1981 Winter Simulation Conference Proceedings
T.I. Oren, C.M. Delfosse, C.M. Shub (Eds.)

Further Investigation into Spectral Analysis for Confidence Intervals in Steady State Simulations

Neil B. Marks
Department of Mathematics
Babson College
Babson Park, Massachusetts 02157

In using spectral analysis for confidence intervals in steady state simulations, the parameter l must be specified. This quantity determines the order of the covariances which are needed in computation of the standard error. This paper investigates the effect of l on the standard error using a small experiment conducted on four queueing systems.

1. INTRODUCTION

The construction of confidence intervals based on simulation output has been a subject of great interest for some time. Confidence intervals computed by ordinary methods often fall far short of desired coverage levels. This problem stems from an underestimated sample variance often. This in turn is the result of the simulation output's tendency to violate assumptions of classical statistics. This is, the data are often not independent observations of a process and are not identically distributed. Hence, under these conditions use of the customary t distribution and the simple sample variance are very difficult to justify.

Fortunately, several methods are available for confronting the variance estimation problem in simulation. The ones mentioned concern simulations of fixed sample sizes, and each will be described briefly below.

In the method of replications a total of n observations is collected (Law, 1980). This sample is obtained by k independent replications of size m ($k \cdot m=n$). The mean of each replication is computed, and from these a grand mean is calculated. Then, the standard error of the grand mean is figured by standard procedure, leading to a confidence interval of the following form:

$$\bar{x} \pm t_{k-1} s.\ e.(\bar{x}) \qquad (1)$$

This method is based on the central limit theorem for justification, but in practice the sample means and variances are still biased, producing unreliable interval estimates (Blomqvist, 1970; Fishman, 1972; Turnquist and Sussman, 1977).

The technique of batch means involves a single simulation run of length n. This sample is divided into k subsamples (batches) of length m ($k \cdot m=n$). Then, construction of the confidence interval proceeds as in replications. Under well prescribed conditions, this method produces unbiased sample means and highly acceptable estimated variances (Fishman, 1978; Fraser, 1957; Law, 1977). In addition, the batch means will be approximately normally distributed. However, for highly autocorrelated data or for m chosen too small, this technique is unacceptable.

The autoregressive procedure is not concerned with producing independent random variables from correlated raw data. Here the data are assumed to be a covariance stationary time series. Developed by Fishman (Fishman, 1971; Fishman, 1973; Fishman, 1978), this technique involves estimation of the variance using covariances of assorted orders. While theoretically elegant, this method fails often because the stationarity assumption is violated, and the use of the t-distribution for the interval itself is sometimes not justified. Spectral analysis builds from this based with modification to provide a more cogent model. It too has limitations but in some situations is superior to the other three techniques mentioned. Since this paper is concerned with one aspect of spectral analysis, the next section will describe the method and the specific problem to be analyzed.

2. SPECTRAL ANALYSIS

As with the autoregressive approach, the data are assumed to comprise a covariance stationary time series. However, some problems encountered in connection with estimation of parameters of the autoregressive model are avoided in spectral analysis.

In this technique a single sample of size n, $x_1, x_2, \ldots, x_n$ is collected. The sample mean, $\bar{x}$, is computed from all n observations. As usual, however, the variance estimation is a complicated matter. One form which has met with popular acceptance is the following:

$$\mathrm{Var}(\bar{x}) = \frac{C_0 + 2\sum_{s=1}^{l-1} w_l(s) C_s}{n - l} \quad (2)$$

where

$$C_s = \frac{\sum_{t=1}^{n-s} (x_t - \bar{x})(x_{t+s} - \bar{x})}{n} \quad (3)$$

C_0 is seen to be the ordinary sample variance, while the C_s are covariances of different orders. The $w_l(s)$ term is known as the spectral window. Though it can assume several functional forms, the Tukey window

$$\tfrac{1}{2}(1 + \cos(\pi s/l)) \quad (4)$$

appears to be superior to the rest. The choice of l is obviously very important, but its discussion will be delayed briefly. Fishman (1969; 1973) shows that the t-distribution is appropriate for this situation, so that the confidence interval is

$$\bar{x} \pm t_{k-1;\alpha/2}(\mathrm{Var}(\bar{x}))^{1/2}$$

The degrees of freedom, k-1, are equal to cn/l, where c=1.33 for the Tukey window.

The choice of l is a wide open matter. If one wishes to establish k-1 in the manner of preparing for batch means, then l is found to be cn/(k-1). Some have suggested that l should be a fraction of sample size, e.g., n/10, n/6, n/4, but no consensus has been reached on the correct proportion. Clark (1978) has developed an heuristic procedure for choosing l, but his method was complicated parameter estimations which make practical implementation difficult.

The purpose of this paper is to examine empirically the choice of l as a function of sample size. The next section describes the methodology for the study, while the following section pertains to the analysis of results.

3. Methodology

Data was obtained for transit times through four simple queueing systems: M/M/1, M/M/2, $M/E_2/1$, and $E_2/M/1$. The traffic intensity in each case was set to .8 to insure congestion and a reasonable amount of autocorrelation in the output. A GPSS program was utilized to generate a 15 transaction warmup run followed by eight batches of size twenty for each system. For each of the eight batches the variance was computed according to equations (2), (3) and (4) above, obtaining results for l=2,3,4, and 5. By doing this l is ranged from n/10 to n/4, the suggested span in the literature.

Thus, for each system 32 variances were calculated, four for each sample. To test for differences attributable to the values of l, the appropriate technique is two-way analysis of variance. Due to the distinct possibility that the assumptions of the F-test would be violated with data on variances, the nonparametric Friedman two-way ANOVA test was employed in analysis. Where significant differences across l values were found, the Wilcoxon matched pairs signed ranks test was used to compare adjacent values. Both tests are nearly as powerful as their parametric analogues, especially when distributional assumptions cannot be proven. The methods will be described in the next section as the results are analyzed.

4. ANALYSIS OF RESULTS

Below in tabular form are the variances computed for each system, sample and l value.

Table 1

Variances

Sample	l= 2	3	4	5
M/M/1				
1	72.9479	80.6970	77.2925	71.7444
2	65.1718	71.4575	70.0003	66.8403
3	65.0607	72.3338	73.5866	72.8412
4	56.7841	60.5267	56.7350	49.7624
5	36.1882	37.4067	32.8029	24.9731
6	36.1364	42.0250	42.6164	40.0403
7	38.2172	43.8075	43.3907	39.0726
8	35.6607	37.9211	35.5355	31.5604
M/M/2				
1	32.4078	35.4548	34.4197	32.1798
2	25.4109	22.6155	17.5507	13.2147
3	31.6356	33.6014	30.6525	26.0418
4	28.2329	27.4240	23.3258	18.1594
5	39.6358	42.2308	29.5964	34.4480
6	47.8247	47.6985	38.4600	26.7411
7	51.7861	49.6159	38.2029	25.9471
8	50.1340	49.3730	42.4463	35.8855
$M/E_2/1$				
1	44.3204	46.2087	43.4277	41.3175
2	40.1241	38.5791	32.4486	27.5701
3	38.1541	36.4604	31.2896	27.7513
4	29.3937	32.1985	33.7829	35.3755

5	32.0582	37.3123	40.5793	43.4978
6	42.9156	50.9449	55.1495	57.1487
7	43.7709	52.3299	58.2252	64.9937
8	44.1215	51.5753	59.0730	67.1271

$E_2/M/1$

1	32.0085	44.4802	50.7552	59.7101
2	31.1118	35.5935	39.2775	42.4101
3	46.9553	53.1208	55.0609	56.1537
4	40.5819	46.5028	48.6149	50.2785
5	38.9881	40.9996	40.1350	39.4595
6	30.2836	32.5563	32.1464	30.4153
7	48.5517	52.8434	49.8013	43.7006
8	75.9960	83.8758	81.9040	77.4233

The Friedman test requires ranking of observations in each row and then summation of ranks in each column as follows:

Table 2

Row Ranks

M/M/1

Sample	1= 2	3	4	5
1	2	4	3	1
2	1	4	3	2
3	1	2	4	3
4	3	4	2	1
5	3	4	2	1
6	1	3	4	2
7	1	4	3	2
8	3	4	2	1
	15	29	23	13

M/M/2

Sample	1= 2	3	4	5
1	2	4	3	1
2	4	3	2	1
3	3	4	2	1
4	4	3	2	1
5	3	4	2	1
6	4	3	2	1
7	4	3	2	1
8	4	3	2	1
	28	27	17	8

$M/E_2/1$

Sample	1= 2	3	4	5
1	3	4	2	1
2	4	3	2	1
3	4	3	2	1
4	1	2	3	4
5	1	2	3	4
6	1	2	3	4
7	1	2	3	4
8	1	2	3	4
	16	20	21	23

$E_2/M/1$

Sample	1= 2	3	4	5
1	1	2	3	4
2	1	2	3	4
3	1	2	3	4
4	1	2	3	4
5	1	4	3	2
6	1	4	3	2
7	2	4	3	1
8	1	4	3	2
	9	24	24	23

The test statistic is computed as follows:

$$X_r^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^{k} (R_j)^2 - 3N(k+1)$$

where N = no. of rows
k = no. of columns
R_j = sum of ranks in j^{th} column

For each case N = 8 and k = 4. The testing results are found in the following table:

Table 3

Friedman ANOVA Results

System	X_r^2	p value
M/M/1	12.3	.008
M/M/2	19.95	<.001
$M/E_2/1$	1.95	.98
$E_2/M/1$	12.15	.008

In three of the four cases, the null hypothesis of no difference between columns (1 values) can be rejected at the 1% level of significance. For these systems obviously judicious care must be taken in the selection of the 1 value. There is no obvious explanation for the unusual behavior of the $M/E_2/1$ system.

The Wilcoxon text will be demonstrated for the middle two columns of the M/M/1 results, and then all significance tests will be summarized in Table 4. For this method pairwise differences are computed in each row. Then, the absolute values of these differences are ranked, but the ranks are given the sign of the difference. The ranks associated with the lesser number of signs are summed, and this number is compared with a table value to establish statistical significance. The critical values for the sum (T) are 4 at the 5% level of significance, 2 at 2%, and 0 at 1%. For the second and third columns of M/M/1 data, we have

1=3	1=4	d	Ranks (signed)
80.6970	77.2925	3.4045	6
71.4575	70.0003	1.4572	4
72.3338	73.5866	-1.2528	-3
60.5267	56.7350	3.7917	7
37.4067	32.8029	4.6038	8
42.0250	42.6164	- .5914	-2
43.8075	43.3907	.4168	1
37.9211	35.5355	2.3856	5

Thus, T = 3 + 2 = 5, which causes acceptance of the null hypothesis of no difference between columns (1 values).

Table 4

Wilcoxon Test for Significant Differences

M/M/1

Column Pair	2-3	3-4	4-5
Test Result	.01	NS	.01

M/M/2

Column Pair	2-3	3-4	4-5
Test Result	NS	.01	.01

E_2/M/1

Column Pair	2-3	3-4	4-5
Test Result	.01	NS	NS

Legend: NS--No significant difference
.01--Significantly different at 1% level

4. Conclusion

The effects of varying the parameter 1 in the spectral analysis of simulation data were examined in this paper. In three of the four queueing systems modelled, highly significant differences in variances attributable to 1 were found. The effects of adjacent values of 1 were less obvious. In nearly half the cases tested, no significant difference was found due to 1, but from system to system the placement of this insignificance varied. Thus, no pattern could be established regarding the sensitivity of this parameter. Perhaps larger sample studies will be able to shed further light on the subject.

5. References

Blomqvist, N. (1970), "On the Transient Behaviour of GI/G/1 Waiting-times," Skand. Aktuar Tidskr., 118-129.

Clark, G.M. (1978), "Confidence Intervals using Sample Autocovariances for Simulation Performance Measures," Department of Industrial Engineering, The Ohio State University.

Fishman, G.S. (1969), Spectral Methods in Econometrics, Harvard University Press, Cambridge, Mass.

Fishman, G.S. (1971), "Estimating Sample Size in Computer Simulation Experiments," Management Science 18, 21-38.

Fishman, G.S. (1972), "Bias Considerations in Simulation Experiments," Operations Research 20, 785-790.

Fishman, G.S. (1973), Concepts and Methods in Discrete Event Digital Simulation, Wiley, New York.

Fishman, G.S. (1978), "Grouping Observations in Digital Simulation," Management Science 24, 510-521.

Fishman, G.S. (1978), Principles of Discrete Event Simulation, Wiley, New York.

Fraser, D.A.S. (1957), Nonparametric Methods in Statistics, Wiley, New York.

Law, A.M. (1977), "Confidence Intervals in Discrete Event Simulation: A Comparison of Replications and Batch Means," Naval Res. Logist. Quarterly 24, 667-678.

Law, A.M., Kelton, W.D. (1978), "Confidence Intervals for Steady-State Simulations, II: A Survey of Sequential Procedures," Technical Report No. 78-6, Department of Industrial Engineering, University of Wisconsin.

Turnquist, M.A., Sussman, J.M. (1977), "Toward Guidelines for Designing Experiments in Queuing Simulation," Simulation 28, 137-144.