THE PROBLEM OF THE INITIAL TRANSIENT IN DIGITAL COMPUTER SIMULATION

A.V. Gafarian, C.J. Ancker, Jr. and
T. Morisaku

University of Southern California
Dept. of Industrial & Systems Engineering

ABSTRACT

The problem of the initial transient has received much attention in the simulation literature. However, much of it is replete with ambiguous admonitions rendering it virtually useless to the average practitioner. In this paper we present a comprehensive framework for this problem within which a systematic study may be pursued and eventually produce some well-defined, comprehensible, and meaningful methods that a simulator can apply to his particular problem. More specifically, we give 1) a precise definition of the problem, 2) a set of criteria by which solutions may be evaluated, and 3) a description of the studies underway within the framework established in 1) and 2).

INTRODUCTION

In the literature of simulation there is much discussion of studying a system in either 1) steady state (equilibrium conditions), or 2) in the startup or transient period. With regard to the problem of steady state, for example, usually three approaches are considered. They are 1) using long enough computer runs so that the data from the transient period are insignificant relative to the data from the steady state condition (the reader is reminded that steady state is an unattainable limiting condition), 2) excluding some appropriate part of the initial period from consideration, and 3) choosing initial starting conditions more typical of steady state and thus reducing the transient period. A critical examination of any of these approaches reveals that one must really know how close to steady state the system is as it evolves, say in time. Without having this knowledge, it is impossible to pin down system conditions for which the conclusions, based on the gathered data, are valid. For example, if all data is collected far from equilibrium it cannot be used to produce good estimates of the steady state mean. The disastrous consequences of not properly accounting for the initial transient have been illustrated very well by Law(1) in the construction of a confidence interval for the steady state mean using either the method of replication or method of batch means.

Many intuitive rules have beeen given in the literature purporting to establish when near equilibrium has been attained. Up to now, however, no attempt has been made to evaluate their effectiveness or cost. It is the purpose of this paper to develop a framework within which such an evaluation may be made and ultimately lead to well defined, meaningful rules that the average practitioner can apply to his problem and know with some assurance the regime under which his observations are made.

In the next two sections a definition of the problem is given and criteria are established for the evaluation of proposed solutions. Finally, in the last section, we give a description of a study that is being conducted within the framework established in this paper.

PROBLEM DEFINITION

In this section we describe what we mean by the problem of the initial transient. Crudely speaking, the problem is to determine that point in simulated time when near equilibrium has been achieved. To make this more precise we suppose that a discrete parameter stochastic process $\{X_t, t=1,2,\ldots\}$ is being observed for which a set of initial conditions, denoted by I, exist at $t=0$. X_t may arise by sampling, at equidistant time intervals, a continuous time series such as the number of jobs in a system; or X_t may be inherently discrete. For example, it may be the waiting time of the t^{th} customer arriving to a queuing system after it opens with some initial number of customers. In any case, we suppose that the first moments of these random variables exist and tend to an asymptotic limit independent of I, i.e.,

$$\lim_{t\to\infty} E(X_t | I) = \mu_\infty ,$$

where μ_∞ is defined as the steady state or equilibrium mean value.

We define the principal problem of the initial transient as that of determining the minimum t (call it t^*) such that the expectation of the random variables X_t, $t \geq t^*$ is as close as one desires to to the limiting expectation. Symbolically, t^* is the smallest t for which

$$1-\varepsilon \leq E(X_t)/E(X_\infty) \leq 1+\varepsilon, t \geq t^*,$$

where $E(X_\infty) = \mu_\infty$ is the steady state expected value and $\varepsilon > 0$ is a preassigned number. Thus, for example $\varepsilon = .05$ if one desires to be within 5% of the steady state value. Figure 1 below illustrates examples of some ways that $E(X_t)/E(X_\infty)$ converges and also the t^* corresponding to the ε. Though our notation does not show it explicitly, it should be noted that t^* depends on ε.

EVALUATION CRITERIA

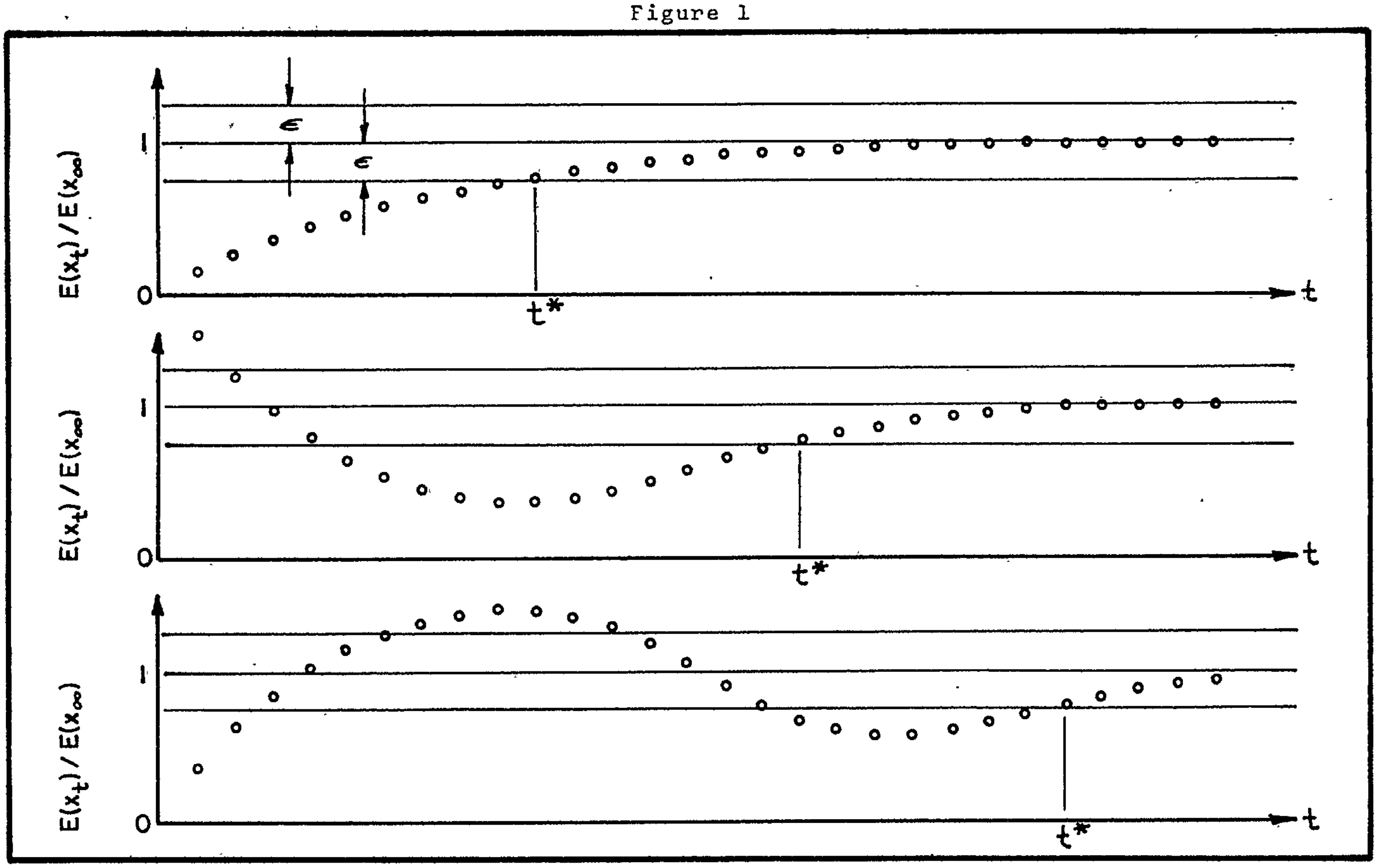
In assessing the goodness of a rule for dedetecting t^*, or in statistical language, estimating t^*, several desirable characteristics were identified. These are accuracy, precision, generality, simplicity, and cost. Each of these will be discussed in turn but before doing so, the reader should be reminded that given a rule, we are merely determining an estimate, $\widehat{t^*}$, of a well defined number t^*. Thus, many of the concepts associated with estimation theory apply such as unbiasedness, mean square error, etc.

1. Accuracy. A rule is a statement which says how to obtain a t^*. Of course, this is just one possible value of a random variable $\widehat{T^*}$, i.e., $\widehat{T^*}$ in an estimator of t^*. Accuracy will be used to define a measure of location. It thus seems that an appropriate definition of accuracy is

$$a = E(\widehat{T^*})/t^*.$$

If this ratio is close to one, then we say the estimator is accurate; greater than one implies a positive bias, less than one a negative bias.

Figure 1



Evaluation Criteria (continued)

2. Precision. Precision will be used as a measure of variation; more specifically the coefficient of variation of T^*, namely,

$$p = \sqrt{\mathrm{Var}(\widehat{T^*})}/E(\widehat{T^*}),$$

will be defined as the precision of the estimator. Clearly, a small value is desirable, so that when it is close to zero we say that the estimator is precise.

3. Generality. This is a property which means that the rule performs well across a broad range of parameters within a system and a broad range of systems.

4. Simplicity. This is a characteristic of a rule which makes it accessible to the average practitioner of large scale system simulations. A rule utilizing abstruse mathematical or statistical results is nearly incomprehensible and virtually useless to the average person who needs to know how to get statistically reliable results from a simulation.

5. Cost. By cost we mean the expense in computer time involved in implementing a given detection rule. There are three factors which combine to arrive at a total cost. Not all factors appear in every rule. These factors are:

i) Computer time associated with the algorithm itself.

ii) Computer time associated with collecting data only for estimating t^* and subsequently discarding this data.

iii) Computer time associated with an inaccurate positively biased rule. Thus, if $E(\widehat{T^*}) >> t^*$, then on any replication of the simulation experiment, the useful data generated between t^* and the smallest integer greater than or equal to $E(\widehat{T^*})$ would not be used.

EVALUATION STUDIES

Evaluation studies are presently underway within the framework developed here. Some of the most common rules that have appeared in the literature, variants of them and some completely new ones are being carefully evaluated in terms of the criteria discussed above. The procedure is empirical and will be carried on for some time.

More specifically, we have begun with the simple M/M/1 queuing system. We chose this for two reasons: 1)Queuing simulations are very common in the real world. 2) The theoretical values of the quantities of interest are known. The basic idea is to apply the rule to this system, which is easy to simulate, and estimate the performance parameters for various values of ε, initial conditions, and ratios of arrival to service rates.

Obviously, it is not sufficient to evaluate rules in this one situation only. The more promising ones are being subjected to tests in other systems for which analytical results are available.

BIBLIOGRAPHY

1. Law, Averill M. "A Comparison of Two Techniques for Determining the Accuracy of Simulation Output," No.75-11, Department of Industrial Engineering, University of Wisconsin at Madison, 1975.