

604 958

604958

1

COPY	1	OF	1872
HARD COPY	\$.	1.00	
MICROFICHE	\$.	0.50	

AN INTRODUCTORY NOTE

For the Proceedings of the
Gainesville, Florida, Symposium on
Monte Carlo Methods

A. W. Marshall

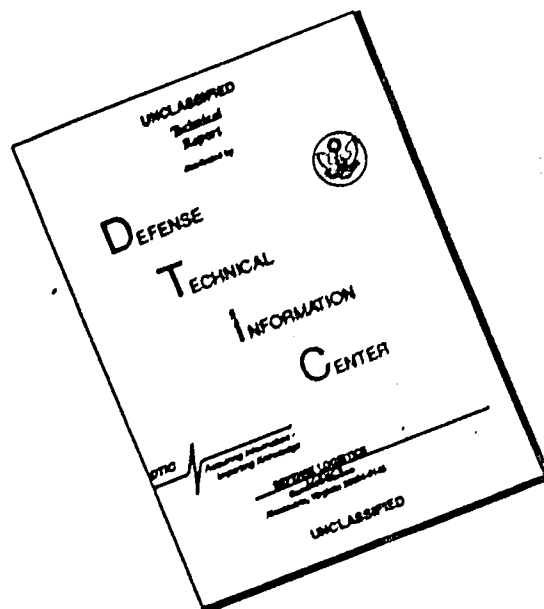
P-756

24 October 1955

Approved for OTS release

DDC
RECEIVED
 AUG 27 1964
DDC-IRA D

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

AN INTRODUCTORY NOTE^{1/}

A. W. Marshall

I. Introduction

This introductory note is intended to fulfill several objectives. First, the symposium itself does not contain a paper summarizing progress and practice in the field of Monte Carlo during the period 1949 (the date of the last Monte Carlo symposium) to the present. This introduction is devoted in part to filling this gap. Second, symposia usually present a somewhat biased view of the state of any field because papers are devoted to new ideas rather than description of standard practice. This would not be an important defect if only experts in the field were the audience but there is considerable evidence of a new, wider interest in Monte Carlo. In the case of the present symposium the bias is somewhat less than is usual. Many of the papers give applications which should by now be standard practice in Monte Carlo computations, but are not yet so, rather than new ideas or advanced techniques. However, there is just enough of the flavor of experts talking to experts to suggest that some guidance for other than expert audiences would be worthwhile. This introductory note is primarily addressed to that growing class of persons interested in Monte Carlo as a tool of numerical analysis rather than as an area of professional work. In other words it is addressed to those who are not now experts in the field but have problems which they believe Monte Carlo techniques may help to solve.

^{1/} The author wishes to acknowledge the large debt he owes to Mr. Herman Kahn for long discussions of material in this note, and to Mr. R. T. Nichols for many helpful suggestions.

Within the compass of so short a note it is not possible to completely review the progress of Monte Carlo or to point up the important aspects of each paper. The first section following upon this one is a commentary on the progress of Monte Carlo from the late forties to the present. The second section following is devoted to some explanatory remarks on the papers collected in this volume.

Before going on, a word about the definition of Monte Carlo, about which there is a good deal of disagreement. The original von Neumann-Ulam concept seems to have been that Monte Carlo specifically designated the use of random sampling procedures for treating deterministic mathematical problems. Some define Monte Carlo to be the use of random sampling to treat problems, whether of a deterministic or probabilistic sort. Others demand that the sampling be sophisticated (involves the use of some variance reducing technique or "swindle") in order to qualify as Monte Carlo; they reserve the names straightforward sampling, experimental sampling, or model sampling for the cases where purely random sampling is used. The writer's sympathies are with the latter view. However the economics of computing is changing so rapidly with the advent of faster and faster machines that sophistication sometimes has a high price. The test of whether it is Monte Carlo rather than model sampling, perhaps, might be whether the sampling design was rationally chosen; the possible application of the most useful variance reducing techniques always being considered in planning the treatment of any problem. As far as the definitional problem is concerned, however, most all discussion is beside the point. Because of its picturesque character and charm the use of the name Monte Carlo has a momentum all its own; the situation now is that in common usage Monte Carlo is synonymous with any use of random sampling in treatment of either deterministic or probabilistic problems.

II. The Progress of Monte Carlo, 1949-1954

The following history and commentary makes no pretense of completeness. Dates are only approximate, as is sufficient for the purpose of this section -- to describe the general progress of Monte Carlo from the late forties and its current state.

A. Early Dominance of the Analogue Idea

If one looks at the Monte Carlo literature of the late forties he finds that it is dominated by the intriguing idea discovered or promulgated by von Neumann and Ulam, that one can use random sampling methods to solve determinate mathematical problems. Others had earlier seen that such things could be done in some cases, for example, the calculation (estimation) of π by random drops of a needle. But they were certainly the first to advocate the idea of systematically inverting the usual situation and treating mathematical problems by finding a probabilistic analogue and then obtaining approximate answers to this analogue by some experimental sampling procedure. Their work and discussion of the basic idea stimulated widespread interest in the use of random sampling procedures in physics and engineering problems. They were also the first to point out the special adaptability of the high speed computing machines, soon to be available, to Monte Carlo type computations.

The statisticians had, of course, been using model sampling methods to investigate some of their problems, for example, the effect of non-normality on statistical test procedures devised for samples from normal (Gaussian) populations, since the early 1900's. Their use of sampling reached a peak in the period 1925-1935 and then died off. However their work was concerned with probabilistic problems so that they were not interested in the sort of thing which might lead to the original von Neumann-Ulam

idea. Moreover, despite their injunctions to all other experimental professions that sampling design was crucial in maximizing the amount of information in a sample for a fixed expenditure of effort, their own work shows no attempt at design at all. Whatever ideas in sample survey design that might have been carried over at this time were not used until the past few years. In any case, the statisticians did not have the analogue idea and this is what got Monte Carlo in its current form started. In addition, they did not initiate the use of variance reducing techniques in the area of artificial sampling or model sampling. Their work in this area represents a separate stream that only today is merging with the much richer and larger one generated by the von Neumann-Ulam Monte Carlo ideas. It is always hard to decide questions of priority and direction of influence, but in the case of the current development of Monte Carlo the situation is relatively clear.

In spite of the fact that the von Neumann-Ulam ideas have in a way been very fruitful, it is possible to get an exaggerated notion of the practical importance of the analogue idea from what has just been written above and especially from the literature of the late forties. For some problems, for example, particle diffusion problems, the application of this idea meant a retranslation of problems back to the probabilistic basis they had in physical theory in the first place. The success of the idea was no accident in these cases. Few, if any, basically deterministic problems have really been treated by Monte Carlo methods. Actually the dominance of the literature by the analogue idea is probably due more to its novelty and its great potentialities than to its historical achievements, and to the professional interest of most of the people contributing papers -- theoretical numerical analysis rather than applied problems of immediate practical importance. The use of Monte Carlo methods of computation have been influenced by

the analogue idea mainly in an indirect way. The most important practical applications thus far have had a probabilistic basis; the influence of the original Monte Carlo idea has been to suggest treating them directly as probabilistic problems rather than attempting a difficult, if not impossible, analytical solution. The translation and later retranslation of problems from probabilistic terms to non-probabilistic mathematical problems and back again has been by-passed.

B. Early Applications of Monte Carlo

Because the most important applications have been on basically probabilistic problems, there is a special interest in mentioning the important problems where the pure analogue idea has already found some application. Around 1947-49 Fermi, with Metropolis and Ulam, used it for obtaining estimates of the eigen values associated with the Schrodinger equation. A little later Kacs also used Monte Carlo methods when working on the same problem. Work by Metropolis and Ulam is continuing at Los Alamos on the application of Monte Carlo methods to the Schrodinger equation. However, thus far no published results show any marked improvements over results attainable by classical techniques.

About 1950 matrix inversion and solution of partial differential equations by Monte Carlo methods were developed. Here again, although these are areas in which the analogue idea would find use, no results in cases with important applications have been published.

The first practical problems treated by the Monte Carlo method were connected with the design of atomic weapons at Los Alamos during the war. In these problems nature was directly modeled in its probabilistic aspects and many problems in particle diffusion were solved. Since these problems were first posed in terms of solving complicated mathematical equations to obtain expected

values the use of nature's model of the same diffusion process as the analogue problem might be considered as an application of the analogue idea. It is the author's understanding that Ulam does consider it so. However, once one remembers that the answers to the problem at hand are the expected values of the stochastic process of central interest in the problem, the use of that process in obtaining estimates of the answer does not seem to involve anything sufficiently removed from the problem, as nature poses it, to be called an analogue. No matter what it is called, however, given that the abstract formulation of problems often obscures their true basis the use of nature's model was a great discovery and very successful. Later the physicists extended the use of these methods of calculation to other areas, such as particle diffusion in shields, meson cascade problems, etc.

At an early date people working in the general area of operations analysis also took up the use of the Monte Carlo method in their problems. This use is now widely developed and will be described in more detail below.

C. Initial Development of Sampling Technique

Given that one is to do problems by random sampling methods, attention naturally turns to three topics: (1) choosing or modeling the probability process to be sampled (in some cases this means choice of the analogue; in others a choice between alternative probability models of the same process), (2) deciding how to generate random variables from given probability distributions in some efficient way, and (3) variance reducing techniques, i.e., ways of increasing the efficiency of the estimates obtained from the sampling process. On the first topic little will be said. With regard to the second topic von Neumann was very ingenious from the start but general techniques had not been worked out in the early years. Indeed the literature contains some incorrect statements of the sort -- it is impossible

to do such and such -- things which actually can be easily done. Work is continuing in this area and good methods for many distributions are available. Of course with the increased speed of the computing machines the problem has shifted a little; the emphasis is less on efficiency and more on just having some way of generating the necessary random variables.

In the case of the third topic, variance reducing techniques, von Neumann and Ulam invented the Russian Roulette and splitting ideas about 1945 (see Kahn's paper) but there was no systematic attempt to investigate this problem until T. E. Harris and Kahn began their work in 1948. Since then a great deal of progress has been made but the results are not well known, nor widely used. One hopes that the publication of this volume will do something toward correcting this deficiency in the current practice of Monte Carlo.

D. The Current State of Monte Carlo

Since the late 40's the main trends have been:

1. A relative decline of interest among people purely concerned with technique and especially in the interest in the exploration of the pure analogue idea. The development of some first rate applications of the idea is probably required to stimulate further work in this now more or less dormant area.
2. A very great absolute and relative increase in applied use of Monte Carlo, mostly, as has been remarked, to problems with a probabilistic basis. Overt interest in Monte Carlo probably passed through a minimum in 1951-52; the novelty had worn off in the area of theoretical interest and applied use was not as widespread as it now is.
3. The statisticians, who all along have been by professional training eminently qualified to make important contributions, have now become interested in the field, both for use on their own problems and

as a separate professional area of sampling technique. Harris and a few others were interested in and working in the field almost from the very beginning, but it is only lately that articles on Monte Carlo have begun to appear in statistics journals.¹

¹ The development of Monte Carlo independently of the statisticians and particularly their poor showing in applying sampling techniques to their own problems may be surprising to some; especially to those who think Student invented Monte Carlo. Some writers in the field of Monte Carlo have been excessively polite and/or generous in indicating that many or all of the variance reducing ideas come from survey sampling theory. This is not true since most of the important ideas exploit the extra degrees of freedom available in the design of sampling procedures that result from the fact that one has complete control of the generation and selection of the random variables he uses. What is surprising is that the statisticians did not take the advice they gave to others on the use of efficient designs.

As to the invention of Monte Carlo by Student, it is a strange case. Student did something slightly different, but much better than most Monte Carlo calculations. In the case of his discovery of the t distribution he had derived the correct distribution analytically, but he had made some jumps in the logic. Being unsure, but not very unsure, of the derived distribution he tested it upon some samples he had drawn for exploratory purposes, computing t, and testing the sample distribution against the theoretical using a χ^2 test. He states that he first tried to get the distribution by sampling, but how he used the sampling results to suggest the right answer is not indicated. In his paper the sample results are used only as a check.

In the case of the distribution of the correlation coefficient Student knew that for $n=2$ the distribution was degenerate and, in addition, he had some samples for $\rho = 0$ of size 4 and 8 and for $\rho = .66$ of size 4, 8, and 30. From the same distributions and moments for the case $\rho = 0$ he conjectured that a Pearson distribution of the form $f(x,n) = k \left[1 - \frac{x^2}{\beta(n)} \right]^{\alpha(n)}$ would approximate the distribution. Fitting α and β for $n=4, 8$ he rounded and guessed that $\alpha = \frac{n-4}{2}$ and $\beta = 1$, checking that for $n=2$ the distribution fitted exactly. Thus in the cases of both t and r, Student made magnificent guesses, partly suggested by prior samples, and supported by checking against actual samples and analytic results for large sample approximations of the moments of the distributions, etc. All of this may or may not be Monte Carlo, it is different from most applications, and in any case an isolated instance of first rate use of sampling for statistical purposes.

To the extent the later sampling experiments of the statisticians were multi-purposed this would, of course, inhibit the application of variance reducing techniques, since this requires one to select what it is that is of most interest and to concentrate upon the reduction of the variance of that particular characteristic, often at the expense of increasing the variance of the others. Interest in generating data giving unbiased information about many aspects of a problem also makes it difficult to apply variance reducing techniques to many operations research problems as they tend to be stated. But in investigating such things as the effect of non-normality or comparisons of the power functions of tests powerful techniques have long been available.

A few new areas for the application of Monte Carlo have been opened up in the past few years. The most important of these is statistical mechanics. Here Metropolis, Rosenbluth and Teller pointed out that one can exploit sampling processes possessing the right kind of ergodic properties. These sampling processes allow one to reproduce (converge to) the correct equilibrium distribution of various quantities such as the total energy of a cubic lattice system. At least two groups are now working to exploit these ideas.

There has also been an increased use of Monte Carlo methods in operations analysis problems especially as these have become more comprehensive and richer in detail and, therefore, beyond the scope of analytic and classical numerical methods. In this case Monte Carlo is used almost by default, since the only feasibility limit to Monte Carlo is one of energy or computing power. As has been mentioned, the advent of more and faster computing machines has had its impact on Monte Carlo just as it has on other numerical analysis techniques. This technological growth has led to a tendency to do larger and larger problems, which if they have a probabilistic basis has tended to make Monte Carlo a preferred method of analysis.

The increase in the speed of the machines has also tended to make variance reducing techniques relatively less interesting, but has by no means eliminated their usefulness. The effect of increased computing speed in the newer machines is to make the cost of designing and coding a problem increase relative to the cost of machine running time. The use of variance reducing techniques shortens running time but at the expense of (1) increasing the time spent in designing the computations so as to adapt the classical techniques to the particular problem or in the invention of a new, more suitable technique, and (2) complicating the coding because of the more elaborate bookkeeping and calculations these techniques usually require. On the whole, however, if

there is one thing that would generally increase the usefulness of Monte Carlo it is the discovery of new variance reducing techniques, or the application of known variance reducing techniques as a matter of course to the ordinary run of problems. Not only should these techniques be used whenever it is economical to do so but, in addition, since the variance reducing techniques are not yet well known there should be a bias toward using them even when they are not economical for the problem at hand in order to learn about them for use in later problems. The use of new techniques in marginal cases is almost always justified as a method of building intellectual capital. In the long run one would suppose that real thought on the design of Monte Carlo problems will be confined to problems of a basically new type whenever they first appear; standard variance reducing techniques will be available, and used, for other problems on the basis that sub-routines for computing common functions now are.

III. Comments on the Papers in the Volume

In this section comments upon the papers contained in this volume will be given. The aim of the comments is mainly to assist those not now expert in the field of Monte Carlo to get a little more than they might otherwise out of reading this collection of papers.

The papers are grouped for discussion into three groups: (1) those dealing with the generation of random numbers and more general random variables; (2) theoretical papers; and (3) applied papers. Not all papers will be commented upon, in particular papers included only as abstracts will not be commented upon.

A. The Generation of Random Numbers and More General Random Variables

One of the first problems faced by anyone doing Monte Carlo calculations is the generation of the sequences of random variables required

by the problem. This generally breaks down, if the problem is to be run on a high speed computing machine, into two problems: (1) how to generate sequences of random numbers uniformly distributed on the interval (0, 1); and (2) how to transform the random numbers into random variables having specified probability distributions. Previously prepared tables of random numbers and special random variables, e.g., random normal deviates, are also useful, especially if they are on punch cards or tapes suitable for feeding into computing machines. However, first rate methods are now available for the generation of random variables in the very large numbers needed for the types of problems run on high speed computing machines; tables are of real help only with hand computations (or computations with traditional IBM type equipment).

The papers of Butler, Lytle, Metropolis, and Taussky-Todd deal with various aspects of the problem of supplying random variables for use in Monte Carlo computations. The Taussky-Todd paper is an especially good summary of the currently available and tested methods of generating sequences of pseudo-random numbers in high speed computing machines. The bogus character of the randomness of the generated sequences does not seem to affect the Monte Carlo calculations, just as it is not detected by the various statistical tests to which these sequences have been subjected, and through which they have passed with flying colors. The congruential methods ($X_{n+1} = K X_n \pmod{M}$) of generating pseudo-random numbers are so good for practical computational purposes that there no longer exists any problem of having an adequate supply of random numbers. Thus the discovery of newer methods, such as the use of the Fibonacci sequence, is no longer practically important, although it may be of great theoretical interest.

The Metropolis paper contains a very interesting section on the mid-square method of generating random numbers. This was the first random

number generating method adapted for use on high speed machines. It is not, however, as good a method of generating random numbers as the newer congruential methods: the two methods take about the same time to produce a random number (cost the same), but for the newer machines sequences of period on the order of 10^{12} are possible with the congruent methods no matter what starting value is used. On the other hand, Metropolis' best starting value gives a sequence of length (or period?) on the order of 10^6 and for most other starting values the sequences are probably much shorter, on the order of 10^4 . It is a substantial practical advantage of the congruent method that one does not have to pick a special initial value in order to get long sequences of usable numbers. Nonetheless the study of the characteristics of sequences of random numbers generated by the mid-square method is of classical interest.

The Lytle paper provides an example of the second problem mentioned above -- the transformation of random numbers into random variables with the appropriate probability distributions. The Butler paper deals more generally with this problem and describes in some detail a few of the more interesting methods of achieving the required transformations. This is an area in which ingenuity is often required for success. It is only gradually that there has developed a practical capacity to generate random variables from a wide class of probability distributions. The choice of any particular scheme of course depends upon the economics of the computation situation. For example, at RAND we currently generate random variables having the exponential distribution by taking the logarithm of random numbers, i.e., by the direct method. Excellent approximations of $\ln X$ are available for use on modern high speed machines that make this scheme competitive with, and probably more efficient than, rejection techniques.

B. Theoretical Papers on Monte Carlo

The papers concerned mainly with theoretical aspects of Monte Carlo methods are those of Albert, Curtiss, Kahn, Marshall (with Walsh's comment), Motzkin, Trotter and Tukey, and Walsh. No comment will be made about the Motzkin or Walsh papers. The remaining papers, with the exception of Curtiss' paper, deal with the central problem of variance reduction.

Albert's paper is concerned with the application of Monte Carlo methods to certain integral equation problems. Several alternative methods of producing estimates are suggested and discussed from the point of view of the relative size of their variances. The extent to which these particular estimates have been used is not mentioned in the paper.

Kahn's paper reviews the techniques of variance reduction that have thus far proved to be most useful in Monte Carlo computations. Many of these techniques were first developed by Kahn, or at least perfected and used in important applied Monte Carlo problems for the first time by him.

My own paper investigates a more or less mechanical, two-stage sampling method for choosing good importance sampling schemes. The technique discussed is fairly straightforward but must be considered as wholly untried. Some technical problems arise in the paper because the expected value of the variance of the estimate derived from the second stage in the sampling scheme is infinite, at least in a wide class of cases. Depending on how large the first sample is, however, the variance of the second stage estimate is smaller than that of the first stage with high probability. Walsh makes the very reasonable comment that in this case one might work with other measures of variation than the variance in order to avoid analytic problems. This is not the only way out, but is a good one if the other measures prove more tractable analytically and in applied numerical work.

The Trotter and Tukey paper, with its companion applied paper mentioned in the next section, is probably the most exciting paper in the volume. First, the authors introduce a new variance reducing technique (conditional Monte Carlo) and, second, in the applied paper they show that it works very well. It is heartening to see mathematical statisticians really becoming interested in Monte Carlo techniques and using them on their own problems. This paper deserves close study (and needs it) to get the conditional Monte Carlo trick straight. It seems to be more complicated than, for example, the methods discussed by Kahn. The paper also contains, besides the technical gem already mentioned, many wise words on the good practice of Monte Carlo which reflect an attitude that should be more widespread.

Curtiss' paper is primarily an examination of the relative accuracy of two classical, non-stochastic methods and the Monte Carlo method of computing one component of the solution of a linear equation system. The paper gives a first rate and fair comparison of the methods. One ends with the feeling that unless something can be done to increase the accuracy of the Monte Carlo method in this area that it will remain merely an interesting novelty, especially when as Curtiss suggests it is so seldom the case in actual applied problems that only one component of the solution is required.

C. Applied Monte Carlo Papers

The applied papers are those of Arnold, Bucher, Trotter and Tukey; Beach and Theus; Berger; and Dismuke. The paper by Arnold, et al, is one of the first papers to be published in which substantial and impressive reduction in variance is achieved. Other problems have been done as well, and indeed factors of 10^6 in variance reduction have been achieved in some cases, but unfortunately none of these results have been published in a form that is both convincing and instructive.

The paper of Beach and Theus also helps to fill the same gap in the literature. It, along with the Berger paper, is the best thing yet published in the field of the application of Monte Carlo to nuclear shielding problems. This is an area where some of the best Monte Carlo has been done. Kahn's Nucleonics paper describes the variance reducing techniques but numerical examples of their use were not presented. In some sense however the work behind the paper of Beach and Theus is better than the paper itself. It is not as educational as it could be. For example, much of the success in choosing good variance reducing techniques in applied Monte Carlo calculations has been due to a physical understanding of the problem or intuition as to the processes in the problem which largely determine the answer. Thus in the present paper it would be instructive to hear more about (1) the rationale behind the choice of the various sampling schemes that were tried, (2) why those that worked well did so, (3) why some combination of energy and scattering biasing and the exponential transformation would probably be required for the treatment of problems with gamma rays of lower energy, say 2 Mev, etc.

Berger uses somewhat different variance reducing techniques than those of Beach and Theus, i.e., correlated sampling and more sophisticated methods of extracting information from the random walks. These techniques are best classified under the heading of the method of expected values discussed in Kahn's paper. Unfortunately while the paper illustrates the use of the method of expected values it does not show any real advantage over straightforward sampling in some cases, for example, in the estimation of the transmission coefficients. This method, which often does very well, probably gives best results in this type of problem in the range of 4-6 mean free paths. Below this (i-4 mean free paths) it does not

do too well as the paper shows. Beyond 6 mean free paths the method continues to increase in efficiency relative to straightforward sampling; but not fast enough to handle the problem which is growing in difficulty by leaps and bounds. Both methods are swamped in this case. On the other hand, while the estimates of at least some of the coefficients are not improved the estimates of the angular spectrum of the emerging gamma rays are probably very much improved. Thus any judgment of the value of the method of expected values in this problem depends upon what is of most interest, transmission coefficients or the angular spectrum. Data on the use of correlated sampling is not given although Berger has been using it but it is undoubtedly very effective in increasing the efficiency with which the effects of changing geometries can be studied.

The Dismuke paper is completely straightforward. It describes the setup of an applied problem of the level of complexity where Monte Carlo really becomes interesting as a numerical technique.

IV. Conclusion

In conclusion, it may be noted that there is considerable evidence of the existence of a large non-expert interest in Monte Carlo. Unfortunately, the literature on Monte Carlo is not very helpful to anyone who wishes to do Monte Carlo problems of his own and who looks for guidance, if only through example, as to how he should best proceed. There are now a number of papers, including some in the present volume, which discuss the general principles of good Monte Carlo practice -- the most usually applicable variance reducing techniques, what to look for in general, etc. However, the Monte Carlo literature is almost completely devoid of well done applied problems, which are essential to the widespread use of the techniques. Well

done problems would probably be useful even if they merely raised standards of aspiration without teaching much. Technique can be learned by the study of well done problems, whether trivial or not. One of the good things about the present volume is that it does contribute a few more examples of applied Monte Carlo.

In most cases a little thought pays big dividends in variance reduction and the use of straightforward sampling is presumptive evidence of a lack of thought or ingenuity in design of the sampling. Some thought is required, however, since only a few people have built up the necessary intellectual capital required to do problems well without much effort. It is suggested that readers who wish to really understand the various techniques (importance sampling, systematic sampling, stratified sampling, and Russian Roulette and splitting) work through simple problems using them. The basic ideas are almost trivial, but long and disappointing experience has shown that they seem very hard to communicate in a useful way. Contact with problems seems essential to understanding for most people. Because one seems to be getting something for nothing, it is necessary to keep straight the process by which everything comes out all right in the end; the efficiency of the methods in particular cases seems unbelievable. The results quite literally have to be seen, and seen through, to be believed.