SIMULATION IN HIGH-ENERGY PHYSICS

F. James
CERN, Geneva, Switzerland.

1. INTRODUCTION

These lectures, for a mixed audience of physicists and computer scientists, offer a good opportunity for stepping back and taking a general view of a calculation technique which is steadily growing in importance in high-energy physics as well as in other sciences.

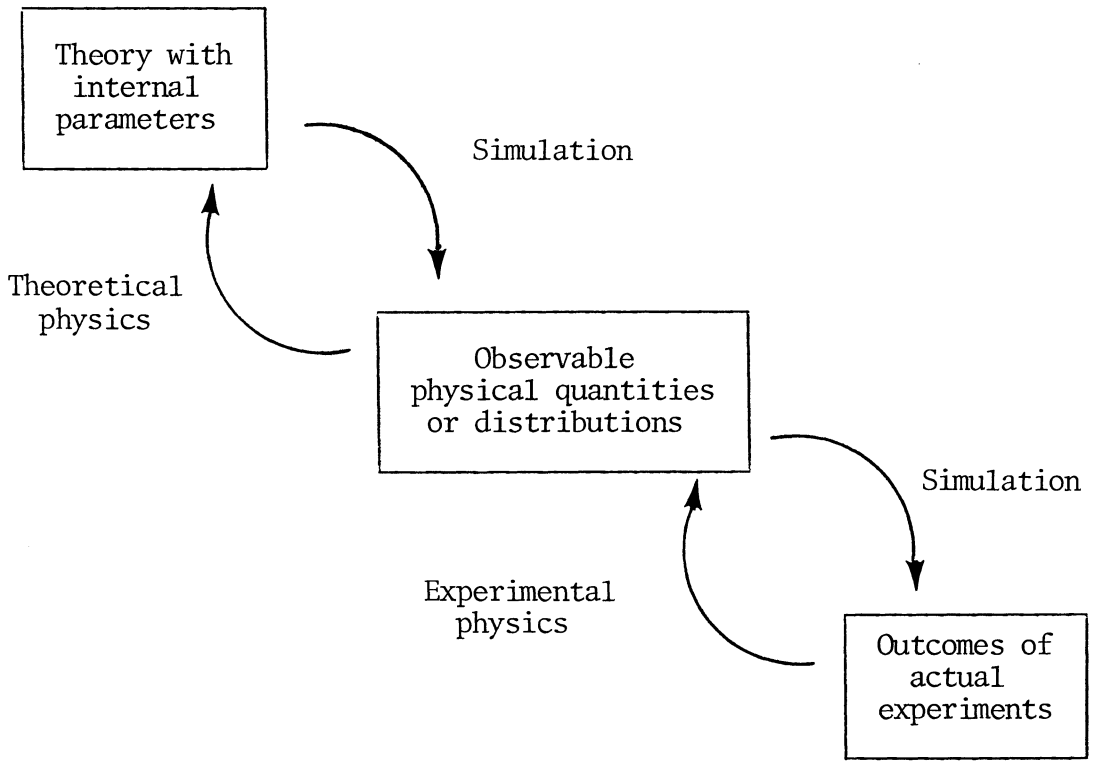
From the physicist's point of view, simulation is a convenient technique, made necessary by the increasing complexity of his theories and experiments, and made possible by the increasing speed of computers. From the point of view of the computer scientist or mathematician, simulation must be considered in the larger context of Monte Carlo methods, generally made useful by the existence of larger computers, but interesting in their own right for the variety of purely mathematical disciplines involved.

In fact the essential problem in preparing these lectures is not so much the excessive amount of material to be covered as the variety of quite different approaches which may be adopted in exposing the subject. When in addition neither the speaker nor the audience may be conveniently classified into some pigeon-hole which could help to define the orientation of the lectures, the only possibility is to try to indicate different approaches as we go along.

In view of the increasing quality and quantity of published works on Monte Carlo methods in the past few years, I will spend a minimum amount of time discussing particular numerical techniques which can now be found abundantly in the literature. Since the audience is composed largely of young people who may themselves make major contributions to this field, it seems appropriate to spend as much time on unsolved problems as on established methods.

2. PROBLEMS IN HIGH-ENERGY PHYSICS

To help in visualizing the place of simulation in the process of "doing high-energy physics", let us consider the following simplified diagram:



The important part of the above diagram is the middle box, containing the observable (although idealized) physical phenomena usually referred to in the jargon of the trade as "the physics of what's going on". In high-energy physics this consists of (i) the fundamental properties of the elementary particles, and (ii) the interactions between them. All of the input information comes essentially from observation of scattering and decay reactions of particles. It is these phenomena which the experimentalists are trying to measure (the lower box), and the theoreticians are trying to explain and predict (the upper box).

As in other branches of science which have reached a relatively advanced stage, a perennial problem in high-energy physics is the inability to observe directly the phenomena of interest. This difficulty of observation usually is due to some combination of the following reasons:

i) The fundamental nature of matter on the microscopic level is statistical rather than deterministic. One does not predict the outcome of a given experiment, but only the probabilities of observing different outcomes. (The apparently deterministic behaviour of macroscopic objects comes about only because one experimental outcome is overwhelmingly more probable than all others.)

ii) The increasingly complex theories put forward to explain high-energy phenomena invoke, more often than not, some unobservable object or process whose properties determine the (in principle) observable quantities.

iii) Even the objects which are in principle observable are usually too small, too fast, or too weak to be detected "directly" and we are forced to make use of complicated devices which introduce detection biases, measurement errors, and background noise, all of which may be difficult to determine accurately.

iv) The number of independent variables necessary to describe completely one high-energy event (scattering or decay) with n outgoing particles is 3n - 4. The problems then arise, especially for large n, of what choice of independent variables to use, what to do about the unobserved variables if any, and the general problem of how to compare two m-dimensional distributions.

In view of the above difficulties, it is a major computational problem (i) for the theoretical physicist, once he has a theory, to predict observable distributions, and (ii) for the experimental physicist, given such a hypothetical physical distribution, to determine what outcome would result in a particular experimental situation. These two kinds of computations are denoted on our diagram above by the word "simulation" to indicate that this is a common and increasingly important technique, although it is certainly not the only one, and often not the best one.

Notice that the role of simulation can never replace that of the physicist since it works essentially in the opposite direction. But by helping him to perform the more straightforward deductive calculations, it frees him for the more creative inductive work indicated on the diagram.

3. SIMULATION AND MONTE CARLO INTEGRATION

Let us define a Monte Carlo calculation as any calculation making use of random numbers. We will discuss "randomness" in some detail later, but for the moment let us consider just a sequence of numbers $r_1 \ldots r_n$ each distributed randomly (in the usual classical sense) between zero and one. Now suppose that the result of the calculation is a number, or set of numbers, F. We have

$$F = F(r_1, r_2 \ldots r_n)$$

and this we know from statistics to be an unbiased estimate of the integral

$$I = \int_0^1 \int_0^1 \ldots \int_0^1 F(x_1, x_2 \ldots x_n)\, dx_1\, dx_2 \ldots dx_n \,.$$

In this rather formal sense then, all Monte Carlo calculations may be considered as integrations. But simulation, an example of a Monte Carlo calculation, is integration in a much more direct sense.

Let us define simulation as any calculation involving a series of n trials (or events), the outcome of each trial being a function of k random numbers. In this case the calculation involves $n \cdot k$ random numbers in all, but is clearly equivalent to an integration over a k-dimensional space, where the estimate of the integral is now the average over the n points.

3.1 Mathematical foundation of Monte Carlo integration

Consider a function of k variables $F(x_1 \ldots x_k)$. We define the expectation of the function F as

$$E(F) = \int_{\Omega} \int \ldots \int F(x_1 \ldots x_k)\, g(x_1 \ldots x_k)\, dx_1 \ldots dx_k \,,$$

where Ω denotes the full range of the $x_1 \ldots x_k$ and $g(x_1 \ldots x_k)$ is the probability density function describing the distribution of the $x_1 \ldots x_k$.

We will be interested in the special case where the $x_1 \ldots x_k$ have all been "normalized" to fall into the range $0 \leq x \leq 1$ and where the $x_1 \ldots x_k$ are all uniformly distributed, that is $g(x_1 \ldots x_k) = 1$. In this case,

$$E(F) = \int_0^1 \int_0^1 \cdots \int_0^1 F(x_1 \ldots x_k)\, dx_1 \ldots dx_k ,$$

so that the expectation is just the mean, or average value of F. Now let us define the variance as

$$\text{var}(F) = E\{F - E(F)\}^2$$

which can be written, for our special case, as

$$\text{var}(F) = \int \int \cdots \int \{F - E(F)\}^2\, dx_1 \ldots dx_k .$$

The variance is thus the mean squared deviation of F from its mean E(F).

Now let us write X the vector $x_1 \ldots x_k$, and let X_i represent the random vector X corresponding to independent, truly random choices for each $x_1 \ldots x_k$. The law of large numbers then says that, if var(F) is finite,

$$\lim_{n \to \infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} F(X_i) \right\} = E(F) .$$

Since the quantity in curly brackets above is just the n-point Monte Carlo estimate of the integral E(F), we can restate the law of large numbers as follows: the Monte Carlo estimate of the integral of a function F is a consistent estimator whenever var(F) is finite.

Having seen that the Monte Carlo estimate is consistent, we may now see that it is unbiased. We recall that an unbiased estimator is one whose expectation is equal to the true value of the quantity estimated. The Monte Carlo estimate is the average of n terms, and since the expectation

of each term is equal to the true value of the integral, the unbiasedness of the estimate follows trivially from the linearity of the expectation operator.

It remains now to find the complete distribution of the Monte Carlo estimate. Unfortunately, this distribution depends on F and on n in a complicated way which we cannot in general write down for small n. Limiting ourselves to the asymptotic distribution (for n large), we can, however, use the Central Limit Theorem which says that, given a sequence of random variables y_i from distributions of expectations μ_i and finite variances σ_i^2, the sum

$$Y_n = \sum_{i=1}^{n} y_i$$

will be asymptotically normally distributed, with mean

$$M = \sum_{i=1}^{n} \mu_i$$

and variance

$$\delta^2 = \sum_{i=1}^{n} \sigma_i^2 .$$

Since each term in a Monte Carlo integration is a random variable from the same distribution, the estimate

$$\bar{I} = \frac{1}{n} \sum_{i=1}^{n} F(X_i)$$

will be asymptotically normally distributed, with mean equal to the true value of the integral and variance

$$\mathrm{var}(\bar{I}) = \frac{1}{n} \mathrm{var}(F) .$$

Since the standard error on $\bar{I}$ [the root-mean-square deviation of $\bar{I}$ from E(F)] is the square root of var($\bar{I}$), this gives directly the well-known result that Monte Carlo estimates converge as $n^{-\frac{1}{2}}$.

3.2 Statistical efficiency

From the last equation of the preceding section it is clear that any desired accuracy can be attained in simulation calculations by increasing the number of trials n to the necessary value. The variance F can be estimated by the sample variance, giving the required number

$$n \geq \frac{\sum_{i=1}^{n} (F - \bar{I})^2}{\sigma^2} = \frac{\left(\sum_{i=1}^{n} F^2\right) - n\bar{I}^2}{\sigma^2} ,$$

where σ is the desired standard error in $\bar{I}$.

For practical calculations, however, one often encounters functions whose variances are so large that the required n is unreasonable, even for the fastest computers. The most striking example of this is the delta function, defined to be zero everywhere except for one infinitely sharp peak whose integral is unity. It is not a proper mathematical function since it cannot be defined completely by giving its value everywhere, but it is of great importance in physics where it is often used to express laws in the form of constraints, such as the conservation of energy-momentum. It can be defined as the limit of a normal distribution

$$\delta(x) = \lim_{\alpha \to 0} N(x,\alpha^2) = \lim_{\alpha \to 0} \frac{1}{\alpha\sqrt{2\pi}} e^{-x^2/2\alpha^2} .$$

The variance of this function (not to be confused with the variance of a normally-distributed variable) may be calculated straightforwardly

$$\mathrm{var}\{N(x,\alpha^2)\} = \int_{-\infty}^{\infty} N^2(x,\alpha^2)\, dx = \frac{1}{\alpha\sqrt{4\pi}}$$

We see, therefore, that the variance of a Gaussian (normal function) is proportional to the sharpness of the peaking, as expressed by the parameter $1/\alpha$, and in the limit of the delta function this variance is infinite. The rule is easily extended to Breit-Wigner (truncated Cauchy) functions, which are of great importance in high-energy physics and also have variances which become very large as the peaks become narrow.

Historically, the delta function has been handled by approximating it by a narrow peak of non-zero width. This means introducing an approximation into the theory which is essentially different from the approximation resulting from the statistical fluctuations of Monte Carlo results. The "wide delta-function" approximation can lead to completely unphysical answers unless the results are interpreted carefully. For example, this approximation may mean allowing momentum and energy conservation to be violated by a small amount in each event.

The effect of a delta function is to reduce the dimensionality of the integration space by one. It is clear that, except in the approximation mentioned above, this integration cannot be handled by Monte Carlo, but must be performed analytically.

3.3 Variance-reducing techniques

As we have seen, the error on a Monte Carlo estimate is of the form

$$\delta = \frac{\sigma}{\sqrt{n}} ,$$

where σ is the square root of the variance of the function involved, and n is the number of points or events used. As long as we use truly random or pseudorandom numbers, there is nothing we can do to increase $\sqrt{n}$ except to run our programs longer (in the next section we discuss the possibility of using so-called "quasi-random" numbers to improve the convergence). The optimization of Monte Carlo solutions consists therefore in reducing the variance of the sample. A large and growing number of variance-reducing techniques exist, and rather than try to mention all of them, we will concentrate on two techniques felt to be most useful and representative.

3.3.1 Stratified sampling

It seems intuitively obvious that the random fluctuations in our Monte Carlo integrations would be reduced if we could somehow assure that points were more evenly distributed over the space. One clear way of doing this is to divide the space up into equal cells and to choose an equal number of points randomly within each cell. Indeed this usually will improve the variance, but in actual practice the situation is somewhat complicated by the following considerations:

i) If we are working in many dimensions the division of the space into equal hypercubes becomes unwieldy since the *minimum* number of such cubes is 2^k for a k-dimensional space.

ii) If we write down the expression for the variance in the case of stratified sampling (it is simply a complicated sum of squares of integrals and integrals of squares) we see that the amount gained -- or in fact whether we gain at all -- depends on the behaviour of the function, on the way we choose the cells, and on the way we distribute points among the different cells.

iii) Rather than choosing equal-sized cells, a better way is to choose them so that the variation of the function is about the same in each cell. Indeed we will gain if the differences between mean values of the function in the different cells are greater than the variations within the cells.

iv) Given the division into cells, one should choose the number of points in the following way: in the j^{th} cell choose

$$n_j^2 \propto \left[\Delta j \int_{\Delta_j} g^2(x)\, dx - \left\{ \int_{\Delta_j} g(x)\, dx \right\}^2 \right],$$

where Δ_j represents the volume of the j^{th} cell in the k-dimensional space.

v) If nothing is known *a priori* about the function the only natural way to choose the cells is to divide the space equally. It can be shown that if we do this, and if we choose equal numbers of points for each

cell, we cannot lose by such a stratification (that is, the variance cannot be larger than for the unstratified estimate). In the worst case, where the mean value of the function within each cell is the same, the variance is not changed by the stratification. The next question is how many cells should we choose? Since we have seen that we cannot lose by subdividing a cell, we should choose as many cells as possible. However, if there is only one point per cell, we no longer have a reliable error estimate since we have no estimate of the variance within a cell. But we have an upper limit on the error, namely the error calculated as though the space had not been stratified.

3.3.2 Importance sampling

We have seen that a large variation in the value of the function F leads to a large error in the Monte Carlo estimate of its integral. Conversely, our simulation calculations will be most efficient when each trial (event) has nearly the same function value (weight). This can be arranged by choosing a large number of trials in regions of the sampling space where the function value is largest and reducing the function value in this region to compensate for the fact that a larger number of points was chosen. In this way the "weighted" function becomes more nearly constant, and the variance is reduced. Such a technique is known as importance sampling since it consists of increasing the sampling density in the more important regions.

To apply importance sampling to a function F(X), a function g(X) must be found such that

i) The integral $G(X) = \int_0^X g(X')\,dX'$ is known analytically.

ii) The integral G(X) can be inverted (solved for X) analytically.

iii) The ratio F(X)/g(X) is as nearly constant as possible, or at least has a smaller variance than F(X) alone.

Once such a function has been found, a transformation is made:

$$F(X)\,dX \rightarrow F(X)\,\frac{dG(X)}{g(X)}\,,$$

that is, instead of choosing X randomly and uniformly in its range, choose G randomly and uniformly in the corresponding range, solve for X, and calculate the new function value as F(X)/g(X). The variance of the estimate will now be proportional to the variance of the *ratio* F(X)/g(X), which should be smaller than that of F(X) alone.

3.4 Limits of integration

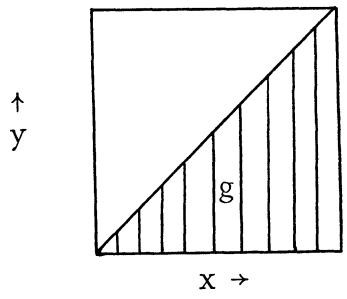
One of the fundamental advantages of the Monte Carlo method is that it can handle easily problems in which the limits of integration are awkward and interdependent. However, we must be careful about how these limits are chosen, as the following example illustrates:

$$I = \int_{x=0}^{1} \int_{y=0}^{x} g(x,y)\, dy\, dx \ .$$

It is tempting to solve this problem in the following way:

i) choose a random number (x_i) between 0 and 1;
ii) choose another random number (y_i) between 0 and x_i;
iii) take the sum of $g(x_i,y_i)$, repeating (i) and (ii).

A simple graphical representation of what we have done here shows that it gives the wrong answer:



While it is true that this procedure would yield points only in the allowed region (the lower triangle), it would give the same number of points along each vertical line. This gives a much higher density of points on the left-hand side than the right-hand side.

A correct way to integrate the above problem is:

i) choose a random number (x_i) between 0 and 1;
ii) choose another random number (y_i) between 0 and 1;
iii) if $y_i > x_i$, reject the point;
iv) take the sum of $g(x_i,y_i)$ over the remaining points.

This method, although correct, has the disadvantage of using only half the points generated. That is, it is equivalent to integrating over the whole square, but considering the function value is equal to zero in the upper triangle. A better way to handle this problem is the following:

i) choose <u>two</u> independent random numbers, r_1,r_2;
ii) set x_i = larger of r_1,r_2;
iii) set y_i = smaller of r_1,r_2;
iv) sum up $g(x_i,y_i)$ as before.

Graphically, this is equivalent to choosing points randomly over the square, then folding the square about the diagonal so that all points fall into the lower triangle. It is clear that this now results in a constant density of points without rejecting any points. However, if we think of the calculation on a large computer, we see that if all points in the upper triangle were rejected, this would amount to very little extra time if the time required for the test (is x_i greater than y_i?) is short compared with the time necessary to evaluate the function. That is, we are still wasting our time in the upper triangle, but we may only be wasting a negligible amount of *time*, even though we are throwing away half the points.

A third correct method for integrating over the triangle consists in choosing points as in the wrong method described first, but then weighting the function values to account for the bias introduced as follows:

i) choose a random number (x_i) between 0 and 1;
ii) choose another random number (y_i) between 0 and x_i;
iii) take the sum of $2x_i\ g(x_i,y_i)$, repeating (i) and (ii).

This last method will, of course, be superior if the variance of $x \cdot g$ is smaller than that of g.

3.5 Comparison with numerical quadrature

The simplest method for numerical integration (*not* Monte Carlo) is the trapezoid rule, which consists of dividing the required interval into n bands and approximating the integral over each band by the area of the trapezoid inscribed under (or over) the curve. This reduces to estimating the integral by taking the average value of the function as determined from n equally-spaced points and multiplying by the total interval. For large n, we can think of the function as being expressed by a Taylor's series expansion about each of the n points; then the constant terms and the first derivative terms will be integrated exactly by the trapezoid rule, and the largest contribution to the error will come from the second derivative (constant curvature) terms. This error will be proportional to the sagittas of the curve segments over each band, and these sagittas will each be proportional to the square of the distance between successive points where the function is evaluated. Therefore, if the function is evaluated at n equally-spaced points, the error on the integral will be proportional to $1/n^2$ for large n. In fact, if we push the method a little farther we can obtain even faster convergence. Consider an estimate of an integral found by using the trapezoid rule at n points:

$$T_n = I + \phi(n^{-2}) + \phi(n^{-4}) + \ldots ,$$

where I is the "true" value of the integral and $\phi(n^{-2})$ means an error of the order of $1/n^2$, or this may be considered to be the integral of the third term in the Taylor's series expansion mentioned above. If we make another estimate based on 2n points we will have:

$$T_{2n} = I + \frac{1}{4}\phi(n^{-2}) + \frac{1}{16}\phi(n^{-4}) + \ldots .$$

We now consider the following linear combination:

$$T' = \frac{4T_{2n} - T_n}{3} = I - \frac{1}{4}\phi(n^{-4}) + \ldots .$$

Here the term in $1/n^2$ has dropped out, so that this method will converge as $1/n^4$. This process can then be extended in an obvious way in order to eliminate the $1/n^4$ term, etc.

But rather than extend this method to its limit, we can use the fully optimized formula due to Gauss, which in fact represents the limit of this kind of method in the sense that the n-point Gauss formula is exact for all polynomials up to order 2n - 1.

We remind the reader here that Monte Carlo integration using standard random numbers converges always as $\sqrt{n}$. The following table may, therefore, be established comparing the dependence of errors on number of points for different methods of numerical estimation of a one-dimensional integral in order of increasingly fast convergence:

Monte Carlo	$\sigma \sim n^{-\frac{1}{2}}$
Trapezoid	$\sigma \sim n^{-2}$
"Second-order rule"	$\sigma \sim n^{-4}$
m-point Gauss rule	$\sigma \sim n^{-2m+1}$.

This makes the Monte Carlo method look pretty bad, even if we notice that it is the only one of these for which the error estimate does not depend on any assumed "smooth" behaviour of the function.

However, the future for Monte Carlo looks considerably brighter when we go to multi-dimensional integration. If we modify the above formulae to hold for integration over k dimensions, we see that the Monte Carlo expression remains unchanged, whereas for the others the exponent must be divided by k. This means that Monte Carlo converges faster than the trapezoid rule in five or more dimensions, and when we go above five dimensions the *non*-Monte Carlo methods begin to break down completely because of the outrageous number of points required. As we have seen, these error formulae are only valid for large n (that is, small spacing between points) which certainly means at least 10 points per dimension, for example. But 10 points in five dimensions means already 10^5 points in the whole space, and this kind of number must be considered as the limit of the possibilities of today's computers. We remark here that a complete description of an ℓ-particle final state requires $3\ell - 4$

parameters, so that for four-body phase space we are already operating in eight dimensions (although some of these, such as the azimuthal orientation of the event about the beam direction, may be of no interest).

One would be justified, at this point, in wondering how it can be that a random selection of point can be more efficient than a uniform or carefully optimized distribution. Intuitively it seems that most of the error in a Monte Carlo integral comes from the non-uniformity of the point distribution, and that we could reduce the error by choosing points more uniformly. In fact, this is not at all true (but see Section 3.3.1, Stratified sampling) as can be seen by considering the Gauss formula, which is optimum in one dimension. The Gauss rule prescribes points which correspond to the zeros of Legendre polynomials, which are not uniformly spaced for small n. In many dimensions we always have effectively a small n, and we have seen that the uniform (rectangular grid) distribution gives very poor convergence.

There are methods which are better than Monte Carlo in many dimensions but these do not correspond to straightforward extensions of good one-dimensional methods, and in fact no general method is known to be optimum for n dimensions. The best that has been done in this domain is to establish limits on the lowest possible error that may be achieved, and it turns out that Monte Carlo is not very far from these limits. The conclusion is then, that in many dimensions you cannot do much better than Monte Carlo, but you can do a lot worse.

4. RANDOM NUMBERS

In mathematical statistics a random variable is simply a variable which may take on different values. It is not, in general, possible to predict in advance the value of a random variable; we can only give the probability that it will take on values in a given range. In the degenerate case where the probability of taking on a given value becomes unity, the random variable becomes a certain variable, but otherwise one cannot say that one variable is, for example, more random than another.

In Monte Carlo studies one often uses the word "random" to mean something quite different. Here the term is applied to sequences of numbers which, once they have been determined, are not at all random in

the statistical sense, but may have some properties which are similar to the properties of a truly random sequence. In fact, it is customary to distinguish three different types of "random" sequences, namely:

i) Truly random numbers, random in the statistical sense that at any given point in the sequence, all numbers are equally probable and independent of the preceding numbers.

ii) Pseudo-random numbers are generated according to an arithmetic prescription so that each number depends on the preceding one, but in such a way that any sufficiently short sequence appears to be truly random in a sense described below.

iii) Quasi-random numbers are not meant to appear random at all, but any sufficiently long sequence should have the property of being more advantageous than truly random numbers for the particular problem being treated.

We have seen that any Monte Carlo calculation involves the use of a sequence of "random" numbers; in fact this may be considered as the definition of a Monte Carlo calculation. The choice of a proper sequence of "random" numbers is therefore central to the method, and, in spite of great progress in the field, this choice is unfortunately not at all trivial. We shall discuss in more detail the three main classes of "random" sequences.

4.1 Truly random numbers

A sequence of truly random numbers is unpredictable and unreproducible. Such a sequence can only be generated by a random physical process, for example radioactive decay, thermal noise in electronic devices, cosmic ray arrival times, etc. If such a physical process is used to generate the random numbers, then there is no theoretical problem since the theory of Monte Carlo outlined earlier in these lectures will certainly be valid provided there is no defect in the physical apparatus which might introduce a bias in the generator.

In practice, however, it turns out to be very difficult to construct physical generators of this type which are fast enough (one needs typically hundreds of numbers per second) and at the same time accurate and unbiased. Faced with these practical difficulties, the development of truly random

generators has largely been abandoned, and the technology has not advanced significantly in recent years. We therefore find ourselves in the unfortunate situation that the truly random numbers demanded by the theory are not feasible for practical reasons. Two possible ways out of this dilemma are discussed below.

4.2 Pseudo-random numbers

It is in the field of pseudo-random numbers that recent progress has been the greatest, but the situation is still far from ideal. The great advantage with pseudo-random numbers is that calculations may be checked or compared by regenerating the identical sequence. Also, the pseudo-random sequences used for large computers are very fast (usually much faster than is really necessary) and require very little storage and no special hardware (a slight exception is the generators which can be initialized by the computer's internal clock, thereby allowing "random" entry points if "independent" sequences are desired). However, pseudo-random generators are plagued by certain fundamental problems:

i) Since each number is generated from the preceding one, the recurrence of one number results in the recurrence of an entire sequence. Each generator is therefore characterized by a certain *period* and much of the progress in finding better generators consists in finding generators with longer periods. Nowadays one can generally achieve the maximum period (limited by the word length in the computer) but for machines like the CDC 6600 this period is so long that there is no point in going to the limit.

ii) Again since each number is generated from the preceding number by a fixed formula, if any number ξ_j is arbitrarily close to some preceding number ξ_i, then ξ_{j+1} will be arbitrarily close to ξ_{i+1}. Exactly what is meant by arbitrarily close depends on the generator and some are better than others in this respect.

iii) In order to show that a particular pseudo-random generator is acceptable, it must be submitted to an infinite number of statistical tests, which is clearly impossible. In practice the generators are submitted to increasingly complicated tests until one finds the lowest order correlation existing (i.e. the simplest test which

is failed) at which time it is said that the principal correlations (unrandomness) are understood. It is generally agreed that it is better to use a generator whose correlations are understood than an unknown one which may be better. Generally good pseudo-random generators exist in the libraries of all large computer centers, but they are not always suitable for all Monte Carlo problems.

iv) The most fundamental problem is how to justify at all the use of pseudo-random numbers in Monte Carlo calculations, even if the sequence in question passes all the statistical tests that can be devised in a finite time. A statistical test after all consists in forming some function of the random numbers and seeing whether it lies in a certain acceptable region around the value expected from a truly random sequence. Since a Monte Carlo estimate is also a function of the (pseudo-) random numbers, one such "randomness" test is to compare the Monte Carlo estimate with the exact value. But the reason for the calculation of the estimate is that we do not know the exact value, so that we must accept the estimate as valid, which is seen to be equivalent to assuming that the additional statistical test will be passed simply because a certain number of other tests were passed.

Although I personally tend to be optimistic about the usefulness of pseudo-random sequences, it is interesting to consider what must be the extreme pessimistic point of view expressed by Zaremba (see bibliography):

> "As far as pseudo-random numbers are concerned, the traditional term 'tests of randomness' is certainly a misnomer. Surely, in contrast to their name, the object af such tests is not the random origin of the sequences, since this would amount to testing a hypothesis *known* to be false ... . The only reasonable object of such tests can be the verification of those properties of the sequences concerned which promise a satisfactory accuracy of the results of computations carried out with their help."

The last sentence of the above quotation suggests a totally different philosophy of "random" number generation, which consists in abandoning completely the idea that the sequence should in some sense

appear to be random. This alternative philosophy is usually referred to as *quasi-Monte Carlo*.

4.3 Quasi-random numbers

The principle of quasi-random generators is that certain correlations are harmless (depending, of course, on the problem) and indeed they can sometimes be helpful. In fact the development of suitable quasi-random generators probably constitutes the most important recent progress in multidimensional numerical integration, since it is only in this way that one can improve on the $n^{-\frac{1}{2}}$ convergence of Monte Carlo estimates.

For an integration in k dimensions we require sets of k random numbers, each set giving rise to a point in the space. The k numbers within each set must be uncorrelated, for if any number in a given set depends on another number in the same set, this restricts the possible combinations, making certain regions of the space inaccessible. However, certain correlations between successive sets of numbers do not bother us. For example, if each point is chosen close to the preceding point (an obvious non-randomness), this does not matter as long as the final sample is evenly distributed over the space. This is equivalent to saying that the order in which the points are chosen is of no importance since this does not normally enter into such calculations.

This leads us to study uniformly-distributed numbers, a subset of quasi-random numbers which have the property that the density of points is more uniform than that of truly random numbers. There is a very strong serial correlation between successive numbers produced by the same generator, so we need in general k different generators for an integration in k dimensions. A point in phase space is then chosen by taking one "random" number from each of the k generators.

Two different types of quasi-random generators have been studied in detail and seem well-suited to Monte Carlo integration problems. Each type has as many different generators as there are prime numbers. Integrals calculated with these generators converge faster than 1/n, for large n, a considerable improvement over pseudo-random numbers.

i) The Richtmyer formula gives, for the i^{th} "random" number of the j^{th} generator:

$$r_{ij} = iS_j, \text{ modulo } 1 ,$$

where S_j is the square root of the j^{th} prime number. This means that the difference between successive numbers of the same generator is always S_j or $S_j - 1$. This also leads unfortunately to strong short-term correlations (i.e. correlations which go away as $n \to \infty$) between corresponding numbers produced by different generators.

ii) The Van der Corput formula consists of expressing the integers in a system of base P, then reversing the digits, preceding them by a point, and interpreting the resulting numbers as fractions in a system of base P. P is any prime number. For example, in the binary system (P = 2):

Decimal	Binary	Binary fraction	Decimal fraction
n = 1	1	0.1	0.5
2	10	0.01	0.25
3	11	0.11	0.75
4	100	0.001	0.125
5	101	0.101	0.625
6	110	0.011	0.375
7	111	0.111	0.875
8	1000	0.0001	0.0625

In spite of the apparent complexity of this method, there is an algorithm which is simple and quite fast for generating such numbers (see bibliography) and it is probably superior to the Richtmyer formula for reasonable sample sizes.

To use either of the above formulas in practical cases where the sample size is limited to the order of thousands of points, it is found necessary to reduce the short-term correlation between different genera-

tors. In order to do this and still retain the uniformity of the individual distribution, a shuffling technique is used as follows.

Before the generation of the first point, a buffer is filled with the first m numbers generated by each of the k generators ($m \sim 20$). Then each time a point is to be generated, the k numbers required are chosen at random one from each set of m in the buffer, and the used members are replaced in the buffer by the next number generated by each of the k generators. This assures that eventually all numbers will be used, but in an order different from that in which they were generated.

5. THE PROBLEM OF MANY-DIMENSIONAL DISTRIBUTIONS

Having seen what are the fundamental problems in Monte Carlo integration, one can look still deeper and ask what is the underlying problem in our scientific method which forces us to use this technique in the first place. Essentially it is the lack of satisfactory statistical methods for comparing two multi-dimensional distributions that forces us to integrate out most of the variables to reduce the dimensionality of the space either to zero (a cross-section, reaction amplitude, phase shift, lifetime, etc.), one (a histogram), or two (a scatter-plot).

For the estimation of parameters from experimental data, the usual maximum likelihood method allows one to use the full multi-dimensional space with no loss of information and without making unreasonable demands on computers (except sometimes for a lack of memory space if the amount of data is large) the power of the maximum likelihood method for estimating parameters is essentially independent of the dimensionality of the space involved.

For goodness-of-fit tests, however, the situation is radically different. The usual chi-square method is fine for a one-dimensional distribution, but since it requires grouping events into bins of at least five events each, this method begins to require bigger and bigger boxes as the dimensionality of the space increases, so that one is essentially forced to throw away the information contained in the precision if the measurements. In eight dimensions for example, if the range of each variable is divided into only two parts, one has already $2^8 = 256$ bins.

If it is desired to divide each variable into 20 regions, then in only three dimensions this makes 8000 bins. It is therefore clear that any method with binning requiring a minimum number of events per bin will be limited to two or possibly three dimensions in practical cases.

The empty cell test also uses binning but does not require so many events per bin, and so will be applicable to spaces of somewhat higher dimensionality, but it is not so powerful. In particular, its power becomes zero in the limit where there is never more than one event per bin.

It is possible that tests based on the sample distribution function, such as the Kolmogorov test, may be extendable to many dimensions, but as yet no satisfactory formulation of such an extension is known to the author.

One is tempted to look to the methods of parametric estimation which are good in many dimensions in order to find a suitable goodness-of-fit test. Such a method is that of maximum likelihood. Thus we may consider the value of the likelihood function itself as a goodness-of-fit statistic. Let $f(\underline{X})$ be the probability density function over the n-dimensional $\underline{X}$-space and let us assume that no parameters have been estimated from the data. Then the likelihood function L is just the joint probability density and we may calculate for N events in the space Ω

$$\ln L = \sum_{I=1}^{N} \ln f(\underline{X}_i)$$

$$E_X(\ln L) = N \int_{\Omega} \ln f(\underline{X})\, f(\underline{X})\, dx$$

$$V_X(\ln L) = N \int_{\Omega} \left[\ln f(\underline{X}) - \frac{E_X(\ln L)}{N} \right]^2 f(\underline{X})\, dx \ .$$

We may then compare the observed value of the log likelihood with its expectation and variance as a goodness-of-fit test. Two major difficulties arise:

i) If parameters have been estimated from the data it is not straightforward to calculate the expectation and variance of ln L.

ii) The test may not be very powerful. In particular, if we make a transformation of variables $\underline{X} \rightarrow \underline{X}'(X)$ such that $f(\underline{X}')$ is constant, V(ln L) becomes zero, and the value of ln L becomes independent of the observations, making it in this case a useless test.

SELECTED BIBLIOGRAPHY

Some good general books on statistics:

H. Cramer, Mathematical Methods of Statistics (Princeton Univ. Press, Princeton, N.J., 1946).

H.G. Kendall and A. Stuart, The Advanced Theory of Statistics (Charles Griffin and Co., London, 1961), Vols. I-III.

S. Wilks, Mathematical Statistics (John Wiley and Sons, N.Y., 1962).

General Monte Carlo and random numbers, with bibliography:

J.M. Hammersley and D.C. Handscomb, Monte Carlo Methods (Methuen, London, 1964).

J.H. Halton, SIAM Review 12, 1 (1970).

Quasi-Monte Carlo methods:

S.K. Zaremba, SIAM Review 10, 303 (1968).

Quasi-random numbers (Van der Corput method):

J.H. Halton, Numerische Mathematik 2, 84 (1960).

Numerical integration:

P.J. Davis and P. Rabinowitz, Numerical Integration (Blaisdell, Waltham, Mass., 1967).

C.B. Haselgrove, Math. Comp. 15, 323 (1960).

A.H. Stroud and D. Secrest, Gaussian quadrature formulas (Prentice Hall Inc., Englewood Cliffs, N.J., 1966).

J.N. Lyness, Math. Comp. 19, 625 (1965).

Application to high-energy physics:

F. James, Monte Carlo Phase Space, CERN Report 68-15 (1968).