SEQUENTIAL STEADY-STATE SIMULATION: RULES OF THUMB FOR IMPROVING ACCURACY OF THE FINAL RESULTS

Jong-Suk R. Lee[†], Krzysztof Pawlikowski[†] and Donald McNickle[‡]

[†]Department of Computer Science and [‡]Department of Management
University of Canterbury, Christchurch, New Zealand

e-mail: {ruth, krys}@cosc.canterbury.ac.nz

KEYWORDS

Sequential stochastic simulation, steady-state simulation, simulation output data analysis, statistical error of results.

ABSTRACT

It is generally accepted that sequential stochastic simulation is the only practical approach allowing control of the statistical error of the results of steady-state stochastic simulation. A commonly used stopping rule of such stochastic simulation is based on the relative statistical error, since then the magnitude of the point estimates does not need to be known beforehand. This relative statistical error of the simulation results can be measured by the ratio of the half-width of the confidence interval and the point estimate of an analysed performance measure. Sequential simulation is stopped when this ratio assumes a satisfactorily low value.

One of the problems of the sequential scenario is that the inherently random nature of the output data generated by any stochastic simulation can cause an accidental, temporary satisfaction of the stopping rule, resulting in acceptance of a wrong point estimate as the final simulation results. In this paper we consider two rules of thumb which could be used for improving the quality of the final results in practical applications of fully automated sequential simulation, and study their performance in three methods of sequential output data analysis.

1 INTRODUCTION

Any stochastic discrete-event simulation should be regarded as a (simulated) statistical experiment. Hence, statistical analysis of simulation output is mandatory. Otherwise, "... *computer runs yield a mass of data but this mass may turn into a mess* <if the random nature of such output data is ignored, and then> ... *instead of an expensive simulation model, a toss of the coin had better be used*" (Kleijnen 1979).

There are two different scenarios for determining the run-length of a steady-state stochastic simulation. Traditionally, the length of such a simulation experiment was set as an input to simulation programs. In such a *fixed-sample-size scenario* the final statistical error of the results is a matter of luck. This is no longer an acceptable approach. Modern methodology of steady-state simulation offers an attractive alternative, known as the *sequential scenario* of simulation or, simply, *sequential simulation*. Today, this scenario is recognised as the only practical approach allowing control of the statistical error of the final simulation results, since "... *no procedure in which the run length is fixed before the simulation begins can be relied upon to produce a confidence interval that covers the true theoretical value with the desired probability* ..." (Law and Kelton 1991). The error control available through sequential simulation also makes a further step towards automating the simulation process, for use by people who are not experts in simulation methodology.

Statistical errors associated with the final results of sequential stochastic simulation are commonly measured by the *relative statistical error*, defined as the ratio of the half-width of the confidence interval (CI) and the point estimate of an analysed performance measure. The simulation follows a sequence of checkpoints at which the relative statistical error of the current estimates is assessed. For example, in the case of simulation during which a mean value μ is estimated, when n observations (or output data) are available at a given checkpoint and the estimate of μ equals $\bar{X}(n)$, the relative statistical error of the mean is measured by $\epsilon(n) = \frac{\Delta(n)}{\bar{X}(n)}$, where $\Delta(n)$ is the current half-width of the CI for μ at the $(1-\alpha)$ confidence level; $0 < \alpha < 1$. If $\epsilon(n) \leq \epsilon_{max}$, where ϵ_{max} is the worst acceptable relative statistical error of the final results at the $(1-\alpha)$ confidence level, $0 < \epsilon_{max} < 1$, then the simulation can be stopped. Otherwise, the simulation is continued, and the relative statistical error of results is analysed again when the next checkpoint is reached. The advantage of using such relative measure of statistical errors is that simulators do not need to know the magnitude of the point estimates of performance measures they analyse.

The problem is that the inherently random nature of output data generated during any stochastic simulation can cause an accidental, temporary satisfaction of the stopping rule, resulting in acceptance of a wrong point estimate as the final result. This problem, and rules of thumb which can protect against the degradation of the coverage [1] of the final results in practical applications of fully automated sequential simulations, are discussed in Section 2. The performances of two heuristic rules assessed in different simulation output data analysis methods are presented in Section 3.

2 PROBLEM AND PROPOSED SOLUTION

As mentioned, the problem faced in practical applications of sequential steady-state simulations is that an assumed stopping criterion, for example, based on the relative statistical error, can be accidently satisfied too early, giving very inaccurate estimates of the analysed parameters. This happens due to random fluctuations in the estimated relative statistical error; see, for example, (Pawlikowski and de Vere 1993). At least a dozen methods have been proposed for analysing the CI of autocorrelated time-series of observations collected for studying the steady-state performance of various stochastic dynamic systems. A survey of methods used until 1990 can be found in (Pawlikowski 1990). Newer methods have appeared in (Charnes and Chen 1994; Fox et al. 1991; Goldsman and Kang 1991; Howard et al. 1992).

One of these methods, known as SA/HW (the method of spectral analysis as proposed in (Heidelberger and Welch 1981)), has been considered to be a good candidate for fully automated implementations of steady-state simulation, both in the case of a traditional simulation executed by a single computer (Pawlikowski 1990) and in the case of a distributed stochastic simulation executed on multiple computers of a LAN; see, for example, (Ewing et al. 1999). As with any other method of sequential analysis of simulation output data, SA/HW incorporates various approximations. Thus, to find the best tuning, one has to know its properties.

The theoretical studies of the properties of the CI generated by a given method of simulation output data analysis can reveal general conditions which have to be satisfied to secure the correct coverage, but correctness of any practical implementation of a specific method also has to be tested experimentally. The results of analysis of coverage of SA/HW have been reported in (Pawlikowski et al. 1998), together with the rules for *sequential coverage analysis*. These preliminary results of the coverage analysis under the SA/HW have shown that a satisfactory level of coverage of the final results produced by the SA/HW can be obtained if one ensures that the results produced by 'too short' simulation runs are discarded. In (Pawlikowski et al.1998), a simulation run was regarded as being 'too short' if the sequence of simulation output data it generated was shorter than the mean simulation run-length minus one standard deviation of the simulation run-length.

[1]In this paper, the coverage is defined as the experimental frequency with which the final confidence intervals contain the true (estimated) value. In the ideal situation, it should be equal to the assumed confidence level.

A closer look at the results associated with such 'too short' simulation runs has revealed that the coverage of simulation results obtained during such 'too short' simulation runs can be very poor indeed; see the third column in Table 1. Each of the results reported there was obtained on the basis of 3,000 independent replications of the steady-state simulation of the $M/M/1/\infty$ queueing system. The second and fourth column gives, respectively, the absolute and the relative number of 'too short' simulation runs in the total number of simulations executed at each load level of this queueing system. The next columns of this table report the minimum acceptable run-length of simulation (column five), and the mean simulation run-length, each over 3,000 replications (column six).

Table 1 clearly shows that the results obtained from 'too short' simulation runs, when simulation output data are analysed by means of the SA/HW, can be very wrong. One can see, too, that the problem becomes more critical in the case of higher loaded queueing systems, or, equivalently, in the case of processes with stronger autocorrelations.

The question is how, in practical applications of sequential simulation, one could try to avoid using the results obtained by 'too short' simulation runs. Therefore, we propose a simple rule of thumb which could help to eliminate acceptance of results from 'too short' simulation runs. Namely:

Rule 1

- execute R independent replications of a given sequential simulation and record their run-lengths (measured by the size of the sample of simulation output data)
- accept the results produced by the longest simulation run only.

Using the results presented in Table 1, one can assess the probability of a wrong decision, i.e. the probability that having applied Rule 1, one would still have the final results coming from a 'too short' simulation. Namely, if one executes R independent simulation replications, $R \geq 1$, then the resulting probability of a wrong decision will be P_{short}^R, with P_{short} being the probability that a simulation run is 'too short'. P_{short}^R is the probability of all R replications belonging to the class of 'too short' simulations. Generally speaking, P_{short} is hardly obtainable in simulation practices of real applications. But, for example, in the case of the $M/M/1/\infty$ queueing system our experimental results allow us to assume that a 'too short' simulation run in sequential steady-state simulations, with output data analysed by the SA/HW, can occur with the probability $P_{short} = 0.167$ or less; see Table 1, column 4, for $\rho = 0.4$. Thus, one can claim that the probability of using the final results originating from a still 'too short' simulation, when applying this rule of thumb with $R = 3$, is not larger than $0.167^3 = 0.005$. It drops to $0.167^5 = 0.00013$, or less, if one repeats the sequential simulation $R = 5$ times. Note that these results are similar to the one reported in (Lee et al. 1999a), in the context of sequential regenerative simulations. One can expect that similar results will characterize

Table 1: Run-length statistics from 3,000 independent simulation replications: SA/HW, M/M/1/∞, theoretical confidence level = 0.95

Load	*Number of short runs*	*Coverage*	*Probability(short)*	*Threshold*	*Mean of lengths*
0.1	0	N/A	0.0%	1341	1723
0.2	0	N/A	0.0%	1377	2002
0.3	133	83.5%	4.4%	1538	2475
0.4	500	80.2%	16.7%	1813	3278
0.5	297	69.7%	9.9%	2383	4670
0.6	364	63.7%	12.1%	3415	7247
0.7	307	53.1%	10.2%	5214	12727
0.8	236	46.2%	7.9%	9743	27906
0.9	263	38.0%	8.8%	33461	107049

simulation of any stochastic process with correlations similar to those of $M/M/1/\infty$ queueing system.

Although in any stochastic simulation, the relative width of the generated CI randomly changes with the number of collected observations, it has a general trend of shrinking. Thus, if information about the number of collected simulation output data is not available, one could try to order the lengths of simulation runs on the basis of the final relative statistical error of replicated simulation experiments. This gives us the following, alternative rule of thumb:

Rule 2

- execute R independent replications of a given sequential simulation and record the final relative statistical error of results
- accept the results produced with the smallest relative statistical error only, i.e., take the apparently most accurate result out of R results obtained.

Note that both rules suggest discarding R - 1 replications of R replications executed. This is certainly a significant diversion from the main concept of an automated sequential simulation that each such a simulation is run only once, even without a pilot run (Heidelberger and Welch 1983). On the other hand, such a rule, although in a different context, was proposed by D. Knuth in 1969, when he wrote that "... *the most prudent policy for a person to follow is to run each Monte Carlo program at least twice, using quite different sources of pseudo-random numbers, before taking the answers of the program seriously*" (Knuth 1969).

Of course, no rules of thumb can ensure that the final CI from a stochastic simulation will contain the theoretical value with a probability equal to the assumed confidence level. One of the ongoing research problems in the area of a sequential steady-state simulation is to find a method of simulation output data analysis which would be valid (in the sense of coverage) when one also applies it in a simulation of highly dynamic stochastic processes. This problem has been identified for all methods of simulation output data analysis whose coverage has been so far analysed sequentially, i.e. it characterises various versions of the method of batch means, the method of spectral analysis in its SA/HW version, and the regenerative simulation; see, for example, (Lee et al. 1999; Mota et al. 1999; Pawlikowski et al. 1998).

In the next section, we will use the results of our exhaustive studies of coverage produced by SA/HW, the non-overlapping batch means (NOBM), and the regenerative method (RM), for studying the effects of Rules 1 and 2 on the quality of the final results from the sequential steady-state simulation.

3 NUMERICAL RESULTS

Since an experimental investigation of the consequences of 'too short' simulation runs requires that the exact values of analysed parameters are known, here, as an example, we have selected the $M/M/1/\infty$ queueing system as the reference simulation model of our studies. This queueing system is notorious for strong autocorrelations of data in output sequences and long simulation runs required for achieving satisfactorily low level of statistical errors, and, because of this, it is commonly used as the reference simulation model in research on methods of simulation output data analysis (Schriber and Andrews 1981). All the numerical results in this section were obtained from sequential steady-state simulation runs of the $M/M/1/\infty$ queueing system, estimating the mean response time, with $\epsilon_{max} \cdot 100\% = 10\%$ as the upper level of the acceptable relative statistical error of the final results, at a confidence level of 0.95. In analysed cases each reported results was obtained from 2,000 independent simulation replications. For example, in the case of R = 5, we studied results of 10,000 replications.

The results presented in Figure 1 clearly show that Rule 1 (taking into account only the longest of R executed replications; R = 2, 3 and 5) is a viable policy, and the larger R is, the better the quality of the final results in terms of coverage. On the other hand, at least in the cases considered here, there is no need to assume R is larger than 3, since already for R = 3 the resulted coverage reaches its ideal level. As the statistical data of Table 1 show, the probability that the replication is still in the class of 'too short' simulation runs, after discarding two shorter replications out of three replications executed, drops to 0.005.

The same simulation results, but screened by applying Rule 2, are depicted in Figure 2. One can see that discarding results with larger (but acceptable) levels of relative statistical error *worsens* the coverage of the final results, regardless the number of replications executed. Similar results one can obtained for both the NOBM and the RM. This suggests that, due to the randomness of the simulation output data, a simulation stopped with a higher (but acceptable) relative statistical error is not necessarily a shorter one. Typically, in long simulation runs the convergence of the relative statistical error to its threshold value is slow but persistent. It is quite likely that a sudden and significant drop of the relative statistical error during a sequential simulation, temporarily causing satisfaction of the simulation stopping rule, can be associated with a 'too short' simulation run. Thus, Rule 2 should not be applied in simulation practise since it may lead to the acceptance of the results from a 'too short' simulation run, instead of their discarding.

The results of experimental studies of Rule 1 with different simulation output data analysis methods: the NOBM and the RM are depicted in Figures 3 and 4. The three methods of output data analysis appear to produce similar results.

4 CONCLUSIONS

This paper addresses the problem of an automated sequential stochastic simulation, caused by the sequential simulation being stopped too early, before a sufficiently long sample of output data has been collected. Having discussed the problem, its genesis and significance, we propose and study two rules of thumb that, if applied in simulation practise, can diminish the probability of having the results coming from such a prematurely finished simulation to a negligible level. Our experimental studies show that Rule 1 is a viable heuristic rule, which can be supported by the results of coverage analysis for the three methods of output data analysis in sequential steady-state simulation.

This rule of thumb can be easily implemented in simulation packages offering an automated control of the statistical error of the final results in sequential steady-state simulation, including a distributed sequential simulation under the Multiple Replications in Parallel (MRIP) scenario, as implemented in Akaroa-2 (Ewing et al. 1999), designed at the University of Canterbury, Christchurch, New Zealand.

REFERENCES

Charnes, J. M. and E. I. Chen. 1994. "Vector-autoregressive Inference for Equally Spaced Time-Averaged Multiple Queue Length Processes." *Proc. 1994 Winter Simulation Conf.*, (Lake Buena Vista, Florida, Dec. 1994): 312-315.

Ewing, G.; K. Pawlikowski; and D. McNickle. 1999. "Akaroa-2: Exploiting Network Computing by Distributed Stochastic Simulation." *13th European Simulation Multiconf.*, (Warsaw, Poland, June 1999): 175-181.

Fox, B. L.; D. Goldsman; and J. Swain. 1991. "Spaced Batch Means." *Operations Research Letters*, 10: 255-263.

Goldsman, D. and K. Kang. 1991. " Cramer-von Mises Variance Estimators for Simulations." *Proc. 1991 Winter Simulation Conf.*, (Phoenix, Arizona, 1991): 916-920.

Heidelberger, D. and P. A. Welch. 1981. "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations." *Comms. of the ACM*, 25: 175-181.

Heidelberger, D. and P. A. Welch. 1983. "Simulation Run Length Control in the Presence of an Initial Transient." *Operations Research*, 31: 1109-1144.

Howard, R. B.; M. A. Gallagher; K. W. Bauer; and P. S. Maybeck. 1992. "Vector-autoregressive Inference for Equally Spaced Time-Averaged Multiple Queue Length Processes." *Proc. 1992 Winter Simulation Conf.*: 586-593.

Kleijnen, J. P. C. 1979. "The Role of Statistical Methodology in Simulation." In *Methodology in Systems Modelling and Simulation*, B.P.Zeigler et al. (eds), North-Holland, Amsterdam.

Knuth, D. E. 1969. *Art of Programming*. Vol. 2. Addison-Wesley.

Law, A. M. and W. D. Kelton. 1991. *Simulation Modelling and Analysis.* McGraw-Hill.

Lee, J. R.; D. McNickle; and K. Pawlikowski. 1999. "Quality of Sequential Regenerative Simulation." *13th European Simulation Multiconf.*, (Warsaw, Poland, June 1999): 161-167.

Lee, J. R.; K. Pawlikowski; and D. McNickle. 1999a. "Do Not Trust Too Short Sequential Simulation." *1999 Summer Computer Simulation Conf.*, (Chicago, Illinois, July 1999): 97-102.

Mota, E.; A. Wolisz; and K. Pawlikowski. 1999. "Sequential Batch Means Techniques for Mean Value Analysis in Distributed Simulation." *13th European Simulation Multiconf.*, (Warsaw, Poland, June 1999): 129-134.

Pawlikowski, K. 1990. "Steady-State Simulation of Queueing Processes: A Survey of Problems and Solutions." *ACM Computing Surveys*, **2**: 123-170.

Pawlikowski, K. and L. de Vere. 1993. "Speeding up Sequential Simulation by Relative Variance Reduction." In *Proc. 8th Australian Teletraffic Research Seminar, ATRS'93*, RMIT, Telstra, 203-212.

Pawlikowski, K.; V. Yau; and D. McNickle. 1994. "Distributed Stochastic Discrete-Event Simulation in Parallel Time Streams." *1994 Winter Simulation Conf.*, (Lake Buena Vista, Florida, December 1994): 723-730.

Pawlikowski, K.; D. McNickle; and G. Ewing. 1998. "Coverage of Confidence Intervals in Sequential Steady-State Simulation." *Simulation Practice and Theory*, **6**(2): 255-267.

Schriber, T. J. and R. W. Andrews. 1981. "A Conceptual Framework for Research in the Analysis of Simulation Output." *Comms. of the ACM*, **24**(4) (April): 218-232.

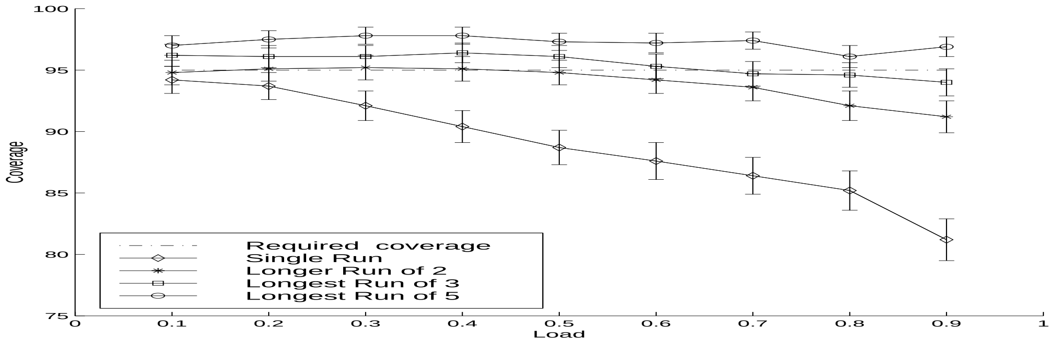

Figure 1: Coverage of the final simulation results with Rule 1, for R = 1, 2, 3 and 5. SA/HW, analysis of the mean response time in the $M/M/1/\infty$ queueing system

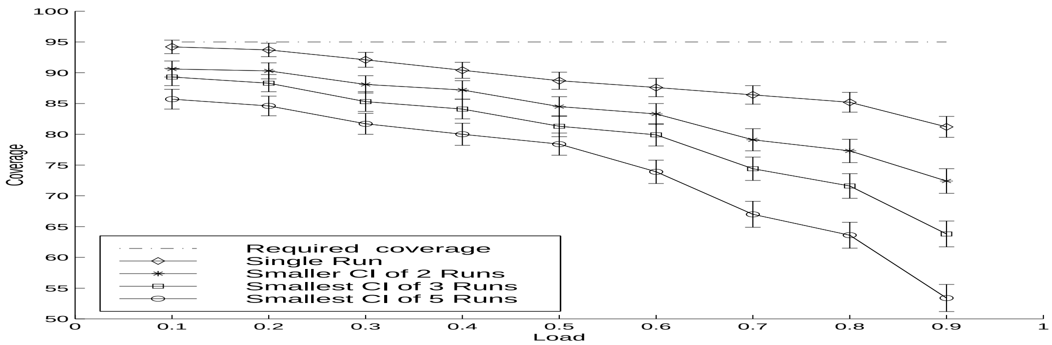

Figure 2: Coverage of the final simulation results with Rule 2, for R = 1, 2, 3 and 5. SA/HW, analysis of the mean response time in the $M/M/1/\infty$ queueing system

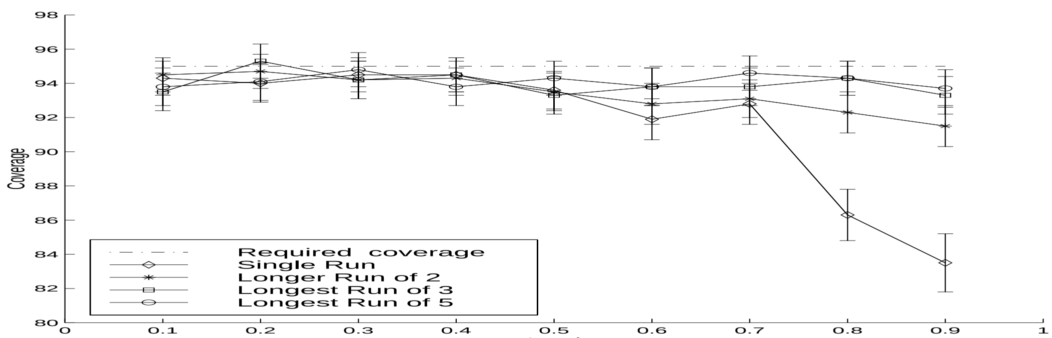

Figure 3: Coverage of the final simulation results with Rule 1, for R = 1, 2, 3 and 5. NOBM, analysis of the mean response time in the $M/M/1/\infty$ queueing system

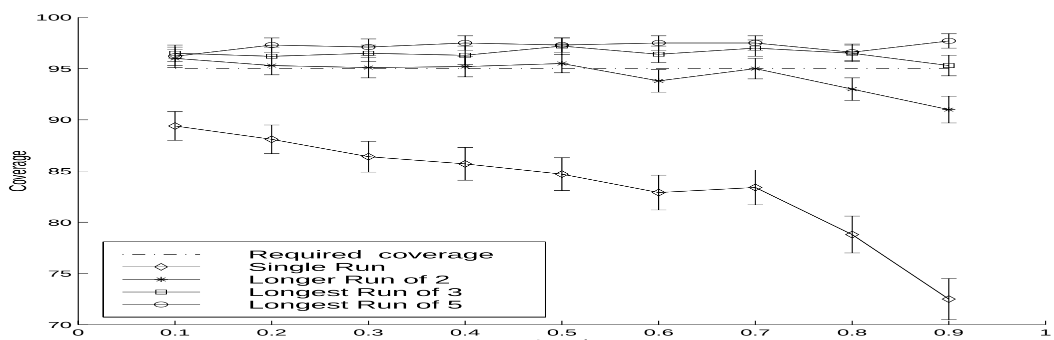

Figure 4: Coverage of the final simulation results with Rule 1, for R = 1, 2, 3 and 5. RM, analysis of the mean response time in the $M/M/1/\infty$ queueing system