A COMPARISON OF TWO TECHNIQUES FOR COLLECTING STATISTICS IN SIMULATION

Boo Ho Rho*

Ⅰ. INTRODUCTION

The four essential elements in the study of simulation are model representation, need for random behavior, statistical analysis and reporting of output. The purpose of this paper is to deal with the area of statistical analysis. The paper specifically compares two techniques of collecting statistics: (1) the method of batch means and (2) the regenerative process method. The comparison will be in terms of accuracy and precision of sample performance measurements and execution time.

The need for effective statistical methods to analyze discrete event digital simulation results has been emphasized for more than 20 years. Conway [1] dealt with the two most important problems in the statistical analysis: the problem of initial transient conditions and the consideration of variability of sample performance measurements and sample size. Fishman [7] also identified fou rcritical problems that all stochastic simulations encounter: (1) estimation of variance; (2) initial transient conditions; (3) distribution theory of stochastic sequences; and (4) design of samplingplans.

To deal with the above statistical problems, three methods of collecting statistics have been devised. They are the method of replication, the method of batch means, and the regenerative process method.

Ⅱ. THREE METHODS OF COLLECTING STATISTICS

A. Method of replication

The method of replication is to run Kindependent simulations each of length m.

Let X_{ij} be the jth observation of the ith run. Then, we will have the following data:

$$\begin{matrix} X_{11}, & X_{12}, & \cdots, & X_{1m} \\ X_{21}, & X_{22}, & \cdots, & X_{2m} \\ \vdots & \vdots & & \vdots \\ X_{k1}, & X_{k2}, & \cdots, & X_{km} \end{matrix}$$

Furthermore, let$\bar{X}_i(m)$ $(i=1,2,\cdots,k)$ be the sample mean of the m observation in the ith run and $X(k,m)$ be the sample mean of the $\bar{X}_i(m)$'s. That is,

$$X(k,m)=\frac{\sum\bar{X}_i(m)}{k}=\frac{\sum\sum X_{ij}}{km}$$

* The University of Nebraska-Lincoln

The estimated variance of $\overline{X}(k,m)$, $\sigma^2[\overline{X}(k,m)]$ is:

$$\sigma^2[X(k,m)]=\frac{\sum[\bar{X}_i(m)-\overline{X}(k,m)]^2}{k(k-1)}$$

If $\bar{X}_i(m)$'s are normally distributed, then

$$\frac{\bar{X}(k,m)-\mu}{\sigma[\overline{X}(k,m)]}$$

has the t distribution with $k-1$ degrees of freedom. Thus, the $100(1-\alpha)\%$ confidence interval for μ is given by

$$\overline{X}(k,m)+t_{k-1},\ \ 1-\alpha/2\sqrt{\sigma^2[\bar{X}(k,m)]}.$$

There are two potential sources of error when using replication to construct confidence interval for a steady state mean [13]: (1) the nonnormality of the $\bar{X}_i(m)$'s, and (2) the initial transient condition.

B. Method of batch means

The method of batch means is to run a simulation of length n and divide these n observations into k batches each of length m.

There are three potential sources of error when using batch means to construct a confidence interval for a steady state mean [13]: (1) the nonnormality of batch means $\bar{X}_i(m)$'s; (2) the correlation between the batch means; and (3) the problem of the initial transient condition.

The method of batch means was first proposed by Conway [1] in dealing with the problem of variability of sample performance measurements. Conway considered two types of measurements; permanent entities and temporary entities. In collecting statistics on permanent entities, the predetermined time period determined the length of each batch, while for temporary entities the length of each batch was determined by the predetermined number of observations in collecting statistics. In proposing the method of batch means, Conway divided the entire run into alternating measurement and nonmeasurement periods in order to overcome the problem of covariance. However, it turned out that the variance became smaller by pooling the estimates of both periods. Thus, he concluded that the estimation should be based on the entire run without discarding results from intervening intervals. Law [13] also used this concept in his paper on comparing the accuracy of the methods of batch means and replication.

The comparison between the method of replication and that of batch means has been extensively explored by Law [13]. Using accuracy (proportion of 90% confidence intervals which actually contained population mean) as the criterion, he concluded that the method of batch means is better than that of replication. The main reason for this conclusion was that the effect of initial transient condition present in the method of replication produced biased sample means. Confidence intervals around these biased sample means naturally does not give high accuracy. The problem of the initial transient condition in the method of batch means is less serious because the problem exists only in the beginning of the simulation.

The problem of the initial transient condition has been analyzed by Conway [1] and Law [13]. Law [13] summarized three general approaches to the problem of initial transient condition: (1) make each run long enough so that the bias in sample mean will be diluted; (2) delete some observations from the beginning; and (3) start the simulation in a state which is more representative of the steady state distribution. However, these approaches are not simple and easy tasks to undertake. First of all, there is no technique for deciding "how long is long enough". Neither is there any definitive techniques for deciding how many observations to delete. Conway [1] gives the following rule of thumb: "truncate a sequence of measurements until the first set of the series is neither the maximum nor minimum of the remaining set." This is, however, only a rule of thumb, and there is no guarantee that it will work effectively all the time. There are some techniques suggested for starting the simulation in a state which is representative of the steady state distribution [13]. One of them is to estimate the state probability distribution from pilot runs. However, it is mentioned by Law [13] that in the case of a complex real world simulation, this procedure would appear to be a formidable task and furthermore would introduce the problem of correlation between the pilot runs.

C. Regenerative process method

The regenerative process method has been introduced by Cox and Smith [2], and extensively tested by Fishman [7] [8] [9] and Crane and Iglehart [3] [4] [5]. This method is based is based on the observation that a queueing process with certain specifiactions can be viewed as a regenerative process. Cox and Smith [2] introduced the concept of tours. Let's define the state of the system as the number of jobs in the system. When an event causes the system to be in state j, a tour is begun. A departure from state j and a subsequent return completes the tour. Kabak [12] made the following statements concerning this concept: When the service times are exponential and the arrivals are Poisson, then any arrival or departure that causes the system in state N (an arbitrary positive integer) starts an independent tour. When service times are nonexponential and the arrivals are poisson, then any departure that leaves the system empty causes an independent tour to begin. When neither the interarrival times nor the service times are exponential, an arrival to an empty system begins a tour.

In the experimental design of simulation, simulation users usually specify that simulation run either until the completion of n jobs or t time units. This specification of either n or t leaves open the possibility of bias due to both starting conditions and ending conditions. This bias can be removed if we get statistics on the busy period[1)] of the queueing system. Notice that every time the system leaves state zero, passage through different system states in the future is independent of the system's past behavior. It is this very concept of state independence that removes the covariances between the busy periods and that facilitates the estimation of variance in the regenerative process method.

Ⅲ. EXPERIMENTAL DESING

The method of batch means and the regenerative method are applied to two different queueing systems M/M/1 and M/M/4 with trafficintensity factors of .1, .5, and .9 for both systems. The following table shows the resultant six different queueing systems with specifications of service times (μ), interarrival times (λ) and traffic intensity factors (ρ).

1) For the definition of busy period, see Gross and Harris [11], p.84 and p.103.

	M/M/1			M/M/4		
μ	45	40	45	45	45	45
λ	50	90	450	12.5	22.5	112.5
ρ	.9	.5	.1	.9	.5	.1

In the design of sampling the GASP simulation language is used and five sets of runs are made to account for sampling errors using five different random number streams.[2] For each run, statistics are collected for 15 batches (in the case of the methed of batch means) and at most 15 busy periods (in the case of the regenerative process method). The sizes of each batch for the method of batch means are 2400, 1200, and 600 time units for the queuing system with traffic intensity factors of .1, .5, and .9 respectively. The size of each batch is determined as the above in order to give reasonable amount of interaction between the temporary entities and the system. For the regenerative process method, there are times when busy periods are so extremely short that there is very little interaction between the temporary entities and the system. This is especially true with the queueing system having a traffic intensity factor of .1. Thus, in order to get away with these extremely short busy period with little interaction, statistics for each period are based on more than one busy period. Statistics are collected for a number of busy periods until the number of temporary entities entering the system exceeds the predetermined number, thus one independent period in the regenerative process method consists of more than one busy period. The predetermined number of temporary entities are 10, 20, and 50 for the queueing systems with traffic intensity factors of .1, .5, and .9 respectively. Statistic are collected for up to 15 periods defined as above or up to 48,000 time units, whichever is smaller, in order to give some meaningful comparison with the method of batch means. The 48,000 time units are the maximum possible time period for the method of batch means.

Ⅳ. ANALYSIS

The comparison will be made on accuracy, precision, and execution time. Accuracy represents the proportion of confidence intervals that contains the population mean and is used by Law [13] to compare the performance of the methods of replicat ion and batch means. However, high accuracy with low precision will not be meaningful in estimating sample performance measurements. Precision is represented by the width of confidence interval and will be used as an additional criterion to compare the methods of collecting statistics.

The statistical results are shown in Table 1, Table 2, and Table 3. As shown in Table 1, there is not much difference in terms of accuracy between the method of batch means and the regener ative process method. For the M/M/1 queue with traffic intensity factor of .9, the regenerative process is generally better than the method of batch means, but for the M/M/4 queue with traffic intensity factor of .9, it is the ot her way around. In all other cases,the accuracy is 100% for both methods. Thus, no definite conclusions can be made.

2) Cemputer programs will be provided upon requests.

However, in terms of precision, the regenerative process method in general gives better results with some exceptions. Table 2 presents the results on precision. The measure of precision is based on one half of the average width of five confidence intervals for five different population means: mean time in the system (W), mean time in the queue (Wq), mean utilization rate (ρ), mean number in the system (L) and mean number in the queue (Lq). As shown in Table 2, for M/M/1 queue the precision is better in the case of the regenerative process method than in the case of the method of batch means. For M/M/4 queue the regenerative process method shows a better result in the case of .9 traffic intensity factor, but the method of batch means is better in the case of .1 and .5 traffic intensity factor for the M/M/4 queue. We may note however that the regenerative process method is much better in the case of the case of the M/M/4 queue with traffic intensity factor of 9 and very close to the performance of the method of batch means in the case of M/M/1 queue with traffic intensity factor of .1 and .5. Thus, we may conclude that the regenerative process method gives better precision.

All these discussions will no be complete without discnssing the exe cution time required for sampling statistics. Table 3 shows the required exeution time. This is obtained by using system subroutines TIMEON and TIMECK. For the traffic intensity factors of .1 and .5, the regenerative process method requires less time, but for the .9 traffic intinsty factor, the required time for the regenerative method more than two times what is required for the method of batch means. It may be recalled that even with more execution time the accuracy of the regenerative process method is lower than the method of batch means in the case of M/M/4 queue with $\rho=.9$. This may mean that in the high traffic in tensity factors, the method of batch means performs better than the regenerative method. It my be inthe case of low intensity factors that we should turn to the regenerative process method.

V. CONCLUSIONS AND SUGGESTIONS

The method of batch means and the regenerative process method are compared in terms of accuracy, precision and execution time. There is no definite conclusion that can be made about accuracy since one is better than the other in one case, but the other way around in the other case. In terms of precision, the regenerative process method in general shows better results. The regenerative procəss method required less execution time in the low (.1 and .5) traffic intensity factors, but more execution time in the high (.9) traffic intensity factors. It may be concluded that the method of batch means performs comparatively better in the higher traffic intensity factors than in lower ones. Overall, the regenerative process method indicates better performance than the method of batch means.

The above results are not conlcusive as only five sets of runs are made for each queueing system. Sampling errors will be smaller with more sets of runs. This extension will be the area of future research. In the design of sampling, number of periods (batch or busy period) de termined the length of each run. Howver, it will be interesting to fix time period of each run and to see the effect of the different length of batches in the method of batch meanson the statistical results as compared to the regenerative method. Law [13] examined [this effect between the method of batch means and the method of replication.

Table 1. MASURE OF ACCURACY (%)

		M/M/1					M/M/4				
		W	Wq	ρ	L	Lq	W	Wq	ρ	L	Lq
Butch Means	$\rho=.1$	100	100	100	100	100	100	100	100	100	100
	$\rho=.5$	100	100	100	100	100	100	100	100	100	100
	$\rho=.9$	60	60	100	60	60	80	80	100	80	80
Regenerative Method	$\rho=.1$	100	100	100	100	100	100	100	100	100	100
	$\rho=.5$	100	100	100	100	100	100	100	100	100	100
	$\rho=.9$	80	80	100	80	80	60	40	100	60	40

Table 2. MEASURE OF PRECISION

		M/M/1					M/M/4				
		W	Wq	ρ	L	Lq	W	Wq	ρ	L	Lq
Batch Means	$\rho=.1$	26.89	8.81	0.05	0.07	0.02	8.25	0.04	0.03	0.11	0.0004
	$\rho=.5$	45.79	38.12	0.17	0.62	0.48	6.98	3.60	0.13	0.47	0.16
	$\rho=.9$	169.62	164.53	0.36	3.75	3.63	72.3	69.36	0.09	7.43	5.39
Regenerative Process	$\rho=.1$	18.13	5.63	0.08		0.03	12.33	0.02	0.05	0.22	0.0004
	$\rho=.5$	37.45	31.43	0.15	0.56	0.43	10.94	4.72	0.13	0.72	0.47
	$\rho=.9$	149.95	148.55	0.09	3.20	3.12	31.46	28.85	0.08	2.63	2.36

Table 3. EXECUTION TIME REQUIRED (in 1/100 second)

		Batch	Regenerative
M/M/1	$\rho=.1$	232	168
	$\rho=.5$	267	223
	$\rho=.9$	299	652
M/M/4	$\rho=.1$	341	191
	$\rho=.5$	556	275
	$\rho=.9$	582	1,200

REFERENCES

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

2. D.R. Cox and W.L. Smith, *Queues*, Methuen, London, 1961. 3. M.A. Crane and D.L. Iglehart' "Simulating Stable Stochastic Systems, I : General Multiserver Queues," *Journal of Association of Computing Machinery*, 21, 103-113, (1974).

4. and "Simulating Stable Stochastic Systems, Ⅱ : Marko Chains," *Journal of Association of Computing Machinery*, 21, 114-123, (1974).
5. and "Simulating Stable Stochastic Systems, Ⅲ : Regenerative Processes and Discrete Event Simulations," *Operations Research*, 23, 33-45, (1975).
6. G.S. Fishman, *Concepts and Methodsin Discrete Event Simulation*, Wiley, New York, 1973.
7. , "Statistical Analysis for Queueing Simulation," *Management Science*, 20, 363-369, (1973).
8. , "Estimation in Multiserver Queueing Simulation," *Operations Research*, 22, 72-78, (1974).
9. , "Statistical Analysis of Multiserver Queueing Simulations," Technical Report 28-1, Department of Administrative Science, Yale University, September 1973.
10. "Estimating Sample size in Computer Simulation Experiments, *Management Science*, 18, 21-38, (1971).
11. D. Gross and C.M. Harris, *Fundamentals of Queueing Theory*, Wiley, New York, 1974.
12. I.W. Kabak, "Stopping Rules for Queueing Simulatinos," *Operations Research*, 16, 431-437, (1968).
13. Averill M. Law, "A Comparison of Two Techniques for Determining the Accuracy of Simulation Output," University of Wisconsin, June 1975.