JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 247 SEPTEMBER 1949 *Volume 44*

THE MONTE CARLO METHOD

NICHOLAS METROPOLIS AND S. ULAM
Los Alamos Laboratory

We shall present here the motivation and a general description of a method dealing with a class of problems in mathematical physics. The method is, essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences.

ALREADY in the nineteenth century a sharp distinction began to appear between two different mathematical methods of treating physical phenomena. Problems involving only a few particles were studied in classical mechanics, through the study of systems of ordinary differential equations. For the description of systems with very many particles, an entirely different technique was used, namely, the method of statistical mechanics. In this latter approach, one does not concentrate on the individual particles but studies the properties of *sets of particles*. In pure mathematics an intensive study of the properties of sets of points was the subject of a new field. This is the so-called theory of sets, the basic theory of integration, and the twentieth century development of the theory of probabilities prepared the formal apparatus for the use of such models in theoretical physics, i.e., description of properties of aggregates of points rather than of individual points and their coordinates.

Soon after the development of the calculus, the mathematical apparatus of partial differential equations was used for dealing with the problems of the physics of the continuum. Hydrodynamics is the most widely known field formulated in this fashion. A little later came the treatment of the problems of heat conduction and still later the field theories, like the electromagnetic theory of Maxwell. All this is very well known. It is of course important to remember that the study of the

physics of the continuum was paralleled through "kinetic theories." These consist in approximating the continuum by very large, but finite, numbers of interacting particles.

I

When a physical problem involves an intermediate situation, i.e., a system with a moderate number of parts, neither of the two approaches is very practical. The methods of analytical mechanics do not even give a qualitative survey of the behavior of a system of three mutually attractive bodies. Obviously the statistical-mechanical approach would also be unrealistic.

An analogous situation exists in problems of combinatorial analysis and of the theory of probabilities. To calculate the probability of a successful outcome of a game of solitaire (we understand here only such games where skill plays no role) is a completely intractable task. On the other hand, the laws of large numbers and the asymptotic theorems of the theory of probabilities will not throw much light even on qualitative questions concerning such probabilities. Obviously the practical procedure is to produce a large number of examples of any given game and then to examine the relative proportion of successes. The "solitaire" is meant here merely as an illustration for the whole class of combinatorial problems occurring in both pure mathematics and the applied sciences. We can see at once that the estimate will never be confined within given limits with certainty, but only—if the number of trials is great—with great probability. Even to establish this much we must have recourse to the laws of large numbers and other results of the theory of probabilities.

Another case illustrating this situation is as follows: Consider the problem of evaluating the volume of a region in, say, a twenty-dimensional space. The region is defined by a set of inequalities

$$f_1(x_1, x_2 \cdots x_{20}) < 0; f_2(x_1, x_2 \cdots x_{20}) < 0; \cdots f_{20}(x_1, x_2 \cdots x_{20}) < 0.$$

This means that we consider all points $(x_1, x_2, x_3, \cdots x_{20})$ satisfying the given inequalities. Suppose further that we know that the region is located in the unit cube and we know that its volume is not vanishingly small in general. The multiple integrals will be hardly evaluable. The procedure based on the definition of a volume or the definition of an integral, i.e., the subdivision of the whole unit cube, for example, each coordinate x_1 into ten parts, leads to an examination of 10^{20} lattice points in the unit cube. It is obviously impossible to count all of them. Here again the more sensible approach would be to take, say 10^4 points

at random from this ensemble and examine those only; i.e., we should count how many of the selected points satisfy all the given inequalities. It follows from simple application of ergodic theorems that the estimate should be, *with great probability*, valid within a few per cent.

As another illustration, certain problems in the study of cosmic rays are of the following form. An incoming particle with great energy entering the atmosphere starts a whole chain of nuclear events. New particles are produced from the target nuclei, these in turn produce new reactions. This cascade process continues with more and more particles created until the available individual energies become too small to produce further nuclear events. The particles in question are protons, neutrons, electrons, gamma rays and mesons. The probability of producing a given particle with a given energy in any given collision is dependent on the energy of the incoming particle. A further complication is that there is a probability distribution for the direction of motions. Mathematically, this complicated process is an illustration of a so-called Markoff chain. The mathematical tool for the study of such chains is matrix theory. It is obvious that in order to obtain a mathematical analysis, one would have to multiply a large number of $(n \times n)$ matrices, where n is quite great.

Here again one might try to perform a finite number of "experiments" and obtain a class or sample of possible genealogies. These experiments will of course be performed not with any physical apparatus, but theoretically. If we assume that the probability of each possible event is given, we can then play a great number of games of chance, with chances corresponding to the assumed probability distributions. In this fashion one can study empirically the asymptotic properties of powers of matrices with positive coefficients, interpreted as transition probabilities.

II

Finally let us consider more generally the group of problems which gave rise to the development of the method to which this article is devoted. Imagine that we have a medium in which a nuclear particle is introduced, capable of producing other nuclear particles with a distribution of energy and direction of motion. Assume for simplicity that all particles are of the same nature. Their procreative powers depend, however, on their position in the medium and on their energy. The problem of the behavior of such a system is formulated by a set of integro-differential equations. Such equations are known in the kinetic theory of gases as the Boltzmann equations. In the theory of probabilities one

has somewhat similar situations described by the Fokker-Planck equations. A very simplified version of such a problem would lead to the equation:

$$\frac{\partial u(x, y, z)}{\partial t} = a(x, y, z)\Delta u + b(x, y, z)u(x, y, z) \qquad (1)$$

where $u(x, y, z)$ represents the density of the particles at the point (x, y, z). The Laplacian term, $a\Delta u$ on the right hand side corresponds to the diffusion of the particles, and bu to the particle procreation, or multiplication. [In reality, the equation describing the physical situation stated above is much more complicated. It involves more independent variables, inasmuch as one is interested in the density $w(x, y, z; v_x, v_y, v_z)$ of particles in phase space, v being the velocity vector.] The classical methods for dealing with these equations are extremely laborious and incomplete in the sense that solutions in "closed form" are unobtainable. The idea of using a statistical approach at which we hinted in the preceding examples is sometimes referred to as the Monte Carlo method.

The mathematical description is the study of a flow which consists of a mixture of deterministic and stochastic processes.[1] It requires its own laws of large numbers and asymptotic theorems, the study of which has only begun. The computational procedure looks in practice as follows: we imagine that we have an ensemble of particles each represented by a set of numbers. These numbers specify the time, components of position and velocity vectors, also an index identifying the nature of the particle. With each of these sets of numbers, random processes are initiated which lead to the determination of a new set of values. There exists indeed a set of probability distributions for the new values of the parameters after a specified time interval Δt. Imagine that we draw at random and *independently*, values from a prepared collection possessing such distributions. Here a distinction must be made between those parameters which we believe vary independently of each other, and those values which are strictly determined by the values of other parameters. To illustrate this point: assume for instance that in the fission process the direction of the emitted neutron is independent of its velocity. Or again, the direction of a neutron in a homogeneous medium does not influence the distance between its origin and the site of its first collision. On the other hand, having "drawn" from appropri-

[1] von Neumann, J., and Ulam, S., *Bulletin A.M.S.*, Abstract 51-9-165 (1945).

ate distributions the velocity of a new-born particle and the distance to its first collision, the time elapsed in travel is completely determined and has to be calculated accordingly. By considering a large number of particles with their corresponding sets of parameters we obtain in this fashion another collection of particles and a new class of sets of values of their parameters. The hope is, of course, that in this manner we obtain a good sample of the distributions at the time $t+\Delta t$. This procedure is repeated as many times as required for the duration of the real process or else, in problems where we believe a stationary distribution exists, until our "experimental" distributions do not show significant changes from one step to the next.

The essential feature of the process is that we avoid dealing with multiple integrations or multiplications of the probability matrices, but instead sample single chains of events. We obtain a sample of the set of all such possible chains, and on it we can make a statistical study of both the genealogical properties and various distributions at a given time.

III

We want now to point out that modern computing machines are extremely well suited to perform the procedures described. In practice, the set of values of parameters characterizing a particle is represented, for example, by a set of numbers punched on a card. We have at the outset a large number of particles (or cards) with parameters reflecting given initial distributions. The step in time consists in the production of a new such set of cards. The original set is processed one by one by a computing machine somewhat as follows: The machine has been set up in advance with a particular sequence of prescribed operations. These divide roughly into two classes: (1) production of "random" values with their frequency distribution equal to those which govern the change of each parameter, (2) calculation of the values of those parameters which are deterministic, i.e., obtained algebraically from the others. It may seem strange that the machine can simulate the production of a series of random numbers, but this is indeed possible. In fact, it suffices to produce a sequence of numbers between 0 and 1 which have a uniform distribution in this interval but otherwise are uncorrelated, i.e., pairs will have uniform distribution in the unit square, triplets uniformly distributed in the unit cube, etc., as far as practically feasible. This can be achieved with errors as small as desired or practical. What is more, it is not necessary to store a collection of such numbers in the machine itself, but paradoxically enough the machine can

be made to produce numbers simulating the above properties by iterating a well-defined arithmetical operation.

Once a uniformly distributed random set is available, sets with a prescribed probability distribution $f(x)$ can be obtained from it by first drawing from a uniform uncorrelated distribution, and then using, instead of the number x which was drawn, another value $y=g(x)$ where $g(x)$ was computed in advance so that the values y possess the distribution $f(y)$.

Regarding the sequence of operations on a machine, more can be and has been done. The choice of the *kind* of step to be performed by the machine can be made to depend on the values of certain parameters just obtained. In this fashion even dependent probabilistic processes can be performed. Quite apart from mechanized computations, let us point out one feature of the method which makes it advantageous with, say, stepwise integration of differential equations. In order to find a particular solution, the usual method consists in iterating an algebraical step, which involves in the nth stage values obtained from the $(n-1)$th step. The procedure is thus serial, and in general one does not shorten the time required for a solution of the problem by the use of more than one computer. On the other hand, the statistical methods can be applied by many computers working in parallel and independently. Several such calculations have already been performed for problems of types discussed above.[2]

IV

Let us indicate now how other equations could be dealt with in a similar manner. The first, purely mathematical, step is to transform the given equation into an equivalent one, possessing the form of a diffusion equation with possible multiplication of the particles involved. For example as suggested by Fermi, the time-independent Schrödinger equation

$$\Delta\psi(x, y, z) = (E - V)\psi(x, y, z)$$

could be studied as follows. Re-introduce time dependence by considering

$$u(x, y, z, t) = \psi(x, y, z)e^{-Et}$$

u will obey the equation

$$\frac{\partial u}{\partial t} = \Delta u - Vu.$$

[2] Among others, problems of diffusion of neutrons, gamma rays, etc. To cite an example involving the study of matrices, there is a recent paper by Goldberger, *Phys. Rev.* 74, 1269 (1948), on the interaction of high energy neutrons with heavy nuclei.

This last equation can be interpreted however as describing the behavior of a system of particles each of which performs a random walk, i.e., diffuses isotropically and at the same time is subject to multiplication, which is determined by the value of the point function V. If the solution of the latter equation corresponds to a spatial mode multiplying exponentially in time, the examination of the spatial part will give the desired $\psi(x, y, z)$—corresponding to the lowest "eigenvalue" E.

The mathematical theory behind our computational method may be briefly sketched as follows: As mentioned above and indicated by the examples, the process is a combination of stochastic and deterministic flows.[1] In more technical terms, it consists of repeated applications of matrices—like in Markoff chains—and completely specified transformations, e.g., the transformation of phase space as given by the Hamilton differential equations.

One interesting feature of the method is that it allows one to obtain the values of certain given operators on functions obeying a differential equation, without the point-by-point knowledge of the functions which are solutions of the equation. Thus we can get directly the values of the first few moments of a distribution, or the first few coefficients in the expansion of a solution into, for example, a Fourier series without the necessity of first "obtaining" the function itself. "Symbolically" if one is interested in the value of $U(f)$ where U is a functional like the above, and f satisfies a certain operator equation $\psi(f)=0$, we can in many cases obtain an idea of the value of $U(f)$ directly, without "knowing" f at each point.

The asymptotic theorems so far established provide the analogues of the laws of large numbers, such as the generalizations of the weak and strong theorems of Bernoulli, Cantelli-Borel.[3] The more precise information corresponding to that given in the Laplace-Liapounoff theory of additive processes has not yet been obtained for our more general case. In particular it seems very difficult to estimate in a precise fashion the probability of the error due to the finiteness of the sample. This estimate would be of great practical importance, since it alone would allow us to suit the size of the sample to the desired accuracy.

The "space" in which our process takes place is the collection of all possible chains of events, or infinite branching graphs.[4] The general properties of such a phase space have been considered but much work remains to be done on the specific properties of such spaces, each corresponding to a given physical problem.

---

[3] Everett, C. J. and Ulam, S., U.S.A.E.C., Los Alamos reports LADC-533 and LADC-534. Declassified, 1948.

[4] Everett, C. J. and Ulam, S., *Proc. Nat. Acad. Sciences*, 34,403 (1948).