1981 Winter Simulation Conference Proceedings
T.I. Ören, C.M. Delfosse, C.M. Shub (Eds.)

THE INITIAL TRANSIENT IN STEADY STATE SIMULATION (PANEL DISCUSSION)

Stephen S. Lavenberg (Chairperson)
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

PANEL MEMBERS

George S. Fishman
Curriculum in Operations Research
and Systems Analysis
University of North Carolina
Chapel Hill, North Carolina 27514

W. David Kelton
Department of Administrative Sciences
Kent State University
Kent, Ohio 44242

Lee W. Schruben
School of Operations Research
and Industrial Engineering
Cornell University
Ithaca, New York 14853

Peter D. Welch
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

The purpose of the panel is to provide an up-to-date perspective on the problem of the initial transient in steady state simulation. A steady state simulation is one whose purpose is to estimate the steady state characteristics of a model or to compare the steady characteristics of several models. The initial transient is the initial portion of a simulation run whose characteristics may differ from those of the steady state. It may be desirable to identify the extent of this initial portion and to delete the outputs produced during this portion when estimating steady state characteristics. Each panel member is active in developing methods for the statistical analysis of simulation outputs. The panel members have been asked to provide their perspectives on the problem of the initial transient including a description of the problem, a discussion of their approaches to handling the problem, recommendations for practitioners and recommendations for future research. Position papers by the panel members follow below.

George S. Fishman

Abstract

This paper presents a perspective on the initial transient problem in steady-state simulation. In particular, it enumerates five generally accepted facts: 1) Conditions prevailing at the beginning of a simulation influence sample paths. 2) The extent of influence is a function of the strength of autocorrelation. 3) Some initial conditions are less detrimental than others are. 4) Truncation reduces bias but usually increases variance. 5) So far no complete solution exists. The remainder of the paper describes a proposal for solving the problem. It relies on the relatively weak assumption that the conditional means in a stochastic process of interest are related linearly. An estimator of the steady-state mean is described which has considerably less bias than one can achieve via conventional trunction. An interval estimator is also described which follows from standard regression theory. A test for residual bias is presented which enables a user to judge whether or not sample data meet the minimal requirements for the proposed technique to apply. A second test allows a user to judge whether or not a more efficient estimation technique can be used.

1. *Perspective*

As you the audience know, the chairman's charge to the invited speakers today is to provide an up-to-date perspective on the problem of the initial transient in steady-state simulations. Although each speaker has his own distinct view of this problem, one hopes that all can agree on a skeletal characterization of it. In particular, in a simulation:

1. The initial conditions that prevail at the beginning of a run influence the sample path that each stochastic process, represented in the simulation, follows.

2. The extent to which the initial conditions affect a stochastic process at a given point in a run is a function of the degree of autocorrelation in the process.

3. Some initial conditions influence a stochastic process at a given point in a run to a lesser extent than other initial conditions do.

4. Truncation of observations near the beginning of a run

reduces the bias in the sample mean as an estimator of the steady-state mean but generally increases its variance.

5. No completely satisfactory procedure for resolving the problem has appeared yet.

In support of Point 5, note that during the past year three of the panelists have proposed solutions. See Adlakha and Fishman (1980), Kelton (1980) and Schruben (1979, 1980). However, I doubt anyone would claim a definitive solution.

Perhaps, one of the reasons that a solution has alluded us is that while the published characterization of the underlying structures in the stochastic processes of interest has enabled us to conceptualize the problem it has been too narrow to allow a satisfactory solution for a wide range of cases. In particular, I refer to the first-order autoregressive representation used in Fishman (1972) and more recently in Kelton (1980) and Schruben (1980). To overcome this inadequacy we describe a generalization of that model and show how it leads directly to an estimator of the steady-state mean that is relatively free of contamination from an initial transient. The remainder of this section provides a concise description of the essential points of the proposal. The Appendix which is available from the author contains a more detailed exposition.

2. *Proposal*

Let I be the initial conditions that prevail in a simulation and let $X_1^{(i)},\ldots,X_{n_i}^{(i)}$ be the sample record collected on the ith of 2m′ independent replications. The objective is to estimate the steady-state mean

$$\mu = \lim_{j\to\infty} EX_j^{(i)} \qquad i = 1,\ldots,m = 2m'$$

For convenience of exposition we take

$$n_{2i-1} = N_1 + k$$

$$n_{2i} = N_2 + k \qquad N_1 < N_2 \quad {}_{i=1,\ldots,m'}$$

The more general case of arbitrary $n_1\ldots,n_m$ is described in the Appendix.

We now impose a restriction on $\{X_j^{(i)}\ j=1,2,\ldots..\}$ that, while weak, has relatively profound implications for the extimation of μ. If the regression of X_{j-1}, X_{j-2},... on X_j is linear and of the form

$$E(X_j\,|\,X_1,\ldots,X_{j-1},I) = b + \sum_{s=1}^{p} a_s X_{j-s} \qquad (*)$$

then

$$\hat{\mu}_{k,n_i}^{(i)} = \frac{1}{n_i - k}\sum_{j=k+1}^{n_i} X_j^{(i)} \qquad p < k < n_i \quad {}_{i=1,\ldots,m}$$

has expectation

$$E\hat{\mu}_{k,n_i}^{(i)} = \mu + \frac{g_k}{n_i-k} - \frac{g_{n_i}}{n_i-k} \qquad (**)$$

where

$$g_t = 0(\gamma^{t-p}) \qquad 0 < \gamma < 1, p < t.$$

Here k is the *truncation* parameter. More importantly, the estimator

$$\tilde{\mu} = \frac{2}{m(N_2-N_1)}\sum_{i=1}^{m'} (N_2\hat{\mu}_{k,N_2+k}^{(2i)} - N_1\hat{\mu}_{N_1+k}^{(2i-1)}) \qquad (***)$$

has expectation

$$E\tilde{\mu} = \mu + (g_{N_1+k} - g_{N_2+k})/(N_2-N_1)$$

$$= \mu + 0(\gamma^{N_1+k-p}/(N_2-N_1)).$$

Observe that whereas the *within-replication* truncated sample means (**) have bias $0(\gamma^{k-p}/N_1)$ the new *across-replication* estimator (***) dilutes this bias to $0(\gamma^{N_1+k-p}/(N_2-N_1))$, provided that $N_2 \geq 2N_1$. Moreover,

$$\hat{\text{var}}\tilde{\mu} = 2\hat{\sigma}^2(N_1 + N_2)/m(N_2-N_1)^2,$$

where $\hat{\sigma}^2$ is defined in (19) in the Appendix, gives an asymptotically ($N_1\to\infty$) unbiased estimator of var $\tilde{\mu}$. Provided that $\{\hat{\mu}_{k,n_i}^{(i)};$ $i = 1,\ldots,m\}$ are normal, one can treat $(\hat{\mu}-\mu)/\sqrt{\hat{\text{var}}\tilde{\mu}}$ as t distributed with m-2 degrees of freedom.

Let us now concentrate on the plausibility of (*) as an underlying characterization. Clearly, (*) hold for autoregressive processes with normal disturbances. It also holds for a variety of stationary sequences of nonnegative random variables with gamma marginal distributions. See Lewis (1979). It also holds for Markov chains. More generally, provided that the expectation on the left is bounded for all j, one can treat the right side of (*) as an approximation to the left side that either becomes exact for some p or whose error of approximation can be restricted by making p suitably large. If the analogy with fitting a pth-order polynomial is kept in mind, p needs to accomodate the smoothness requirements of $E(X_j\,|\,X_1,\ldots,X_{j-1},I)$ for j = p + 1, p + 2,... .

To assure oneself of the relative insignificance of the bias in $\tilde{\mu}$ a test based on $\{X_{N_1+k}^{(2i-1)}, X_{N_2+k}^{(2i)};\ i = 1,\ldots,m'\}$ is provided in the Appendix. Briefly, if $EX_{N_1+k}^{(2i-1)} - EX_{N_2=k}^{(2i)} \cong 0$, then $g_{N_1+k} \cong 0$. Although it is difficult to choose a k such that $g_k = 0$, one would hope to be capable of choosing N_1 + k so that $X_{N_1+k}^{(2i-1)}$ the last observation collected in each other the shorter replications, is unbiased. If no significance is found, then $\tilde{\mu}$ is the best linear unbiased estimator of μ based on $\{\hat{\mu}_{k,N_1+k}^{(2i-1)}, \hat{mu}_{k,N_2+k}^{(2i)};$ i= 1,...,m′}.

Occasionally one may pick k sufficiently large so that $g_k/N_1\mu$ is relatively incidental. In this case the estimator

$$\hat{\mu} = \frac{2\sum_{i=1}^{m'} (\sqrt{N_1}\hat{\mu}_{k,N_1+k}^{(2i-1)} + \sqrt{N_2}\hat{\mu}_{k,N_2+k}^{(2i)})}{m(\sqrt{N_1}+\sqrt{N_2})}$$

gives smaller variance than $\tilde{\mu}$. The Appendix provides a method of computing a confidence interval for $g_k/N_1\mu$.

Figure 1 shows the essential steps to follow and procedure M in the Appendix contains all required computational expressions. We remark that the choice of two sample sizes N_1 + k and N_2 + k leads to considerable convenience and simplication with regard to estimating μ

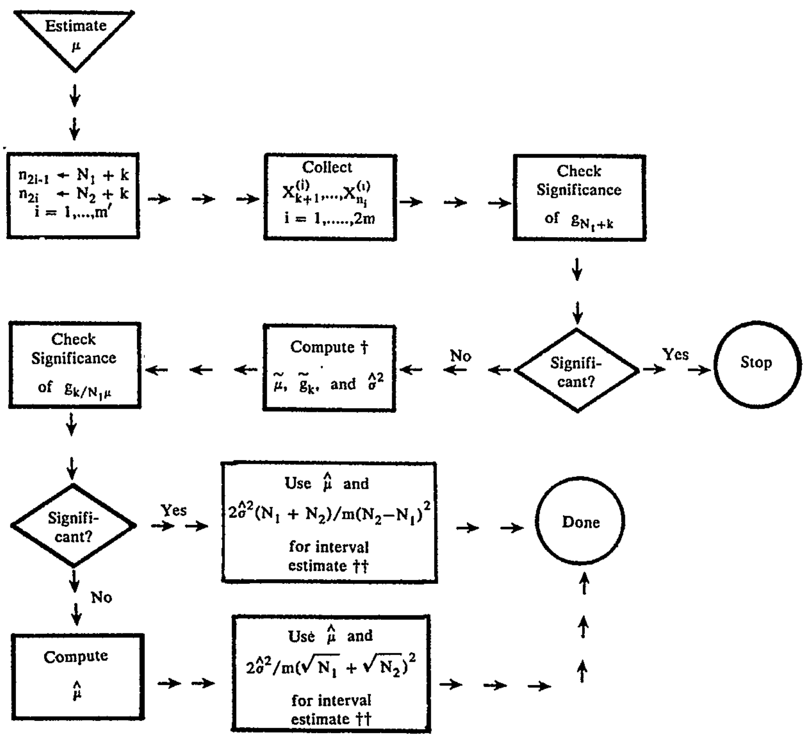

Figure 1. Estimation of μ

† See (14) in Appendix for $\tilde{g}_k$.

†† t distribution with m-2 degrees of freedom.

3. *References*

Adlakha, Veena and George S. Fishman (1980). "Starting and Stopping Rules for Simulations Using *a priori* Information," Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill.

Fieller, E. C. (1954). "Some Problems in Interval Estimation," *Journal of the Royal Statistical Society,* Series B, No. 2, 175-85.

Fishman, George S. (1972). "Bias Considerations in Simulation Experiments," *Operations Research,* 20, 785-91.

Kelton, W. David (1980). "The Startup Problems in Discrete-Event Simulation," unpublished Ph.D. thesis, Department of Industrial Engineering, University of Wisconsin at Madison.

Lewis, Peter A. W. (1979). "Simple Models for Positive-Valued and Discrete-Valued Times Series with ARMA Correlation Structure," *Proceedings of the Fifth International Symposium on Multivariate Analysis,* Amsterdam: North-Holland.

Schruben, Lee (1979). "Controlling Initialization Bias in Simulation Experiments," Technical Report 391, School of Operations Research and Industrial Engineering, Cornell University.

Schruben, Lee (1980). "A Significance Test for Initialization Bias in Simulation Output," Technical Report 444, School of Operations Research and Industrial Engineering, Cornell University.

W. David Kelton

I. Introduction

In a steady-state simulation of a discrete-time stochastic process $\{X_i,\ i = 1,2,\ldots\}$, one usually tries to estimate the steady-state mean

$$\mu = \lim_{i\to\infty} E(X_i), \tag{1}$$

assuming that the limit exists; the continuous-time case is analogous. (Alternative definitions of steady-state means are sometimes given; these are often equivalent to (1).) More generally, a simulator's goal is (or should be) to draw some kind of statistical inference about μ, rather than merely obtaining a simple point estimate. Furthermore, many simulation studies involve more than a single simulated system, in which case we are faced with a statistical inferential problem concerning steady-state means of several different stochastic processes.

A review of the statistics literature reveals that nearly all statistical procedures designed for these various inferential problems assume the ability to "sample" from an appropriate "population," i.e., to obtain independent and identically distributed (i.i.d.) random variables (r.v.'s) Y_1, Y_2,... with expectation equal to the parameter of interest. In our case, then, the ability to obtain i.i.d. Y_j's with $E(Y_j) = \mu$ would allow us to tap this large literature of "classical" statistical analysis problems and procedures for application in steady-state simulation. (See [7] for a survey emphasizing multiple comparison and selection problems for simulations of alternative systems.) Thus, developing a way to obtain observations which can safely be treated as being i.i.d. and unbiased for μ would seem to be the key problem in statistical analysis of steady-state simulations.

Probably the simplest way of collecting i.i.d. data in simulation is to make independent replications of the model; each replication is initialized and terminated in the same way, and uses an independent stream of psuedo-random numbers. Assuming a properly operating random-number generator, a r.v. obtained from a replication is an i.i.d. replicate of this r.v. obtained from any other replication. More specifically, if $X_i(j)$ is the (simulated) value of X_i on the jth independent replication of length m, a common way of obtaining i.i.d. estimates $Y_1, Y_2, \ldots$ of μ is to let $Y_j = \overline{X}_{lm}(j)$, where

$$\overline{X}_{lm}(j) = \sum_{i=l+1}^{m} X_i(j)/(m-l)$$

is the average of the last m-l values of the X_i's in the jth replication. The rationale for this estimtor is that by choosing l and m "properly," the biasing (for μ) effect of the artificial initial conditions will not induce appreciable bias in $\overline{X}_{lm}(j)$ as an estimator of μ, since the run is "warmed up" for the first l points before observation begins. This idea of deleting the initial l points from a simulation and then using the average of some additional number m-l of points has been around for some time (see, e.g., Conway [3] or Gafarian, Ancker, and Morisaku [5]), but its efficacy has been subject to considerable controversy (see, e.g., Blomqvist [2], Fishman [4], Turnquist and Sussman [10], or Wilson and Pritsker [11]). Furthermore, "proper" choice of l and m has proven to be a notoriously difficult problem (see [5], [6], Schruben [9], and [11]). Nevertheless, the idea deserves serious consideration, since identification of l and m such that $\overline{X}_{lm}(j)$ has mean near μ would put us in the enviable position described in the preceeding paragraph: that of having i.i.d. observations which can be regarded as having expectation μ. With this in mind, then, a utilitarian definition of the initial transient problem might simply be to find a practical and reliable method of determining l and m such that $E[\overline{X}_{lm}(j)] = \mu$. Given such a method, the simple approach of independent replication would become a viable overall tactic for steady-state analysis.

Two comments on the above definition of the problem should be made. First, as pointed out by Schmeiser [8], this formulation is not entirely well-posed, since many different (l,m) combinations could be found which would reduce the bias in $\overline{X}_{lm}(j)$ to the same level. Thus, we might wish to find some sort of optimal combination, e.g., find l and m to minimize the variance of a final point estimator, subject to low-bias and budget constraints. As mentioned above, however, it is hard enough to find *any* reasonable (l,m) pair resulting in low bias, so that questions of optimality probably must be left for later. The second comment is that the above definition of the problem differs considerably from most others, such as finding the minimal i* such that $E(X_i) \cong \mu$ for $i \geq i^*$. This latter definition of the problem is aimed at identifying a point in time beyond which the simulation is operating "in steady state," and is more restrictive than the one proposed in the previous paragraph, since it is likely that i* will be quite a bit larger than l if $E(X_i)$ approaches μ monotonically. Again, the proposed definition is aimed at obtaining i.i.d. observations which are unbiased for μ.

II. An Approach to the Problem

This section briefly describes a method developed and tested in [6] for dealing with the initial transient problem as defined in Section I. For a complete account, see [6].

The procedure basically forms estimates of the $E(X_i)$'s for i = 1, 2, ..., and then attempts to use these data to identify time indices l and m between which $E(X_i)$ appears to be flat, i.e., independent of i. Thus, between l and m, the data do not exhibit a "trend" or "drift," which would be characteristic of steady-state operation.

The estimates of the $E(X_i)$'s are obtained by making some number k of independent replications, each of some initial length m, and letting

$$\overline{X}_i = \sum_{j=1}^{k} X_i(j)/k,$$

the average (over replications) of the k values of $X_i(j)$. Thus, $E(\overline{X}_i) = E[X_i(j)] = E(X_i)$, but $Var(\overline{X}_i) = Var[X_i(j)]/k$, so that $\overline{X}_i$ is an unbiased estimator of $E(X_i)$ but has variance k times smaller than $X_i(j)$, the estimate of $E(X_i)$ from a single replication.

This averaged time series $\{\overline{X}_i\}$ is then examined to identify, if possible, a suitable starting index l. Starting at the *end* of the $\overline{X}_i$'s (i.e., with latest time indices), a simple straight line is fitted to a segment of the $\overline{X}_i$'s by means of a time-series regression technique (see Amemiya [1]), and the null hypothesis of zero slope is formally tested. If rejection occurs (i.e., even the final segment of the $\overline{X}_i$'s still exhibits a drift), then the trial value of m is increased (additively), all k replications are extended out to this greater length, and we try again. If, on the other hand, the initially fitted line has slope which cannot be distinguished from zero, we begin "backing up" toward the beginning of the $\overline{X}_i$'s and refit straight lines until it appears that the transient drift is being detected, as evidenced by rejection in the zero-slope hypothesis test. Finally, l is taken to be the index corresponding to the left end of the leftmost fitted line which appeared to be flat.

This procedure was empirically tested by applying it repeatedly to stochastic models with known μ. For each of thirteen models, 100 (in some cases 150) independent experiments were carried out (using k = 5 replications in each experiment), to determine values for l and m. Performance measures included estimates of absolute values of point estimator bias, and estimated coverage probabilities of nominal 90% confidence intervals (c.i.'s) for μ formed by treating the five $\overline{X}_{lm}(j)$'s as i.i.d. normal r.v.'s with expectation μ. Results were encouraging; the average absolute bias (as a percentage of μ) over the thirteen models was less than 2%, and the average c.i. coverage was 84%. Application of the procedure to a sequential c.i. formation problem (with a prespecified smallness condition) also yielded good results. Finally, the procedure was successfully used in the context of a multiple selection problem involving three alternative systems to be compared on the basis of steady-state means. Thus, at least for the models considered, this method would appear to yield $\overline{X}_{lm}(j)$'s which can be regarded as i.i.d. and unbiased for μ.

Presently, the procedure has at least two drawbacks. First, it appears to perform well only if the transient is monotone, i.e., if $E(X_i)$ converges to μ monotonically. While this exclusion of oscillatory approach to μ does restrict the current range of applicability, it still admits a broad class of models, such as many queueing systems initialized in the empty-and-idle state. Secondly, the procedure's operation requires the ability to restart several replications which were all terminated earlier. This might be accomplished, for example, by saving a "snapshot" of all necessary state variables at the end of each replication. The difficulty of programming this capability depends on model structure and complexity, and on the language used.

III. Recommendations and Future Prospects

The initial transient problem will probably never be "solved" to the satisfaction of everyone, and for all models and all purposes. At present, it might be possible to summarize the situation from the standpoint of practitioners and researchers, as follows:

The most important question for the practitioner to answer with regard to questions of steady-state analysis is whether a steady-state simulation is *really* what is desired. If the goal of the study can be rephrased such that the simulation is of the terminating type (relative to specific starting and stopping conditions), then simple replication leads to the desired i.i.d. unbiased data. Thus, with a mere restating of a study's goals, difficult problems of steady-state analysis can simpley be avoided altogether.

Provided that the practitioner really does wish to carry out a steady-state simulation, the procedure described in Section II could be applied if it is felt that the transient is monotone, and if it is possible to restart the replications. (FORTRAN programs to carry out the method are given in [6].) The prudent simulator should augment this with inspection of graphics displays of the output (e.g., plots of $\bar{X}_i$ vs. i), as well as with any special knowledge of or experience with the model.

If one accepts the definition of the problem given in Section I, then future research should be aimed at developing or improving methods for obtaining observations which can be safely regarded as being i.i.d. and unbiased for μ. Several more specific areas for research might be:

A. Improved ways of determining starting (l) and stopping (m) values for data collection in individual replications.

B. Determining "optimal" values for l and m.

C. Finding ways of specifying starting conditions which are "better," i.e., more representative of steady-state behavior.

D. Improved batch means procedures. (Since one "long" run is made, the initial transient problem is less severe; the problem lies in obtaining uncorrelated "batch means." Possible approaches would include separating the batches, or using weighted batch means with greater weights near the center of a batch.)

E. Studies of the effect of *not* eliminating the startup bias in comparative simulations of several similar, alternative systems. (If the individual transients are monotone in the same direction, the problem might not be as severe as in the single-system case.)

In general, the prospects for more reliable and complete steady-state analyses would appear to be bright (as for simulation in general), if the cost of computing continues to fall.

References

1. Amemiya, T., "Generalized Least Squares with an Estimated Autocovariance Matrix," *Econometrica 41,* 723-732 (1973).

2. Blomqvist, N., "On the Transient Behaviour of the GI/G/1 Waiting-times," *Skand. AktuarTidskr.,* 118-129 (1970).

3. Conway, R. W., "Some Tactical Problems in Digital Simulation," *Management Science 10,* 47-61 (1963).

4. Fishman, G. S., "Bias Considerations in Simulaiton Experiments," *Operations Res. 20,* 785-790 (1972).

5. Gafarian, A. V., C. J. Ancker, Jr., and T. Morisaku, "Evaluation of Commonly Used Rules for Detecting 'Steady State' in Computer Simulation," *Naval Res. Logist. Quart. 25,* 511-529 (1978).

6. Kelton, W. D., "The Startup Problem in Discrete-Event Simulation," Technical Report No. 80-1, Department of Industrial Engineering, University of Wisconsin (1980).

7. Kelton, W. D. and A. M. Law, "Statistical procedures for Comparison and Selection of Alternative Systems Designs by Simulation," Technical Report No. 79-3, Department of Industrial Engineering, University of Wisconsin (1979).

8. Schmeiser, B. W., personal communication (1980).

9. Schruben, L. W., "Controlling Initialization Bias in Simulation Experiments," Technical Report No. 391, School of Operations Research and Industrial Engineering, Cornell University (1979).

10. Turnquist, M. A. and J. M. Sussman, "Toward Guidelines for Designing Experiments in Queueing Simulation," *Simulation 28,* 137-144 (1977).

11. Wilson, J. R. and A. A. B. Pritsker, "Evaluation of Startup Policies in Simulation Experiments," *Simulation 31,* 79-89 (1978).

Lee W. Schruben

The initial conditions specified for a computer simulation can influence the results in a variety of ways. The mean of a system performance measure might be affected by the starting state. The variance and serial correlation structure of the output might change as the initial conditions are altered (Duket, 1974; Fishman, 1973). Also, dependencies between different performance measures in multiple response simulation studies might be influenced by the start-up strategy (Schruben; 1980A).

Proposition 1:

Of the potential problems due to simulation initialization, dependence of output sample means on the starting state is the most important.

An implication of this proposition is that *bias* in the mean is the most important statistical criterion for judging the effectiveness of a particular run initialization strategy. Bias by itself does not measure the efficiency of a technique; however, the mean squared error criterion is perhaps not appropriate. The estimator mean squared error should not be the only criterion (Fishman, 1973). Explicit recognition of the trade-off between bias and estimator variance should be made; both measures need to be included separately. Some interesting criteria are presented in (Gafarian, et al.; 1978), (Kelton, 1980), and (Adlakha and Fishman; 1979). Some of these criteria are difficult to compute for complex systems and are perhaps redundant. A few reasonable and easy to evaluate criteria are needed.

Proposition 2:

The statistical effectiveness of a particular simulation utilization technique should be evaluated based on a few established criteria.

It is proposed here that the sample bias and the sample (generalized) variance be used; however, *agreement* on meaningful and widely applicable (not requiring extensive analytical results) criteria is the goal.

Related to proposition 2 is

Proposition 3:

A set of inexpensive widely available simulation models needs to be identified for testing procedures.

Again *agreement* is sought. To initiate discussion, it is proposed that the M/M/1 queue and the inventory model in (Law, 1977), along with the Computer Time Sharing model in (Adiri, Avi-Itzak, 1969), and some non-normal stationary time series (Lewis, 1980) be used in the evaluation of new methods for controlling initialization bias.

The most popular method for controlling initialization bias is to allow the simulation to "warm-up" before output data are retained for analysis (Wilson and Pritsker, 1978), (equivalent to truncation of the output record). The more information that is available on which to base a truncation decision the more likely that the decision will be satisfactory.

Proposition 4

The entire output record should be saved and used in selecting a truncation point.

Final decisions on truncating output should be based on all the information available. In most simulation studies the cost of generating data is high enough to off-set temporary storage costs. Procedures that prospectively determine a data collection starting time may be useful but these truncation points should be preliminary. While some procedures based on partial output records may in the main be effective there appear to be occasional severe aborations that in retrospect might have been avoided. Using the entire output record (or at least not just the output up to the truncation decision point) was suggested some time ago. The success of this strategy is demonstrated in (Kelton, 1980.)

Another criterion for procedure evaluation must therefore be included: data storage requirements. Storage requirements and computational overhead (CPU Time?) reflect the cost added by a particular procedure.

The wide variety of simulation models make a fifth proposition appealing.

Propostion 5:

A run initialization procedure should be applicable to a variety of situations.

Unfortunately, it will most likely be impossible to develop an effective procedure that is not based on some restrictive assumptions concerning the process being simulated or on the behavior of the initial transients.

Many initialization strategies are based on a test for the presence of significant initial transient effects. Gafarian et al. (1978) have performed studies of many such proposals. Several new ideas have been offered since these studies. Three that are probably most interesting to the panel are in (Adlakha and Fishman, 1979; Kelton, 1980; Schruben 1980B; and Schruben, Singh and Tierney, 1980). Briefly the proposals are as follows:

Method 1

(Adlakha and Fishman): Significant transient effects have dissipated once $\hat{\rho}$, the estimated traffic intensity in a queueing simulation, is representative of the true traffic intensity.

Method 2

(Kelton): Transient effects have dissipated once the slope of the output record regressed on time is not significantly (at the .5 level) different from zero.

Method 3

(Schruben, Singh, and Tierney): Transient effects have dissipated if partial sums of deviations about the sample mean exhibit behavior consistent with the limiting stochastic process (a Brownian bridge).

Of these three proposed techniques, the third, based on weak convergence theory, is most generally applicable. None of the methods apparently require an excessive computational overhead. It is reasonable to presume that one of these approaches may provide a solution to the long standing problem of initialization bias. Comparative studies are in order.

References

Adiri, I. and Avi-Itzhak, B. (1969) "A Time-Sharing Queue with a Finite Number of Customers," *Journal of the A.C.M.*, 16, 315-323.

Adlakha, V. G. and Fishman, G. S. (1979) "A Starting Rule for Data Collection in Queueing Simulations" Technical Report 79-3, Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill, NC.

Duket, S. D. (1974) "Simulation Output Analysis," M.S.I.E. Thesis, Purdue University.

Fishman, G. S. (1973) "Bias Considerations in Simulation Experiments" *Operations Research*, 20, 785-790.

Gafarian, A. V., Ancker, C. J. Jr. and Morisaku, T. (1978) "Evaluation of Commonly Used Rules for Detecting 'Steady State' in Computer Simulation," *Naval Research Logistics Quarterly*, 25, 511-529.

Kelton, D. W. IV (1980) The Start-up Problem in Discrete Event Simulation, Ph.D. Dissertation, University of Wisconsin-Madison, WI.

Law, A. M. (1977) "Confidence Intervals in Discrete Event Simulation: A Comparison of Replication and Batch Means," *Naval Research Logistics Quarterly*, 24, 667-678.

Lewis, P. A. W. (1980) "Simple Models for Positive-Valued and Discrete Valued Time Series with ARMA Correlation Structure" in *Multivariate Analysis - V* (Krishnaiah, P. R., Editor), North-Holland Publishing Company, 151-166.

Schruben, L. W. (1980A) "Control of Initialization Bias in Multivariate Simulation Response" Technical Report 455, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY. (to appear in *Comm. of the A.C.M.*).

Schruben, L. W. (1980B) "Detecting Initialization Bias in Simulation Output" Technical Report 444, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

Schruben, L. W., Singh, H. and Tierney, L. J. (1980) "A Test of Initialization Bias Hypotheses in Simulation Output" Technical Report 471, School of Operations Research

and Industrial Engineering, Cornell University, Ithaca, NY.

Peter D. Welch

Suppose a simulation generates a sequence of output random variables X_1, X_2,... which converge to a steady state random variable X. Specifically, we let $F_n(x)$ be the distribution function of X_n and assume

$$\lim_{n \to \infty} F_n(x) = F(x)$$

where F(x) is the distribution function of X. Suppose further that we are interested in estimating some characteristic of this limiting random variable.

In this situation, one generally applies a statistical procedure which assumes a sequence of random variables Y_1, Y_2,... from the steady state portion of the output sequence. This is approximately achieved in practice by dropping some number n_o from the initial portion of $\{X_n\}$ and letting $\{Y_n\} = \{X_{n+n_o}\}$. The question then becomes how to pick n_o.

We propose a very straight-forward and conservative approach to the selection of n_o. We propose an independent statistical study of the transient phase aided by simple graphics. This requires repeated short runs of the simulation. The most direct approach would be to estimate $F_n(x)$ as a function of n and x over a range of n sufficient to observe the convergence of $F_n(x)$. In practice, however, such an investigation would require more computing time than is reasonable. Hence, we propose the investigation of the transient behavior of the *particular* characteristic of interest. As an example suppose we are interested in estimating the steady state mean, E[X]. In this case we estimate the point n_o at which the transient mean

$$\mu_n = E[X_n]$$

is approximately equal to the steady state mean. We assume that the resulting sequence $\{Y_n\} = \{X_{n+n_o}\}$ will be adequate for the statistical procedures to be applied; e.g., adequate to generate a valid confidence interval for E[X].

As an example consider the simulation of the queueing network of Figure 1. This is a model of a simple terminal driven computer system. Suppose there are 25 customers in the network. These customers represent 25 users at terminals. There is no queueing at the terminals but there is a random think time. At the queues indicated, service is on a first come, first served basis. We assume a central processing facility with a maximum multiprogramming level of 5. This is represented in Figure 1 by the dashed enclosure. Hence, the number of customers in this dashed enclosure cannot exceed 5. Customers wait in Queue 2 for entry into the multiprogramming set. Queue 3 is in front of the processor and Queues 4 through 7 are in front of backing store devices. When customers depart from the processor, Queue 3, they choose a route according to the branching probabilities indicated. We will consider the sequence of waiting times $W_{2,n}$:n=1,2,... of successively departing customers from Queue 2. $W_{2,n} \to W_2$ and we assume we are interested in estimating $E[W_2]$

Notice first that a "reasonable" estimate of n_o *cannot* in general be drawn from a single realization. In Figure 2 we have plotted five independent realizations of $\{W_{2,n}$:n=1,...,300$\}$ for a particular model of the type described by Figure 1. By looking at any one of these sequences alone, one would draw quite different conclusions as to the end of the transient phase. However, if we let $\{X_{m,n}$:n=1,.....,300$\}$ be the m'th realization and estimate $\mu_n = E[W_{2,n}]$ by the average of these five sequences, i.e., by

$$\hat{\mu}_n = (1/5) \sum_{m=1}^{5} X_{m,n}$$

a reasonable estimate of n_o can be made. This is illustrated in Figure 3. In this case, a reasonable estimate of n_o would be 50 waiting times. Of course the larger the number of realizations M, the more stable is the sequence $\{\hat{\mu}_n\}$ and hence, the more reliable the judgement as to the point of convergence. We illustrate this in Figure 4 where we have averaged over M=25 and M=100 output sequences.

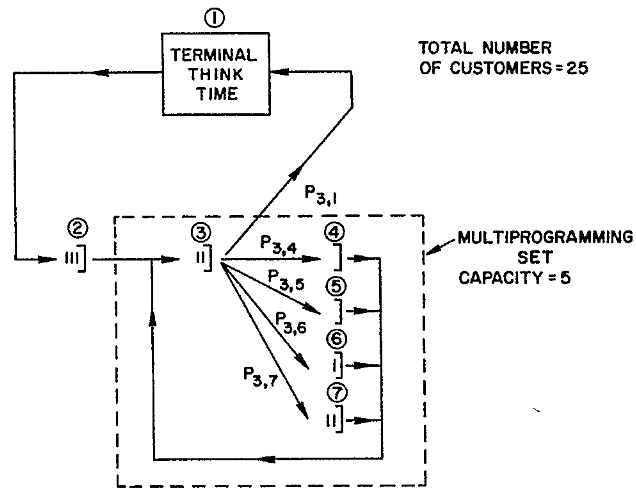

Figure 1
A Closed Queueing Model with Blocking

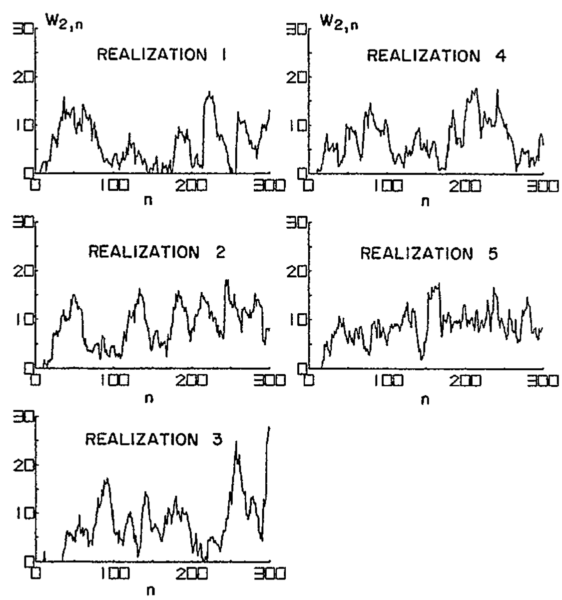

Figure 2
Five Sample Realizations of the Sequence $\{W_{2,n}\}$

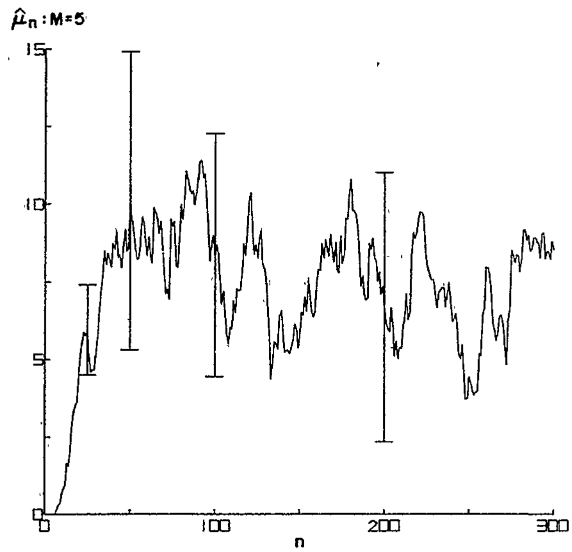

Figure 3

A Realization of $\{\hat{\mu}_n\}$

Obtained from the Five Sequences of Figure 2

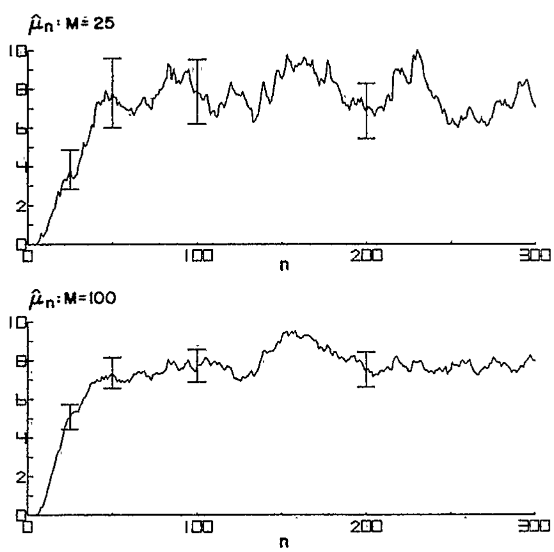

Figure 4

Realizations of $\{\hat{\mu}_n\}$ Obtained from 25 and 100

Independent Realizations of $\{W_{2,n}\}$

Some notion as to the significance of the convergence features of the sequence $\{\hat{\mu}_n\}$ can be obtained by plotting confidence intervals for a selected set of the μ_n. These confidence intervals can be obtained via the t-distribution. A set of 90% confidence intervals were calculated and are plotted in Figures 3 and 4. They were done for μ_{25}, μ_{50}, μ_{100} and μ_{200}. It is important to realize that these individual confidence intervals can only be used as a rough guideline for the sequence $\{\mu_n\}$. From the Bonferroni inequality the confidence level for all four jointly is greater than or equal to .6.

Finally, sometimes these judgements about the long term trends in $\{\mu_n\}$ are easier to make if an explicit attempt is made to smooth out the short term (high frequency) fluctuations in $\{\hat{\mu}_n\}$. The simplest way to do this is to take a moving average over an interval long enough to remove short term fluctuations but not so long as to distort the long term trend. A moving average of length 2K+1 is defined as

$$\hat{\mu}(n;K) = \frac{1}{2K+1}\sum_{k=-K}^{K} \hat{\mu}_{n+k} \quad \text{if } n \geq K+1$$

$$= \frac{1}{2K+1}\sum_{k=(n-1)}^{n-1} \hat{\mu}_{n+k} \quad \text{if } n < K+1.$$

In Figure 5 we plot $\hat{\mu}(n,10)$ for the $\{\hat{\mu}_n\}$ of Figures 3 and 4. Since the moving average of the average is also the average of the moving averages confidence intervals could also be placed on these plots.

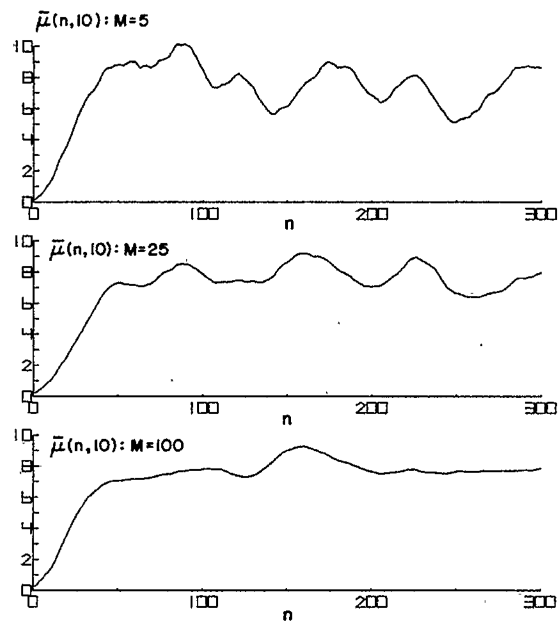

Figure 5

Moving Averages of $\hat{\mu}_n$ for M = 5, 25 and 100

If the experimentor was interested in some other characteristic such as the 90% point of F(x) or the variance of X he would estimate the transient behavior of that quantity in an analogous fashion. These would in general result in different values of n_o than obtained for E{X}.