

REFINING SPECTRAL ANALYSIS FOR CONFIDENCE INTERVAL ESTIMATION IN SEQUENTIAL SIMULATION

Don McNickle
Department of Management
University of Canterbury
Christchurch, New Zealand
Email:Don.McNickle@canterbury.ac.nz

Gregory C. Ewing,
Krzysztof Pawlikowski
Computer Science and Software Engineering
University of Canterbury
Christchurch, New Zealand

KEYWORDS

Sequential simulation, confidence intervals, spectral analysis

ABSTRACT

The method of Spectral Analysis proposed by Heidelberger and Welch (SA/HW) is an effective and efficient way of calculating the error of a sample mean in sequential simulation. A simple modification to the method improves the coverage of the resulting estimators in the case of sequential simulation.

INTRODUCTION

Sequential stochastic discrete-event simulation, i.e. stochastic simulation with on-line analysis of output data, is generally accepted as the most effective way to secure representativeness of samples of observations collected during simulation (Law and Kelton 2000). In this scenario, a simulation experiment is stopped when the statistical error of the estimate(s) reaches a required (low) level.

The method of Spectral Analysis proposed by Heidelberger and Welch (1981), which we abbreviate to SA/HW, has proved to be an effective and efficient way of calculating the statistical error. While more sophisticated spectral methods have been proposed (e.g. Lada, Wilson and Steiger, 2003), SA/HW is the only currently known method of sequential estimation of steady-state mean values in which designers have large freedom for deciding about the granularity of sequential data analysis, since SA/HW can be applied after grouping data in blocks of arbitrary size. This makes it an attractive choice for parallel simulation executed under the Multiple Replications in Parallel (MRIP) scenario (see Ewing, Pawlikowski, and McNickle, 2002.) In this paper we consider a simple modification of the SA/HW algorithm which improves the coverage of the estimators still further.

THE SPECTRAL ANALYSIS METHOD

Simulation output often consists of highly correlated sequences of observations, for example waiting times of successive customers in a queue. Estimating the error in the mean waiting time thus requires techniques that account for this correlation. The best known method is that of Batch Means, where the means of batches of observations chosen large enough to be almost

independent are used to construct a confidence interval. Two problems with Batch Means are: in sequential simulation the granularity imposed by the batch length means that runs may be longer than needed; and finding an easy algorithm to reliably determine the size of the batch length is difficult. Figure 1 illustrates this problem for a particular algorithm. Here Batch Means has been used to produce supposedly 95% confidence intervals (the horizontal line.) However the actual coverage (see Section 2) of the confidence intervals drops off as the traffic intensity (and hence correlation of waiting times) in an M/M/1 queue increases.

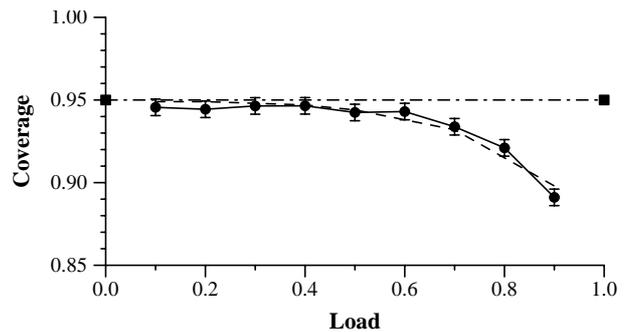


Figure 1: Coverage produced by an automated Batch Means Method. M/M/1 queue

The reduction in coverage turns out to be almost entirely due to the fact that the algorithm for determining batch length has produced batches that are too short. Daley (1968) gives formulas for the serial correlation of M/M/1 waiting times. Law (1977) outlines the steps needed to calculate the serial correlations between the batch means from these correlations. Using this method we can estimate the expected coverage (plotted as a dashed line) from the average batch lengths that the algorithm has produced. Thus almost all of the reduction in coverage appears able to be explained by the fact that the batches are too short. Since reliably estimating small correlations is difficult, the method of Batch Means always carries this risk: that the batches will be too short and hence batch means will remain significantly correlated.

On the other hand the Spectral Analysis method of estimation of the variance of a steady-state mean from a correlated sequence of observations x_0, x_1, \dots explicitly takes account of correlation between the observations. It was originally proposed as a simulation output analysis

method in Heidelberger and Welch, (1981). The variance is obtained as the value of the periodogram $P(f)$ (of the analysed sequence of observations) at frequency $f=0$. Because of high variability of a typical periodogram at low frequencies, in SA/HW its value at $f=0$ is obtained through a regression fit to the logarithm of the averaged periodogram, where fitting is done using a polynomial of degree d (typically $d = 1$ or 2). The fitting is done using K fixed points of the periodogram. As shown in Heidelberger and Welch (1981), if $d=2$, then the confidence interval of the sample mean can be obtained using quantiles of the Student t-distribution with 7 degrees of freedom (if $K=25$). The periodogram can be calculated either over the sequence of individual observations or over the sequence of their batch means. Thus SA/HW can be also applied to sequences of batch means of arbitrary size, instead of individual observations, greatly reducing storage and processing costs. In a subsequent paper, Heidelberger and Welch (1981b) considered a range of alternative values for d , and adaptive smoothing techniques. However they concluded that for both fixed-length and sequential simulation, the modifications offered no substantial improvement over their original recommendation of $d=2$ and $K=25$ points.

COVERAGE ANALYSIS

Coverage analysis is widely used for assessing the quality of different methods used for constructing confidence intervals on the basis of simulation output data. By performing a large number of experiments we estimate the fraction of the generated confidence intervals which actually contain the true value of the parameter. If the method is accurate then when the theoretical confidence level has been set for example to 95% this fraction should also be close to 95%.

We performed sequential analysis of coverage, using the methodology described in Pawlikowski, Ewing and McNickle (1998), to produce coverage of SA/HW estimates with a relative precision of 0.01 at the 95% confidence level. It is worth noting that for each setting of the parameters of the reference models, getting coverage results with the statistical accuracy required meant that up to 14,000 separate experiments were needed.

Experiments were conducted for a number of reference models: M/M/1, M/D/1 and M/H₂/1 and some simple network models. Here we give only the results for the queueing models, with traffic intensities ranging from 0.1 to 0.9. Figure 2 shows the coverage produced by the original SA/HW algorithm in sequential simulation for estimating the mean waiting time in the queue, plotted against the load.

There are two effects that can be noted. The first is that the coverage becomes poorer as the models become more variable. And the second effect is that the coverage reduces slightly, but steadily as the load in each of the queues increases.

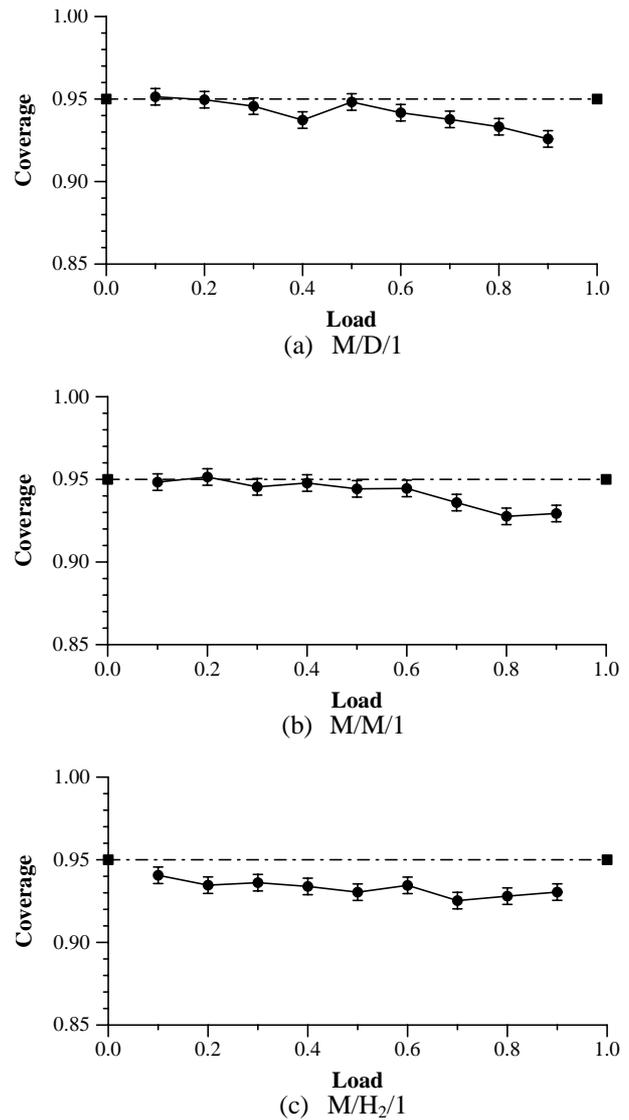


Figure 2: Coverage for SA/HW, M/D/1, M/M/1, M/H₂/1 Queues

We are using the sequential version of SA/HW described in Pawlikowski (1990) in which the observations are grouped into a number of batches, and only the batch means are used as data. Our hypothesis is that that this fall-off in coverage is due to the increased run lengths required for more variable models, or as the traffic intensity increases, which in turn have resulted in batches of larger size. Large batches, we claim, may result in an inappropriate but easily fixed shape of the fitting polynomial.

A MODIFICATION TO SA/HW

As mentioned previously, one attraction of the method is that it can be applied to grouped data, with essentially no change in the algorithm. Grouping the data reduces storage and network costs, so this is an attractive option. As the batch length increases the spectrum becomes flatter, tending towards the constant needed to estimate the variance of the overall mean. Heidelberger and Welch recommend approximating the log of the

periodogram by a low order polynomial, preferably of order $d=2$, in order to estimate the log of the periodogram at zero. For problems where the acceptable relative error is fairly high (e.g. greater than 10%) we have found that this works reasonably well, because the spectrum does have a shape that decays away from zero frequency. However where a very small degree of relative error is required we have found that the simulation can stop too early, with coverage well below the specified level.

The reason for this appears to be that the fitted polynomial is often convex upward when the simulation stops. In fact over the range of queueing models we have observed that about 90% of the simulations using SA/HW with $d=2$ stopped with a convex upward quadratic.

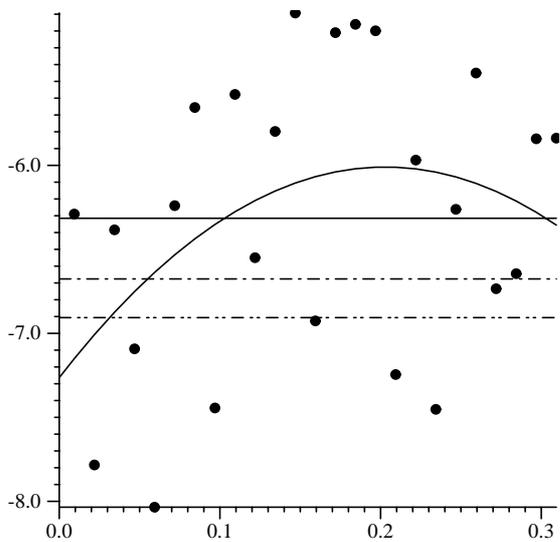


Figure 3: Typical Quadratic and Average Fits to the Log of the Averaged Periodogram at Stopping Time

For example Figure 3 shows the average and quadratic fits to grouped data, with a batch length of 1024, of waiting times for an M/M/1 queue with a traffic intensity of 0.8, at the time when the simulation stopped with an estimated relative error of 0.05, for a 95% confidence interval.

The upper and lower dashed lines show the values that must be reached for the simulation to stop for $d=2$, and $d=0$ respectively. Thus this simulation will stop if a quadratic fit is used, but will not stop if $d=0$.

Since the stopping criterion is satisfied when the y intercept falls below a prespecified level it is clear why this form of fitting polynomial is most likely to occur at the stopping time. However a quadratic ($d=2$) with a positive slope at zero is unrealistic, since the periodogram from simulation output should be a reducing function of frequency, especially after batching. Heidelberger and Welch (1981b) commented on the relative values of $d = 0, 1$ and 2 , and suggested three adaptive methods for picking or altering the degree of the polynomial during the run. They concluded that for both fixed-length and sequential simulation, they offered no substantial improvement

over their original recommendation. However the fraction of sequential simulations which actually stop with a convex upward quadratic suggest a simpler approach which appears to work well.

Since grouping has reduced the periodogram to close to that of an independent process, an obvious modification is to replace the polynomial by simply averaging the values in order to estimate the intercept, in cases where an inappropriate (i.e. increasing at zero) polynomial occurs. This is equivalent to fitting a polynomial of degree zero.

Thus if: $d=2$ and the slope of the quadratic at $f=0$ is positive, we use the average of the periodogram points as the estimate of $P(0)$. The Heidelberger and Welch method requires two constants: $C1(K,d)$ to produce an unbiased estimate of $P(0)$, and $C2(K,d)$ to give the approximate degrees of freedom of the t-distribution. For $d=0$, the values, which were not included in the original paper, are:

Table 1: Constants for the Average Fit

K	D	C1(K,d)	C2(K,d)
25	0	.987	76
50	0	.994	154

Thus if the stopping criterion appears to have been met and the slope at $f=0$ is positive, we use the average value and the parameters in Table 1 to re-estimate the variance. The simulation only stops if this estimate of the error is small enough.

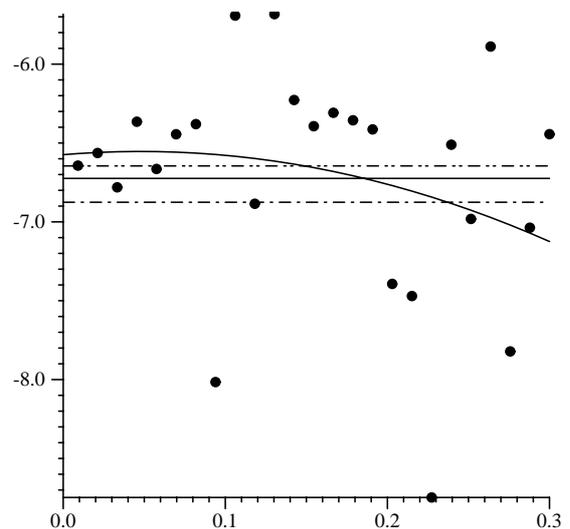


Figure 4: A Case where the Average Fit results in a Reduced Variance Estimate

It should be noted that if the quadratic fit has a positive slope at $f=0$ this does not guarantee that the average will produce a larger variance estimate, as Figure 4 shows. In this example the upper dashed line is the

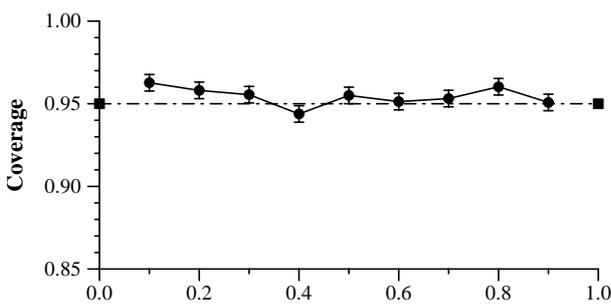
stopping criterion for $d=0$, while the lower line is that for $d=2$. Thus in this case the simulation will stop if an average is used, but will continue if we use the quadratic fit, in spite of the quadratic having a positive slope at $f=0$.

Thus we consider two versions of the modification: if the slope of the quadratic at $f=0$ is positive, we use the average unconditionally to re-calculate the variance, (“Slope Protection”), and using the average only if it provides a larger estimate of variance than the quadratic (“Conditional Slope Protection”).

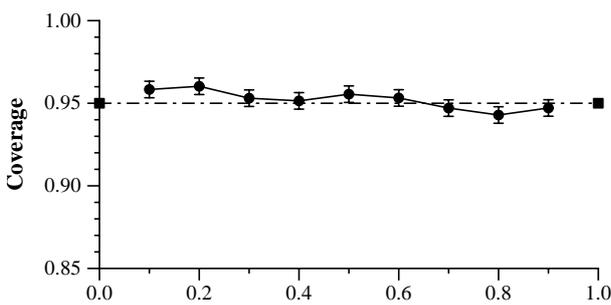
RESULTS

The simulation were carried out using the Akaroa2 Simulation package (Ewing, Pawlikowski, and McNickle, 1999.) The implementation of SA/HW, except for the modification as above, is as described in Pawlikowski, (1990)

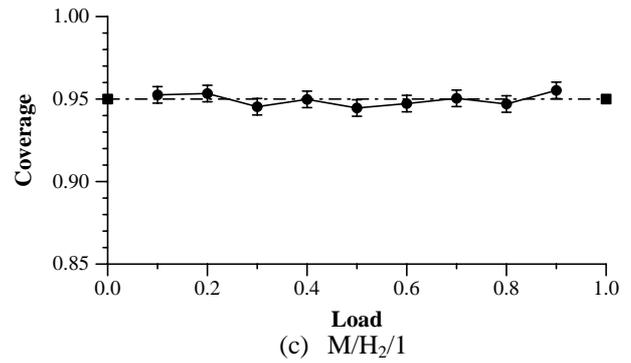
Figures 5 and 6 show the effects of these two schemes on the three queueing models. The coverage is uniformly increased, with the larger increase coming from the conditional scheme. The results for other reference models were consistent with those for simple queueing models presented here.



(a) M/D/1

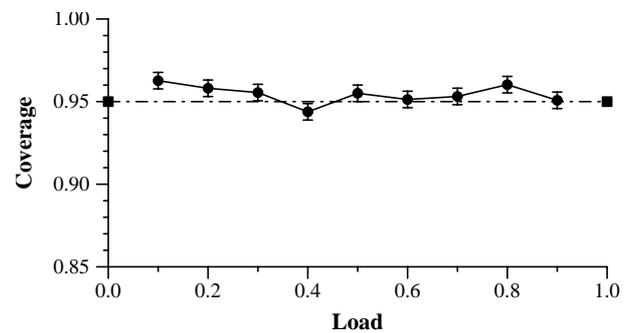


(b) M/M/1

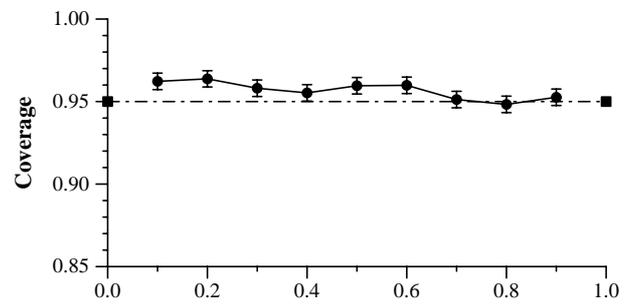


(c) M/H₂/1

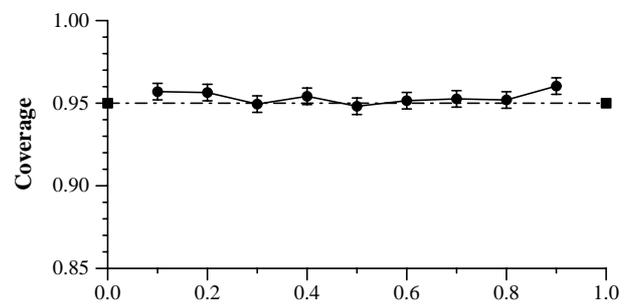
Figure 5: SA/HW with Slope Protection



(a) M/D/1



(b) M/M/1



(c) M/H₂/1

Figure 6: SA/HW with Conditional Slope Protection

CONCLUSIONS

The method of SA/HW has been found experimentally to produce coverage values which agree well with those expected. Further improvements in coverage of SA/HW in sequential simulation can be obtained by adding a simple extra step to the calculation of the stopping criterion to check if the fitted quadratic is increasing at zero. When this happens using the average value of the periodogram instead of a fitted quadratic to estimate the variance of the mean provided coverage levels that were almost exactly those required. The conditional use of the average only if it gave a larger estimate of the error, produced results which were typically slightly above the specified level of coverage and could be considered as providing an additional margin of accuracy.

REFERENCES

- Daley, D. J. 1968 "The serial correlation coefficients of waiting times in a stationary single server queue." *Journal of the Australian Mathematical Society*, vol. 8, 683-699.
- Ewing, G., K. Pawlikowski, and D. McNickle. 1999. "Akaroa2: Exploiting Network Computing by Distributing Stochastic Simulation". *Proceedings of the European Simulation Multiconference ESM'99, Warsaw*. International Society for Computer Simulation. 175-181
- Ewing, G., Pawlikowski, K. and McNickle, D. 2002. "Spectral Analysis for Confidence Interval Estimation under Multiple Replications in Parallel". *Dresden, Germany: Proceedings of the 14th European Simulation Symposium, ESS'2002*. 52-55 & 61. October 2002.
- Heidelberger, P. and P. D. Welch. 1981. "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations." *Communications of the ACM*, vol. 24, no. 4 (April), pp. 233-245
- Heidelberger, P. and P. D. Welch. 1981b. "Adaptive Spectral Methods for Simulation Output Analysis." *IBM Journal of Research and Development*, vol. 25, 860-876
- Lada, E. K., J. R. Wilson, and N.M. Steiger. 2003. "A Wavelet-based Spectral Method for Steady-state Simulation Analysis." *Proceeding of the Winter Simulation Conference*, 422-430.
- Law, A. M. 1977. "Confidence intervals in discrete event simulation: a comparison of replication and batch means." *Naval Research. Logistics Quarterly*, vol. 24, 667-678.
- Law, A. M. 1983. "Statistical Analysis of Simulation Output Data". *Operations Research*, vol. 31, no. 6, 983-1029
- Law, A. M. and Kelton, W. D. 2000. *Simulation Modelling and Analysis*, 3rd Edition. New York: McGraw-Hill.
- Pawlikowski, K. 1990. "Steady State Simulation of Queueing Processes: A Survey of Problems and Solutions". *ACM Computing Surveys*, vol. 22, no. 2, 123-170
- Pawlikowski, K., G. Ewing and D. McNickle. 1998. "Coverage of Confidence Intervals in Sequential Steady-State Simulation". *Journal of Simulation Practise and Theory*, vol. 6, 255-267

BIOGRAPHIES

DONALD 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 member of INFORMS. His email address is <don.mcnickle@canterbury.ac.nz>.

GREG EWING is a research associate in the Department of Computer Science and Software Engineering at Canterbury; where received a Ph.D. His research interests include simulation; distributed systems; programming languages, 3D graphics and graphical user interfaces. He has made contributions to the Python programming language; and has recently been nominated for membership of the Python Software Foundation. His email address is <greg@cosc.canterbury.ac.nz>

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. He is a Senior Member of IEEE and a member of SCS and ACM. His email address is <krys@cosc.canterbury.ac.nz> and his web page is <<http://www.cosc.canterbury.ac.nz/~krys/>>.