TWO-STAGE STOPPING PROCEDURES BASED ON STANDARDIZED TIME SERIES

MARVIN K. NAKAYAMA
Computer and Information Science Department
New Jersey Institute of Technology
Newark, NJ 07102

We propose some new two-stage stopping procedures to construct absolute-width and relative-width confidence intervals for a simulation estimator of the steady-state mean of a stochastic process. The procedures are based on the method of standardized time series proposed by Schruben and on Stein's two-stage sampling scheme. We prove that our two-stage procedures give rise to asymptotically valid confidence intervals (as the prescribed length of the confidence interval approaches zero and the size of the first stage grows to infinity). The sole assumption required is that the stochastic process satisfy a functional central limit theorem.
(SIMULATION; OUTPUT ANALYSIS; STOPPING RULES; DIFFUSION APPROXIMATIONS)

1 Introduction

When running a simulation to estimate the steady-state mean of some stochastic process, we often would like a method for determining the run length needed so that the resulting confidence interval is of a prespecified (absolute or relative) width. In the absence of such a method, we may end up just fixing the total run length prior to running the simulation. The main disadvantange of such an approach is that the constructed confidence interval may not be of the desired width. Since the size of the confidence interval is usually unknown in advance, an inappropriately long run length wastes computer resources refining the estimator beyond the accuracy needed. On the other hand, if the specified run length is too short, the resulting confidence intervals may have widths that are too large to be of practical use.

One method for determining an appropriate simulation run length is to use a fully sequential stopping procedure. In this approach, we terminate the simulation once the confidence interval achieves the predetermined width. Schemes of this type are typically based on the initial work of Chow and Robbins (1965), which established the asymptotic validity (as the confidence interval width approaches zero) of a fully sequential procedure when the output sequence consisted of i.i.d. random variables. Glynn and Whitt (1992) extended their work by proving the asymptotic validity of fully sequential stopping rules for certain dependent processes, which typically arise in the simulation context. Others have also proposed and empirically studied sequential procedures for use in simulations. In particular, Fishman (1977), Adam (1983), Law and Carson (1979), and Law and Kelton (1982) all consider sequential schemes using batch means; Lavenberg and Sauer (1977) investigate sequential procedures in regenerative simulations; Heidelberger and Welch (1981a, 1981b, 1983) use a spectral approach; and Iglehart (1977), in a different kind of application, presents selection procedures based on sequential methods. For an overview of many of these methods, see pp. 81–103 of Bratley, Fox, and Schrage (1987).

However, there are certain drawbacks to using a sequential stopping procedure. First, problems arise because the run length is now randomly determined. Since we do not have direct control over the simulation, the run length may turn out be very long, thus using more computer resources than desired. On the other hand, the simulation may terminate inappropriately early due to statistical variability, which may cause difficulties. For example, in many statistical settings, the point estimate and the variance estimate are positively correlated. Hence, the confidence interval width (which is typically determined by the estimate of the variance) is likely to be small if the point estimate is small. Consequently, the sequential stopping procedure will tend to terminate early when the point estimate is small, possibly leading to significant problems in the coverage of the resulting confidence interval. This problem seems to be endemic to virtually all Chow-Robbins type procedures. The most significant disadvantage, though, is that in order to establish the asymptotic validity of their fully sequential procedures, Glynn and Whitt (1992) require the estimator of the asymptotic variance to be strongly consistent; i.e., it converges with probability one. (The strong consistency assumption can be replaced by assuming that the variance estimator satisfy a functional weak law of large numbers.) The requirement that the estimator of the asymptotic variance is

strongly consistent is restrictive. In nonregenerative settings, constructing such estimators often requires the knowledge of more advanced statistical techniques, such as spectral and autoregressive methods, which makes implementation difficult. These techniques are also somewhat problem dependent. Thus, the need arises for robust procedures to determine appropriate simulation run lengths without such a restrictive requirement.

To this end, we propose two-stage stopping rules that combine a scheme by Stein (1945) with standardized time series. Stein developed a two-stage procedure for determining the sample size needed to construct a confidence interval of predetermined absolute length and confidence coefficient for the mean of i.i.d. normal random variables when the variance is unknown. Stein's first stage consists of a fixed number of samples from which an estimate of the variance of the distribution is formed. Using this, Stein then determines the total number of samples needed. Standardized time series is a class of methods used to construct confidence intervals (in a nonsequential setting) without consistently estimating the asymptotic variance constant; see Schruben (1983) or Glynn and Iglehart (1990). These techniques "cancel out" the variance constant in a manner reminiscent of the t-statistic instead of trying to consistently estimate it. Thus, in our two-stage procedures, the first stage amounts to simulating a pilot run having a fixed number of equal-sized batches. We apply a standardized time series methodology to the first stage and then determine the additional number of batches needed in order to construct a confidence interval having the desired relative or absolute length for the steady-state mean of our process.

As with the fully sequential methods described by Glynn and Whitt (1992), our two-stage stopping rules produce asymptotically valid confidence intervals. The asymptotic validity of our procedures occurs as the prescribed (relative or absolute) length of the confidence interval approaches zero and the size of each batch grows to infinity. The only condition we require is that the stochastic process obey a functional central limit theorem. This mild assumption, also used by Glynn and Whitt (1992), is satisfied in virtually all practical settings. The main advantage of our approach is that by using a standardized time series methodology, we avoid requiring a strongly consistent estimate of the variance constant. Also, implementing the method of batch means with a fixed number of batches, which is an example of a standardized time series, does not require any knowledge of more advanced statistical theory, thus making implementation a relatively simple task. Furthermore, we prove that the total number of batches collected converges to some *proper* limiting random variable. Hence, our procedure is in the spirit of the work of Schmeiser (1982), which suggests that in most applications the total number of batches in a (nonsequential) batch means procedure should not grow to infinity as the run length goes to infinity, but rather should be kept fairly small.

Although our two-stage procedures have many desirable properties, there are also some drawbacks. First, our methods also suffer from one of the problems that afflict fully sequential stopping rules; viz., the run length of the simulation is now randomly determined and so it may be inappropriately long or short. However, by specifying an appropriate length for the first stage, we can avoid this problem. Determining a suitable size for the first stage, though, is somewhat difficult.

The complication arises from the fact that there is an interdependence between the batch size and the desired half-width which must be satisfied in order for our methods to be asymptotically valid. More specifically, the batch size must grow at a rate which is proportional to the inverse of the square of the half-width, but the exact rate needed in practice depends on the stochastic process being simulated. In Section 5 we give some suggestions on how one might deal with this in certain contexts.

Finally, we should mention that many aspects of standardized time series have been studied previously. Glynn and Iglehart (1990) show that the methods of standardized time series with a fixed number of batches yield asymptotically valid confidence intervals when the stochastic process satisfies a functional central limit theorem. Schruben (1983) shows a similar result but under the assumption that the process is stationary and satisfies a mixing condition. Goldsman and Schruben (1984) focus on some asymptotic properties of the confidence intervals produced by different standardized time series schemes. Sargent, Kang, and Goldsman (1992) investigate the small sample behavior and convergence properties of these confidence intervals.

A special case of standardized time series is the method of batch means with a fixed number of batches, a procedure which has been studied extensively in the literature; e.g., see Mechanic and McKay (1966) or Bratley, Fox, and Schrage (1987). However, Glynn and Whitt (1991) show that the asymptotic variance constant cannot be consistently estimated from a batch means procedure in which the number of batches is kept fixed while the size of each batch grows to infinity. Hence, Glynn and Whitt's (1992) results for constructing fully sequential stopping procedures are not applicable in this setting.

On the other hand, Damerdji (1989) has shown (under certain conditions) that by letting the number of batches in a nonsequential batch means scheme grow to infinity, we can construct a strongly consistent estimate of the variance. Thus, the approach can be used in a fully sequential procedure. However, Damerdji's procedure is computationally complicated and not completely robust (since the rate at which the number of batches needs to grow depends on the stochastic process being simulated).

The rest of the paper is organized as follows. In Section 2 we first review Stein's (1945) original two-stage scheme applied to normal random variables and then describe how to extend it to work with the method of batch means. We generalize the two-stage batch means procedure to use any standardized time series methodology in Section 3. In Section 3.1 we impose a condition on the stochastic process being simulated and discuss some of its ramifications; Section 3.2 provides theorems showing the asymptotic validity of our procedures. Section 4 contains examples of standardized time series as used in our context. We discuss various implementation issues in Section 5. Experimental results are given in Section 6, and Section 7 concludes with some directions for future work. Finally, the appendix contains all of the proofs.

2 Modifying Stein's Two-Stage Procedure to Work With the Method of Batch Means

Stein (1945) proposed the following two-stage procedure for constructing confidence intervals having coverage probability $1-\delta$ and predetermined half-width ϵ for the mean μ of a normal random variable with unknown variance σ^2. Fix $m \geq 2$, and in the first stage collect $\tilde{Z}_1, \ldots, \tilde{Z}_m$, which are m independent random samples from a $N(\mu, \sigma^2)$ distribution. Then compute $\tilde{s}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(\tilde{Z}_i - \frac{1}{m}\sum_{k=1}^{m}\tilde{Z}_k)^2$, which is the sample variance of the first m samples. The total number of observations needed is then given by

$$\tilde{N} = \max\left\{m, \left\lceil \frac{\tilde{s}^2 t^2_{m-1,\delta}}{\epsilon^2} \right\rceil \right\}, \tag{1}$$

where $t_{m-1,\delta}$ is the upper $100(1-\delta/2)\%$ quantile point of a Student's t-distribution with $m-1$ degrees of freedom, and $\lceil q \rceil$ denotes the smallest integer greater than or equal to q. After computing $\tilde{N}$, we collect additional samples $\tilde{Z}_{m+1}, \ldots, \tilde{Z}_{\tilde{N}}$ from the $N(\mu, \sigma^2)$ distribution in the second stage and use the sample mean of all $\tilde{N}$ observations as our point estimate. We then form the confidence interval $[\frac{1}{\tilde{N}}\sum_{i=1}^{\tilde{N}}\tilde{Z}_i - \epsilon, \frac{1}{\tilde{N}}\sum_{i=1}^{\tilde{N}}\tilde{Z}_i + \epsilon]$. Stein proved the following result, which shows the validity of the procedure.

Proposition 1 (Stein) $P\left\{\mu \in \left[\frac{1}{\tilde{N}}\sum_{i=1}^{\tilde{N}}\tilde{Z}_i - \epsilon, \frac{1}{\tilde{N}}\sum_{i=1}^{\tilde{N}}\tilde{Z}_i + \epsilon\right]\right\} \geq 1-\delta.$

The proof is given in Stein (1945).

We can modify Stein's method to create a two-stage stopping procedure that will determine an appropriate run length of a simulation so that the resulting $100(1-\delta)\%$ confidence interval for the steady-state mean μ of the stochastic process is of prespecified width ϵ. We accomplish this by combining Stein's procedure with the method of batch means. (Actually, we can replace the method of batch means with any standardized time series methodology; see Section 3.) In our two-stage scheme the first stage amounts to simulating an initial pilot run having $m \geq 2$ equal-sized batches. We then apply (1) with the means of the batches from the first stage in the place of the i.i.d. normal random variables. This gives us the total number of batches that need to be simulated in order for the resulting confidence interval to be of appropriate size.

More specifically, let $\mathbf{Y} = \{Y(t) : t \geq 0\}$ denote the simulation output of some stochastic process having steady-state mean μ. (There are certain regularity conditions which we require of the process. These will be discussed in Section 3.1.) First, simulate an initial pilot run of length $1/\epsilon^2$ and divide it into $m \geq 2$ equal-sized batches. (It is important to note that the batch size is inversely proportional to the square of the desired half-width.) Let

$$Z_i(\epsilon) = \frac{\int_{(i-1)/(m\epsilon^2)}^{i/(m\epsilon^2)} Y(s)ds}{1/(m\epsilon^2)}, \quad i \geq 1, \tag{2}$$

which is the mean of the i^{th} batch of size $1/(m\epsilon^2)$ of the process $\mathbf{Y}$, and define

$$s^2(\epsilon) = \frac{1}{m-1}\sum_{i=1}^{m}\left(Z_i(\epsilon) - \frac{1}{m}\sum_{k=1}^{m}Z_k(\epsilon)\right)^2, \tag{3}$$

which is the sample variance of the first m batch means. The total number of batches needed in order to construct our desired confidence interval is consequently given by

$$N_a(\epsilon) = \max\left\{m, \left\lceil \frac{s^2(\epsilon)t^2_{m-1,\delta}}{\epsilon^2} \right\rceil\right\}.$$

Then simulate the second stage, which consists of $N_a(\epsilon) - m$ batches of size $1/(m\epsilon^2)$, and form the absolute-precision confidence interval

$$I_a(\epsilon) = \left[\frac{1}{N_a(\epsilon)}\sum_{i=1}^{N_a(\epsilon)} Z_i(\epsilon) - \epsilon, \frac{1}{N_a(\epsilon)}\sum_{i=1}^{N_a(\epsilon)} Z_i(\epsilon) + \epsilon\right].$$

The resulting confidence interval $I_a(\epsilon)$ is asymptotically valid in the sense that

$$\lim_{\epsilon\to 0} P\{\mu \in I_a(\epsilon)\} \geq 1 - \delta;$$

see Theorem 1 in Section 3.2 for more details.

Figure 1 gives the algorithm for our two-stage stopping procedure for producing a confidence interval of absolute half-width ϵ for the steady-state mean μ of a discrete-time process $\mathbf{Y} = \{Y_n : n \geq 0\}$. (We can easily modify the algorithm to work with continuous-time processes.) The variable N_a in the algorithm is the total number of batches collected, and L and R, respectively, represent the resulting left and right endpoints of the confidence interval.

The method just described gives rise to absolute-precision confidence intervals. However, we often would like to construct confidence intervals having a given relative precision. For example, we may want the resulting confidence interval to have total width which is, say, 10% of the point estimate. So now we extend the previous methods in order to have a two-stage procedure that produces relative-precision confidence intervals. Again, we first simulate an initial pilot run of m batches, each of size $1/(m\epsilon^2)$. Using the initial m batch means, we compute the sample variance $s^2(\epsilon)$ and

$$N_r(\epsilon) = \max\left\{m, \left\lceil \frac{s^2(\epsilon)t^2_{m-1,\delta}}{\epsilon^2\left(\frac{1}{m}\sum_{i=1}^m Z_i(\epsilon)\right)^2} \right\rceil\right\}.$$

Note that the second term on the right side of the above equation contains the sample mean of the first m batch means in the denominator. Then simulate $N_r(\epsilon) - m$ batches of size $1/(m\epsilon^2)$ for the second stage, and form the relative-precision confidence interval

$$I_r(\epsilon) = \left[\hat{\mu}_r(\epsilon) - \epsilon|\hat{\mu}_r(\epsilon)|,\ \hat{\mu}_r(\epsilon) + \epsilon|\hat{\mu}_r(\epsilon)|\right],$$

where

$$\hat{\mu}_r(\epsilon) = \frac{1}{N_r(\epsilon)}\sum_{i=1}^{N_r(\epsilon)} Z_i(\epsilon), \tag{4}$$

which is the point estimate for the steady-state mean obtained from the entire simulation. The resulting confidence interval $I_r(\epsilon)$ satisfies

$$\lim_{\epsilon\to 0} P\{\mu \in I_r(\epsilon)\} \geq 1 - \delta$$

$m \leftarrow$ number of batches in first stage ($m \geq 2$);
$\epsilon \leftarrow$ absolute width of confidence interval desired;
$b \leftarrow$ batch size ($b \sim 1/\epsilon^2$);
$t_{m-1,\delta} \leftarrow 100(1-\delta/2)\%$ quantile point of a Student's
t-distribution with $m-1$ degrees of freedom;

begin
for $i \leftarrow 1, \ldots, m$, **do begin**
$Z_i \leftarrow \frac{1}{b} \sum_{j=1}^{b} Y_{j+(i-1)b}$;
end
$s \leftarrow \left[\frac{1}{m-1} \sum_{i=1}^{m} \left(Z_i - \frac{1}{m} \sum_{j=1}^{m} Z_j \right)^2 \right]^{1/2}$;

$N_a \leftarrow \max \left\{ m, \left\lceil \frac{s^2 t^2_{m-1,\delta}}{\epsilon^2} \right\rceil \right\}$;
for $i \leftarrow m+1, \ldots, N_a$, **do begin**
$Z_i \leftarrow \frac{1}{b} \sum_{j=1}^{b} Y_{j+(i-1)b}$;
end;

$L \leftarrow \frac{1}{N_a} \sum_{i=1}^{N_a} Z_i - \epsilon$;
$R \leftarrow \frac{1}{N_a} \sum_{i=1}^{N_a} Z_i + \epsilon$;
end.

Figure 1: Absolute precision two-stage stopping procedure

$m \leftarrow$ number of batches in first stage ($m \geq 2$);
$\epsilon \leftarrow$ relative width of confidence interval desired;
$b \leftarrow$ batch size ($b \sim 1/\epsilon^2$);
$t_{m-1,\delta} \leftarrow 100(1-\delta/2)\%$ quantile point of a Student's
t-distribution with $m-1$ degrees of freedom;

begin
for $i \leftarrow 1, \ldots, m$, **do begin**
$Z_i \leftarrow \frac{1}{b}\sum_{j=1}^{b} Y_{j+(i-1)b}$;
end
$s \leftarrow \left[\frac{1}{m-1}\sum_{i=1}^{m}\left(Z_i - \frac{1}{m}\sum_{j=1}^{m} Z_j\right)^2\right]^{1/2}$;

$N_r \leftarrow \max\left\{m, \left\lceil \frac{s^2 t^2_{m-1,\delta}}{\epsilon^2\left(\frac{1}{m}\sum_{i=1}^{m} Z_i\right)^2}\right\rceil\right\}$;
for $i \leftarrow m+1, \ldots, N_a$, **do begin**
$Z_i \leftarrow \frac{1}{b}\sum_{j=1}^{b} Y_{j+(i-1)b}$;
end;

$L \leftarrow \frac{1-\epsilon}{N_r}\sum_{i=1}^{N_a} Z_i$;
$R \leftarrow \frac{1+\epsilon}{N_r}\sum_{i=1}^{N_a} Z_i$;
end.

Figure 2: Relative precision two-stage stopping procedure

provided $\mu \neq 0$; see Theorem 3 in Section 3.2 for more details.

Figure 2 gives the algorithm for our two-stage stopping procedure to produce a confidence interval of relative half-width ϵ for the steady-state mean μ of a discrete-time process $\mathbf{Y} = \{Y_n : n \geq 0\}$.

3 Generalizing Our Two-Stage Batch Means Procedure

In the previous section we described how Stein's two-stage procedure can be combined with the method of batch means. The method of batch means was primarily used to form an "estimate" of the asymptotic variance of the stochastic process from the output of the first stage. (As Glynn and Whitt 1991 have shown, the "estimate" of the variance is not consistent.) We can generalize the schemes by using any standardized time series methodology in the first stage to "estimate" the variance. Before describing the new methods, we first discuss a condition imposed on the stochastic process being simulated.

3.1 An Assumption on the Stochastic Process Being Simulated

We now review some of the mathematical machinery that will be used. Let $D[0,\infty)$ denote the space of real-valued functions x on $[0,\infty)$ which are right continuous and have left limits (i.e., for each

$t \geq 0$, $\lim_{s \to t^+} x(s) = x(t)$ and $\lim_{s \to t^-} x(s) \equiv x(t-)$ exists), and let d denote the Skorohod metric on the space $D[0, \infty)$; i.e., $d(x, y)$ measures the distance between two elements $x, y \in D[0, \infty)$. Virtually all stochastic processes arising in operations research applications lie in $D[0, \infty)$. Also, let $C[0, \infty)$ denote the space of continuous functions x on $[0, \infty)$, and note that $C[0, \infty) \subset D[0, \infty)$. For more details on the spaces $D[0, \infty)$ and $C[0, \infty)$, the reader is referred to Ethier and Kurtz (1986) and Glynn (1989).

We define a function $h : C[0, \infty) \to \Re$ to be *continuous* at $x \in C[0, \infty)$ if $h(x_\epsilon) \to h(x)$ as $\epsilon \to 0$, whenever $d(x_\epsilon, x) \to 0$ as $\epsilon \to 0$, where $x_\epsilon \in C[0, \infty)$ for all $\epsilon > 0$. In this paper we will consider functions $h : C[0, \infty) \to \Re$ which are typically not continuous, and we let $D(h)$ denote the set of points $x \in C[0, \infty)$ at which h is not continuous.

Let $\{X_\epsilon : \epsilon > 0\}$ be a family of random elements taking values in $C[0, \infty)$; i.e., the X_ϵ correspond to stochastic processes with sample paths in $C[0, \infty)$. If X is a random element of $C[0, \infty)$, then the X_ϵ are said to converge *weakly* to X (written $X_\epsilon \Rightarrow X$ as $\epsilon \to 0$) if

$$Ef(X_\epsilon) \to Ef(X)$$

as $\epsilon \to 0$, for every bounded, continuous function $f : C[0, \infty) \to \Re$; see the appendix for more details.

In our development we will only consider random elements X_ϵ having a particular form. More specifically, let $\mathbf{Y} = \{Y(t) : t \geq 0\} \in D[0, \infty)$ be a real-valued (measurable) stochastic process representing the output of a simulation. We can deal with discrete-time processes by taking $Y(t) = Y_{\lfloor t \rfloor}$, where $\lfloor q \rfloor$ denotes the greatest integer less than or equal to q. We define

$$X_\epsilon(t) = \frac{1}{\epsilon}(\bar{Y}_\epsilon(t) - \mu t) \tag{5}$$

with

$$\bar{Y}_\epsilon(t) = \frac{\int_0^{t/\epsilon^2} Y(s)ds}{1/\epsilon^2}, \quad \text{for } t \geq 0.$$

X_ϵ represents a cumulative simulation-generated estimator of the original process $\mathbf{Y}$. Note that $X_\epsilon \in C[0, \infty)$.

To obtain our results, we need to assume that our stochastic process $\mathbf{Y}$ satisfies the following functional central limit theorem (FCLT):

A1. There exist finite constants μ and σ ($\sigma > 0$) such that

$$X_\epsilon \Rightarrow \sigma B$$

as $\epsilon \to 0$, where B is a standard Brownian motion and X_ϵ is defined in (5).

A similar assumption is used by Glynn and Whitt (1992) when proving the validity of their fully sequential stopping procedures. In addition Glynn and Iglehart (1990) assume a FCLT to establish the validity of standardized time series.

Recall that every sample path of a Brownian motion is continuous and so $B \in C[0,\infty)$. Thus, since $X_\epsilon \in C[0,\infty)$ also, we are working with weak convergence in the space $C[0,\infty)$. In addition note that the time parameter of the processes $X_\epsilon(\cdot)$ and $\bar{Y}_\epsilon(\cdot)$ are rescaled by $1/\epsilon^2$ as compared to the time parameter of the original process $Y(\cdot)$.

Assumption A1 guarantees that

$$\frac{\int_0^t Y(s)ds}{t} - \mu = \frac{1}{\sqrt{t}} X_{1/\sqrt{t}}(1) \Rightarrow 0 \cdot \sigma B(1) = 0$$

as $t \to \infty$, and so $\int_0^t Y(s)ds/t \Rightarrow \mu$ as $t \to \infty$. Thus, μ is the steady-state mean that the simulator wishes to estimate. Also, Assumption A1 gives us

$$\sqrt{t}\left[\frac{\int_0^t Y(s)ds}{t} - \mu\right] = X_{1/\sqrt{t}}(1) \Rightarrow \sigma B(1)$$

as $t \to \infty$. Recalling that $B(1)$ is normally distributed with mean 0 and variance 1, we see that σ^2 is the asymptotic variance constant of the process $\mathbf{Y}$.

Virtually all "real-world" steady-state simulations satisfy Assumption A1. In particular a wide spectrum of different assumptions on the structure of the output process $\mathbf{Y}$ give rise to FCLTs of the form in Assumption A1. For example, Assumption A1 holds when any one of the following is in force:

(i) $\mathbf{Y}$ is regenerative and satisfies suitable moment conditions (see Glynn and Whitt 1987);

(ii) $\mathbf{Y}$ is a martingale process (see Chapter 7 of Ethier and Kurtz 1986);

(iii) $\mathbf{Y}$ satisfies appropriate mixing conditions (see Chapter 7 of Ethier and Kurtz 1986);

(iv) there is appropriate positive dependence in the process $\mathbf{Y}$ (specifically, when the $Y(t)$ are associated; see Newman and Wright 1981).

3.2 Two-stage procedures using Standardized Time Series

Now we are in a position to describe how to generalize the two-stage batch means procedure described in Section 2 to work with any standardized time series methodology. An overview of the scheme is as follows. We first simulate an initial pilot run (the first stage) of the original process $\mathbf{Y}$ of length $1/\epsilon^2$ and divide this into m batches, each of length $1/(m\epsilon^2)$, where $m \geq 1$ is fixed. Now instead of forming the sample variance of the first m batch means as done in Section 2, we apply a standardized time series technique to the output of the first stage to form an "estimate" of the asymptotic variance. (As noted by Glynn and Iglehart 1990, the "estimate" of the variance is not consistent.) Using this "estimate" of the variance, we determine the total number of batches needed in the same way as done in Section 2.

We now discuss how the estimate of the variance is formed using a standardized time series. The basis of standardized time series is a class of functions g developed by Glynn and Iglehart (1990) which are applied to an integrated (accumulated) version of the entire simulation output $\mathbf{Y}$, namely,

X_ϵ. In our two-stage procedures we slightly modify the definition of the functions so that they only depend on the first stage of the simulation. The functions g actually depend on the number of batches m in the first stage, but for notational convenience, we do not show explicitly the dependence. We define the class of functions in the next assumption:

A2. The (measurable) function $g : C[0,\infty) \to \Re$ satisfies the following conditions:

- **(i)** For every $x \in C[0,\infty)$, the quantity $g(x)$ depends only on $\{x(s) : 0 \le s \le 1\}$. Thus, if $x, y \in C[0,\infty)$ and $x(s) = y(s)$ for all $0 \le s \le 1$, then $g(x) = g(y)$.
- **(ii)** $g(\alpha x) = \alpha g(x)$ for $\alpha > 0$, $x \in C[0,\infty)$.
- **(iii)** $g(x - \beta k) = g(x)$ for $\beta \in \Re$ and $x \in C[0,\infty)$, where $k(t) = t$.
- **(iv)** $P\{g(B) > 0\} = 1$.
- **(v)** $P\{B \in D(g)\} = 0$.
- **(vi)** $g(B)$ has a continuous distribution function.

Let $\mathcal{M}'$ be the class of functions g satisfying Assumption A2. Condition (i) ensures that the function g only depends on the evolution of the process up to time 1. Applying the function g to the process X_ϵ, we see that $g(X_\epsilon)$ is solely determined by $\{X_\epsilon(s) : 0 \le s \le 1\}$, which corresponds to the time interval from 0 to $1/\epsilon^2$ of the original process $\mathbf{Y}$. Hence, g depends only on the evolution of the process in the first stage. Condition (ii) ensures that, in some sense, $g(X_\epsilon)$ is a well-behaved "estimator" of the parameter σ. More specifically, (ii) guarantees that if we change the basic units of measurement of the simulation output so that all observations are effectively multiplied by some positive factor α and the corresponding variance parameter is $\alpha\sigma$, then the estimator based on the function g will also properly reflect this change in the units of measurement. This property will allow us to "cancel out" the asymptotic variance constant σ; for more details, see the proof of Theorem 1 in the appendix. Condition (iii) guarantees that $g(X_\epsilon)$ does not depend on the unknown parameter μ. Conditions (iv)–(vi) are technical assumptions required to invoke the continuous mapping principle.

The class $\mathcal{M}'$ is the same as the class $\mathcal{M}$ described by Glynn and Iglehart (1990) upon which the theory of standardized time series is built, except that we have the additional conditions (i) and (vi). Condition (i) of Assumption A2 ensures that functions in $\mathcal{M}'$ have essentially the same domain as the mappings in $\mathcal{M}$ possess, and so it does not really affect the class of functions being considered. On the other hand, condition (vi) is an additional restriction. We should note though that condition (vi) is satisfied in all of the applications that we have in mind. Hence, condition (vi) does not seem to be a restrictive assumption in practice. In fact, it is probably true that conditions (i)–(v) imply condition (vi), but we have not been able to verify this. However, by slightly modifying the proof of Proposition 4.26 of Glynn and Iglehart (1990), it can be shown that conditions (i)–(v) imply that the distribution of $g(B)$ cannot solely consist of a finite or countable number of atoms that are bounded away from 0. This is necessary for condition (vi) to hold.

As discussed in Glynn and Iglehart (1990), the method of batch means (with a fixed number of batches) is an example of a standardized time series methodology. Hence, we can define a function g that corresponds to this procedure; i.e., there exists a g that will give rise to (3). For more details on this and other g functions, see Section 4.

Now we describe our new two-stage procedures based on standardized time series. First, simulate a first stage of length $1/\epsilon^2$ divided into $m \geq 1$ batches of equal size, and then define

$$s(\epsilon) = m^{1/2}\epsilon g(X_\epsilon),$$

where $g\ (= g_m) \in \mathcal{M}'$. The quantity $s(\epsilon)$ is our "estimate" of the variance parameter σ obtained using the standardized time series method corresponding to g. From Assumption A2(i), $s(\epsilon)$ only depends on the evolution of the process X_ϵ in the first stage. Next we define the total number of batches needed as

$$N_a(\epsilon) = \max\left\{m, \left\lceil \frac{s^2(\epsilon)a_\delta^2}{\epsilon^2} \right\rceil\right\}, \tag{6}$$

where a_δ is the $100(1-\delta/2)\%$ quantile point of the random variable $B(1)/g(B)$. Glynn and Iglehart (1990) show that the distribution function of $B(1)/g(B)$ is continuous and strictly increasing, so the quantile exists. In the second stage simulate the remaining $N_a(\epsilon) - m$ batches, each of size $1/(m\epsilon^2)$, and form the absolute-precision confidence interval

$$I_a(\epsilon) = \left[\frac{1}{N_a(\epsilon)}\sum_{i=1}^{N_a(\epsilon)} Z_i(\epsilon) - \epsilon, \frac{1}{N_a(\epsilon)}\sum_{i=1}^{N_a(\epsilon)} Z_i(\epsilon) + \epsilon\right],$$

where the $Z_i(\epsilon)$ are defined in (2). Then we have the following result.

Theorem 1 *Suppose that $g \in \mathcal{M}'$. If Assumption A1 holds, then*

(i) *$N_a(\epsilon) \Rightarrow N_a$ as $\epsilon \to 0$, where N_a is some proper limiting random variable,*

(ii) *$\lim_{\epsilon\to 0} P\{\mu \in I_a(\epsilon)\} \geq 1 - \delta$.*

Part (i) shows that the total number of batches needed converges to some proper limiting random variable. The asymptotic validity of the confidence intervals produced by the procedure is shown in part (ii). The proof is given in the appendix.

Because of the presence of the ceiling function in (6), the previous method often requires the simulation to be run longer than necessary. This occurs when a second stage is needed, and the simulation is run until the size of the last batch is the same as the others. In order to alleviate this problem, we now describe another two-stage procedure in which the ceiling function when determining the total run length is eliminated, thus creating a method which gives rise to a total run length that is never longer, and often shorter, than that of the previous procedure.

To this end, define

$$Q_a(\epsilon) = \max\left\{m, \frac{s^2(\epsilon)a_\delta^2}{\epsilon^2}\right\}. \tag{7}$$

We use $Q_a(\epsilon)$ as a stopping rule in the following manner. First simulate the m initial batches of size $1/(m\epsilon^2)$ in the first stage as before. However, in the second stage, we simulate from $1/\epsilon^2$ to $Q_a(\epsilon)/(m\epsilon^2)$ and consider the whole segment as one final batch. So define $Z_i(\epsilon)$, for $i = 1, \ldots, m$, as in (2), and let

$$Z'_{m+1}(\epsilon) = \frac{\int_{m/(m\epsilon^2)}^{Q_a(\epsilon)/(m\epsilon^2)} Y(s)ds}{(Q_a(\epsilon) - m)/(m\epsilon^2)}$$

be the sample mean of the process in the second stage. Finally, we define a new absolute-precision confidence interval

$$I'_a(\epsilon) = [\hat{\mu}_a(\epsilon) - \epsilon, \hat{\mu}_a(\epsilon) + \epsilon],$$

where

$$\hat{\mu}_a(\epsilon) = \frac{1}{Q_a(\epsilon)} \left[\sum_{i=1}^{m} Z_i(\epsilon) + (Q_a(\epsilon) - m) Z'_{m+1}(\epsilon) \right].$$

Note that the batch corresponding to $Z'_{m+1}(\epsilon)$ is not necessarily the same size as the first m batches. Thus, we have scaled $Z'_{m+1}(\epsilon)$ by $Q_a(\epsilon) - m$. The procedure satisfies the following result.

Theorem 2 *Suppose that $g \in \mathcal{M}'$. If Assumption A1 holds, then*

(i) *$Q_a(\epsilon) \Rightarrow Q_a$ as $\epsilon \to 0$, where Q_a is some proper limiting random variable,*

(ii) $\lim_{\epsilon \to 0} P\{\mu \in I'_a(\epsilon)\} \geq 1 - \delta$.

The proof is omitted as this result can be shown following an argument similar to that used to establish Theorem 1.

On the set $\{N_a(\epsilon) > m\}$, we have $N_a(\epsilon) = \lceil Q_a(\epsilon) \rceil$. Thus, by using $Q_a(\epsilon)$ rather than $N_a(\epsilon)$ to determine the total run length, we avoid having to simulate from $Q_a(\epsilon)/\epsilon^2$ to $\lceil Q_a(\epsilon) \rceil/\epsilon^2$ when $N_a(\epsilon) > m$. For small values of ϵ, this can lead to significant savings in the total run length.

Now we extend the previous methods in order to have a two-stage procedure that produces relative-precision confidence intervals. Thus, define

$$N_r(\epsilon) = \max \left\{ m, \left\lceil \frac{s^2(\epsilon) a_\delta^2}{\epsilon^2 \left(\frac{1}{m} \sum_{i=1}^{m} Z_i(\epsilon) \right)^2} \right\rceil \right\}, \tag{8}$$

which is the total number of batches needed. Also, let

$$\hat{\mu}_r(\epsilon) = \frac{1}{N_r(\epsilon)} \sum_{i=1}^{N_r(\epsilon)} Z_i(\epsilon),$$

be the point estimate of the steady-state mean obtained from the entire simulation. Finally, define the relative-precision confidence interval

$$I_r(\epsilon) = [\hat{\mu}_r(\epsilon) - \epsilon|\hat{\mu}_r(\epsilon)|,\ \hat{\mu}_r(\epsilon) + \epsilon|\hat{\mu}_r(\epsilon)|].$$

Then we have the following result.

Theorem 3 *Suppose that $g \in \mathcal{M}'$. If Assumption A1 holds and $\mu \neq 0$, then*

(i) *$N_r(\epsilon) \Rightarrow N_r$ as $\epsilon \to 0$, where N_r is some proper limiting random variable,*

(ii) *$\lim_{\epsilon \to 0} P\{\mu \in I_r(\epsilon)\} \geq 1 - \delta$.*

The proof is given in the appendix.

It should be pointed out that Stein (1945) only discusses absolute-precision confidence intervals. The reason that we are able to derive a relative-precision confidence interval procedure is that for fixed m, $\frac{1}{m}\sum_{i=1}^{m} Z_i(\epsilon) = \bar{Y}_\epsilon(1) \Rightarrow \mu$ as $\epsilon \to 0$, and so we have an asymptotically consistent estimator of the steady-state mean. However, in Stein's procedure, since m is fixed, there is no consistent estimate of the mean of the distribution available from the first stage.

As in the case for absolute-precision two-stage methods, we can define a relative-precision two-stage procedure in which the total run length is never longer, and often shorter, than that of the previous method by eliminating the ceiling function from (8). To do this, we define

$$Q_r(\epsilon) = \max\left\{m, \frac{s^2(\epsilon)a_\delta^2}{\epsilon^2\left(\frac{1}{m}\sum_{i=1}^{m} Z_i(\epsilon)\right)^2}\right\}. \tag{9}$$

Also, define the relative-precision confidence interval

$$I_r'(\epsilon) = \left[\hat{\mu}_r'(\epsilon) - \epsilon|\hat{\mu}_r'(\epsilon)|,\ \hat{\mu}_r'(\epsilon) + \epsilon|\hat{\mu}_r'(\epsilon)|\right],$$

where

$$\hat{\mu}_r'(\epsilon) = \frac{1}{Q_r(\epsilon)}\left[\sum_{i=1}^{m} Z_i(\epsilon) + (Q_r(\epsilon) - m)Z_{m+1}'(\epsilon)\right].$$

Then we have the following theorem.

Theorem 4 *Suppose that $g \in \mathcal{M}'$. If Assumption A1 holds, then*

(i) *$Q_r(\epsilon) \Rightarrow Q_r$ as $\epsilon \to 0$, where Q_r is some proper limiting random variable,*

(ii) *$\lim_{\epsilon \to 0} P\{\mu \in I_r'(\epsilon)\} \geq 1 - \delta$.*

We omit the proof of this result as it is very similar to that of Theorem 3.

4 Examples of Standardized Time Series

In this section we describe various functions $g \in \mathcal{M}'$ and the resulting standardized time series. All of these examples are taken directly from Glynn and Iglehart (1990).

Example 1. The first function g that we describe corresponds to the method of batch means, which we considered in Section 2. For this example we require that the number of initial batches m is at least 2. Define the function $g : C[0,\infty) \to \Re$ as

$$g(x) = \left[\frac{m}{m-1}\sum_{i=1}^{m}\left(\Delta_m x\left(\frac{i}{m}\right) - \frac{x(1)}{m}\right)^2\right]^{1/2},$$

where

$$\Delta_h x(t) = x(t) - x\left(t - \frac{1}{h}\right). \tag{10}$$

Thus, we have

$$\begin{aligned} g(X_\epsilon) &= \frac{1}{m^{1/2}\epsilon}\left[\frac{1}{m-1}\sum_{i=1}^{m}\left(Z_i(\epsilon) - \frac{1}{m}\sum_{j=1}^{m} Z_j(\epsilon)\right)^2\right]^{1/2}, \\ g(\sigma B) &= m^{1/2}\left[\frac{1}{m-1}\sum_{i=1}^{m}\left(\Delta_m \sigma B\left(\frac{i}{m}\right) - \frac{\sigma B(1)}{m}\right)^2\right]^{1/2}, \end{aligned}$$

where $Z_i(\epsilon)$ is defined in (2). Note that $g(X_\epsilon)$ is equal to the sample standard deviation of the batch means divided by $m^{1/2}\epsilon$. Conditions (i)–(v) of $\mathcal{M}'$ are easily verified. Condition (vi) of $\mathcal{M}'$ can be demonstrated to hold as follows. Note that $\Delta_m \sigma B(i/m), i = 1, \ldots, m$, are increments of Brownian motion and so are independent normally distributed random variables with mean 0 and variance σ^2/m. Also, $\sigma B(1)/m$ is the sample mean of the m increments. Thus, $(m-1)g^2(\sigma B)/\sigma^2$ has a χ^2 distribution with $m-1$ degrees of freedom. Hence, $g(B)$ has a continuous distribution function, and so condition (vi) holds. Furthermore, from Proposition 2.8 of Glynn and Iglehart (1990), $B(1)$ is independent of $g(B)$, and so $B(1)/g(B)$ has a Student's t-distribution with $m-1$ degrees of freedom; therefore, the quantile point is given by $a_\delta = t_{m-1,\delta}$, where $t_{m-1,\delta}$ is the $100(1-\delta/2)\%$ quantile point of a Student's t-distribution having $m-1$ degrees of freedom. The algorithms for the two-stage procedures corresponding to Theorems 1 and 3 when using this function g (i.e., the method of batch means) are given in Figures 1 and 2, respectively.

Before discussing our next example, we first give some motivation. As noted in Glynn and Iglehart (1990), the basic assumption necessary for the validity of the standardized time series procedures is that we can approximate the output process by a Brownian motion. Thus, it seems reasonable that we can approximate the increments of our first stage by increments of a Brownian motion. This suggests that in order to obtain more powerful procedures, we can apply one of the standardized time series methods to each of the initial batches and then combine them.

In order to do this, we first make some definitions. Let $x \in C[0,\infty)$. We define the functions $\Gamma : C[0,\infty) \to C[0,\infty)$ and $\Lambda_i : C[0,\infty) \to C[0,\infty)$ for $i = 0, \ldots, m-1$, as

$$\begin{aligned} \Gamma(x) &= x - kx(1) \\ \Lambda_i(x) &= x\left(\frac{i+k}{m}\right) - x\left(\frac{i}{m}\right), \end{aligned}$$

where $k(t) = t$. Note that

$$(\Gamma \circ \Lambda_i)(x) = x\left(\frac{i+k}{m}\right) - x\left(\frac{i}{m}\right) - kx\left(\frac{i+1}{m}\right) + kx\left(\frac{i}{m}\right).$$

Now we define $\mathcal{N}'$ to be the class of functions $b : C[0,\infty) \to \Re$ which satisfy

(i) For every $x \in C[0,\infty)$, the quantity $b(x)$ only depends on $\{x(s) : 0 \le s \le 1\}$.

(ii) $b(\alpha x) = \alpha b(x)$ for $\alpha > 0$, $x \in C[0, \infty)$.

(iii) $P\{(b \circ \Gamma)(B) > 0\} = 1$.

(iv) $P\{B \in D(b \circ \Gamma)\} = 0$.

(v) $(b \circ \Gamma)(B)$ has a continuous distribution function.

Finally, we define our function $g : C[0, \infty) \to \Re$ as

$$g(x) = \left(\sum_{i=0}^{m-1} (b^2 \circ \Gamma \circ \Lambda_i)(x) \right)^{1/2},$$

where $b^2(x) = b(x) \cdot b(x)$. Condition (i) of $\mathcal{N}'$ ensures that Assumption A2(i) is satisfied. Also, it can be easily shown that item (v) holds if and only if Assumption A2(vi) is satisfied. Thus, the class of functions $\mathcal{M}^* = \{g : g = b \circ \Gamma, b \in \mathcal{N}'\}$ is equivalent to the class $\mathcal{M}'$; see Glynn and Iglehart (1990).

Example 2. Let $m \geq 1$. We define the functions $b : C[0, \infty) \to \Re$ as

$$b(x) = \left| \int_0^1 x(t)dt \right|.$$

Thus, we have

$$\begin{aligned} g(x) &= \left[\sum_{i=0}^{m-1} \left| \int_0^1 x\left(\frac{i+t}{m} \right) dt - \frac{1}{2} \left(x\left(\frac{i+1}{m} \right) + x\left(\frac{i}{m} \right) \right) \right|^2 \right]^{1/2} \\ g(X_\epsilon) &= \epsilon \left[\sum_{i=0}^{m-1} \left(\int_0^1 \int_{i/(m\epsilon^2)}^{(i+t)/(m\epsilon^2)} Y(s)ds \; dt - \frac{1}{2} \int_{i/(m\epsilon^2)}^{(i+1)/(m\epsilon^2)} Y(s)ds \right)^2 \right]^{1/2}. \end{aligned}$$

Glynn and Iglehart (1990) prove that $\int_0^1 (\Gamma B)(t)dt$ is normally distributed with mean 0 and variance $1/12$, and so $(b \circ \Gamma)(B)$ has a continuous distribution. Hence, it can be shown that $b \in \mathcal{N}'$, and so $g \in \mathcal{M}'$; see Glynn and Iglehart (1990). Finally, Glynn and Iglehart (1990) show that the random variable $12mg^2(B)$ has a χ^2 distribution with m degrees of freedom and that $B(1)$ is always independent of $g(B)$. Hence, $B(1)/((12m)^{1/2}g(B))$ has a Student's t-distribution with m degrees of freedom, and so the quantile point $a_\delta = (12m)^{1/2}t_{m,\delta}$. The resulting function g gives rise to the *standardized sum process* described by Schruben (1983). The algorithms for the resulting two-stage procedures are explicitly given in Nakayama (1992).

In a similar manner we can define a function g that corresponds to the *standardized maximum intervals* method developed by Schruben (1983). A description of this function and algorithms for the resulting two-stage procedures are given in Nakayama (1992).

5 Implementation Issues

The algorithms given in Figures 1 and 2 for implementing the two-stage batch means procedures have parameters that need to be specified prior to running the simulation. These parameters are the desired (absolute or relative) width of the resulting confidence interval (ϵ), the size of the initial batches (b), and the number of initial batches (m). Theorems 1–4 all require that $\epsilon \to 0$ and $b \to \infty$, with b proportional to $1/\epsilon^2$, in order for the results to hold. In practice though, the asymptotics start taking affect for "reasonable" values of ϵ and b.

However, we still need to determine appropriate values for these parameters. Lavenberg and Sauer (1977) state that selecting $\epsilon \leq 0.025$ seems to work well in practice for their fully sequential stopping procedure for constructing relative-precision confidence intervals. It is probably reasonble to assume that we can also choose ϵ in this range for our methods. Determining an appropriate batch size for a given half-width, though, is more difficult. However, when we are simulating a queueing system, the results of Whitt (1989a, 1989b) may be applied in order to obtain a rough idea of how to select an appropriate batch size. Whitt calculates approximate values of the asymptotic variance constants by using heavy-traffic limits for queues and associated diffusion approximations. Using these values, Whitt estimates the total run length t needed so that the resulting confidence interval is roughly of desired (absolute or relative) width. In our context, we might, for example, let the simulation corresponding to the first stage be of total length $t/2$ by letting each of the first stage batches be of size $t/(2m)$, when the number of initial batches m is fairly small. By doing this, we (hopefully) ensure that the size of the initial batches is sufficiently large so that they are approximately independent and normally distributed. Also, the second stage, when needed, guarantees that the total run length is of appropriate size.

We must also select the total number, m, of batches in the first stage. When using the procedure corresponding to Theorem 1 in conjunction with the method of batch means (see Section 2), we can refer to previous results on Stein's two-stage procedure, since our procedure in this case and Stein's (1945) two-stage procedure are asymptotically equivalent (in the sense that the batch means are asymptotically i.i.d. normally distributed random variables and the function g used for batch means gives rise to a quantity that corresponds exactly to Stein's estimate of the variance that he forms using the first stage observations). Seelbinder (1953) computes the expected total sample size for various values of m and gives tables of this. Moshman (1958) suggests that we should consider both the resulting expected total sample size and an upper percentage point of the distribution of the total sample size when determining m. However, they both carry through their analyses under the assumption that we have an idea of the value of the variance constant σ^2 before taking any samples, which may be an unrealistic assumption. On the other hand, we may follow Schmeiser's (1982) suggestions for selecting the total number of batches to be used in a nonsequential batch means procedure. Schmeiser suggests that for a fixed run length, the total number of batches in a batch means procedure should be kept fairly small (between 10 and 30), thus allowing the size of each batch to be fairly large. Hence, the implicit assumption of the method of batch means that the batches are almost independent and normally distributed is more likely to be satisfied, making the

Batch Sizes	Customers in Stage 1	m	ϵ	Abs/Rel Precision	Coverage	Avg Total Batches	Avg Total Customers
Equal	6720	7	0.1	abs	92.7%	12.88	12364.8
Equal	6720	7	0.05	abs	88.9%	47.45	45551.0
Equal	6720	7	0.025	abs	88.2%	188.17	180646.6
Unequal	6720	7	0.1	abs	92.0%	12.55	12045.1
Unequal	6720	7	0.05	abs	88.0%	46.95	45074.2
Unequal	6720	7	0.025	abs	88.5%	187.67	180166.0
Equal	6720	7	0.1	rel	92.2%	12.50	11999.5
Equal	6720	7	0.05	rel	89.6%	46.11	44269.4
Equal	6720	7	0.025	rel	89.4%	182.89	175570.1
Unequal	6720	7	0.1	rel	91.7%	12.15	11663.2
Unequal	6720	7	0.05	rel	88.8%	45.63	43803.3
Unequal	6720	7	0.025	rel	89.4%	182.39	175097.3
Equal	1680	7	0.1	abs	85.8%	43.75	10500.8
Equal	1680	7	0.05	abs	84.9%	173.31	41594.0
Equal	1680	7	0.025	abs	85.4%	691.74	166016.8
Unequal	1680	7	0.1	abs	85.3%	43.26	10383.5
Unequal	1680	7	0.05	abs	84.8%	172.81	41474.5
Unequal	1680	7	0.025	abs	85.3%	691.24	165896.6
Equal	1680	7	0.1	rel	87.7%	40.57	9736.4
Equal	1680	7	0.05	rel	86.2%	160.69	38564.5
Equal	1680	7	0.025	rel	86.4%	641.27	153904.8
Unequal	1680	7	0.1	rel	86.8%	40.09	9620.5
Unequal	1680	7	0.05	rel	86.3%	160.19	38446.5
Unequal	1680	7	0.025	rel	86.4%	640.78	153786.0

Table 1: Coverage results using two-stage batch means stopping procedure from 2000 independent replications of estimating the expected waiting time in an M/M/1 queue with $\lambda = .5$, $\mu = 1$.

procedure more robust. In our context, we should probably select the number of initial batches m to be on the lower end of Schmeiser's suggested range since there is still the second stage in our procedure. Thus, m should be chosen to be between 5 and 15 for our two-stage procedures.

6 Experimental Results

In this section we present some empirical results obtained from simulations using the two-stage stopping procedures proposed in this paper. The purpose of the experiments was not so much to thoroughly test the methods, but rather to demonstrate the interdependence between the batch size and the desired confidence interval half-width and how this changes for different models. Furthermore, we show how Whitt's (1989a) calculations (referred to in Section 5) can be employed to determine an appropriate batch size for a given half-width in a simulation of a queueing system.

We ran simulations to estimate the expected waiting time in an M/M/1 queue. The method of

Batch Sizes	Customers in Stage 1	m	ϵ	Abs/Rel Precision	Coverage	Avg Total Batches	Avg Total Customers
Equal	6720	6	0.1	abs	92.0%	11.02	12346.3
Equal	6720	6	0.05	abs	88.1%	40.06	44867.2
Equal	6720	6	0.025	abs	87.6%	158.61	177647.7
Unequal	6720	6	0.1	abs	91.7%	10.70	11981.7
Unequal	6720	6	0.05	abs	87.4%	39.57	44317.3
Unequal	6720	6	0.025	abs	88.5%	158.11	177079.2
Equal	6720	6	0.1	rel	91.8%	10.67	10880.9
Equal	6720	6	0.05	rel	88.0%	38.82	43477.8
Equal	6720	6	0.025	rel	88.7%	153.66	172097.0
Unequal	6720	6	0.1	rel	91.5%	10.34	11576.3
Unequal	6720	6	0.05	rel	88.1%	38.32	42920.4
Unequal	6720	6	0.025	rel	89.1%	153.15	171529.2
Equal	1680	6	0.1	abs	84.4%	35.66	9984.1
Equal	1680	6	0.05	abs	82.6%	140.85	39438.8
Equal	1680	6	0.025	abs	82.8%	561.92	157337.0
Unequal	1680	6	0.1	abs	84.0%	35.18	9851.0
Unequal	1680	6	0.05	abs	82.4%	140.35	39299.2
Unequal	1680	6	0.025	abs	82.7%	561.42	157196.9
Equal	1680	6	0.1	rel	85.4%	33.41	9353.4
Equal	1680	6	0.05	rel	83.9%	131.97	36951.6
Equal	1680	6	0.025	rel	84.6%	526.39	147388.6
Unequal	1680	6	0.1	rel	84.8%	32.93	9219.9
Unequal	1680	6	0.05	rel	83.6%	131.47	36811.9
Unequal	1680	6	0.025	rel	84.2%	525.88	147247.5

Table 2: Coverage results using two-stage standardized sum stopping procedure from 2000 independent replications of estimating the expected waiting time in an M/M/1 queue with $\lambda = .5$, $\mu = 1$.

batch means (BM) and the standardized sum procedure (SS) were each used in different experiments as the standardized time series method in the first stage (see Examples 1 and 2 in Section 4, respectively). Also, each of the experiments was run using one of the two-stage procedures corresponding to Theorems 1–4; i.e., we used procedures with either equal sized batches or the last batch being of different size in order to construct absolute or relative precision confidence intervals. Finally, we varied the desired (absolute or relative) half-width ϵ of the confidence interval from 0.025 to 0.1. Tables 1 and 2 contain the results from using the two-stage batch means stopping procedure and two-stage standardized sum stopping procedure, respectively, with $\lambda = 0.5$ and $\mu = 1.0$. Both of these tables contain the results from a total number of 2000 independent replications, thus ensuring that the coverages are accurate to 2 decimal places with 90% probability. Tables 3 and 4 contain the results from using the two-stage batch means stopping procedure and two-stage standardized sum stopping procedure, respectively, with $\lambda = 0.8$ and $\mu = 1.0$. These tables contain the results from a total number of 1000 independent replications, thus ensuring that the coverages are accurate to $\pm 2\%$ with 90% probability. For all of the cases, the numbers of initial batches m using the different standardized time series methodologies were selected so that the resulting Student's t-statistic had the same number of degrees of freedom. Thus, we let $m = 7$ when we used the method of batch means and $m = 6$ when the standardized sum method was implemented.

We selected the number of customers in the first stage as follows. The asymptotic variance of the waiting time process of an M/M/1 queue is

$$\sigma_\rho^2 = \frac{\rho(2 + 5\rho - 4\rho^2 + \rho^3)}{(1-\rho)^4},$$

where $\rho = \lambda/\mu$ is the traffic intensity (see Law 1975). Using Whitt's (1989a) calculations, in order to have the resulting $100(1-\delta)\%$ confidence interval be of absolute half-width ϵ, the total number of customers simulated should be

$$t_a(\rho, \epsilon, \delta) = \frac{\sigma_\rho^2 z_\delta^2}{\epsilon^2},$$

where z_δ is the $100(1-\delta/2)\%$ quantile point of a standard normal distribution. Thus, for $\rho = 0.5$, we have $\sigma_{0.5}^2 = 29$, and so we should simulate a total of $t_a(0.5, 0.1, 0.9) = 11140$ customers when we want to construct 90% confidence intervals having half-width $\epsilon = 0.1$. As suggested in Section 5, the number of customers in the first stage of our two-stage procedures should be approximately $t_a(0.5, 0.1, 0.9)/2 = 5570$. In addition we must select the batch sizes so that they are divisible by $m = 7$ and $m = 6$, which are the number of batches in the first stage when using BM and SS, respectively. Hence, to be on the safe side, we chose to have 6720 customers in the first stage in half of the experiments with $\rho = 0.5$. In order to show that the results only hold when the batch sizes are sufficiently large, we ran the rest of the experiments for $\rho = 0.5$ using 1680 customers in the first stage. When $\rho = 0.8$ as in our other experiments, we now have $\sigma_{0.8}^2 = 1976$, and so we should simulate a total of $t_a(0.5, 0.1, 0.9) = 759100$ customers. However, we let the first stage consist of 6720 customers when $\rho = 0.8$ in order to show that the number of customers that should be observed in the first stage depends on the stochastic process being simulated.

Batch Sizes	Customers in Stage 1	m	ϵ	Abs/Rel Precision	Coverage	Avg Total Batches	Avg Total Customers
Equal	6720	7	0.4	abs	83.6%	48.07	46144.3
Equal	6720	7	0.2	abs	81.5%	190.55	182927.0
Equal	6720	7	0.1	abs	83.2%	760.65	730225.0
Unequal	6720	7	0.4	abs	83.4%	47.60	45693.1
Unequal	6720	7	0.2	abs	81.5%	190.05	182443.3
Unequal	6720	7	0.1	abs	83.6%	760.16	729751.2
Equal	6720	7	0.1	rel	86.5%	41.64	39977.3
Equal	6720	7	0.05	rel	84.7%	164.99	158393.3
Equal	6720	7	0.025	rel	84.6%	658.51	632166.7
Unequal	6720	7	0.1	rel	85.8%	41.17	39522.2
Unequal	6720	7	0.05	rel	84.7%	164.50	157923.9
Unequal	6720	7	0.025	rel	84.3%	658.00	631680.2

Table 3: Coverage results using two-stage batch means stopping procedure from 1000 independent replications of estimating the expected waiting time in an M/M/1 queue with $\lambda = .8$, $\mu = 1$.

In Tables 1 and 2, we can see that when the number of customers in the first stage is 6720, the coverages for all of the methods are close to the nominal value of 90% when $\epsilon = 0.1$. However, when $\epsilon = 0.05$ or 0.025, the coverages start falling off slightly. Thus, we can see that the batch size and desired half-width do depend on each other. Also, when the number of customers in the first stage is 1680, all of the coverages are below 90%. Therefore, the sizes of the batches are too small for the corresponding values of ϵ in order for the procedures to be valid.

Also, the average total number of batches when $\epsilon = 0.1$ is relatively small, which agrees with part (i) of Theorems 1–4. (The total number of batches when the last batch is not of the same size as in the first stage (as in Theorems 2 and 4) is calculated by the total number of customers divided by the batch size of the first stage batches.)

Examining Tables 3 and 4, we see that all of the coverages are below 90%. Thus, the sizes of the batches are too small for the corresponding values of ϵ in order for the procedures to be valid, even though there are 6720 customers in the first stage (the same number for which the procedures had good coverage when $\epsilon = 0.1$). This demonstrates that the exact rate at which the batch size needs to grow with respect to ϵ depends on the problem being considered.

7 Conclusion

We have shown the asymptotic validity of our proposed two-stage stopping procedures. The main advantage of our method over fully sequential stopping procedures is that we do not require a strongly consistent estimate of the asymptotic variance constant, which may be difficult to obtain. We avoid this issue by using the method of standardized time series in the first stage of the procedure.

Batch Sizes	Customers in Stage 1	m	ϵ	Abs/Rel Precision	Coverage	Avg Total Batches	Avg Total Customers
Equal	6720	7	0.4	abs	82.4%	36.93	41362.7
Equal	6720	7	0.2	abs	81.5%	145.76	163251.2
Equal	6720	7	0.1	abs	79.0%	581.56	651342.7
Unequal	6720	7	0.4	abs	82.8%	36.47	40847.7
Unequal	6720	7	0.2	abs	81.2%	145.28	162709.2
Unequal	6720	7	0.1	abs	78.8%	581.07	650794.5
Equal	6720	7	0.1	rel	83.3%	32.30	36174.9
Equal	6720	7	0.05	rel	82.1%	127.40	142686.9
Equal	6720	7	0.025	rel	80.0%	508.07	569042.9
Unequal	6720	7	0.1	rel	83.2%	31.82	35637.7
Unequal	6720	7	0.05	rel	81.6%	126.90	142126.2
Unequal	6720	7	0.025	rel	80.0%	507.57	568482.4

Table 4: Coverage results using two-stage standardized sum stopping procedure from 1000 independent replications of estimating the expected waiting time in an M/M/1 queue with $\lambda = .8$, $\mu = 1$.

One theoretical problem still needs to be resolved. As stated in Section 3.2, property (vi) of Assumption A2 probably follows from properties (i)–(v), but we have not been able to show this. We should note though that property (vi) is satisfied in all of the applications that we have in mind, and so it does not seem to be a restrictive assumption in practice.

Some practical issues are still open for further investigation. Most importantly, the question of how to determine an appropriate batch size to correspond to a desired half-width in general simulations is unanswered. It is probably the case that there is not a single formula that can be applied in all contexts. However, in queueing simulations, we have given some suggestions on how to deal with this problem (see Section 5). In other settings, more experimentation should be carried out in order to determine the sensitivity of the coverage with respect to these two parameters. This may lead to the development of different heuristics that can be applied in the various situations.[1]

Appendix

Here we will provide the proofs for Theorems 1 and 3 from Section 3.2. In order to prove our results, we will apply to the accumulated process X_ϵ a function which is "suitably well-behaved" when evaluated at the limiting process B. More precisely, consider a function $h : C[0,\infty) \to \Re$, and let $D(h)$ be the set of elements $x \in C[0,\infty)$ at which the function h is discontinuous. Then the exact description of "suitably well-behaved" is given in the following proposition, known as the

[1]This paper is based directly on material from the author's Ph.D. dissertation which was written while he was in the Operations Research Department at Stanford University. The research there was supported by IBM contract 12480042 and NSF grant ECS 86-17905. The author would like to thank his dissertation advisor Peter W. Glynn for his guidance and the department editor, the associate editor, and the referees for their helpful comments.

continuous mapping principle.

Proposition 2 *Suppose $X_\epsilon, X \in C[0,\infty)$ are random elements such that $X_\epsilon \Rightarrow X$ as $\epsilon \to 0$. If $P\{X \in D(h)\} = 0$, then $h(X_\epsilon) \Rightarrow h(X)$ as $\epsilon \to 0$.*

See Ethier and Kurtz (1986), Billingsley (1968), or Glynn (1989) for a proof of this result.

Because of the importance of Proposition 2, we now describe how it can be applied to develop our results. It turns out that each formula we use to determine the final number of batches or to estimate μ boils down to a (measurable) function of the form $h(X_\epsilon)$. ¿From Assumption A1, X_ϵ converges weakly to a standard Brownian motion process B, which we recall only takes on sample paths in $C[0,\infty)$. The continuous mapping principle will therefore ensure that every limiting random variable of interest will have the form $h(B)$, i.e., $h(X_\epsilon) \Rightarrow h(B)$, provided that $P\{B \in D(h)\} = 0$. In general, it is difficult to establish that $P\{B \in D(h)\} = 0$ by working directly with the definition of continuity and the Skorohod metric. However, we will be able to avoid doing this by appealing to Assumption A2.

We now present two preparatory lemmas that will be useful for proving Theorems 1 and 3.

Lemma 1 *Let $r : \Re \to \Re$ be a measurable function such that $r(\alpha) \geq 1$ for all $\alpha \in \Re$. Also, define $R = r(\sigma g(B))$, where g satisfies Assumption A2. Then, given $g(B)$, the random variable $B(R)$ is conditionally normally distributed with mean 0 and variance R.*

Proof. First, we will show that $B(R) - B(1)$ is independent of $B(1)$. Define

$$z_A(x) = P\{B(R) - B(1) \in A \mid g(B) = x\},$$

where A is some (measurable) subset of $\Re$. Note $R \geq 1$ and R is known when given $g(B)$. ¿From Proposition 2.8 of Glynn and Iglehart (1990), $B(1)$ is independent of $g(B)$. Thus, using Assumption A2(i) and the fact that increments of Brownian motion are independent and normally distributed with mean 0 and variance equal to the length of the increment, $z_A(x)$ is the probability that a normal random variable with mean 0 and variance $R-1$ is in the set A. Hence,

$$\begin{aligned} & P\{B(R) - B(1) \in A \mid B(1)\} \\ &= E\left[P\{B(R) - B(1) \in A \mid B(1),\ \sigma g(B)\} \mid B(1)\right] \\ &= E[z_A(\sigma g(B)) \mid B(1)] \end{aligned}$$

by the independent increments property of Brownian motion. Since $B(1)$ is independent of $g(B)$, $E[z_A(\sigma g(B)) \mid B(1)] = E[z_A(\sigma g(B))]$, and so $B(R) - B(1)$ is independent of $B(1)$. Again using the fact that $B(1)$ and $g(B)$ are independent leads us to conclude that given $g(B)$, $B(R) - B(1)$ and $B(1)$ are conditionally independent. Therefore, since $B(1)$ is normally distributed with mean 0 and variance 1, given $g(B)$, $B(R)$ is conditionally normally distributed with mean 0 and variance R. ■

Lemma 2 *Let $r : \Re \to \Re$ be a measurable function such that $r(\alpha) \geq 1$ for all $\alpha \in \Re$. Also, let $R = r(g(\sigma B))$, where g satisfies Assumption A2. Then, $B(R)/(g(B)\sqrt{R})$ has the same distribution as $B(1)/g(B)$.*

Proof. Letting $G(x) = P\{g(B) \leq x\}$ and letting $N(a, b)$ denote a normally distributed random variable with mean a and variance b, we have

$$\begin{aligned} P\left\{\frac{B(R)}{g(B)\sqrt{R}} \leq x\right\} &= \int_0^\infty P\left\{\frac{B(R)}{\sqrt{R}} \leq xg(B) \;\middle|\; g(B) = y\right\} G(dy) \\ &= \int_0^\infty P\left\{\frac{N(0,R)}{\sqrt{R}} \leq xg(B) \;\middle|\; g(B) = y\right\} G(dy) \end{aligned}$$

by Lemma 1. Hence,

$$P\left\{\frac{B(R)}{g(B)\sqrt{R}} \leq x\right\} = \int_0^\infty P\left\{N(0,1) \leq xg(B) \mid g(B) = y\right\} G(dy) = P\left\{\frac{B(1)}{g(B)} \leq x\right\},$$

which proves the result. ∎

We are now in a position to prove the theorems of Section 3.2.

Proof of Theorem 1. First, we prove part (i). Define the function $h_a : C[0, \infty) \to \Re$ as

$$h_a(x) = \frac{1}{m} \max\left\{m, \left\lceil mg^2(x)a_\delta^2 \right\rceil\right\}. \tag{11}$$

Note that $h_a(X_\epsilon) = N_a(\epsilon)/m$ and $D(h_a) = D(g) \cup \{x : mg^2(x)a_\delta^2 \in \{m, m+1, m+2, \ldots\}\}$. Assumption A2(v) implies $P\{\sigma B \in D(g)\} = 0$ since σ is finite. Also, using Assumptions A2(ii) and A2(vi), we have

$$\begin{aligned} &P\{mg^2(\sigma B)a_\delta^2 \in \{m, m+1, m+2, \ldots\}\} \\ &\quad = P\left\{g(B) \in \left\{\frac{m^{1/2}}{m^{1/2}\sigma a_\delta}, \frac{(m+1)^{1/2}}{m^{1/2}\sigma a_\delta}, \frac{(m+2)^{1/2}}{m^{1/2}\sigma a_\delta}, \ldots\right\}\right\} \\ &\quad = 0. \end{aligned}$$

Thus, $P\{\sigma B \in D(h_a)\} = 0$, and so $h_a(X_\epsilon) \Rightarrow h_a(\sigma B)$ as $\epsilon \to 0$ by the continuous mapping principle. Hence, we have shown that $N_a(\epsilon) \Rightarrow N_a$ as $\epsilon \to 0$, where $N_a = mh_a(\sigma B)$. Now we have to show that $P\{N_a < \infty\} = 1$. Since $0 < \sigma < \infty$, $\{g(\sigma B) < \infty\} = \{g(B) < \infty\}$ by Assumption A2(ii). Assumption A2(v) states that g is continuous at B with probability 1 so $P\{g(B) < \infty\} = 1$ (since a continuous function must be finite). Therefore, $P\{h_a(\sigma B) < \infty\} = 1$ and N_a is a proper random variable.

Now we prove part (ii). First, define the function $u_a : C[0, \infty) \to \Re$ as

$$u_a(x) = \frac{x(h_a(x))}{g(x)\sqrt{h_a(x)}}. \tag{12}$$

Since x is assumed to be continuous, the numerator in (12) is discontinuous at the discontinuity points of h_a. Hence, $D(u_a) \subset D(h_a) \cup D(g) \cup A_1$, where $A_1 = \{x : g(x)\sqrt{h_a(x)} = 0\}$. ¿From the

proof of part (i) we have $P\{\sigma B \in D(h_a)\} = 0$ and $P\{\sigma B \in D(g)\} = 0$. Also, using (11), we have $h_a(x) \geq 1$ for all x, and $P\{g(\sigma B) = 0\} = 0$ from Assumption A2(iv), and so $P\{\sigma B \in A_1\} = 0$. Therefore, $P\{\sigma B \in D(u_a)\} = 0$, and so $u_a(X_\epsilon) \Rightarrow u_a(\sigma B)$ as $\epsilon \to 0$ by the continuous mapping principle.

By Assumption A2(ii), we have

$$u_a(\sigma B) = \frac{\sigma B(h_a(\sigma B))}{g(\sigma B)\sqrt{h_a(\sigma B)}} = \frac{B(h_a(\sigma B))}{g(B)\sqrt{h_a(\sigma B)}}.$$

Hence, $u_a(\sigma B)$ has the same distribution as $B(1)/g(B)$ by Lemma 2. We can write

$$u_a(X_\epsilon) = \frac{\sum_{i=1}^{N_a(\epsilon)}(Z_i(\epsilon) - \mu)}{s(\epsilon)\sqrt{N_a(\epsilon)}}.$$

¿From (6), we have $\epsilon \geq s(\epsilon)a_\delta/\sqrt{N_a(\epsilon)}$, and so

$$\begin{aligned} P\{\mu \in I_a(\epsilon)\} &= P\left\{\frac{1}{N_a(\epsilon)}\sum_{i=1}^{N_a(\epsilon)} Z_i(\epsilon) - \epsilon \leq \mu \leq \frac{1}{N_a(\epsilon)}\sum_{i=1}^{N_a(\epsilon)} Z_i(\epsilon) + \epsilon\right\} \\ &\geq P\left\{\frac{1}{N_a(\epsilon)}\sum_{i=1}^{N_a(\epsilon)} Z_i(\epsilon) - \frac{a_\delta s(\epsilon)}{\sqrt{N_a(\epsilon)}} \leq \mu \leq \frac{1}{N_a(\epsilon)}\sum_{i=1}^{N_a(\epsilon)} Z_i(\epsilon) + \frac{a_\delta s(\epsilon)}{\sqrt{N_a(\epsilon)}}\right\} \\ &= P\left\{-a_\delta \leq u_a(X_\epsilon) \leq a_\delta\right\}. \end{aligned}$$

Thus, since the random variable $B(1)/g(B)$ has a continuous distribution function (see Glynn and Iglehart 1990) and $u_a(X_\epsilon) \Rightarrow u_a(\sigma B)$ as $\epsilon \to 0$,

$$P\{-a_\delta \leq u_a(X_\epsilon) \leq a_\delta\} \to P\{-a_\delta \leq u_a(\sigma B) \leq a_\delta\} = 1 - \delta,$$

as $\epsilon \to 0$, by Theorem 2.1 of Billingsley (1968). The last equality follows from the fact that $u_a(\sigma B)$ has the same distribution as $B(1)/g(B)$ by Lemma 2. Hence, we have our result. ∎

Proof of Theorem 3. First, we prove part (i). Define the function $k \in C[0,\infty)$ as $k(t) = t$. By Assumption A1, $\bar{Y}_\epsilon - \mu k = \epsilon X_\epsilon \Rightarrow 0 \cdot \sigma B = 0$ as $\epsilon \to 0$, and so $\bar{Y}_\epsilon \Rightarrow \mu k$ as $\epsilon \to 0$. Since μk is deterministic, $(X_\epsilon, \bar{Y}_\epsilon) \Rightarrow (\sigma B, \mu k)$ as $\epsilon \to 0$ by Theorem 4.4 of Billingsley (1968).

Now define the function $h_r : C[0,\infty) \times C[0,\infty) \to \Re$ as

$$h_r(x,y) = \frac{1}{m}\max\left\{m, \left\lceil \frac{mg^2(x)a_\delta^2}{y^2(1)} \right\rceil\right\}.$$

Note that $h_r(X_\epsilon, \bar{Y}_\epsilon) = N_r(\epsilon)/m$. Also, since we are working in $C[0,\infty) \times C[0,\infty)$, y is continuous at 1 by assumption. Hence, $D(h_r) \subset A_1 \cup A_2 \cup A_3$, where

$$\begin{aligned} A_1 &= \{(x,y) : x \in D(g)\} \\ A_2 &= \{(x,y) : y(1) = 0\} \\ A_3 &= \left\{(x,y) : \frac{mg^2(x)a_\delta^2}{y^2(1)} \in \{m, m+1, m+1, \ldots\}\right\}. \end{aligned}$$

Assumption A2(v) implies $P\{\sigma B \in D(g)\} = 0$, and so $P\{(\sigma B, \mu k) \in A_1\} = 0$. Since $\mu \neq 0$, $P\{(\sigma B, \mu k) \in A_2\} = 0$. Also, using Assumptions A2(ii) and A2(vi), we have that

$$\begin{aligned} & P\left\{ \frac{mg^2(\sigma B)a_\delta^2}{\mu^2} \in \{m, m+1, m+1, \ldots\} \right\} \\ = \; & P\left\{ g(B) \in \left\{ \frac{|\mu| m^{1/2}}{m^{1/2}\sigma a_\delta}, \frac{|\mu|(m+1)^{1/2}}{m^{1/2}\sigma a_\delta}, \frac{|\mu|(m+2)^{1/2}}{m^{1/2}\sigma a_\delta}, \ldots \right\} \right\} \\ = \; & 0. \end{aligned}$$

Hence, $P\{(\sigma B, \mu k) \in D(h_r)\} = 0$, and so $h_r(X_\epsilon, \bar{Y}_\epsilon) \Rightarrow h_a(\sigma B, \mu k)$ as $\epsilon \to 0$ by the continuous mapping principle. Hence, we have shown that $N_r(\epsilon) \Rightarrow N_r$ as $\epsilon \to 0$, where $N_r = m h_r(\sigma B, \mu k)$. Also, we can write

$$N_r = \max\left\{ m, \left\lceil \frac{mg^2(\sigma B)a_\delta^2}{\mu^2} \right\rceil \right\}. \tag{13}$$

Since $P\{g \in D(g)\} = 0$ by Assumption A2(v), we have $P\{g(\sigma B) < \infty\} = 1$ since $\sigma < \infty$. Furthermore, $\mu \neq 0$ implies $P\{N_r < \infty\} = 1$, thus showing N_r is a proper random variable.

Now we will prove part (ii). First, define the function $w_r : C[0,\infty) \times C[0,\infty) \to \Re$ as

$$w_r(x, y) = \frac{x(h_r(x, y))}{|y(h_r(x, y))|}.$$

Note that $D(w_r) \subset D(h_r) \cup A_4$, where $A_4 = \{(x, y) : y(h_r(x, y)) = 0\}$. We showed $P\{(\sigma B, \mu k) \in D(h_r)\} = 0$ in the proof of part (i). Finally, since $\mu \neq 0$ and since $h_r(x, y) \geq 1$ for all (x, y) by (13), we have $P\{(\sigma B, \mu k) \in A_4\} = 0$. Thus, $P\{(\sigma B, \mu k) \in D(w_r)\} = 0$, and so

$$w_r(X_\epsilon, \bar{Y}_\epsilon) \Rightarrow w_r(\sigma B, \mu k) = \frac{\sigma B(N_r/m)}{|\mu| N_r/m}$$

as $\epsilon \to 0$ by the continuous mapping principle.

Now we will show that $w_r(\sigma B, \mu k)$ has a continuous distribution function H. Let x be some real-valued constant. Then

$$\begin{aligned} H(x) \; \equiv \; & P\{w_r(\sigma B, \mu k) \leq x\} \\ = \; & P\left\{ \frac{\sigma B(h_r(\sigma B, \mu k))}{|\mu| h_r(\sigma B, \mu k)} \leq x \right\} \\ = \; & \sum_{k=m}^{\infty} P\left\{ \frac{\sigma B(h_r(\sigma B, \mu k))}{|\mu| h_r(\sigma B, \mu k)} \leq x \;\middle|\; h_r(\sigma B, \mu k) = \frac{k}{m} \right\} P\left\{ h_r(\sigma B, \mu k) = \frac{k}{m} \right\} \\ = \; & \sum_{k=m}^{\infty} P\left\{ \frac{\sigma B(N_r/m)}{|\mu| N_r/m} \leq x \;\middle|\; N_r = k \right\} P\left\{ N_r = k \right\}. \end{aligned}$$

¿From Lemma 1, given $g(\sigma B)$, $\sigma B(N_r/m)$ is conditionally normally distributed with mean 0 and variance $N_r\sigma^2/n$. Therefore, letting $N(a, b)$ denote a normal random variable with mean a and variance b, we have

$$\begin{aligned} H(x) \; = \; & \sum_{k=m}^{\infty} P\left\{ \frac{N(0, \sigma^2 N_r/m)}{|\mu| N_r/m} \leq x \;\middle|\; N_r = k \right\} P\left\{ N_r = k \right\} \\ = \; & \sum_{k=m}^{\infty} P\left\{ \frac{N(0,1)}{N_r^{1/2}} \leq \frac{|\mu| x}{m^{1/2}\sigma} \;\middle|\; N_r = k \right\} P\left\{ N_r = k \right\}. \end{aligned}$$

Note that $N(0,1)$ is independent of N_r since $B(1)$ is independent of $g(B)$. Thus, the continuity of the distribution of $N(0,1)$ and the bounded convergence theorem imply that H is continuous everywhere in x.

Now note that

$$w_r(X_\epsilon, \bar{Y}_\epsilon) = \frac{\sum_{i=1}^{N_r(\epsilon)} (Z_i(\epsilon) - \mu)}{\epsilon \left| \sum_{i=1}^{N_r(\epsilon)} Z_i(\epsilon) \right|}.$$

Recalling our definition of $\hat{\mu}_r(\epsilon)$ given in (4), we have

$$\begin{aligned}
P\{\mu \in I_r(\epsilon)\} &= P\left\{\hat{\mu}_r(\epsilon) - \epsilon|\hat{\mu}_r(\epsilon)| \le \mu \le \hat{\mu}_r(\epsilon) + \epsilon|\hat{\mu}_r(\epsilon)|\right\} \\
&= P\left\{-1 \le \frac{\sum_{i=1}^{N_r(\epsilon)} (Z_i(\epsilon) - \mu)}{\epsilon \left| \sum_{i=1}^{N_r(\epsilon)} Z_i(\epsilon) \right|} \le 1\right\} \\
&= P\left\{-1 \le w_r(X_\epsilon, \bar{Y}_\epsilon) \le 1\right\} \\
&\to P\left\{-1 \le w_r(\sigma B, \mu k) \le 1\right\},
\end{aligned}$$

as $\epsilon \to 0$, since $w_r(X_\epsilon, \bar{Y}_\epsilon) \Rightarrow w_r(\sigma B, \mu k)$ as $\epsilon \to 0$ and $w_r(\sigma B, \mu k)$ has a continuous distribution function; see Theorem 2.1 of Billingsley (1968). Also, from (13), $|\mu| \ge g(\sigma B) a_\delta / \sqrt{N_r}$, and so

$$\begin{aligned}
P\left\{-1 \le w_r(\sigma B, \mu k) \le 1\right\} &= P\left\{-|\mu| \le \frac{\sigma B(N_r/m)}{N_r/m} \le |\mu|\right\} \\
&\ge P\left\{-\frac{g(\sigma B) a_\delta}{\sqrt{N_r/m}} \le \frac{\sigma B(N_r/m)}{N_r/m} \le \frac{g(\sigma B) a_\delta}{\sqrt{N_r/m}}\right\} \\
&= P\left\{-a_\delta \le \frac{\sigma B(N_r/m)}{g(\sigma B)\sqrt{N_r/m}} \le a_\delta\right\} \\
&= P\left\{-a_\delta \le u_r(\sigma B, \mu k) \le a_\delta\right\},
\end{aligned}$$

where we define the function $u_r : C[0,\infty) \times C[0,\infty) \to \Re$ as

$$u_r(x,y) = \frac{x(h_r(x,y))}{g(x)\sqrt{h_r(x,y)}}.$$

By Lemma 2, $u_r(\sigma B, \mu k)$ has the same distribution as $B(1)/g(B)$ so $P\left\{-a_\delta \le u_r(\sigma B, \mu k) \le a_\delta\right\} = 1 - \delta$, which proves part (ii). ∎

References

Adam, N., "Achieving a Confidence Interval for Parameters Estimated by Simulation," *Management Sci.*, 27 (1983), 856–866.

Billingsley, P., *Convergence of Probability Measures*, John Wiley, New York, 1968.

Bratley, P., B. L. Fox, and L. E. Schrage, *A Guide to Simulation*, Second edition, Springer-Verlag, New York, 1987.

Chow, Y. S. and H. Robbins, "On the Asymptotic Theory of Fixed-Width Sequential Confidence Intervals for the Mean," *Ann. Math. Statist.*, 36 (1965), 457–462.

Damerdji, H., "Strong Consistency of the Variance Estimator in Steady-State Simulation Output Analysis," *Math. Oper. Res*, to appear, (1989).

Ethier, S. N. and T. C. Kurtz, *Markov Processes: Characterization and Convergence*, John Wiley, New York, 1986.

Fishman, G. S., "Achieving Specific Accuracy in Simulation Output Analysis," *Commun. ACM*, 20 (1977), 310–315.

Glynn, P. W., "Diffusion Approximations," Chapter 4 of *Handbooks in Operations Research and Management Science, Vol. 2, Stochastic Models* (edited by D. Heyman and M. Sobel), Elsevier Science Publishers B.V. (North-Holland), 1990.

Glynn, P. W. and D. L. Iglehart, "Simulation Output Analysis Using Standardized Time Series," *Math. Oper. Res.*, 15 (1990), 1–16.

Glynn, P. W. and W. Whitt, "Sufficient Conditions for Functional-Limit-Theorem Versions of $L = \lambda W$," *Queueing Systems*, 1 (1987), 279–287.

Glynn, P. W. and W. Whitt, "Estimating the Asymptotic Variance with Batch Means," *Oper. Res. Letters*, 10 (1991), 431–435.

Glynn, P. W. and W. Whitt, "The Asymptotic Validity of Sequential Stopping Rules for Stochastic Simulations," *Ann. Applied Probability*, 2 (1992), 180–198.

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

Heidelberger, P. and P. D. Welch, "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations," *Commun. ACM*, 24 (1981a), 233–245.

Heidelberger, P. and P. D. Welch, "Adaptive Spectral Methods for Simulation Output Analysis," *IBM J. Res. Develop.*, 25 (1981b), 860–876.

Heidelberger, P. and P. D. Welch, "Simulation Run Length Control in the Presence of an Initial Transient," *Oper. Res.*, 31 (1983), 1109–1144.

Iglehart, D. L., "Simulating Stable Stochastic Systems, VII: Selecting the Best System," *TIMS Studies in Management Sci.*, 7 (1977), 37–49.

Lavenberg, S. S. and C. H. Sauer, "Sequential Stopping Rules for the Regenerative Method of Simulation," *IBM J. Res. Develop.*, 21 (1977), 545–558.

Law, A. M., "Efficient Estimators for Simulated Queueing Systems," *Management Sci.*, 22 (1975), 30–41.

Law, A. M. and J. S. Carson, "A Sequential Procedure for Determining the Length of a Steady-State Simulation," *Oper. Res.*, 27 (1979), 1011–1025.

Law, A. M. and W. D. Kelton, "Confidence Intervals for Steady-State Simulation II: A Survey of Sequential Procedures," *Management Sci.*, 28 (1982), 550–562.

Mechanic, H. and W. McKay, "Confidence Intervals for Averages of Dependent Data in Simulations II," IBM Advanced Systems, Development Division. Technical Report 17-202, 1966.

Moshman, J., "A Method for Selecting the Size of the Initial Sample in Stein's Two Sample Procedure," *Ann. Math. Stat.*, 29 (1958), 1271–1275.

Newman, C. M. and A. L. Wright, "An Invariance Principle for Certain Dependent Sequences," *Ann. Probability*, 9 (1981), 671–675.

Nakayama, M. K., "Algorithms for Two-Stage Stopping Procedures Based on Standardized Time Series," IBM Research Report, Yorktown Heights, forthcoming, (1992).

Sargent, R. G., K. Kang, and D. Goldsman, "An Investigation of Finite Sample Behavior of Confidence Interval Estimation Procedures," *Oper. Res.*, 40 (1992), 898–913.

Schmeiser, B., "Batch Size Effects in the Analysis of Simulation Output," *Oper. Res.*, 30 (1982), 556–568.

Schruben, L. W., "Detecting Initialization Bias in Simulation Output," *Oper. Res.*, 30 (1982), 569–590.

Schruben, L. W., "Confidence Interval Estimation Using Standardized Time Series," *Oper. Res.*, 31 (1983), 1090–1108.

Seelbinder, B. M., "On Stein's Two-Stage Sampling Scheme," *Ann. Math. Stat.*, 24 (1953), 640–649.

Stein, C., "A Two-Sample Test for a Linear Hypothesis Whose Power is Independent of the Variance," *Ann. Math. Stat.*, 16 (1945), 243–258.

Whitt, W., "Planning Queueing Simulations," *Management Sci.*, 35 (1989a), 1341–1366.

Whitt, W., "Simulation Run Length Planning," *Proceedings of the 1989 Winter Simulation Conference*, E. A. MacNair, K. J. Musselman, P. Heidelberger (Eds.), 1989b, 106–112.