaxation time approach for reducing initialization bias in simulation. In *Record of Proceedings, Annual Simulation Symp.*, pages 189–203, Tampa, Florida, 1985. IEEE Computer Society.

[178] E. Roth. A relaxation time heuristic for reducing initialization bias in simulation of m/ek/1 queueing systems. In *Proc. 1985 Winter Simulation Conf.*, pages 263–268. IEEE, 1985.

[179] R.Y. Rubinstein and R. Marcus. Efficiency of multivariate control variates in monte carlo simulation. In *Operations Res.*, pages 661–677, 1985.

[180] R.Y. Rubinstein. Sensitivity analysis and performance extrapolation for computer simulation models. In *Operations Res.*, pages 72–81, 1989.

[181] S. Ruiz-Mier and J.Talavage. A hybrid paradigm for modeling of complex systems. In *Simulation*, pages 135–141, 1987.

[182] R.G. Sargent. The use of graphical models in model validation. In *Proc. 1986 Winter Simulation Conf.*, pages 237–241. IEEE, 1986.

[183] E.A. MacNair Sauer, C.H. and J.K. Kurose. Queueing network simulations of computer communications. In *IEEE J. Select Areas Commun.*, pages 203–220, 1984.

[184] B. Schmeiser. Batch size effects in the analysis of simulation output. In *Operations Res.*, pages 556–568, 1982.

[185] B. Schmeiser and W.T. Song. orrelation among estimators of the variance of the sample mean. In *Proc. 1987 Winter Simulation Conf.*, pages 309–317. IEEE, 1987.

[186] J.W. Schmidt and C.Ho. An algorithm for testing dependence of simulation output data. In *Proc. 1988 Winter Simulation Conf.*, pages 532–539. IEEE, 1988.

[187] T. Schriber. Simulation using gpss. Wiley, N.Y., 1974.

[188] T.J. Schriber and R.W. Andrews. Interactive analysis of simulation output by the method of batch means. In *Proc. 1979 Winter Simulation Conf.*, pages 512–524. IEEE, 1979.

[189] T.J. Schriber and R.W. Andrews. A conceptual framework for research in the analysis of simulation output. In *Comm. of the ACM*, pages 218–232, 1981.

[190] L.W. Schruben. A coverage function for interval estimators of simulation response. In *Management Sci.*, pages 18–27, 1980.

[191] L.W. Schruben. Control of initialization bias in multivariate simulation response. In *Comm. of the ACM*, pages 246–252, 1981.

[192] L.W. Schruben. Detecting initialisation bias in simulation output. In *Operations Res.*, pages 569–590, 1982.

[193] H. Singh Schruben, L.W. and L. Tierney. Optimal tests for initialisation bias in simulation output. In *Operations Res.*, pages 1167–1178, 1983.

[194] L. Schruben. Confidence interval estimation using standardised time series. In *Operations Res.*, pages 1090–1108, 1983.

[195] L. Schruben. Overview of standardized time series. In *Proc. 1985 Winter Simulation Conf.*, pages 115–118. IEEE, 1985.

[196] L. Schruben. Using simulation to solve problems: A tutorial on the analysis of simulation output. In *Proc. 1987 Winter Simulation Conf.*, pages 40–42. IEEE, 1987.

[197] A.F. Seila. A batching approach to quantile estimation in regenerative simulation. In *Management Sci.*, pages 573–581, 1982.

[198] A.F. Seila. Multivariate estimation in regenerative simulation. In *Operations Res. Letters*, pages 153–156, 1982.

[199] A.F. Seila. Multivariate estimation in simulation. In *Proc. 1983 Winter Simulation Conf.*, pages 693–696. IEEE, 1983.

[200] et al. Shahabuddin, P. Variance reduction in mean time to failure simulation. In *Proc. 1988 Winter Simulation Conf.*, pages 491–499. IEEE, 1988.

[201] R.E. Shannon. Test for the verification and validation of computer simulation models. In *Proc. 1981 Winter Simulation Conf.*, pages 573–577. IEEE, 1981.

[202] S.S. Shapiro and M.B. Wilk. An analysis of variance test for normality (complete) samples. In *Biometrika*, pages 591–611, 1965.

[203] J. Shanthikumar and R. Sargent. A unifying view of hybrid simulation/analytic models and modeling. In *Operations Res.*, pages 1030–1052, 1983.

[204] G.S. Shedler. Regeneration and networks of queues. Springer-Verlag, N.Y., 1987.

[205] S.L. Solomon. Simulation of waiting-line systems. 1983.

[206] W.T. Song. Estimators of the variance of the sample mean: Quadratic forms, optimal batch sizes, and linear combinations. In *Ph.D. Dissertation*, School of Industrial Engineering, Purdue University, West Lafayette, Indiana, 1988.

[207] W.T. Song and B.Schmeiser. inimal mse linear combinations of variance estimators of the sample mean. In *Proc. 1988 Winter Simulation Conf.*, pages 414–421. IEEE, 1988.

[208] M.C. Stairmand and W.Kreutzer. Pose: A process-oriented simulation environment embedded in scheme. In *Simulation*, pages 143–153, 1988.

[209] R. Suri and M.A. Zazanis. Perturbation analysis gives strongly consistent sensitivity estimates for the m/g/1 queue. In *Management Sci.*, pages 39–64, 1988.

[210] K.D. Tocher. The art of simulation. In *The English Universities Press Ltd.*, 1963.

[211] K.S. Trivedi. Probability and statistics with reliability, queueing, and computer science applications. Prentice-Hall, Englewood Cliffs, 1982.

[212] M.A. Turnquist and J. M. Sussman. Toward guidelines for designing experiments in queuing simulation. In *Simulation*, pages 137–144, 1977.

[213] B.W. Unger. Distributed simulation. In *Proc. 1988 Winter Simulation Conf.*, pages 198–205. IEEE, 1988.

[214] G. Vassilacopoulos. Testing for initialization bias in simulation output. In *Simulation*, pages 151–153, 1989.

[215] J .M. Velayas and R.R.Levary. Validation of simulation models using decision theory. In *Simulation*, pages 87–92, 1987.

[216] S. Venkatraman and J.R. Wilson. Using path control variates in activity network simulation. In *Proc. 1985 Winter Simulation Conf.*, pages 217–222. IEEE, 1985.

[217] S. Venkatraman and J.R. Wilson. The efficiency of control variates in multiresponse simulation. In *Operations Res. Letters*, volume 5, pages 37–42, 1986.

[218] J. von Neumann. Distribution of the ratio of the mean square successive difference to the variance. In *Ann.Math.Stat.*, pages 367–395, 1941.

[219] D.B. Wagner and E.D.Lazowska. Parallel simulation of queueing networks: Limitations and potentials. In *Performance Evaluation Review*, pages 146–155, 1989.

[220] G. Wahba. Automatic smoothing of the log periodogram. In *J. Amer. Statist. Association*, pages 122–132, 1980.

[221] J. Walrand. Quick simulation of queueing networks. In *Proc. 2nd Workshop on Appl. Math. and Performance/Reliability Models of Comput./Commun. Systems*, pages 275–286, Rome, Italy, 1987.

[222] P.D. Welch. Statistical analysis of simulation results. In *Computer Performance Modelling Handbook*, pages 287–329, S.S. Lavenberg (ed.), Academic Press, 1983.

[223] P.D. Welch. On the relationship between batch means, overlapping batch means and spectral estimation. In *Proc. 1987 Winter Simulation Conf.*, pages 320–323. IEEE, 1987.

[224] J.R. Wilson and A.A.B. Pritsker. A survey of research on the simulation startup problem. In *Simulation*, pages 55–58, 1978.

[225] J.R. Wilson and A.A.B. Pritsker. Evaluation of startup policies in simulation experiments. In *Simulation*, pages 79–89, 1978.

[226] J.R. Wilson. Variance reduction: The current state. In *Math. Comput.Simul.*, pages 55–59, 1983.

[227] J.R. Wilson. Variance reduction techniques for digital simulation. In *American J. of Mathematical and Management Science*, pages 277–312, 1984.

[228] W. Yang and B.L.Nelson. Multivariate estimation and variance reduction in terminating and steady state simulation. In *Proc. 1988 Winter Simulation Conf.*, pages 466–472. IEEE, 1988.

[229] B.P. Zeigler. Hierarchical, modular discrete-event modelling in an object-oriented environment. In *Simulation*, pages 219–230, 1987.