DOCUMENT RESUME

ED 121 850 TM 005 282

AUTHOR Halperin, Silas
TITLE Design of Monte Carlo Studies.
PUB DATE [Apr 76]
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS *Computer Programs; Hypothesis Testing; Research Methodology; *Simulation; *Statistical Analysis; Tests of Significance
IDENTIFIERS *Monte Carlo Methods

ABSTRACT

There are good reasons for the growing popularity of Monte Carlo procedures; but with increasing use comes increasing misuse. A variety of exact and approximate alternatives should be considered before one chooses to approach a problem with Monte Carlo methods. Once it has been decided that simulation is desirable, consideration should be given to making the study as efficient and general as possible. A simpler model or a canonical form can often make results more general while at the same time reducing the cost of the study. (Author)

ED121850

Design of Monte Carlo Studies

Silas Halperin
Syracuse University

TM005 282

Paper presented at SIG/ES symposium on Monte Carlo Methods.
AERA Annual Meeting, San Francisco, April, 1976

Abstract

There are good reasons for the growing popularity of Monte Carlo procedures; but with increasing use comes increasing misuse. A variety of exact and approximate alternatives should be considered before one chooses to approach a problem with Monte Carlo methods. Once it has been decided that simulation is desirable, consideration should be given to making the study as efficient and general as possible. A simpler model or a canonical form can often make results more general while at the same time reducing the cost of the study.

Design of Monte Carlo Studies

The title is at the same time too broad and too narrow. It is too broad in the sense that Monte Carlo procedures have application far beyond research in statistics; yet the paper will be constrained to consider only statistical applications. It is too narrow in the sense that we will consider not only the design of Monte Carlo studies, but also the thought process which should precede the decision to rely on Monte Carlo methods.

It is clear that no one starts out to design a Monte Carlo study, or at least no one should. Any study is the consequence of a series of questions which can be answered by sufficient effort aimed in the proper direction. In statistics, that effort has traditionally been directed towards finding analytic solutions to the problem. Little needs be said about the advantages of solving a problem analytically; all recognize the exactness, generalizability and elegance of such results. It is also well understood that, in certain applications, mathematical analysis either yields results which are too restrictive to be of real use or is sufficiently complex so as to deter solution. In either of these cases, some substitute for analysis must be found. Monte Carlo procedure should be viewed as but one of the many possible such substitutes.

Once analysis has been eliminated as a means to answer the question of interest, a variety of alternatives should next be considered. One such possibility has great potential but is only rarely found in the literature of educational statistics: enumeration. When only discrete variables are considered, the "brute force" approach of direct enumeration will often suffice as a substitute for analysis. It is not as elegant and will yield generalizability only grudgingly, but it is often capable of giving results which allow us to answer particular questions. Illustrations of enumeration will be given

in later sections.

Enumeration shares one desirable property with analysis: it is exact. If exact solutions are not feasible, some approximation technique must be employed. However, approximation techniques refer to a broad set of procedures of which Monte Carlo procedures represent but one small part. The choice of a particular approximation technique should be made on the basis of its accuracy, efficiency and utility. Illustrations will be given of approximations other than Monte Carlo. As these are discussed, it will become apparent that Monte Carlo techniques have certain advantages which make them attractive as supplements to other approximations. The attractiveness of Monte Carlo procedures, when combined with their inherent ease of application, is often interpreted as license for routine use. It is partly the purpose of this paper to dissuade researchers of this attitude. It is further a purpose to introduce, through illustrations, a variety of ways in which Monte Carlo techniques can be improved to yield more general results with greater efficiency. Before embarking on these endeavors, it is a useful diversion to consider some of the history of Monte Carlo methods.

The Monte Carlo procedure, although based on statistical principles, was not motivated by statistical topics. As observed by Hammersley & Handscomb (1964): "The real use of Monte Carlo methods as a research tool stems from work on the atomic bomb during the second world war." The name was first used by Metropolis & Ulam (1949) and the procedure is usually attributed to Ulam, von Neumann and Fermi. However, informal use of empirical sampling procedures predates these applications by almost a century. During the Civil War, officers found a diversion from more serious matters by tossing needles on a board ruled with parallel straight lines to infer the value of π. In 1901, Lord Kelvin studied Boltzman equations by Monte Carlo methods. In the early part of the

twentieth century, Monte Carlo procedures were used by the British statistical schools for their pedagogical value. Gossett (Student, 1908) was led to the distribution of the correlation coefficient partly through the use of empirical random sampling.

Within the past decade, the application of Monte Carlo procedures to statistical problems has enjoyed a popularity not possible prior to the advent of high speed computers. In a recent survey (Hoaglin and Andrews, 1975) it was found that roughly one of every five papers published in *JASA* and *Biometrika* in 1973 contained results from computer simulation. It is clear that a technique which was once relegated to second class status is now accepted in the most prestigious journals as a legitimate tool of the statistician. Unfortunately, as the number and variety of legitimate applications grows, concomitantly the number of misuses also grows. For example, Games (1971) pointed out that portions of a study by Petrinovitch and Hardyck (1969) could have been approached analytically rather than by Monte Carlo approximations. This illustrates the need for careful thought prior to selecting an approach to a research problem.

Alternatives to Mathematical Analysis

Many statisticians, especially those working with the small sample characteristics of test statistics which have known asymptotic distributions, find enumeration a useful alternative to analysis. For example, Odoroff (1970) studied the properties of a variety of statistics and estimators used to test interaction in RxCx2 contingency tables. He calculated the small sample error rates of these tests by enumerating all possible samples for selected parameter sets and resorted to Monte Carlo approximations only when his sample

size grew beyond the point that enumeration became unwieldly. Although the results of enumeration depend upon the selection of parameters, Odoroff was able to demonstrate a certain degree of invariance to the selection of a particular parameter set. He was thus able to overcome one disadvantage of this non-analytic approach, its lack of generalizability. Margolin and Light (1974) approached the comparison of three statistics to test homogeneity in contingency tables in a similar fashion. They again demonstrate that enumeration can provide solutions to questions concerning tests on categorical data.

When no exact formulation of the problem is possible, it is necessary to resort to approximations. Gosslee and Lucas (1965), studying the properties of tests based on the additive sums of squares methods applied to disproportionate data, found exact distributions available only when the null hypothesis was true; approximations were necessary to study comparative power of the tests. For this reason, they decided to study the approximations under the null hypothesis as well as under alternatives so that relative precision could be determined. Two types of approximations were employed. The first, developed by Box (1954), used a chi-square distribution with degrees of freedom adjusted to be consistent with certain moments of the quadratic form under study. The second was Monte Carlo approximation based on 400 replications. A comparison indicated that the results of the chi-square approximation were closer to the exact results. One might feel this stems from the very small number of replications used (the original experiment was performed by Gosslee in 1956); however DuPuis (1974), in a similar study, found the chi-square approximation to be superior to Monte Carlo results based on 5000 replications. This should not be interpreted as a total repudiation of Monte Carlo procedures. In that same study, DuPuis found a feature of Monte Carlo

procedures which could not be duplicated by the other approximation. It was of interest to study the joint distribution of the various tests as well as their marginal distributions. Monte Carlo procedures were well suited to this purpose and their results were a valuable supplement to those from the chi-square approximation.

The Use of Monte Carlo Techniques

When it becomes clear that Monte Carlo techniques should be used either as a supplement to other approaches or by themselves, the design of the experiment must be carefully planned. Monte Carlo techniques require the total specification of the population and conditions of sampling. The results may depend upon a variety of such exogenous variable and, for that reason, these variables should be varied (or fixed) in such a manner that the results possess some generalizability. Studies relying on Monte Carlo procedures typically specify the levels of the various exogenous variables and consider all combinations of them in a completely crossed design. If adhered to too rigidly, this approach can limit the value of the study.

It is necessary to start with a plan of action and the completely crossed layout is usually a reasonable choice. Once into the study it is often the case that, due to unforseen intermediate results, the plan becomes inappropriate. There are several types of intermediate results which should dictate a change of plans in the middle of a study. We may, for example, have interest in the effects of sample size (N) and number of variables (p) on a certain statistic. If we find that varying N within some selected range has little effect, irrespective of p, we should try values of N other than those specified by our plan. We must have the flexibility to make such

revisions. A change of plans is also dictated by excessive predictability. If, on the basis of intermediate outcomes, the results from successive cells become obvious, those cells should not be filled only for the sake of completing the crossed layout. Monte Carlo procedures tend to be expensive and our resources should be allocated to situations which provide maximum information. In the same vein, the "cheapest" cells (e.g., those with smallest N, smallest p) should be filled first to avoid possible overspending on non-useful cases.

It is not reasonable to attempt to dictate how many replications are needed to make Monte Carlo results sufficiently stable. It is reasonable to suggest that question be given some priority at the planning stage. There will be studies where minimum differences may be of value and the number of replications must be huge. In other situations, a fairly small number of replications (e.g., 400 in the Gosslee (1956) study) may suffice. In any case, we might be well advised to adopt a Tukey philosophy toward Monte Carlo results and get as much detail as possible output. If more replications are found to be needed, the extra output may allow us to do more running and combine runs by hand.

Because of the general availability of random number generators and the relative ease of approaching problems with Monte Carlo techniques, we can readily be lulled into a philosophy of letting the computer do our thinking for us. The careful planning of a Monte Carlo study is important not only to insure its generalizability, but also to limit its expense. At first glance, expense and generalizability appear to be competing goals. That need not be the case however. Certain general ideas exist in the Monte Carlo and statistics literature which allow us to increase the generalizability of results while at the same time limiting the costs. These techniques fall under the general ephithet of variance reduction.

Handscomb (1969) suggests that "... we can regard variance and work as mutually convertible, and so are led to define efficiency with which a simulation estimates a parameter by

$$\text{efficiency} = 1/(\text{variance} \times \text{work}).$$

A technique is called variance-reducing in the Monte Carlo sense if it increases the efficiency; that is if it reduces the variance proportionately more than it increases the work involved." For some reason, simulation studies in areas other than statistics have taken fuller advantage of variance reduction techniques than have statistical studies.

Let us begin an investigation of variance reduction techniques with a simple example. Prowda (1975) applied a nested analysis of variance model to item sampling in test construction. He wished to generate data according to the normal ogive model but recognized the expense of such generation. Substituting the logistic model for the normal ogive model gave comparable results with less computer time and thus qualifies as a variance reduction technique.

An even more attractive variance reduction technique was employed by Marks and Dunn (1974) and Curlette (1975) in studies of classification procedures. As with many multivariate models, classification using discriminant functions is invariant to linear transformations. For this reason, studies such as these can use a canonical form to good advantage. It can be shown that, in a two group case, the vector variables $\underline{x}_1 \sim (\underline{\mu}_1, \Sigma_1)$ and $\underline{x}_2 \sim (\underline{\mu}_2, \Sigma_2)$ can be linearly transformed into the vector variables $\underline{y}_1 \sim (\underline{o}, I)$ and $\underline{y}_2 \sim (\underline{\nu}, \Lambda)$, where Λ is a diagonal matrix. This is accomplished through the use of a theorem by Rao (1973, p.41) which establishes a transformation matrix T such

that $T'\Sigma_1 T = I$ and $T'\Sigma_2 T = \Lambda$. With T defined by

$$T = P_1^{-1} P_2 ,$$

where $\Sigma_1 = P_1' P_1$, $P_1^{-1'} \Sigma_2 P_1^{-1} = P_2 \Lambda P_2'$ and $P_2' P_2 = I$, our canonical vector variables are simply defined as

$$\underline{y} = T' (\underline{x} - \underline{\mu}_1).$$

There are several important reasons for preferring to work with these canonical vector variables. First, they are generated with less effort than the original variables. If we wish to generate samples from multivariate normals, the elements of the vector $\underline{y}$ are independent and can be generated individually. Second, selection of $\underline{\nu}$ and Λ defines an infinite equivalence class of pairs of populations, each pair possessing the identical canonical form. This fact appears to have been first recognized by Curlette (1975) and allows us to generalize the results from one canonical selection to an infinite number of equivalent situations. This provides an excellent illustration of how a variance reduction technique can improve the generalizability of our results while simultaneously reducing the amount of work required.

Another variance reduction technique which has great potential but remains relatively obscure utilizes a procedure developed by Odell and Feiveson (1966). In their paper, they develop a quick procedure for generating sample covariance matrices from a multivariate normal distribution. The advantage of their procedure over that suggested by Kaiser and Dickman (1962) stems from eliminating the need to generate score vectors. This is critical when sample size becomes large.

The direct generation scheme relies on the ability to generate chi-square

variables. Odell and Feiveson suggest an algorithm to generate chi-square variables, but the method is approximate and, according to Zelen and Severo (1964) is best for large degrees of freedom. It might be wise to resort to this approximation when df are large, but use the exponential generation scheme described by Knuth (1969, pp. 113-115) for smaller df.

Direct generation of covariance matrices could be of real utility for a wide class of problems; often it is possible to calculate everything of interest from the covariance matrix without ever requiring the score vectors. If centroids are additionally required, we can easily generate them separately, since means are independent of covariances in samples from normal populations. Use of this scheme could improve the efficiency of studies concerned with multiple regression on random predictors, as well as a wide class of multivariate studies.

The three variance reduction techniques given above are meant to be illustrations and should not be viewed as exhaustive. Given a specific problem, it is often possible to use ideas such as the three presented to improve the study, while simultaneously reducing computer time. However, variance reduction is not the only means to improve Monte Carlo applications. One final example will be given to illustrate how Monte Carlo procedures may be used to supplement an analytic approach.

In an article by Gleason and Halperin (1975), a model for data from a round-robin experiment is developed. Under certain conditions, tests of hypotheses about the parameters could be derived analytically. Under more realistic conditions, however, exact tests which are free of nuisance parameters could not be developed for certain hypotheses. For this reason, a number of approximate tests of the quasi-F variety were suggested. In an attempt to

discover which approximate test was to be preferred, a small Monte Carlo experiment was performed. Since the Monte Carlo procedure required total specification of all parameters, the exact test which depended upon nuisance parameters could be calculated and used as a "benchmark" to compare with the various quasi-F statistics. These comparisons provided an extra criterion for deciding which approximate statistic to recommend. Here, the Monte Carlo procedure allowed approximate tests, which could be calculated on real data, to be compared with an exact test which could not be calculated unless parameters were known. Since parameters are unknown in actual application, the exact test was of no practical value, but possessed significant theoretical value within the confines of the Monte Carlo experiment. It is reasonable to believe that this idea is applicable to a wide variety of other situations where nuisance parameters deter development of exact tests.

Summary

Two major themes ran through this paper:

1. Every attempt should be made to solve our problems analytically. When this attempt clearly is not productive, a range of alternatives should be considered. We should not view Monte Carlo techniques merely as a crutch.
2. When Monte Carlo techniques are clearly indicated, we should make every effort to use such techniques creatively. Our literature abounds with methods which can be used to our advantage and, as statisticians, we should be in the forefront of such application.

The illustrations given in this paper are not original; they all reside in our literature. It is interesting that most techniques meant to improve the

the general art of computer simulation are statistical in nature and yet statisticians seldom avail themselves of these procedures. This may be because our problems are so readily congruent to a brute force Monte Carlo approach that we fail to appreciate refinements which are easily made. It is hoped that this discussion will alert those statisticians who do use Monte Carlo techniques to some new possibilities and that we will all be more sensitive in the future to the possible applications of our own discipline.

References

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification. Annals of Mathematical Statistics, 1954, 25, 290-302.

Curlette, W. L. Some classification techniques appropriate for unequal co-variance matrices: a comparison with existing methods. Unpublished doctoral dissertation, Syracuse University, 1975.

DuPuis, F. A. Investigation of the test of significance of interaction in the method of unweighted means. Unpublished doctoral dissertation, Syracuse University, 1974.

Games, P. A. Inverse relation between the risk of type I and type II errors and suggestions for the unequal n case in multiple comparisons. Psychological Bulletin, 1971, 75, 97-102.

Gleason, J. R. & Halperin, S. A paired compositions model for round-robin experiments. Psychometrika, 1975, 40, 433-454.

Gosslee, D. G. The level of significance and power of unweighted means test. Unpublished doctoral dissertation, North Carolina State College, Raleigh, N.C., 1956.

Gosslee, D. G. & Lucas, H. L. Analysis of variance of disproportionate data when interaction is present. Biometrics, 1965, 21, 115-133.

Hammersly, J. M. & Handscomb, D. C. Monte Carlo methods. London: Methuen, 1964.

Handscomb, D. C. Monte Carlo techniques: Theoretical. In T. H. Naylor (Ed.), The design of computer simulation experiments. Durham: Duke University Press, 1969.

Hoaglin, D. C. & Andrews, D. F. The reporting of computation-based results in statistics. The American Statistician, 1975, 29, 122-126.

Kaiser, H. F. & Dickman, K. Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. Psychometrika, 1962, 27, 179-182.

Kelvin, W. T. Nineteenth century clouds over the dynamic theory of heat and light. Philosophical Magazine, 1901, 2(6), 1-40.

Knuth, D. E. The art of computer programming (Vol. 2/Seminumerical algorithms). Reading, Massachusetts: Addison-Wesley, 1969.

Margolin, B. H. and Light, R. J. An analysis of variance for categorical data, II: small sample comparisons with chi-square and other competitors. Journal of the American Statistical Association, 1974, 69, 755-764.

Marks, S. & Dunn, O.J. Discriminant functions when covariance matrices are unequal. Journal of the American Statistical Association, 1974, 69, 555-559.

Metropolis, N. & Ulan, S. The Monte Carlo method. Journal of the American Statistical Association, 1949, 44, 335-341.

Odell, P.L. & Feiveson, A. H. A numerical procedure to generate a sample covariance matrix. Journal of the American Statisticai Association, 1966, 61, 199-203.

Odoroff, C. L. A comparison of minimum logit chi-square estimation and maximum likelihood estimation in 2x2x2 and 3x2x2 contingency tables: tests for interaction. Journal of the American Statistical Association, 1970, 65, 1617-1631.

Petrinovich, L. & Hardyck, C. D. Error rates for multiple comparison methods: some evidence concerning the frequency of erroneous conclusions. Psychologicai Bulletin, 1969, 71, 43-54.

Prowda, P. P. Estimating individual total-test scores using partially matched sampling. Unpublished doctoral dissertation, Syracuse University, 1975.

Rao, C. R. Linear statistical inference and its applications (2nd ed.) New York: John Wiley & Sons, 1973.

"Student." On the probable error of a correlation coefficient. Biometrika, 1907, 6, 302-310.

Zelen, M. & Severo, N.C. Probability functions. In M. Abramowitz & I. A. Stegun (Eds.), Handbook of Mathematical functions with formulas, graphs, and mathematical tables (National Bureau of Standards Applied Mathematics Series .55). Washington, D. C.: U.S. Government Printing Office, 1964.