Chapter -7

ANYLISIS OF SIMULATION OUTPUT

In the previous chapters we have learn how to produce a model of a system, how to simulate and produce a computer program for simulating the system such as predator prey system. By which of this program, we can perform the desired experiment on the system being studied and can analyze the output. In those programs, we have a freedom to specify initial condition, run length or replication runs in order to analyze the output simulation.

A large body statistical method has been developing over the years to analyze results in science, engineering and other fields where experimental observation are made. So, because of experimental measure of system of simulation for these statistical methods can be adapted to simulation results to analyze.

The newly developing statistical methodology concerns;

1) To ensure that the statistical estimates are consistent, meaning that as the sample size increases the estimate tends to true value.
2) To control biasing in measure of both new values of variance. Bias causes the distinction of estimate to differ significantly from the true population statistics, even though the estimate may be consistent.
3) To develop sequential testing methods, to determine how long a simulation should be run in order to obtain a confidence in its return.

ESTIMATION METHOD

- Statistical methods are commonly used on random variable. Usually a random variable is drawn from an infinite population with a finite mean 'μ' and finite variance 'σ^2'.
- These random variable are independently and identically distributed (i.e. iidvariables).

Let,x_i=iid random variables. (i=1, 2,......, n), then according central limit theorem and applying transformation, approximate normal variance,

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

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

Where $\bar{x}$= sample mean

- It can be shown to be a consistent estimator for the mean of population from which the sample is done.
- Since the sample mean is some of random variable, it is itself a random variable. So a confidence interval about its computed value needs to be established.
- The probability density function on the standard normal variable (Z) is shown in figure (1).

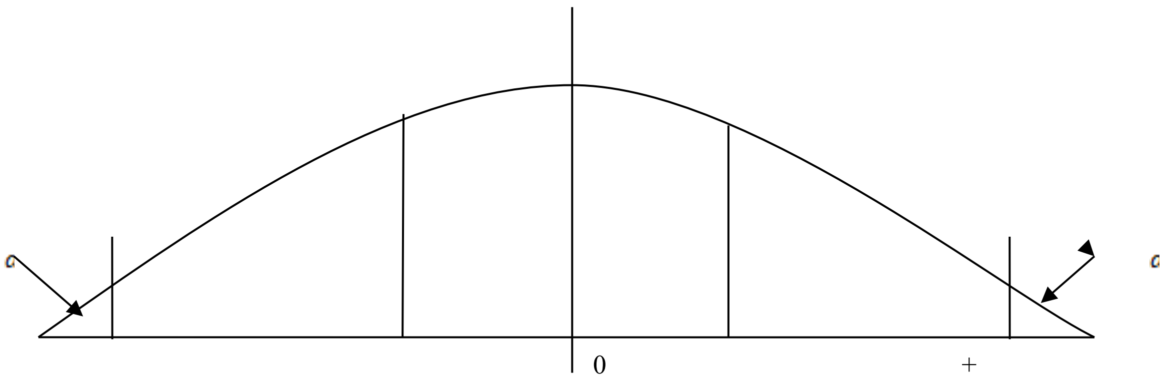


Fig. (1)

- The integral formed $-\infty$ to μ is the probability that Z is less than or equal to μ and is denoted by ϕ (u).
- Let us consider the value of u ($u_{\alpha/2}$) such that ϕ(u) = 1-α/2 where α = some constant < 1.
- Then probability of Z for $Z>u_{\alpha/2}= \alpha/2$.
- Probability of Z for $(-u_{\alpha/2} \leq Z \leq +u_{\alpha/2}) = 1-\alpha$.
- In terms of sample mean μ, the probability statement can be written as
- Prob$\left\{\bar{x}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2} \geq \mu \geq \bar{x}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right\} = 1-\alpha$

Here the constant $1-\alpha$ is a confidence level (usually expressed in %) and the interval $\bar{x} \pm \frac{\sigma}{\sqrt{n}}u_{\alpha/2}$ is the confidence interval.

Estimation population variance s2 (not σ^2 which is actual population variance) is given by:

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

Replacing $u_{\alpha/2}$by $t_{n-2,\alpha/2}$, then the estimated variance s2, the confidence interval for $\bar{x}$ is given by

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

SIMULATION RUN STATISTICS

On every simulation run, some statistic are measure based on some assumption ; for example – on establishing confidence interval it is assume that the observation are mutually independent and distinction from which they are drawn is stationary. But many statistics are interest in a simulation don't meet this condition.

Let us illustrate the problems that arise in measuring statistic from simulation run with the example of single server system.

Consider occurrence of arrivals has a Poisson distribution.

- The service time has an exponential distribution.
- The queuing discipline is FIFO
- The inter-arrival time is distributed exponentially
- System has a single server.

Then in a simulation run, the simplest way to estimate the mean waiting is to accumulate the waiting time of n successive entities and dividing it by 'n'. This gives sample mean denoted by $\bar{x}(n)$.

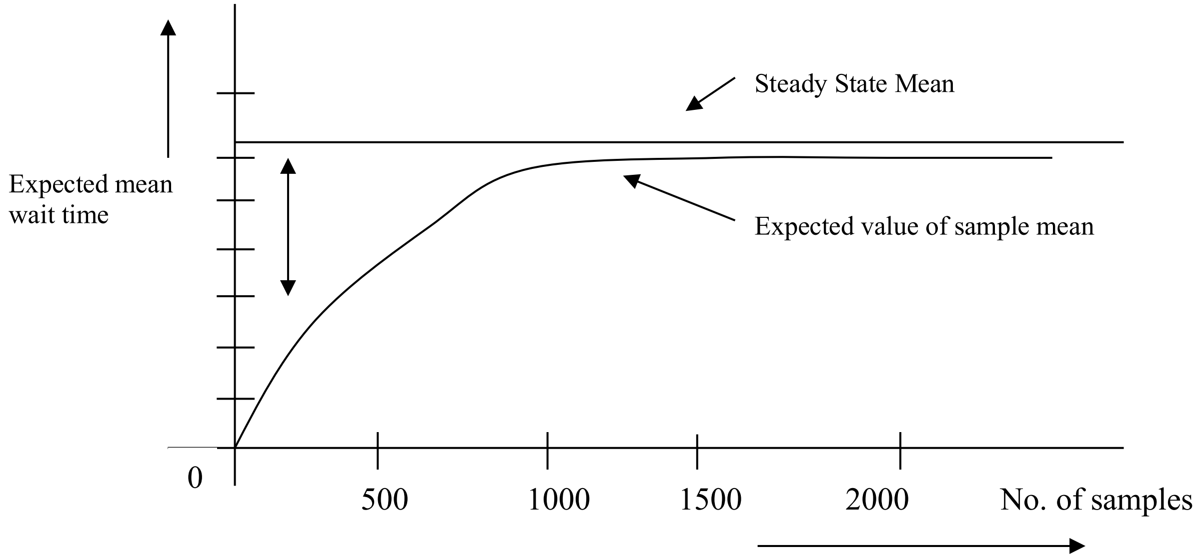
If x_i (i = 1, 2, 3, ……….,n) are the individual waiting times, then

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

The 1st problem, here is that, the waiting times measure this way are not independent because whenever a waiting line forms, the waiting time of each entity on the line depends upon the waiting time of its predecessor (i.e. the entities are auto co-related).

The usual formula for estimating the mean value of the distribution remains on satisfactory estimate for the mean of auto co-related data. However the variance of auto-correlated data is not related to the population variance by simple expression $\frac{\sigma^2}{n}$ as occurs for independent data.

The 2nd problem is that the distribution may not be stationary; it is because 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, thus the early arrivals have a more probability of obtaining service quickly. So a sample mean that includes the early arrivals will be biased. As the length of simulation run extended and the sample size increases, the effect of bias will be minimum. This is shown in below fig.



REPLICATION OF RUNS

One problem in measuring the statistic in the simulation run is that the results are dependent. But it is required, in simulation, to get the independent result. The one way of obtaining independent result is to repeat the simulation.

Repeating the experiment with different random numbers for the same sample size 'n' gives a set of independent determination of sample mean $\bar{x}(n)$.

Even though the distribution of sample mean depends upon the degree of auto correlation, this independent determination of sample mean can be used estimate the variance of distribution.

Suppose,

- Experiment is repeated p-times with independent random numbers.
- $x_{ij} = i^{th}$observation of j^{th} run. Then estimates for:
- Sample mean $\bar{x}_j(n) = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$
- Then estimate for variance for j^{th} run $s^2{}_j(n)$ is given by

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

Now, combining the results of p independent measurement gives the following estimate for the mean waiting time $\bar{x}$, and variance s^2:

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

$$s^2 = \frac{1}{p}\sum_{j=1}^{p} s^2{}_j\,(n)$$

Here, the value of $\bar{x}$ is an estimate for mean waiting time and the value of s^2 can be used to establish a confidence of intervals.

ELIMINATION OF INITIAL BIAS

There two general approaches that can be used to remove the initial bias;

1) The system can be started in more representative states rather than in the empty state.
2) The first part of the simulation run can be ignored.

- In the first approach, it is necessary to know the steady state distinction for the system and we then select the initial state distinction. In the study of simulation, particularly existing system, there may be information available on the expected condition which makes it feasible to select better initial condition and thus eliminating the initial bias.
- The second approach that is used to remove the initial bias is the most common approach. In this method, the initial section of the run which has highly bias (simulation) result is eliminated. First, the run is started from an idle state and stopped after a certain period of time (the time at which the bias is satisfactory). The entities existing in the system at that are left as they are and this point is the point of restart for other repeating simulation run. Then the run is restarted with statistics being gathered from the point of restarted. These approaches have following difficulties:

1) No simple rules can be given to decide how long an interval should be eliminated. For this we have to use some pilot run starting from the ideal state to judge how long the initial bias remains. These can be done by plotting the measured, statistics against the run length.

2) Another disadvantage of eliminating the first part of simulation run is that the estimate of variance will be based on less information affecting the establishing of confidence limit. These will then cause to increase in confidence internal size.