

## SO YOU HAVE YOUR MODEL: WHAT TO DO NEXT A TUTORIAL ON SIMULATION OUTPUT ANALYSIS

Martha A. Centeno  
M. Florencia Reyes

Industrial and Systems Engineering Department  
Florida International University  
Miami, Florida 33199, U.S.A.

### ABSTRACT

Simulation models are built with the intent of studying the behavior of the real system represented by the model. However, a simulation model generates random outputs; thus, the data generated by it can only be used to *estimate* the true measure of performance. In this tutorial, we introduce several concepts and techniques to analyze such output. Additional examples will be given during the presentation of the tutorial.

### 1 INTRODUCTION

Simulation modeling enables the study of the stochastic behavior of systems, the testing of hypotheses that account for the observed behavior, and the use of these theories to predict future behavior. The simulation modeling methodology has several stages that begin with defining the objective of the study, model abstraction, model verification, and model validation. At the end of model validation, you finally have a working model. However, your job as a systems analyst is far from complete.

The actual analysis of the data provided by the simulation modeling will always depend on the initial objectives of the study, and on the type of system being modeled (Centeno, 1996; Sadowski, 1993). There are two main types of systems: *terminating* and *non-terminating*. Terminating systems have a natural *starting point* (operations begin) and a natural *ending point* (operations end), whereas non-terminating systems have a natural beginning, but they do not have a natural ending (Law, 1990). Typical terminating systems are most fast food restaurants, dental clinics, department stores, public-oriented government offices, and the stock market. There are other systems for which it is not obvious that they are terminating, but they are. For instance, a company producing the external tank for the space shuttle may be interested in studying several configurations to produce 12 tanks in 6 months. In this case, the system is terminating

because the natural ending of it is the production of the last external tank requested. Typical non-terminating systems are emergency rooms, some operations at hospitals, airports at large cities, and petrochemical plants. A manufacturing company that only works one shift may still be considered a non-terminating system if the ending conditions for the shift are the initial conditions for the next shift (Law, 1990).

Depending on the type of objectives, the analysis may call for the comparison of various alternatives, or for the thorough analysis of the behavior of the systems under a specific configuration, or a quick analysis of a factor that may affect the performance of the system.

Every simulation model has various components, including *dynamic entities*, *resources*, and the *state of the system*. For the beneficiary of the simulation analysis, the important component is the state of the system because it is the collection of variables needed to describe the system's performance. An introduction to the analysis of these variables is the focus of this paper.

Section 2 discusses the importance of experimental design. A detailed example will be given when the tutorial is presented. Section 3 explains how to analyze terminating systems, whereas section 4 explains how to analyze non-terminating systems. Section 5 discusses how to compare several alternatives. Section 6 briefly presents what to do when dealing with a single replication. Finally, section 7 summarizes the tutorial.

### 2 EXPERIMENTAL DESIGN

When simulating a system, you have to think ahead about the scenarios of the system that you would like to evaluate using the simulation model (Kelton, 1994). In some cases, the scenarios are a natural consequence of the objectives and expectations set by management. In other cases, you need to find the right combination of the input parameters. If the number of inputs is small, exhaustive enumeration may be an easy and inexpensive way of finding the right

combination. But if the number of input factors is large, you should use some form of design of experiments to reduce the size of the search space (combinations to explore).

In the context of simulation, the *factors* are the various inputs to the model, the *levels* are the various options for each input parameter, and the *responses* are the outputs of the simulation model. Once these elements are clearly identified, you can analyze the experimental design in terms of measuring effects of the factors and the interactions among them. Donohue (1994) summarizes the various steps of using a statistical approach to the design and analysis of experiments, which also address tactical and strategic issues. A modified version of these steps is as follows:

1. Choose the factors (controllable input variables) and the response variables (uncontrollable output variables).
2. Define the region of operability (factor levels) and the regions of interest.
3. Select proper statistical analysis (ANOVA, regression, etc.), criteria to choose a "best" design, and appropriate experimental class design (e.g. factorial, Latin square, etc.)
4. Perform experiments, collect data, and analyze and summarize it. Incorporate here decisions regarding the type of system (terminating, non-terminating), simulation length, initial conditions, number of replications, and random seeds.
5. Draw inferences and conclusions

Your choices for the experiment design include complete factorials, fractional factorials, artificial factors, frequency domains, and correlated factors among others. Regardless of the chosen design, once you have collected the outputs, you must turn your attention to the response metamodels. The type of objectives set forth will dictate the selection of the appropriate metamodels (Donohue, 1994). At this stage, one may be interested in sensitivity analysis, in prediction, or in optimizing the response variables. Thus, we are interested in devising a model that characterizes the behavior of the system as a function of its outputs. Regression is commonly used to fit the behavior of the average of the performance measure and the variance of it. To add stability to the analysis, logarithmic transformations may be used (Sanchez, 1994).

Kelton (1994) provides an extensive list of references that give more details on designing the experiments. Sanchez (1994) and Donohue (1994) provide detailed examinations of the subject.

### 3 TERMINATING SYSTEMS

Terminating systems are systems that have a clear point in time when they start operations and a clear point in time when they end operations. For this type of systems, it is necessary to decide two things: the sample size and the simulation length. The simulation length is typically established by the context of the problem. For a car rental operation, it may be an entire day of operations, or it may just be the morning rush. The sample size is established based on the accuracy, reliability, and variation desired for the study, using the equation (1)

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{d^2} \quad (1)$$

where  $d$  is the accuracy expressed in the same units as those of the measure of performance (e.g. within 2 unit),  $z$  is the critical value from the standard normal table at a given reliability level,  $1 - \alpha$ , (e.g. 95% reliability yields  $\alpha = 0.05$ ), and  $\sigma$  is the standard deviation desired.

The resulting value of  $n$  is the minimum number of *replications* (not runs) needed to obtain statistically valid results. It is very common for the novice to confuse a replication with a simulation run. A *run* is what happens from the moment the user clicks on the run option of the main menu to the moment in which the software finishes outputting data and comes back to the main menu. A *replication*, on the other hand, is what happens from the simulated start time to the closing simulation time. In other words, *replication* is the repetition of a simulation with fixed inputs but different outputs due to different random numbers replications (Centeno, 1996; Clark, 1988). For a terminating system, a simulation run has  $n$  replications.

It is important to point out that  $n$  refers to the number of replications and not to the number of observations per replication. Why is this? Very simple! Assume that the measure of performance of interest is the time in the system ( $X_i$ ), and that that you ran the simulation model so that 75 customers were processed. At the end of the replication run, there would be 75 values of  $X_i$ , one for each customer. These values would have been used to establish the replication's average time in the system ( $\bar{X}_{within}$ ).  $\bar{X}_{within}$  is an unbiased estimator of the true measure of performance's average ( $\alpha$ ), but the variance of the observations within the replication ( $S_{within}^2$ ) is biased because the  $X_i$  are not necessarily independent and identically distributed random variables; thus, it may happen that  $\text{var}(\bar{X}_{within}) \neq \text{var}(X_i) / 75$ . To avoid this, we use the method of independent replications (Goldsman, 1992). This method yields **one observation** per replication. So, at the end of the first replication where 75

customers were processed, you have one statistically independent observation of the time in the system ( $Y_i = X_{within}$ ). Usually, terminating systems are analyzed using the method of independent replications. However, there are instances in which this method may be inefficient to use (see section 6).

Let us look at an example of the method of independent replications. Supposed that there is a driver licensing office that has seven officers to examine your driving skills. Customers arrive according to a normal distribution with mean 5 minutes and a standard deviation of 1 minute. It takes approximately 25 minutes to take the exam, normally distributed with standard deviation of 3.2 minutes. Further assume that you have generated a model of it, verify it and validate it. These are the things that you need to do now:

Establish the measures of performance for the analysis. From your objectives, you already know which measures of performance are important. Supposed that you are only interested on the average time in the system.

Decide the type of accuracy and reliability that you seek. Supposed that you want a 95% reliability ( $\alpha = 0.05$ ), and an accuracy that yields an ideal half width confidence interval ( $h$ ) equal to 3.

Run the model for a small number of replications. Depending on the size of the model, and the time it takes to execute it, small may mean 5, 10 or 15 replications. For this example, small means 10 replications. After running your model, you will obtain outputs similar to those given in Table 1.

Table 1: Output from a Terminating System

Replica- tion	Average Tsys	Replica- tion	Average Tsys
1	257.43	6	269.16
2	264.96	7	250.09
3	254.22	8	248.54
4	265.84	9	254.31
5	252.75	10	257.10

Compute a (1- $\alpha$ )% confidence interval. This is done for the average of the measure of performance of interest, using equation (2).

$$\bar{x} - t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}} \leq \alpha \leq \bar{x} + t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}} \quad (2)$$

where  $t_{n-1, 1-\alpha/2}$  is the (1- $\alpha/2$ ) percentile of the t-student distribution with n-1 degrees of freedom. The example data yields

$$\bar{x} = 253.44 \quad s = 3.21 \quad t_{10-1, 1-0.05/2} = 2.26$$

$$252.44 \leq \alpha \leq 262.44$$

which means that the half width ( $h$ ) is equal to 5. Recall that you are trying to answer the question of *how many replications to make, so that you obtain the appropriate resolution.* In other words, you are trying to find out if  $h < h^*$ . If after doing the preliminary run, the answer is yes, you are done. On the other hand, if the answer is no, you need to compute  $n^*$  using equation (3). This equation is the same as (1), but expressed in terms of the half widths of the confidence interval.

$$n^* = \text{Round} \left[ n \cdot \left( \frac{h}{h^*} \right)^2 \right] = 28 \quad (3)$$

Modify simulation model to reflect the ideal  $n^*$ .

Change the number of replications, so that you can execute a full production run, and run the model.

Change the initial seed of the random number streams.

There are a couple of things that you need to decide before executing the production run.

You already ran 10 replications, but you need 28 replications. The questions that need to be answered are *Do you run the model again for 28 replications?*, or *Do you run the model for only 18 more replications?* If you run the model for 28 replications, the first 10 replications will be exactly the same 10 replications plus 18 new ones. This implies that you wasted some time. The first 10 replications are exactly the same because the random number streams begin exactly at the same point every time you run the model. So, to avoid this situation, what you need to do is to change the initial seed of the various random number streams that your model uses.

Computes the final (1- $\alpha$ )% confidence interval as before. This yields the final confidence interval of the various measures of performance.

The method of independent replications can be used to build confidence intervals for statistics other than the mean value of the measure of performance. Other statistics of interests are the various percentiles. More details and examples can be found in Banks, Carson, and Nelson (1996).

#### 4 NON TERMINATING SYTEMS

For non-terminating systems the fundamental question to answer is *for how long should the simulation be run?* To answer this question, you need two address two critical issues: 1) achieving steady state conditions, and 2) obtaining statistically independent observations. Once these questions are answered, it would be possible to obtain the confidence intervals as in the case of terminating systems.

As an example, consider the processing of the space shuttle which requires a large number of assembly and repair operations. Technicians working on the shuttle receive the necessary assembly pieces and tools in kits. These kits are customized for the various operations, but