MONTE CARLO METHODS

by

THOMAS MICHAEL KEEGAN
B. S., Kansas State University, 1962

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1963

Approved by:

A. Thomas Parker
Major Professor

TABLE OF CONTENTS

INTRODUCTION

Essentially the term "Monte Carlo Method" refers to any computational technique which makes use of stochastic processes. Situations which lend themselves to such techniques tend to have a probabilistic background. However, the solution of many problems of a physical nature has been found to be easily adaptable to this approach. In most cases where a Monte Carlo method is used, there are analytic techniques which could give the solution. The reason for not using the analytic technique is usually because of the cumbersomeness of this more standard method for the particular case at hand.

In this paper many references are made to random numbers, randomness, and similar concepts. From a practical standpoint, there is no such thing as a random number. That is, a sequence of numbers, although referred to as random, will only appear to have random characteristics. By definition, a sequence of numbers is random if each number in the sequence is statistically independent from every other number in the sequence, and all possible distinct values in the domain of the sequence are distributed identically. To get a sequence which satisfies this definition is impossible, from a practical standpoint. This makes it necessary to devise techniques which give random-appearing sequences, so-called "pseudo-random" sequences.

In Section II of this paper, some methods for developing such sequences are given and compared. Methods of testing sequences are also given. The term random number, as used in this paper, refers to a generated number from the uniform or rectangular distribution.

Lehmer is quoted in [3] as describing a pseudo-random sequence as "a vague notion embodying the idea of a sequence in which each term is unpredictable to the uninitiated and whose digits pass a certain number of tests, traditional to statisticians and depending somewhat on the uses to which the sequence is to be put." This is about as concise a definition as one can find in the literature on this topic.

Of the various methods for generation of random sequences, the most interesting are those in the section called Ergodic Theorems. The name ergodic has been applied "to those mechanical systems which have the property that each particular motion, when continued indefinitely, passes through every configuration and state of motion of the system which is compatible with the value of the total energy", page 2182 in [9] . This idea is not unlike the idea of a random sequence, which is probably why Weyl used the term, in [3].

Section III, on sampling, contains methods of transforming the uniform distribution into some other distribution which may be of use in a specialized instance. Other distributions may be needed often in quadrature applications, random walks, simulation situations, sampling studies, etc. Several general methods are listed in addition to a few special cases which are intended to indicate a possible approach to be taken when an obscure distribution occurs in practice.

Quadrature techniques are discussed in Section IV. This is the classical application of Monte Carlo methods dating back to Buffon and others, page 2381 in [9]. The solution of partial differential equations, discussed in Section V, is in one sense a type of quadrature. The approach given introduces the ideas of random walks. A random walk

shall be defined as the path described by a particle when moving between two points of a grid in a finite number of steps. Random walks are also used in connection with simulation of small particle transmission, which has widespread usage in the area of nuclear physics.

Again it must be mentioned that not all the applications of Monte Carlo methods are covered in this paper. The only applications which are included are sampling, quadrature, and random walks. Actually these are quite inter-related in that the distributions listed in the section on sampling could be of interest in the uses of Monte Carlo quadrature to suggest factorizations of integrands and when a random walk happens to have probabilities from some non-uniform distribution.

II. Generation of Random Sequences

A. Methods of Generation

1. Electronic noise machines

These devices are, figuratively speaking, electronic roulette wheels. They are probably the best possible devices for generating random numbers. Their main drawbacks are their cost and lack of adaptability.

The Rand tables [11] were generated by an electronic device by passing random frequency pulses through a binary counter. The resulting binary numbers were converted to decimal numbers and punched out on Hollerith cards. The table was then transformed by adding corresponding digits on adjacent punched cards modulo ten. This transformation improved the uniformity of the distribution of the digits. The original production of digits was tested statistically and the machine refined until the numbers it produced were considered satisfactory.

Another device of a similar nature is ERNIE [3]. ERNIE is a random digit generator used by the British Post Office for the sole purpose of determining numbers for a lottery. It is not part of a computer, however, and only produces digits at a rate of about fifty per second.

The maintenance of these devices seems to be an important factor in their adoption. The machine used for the Rand tables was said to "run down" after repeated usage, that is, give statistically biased numbers even though periodic electronic checks seemed to indicate that it was in order. This fact indicates that continual tests must be made on the numbers and that electronic adjustments must be effected when necessary.

Since the development of efficient and reliable arithmetic generators, the use of electronic generators has lapsed. The present arithmetic generators are at least sufficient for most purposes, but it is quite doubtful that they can ever achieve the random qualities of the electronic generators.

2. Ergodic theorems

Work has been done examining the randomness of the digits in the decimal expansions of π and e and various functions of them [3]. For example, Weyl in 1916 proved many ergodic theorems when he showed that the fractions $\pi n^2 - [\pi n^2]$, for n = 1, 2, 3, ... , are uniformly distributed between zero and one (where $[x]$ is the so-called "bracket function", the largest integer contained in x). Some work has been published in this area by others. The obvious drawback to these procedures is the cumbersome computational aspect.

3. Congruential methods

Congruential methods are arithmetical procedures which are used to manufacture a sequence of numbers which appear random. Obviously these procedures are completely deterministic, yet they produce numbers which can be used as though they were in fact truly random. The main consideration when developing these methods are periodicity, the occurrence of runs, and possible degeneration of the sequence. For purposes of this paper a run is the distance between the successive occurrences of the same value in the sequence.

a. Multiplicative congruential methods

Multiplicative congruential methods of generation use a form as follows: $u_{n+1} = u_n X + C \pmod m$ where X is a constant multiplier, m is the size of the word (an integral power of 10), u_n is the desired random number from a rectangular distribution such that $0 \leq u_n < 1$, and C is an additive constant. As an example, if one picks the following starting values: $m = 100$, $X = 11$, $C = 7$, and $u_o = 13$; one would have: $u_{n+1} = u_n \cdot 11 + 7 \pmod{100}$. So the first few terms of the sequence would be:: $u_o = 13$, $u_1 = 50$, $u_2 = 57$, $u_3 = 34$ . . . . The special case where $C = 0$ gives rise to the power residue method of generation.

The primary concern of this section is to develop the characteristics of these methods from the standpoint of number theory. By judicious choice of parameters, the sequences which are produced have a "more random appearance". From a statistical investigation of the produced sequence it turns out that as long as the sequences do not repeat too readily, the results are satisfactorily random appearing. It is easily

seen that the sequence will repeat at least after m numbers, as there will be only m possible values of u_n (mod m). For the power residue method the sequence repeats before the full period, so additional restrictions must be placed upon the parameters.

For the case of C not zero, a theorem giving conditions on the parameters to insure a full period is given in [3]. The conditions are: 1.) C must be relatively prime to m, thus C cannot be a multiple of 2 or 5 as m is a power of 10. 2.) $X \equiv 1 \pmod{p}$, where p is a prime factor of m; or $X \equiv 1 \pmod{4}$ if four is a factor of m.

For the case of C = 0, conditions are given in [4] for the choice of parameters which maximize the period. The starting values for u_o must not be a multiple of 2 or 5 and must be k digits in length, where $m = 10^k$. For $k \geq 3$ these conditions will yield $5 \cdot m \cdot 10^{-2}$ terms before repeating. The constant multiplier should be of the form $x = 200\ t \pm r$, where t is any integer and plus or minus r represents an element from one of the 32 residue classes modulo 200. The values for r are as follows: 3, 11, 13, 19, 21, 27, 29, 37, 53, 59, 61, 67, 69, 77, 83, 91. It has usually been found to be best to choose an X whose value is close to $\sqrt{m}$.

Once having a sequence of u_n, a sequence of d-digit random numbers from the rectangular distribution over [0,1) is obtained by discarding the lower order digits of each term of the sequence of u_n. These remaining higher order digits will have the greatest period of any of the digits in each term.

From this discussion it is seen that the same period can be obtained by using the power residue method as from the regular multiplicative method only at the expense of using a larger value of m. If one were generating this sequence on a digital computer with fixed word length, this could be an important consideration.

The primary difference in the properties of the sequences obtained from the two given methods is the statistical behavior of the terms. Both methods are generally quite good; however, the regular multiplicative method seems to be completely unacceptable for some particular choices of parameters. There are only a very few cases of this [3], but there are no such exceptional cases for the power residue method.

b. Mid-square method

The mid-square method of generating a sequence of random rectangular deviates was developed by von Neumann and Metropolis about 1946. These men were the first to consider the idea of developing an arithmetic generator, [3]; [4]; articles by Taussky, Todd, and Metropolis in [6]; articles by Brown, Forsythe, and Hammer in [8]. The procedure is to take for a starting value an m-digit number, square it, and take for the next term in the sequence, the middle m digits of the 2m-digit product.

This method is very difficult to analyze and has not always produced satisfactory results. The period must be found by trial using the particular starting value. The procedure generally does not return to the starting value to repeat. This method of calculation is a trifle cumbersome to use and thus computing is relatively slow. This procedure was used successfully at Los Alamos for many years to obtain sequences of about 750,000 terms [3].

c. Additive

The additive congruential methods of generation, often referred to as Fibonacci series, page 718 in [9], have as a general form: $u_{n+1} = u_n + u_{n-1} \pmod m$. The use of this procedure is much quicker for computational purposes and has a long period but is statistically unsound. It is essentially a power residue method with the constant multiplier, X, taken to be equal to $(1 + \sqrt{5})/2$, which is too small for good results [4]. Many variations of this procedure have been developed, but they cause the use of this procedure to lose its advantage of speed of calculation. Many other types of additive generators have been developed and considered in recent years [3].

B. Tests for Randomness

There are two basic classifications of statistical tests for randomness: runs and frequency. Run tests are primarily concerned with the independence of the successive numbers. Frequency tests determine whether the numbers actually fit the proper distribution. All of these tests assume the sequences are independent and uniformly distributed in order to arrive at expected frequencies.

Frequency tests are the easiest to use. One generates a series of large blocks of digits and counts the number of occurrences of each digit. Then one applies a chi-square "goodness-of-fit" test [3] on the block of digits to get a probability statement on the likeliness of such a block. It is advisable to apply these tests on blocks which are not too large and then combine blocks and apply the tests on the

larger blocks. In this manner one would have a measure of the over-all periodicity. One also could check the frequency of odd and even digits, which are assumed to be binomially distributed.

Run tests include poker tests and serial tests. Poker tests consider sets of five digits and consider the probable occurrence of various arrangements of digits in the particular chosen sets. As examples of such five digit sets to consider: busts, abcde; pairs, aabcd; two pairs, aabbc; three of a kind, aaabc; full house, aaabb; four of a kind, aaaab; and five of a kind, aaaaa. The Rand Corporation [11] applied this technique to blocks of 1000 poker hands by calculating the expected frequency and running a chi-square "goodness-of-fit" test against the actual observed frequencies. As with the frequency tests, the poker tests can be applied to a seemingly endless variety of blocks of digits.

Serial tests investigate the order of the digits in the sequences generated. For example, one could check the frequencies with which each digit followed a certain digit and test whether the orderings were all equally likely. Again a chi-square "goodness-of-fit" test would apply. Also one could count the number of runs of various lengths which occurred. A certain number of runs would be expected to occur and this could be compared to how many actually occurred.

Another approach to testing relationships among the elements of the sequence is to use the serial correlation coefficient [3]. A measure of serial correlation is essentially a correlation coefficient calculated between two sequences of numbers, where the i th element in the first sequence is the $(i+k)$ th element in the second sequence. The first sequence is then said to have a lag of k. The serial correlation

coefficient will therefore measure the relationship between digits which are k terms apart. This coefficient is especially useful when using deterministic generators, as the independence of the elements implies a correlation zero. It will, therefore, be a measure of how "good" the starting values of the parameters are, where "good" refers to the appearance of the generated sequence with respect to randomness.

This by no means exhausts the possible list of tests for randomness. If the numbers are to be used for a specific purpose, some tests may be more meaningful than others. This situation has caused the designation of "pseudo-random sequence" by Lehmer which was referred to in the introduction of this paper.

III. Sampling Techniques

A. General Cases

1. Table look-up

The most straight-forward and the easiest approach to arriving at a random-sampling distribution is to prepare a table of values and randomly pick a set of values from the table. To set up the table, first list the values of F(X), the cumulative distribution function under consideration, taken at a constant increment of F(X) where the range of F(X) has been standardized to the unit interval. The table will then be taken to contain the values of X corresponding to the values of F(X) in the list. The degree of precision for this method is determined by the value of k, where there are 10^k items in the table.

It is thus most convenient to use 10^{-k} for the increment on F(X) and then interpret a k-digit uniform deviate as the location in the table of a particular randomly drawn value of X. By taking advantage of symmetry and other peculiarities of the distribution under study, one can decrease the size of the table necessary for the sampling. In many instances this reduction is an important consideration.

The use of this method, in reference to applications on digital computers, is usually faster than other methods of sampling. That is to say, it would take less computer time per random value of X than most other methods.

The extension of the use of this method to three or more dimensions requires too many values for a reasonable table. This is due to the fact that the number of tabled values increases exponentially as the number of dimensions increases linearly.

2. Modified table look-up

Marsaglia [5] recommends a modification of the table look-up method given in the preceding section. The use of this modification is generally slower per average variate but requires a much smaller table. The method will be described by giving an example so as to alleviate some notational problems. This example has been taken directly from [5] with supplementary remarks inserted for added clarity.

It is assumed that a source of uniform deviates [0,1) is available. Suppose one wishes to sample from the distribution of a discrete variable whose distribution is as follows:

Value of X	Probability of such an X
a	.023
b	.038
c	.074
d	.103
e	.148
f	.206
g	.140
h	.101
i	.093
j	.037
k	.026
m	.011

Consider setting up a table of values of X as follows:

n	X	n	X	n	X
0	d	30	i	60	e
1	e	31	i	61	e
2	f	32	i	62	e
3	f	33	i	63	e
4	g	34	i	64	e
5	h	35	j	65	e
6	a	36	j	66	e
7	a	37	j	67	f
8	b	38	k	68	f
9	b	39	k	69	f
10	b	40	m	70	f
11	c	41	a	71	f
12	c	42	a	72	f
13	c	43	a	73	h
14	c	44	b	74	i
15	c	45	b	75	i
16	c	46	b	76	i
17	c	47	b	77	j
18	e	48	b	78	j
19	e	49	b	79	j
20	e	50	b	80	j
21	e	51	b	81	j
22	g	52	c	82	j
23	g	53	c	83	j
24	g	54	c	84	k
25	g	55	c	85	k
26	i	56	d	86	k
27	i	57	d	87	k
28	i	58	d	88	k
29	i	59	e	89	k
				90	m

Note that the first six X's in the table are the values which have a probability of occurrence greater than .099. In other words, if the given probabilities are truncated after the first decimal digit there results:

Value of X	Probability of such an X
d	.1
e	.1
f	.2
g	.1
h	.1

These correspond to the first six values in the table.

In a similar manner for the next thirty-five items in the table there are values of X which would occur if one considered only the second digits of the given probabilities as probabilities of occurrence. These values of X are placed in the table with the frequency associated with their respective second digits. Likewise, the last fifty items in the table are the values of X associated with the frequency of the third digit of the given probabilities.

This procedure, once the table is properly set up, requires one to acquire a three-digit uniform deviate, called $d_1d_2d_3$. If $d_1 < 6$, pick the value of X in location $n = d_1$. If $60 \leq d_1d_2 < 95$, pick the value of X in location $n = d_1d_2-(60-6)$. If $950 \leq d_1d_2d_3$, pick the value of X in location $n = d_1d_2d_3-(950-35-6)$. Note that six is the sum of the first digits in the given probabilities and thirty-five is the sum of the second digits in the given probabilities.

Thus the sum of all the individual digits in the given probabilities is the number of values needed for the table, and is not one thousand, as would be the case when using the method of the preceding section.

The use of this method, although using far fewer values in the table, gives rise to a slower procedure than the previous method.

It should be remarked that this method can be used for some continuous distributions if one can be satisfied with using discrete points for the values of X. This is usually the case in any applied situation anyway.

3. Inverse

In a few instances it is possible to determine analytically the actual inverse of the distribution function, as an example $Y = F(X) = X^n$, $X = \sqrt[n]{Y}$. The sampling problem is then reduced to the evaluation of a given function for points randomly drawn from the domain of definition.

Too often when the inverse can be found, it is a function which is not in a form which can be conveniently evaluated. It may then be easier to use one of the given general methods or to try to find a special technique to fit the situation.

4. Distribution function as an integral

Another general method which is of interest makes use of the fact that the distribution function, $Y = F(X)$, is the integral of the density function, $f(X)$. It requires the use of two uniform deviates, one for the X axis and one for the Y axis. The X deviate is taken uniformly from the interval between the limits of definition of the function. The Y deviate is taken uniformly from the interval from zero to the upper bound on $f(X)$.

The procedure is as follows: pick an X and a Y; compute f(X); if $f(X) \leq Y$, accept X; if $f(X) > Y$, reject X and pick a new pair of uniform deviates and repeat the process. The accepted values of X will have the desired distribution [1].

This procedure is readily extended to a multi-dimensional situation. This is done simply by picking a uniform deviate for each dimension, thus giving a set of several X's and one Y.

The main disadvantage of this method is that it "wastes" or does not use so many of the required uniform deviates. Usually this is not of much significance as uniform deviates are readily available.

B. Special Cases

1. Normal Distribution

a. The central limit theorem

One of the simplest techniques for generating normal deviates makes use of The Central Limit Theorem from probability theory, page 257 in [13]. This theorem states that the distribution of the mean of independent deviates from any distribution will approach the normal distribution as the sample size becomes large. Thus using $Y = (S - \frac{1}{2}) \sqrt{12n}$, where S is the mean of n uniform deviates, one can say that the distribution of Y will approximate a normal distribution with a mean of zero and a variance of one.

In regard to the actual distribution of Y, the mean is zero and the variance is one. Unfortunately the higher even moments of Y are dependent upon n. The central moments for the normal are $\mu_k = 0$ for $k = 1, 3, 5, 7, \ldots$ and $\mu_{2k} = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2k-1)$ for $k = 1, 2, 3, \ldots$.

The higher odd central moments of Y are zero. The fourth central moment of Y is $3 - 6/5n$ and the sixth central moment of Y is $15 + \frac{48}{7n^2} - \frac{18}{n}$. The other higher even moments become increasingly dependent upon n. This factor results in the need for n to be larger for a satisfactory approximation.

For studies which may be sensitive with regard to higher moments, this method is not desirable. The use of this approach will, therefore, depend upon the problem.

In [4] the authors imply that n = 6 will be completely satisfactory. However, Muller [7] discourages the use of this method altogether. Its primary advantage is its ease of computation.

Muller [7] shows that for values within plus or minus three standard deviations of the mean, the approximated normal deviates are fairly good. However for values beyond these limits, the approximation is extremely poor.

b. Hasting's approximation

Hasting's [2] has developed a rational approximation for random normal deviates which is very reliable for all values of these deviates. However, the use of this method is not as fast as the use of most other methods. The procedure is as follows: pick values of u which are uniformly distributed on the interval $(0,.5]$, compute $r = -2 \ln u$, evaluate

$$X = r - \frac{a_o + a_1 r + a_2 r^2}{1 + b_1 r + b_2 r^2 + b_2 r^3}$$

where $a_0 = 2.515\ 517$ $\quad b_1 = 1.432\ 788$

$a_1 = 0.802\ 853$ $\quad b_2 = 0.189\ 269$

$a_2 = 0.010\ 328$ $\quad b_3 = 0.001\ 308$

By randomly assigning a plus or minus sign to the computed X's, one will get approximately normally distributed values.

Muller [7] reports that this method is exceptionally reliable but much too cumbersome for ease of computation.

c. Teichroew method

An approximation method for the normal distribution was developed by Teichroew for his doctoral dissertation (see [7] and [12]). His method may be called "an approximation by curve fitting."

The procedure is as follows: find a value $Y = m(S)$ such that $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y} e^{-t^2/2}\, dt = \int_{0}^{S} f_j(t)\, dt$, where S is a sum of j uniform deviates. Now $f_j(S)$ is the density function of S. One must require that $S_L \leq S \leq S_u$. These bounds are determined by the accuracy required by the problem. An interpolating polynomial is to be determined to approximate $Y = m(S)$. A Chebycheff polynomial of degree (k-1) is fitted so that it coincides with m(S) at the points whose abscissas are the k zeros of the Chebycheff polynomial of degree k. If j is chosen to be equal to twelve, for example, this method gives good results. Since $0 \leq S \leq 12$, one picks $S_L = 2$ and $S_u = 10$. This will exclude only 0.002 percent of the desired distribution and will simplify the approximation greatly. On setting $r = (S-6)/4$, there results $X = \sum_{j=0}^{9} d_{2j+1}\, T_{2j+1}(r)$, where $T_i(r)$ is the i th Chebycheff polynomial.

The d_i are the coefficients which arise in the approximation. In practice this series can be truncated at $T_9(r)$ due to the negligible coefficients of the higher order terms. This simplified to:

$$X = a_1 r + a_3 r^3 + a_5 r^5 + a_7 r^7 + a_9 r^9 ,$$

where $a_1 = 3.949\ 846\ 138$

$a_3 = 0.252\ 408\ 784$

$a_5 = 0.076\ 542\ 912$

$a_7 = 0.008\ 355\ 968$

$a_9 = 0.029\ 899\ 776$

Thus this polynomial transforms the set of $r = (S-6)/4$, where S is the sum of twelve uniform deviates, into a set of X which are approximately normally distributed with a mean of zero and a variance of one.

2. Chi-square

Teichroew and Sitgreaves [12] mention a method for sampling from a chi-square distribution. The procedure is to compute a set of $Y = -2 \ln \left(\prod_{i=1}^{k} X_i \right)$ where the X_i are uniform deviates from $[0,1)$. These Y will have a chi-square distribution with 2k degrees of freedom.

It is shown on page 189 of [13] that this function is an exact chi-aquare. This seems to show that if one is interested in doing some sampling from a particular distribution, it would be worthwhile to browse through a good book on mathematical statistics.

3. Distribution function of X^n

If one should want to sample from a distribution of the form, $F(X) = X^n$, one would need to compute the n th root of a uniform $[0,1)$ deviate to get the desired X value. One finds on page 236 in [13] that the given distribution function is the distribution function of the maximum value of n uniform deviates. Thus instead of computing an n th root, one only needs to generate n uniform deviates and pick the largest.

The use of this procedure is usually much quicker than computing the n th root, although one might consider the disadvantage of discarding of the (n-1) uniform deviates each time.

IV. Quadrature Techniques

Quadrature by Monte Carlo techniques goes back at least to the eighteenth century when Buffon considered his famous needle problem, page 2381 [9]. The problem is to consider a series of equidistant parallel lines and a straight wire which is shorter than the distance between the lines. If the wire is tossed "at random" upon the lines, it will either cross a line or it will lie between two lines. Buffon and Laplace proved that, under repeated trials, the fraction of trials in which a line is crossed will approach the ratio of twice the length of the wire to the circumference of a circle with diameter equal to the distance between two of the lines.

This fact was used in the Monte Carlo determination of the value of π. The method was used to achieve surprisingly good results. Actually the validity of the method was not believed until the experiment had been repeated so often that "there could never have been any doubt about it." [9].

The sampling method in 4 of III A which considered the distribution function as an integral, indicates one approach to Monte Carlo quadrature procedures. One could randomly pick values of X which were uniformly distributed over the interval of definition of X and calculate the integrand, f(X). If the calculated value of f(X) was less than a value of Y which had been randomly picked from the range of values for Y = f(X), the X would be accepted. If f(X) had been greater than Y, the value of X would have been rejected. The fraction of acceptances from the total number of trials, when multiplied by the range of Y and the interval of X, would approximate the value of the integral of f(X) over the interval of X.

This method has several definite disadvantages. One of them is the non-use of the required random numbers. The computation time may also be quite lengthy. If the function happens to have a long interval of definition or is unbounded, or has a large bound, this method will not be satisfactory.

A more desirable approach takes advantage of the statistical theory of expectations [13]. Consider the integral $\int_a^b g(X)\, f(X)\, dx$, where the integrand has been arbitrarily factored. If one obtains N values of X which have a distribution of f(X) on [a,b], an estimate of the value of the integral can be obtained by taking the arithmetic mean of the values of g(X). The error of this method is of the order of $(N^{-\frac{1}{2}})$, which means that for one more decimal place of accuracy one must use one hundred times the number of values of X, that is 100 N instead of N values.

For multiple quadrature one proceeds in the same manner. However, if there are k dimensions, then k random numbers must be picked so as to have a distribution of $f(X)$ for the calculation of each $g(X_1, X_2, \ldots, X_k)$.

In order to increase the accuracy of these procedures, one could partition the interval of X and concentrate on the more important parts. Thus if the function is "well-behaved" in some parts of the domain of definition, one would not need to do much sampling there. For sections of the domain in which the function is not as nicely behaved, one should do more extensive sampling. This technique is called importance, or stratified, sampling.

Most integrals which are easily adapted to Monte Carlo quadrature arise from a probabilistic background. It is hard to say whether this is coincidental or symptomatic, Kahn in [10]. Too often the reason for using a Monte Carlo approach to quadrature occurs when the more common techniques would require the evaluation of the function at a large number of points. For instance, with the trapezoidal rule and m intervals each in n dimensions, one must compute m^n values of the function. With very many intervals the accuracy of this approach is often surpassed by a Monte Carlo approach with the calculation of far fewer points, when n is large.

V. Partial Differential Equations

Consider a rectangular grid with coordinates (x,y) and a particle which is restricted in movement between coordinate points of the grid. That is, for each step this particle can move in any direction only between adjacent points of the grid. A random walk shall be defined as the path described by a particle when moving between two points in the grid in a finite number of steps. Thus if the direction of movement for

each step is arbitrary, the probability of movement to an adjacent position in a particular direction is one-fourth. Hence, there exist several possible paths, each equally likely.

Consider the probability that a particle will be at position (i,j) after S steps from position (0,0) and call this probability V(i,j,S). This probability must satisfy the following difference equation:

$$V(i,j,S+1) = \frac{1}{4}\left\{ V(i+1,j,S) + V(i-1,j,S) + V(i,j+1,S) + V(i,j-1,S)\right\} .$$

This is due to the fact that after S steps, the particle must have been at one of the four positions adjacent to position (i,j). This equation can be written as follows:

$$V(i,j,S+1) - V(i,j,S) = \frac{1}{4}\left\{\left[V(i+1,j,S) - 2V(i,j,S) + V(i-1,j,S)\right] + \left[V(i,j+1,S) - 2V(i,j,S) + V(i,j-1,S)\right]\right\}.$$

This form relates the first difference in probabilities with respect to time with the second differences with respect to position. Thus it is similar in form to the following partial differential equation, the heat equation: $\frac{\partial V}{\partial t} = K\left[\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2}\right]$. This similarity suggests that a limiting process could lead to a formal relationship.

For the case where $\frac{\partial V}{\partial t} = 0$, this equation becomes Laplace's equation which occurs often in boundary value problems. It is useful to be able to solve for a solution to Laplace's equation at a point interior to the given boundary rather than to find an analytic expression for the solution and evaluate it at the particular point, particularly since the latter cannot always be done.

A procedure for finding the solution at a point subject to given boundary values is as follows. Set up a rectangular grid over a region containing the boundary. Construct a sequence of random walks from the point of interest (x_o, y_o) using random numbers to determine the choice of direction. Terminate each random walk when it crosses the boundary and note the value of the function at this point on the boundary. The walks will always cross the boundary in that the probability of remaining in the region forever is zero. The average of these terminating values will be the value assigned to the function at (x_o, y_o). The accuracy of this estimate obviously depends upon the number of random walks considered. Accuracy can also be improved by diminishing the mesh size of the grid.

Although this procedure has been outlined for two dimensions only, it can be readily extended to higher dimensional spaces. This modification merely increases the complexity of the grid and increases the number of possible moves at each step of the random walk.

This procedure is adaptable to many second order partial differential equations of the elliptic type, Klahn in [10]. Consider an equation of the following form: $F(X,Y) = B_{11} \frac{\partial^2 V}{\partial X^2} + 2B_{12} \frac{\partial^2 V}{\partial X \partial Y} + B_{22} \frac{\partial^2 V}{\partial Y^2} + 2A_1 \frac{\partial V}{\partial X} + 2A_2 \frac{\partial V}{\partial Y}$. If the mesh size of the grid is h, the following differences will be used to represent the partial derivatives:

$\left[V(P_1) - V(P_o)\right] /h$ for $\frac{\partial V}{\partial X}$

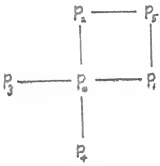
$\left[V(P_2) - V(P_o)\right] /h$ for $\frac{\partial V}{\partial Y}$

$\left[V(P_1) - 2V(P_o) + V(P_3)\right] /h^2$ for $\frac{\partial^2 V}{\partial X^2}$

$\left[V(P_5) - V(P_1) - V(P_2) + V(P_o)\right] /h^2$ for $\frac{\partial^2 V}{\partial X \partial Y}$

and $\left[V(P_2) - 2V(P_o) + V(P_4)\right] /h^2$ for $\frac{\partial^2 V}{\partial Y^2}$.

The symbol P_o designates the point (x,y). The P_i are for the points adjacent to P_o and starting at (x+h,y) and proceding counter-clockwise, and P_5 denotes (x+h, y+h); that is:



Define coefficients as:

$p_1(P) = (B_{11} - 2B_{12} + 2hA_1)/D,$

$p_2(P) = (B_{22} - 2B_{12} + 2hA_2)/D,$

$p_3(P) = B_{11}/D,$

$p_4(P) = B_{22}/D,$

$p_5(P) = 2B_{12}/D,$ and

$D = 2B_{11} + 2B_{22} - 2B_{12} + 2h(A_1 + A_2).$

These are assumed positive. These conditions of positivity are satisfied for small h, $B_{12} = 0$, and $B_{11} \cdot B_{22} > 0$. These also ensure that the equation be an elliptic partial differential equation. Independent of the B_{ij} or the A_i, the sum of the $p_i(P)$ must be identically one. The $p_i(P)$ can thus be seen to be the transition probabilities at P. The general elliptic equation, when written in difference form will be as follows: $V(P) = \sum_{i=1}^{5} p_i(P)\, V(P_i) - h^2 F(P)/D.$

As in the special case, a sequence of random walks is considered and the average of the tallies for all the walks is computed. For the general elliptic equation the calculation of each tally is more complex.

The required tally will be: $\xi_i = Q_i - \sum_j h^2 F(P_j)/D(P_j)$, where Q_i is the value of the function at the boundary points which terminated the walk, the P_j are the coordinates of each step in the random walk, and the sum is taken over all the stepping points in the walk.

The justification of this expression as a solution to the difference equation relies on the fact that there will be a finite number of steps in the walk and that all the Q_i are positive. The fact that the number of steps is finite has been previously discussed. The presumption that the Q_i are positive is not too restrictive in that some constant can be added to each value in order to meet this requirement. This will merely give a solution which will be greater than the required solution by an additive factor of this constant.

If $W_m(P)$ is taken to be the average value of the tallies which are for walks of length m, then it satisfies the following equation for each P of the grid: $W_m(P) = \sum_{i=1}^{5} p_i(P)\, W_{m-1}(P_i) - h^2 F(P)/D(P)$. The last term is added to $W_m(P)$ at every step, as $W_m(P)$ is defined as the average of the tallies and it will therefore be contributed from each P. The summation represents the transition probabilities possible at each step. Now the expression for $W_{m+1}(P)$ will be similar. Since $W_{m+1}(P) - W_m(P) = \sum_{i=1}^{5} p_i(P)\left[W_m(P_i) - W_{m-1}(P_i)\right]$, the property $W_{m+1} \geq W_m$ will always be true due to the transition probabilities. Hence the $W_m(P)$'s form a monotonic sequence with an upper bound of $R + S \cdot \bar{k}_n$. In this expression R is the maximum boundary value, S is the maximum value of $-h^2 F(P)/D(P)$ within the boundary, and $\bar{k}_n$ is the average of the number of steps in the random walks. With a bounded region of finite area, $\bar{k}_n$ is uniformly bounded, independent of n; Klahr in [10]. By taking the limit as m approaches infinity, it develops that the given expression for the tally is in fact a solution to the given difference equation.

Many variations are possible for this type of problem. Much literature is available on special cases and particular situations. Too often a special technique is more advisable to use for a given situation. Again, as in most Monte Carlo approaches, much ingenuity must be applied in order to arrive at a "best" method of attack.

ACKNOWLEDGEMENTS

The author would like to thank Dr. S. T. Parker for his invaluable help and advice.

BIBLIOGRAPHY

1. Brown, George W., article in "Modern Mathematics for Engineers", edited by E. F. Beckenbach, New York, McGraw-Hill, 1956, 279-303.

2. Hastings, Cecil, "Approximations for Digital Computers", Princeton, Princeton Univ. Press, 1955, sheets 67-68.

3. Hull, T. E., and A. R. Dobell, "Random Number Generators", Soc. Ind. Appl. Math. Review, July 1962, 4:3, 230-254.

4. International Business Machines Corporation, "Random Number Generation and Testing", Reference manual C20-8011, New York, 1961.

5. Marsaglia, G., "Generating Discrete Random Variables in a Computer", Comm. Assoc. Comp. Mach., Jan. 1963, 6:1, 37-38.

6. Meyer, H., editor "Symposium on Monte Carlo Methods", New York, Wiley, 1956.

7. Muller, Mervin E., "A Comparison of Methods for Generating Normal Deviates on Digital Computers", Assoc. Comp. Mach. Journ., 1959, 6, 376-380.

8. National Bureau of Standards, "Monte Carlo Method", Applied Math. Series 12, Washington, 1951.

9. Newman, J. R., editor "The World of Mathematics", New York, Simon and Schuster, 1956.

10. Ralston, A., and H. Wilf, editors "Mathematical Methods for Digital Computers", New York, Wiley, 1960.

11. Rand Corporation, "A Million Random Digits with 100,000 Normal Deviates", Glencoe, Free Press, 1955.

12. Teichroew, D., and Sitgreaves, article in "Studies in Item Analysis and Prediction", edited by H. Solomon, Stanford, Stanford Univ. Press, 1961.

13. Wilks, S. S., "Mathematical Statistics", New York, Wiley, 1962.

MONTE CARLO METHODS

by

THOMAS MICHAEL KEEGAN

B. S., Kansas State University, 1962

ABSTRACT
OF
A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1963

ABSTRACT

Monte Carlo Methods are computational procedures which use random numbers to arrive at solutions for mathematical problems. These procedures accomplish this by taking advantage of various stochastic properties of random numbers in the evaluation of functions. This paper discusses some of the requirements of a Monte Carlo technique. The most important requirement is a method for selection of random numbers from a uniform distribution. Methods of testing the randomness of sequences of numbers and methods of arriving at sequences of non-uniform random deviates are also given.

The classical mathematical application of Monte Carlo Methods is quadrature. Quadrature is discussed from the standpoint of using the random sampling distributions discussed in the previous sections. The final section discusses the solution of elliptic partial differential equations by the use of random walks.

This paper is intended to give some methods or procedures which are important in a good application of Monte Carlo techniques. If the given methods are not found to be essential in a particular application, it is hoped that they will be found to be quite helpful.