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

USING CONTROL VARIATES TO ESTIMATE DISTRIBUTION FUNCTIONS

George S. Fishman
Curriculum in Operations Research and Systems Analysis
University of North Carolina
Chapel Hill, North Carolina 27514

EXTENDED ABSTRACT

When estimating a parameter of a problem by the Monte Carlo method, one can usually improve the statistical efficiency of the estimation procedure by using prior information about the problem. Techniques for achieving this improvement are called *variance reduction methods* and they differ considerably in the way they gain their advantages. For example, a user of the *importance sampling technique* draws data from a sampling distribution designed, on the basis of prior information, to reduce the variance of each observation while preserving its mean. The user of the *stratified sampling technique* draws observations from *partitions* of the sample space and then forms a linear combination of the resulting sample means as the estimator. By using prior information to determine the optimal relative number of observations per partition, one achieves a smaller variance for the estimator than the variance that crude Monte Carlo allows. The *antithetic variate technique* derives its benefit by inducing negative correlation between sample outcomes taken in pairs, when it is known *a priori* that certain monotone relationships hold.

By contrast, the *control variate technique* does not gain its advantage by modifying the sampling procedure. Instead, on each trial it collects ancillary sample data on phenomena whose true means are known and then uses a regression method to derive an estimator of the parameter of interest with reduced variance. Since the only additional work in using the technique is the collection of the additional data and, at the end of the sample, to derive the control variate estimator, the incremental cost is usually relatively small.

Originally, Fieller and Hartley (1954) proposed using the control variate technique in a Monte Carlo study designed to estimate the relative frequencies in an unknown population. More recently, Wilson (1984) has summarized the known theoretical results for the method, noting that for the finite sample size case control variate estimators are generally not unbiased and only in the case of normally distributed observations does an exact distribution theory exist for deriving confidence interval for the parameter of interest.

The present paper represents a contribution to the theory of control variates in both the finite sample size and asymptotic cases when estimating a proportion, or more generally a distribution function, and when information on stochastic orderings between the phenomenon of interest and ancillary phenomena with known population parameters is available to the experimenter prior to sampling. In particular, the paper derives an unbiased point estimator (Section 2) and $100(1-\alpha)$ percent confidence interval (Section 3), for the parameter, that hold for every sample size K.

Section 2 also uses this prior information to derive upper bounds on the variance and coefficient of variation of the estimator and a lower bound on the achievable variance reduction. This information is especially valuable before sampling begins. It tells the experimenter what the least possible benefit of the control variate technique is and it enables him to achieve, say, a specified variance or coefficient of variation.

The results for a single parameter extend easily to the multiparameter case. In particular, Section 4 describes how they apply to the estimation of a distribution function (d.f.). A variance reduction is achieved for all estimated ordinates of the d.f., a notable improvement over most earlier applications of Monte Carlo variance reducing methods that focused on estimating a single ordinate of the d.f.

The proposed technique offers yet an additional benefit. For a discrete d.f., Section 5 shows how one can use the variance reduced ordinates of the d.f. to derive an unbiased estimator of the population mean. It also gives the variance of this estimator to order 1/K. To illustrate this technique, Section 6 describes the estimation of the complementary distribution function of maximal flow in a flow network of 10 nodes and 25 arcs where the capacities of the arcs are subject to stochastic variation.

The complete paper is available from the author upon request.

REFERENCES

Fieller, E. C. and Hartley, H. O. (1954). Sampling with control variables. *Biometrika* **41**, 494-501.

Wilson, J. R. (1984). Variance reduction techniques for digital simulation. *American Journal of Mathematical and Management Sciences* **4**, 277-312.

AUTHOR'S BIOGRAPHY

George S. Fishman is professor and chairman of the Curriculum in Operations Research and Systems Analysis at the University of North Carolina at Chapel Hill. His principal interest is the development of statistical methodology applicable to the analysis of output from discrete event digital simulation models. He is the author of *Concepts and Methods in Discrete Event Digital Simulation* published by Wiley in 1973, and of *Principles of Discrete Event Simulation* published by Wiley in 1978. He is a frequent contributor to the operations research and statistical literature on this topic. At present, he is working on variance reducing methods for network reliability estimation and on the influence of concurrent processing on simulation program structure. Professor Fishman is simulation departmental editor for MANAGEMENT SCIENCE and is a member of the Operations Research Society of America, The Institute of Management Sciences and the American Statistical Association.

Professor George S. Fishman
Operations Research and Systems Analysis
209 Smith Building
University of North Carolina
Chapel Hill, North Carolina 27514.
(919) 962-6997