

FAST AUTOMATED ESTIMATION OF VARIANCE IN SEQUENTIAL DISCRETE EVENT STOCHASTIC SIMULATION

Don McNickle
Management Department
University of Canterbury
Private Bag 4800, Christchurch
New Zealand
Don.McNickle@canterbury.ac.nz

Krzysztof Pawlikowski and Nelson Shaw
Computer Science and Software Engineering
University of Canterbury
Private Bag 4800, Christchurch
New Zealand
Krys.Pawlikowski@canterbury.ac.nz
Nelson.Shaw@pg.canterbury.ac.nz

KEYWORDS

On-line variance estimation; Quantitative discrete event simulation; Sequential simulation.

ABSTRACT

On-line analysis of output data from discrete event stochastic simulation focuses almost entirely on estimation of means. Most “variance estimation” research in simulation refers to the estimation of the variance of the mean, to construct confidence intervals for mean values. There has been little research on the estimation of variance in simulation. We investigate three methods for point and interval estimates of variance and discuss an implementation of the best technique in an extended version of Akaroa2, a quantitative stochastic simulation controller.

1. INTRODUCTION

Current methodology for the analysis of output data from stochastic simulation focuses mainly on the estimation of means, and to a much lesser extent, on quantiles. Higher moments have not been considered. Research on “variance estimation” in simulation almost always refers to the estimation of the variance of the mean, in order to construct confidence intervals for mean values. The only previous research we could find on the direct estimation of the variance of correlated simulation data is an unpublished report by Deuermeyer, Feldman and Yang (1996). In communication networks variance of packet delays is an important performance measure e.g. for Voice-over Internet, as changes in the delay of packet delivery can have a significant degenerative effect on quality of service. Applying stochastic simulation in studies of new solutions for next generations of telecommunication networks will require reliable estimates of variance and its statistical error.

In Schmidt, Pawlikowski and McNickle (2008) (see also Schmidt, Pawlikowski and McNickle (2009)), three possible methods of variance estimation were identified. Here we focus on the testing and development of a reliable, accurate and efficient algorithm for sequential assessment of variance from simulation output data. The

possible estimators of steady-state variance are described in Section 2. Section 3 reports the results of their testing, which involved running over 320,000 simulations, each involving up to 36,000,000 observations. Finally, the technical issues of implementing the selected estimators of variance in an extended version of Akaroa2, a quantitative stochastic simulation controller (McNickle, Pawlikowski and Ewing 2010), are discussed in Section 4.

1.1. Variance Estimation

We want to find point and interval estimates for the variance of a simulated process $X_1, X_2, \dots, X_j, \dots$. If the simulation output data represents *independent* and *identically distributed* random variables, the usual well-behaved point estimate for the variance is:

$$S^2(n) = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}(n))^2, \quad (1)$$

where $\bar{X}(n)$ is the average of the first n observations.

From [5] the variance of $S^2(n)$ is:

$$\text{Var}[S^2(n)] = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right), \quad (2)$$

where σ^2 is the steady-state variance and μ_4 the fourth central moment of the steady-state distribution. Thus, provided μ_4 exists, the point estimator will converge to its theoretical value.

In simulation, however, observations are often highly correlated. In an M/M/1 queue with a traffic intensity of 0.9, for example, the correlation between successive waiting times is 0.99043 (Daley, 1968). Now $S^2(n)$ is only asymptotically unbiased, and for interval estimates Equation (2) no longer applies. Alternative estimators are needed, which are discussed in Section 2.

1.1.1. Coverage Analysis

The method we will use to measure the quality of the estimates is Coverage Analysis. Coverage is the relative frequency with which an estimated confidence interval, supposedly of a certain size (e.g. 95%), actually contains the true value of the parameter of interest. It has the advantage of simultaneously capturing the quality of both point and interval estimates. Thus good estimators should produce coverages close to (say) 95%. In this research, we followed the principles and guidelines of coverage analysis described in Pawlikowski, McNickle and Ewing (1998).

1.2. Akaroa2

Akaroa2 is an automated controller of both terminating and steady-state quantitative stochastic simulations, using the Multiple Replications in Parallel (MRIP) simulation scenario. In MRIP, identical simulation programs are run in parallel on multiple processors, which work as simulation engines producing multiple streams of output data. As shown in Pawlikowski and McNickle (2001), the resulted speedup of simulation is governed by a truncated version of Amdahl's Law. The current version of Akaroa2 guarantees fast and accurate on-line analysis of arbitrary numbers of means. It is downloadable, at no charge for academic purposes, from www.akaroa.canterbury.ac.nz. The results of this research will expand its functionality by including automated estimation of variances.

Users of Akaroa2 need to provide a simulation program of the system they are studying. The default language is C++. However, interfaces between Akaroa2 and a number of other languages and simulation packages are provided too. Observations collected during simulation are automatically submitted to the controller that is in charge of running the simulation and stopping it once a pre-defined stopping criterion (the acceptable level of statistical error of final results) is met.

Control parameters (such as the acceptable level of the final statistical error of results, and the number of computers involved in MRIP) are set by the user before the simulation is run.

In the case of terminating simulations only one method of variance estimation can be generally defined, that of independent replications, where a single value is taken from each replication of the system of interest, and that will be implemented in an extended version of Akaroa2.

2. STEADY-STATE SIMULATION

Finding accurate point and interval estimators of variance in steady-state simulation is much more challenging. Possible estimators have been described in Schmidt, McNickle and Pawlikowski (2008). Here we report on exhaustive testing of these over a wider range of models and parameters, and in particular we investigate the problem of batch mean sizes, which proves to be the Achilles Heel of an otherwise interesting method.

The first estimator discussed is based on independent replications, similar to the approach used in terminating simulation. The second estimator is based on variance as an expected value, and the last estimator is based on batch means. Each estimator is discussed further in its section.

One issue in steady-state simulation is that of dealing with the initial transient period. In the case of variance

analysis, no generally applicable method for initial transient period detection has been proposed. One method that could be applied after further research is the method proposed by Eickhoff (2008). Here we simply discard a very generous, fixed number of observations. Development of a sequential rule for detecting the length of initial transient period for estimators of steady-state variance has been left for further research.

2.1. Independent Replications

If somehow a sample of uncorrelated observations can be collected then $S^2(n)$ is an unbiased estimator of the variance and Equation (2) can be used to construct the corresponding confidence interval.

One way to ensure that the observations used are uncorrelated is to use independent replications of the simulation. From each replication only a single observation is taken from the steady-state phase of simulation. The very obvious disadvantage of this method is that many observations are discarded; hence the runtime of the simulation is likely to be very large. However we include the method for completeness.

2.2. Variance as an Expected Value

This estimator comes from the definition of variance as

$$\text{Var}[X] = E[(X - E[X])^2].$$

Thus, before analysis, the observations are first converted to:

$$Y_i = (X_i - \bar{X}(i))^2.$$

We rely on the fact that the sample variance, $S^2(n)$ from Equation (1), is at least an asymptotically unbiased estimator. By treating the required parameter as an expected value of some random variable, we can use existing mean value estimation techniques to calculate the variance. For confidence intervals we use *spectral analysis*, as described in Heidelberger and Welch (1981). This method of variance estimation produces good coverage results for mean values (see Pawlikowski, McNickle and Ewing (1998)). The conditions for the use of spectral analysis for variance are that the sum of the autocorrelations of $\{Y_i\}$ is finite, along with, of course σ^2 and μ_4 .

2.3. Batch Means

This technique is described in Deuermeyer, Feldman and Yang (1996) (a more accessible reference is Schmidt, Pawlikowski and McNickle (2008)). It provides a method of compensating for the bias of sample variance $S^2(n)$, which also lends itself to calculating a confidence interval. The basic idea behind calculating variance via batch means is to split the variance into two components: Local Variance – Mean variance inside batches, and Global Variance – Variance of means of the batches. The mean batch variance corrects the variance of the means of the batches to

produce an unbiased estimate. A curious property of this estimator is that unbiased estimators are obtained regardless of the size of the batches. Thus, the batch-size parameter can be left free to be determined by the requirements of the confidence interval calculation. The question is: does this independence of batch sizes extend, at least to some extent, to the calculation of confidence intervals?

2.4. Evaluation of the Estimators

2.4.1. Reference Models

We considered the single server queues: M/M/1, M/E₂/1 and M/H₂/1. For M/H₂/1, squared coefficients of variation for the service times, CV² = 5 and 50 were used. Use of the M/H₂/1 queue with CV² = 50 allows some investigation of how heavy-tailed service distributions can affect the performance of estimators. The property of each queue investigated is the waiting time. Generous fixed initial transient periods (up to 20,000 observations) were used to ensure that the output data represented steady state.

The Pollaczek-Khintchine transform formula $W_q(s)$ for the waiting time distribution (Gross and Harris 1985) was used to obtain the theoretical values of the variance of the waiting time. The n th moment of the waiting time distribution is found by differentiating $W_q(s)$ n times using Maple and evaluating the derivatives at $s = 0$.

2.4.2. Batch Means – The effect of batch size

The key issue for the batch means method is, as always, how to select appropriate batch sizes. It seems reasonable that a practitioner estimating variances will also be interested in mean values. So, as a starting point, we calculated batch sizes for the M/M/1 queue which result in batch means with an arbitrarily low serial correlation of 0.01. They were calculated from the results of Law (1977) and Daley (1968). Due to the need to calculate large numbers of autocorrelations when deriving batch sizes for given queuing systems, Law’s method can be very lengthy and prone to failure, taking up to an hour in Maple when it worked. Also repeated runs were necessary to find a batch size giving the required correlation between batch means.

Load	Batch Size	Correlation between successive batch means
0.1	30	0.0085
0.2	60	0.0093
0.3	100	0.0099
0.4	170	0.0097
0.5	300	0.0093
0.6	550	0.0092
0.7	1100	0.0093
0.8	2600	0.0099
0.9	11000	0.0099

Table 1: Initial Batch Sizes for the M/M/1 Queue

For this reason, the other queues used these batch sizes as initial values, but they were increased or decreased depending on the expected correlation. Table 1 shows the calculated batch sizes for the M/M/1 queue.

3. RESULTS

Here we give a few representative results of coverage analysis for the steady-state estimators of variance. We concentrate on the most challenging model, M/H₂/1. All coverage analysis results are calculated for 95% confidence intervals (a Required Coverage of 0.95) with a relative error of 5%. The simulations are repeated until this level of accuracy is achieved (this may require up to 15,000 simulations.) A brief discussion is also given of the specific performance aspects of each estimator.

3.1. Independent Replications

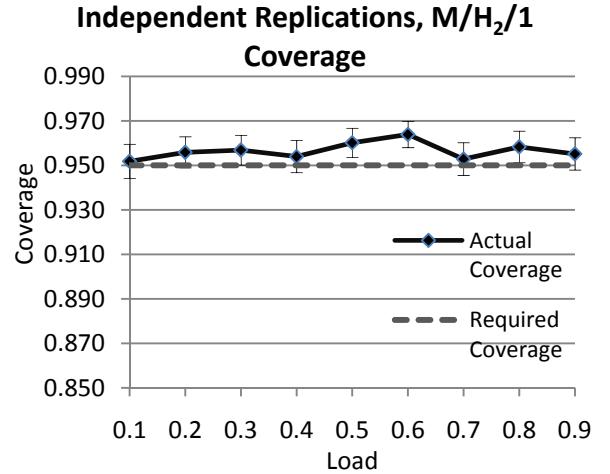


Figure 1: Coverage of the $\hat{\sigma}_{IR}^2$ estimator on an M/H₂/1 queue

This estimator provides satisfactory coverage for all loads investigated. The main issue with this estimator is its inefficiency. Because of the number of observations it discards (see Section 3.4) it cannot be recommended for practical use. Both its speed and accuracy are also totally reliant on the accurate detection of the initial transient period, and this problem has not been satisfactorily solved yet.

3.2. Variance as an Expected Value or Mean

The coverage analysis results for the Variance as an Expected Value estimator ($\hat{\sigma}_M^2$) are shown in Figure 2 and Figure 3. 20,000 observations were deleted to make sure that observations from the initial transient period are not included. Figure 2 and Figure 3 show that $\hat{\sigma}_M^2$ performs well in terms of coverage, for all loads. The apparent deterioration of coverage at the load of 0.7 in Figure 3 is probably due to the extremely high coefficient of variation of service times in this case producing erratic results. Other than in this case, the estimator performs well.

Another feature of this estimator, apart from its good coverage, is that it is easy to implement as it makes use of mean value estimators which have been thoroughly investigated. Run times of its sequential implementation are similar to the Batch Means method, with no substantial difference between them detected.

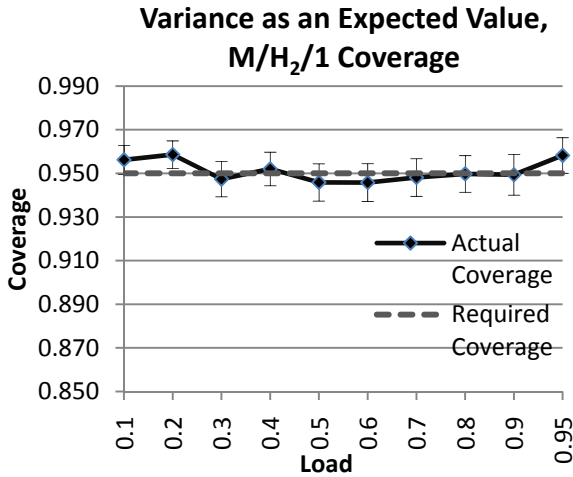


Figure 2: Coverage of the $\hat{\sigma}_M^2$ estimator on an M/H₂/1 queue, with CV² = 5

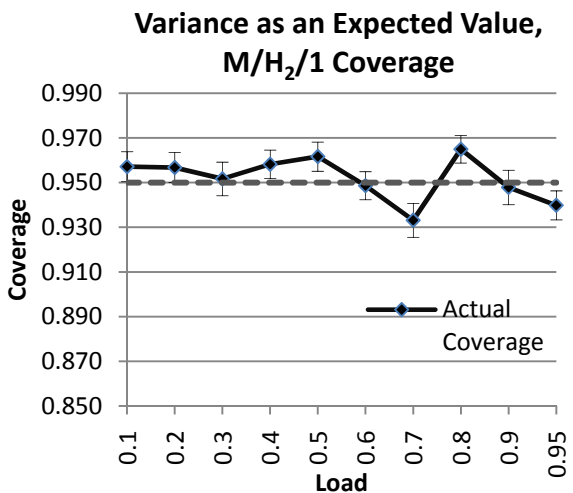


Figure 3: Coverage of the $\hat{\sigma}_M^2$ estimator on an M/H₂/1 queue, with CV² = 50

3.3. Batch Means

In this section, we consider the coverage of the Batch Means ($\hat{\sigma}_{BM}^2$) estimator, as a function of batch size. The results show that the estimator can perform well for every load and queue investigated, if the correct batch size is known. The results confirm the initial results given in Schmidt, McNickle and Pawlikowski (2009) and also show that this estimator provides good coverage for queues with larger coefficients of variation, again under the proviso that an appropriate batch size can be found.

3.3.1. M/E₂/1 Queue

The coverage of the Batch Means ($\hat{\sigma}_{BM}^2$) estimator for the M/E₂/1 queue with a load of 0.9 and varying batch sizes is shown in Figure 4. Good coverage is obtained if the batch size is large enough.

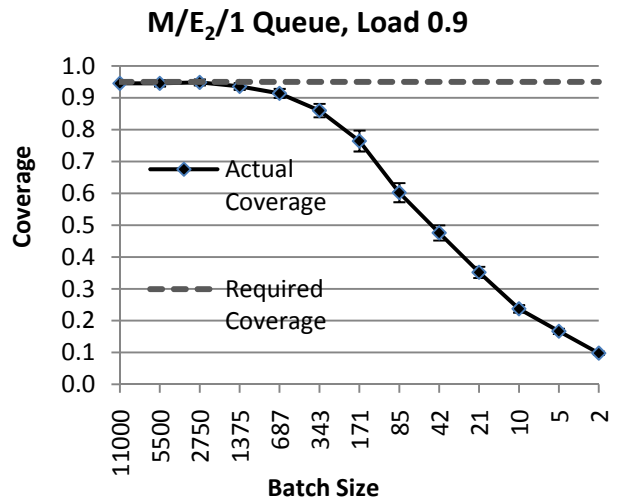


Figure 4: Coverage of the $\hat{\sigma}_{BM}^2$ estimator on an M/E₂/1 queue with load 0.9

3.3.2. M/M/1 Queue

The coverage of the Batch Means ($\hat{\sigma}_{BM}^2$) estimator on the M/M/1 queue with a load of 0.9 and varying batch sizes is shown in Figure 5. Again acceptable coverage can be found if the batch size is large enough, although the required batch size has now increased considerably.

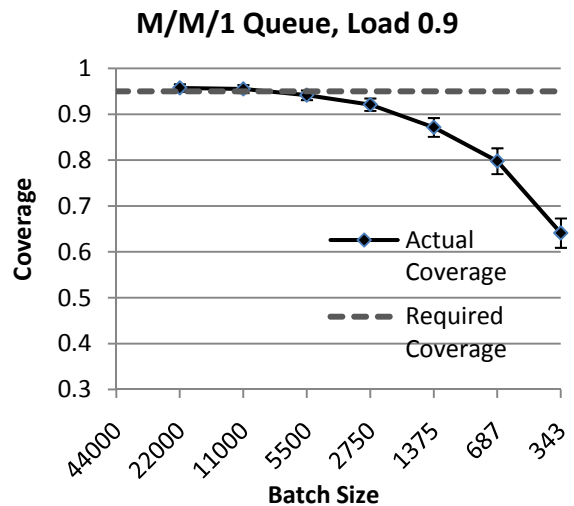


Figure 5: Coverage of the $\hat{\sigma}_{BM}^2$ estimator with varying batch sizes on an M/M/1 queue with load 0.9

3.3.3. M/H₂/1 Queue

The coverage of the Batch Means ($\hat{\sigma}_{BM}^2$) estimator on the M/H₂/1 queue with a load of 0.8 and varying batch sizes is shown in Figure 6. The two curves in this figure are for CV² of the service times of 5 and 50, respectively.

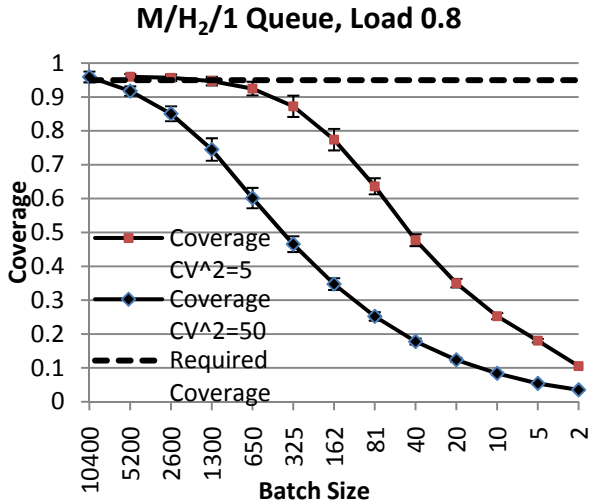


Figure 6: Coverage of $\hat{\sigma}_{BM}^2$ for M/H₂/1 queues with CV² = 5, 50 and load 0.8

In each of Figures 4, 5, and 6, there is a batch size at which the Batch Means estimator providing satisfactory coverage. The coverage drops off much more quickly for the queue with a higher coefficient of variation. This is because the correlation between batch means is higher, which causes the generated confidence intervals to be worse. Note that Figure 6 uses a traffic intensity of 0.8, as a traffic intensity of 0.9 turned out to require several weeks of computer time for coverage analysis of a single batch size. For a CV² of 50, even for a load of 0.8, however, satisfactory coverage is only just achieved within the range of batch sizes studied.

The key conclusion of this part of the study is that unlike the point estimator, the performance of a Batch-Means based interval estimator for variance does depend critically on getting the batch size right.

3.4 Run Lengths

Table 2 shows mean lengths of sequential simulations needed for obtaining the final results with statistical error of 5%, measured by the number of observations (or the number of replications in the case of Independent Replications)

	Independent Replications	Variance as Expected Value	Batch Means
M/E ₂ /1	12685	1257700	945447
M/M/1	12572	1684400	1333580
M/H ₂ /1(5)		5556790	4204000
M/H ₂ /1(50)		52140400	36281300

Table 2. Run lengths for a traffic intensity of 0.8

The first conclusion is that Independent Replications is going to be very inefficient. Even if we take a (very small) number of 1,000 observations as the length of initial transient period, the total required number of observations is an order of magnitude greater than those

for the other two methods. In addition determining when steady state has been reached remains a largely open question for estimators other than means. Accurate answers for the M/H₂/1 queues could not be obtained in reasonable time, so these runs were abandoned. Secondly, the numbers of observations for the other two methods are similar. They increase with CV of service times, and hence with the variance of waiting times. Thus, it is not surprising that these numbers are more than an order of magnitude bigger than the numbers required for estimating mean waiting times, illustrating the extra difficulty of variance estimation.

4. IMPLEMENTATING THE ESTIMATORS IN AKAROA2

For terminating simulations the Independent Replications estimator is the only feasible point and interval-estimator for the variance. In steady-state, however, although the estimator $\hat{\sigma}_{IR}^2$, based on Independent Replications produced good coverage, it was eliminated due to its poor efficiency and prohibitively long run lengths. This left us with two estimators, based on Variance as an Expected Value ($\hat{\sigma}_M^2$) and on Batch Means ($\hat{\sigma}_{BM}^2$). The $\hat{\sigma}_{BM}^2$ estimator was found to produce good coverage on every queue and load tested, if a good batch size could be found. This is the main problem with any batch means-based estimator, as determination of an appropriate batch size is crucially important to the quality of the results. While a number of batch size selection algorithms have been proposed, they have only been tested on mean values. The $\hat{\sigma}_M^2$ estimator also produced good coverage on each queue investigated, even on queues with high loads. One advantage of this estimator is that it is easy to implement as it makes use of existing mean value analysis modules. The main difference between the two methods is that Variance as Expected Value does not rely on the correct setting of an arbitrary parameter.

Once implemented into Akaroa2, the estimator must be able to work on complex systems as well, and it is difficult to see how a good batch size could be determined in such cases for the $\hat{\sigma}_{BM}^2$ estimator. As a result of this, the $\hat{\sigma}_M^2$ estimator was chosen to be implemented into an extended version of Akaroa2.

4.1. Issues of Implementation

In order to implement the variance analysis in Akaroa2, the global analyser needs to be modified; see Ewing, Pawlikowski and McNickle (2010). Fortunately, the estimation of variance by $\hat{\sigma}_M^2$ requires similar processes to those already implemented for estimating means. So the variance can be estimated by the ratio:

$$\hat{\sigma}_M^2 = \frac{\sum_i n_i \hat{\sigma}_i^2}{\sum_i n_i},$$

where $\hat{\sigma}_i^2$ is the local point estimate of the variance from the i^{th} simulation engine, and n_i the number of

observations used by that engine. The variance of the global estimate is calculated from:

$$Var[\hat{\sigma}^2] = \frac{\sum_i n_i^2 Var[\hat{\sigma}_i^2]}{(\sum_i n_i)^2},$$

where $Var[\hat{\sigma}_i^2]$ is the variance of the local variance estimate at the i^{th} simulation engine

Screenshots of a prototype graphical user interface of an extended Akaroa2, incorporating variance analysis, are shown in Figures 7 and 8. Figure 7 shows the “New Simulation” window, where the parameters for the simulation are set. In this case, the simulation will run until the mean and variance converge to the point where their 95% confidence intervals have a 5% relative error.

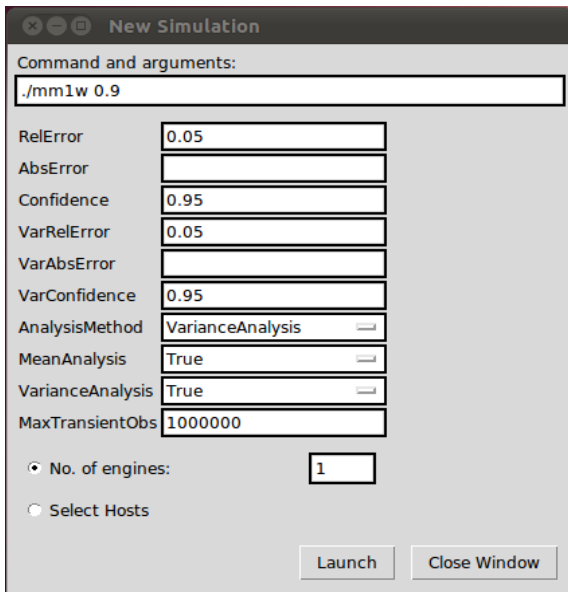


Figure 7: Prototype “New Simulation” window

Figure 8 shows the simulation in progress. The two bars show the mean and variance’s respective convergence to their required relative errors. The Global Estimates table gives information regarding the current global estimate of parameters: in this case: the mean-value and its relative error, and the variance and its relative error. The output of the simulation gives information on the mean (Mean) and its confidence interval half-width (Delta), as well as the variance (Variance) and its confidence interval half-width (Delta Variance). As usual with Akaroa2, additional simulation engines can be added at any time if the simulation of any of the parameters is converging too slowly (“Add Engines”). The system is also largely immune from the effects of processor failures, in that a failed processor is no different to one which is responding slowly. This property is particularly important given the long run-times required for variance estimation.

5. CONCLUSIONS

Three different sequential estimators for variance were tested extensively. The method of Independent Replications was eliminated as too inefficient. The remaining two estimators, based on Variance as an Expected Value, and Batch Means respectively, were thoroughly tested. Variance as an Expected Value produced satisfactory coverage and was noted as being particularly easy to implement. The Batch Means estimator was found to produce good coverage if a suitable batch size could be found. The inherent problem of Batch Means lies in determining the batch size, which will inevitably become more complicated as the complexity of the simulated model increases. As the Variance as an Expected Value estimator produces good coverage and does not require determination of any additional parameters, it has been selected to implement in an extended version of Akaroa2.

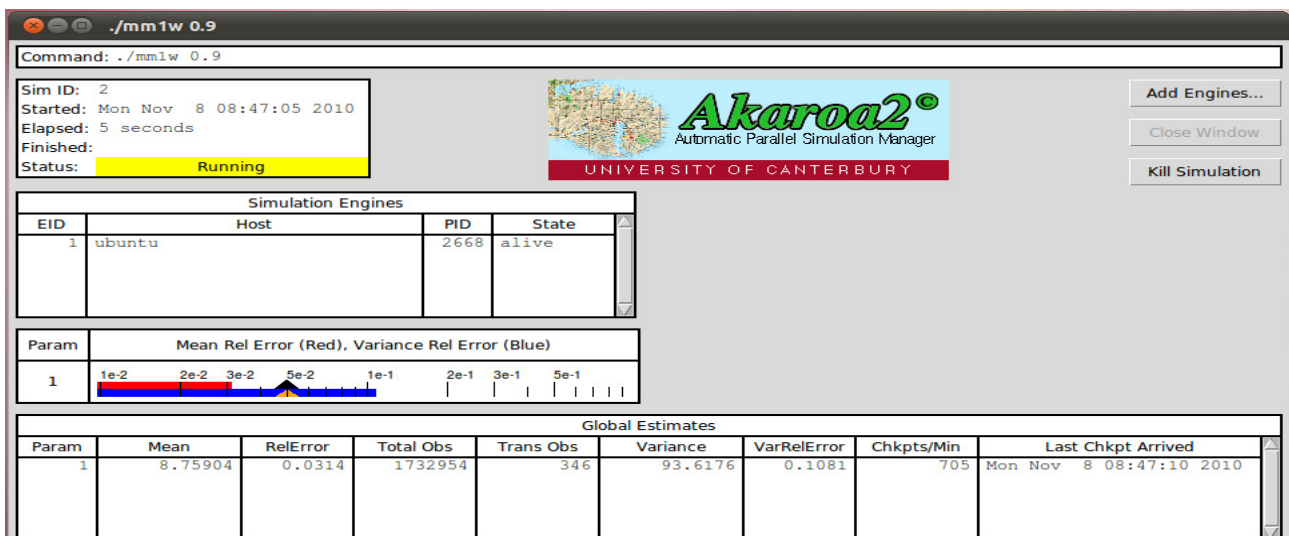


Figure 8: Akaroa2 Prototype Simulation Running Window

6. REFERENCES

- Daley, D. J. (1968) The serial correlation coefficients of waiting times in a stationary single server queue. *J. Aust. Math. Soc.* 8, 683-699.
- Deuermeyer, B.L., R.M Feldman and Y. Yang. (1996) Estimation of process variance by simulation. Unpublished paper, Texas A&M University.
- Gross, D. and C.M. Harris. (1985) *Fundamentals of Queuing Theory*. John Wiley & Sons, 2nd edition.
- Eickhoff, M. (2008) Sequential Analysis of Quantiles and Probability Distributions by Replicated Simulations PhD 02/08. Department of Computer Science and Software Engineering, University of Canterbury, New Zealand.
- Eickhoff, M., K. Pawlikowski and McNickle, D. (2007) "Detecting the Duration of Initial Transient in Steady State Simulation of Arbitrary Performance Measures", in Proc. *ACM ValueTools'07* (Nantes, France, Oct 2007).
- Ewing, G., K. Pawlikowski, D. McNickle. (2010) Akaroa 2.7.9 User's Manual, available at www.akaroa.canterbury.ac.nz
- Ewing, G., D. McNickle and K. Pawlikowski. (1999) "Akaroa2: Exploiting Network Computing by Distributed Stochastic Simulation". Proc. European Simulation Multiconference, ESM'99 (Warsaw, Poland, June 1999). ISCS, 1999, 175-181.
- Ewing, G., D. McNickle and K. Pawlikowski (2002). "Spectral Analysis for Confidence Interval Estimation under Multiple Replications in Parallel", in Proc. *14th European Simulation Symposium* (Dresden, Germany, Oct 2002), 52-61.
- Heidelberger, P., and P.D. Welch. (1981) A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM*, 25:233-245.
- Law, A. M. (1977) Confidence intervals in discrete event simulation: a comparison of replication and batch means. *Naval Res. Logist. Quart.*, 24, 667-678.
- McNickle, D., K. Pawlikowski and G. Ewing. (2010) AKAROA2: A Controller of Discrete-Event Simulation which Exploits the Distributed Computing Resource of Networks, in Proc. 24th European Conf. Modelling and Simulation, ECMS 2010 (Kuala Lumpur, Malaysia, June 2010).
- Pawlikowski, K., D. McNickle and G. Ewing. (1998) Coverage of confidence intervals in sequential steady-state simulation. *Journal of Simulation Practice and Theory*, 6(3):255-267; with "Erratum", *Journal of Simulation Practice and Theory*, 7(1): 1.
- Pawlikowski, K., and D. McNickle. (2001) Speeding Up Stochastic Discrete-Event Simulation. Proc. European Simulation Symposium, ESS 2001 (Marseille, France, Oct. 2001), ISCS Press, 132-8.
- Schmidt, A., K. Pawlikowski and D. McNickle (2008) Sequential Estimation of Variance in Steady-State Simulation TR-COSC 03/08. Department of Computer Science and Software Engineering, University of Canterbury, New Zealand, 2008.
- Schmidt, A., K. Pawlikowski and D. McNickle (2009) Sequential Estimation of the Steady-State Variance in Discrete Event Simulation, in Proc. 23rd European Conference on Modelling and Simulation, ECMS 2009 (Madrid, Spain, June 2009)
- Wilks, S (1962) *Mathematical Statistics*. John Wiley & Sons, 2nd edition

AUTHOR BIOGRAPHIES

DON MCNICKLE is an Associate Professor of Management Science in the Department of Management at the University of Canterbury. His research interests include queueing theory; networks of queues and statistical aspects of stochastic simulation. He is a full member of INFORMS.



KRZYSZTOF PAWLIKOWSKI is a Professor of Computer Science at the University of Canterbury. His research interests include quantitative stochastic simulation; and performance modelling of telecommunication networks. He received a PhD in Computer Engineering from the Technical University of Gdansk, Poland.



NELSON SHAW is a Ph.D. student in Computer Science at the University of Canterbury. His research interests include on-line analysis of output data from quantitative stochastic simulation.

