Proceedings of the 1988 Winter Simulation Conference
M. Abrams, P. Haigh, and J. Comfort (eds.)

Initializing for bias reduction: Some analytical considerations

Joseph R. Murray
Dept. of Industrial and Operations Engineering
The University of Michigan
Ann Arbor, MI 48109

and

W. David Kelton
Dept. of Management Sciences
University of Minnesota
Minneapolis, MN 55455

ABSTRACT

We are concerned with simulation studies where the simple method of independent replications and classical statistical techniques are used to construct estimates (point and interval) for a steady-state parameter of interest. This paper describes the results of an analytical study on the effectiveness of *stochastic initialization*, where the initial state of each replication is picked randomly from an initial state distribution. The initial-state distribution is constructed from data observed during a "pilot phase". Using an AR(1) process, and assuming a fixed simulation budget, we show that stochastic initialization is effective in reducing bias in the point estimate and increasing coverage of the interval estimate without unduly increasing variance or mean square error.

INTRODUCTION

The problem of initialization bias is well known. In particular, when using replications and classical i.i.d. statistical methods to construct estimates for a steady-state parameter, the artificial initial conditions used to begin each replication cause the estimates to be biased. The usual technique to deal with initialization bias, within the context of the method of replications, is deletion; a portion of the output data is deleted from the beginning of each replication, and the deleted portion is assumed to account for the warming-up of the system to steady-state conditions.

In general practice, the initial state for each replication is arbitrarily specified, and the same initial state is used to initialize each replication. In this paper, we examine a different technique for initialization, termed *stochastic initialization*. Here, the initial state for each replication is picked randomly from an initial-state distribution. The initial states are picked independently from the same distribution, and thus, the replications remain i.i.d. and classical statistical methods are still applicable. The initial-state distribution is constructed by observing the system (via simulation) during a pilot phase; thus, the proposed technique is feasible. We assume a fixed simulation budget, and divide this budget into two parts, one part being used for the pilot phase, and the other part being used for the production phase.

After defining the criteria used to determine estimator quality (in the next section), we examine the effects of various procedural parameters on these criteria, and draw some conclusions regarding the effectiveness of stochastic initialization.

NOTATION AND CRITERIA

We assume a fixed simulation budget that allows the generation of n observations of the output process to obtain point and interval estimates for μ, the steady-state parameter of interest. The budget for the pilot phase is n_1, broken up into k_1 replications of length m_1 each. Each pilot phase replication is initialized at the same deterministic state x_0^d.

Pictorially, the pilot phase can be represented as follows:

$$
\begin{array}{llllll}
\text{Replication} & & & & & \\
1 & x_0^d & X_1^d(1) & X_2^d(1) & \ldots & X_{m_1}^d(1) \\
2 & x_0^d & X_1^d(2) & X_2^d(2) & \ldots & X_{m_1}^d(2) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
k_1 & x_0^d & X_1^d(k_1) & X_2^d(k_1) & \ldots & X_{m_1}^d(k_1)
\end{array}
$$

where $X_i^d(j)$ represents the i^{th} observation in the j^{th} deterministically initialized replication.

The data from the pilot phase are used to construct an initial-state distribution. The remaining budget, n_2, is used for the production phase, consisting of k_2 replications of length m_2 each, with each replication initialized at a state picked independently from the initial-state distribution.

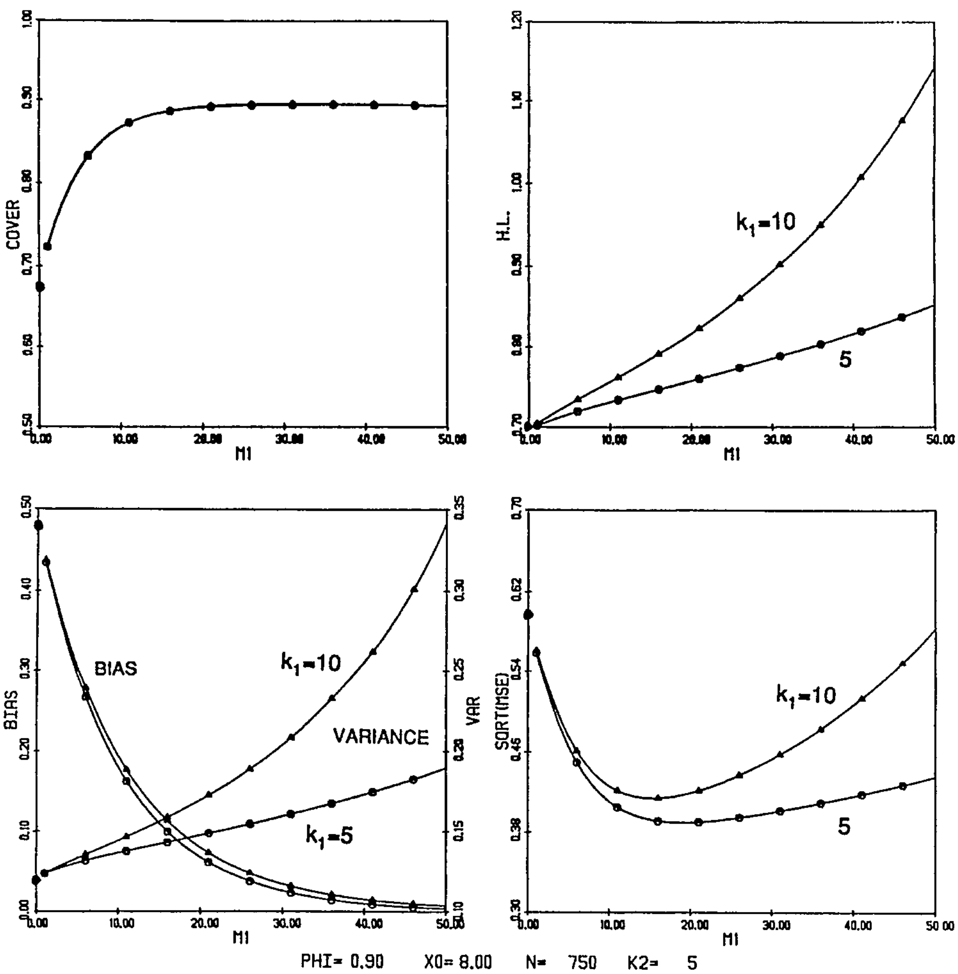

Figure 1. Bias, Variance, $\sqrt{\text{MSE}}$, Coverage and Half-Length as functions of m_1 and k_1, for $\phi = 0.9$, $x_0^d = 8.0$, $n = 750$, and $k_2 = 5$.

If $X_i(j)$ represents the i^{th} observation of the j^{th} replication in the production phase, (e.g., the delay of customer i in replication j), then the production phase can be pictorially represented as:

Replication						
1	$X_0(1)$	$X_1(1)$	$\ldots$	$X_{m_2}(1)$	$\longrightarrow$	$\overline{X}_{m_2}(1)$
2	$X_0(2)$	$X_1(2)$	$\ldots$	$X_{m_2}(2)$	$\longrightarrow$	$\overline{X}_{m_2}(2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k_2	$X_0(k_2)$	$X_1(k_2)$	$\ldots$	$X_{m_2}(k_2)$	$\longrightarrow$	$\overline{X}_{m_2}(k_2)$

where $\overline{X}_{m_2}(j) = {m_2}^{-1}\sum_{i=1}^{m_2} X_i(j)$ represents the mean within the j^{th} replication. Let $\overline{\overline{X}}_{m_2k_2} = (1/k_2)\sum_{j=1}^{k_2}\overline{X}_{m_2}(j)$ and $S^2_{m_2k_2} = (1/(k_2-1))\sum_{j=1}^{k_2}\left(\overline{X}_{m_2}(j) - \overline{\overline{X}}_{m_2k_2}\right)^2$ be the sample mean and variance of the $\overline{X}_{m_2}(j)$'s.

The quality of the point estimator, $\overline{\overline{X}}_{m_2k_2}$, is evaluated using the following criteria: Bias, Variance, and Mean Square Error. The quality of the interval estimator, $\overline{\overline{X}}_{m_2k_2} \pm t_{k_2-1,1-\alpha/2}\sqrt{S^2_{m_2k_2}/k_2}$, is evaluated using the following criteria: Coverage and Half-Length. Although it is generally impossible to evaluate these criteria analytically, we can do so by assuming the following AR(1) model.

AR(1) MODEL AND RESULTS

The AR(1) process is defined by

$$X_i = \mu + \phi(X_{i-1} - \mu) + \epsilon_i, \quad \text{for } i = 1, 2, \ldots$$

where the ϵ_i's are i.i.d. $N(0, \sigma^2)$, and $0 < \phi < 1$.

To construct an initial-state distribution from the pilot phase data, we appeal to the *principle of maximum entropy* (Jaynes 1957]). The maximum entropy distribution is a normal distribution in this case, with mean $\overline{X}_{m_1 k_1}$ and variance $S^2_{m_1 k_1}$, where $\overline{X}_{m_1 k_1} = (1/k_1)\sum_{j=1}^{k_1} X^d_{m_1}(j)$ and $S^2_{m_1 k_1} = (1/(k_1-1))\sum_{j=1}^{k_1}(X^d_{m_1}(j)-\overline{X}_{m_1 k_1})^2$, the sample mean and sample variance of the final observation in each pilot phase replication.

Using this initial-state distribution, we can establish expressions for the various criteria in terms of k_1, m_1, k_2, m_2, and x^d_0; the details are in Murray [1988]. Figure 1 shows an example of the effect of m_1 and k_1 on the criteria, for the settings stated in the caption of the figure. Note the decrease in Bias and Mean Square Error and the increase in Coverage as the length of the pilot replications, m_1, is increased. Also, note that a low value of k_1 is better than a high value.

CONCLUSIONS

The results from our study, an example of which is shown in Figure 1, lead us to conclude that stochastic initialization is effective in reducing bias and increasing coverage. There is an increase in the variance of the point estimator, and in the half-length of the interval estimator, but this increase is not severe. Mean Square Error, an indicator of the trade-off between reduced bias and increased variance, actually decreased for low values of m_1 for many of the cases studied.

REFERENCES

Jaynes, E.T. (1957). Information Theory and Statistical Mechanics. *Physical Review 106*, 620-630.

Murray, J.R. (1988). Stochastic Initialization in Steady-State Simulations. Unpublished Ph. D. Thesis, Industrial and Operations Engineering Department, University of Michigan, Ann Arbor, Michigan.

AUTHORS' BIOGRAPHIES

JOSEPH R. MURRAY is a student in the Industrial and Operations Department at the University of Michigan, Ann Arbor. He received a B.S. degree in Mechanical Engineering from the Indian Institute of Technology, Madras, India, and M.S. and Ph.D. degrees in Industrial and Operations Engineering from the University of Michigan, Ann Arbor. His research interests are in simulation methodology, operations management and manufacturing engineering. He is a member of TIMS and ORSA.

Joseph R. Murray

Dept. of Industrial Operations Engineering

The University of Michigan

Ann Arbor, MI 48109

(313) 764-6474

joseph_murray@ub.cc.umich.edu

W. DAVID KELTON is an Associate Professor in the Management Sciences Department at the Carlson School of Management, University of Minnesota, in Minneapolis. He received a B.A. in Mathematics from the University of Wisconsin-Madison, an M.S. in Mathematics from Ohio University, and M.S. and Ph.D. degrees in Industrial Engineering from the University of Wisconsin- Madison. His research interests are in simulation methodology, stochastic modeling and estimation, and quality control. He is a member of ORSA, TIMS, ASA, and SCS. He served as Program Chair for the 1987 Winter Simulation Conference.

W. David Kelton

Department of Management Sciences

Carlson School of Management

University of Minnesota

Minneapolis, Minnesota 55455

612/624-8503

dkelton@umnacvx.bitnet