Whether a simulation is considered to be terminating or non-terminating depends on both the objective of the simulation study and the nature of the system

## Nature of the Problem

Once a stochastic variable has been introduced into a simulation model, almost all the system variable describing the system's behavior also become stochastic, because of the way endogenous events make one variable depend upon another. The values of most, if not all, of the system variable will fluctuates the simulation proceeds, so that no one measurement can be arbitrarily taken to represent the value of a variable. Instead, many observations of the variable's value must be made, in order to make a statistical estimate of its true value. Some statement must be made, in order to make a statistical estimate of its true value. Some statement must also be made about the probability of the true value falling within a given interval about the estimated value. Such as statement defines a confidence interval. Without it, simulation results are of little value to a system analyst.

A large body of statistical methods has been developed over the years to analyze results in science, engineering, and other fields where experimental observation are made. Because of the experimental nature of system simulation, it seems natural to attempt applying these methods to simulation results. Unfortunately, most of them presuppose that the observations being made are mutually independent-a reasonable assumption when an experiment is being repeated, or independent samples are being selected.

Simulation results, however, are not likely to be mutually independent. A single simulation run will produce many "observed" values of a variable, but the value observed at one time is likely to be influenced by influenced by the value at some earlier time. For example, in the simulation of a waiting line, the time one entity spends waiting depends upon the



number of entities that happened to be on the waiting careful application of the established statistical method. This need has also lead to the development of new statistical methods, and it is the subject of must current research work.

One concern of the newly developing statistical methodology is to ensure that the statistical estimate are consistent, meaning that, as the sample size increase, the estimate tends to the true value. Another concern is to control bias in measures of both mean values and variances. Bias causes the distribution of estimates based on finite sample to differ significantly from the true population statistics, even though the estimates may be consistent. A third, practical aspect of current research work is the attempt to develop sequential testing methods that will allow automatic controls to determine how long a simulation should be run in order to obtain a given level of confidence in its result.

## Measures of Performance and their Estimation

1. Consider the estimation of a performance parameter, $\theta$ (or $\phi$) of a simulated system.
   - Discrete time data: $(Y_1, Y_2, \ldots, Y_n)$ with ordinary mean: $\theta$
   - Continuous-time data: $\{Y(t), 0 \le t \le T_E\}$ with time-weighted mean: $\phi$
2. Point estimation for discrete time data
   - The point estimator:

$$\hat{\theta} = \frac{1}{n}\sum_{t=1}^{n} Y_t$$

   - Is unbiased if its expected value is $\theta$, that is if: $E(\hat{\theta}) = \theta$
   - Is biased if: $E(\hat{\theta}) \neq \theta$ and $E(\hat{\theta}) - \theta$ is called bias of $\hat{\theta}$.

**Point Estimation**

3. Point estimation for continuous-time data.

   – The point estimator:

$$\hat{\phi} = \frac{1}{T_E}\int_0^{T_E} Y(t)\,dt$$

   – Is biased in general where: $E(\hat{\phi}) \neq \phi$

   – An unbiased or low-bias estimator is desired.

4. Usually, system performance measures can be put into the common framework of $\theta$ or $\phi$:

   – The proportion of days on which sales are lost through an out-of-stock situation, let:

$$Y(i) = \begin{cases} 1, & \text{if out of stock on day } i \\ 0, & \text{otherwise} \end{cases}$$

5. Performance measure that does not fit: quantile or percentile: $\Pr\{Y \leq \theta\} = p$

   – Estimating quantiles: the inverse of the problem of estimating a proportion or probability.

   – Consider a histogram of the observed values Y.

   Find $\hat{\theta}$ such that 100% of the histogram is to the left of (smaller than) $\hat{\theta}$.

   – A widely used performance measure is the median, which is the 0.5 quantile or 50th percentile.

**Interval Estimation**

6. Confidence Interval (CI)

   – A measure of error.

   – Where, $Y_i$ are normally distributed.

$$\bar{Y}_{..} \pm t_{\frac{\alpha}{2},\, R-1} \frac{S}{\sqrt{R}}$$



   – We cannot know for certain how far $\bar{Y}_{..}$ is from $\theta$ but CI attempts to bound that error.

   – A CI, such as 95%, tells us how much we can trust the interval to actually bound the error between $\bar{Y}_{..}$ and $\theta$.

   – The more replications we make, the less error there is in $\bar{Y}_{..}$ (converging to 0 as R goes to infinity)

7. Prediction Interval (PI):

   – A measure of risk.

   – A good guess for the average cycle time on a particular day is our estimator but it is unlikely to be exactly right.

   – PI is the designed to be wide enough to contain the actual average cycle time on any particular day with high probability.

   – Normal-theory prediction interval:

$$\bar{Y}_{..} \pm t_{\frac{\alpha}{2},\, R-1}\, S\sqrt{1 + \frac{1}{R}}$$

   – The length of PI will not go to 0 as R increases because we can never simulate away risk.

   – PI's limit: $\theta \pm Z_{\alpha/2}\, \sigma$

## ESTIMATION METHOD

We first review some of the statistical methods commonly used to estimate parameters from observations on random variable. Usually, a random variable is drawn from an infinite population that has a stationary probability variable is drawn from an infinite population that has a stationary probability distribution with a finite mean, $\mu$, and finite variance, $\sigma^2$. This means that the population distribution is not affected by the number of samples already made, nor does it change with time. If, further, the value of one sample is not affected in any way by the value of any other sample, the random variables are mutually independent. Random variable that meet all these conditions are said to be independently and dietetically distributed, usually abbreviated to i.i.d. Under broad condition that can be expected to hold for simulation data, the central limit theorem can be applied to i.i.d. data. The theorem states that the sum of $n$ i.i.d. variable, drawn from a population that has a mean of $\mu$ and a variance of $\sigma^2$, is approximately distributed as a normal variable with a mean of $n\mu$ and a variance of $n\sigma^2$.

Any normal distribution can be transformed into a standard normal distribution, that has a mean of 0 and a variance of 1. Let $x_i$ (i = 1, 2, . . . , n) be the n i.i.d. random variables. Using the central limit theorem and, applying the transformation, gives the following (approximate) normal variate:

$$z = \frac{\sum_{i=1}^{n} x_i - n\mu}{\sqrt{n}\sigma}$$

Dividing top and bottom of the fraction by n, and defining $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ we have

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

The variable $\bar{x}$ is the sample mean. It can be shown to be a consistent estimator for the mean of the population from which the sample is drawn. Since the sample mean is the sum of random variable, it is itself a random variable. As a result, a confidence interval about its computed value, needs to be established.



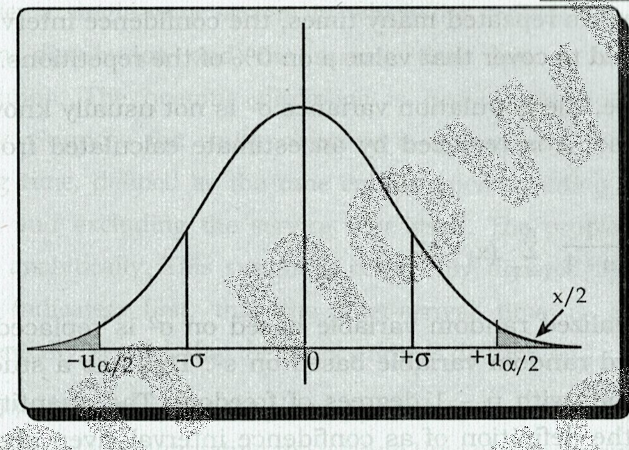



**Figure 7.1: Probability density function of the standard normal variate**

The probability density function of the standard normal variate is illustrate in Fig 7.1. The integral from $-\infty$ to a value $u$ is the probability that $z$ is less than or available. Suppose the value of us is chosen so that $\Phi(u)$ and table of its value are widely available. Suppose the value of u is chosen so that $\Phi(u) = 1 - a/2$, where $\alpha$ is some constant less than 1, and denote that value of u by $u_{\alpha/2}$ is also $\alpha/2$. Consequently, the probability that z lies between $-u_{\alpha/2}$ and $u_{\alpha/2}$ is $1 - \alpha$. That is,

$$\text{Prob}\{-u_{a/2} \le z \le u_{\alpha/2}\} = 1 - \alpha$$

In terms of the sample means, this probability statement can be written

$$\text{Prob}\left\{\bar{x} + \frac{\sigma}{\sqrt{n}} u_{\alpha/2} \ge \mu \ge \bar{x} - \frac{\sigma}{\sqrt{n}} u_{\alpha/2}\right\} = 1 - \alpha$$

The constant $1 - \alpha$ (usually expressed as a percentage) is he confidence level and the interval $\bar{x} \pm \frac{\sigma}{\sqrt{n}} u_{\alpha/2}$

is the confidence interval. The size of the confidence interval depends upon the confidence level chosen. Typically, the confidence level might by 90%, in which case $u_{\alpha/2}$ is 1.65. The statement then says the $\mu$ will be covered by the confidence interval $\bar{x} \pm 1.65 \frac{\sigma}{\sqrt{n}}$ with probability 0.9; meaning that, if the experiment is repeated many times, the confidence interval can be expected to cover that value $\mu$ on 0% of the repetitions.

In practice, the population variance $\sigma^2$ is not usually known; in which case, it is replaced by an estimate calculated from the formula.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_I - \bar{x})^2$$

The normalized random variable based on $\sigma^2$ is replaced by a normalized random variable based on $s^2$. This has a student-t distribution, with n – 1 degrees of freedom. The quantity $u_{\alpha/2}$ used in the definition of as confidence interval given above, is replaced by a similar quantity, $t_{n-1, \alpha/2}$, based on the students-*t* distribution, for which tables are also readily available.

The Student-*t* distribution is strictly accurate only when the population from which the sample are drawn is normally distributed. It is common practice, however, to rely when the central limit theorem is being invoked.

Expressed in terms of the estimated variance, $s^2$, the confidence interval for $\bar{x}$ is defined by

$$\bar{x} \pm \frac{S}{\sqrt{n}} t_{n-1,\alpha/2}$$

## SIMULATION RUN STATISTIC

In addition to the assumption of normality inherent in these of the central limit theorem, the method of establishing confidence intervals, outlined in the previous section, is based on two other assumptions. It is assumed that the observation are mutually independent, and it is assumed that the distribution from which they are drawn is stationary. Unfortunately, many statistics of interest in a simulation do not meet these conditions. To illustrate the problems that arise in measuring statistics from simulation runs, specific example with be discussed.



Consider a single-sever system in which the arrivals occur with a Poisson distribution and the service time has an exponential distribution. The queuing discipline is first-in, first-out with no priority. Suppose the study objective is to measure the meaning waiting time, defined as the time entities spend waiting to receive service and excluding the service time itself. The problem can be solved analytically. This system is commonly denoted by M/M/1, which indicates; first, that the inter-arrival time is distributed exponentially; second, that the service time is distributed.)

In a simulation run, the simplest approach is to estimate the mean waiting time by accumulating the waiting time of n successive entities and dividing by n. This measure, the sample mean, is denoted by $\bar{x}$ (n) to emphasize the fact that its value depends upon the number of observations taken. If $x_1$ (i = 1, 2, . . . , n) are the individual waiting times (including the value 0 for those entities that do not have to wait), then

$$\bar{x}(n) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Waiting times measured this way are not independent. Whenever a waiting line forms, the waiting time of each entity on the line clearly depends upon the waiting time of ties predecessors. Any series of data that has this property of having one value affect other values is said to be auto correlated. The degree to which the data are auto correlated can be measured in ways that will be briefly described in a later section.

Under broad conditions that can normally be expected to hold in a simulation run, the sample mean of auto correlated data can be shown to approximate a normal distribution as the sample size increase. The usual formula for estimating the mean value of the distribution, remains a satisfactory estimate for the mean of auto correlated data. However, the variance of the auto correlated data is not related to the population variance by the simple expression $\sigma^2/n$, as occurs for independent data. A term must be added to account for the autocorrelation. The term is usually positive in the case M/M/1 system, so that, if it is ignored, the variance is underestimated, but, in other system, it can be negative, resulting in an overestimated.

Another problems that must be faced is that the distributions may not be stationary. In particular, a simulation run is started with the system in some initial state, frequently the idle state, in which no service is being given and no entities are waiting. The early arrivals then have a more than normal probability of obtaining service quickly, so a sample mean that includes that early arrivals will be biased. As the length of the simulation run is extended, and the sample size increase, the effect of the bias will die out. For a given sample size starting from a given initial condition, the sample mean distribution is stationary; but, if the distributions could be compared for different sample sizes, the distributions would be slightly different. The analytical solutions previously quoted are for the steady state values to which the distributions converge as the sample size increase.

Figure 7.2, shows how the expected value of the sample mean depends upon the sample length, for the M/M/1 system, starting from an initial empty state, with a server utilization of 0.. It is known that the steady state mean in this case is 8.1. It can be seen that the mean value is biased below the steady state value. The bias diminishes as the sample size increase but, even with a sample size of 2000, the mean has still only reached about 95% of the steady state value. The steady state value will be approached more rapidly for lower levels of server utilization, but, unfortunately, high server utilization cases are usually the ones of interest in simulation studies.



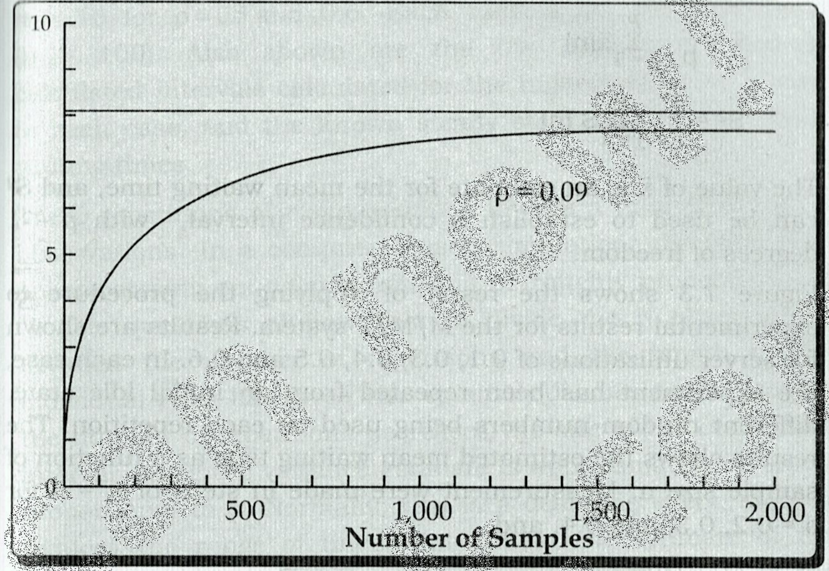



**Figure 7.2: Mean wait time in M/M/1 system for different sample sizes.**

## REPLICATION OF RUNS

One way of obtaining independent results is to repeat the simulation. Repeating the experiment with different random numbers for the same sample size n gives a set of independent determinations of the sample mean $\bar{x}$ (n). Even through the distribution of the sample mean depends upon the degree of autocorrelation, these independent determinations of the sample mean can properly be used to estimate the variance of the distribution. Suppose the experiment is repeated *p* times with independent random number series. Let $x_{ij}$ be the $i^{th}$ observation in the $j^{th}$ observation in the $j^{th}$ run, and let the sample mean and variance for the $j^{th}$ run be denoted by $\bar{x}_j(n)$ and $S_j^2$, respectively. For that $j^{th}$ run, the estimates are:

$$\bar{x}_n(n) = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

$$S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}[x_{ij} - \bar{x}_j(n)]^2$$

Combining the results of *p* independent measurements gives the following estimates for the mean writing time, $\bar{x}$, and the variance, $S^2$, of the population:

$$\bar{x} = \frac{1}{p} \cdot \sum_{j=1}^{p} \bar{x}(n)$$

$$S^2 = \frac{1}{p} \cdot \sum_{j=1}^{p} S(n)$$

The value of $\bar{x}$ is an estimate for the mean waiting time, and $S^2$ can be used to establish a confidence interval, with $p - 1$ degrees of freedom.

Figure 7.3 shows the result of applying the procedure to experimental results for the M/M/1 system. Results are shown for server utilizations of 0.1, 0.3, 0.4, 0.5 and 0.6. In each case, the experiment has been repeated from an initial idle state, different random numbers being used on each repetition. The results shows the estimated mean waiting time as a function of sample size n. Measurement were made in steps of n = 5 for ρ = 0.2, 0.3, and 0.4, and

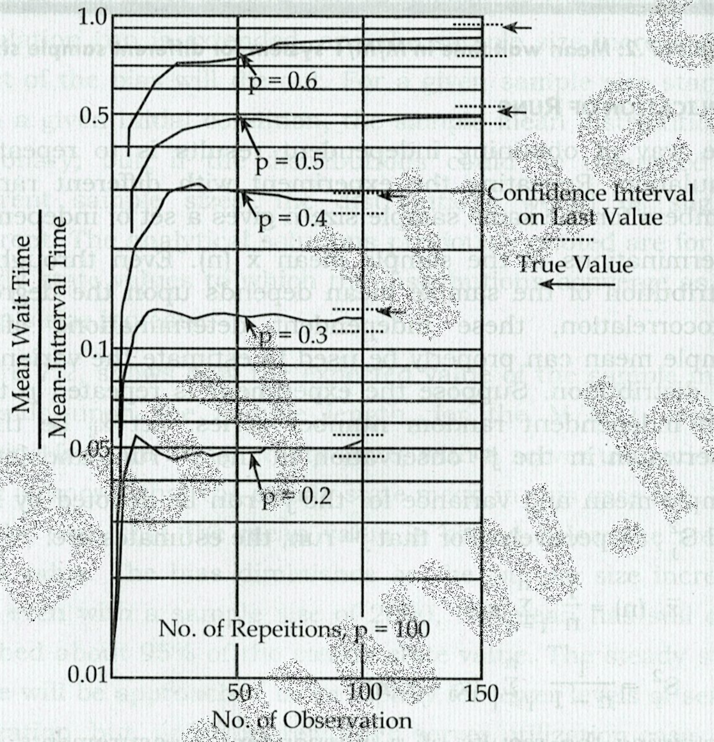



**Figure 7.3: Experimentally measured wait time in M/M/ system for different sample sizes.**



$n$ = 10 for ρ = 0.5 and 0.6. Each case is for 100 repetitions (p = 100). Also shown are the 0% confidence intervals calculated intervals calculated for the highest value of n used in each case, and the known steady state value of the mean waiting times.

The $p$ repetitions of n observations involve a total of N = $p$. n observations. In a computer based simulation, the total time spent carrying out calculations will be roughly proportional to N. The questions of how best to divide the N observations between the number of repetitions and the length of the individual runs has given rise to much discussion. Increasing the number of repetitions decreases the size of the confidence interval, since the variance estimate is approximately inversely proportional to p. Normally, this is a desirable effect, since it reduces the range of uncertainty about the estimate of the mean. However, the desirability assumes the estimate of the means unbiased. A short confidence limit, centered on a biased estimate of the mean is unbiased. A short confidence limit, centered on a biased estimate, can easily fail to cover the true value being estimated. In the replication of simulation runs, (in order to keep N constant), the estimate of the mean will be more biased, as a result of the initial empty state.

For example, reports results of experiments with the M/M/1 system at a server utilization of 0.9. The entire process of computing a confidence interval by replication of runs starting from an empty state, in this case at a 90% confidence level, was repeated 400 times for different combinations of n and $p$. For each combination of n and p, the total number of observations used in each determination of a single confidence interval, N, was kept at 12,800.

At a 90% confidence level, it is to be expected that 90% of the 400 independently determined confidence intervals will cover the true mean. Taking the steady state samples (n = 2,560 and p = 5), 83% of the confidence intervals covered the true mean. At the other extreme of the experiment, 40 replications, each of 320 samples (n = 320 and p = 40), the figure dropped to 9%. It can be estimated from Fig. 7.3 that, at steady state mean. At a sample size of 320, the mean is about 6, which is 2.1 below the steady state mean. The shorter confidence interval, achieved with the greater number of replications, is bought at the cost of a much greater bias-resulting in the large difference in the accuracy.

From this evidence, we see that it is preferable to keep the number of repetitions as low as possible, bearing in mind the need to approximate a normal distributions with the sample means. These results, of course, are for a simple system. However, they are indicative of the type of research work being conducted in the study of the simulation process.

## ELIMINATION OF INITIAL BIAS

The experimental results given in figure 7.3, clearly show the need to remove the initial bias, or reduce its effects. Two general approaches can be taken to remove the bias: the system can be started in a more representative state than the empty state, or the first part of the simulation run can be ignored.

In some simulation studies, particularly of existing systems, there may be information available on the expected conditions that makes it feasible to select better initial conditions. The ideal situation is to know the steady state distribution. In the study previously discussed, Law repeated the experiments on the M/M/1 system, supplying an initial waiting line for each run, selecte4d at random from the known steady state distribution of the waiting line. The case of 40 repetitions of 320 samples, which previously resulted in a coverage of only %, was improved to a coverage of 88%. Of course, the theoretical knowledge on which this technique is based is not usually available. However, experience with an existing system, or similar type of system, could provide a reasonable approximation.



The more common approach to removing initial bias is to eliminate an initial section of the run. The run is started from an idle state and stopped after a certain period of time. The entities existing in the system at that time are left as they are. The run is then restarted with statistics being gathered from the point of restart. As a practical matter, it is usual to program the simulation so that statistics are gathered from the beginning, and simply wipe out the statistics gathered up to the point of restart. No simple rules can be given to decide how long an interval should be eliminated. It is advisable to use some pilot runs starting from the idle state to judge how long the initial bias remains. The can be done by plotting the measured statistic against run length as has been done in figure 7.3.

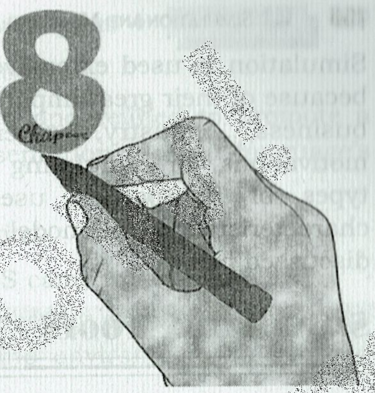
Another disadvantage of eliminating the first part of a simulation run is that the estimate of the variance, needed to establish a confidence limit, must be based on less information. The reduction in bias, therefore, is obtained at the price of increasing the confidence interval size.

## DISCUSSION EXERCISE

1. What are the types of simulations with respect to output analysis?
2. Explain stochastic nature of output data with example.
3. Explain measure of performance and their estimation.
4. Explain output analysis of terminating simulations with examples.
5. With illustrative examples explain output analysis of steady-state simulations.
6. Explain how probabilities and quantiles can be estimated from summary data?

7. Describe initialization bias in steady state simulation
8. Explain batch means for interval estimation in steady state simulation
9. What is the nature of problem in analysis of simulation output? What is internal bias and how can you eliminate that?
10. Why do we perform the analysis of simulation output? Explain how you use simulation run statistics in the output analysis.
11. Why it is necessary to analysis the simulation output. Explain different estimation methods which are used in simulation output analysis.

❑❑❑




# Simulation of Computer Systems

**Chapter Outline**

**After studying this chapter, the students will be able to understand the**

- Simulation Tools
- Simulation Languages: GPSS
- Case Studies of different types of Simulation Models and Construction of sample mathematical models