Proceedings of the 1991 Winter Simulation Conference
Barry L. Nelson, W. David Kelton, Gordon M. Clark (eds.)

AVERAGE REGRESSION-ADJUSTED CONTROLLED REGENERATIVE ESTIMATES

Peter A. W. Lewis
Richard L. Ressler

Naval Postgraduate School
Monterey, CA 93943

ABSTRACT

One often uses computer simulations of queueing systems to generate estimates of system characteristics along with estimates of their precision. Obtaining precise estimates, especially for high traffic intensities, can require large amounts of computer time. Average regression-adjusted controlled regenerative estimates result from combining the two techniques of controlled regenerative estimates and average regression-adjusted regenerative estimates. Combining these two techniques can create estimates whose estimated mean-square error is much lower than can be obtained through using either technique alone.

1 INTRODUCTION

When simulating queueing systems one would like to use a variance reduction technique to minimize the resources necessary for the simulation. Iglehart and Lewis (1979) developed several internal linear controls for the estimate of the stationary waiting time in a regenerative simulation of the M/M/1 queue. The linear control they identified as the most suitable reduced the estimated variance of the controlled regenerative estimate standard deviation to 54% of that of the uncontrolled estimate. Iglehart and Lewis's results and notation are summarized in Section 2.

An asymptotic formula exists for estimating the variance of a regenerative estimate. Unfortunately, with this asymptotic formula the estimate of the variance of the point estimate is correlated with the original regenerative estimate. An alternative method for estimating the variance of the regenerative estimates is to use multiple replications, or equivalently, section one long replication of N cycles into m smaller "replications" of n cycles each. While sectioning eliminates the correlation between the regenerative estimate and the estimate of its variance, one must determine the sectioning parameters, m and n. Heidelberger and Lewis (1981) developed a method incorporating regression-adjusted and graphical techniques to assist the experimenter in choosing these parameters. Their method produces a reduced-bias regenerative estimate along with a stable estimate of its variance. Section 3.3.1 briefly describes their average regression-adjusted regenerative estimate.

Section 3.3.2 shows how one can use the regression-adjusted technique of Heidelberger and Lewis (1981) with the controlled regenerative estimates of Iglehart and Lewis (1979). This combination of techniques allows one to obtain estimates of the stationary waiting time for the nth customer with much lower estimated mean square error than by using either technique alone. Section 4 provides an example using data from a simulation of an M/M/1 queue. In this example the estimated mean square error for the average regression-adjusted controlled regenerative estimate is just 10% of the mean square error estimate for the straightforward regenerative estimate.

2 THE CONTROL OF REGENERATIVE ESTIMATES FOR VARIANCE REDUCTION

2.1 The Regenerative Estimate of the Stationary Waiting Time in an M/M/1 Queue

This section is condensed from Iglehart and Lewis (1979) to provide the basis for the controlled regenerative estimator. Although the M/M/1 queue will be used as the example, the technique can be applied to more general regenerative simulations.

Define the waiting time of the nth customer in an M/M/1 queue, namely W_n, as the time from the customer's arrival until the commencement of service. One can show that under certain conditions, the waiting time process $\{W_n : n \geq 0\}$ is a regenerative process. When the queue is stable, $W_n \Rightarrow W$ as $n \to \infty$.

Assume the zeroth customer arrives at time $t_0 = 0$, finds the server free, and has a service time of ν_0. The nth customer arrives at time t_n and has a service time of ν_n. Define the interarrival times u_n as $u_n = t_n - t_{n-1}$ for $n \geq 1$. Assume that the ν_n and u_n sequences are independent of each other and that each consists of i.i.d. random variables. Let $\mathrm{E}[\nu_n] = \mu^{-1}$, and let $\mathrm{E}[u_n] = \lambda^{-1}$. Denote the *traffic intensity* by ρ where $\rho = \lambda/\mu$, assuming that λ and μ are both positive and μ is finite. Assume that the traffic intensity ρ is less than one so that the system is stable.

Since the queue is stable, one can show that there exists a sequence of integer-valued random variables $\{T_k : k \geq 0\}$ such that the customers numbered T_k arrive to find the server free and experience no waiting in the queue. These customers start a new cycle or *busy period* for the system. Let $\tau_k = T_k - T_{k-1}$ for $k \geq 1$. Thus τ_k represents the number of customers served in the kth busy period (the length of the cycle). Now define the sequence $\{Y_k : k \geq 1\}$ by

$$Y_k = \sum_{j=T_{k-1}}^{(T_k)-1} W_j, \quad \text{for } k \geq 1.$$

The random variable Y_k is sum of the waiting times in the kth busy period (the area under the function $f(\cdot)$ for the cycle).

Given that the queue is a regenerative process and is stable, one can use results on regenerative estimators from Iglehart and Crane (1975, App. A) to establish that a strongly consistent point estimator for W, based on n cycles, is

$$\widehat{W}(n) = \frac{\overline{Y}(n)}{\bar{\tau}(n)} \tag{1}$$

where $\overline{Y}(n) = n^{-1}\sum_{k=1}^{n} Y_k$ and $\bar{\tau}(n) = n^{-1}\sum_{k=1}^{n} \tau_k$.

In practice, one does not have to estimate W for the M/M/1 queue as when $\rho < 1$, the expected value of W is known i.e.,

$$\mathrm{E}[W] = \frac{\rho^2}{\lambda(1-\rho)}.$$

However, a known value for $\mathrm{E}[W]$ provides a basis for comparing the bias of different estimators via the estimated mean square error.

When estimating the variance of $\widehat{W}(n)$ using sectioned multiple, independent, replications, the point estimate, based on m replications of n busy periods each, would be

$$\overline{\overline{W}}(m,n) = \frac{1}{m}\sum_{j=1}^{m} \widehat{W}_j(n). \tag{2}$$

One would use the variance of the sample mean of the $\widehat{W}_j(n)$ to estimate the variance of $\overline{\overline{W}}(m,n)$, namely $\mathrm{Var}\left[\overline{\overline{W}}(m,n)\right]$ i.e.,

$$s^2_{\widehat{W}(m,n)} = \frac{1}{m(m-1)}\sum_{j=1}^{m}\left(\widehat{W}_j(n) - \overline{\overline{W}}(m,n)\right)^2.$$

The estimate of the standard deviation of $\overline{\overline{W}}(m,n)$ would simply be the square root of $s^2_{\widehat{W}(m,n)}$.

2.2 The Linear Control of Iglehart and Lewis

Now that one has an estimator for the variance of $\widehat{W}(n)$, one would like to reduce the variance of $\widehat{W}(n)$ for a given number of busy periods n. Iglehart and Lewis developed a controlled regenerative estimator by applying a linear control to the Y on top of the ratio in (1). One can write the controlled estimator $\widehat{W}'(n)$, as

$$\widehat{W}'(n) = \frac{\overline{Y'}(n)}{\bar{\tau}(n)} = \frac{(1/n)\sum_{i=1}^{n}\{Y_i - \theta\,(C_i - \mathrm{E}[C])\}}{\bar{\tau}(n)}$$

where C_i represents the value of an i.i.d. random variable that is the control for the ith cycle and θ is a coefficient chosen so as to minimize the variance of $\widehat{W}'(n)$.

One can show that asymptotically

$$\mathrm{Var}\left[\widehat{W}'(n)\right] = \mathrm{Var}\left[\widehat{W}(n)\right]\left(1 - \mathrm{Cor}[C, Y - W\tau]^2\right)$$

where C represents C_k. Thus one would like to choose a control C that is highly correlated with $Z = Y - W\tau$. Without going into all the details, Iglehart and Lewis (1979) chose to use as a control the quantity $C = D - W/\tau$ where D was selected so as to mimic the behavior of Y for the first two customers in each busy period. Using this control scheme, Iglehart and Lewis were able to reduce the estimated variance of the controlled regenerative estimate to 54% of the estimated variance of the crude estimate.

3 THE AVERAGE REGRESSION-ADJUSTED CONTROLLED REGENERATIVE ESTIMATE

3.1 The Average Regression-adjusted Regenerative Estimate

Heidelberger and Lewis (1981) proposed the regression-adjusted technique in order to improve the analyst's ability to reduce the bias of a regenerative estimate (*re*) while assessing the normality/symmetry of the regenerative estimate. Their regression-adjusted technique exploits two aspects of the structure of regenerative simulations.

The first aspect of the structure is the i.i.d. nature of the busy periods. Since the busy periods are i.i.d., one can section a single simulation of $N = m \times n$ busy periods into m i.i.d. simulations of n busy periods each. Thus one can average the m estimates of $\mathrm{E}\left[\widehat{W}(n)\right]$, namely $\widehat{W}_j(n)$ for $j = 1, \ldots, m$, to get the *average regenerative estimate* (the *are*(m_k, n_k) of Heidelberger and Lewis 1981). The average *re* is nothing more than $\overline{\overline{W}}(m,n)$ from (2). Heidelberger and Lewis's (1981) idea was to compute estimates $\overline{\overline{W}}(m,n)$ for different values of n. Let n_k, for $k = 1, \ldots, p$, represent p different values of n. If for a given simulation of N busy periods one estimates $\overline{\overline{W}}(m_k, n_k)$ for each of the p values of n_k, one gets p unbiased but correlated estimates of $\mathrm{E}\left[\widehat{W}(n_k)\right]$.

The second aspect that the regression-adjusted technique exploits is the known bias structure of the regenerative estimate i.e.,

$$\mathrm{E}\left[\widehat{W}(n)\right] = \beta_0 + \beta_1/n + \beta_2/n^2 + \cdots + \beta_d/n^d + \cdots. \quad (3)$$

Estimating the coefficients in (3) to eliminate some of the bias in the regenerative estimate leads one to the regression-adjusted regenerative estimate.

Let $\widehat{W}\mathrm{ra}(N)$ represent the *regression-adjusted regenerative estimate* of the stationary waiting time based on a simulation of N busy periods (the *rare*(N) of Heidelberger and Lewis 1981). The estimate $\widehat{W}\mathrm{ra}(N)$ is defined as the estimate of β_0 in (3). To estimate β_0, the p average regenerative estimates $\overline{\overline{W}}(m_k, n_k)$ are used as dependent variables in an unweighted least-squares linear regression on $\beta_0 + \beta_1/n + \cdots + \beta_d/n^d$. The regression can be to order d=1, 2, or 3 or more. For a given order d, the regression-adjusted regenerative estimate $\widehat{W}\mathrm{ra}(N)$ is unbiased out to terms of order $1/n^d$.

One needs an estimate of the variance of the regression-adjusted estimate though. Given that one can calculate a regression-adjusted regenerative estimate from a simulation of N busy periods, the final step of obtaining a variance estimate requires M independent replications of the regression-adjusted regenerative simulation. Thus in essence, one runs the simulation until a total of $M \times N$ busy periods are completed. Let $\overline{\widehat{W}\mathrm{ra}}(M,N)$ denote the *averaged regression-adjusted regenerative estimate* formed from M replications of N busy periods each (the *arare*(m,n) of Heidelberger and Lewis, 1981). The estimate $\overline{\widehat{W}\mathrm{ra}}(M,N)$ is simply the average of the M independent regression-adjusted estimates. Since $\overline{\widehat{W}\mathrm{ra}}(M,N)$ is a sample mean, one can also estimate the variance of $\overline{\widehat{W}\mathrm{ra}}(M,N)$ as the sample variance of the M regression-adjusted estimates divided by M.

An immediate concern with forming average regression-adjusted regenerative estimates is determining appropriate values for the various parameters such as M, N, p, the n_k and d. Heidelberger and Lewis (1981) describe a graphical protocol which can assist the analyst in selecting some of these values. For the remainder of this chapter, assume that the total number of busy periods in the simulation, namely $M \times N$, has been set at 200,000. The next subsection will discuss the methods for using the regression-adjusted technique with controlled regenerative estimates and the impact of ridge regression in lieu of least-squares regression.

3.2 Using the Regression-adjusted Technique with Controlled Estimates

Average regression-adjusted controlled regenerative estimates result from applying the regression-adjusted technique to controlled *re*'s. The overall procedure is the same as described above in Section 3.3.1. However, instead of using $\widehat{W}(n)$ to calculate the average *re*, one uses $\widehat{W}'(n)$ to calculate the average controlled *re*. The notation for the average regression-adjusted controlled regenerative estimate is simply $\overline{\widehat{W}'\mathrm{ra}}(M,N)$.

A potential difficulty with the regression-adjusted technique is the tendency for the least-squares regression matrix columns, composed of k rows of 1, $1/n$, $1/n^2$, ..., $1/n^d$, to be collinear. The collinearity can increase the variance of the regression-adjusted regenerative estimates. Johnson and Lewis (1989) presented results demonstrating that using ridge regression in lieu of least squares regression can diminish the impact of the collinearity and produce estimates with lower estimated mean square error. Ridge regression developed from the realization that although least-squares estimators are the minimum variance among linear estimators, "they are not in general minimum-mean-square-error estimators in that class." (Kendall and Stuart, 1979, p.92) In the example that follows, average *ridge* regression-adjusted estimates were computed using the ridge regression technique of Dempster, Schatzoff and Wermuth (1972).

4 THE M/M/1 QUEUE WITH TRAFFIC INTENSITY OF .99

A simulation experiment was conducted with the parameters chosen so that the traffic intensity would be .99 while the expected value of W was 10. The simulation was run until 200,000 busy periods were completed. In what follows, the term "best" estimate will refer to the estimate which has the smallest estimated mean square error (MSE). While the data will not be able to establish which particular parameters are optimal, it will establish trends that demonstrate the effectiveness of using the regression-adjusted technique in combination with controlled regenerative estimates.

The first part of the evaluation consisted of using the 200,000 busy periods to estimate $\widehat{\widehat{W}}'(m,n)$, its variance and mean square error, for different n where $m \times n = 200,000$. A sample of the results is in Table 1. Table 1 shows the best section crude estimate was $\widehat{\widehat{W}}(40,5000)$ with an estimated MSE of .220. It also shows that the best sectioned controlled estimate was $\widehat{\widehat{W}}(20,10000)$ with an estimated MSE of .057.

Table 1 demonstrates that one can not rely solely on the reduction on the standard deviation as a measure of effectiveness of a control. The s'/s rows contain the ratio of the estimated standard deviation of the controlled estimate to that of the crude estimate. When evaluating controls for biased estimators, one must consider the effect of the control on the estimated mean square error in addition to its effect on the estimated standard deviation.

Table 1 also shows the importance of selecting the proper number of busy periods n to use to calculate $\widehat{W}'(n)$. Iglehart and Lewis (1979) chose $n = 2000$ for their estimates. They noted that for $\rho = .99$ and $n = 2000$, the baseline linear control estimates $\widehat{W}'(n)_j$ were nonnormal and $\widehat{\widehat{W}}'(m,n)$ had substantial bias. They recommended that n be increased beyond 2000 to alleviate these problems.

The 200,000 busy periods from the same simulation of the M/M/1 queue with ρ=.99 was used to evaluate the performance of the regression-adjusted controlled regenerative estimate, $\overline{\widehat{W}'\mathrm{ra}}(M,N)$, against both the section controlled estimate $\overline{\widehat{W}'}(m,n)$ and the average regression-adjusted crude estimate $\overline{\widehat{W}\mathrm{ra}}(M,N)$. Other factors in the evaluation were the degree, $d = 1$ and $d = 2$, the type of regression, least-squares versus ridge regression, and N, the number of busy periods used for computing each regression-adjusted controlled estimate.

Table 2 and Table 3 contain average regression-adjusted estimates of the stationary waiting time, crude and controlled respectively, along with estimates of their standard deviation (SD) and mean square error (MSE). Both tables indicate that for a fixed number of busy periods equal to $M \times N$, where M is the number of replications of length N, the choice of large M versus large N is important. In both tables, the estimates of the MSE in the row for $M = 8$ are each lower than the estimates in the rows for $M = 5$ and $M = 4$. This indicates that it is more important to have multiple regression-adjusted estimates (large M) than to have many regenerative estimates for forming the average regenerative estimates used in the regression (large N).

A second trend in the two tables is that for both the least squares and the ridge regression estimates, the degree $d = 1$ regressions produce better MSE estimates than the degree $d = 2$ regressions. For example, in the $M = 8$ row in Table 2, the least squares estimate of the MSE goes from .292 to .457 as d goes from 1 to 2 and the ridge regression estimated MSE in Table 2 goes from .265 to .267. These estimates also show that increasing the degree of regression from 1 to 2 caused a much larger increase in the estimated MSE for the least square regression estimate than for the ridge regression estimate.

Finally, in both tables the ridge regression at degree $d = 1$ produced the best average regression-adjusted estimate. For the average regression-adjusted (crude) estimate, the $M = 8$ row in Table 2 had the best estimated MSE of .265. This was larger than the best section crude estimate from Table 1 of .220. However, the best average regression-adjusted linearly controlled estimate, the $M = 8$ row in Table 3, had an estimated MSE of .017. *This estimate is just 8% of the best sectioned crude estimate.* The average (least-squares) regression-adjusted estimate from the same row has an estimated MSE of .02, again less than 10% of the sectioned crude estimate.

In summary, as demonstrated by this simulation of the M/M/1 queue with traffic intensity of .99, combining the regression-adjusted technique with the technique of linearly controlled regenerative estimates can produce dramatic decreases in the estimated mean square error for the estimates of the stationary waiting time.

Table 1: Section estimates based on 200,000 busy periods for the stationary waiting time in an M/M/1 queue with traffic intensity of .99 for different sample sizes n.

The Crude Regenerative Estimate

n	500	1000	2000	4000	5000	7000	8000	10,000
$\widehat{W}(m,n)$	7.54	8.55	9.32	9.77	9.88	9.98	10.0	10.0
S.D.	.248	.336	.405	.485	.455	.526	.527	.550
MSE	6.10	2.50	.621	.287	.220	.277	.279	.304
The Controlled Regenrative Estimate								
$\widehat{W}'(m,n)$	8.09	8.78	9.49	9.81	9.77	9.97	9.99	9.93
S.D.	.189	.232	.254	.308	.223	.268	.282	.228
MSE	3.67	1.54	.323	.133	.102	.072	.079	.057
s'/s	.76	.69	.63	.64	.49	.51	.54	.41

Table 2: Average regression-adjusted crude estimates based on $M \times N = 200,000$ busy periods for the stationary waiting time in an M/M/1 queue with traffic intensity of .99 for n= 500 1000 2000 4000 5000 7000 8000 .

		Least Squares		Ridge	
M, N		$d=1$	$d=2$	$d=1$	$d=2$
8, 25,000	$\widehat{W}\mathrm{ra}(M,N)$	10.2	10.4	10.1	10.2
	S.D.	.501	.509	.500	.490
	MSE	.292	.457	.265	.267
5, 40,000	$\widehat{W}\mathrm{ra}(M,N)$	10.1	10.4	10.1	10.2
	S.D.	.593	.633	.589	.606
	MSE	.369	.532	.358	.415
4, 50,000	$\widehat{W}\mathrm{ra}(M,N)$	10.1	10.4	10.1	10.3
	S.D.	.607	.645	.600	.620
	MSE	.387	.548	.379	.455

Table 3: Average regression-adjusted linearly controlled estimates based on $M \times N = 200,000$ busy periods for the stationary waiting time in an M/M/1 queue with traffic intensity of .99 for n= 500 1000 2000 4000 5000 7000 8000 .

		Least Squares		Ridge	
M, N		$d=1$	$d=2$	$d=1$	$d=2$
8, 25,000	$\widehat{W}'\mathrm{ra}(M,N)$	10.0	10.2	10.0	10.1
	S.D.	.137	.149	.130	.140
	MSE	.020	.061	.017	.024
5, 40,000	$\widehat{W}'\mathrm{ra}(M,N)$	10.03	10.2	10.0	10.1
	S.D.	.180	.209	.176	.201
	MSE	.033	.073	.031	.054
4, 50,000	$\widehat{W}'\mathrm{ra}(M,N)$	10.0	10.2	10.0	10.0
	S.D.	.175	.220	.175	.230
	MSE	.032	.080	.031	.065

ACKNOWLEDGMENTS

The research of P.A.W.Lewis was supported at the Naval Postgraduate School under an Office of Naval Research Grant.

REFERENCES

Crane, M. A. and Iglehart, D. L., "Simulating Stable Stochastic Systems: III. Regenerative Processes and Discrete-Event Simulation", *Operations Research*, v. 23, no. 1, pp. 33–45, January-February 1975.

Dempster, A. P., Schatzoff, M., and Wermuth, N., "A Simulation Study of Alternative to Ordinary Least Squares", *Journal of the American Statistical Association*, v. 72, pp. 77–106, 1972.

Heidelberger, P. and Lewis, P. A. W., "Regression-adjusted Estimates for Regenerative Simulations, with Graphics", *Communications of the ACM*, v. 24, no. 4, pp. 260–273, April 1981.

Iglehart, D. L. and Lewis, P. A. W., "Regenerative Simulation with Internal Controls", *Journal of the Association for Computing Machinery*, v. 26, no. 2, pp. 271–282, April 1979.

Johnson, L. D. and Lewis, P. A., Simulation and the Delta Method: Viable Estimation in Asymptotic Expansions, The 21st Symposium on the Interface, 1989.

Kendall, M. and Stuart, A., *The Advanced Theory of Statistics*, 4th ed., v. 2, Griffin, 1979.

AUTHORS' BIOGRAPHIES

PETER A. W. LEWIS is Distinguished Professor of Statistics and Operations Research in the Department of Operations Research at the Naval Postgraduate School in Monterey, California. He holds an A.B. degree (1954) from Columbia University, B.S. and M.S. degrees (1955 and 1957) from Columbia School of Engineering, and a Ph. D. in Statistics from the University of London in 1964. He is co-author, with D.R. Cox, of the book 'The Statistical Analysis of Series of Events' (1966), editor of the Volume entitled 'Stochastic Point Processes: Statistical Analysis, Theory and Applications' (1972) and co-author, with E.J.Orav, of the book 'Simulation Methodology for Statisticians, Operations Analysts and Engineers' (1989). He has also published a software package 'Enhanced Simulation and Statistics Package' (1989) with E.J.Orav and L.Uribe. He is the author of over eighty refereed papers in technical journals, and is a Fellow of the American Statistical Association and the Institute of Mathematical Statistics.

RICHARD L. RESSLER received a M.S. and a Ph.D. in Operations Research from the Naval Postgraduate School in 1987 and 1991 respectively. He received a B.A. in Biochemistry from the University of Pennsylvania in 1978.