1961

Monte Carlo methods

Peter Anthony Lachenbruch
Lehigh University

Photoduplication Service
Lehigh University
Bethlehem, Pa.

Job No. _____

REQUISITION FOR PHOTODUPLICATION SERVICES

Item No.	Type of Reproduction	Complete Description	Unit Price	Item Price
1	Microfilm (negative)	Lachenbruch, Peter Anthony Monte Carlo methods. (Thesis - M.S.)		

Microfilm reel and box _____
Microfilm minimum charge _____
Postage, as required _____

TOTAL CHARGE _____

This work is done under the conditions stated on the back of this sheet, to which I agree.

Authorized Signature _____ Department _____

Account Number 1087 Date _____

Personal payment _____ (PREPAYMENT WILL NOT BE ACCEPTED. Applicant will be billed by ~~the~~ University on completion of work.)

If non-Lehigh request, name and address of institution:

_____

_____

Is copyright agreement on file? _____

MONTE CARLO METHODS

by

Peter Anthony Lachenbruch

A Thesis

Presented to the Graduate Faculty

of Lehigh University

in Candidacy for the Degree of

Master of Science

Lehigh University

1961

CERTIFICATE OF APPROVAL

This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science.

Jan. 12, 1961
(date)

Doris V. Latshaw
Professor in charge

Everett Peters
Head of the Department

ACKNOWLEDGEMENTS

The author wishes to thank Dr. V. V. Latshaw, Mr. S. Monro, and Mr. J. A. Ingram for their valuable assistance.

TABLE OF CONTENTS

Chapter I

INTRODUCTION AND HISTORICAL BACKGROUND

The purpose of this paper is to introduce Monte Carlo methods and to describe some significant applications of these methods. By computations on random numbers, Monte Carlo methods provide solutions of analytic and probabilistic problems. The following example illustrates the use of Monte Carlo in solving a familiar problem.

In elementary calculus, one of the first problems the student learns to solve is the evaluation of definite integrals which are often interpreted to be the area under curves such as $y=x^2$, between the values $x=0$ and $x=1$. We know $\int_0^1 x^2 dx = x^3/3\Big|_0^1 = 1/3$ by standard methods (see fig. 1). If a random sampling of points in the unit square is taken, about 1/3 of the points should lie below the curve $y=x^2$. Chapter III describes several ways in which sampling can be done. Because of the simplicity of the example, a primitive technique was employed here. Rather than consider the closed interval $[0,1]$ as a continuum of points, the sample space was considered to be discrete and consisted of the two-decimal numbers between 0 and 1. These numbers (and their squares) were recorded and placed in a container. To determine if a point (x,y) lay below the curve, the contents of the container were shaken and a number, r_1, was picked. This was our x value. Squaring the value of r_1 yielded x^2. We then obtained another number, r_2, which

was our y value. The values of x^2 and y were compared. If $x^2 \geqq y$ we considered the trial a success and recorded it as such; if not, the trial was recorded as a failure. After a large number of trials, we formed the ratio (number of successes)/(number of trials) which is an estimate of $\int_0^1 x^2 dx$.

It is intuitively clear that if a large enough number of trials is taken, the above ratio will approximate the value of the integral (in this case 1/3) since 1/3 of the unit square lies below $y = x^2$ and 2/3 of the unit square lies above $y = x^2$.

In performing this experiment, 38 successes out of 120 trials were obtained giving a ratio of 38/120 = .31666... This deviates from the true value of the integral (.333...) by -.01666... The sampling in this problem was done in a naive manner. By various sampling improvements equivalent results can be obtained with less effort. The more important of these sampling improvements are discussed in Chapter IV.

A simple generalization of the above problem is to determine the area within a closed curve $x = \phi(t)$, $y = \psi(t)$. In principle, the computation is the same as that involved in the evaluation of the area under $y = x^2$. We obtain a point (x,y) and determine whether it lies within the area bounded by the curve. This calculation requires more effort than the previous example due to the more complex boundary conditions involved in this problem (see figure 2).

Historical Background

Monte Carlo has its origins in a statistical method known as model sampling. Model sampling begins with a model of a stochastic phenomenon after which various elements of the population are chosen at random. Marshall notes that statisticians have been using these sampling methods in the investigation of probabilistic relationships. An example of this is ". . . the effect of non-normality on statistical test procedures devised for samples from normal (Gaussian) populations . . . ." (14a) The term will probably continue to be used; however, it has since been embodied into what we here refer to as Monte Carlo Methods.

Model sampling was generally applied to stochastic problems rather than problems arising out of an analytic process. However, in discussing Monte Carlo methods, Householder (15c) states:

> The novelty. . . in the suggestion that where an equation arising in a non-probabilistic context demands a numerical solution not easily obtainable by standard numerical methods, there may exist a stochastic process with distributions or parameters which satisfy the equation, and it may actually be more efficient to construct such a process and compute the statistics than to attempt to use those standard methods. (16a)

This concept was discovered by Ulam and Von Neumann and dominated much of the early literature on Monte Carlo.

Thus, Monte Carlo is a method of studying an stochastic model of an event. However, some mathematicians feel that

the term "Monte Carlo" should not be used unless some sort of variance reduction technique is used in the calculation. The name, "model sampling", should be used when purely random sampling is invplved. As Marshall points out, ". . . sophistication sometimes has a high price. . ." due to lengthy routines on the computers and the extra time which must be spent in analyzing the problem. He states, however, ". . . in common usage Monte Carlo is synonymous with any use of random sampling in treatment of either deterministic [analytic] or probabilistic problems." (14a)

The earliest use of Monte Carlo was in connection with the development of atomic weapons at Los Alamos. Problems in particle diffusion were modeled in their probabilistic character (see, for example, Cashwell and Everett, (1)).

Another early use of Monte Carlo occurred in the period 1947-49 when Fermi, Metropolis, and Ulam used it to obtain estimates of the eigenvalues associated with the Schroedinger equation. Kac also used Monte Carlo on this problem. Marshall noted (in 1954) "Thus far, no published results show any marked improvement over results attainable by classical techniques." (14a)

In an expository paper in 1949, Ulam expressed the concept of forming a stochastic analogue of a problem whose direct solution would be much too laborious. (20) He suggested that digital computers may be used to advantage in the field of combinatorial analysis. Ulam further stated

> ". . . the theory of probabilities, which from one point of view is a branch of combinatorial analysis is a case in point. The so-called Monte Carlo method may be said to consist of a 'physical' production of models of combinatorial situations."

For example, estimate the production of n! permutations of n objects possessing a given property P (in practice, complicated). For large values of n, it is impossible to consider all cases, so the procedure is to check a large number of permutations at random and observe the proportion having property P.

Matrix inversion by Monte Carlo has also been studied. The basic idea was devised by Von Neumann and Ulam and is based on the fact that $A^{-1}=\sum_{k=1}^{\infty} A^k$. Standard methods have proved faster and more accurate for this problem. (4,6)

Reliability estimation has been studied by S. I. Firstman. (5) Suppose an estimate of the reliability, R, of a piece of equipment is desired. The distributions and parameters are specified and estimates are made. The calculation is repeated a large number of times (depending on the accuracy desired), and the results are tabulated. A graph may be drawn showing the cumulative distribution as shown in figure 3. Figure 3 indicates that a represenatative machine has the probability $P(R \leq X)$ that its reliability is less than X.

This problem may include down times (inoperative due to some malfunction) following a breakdown, temperature, voltage and humidity variations. For an example of a

down time calculation, after having determined R for a specific machine, choose a random number S. If $S < R$, the machine has not failed. If $S \geq R$ the machine has failed and can be repaired with some probability, say p. By choosing additional random numbers, we may determine whether or not the machine is repairable. If so, the down time may be determined by selecting another random number. Consequently, by running a large number of trials, an estimation of operating time can be made.

The early theoretical treatments of Monte Carlo concentrated largely on the solution of the above type of deterministic problems. In practice, few of these have been solved by Monte Carlo due to the amount of computer time needed. The dominance of this type of problem in the literature is due to its potentialities and the fact that many of the contributors have been working in the field of theoretical numerical analysis than on applied problems.

Requirements of a Monte Carlo Calculation

The requirements of a Monte Carlo calculation are few and relatively simple. Firstly, the distributions and parameters of the stochastic process must be specified. These may be varied during different runs of the same experiment, but they must be specified each tiem. Secondly, the problem should not be given to a quick solution by usual numerical or deterministic methods. That is, it may be solvable by analytic methods, but the solution is more

economical if obtained by a Monte Carlo calculation. Thirdly, a supply of random numbers (usually uniformly distributed) is required. To paraphrase Saaty, a random number is a real variable whose value is determined by chance. (18) The random numbers may be supplied directly as values of some probability density function or they may be supplied as the values of some distribution function ($r = P(X \leqq t_o)$). It may be noted here that the distribution function (which is sometimes called the cumulative distribution) $P(X \leqq t_o) = \int_{-\infty}^{t_o} p(s)ds$.

A natural question arises at this point. Assuming the distributions and parameters have been specified correctly, and the random numbers correctly chosen, what error will be present in the final answer? It is improbable that it will be exactly correct, but with some sophistication, satisfactory results can be obtained. The error is statistical in nature, so, fortunately, there is a good body of theory to refer to.

Suppose the quantity it is desired to obtain is Z^* (which is, of course, unknown). Let the Monte Carlo estimate of Z^* be denoted by Z' and suppose there have been N trials to obtain Z'. That is, $Z' = (1/N)\sum Z_i$. An estimate of the variance is $V = (1/N)\sum(Z_i - Z')^2$. By referring to the Central Limit Theorem it is seen that $P\left(\left|\frac{Z^* - Z'}{\sigma / N^{\frac{1}{2}}}\right| < K\right) = \phi(K) - \phi(-K)$ where $\sigma^2 = V$, K is some arbitrary number > 0, and $\phi(K)$ is the cumulative normal distribution. It seems apparent that in performing a Monte Carlo calculation either of two parameters can be manipulated to minimize the error. Either the number of trials may be increased or some variance reducing de-

vices may be sought. The variance is an inverse function of the number of trials, so by doubling the number of trials, we effectively halve the variance. This, unfortunately, can prove to be quite expensive; for example, in a problem calling for a minimum of 10,000 separate trials. There are numerous variance reducing sampling techniques, a few of which are listed here:

1) Importance Sampling
2) Russian Roulette and Splitting
3) Use of Expected Values (combination of analytic and probabilistic methods)
4) Correlation and Regression
5) Systematic Sampling
6) Stratified Sampling (Quota Sampling)

These methods are considered more fully in Chapter IV.

When a variance reducing device is used, care must be taken to include this fact in the calculations; otherwise, the results will be biased. The list of techniques given here is by no means a complete one. Indeed, one can design and use one's own variance reducing device as the need arises.

The Variance of an Estimate of a Probability p

If a probability p is estimated by $p'=n/N$, the variance is given by $(1/N)(p(1-p))$. Proof: Consider the event A. Let X be the number of successes in N trials. Then $E(X)=\sum x\,b(X,N,p)$ where $b(X,N,p)$ denotes the binomial distribution, X the number of successes, N the number of trials, and p the probability of success on any one trial. It is well known that $E(X)=Np$

and $Var(X)=E((X-Np)^2)=E(X^2)-(E(X))^2$. By a standard result we have $Var(X)=Np(1-p)$. If, however, we multiply X by $1/N$ we then would obtain the expected value of p and $Var(X/N)=Var(p)=E((1/N(X-Np))^2)=E(X^2/N^2)-(E(X/N))^2=(1/N^2)Var(X)=(1/N^2)Np(1-p)=(1/N)p(1-p)$ which is the desired result.

Chapter II

APPLICATIONS

Introduction

The five applications of Monte Carlo to be discussed in this section are: (a) The Evaluation of Integrals;
(b) Nuclear Diffusion Problems;
(c) Length of Service Problems;
(d) Partial Differential Equations;
(e) Queueing Theory.

The topics are briefly summarized here. The evaluation of integrals is the most widely known theoretical application of Monte Carlo. It is not suggested that this technique be used in two-dimensional cases where Simpson's Rule or Gaussian Quadrature Formulas can be used easily. It is in the mult-dimensional cases where the integrand is not of an elementary type that a Monte Carlo method is beneficial. Next, Nuclear Diffusion problems will be considered. In practice, Monte Carlo has been applied more often in this field than in others because it has been easier to formulate the problems. No probabilistic analogue of an analytic concept occurs in this type of problem. The method of solution of Nuclear Diffusion problems uses random numbers to estimate random processes. The third case, Length of Service problems, has military applications. One specific case, not discussed in detail, is determining the range of a nuclear powered submarine. The solution of Partial Differential Equations will then be considered. The specific case of Laplace's equation is discussed and brief mention is given to certain research performed on the determination of

the lowest eigenvalues of Schroedinger's equation. Finally, a short resume of the applications of Monte Carlo to Queueing Theory is given.

The Evaluation of Integrals

As an introductory example, consider the problem given in the first chapter: the evaluation of $\int_0^1 x^2 dx$. The unit square was divided by a lattice of 10,000 points, (x,y), and values of x and y were chosen from a sample set consisting of the two-decimal numbers between 0 and 1. The estimate of the integral was made by forming the ratio of the number of these "lattice points" which were chosen. A generalization of this procedure to multi-dimensional integrals follows.

Let X be an n-dimensional vector, R some n-dimensional region and f some function of X. The problem given above becomes

$$J=\int_R f(X)dX.$$

J will be evaluated by reducing J to $J'=\int_{R'} dX'$; where X' is an (n+1)-dimensional vector and R' is defined by the inequalities: $G_1(X')<0$, ...$G_m(X')<0$. The $G_i(X')$ are evaluated at each point. If a point satisfies these m inequalities, it is classified as a success; if not, as a failure. The total of successes divided by the total trials gives an estimate of the proportion of the sample space which the region R' occupies. More compactly, a large number of lattice points (not necessarily of the order of the total of all lattice points) are taken at random and

it is determined if they satisfy the inequalities $G_i(X')$. The Monte Carlo ratio is then calculated.

Instead of determining a set of inequalities, G_i, which may prove difficult or impossible to do, a more direct method may be attempted. As before, let $J=\int_R f(X)dX$. The quantity $\sum_j w_j f(Z_j)$ is an unbiased estimate of J where the Z_j are chosen at random from R and the w_j are weighting factors with the property $\sum_j w_j=1$. If enough information is known about the function f, the error can be reduced by suitable sampling modifications (see Chapter IV).

Monte Carlo methods are rarely used to evaluate integrals because they are lengthy and cumbersome. Usually, a large number of samples must be taken to obtain any degree of accuracy. Frequently, the functions which must be evaluated in the course of the computation are complicated and difficult to work with. In computers, for example, the extraction of roots is a time consuming process. In certain complicated problems Monte Carlo methods, although time-consuming, are the only method available and should be used. Hammersley has published some results of his investigations of the evaluation of integrals. (8) (Also, see bibliography in (8))

Nuclear Diffusion Problems

The Monte Carlo technique in solving this class of problems is very realistic since the processes which are simulated by Monte Carlo are, as far as is known, random in nature. In this class of problems, we determine what percentage of particles starting at a given source terminate in various categories after passing through a given medium with a known configuration. Some of these

categories might be: escaped from the system, absorbed within the system, and energy too low to be of interest. Each trial traces a particle from the time it leaves the source until it can be classified. These trials are called histories. If the probabilities of the elementary events are known, the Monte Carlo method is used at each stage in a particle's history to determine its behavior. (1)

A set of parameters is needed to determine the initial conditions of the problem. The set will consist of the particle's position, direction, and, when the particles are not from a monoenergetic source, its energy. The first step in the computation determines if a particle, starting at an arbitrary point within a region, has a collision in the region or escapes from the region. The new parameters of the particle are computed before continuing. If the particle escapes from the region, it is next determined if the particle has escaped from the system. If it has, the fact is noted and a new particle is examined. If it has not escaped from the system, the region it has passed into is determined by the new position parameters. The first step is repeated---it is determined whether the particle has a collision or escapes from the new region.

If the particle has a collision in the original region, the type of collision (such as elastic scattering, inelastic scattering) is calculated. It is decided whether the particle was captured or scattered; if scattered, the angle of scattering and the particle's new energy must be

calculated. After the new parameters have been calculated, the first step is repeated using the new parameters. In some cases, the scattered particle will not have enough energy to be of further interest. If the particle has been captured, the terminal category (such as loss to weight cutoff, loss to energy cutoff) is selected and a new particle is studied.

As an example, consider the problem of the effects of neutron irradiation of human tissue. This has been studied by Nancy M. Dismuke of the Oak Ridge National Laboratories. (14d) In the problem, random processes arise naturally from the physical phenomena being investigated. "Whenever a random selection must be made in our experiment, the corresponding physical situation seems to be a matter of random choice." (14d)

The problem was to estimate the damage done by a broad beam of neutrons at energy level E_0, to a slab of human tissue. The value E_0 was fixed during the problem. A sample of 10,000 neutrons was examined and several functions were calculated from which the damage was determined. These functions are:

1. $E_T(X)$ Average energy imparted to tissue atoms by interation type T (one of six possible types)
2. $C(X)$ Collision density of thermal neutrons
3. $S_1(X)$ Source created by neutrons slowing past 1 ev level
4. $S_{th}(X)$ Source created by neutrons slowing past 1/40 ev level (14d)

The effective target area that a nucleus presents is called its cross-section and is denoted by σ. It is a function of the type of particle bombarding the nucleus and of the energy of the particle. The probability that a neutron of energy E_0 is scattered by an atom of a certain element is σA, where A is the number of atoms of that element present.

In this problem it was assumed that biological damage was due to energy dissipation of γ-rays, protons emitted in certain absorption reactions, and the energy dissipation of the recoil ions which were produced. The protons and recoil ions caused only local damage but the γ-rays were damaging at some distance from their origin.

The slab of tissue lay at $0 \leq X \leq 30$ cm. The program was begun by starting a neutron, with energy E_0, at the origin in the positive X-direction. Then a distance L_1, travelled by the neutron before the first collision, was picked at random. The coordinates of the point at which the collision took place were computed. If the collision was inside the slab, the type of interaction was chosen, the new parameters computed, and an increment was added to the appropriate $E_T(X)$. A thermal neutron's contribution would also be added to $C(X)$. If an energy level was crossed, this was noted, along with the coordinates at the time. The process was repeated until the neutron was absorbed or had passed through the slab (escaped from the system).

After 10,000 histories had been examined, and the functions $E_T(X)$, $C(X)$, $S_1(X)$, $S_{th}(X)$ evaluated, the average damage to the tissue was determined by assuming that this damage was proportional to the energy gained by the tissue. These proportionality factors are called RBE (Relative Biological Effectiveness) factors.

Length of Service Problems

"Length of Service problems are . . .multivariate problems in sequence." (15) Some examples of this type of problem are: 1. A manufacturer of a new product desires to learn the average lifetime of this new product. By using a Monte Carlo method he can obtain estimates of these data. 2. Suppose we want to determine if a missile is capable of travelling a certain minimum distance. Some of the phenomena involved are random in nature (such as thrust, specific impulse, residual fuel) so the Monte Carlo method has been used in this field. 3. The Navy may desire performance estimates of a nuclear submarine. For this problem, an additional Monte Carlo run may be desired to compute characteristics of the nuclear power plant. 4. An automobile company may want gasoline mileage estimates of a new model automobile. This problem will be discussed in more detail.

Initially, the parameters and distributions which characterize the problem are specified. These may take the form of design specifications, theoretical calculations, or simply, given functions of empirical data.

The more accurately the initial conditions are known, the more confidence may be placed in the result.

The answer is computed by evaluating the function of the random variables for many arguments and the values averaged. Some values may be weighted more heavily than others. This is because more information may be available about some points and less about others.

Consider the problem of establishing a gas consumption rate for a new model automobile. Suppose the designers estimate the mileage as X miles per gallon in the city, Y miles per gallon in suburban zones, and Z miles per gallon in the country, with standard deviations v_x, v_y, v_z, respectively. Sometimes, when the tank will not feed gas any more, there is a small amount of residual fuel left. This amount varies between 0 and W. For simplicity, suppose these quantities are uniformly distributed. Finally, suppose that p_1 is the proportion of city driving, p_2 is the proportion of suburban driving, and p_3 is the proportion of country driving. Choose four random numbers: r_1, r_2, r_3, from the uniform distribution between -1 and 1, and r_4 from the uniform distribution between 0 and 1. Since the former distribution has mean 0 and standard deviation $2/\sqrt{3}$, the distribution must be adjusted to have the appropriate mean and standard deviation (gasoline consumption for city, suburban, and country driving). This is accomplished by multiplying the random numbers by $v_x\sqrt{3}/2$, $v_y\sqrt{3}/2$, or $v_z\sqrt{3}/2$, giving r_1', r_2', and r_3'; and

then adding r_1', r_2', r_3' to X, Y, or Z, respectively. The computation of overall miles per gallon, M, is now quite simple:

$$M=(X+r_1')p_1+(Y+r_2')p_2+(Z+r_3')p_3.$$

The residual fuel is Wr_4. The total gasoline used is $C-Wr_4$. where C is the capacity of the tank. The total miles travelled is $M(C-Wr_4)$.

Repeating these calculations a large number of times provides good estimates of gasoline mileage. These data can be represented as an arithmetic average, a table giving results of each trial, or a graph ($y=P(\text{miles per gallon} \leq x)$). The problem can be made more general or complex in a number of ways. The distributions of the mileages (X,Y,Z,) need not be uniform, or the proportions (p_1, p_2, p_3) can be chosen randomly.

The process given above is merely a specific example of the more general class of length of service problems. An entirely analogous procedure is followed to solve these questions.

Partial Differential Equations

Two examples of Monte Carlo applications in the field of Partial Differential Equations are given in this section. Laplace's equation in two dimensions, $\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0$, has been solved, using Monte Carlo, independently by E. C. Yowell and by J. Todd. (22, 23)

> "There is a structural relation between simple differential equations and stochastic approximation methods involving sequentially dependent variables. The lack of a substantial body of statistical theory dealing with such variables makes Monte Carlo indispensable in this area." (15)

Suppose it is desired to solve the Laplace equation $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$ for $0 \leq x \leq 1$, $0 \leq y \leq 1$ with boundary conditions $U(x,1) = \sin \pi x$, $U(x,0) = U(0,y) = U(1,y) = 0$. Todd gives the exact solution as $U(x,y) = (\sin \pi x \sin h \pi y)/\sin h \pi$. The Monte Carlo procedure in this problem is to let a particle begin a random walk at a point P. When the particle reaches a point Q on the boundary, the boundary function is evaluated at Q. It has been shown that the expected value of $U(Q) = U(x_Q, y_Q)$ is U(P), the value of U at P. (2)

The random walk is described as follows: start a particle at some point P_o interior to the region. Then if the particle is at some point (x_n, y_n), at the n^{th} step, it will move to one of the points (x_n, y_n-h), (x_n-h, y_n), (x_n, y_n+h), (x_n+h, y_n) where h is the length of a step; the probability of moving to any one of these points is 1/4. The random walk is continued until the boundary is reached; the function is then evaluated at this boundary point. The average of these boundary values will converge to the functional value at the starting point. In the case under consideration, a walk terminates if either x_n or y_n assumes the value 0 or the value 1. In this problem the value 1/16 was used for h.

Todd described an importance sampling technique (Chapter IV) for use in this problem. He assumed that an approximation U^* to U was known and then used a walk biased towards points at which U^* was large. The bias was removed by use of a weighting factor. He noted that the length of time required to compute 64 walks was "about five minutes . . . compared to 20 seconds required in the simple case." (20) The simple case, was, of course, the one in which there was no variance reducing techniques used.

It is apparent that importance sampling was not advantageous in this case. Indeed, in the problem of solving partial differential equations in two dimensions, the Monte Carlo method is not an economical one since more accurate results can be obtained by relaxation methods. Although in two dimensions, Monte Carlo is not practical, in higher dimensions Monte Carlo is more advantageous than relaxation methods since comparable results can be found by Monte Carlo without large increases in cost over the two-dimensional case, whereas the application of relaxation methods increases in difficulty rapidly as the number of dimensions increases.

The solution of eigenfunction-eigenvalue problems is another area of study which offers great theoretical potentialities. Donsker and Kac used the method to obtain estimates of the lowest eigenvalue and principal eigen-

function of Schroedinger's equation. (3) Wasow used a similar technique to determine the eigenvalues of elliptic difference equations. (21)

The one-dimensional Schroedinger equation is:

$$\frac{1}{2}\frac{d^2\psi}{dx^2} - V(x)\psi(x) = -\lambda\psi \qquad \psi > 0$$

In the solution (3), the first eigenvalue agrees fairly well with the exact value, but the second eigenvalue does not show comparable behavior.

In general, Monte Carlo does not bring great improvement over conventional methods in the solution of partial differential equations. If improvements in the techniques of handling these equations and faster and larger machines are developed, a much larger class of partial differential equations will be able to be solved than at the present time. Indeed, machines such as LARC and STRETCH are probably able to handle the computations now. Monte Carlo Methods should be reserved for those problems in this field which are of such a magnitude that they cannot be solved by conventional methods.

Queueing Theory

There are many places in Queueing Theory where random factors can arise. For example, the number of arrivals per unit time, mean service rate per channel, and mean number of free channels all have a random character.

A simple application of Monte Carlo is given below. The time intervals between arrivals (I) are obtained by random

selection; similarly, corresponding service times (S) are determined. The waiting time (W) for each unit is computed by the following formula: W(present unit)= W(previous unit)+S(previous unit)-I. If this quantity is<0, the waiting time is zero.

Saaty gives a list of typical problems in Queueing Theory. (18) The list is approximately as follows:

a) activity at a telephone switchboard
b) the loading of aircraft (stacking, etc.)
c) the loading and unloading of ships
d) scheduling of patients in clinics
e) customers and taxis at a stand
f) customers at a store (random selection for service)
g) items in a production line.

This is only a very brief list of problems which can be solved by applying Monte Carlo techniques to Queueing Theory. The interested reader can obtain further references by consulting Saaty (18), or any of the Operations Research journals.

Chapter III

TECHNIQUES WITH RANDOM NUMBERS

All Monte Carlo methods require a supply of random numbers. This section explains the following:

(a) How to obtain random numbers

(b) Important tests for randomness

(c) Conversion of random numbers from a Uniform distribution to a given distribution

Methods of Obtaining Random Numbers

Tables of random numbers are available and provide the most direct means of obtaining random numbers. The major work in this field is A Million Random Digits with 100,000 Normal Deviates by the Rand Corporation. (17) These numbers were produced by the electronic equivalent of a 32-place roulette wheel. Additional information concerning the generation and testing of these numbers can be found in the work itself or in (16b)

Unfortunately, tables of random numbers can be used only in a limited number of cases. In almost any case of interest, performing the computations by hand proves cumbersome and, in most instances, impossible. The entire calculation can be performed on a computer by feeding in random numbers one by one and storing them within the computer. The Rand table is available on punched cards for this purpose. However, in a high-speed digital computer this would be a great waste of storage space if anything more than a small quantity of numbers was needed. Since the input-output function is the slowest mode of operation

for a large-scale computer, it would be a waste of time if the random numbers were fed in only when needed. The calculations we are concerned with are usually of such a magnitude as to require a large supply of random numbers. Thus, we see that for computer applications, a method of internally generating random numbers is necessary.

Numbers can be generated which pass certain tests for randomness; in this paper, these sets of numbers will be called pseudo-random numbers.

The Mid-square Method provides a means of generating pseudo-random numbers. The square of an n-digit number will contain 2n digits (some digits at the beginning or end may be zero). Long sequences of random numbers can be obtained by taking the middle n digits of the square of some n-digit number as the next member of the set of pseudo-random numbers. For example, let r_0 equal 1234, then r_0^2=01522756 and r_1=5227. Similarly, r_1^2=27321529 and r_2=3215. When a number appears for the second time, all numbers which follow it will appear for the second time. To obtain the longest possible sequence, the starting value, r_0, must be chosen carefully. The process can be modified by multiplying two n-digit numbers and extracting the middle n digits of the product.

The Congruential Multiplicative Method is most commonly used because sequences of extremely long period can be generated which have desirable randomness properties. The numbers are produced according to the formula r_{n+1}=

$kr_n \pmod p$. The choice of k, p, and r_0 is quite important; if k, p, and r_0 are not properly selected, a sequence may be generated which has too short a period. However, it is possible to obtain long sequences of pseudo-random numbers with this method. Lehmer used k=23, $p=10^8+1$, $r_0=1$ and obtained a sequence of length 5882352. This type of pseudo-random number sequence terminates in the same way as the Mid-Square Method sequence does; a number which has appeared before is obtained. L. E. Cunningham performed randomness tests on 5000 numbers generated in this way and the numbers were found to be satisfactory.

A similar approach is the Congruential Additive Method. The paper of Taussky and Todd (14b) states,

> "The only practical reason to search further for processes to generate random numbers is to gain speed. The obvious suggestion is to try using addition instead of multiplication."

That is, construct a Fibonacci sequence modulo P. P-1 is usually the largest number acceptable by the computer. Then by using fixed-point addition and disregarding overflow, the reduction mod P can be made. Unfortunately, the numbers are not independent; hence, not truly random.

In summary, we see that random number tables have good randomness properties, but are inconvenient to use in practise. Mid-Square Methods are easier to use in practise than random number tables. Although the programming of Mid-Square Methods is difficult, they are more readily adaptable to computers. The numbers possess the desired randomness properties,

but are sensitive to the choice of the initial value. Congruential Multiplicative Methods are the easiest pseudo-random number generating techniques to use in a computer. They posses desirable randomness properties but are sensitive to the choice of k, r_0, and p. Finally, Congruential Additive Methods, while faster than Congruential Multiplicative Methods, do not furnish the required random numbers.

Tests for Randomness

Several tests of the randomness of number-sets will be described next. These tests are necessary to check on the randomness of a sequence of numbers which shall be used in our calculations. The most important work in this field is "Randomness and Random Sampling Numbers" by Kendall and Smith. (10) The four tests described by Kendall and Smith are as follows (In all cases suppose that the set being tested has N members):

1. Frequency Test All digits should appear an approximately equal number of times (N/10).

2. Serial Test. No digit should tend to be followed by any other digit. That is,

> "if. . . we form a bivariate table..... arranged in rows according to the first digit, and in columns according to the second digit, we should get frequencies which are approximately equal in all the cells." (10)

3. Poker Test. If the digits are chosen in groups of five, certain poker-type configurations should be expected: a bust,

a pair, three of a kind, a full house, four of a kind, five of a kind.

4. Gap Test.

> "Finally, there are certain expectations in regard to the gaps occuring between the same digits in the series. For instance, if we take one digit, say, zero, in about one-tenth of the cases the first zero will be followed immediately by a second zero and there will be no gap. In about nine-hundredths of the cases there will be one digit between the two zeros. In about eighty-one-thousandths of the cases there will be a gap of two digits between successive zeros and so on." (10)

Kendall and Smith state that it is very difficult to find a non-random sequence which evades all of the above tests. A locally random sequence is defined to be one which passes these tests. A sequence may be acceptable when taken as a whole, but a part of it may be decidedly non-random. As an extreme case of a very large sample, consider a locally random set of $10^{10^{10}}$ digits. It is practically certain that a block of a million zeros occurs. If we were to sample from this block we would decide that the set was definitely non-random. Thus, care should be taken that the sample set is truly random.

In all of the above tests, the expected frequency is calculated, a sampling is made, and a χ^2 is computed with the appropriate degrees of freedom. The usual procedures in applying a χ^2 test are then followed.

5. The d^2 Test. Random digits are used to select random points in the unit square. The propability that the square

of the distance between two points is less than α^2 is given by:

$$P=\pi\alpha^2-8\alpha^2/3+\alpha^4/2 \text{ for } 0\leq\alpha^2\leq 1.0$$

$$P=1/3+(\pi-2)2+4(\alpha^2-1)^{\frac{1}{2}}+(8/3)(\alpha^2-1)^{3/2}-\alpha^4/2-$$

$$4\alpha^2\sec^{-1}\alpha \text{ for } 1.0<\alpha^2<2.0$$

The probabilities are given in (5) for α^2 varying in increments of .1 from 0.0 to 2.0. It is noted that three digits suffice to represent one coordinate. The authors, Gruenberger and Mark, remark that this test is more stringent than those of Kendall and Smith.

Several other randomness tests may be used although they are not as powerful as the previous five. They are given below.

Periodicity Test. If a pseudo-random number generator is used, the original sequence (or some subset of it) will eventually be repeated. This repeated sequence should not have a small period; the minimum acceptable length depends on the user's needs.

Serial Correlation Test. The set of pseudo-random numbers may have some correlation of digits with lags of R positions (i.e., n_i may be correlated with the digits in position n_i+R). Serial correlation with a lag of R numbers may be tested on a total of N+R numbers. The correlation is computed according to the formula:

$$C_k=\frac{N\sum X_i X_{i+k}-(\sum X_i)^2}{N\sum X_i^2-(\sum X_i)^2}$$

where C_k is the correlation coefficient with a lag of k positions.

Here $E(C_k)=-1/N-1$; $Var(C_k)=1/N-2$.

Of the last two tests given above, only the Serial

Correlation Test is seen with any regularity.

Random Number Conversion Methods

Generally, the random numbers which are obtained in the course of a Monte Carlo calculation are uniformly distributed between 0 and 1 (this distribution has mean = 1/2 and variance = 1/3). The numbers which have been obtained are converted to non-uniform distributions which arise in the problems. One means of accomplishing this is through the use of mappings. We apply the transformation $y = T(x)$, where x is a uniform random number. In practice this is not easily done, as the transformations may be quite complicated.

Another method of conversion is through the use of the Rejection Technique. A simple example of the Rejection Technique was given at the beginning of this paper in the evaluation of $\int_0^1 x^2 dx$. What really is going on is this: if y lies below f(x) accept it, otherwise pick another set of values (x,y) and try again. Suppose the random variable, x, has the probability density function f(x) given by:

$$f(x) = 0 \quad x < a \text{ or } x > a + b$$

$$0 \leqq f(x) \leqq M; \quad a \leq x \leq a + b$$

(necessarily $\int_{-\infty}^{\infty} f(s)ds = \int_a^{a+b} f(s)ds = 1$)

Kahn formalizes the process in 2-space as follows:

1. Obtain two random numbers, R_1 and R_2.
2. If $R_1 \leqq f(a+bR_2)/M$ let $x = a+bR_2$.
3. If $R_1 > f(a+bR_2)/M$, pick two random numbers R_1' and R_2' and try again.

The probability of accepting an x value is f(x)/M. (9)

More generally, let n(x) and m(y) be probability density functions and H(x) be an arbitrary function. Then:

1. Select x from n(x).
2. Independently choose y from m(y).
3. If $y \leq H(x)$, accept s, otherwise repeat steps one and two.

In step three above, it may prove easier to check an inequality of the form $g(y) \leq h(x)$. This is equivalent to the original formulation with $H(x) = g^{-1}(h(x))$. There are many variations of this technique, some of which may be seen by referring to Kahn. (9) As examples of the rejection technique, consider how to obtain three non-uniform distributions from a uniform distribution. Notation: X is a random variable with the desired distribution, U(a,b) is the uniform distribution between a and b, and f(X) is the probability distribution of X.

1. $f(x) = e^{-x^2/2}$, $0 \leq x < \infty$. This is a half-Gaussian distribution in the right half-plane.

Choose R_1 and R_2 from U(0,1) and let $y = -\ln R_1$, $z = -\ln R_2$. Then if $(y-1)^2 \leq 2z$, let $x = y$; if not, pick two new random numbers R_1, R_2 from U(0,1) and repeat the above processes.

Another method of sampling (from an approximate Gaussian distribution) is by using the Central Limit Theorem, which states that the random variable $x = \sum R_i$, where R_i is a member of U(-1,1), is normally distributed with zero mean and variance $2n/\sqrt{3}$ as $n \to \infty$. The members of U(-1,1), are obtained by applying the mapping $R' = 2R-1$, to members of U(0,1). Then $x = 2\sum R_i - n$, where R_i is a member of U(0,1).

2. $f(i) = a^i e^{-a}/i!$ $0 < a$, $i = 0, 1, 2 \ldots$ (Poisson distribution). Initially choose $k = 0$, $y_0 = R_0$. The testing procedure consists

of determining if $y_k < e^{-a}$. If so, let i = k. If not, replace k by k + 1 and y_k by $y_{k+1} = R_k y_k$ and retest. Since $R_k \leq 1$ and y_k is monotonically decreasing, the probability that $y_k > e^{-a}$ (i.e., fails the test) is given by $1 - e^{-a} \sum a^j/j!$. The average number of R_k used for each i is a + 1.

3. $f(x) = e^{-x}$ $0 \leq x < \infty$. (Exponential distribution) Choose R_o, R_1. The testing procedure is as follows: if $R_{i-1} \leq R_i$, pick R_{i+1} and repeat the procedure (i.e., inquire if $R_i \leq R_{i+1}$). If not, then $R_{i-1} > R_i$ and let $Z = R_o$. Choose x = Z + L, where L number of failures. The probability a trial will be rejected is e^{-1}.

These three conversions are specific examples of the rejection technique. The rejection technique can be used to sample from any desired frequency distribution. Thus, one does not need to use mappings or to have a separate generating function for each distribution of interest.

Summary

In this chapter, methods of obtaining sets of random numbers have been discussed. Tables of random numbers were mentioned along with algorithms for generating pseudo-random numbers, and their advantages and disadvantages noted. The main advantage of tables of random numbers is that they are obtained from random processes such as neutron emission from a radiative source or a random noise generator (thus truly random), rather than generated by an algebraic process. Unfortunately, they are very clumsy to use in digital computers. Pseudo-random numbers appear to be random when subjected to a series of

tests and may be used in place of tables of random numbers. They are much easier to use than tables since the generating functions take up very little space in the memory of the computer. A possible disadvantage to the use of pseudo-random numbers is that they are not truly random in the sense mentioned above, but are generated by an algebraic process. Because the pseudo-random numbers might not possess the desired randomness properties, seven tests of randomness have been described. These tests can be performed by the computer. It may be necessary to convert a set of uniform random numbers to a set of non-uniform random numbers. This need arises in practice when the distribution of the problem is non-uniform. The methods of conversion by mappings and by the rejection technique were described.

Chapter IV

VARIANCE REDUCTION TECHNIQUES

In this chapter, Straightforward Sampling shall be discussed in addition to the six sampling methods which were mentioned in the first chapter. The following notation conventions shall be used in the description of the various techniques. The expected value of a random variable, say U, is denoted by $\bar{U}$. The space, S, in which sampling is done is partitioned into M mutually exclusive regions, S_1, S_2, . . . S_M. For each value of k, $1 \leq k \leq M$, a value of the mean, $\bar{Z}_k$, and sample variance, $V_k = \overline{Z_k^2} - \bar{Z}_k^2$, of the quantity being estimated is obtained. The mean over the entire space, $\bar{Z} = \sum_k \bar{Z}_k g(k)$, where g(k) is the probability that a random point is in the region S_k.

a. Straightforward Sampling

In many cases, the simplest way to attempt sampling is to construct the desired distribution (e.g., by using techniques described in Chapter III) and pick samples from it. This method is known as Straightforward Sampling (also known as Model Sampling). The estimate of the mean and variance are given by $Z^* = (1/N)\sum Z_i$ and $V_a = (1/N)(\sum V_j + \overline{(\bar{Z}_j - \bar{Z})^2})$. This is often a slow and inaccurate process.

We can improve on the results obtained by this method by attempting to reduce the variance.

A short example will serve to illustrate this point. As in the first chapter, suppose we are estimating a probability

p by n/N. The variance was given by V=p(1-p)/N. One way of reducing V would be to increase N. However, if the value of p is increased (decreased) when $p \geq \frac{1}{2}$ (when $p \leq \frac{1}{2}$), the variance will be decreased. We shall see in the following paragraphs how the value of p may be increased or decreased.

b. Importance Sampling

Importance sampling may be described briefly as picking samples from regions of interest. One draws "samples from a distribution other than the one suggested by the problem and then . . . applies an appropriate weighting factor which . . . corrects for having used the wrong distribution." (14c) Thus, the chance that the sample is drawn from an interesting region can be increased. Kahn notes "In principle, . . . it is always possible to design an Importance Sampling scheme that has zero variance." (14c) The significance of this statement is that if one is fortunate, or has good intuition, one may be able to choose a very efficient sampling scheme form the infinite number available.

Formalizing this process:

1. A "k" is chosen from $g^*(k)$ instead of $g(k)$. $g^*(k)$ is the function which changes the probability that the sampling is performed in the region S_k.

2. Select a point from the distribution function for the region S_k.

3. With the coordinates of points chosen in this way, evaluate $Z^* = \sum Z_k^*$ where $Z_k^* = \frac{g(k)}{g^*(k)} Z_k$

Three conditions must be met:

a. $g^*(k) \neq 0$ unless $g(k) = 0$

b. $0 \leq g^*(k) \leq 1$

c. $\sum g^*(k) = 1$

Outside of these restrictions, the $g^*(k)$ are arbitrary. It can be shown that the expected value of Z^* is $\bar{Z}$. However, $\overline{Z^{*2}}$ is not equal to $\overline{Z^2}$. Kahn (9) derives an expression for the variance of Z^*: $V_b = (1/N)(\overline{Z^{*2}} - \bar{Z}^2) = (1/N)\sum \frac{g(k)^2 \overline{Z_k^2}}{g^*(k)} - \bar{Z}^2$. To minimize V_b, the quantity $\sum \frac{g(k)^2 \overline{Z_k^2}}{g^*(k)}$ must be minimized since $\bar{Z}$ is independent of the manner in which the sampling was performed. By using standard Calculus of Variations techniques the optimum $g^*(k)$ is found to be $g^*(k) = (1/D)g(k)\sqrt{\overline{Z_k^2}}$ where $D = \sum g(k)\sqrt{\overline{Z_k^2}}$. (9) The minimum $V_b = (1/N)\left(\sum g(k) \sqrt{\overline{Z_k^2}}\right)^2 - \bar{Z}^2$. The difference $V_a - V_b$ is $\left[\sqrt{\overline{Z_j^2}} - \sqrt{\overline{Z_k^2}}\right]^2$. That is, V_a is reduced by the variance of random variable whose probability of assuming the value $\sqrt{\overline{Z_k^2}}$ is $g(k)$. It is usually easier to construct $g^*(k)$ proportional to Z_k. In this case, $g^*(k) = g(k)\bar{Z}_k/\bar{Z}$. The variance reduction is then $\left[\overline{Z_k^2}(1-\bar{Z}/\bar{Z}_k)\right]$.

c. Russian Roulette and Splitting

This is a sequential sampling technique which is used in multi-stage sampling. The value of X_1 is selected, then a value X_2, and so on.

> "For some regions of X_1 the error is small enough or the expense of picking . . . [X_2, X_3, . . .] and evaluating Z large enough that it isn't efficient to pick a . . . [set of values X_2, X_3, . . .] and evaluate Z for every such X_1

> that has been sampled, but only a portion of them. In other regions of the X_1-space the error in the . . . $[X_2\text{-}X_3\text{-}. . .]$ space is large enough or the additional cost small enough that it pays to sample many values of . . . $[X_2, X_3, . . .]$ for each X_1 value that has been picked." (14c)

Stepwise, the process is as follows:

1. Every region S_k is classified as being of type I or type II. In a type I region it is not desirable to select many samples for one or both of the reasons mentioned above. In a type II region, for each X_1 choose one or more sets $(X_2, X_3, . . .)$ and evaluate Z.

2. Russian Roulette is the name of the sampling done in a type I region. If X_1 is in a type I region, S_k, we play an additional game of chance. Suppose we only wish to consider the samples from type I regions, S_k, q_k of the time. Select a random number; if it is less than q_k pick $(X_2, X_3, . . .)$, evaluate Z, and record Z/q_k for the sample. If it is not less than q_k (with probability $1\text{-}q_k$), do not select $(X_2, X_3, . . .)$; record zero for the sample. This is especially useful when Z is a complicated function and there are a number of type I regions. The variance in a type I region S_k is $V_{ck} = (\bar{z}_k^{2/q_k}) - \bar{z}_k^2$.

3. Splitting is used in a type II region S_t. If X_1 is a member of such a region, choose n_t sets $(X_2, X_3, . . .)$ where n_t depends on X_1. Then the sample estimate is the mean of the values $Z(X_1, X_{2i}, . . .)$. The variance is given by $V_{ct} = (1/n_t)v_t$, where v_t is the variance of the sample in region S_t.

For the variance of the entire sample (all type I and all type II regions) Kahn (9) shows: $V_c = V' + \sum_I g(k)\bar{Z}_k^2/q_k + \sum_{II} g(k)v_k/n_k$ where $V' = \sum_{II} g(k)\bar{Z}_k^2 - \bar{Z}^2$.

The interested reader may refer to Kahn who gives methods for determining the type I regions, type II regions, q_k, and n_k. (9)

Russian Roulette and Splitting has been used with success on nuclear diffusion problems. For example it is desired to calculate the probability that a particle will pass through a slab. The particle starts at one side of the slab and has collisions which send it backwards, absorb it, or transmit it. The particles of interest are the ones which are transmitted, and only these particles are considered.

d. Use of Expected Values

Occasionally, the values of $\bar{Z}_k$ are known (or can be evaluated analytically without great difficulty), but the values of g(k) are not known. It is senseless to use Monte Carlo on the part which is analytically tractable. Sample X_1 to determine the region, and use the value Z_k as the sample estimate: $\bar{Z}_d = (1/N)\sum g(k)\bar{Z}_{ki}$, where $\bar{Z}_{ki}$ is the analytically calculated expected value of Z in S_k chosen on the i^{th} sample. The variance is given by $V_d = (1/N\sum g(k)(\bar{Z}_k - \bar{Z})^2$. The reduction of variance is $V_a - V_d = \sum g(k)v_k$. "In many cases most of the variance of a problem can be eliminated by using only a little bit of analysis." (14c) It is obvious that if Monte Carlo is used on all of the probabilistic parts of a problem (instead of doing analytically what can

be done analytically) the variance will be larger than if expected values are used.

e. Correlation and Regression

Suppose two or more problems are to be solved. Occasionally, it will prove advantageous to solve them simultaneously. For example, we might attempt to optimize certain parameters in an equation. Instead of optimizing each one separately, it may be possible to combine certain portions of the problem, resulting in a savings of both time and money. The topic of interest might be the relationship between the problems. If we solve both at the same time, it is usually possible to determine the relationship more precisely than if the problems were attacked independently and results compared. Still another reason for using correlated sampling is that the answer to one of the problems may be known; it is simpler to add the estimate of the difference to the known quantity than to attempt the solution directly.

In the last instance, we obtain estimates of a quantity (two-dimensions) $W(X, Y, r, s)=Z(X,Y)-a(U(r, s)-\bar{U})$, where Z is the unknown quantity, and U is the known quantity with r, s as parameters. In the course of the calculation some of the random numbers used to obtain (X, Y) will also be used to obtain (r, s), so the two samples are not independent. Note that the expected value of W is $\bar{Z}$, the expected value of Z. It is shown in (14c) that the variance of W is given by: $V_e=(1/N)\ (s_1^2-2aps_1s_2\ a^2s_2^2)$ where s_1^2= variance of Z, s_2^2=

variance of U, and p=correlation coefficient of the two processes U and Z. The correlation coefficient, p, measures the amount of similarity in the two processes. Often it is possible to choose a=1. If the known problem is similar to the unknown one, s_1 and s_2 are nearly equal, and p near 1, V_e can be less than s_1^2. Indeed, it has been shown that for optimum $a=ps_1/s_2$, $V_e=(1/N)s_1^2(1-p^2)$. (9)

In general, $a=ps_1/s_2$ is unknown; hence the need for some means of estimating a. Kahn suggests dividing the sample into two parts, obtaining estimates for "a" form each part, and using these values on the other halves of the sample to obtain two values of $\overline{Z}$, which are then averaged. (14c) This may occasionally increase the variance, but it is unbiased, which at times is very important.

f. Systematic Sampling

Suppose the function g(k), that is the probability that the point $(X_1, X_2, \ldots, X_n)$ lies in the region S_k, is known. Suppose it has been previously decided that exactly T samples will be picked. Then a very convenient device is available. Simply choose samples until $t_k=g(k)T$ samples have been picked form the k^{th} region S_k, This eliminates the variance which arises in the choice of regions. Little extra work is needed to employ this device. In general, it will not bring great improvements in efficiency. Kahn estimates the improvement between five and thirty percent, depending on the problem. (9)

The variance associated with this method is given by:

$V_f=(1/N)\sum v_k$.

The improvement over straightforward sampling is:

$$v_a - V_f = (1/N) \sum g(k) \ (\bar{z}_k - \bar{z})^2.$$

Thus the variance is decreased by the variance of the average values of Z_k over the M region.

g. Stratified Sampling (Quota Sampling)

This method of sampling is similar to systematic sampling. In each region S_k a specified number of samples is chosen. In this scheme, instead of choosing $g(k)T = t_k$ samples in the region S_k, t_k^* samples are selected, where t_k^* is chosen to make the variance of the estimate a minimum. It has been shown that this number $t_k^* = Tg(k)\sqrt{v_k}/\sqrt{\bar{v}_k}$. (9) In this case $V_g = (1/T)\bar{v}_k$.

The reduction is $V_a - V_g = (1/T)(\sum g(k)(\sqrt{v_k} - \sqrt{\bar{v}_k})^2 + \sum g(k) \cdot (\bar{z}_k - \bar{z})^2)$. Not only has the variance due to the variation in choosing samples from the M regions been eliminated, but a variance due to the variation in v_k has also been eliminated. This is more accurate (i.e., the variance is less) than Importance Sampling.

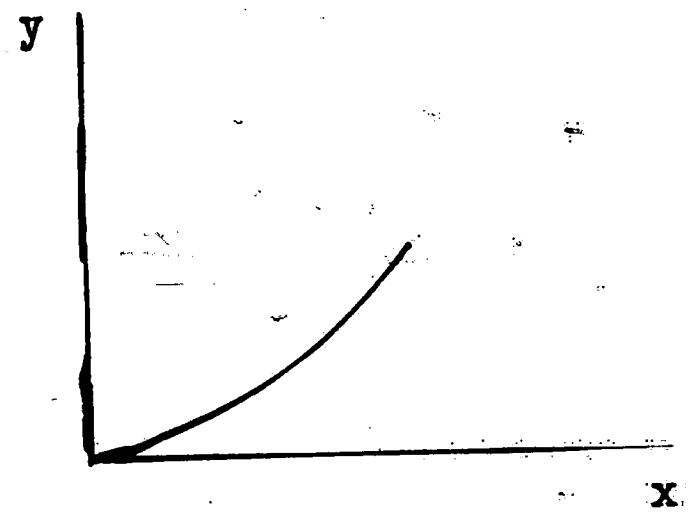
If the values of $g(k)$ are not known in advance, but approximations to the t_k^* are known, points may be chosen at random and each point identified with the region in which it lies. If t_k^* points have already been chosen, disregard that trial and choose a new point. If t_k^* points have not been chosen, evaluate $Z(X_1, X_2, \ldots)$ and record it. This process is of value only if the cost of picking $(X_1, X_2, \ldots)$ is negligible as compared with the cost of evaluating Z. If the cost cannot be neglected with respect

to the evaluation of Z, it may be decided to ignore a point, depending on the relative costs of choosing $(X_1, X_2, \ldots)$ and evaluating Z.

Finally, it is of interest to consider the following statement by Herman Kahn (14c):

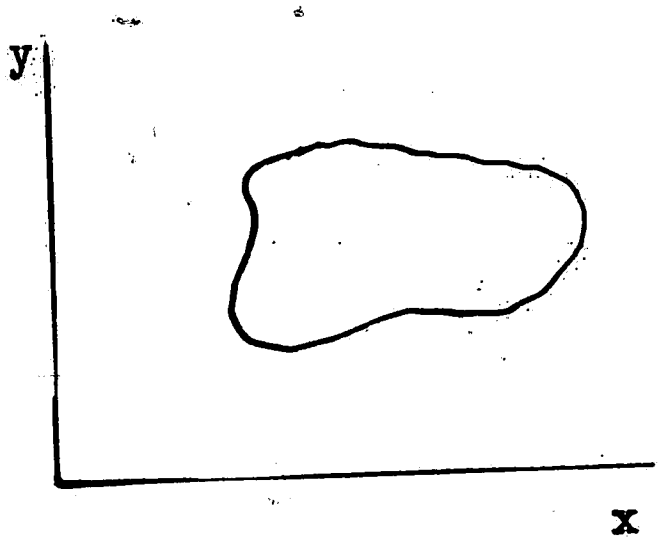
> It is probably clear to the reader that the problems faced by the Monte Carlo experimenter in trying to cut down his statistical fluctuations are quite similar to those that are faced in almost any application of sampling. Therfore, much of the literature of statistics is relevant to the problems we have been considering. . . It is valuable to have professional statistical help in designing these calculations. However, if one has to choose between a person who is mainly interested in statistics and one who is mainly interested in the problem itself, experience has shown, in this field at least, the latter is preferable. This last remark is not intended as a slur on statiscians, but simply to amplify a comment made earlier, that, 'the greatest gains in variance reduction are often made by exploiting specific details of the problem, rather than by routine application of general principles.'

fig. 1



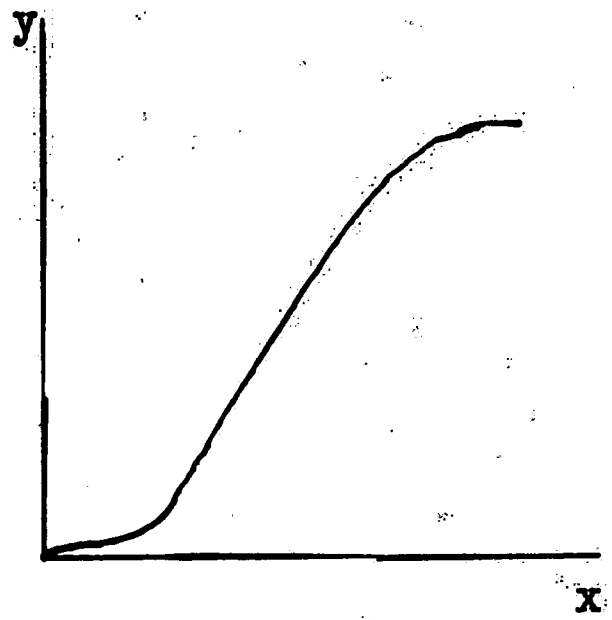
$y=x^2$

fig. 2.



$x=\phi(t) \quad y=\psi(t)$

fig. 3



$y = P(R \leq x)$

REFERENCES

1. Cashwell, E. D., and Everett, C. J. A Practical Manual on the Monte Carlo Method for Random Walk Problems Pergamon Press, N.Y., 1959

2. Curtiss, J. H. "Sampling Methods Applied to Differential and Difference Equations" Proceedings, Seminar on Scientific Computation IBM Corp. N.Y., 1949

3. Donsker, M. D., and Kac, M. "A Sampling Method for Determining the Lowest Eigenvalue and the Principal Eigenfunction of Schroedinger's Equation" J. Res. of Nat. Bur. Standards 44:551-557, 1950

4. Edmundson, H. P. "Monte Carlo Matrix Inversion and Recurrent Events" MTAC 7, 41:18-21, 1953

5. Firstman, S. I. "Reliability Estimation by Monte Carlo Simulation" Master's Thesis, UCLA, Los Angeles, 1958

6. Forsythe, G. E., and Leibler, R. A. "Matrix Inversion by a Monte Carlo Method" MTAC 4, 31:127-129, 1950

7. Gruenberger, F., and Mark, A. M. "The d^2 test of Random Digits" MTAC 5, 34:109-110, 1951

8. Hammersley, J.M. "Monte Carlo Methods for Solving Multivariable Problems" Annals, N.Y. Acad. of Sci., 1960

9. Kahn, H. Applications of Monte Carlo RM-1237-AEC RAND Corp., Santa Monica, 1956

10. Kendall, M. G., and Smith, B. B. "Randomness and Random Sampling Numbers" J. Roy. Stat. Soc. 101: 147-166, 1938

11. Kendall, M. G., and Smith, B. B. Random Sampling Numbers Cambridge University Press, Tracts for Computers XXIV, 1951

12. Mantel, N. "An Extension of the Buffon Needle Problem" Ann. Math. Stat. 24:674-677, 1953

13. Metropolis, N., and Ulam, S. "The Monte Carlo Method" J. Amer. Stat. Assoc. 44:335-341, 1949

14. Meyer, H. A., ed. Symposium on Monte Carlo Methods J. Wiley and Sons, N.Y., 1956. The following papers appeared in this work and are referenced as follows:
    14a) Marshall, A. W. "An Introductory Note"
    14b) Taussky, O., and Todd, J. "Generation of Pseudo-Random Numbers"
    14c) Kahn, H. "Use of Different Monte Carlo Sampling Techniques"
    14d) Dismuke, N. "Monte Carlo Computations"

15. Monro, S. Personal Communication Dec. 28, 1960

16. National Bureau of Standards, AMS 12 Monte Carlo Method U. S. Government Printing Office, Washington, 1951. The following papers appeared in this work and are referenced as follows:
    15a) Householder, A. S. Foreword
    15b) Brown, G. W. "History of RAND's Random Digits"
    15c) Hammer, P. C. "The Mid-Square Method of Generating Digits"
    15d) Forsythe, G. E. "Generation and Testing of Random Digits at the National Bureau of Standards, Los Angeles"
    15e) Von Neumann, J. "Various Techniques Used in Connection with Random Digits"

17. RAND Corp. A Million Random Digits with 100,000 Normal Deviates The Free Press, Glencoe, Illinois, 1955

18. Saaty, T. L. Mathematical Methods of Operations Research McGraw-Hill, N.Y., 1959

19. Tippett, L. H. C. Random Sampling Numbers Cambridge University Press, Tracts for Computers XV, 1950

20. Todd, J. "Experiments in the Solution of Differential Equations by Monte Carlo Methods" J. Wash. Acad. Sci. 44; 12:377-381, 1954

21. Ulam, S. "On the Monte Carlo Method" Proceedings of a Second Symposium on Large Scale Digital Calculating Machinery pp. 207-212

22. Wasow, W. "Random Walks and the Eigenvalues of Elliptic Difference Equations" J. Res. Nat. Bur. Stds. 46:65-73, 1951

23. Yowell, E. C. "A Monte Carlo Method of Solving Laplace's Equation" Proceedings, Computation Seminar IBM Corp. N.Y., 1949

Peter Anthony Lachenbruch was born on February 5, 1937 in Los Angeles, California. He is the son of Mr. and Mrs. Jerome Lachenbruch of Los Angeles, California.

His secondary education was completed in Los Angeles where he graduated from University High School in 1954. He then attended Occidental College in Los Angeles for one year where he participated in two sports. In 1955, Mr. Lachenbruch entered the University of California at Los Angeles (UCLA) as a sophomore. He graduated from that institution in 1958. In the fall of that year he accepted a position at Lehigh University as a graduate assistant in Mathematics. During the period from June 1959 to September 1960 Mr. Lachenbruch was employed by the Douglas Aircraft Company as a computer analyst and the System Development Corporation as a Senior Operations Research Analyst. At the present time he is on leave of absence from the System Development Corporation to complete his Master's Degree studies.