

CHAPTER

10

Estimation of Absolute Performance

Module 4

10.1 Introduction

- Q. Write a note on Need for output analysis in simulations.

MU - Dec. 10, Dec. 14

When a simulation is stochastic (its behavior as seen in its outputs is at least partially determined by one or more random variables), we have the problem of deciding just what to do with those outputs. We know, or should know, the performance metrics that are important – what the simulation outputs should tell us about the behavior of the real system being simulated.

To be precise, the simulation outputs tell us about the behavior of the model; it is the task of the modeler to ensure that the model is sufficiently “close” to the system being modeled that the results obtained from the model can be applied to the system.

Each output from a simulation of the model can be thought of as a measurement of its behavior. We need to consider two issues. First, how do we use these measurements to generate the performance metrics of interest? Second, how good are the performance metrics we generate? That is, how much faith can we put in them? The field of statistics provides answers to both questions.

The difference between a good simulation and a bad one is often determined by how well we apply the appropriate techniques from statistics to analyze the data produced by the simulation.

10.2 Types of Simulations with respect to Output Analysis

- Q. What is Terminating Simulation and Non terminating Simulation give examples ?

MU - Dec. 06, May 09

- Q. Explain different types of simulation with respect to output analysis.

MU - May 10, May 11, Dec. 11, May 12, Dec. 13, May 14, May 15

- There are two types of simulations with respect to output analysis :
 - Terminating Simulation
 - Non terminating Simulation

10.2.1 Terminating Simulation

- Terminating systems are characterized by having a fixed starting condition and a naturally occurring event that marks the end of the system.
- An example of a terminating system is a work day that starts at 8 am and ends at 4 pm at a bank.
- For terminating systems the initial conditions of the system generally affect the desired measures of performance.
- The purpose of simulating terminating systems is to understand their behaviour during a certain period of time, and this is also referred to as studying the transient behaviour of the system.
- Terminating simulations are used to simulate terminating systems. The length of a terminating simulation is determined either by the system itself, if the system is a terminating system, or by the objective of a simulation study.
- For example, the goal of a performance study may be to analyze the performance of either the entire work day at the bank or only during the time it takes to assist the first 10 customers.
- The length of a terminating simulation can be determined by a fixed amount of time, e.g. 8 hours, or it can be determined by some condition, e.g. the departure of the tenth customer.
- Thus terminating (transient) simulation runs for some duration of the time T_E and the initial condition and stopping event is specified.

10.2.2 Non-terminating Simulation

- In a non-terminating system, the duration of the system is not finite. The Internet exemplifies a non-terminating system.
- Non-terminating simulations are used to simulate non-terminating systems. In a non-terminating simulation, there is no event to signal the end of a simulation, and such simulations are typically used to investigate the long-term behaviour of a system.
- Non-terminating simulations must, of course, stop at some point, and it is a non-trivial problem to determine the proper duration of a non-terminating simulation.

- If the behaviour of the system becomes fairly stable at some point, then there are techniques for analyzing the steady-state behaviour of the system using non-terminating simulations.
- When analyzing steady-state behaviour using non-terminating simulations, it is often useful to be able to specify a warm-up period in which data is not collected because the model has not yet reached a steady state. Determining when, or if, a model reaches steady state is also a complicated issue.
- Example of non terminating simulation is continuous production system, telephone system, and assembly line.

10.3 Stochastic Nature of Output Data

- Simulation result is generated for a given input data. Essentially a model is an input/output transformation.
- Since the input variables are random variables, the output variables are random variables as well, thus they are stochastic (probabilistic).
- Model output consists of one or more random variables because the model is an input output transformation and the input variables are random variables.
- The simulation model is run with different input data sets for n times.
- Two issues that need to be addressed with regard to the statistical analysis. They are :
 - Output data from various runs are independent and identically distributed.
 - Estimation of the true utilization: Point estimate $\hat{\theta} = E(\hat{\theta})$.
 - Estimation of the error in our point estimate as standard error or a confidence interval.
 - The effects of correlation and initial conditions of the estimation of the long run mean measures of performance.
 - The batching transforms the continuous time queue length process into a discrete time batch mean process.
 - The sequence of the batches is auto correlated as all the data is obtained using one replication only.
 - Avoid direct statistical analysis of the within replication output, as the sequence is a non-stationary auto correlated stochastic process.
 - Here non-stationary means not identically distributed.

10.4 Measure of Performance and Their Estimation

10.4.1 Estimators

- When a simulation uses random variables in modeling the workload of a system, the outputs of the simulation are themselves random variables. The performance metrics for the system typically are characteristics of the underlying distributions of these random variables. By far the most popular type of metric is the mean, but we might also (or instead) be interested in the variance or standard deviation, the median, a percentile of the distribution, etc.
- In general, these characteristics are deterministic (non-random) quantities. We attempt to calculate them from the sequence of output values produced by the simulation, using some function or functions. Since functions of random variables are themselves random variables, we are in fact trying to estimate fixed but unknown quantities by random variables. For example, we may want to know the mean value of an output. Let $\{X_i\}$ be a sequence of n identically distributed random variables.

Define

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- \bar{X} , the sample mean, is an estimator for $\mu = E[X_i] = E[X]$, where X is a representative random variable with the same distribution as each of the X_i 's and μ is their common mean. In general, an estimator $\Theta = \Theta(X_1, X_2, \dots, X_n)$ is an unbiased estimator of a characteristic θ if $E[\Theta(X_1, X_2, \dots, X_n)] = \theta$
- The sample mean is an unbiased estimator of μ , as we can easily show.

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] \\ &= \frac{1}{n} \sum_{i=1}^n E[X] = \frac{1}{n} \cdot n \cdot \mu = \mu \end{aligned}$$

- This is always true, even if the X_i 's are not independent of one another.
- As another example of an unbiased estimator, consider the sample variance, defined by:

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

To show that s^2 is unbiased, let us first rewrite the expression for s^2 as

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} \bar{X}^2 \end{aligned}$$

The expectation of the sample variance is

$$E[s^2] = \frac{1}{n-1} \sum_{i=1}^n E[X_i^2] - \frac{n}{n-1} E[\bar{X}^2]$$

If σ^2 is the variance of X ,

$$E[X_i^2] = \text{var}[X_i] + (E[X_i])^2 = \sigma^2 + \mu^2$$

and if the X_i 's are i.i.d.,

$$E[\bar{X}^2] = \text{var}[\bar{X}] + (E[\bar{X}])^2 = \frac{\sigma^2}{n} + \mu^2$$

Substituting equation gives

$$E[s^2] = \frac{1}{n-1} \cdot n \cdot (\sigma^2 + \mu^2) - \frac{n}{n-1} \left(\frac{\sigma^2}{n} + \mu^2 \right) = \sigma^2$$

- More than one estimator of a particular characteristic θ may be unbiased. In this case, we would like to choose an estimator Θ that is "good" in the sense that most of the probability mass in the estimator's distribution is tightly packed about θ . To be more precise, we usually want to select a Θ such that $\Pr[|\Theta - \theta| \geq \epsilon]$ is small. The quantity $|\Theta - \theta|$ is sometimes called the dispersion of Θ . From Chebychev's Inequality, we know that for a random variable Y with mean μ and variance σ^2 ,

$$\Pr[|Y - \mu| \geq t] \leq \frac{\sigma^2}{t^2}, t > 0$$

- In our case, this means that the probability that the dispersion of the estimator Θ exceeds a value ϵ can be bounded as follows:

$$\Pr[|\Theta - \theta| \geq \epsilon] \leq \frac{\text{var}[\Theta]}{\epsilon^2}, \epsilon > 0$$

- Based on this result, we say that an unbiased estimator Θ_1 is more efficient than an unbiased estimator Θ_2 if $\text{var}[\Theta_1] < \text{var}[\Theta_2]$.

- As an example, consider the following general estimator for the mean of a sequence of n i.i.d. random variables

$$\Theta = \sum_{i=1}^n a_i X_i$$

with the restrictions that all of the a_i s are non-negative and

$$\sum_{i=1}^n a_i = 1$$

Then

$$E\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i E[X_i] = \sum_{i=1}^n a_i \mu = \mu \cdot \sum_{i=1}^n a_i = \mu$$

- So that any estimator of this form is an unbiased estimator of μ . Which of these estimators (called linear estimators) is the most efficient estimator of μ ? To answer this question, we must first find the variance of a linear estimator.

$$\text{var}\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i^2 \cdot \text{var}[X_i] = \sum_{i=1}^n a_i^2 \sigma^2$$

- We want to find the values for the a_i s which minimize the variance, subject to the constraints that the a_i s are non-negative and sum to 1. We can solve this problem using Lagrangian multipliers. Let

$$G = \sigma^2 \sum_{i=1}^n a_i^2 + \beta \left[1 - \sum_{i=1}^n a_i \right]$$

- That is, G is the variance of a linear estimator plus β times a term that is zero everywhere in the region of the $(a_1; a_2; \dots; a_n)$ space in which the solution to our problem must lie. To minimize the variance of Θ , we need to find an extremum of G .

$$\frac{\partial G}{\partial a_i} = 2\sigma^2 a_i - \beta = 0 \Rightarrow \beta = 2\sigma^2 a_i$$

- Next, we form the constraint:

$$1 = \sum_{i=1}^n a_i = \sum_{i=1}^n \frac{\beta}{2\sigma^2} = \frac{n\beta}{2\sigma^2}$$

which gives

$$\beta = \frac{2\sigma^2}{n}$$

and hence

$$a_i = \frac{1}{n}$$

That is, the sample mean is the most efficient linear estimator of μ . From the above discussion, we also see that the variance of the sample mean is σ^2/n , which decreases as the number of samples increases.

Table 10.4.1 : Observations of changes in the number of active processes in a multi-programmed CPU

Time	# of active processes
0	0
10	1
15	2
30	3
40	2
70	3
80	4
95	3

The sample mean of the previous section is an example of a point estimator. Each sample is weighted by $1/n$ where n is the total number of samples. The length of time that the quantity being measured was each value is unimportant.

For an interval estimator, the weight given to a particular value depends on the time that the quantity being measured had this value. For example, suppose we want to obtain information about the number of active processes in a multi-programmed CPU. We could observe the system and note all of the instances at which the number of active processes changes and the values just after each change. Table shows one possible set of such measurements made over a 100-second interval.

One statistic in which we might be interested is the average number of active processes just after a change. For this, we use a point estimator, since the average does not depend on how long the number of active processes was 1 or 2 or 3 or 4. If we consider the initial value 0 to be a "change," we have eight samples and the average is $(1+2+3+2+3+4+3)/8 = 2.25$.

On the other hand, we might want to know the average number of active processes during $(0; 100)$, in which case we need to use the interval estimator $[5(1) + 45(2) + 25(3) + 15(4)]/100 = 2.3$.

What if we want to know how long a job was in the system on the average? This is sometimes referred to as the average residence time. If we knew the time that each job

- spent in the system in the interval $(0, 100)$, we would use a point estimator, since the weights for all residence times should be the same.
- Assuming that jobs could never arrive or depart simultaneously, there were five total jobs in the system in the interval $(0, 100)$, three of which were still in the system at the end of the interval. Let T_i be the residence time for the i th job that arrives. Then the average residence time is $\frac{1}{5} \sum_{i=1}^5 T_i$. The summation in the numerator is just the total

number of seconds spent in the system by all jobs in the interval $(0, 100)$. This must be equal to $5(1) + 45(2) + 25(3) + 15(4) = 230$, which can be seen by looking at Fig. 10.4.1.

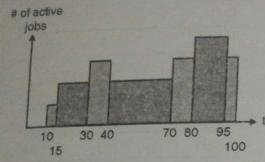


Fig. 10.4.1 : Number of active jobs in a multiprogramming system

- The shaded area under the curve is the total job-seconds in the $(0, 100)$ interval. Hence, $\sum_{i=1}^5 T_i = 230$ and the average residence time is $230 / 5 = 46$ seconds.
- What kind of estimate is this? The values in the numerator are weighted by the lengths of intervals, as in an interval estimator but the denominator is not the sum of the weights, as it should be for an interval estimator. The answer is that it is still a point estimator, as we originally stated, but we have chosen to evaluate the sum of the residence times in a way that does not require us to know the individual residence times, which we do not have.
- This last example illustrates an issue that arises from time to time in the analysis of measurement data, including simulation outputs. Three of the five jobs are still in the system at the end of the measurement interval, and we have in essence assumed that the residence time for each of these three jobs is just the time it has been in the system by the end of the interval.
- Unless all three jobs leave the system at $t = 100$, however, we are under-estimating the average job residence time by including these three in our calculations. This problem only arises because of the way we computed the sum of the job residence times.

- Had we known the residence times of the individual jobs, we could have used only the residence times for the jobs which actually completed in the interval $(0, 100)$.
- It is tempting to try to avoid this issue by running the simulation for a long interval of simulation time, so long that most of the jobs which arrived during the interval completed, and the effect of the jobs remaining in the system at the end of the interval on the average response time is minimal. Unfortunately, there is always the risk that one or more of the remaining jobs have been in the system for a long time.
- Consider Fig. 10.4.2, describing the activity in a multiprogrammed computer which always has two active jobs over the interval $(0, 100)$. Blocks represent individual jobs; shading indicates that the job has not completed by the end of the measurement interval. Ignoring the remaining job in (a) will result in an average residence time of 10, while ignoring the remaining job in (b) leads to an average residence time of a little less than 20.
- This does not seem unreasonable for (b), but underestimates the average by at least a factor of almost 2 for (a). Situations like this can arise for a variety of reasons, one of the most common of which is some form of priority scheduling that causes starvation for one or more classes of jobs when higher priority jobs utilize a large fraction of the system's capacity.

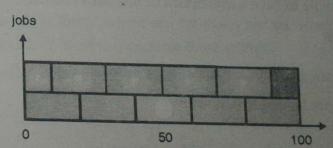
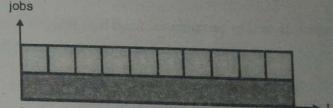


Fig. 10.4.2

10.4.2 Point Estimators

- Q. Explain point estimation and interval estimation of θ . MU - Dec 04

- The basic aim here is to have a point estimate and an interval estimate and an interval estimate of a performance parameter θ (or ϕ)

- The length of the interval estimate is a measure of the error in the point estimate.
- If the simulation output data is of the form $\{Y_1, Y_2, \dots, Y_n\}$, it is called as discrete-time data, since n is discrete valued. Here θ is an ordinary mean.
- For continuous time data = $f\{Y_1, Y_2, \dots, Y_n\}$, is a time weighted mean.
- Here Y_i presents the customer delay.
- And Y_i represents the length of the queue at time t .
- Consider a set of output values for the same measure Y_1, Y_2, \dots, Y_n (e.g. delays of n different runs, or waiting times of n different runs). We want to have
- A point estimate to approximate the true value of Y_p and
- An interval estimate to outline the range where the true value lies.

Point Estimation

Discrete time data

- The point estimator of θ based on the discrete data Y_1, Y_2, \dots, Y_n is defined by :

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n Y_i$$

- The point estimator $\hat{\theta}$ is said to be unbiased for θ if :

$$E(\hat{\theta}) = \theta$$

- In general $E(\hat{\theta}) = \theta + b$ i.e. there is a drifting, or bias.
- It is always desirable to have a low biased or unbiased estimator.

Continuous time data

- For continuous data, the point estimator of ϕ based on data $\{Y(t), 0 \leq t \leq T_E\}$, where T_E is the simulation run length, is defined by

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

and is called a time average of $Y(t)$ over $[0, T_E]$.

- In general $E(\hat{\phi}) = \phi + b$ if $b = 0$, $\hat{\phi}$ is said to be unbiased for ϕ
- It is always desirable to have a low biased or unbiased estimator.

10.4.3 Interval Estimation

- The point estimate rarely matches the actual value of the parameter being estimated.
- Hence there is a need to find an interval estimate.
- For this we need to estimate the variance of the point estimator for the given number of samples we can determine how many samples we need for a given desired variance.
- We construct an interval called as the confidence interval, in which we have definite confidence that it contains the true value of the unknown parameter.
- Consider the sample data point $\{Y_1, Y_2, \dots, Y_n\}$.
- The true variance of the point estimator is $\sigma^2(\hat{\theta}) = \text{var}(\hat{\theta})$.
- The estimate of the variance of point estimator is :

$$\hat{\sigma}^2(\hat{\theta}) = \text{var}(\hat{\theta})$$

- If $E[\hat{\sigma}^2(\hat{\theta})] \neq \sigma^2(\hat{\theta})$, the estimator is then considered to be biased.

- But if the variance of point estimator $\hat{\sigma}^2(\hat{\theta})$ is nearly unbiased, the statistic, $t = \frac{\hat{\theta} - \theta}{\hat{\sigma}(\hat{\theta})}$

is approximately t -distributed with f degree of freedom and can be bounded by $100(1 - \alpha)$ confidence interval for and it is given by :

$$\hat{\theta} \pm t_{\alpha/2, f} \hat{\sigma}(\hat{\theta}) \text{ or } \hat{\theta} - t_{\alpha/2, f} \hat{\sigma}(\hat{\theta}) \leq \hat{\theta} \leq \hat{\theta} + t_{\alpha/2, f} \hat{\sigma}(\hat{\theta})$$

- It is very difficult to obtain the unbiased estimate of $\hat{\sigma}^2(\hat{\theta})$.
- This relation involves three parameters, estimator for mean, estimator for variance, and the degree of freedom. How to determine these values?
- Estimator for mean is calculated as above as a point estimator

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n Y_i$$

- Estimator for the variance and for the degree of freedom has to consider two separate cases

- If Y_i 's are statistically independent observations then use

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \hat{\theta})^2$$

to calculate

$$\hat{\sigma}^2(\hat{\theta}) = \frac{S^2}{n} \text{ with the degree of freedom } f = n - 1.$$

- o If Y_i 's are not statistically independent, then the above estimator for variance is biased. Y_i 's is an autocorrelated sequence, sometimes called a time series. In this case,

$$\hat{\sigma}^2(\hat{\theta}) = \text{var}(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \text{cov}(Y_i, Y_j)$$

i.e. one needs to calculate co-variance for every possible pair of observations.

- o If the simulation is long enough to have passed the transient phase, the output is approximately covariance stationary. That is Y_{i+k} depends on Y_i in the same way as Y_k depends on Y_1 . For a covariance stationary time series Y_i 's, define the lag k auto covariance by

$$\gamma_k = \text{cov}(Y_i, Y_{i+k}) = \text{cov}(Y_i, Y_{i+k})$$

- o If a time series is covariance stationary, then the calculation of sample variance can be substantially simplified.

$$\hat{\sigma}^2(\hat{\theta}) = \frac{y_0}{n} \left[1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) p_k \right]$$

- o So all we need is to calculate covariance between one sample and every other samples, but not every sample with every other samples.

10.5 Transient vs. Steady-state Behaviour

- A typical simulation exhibits two patterns of behavior. During the transient phase, the simulation activity is influenced by the initial state of the simulation model, which is usually chosen in an arbitrary fashion. In the steady-state phase, the model behavior is not influenced by the initial state, and the probability distributions for the model's outputs are no longer functions of the simulation time. The transient phase may be brief or long, depending on how close the chosen initial state is to the model's steady-state behavior and how fast the system "move" towards steady state.
- To be more precise, let $f(X_1; X_2; \dots; X_n)$ be a (simulation) time-ordered sequence of random variables corresponding to model outputs, and let $F_i(x)$ be the cumulative distribution function for X_i . The model has a steady-state phase if and only if

$$\lim_{i \rightarrow \infty} \Pr[X_i \leq x] = \lim_{i \rightarrow \infty} F_i(x) = F(x)$$

where $F(x)$ is the cumulative distribution function for the limiting or steady-state random variable X .

Although sometimes we are interested in the system behavior during the transient phase, we normally want to study the system in steady state. This presents us with two problems in dealing with the simulation outputs.

- (1) How do we determine when the model has reached its steady-state behavior, so that the output random variables (the model measurements) are identically distributed?
 - (2) How do we compensate for correlation among the X_i 's?
- Question (1) arises because $F_i(x) \neq F(x)$ for $i \neq j$ and either X_i or X_j an observation from the transient phase. Question (2) is due to the dependence of successive observations. For instance, if we are measuring the number of messages queued for transmission at an output port of a switch in a computer network; knowledge that the i th observation is large (small) tells us that it is highly probable that the next observation will also be large (small).
- This is important because standard statistical techniques for judging how close an estimator is likely to be to the characteristic it is trying to estimate only deal with sequences of independent, identically distributed random variables.

The transient phase

- In a study of steady – state behavior, we want to omit those early elements of $\{X_i\}$ such that $F_i(x) \neq F(x)$. One possible approach would be to run the simulation many times (say, m times), and for each value of i , compute a normalized histogram of the m values of X_i , one value from each simulation run. This approach is illustrated in Fig. 10.5.1 x_jⁱ is the j th sample taken from the i th simulation run. The histogram for X_i is computed using all the samples in the i th column.
- The histogram for one column might look like Fig. 10.5.2. Each vertical bar represents a sample value or set of sample values, usually referred to as a bin. The height of the bar is equal to the number of samples in that bin, normalized by the total number of samples. If you were to take more samples and make the bin sizes smaller (i.e. make the number of sample values in each set smaller), the histogram might look like Fig. 10.5.3.
- Generally, we expect the new histogram to resemble the old, only more refined. In Fig. 10.5.3, the histogram of Fig. 10.5.2 has been rescaled and shown in gray to illustrate the similarities between the two.

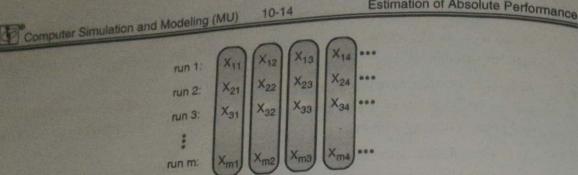


Fig. 10.5.1 : Sets of identically distributed cross-simulation samples from m independent simulations

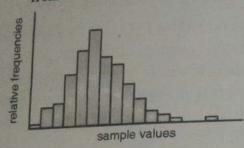


Fig. 10.5.2 : Normalized histogram of i th cross-simulation samples

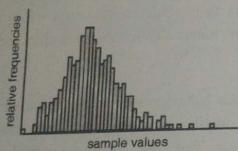


Fig. 10.5.3 : Normalized histogram with additional samples, smaller bin sizes

If the number of runs is very large and the bin size for the histogram is suitably chosen, the histogram, appropriately scaled, should closely approximate the probability density function f_i of X_i . The histograms for the sequence of X_i s could be compared to determine when the distributions (or, as shown here, the density functions) approximately converge. However, this approach is generally much too expensive computationally because of the large number of trials necessary to approximate closely a density function by a histogram.

- An alternative procedure is to select one or more important characteristics of f_i and observe it (them) as a function of i , testing for convergence. For example, if $\mu_i = E[X_i]$, we might wish to examine the sequence $\{\mu_1; \mu_2; \dots\}$. It will converge to $\mu = E[X]$ if the distribution functions converge.
- To apply this to finding the end of the transient phase, we assume that we can reverse the implication; i.e., convergence of the means implies convergence of the distribution

functions. Although this cannot be rigorously defended, it is an adequate heuristic for most cases.

- There is a practical problem with implementing a test for the end of the transient phase by looking for convergence of the means: we don't know what the means are. The best that we can do is to construct unbiased estimators of the means μ_i and then look for convergence of the sample means. We must assume (hope) that convergence of the sample means implies that the true means converge, and that in turn this implies that the distributions converge.
- The following is a more detailed description of this ad hoc procedure for finding the end of the transient phase. Run m simulations, with independent initial states.
- Let X_{ij} be the j th output of the i th simulation run, as before. Compute sample means for each set of random variables $X_{:j}$ across the m simulation runs (not outputs from within the same simulation run).

$$X_j = \frac{1}{m} \sum_{i=1}^m X_{ij}$$

- The sequence of distribution means $\{\mu_1; \mu_2; \dots\}$ will be approximated by the sequence of sample means $\{X_1; X_2; \dots\}$. Plot the sequence of sample means as a function of the sample number. Look for a point on the plot at which the sample means appear to have stabilized around some relatively constant value.
- Fig. 10.5.4 is a plot of sample means taken from a series of simulations of a model of a multiprogrammed computer system. The cross-simulation sample mean of the response time at the CPU is plotted against the sample number. The plot suggests that by about the 60th sample, the simulations have reached the end of the transient phase.
- Determining the end of the transient phase may not be quite so clear-cut in some simulations. The X_j s are random variables with unknown distributions, including unknown means $\mu_j = E[X_j]$ and unknown variances.
- Consequently, the computed values for the random variables $\{X_j\}$ fluctuate about the converging sequence of means $\{\mu_j\}$. Even though the sample mean has the lowest variance of all linear estimators of the mean, the random fluctuations may be large enough to make detecting the end of the transient phase difficult.
- Variance-reduction techniques can be used to reduce the problem. For example, as m increases, the variance of each X_j decreases, and thus the fluctuation of the sample means about $\{\mu_j\}$ should also tend to decrease. The reduction in variance is fairly slow with increasing number of runs. It may take dozens or even hundreds of runs to smooth the sequence of cross-simulation sample means sufficiently. If the simulation reaches steady-state fairly quickly, this may not be a problem.

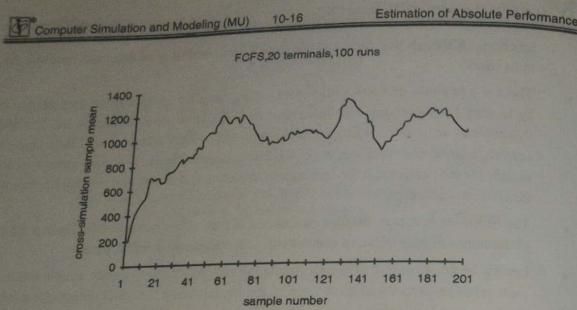


Fig. 10.5.4 : Cross-simulation sample means for multiprogrammed computer system simulation : FCFS, 20 terminals, 100 runs

- Alternatively, smoothing (removing high frequency components in $\{X_i\}$) may make establishing convergence easier. One way of doing this is to use a moving average with appropriate adjustments for the first K elements of the sequence, or only compute the moving averages beginning with the $(K + 1)$ st sample mean. This replaces each point by the average of itself and the K sample points on either side of it.
- Use of this or any other smoothing technique should be done cautiously, however; too much smoothing may obscure the transient phase data.

$$\hat{X}_j(K) = \frac{1}{2K+1} \sum_{k=-K}^{+K} X_{j+k}$$

10.6 Output Analysis of Terminating Simulation

Q. Discuss the Output analysis of terminating simulations.

MU Dec 04

- Consider a terminating simulation runs over a simulated time interval $[0, T]$.
- Consider n as sample size and Y_1, Y_2, \dots, Y_n be the output observations
- A common goal is to estimate

$$\theta = E \left(\frac{1}{n} \sum_{i=1}^n Y_i \right)$$

Computer Simulation and Modeling (MU) 10-17 Estimation of Absolute Performance

$$\text{or } \phi = E \left(\frac{1}{T_e} \int_0^{T_e} Y(t) dt \right)$$

Using different random stream, the simulation is repeated a total of R times.

- Let Y_n be the observation within replication r then, Y_{r1}, Y_{r2}, \dots are correlated.
- But for different replications r and s , Y_{ri} and Y_{sj} are statistically independent for all i and j .

$$\text{Let } \hat{\theta}_r = \frac{1}{n_r} \sum_{i=1}^{n_r} Y_{ri}$$

- Here all the samples $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_R$ are statistically independent and identically distributed and are the unbiased estimators of θ .
- We can now finally apply the classical methods of interval estimation.

10.6.1 Statistics Review

- Let $\{Y_1, Y_2, \dots, Y_n\}$ be the statistically independent observations.
- Then, sample mean $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n Y_i$
- Sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 = \frac{\sum_{i=1}^n Y_i^2 - n\bar{Y}^2}{n-1}$
- If Y_i are independent and identically distributed, the sample variance S^2 is an unbiased estimator of the population variance $\sigma^2(\theta) = \text{var}(Y_i)$.
- $\hat{\sigma}(\hat{\theta}) = \frac{S}{\sqrt{n}}$ is also denoted by s.e. ($\hat{\theta}$) and is called as the Standard Error of the point estimator $\hat{\theta}$.
- The standard error is the measure of :
 - The precision of a point estimator.
 - The average deviation to be expected between the point estimator and the true mean.

10.6.2 Confidence Interval Estimation for a Fixed Number of Replications

10.6.2.1 Discrete Time Data

- Consider R independent replications of simulation are made.
- If results in :

- Independent estimates that are given by

$$\hat{\theta}_r = \frac{1}{n_r} \sum_{i=1}^{n_r} Y_{ri} \quad r = 1, 2, \dots, R$$

- The overall point estimate, $\hat{\theta}$ is given by

$$\hat{\theta} = \frac{1}{R} \sum_{r=1}^R \hat{\theta}_r$$

- The estimate of the variance is given by,

$$\hat{\sigma}^2(\hat{\theta}) = \frac{S^2}{n} = \frac{1}{(R-1)R} \sum_{r=1}^R (\hat{\theta}_r - \hat{\theta})^2$$

- A 100 $(1-\alpha)\%$ confidence interval with $f = R - 1$ degree of freedom is given by,

$$\hat{\theta} - t_{\alpha/2, f} \hat{\sigma}(\hat{\theta}) \leq \hat{\theta} \leq \hat{\theta} + t_{\alpha/2, f} \hat{\sigma}(\hat{\theta})$$

- The standard error of the point estimator $\hat{\theta}$ is given by

$$\hat{\sigma}(\hat{\theta}) = \sqrt{\hat{\sigma}^2(\hat{\theta})}$$

- As R increases the standard error becomes smaller and finally reduces to zero.

10.6.2.2 Continuous Time Data

- Consider the output data, $\{Y_r(t), 0 \leq t \leq T_E\}$ with $r = 1, 2, \dots, R$ independent replications.

- Then it results in :

- Independent estimates those are given by,

$$\hat{\phi}_r = \frac{1}{T_E} \int_0^{T_E} Y_r(t) dt, \quad r = 1, 2, \dots, R$$

- The overall point estimate is given by,

$$\hat{\phi} = \frac{1}{R} \sum_{r=1}^R \hat{\phi}_r$$

- The estimate of variance is given by :

$$\hat{\sigma}^2(\hat{\phi}) = \frac{1}{R} \sum_{r=1}^R \frac{(\hat{\phi}_r - \hat{\phi})^2}{R-1}$$

The sample variance is essentially a measure of error. In statistics, we use its square root as the standard error.

The confidence interval is the same as in case of discrete data.

10.6.3 Interval Estimate with Specified Precision

- The half-length (h.i.) of a 100 $(1-\alpha)\%$ confidence interval for a mean θ , based on the t distribution, is

$$\text{h.i.} = t_{\alpha/2, R-1} * \hat{\sigma}(\hat{\theta}) (*)$$

where $\hat{\sigma}(\hat{\theta}) = S/\sqrt{R}$, S is the sample standard deviation, R is the number of replications.

- Assume an error criterion ϵ is specified with a confidence level $1 - \alpha$, it is desired that a sufficiently large sample size R be taken such that

$$P(|\hat{\theta} - \theta| < \epsilon) \geq 1 - \alpha$$

- Since we have the relation (*), the desired the error control condition can be written as

$$\text{h.i.} = \frac{t_{\alpha/2, R-1} S_0}{\sqrt{R}} \leq \epsilon$$

- Solve the above relation, we have

$$R \geq \left(\frac{t_{\alpha/2, R-1} S_0}{\epsilon} \right)^2$$

since $t_{\alpha/2, R-1} \geq z_{\alpha/2}$ the above relation can be written.

$$R \geq \left(\frac{z_{\alpha/2} S_0}{\epsilon} \right)^2$$

For $R \geq 50$, $t_{\alpha/2, R-1} \approx z_{\alpha/2}$ the inequality with standard normal distribution holds.

- This says we need to run that many (R) replications to satisfy the error requirement.
- The true value of θ is in the following range with probability of $100(1 - \alpha)\%$.

10.6.4 Confidence Intervals for Quantiles

- Here, a proportion or probability is treated as a special case of a mean.
- We consider the number of independent replications Y_1, Y_2, \dots, Y_r to be large enough so that

$$t_{\alpha/2, R-1} = z_{\alpha/2}$$

- So the confidence interval for a probability p is often written as :

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{R-1}}$$

- A quantile is the inverse of the probability to the probability estimation problem.

Steps :

- Obtain θ such that $P_r(Y \leq \theta) = p$.
- To estimate the ' p ' quantile, find such that $100p\%$ of the data in a histogram of Y is to the left of θ or the R_p^{th} smallest value of Y_1, Y_2, \dots, Y_r .
- Obtain an appropriate $(1 - \alpha)$ 100% confidence interval for by calculating :
 - θ_1 that cuts off $100 P_1\%$ of the histogram (the R_{P_1} smallest value of the sorted data).
 - θ_μ that cuts off $100 P_\mu\%$ of the histogram (the R_{P_μ} smallest value of the sorted data).

Where $P_1 = p - z_{\alpha/2} \sqrt{\frac{p(1-p)}{R-1}}$

$$P_\mu = p + z_{\alpha/2} \sqrt{\frac{p(1-p)}{R-1}}$$

10.7 Output Analysis for Steady State Simulation

Q. Discuss Output analysis for steady state simulation. MU - May 05, Dec 05, Dec 08

- The main aim here is to estimate a steady state or long run characteristics of the system.
- Consider a single run of a simulation model. The single run produces observation Y_1, Y_2, \dots which are generally the samples of an autocorrelated times series.

The steady state performance measure is given by

$$\theta = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_i \quad \text{for discrete measure (with probability 1)}$$

$$\theta = \lim_{n \rightarrow \infty} \frac{1}{T_E} \int_0^{T_E} Y_i \quad \text{for discrete measure (with probability 1)}$$

Value of θ is independent of the initial conditions.

- The steady state simulation stops after time T_{SE} , the stopping time.
- The sample size is a design choice, independent of the nature of problem.
- But there are several considerations that are to be kept in mind. These are :
 - Any bias in the point estimator that is due to artificial or arbitrary initial conditions.
 - Desired precision of point estimator.
 - Budget constraints on computer resources.
- Notations : The estimation of θ from a discrete-time output process.
 - One replication, output data : Y_1, Y_2, Y_3
 - With several replications, the output data for replication r : Y_{r1}, Y_{r2}, Y_{r3}

10.7.1 Initialization Bias in Steady State Simulation

Q. Discuss the initialization bias in steady state simulation.

MU - Dec. 06, Dec. 07, Dec. 08, May 08, May 11, Dec. 11, May 14

There are several methods of reducing the point estimator bias which is caused by using artificial and unrealistic initial conditions in a steady-state simulation.

- Initialize the simulation in a state that is more representative of long-run conditions. E.g. use a set of real data as initial condition.
- Divide the simulation into two phases, warm-up phase and steady state phase. Data collection doesn't start until the simulation passes the warm-up phase.

Intelligent Initialization

- Divide simulation into an initialization phase and data collection phase.
- Initialize the simulation in a state that is more representative of long run conditions.
- Give some time to the system to stabilize.

- If the system exists, collect data on it and use these data to specify more nearly typical initial conditions.
- If system does not exist, then use any data to make a simplified model so as to make it mathematically solvable.

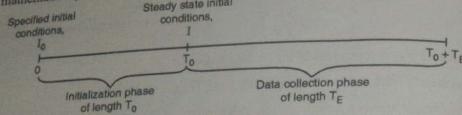


Fig. 10.7.1

Divide each simulation into two phases as shown above :

- An initialization phase, from time 0 to time T_0 .
- A data collection phase from T_0 to the stopping time $T_0 + T_E$.
- The choice of T_0 is important because after T_0 the system should be more nearly representative of steady state behaviour.
- After T_0 the system has finally reached the steady state where the probability distribution of the system state is close to the steady state probability distribution.
- Hence the bias point estimates of response variables are made negligible.
- To decide how much data to be deleted to reduce initialization bias to negligible level, we don't have proven technique. Following point you should keep in mind
 - Cumulative averages become less variable as more data are averaged. Thus, it is expected that the curve at left side (the starting of the simulation) of the plotting is less smooth than the right side.
 - Simulation data, especially from queueing models, usually exhibits positive autocorrelation. The more correlation present, the longer it takes for the average to approach steady state.
 - In most simulation studies the analyst is interested in several measures such as queue length, waiting time, utilization, etc. Different performance measures may approach steady state at different rates. Thus it is important to examine each performance measure individually for initialization bias and use a deletion point that is adequate for all of them.

10.7.2 Replication Method for Steady-State Simulations

- If initialization bias in the point estimator has been reduced to a negligible level, the method of independent replications can be used to estimate point-estimator variability and to construct a confidence interval.
- If significant bias remains in the point estimator and a large number of replications are used to reduce point estimator variability, the result confidence interval can be misleading.
 - The bias is not affected by the number of replications R , but by deleting more data (i.e. increasing T_0) or extending the length of each run (i.e. increasing T_E).
 - Increasing the number of replications R may produce shorter confidence intervals around a "wrong point" $\theta + b$, rather than θ .
- If d is the number of observations to delete from a total of n observations, a rough rule is $n-d$ should be at least $10d$, or T_E should be at least $10T_0$.
- Given the run length, the number of replications should be as many as possible. Kelton in 1986 established that there is little value to run more than 25 replications. So if time is available, make the simulation longer, instead of making more replications.

10.7.3 Batch Means for Interval Estimation in Steady-State Simulations

- Q.** Explain in detail the method of batch means for interval estimation in steady state simulation. MU - Dec. 10, May 12

- One disadvantage of replication is that data must be deleted on each replication.
- One disadvantage of a single-replication is its data tend to be autocorrelated.
- The method of *batch mean* divides the output data from one replication into a few large batches.
- Treat the means of these batches as if they were independent.
- The key issue is that no commonly accepted method for choosing an acceptable batch size m . This is actually one of the research areas in simulation.
 - Schmeiser found for a fixed total sample size there is little benefit from dividing it into more than $k = 30$ batches.
 - Although there is typically autocorrelation between batch means at all lags, the lag-1 autocorrelation $\text{corr}(\hat{Y}_j, \hat{Y}_{j+1})$ is usually studied to assess the dependence between batch means.

- The lag-1 autocorrelation between batch means can be estimated using the method described earlier. They should not be estimated from a small number of batch means, i.e. we need to have large number of batches, though the size of batches could be small.
- If the total sample size is to be chosen sequentially (i.e. choose one for one experiment, choose another one for improvement etc.), then it is helpful to allow the batch size and number of batches to grow as the run length increases.

10.8 Statistical Analysis of Steady-State Behavior

- Once we have eliminated those samples taken during the transient phase of the simulation, we would like to compute estimators for performance metrics based on the remaining samples. Because the estimators are random variables, we also like to have an idea of just how good each estimator is. The standard approach to doing this is to compute an interval, based on the simulation outputs, which has a given probability of actually including the value of the metric being estimated.
- To be somewhat more precise, we compute values for the random variables which are the endpoints of an interval as functions of the probability that such a random interval will include the value of the performance metric. Such an interval is called a confidence interval. One problem we must address is that the outputs of a single simulation run typically are correlated, especially if the samples are taken close together. Because the outputs are not independent, standard statistical techniques for computing confidence intervals cannot be applied.
- There are several methods for dealing with the correlation among the simulation outputs. We will focus on one of the most widely used methods, that of independent replications. Before presenting the method, we will review the basics of generating confidence intervals.

10.8.1 Confidence Levels and Intervals for the Mean

- Suppose that we compute the sample mean of a set of i.i.d. random variables, and we want to know just how likely it is that the mean of the distribution of these random variables is "close" to the sample mean. Let $\{X_1; X_2; \dots; X_n\}$ be the set of n independent and identically distributed random variables. Their sample mean X is an unbiased estimator of the true mean $\mu = E[X_i]$, $1 \leq i \leq n$. We have seen that

$$\text{var}(X) = \frac{\sigma^2}{n}$$

where $\sigma^2 = \text{var}(X_i)$, $1 \leq i \leq n$. Also, recall that s^2 is the sample variance.

If the random variables have normal distributions,

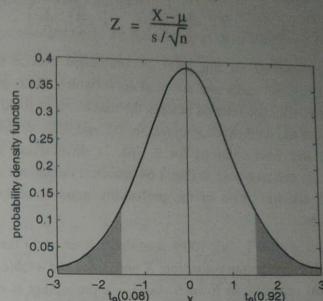


Fig. 10.8.1

- Has a Student's t-distribution with $n - 1$ degrees of freedom (d.o.f.). Even if the $\{X_i\}$ are not normally distributed, Z is approximately t-distributed. For n reasonably large ($n \geq 25$), the t-distribution is very close to the normal distribution with zero mean and unit variance.
- Let $t_{n-1}(y)$ be the $100y$ th percentile of the t-distribution with $n-1$ d.o.f. That is, $t_{n-1}(y)$ is the value z of the random variable Z with an $(n-1)$ -d.o.f. t-distribution F_{n-1} such that $F_{n-1}(z) = \Pr[Z \leq z] = y$; $t_{n-1}(y) = F_{n-1}^{-1}(y)$. Hence,

$$\Pr\left[t_{n-1}\left(\frac{\alpha}{2}\right) \leq \frac{X - \mu}{s / \sqrt{n}} \leq t_{n-1}\left(1 - \frac{\alpha}{2}\right)\right] = 1 - \alpha$$

- The interval between $t_{n-1}(y)$ and $t_{n-1}(1 - y)$ contains a total probability mass of $1 - 2y$; there is y total probability in each of the tails. The t-distribution is symmetric about 0:

$$\begin{aligned} t_{n-1}\left(\frac{\alpha}{2}\right) &= -t_{n-1}\left(1 - \frac{\alpha}{2}\right) \\ \text{Hence, } \Pr\left[t_{n-1}\left(\frac{\alpha}{2}\right) \leq \frac{X - \mu}{s / \sqrt{n}} \leq t_{n-1}\left(1 - \frac{\alpha}{2}\right)\right] &= \Pr\left[-t_{n-1}\left(1 - \frac{\alpha}{2}\right) \leq \frac{X - \mu}{s / \sqrt{n}} \leq t_{n-1}\left(1 - \frac{\alpha}{2}\right)\right] \\ &= \Pr\left[X - t_{n-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}} \leq \mu \leq X + t_{n-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}}\right] \end{aligned}$$

- The quantity $1 - \alpha$ is the confidence level, and the interval
$$\left(X - t_{n-1} \left(1 - \frac{\alpha}{2} \right) \cdot \frac{s}{\sqrt{n}}, X + t_{n-1} \left(1 - \frac{\alpha}{2} \right) \cdot \frac{s}{\sqrt{n}} \right)$$

is the confidence interval for this level of confidence. Notice that the interval is a random interval; i.e., its endpoints are random variables. With probability $1 - \alpha$, the endpoints of the confidence interval bracket the true mean of the distribution. As n gets large, the widths of the confidence intervals for 0.9 and 0.95 confidence levels approach the values 1.64 and 1.96 obtained for a $N(0, 1)$ distribution. That is, 90% of the probability mass is concentrated within 1.64 standard deviations of the mean (0) of a $N(0, 1)$ random variable. 95% of the probability mass is concentrated within 1.96 standard deviations.

- To determine the size of a $(1 - \alpha)$; 100% confidence interval for the mean μ , given the n observations $\{X_i\}$, compute the estimators X and s , and use a table for the t-distribution to find the interval centered on X such that $1 - \alpha$ probability is contained within this interval. This approach is exact if the $\{X_i\}$ are normally distributed; and close under general conditions if they are not, but does require that the random variables be independent.

Example

- Suppose that the following values are obtained from a sequence of i.i.d. random variables :

$$12, 20, 17, 5, 14, 8, 10, 15, 12, 9$$

- From these values compute the sample mean $X = 12.2$ and the sample standard deviation $s = 4.467$. Then

$$\frac{s}{\sqrt{n}} = \frac{4.467}{\sqrt{10}} = 1.413$$

To get a 90% confidence interval ($\alpha = 0.1$) for the unknown mean μ , use the table

- in Appendix A to find

$$t_{n-1} \left(1 - \frac{\alpha}{2} \right) = t_9 (0.95) = 1.833$$

- The 90% confidence interval is

$$(12.2 - 1.833 \times 1.413; 12.2 + 1.833 \times 1.413) = (9.610; 14.790)$$

90% of the confidence intervals that are generated in this way will contain the mean μ .

10.8.2 More on Confidence Intervals

- In some situations, we may be interested in confidence intervals for a distribution parameter θ other than the mean. We can follow the same approach used for the mean. Find a random variable $Y = g(\theta)$ which is a function only of the parameter θ and the values of the samples and which has a known distribution $F_Y(y)$. To compute a $(1 - \alpha) \times 100\%$ confidence interval for θ , determine values y_1 and y_2 such that $F_Y(y_2) - F_Y(y_1) = 1 - \alpha$.

$$\Pr [y_1 \leq Y \leq y_2] = 1 - \alpha$$

or

$$\Pr [g^{-1}(y_1) \leq \theta \leq g^{-1}(y_2)] = 1 - \alpha$$

- For example, we might want to find a confidence interval for the variance of the distribution from which we have obtained a set $\{X_1; X_2; \dots; X_n\}$ of n i.i.d. random variables. If the $\{X_i\}$ are normally distributed with the same mean μ and variance σ^2 , and s^2 is the sample variance of the values of these random variables, then the random variable ns^2/σ^2 has a χ^2 distribution with $n - 1$ degrees of freedom. The χ^2 distribution is not symmetric about zero, as is Student's t-distribution. To get a $(1 - \alpha) \times 100\%$ confidence interval, find two values and χ^2 of a random variable with $n - 1$ degrees of freedom such that probability $1 - \alpha$ lies in the interval $(\chi^2_1; \chi^2_2)$. That is,

$$\Pr \left[\chi^2_1 \leq \frac{ns^2}{\sigma^2} \leq \chi^2_2 \right] = 1 - \alpha$$

This gives us

$$\begin{aligned} \Pr \left[\chi^2_1 \leq \frac{ns^2}{\sigma^2} \leq \chi^2_2 \right] &= \Pr \left[\frac{1}{\chi^2_2} \leq \frac{\sigma^2}{ns^2} \leq \frac{1}{\chi^2_1} \right] \\ &= \Pr \left[\frac{ns^2}{\chi^2_2} \leq \sigma^2 \leq \frac{ns^2}{\chi^2_1} \right] \\ &= 1 - \alpha \end{aligned}$$

- ns^2/χ^2_1 and ns^2/χ^2_2 are the lower and upper endpoints, respectively, of a $(1 - \alpha) \times 100\%$ confidence interval for the variance σ^2 .

Example

- Assume that the numbers listed below are values for ten independent random variables, all of which have the same $N(\mu, \sigma)$ distribution.

10.8248	8.1342	5.3754	6.7580	1.5174
13.4807	5.2953	13.9854	6.3203	9.3584

- The sample variance for these values is about 15.1319. Appendix B gives values for the inverse χ^2 distribution. To compute a 95% confidence interval, we can choose probabilities 0.025 and 0.975, since their difference is 0.95%. For 9 degrees of freedom, the endpoints are $\chi_1 = 2.7004$ (probability 0.025) and $\chi_2 = 19.0228$ (probability 0.975). Hence, a 95% confidence interval for the variance is

$$(10 \times 15.1319 / 19.0228, 10 \times 15.1319 / 2.7004) = (7.9546, 56.0360)$$

or, equivalently, $(2.8204, 7.4857)$ is a 95% confidence interval for the standard deviation. The size of this confidence interval compared to the sample standard deviation (about 3.8900) suggests that we might require many more samples to reduce the interval to a reasonable width.

10.8.3 Independent Replications

- When we try to use confidence intervals for the data obtained from a simulation, we have to deal with the problem that the $\{X_i\}$ obtained from a single simulation run may be (and usually are) correlated. One way to get around this problem is to use the method of independent replications. The method requires the use of $m > 1$ independent simulation runs. The runs are independent of the starting states and the random number generator seeds of the runs are independent.
- Let X_{ij} be the j th steady-state output from the i th simulation run. For each replication (independent simulation run), form the sample mean

$$\hat{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}$$

where n is the number of samples taken from the steady state phase in each simulation run.

- The estimators in the set $\{\hat{X}_1, \hat{X}_2, \dots, \hat{X}_m\}$ are independent random variables with

$$E[\hat{X}_i] = \mu, \quad 1 \leq i \leq m$$

- Hence,

$$\hat{X} = \frac{1}{m} \sum_{i=1}^m \hat{X}_i$$

is an unbiased estimator of μ . Let

$$\hat{s}^2(\hat{X}_1, \dots, \hat{X}_m) = \frac{1}{m-1} \sum_{i=1}^m (\hat{X}_i - \hat{X})^2$$

$$= \frac{1}{m-1} \sum_{i=1}^m \hat{X}_i^2 - \frac{m}{m-1} \hat{X}^2$$

be the sample variance of the set of X_i s. Then $(\hat{X} - \mu/\sqrt{m})$ is approximately t -distributed with $m-1$ degrees of freedom, and therefore

$$\Pr\left[\hat{X} - t_{m-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\hat{s}}{\sqrt{m}} \leq \mu \leq \hat{X} + t_{m-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\hat{s}}{\sqrt{m}}\right] = 1 - \alpha$$

The width of the $(1 - \alpha)$ confidence interval is $2 \times t_{m-1}(1 - \alpha/2) \times (\hat{s}/\sqrt{m})$. For $m \geq 10$, $t_{m-1}(y)$ is virtually constant as a function of m . Also, $E[\hat{s}^2] / (m-1) \propto \sqrt{n/m}$ and hence the expected width of the confidence interval is proportional to $1/\sqrt{nm}$. If each simulation run has no initial samples in the transient phase before the n steady state samples, then the total number of samples from all simulation runs is $m(n_0 + n)$, of which $m n_0$ are unused in the steady-state analysis.

- For a fixed α , we can reduce the width of the confidence interval by increasing either m or n ; however, increasing m increases the amount of unused data. Hence, it is more efficient to keep m small and use n to decrease the width of the confidence interval. However, this approach makes managing multiple simulations more difficult. The simpler method is to run several simulations, check to see if the confidence interval is small enough, and if not, run additional simulations until it is.

Example

- A computer system designer ran four simulations of a proposed system and obtained four samples from each simulation. The sample values and corresponding sample means from each of the simulations were

$$7, 2, 3, 4 \Rightarrow \hat{X}_1 = \frac{1}{4} \cdot 16 = 4$$

$$6, 8, 1, 3 \Rightarrow \hat{X}_2 = \frac{1}{4} \cdot 18 = 4.5$$

$$5, 5, 7, 5 \Rightarrow \hat{X}_3 = \frac{1}{4} \cdot 22 = 5.5$$

$$4, 9, 2, 2 \Rightarrow \hat{X}_4 = \frac{1}{4} \cdot 17 = 4.25$$

- Use the sample means from the individual simulations to compute a sample mean and sample variance:

$$\hat{X} = \frac{1}{4} \sum_{i=1}^4 \hat{X}_i = \frac{18.25}{4} = 4.5625$$

$$\hat{s}^2 = \frac{1}{3} \sum_{i=1}^4 (\hat{X}_i - \hat{X})^2 \\ = \frac{(0.36140 + 0.00390 + 0.87890 + 0.9765)}{3} = 0.44728$$

- To obtain a 90% confidence interval ($\alpha = 0.1$), use $t_3(0.95) = 2.3534$. The 90% confidence interval is

$$\left(\hat{X} - t_3(0.95) \cdot \frac{s}{\sqrt{4}}, \hat{X} + t_3(0.95) \cdot \frac{s}{\sqrt{4}} \right) \\ = \left(4.5625 - 2.3534 \cdot \frac{0.66789}{2}, 4.5625 + 2.3534 \cdot \frac{0.66789}{2} \right) \\ = (3.7766, 5.3484)$$

- If the designer wants a higher degree of confidence, she must be willing to accept a larger confidence interval. For example, to reach a 0.95% level of confidence ($\alpha = 0.05$), she must use $t_3(0.975) = 3.1824$. This results in a confidence interval (3.4998, 5.6252).

Review Questions

- Q. 1 Explain point estimation and interval estimation of θ .
- Q. 2 What is Terminating Simulation and Non terminating Simulation give examples ?
- Q. 3 Discuss the Output analysis of terminating simulations.
- Q. 4 Discuss Output analysis for steady state simulation.
- Q. 5 State the effect of initialization bias in steady state simulation and how the effect can be reduced.
- Q. 6 How can the initialization bias be reduced in steady state simulation?
- Q. 7 Discuss the initialization bias in steady state simulation.
- Q. 8 State the effect of initialization bias in steady state simulation and how the effect can be reduced.

Q. 9 How can the initialization bias be reduced in steady state simulation?

Q. 10 Give examples of terminating and non-terminating simulation. Estimate confidence interval.

10.9 University Questions and Answers

May 2010

Q. 1 Explain the following with examples :

- (i) Terminating simulation (ii) Non-terminating simulation.
(Section 10.2)

(10 Marks)

Dec. 2010

Q. 2 Explain in detail the method of batch means for interval estimation is steady state simulation. (Section 10.7.3)
(12 Marks)

Q. 3 What is output analysis ? State when it is used (Section 10.1)
(6 Marks)

May 2011

Q. 4 Explain Naylor and Finger validation approach. (Section 10.7.1)
(10 Marks)

Q. 5 Types of simulations with respect to output analysis. (Section 10.2)
(10 Marks)

Dec. 2011

Q. 6 Describe initialization bias in steady-state simulation. (Section 10.7.1)
(10 Marks)

Q. 7 Distinguish between Terminating and non terminating simulations.
(Section 10.2)
(3 Marks)

May 2012

Q. 8 Explain different types of simulation with respect to output analysis.
(Section 10.2)
(10 Marks)

Q. 9 Explain in detail the method of batch means for interval estimation in steady state simulation. (Section 10.7.3)
(10 Marks)

Q. 10 Write short notes on Terminating and Non-Terminating Simulation.
(Section 10.2)
(5 Marks)

Dec. 2013

Q. 11 Write short note on : Terminating and non terminating simulation.
(Section 10.2)
(5 Marks)