TUTORIAL ON SIMULATION OUTPUT ANALYSIS

Lee W. Schruben
Cornell University

Technical details of various approaches to the analysis of simulation output are presented in other sessions of this conference and have been the topic of previous Winter Simulation Conference sessions. This tutorial will include an introduction to many of these techniques but the focus is on particular problems in translating classical statistical methodologies to the simulation environment.

A simulation experiment is often conducted to estimate some unknown parameter (scalar or vector) on which a decision is to be based. Furthermore the decision is likely to involve considerable risk; otherwise, the relatively expensive and time consuming methodology of computer simulation should not be employed. From this it follows that estimates are of little value unless their accuracy can be assessed. So-called "point estimates," while useful in other phases of a simulation project such as distribution fitting and testing, are not nearly as meaningful in summarizing output. At this stage of a simulation study (when the model influences actual decisions) interval estimates or confidence intervals become an important form for presenting results.

The accuracy of the assumptions supporting an interval estimation procedure depend not only on the structure of a simulation program but on how the data is generated, collected, and processed. The effects of incorrect procedures will be illustrated by considering a random variable, η^*, defined as

$$\eta^* = \inf\{\eta \in [0,1];\ \text{Prob}\{\theta \in R(\eta,X)\} = 1\}.$$

That is, η^* is the smallest confidence coefficient that insures that the unknown parameter θ is within the region $R(\eta,\underset{\sim}{X})$, constructed from the data $\underset{\sim}{X}$.

The usefulness of the random variable η^* arises from the observation; when the interval estimation procedure is based on true assumptions, η^* is uniformly distributed [1]. Three common causes of interval estimation errors are presented. These included unwarrented estimator distribution assumptions, initialization bias, and autocorrelation.

The literature on statistical methods for analysis of simulation output deals almost exclusively with parameter estimation (and perhaps hypothesis testing). The relationship of the information gathered from a simulation to the decisions which motivated the study is at best implicit. An understanding of the decision process and a careful analysis of the interaction between the simulation study and the decision maker is an important but often neglected preliminary to designing simulation experiments [2].

Most of the texts on statistical simulation methodologies do not include material on statistical decision theory. While the structure of the problem may preclude the formulation of an explicit loss function, there are many useful concepts in statistical decision theory. An introduction to this area of statistical analysis will be presented.

Many of the difficulties in output analysis can be alleviated by particular data collection strategies. Several approaches to simulation initialization will be discussed. Batch sampling, state-time sampling, and the use of regeneration cycles will also be introduced [3], [6].

There are many approaches to simulation output analysis. These include spectral analysis [4], autoregressive model fitting [5], and multiple comparison and ranking. The appropriateness of these methods will be discussed as time permits. A good reference on various analysis techniques is Chapter 10 in the book by G.S. Fishman [5].

BIBLIOGRAPHY

Source material for particular problems and methods will be given during the tutorial.

[1] Schruben, L.W., "Problems in Interval Estimators From Simulation Output," Technical Report No. 359, School of Operations Research and Industrial Engineering, Cornell University.

[2] Shannon, R.E., Simulation: The Art and Science, Prentice-Hall Inc., Englewood Cliffs, N.J., 1975.

[3] Law, A., "Confidence Intervals in Discrete Event Simulation: A Comparison of Replication and Batch Means," to appear in Naval Research Logistics Quarterly.

[4] Fishman, G.S., and Kiviat, P.J., "The Analysis of Simulation-Generated Time Series," Management Science, Vol. 13, No. 7, March 1967.
[5] Fishman, G.S., Concepts and Methods in Discrete Event Digital Simulation, John Wiley and Sons, New York, 1973.
[6]. Crane, M.A., and Iglehart, D.L., "Statistical Analysis of Discrete-Event Simulations," Proceedings of the 1974 Winter Simulation Conference, Washington, D.C., January 1974.