Simulation Modeling
and Statistical Computing

N. Adam
Guest Editor

A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations

Philip Heidelberger and Peter D. Welch
IBM Thomas J. Watson Research Center

This paper discusses a method for placing confidence limits on the steady state mean of an output sequence generated by a discrete event simulation. An estimate of the variance is obtained by estimating the spectral density at zero frequency. This estimation is accomplished through a regression analysis of the logarithm of the averaged periodogram. By batching the output sequence the storage and computational requirements of the method remain low. A run length control procedure is developed that uses the relative width of the generated confidence interval as a stopping criterion. Experimental results for several queueing models of an interactive computer system are reported.

Key Words and Phrases: simulation, confidence interval, variance estimation, spectral analysis, batch means

CR Categories: 5.5, 8.1

Introduction

This paper is concerned with the problem of generating confidence intervals for the steady state mean of an output sequence from a discrete event simulation. More specifically, it is concerned with generating confidence intervals from a single simulation run and using them to control the run length in order to achieve estimates of a prespecified accuracy. We assume only that the output sequence converges to a steady state and that in this steady state it can be reasonably modeled as a covariance stationary process. The method we describe is based on spectral analysis techniques. For fixed length simulations, it is an extension of the method described in [12]. It has only modest storage and computational requirements and is easy to apply in practice.

In generating confidence intervals for steady state characteristics there are two fundamental problems:

(1) There is a transient phase during which the characteristics of the output sequence do not approximate the steady state characteristics.

(2) The output sequence is, in general, correlated and hence standard statistical techniques based on uncorrelated observations do not directly apply.

We are only concerned with the second of these difficulties; we assume the output sequence is already in steady state.

Other approaches to the problem of generating confidence intervals for the mean of a dependent sequence have typically required that the sequence can be sectioned into independent (or approximately independent) blocks. These approaches include independent replications (see [7]), the method of batch means (see [11]), and the regenerative method (see [8]).

The method described here requires no such independence assumption. In fact, the problems of dealing with dependent data are largely avoided by working in the frequency domain with the sample spectrum (periodogram) of the process rather than in the time domain with the process itself.

Section 2 contains a detailed development of the confidence interval method. The method is based on the well known result that the variance of the sample mean of a covariance stationary sequence is given approximately by $p(0)/N$, the spectral density at zero frequency divided by the sample size. The main novelty in the method involves the technique used to estimate $p(0)$. The estimate of $p(0)$ is obtained through the application of polynomial regression to the logarithm of the averaged periodogram. Section 3 discusses the application of the method to batch means; the averages of contiguous, equal size blocks of the original sequence. Batching the data is desirable because it keeps the storage requirements of the method low. However, unlike the method of batch means (e.g., see [11]), we do not require the batch means to be uncorrelated. Section 4 develops a run length control procedure which is designed to terminate the simulation when a confidence interval of a prespecified relative width has been generated or to continue the run to a maximum length and then generate a confidence interval. Section 5 contains experimental results on the application of the confidence interval method in both fixed length simulations and in the run length control procedure. These experimental results involve waiting times and system response times of two simple

Authors' present addresses: P. Heidelberger and P. D. Welch, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598.

queueing network models of interactive computer systems. Finally, Sec. 6 summarizes the results of the paper.

2. Derivation of the Method

2.1 Introduction

We assume that the simulation generates a sample $X(1), \ldots, X(N)$ from a covariance stationary sequence. We are interested in estimating the mean $\mu = E[X(j)]$. We let $\gamma(k)$ be the covariance function, defined by

$$\gamma(k) = E[(X(j) - \mu)(X(j+k) - \gamma)] \tag{1}$$

and assume that

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty \tag{2}$$

so that the process has a spectral density, $p(f)$ (see [3]). The functions $\gamma(k)$ and $p(f)$ are a Fourier pair and are related by the equations

$$p(f) = \sum_{k=-\infty}^{\infty} \gamma(k)\cos(2\pi f k) \tag{3}$$

$$\gamma(k) = \int_{-1/2}^{1/2} p(f)\cos(2\pi f k)\, df. \tag{4}$$

The standard estimate of μ is the sample mean $\hat{\mu}$ defined by

$$\hat{\mu} = (1/N) \sum_{j=1}^{N} X(j). \tag{5}$$

The relationship between the variance of $\hat{\mu}$, $\sigma^2(\hat{\mu})$, and the variance of $X(j)$, $\sigma^2(X)$, which holds for uncorrelated samples, does not hold for correlated data, i.e., $\sigma^2(\hat{\mu}) \neq \sigma^2(X)/N = \gamma(0)/N$. The variance of $\hat{\mu}$, which can be expressed in terms of the covariance function, is given by

$$\sigma^2(\hat{\mu}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}[X(i), X(j)] = (1/N) \sum_{k=-(N-1)}^{N-1} ((N - |k|)/N)\gamma(k). \tag{6}$$

Hence although $\sigma^2(X)$ can be estimated by

$$\hat{\sigma}^2(X) = (1/(N-1)) \sum_{j=1}^{N} (X(j) - \bar{X})^2, \tag{7}$$

the quantity $\hat{\sigma}^2(X)/N$ is not a good estimate of $\hat{\sigma}^2(\hat{\mu})$. In most queueing simulations $\gamma(k) \geq 0$, so that $\hat{\sigma}^2(X)/N$ typically underestimates $\sigma^2(\hat{\mu})$.

From Eqs. (2) and (6) we have that

$$\lim_{N\to\infty} N\sigma^2(\hat{\mu}) = \sum_{k=-\infty}^{\infty} \gamma(k) = p(0). \tag{8}$$

Hence as N becomes large, the variance of $\hat{\mu}$ can be approximated by

$$\sigma^2(\hat{\mu}) \approx p(0)/N. \tag{9}$$

Thus, for large N, to estimate $\sigma^2(\hat{\mu})$ we need only estimate $p(0)$, the value of the spectral density at zero frequency.

In our method we obtain an approximately unbiased estimate $\hat{p}(0)$ of $p(0)$, and derive the coefficient of variation of $\hat{p}(0)$. We then obtain a confidence interval for μ by assuming that $(\hat{\mu} - \mu)/(\hat{p}(0)/N)^{1/2}$ is approximately distributed as a t random variable with m degrees of freedom where m is the number of degrees of freedom of a χ^2 random variable with the same coefficient of variation as $\hat{p}(0)$. This is the t random variable whose denominator squared has the same coefficient of variation as $\hat{p}(0)/N$.

2.2 Bias Problems Associated With the Spectral Windows of the Usual Spectral Estimators.

As mentioned previously, we are interested in estimating the spectrum at zero frequency. In this section we show why classical methods of spectral estimation (e.g., see [9]) are not appropriate for this problem.

The classical spectral estimates have an expectation that is a weighted average of $p(f)$. This weighting is necessary in order to produce stable estimates. Specifically if $\hat{p}(g)$ is an estimate of $p(g)$ at frequency g, then

$$E[\hat{p}(g)] = \int_{-1/2}^{1/2} H(f-g)p(f)\, df. \tag{10}$$

The weighting function $H(f)$ is called the spectral window. It is a positive, even function that integrates to one and whose area is concentrated about zero. If the spectrum is approximately linear in the region about g, then $E[\hat{p}(g)] \approx p(g)$, i.e., the estimate is approximately unbiased.

The spectrum is an even function, i.e., symmetric about zero. Therefore, in general, at zero it has either a peak or a valley and is *not* approximately linear. In fact, for many of the output sequences encountered in queueing simulations, the spectrum is sharply peaked at zero. Hence any weighted average of the spectrum about the point zero will result in a biased estimate of $p(0)$ and the larger the region of averaging, i.e., the wider the spectral window, the more biased the estimate will be.

The width of the spectral window is one of the parameters of any classical spectral estimation procedure. The observation in the previous paragraph indicates that for estimating $p(0)$ the window should be narrow so that the bias is small. However, the variance of any estimator $\hat{p}(g)$ increases as the width of the window decreases. In fact, the variance is proportional to the inverse of the window width. Consequently one is faced with a dilemma. One obtains either a highly variable estimate of $p(0)$ which is approximately unbiased or a stable estimate which can be strongly biased. Furthermore in the case of a spectrum peaked at zero this bias is negative, leading to overly narrow confidence intervals.

This dilemma *cannot* be avoided by approaching the problem through the covariance function and Eq. (6).

The sample covariance function is defined by

$$\hat{\gamma}(k) = (1/(N-k)) \sum_{j=1}^{N-k} (X(j) - \bar{X})(X(j+k) - \bar{X}) \qquad 0 \le k \le N-1$$

$$\hat{\gamma}(k) = \hat{\gamma}(-k) \qquad 0 > k \ge -(N-1). \tag{11}$$

Sometimes a multiplier $1/N$ rather than $1/(N-k)$ is used in the above equation (see [9]). Summations over the covariance function can be shown to be classical spectral estimates of $p(0)$ (see [9]). Therefore, all the remarks made above apply. In fact, the wider the summation over the sample covariance function, the narrower the spectral window. Summations over $\hat{\gamma}(k)$ for a wide range of k tend to be less biased but highly variable whereas summations over $\hat{\gamma}(k)$ for a narrow range of k are more stable but can be strongly biased.

The method we describe next approaches the problem from a somewhat different tack. It achieves variance stability in a fashion similar to classical spectral estimates but at the same time yields an unbiased estimate of $p(0)$ under quite general conditions. The use of classical spectral estimates to estimate $p(0)$ in this context is discussed in [7] and negative results are reported in [6]. An entirely different approach which estimates $p(0)$ by fitting an autoregressive model to $\{X(j)\}$ is described in [7].

2.3 Distributional Properties of the Periodogram

This method for estimating $p(0)$ proceeds (as do the classical methods discussed earlier) through the periodogram. The periodogram, $\{I(n/N)\}$, is defined by

$$I(n/N) = \left| \sum_{j=1}^{N} X(j) e^{-2\pi i (j-1)n/N} \right|^2 \Big/ N \tag{12}$$

where $i = (-1)^{1/2}$. It can be efficiently computed using fast Fourier algorithms; in fact, $I(n/N) = |a(n)|^2/N$ where $a(n)$ is the discrete Fourier transform of $X(1), \ldots, X(N)$ (see [3]).

Under very general conditions (see [3], [14], and [17]) the periodogram has the following approximate properties;

$$\begin{aligned} &E\ [I(n/N)] \approx p(n/N) && 0 < n < N/2, \\ &\mathrm{Var}[I(n/N)] \approx p^2(n/N) && 0 < n < N/2, \\ &\mathrm{Cov}[I(n/N), I(m/N)] \approx 0 && 0 < n \ne m < N/2. \end{aligned} \tag{13}$$

Furthermore $I(n/N)$ (for $0 < n < N/2$) is distributed approximately as a multiple of a χ^2 random variable with two degrees of freedom, i.e., as an exponential random variable.

Equations (13) provide the key to understanding the advantages of working in the frequency domain with the periodogram rather than working in the time domain with the original output sequence. While $\{X(j)\}$ may be a highly correlated (but stationary) sequence, the periodogram $\{I(n/N)\}$ is an approximately uncorrelated (but nonstationary) sequence. Furthermore the distributional form of the periodogram is known. The problem of working with correlated data thus translates into a problem of function estimation with uncorrelated data, the latter problem being more tractable.

At $n = 0$ we have $I(0) = N\bar{X}^2$. Therefore, the zero value of the periodogram depends only on the average value of the process and contains no information about $p(0)$. Conversely the other values of the periodogram, i.e., $I(n/N)$ for $0 < n < N/2$, do not depend on the value of $\bar{X}$.

A typical periodogram is displayed in Figure 1(a). Notice the uncorrelated appearance, the increase in variance with expected value, and the positive skewness of the exponential distribution.

We want to obtain an estimate of $p(0)$, the value of the spectrum at zero, from the values of the periodogram in the region near zero. Given the uncorrelated nature of the periodogram values and the fact that their expected values are $p(n/N)$, a reasonable approach would be to assume the spectrum is a smooth function in this region and apply regression techniques. The spectral estimate would be the value at zero (the y axis intercept) of the fitted regression function. The regression would provide the stability of averaging over a number of periodogram values and the flexibility of the family of fitted functions would avoid the bias associated with the simple weighted average provided by a spectral window. The spectral window approach also averages over the periodogram but only yields an unbiased estimate of $p(0)$ when the spectrum is flat in the region about zero. Regression analysis of the periodogram has been considered previously (see, e.g., [5]).

However, there are two problems associated with the application of regression techniques to the periodogram. First, the variance is not constant and second, the exponential distribution is very positively skewed. The variance problem can be solved by considering $\log_e I(n/N)$. The logarithm (henceforth all logs will be base e unless otherwise stated) of the periodogram has approximately the following properties for $0 < n, m < N/2$ (see [1]);

$$\begin{aligned} &E\ [\log(I(n/N))] \approx \log(p(n/N)) - 0.577 \\ &\mathrm{Var}[\log(I(n/N))] \approx 1.645 \\ &\mathrm{Cov}[\log(I(n/N)), \log(I(m/N))] \approx 0 \quad n \ne m. \end{aligned} \tag{14}$$

Hence $\log(I(n/N)) + 0.577$ has expectation $\log(p(n/N))$, a known, constant variance, and forms an uncorrelated sequence. Further $\log(p(f))$ is a smoother function than $p(f)$ for the type of peaked spectra usually encountered in queueing simulations. However, whereas the distribution of $I(n/N)$ is strongly positively skewed, the distribution of $\log(I(n/N))$ is strongly negatively skewed. The logarithm of the periodogram in Figure 1(a) has been plotted in Figure 1(b). Notice the constant variance and negative skewness. Fortunately the negative skewness can be reduced by averaging over adjacent periodogram values before taking logarithms. The average of adjacent periodogram values is distributed approximately as a χ^2 random variable with four degrees of freedom. The distribution of $\log(\chi^2_4)$ is much more sym-

Fig. 1. Illustration of the Distributional Properties of the Periodogram.

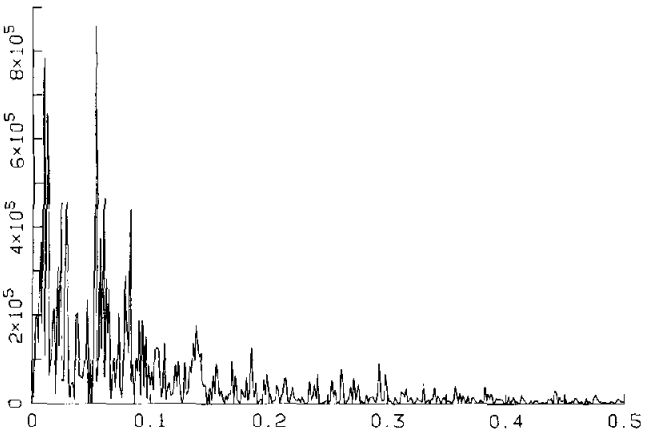


Fig. 1(a). Sample Periodogram

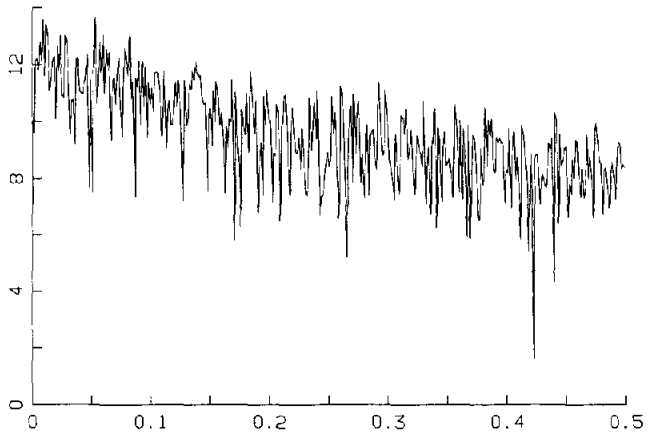


Fig. 1(b). Logarithm of Periodogram

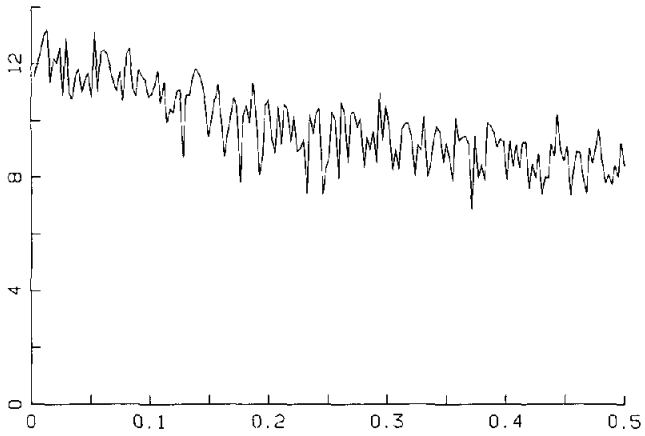


Fig. 1(c). Logarithm of Averaged Periodogram

metric than that of $\log(\chi^2_2)$ (see [1]). Let $f_n = (4n - 1)/2N$ and define

$$J(f_n) = \log(\{I((2n-1)/N) + I(2n/N)\}/2). \tag{15}$$

The quantity $J(f_1)$ is the log of the average of $I(1/N)$ and $I(2/N)$, $J(f_2)$ is the log of the average of $I(3/N)$ and $I(4/N)$, etc. For $0 < n, m < N/4$, $J(f_n)$ has the following approximate properties (see [1]):

$$\begin{aligned} &E\ [J(f_n)] \approx \log(p(f_n)) - 0.270 \\ &\mathrm{Var}[J(f_n)] \approx 0.645 \\ &\mathrm{Cov}[J(f_n), J(f_m)] \approx 0 \quad n \neq m. \end{aligned} \tag{16}$$

Notice also

$$\begin{aligned} \mathrm{Var}[J(f_n)] < \{&\mathrm{Var}[\log(I((2n-1)/N))] \\ &+ \mathrm{Var}[\log(I(2n/N))]\}/2, \end{aligned} \tag{17}$$

so that a reduction in variance is obtained by this averaging process. In Figure 1(c) $\{J(f_n)\}$ for the periodogram of Figure 1(a) has been plotted. Notice the absence of negative spikes. The sequence $\{J(f_n)\}$ is therefore well suited for the application of regression methods to estimate $p(0)$. This method is described in detail below.

2.4 The Method and its Application to Unbatched Data

We will consider fitting polynomials to $J(f_n)$ in the low frequency region. The method has two parameters, K and d. The quantity K is the number of points of $J(f_n)$ that are used to obtain the polynomial fit and d is the degree of the polynomial. The parameter K determines the frequency range over which the fit is made. The maximum number of points f_n available is $N/4$. The range $0 < n < N/4$ covers the total frequency range $0 < f < \frac{1}{2}$. Hence, we assume that the number of observations N is such that $N/4 > K$. As a result of this, the method makes the fit in the region $(0, 2K/N)$. It thus assumes that $\log(p(f))$ can be approximated by a polynomial of degree d in this region. Notice that the region becomes more and more concentrated about zero as N increases. Specifically, the method is as follows.

(1) Calculate $I(n/N)$ for $n = 1, \ldots, 2K$ and $J(f_n)$ for $n = 1, \ldots, K$.

(2) Using ordinary least squares, fit a polynomial

$$g(f_n) = \sum_{k=0}^{d} a_k f_n^k \tag{18}$$

to $J(f_n) + 0.270$ for $n = 1, \ldots, K$.

(3) Let the resulting least squares estimate of a_0 be $\hat{a}_0$. Under the assumption that $\log(p(f))$ is a polynomial of degree d for $0 \leq f \leq 2K/N$, $\hat{a}_0$ is an unbiased estimate of $\log(p(0))$.

(4) Estimate $p(0)$ by $\hat{p}(0) = C_1 e^{\hat{a}_0}$ where $C_1 = C_1(K,$

d) is a constant chosen to make $\hat{p}(0)$ approximately unbiased. The function $C_1(K, d)$ is discussed in the Appendix.

(5) Finally a confidence interval is obtained by assuming that

$$(\hat{\mu} - \mu)/(\hat{p}(0)/N)^{1/2} \tag{19}$$

has a t-distribution with $C_2(K, d)$ degrees of freedom where $C_2(K, d)$ is also discussed in the Appendix. This is the distribution of a t random variable whose denominator squared has the same coefficient of variation as $\hat{p}(0)/N$.

The choice of the parameters d and K will be discussed in the next section. The functions $C_1(K, d)$ and $C_2(K, d)$ for the values of K and d considered in this paper are given in Table I.

2.5 The Selection of Parameters

The parameter d is the degree of the polynomial used in the least squares procedure. If d is chosen large, then the method will be better able to adapt itself to complexities in the shape of the spectrum. However, the larger d, the larger the variance of $\hat{p}(0)$, i.e., the greater the instability of the spectral estimate. The parameter K, the number of points used in the polynomial fit, determines the frequency range over which the fit is made. This range is $(0, 2K/N)$. Hence as K decreases, the frequency range of the polynomial fit becomes narrower and the polynomial can more closely approximate $\log(p(f))$. However, as K decreases, the variance of $\hat{p}(0)$ increases.

Thus a compromise is necessary. To achieve an approximately unbiased estimate of $p(0)$ under the widest range of conditions, a large d and small K would be chosen. To achieve a highly stable estimate, under conditions where $\log(p(f))$ is very smooth, a small d and large K would be chosen.

The choice of these parameters was empirically investigated both in this study and the earlier study concerned with fixed length runs (see [12]). The present study also resulted in a theoretical measure of the variance of $\hat{p}(0)$ in terms of K and d. Basically the tradeoff is between good small-sample behavior when flexibility is required to adapt to the changes in the spectrum over a wide frequency range and good large-sample behavior when the spectrum over a narrow low frequency range requires less flexibility and hence can yield a more stable estimate. The behavior of the method using two sets of parameters is reported in Sec. 4. These two sets are $d = 2$, $K = 25$ and $d = 2$, $K = 50$. The first set gives better small-sample results while the second yields greater large-sample stability. Polynomials of degree 1 were inadequate for practical purposes, particularly for small samples.

3. The Application of the Method to Batched Data

Consider the sample $X(1), \ldots, X(N)$. Suppose $N = MB$ and that we batch the data into contiguous, non-overlapping batches of size B in the usual fashion. Let $\bar{X}_B(m)$ denote the mean of the mth batch, i.e.,

$$\bar{X}_B(m) = (1/B) \sum_{j=1}^{B} X((m-1)B + j) \qquad m = 1, \ldots, M. \tag{20}$$

Now the mean of the batch means is the overall mean, i.e.,

$$\hat{\mu} = (1/N) \sum_{j=1}^{N} X(j) = (1/M) \sum_{m=1}^{M} \bar{X}_B(m). \tag{21}$$

Hence, conceptually, working with the batch means is identical to working with the basic observations $X(j)$. If the number of batches M is large, then as with Eq. (9)

$$\sigma^2(\hat{\mu}) \approx p_B(0)/M \tag{22}$$

where $p_B(f)$ is the spectrum of the sequence of batch means $\{\bar{X}_B(m)\}$.

Now assume that $N > 4K$ so that the assumption of the method as applied to the original unbatched sequence is that $\log(p(f))$ can be approximated by a d degree polynomial in the frequency region $(0, 2K/N)$. Further assume that the sequence is partitioned into batches of size B and that $M = N/B > 4K$. In the batched case the assumption is that $\log(p_B(f))$ is smooth in the frequency region $(0, 2K/M) = (0, 2KB/N)$. Therefore the assumption is that the spectrum of the batched process is smooth over B times the frequency range required for the spectrum of the original process. We next determine the relationship between $p(f)$ and $p_B(f)$ so that the significance of this can be understood.

The batch means $\bar{X}_B(m)$ are obtained by passing the original sequence through an averager which averages over B consecutive elements and then through a sampler that samples the output of the averager every B elements. The averager is a linear system with weighting function

$$h(k) = \begin{cases} 1/B & k = 0, \ldots, B-1 \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

We let $Y(j)$ be the output of this averager, i.e.,

$$Y(j) = \sum_{k=0}^{B-1} X(j-k)h(k) \qquad j \geq B, \tag{24}$$

then

$$\bar{X}_B(j) = Y(jB). \tag{25}$$

From standard Fourier theory, the spectrum of $\{Y(j)\}$ is

$$p_Y(f) = T(f; B)p(f) \tag{26}$$

Table I. Constants Necessary for Application of the Method.

K	d	$C_1(K, d)$	$C_2(K, d)$
25	1	0.948	18
25	2	0.882	7
25	3	0.784	3
50	1	0.974	37
50	2	0.941	16
50	3	0.895	8

where $T(f; B)$ is the power transfer function of the averager and is given by

$$T(f; B) = \sin^2(B\pi f)/B^2 \sin^2(\pi f). \quad (27)$$

The spectrum of $\bar{X}_B(j)$ is

$$p_B(f) = (1/B) \sum_{k=0}^{B-1} p_Y((f + k)/B). \quad (28)$$

The averager producing $Y(j)$ is a low pass filter and $T(f; B)$ has a main lobe which extends out to $1/B$. The spectrum $p_B(f)$ is the spectrum $p_Y(f)$ stretched out by a factor of B and then folded (aliased) over itself B times.

Since $T(0; B) = 1$ and $T(k/B; B) = 0$ for $k = 1, \ldots, B - 1$ we have, by Eqs. (27) and (28)

$$p_B(0) = p(0)/B \quad (29)$$

and

$$p_B(0)/M = p(0)/MB = p(0)/N. \quad (30)$$

Hence, in working with the batched data, we are estimating *exactly* the same quantity we were estimating with the original sequence. As B, the batch size, becomes large, the spectrum $p_B(f)$ becomes progressively flatter. To see this, notice that as B becomes larger $T(f; B)$ becomes more and more concentrated about zero. Further the stretched and aliased $T(f; B)$ is identically equal to one, i.e.,

$$\sum_{k=0}^{B-1} T((f + k)/B; B) = 1. \quad (31)$$

Thus as B increases the spectrum $p_B(f)$ becomes flatter and is asymptotically constant. That is, for all frequencies f,

$$\lim_{B\to\infty} Bp_B(f) = p(0). \quad (32)$$

Hence, as B increases the spectrum $p_B(f)$ can be approximated by a d degree polynomial over the entire frequency range $0 \le f \le \frac{1}{2}$. In our application of the method, we will, to conserve storage, continually rebatch the data into larger and larger batch sizes and apply the method over frequency ranges from $0 < f \le \frac{1}{4}$ to $0 < f \le \frac{1}{2}$.

Examples of the convergence of $p_B(f)$ to a flat spectrum are shown in Figures 4 and 5. A flat spectrum corresponds to uncorrelated observations and hence this convergence is equivalent to the well known phenomenon that the correlation between batch means goes to zero as the batch size increases (see [11]). A result of this convergence is that as the batch size increases, only information about $p(0)$ is retained. The shape of the spectrum $p(f)$ is lost.

It is interesting to note that Parseval's theorem (see, e.g., [4]) for discrete sequences takes the form

$$\sum_{m=1}^{M} \bar{X}_B^2(m) = \sum_{m=0}^{M-1} I_B(m/M) \quad (33)$$

where $I_B(m/M)$ is the periodogram of the batch means. Since $I_B(0) = M\bar{X}^2$, we have

$$(1/(M-1)) \sum_{m=1}^{M} (\bar{X}_B(m) - \bar{X})^2 = (1/(M-1)) \sum_{m=1}^{M-1} I_B(m/M) \quad (34)$$

and the variance estimate that is obtained from the method of batch means is *identical* to the estimate obtained by fitting a zero degree polynomial (horizontal line) to the entire periodogram of the batch means.

4. A Method of Run Length Control

4.1 Introduction

An estimate produced by a simulation typically has an accuracy requirement determined by the application to which the estimate will be put. The experimenter wishes to run the simulation only until an estimate is obtained that meets this accuracy requirement. Running the simulation less than this length would not provide the information needed, while running it longer would be a waste of computing time. We will assume that this accuracy requirement is specified as the relative half-width of the confidence interval; the half-width of the confidence interval divided by the point estimate. In our case, the relative half-width for a confidence interval of level $1 - \alpha$ is

$$\text{relative half-width} = \frac{t_m(1 - \alpha/2)(\hat{p}_B(0)/M)^{1/2}}{\hat{\mu}} \quad (35)$$

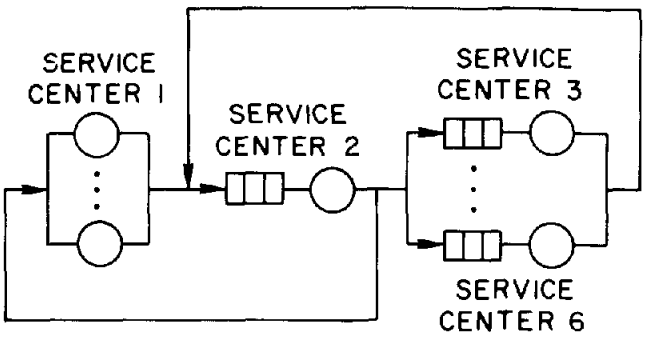
where M is the number of batched means generated by the simulation and $m = C_2(k, d)$. An analogous accuracy criterion for controlling the length of regenerative simulations was developed in [10]. In addition to the accuracy requirement, we assume there is a cost constraint represented by a maximum limit on the length of the simulation run. That is, the run is continued until either the accuracy requirement is met or until this upper limit is reached. In either case a confidence interval is generated.

In this section we will describe the incorporation of the method of Sec. 3 into such a run length control procedure. The method is applied to batched data produced by a particular batching procedure which will be described in Sec. 4.2. This batching procedure places a reasonable bound on the amount of storage required by the overall procedure and requires only a small amount of computing per output element, i.e., per $X(j)$. This batched data is then tested periodically using the method of Sec. 3 to generate the confidence interval. The precise testing procedure will be described in Sec. 4.3.

4.2 The Batching Procedure

We will keep a number of batches ranging from L to $2L$ (L would typically be 100 to 200). We thus require $2L$ storage locations. We first store $X(j)$ in storage location j for $j = 1, \ldots, 2L$. After the occurrence of $X(2L + 1)$ we consolidate the data into L batches, each

Fig. 2. A Queueing Network Model of an Interactive Computer System.



composed of two elements, by storing $X(1) + X(2)$ in location 1, $X(3) + X(4)$ in location 2, etc. After this consolidation we have L batches of size two in the first L locations. We then store the next $2L$ observations in batches of size two in the last L locations. After the occurrence of $X(4L + 1)$ we consolidate the data again into L batches of size four and store the next $4L$ observations in batches of size four in the last L locations and so on. Hence, after the occurrence of $X(N)$, where $N \geq L$, we have between L and $2L$ batches of size 2^k where k is the greatest integer less than or equal to $\log_2((N - 1)/L)$.

With the generation of each new element in the sequence, all that is necessary is the addition of this new element to the contents of the appropriate location and the update (if necessary) of the index which points to the location into which the next element will be added. At points when $\log_2((N - 1)/L)$ is an integer, consolidation is done. Hence the statistics gathering routine uses very little computation. In addition, at most $2L$ storage locations are required regardless of the length of the original output sequence.

4.3 The Run Length Control Procedure

As we mentioned in Sec. 1 we assume that the sequence $\{X(j), j \geq 1\}$ is taken from the steady state portion of the output.

The procedure requires specification of the following parameters:

(1) The point j_1 at which the testing procedure begins.
(2) The maximum run length, $j_{max}(j_{max} > j_1)$.
(3) The required relative half-width ϵ.
(4) A multiplier I by which the run length is increased if the relative half-width requirement is not met.

The run length control procedure employs these parameters as follows: The testing begins at time j_1, i.e., after the occurrence of $X(j_1)$. At this point the confidence interval generation method is applied to the set of available batches. If the relative half-width of the resulting confidence interval is less than or equal to ϵ, the simulation is terminated. If not, the simulation is continued to the point $j_2 = \min(I \times j_1, j_{max})$. A confidence interval is generated at time j_2. If the resulting relative half-width is less than or equal to ϵ or if $j_2 = j_{max}$, the simulation is terminated. Otherwise it is continued to the point $j_3 = \min(I \times j_2, j_{max})$. This testing and incrementing is continued until either the relative half-width is less than or equal to ϵ or until j_{max} is reached. An increase in run length proportional to the existing sample size is taken at each step in order to make the changes in the relative half-width uniformly significant as the check points progress.

5. Experimental Results

5.1 Models Studied

The choice of models and output sequences studied in this paper directly reflects our interest in the performance modeling of computer systems. We conducted experimental studies on models of the general form depicted in Figure 2. These are simple, closed queueing network models of interactive computer systems. For a detailed discussion of such models, see [2].

There are N customers in the network and these customers represent users of the system. Service center 1 represents the terminals. A service time at this service center represents a user's think time. Service center 1 has N servers, one for each customer in the network, so that there is no queueing at this service center. Service centers 2 through 6 are single server queues with customers served in order of arrival. Service center 2 represents the central processing unit (cpu) and service centers 3 through 6 represent secondary storage devices. The routing of customers through the network is as indicated in Figure 2. It is deterministic except upon departure from the cpu at which time the customer enters service center i with probability $p_i (i = 1, 3, 4, 5, 6)$. Successive service times at center i are independent exponentials with mean $E(S_i)$. The routing of customers and the service time sequenes are mutually independent.

We will describe experiments using two models of this type, Models A and B. Their parameters are listed in Table II. The stationary distributions of these models have a product form (see [2]) and the stationary utiliza-

Table II. Parameters for Queuing Systems.

		Model A	Model B
Model population	N	25	25
branching probabilities	p_1	0.20	0.20
	$p_3 = p_4$	0.36	0.36
	$p_5 = p_6$	0.04	0.04
Service times	$E(S_1)$	100	100
	$E(S_2)$	1.00	1.00
	$E(S_3) = E(S_4)$	1.39	5.56
	$E(S_6) = E(S_6)$	12.5	25.0
Utilizations	U_2	0.89	0.46
	$U_3 = U_4$	0.44	0.92
	$U_6 = U_6$	0.44	0.46
Waiting times	$E(W_2)$	3.77	0.84
	$E(W_3) = E(W_4)$	1.02	34.2
	$E(W_5) = E(W_6)$	9.2	20.9
Response time	$E(R)$	41.2	171

Fig. 3. Estimated Spectra and Logarithms of Spectra for Waiting and Response Time Sequences in Model B.

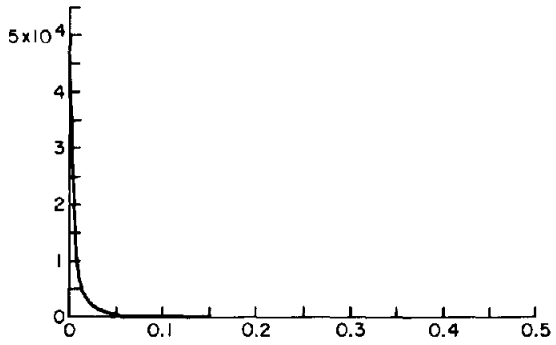


Fig. 3(a). Waiting Time Spectrum

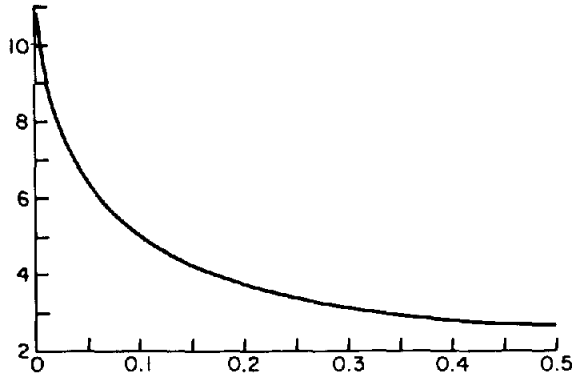


Fig. 3(b). Logarithm of Waiting Time Spectrum

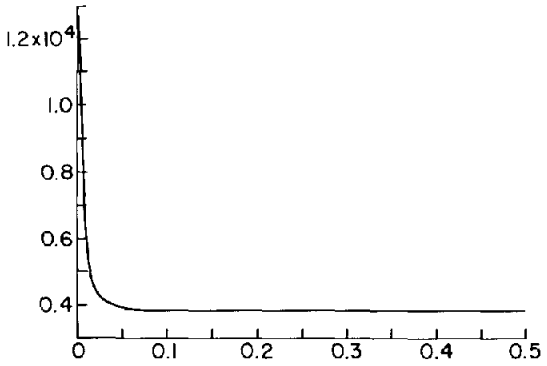


Fig. 3(c). Response Time Spectrum

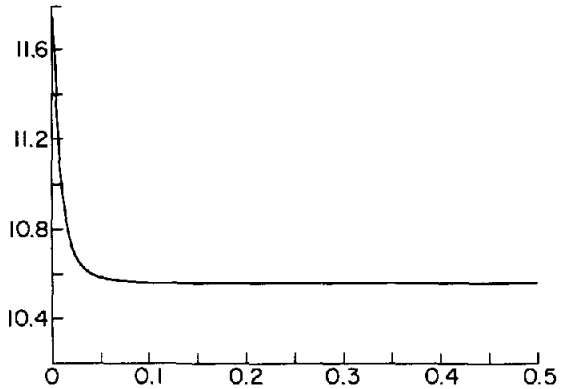


Fig. 3(d). Logarithm of Response Time Spectrum

tions, $\{U_i, i = 1, \ldots, 6\}$, expected waiting times, $\{E(W_i), i = 1, \ldots, 6\}$, and expected response times $E(R)$ given in Table II were obtained using standard techniques. The system response time is defined to be the time from when a customer departs the terminals until he next returns to the terminals. In Model A the most heavily utilized service center is the cpu, while in Model B it is one of the secondary storage devices.

For each of these models we studied the problems of estimating the expected steady state response time and the expected waiting time at the most heavily utilized device. These models were simulated using RESQ, the Research Queueing package (see [15]). For each of the four output sequences, two each for Models A and B, we made 50 independent simulation runs each 14 000 elements long. The first 500 elements of each sequence were removed to control the effect of the initial conditions. Hence for each of the four output sequences we obtained 50 independent, approximately stationary sample sequences each 13 500 elements long. These 200 sample sequences formed the basis for experiments illustrating the application of the procedures described in Sec. 3 and 4 and comparing alternative versions of these procedures. These experiments will be described below.

Estimates of the spectrum and the log of the spectrum of the waiting time and system response time sequences for Model B are plotted in Figure 3. These estimates were obtained by smoothing in frequency the average of the 50 periodograms. The spectra are typical of waiting time spectra for a first-come first-served queue and of system response time spectra. The waiting time spectra have their variance concentrated in the low frequency region with a spectrum that decreases monotonically as the frequency increases to 0.5. Because of this, the logarithms of waiting time spectra are quite amenable to polynomial approximation over the entire frequency range $0 \leq f \leq \frac{1}{2}$. The response time spectra also have a low frequency peak caused by low frequency fluctuations in the system's congestion. However, these spectra do not continually decrease as the frequency increases but reach a constant value. There is a white noise component in the response time caused by uncorrelated fluctuations in the routing and service times of successive customers. Because the response time spectrum does not steadily decrease to zero, the logarithm of this spectrum does not steadily decrease and hence is less suitable for polynomial approximation over $0 \leq f \leq \frac{1}{2}$.

Figures 4 and 5 contain estimates of the logarithms of the spectra of the batched data for batches of size 1, 4, 8, and 32. Notice how the spectra get stretched and flattened as the batch size increases. This is the phenomenon discussed in Sec. 3. In particular, notice that for the response time spectrum, the batching procedure generates a spectrum that is quite amenable to polynomial approximation over the entire frequency range $0 \leq f \leq \frac{1}{2}$.

Fig. 4. Effect of Batch Size on the Logarithm of the Waiting Time Spectrum in Model B.

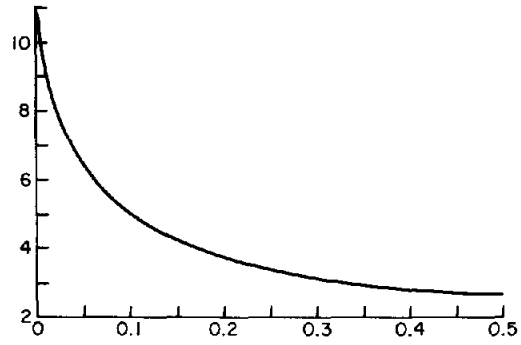


Fig. 4(a). Batch Size = 1

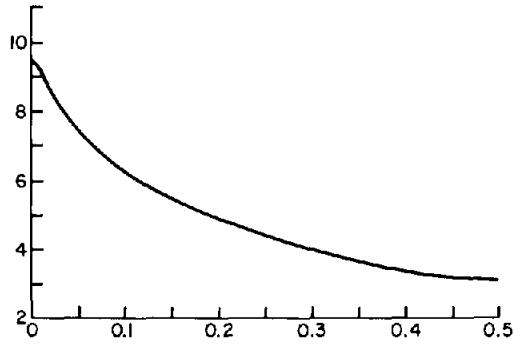


Fig. 4(b). Batch Size = 4

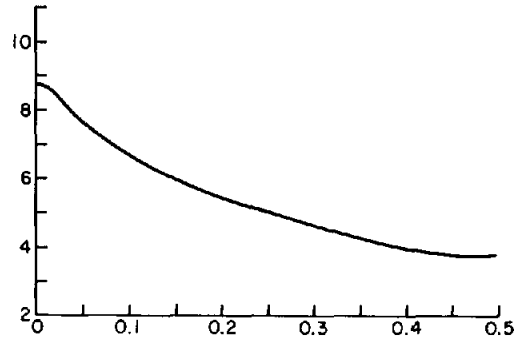


Fig. 4(c). Batch Size = 8

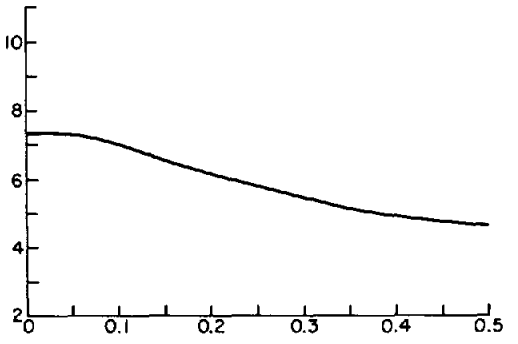


Fig. 4(d). Batch Size = 32

Fig. 5. Effect of Batch Size on the Logarithm of the Response Time Spectrum in Model B.

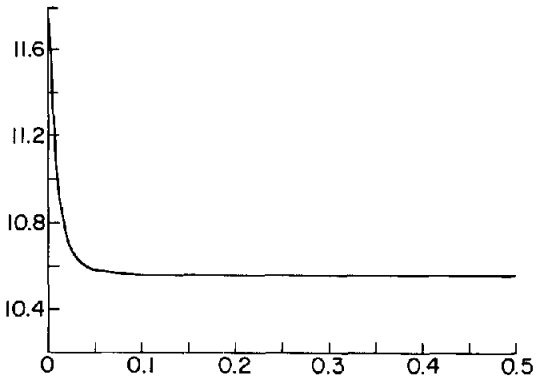


Fig. 5(a). Batch Size = 1

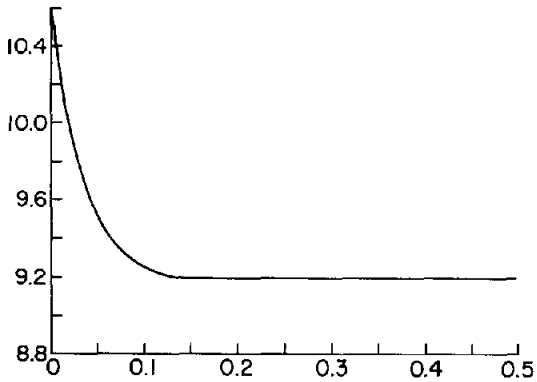


Fig. 5(b). Batch Size = 4

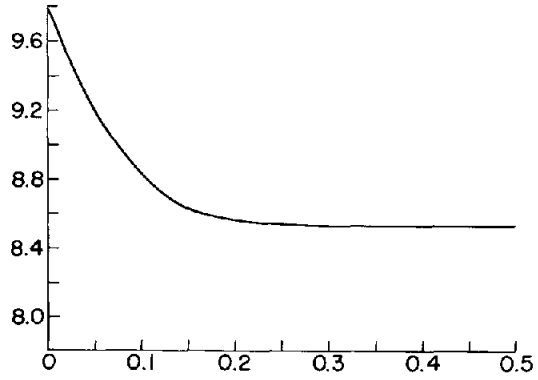


Fig. 5(c). Batch Size = 8

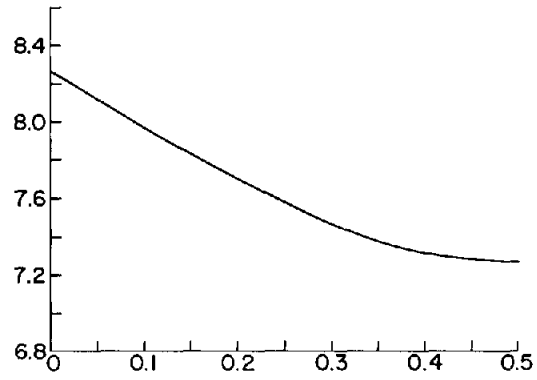


Fig. 5(d). Batch Size = 32

5.2 Experimental Results

The method of confidence interval generation described in Sec. 2.4 has two parameters: d, the degree of the polynomial fit to the log of the averaged periodogram and K, the number of these values to which the fit is made. The selection of these parameters is discussed in Sec. 2.5. We will now discuss the application of the method with two of these parameter sets: $d = 2$, $K = 25$ and $d = 2$, $K = 50$.

The batching procedure discussed in Sec. 4.2 has a single parameter L. In our experiments we set $L = 4K$. Thus the number of batches fluctuated between $4K$ and $8K$. This is the minimum value of L that allows at least K points to be used in the regression. Thus, the region over which the fit was made ranged from $0 < f \leq \frac{1}{4}$ to $0 < f \leq \frac{1}{2}$.

The run length control procedure described in Sec. 4.3 has three parameters: j_1, the initial checkpoint, I, the multiplier yielding the next check point, and $j_{\max}$, the limit on the run length. In our experiments we chose $j_1 = 500$, $I = 1.5$, and $j_{\max} = 13\,500$. This set of parameters generates a set of ten checkpoints that are listed in Table III. We first consider the performance of the method for fixed sample sizes, i.e., when the simulation is run for a fixed, prespecified, number of output values and then a confidence interval is produced. This situation is of interest in its own right and results for it are necessary for the interpretation of the run length control experiments. We considered the sample sizes corresponding to the ten check points in the run length control procedure mentioned above. The results of these experiments are reported in Table III. At each check point we formed 90 percent confidence intervals for μ. In Table III, for each check point we report the fraction of the 50 confidence intervals generated (one for each independent replication) that actually contained μ. This fraction is called a (90 percent) coverage and it should be close to 0.90 if valid confidence intervals are being formed. Coverages less than 0.82 are significantly lower than 0.90 at the 0.90 level. Table III also gives an estimate of the expected relative half-widths of the confidence intervals (the average of the 50 confidence interval half-widths divided by μ) and the sample variance of the confidence interval half-widths.

Notice that as predicted in Sec. 2.5, parameter set 1 has better small-sample coverage than does set 2. On the other hand, set 2 has somewhat better large-sample behavior; the confidence intervals are slightly narrower and less variable. The variance of the confidence interval half-width is roughly half what it is for set 1. This is as predicted by the doubling of the effective degrees of freedom $C_2(K, d)$.

We ran the run length control procedure for values of the desired relative half-width ϵ, equal to 0.20, 0.15, 0.10, and 0.05. This corresponds roughly to a relative accuracy of ±20 percent, ±15 percent, ±10 percent, and ±5 percent, respectively. This range of accuracies was considered reasonable for most practical applications.

The positive side of a sequential procedure is that even though the confidence interval method may have small-sample problems, the relative width criterion will force it to run into the large-sample region where it works well. If we look at the estimated expected relative half-widths of the confidence intervals given in Table III, we notice that this looks to be the case for parameter set 1 but not for parameter set 2. For parameter set 1 the small-sample sizes with low coverage generate confidence intervals with expected relative half-widths much greater than 0.2, the largest value of ϵ considered. Hence we would expect a sequential procedure to force the simulation to run into the region where valid confidence intervals are generated. However, for parameter set 2 the small-sample sizes with low coverage generate confidence intervals with expected relative half-widths as small as 0.1. Thus one would not expect a sequential procedure to work well in these cases for the larger values of ϵ. We will see below that this is, in fact, the case for our procedure.

The negative side of a sequential procedure is that those samples which underestimate the variance, i.e., underestimate $p(0)$, produce small relative half-widths and therefore tend to both pass the relative half-width test and not cover the true value. Thus there is a danger that a sequential procedure will tend to stop each simulation run at a point where the coverage would tend to be low. A method could therefore have good fixed sample coverage and yet have poor coverage when embodied in a sequential procedure. We did observe this type of behavior: Those samples passing the test early tended to have very low coverage and the coverage of the sequential procedure tended to be less than what would be expected from a random selection with the same run length distribution. To keep this effect under control we used a version of the sequential procedure that checked relatively infrequently, i.e., used a large value of I (1.5) and began checking only after a fairly large number of observations had occurred, i.e., at a fairly large value of j_1 (500). The proportional nature of the increments also ensured that each estimate had the same proportion of new data and hence that the checking is done only after significant changes have occurred. Relative to the rate at which the estimates are changing, the rate of checking is uniform.

The above procedure was applied to the four estimation problems with both parameter sets. The results are given in Table IV and V. The first column gives ϵ, the desired relative half-width. The second column gives the overall coverage, the coverage of all the confidence intervals produced including those produced at $j_{\max}$ that did not meet the accuracy criterion. Column 3 gives the coverage of those that met the accuracy criterion. This includes those that ran to $j_{\max}$ and then produced relative half-widths less than ϵ. Column 4 gives the coverage of those runs that were unable to meet the criterion by $j_{\max}$.

Table III. Simulation Results for Fixed Run Lengths.

	Run length	$K = 25$, $100 \leq$ Number of Batches ≤ 200			$K = 50$, $200 \leq$ Number of Batches ≤ 400		
		Coverage	Average interval half-width $\div \mu$	Variance of interval half-width	Coverage	Average interval half-width $\div \mu$	Variance of interval half-width
Model A system response time	500	0.92	0.19	4.48	0.78	0.14	1.01
	750	0.94	0.17	3.32	0.84	0.13	0.84
	1125	0.96	0.15	2.00	0.86	0.13	0.48
	1687	0.92	0.13	1.78	0.90	0.11	0.54
	2530	0.88	0.09	0.97	0.86	0.10	0.69
	3795	0.94	0.08	0.49	0.92	0.08	0.25
	5692	0.88	0.06	0.27	0.88	0.06	0.13
	8538	0.92	0.05	0.25	0.90	0.05	0.11
	12807	0.94	0.04	0.12	0.96	0.04	0.06
	13500	0.90	0.04	0.11	0.94	0.04	0.0
Model A waiting time at queue 2	500	0.76	0.38	0.29	0.60	0.29	0.11
	750	0.74	0.35	0.20	0.66	0.27	0.09
	1125	0.86	0.33	0.12	0.84	0.25	0.05
	1687	0.94	0.32	0.13	0.88	0.24	0.05
	2530	0.96	0.28	0.09	0.92	0.22	0.04
	3795	0.94	0.24	0.06	0.90	0.21	0.03
	5692	0.90	0.19	0.05	0.92	0.18	0.02
	8538	0.90	0.15	0.04	0.94	0.15	0.02
	12807	0.84	0.13	0.02	0.84	0.13	0.01
	13500	0.90	0.13	0.02	0.88	0.12	0.01
Model B system response time	500	0.86	0.14	27.89	0.76	0.11	7.38
	750	0.90	0.12	21.23	0.80	0.10	6.67
	1125	0.90	0.11	16.08	0.88	0.09	7.18
	1687	0.96	0.09	16.71	0.94	0.08	4.51
	2530	0.94	0.08	10.38	0.94	0.07	3.85
	3795	0.90	0.06	10.05	0.92	0.06	3.16
	5692	0.92	0.05	5.91	0.94	0.05	2.26
	8538	0.96	0.04	2.57	0.96	0.04	1.00
	12807	0.90	0.03	1.70	0.96	0.03	0.87
	13500	0.92	0.03	1.69	0.96	0.03	0.86
Model B waiting time at queue 3	500	0.82	0.37	14.41	0.74	0.25	3.64
	750	0.88	0.35	10.33	0.74	0.25	3.90
	1125	0.90	0.33	8.75	0.80	0.25	3.15
	1687	0.92	0.33	6.75	0.94	0.25	2.39
	2530	0.92	0.27	5.14	0.92	0.23	1.63
	3795	0.96	0.21	4.00	0.94	0.20	1.27
	5692	0.90	0.16	2.04	0.92	0.17	0.87
	8538	0.94	0.13	0.97	0.94	0.13	0.64
	12807	0.88	0.11	0.97	0.94	0.11	0.57
	13500	0.88	0.10	0.70	0.90	0.10	0.47

Column 5 gives the fraction of runs that met the accuracy criterion. Column 6 gives the average run length and Column 7 gives the average relative half-width.

The method with parameter set 1 has good overall coverage for all four situations. For response times the procedure yielded confidence intervals that met the accuracy requirements in almost all cases. Notice that the average relative half-widths are somewhat less than those required. This is the price paid for proportionally increasing the run length and avoiding a substantial loss in coverage. For the waiting times and the smaller ϵ's (more rigid accuracy requirements), the procedure ran to j_{max} a large percentage of the time. As an example of the low coverage for the early stoppers, see the data for the expected waiting time at queue 2 of Model A.

For the system response times the method with parameter set 2 has low coverage for the larger values of ϵ. As discussed earlier, this is caused by the small-sample low coverage for this parameter set and the fact that, for these values of ϵ, the accuracy criterion does not force the simulation to run into the large-sample region where the coverage is adequate. For the waiting times the coverage is adequate even though the set has even lower small-sample coverage since the accuracy criterion forces it into the large-sample region.

A more recent study[1] investigated adaptive procedures, including the application of spline techniques (see

[1] Heidelberger, P. and Welch, P.D. On the statistical control of simulation run length. IBM Research Rept. RC 8571 Yorktown Heights, N.Y. 1980.

Table IV. Simulation Results for Sequential Method with $K = 25$ and $100 \leq$ Number of Batches ≤ 200.

	Relative half-width	Overall coverage	Coverage given stopped	Coverage at end	Probability of stopping	Mean run length	Mean relative half-width
Model A	0.20	0.92	0.92	*	1.00	669	0.16
system	0.15	0.92	0.92	*	1.00	1 198	0.13
response	0.10	0.84	0.84	*	1.00	2 727	0.08
time	0.05	0.88	0.88	1.00	0.98	9 849	0.04
Model A	0.20	0.90	0.90	1.00	0.98	5 660	0.17
waiting	0.15	0.86	0.84	1.00	0.88	9 506	0.13
time at	0.10	0.84	0.54	0.95	0.26	12 934	0.13
queue 2	0.05	0.90	*	0.90	0.00	13 500	0.13
Model B	0.20	0.86	0.86	*	1.00	510	0.13
system	0.15	0.84	0.84	*	1.00	590	0.12
response	0.10	0.80	0.80	*	1.00	1 350	0.09
time	0.05	0.84	0.84	*	1.00	6 566	0.04
Model B	0.20	0.88	0.88	*	1.00	4 913	0.17
waiting	0.15	0.92	0.92	1.00	0.98	8 169	0.12
time at	0.10	0.90	0.84	0.96	0.50	12 569	0.10
queue 3	0.05	0.88	1.00	0.88	0.02	13 401	0.10

* Event did not occur

Table V. Simulation Results for Sequential Method with $K = 50$ and $200 \leq$ Number of Batches ≤ 400.

	Realtive half-width	Overall coverage	Coverage given stopped	Coverage at end	Probability of stopping	Mean run length	Mean relative half-idth
Model A	0.20	0.78	0.78	*	1.00	500	0.14
system	0.15	0.78	0.78	*	1.00	586	0.13
response	0.10	0.88	0.88	*	1.00	2633	0.09
time	0.05	0.92	0.92	*	1.00	9695	0.04
Model A	0.20	0.86	0.86	*	1.00	4247	0.18
waiting	0.15	0.86	0.85	1.00	0.92	10091	0.13
time at	0.10	0.86	0.75	0.89	0.24	13289	0.12
queue 2	0.05	0.88	*	0.88	0.00	13500	0.12
Model B	0.20	0.76	0.76	*	1.00	500	0.11
system	0.15	0.76	0.76	*	1.00	500	0.11
response	0.10	0.74	0.74	*	1.00	852	0.09
time	0.05	0.88	0.88	*	1.00	6043	0.04
Model B	0.20	0.90	0.90	*	1.00	4246	0.18
waiting	0.15	0.92	0.92	*	1.00	9040	0.13
time at	0.10	0.94	0.86	1.00	0.44	13151	0.10
queue 3	0.05	0.90	*	0.90	0.00	13500	0.10

* Event did not occur.

[16]); which attempt to automate the conversion from low stability to high stability estimates as the sample size increases. However, these adaptive methods did not perform successfully within the run length control procedure. On the basis of that study, the current study and other experiments we have performed, we recommend that parameter set 1 be used for general applications where the method is part of an automated procedure. We also recommend that j_1 be set relatively large, say between $0.1\, j_{max}$ and $0.2\, j_{max}$, and that $I = 1.5$. However, in situations where the experimenter knows *a priori* that a very long simulation run is required and sets j_1 accordingly, it would be appropriate to consider the use of parameter set 2.

6. Summary

We have described a method for generating confidence intervals on the steady state expectations of output sequences from discrete event simulations and have incorporated it into a procedure for run length control. The method operates on a single simulation run and is applicable to either the output sequence itself or to batched means generated from this sequence. It assumes only that the output sequence can be reasonably modeled as a covariance stationary process with a smooth log spectrum. The variance estimate required for the confidence interval is generated by estimating the spectrum at zero frequency. This estimation is accomplished by

the application of polynomial regression to the logarithm of the averaged periodogram. The advantage of the spectral approach is that it enables one to work with uncorrelated periodogram values rather than with a highly correlated output sequence. The regression provides flexibility and stability in estimating the variance.

The run length control procedure examines the confidence interval width at a succession of check points. If a prespecified relative width criterion is satisfied, the simulation is terminated, otherwise it is continued to the next check point or the maximum length. In any event, a confidence interval is generated. This procedure operates on batched data. It requires a small, fixed amount of storage and has minimal computational overhead. A specific version is recommended for the automatic control of simulation experiments.

Experiments were described illustrating the application of the procedure to the estimation of mean waiting times and system response times in simple closed queueing network models of an interactive computer system. Two alternative versions of the method were tested and one of these versions is recommended for most practical applications.

Appendix: Derivation of the Functions $C_1(K, d)$ and $C_2(K, d)$

Through the application of regression methods to $Y_n = J(f_n) + 0.270$ we obtain, under our assumptions, an unbiased estimate of log $(p(0))$. The regression model is of the form

$$Y = \mathbf{X}\alpha + \epsilon \tag{A1}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & f_1 & \cdots & f_1^d \\ \cdot & \cdot & \cdots & \cdot \\ 1 & f_K & \cdots & f_K^d \end{bmatrix} \tag{A2}$$

$$\alpha' = (a_0, \ldots, a_d), \qquad Y' = (Y_1, \ldots, Y_K) \tag{A3}$$

and ϵ is a vector of uncorrelated random variables with a common variance equal to 0.645. We let $\hat{a}_0$ be the least squares estimate of a_0. Then $\hat{a}_0$ is an unbiased estimate of $\log(p(0))$. It is approximately normally distributed with variance σ^2 given by

$$\sigma^2 = 0.645\, s_{11}(K, d) \tag{A4}$$

where $s_{11}(K, d)$ is the upper leftmost element of $(\mathbf{X}'\mathbf{X})^{-1}$.

Hence $e^{\hat{a}_0}$ has approximately a log normal distribution (see [13]) with

$$E[e^{\hat{a}_0}] = e^{\sigma^2/2} p(0) \tag{A5}$$

$$\mathrm{Var}[e^{\hat{a}_0}] = (e^{2\sigma^2} - e^{\sigma^2}) p^2(0). \tag{A6}$$

Thus

$$\hat{p}(0) = e^{-\sigma^2/2} e^{\hat{a}_0} \tag{A7}$$

is an approximately unbiased estimate of $p(0)$. In the notation of Sec. 2.4

$$C_1(K, d) = e^{-\sigma^2/2} \tag{A8}$$

where σ^2 is given by Eq. (A4).

From Eqs. (A5), (A6), and (A7), $\hat{p}(0)$ has a squared coefficient of variation C^2 given by

$$C^2 = \mathrm{Var}[\hat{p}(0)]/E^2[\hat{p}(0)] = e^{\sigma^2} - 1. \tag{A9}$$

Now for the t-distribution the estimate of the variance, the square of the denominator, is distributed as a multiple of a χ^2 random variable with some number, say m, degrees of freedom. The squared coefficient of variation of this distribution is $2/m$. Hence the χ^2 distribution with the same coefficient of variation as $\hat{p}(0)$ would have, in the notation of Sec. 2.4,

$$C_2(K, d) = 2/(e^{\sigma^2} - 1) \tag{A10}$$

degrees of freedom where σ^2 is given by Eq. (A4).

Received 4/80; revised 9/80; accepted 11/80

References

1. Bartlett, M.S. and Kendall, D.G. The statistical analysis of variance heterogeneity and the logarithmic transformation. *J. Roy. Statist. Soc. (Suppl.), 8*, 1 (1946), 128–138.
2. Baskett, F., Chandy, K.M., Muntz, R.R., and Palacios, F.G. Open, closed, and mixed networks of queues with different classes of jobs. *J. ACM 22*, 2 (April 1975), 248–260.
3. Brillinger, D.R. *Time Series, Data Analysis and Theory.* Holt, Rinehart and Winston, New York, 1975.
4. Cooley, J.W., Lewis, P.A.W., and Welch, P.D. The finite Fourier transform. *IEEE Trans. Audio Electroacoust. 17*, 2 (June 1969), 77–85.
5. Cooley, J.W., Lewis, P.A.W., and Welch, P.D. The application of the fast Fourier transform to the estimation of spectra and cross spectra. *J. Sound Vib. 12*, 3 (1970), 339–352.
6. Duket, S.D. and Pritsker, A.A.B. Examination of simulation output using spectral methods. *Math. Compt. Simulation. 20*, (1978), 53–60.
7. Fishman, G.S. *Principles of Discrete Event Simulation.* John Wiley and Sons, New York, 1978.
8. Iglehart, D.L. The regenerative method for simulation analysis. In *Current Trends in Programming Methodology, Vol. III: Software Engineering.* K.M. Chandy and R.T. Yeh (Eds.), Prentice-Hall, Englewood Cliffs, N.J., 1978.
9. Jenkins, G.M. and Watts, D.G. *Spectral Analysis and Its Applications.* Holden Day, San Francisco, 1968.
10. Lavenberg, S.S. and Sauer, C.H. Sequential stopping rules for the regenerative method of simulation. *IBM J. Res. Develop. 21*, 6 (Nov. 1977), 545–558.
11. Law, A.M. and Carson, J.S. A sequential procedure for determining the length of a steady-state simulation. *Operations Res. 27*, 5 (Sept.–Oct. 1979), 1011–1025.
12. Moeller, T.L. and Welch, P.D. A spectral based technique for generating confidence intervals from simulation outputs. Proc. 1977 Winter Simulation Conf. Vol. 1, (1977) 177–184.
13. Mood, A.M., Graybill, F.A., and Boes, D.C. *Introduction to the Theory of Statistics*, 3rd Ed. McGraw-Hill, New York, 1974.
14. Olshen, R.A. Asymptotic properties of the periodogram of a discrete stationary process. *J. Appl. Probability 4*, (1967), 508–528.
15. Sauer, C.H. and MacNair, E.A. Queueing network software for systems modelling. *Software Practice and Experience 9*, 5 (May 1979), 360–380.
16. Wahba, G. Automatic smoothing of the log periodogram. *J. Am. Statist. Assoc. 75*, 369 (March 1980), 122–132.
17. Walker, A.M. Some asymptotic results for the periodogram of a stationary time series. *J. Austral. Math. Soc. 5*, (1965), 107–128.