On the Monte Carlo Method

S. ULAM

Los Alamos Scientific Laboratory

The advent of fast computing machines is greeted as a way to "settle" a great many questions existing in applied mathematics and physics. It is hoped that the formulation of many problems in physical sciences is now correct, but the obtaining of solutions through methods that are perfectly well understood in principle is much too laborious and tedious to be attempted by paper and pencil. The mere number of arithmetical steps necessary to perform the procedures for numerical solution is such as to require, in some cases, years, even hundreds of years, of computer's time. The machines, especially the electronic calculators, can shorten this time very greatly, since the elementary steps, e.g., the multiplications, may take less than 1/1000 of the time required by a person, even if provided with a desk multiplier.

It seems that in addition to this program of testing and verification of existing mathematical formulations of physical problems, some other possibilities will be open. One—a rather general program—will be to use the calculations of electronic calculators for heuristic purposes. This may be equally possible in exploring new physical models and in pure mathematics itself. The latter possibility, perhaps less evident a priori, should be clear if one remembers the role of *examples* in abstract mathematics. It is sufficient to point out their importance in such parts of mathematics as geometry, topology, theory of functions, and abstract algebra. The field of combinatorial analysis, which is so hard to define just because it consists of a variety of special problems or examples not yet embraced by a simple general theory, seems the clearest case. The combinatorial problems studied (many of them problems of enumeration), drawn, as it were, from outside of mathematics, often have their origin in problems about configurations of physically existing objects and relations. In a field like this it is clear that the ability of machines to survey all the possibilities of specified arrangements will provide the material suggesting future theories.

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.

A simple example would be this: The problem consists of estimating the proportion of the 52! permutations of objects (cards) possessing a given—in practice always a complicated—property. One should then consider all these permutations, counting the number of those among them that possess this property. This would of course be impossible, even granting a continuous development of the speed in computing for the next hundred years. A way to get the proportion with good probability is to produce a "large" number of permutations, say 10,000, at random and count the proportion of the permutations possessing the given property.

It is perhaps surprising how many mathematical problems have in practice a structure logically similar to the one of this example. The evaluation of a definite integral may be thought of as a task somewhat analogous to the one of the previous example. The problem consists in finding the value, on a region S of the unit cube defined by inequalities, of the number

$$\int\int_{(S)} \cdots \int f(x_1 \cdots x_n) dx_1 \cdots dx_n,$$

f being a given function. This will be reduced to finding, say,

$$\int\int_{(R)} \cdots \int dx_1 \cdots dx_{n+1}$$

where the region R is defined by a class of inequalities

$$\varphi_1(x_1 \cdots x_{n+1}) < 0, \varphi_2(\cdots) < 0, \varphi_k(\cdots) < 0.$$

The procedure of elementary calculus will consist in counting the lattice points of a subdivision in a space of n + 1 dimensions, and ascertaining the proportion of these points that satisfy the given inequalities. The Monte Carlo procedure would be to take a large number of lattice points *at random* and examine these only. This number need not be of the order of he total of all lattice points.

The problem of estimating very "small" volumes requires special tricks. To illustrate the nature of possible devices, let us again take a more purely combinatorial problem. In a solitaire (game of cards for one), one desires to estimate the probability of a successful outcome. (We assume that skill plays no role, so that it is purely a game of chance.) In cases where the game has a very small probability of success most actual plays will end in failure and only an upper limit will be obtainable for this probability. How can one get some idea of a lower limit > 0? Suppose that one obtains, still obeying the rules of the game, in a noticeable proportion a of tries, a situation A where only, say, ten cards are left uncovered; after that, however, we meet with "failure." It might be justifiable to restore the ten cards to their positions in a *different* permutation and try from the situation A again. By examining a large number of the 10! permutations we might obtain the number β expressing the chance that starting with A we "win" B. A reasonable guess for the chance of success from the beginning without "cheating" would then be greater than $\alpha\beta$. Of course A should he really a class of positions, not a possiblc or a very special one. It seems, however, that if the playing of the whole game is decomposed into two or more stages, there will be a saving in the number of experiments compared with the number necessary to play to the end each time and beginning anew after each failure from the start, that is, a new permutation of the 52 cards.

The validity of such a procedure can be established in some cases. One has to prove independence, or estimate from above the correlation between the classes of events A and success B.

It is of course obvious that one can study "experimentally" the behavior of solutions of equations which themselves describe a random process, by using the digital computer as an analogy machine, as it were.' This experimental—that is, statistical—approach by Monte

Carlo techniques has been applied by various authors to linear partial differential equations., In the case of equations that are quadratic or of higher order in the unknown' functions and their derivatives, the obvious Monte Carlo procedure would be much more cumbersome, but may still have heuristic value. As an example, let us take a, bilinear system of two partial differential equations

$$\left.\begin{aligned}\frac{\partial u_1}{\partial t} &= \alpha_1 \Delta u_1 + \beta_1(u_2)u_1 \\ \frac{\partial u_2}{\partial t} &= \alpha_2 \Delta u_2 + \beta_2(u_1)u_2\end{aligned}\right\},$$

where u_1 and u_2 are unknown functions of coordinates x, y, 2, and t; a, and α_2 are given constants; β_1 and β_2 are given functions, for simplicity linear in u_1 and u_2 and also involving the independent variables x, y, z. One would like to know the asymptotic form of u_1 and u_2 (for large values of t). This problem may be looked upon as a straightforward generalization of the diffusion model (Fermi) of the Schrodinger equation.' It would correspond to a model of a system of two particles with the potential function for u_i replaced by the corresponding u function of the other particle. This linked system, treated then somewhat in the spirit of a field theory, is nonlinear. There will not be in general eigenfunctions—the separation into a time-independent equation will not be possible; yet for large values of the parameter t the space part of u may have an almost periodic or summable (by the first mean) behavior. A numerical approach to the study of such systems could again be a Monte Carlo procedure. One would diffuse and multiply the (fictitious) particles corresponding to u_1 and u_2 according to their numbers, instead of a given function V of coordinates. Since these numbers change in t, it will be necessary to make frequent censuses—as it were, to interrupt the calculation periodically—in order to ascertain the values to be used for "potentials."

The problem of transforming first purely formally, an equation not of a diffusion or Boltzmann type into one of the above type thus becomes of practical importance. Let us indicate some possibilities in this direction. The equation of Hamilton-Jacobi in one dimension has the form

$$\left(\frac{\partial S}{\partial x}\right)^2 = \frac{1}{v^2(x)}. \tag{1}$$

On the other hand, consider the equation

$$\frac{\partial W}{\partial t} = \frac{\partial}{\partial x}\left[v\,\frac{\partial (vW)}{\partial x}\right]. \tag{2}$$

This latter equation will describe the probability behavior of a particle starting, say, from the origin, and performing a random walk on the line, steps being equally probable to the right or to the left. However, the length of the steps in the position x is proportional to the value of $v(x)$. If we perform the passage to the limit with the length of the step tending to zero the resulting continuous process gives a distribution of position in time t obeying Eq. (2). It can be proved3 that the crest of the distribution, that is, the place x where $\partial W/\partial t = 0$, will satisfy a relation $S(x) = t^{\frac{1}{2}}$, where S is the solution of Eq. (1).

It is of course quite unnecessary to take recourse to such methods for a one-dimensional equation that is easily solved explicitly by quadratures. The example here given is meant merely to indicate the possibility of relations between two seemingly very different processes. One is a strictly deterministic one, described by the equation of geometric optics (or the equations of mechanics), the characteristic equation of Hamilton. The other is a continuous random-walk process with the length of the elementary step a given function of position. It turns out that at least in one dimension the locus of the points where the first derivative with respect to time of the probability distribution is equal to zero coincides with the locus of the points where the value of the Hamilton function $S = \sqrt{t}$. In two or more dimensions, the two loci are probably at least asymptotically equal, that is, for large values of t.

In the first examples of application of the Monte Carlo method to empirical evaluation of properties of solutions of differential equations, one studied the density of the diffusing and branching, that is, multiplying and transmuting, particles. This density as a function of the independent variables obeyed a linear partial differential equation of a parabolic or elliptic type. It is clear that for nonlinear equations one will have to examine, not this density directly, but appropriately chosen junctionals of this function.

The diffusion process can be described, of course, as a Markoff chain, and this in turn by a study of the interaction of matrices with nonnegative coefficients. Let us indicate a way to study "experimentally" the behavior of powers of matrices with arbitrary real terms. This possibility rests on the fact that real numbers can be considered as matrices, with positive terms; for example, -2 corresponds to $\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$. This correspondence obviously preserves both addition and multiplication. Any system described by any n-by-n matrix giving the transition moments as real numbers can be interpreted probabilistically by using $2n$-by-$2n$ matrices with nonnegative terms. The diffusion and branching or multiplication are performed by two kinds of particles—black and red—with the transformation rates given by the matrices above.

In having four kinds of particles one can then realize stochastic models for matrices with complex terms; more generally, with an appropriate number of kinds of particles, one can realize stochastic models for more general algebras over real numbers.*

The possibility of a statistical or probabilistic evaluation of definite integrals in n-dimensional space affords merely one example of an attempt to gain insight into a situation involving a system of n particles. Let us think here of n as having a value of the order of 10 or 20. The "appearance" of a set of points in a euclidean space of this dimension, if the set is defined as above by many inequalities, cannot of course be studied on graphs directly, or very well by projections of the set into three-dimensional component spaces. Now, in physical chemistry, for instance, the occurrence of this situation and its importance are well known. The properties of a molecule with a large number of atoms depends on characteristics of configurations of certain n-dimensional sets. The evaluation of various functionals of these configurations can, probably, be done best by a Monte Carlo procedure, that is, by testing a large number of

n-tuples, chosen at random with appropriate distribution, for the values of these given functionals.

It is rather curious that one meets with an analogous situation in pure mathematics itself. Let us describe it very briefly: a *formal* system in mathematics involve? in addition to the Boolean operations of elementary logic or set theory (the addition and intersection of sets of points), the so-called quantifiers, the two symbols $\sum_x \varphi(x)$, meaning that there exists an x for which $\varphi(x)$, a propositional function, is true, and $\prod_x \psi(x)$, meaning that $\psi(x)$ holds for all x.

A large part of the study of mathematics as a formal system involves the study of classes of sets on which one performs these operations. One knows that a "geometric" interpretation of these operators is particularly simple.[5] The existence quantifier corresponds to taking an orthogonal projection parallel to one or more axes of a given set of points in n dimensions on a space of fewer dimensions. The other quantifiers can be expressed by means of the first and the Boolean operations.

Even the simplest mathematical definitions lead to sets defined in a higher number of dimensions. The problem presents this appearance: there are given in a space of n dimensions several "primitive" sets of points. Starting with these sets, one obtains new ones by adding them, intersecting them with one another; these are the Boolean operations. One also takes projections of the sets obtained and conversely, having sets in spaces of fewer dimensions, erects cylinder sets in the full n-dimensional space.

One can in this fashion, starting from two given sets, obtain an infinity of new sets.

The mutual relations of these sets form the object of the logical or metamathematical study of the system.

It is possible that, for heuristic purposes alone, it would be useful to study these constructions on a large number of examples.

A mathematical theorem can be formulated in this language as stating that a certain set of the class obtained is vacuous. In cases where a proof would appear very difficult it might be of value to, so to say, try to construct points of it by random choices of the starting sets or values of "free variables" in the n-dimensional space. The failure to obtain any after a great number of choices would then lead to the belief that if the set is not vacuous it is small. It is clear that a proof will never be obtained in this fashion. However, the heuristic value of such a procedure might not be negligible.

REFERENCES

1. N. Metropolis and S. Ulam, "The Monte Carlo method," *J.* Am. Statist. *Assoc.* 44, 335–341 **(1949).**

2. M. D. Donsker and M. Kac, "**A** sampling method for determining the lowest eigenvalue and the principle eigenfunction of Schrodinger's equation," **J.** Research *Nat.* **Bur.** Standards 44, 551–557 **(1950).**

3. C. J. Everett and S. Ulam, "Random walk and the Hamilton-Jacobi equation," ***Bull.*** *Am. Math. Soc.* (1950), abstract (to be published).

4. C. J. Everett and S. Ulam, "On an application of a correspondence between matrices over real algebras and matrices of positive real numbers," Bull. *Am. Math. Soc.* (1950), abstract (to be published).

5. C. J. Everett and S. Ulam, "Projective algebra. I," *Am. J. Math. 68, 77–88 (1946).*