A BRIEF SURVEY OF STOPPING RULES IN MONTE CARLO SIMULATIONS

Michael J. Gilman

Bradford Computer & Systems, Inc.
New York, N. Y.

Summary

A brief survey of the existing methods for determining when to stop sampling in Monte Carlo simulations is presented. The distinction is made between stopping rules for simulations using independent samples and those using correlated samples. Possible avenues for further research are mentioned.

Introduction

The reputation of the Monte Carlo technique as a powerful tool in the solution of many types of scientific and technical problems has become very widespread in the past several years. However, it should be remembered that a Monte Carlo solution is often computationally expensive, and that an analytic solution might be economically more desirable. For this reason, several authors [8, 9, 14, 15] have discussed methods of increasing the efficiency of the Monte Carlo technique by means of such techniques as stratified sampling, importance sampling, control variates, and antithetic variates. Closely related is the problem of determining the number of samples to take in a Monte Carlo simulation in order to achieve the desired degree of accuracy. The set of samples gathered to estimate a parameter is called the sample record, and the rule that determines its length is the stopping rule. In the large majority of current simulations, the required sample record length is guessed at by using some rule such as "Stop sampling when the parameter to be estimated does not change in the second decimal place when 1000 more samples are taken." The analyst must realize that makeshift rules such as this are very dangerous, since he may be dealing with a parameter whose sample values converge to a steady state solution very slowly. Indeed, his estimate may be several hundred percent in error. Therefore it is necessary that adequate stopping rules be used in all simulations.

The approach to the determination of a stopping rule depends on whether or not the individual samples are correlated, and also on the parameter to be estimated. In this paper we will discuss only the techniques that are applicable to estimation of population means when the samples are either independent or correlated.

Methods applicable when the samples are independent are comparatively straightforward, whereas the determination of stopping rules when samples are correlated is more difficult and no single approach has been universally accepted.

Stopping Rules When Samples Are Independent

Let the samples $X_1, X_2, \ldots, X_n$ be independent identically distributed variables with mean μ and variance σ^2, which constitute the sample record of length n. If, as is the case in most simulations, it is desired to estimate the mean of the sequence, we compute the unbiased estimator $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

and

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2 \quad , n \geq 2 \tag{2}$$

which are the sample mean and variance, respectively. The problem is to find a stopping rule so that we estimate μ by a confidence interval of length d, and with coverage probability $\geq \alpha$. That is,

$$\Pr\{\text{confidence interval contains } \mu\} \geq \alpha$$

It can be shown[7] that we satisfy the above condition if we observe the sequence $\{S_n^2\}$, $n = 1, 2, \ldots$ and stop sampling with $n = N$, where N is the record-length, and is equal to the first $n \geq 2$ such that

$$S_n^2 \leq \frac{nd^2}{Z_\alpha} \qquad (3)$$

where Z_α is defined by the relationship:

$Pr\{-Z_\alpha \leq Z \leq Z_\alpha\} = \alpha$, where Z is a standardized normal random variable (i.e., unity mean and zero variance).

The expected sample size is given by

$$E(N) = \frac{Z_\alpha^2 \sigma^2}{d^2} \qquad (4)$$

which shows the importance of choosing as large an interval length d as possible, consistent with accuracy requirements, since the expected sample size is inversely proportional to the square of d.

The above procedure suffers from the fact that it is troublesome and time consuming to compute S_n^2 after each sample. Of course one might, in some cases, choose to compute S_n^2 after each, say, ten intervals. This approach does not insure minimum record length however. A more computationally convenient stopping procedure in the case of independent samples is that of Anscombe[2]. The procedure consists of computing Y_i after each sample, X_i, where

$$Y_k = \sum_{i=1}^{k} X_i = Y_{k-1} + X_k \text{ and k is the} \qquad (5)$$

current length of the sample record.

The rule of Anscombe is to stop sampling when

$$\sum_{i=1}^{n-1} U_i \leq \frac{d^2}{4Z_\alpha^2} n\left(n - 2.676 - \frac{Z_\alpha^2}{2}\right) \qquad (6)$$

where

$$U_i = \frac{1}{i(i+1)} \left(iX_{i+1} - Y_i\right)^2 \qquad (7)$$

The expected sample size, N, using this stopping rule is:

$$E(N) = \frac{4\sigma^2 Z_\alpha^2}{d^2} + \frac{1 + Z_\alpha^2}{2} \qquad (8)$$

Anscombe's method does not require the computation of the sample variance after each step as in the previous method. However, the expected sample size is more than four times as great, and therefore it is not clear that either method is superior.

A word of caution is appropriate here. The above methods are asymptotic and hence true only in the limit as $n \to \infty$. However, for sample sizes of fifty or greater, the above procedures are accurate enough for virtually all applications. The derivation of both results depends on the fact that $\sum_{i=1}^{n} X_i$ has an approximately normal distribution with mean $n\mu$ and standard deviation $\sqrt{n}\sigma$ for large n, if the X_i are identically distributed random variables with mean μ and standard deviation σ.

Correlated Samples

If the samples are correlated, one cannot apply the central limit theorem in the case of the sample mean, and the problem of determining a stopping rule is consequently more difficult than it is with independent samples. Several approaches have been tried, and no single one has been accepted. Many simulation program designers ignore sample correlation entirely and treat the samples as if they were independent. This leads to the danger of underestimating the variance substantially and consequently taking too few samples to achieve the desired statistical accuracy.

A method that does account for serial correlation merely requires that sampling stop when the variance of the sample mean falls below a specified level. When samples $X_1, X_2, \ldots, X_n$, are correlated, the variance of the sample mean is:

$$\sigma_{\bar{X}}^2 = \mathrm{Var}(\bar{X}) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) =$$

$$\frac{\sigma^2}{n} + \frac{2\sigma^2}{n}\sum_{s=1}^{n-1}\left(1 - \frac{s}{n}\right)\rho(s) \qquad (9)$$

where $\rho(s)$ is the serial correlation coefficient for terms of lag s (i.e., s units apart).

Often, covariance of lags greater than some number, say k, are ignored if it can be assumed that correlation decreases as the samples become more widely separated. Then equation (9) becomes

$$\sigma_{\bar{X}}^2 = \mathrm{Var}(\bar{X}) = \left(\sigma^2 + 2\sum_{s=1}^{k}\mathrm{cov}(s)\right)\Big/ n \qquad (10)$$

where cov(s) is the covariance of terms s units apart.

The stopping rule is: stop sampling when $\sigma_{\bar{X}}^2 \leq M$, where M is some constant. The choice of M and k may often be merely guesswork, and this serial correlation method is therefore rather poor unless one has fairly detailed knowledge of the system he is simulating (which, of course, is a rare event).

Since the serial correlation method is not entirely satisfactory, some simulation designers have suggested that sample record $(X_1, X_2, \ldots, X_n)$, be divided somehow into subgroups of uniform length k; $(X_1, X_2, \ldots, X_k)$, $(X_{k+1}, \ldots, X_{2k})$, ..., $(X_{n-k+1}, \ldots, X_n)$, where the means of the subgroups are independent. The variance of the sample mean is equal to the sum of the variances of the subgroup means. As in the previous method, sampling stops when the variance of the sample mean falls below a predetermined constant.

This subgrouping approach is also far from satisfactory. The choice of a subgroup length is usually not easy, since no simple test will tell whether the means are truly independent. Hauser, et al [10], in a process control simulation, have checked for zero correlation between adjacent subgroups. However, this is a necessary but not sufficient condition.

A promising avenue for further research is the method of spectral analysis [5, 6] in the case where methods applicable to independent observations do not apply. Mathematical models of autocorrelated time series appear to be promising tools to aid in the formulation of stopping rules for correlated samples.

References

1. Anscombe, F.J., "Large Sample Theory of Sequential Estimation," Proc. Camb. Phil. Soc. V48 (1952) pp 600-607. Discussion of asymptotic theory of sequential estimation. Anscombe shows that for large sample sizes the fixed sample size formulas are still approximately applicable for the sequential procedure.

2. Anscombe, F.J., "Sequential Estimation," Journal of the Royal Statistical Society, Series B V15 (V1953, pp 1-29. Examination of the problem of finding confidence

intervals for the mean of a normal population with unknown variance.

3. Baker, A.C., "Properties of Some Tests in Sequential Analysis," Biometrika, V37 (1950), pp 334-346. General paper concerned with sequential theory.

4. Conway, R.W., "Some Tactical Problems In Digital Simulation," Management Science, Vol. 10, No. 1 (1953), pp 47-61. Concerned with problems encountered in running simulation programs. Discusses the effect of initial conditions on final results and sample size related to precision of results.

5. Fishman, G.S., Kiviat, P.J., "The Analysis of Simulation Generated Time Series," Management Science, Vol. 13, No. 7 (1967), pp 525-557. Comprehensive study of application of spectral analysis to the study of simulation generated time series. Covariance stationary mathematical models of autocorrelated time series are treated. The abundance of definitions make this paper self contained.

6. Fishman, George, S., "Problems in the Statistical Analysis of Simulation Experiments: The Comparison of Means and the Length of Sample Records," Communications of the ACM, Vol. 10, No. 2 (1967), pp 94-99. Stopping rules are discussed for simulations which compare sample means between two alternatives.

7. Gleser, L.J., Robins, H., and Star, N., "Some Asymptotic Properties of Fixed Width Sequential Confidence Intervals for the Mean of a Normal Population With Unknown Variance," Department of Mathematical Statistics, unpublished Columbia University Report, 24 April 1964. Discussion of confidence intervals when sampling is sequential. Justification of Anscombe's result[2] concerning sequential confidence intervals.

8. Hammersley, J.M., and Handscomb, D.C., Monte Carlo Methods, John Wiley, New York (1964). Mathematical treatment of Monte Carlo methods. Good section on methods of increasing sampling efficiency (variance reducing techniques). Several chapters on applications.

9. Harling, J., "Simulation Techniques in Operations Research - A Review," Operations Research, pp 307-319, May-June 1958. Introductory paper on simulation techniques. Variance reducing techniques and simulation techniques are treated quantitatively.

10. Hauser, N., Barish, N., and Ehrenfeld, S., "Design Problems in a Process Control Simulation," The Journal of Industrial Engineering 17 (1966), pp 79-86. Good applications paper. A case study of a closed loop simulation is presented and various approaches to the selection of a stopping rule are discussed.

11. Kabak, W.I., "Stopping Rules for Queueing Simulations," Operations Research, March-April 1968. Expressions are presented for computing the mean and variance of the probability of being served immediately in the case of queues with exponential interarrival and service times. Samples are independent and sampling stops when the variance falls below a given level.

12. Ray, W.D., "Sequential Confidence Intervals for the Mean of a Normal Population with Unknown Variance," Journal of the Royal Statistical Society, B, V19 (1957), pp 133-143. Small sample approach to the problem of obtaining a confidence interval for the mean of a normal population when the variance is unknown.

13. Sen, Kumar Pranab, "Some Non-Parametric Tests for m-Dependent Time Series," Journal of the American Statistical Association, 60 (1965), pp 134-147. Several non-parametric tests are modified to apply to moving average schemes. The use of the theory of m-Dependent Time Series might someday prove a useful tool in the determination of stopping rules when samples are correlated.

14. Shreider, Y.A., (Editor) The Monte Carlo Method - The Method of Statistical Trails, Pergamon Press, New York, 1966. Explains methods for increasing efficiency of Monte Carlo simulations. Some discussion of sampling error.

15. Tocher, K.D., The Art of Simulation, John Wiley, New York, 1962. Discusses

the theory of sampling as applicable to Monte Carlo simulation.

16. Wetherill, G.B., Sequential Methods in Statistics, John Wiley, New York, 1966. Comprehensive textbook on sequential statistical methods. Includes discussion of sequential estimation, sequential probability ratio test and double sampling.