CHAPTER 1

The General Nature of Monte Carlo Methods

1.1 Mathematical theory and experiment

We often classify mathematicians as either pure or applied; but there are, of course, many other ways of cataloguing them, and the fashions change as the character of mathematics does. A relatively recent dichotomy contrasts the theoretical mathematician with the experimental mathematician. These designations are like those commonly used for theoretical and experimental physicists, say; they are independent of whether the objectives are pure or applied, and they do not presuppose that the theoretician sits in a bare room before a blank sheet of paper while the experimentalist fiddles with expensive apparatus in a laboratory. Although certain complicated mathematical experiments demand electronic computers, others call for no more than paper and pencil. The essential difference is that theoreticians deduce conclusions from postulates, whereas experimentalists infer conclusions from observations. It is the difference between deduction and induction.

For those who care to hunt around and stretch a point, experimental mathematics is as old as the hills. The Old Testament (1 Kings vii. 23 and 2 Chronicles iv. 2) refers to an early mathematical experiment on the numerical value of the constant π. In this case the apparatus consisted of King Solomon's temple, and the experimenter observed its columns to be about three times as great in girth as breadth. It would be nice to think that he inferred that this was a universal property of circular objects; but the text does not say so. On the other hand, experimental mathematics on anything like a substantial scale is quite a recent innovation, particularly if confined to experiments on mathematical objects, such as numbers or equations or polygons.

Monte Carlo methods comprise that branch of experimental mathematics which is concerned with experiments on random numbers. In the last decade they have found extensive use in the fields of operational research and nuclear physics, where there are a variety of problems beyond the available resources of theoretical mathematics. They are still in an early stage of development; but, even so, they have been employed sporadically in numerous other fields of science, including chemistry, biology, and medicine.

Problems handled by Monte Carlo methods are of two types called probabilistic or deterministic according to whether or not they are directly concerned with the behaviour and outcome of random processes. In the case of a probabilistic problem the simplest Monte Carlo approach is to observe random numbers, chosen in such a way that they directly simulate the physical random processes of the original problem, and to infer the desired solution from the behaviour of these random numbers. Thus we may wish to study the growth of an insect population on the basis of certain assumed vital statistics of survival and reproduction. If analysis fails us, we can at least set up a model with paper entries for the life histories of individual insects. To each such individual we may allot random numbers for its age at the births of its offspring and at its death; and then treat these and succeeding offspring likewise. Handling the random numbers to match the vital statistics, we get what amounts to a random sample from the population, which we can analyse as though it were data collected by the entomologist in the laboratory or in the field. But the artificial data may suit us better if easier to amass, or if it lets us vary the vital statistics to an extent that nature will not permit. This sort of Monte Carlo work need not call for electronic computers: Leslie and Chitty [1]† made a Monte Carlo study of capture-recapture biometrics using no more tools than a tinful of Lotto bits. Design studies of nuclear reactors and of telephone exchanges provide other examples of probabilistic problems. The fundamental particles of nuclear physics seem to obey probabilistic rather than deterministic laws. Hence we can simulate the performance of a nuclear reactor by choosing random numbers which represent the random motions of the neutrons in it. In this way we can experiment with the reactor without incurring the cost, in money, time, and safety, of its actual physical construction. If the geometry of the reactor is at all complicated, which it usually is in practice, one will probably need large-scale computing equipment to trace out the life-histories of individual neutrons in accordance with the random numbers governing them. 'In telephone systems' writes Thomson [2], 'the dependence on chance arises because, so far as the planner is concerned, the demand for service at any one time, dependent as it is on the individual decisions of a large number of subscribers, is quite unpredictable.... The planner works in terms of an average calling rate corresponding to the busy hour of the day, with random fluctuations described by some suitable probability distribution, and he has to decide how many of the individual switches and so on that are used in the course of a telephone call should be installed.... In the early days, there was no theoretical analysis available; as time went on, theory improved, but systems became more complicated and there will always be practical questions which cannot be readily answered from the existing theory.' And Thomson goes on to describe how the British Post Office built an analogue computer in 1949 to simulate the random fluctuations of telephone traffic and to analyse the resulting Monte Carlo experiment.

† Numbers in square brackets refer to the bibliography on pp. 150–158.

One of the main strengths of theoretical mathematics is its concern with abstraction and generality: one can write down symbolic expressions or formal equations which abstract the essence of a problem and reveal its underlying structure. However, this same strength carries with it an inherent weakness: the more general and formal its language, the less is theory ready to provide a numerical solution in a particular application. The idea behind the Monte Carlo approach to deterministic problems is to exploit this strength of theoretical mathematics while avoiding its associated weakness by replacing theory by experiment whenever the former falters. Specifically, suppose we have a deterministic problem which we can formulate in theoretical language but cannot solve by theoretical means. Being deterministic, this problem has no direct association with random processes; but, when theory has exposed its underlying structure, we may perhaps recognize that this structure or formal expression also describes some apparently unrelated random process, and hence we can solve the deterministic problem numerically by a Monte Carlo simulation of this concomitant probabilistic problem. For example, a problem in electromagnetic theory may require the solution of Laplace's equation subject to certain boundary conditions which defeat standard analytical techniques. Now Laplace's equation occurs very widely and, *inter alia*, in the study of particles which diffuse randomly in a region bounded by absorbing barriers. Thus we can solve the electromagnetic problem by performing an experiment, in which we guide the particles by means of random numbers until they are absorbed on barriers specially chosen to represent the prescribed boundary conditions.

This technique of solving a given problem by a Monte Carlo simulation of a different problem has sometimes been called sophisticated Monte Carlo, to distinguish it from straightforward simulation of the original problem. There are various degrees of sophistication: for instance, one may start from a given probabilistic problem, formulate it in theoretical terms, discern a second probabilistic problem described by the resulting theory, and finally solve the first probabilistic problem by simulating the second. The second problem may be a greater or a lesser distortion of the first, or it may even be totally different in character: the only thing that matters is that it shall have the same numerical solution as the first, or more generally that the wanted parts of the two solutions shall differ by a negligible amount, there being no need to ensure agreement between unwanted parts of the two solutions.

There are various reasons for indulging in sophisticated Monte Carlo methods. The main reason springs from the inferential nature of Monte Carlo work. Whenever one is inferring general laws on the basis of particular observations associated with them, the conclusions are uncertain inasmuch as the particular observations are only a more or less representative sample from the totality of all observations which might have been made. Good experimentation tries to ensure that the sample shall be more rather than less representative; and good presentation of the conclusions indicates how likely they are to be wrong by how much. Monte Carlo answers are uncertain

because they arise from raw observational data consisting of random numbers; but they can nevertheless serve a useful purpose if we can manage to make the uncertainty fairly negligible, that is to say to make it unlikely that the answers are wrong by very much. Perhaps it is worth labouring this point a little, because some people feel genuine distress at the idea of a mathematical result which is necessarily not absolutely certain. Applied mathematics is not in any case a black-and-white subject; even in theoretical applied mathematics there is always doubt whether the postulates are adequately valid for the physical situation under consideration, and whenever in a theoretical formula one substitutes the value of some experimentally determined quantity, such as the velocity of light or the constant of gravitation, one gets an uncertain numerical result. But this is no cause for worry if the uncertainty is negligible for practical purposes.

One way of reducing uncertainty in an answer is to collect and base it upon more observations. But often this is not a very economic course of action. Broadly speaking, there is a square law relationship between the error in an answer and the requisite number of observations; to reduce it tenfold calls for a hundredfold increase in the observations, and so on. To escape a formidable or even impracticable amount of experimental labour, it is profitable to change or at least distort the original problem in such a way that the uncertainty in the answers is reduced. Such procedures are known as variance-reducing techniques, because uncertainty can be measured in terms of a quantity called variance. In this direction the mathematical experimentalist is more fortunate than the experimentalist in the physical sciences: his experimental material consists of mathematical objects which can be distorted, controlled, and modified more easily and to a much greater extent than material subject to physical limitations, such as instrumental errors or structural tolerances or phenomena affected by weather and climate and so on.

Although the basic procedure of the Monte Carlo method is the manipulation of random numbers, these should not be employed prodigally. Each random number is a potential source of added uncertainty in the final result, and it will usually pay to scrutinize each part of a Monte Carlo experiment to see whether that part cannot be replaced by exact theoretical analysis contributing no uncertainty.

Moreover, as experimental work provides growing insight into the nature of a problem and suggests appropriate theory, good Monte Carlo practice may be to this extent self-liquidating [3].

Experimental mathematicians have not come to replace theoretical ones. Each relies on the other, as with other scientists. The experimenter needs theory to give structure and purpose to his experiments; and the theoretician needs experiment to assess the value and position of his theory. The Monte Carlo experimentalist needs wide experience of formulae and results in pure mathematics, and especially the theory of probability, in order that he may discern those connexions between apparently dissimilar problems which suggest sophisticated Monte Carlo methods. He has to exercise ingenuity in distorting and modifying problems in the pursuit of variance-reducing techniques. He has to be competent at statistical and inferential procedures in order to extract the most reliable conclusions from his observational data. The tools of his trade include computing machinery, and he must be familiar with numerical analysis. As in all experimental work, a feel for the problem is a great and sometimes essential asset. Finally Monte Carlo methods are recent innovations still under development, general devices are few in number, and a great deal depends upon having enough originality to create special methods to suit individual problems. Despite all these demands, Monte Carlo work is a subject with which some acquaintance is well worth while for anyone who has to deal with mathematical problems encountered in real life, as opposed to ones engineered to exemplify textbook theory. For real-life problems mathematical experiment is a most necessary alternative to theory. In short, Monte Carlo methods constitute a fascinating, exacting, and often indispensable craft with a range of applications that is already very wide yet far from fully explored.

1.2 Brief history of Monte Carlo methods

The name and the systematic development of Monte Carlo methods dates from about 1944. There are however a number of isolated and undeveloped instances on much earlier occasions. For example, in the second half of the nineteenth century a number of people performed experiments, in which they threw a needle in a haphazard manner onto a board ruled with parallel straight lines and inferred

the value of π from observations of the number of intersections between needle and lines. An account of this playful diversion (indulged in by a certain Captain Fox, amongst others, whilst recovering from wounds incurred in the American Civil War) occurs in a paper by Hall [4]. In the early part of the twentieth century, British statistical schools indulged in a fair amount of unsophisticated Monte Carlo work; but most of this seems to have been of a didactic character and rarely used for research or discovery. The belief was that students could not really appreciate the consequences of statistical theory unless they had seen it exemplified with the aid of laboratory apparatus: demonstrators therefore poured lead shot down boards studded with pins, which deflected the shot in a random fashion into several collecting boxes, and the students were required to see that the frequency of the shot in the various boxes conformed more or less to the predictions of theory; they drew numbered counters from jam-jars and pots (called urns for the sake of scientific dignity) and verified that averages of various sets of such numbers behaved as sampling theory said they should, and so on. Only on a few rare occasions was the emphasis on original discovery rather than comforting verification. In 1908 Student (W. S. Gosset) used experimental sampling to help him towards his discovery of the distribution of the 1
correlation coefficient. Apparently he knew some of the moments of the distribution and had conjectured, perhaps on the basis of these or perhaps by way of Occam's razor, that the analytical form would be proportional to $(1-\alpha r^2)^\beta$, one of Pearson's frequency curves, where r is the correlation coefficient and α and β are constants depending upon the sample size n. Having fitted samples for $n = 4$ and $n = 8$ to this conjectured expression, and having rounded the resulting estimates of α and β, he guessed that $\alpha = 1$ and $\beta = \frac{1}{2}(n-4)$, which happens to be the exact theoretical result. This remarkable achievement is nevertheless rather different from, though in a sense better than, the ordinary use of Monte Carlo methods nowadays, in which there is little attempt to guess exact results. In the same year Student also used sampling to bolster his faith in his so-called t-distribution, which he had derived by a somewhat shaky and incomplete theoretical analysis.

One consequence of this didactic and verifying rôle for sampling experiments was that the experiments were deliberately shorn of distracting improvements, which might have been employed to sharpen the accuracy of the results: statisticians were insistent that other experimentalists should design experiments to be as little subject to unwanted error as possible, and had indeed given important and useful help to the experimentalist in this way; but in their own experiments they were singularly inefficient, nay, negligent in this respect.

The real use of Monte Carlo methods as a research tool stems from work on the atomic bomb during the second world war. This work involved a direct simulation of the probabilistic problems concerned with random neutron diffusion in fissile material; but even at an early stage of these investigations, von Neumann and Ulam refined this direct simulation with certain variance-reducing techniques, in particular 'Russian roulette' and 'splitting' methods [5]. However, the systematic development of these ideas had to await the work of Harris and Herman Kahn in 1948.

The possibility of applying Monte Carlo methods to deterministic problems was noticed by Fermi, von Neumann, and Ulam and popularized by them in the immediate post-war years. About 1948 Fermi, Metropolis, and Ulam obtained Monte Carlo estimates for the eigenvalues of the Schrödinger equation. Dr Stephen Brush (of the Radiation Laboratory at Livermore), who has a particular interest in the history of mathematics, has unearthed a paper by Kelvin [6] in which sixty years ago astonishingly modern Monte Carlo techniques appear in a discussion of the Boltzmann equation. But Lord Kelvin was more concerned with his results than with his (to him, no doubt, obvious) methods, and it seems entirely right and proper that Ulam, von Neumann, and Fermi should take the credit for not only independently rediscovering Monte Carlo methods but also ensuring that their scientific colleagues should become aware of the possibilities, potentialities, and physical applications. The dissemination of ideas is always an essential component of their production, and never more so than in today's conditions of prolific discovery and publication.

The ensuing intensive study of Monte Carlo methods in the 1950s, particularly in the U.S.A., served paradoxically enough to discredit

the subject. There was an understandable attempt to solve every problem in sight by Monte Carlo, but not enough attention paid to which of these problems it could solve efficiently and which it could only handle inefficiently; and proponents of conventional numerical methods were not above pointing to those problems where Monte Carlo methods were markedly inferior to numerical analysis. Their case is weakened by their reluctance to discuss advanced techniques, as Morton remarks [7].

In the last few years Monte Carlo methods have come back into favour. This is mainly due to better recognition of those problems in which it is the best, and sometimes the only, available technique. Such problems have grown in number, partly because improved variance-reducing techniques recently discovered have made Monte Carlo efficient where it had previously been inefficient, and partly because Monte Carlo methods tend to flourish on problems that involve a mass of practical complications of the sort encountered more and more frequently as applied mathematics and operational research come to grips with actualities.