A NEW APPROACH TO SIMULATING STABLE STOCHASTIC SYSTEMS*

Michael A. Crane and Donald L. Iglehart
Control Analysis Corporation and Stanford University

Abstract

A technique is introduced for analyzing simulations of stochastic systems in steady-state. Confidence intervals are obtained for a general function of the steady-state distribution.

1. INTRODUCTION

The principal goal of most simulations of stable stochastic systems is to estimate properties of the stationary or steady-state behavior of the system. Two of the major problems in such simulations are the statistical dependence between successive observations and the inability of the simulator to begin the system in the steady-state. The first problem has necessitated using methods of time series analysis rather than classical statistics. The second has inspired many simulators to let the system run for a sufficient length of time so that the initial transient wears off and a steady-state condition obtains. This procedure, of course, requires a judgement on how long to let the system run before making observations.

For many stochastic systems being simulated it is possible to find a random grouping of observations which produces independent identically distributed (i.i.d.) blocks from the start of the simulation. This grouping then enables the simulator to avoid the two problems mentioned above. He has at his disposal the methods of classical statistical analysis such as confidence intervals, hypothesis testing, regression, and sequential estimation since the observations are now i.i.d. Furthermore, information that is useful in estimating the steady-state behavior of the system can be collected from scratch thus eliminating the problem of the initial transient.

The key requirement for obtaining these i.i.d. blocks is that the system being simulated return to a single state infinitely often and that the mean time between such returns is finite. This requirement will be met for many, but not all, stable systems that might be simulated.

In this paper we shall illustrate the main ideas of this approach as applied to Markov chains, in both discrete and continuous time, and to the GI/G/1 queue. The results will only be sketched here as the complete details are available in [1], [2], and [3]. This paper is organized as follows. Section 2 summarizes the

---

* This research was sponsored by Office of Naval Research contract N00014-72-C-0086 [NR-047-106] and prepared for delivery at the 1973 Winter Simulation Conference, January 17-19, 1973, San Francisco.

probabilistic structure of Markov chains with an eye toward using these results in carrying out a simulation. Section 3 does the same for the GI/G/1 queue. In Section 4 a statistical confidence interval is stated for the ratio of two means. Numerical illustrations of this method are given in Section 5 for the repairman problem and the M/M/1 queue. The reader who is only interested in the results and not the underlying theory can turn directly to Section 5 with little loss of continuity.

2. MARKOV CHAINS

Suppose we are interested in simulating a stochastic system evolving as a Markov chain (M.c.). Let $\{X_n : n \geq 0\}$ be a discrete M.c. defined on a probability triple $(\Omega, \mathcal{F}, P)$ with discrete state space $I = \{0, 1, 2, \cdots\}$. Everything we do here can be carried over to the case of I finite. Assume that this M.c. is known to be irreducible, aperiodic, and positive recurrent. Under these conditions there will exist a unique stationary distribution, $\{\pi_i : i \in I\}$, for the M.c.

Select now a fixed state of the M.c. which we shall take for convenience to be the state 0. Now set $X_0 = 0$ with probability one; that is, we shall always begin our M.c. in the 0 state. Since the M.c. is assumed to be positive recurrent, there exists an infinite sequence of random time epochs $\{\beta_i : i \geq 0\}$ such that $X_{\beta_i} = 0$ with probability one. Thus the epochs β_i are the successive times the process returns to 0. We shall speak of the integers $\{\beta_{k-1} + 1, \cdots, \beta_k\}$ as constituting the k^{th} cycle of the M.c. Let $\alpha_i = \beta_i - \beta_{i-1}$, $i \geq 1$ and for $k \geq 1$ form the random vectors

$$\underset{\sim}{Y}_k = \{\alpha_k, X_{\beta_{k-1}+1}, \cdots, X_{\beta_k}\} .$$

As a consequence of the fact that the random variables (r.v.'s) $\{\beta_k : k \geq 1\}$ are optional and finite with probability one it is possible to show the following results.

PROPOSITION 1. *The random vectors* $\{\underset{\sim}{Y}_k : k \geq 1\}$ *are independent and identically distributed.*

This proposition lies at the heart of our method of analyzing simulations.

Now let f be a function from I to $(-\infty, +\infty)$ and suppose the object of our simulation is to estimate $\sum_{j \in I} f(j)\, \pi_j$, the stationary expected value of f. Define new r.v.'s

$$Y_k = \sum_{j=\beta_{k-1}}^{\beta_k - 1} f(X_j), \qquad k \geq 1 .$$

As an immediate corollary of Proposition 1 we state

COROLLARY 1. *The sequences* $\{\alpha_k : k \geq 1\}$ *and* $\{Y_k : k \geq 1\}$ *are independent and identically distributed.*

The second important result is

PROPOSITION 2. *If* $\sum_{j\in I} |f(j)|\pi_j < \infty$, *then the*

$$\sum_{j\in I} f(j)\,\pi_j = E\{Y_k\}/E\{\alpha_k\}\ .$$

Corollary 1 and Proposition 2 form the basis for our method. We mention in passing that all the results of this section carry over to the case where $\{X_n : n \geq 0\}$ is a Markov process with a general state space E, a single ergodic set and no cyclically moving sets, provided there exists a point (singleton set) to which X_n returns infinitely often with probability one and for which the expected length of the cycles is finite.

Suppose now we are interested in simulating a continuous time M.c. Let $\{X(t) : t \geq 0\}$ be a continuous time M.c. defined on a probability triple $(\Omega, \mathcal{F}, P)$ and having discrete state space $I = \{0, 1, 2, \cdots\}$ and standard transition matrix $\{p_{ij}(t) : t \geq 0,\ i,j \in I\}$. Again assume that the M.c. is irreducible and positive recurrent. As in the discrete case, there exists a unique stationary distribution, $\{\pi_i : i \in I\}$, of the M.c. Also the $\lim_{t\to\infty} p_{ij}(t) = \pi_j$, for all $i,j \in I$.

Now set $X(0) = 0$ with probability one. Since the state 0 will be entered an infinite number of times as a consequence of our assumption of positive recurrence, we can define $\rho_i (i \geq 1)$ to be the length of the i^{th} visit to state 0. Then let $\beta_0 = 0$, and

$$\beta_i = \inf\{t > \rho_i + \beta_{i-1} : X(t) = 0\},\quad i \geq 1\ .$$

Thus β_i is the time of i^{th} return to 0. If we let $\alpha_i = \beta_i - \beta_{i-1}$, $i \geq 1$, then α_i is the length of the i^{th} cycle from the state 0 and plays the same role as in the discrete time case.

While the technical details for the continuous case are much harder than the discrete case, the intuitive ideas are the same. Hence for this discussion we shall keep the details brief. Let f be a mapping from I to $(-\infty, +\infty)$ and define the r.v.'s

$$Y_k = \int_{\beta_{k-1}}^{\beta_k} f[X(s)]ds,\quad k \geq 1\ .$$

For this continuous time M.c. Proposition 2 and Corollary 1 continue to hold. These results provide the basis for analyzing simulations of continuous time M.c.'s.

3. QUEUES

Consider now a GI/G/1 queueing system in which the 0^{th} customer arrives at time $t_0 = 0$, finds a free server, and experiences a service time v_0. The n^{th} customer arrives at time t_n and experiences a service time v_n. Let the inter-arrival times $t_n - t_{n-1} = u_n$, $n \geq 1$. Assume that the two sequences $\{v_n : n \geq 0\}$ and $\{u_n : n \geq 1\}$ each consist of i.i.d. r.v.'s and are themselves independent. Let $E\{u_n\} = \lambda^{-1}$, $E\{v_n\} = \mu^{-1}$, and $\rho = \lambda/\mu$ where $0 < \lambda,\ \mu < \infty$. Thus $\mu(\lambda)$ has the interpretation of the mean service (arrival) rate. The parameter ρ is called the traffic intensity and is the natural

measure of congestion for this system. We shall assume that $\rho < 1$, a necessary and sufficient condition for the system to be stable.

The principal system characteristics of interest are $Q(t)$, the number of customers in the system at time t; W_n, the waiting time (time for arrival to commencement of service) of the n^{th} customer; $W(t)$, the work load facing the server at time t; $B(t)$, the amount of time in the interval $[0,t]$ that the server is busy; and $D(t)$, the total number of customers who have been served and have departed from the system in $[0,t]$.

Here we shall review the basic structure of the GI/G/1 queue relevant to our simulation study. Using the notation of optional r.v.'s, it can be shown that there exists a sequence of r.v.'s $\{\beta_k : k \geq 0\}$ such that $\beta_0 = 0$, $\beta_k < \beta_{k+1}$, and $W_{\beta_k} = 0$ with probability one. In other words, the customers numbered β_k are those lucky fellows who arrive to find a free server and experience no waiting in the queue. The fact that there exists an infinite number of such customers is a direct consequence of the assumption that $\rho < 1$. The time axis $R_+^1 = [0,\infty)$ can be divided into alternating intervals during which the server is busy, idle, busy, etc. We call these intervals busy periods (b.p.'s) and idle periods (i.p.'s). An i.p. plus the proceding b.p. is called a busy cycle (b.c.). If we let $\alpha_k = \beta_k - \beta_{k-1}$, $k \geq 1$, then α_k represents the number of customers served in the k^{th} busy period (b.p.) and they are numbered $\{\beta_{k-1}, \beta_{k-1}+1, \cdots, \beta_k - 1\}$. The sequence $\{\beta_k : k \geq 1\}$ plays the same role here as in Section 2 on M.c.'s.

Next define the random vectors $\underset{\sim}{X}_k = (v_{k-1}, u_k)$ and $\underset{\sim}{Y}_k = \{\alpha_k, \underset{\sim}{X}_{\alpha_{k-1}+1}, \cdots, \underset{\sim}{X}_{\beta_k}\}$, $k \geq 1$. Observe that the vector $\underset{\sim}{Y}_1 = \{\alpha_1, \underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_{\alpha_1}\}$ includes all the data required to completely construct the behavior of the system in the first b.p. Let f be a measurable function from $[0,\infty)$ to $(-\infty,\infty)$ and set

$$Y_k = \sum_{j=\beta_{k-1}}^{\beta_k - 1} f(W_j), \qquad k \geq 1.$$

Then Proposition 1 and Corollary 1 continue to hold. Hence we have the intuitively plausible conclusion that comparable r.v.'s in different b.p.'s are i.i.d. However, Proposition 2 must be replaced by

PROPOSITION 3. *If* $E\{|f(W)|\} < \infty$, *then the*

$$E\{f(W)\} = E\{Y_k\}/E\{\alpha_k\} .$$

where W *is the stationary waiting time.*

In addition to obtaining results for $E\{f(W)\}$ we can also handle the expected value of the stationary queue length and virtual waiting time, length of a b.p., b.c., or i.p. Furthermore, this technique can be extended to the queue GI/G/s, $s > 1$; see [1].

4. CONFIDENCE INTERVALS

From Propositions 2 and 3 we are confronted with the need to produce confidence intervals for the ratio of two means. Suppose we observe i.i.d. random (column) vectors $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \cdots, \underset{\sim}{X}_n$, where $\underset{\sim}{X}_k = (Y_k, \alpha_k)$, and assume that $E\{\underset{\sim}{X}_1\} = \underset{\sim}{\mu} = (\mu_1, \mu_2)$ with $\mu_2 \neq 0$. Let the positive definite covariance matrix of $\underset{\sim}{X}_1$ be $\underset{\sim}{\Sigma}$ with elements given by

$$\underset{\sim}{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

Let $\nu = \mu_1/\mu_2$. Our goal is to form a confidence interval for ν based on the observations $\{\underset{\sim}{X}_k : 1 \leq k \leq n\}$ where n is large. This problem was treated by ROY and POTTHOFF (1958) for the case of bivariate normal random vectors.

Let the sample mean of the n observations $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_n$ be denoted by

$$\bar{\underset{\sim}{X}}(n) = \frac{1}{n} \sum_{i=1}^{n} \underset{\sim}{X}_i = \begin{pmatrix} \bar{Y}(n) \\ \bar{\alpha}(n) \end{pmatrix}$$

and the sample covariance matrix by

$$\underset{\sim}{S}(n) = \frac{1}{n-1} \sum_{i=1}^{n} (\underset{\sim}{X}_i - \bar{\underset{\sim}{X}}(n))\,(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}}(n))'$$

with elements

$$\underset{\sim}{S}(n) = \begin{pmatrix} s_{11}(n) & s_{12}(n) \\ s_{12}(n) & s_{22}(n) \end{pmatrix}.$$

Next let $Z_k = Y_k - \alpha\nu_k$, $k = 1, 2, \cdots, n$ and let $\bar{Z}(n) = \frac{1}{n} \sum_{k=1}^{n} Z_k$. Observe that the $E\{Z_k\} = 0$ and the $\sigma^2\{Z_k\} = \sigma^2 = \sigma_{11} - 2\nu\sigma_{12} + \nu^2\sigma_{22}$. The idea of introducing the Z_k's is due to ROY and POTTHOFF (1958) and is the key to the confidence interval we obtain. Let $z_\gamma = \Phi^{-1}(\gamma)$, where

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du .$$

Using the central limit theorem for sums of i.i.d. random variables and the strong law of large numbers we can show that for $0 < \gamma < 1$ and large n the random interval

$$\left[\frac{(\bar{Y}\bar{\alpha} - hs_{12}) - D^{1/2}}{\bar{\alpha}^2 - hs_{22}}, \frac{(\bar{Y}\bar{\alpha} - hs_{12}) + D^{1/2}}{\bar{\alpha}^2 - hs_{22}} \right] \quad (1)$$

surrounds the parameter ν with probability approximately $1-\gamma$, where

$$D = (\bar{Y}\bar{\alpha} - hs_{12})^2 - (\bar{Y}^2 - hs_{11})(\bar{\alpha}^2 - hs_{22})$$

and

$$h = z^2_{1-\gamma/2}/n .$$

5. NUMERICAL EXAMPLES

We illustrate our methods with simulations of two different models. The first is the classical repairman problem, and the second is the M/M/1 queue. The theoretical results for these models are well known and provide a basis for comparison.

The repairman problem is a continuous time Markov chain that can be described as follows. We have $M + N$ identical pieces of equipment which have an expone ial failure time with failure rate λ. At most N of these units operate at one time, the other M units being thought of as spares. When a unit fails, it is sent to a repair facility consisting of S repairmen (servers) having exponential repair rate μ. Let $X(t)$ denote the number of failed units undergoing or waiting for service at the repair facility at time t. With the above assumptions $\{X(t) : t \geq 0\}$ is a birth-death process, a special type of the continuous time M.c. discussed in Section 2.

Let X be a discrete random variable having the stationary distribution $\{\pi_i : i = 0, \cdots, M + N\}$ of the M.c. In other words, X is the random variable to which $X(t)$ converges in distribution. We simulated the repairman problem in order to estimate $E\{f(X)\}$ for various choices of the function f.

Now let $X(0) = 0$. In order to analyze the simulation recall that the process returns to the state 0 infinitely often, and α_k is the length of the k^{th} cycle from the state 0. As in Section 2, let Y_k be the integral of $f[X(t)]$ over the k^{th} cycle. From Proposition 2 and Corollary 1, we know that the random vectors $\{(Y_k, \alpha_k)\ k \geq 1\}$ are i.i.d. and that $E\{f(X)\} = E\{Y_k\}/E\{\alpha_k\}$. We may thus obtain a confidence interval for $E\{f(X)\}$ by simulating the system for a fixed number n cycles and applying the method of Section 4. In particular, let $\bar{Y}$ and $\bar{\alpha}$ denote respectively the sample means for Y_k and α_k in n observations, let s_{11} and s_{22} denote the sample variances, and let s_{12} denote the sample covariance between Y_k and α_k. A $100(1-\gamma)\%$ confidence interval for $E\{f(X)\}$ is then given by equation (1) of Section 4.

To illustrate, we consider seven choices for the function f:

i) $f_1(i) = i, \quad i = 0, \cdots, M + N$;

ii) $f_2(i) = i^2, \quad i = 0, \cdots, M + N$;

iii) $f_3(i) = \begin{cases} 0, & i = 0, \cdots, M \\ 1, & i = M + 1, \cdots, M + N \end{cases}$;

iv) $f_4(i) = \begin{cases} 0, & i = 0, \cdots, S \\ 1, & i = S + 1, \cdots, M + N \end{cases}$;

v) $f_5(i) = \begin{cases} S-i, & i = 0, \cdots, S \\ 0, & i = S + 1, \cdots, M + N \end{cases}$;

vi) $f_6(i) = \begin{cases} 0, & i = 0 \\ 1, & i = 1, \cdots, M + N \end{cases}$;

vii) $f_7(i) = \begin{cases} 1, & i = 0 \\ 0 & i = 1, \cdots, M + N \end{cases}$.

These functions allow us to estimate, respectively, the expected value of X, the second moment of X, the probability that X exceeds M (insufficient spares), the probability that X exceeds S (positive queue length), the expected number of idle servers, the probability that X exceeds zero (at least one server busy), and the probability that X equals zero (all servers idle). There are of course many other functions which would yield useful estimates of the steady-state behavior.

Table 1 shows 90% confidence intervals obtained after a run length of 300 cycles from the state $X(0) = 0$. The parameter settings used for this run were $N = 10$ operating units, $M = 4$ spares, $S = 3$ servers, failure rate $\lambda = 1$, and repair rate $\mu = 4$. Table 2 shows estimates for $E\{X\}$ in ten replications of the simulation.

TABLE 1

SIMULATION RESULTS FOR THE REPAIRMAN PROBLEM

Parameter	Theoretical Value	Point Estimate	90% Confidence Interval
$E\{f_1(X)\} = E\{X\}$	3.471	3.406	[3.205, 3.607]
$E\{f_2(X)\} = E\{X^2\}$	17.278	16.844	[15.094, 18.594]
$E\{f_3(X)\} = P\{X > M\}$	.306	.294	[.206, .328]
$E\{f_4(X)\} = P\{X > S\}$	.438	.429	[.393, .465]
$E\{f_5(X)\} = E\{[S - X]^+\}$	.678	.705	[.637, .773]
$E\{f_6(X)\} = P\{X > 0\}$	.939	.930	[.919, .942]
$E\{f_7(X)\} = P\{X = 0\}$	.061	.070	[.058, .081]

TABLE 2

ESTIMATES FOR $E\{X\}$ IN TEN SIMULATION REPLICATIONS OF THE REPAIRMAN PROBLEM

Replication	Point Estimate	Confidence Interval
1	3.406	[3.205, 3.607]
2	3.386	[3.221, 3.551]
3	3.384	[3.196, 3.571]
4	3.440	[3.260, 3.620]
5	3.234	[3.047, 3.420]
6	3.542	[3.373, 3.712]
7	3.433	[3.246, 3.620]
8	3.382	[3.163, 3.600]
9	3.380	[3.213, 3.548]
10	3.415	[3.234, 3.596]
Average	3.400	[3.216, 3.585]
Average length		0.369

Theoretical value of $E\{X\} = 3.471$

Our second example is the M/M/1 queue. We have Poisson arrivals, exponential service, and a single server. Although the queue length process is a birth-death process and could be treated like the repairman problem, we focus our attention here on the sequence of customer waiting times $\{W_n : n \geq 0\}$. Recalling the discussion of Section 3, the process returns to the state W_0 infinitely often, and the time intervals between returns define busy cycles (b.c.'s). Letting f be a function on the state space, letting Y_k be the sum of $f(W_n)$ over the k^{th} b.c. and letting α_k be the number of customers served in the k^{th} b.c., we once again have $E\{f(W)\} = E\{Y_k\}/E\{\alpha_k\}$, where W is the stationary waiting time and the random vectors $\{(Y_k, \alpha_k)\ k \geq 1\}$ are i.i.d. We may thus proceed exactly as before to obtain confidence intervals for $E\{f(W)\}$.

Table 3 shows 90% confidence intervals in ten replications of the queueing simulation, each consisting of 2000 busy cycles. For these runs, the customer arrival rate was assumed to be 5 and the service rate 10 so that $\rho = .5$. We consider only the function $f(W) = W$, although there are many other interesting possibilities.

TABLE 3

ESTIMATES FOR $E\{W\}$ IN TEN SIMULATION REPLICATIONS OF THE M/M/1 QUEUE

Replication	Point Estimate	Confidence Interval
1	0.110	[.096, .123]
2	0.091	[.080, .102]
3	0.095	[.084, .105]
4	0.111	[.087, .133]
5	0.096	[.083, .109]
6	0.100	[.087, .112]
7	0.092	[.081, .103]
8	0.099	[.084, .114]
9	0.096	[.082, .109]
10	0.090	[.078, .102]
Average	0.098	[.084, .111]
Average length		.027

Theoretical Value of $E\{W\} = .100$

REFERENCES

[1] CRANE, M.A. and IGLEHART, D.L. (1972). A new approach to simulating stable stochastic systems, I: General Multi-server queues. Technical Report No. 86-1, Control Analysis Corporation, Palo Alto, California.

[2] CRANE, M.A. and IGLEHART, D.L. (1972). Confidence intervals for the ratio of two means with application to simulations. Technical Report No. 86-2, Control Analysis Corporation, Palo Alto, California.

[3] CRANE, M.A. and IGLEHART, D.L. (1972). A new approach to simulating stable stochastic systems, II: Markov chains. Technical Report No. 86-3, Control Analysis Corporation, Palo Alto, California.

[4] ROY, S.N. and POTTHOFF, R.F. (1958). Confidence bounds on vector analogues of the "ratio of means" and the "ratio" of variances" for two correlated normal variates and some associated tests. Ann. Math. Statist. 29, 829-841.