Sequential batch means techniques for mean value analysis in distributed simulation

Edjair Mota
Adam Wolisz
Technical University of Berlin
Krzysztof Pawlikowski
University of Canterbury

March 16, 1999

Abstract

In this paper we stress the necessity of design of efficient procedures for constructing confidence intervals of mean values in a simulation environment of multiple replications in parallel. By using the popular batch means method, we implemented and investigated sequential variants that can give us an attractive speedup in simulation experiments of dynamic complex systems, such as communication networks, and at the same time the accuracy is guaranteed. Empirical results showing their statistical properties in such distributed environment are presented.

1 Introduction

Constructing confidence interval for steady-state mean values of a simulated stochastic process has been an attractive research area. At least a dozen of confidence interval procedures (CIPs) have been proposed in the literature (refer to [9] for a survey of these methods), but they were conceived having in mind a single-processor environment. Considering that stochastic discrete-event simulation of complex dynamic systems, such as modern communication networks, require large amount of computer time in order to yield statistically accurate estimates, one needs to make the experiments more efficient either by means of statistical aproaches (e.g. variance reduction techniques), or by making use of more computational resources.

Concerning the latter, Multiple Replications in Paralel (MRIP) is an interesting alternative for speeding up stochastic simulations. However, CIPs under this approach need to be investigated and their performance quantified, since due the effects of parallelisation, linear combination of estimators, which are used in MRIP, can converge to the wrong quantity [5]. In order to find a good CIP under MRIP for a specific range of applications, it is necessary to carry out comparative studies of their performance as the number of processors increases.

2 Sequential CIPs

The necessity of sequential CIPs for analysing simulation output data has been cleared up in [6], where it was stated that "*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 steady-state mean with the desired probability level*". In sequential procedures the run length is a random variable itself as it depends on the outcome of the observations, and one no longer has direct control of the amount of simulation time, which imposes statistical difficulties.

AKAROA.2, a user-friendly package for running distributed quatitative stochastic simulation, is an MRIP implementation designed at the Department of Computer Science of the University of Canterbury, in Christchurch, New Zealand, for full automatic parallelisation of common sequential simulations for concurrent execution on a network of workstations. An instance of a sequential simulation model is launched on a number of workstations connected via a network, and a central processor takes care of collecting asynchronously the observations from each processor and run a CIP for analysing all observations.

Currently, AKAROA.2 supports two sequential CIPs based on spectral analysis and on non-overlapping batch means (NOBM). Performance comparison of these two CIPs under MRIP can be found in [8]. However, spectral analysis methodology requires a considerable background from the user, and non-overlapping batch means, despite of its simplicity of understanding and implementation, offers an additional burden concerning the batch size determination (BSD) phase, that can not be parallelized under MRIP [2].

Thus, one can try to reduce this BSD phase by means of more efficient heuristic procedures for determining the optimal batch length and compensate that burden by making more efficient use of the data during the estimation phase. From this perspective, we look at a sequential version of *Spaced Batch Means* (SBM), proposed by Fox et al. [3], since it tries to diminish the correlation among the batch means by some observations between consecutive batches, and a sequential version of *Overlapping Batch Means*, proposed by Meketon and Schmeiser [7], which reuses some observations of each batch for constructing the next one.

3 Experimental design

The main analytical problem of simulation output data analysis is usually caused by strong correlation between events in typical simulated processes. Neglecting the existence of statitical correlation can result in an excessively optmistic confidence intervals. Batch Means approach divides a simulation output data into groups of observations and computes the batch means for an appropriate batch size, these batch means are supposed to be uncorrelated; a formal justification can be found in [1]. In order to find whether the batch means can be considered uncorrelated at a specified signicance level β ($0 < \beta < 1$), AKAROA.2 estimates all L autocorrelation coefficients of lag k (k = 1, 2, ..., L) by means of jackkknife estimators, which are usually less biased than the ordinary autocorrelation coefficient estimators [9]. L should not exceed 10% of the number of observations.

SBM tries to attenuate the serial correlations among the observations by discarding some of them bewteen contiguous batches. When the amount $\boldsymbol{s}$ of discarded observations is $\boldsymbol{s}$=0, we have the conventional NOBM. Preliminary experiments showed us that the greater $\boldsymbol{s}$, the better is the coverage of results, but that imposes an obvious problem, namely throwing out many observations. We have adopted a spacing equal 20% of the initial batch size but our feeling is that this amount of discarding should be determined according to the underlying stochastic process.

OBM is based on the assumption that batch size is the crucial problem rather than the independence of batches. By grouping observations into a batch, if one reuses some observations from the previous one, one can reduce the variance of the estimator by a factor of 2/3. Preliminary experiments confirmed Welch's assertion [13], that with modest amounts of overlapping most variance reduction can be achieved. Thus, after finding an optimal batch size **M**, we have applied an overlap of 25%.

Schmeiser's studies [11] pointed out that after finding the optimal batch size **M**, for a certain number **N** of observations, it would not be adivisable obtaining very large number of batches and these observations should be grouped into ***B*** batches (10 ≤**B** ≤30). In terms of sequential analysis it means that whenever the CIP collects more observations to acquire a specific relative precision for stopping the simulation experiment, it should rearrange the distribution of the batches in order to keep that range of number of batches. That's the way Akaroa.2 currently implements NOBM. Nevertheless, Glynn and Whitt [4] established that there is no estimation procedure based on a fixed number of batches that is consistent, i.e., sequential CIPs based on batch means should have the number of batches increased as the run length increases. For the sake of comparison, we have implemented a variant of non-overlapping batch means (here called NOBM/GW). Preliminary experiments showed us that one should agroup the the observations according to Schmeiser's findings (e.g. **B**=30), but **B** increases slightly as long as the required precision is not yet achieved. We have found suitable an increase as short as 2.

4 Performance comparison

The main criterion for comparing the performance of CIPs is the coverage, defined as the frequency with which the final confidence intervals contain the true system parameter being estimated.

A common practice is to assess the coverage just for a confidence level, e.g. $\eta = 0.90$, but some important information can be lost since different CIPs perform differently under different values of η. Schruben [12] suggested the construction of an empirical distribution function $\mathrm{G}_{\eta^*}(\eta)$, obtained with the observed values of confidence coefficient η^*, for different values of nominal confidence coefficient η. $\mathrm{G}_{\eta^*}(\eta)$ is an increasing function of the confidence level bounded above by 1, thus, as long as the nominal η is set high enough a robust CIP should produce nearly perfect coverage. Ideally, $\mathrm{G}_{\eta^*}(\eta)=\eta$, but frequently either positive correlation among observations is ignored and the final η is overestimated $(\mathrm{G}_{\eta^*}(\eta) < \eta)$, or negative correlation among observations is ignored and η is understated $(\mathrm{G}_{\eta^*}(\eta) > \eta)$, which can lead to wasting time in sequential simulation output analysis.

Another common practice is the assessment of coverage on a fixed-sample basis, but we are of the opinion that coverage analysis should also be run sequentially. Therefore, we have applied the sequential coverage analysis proposed by Pawlikowski et al. [10], to find the coverage for a set of confidence levels ranging from $\eta = 0.10$ to $\eta = 0.95$. The sequential analysis started after collecting a minimum number of bad confidence intervals, and discarding too short runs; it stopped when the relative precision of the confidence intervals for the coverage is less than 5%. Thus, expected value of the confi-

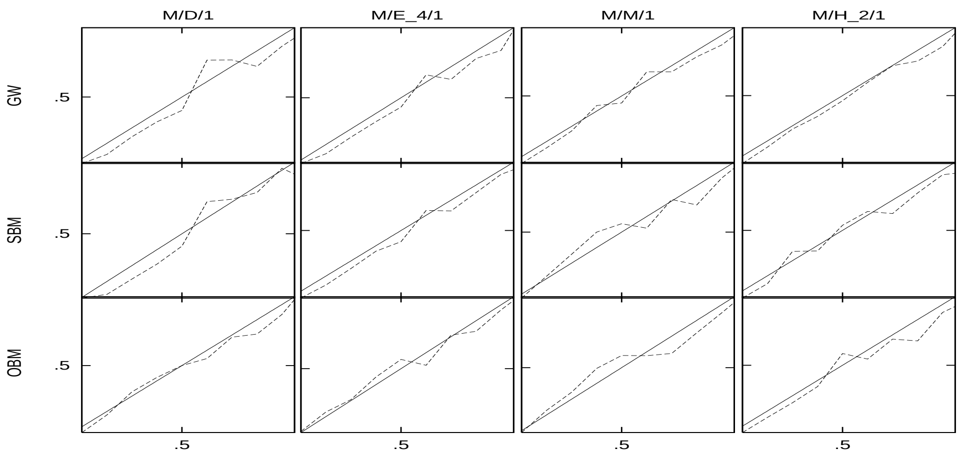

Figure 1: Coverage function

dence interval half-width, which measures the accuracy of the estimation, is not at issue in our experiments.

The coverage analysis can be applied only to sytems with theoretically well-known behavior, therefore, we have used four queueing models with different coefficient of variation (C_X) of the service process, for estimatimating the mean waiting time : M/D/1 ($C_X = 0$), $M/E_4/1$ ($0 < CX < 1$), M/M/1 ($C_X = 1$), $M/H_2/1$ ($C_X > 1$), each one with a FCFS priority rule.

Fig. 1 shows an example of such a coverage function for the three CIPs, for a system load $\rho = 0.90$, using 10 processors. For queueing systems with almost deterministic service process (i.e. $C_X \leftarrow 0$, such as M/D/1 and $M/E_4/1$, NOBM/GW and SBM yielded coverages below the expected, for small and very high values of confidence level, whereas OBM alternated above and below the nominal coverage, but always a little closer. As C_X increases, GW stabilizes around the nominal coverage, while SBM and OBM oscilates around them, but results are still acceptable. Besides that, for very high values of confidence levels, the three CIPs present convergence to the expected coverage, which points them out as promissing CIPs under MRIP, at least in terms of this main performance criterium.

Besides coverage, we are interested in the average run length and average number of observations as the number of processors increases. The following table summarize an experiment which used $\rho = 0.90$ and $\eta = 0.95$, for a number of processors varying from

CIP	NOBM/GW					SBM					OBM				
proc	2	4	6	8	10	2	4	6	8	10	2	4	6	8	10
M	0.89	0.86	0.86	0.85	0.88	0.91	0.91	0.96	0.90	0.88	0.93	0.91	0.92	0.92	0.93
D	0.48	0.51	1.00	1.58	3.88	0.51	0.98	3.18	2.62	1.90	0.55	0.96	1.35	2.59	3.57
1	45.56	22.30	7.29	10.20	4.13	73.53	18.51	6.57	7.48	4.12	28.78	19.78	17.70	6.22	5.15
M	0.86	0.86	0.84	0.91	0.94	0.87	0.92	0.91	0.90	0.90	0.89	0.92	0.94	0.93	0.93
$\boldsymbol{E_4}$	0.45	0.79	1.29	4.28	5.97	0.54	1.86	2.39	2.41	2.97	0.40	1.60	1.89	3.43	4.71
1	55.17	23.56	14.03	12.09	7.72	64.77	16.01	12.72	6.58	6.47	70.91	20.52	13.36	10.76	8.41
M	0.84	0.84	0.90	0.89	0.90	0.85	0.89	0.91	0.90	0.91	0.90	0.90	0.93	0.86	0.91
M	0.65	1.22	3.78	4.07	5.77	0.68	2.44	4.12	3.49	5.50	2.17	1.50	2.77	1.91	5.25
1	104.42	32.46	26.76	22.58	15.54	85.02	43.78	13.48	7.76	7.43	122.99	21.79	19.93	11.03	8.67
M	0.84	0.86	0.92	0.80	0.91	0.84	0.90	0.82	0.87	0.88	0.86	0.84	0.93	0.94	0.89
$\boldsymbol{H_2}$	1.10	5.09	4.82	3.59	8.76	1.04	3.52	2.71	8.51	8.13	0.99	2.13	6.45	7.52	8.94
1	94.57	48.91	24.08	27.86	13.43	88.23	56.20	17.89	13.80	11.06	94.40	61.63	24.76	19.60	16.80

Table 1: Sequential CIPs performance comparison

2 to 10. The first raw stands for the coverage, the second raw stands for the average number of total observations (x10^6) collected until a required precision is achieved, and third raw stands for the average run length (in seconds), measured from the time the simulation experiment initated to the time Akaroa.2 stopped the experiment since the required precision was achieved and, thus, no more observations are required.

Considering that a simulation is always an approximation to the corresponding real-world system, we believe that the estimated coverage $G_{\eta^*}(\eta)$=0.88 is close enough to the desired 0.95 to be useful. According to this somewhat arbitrary criterion, OBM performed near to the expected (just three times,among 20, its coverage was below this low bound), followed by SBM (five times) and NOBM/GW (eleven times).

In terms of the average total number of observations, there was no signifcative out-performance from one of the CIPs, but a detailed analysis can show that, relatively, NOBM/GW required fewer observations, which maybe could explain its poor overall performance in terms of coverage, since it stopped the simulation based on an erroneous decision that the amount of observations was enough. It is worthwhile to note that SBM outperformed the two toher when the number of processors was 10, and that had no serious impact on the coverage.

Concerning the average run length, by all means there was an attractive speedup. SBM behaved somewhat better when C_X=0 (M/D/1), whereas OBM somewhat better when C_X=1 (M/M/1).

5 Conclusions and final remarks

Batch Means has been used frequently in simulations experiments and its popularity dues to its simplicity of conception and ease of implementation. Nevertheless, it offers an aditional burden for determining the optimal batch size that yields acceptable correlation among the batch means. In distributed environment this additional burden can not be parallelized and the analyst should look for ways of relieving it. We have implemented and investigated the performance of three interesting sequential CIPs based on batch means under multiple replications in parallel (MRIP). They have offered an attractive speedup as the number of processors increases, but at the same time the coverage of the results achieved a desired niveau (at least for the CIPs based on *Spaced Batch Means* and on *Overlapping Batch Means*, which suggests they are promising candidates for being robust under MRIP.

Application of computational efficient CIPs to networks with multiple queues and servers is difficult but this is an important topic for research. Another issue is the asymptotic behavior of these CIPs under MRIP, in order to antecipate their behavior as the run length increases.

References

[1] BRILLINGER, D.R., "*Estimation of the mean of a stationary time series by sampling*", Journal of Applied Probability, Vol. 10, 419-431, 1973.

[2] EWING, G., McCNICKLE, M. and PAWLIKOWSKI, K., *"Multiple replications in parallel: Distributed generation of data for speeding up quantitative stochastic simulation*", 15th IMACS, 397-402, Berlin, August 1997.

[3] FOX, B.L., GOLDSMAN, D. and SWAIN, J.J., *"Spaced Batch Means*", Operations Research, Vol.10, 255-263, 1991.

[4] GLYNN, P.W. and WHITT, W., *"Estimating the asymptotic variance with batch means*", Operations Research Letters, 431–435, 1991.

[5] HEILDELBERGER, *"Discrete event simulations and parallel processing : Statistical properties"*, SIAM J. Stat. Comput., Vol. 9, No. 6, 1114-1132, November 1988.

[6] LAW, A.M. and KELTON, W.D., *"Confidence intervals for steady-state simulations, II: A survey of sequential procedures*", Management Science, Vol.28, no.5, May 1982, 550-562.

[7] MEKETON, M.S. and SCHMEISER, B., *"Overlapping batch means: something for nothing ?"*, Proceedings of the 1984 Winter Simulation Conference, 227-230, 1984.

[8] MOTA, E., WOLISZ, A. and PAWLIKOWSKI, K., *"Experience with AKAROA : Speeding up stochastic stochastic simulation by using multiple replications in parallel*, Advanced Simulation Technologies Conference, Boston, pp. 296-301, 1998.

[9] PAWLIKOWSKI, K., *"Steady-state simulation of queueing processes: a survey of problems and solutions"*, ACM Computing Surveys, Vol.22, 123-170, 1990.

[10] PAWLIKOWSKI, K., McCNICKLE, D. and EWING, G., *"Coverage of confidence intervals in sequential steady-state simulation"*, in Simulation Practice and Theory, 6(1998), pp. 255-267.

[11] SCHMEISER, B., *"Batch size effects in the analysis of simulation output",* Operations Research, Vol.30, no.3, 556-598, May-June 1982.

[12] SCHRUBEN, L.W., *"A coverage function for interval estimators of simulation response"*, Management Science, Vol. 26, 18-27, 1980.

[13] WELCH, P., *"On the relationship between batch means and overlapping batch means"*, Proceedings of the 1987 Winter Simulation Conference, 320-323, 1987.