CHAPTER 9

OUTPUT DATA ANALYSIS FOR A SINGLE SYSTEM

Recommended sections for a first reading: 9.1 through 9.3, 9.4.1, 9.4.3, 9.5.1, 9.5.2, 9.8

9.1 INTRODUCTION

In many simulation studies a great deal of time and money is spent on model development and programming, but little effort is made to analyze the simulation output data appropriately. As a matter of fact, a very common mode of operation is to make a single simulation run of somewhat arbitrary length and then to treat the resulting simulation estimates as the "true" model characteristics. Since random samples from probability distributions are typically used to drive a simulation model through time, these estimates are just particular realizations of random variables that may have large variances. As a result, these estimates could, in a particular simulation run, differ greatly from the corresponding true characteristics for the model. The net effect is, of course, that there could be a significant probability of making erroneous inferences about the system under study.

Historically, there are several reasons why output data analyses have not been conducted in an appropriate manner. First, users often have the unfortunate impression that simulation is just an exercise in computer programming,

albeit a complicated one. Consequently, many simulation "studies" begin with heuristic model building and coding, and end with a single run of the program to produce "the answers." In fact, however, a simulation is a computer-based statistical sampling experiment. Thus, if the results of a simulation study are to have any meaning, appropriate statistical techniques must be used to design and analyze the simulation experiments. A second reason for inadequate statistical analyses is that the output processes of virtually all simulations are nonstationary and autocorrelated (see Sec. 5.5.3). Thus, classical statistical techniques based on IID observations are not directly applicable. At present, there are still several output-analysis problems for which there is no completely accepted solution, and the methods that are available are often complicated to apply. Another impediment to obtaining precise estimates of a model's true parameters or characteristics is the cost of the computer time needed to collect the necessary amount of simulation output data. Indeed, there are situations where an appropriate statistical procedure is available, but the cost of collecting the quantity of data dictated by the procedure is prohibitive. This latter difficulty is becoming less severe since many analysts now have their own high-speed microcomputers or engineering work stations. These computers are relatively inexpensive to buy and can be run overnight or on weekends to produce large amounts of simulation output data, at essentially zero marginal cost.

We now describe more precisely the random nature of simulation output. Let $Y_1, Y_2, \ldots$ be an output stochastic process (see Sec. 4.3) from a *single* simulation run. For example, Y_i might be the throughput (production) in the ith hour for a manufacturing system. The Y_i's are random variables that will, in general, be neither independent nor identically distributed. Thus, most of the formulas of Chap. 4, which assume independence [e.g., the confidence interval given by (4.12)], do not apply *directly*.

Let $y_{11}, y_{12}, \ldots, y_{1m}$ be a realization of the random variables $Y_1, Y_2, \ldots, Y_m$ resulting from making a simulation run of length m observations using the random numbers $u_{11}, u_{12}, \ldots$. (The ith random number used in the jth run is denoted u_{ji}.) If we run the simulation with a different set of random numbers $u_{21}, u_{22}, \ldots$, then we will obtain a different realization $y_{21}, y_{22}, \ldots, y_{2m}$ of the random variables $Y_1, Y_2, \ldots, Y_m$. (The two realizations are not the same since the different random numbers used in the two runs produce different samples from the input probability distributions.) In general, suppose that we make n independent replications (runs) of the simulation (i.e., different random numbers are used for each replication, the statistical counters are reset at the beginning of each replication, and each replication uses the same initial conditions; see Sec. 9.4.3) of length m, resulting in the observations:

$$\begin{array}{ccccc} y_{11}, & \ldots, & y_{1i}, & \ldots, & y_{1m} \\ y_{21}, & \ldots, & y_{2i}, & \ldots, & y_{2m} \\ \vdots & & \vdots & & \vdots \\ y_{n1}, & \ldots, & y_{ni}, & \ldots, & y_{nm} \end{array}$$

The observations from a particular replication (row) are clearly not IID. However, note that $y_{1i}, y_{2i}, \ldots, y_{ni}$ (from the ith column) are IID observations of the random variable Y_i, for $i = 1, 2, \ldots, m$. This *independence across runs* (see Prob. 9.1) is the key to the relatively simple output-data-analysis methods described in later sections of this chapter. Then, roughly speaking, the goal of output analysis is to use the observations y_{ji} $(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$ to draw inferences about the (distributions of the) random variables $Y_1, Y_2, \ldots, Y_m$. For example, $\bar{y}_i(n) = \sum_{j=1}^{n} y_{ji}/n$ is an unbiased estimate of $E(Y_i)$.

Example 9.1. Consider a bank with five tellers and one queue, which opens its doors at 9 A.M., closes its doors at 5 P.M., but stays open until all customers in the bank at 5 P.M. have been served. Assume that customers arrive in accordance with a Poisson process at rate 1 per minute (i.e., IID exponential interarrival times with mean 1 minute), that service times are IID exponential random variables with mean 4 minutes, and that customers are served in a FIFO manner. Table 9.1 shows several typical output statistics from 10 independent replications of a simulation of the bank, assuming that no customers are present initially. Note that results from various replications can be quite different. Thus, one run clearly does not produce "the answers."

Our goal in this chapter is to discuss methods for statistical analysis of simulation output data and to present the material with a practical focus that should be accessible to a reader having a basic understanding of probability and statistics. (Reviewing Chap. 4 might be advisable before reading this chapter.) We will discuss what we believe are all the important methods for output analysis; however, the emphasis will be on statistical procedures that are relatively easy to understand and implement, have been shown to perform well in practice, and have applicability to real-world problems.

TABLE 9.1
Results for 10 independent replications of the bank model

Replication	Number served	Finish time (hours)	Average delay in queue (minutes)	Average queue length	Proportion of customers delayed < 5 minutes
1	484	8.12	1.53	1.52	0.917
2	475	8.14	1.66	1.62	0.916
3	484	8.19	1.24	1.23	0.952
4	483	8.03	2.34	2.34	0.822
5	455	8.03	2.00	1.89	0.840
6	461	8.32	1.69	1.56	0.866
7	451	8.09	2.69	2.50	0.783
8	486	8.19	2.86	2.83	0.782
9	502	8.15	1.70	1.74	0.873
10	475	8.24	2.60	2.50	0.779

In Secs. 9.2 and 9.3 we discuss types of simulations with regard to output analysis, and also measures of performance or parameters θ for each type. Sections 9.4 through 9.6 show how to get a point estimator $\hat{\theta}$ and confidence interval for each type of parameter θ, with the confidence interval typically requiring an estimate of the variance of $\hat{\theta}$, namely, $\widehat{\mathrm{Var}}(\hat{\theta})$. Each of the analysis methods discussed may suffer from one or both of the following problems:

1. $\hat{\theta}$ is not an unbiased estimator of θ, that is, $E(\hat{\theta}) \neq \theta$; see, for example, Sec. 9.5.2
2. $\widehat{\mathrm{Var}}(\hat{\theta})$ is not an unbiased estimator of $\mathrm{Var}(\hat{\theta})$; see, for example, Sec. 9.5.3

Section 9.7 extends the above analyses to confidence-interval construction for several different parameters simultaneously. Finally, in Sec. 9.8 we show how time plots of important variables may provide insight into a system's dynamic behavior.

We will not attempt to give every reference on the subject of output data analysis, since a very comprehensive set of references was given in the survey paper by Law (1983). Also see the book chapter by Welch (1983).

9.2 TRANSIENT AND STEADY-STATE BEHAVIOR OF A STOCHASTIC PROCESS

Consider the output stochastic process $Y_1, Y_2, \ldots$. Let $F_i(y|I) = P(Y_i \le y|I)$ for $i = 1, 2, \ldots$, where y is a real number and I represents the initial conditions used to start the simulation at time 0. [The conditional probability $P(Y_i \le y|I)$ is the probability that the event $\{Y_i \le y\}$ occurs *given* the initial conditions I.] For a manufacturing system, I might specify the number of jobs present, and whether each machine is busy or idle, at time 0. We call $F_i(y|I)$ the *transient distribution* of the output process at (discrete) time i for initial conditions I. Note that $F_i(y|I)$ will, in general, be different for each value of i and each set of initial conditions I. The density functions for the transient distributions corresponding to the random variables Y_{i_1}, Y_{i_2}, Y_{i_3}, and Y_{i_4} are shown in Fig. 9.1 for a particular set of initial conditions I and increasing time indices i_1, i_2, i_3, and i_4, where it is assumed that the random variable Y_{i_j} has density function $f_{Y_{i_j}}$. The density $f_{Y_{i_j}}$ specifies how the random variable Y_{i_j} can vary from one replication to another.

For fixed y and I, the probabilities $F_1(y|I), F_2(y|I), \ldots$ are just a sequence of numbers. *If* $F_i(y|I) \to F(y)$ as $i \to \infty$ for all y and for any initial conditions I, then $F(y)$ is called the *steady-state distribution* of the output process $Y_1, Y_2, \ldots$. Strictly speaking, the steady-state distribution $F(y)$ is only obtained in the limit as $i \to \infty$. In practice, however, there will often be a finite time index, say, $k + 1$, such that the distributions from this point on will be approximately the same as each other; "steady state" is figuratively said to start at time $k + 1$ as shown in Fig. 9.1. Note that steady state does *not* mean conditions in the bank at noon. This approach can be carried out in SIMLIB (see Chap. 2) by reinitializing the statistical counters for subroutines SAMPST, TIMEST, and FILEST (see Prob. 2.7) at noon.

An alternative approach is to collect data on the number of customers present in the bank at noon for several different days. Let $\hat{p}_i$ be the proportion of these days that i customers $(i = 0, 1, \ldots)$ are present at noon. Then we simulate the bank from noon to 1 P.M. with the number of customers present at noon being randomly chosen from the distribution $\{\hat{p}_i\}$. (All customers who are being served at noon might be assumed to be just beginning their services. Starting all services fresh at noon results in an approximation to the actual situation in the bank, since the customers who are in the process of being served at noon would have partially completed their services. However, the effect of this approximation should be negligible for a simulation of length 1 hour.)

If more than one simulation run from noon to 1 P.M. is desired, then a different sample from $\{\hat{p}_i\}$ is drawn for each run. The X_j's that result from these runs are still IID, since the initial conditions for each run are chosen independently from the same distribution.

9.5 STATISTICAL ANALYSIS FOR STEADY-STATE PARAMETERS

Let $Y_1, Y_2, \ldots$ be an output stochastic process from a single run of a nonterminating simulation. Suppose that $P(Y_i \le y) = F_i(y) \to F(y) = P(Y \le y)$ as $i \to \infty$, where Y is the steady-state random variable of interest with distribution function F. (We have suppressed in our notation the dependence of F_i on the initial conditions I.) Then ϕ is a steady-state parameter if it is a characteristic of Y such as $E(Y)$, $P(Y \le y)$, or a quantile of Y. One difficulty in estimating ϕ is that the distribution function of Y_i (for $i = 1, 2, \ldots$) is different from F, since it will generally not be possible to choose I to be representative of "steady-state behavior." This causes an estimator of ϕ based on the observations $Y_1, Y_2, \ldots, Y_m$ not to be "representative." For example, the sample mean $\bar{Y}(m)$ will be a biased estimator of $\nu = E(Y)$ for all finite values of m. The problem we have just described is called the *problem of the initial transient* or the *startup problem* in the simulation literature.

Example 9.23. To illustrate the startup problem more succinctly, consider the process of delays $D_1, D_2, \ldots$ for the $M/M/1$ queue with $\rho < 1$ (see Example 9.2). From queueing theory, it is possible to show that

$$P(D_i \le y) \to P(D \le y) = (1-\rho) + \rho[1 - e^{-(\omega-\lambda)y}] \qquad \text{as } i \to \infty$$

If the number of customers s present at time 0 is 0, then $D_1 = 0$ and $E(D_i) \ne E(D) = d$ for any i. On the other hand, if s is chosen in accordance with the steady-state number in system distribution [see, for example, Gross and Harris (1985, p. 65)], then for all i, $P(D_i \le y) = P(D \le y)$ and $E(D_i) = d$ (see Prob. 9.11). Thus, there is no initial transient in this case.

In practice, the steady-state distribution will not be known exactly and the above initialization technique will not be possible. Techniques for dealing with the startup problem in practice are discussed in the next section.

9.5.1 The Problem of the Initial Transient

Suppose that we want to estimate the steady-state mean $\nu = E(Y)$, which is also generally defined by

$$\nu = \lim_{i \to \infty} E(Y_i)$$

Thus, the transient means converge to the steady-state mean. The most serious consequence of the problem of the initial transient is probably that $E[\bar{Y}(m)] \ne \nu$ for any m [see Law (1983, pp. 1010–1012) for further discussion]. The technique most often suggested for dealing with this problem is called *warming up the model* or *initial-data deletion*. The idea is to delete some number of observations from the beginning of a run and to use only the remaining observations to estimate ν. For example, given the observations $Y_1, Y_2, \ldots, Y_m$, it is often suggested to use

$$\bar{Y}(m,l) = \frac{\sum_{i=l+1}^{m} Y_i}{m-l}$$

$(1 \le l \le m-1)$ rather than $\bar{Y}(m)$ as an estimator of ν. In general, one would expect $\bar{Y}(m,l)$ to be less biased than $\bar{Y}(m)$, since the observations near the "beginning" of the simulation may not be very representative of steady-state behavior due to the choice of initial conditions. For example, this is true for the process $D_1, D_2, \ldots$ in the case of an $M/M/1$ queue with $s = 0$, since $E(D_i)$ increases monotonically to d as $i \to \infty$ (see Fig. 9.2).

The question naturally arises as to how to choose the *warmup period* (or deletion amount) l. We would like to pick l (and m) such that $E[\bar{Y}(m,l)] \approx \nu$. If l and m are chosen too small, then $E[\bar{Y}(m,l)]$ may be significantly different from ν. On the other hand, if l is chosen larger than necessary, then $\bar{Y}(m,l)$ will probably have an unnecessarily large variance. There have been a number of methods suggested in the literature for choosing l. However, Gafarian, Ancker, and Morisaku (1978) found that none of the methods available at that time performed well in practice. Kelton and Law (1983) developed an algorithm for choosing l (and m) that worked well {that is, $E[\bar{Y}(m,l)] \approx \nu$} for a wide variety of stochastic models. However, a theoretical limitation of the procedure is that it basically makes the assumption that $E(Y_i)$ is a monotone function of i.

The simplest and most general technique for determining l is a graphical procedure due to Welch (1981, 1983). Its specific goal is to determine a time index l such that $E(Y_i) \approx \nu$ for $i > l$, where l is the warmup period. [This is equivalent to determining when the transient mean curve $E(Y_i)$ (for $i =$

1, 2, . . .) "flattens out" at level ν; see Fig. 9.1.] In general, it is very difficult to determine l from a single replication due to the inherent variability of the process $Y_1, Y_2, \ldots$ (see Fig. 9.7, below). As a result, Welch's procedure is based on making n independent replications of the simulation and employing the following four steps:

1. Make n replications of the simulation ($n \geq 5$), each of length m (where m is large). Let Y_{ji} be the ith observation from the jth replication ($j = 1, 2, \ldots, n$; $i = 1, 2, \ldots, m$), as shown in Fig. 9.5.
2. Let $\bar{Y}_i = \sum_{j=1}^{n} Y_{ji}/n$ for $i = 1, 2, \ldots, m$ (see Fig. 9.5). The averaged process $\bar{Y}_1, \bar{Y}_2, \ldots$ has means $E(\bar{Y}_i) = E(Y_i)$ and variances $\mathrm{Var}(\bar{Y}_i) = \mathrm{Var}(Y_i)/n$ (see Prob. 9.12). Thus, the averaged process has the same transient mean curve as the original process, but its plot has only $(1/n)$th the variance.
3. To smooth out the high-frequency oscillations in $\bar{Y}_1, \bar{Y}_2, \ldots$ (but leave the low-frequency oscillations or long-run trend of interest), we further define the moving average $\bar{Y}_i(w)$ (where w is the *window* and is a positive integer such that $w \leq \lfloor m/2 \rfloor$) as follows:

$$\bar{Y}_i(w) = \begin{cases} \dfrac{\sum_{s=-w}^{w} \bar{Y}_{i+s}}{2w+1} & \text{if } i = w+1, \ldots, m-w \\[2ex] \dfrac{\sum_{s=-(i-1)}^{i-1} \bar{Y}_{i+s}}{2i-1} & \text{if } i = 1, \ldots, w \end{cases}$$

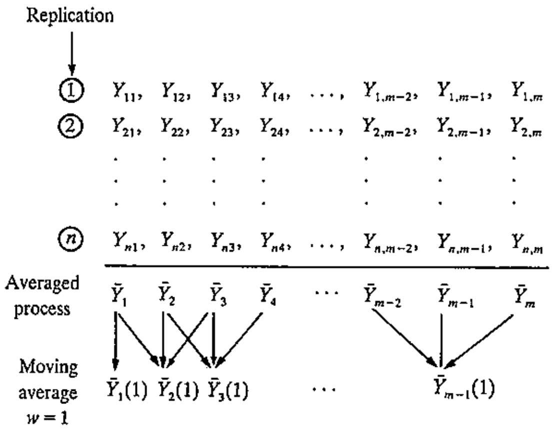

FIGURE 9.5
Averaged process and moving average with $w = 1$ based on n replications of length m.

Thus, if i is not too close to the beginning of the replications, then $\bar{Y}_i(w)$ is just the simple average of $2w+1$ observations of the averaged process centered at observation i (see Fig. 9.5). It is called a moving average since i moves through time.

4. Plot $\bar{Y}_i(w)$ for $i = 1, 2, \ldots, m - w$ and choose l to be that value of i beyond which $\bar{Y}_1(w), \bar{Y}_2(w), \ldots$ appears to have converged. See Welch (1983, p. 292) for an aid in determining convergence.

The following example illustrates the calculation of the moving average.

Example 9.24. For simplicity, assume that $m = 10$, $w = 2$, $\bar{Y}_i = i$ for $i = 1, 2, \ldots, 5$, and $\bar{Y}_i = 6$ for $i = 6, 7, \ldots, 10$. Then

$$\begin{array}{lll} \bar{Y}_1(2) = 1 & \bar{Y}_2(2) = 2 & \bar{Y}_3(2) = 3 \\ \bar{Y}_4(2) = 4 & \bar{Y}_5(2) = 4.8 & \bar{Y}_6(2) = 5.4 \\ \bar{Y}_7(2) = 5.8 & \bar{Y}_8(2) = 6 & \end{array}$$

Before giving examples of applying Welch's procedure to actual stochastic models, we make the following recommendations on choosing the parameters n, m, and w:

- Initially, make $n = 5$ or 10 replications (depending on model execution cost), with m as large as practical. In particular, m should be much larger than the anticipated value of l (see Sec. 9.5.2) and also large enough to allow infrequent events (e.g., machine breakdowns) to occur a reasonable number of times.
- Plot $\bar{Y}_i(w)$ for several values of the window w and choose the smallest value of w (if any) for which the corresponding plot is "reasonably smooth." Use this plot to determine the length of the warmup period l. [Choosing w is like choosing the interval width Δb for a histogram (see Sec. 6.4.2). If w is too small, the plot of $\bar{Y}_i(w)$ will be "ragged." If w is too large, then the $\bar{Y}_i$ observations will be overaggregated and we will not have a good idea of the shape of the transient mean curve, $E(Y_i)$ for $i = 1, 2, \ldots$.]
- If no value of w in step 3 is satisfactory, make 5 or 10 additional replications of length m. Repeat step 2 using all available replications. [For a fixed value of w, the plot of $\bar{Y}_i(w)$ will get "smoother" as the number of replications increases. Why?]

The major difficulty in applying Welch's procedure in practice is that the required number of replications, n, may be relatively large if the process $Y_1, Y_2, \ldots$ is highly variable.

Example 9.25. A small factory consists of a machining center and inspection station in series, as shown in Fig. 9.6. Unfinished parts arrive to the factory with exponential interarrival times having a mean of 1 minute. Processing times at the

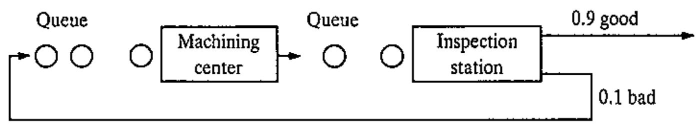

FIGURE 9.6
Small factory consisting of a machining center and an inspection station.

machine are uniform on the interval [0.65, 0.70] minute, and subsequent inspection times at the inspection station are uniform on the interval [0.75, 0.80] minute. Ninety percent of inspected parts are "good" and are sent to shipping; 10 percent of the parts are "bad" and are sent back to the machine for rework. (Both queues are assumed to have infinite capacity.) The machining center is subject to randomly occurring breakdowns. In particular, a new (or freshly repaired) machine will break down after an exponential amount of *calendar* time with a mean of 6 hours (see Sec. 13.4.2). Repair times are uniform on the interval [8,12] minutes. Assume that the factory is initially empty and idle.

Consider the stochastic process $N_1, N_2, \ldots$, where N_i is the number of parts produced in the ith hour. Suppose that we want to determine the warmup period l so that we can eventually estimate the steady-state mean hourly throughput $\nu = E(N)$ (see Example 9.27). We made $n = 10$ independent replications of the simulation each of length $m = 160$ hours (or 20 days). In Fig. 9.7 we plot the

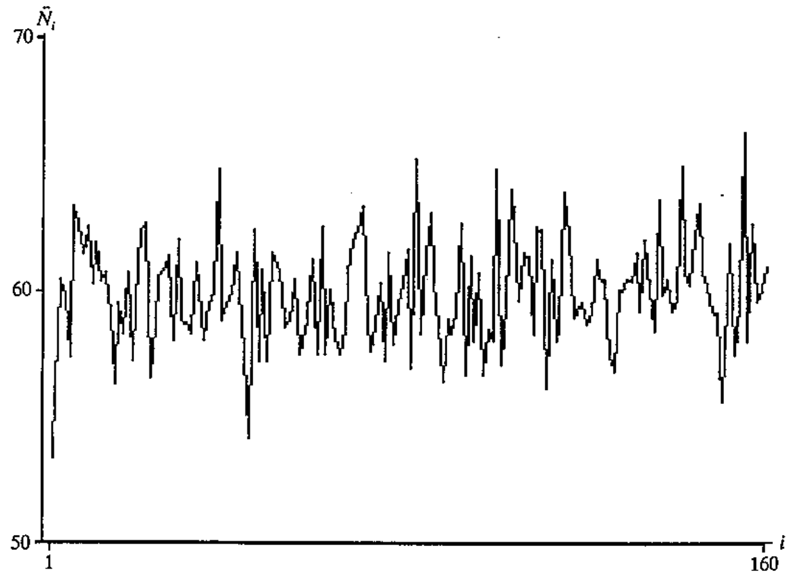

FIGURE 9.7
Averaged process for hourly throughputs, small factory.

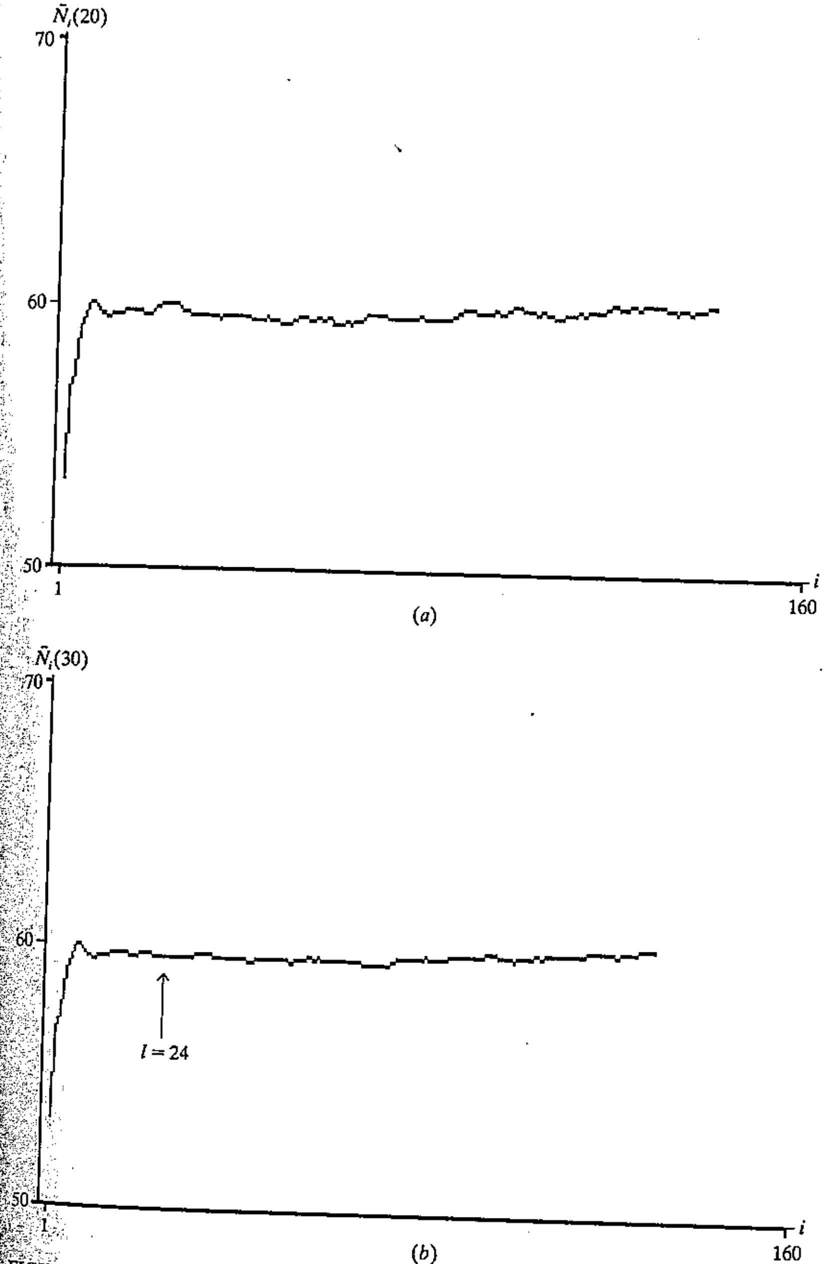

FIGURE 9.8
Moving averages for hourly throughputs, small factory: (a) $w = 20$; (b) $w = 30$.

averaged process $\bar{N}_i$ for $i = 1, 2, \ldots, 160$. It is clear that further smoothing of the plot is necessary, and that one replication, in general, is not sufficient to estimate l. In Figs. 9.8a and 9.8b we plot the moving average $\bar{N}_i(w)$ for both $w = 20$ and $w = 30$. From the plot for $w = 30$ (which is smoother), we chose a warmup period of $l = 24$ hours. Note that it is probably better to choose l too large rather than too small, since our goal is to have $E(Y_i)$ close to ν for $i > l$. (We choose to tolerate slightly higher variance in order to be more certain that our point estimator for ν will have a small bias.)

Example 9.26. Consider the process $C_1, C_2, \ldots$ for the inventory system of Example 9.3. Suppose that we want to determine the warmup period l in order to estimate the steady-state mean cost per month $c = E(C) = 112.11$. We made $n = 10$ independent replications of the simulation of length $m = 100$ months. In Fig. 9.9 we plot the moving average $\bar{C}_i(w)$ for $w = 20$, from which we chose a warmup period of $l = 30$ months.

Schruben (1982), in an important paper, developed a very general procedure based on standardized time series (see Sec. 9.5.3) for determining whether the observations $Y_{s+1}, Y_{s+2}, \ldots, Y_{s+t}$ (where s need not be zero) contain initialization bias with respect to the steady-state mean $\nu = E(Y)$, that is, whether $E(Y_i) \neq \nu$ for at least one i (where $s + 1 \leq i \leq s + t$). As the procedure is now constituted, it is not an algorithm for determining a deletion amount l, but rather a test to determine whether a set of observations contains initialization bias. For example, it could be applied to the truncated averaged process $\bar{Y}_{l+1}, \bar{Y}_{l+2}, \ldots, \bar{Y}_m$ resulting from applying Welch's procedure, in order to determine if there is significant remaining bias. Schruben tested his procedure on several stochastic models with a known value of ν, and found that it had high power in detecting initialization bias. A variation of this initialization-bias test is given in Schruben, Singh, and Tierney (1983).

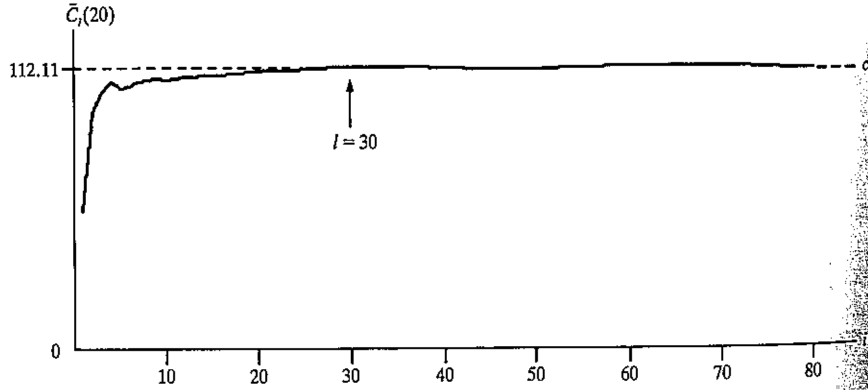

FIGURE 9.9
Moving average with $w = 20$ for monthly costs, inventory system.

In Example 9.23 we saw that initializing the $M/M/1$ queue with the steady-state number in system distribution resulted in the process $D_1, D_2, \ldots$ not having an initial transient. This suggests trying to estimate the steady-state distribution from a "pilot" run, and then independently sampling from this estimated distribution in order to determine the initial conditions for each production run. Kelton (1989) applied this idea to several queueing systems and also a computer model, where in each case the state of the system is an integer-valued random variable. He found that random initialization reduced the severity and duration of the initial transient period as compared with starting the simulation in a fixed state (e.g., no one present in a queueing system). This technique would be harder to apply, however, in the case of many real-world simulations, where the state of the system has a multivariate distribution [see Murray (1988) and Law (1983, p. 1016) for further discussion]. Glynn (1988) discusses a related method where a one-time pass through the "transient period" is used to specify the starting conditions for subsequent replications.

9.5.2 Replication/Deletion Approach for Means

Suppose that we want to estimate the steady-state mean $\nu = E(Y)$ of the process $Y_1, Y_2, \ldots$. There are six fundamental approaches for addressing this problem, which are discussed in this and the next section. We will for the most part, however, concentrate on one of these, the replication/deletion approach, for the following reasons:

1. If properly applied, this approach should give reasonably good statistical performance.
2. It is the easiest approach to understand and implement. (This is very important in practice due to the time constraints of many simulation projects and because many analysts do not have the statistical background necessary to use some of the more complicated analysis approaches.)
3. This approach applies to all types of output parameters (i.e., Secs. 9.4 through 9.6).
4. It can easily be used to estimate several different parameters for the same simulation model (see Sec. 9.7).
5. This approach can be used to compare different system configurations, as discussed in Chap. 10.

We now present the *replication/deletion approach* for obtaining a point estimate and confidence interval for ν. The analysis is similar to that for terminating simulations except that now only those observations beyond the warmup period l in each replication are used to form the estimates. Specifically, suppose that we make n' replications of the simulation each of length m'

observations, where m' is much larger than the warmup period l determined by Welch's graphical method (see Sec. 9.5.1). Let Y_{ji} be as defined before and let X_j be given by

$$X_j = \frac{\sum_{i=l+1}^{m'} Y_{ji}}{(m'-l)} \qquad \text{for } j = 1, 2, \ldots, n'$$

(Note that X_j uses only those observations from the jth replication corresponding to "steady state," namely, $Y_{j,l+1}, Y_{j,l+2}, \ldots, Y_{j,m'}$.) Then the X_j's are IID random variables with $E(X_j) \approx \nu$ (see Prob. 9.15), $\bar{X}(n')$ is an approximately unbiased point estimator for ν, and an approximate $100(1-\alpha)$ percent confidence interval for ν is given by

$$\bar{X}(n') \pm t_{n'-1,1-\alpha/2}\sqrt{\frac{S^2(n')}{n'}} \tag{9.5}$$

where $\bar{X}(n')$ and $S^2(n')$ are computed from Eqs. (4.3) and (4.4), respectively.

One legitimate objection that might be levied against the replication/deletion approach is that it uses one set of n replications (the pilot runs) to determine the warmup period l, and then uses *only* the last $m'-l$ observations from a different set of n' replications (production runs) to perform the actual analyses. However, this is often not a problem due to the relatively low cost of computer time (see Sec. 9.1).

In some situations, it should be possible to use the initial n pilot runs of length m observations both to determine l and to construct a confidence interval. In particular, if m is substantially larger than the selected value of the warmup period l, then it is probably safe to use the "initial" runs for both purposes. Since Welch's graphical method is only approximate, a "small" number of observations beyond the warmup period l might contain significant bias relative to ν. However, if m is much larger than l, these biased observations will have little effect on the overall quality (i.e., lack of bias) of X_j (based on $m-l$ observations) or $\bar{X}(n)$. Strictly speaking, however, it is more correct statistically to base the replication/deletion approach on two independent sets of replications (see Prob. 9.16).

Example 9.27. For the manufacturing system of Example 9.25, suppose that we would like to obtain a point estimate and 90 percent confidence interval for the steady-state mean hourly throughput $\nu = E(N)$. From the $n = 10$ replications of length $m = 160$ hours used there, we specified a warmup period of $l = 24$ hours. Since $m = 160$ is much larger than $l = 24$, we will use these same replications to construct a confidence interval. Let

$$X_j = \frac{\sum_{i=25}^{160} N_{ji}}{136} \qquad \text{for } j = 1, 2, \ldots, 10$$

Then a point estimate and 90 percent confidence interval for ν are given by

$$\hat{\nu} = \bar{X}(10) = 59.97$$

and

$$\bar{X}(10) \pm t_{9,0.95}\sqrt{\frac{0.62}{10}} = 59.97 \pm 0.46$$

Thus, in the long run we would expect the small factory to produce an average of about 60 parts per hour. Does this throughput seem reasonable? (See Prob. 9.17.)

The half-length of the replication/deletion confidence interval given by (9.5) depends on the variance of X_j, $\mathrm{Var}(X_j)$, which will be unknown when the first n replications are made. Therefore, if we make a fixed number of replications of the simulation, the resulting confidence-interval half-length may or may not be small enough for a particular purpose. We know, however, that the half-length can be decreased by a factor of approximately 2 by making four times as many replications. See also the discussion of "Obtaining a Specified Precision" in Sec. 9.4.1.

9.5.3 Other Approaches for Means

In this section we present a more comprehensive discussion of procedures for constructing a point estimate and a confidence interval for the steady-state mean $\nu = E(Y)$ of a simulation output process $Y_1, Y_2, \ldots$. The following definitions of ν are usually equivalent:

$$\nu = \lim_{i\to\infty} E(Y_i)$$

and

$$\nu = \lim_{m\to\infty} \frac{\sum_{i=1}^{m} Y_i}{m} \qquad \text{(w.p. 1)}$$

General references on this subject include Banks and Carson (1984), Bratley, Fox, and Schrage (1987), Fishman (1978), Law (1983), and Welch (1983).

Two general strategies have been suggested in the simulation literature for constructing a point estimate and confidence interval for ν:

1. *Fixed-sample-size procedures*. A single simulation run of an *arbitrary* fixed length is made, and then one of a number of available procedures is used to construct a confidence interval from the available data.
2. *Sequential procedures*. The length of a single simulation run is sequentially increased until an "acceptable" confidence interval can be constructed. There are several techniques for deciding when to stop the simulation run.

Fixed-Sample-Size Procedures. There have been six fixed-sample-size procedures suggested in the literature [see Law (1983) and Law and Kelton (1984) for surveys]. The replication/deletion approach, which was discussed in Sec. 9.5.2, is based on n independent "short" replications of length m observations.