# UNIT-7

## Analysis of Simulation Output

**Q. Why do we perform the analysis of simulation output?**

**Ans:** We perform the analysis of simulation output to predict system performance or compare performance of two or more system designs. It is needed because output data from a simulation show random variability when random number generators are used i.e, two different random number streams will produce two sets of output which probably will differ. The output analysis determines:

→ The estimate of the mean and variance of random variables.

→ The number of observations required to achieve a desired precision of these estimates.

### Confidence Interval:

Confidence interval is a range of possible values that is likely to capture an unknown parameter, given a certain degree of probability (confidence). Confidence interval in short can be denoted by CI and given by the formula:

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

where, $\bar{x}$ = sample mean

$z$ = confidence level value

$s$ = sample standard deviation.

$n$ = sample size.

Confidence level % = $1-\alpha$. Alpha ($\alpha$) is known as the significance level or accepted error. An $\alpha = 0.05$ is typically a good level of accepted risk, but varies depending on the situation.

**I.I.D:** Infinite population has a stationary distribution with a finite mean $\mu$ and finite variance $\sigma^2$. Sample variable and time does not affect population distribution. Random variables that meet all these conditions are called independently and identically distributed (I.I.D).

## ⊛. Hypothesis Testing:

Hypothesis is an assumption or claim or statements. Hypothesis testing is a method by the help of which we are able to test whether the hypothesis is true or not.
There are two types of hypothesis:

i) **Null Hypothesis ($H_0$):** It is symbolized by $H_0$ which is known or accepted fact. It states that there is no statistical significance between two variables and is usually what we are looking to disprove.

For Example: $H_0 = 0$
i.e, there is no significance difference between two hypothesis

ii) **Alternative Hypothesis ($H_1$):** Alternate hypothesis is symbolized by $H_1$. It is the opposition of the null, and is what we are testing for statistical significance.

Example: $H_1 \neq 0$.
i.e, There is significance difference between two hypothesis.

## ⊛. Estimation Methods:

Consider the estimation of a performance parameter, $\theta$ of a simulated system.

→ Discrete time data: $(Y_1, Y_2, \ldots, Y_n)$ with ordinary mean: $\theta$

→ Continuous-time data: $\{Y(t), 0 \leq t \leq T_E\}$ with time weighted mean: $\phi$

## ⊛. Point estimation for discrete time data:

↳ The point estimator is $\hat{\theta} = \frac{1}{n} \sum_{t=1}^{n} Y_t$.

↳ It is unbiased if its expected value is $\theta$,
i.e, if: $E(\hat{\theta}) = \theta$

↳ It is biased if: $E(\hat{\theta}) \neq \theta$ and $E(\hat{\theta}) - \theta$ is called bias of $\hat{\theta}$.

## ⊛. Point estimation for continuous-time data:

↳ The point estimator is: $\tilde{\phi} = \frac{1}{T_E} \int_0^{T_E} Y(t)\, dt.$

↳ It is biased in general where: $E(\hat{\phi}) \neq \phi.$

↳ An unbiased or low-bias estimator is desired.

## ⊛. Interval estimation/Confidence Interval estimation:

Suppose the model is the normal distribution with mean $\theta$, variance $\sigma^2$ and we have a sample of $n$ size, then the variance of sample data is:

$$s^2 = \frac{1}{(n-1)} \sum_{t=1}^{n} (y_t - \bar{y})^2$$

where, $\bar{y}$ is sample mean.

The confidence interval of $t$-distribution with $n-1$ degree of freedom, estimated variance $s^2$, for $\bar{x}$ is defined by:

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

Here the quantity $t_{n-1,\alpha/2}$ is found in $t$-distribution table.

Example: The daily production time of a product in a factory for 120 days is 5.8 hours and sample standard deviation (s) is 1.6. Calculate confidence interval for 95%.

Solution:

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

$$= 5.8 \pm \frac{1.6}{\sqrt{120}} \times 1.98$$

$$= 5.8 \pm 0.29$$

since, $t_{120-1,0.05/2}$ i.e, $t_{119,0.025} = 1.98$

Hence, the estimates between $5.8 \pm 0.29$ can be accepted for 95% confidence interval.

## ⊛. Simulation run statistics: [Imp]

In the estimation method, it is assumed that the observations are mutually independent and the distribution from which they are drawn is stationary. Unfortunately many statistics of interest in simulation do not meet these conditions. An example of such case is queuing system. Correlation is necessary to analyze such scenario. In such cases, simulation run statistics method is used.

**Example:** Consider a system with Kendall's notation M/M/1/FIFO and the objective is to measure the mean waiting time. In simulation run approach, the mean waiting time is estimated by accumulating the waiting time of n successive entities and then it is divided by n. This measures the sample mean such that:

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

Whenever a waiting line forms, the waiting time of each entity on the line clearly depends upon the waiting time of its predecessors. Such series of data in which one value affect other values is said to be autocorrected. The sample mean of autocorrected data can be shown to approximate a normal distribution as the sample size increases.

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 early arrivals will be biased.

**Q.** Why confidence interval is needed in the analysis of simulation output? How can we establish a confidence interval? [Imp]

**Ans:** In the analysis of simulation output, confidence interval is needed because it provide us with an upper and lower limit on sample mean, and with this interval we can then be confident we have captured the population mean. The lower limit and upper limit around our sample mean tells us the range of values that our true population mean is likely to lie within.

We can establish confidence interval as follows:

i) Identify a sample statistic (e.g, sample mean, same proportion) that we will use to estimate a population parameter.

ii) Select a confidence level. We can choose 90%, 95%, 99%.

iii) Find the margin of error. It is calculated as;

margin of error = Critical value * Standard deviation

OR margin of error = Critical value * Standard error.

iv) Specify the confidence interval as;

Confidence interval = sample statistic ± Margin of error.

## ⊛. Replications of Runs:

One way of obtaining independent result is to repeat simulation. Repeating of the experiment with different random numbers for the sample size n gives a set of independent determination of sample mean $\bar{x}$. Suppose the experiment is repeated p times with independent random values of n sample sizes. Let $x_{ij}$ be the ith observation in jth run and let the sample mean and variance for the jth run is denoted by $\bar{x}_j(n)$ and $s_j^2(n)$ respectively. Then for jth run, the estimates are;

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

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

Combining the the result of p independent measurement gives the following estimate for mean $\bar{x}$ and variance $s^2$ of population as:

$$\bar{x} = \frac{1}{p}\sum_{j=1}^{p}\bar{x}_j \qquad \text{and} \qquad s^2 = \frac{1}{p}\sum_{j=1}^{p}s_j^2$$

## ⊛. Elimination of initial bias:

To remove the bias two general approaches can be taken:

i) The system can be started in a more representative state than the empty state.

ii) The first part of the simulation run can be ignored.

In some simulation studies where the information about expected value is available, it is feasible to select better initial conditions. The ideal solution is to know the steady state distribution for the system and select the initial conditions from that distribution. The more common approach to remove initial bias is to illuminate an initial section of run.

The run is started from an idle state and stopped after a certain period of time. The entities existing in that system at that time are left as they are. The run is then restarted with the statistics being gathered from the point of restart. No simple rules can be given to decide how long an interval would be eliminated. It is advisable to use some pilot runs starting from idle state to judge how long the initial bias remains.