Proceedings of the 1986 Winter Simulation Conference
J. Wilson, J. Henriksen, S. Roberts (eds.)

AUTOMATED ESTIMATION AND VARIANCE REDUCTION FOR STEADY-STATE SIMULATIONS

Rowena Añonuevo
Barry L. Nelson
Department of Industrial and Systems Engineering
The Ohio State University
1971 Neil Avenue
Columbus, Ohio 43210, U.S.A.

ABSTRACT

We present an automated procedure that interfaces with SIMSCRIPT II.5 simulation experiments to derive point and interval estimators for steady-state parameters of stochastic simulations. The procedure combines the nonoverlapping batch means method of output analysis and the control variates variance reduction technique. Batch size and control variates are selected automatically.

1. INTRODUCTION

Variance reduction techniques (VRTs) are used to reduce the population variance of estimators from stochastic simulation experiments; see Nelson and Schmeiser (1986) for a review of well-known VRTs. Most VRTs are designed for finite horizon (sometimes called "transient" or "terminating") simulations for which the experimental design is to sample independent and identically distributed (i.i.d.) realizations of the process. However, for infinite horizon (sometimes called "steady-state") simulations a single long realization may be more practical because of the need to model or delete an initial transient period on each realization.

In simulation output analysis, the initial transient problem has lead to the development of single realization methods for point and interval estimation of the steady-state mean of a process. The primary advantage of single realization methods is that more of the simulation budget can be allocated to generating usable outputs, since only one initial transient period must be deleted. Nonoverlapping batch means (Schmeiser, 1982) is one of these methods, and it is simple enough to be almost entirely automated (see for example Mechanic and McKay (1966), Fishman (1978), Law and Carson (1979), and Schriber and Andrews (1979)). Similarly, it is desirable to apply VRTs in single realization designs to further improve the statistical efficiency of the experiment, and to do it automatically.

In this paper we present an automated procedure called BMCV that combines the batch means method of output analysis and the control variates VRT for simulation experiments written in SIMSCRIPT II.5 (C.A.C.I., Inc.). The procedure's design is based on results in Nelson (1986). Section 2 reviews the batch means and control variates methods. Section 3 outlines procedure BMCV and section 4 gives some preliminary experimental results. We close by discussing possible extensions of BMCV in section 5.

2. REVIEW

This section, which is based on Nelson (1986), reviews the theoretical basis for the batch means method and control variates. Let the output of the simulation experiment be represented by a sequence of identically distributed random (column) vectors $Z'_i = [Y_i, X_{1i}, X_{2i}, \ldots, X_{qi}]$, $i = 1, 2, \ldots, n$, implying that initial transient effects have been removed. Let $E[Z'_i] = [\theta, \mu_1, \mu_2, \ldots, \mu_q]$ and $Cov[Z_i] = \Sigma$ where

$$\Sigma = \begin{bmatrix} \sigma_Y^2 & \sigma'_{YX} \\ \sigma_{YX} & \Sigma_X \end{bmatrix}$$

so that $\sigma_Y{}^2 = Var[Y_i]$, Σ_X is the qxq covariance matrix of $[X_{ji}, X_{mi}]$, $j, m = 1, 2, \ldots, q$, and σ_{YX} is the qx1 vector of $Cov[Y_i, X_{ji}]$, $j = 1, 2, \ldots, q$. Thus, the square of the multiple correlation coefficient of Y_i on $[X_{1i}, X_{2i}, \ldots, X_{qi}]$ is

$$R^2_{YX} = \frac{\sigma'_{YX}\ \Sigma_X^{-1}\sigma_{YX}}{\sigma_Y^2}$$

For our purposes, θ is the unknown parameter of interest and $X_{1i}, X_{2i}, \ldots, X_{qi}$ are the q control variates. To be useful as a control variate, X_{ji} must be correlated with Y_i, and $\mu_j = E[X_{ji}]$, $j = 1, 2, \ldots, q$, must be known. For later convenience, define the column vector $(X_i - \mu)' = [X_{1i}-\mu_1, X_{2i}-\mu_2, \ldots, X_{qi}-\mu_q]$, which has expectation $[0, 0, \ldots, 0]$ and covariance matrix Σ_X. Note that, for random variables, our convention is to use single subscripts to denote column vectors and double subscripts to denote scalar elements, with the exception of Y_i which is a scalar random variable.

The idea behind batch means is to transform the n dependent vectors $Z_1, Z_2, \ldots, Z_n$ into fewer (almost) independent and (almost) q+1 variate normally distributed batch vectors

$$\bar{Z}_j(k) = b^{-1} \sum_{i=(j-1)b+1}^{jb} Z_i$$

for $j = 1, 2, ..., k$; $b = n/k$ is called the batch size, and k the number of batches. Here vector addition is component-by-component addition. We use the convention that any random variable with a bar and argument (k) is a batch mean of $b = n/k$ observations; e.g. $\bar{Y}_j(k)$ is the jth batch mean of the Y_i with batch size $b = n/k$. We assume for now that the total sampling budget n is fixed.

Given k batch means, the control variate estimator of θ is

$$\hat{\theta}(k,q) = \bar{Y} - \hat{\beta}'(k,q)(\bar{X}-\mu) \qquad (1)$$

where

$$\bar{Y} = k^{-1} \sum_{j=1}^{k} \bar{Y}_j(k) = n^{-1} \sum_{i=1}^{n} Y_i$$

$$(\bar{X}-\mu) = k^{-1} \sum_{j=1}^{k} (\bar{X}_j(k)-\mu) = n^{-1} \sum_{i=1}^{n} (X_i-\mu)$$

$$\hat{\beta}(k,q) = \hat{\Sigma}_X^{-1}(k,q)\hat{\sigma}_{YX}(k,q) \qquad (2)$$

The quantities on the right-hand side of (2) are the sample versions of $\Sigma_X(k,q) = \mathrm{Cov}[\bar{X}_j(k)]$ and $\sigma_{YX}(k,q) = \mathrm{Cov}[\bar{Y}_j(k), \bar{X}_j(k)]$. Confidence intervals for θ are given in Nelson (1986).

Lavenberg and Welch (1981) considered the case when $k = n$ ($b = 1$), and the Z_i are i.i.d. q+1 variate normal vectors. They showed that $\mathrm{Var}[\hat{\theta}(n,q)] = (1-R_{YX}^2)(\sigma_Y^2/n)(n-2)/(n-q-2)$. This compares to $\mathrm{Var}[\bar{Y}] = \sigma_Y^2/n$, showing that a variance reduction relative to the sample mean can be achieved if $R_{YX}^2 > q/(n-2)$. Schmeiser (1982) considered the case when $q = 0$ (no control variates) and there exists a number of batches $2 \le k^* \le n$ such that for $k \le k^*$ the dependency and nonnormality of the k batch means $\bar{Y}_j(k)$, $j = 1, 2, ..., k$, is negligible. He showed that there is little additional benefit in terms of point and interval estimator performance from $k > 30$ batches, provided $k^* > 30$.

Nelson (1986) examined the joint effect on variance reduction and confidence interval performance of simultaneously applying control variates and batching, so that the results of Lavenberg and Welch and Schmeiser are special cases. He found that as the number of control variates increases from $q = 1$ to 5, $30 \le k \le 60$ batches assure good point and interval estimator performance, provided $k^* > 60$. Also, at least 30 batches are needed to guard against serious deterioration of estimator performance due to selecting ineffective control variates. This is an important result for BMCV, since BMCV automatically selects control variates using statistical procedures. If Z_i, $i = 1, 2,..., n$ cannot be partitioned into at least 30 acceptable batches, then Nelson recommends increasing n.

Up to this point we have been implicitly assuming that the simulation output process can be represented by Z_i, $i = 1, 2,..., n$, as defined above. In some simulation experiments it may be the case that the output process has a continuous time index. Batching by time, rather than by count, is necessary to obtain an output process of the form considered here. For example, if we have a continuous-time process $Z(t)$, $0 \le t \le \tau$, then

$$\bar{Z}_j(k) = b^{-1} \int_{(j-1)b}^{jb} Z(t)dt$$

where $b = \tau/k$ and τ is fixed, rather than n. Since we have both discrete and continuous-time outputs in general simulation experiments, procedure BMCV batches all output variables by time. While this makes the number of outputs per batch a random variable for discrete-time outputs, the expected number of outputs per batch is the same for all batches.

3. PROCEDURE BMCV

In this section we briefly outline how procedure BMCV works. Complete details are given in Añonuevo (1986). BMCV interfaces with SIMSCRIPT II.5, which is a general purpose programming language containing features that support discrete-event and process-interaction simulation models. See, for instance, Russell (1983). Procedure BMCV is written in SIMSCRIPT II.5.

The following declarations are required by BMCV: the output variable(s) of interest (Y_i) and the potential control variates ($X_{1i}, X_{2i},..., X_{qi}$), the expected values of the potential control variates ($\mu_1, \mu_2,..., \mu_q$), and the maximum number of batches (k_{max}). During program execution BMCV collects outputs and manages batching. When program execution ends, tests of independence (Fishman, 1978) and multivariate normality (extended Shapiro-Wilk test due to Malkovich and Afifi, 1973) are applied. BMCV is designed to use sequences of i.i.d. input variables as control variates, so that only dependence within the Y_i sequence is a concern. It is important to note that, even if passed, these tests do not guarantee i.i.d. multivariate normal batch vectors, nor, if failed, do they prove that i.i.d. multivariate normal batch vectors were not obtained. Thus, BMCV computes point and interval estimates even if the tests fail, but it reports that they failed. A stepwise regression procedure selects what appear to be the most effective control variates from the q potential controls. Point and interval estimates are computed and reported in the results.

An algorithm presentation of BMCV follows. BMCV has a routine for clearing statistics after an initial transient period, which is not shown in the algorithm. A listing of BMCV is in Añonuevo (1986).

Procedure BMCV

0. Declarations: Y_i; $X_{1i},..., X_{qi}$; $\mu_1,..., \mu_q$; k_{max} (default 60). Begin with a batch size of $b \leftarrow 1$ time unit.

1. Collect and batch $\bar{Y}_j(k)$, $\bar{X}_{1j}(k),..., \bar{X}_{qj}(k)$ by time. If the current number of batches is $\geq k_{max}$, then double the batch size ($b \leftarrow 2b$), combine the batches collected so far, and continue.

2. If conditions for ending the simulation are satisfied, then perform "clean-up" procedures to insure that Y and X have common batch size, b_0.

3. Perform the test of independence on $\bar{Y}_j(k)$, $j = 1, 2,..., k$. If the test fails, then reduce the number of batches, k, as follows: $k \leftarrow n/mb_0$, $m = 2, 3,...$, where m-1 is the number of times the test is performed. Recompute the batch means and repeat the test. Report if $10 \leq k \leq 30$ but continue. If $k < 10$ then there are insufficient batches, so stop.

4. Perform the test of multivariate normality on $\bar{Y}_j(k)$, $\bar{X}_{1j}(k),..., \bar{X}_{qj}(k)$, $j = 1, 2,..., k$, where k is initially the number of batches that passed the test of independence. Use the same procedure as step 3 if the test fails.

5. Perform stepwise regression of $\bar{Y}_j(k)$ on $\bar{X}_{1j}(k) - \mu_1,..., \bar{X}_{qj}(k) - \mu_q$ to select $q' \leq 5$ control variates from the q potential control variates.

6. Compute and report the results: point estimates $\hat{\theta}(k,q')$, $\bar{Y}$; 90% and 95% confidence intervals for θ; estimated $Var[\hat{\theta}(k,q')]$, $Var[\bar{Y}]$; estimated percentage variance reduction.

Note that if the number of acceptable batches is less than 10, the current version of BMCV stops. Ideally, it should restart the simulation and increase the total sample size.

4. EXPERIMENTAL RESULTS

To test procedure BMCV, we simulated the closed machine-repair system that Wilson and Pritsker (1984ab) used to test their standardized control variates. This experiment was chosen for two reasons: First, the parameters of interest can be determined analytically, so the performance of the confidence interval procedure can be evaluated. Secondly, Wilson and Pritsker's experimental design used multiple independent realizations, deleting an initial transient period on each one, which is an alternative to the single long realization approach of BMCV.

The machine-repair system operates as follows: There are initially 5 machines in operation and 2 idle spares. The time to failure for an operating machine is exponentially distributed with mean μ_1 = 10.0 time units. When a machine fails it needs a major overhaul with probability .25, in which case it waits in a FCFS queue for a single repairman. The time required to do a major overhaul is exponentially distributed with mean μ_2 = 1.5 time units. Those failed machines not requiring a major overhaul receive minor repair on a FCFS basis from a different repairman whose repair time is exponentially distributed with mean μ_3 = 1.0 time unit. Finally, all repaired machines are inspected by a single inspector. Those machines that pass inspection (probability .9) return to the queue of spares if 5 machines are currently operating, or directly into service if less than 5 are operating. Machines that fail inspection are returned to the minor repair facility. The time required for inspection is exponentially distributed with mean μ_4 = .5 time units.

We are interested in estimating steady-state parameters such as the average number of operating machines, the average utilization of the repairmen and the inspector, the expected waiting time for repair and inspection, and the expected number of idle spares. All of these parameters can be determined analytically; see Wilson and Pritsker (1984b) for details. Here, we report results for estimating θ_1, the average number of operating machines, and θ_2, the average utilization of the repairman at the minor repair facility. These represent the typical and the best results obtained by BMCV, respectively.

Our basic experiment was to simulate the machine-repair system for 7400 time units, deleting outputs from the first 1000 time units. This leaves 6400 time units of usable output. Wilson and Pritsker generated 30 independent realizations of 250 time units each, deleting outputs from the first 50 time units of each realization. Thus, they had a total of 7500 time units, of which 6000 were usable. This illustrates the advantage of a single realization method: out of a smaller total budget (7400 vs. 7500 time units) we were able to allocate more effort to removing the initial transient period (1000 vs. 50 time units per realization) while still obtaining more usable output (6400 vs. 6000 time units). Had we allocated only 50 time units to the initial transient period then the increase in usable output would have been even greater.

As in Wilson and Pritsker (1984b), we replicated the entire experiment 50 times to estimate the average performance of the procedure and the confidence interval coverage probability. The potential control variates were the time to machine failure, X_{1i}, the time to do a major overhaul, X_{2i}, the time to do a minor repair, X_{3i}, and the time to inspect a machine, X_{4i}. The index i represents the i^{th} realization of each random variable, and for fixed j, X_{ji}, i = 1, 2,... is a sequence of i.i.d. random variables. The expected values of these random variables were given above. Note that, unlike Wilson and Pritsker who used all four control variates on all replications, the particular control variates selected by the stepwise procedure in BMCV can and did differ over the 50 replications.

Tables 1 and 2 present a portion of the results for estimating θ_1 and θ_2. "Crude" indicates the usual batch means estimator and "CV" indicates the control variate estimator. Both estimators are based on the same number of batches. Our results compare quite favorably to Wilson and Pritsker (1984b), obtaining almost identical variance and confidence interval (c.i.) half width reductions, and confidence interval coverage probabilities. The variance reductions are still worthwhile after taking the extra computational burden of BMCV into account. For complete results see Añonuevo (1986).

Table 1: Average performance of BMCV over 50 replications of the machine-repair system for estimating θ_1.

	Crude		CV
variance	.000269		.000153
% reduction		43%	
90% c.i. halfwidth	.027329		.020600
% reduction		25%	
probability of coverage	.82		.90

Table 2: Average performance of BMCV over 50 replications of the machine-repair system for estimating θ_2.

	Crude		CV
variance	.000117		.000032
% reduction		73%	
90% c.i. halfwidth	.018121		.009386
% reduction		48%	
probability of coverage	.96		.86

5. DISCUSSION

The experimental results in the previous section, and others in Añonuevo (1986), show that BMCV implements a competitive method for estimating steady-state parameters. However, to be more general BMCV should include a run length control procedure to determine the initial deletion amount and the total run length based on the available budget. In addition, since it is possible that no batch size passes the test of multivariate normality, a procedure to form Jackknife estimators (Bratley, Fox and Schrage, 1983) could be included. Nonnormality causes $\hat{\theta}(k,q)$ to be a biased estimator, and Jackknifing is a procedure to reduce this bias. Finally, results in Nelson (1986) indicate that basing the selection of control variates on the change in the multiple correlation coefficient R_{yx}^2 might be better than stepwise regression.

ACKNOWLEDGMENTS

This research was partially supported by a Seed Grant from the Office of Research and Graduate Studies, The Ohio State University. The authors benefited from discussions with Jane M. Fraser and Charles H. Reilly of The Ohio State University.

REFERENCES

Añonuevo, M.R. (1986). Automated Estimation and Variance Reduction via Control Variates for SIMSCRIPT II.5 Simulations. Unpublished M.S. thesis, Department of Industrial and Systems Engineering, The Ohio State University.

Bratley, P., B.L. Fox and L.E. Schrage (1983). *A Guide to Simulation*. Springer-Verlag, NY.

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

Lavenberg, S.S. and P.D. Welch (1981). A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations. *Management Science* 27, 322-335.

Law, A.M. and J.S. Carson (1979). A Sequential Procedure for Determining the Length of a Steady State Simulation. *Operations Research* 27, 1011-1025.

Malkovich, J.F. and A.A. Afifi (1973). On Tests for Multivariate Normality. *Journal of the American Statistical Association* 68, 176-179.

Mechanic, H. and W. McKay (1966). Confidence Intervals for Averages of Dependent Data in Simulations II. Technical Report ASDD 17-202, IBM Corp., Yorktown Heights, NY.

Nelson, B.L. (1986). Batch Size Effects on the Efficiency of Control Variates in Simulation. Working Paper Series No. 1986-001, Department of Industrial and Systems Engineering, The Ohio State University.

Nelson, B.L. and B.W. Schmeiser (1986). Decomposition of Some Well-Known Variance Reduction Techniques. *Journal of Statistical Computation and Simulation* 23, 183-209.

Russell, E.C. (1983). *Building Simulation Models with SIMSCRIPT II.5*. C.A.C.I., Los Angeles.

Schmeiser, B. (1982). Batch Size Effects in the Analysis of Simulation Output. *Operations Research* 30, 556-568.

Schriber, T.J. and R.W. Andrews (1979). Interactive Analysis of Simulation Output by the Method of Batch Means. *Proceedings of the Winter Simulation Conference*, 513-525.

Wilson, J.R. and A.A.B. Pritsker (1984a). Variance Reduction in Queueing Simulation using Generalized Concomitant Variables. *Journal of Statistical Computation and Simulation* 19, 129-153.

Wilson, J.R. and A.A.B. Pritsker (1984b). Experimental Evaluation of Variance Reduction Techniques for Queueing Simulation using Generalized Concomitant Variables. *Management Science* 30, 1459-1472.

AUTHORS' BIOGRAPHIES

ROWENA AÑONUEVO is a graduate student in the Department of Industrial and Systems Engineering at The Ohio State University. She has a B.S. in mathematics from the College of Arts and Sciences of the University of the Philippines, and was employed by the Operations Research Department of the San Miguel Corporation from 1980 - 1984. Her graduate studies emphasized simulation of stochastic systems, and she will be employed by the Operations Research Group at American Airlines in Arlington, Texas following completion of her Master's degree.

Ms. Añonuevo is a student member of the Operations Research Society of America. She can be contacted by writing to the second author at the address below.

BARRY L. NELSON is an Assistant Professor in the Department of Industrial and Systems Engineering at The Ohio State University. He is also associated with the Ohio State University Statistical Consulting Service. His Ph.D. is from the School of Industrial Engineering at Purdue University, and his research interests center on the design and analysis of computer simulation experiments, particularly methods for statistically efficient simulation. He teaches courses in simulation and stochastic processes.

Continuing memberships include the American Statistical Association, the Institute of Industrial Engineers, the Operations Research Society of America, The Institute of Management Sciences and the Society for Computer Simulation. Dr. Nelson is an active member of the TIMS College on Simulation and Gaming, and is editor of its newsletter.

Barry L. Nelson
Department of Industrial and Systems Engineering
The Ohio State University
1971 Neil Avenue
Columbus, OH 43210, U.S.A.
(614) 292-0610