9. Estimation of Particle Transmission by Random Sampling

By Herman Kahn and T. E. Harris

We are here concerned with essentially the same problem that some of the other speakers have spoken about. We wish to estimate the probability that a particle is transmitted through a shield, when this probability is of the order of 10^{-6} to 10^{-10}, and we wish to do this by sampling about a thousand life histories. It's clear that a straightforward approach will not give the results desired.

Essentially there are three techniques that will enable one to attack most of the problems that arise. The first one we will discuss is the splitting technique, mentioned by Dr. von Neumann. In one method of applying this, one definies regions of importance in the space being studied, and, when the sampled particle goes from a less important to a more important region, it is split into two independent particles, each one-half the weight of the original. If you go from a more important to a less important region, you double the weight of the sampled neutron but play a game of chance, with probability one-half of winning to decide if the history is to be continued. The purpose of this is to spend most of the time studying the important rather than the typical particles, but to do it in unbiased fashion.

This immediately raises the question: Why wait for chance to put the neutrons into the more important region? Why not sample in such a fashion that the number of neutrons entering any region of phase space is proportional to the importance of that region. More exactly it will turn out that we will want to choose samples proportional in size to the product of the probability of getting into any region times the importance of that region. It is clear that in order to keep the estimate unbiased something will have to be done to correct for the effect of sampling from the "wrong" distribution function. It turns out that this can be done by carrying along a weighting factor. For convenience, we refer to this second type of sampling where the neutrons are forced into the more important regions, as importance sampling.

The third technique is also close to a suggestion of Dr. von Neumann's. Analytical or numerical techniques can be combined with the random sampling. That part of the problem that is hard to do analytically is done by sampling and that part that is easy to do analytically, is so done. In this way, the statistical fluctuations in the sampling of the easy parts of the problem are eliminated. It turns out, for example, that in some problems, the hard part to calculate is the histories in angle and energy, and most of the fluctuations of the ordinary sampling techniques are due to the history in space. A spatial history can be treated to some extent by purely analytical means, with but a small increase in the difficulty of the computation. Therefore it is often profitable to do this.

I suspect that most of the audience—at least, most of the statisticians and mathematicians—are somewhat unfamiliar with this neutron shielding problem. Therefore before we go on to the problem of calculating shields, let us see how these three techniques would apply, for example, to the numerical evaluation of a double integral. If we do this, I think that the principles involved will be clear.

Let us consider a random variable, Z, which is a function of the random variables X and Y. The problem is to calculate the average value, $\bar{z}$ where

$$\bar{z}=\int\int z(x,y)\,f(x,y)\,dxdy.$$

$f(x,y)\,dxdy$ is a distribution function, giving the joint probability that X lies between x and $x+dx$ and that Y lies between y and $y+dy$. The straightforward technique would correspond to an estimate of the following sort:

$$\tilde{z}=\frac{1}{N}\sum_{i=1}^{N} z(x_i,y_i),$$

where the x and the y are picked out of the population described by $f(x,y)$ and $\tilde{z}$ denotes an estimate of $\bar{z}$.

The next type of estimate (what we have already referred to as importance sampling) would correspond to something like this: We represent $\bar{z}$ in the form

$$\bar{z}=\int\int [z(x,y)\,f(x,y)/f^*(x,y)]\,f^*(x,y)\,dxdy,$$

where $f^*(x,y)$ function is an arbitrary nonzero probability density function. In this way $\bar{z}$ is represented as the expected value of $z\,f/f^*$ with respect to f^*, instead of the expected value of z with respect to f. Values of x and y are picked out of f^* at random and an estimate of $\bar{z}$ is

$$z'=\frac{1}{N}\sum z(x_i,y_i)\,f(x_i,y_i)/f^*(x_i,y_i).$$

The most interesting thing about this last type of estimate is that if f^* is properly selected, it is possible to get, what we may call, a zero-variance estimate; that is, a sample of one always gives the correct value. This is done by taking $f^*=zf/\bar{z}$. In order to do this it is necessary to know the value of $\bar{z}$ in advance. Of course the value of $\bar{z}$ would not be known in advance but it might be approached sequentially in a series of experiments. In other words one would first perform the experiment using an f^* as close as possible to the optimum f^*. The initial selection of f^* would be made on the basis of physical intuition and whatever other information one had about the problem. The knowledge gained from the first sample would then help in selecting a better f^* for the second step and so on.

The third type of sampling can be applied when it is easy to find analytically the expected value of $Z(X,Y)$ for a fixed value $X=x$. One can select N values of X from the marginal distribution of x. The estimate of $\bar{z}$ is then

$$z''=\frac{1}{N}\sum_{i=1}^{N} E(z/x_i),$$

where $E(z/x_i)$ represents the expected value of z, given that $X=x_i$.

In addition to these techniques, which are all known to statisticians, various types of stratified sampling can also be utilized.

Let us now consider an actual physical problem. A beam of particles impinges on the one face of a slab of thickness a. The progress of an individual particle through the slab is a random walk. A particle traveling through the slab has a probability μds of having a collision in an interval ds, where μ is a function of the energy of the particle and of the constitution of the slab. When a collision takes place the particle may be absorbed or it may leave the point of collison with a different direction and energy. Thus a particle that enters the slab will eventually be transmitted, absorbed in the slab, or reflected. We wish to know what fraction of the original particles will be transmitted through the slab and how much energy they will carry with them.

First let us discuss the splitting technique. We shall consider here the time-independent case, where there is a steady flow of neutrons across the slab. The concentration of neutrons at the point in phase space devoted by (x, α, γ) is given by a function $\psi\ (x, \alpha, \gamma)$ where x is the position coordinate, α the energy, and γ the cosine of the angle with the normal to the slab. We define a function $\phi(x,\alpha,\gamma)$ such that $\phi(x,\alpha,\gamma)$ is the probability that a particle (x,α,γ) will eventually be transmitted. The function ϕ represents the importance of the region at (x,α,γ). Finally, there is a transition function $f(x,\alpha,\gamma;x',\alpha',\gamma')$, which measures the rate at which particles flow from (x,α,γ) to (x',α',γ').

The splitting technique, performed in optimum fashion, would be somewhat as follows. A set of surfaces would be defined by the equations:

$$\phi(x,\alpha,\gamma)=2^{-\sigma},$$

where σ has values running from one to some integer n. In other words, surfaces of constant importance are used rather than surfaces of fixed values of x. Whenever a particle passes from a less important to a more important region, it is split in two. Each of the resulting particles is given one-half the weight of the original particle and is treated independently from then on. When the passage is from the more important to a less important region, a game of chance is played in which the particle has a 50-percent chance of being eliminated and a 50-percent chance of being allowed to survive with double its original weight. It is easy to see that all the particles that get through will have the weight 2^{-n} and if N histories are started originally and k particles get through then an estimate of the probability of transmission is $(k/N)\ 2^{-n}$. It is, however, a little difficult to discuss the variance of this estimate. However, if we modify the von Neumann technique in a mild way it is possible to discuss the problem completely. In this modification the splitting operation is performed on a particle going from a region A into a more important region B only if no ancestor of this particle has previously made a transit from A to B. No operations will be performed on a particle going from a more important to a less important region. We then have a simple branching process of a well-known type. It then turns out that for attenuations of the order of 10^{-10}, about 2,000 life histories are required, to give a 10-percent probable error.

The surfaces $\phi=2^{-\sigma}$, rather than $\phi=a^{-\sigma}$ for an arbitrary integer a, are selected because $a=2$ minimizes the variance of the estimate among all integers a. Unfortunately, it is difficult to work with surfaces in x, α, and γ. It is much less difficult to consider surfaces in x and α. Accordingly, it appears more practical to use a function $\bar{\phi}(a, \alpha)$, where γ has been averaged out in some fashion. It should, of course, be clear that in actuality we do not know $\phi(x, \alpha, \gamma)$ but simply guess or approximate it in any calculation.

We now turn to the second method, corresponding to the use of the distribution $f^*(x,y)$ discussed previously. Suppose that a particle history is being traced and that a collision has just occurred at x_i, the energy and direction following the collision being α_i and γ_i.

The quantities x_{i+1}, α_{i+1}, γ_{i+1} would, in the natural process, be selected according to a certain distribution $f(x_{i+1}, \alpha_{i+1}, \gamma_{i+1}|x_i, \alpha_i, \gamma_i)$. Instead, we may select x_{i+1}, α_{i+1}, γ_{i+1} from the distribution

$$f^*(x_{i+1}, \alpha_{i+1}, \gamma_{i+1}|x_i, \alpha_i, \gamma_i),$$

and multiply our estimate by the compensating factor f/f^*. If the function f^* is perfectly selected,

at each step, we can again obtain a zero-variance estimate, but this can be done only if the answer to the problem is known. If we are interested only in the probability that a particle penetrates the slab, and not the amount of energy transmitted, the proper choice for f^* is

$$f^*=f\phi(x_{i+1}, \alpha_{i+1}, \gamma_{i+1})/\phi(x_i, \alpha_i, \gamma_i),$$

for transitions within the slab and $f^*=f/\phi(x_i, \alpha_i, \gamma_i)$ for going through the slab. [$\phi=1$ if particle is through the slab], where as before $\phi(x, \alpha, \gamma)$ is the probability that a particle at x, α, γ will eventually be transmitted. Again, physical intuition, and preliminary sampling work, serve as a guide to a good approximation to f^*. I will not take up at this time exactly how one would go about doing this in practice, but I would like to discuss in a little more detail why this works. For simplicity let us consider a set of discrete states running from 1 to n. Let a_{ij} be the probability that a particle jumps from the i'th to the j'th state. Let P_i be the probability that a particle starts in the i'th state, and let us again consider a time independent case—that is let particles be generated continuously. Furthermore let the k'th state, and certain others, be "trap" states; once the particle enters one of these states it must disappear on the next step. If ψ_i is the expected number of particles in the i'th state then

$$\psi_j=\sum_{i=1}^{n}\psi_i a_{ij}+P_j, \tag{1}$$

or in matrix notation,

$$\psi=\psi A+P,$$

because the expected number in the j'th state must be the expected number in the i'th state times the probability of transition to the j'th state summed over all i, plus the expected number that started in the j'th state.

Similarly if ϕ_i is the probability of eventually getting to the k'th state on starting from the i'th state, then where Q_i is the probability of going directly from i to k:

$$\phi_i=\sum_j a_{ij}\phi_j+Q_{ij},$$

in matrix notation we have

$$\tilde{\phi}=A\tilde{\phi}+\tilde{Q}, \tag{2}$$

where $\tilde{\phi}$ stands for the transpose of ϕ. If the probability that we wish to estimate is denoted by T then

$$T=\sum_i \psi_i Q_i=\sum \phi_i P_i,$$

as can be seen by multiplying eq 1 by Q_j and summing and multiplying eq 2 by P_i and summing, or it can also be seen directly from the definition of the quantities involved.

To evaluate T by a straightforward technique one would pick at random a state i from the probability distribution specified by the P_i. From then on the particle is allowed to jump from the i to the j'th state according to the transition probabilities a_{ij} and so on. If the particle reaches the k'th state it is given a score of 1. If it reaches some other trap state first it is given a score of 0. The total score divided by the total number of particles stated is then an estimate of T.

It is however also possible to estimate T by working with eq 2, the so-called adjoint equation. In this case the distribution from which the initial particle is picked is given by the Q_i. However, since the summation of the Q_i over i is not necessarily one the initial particle is given a weight equal to this summation. In this adjoint problem the probability of transition from the i to the j state is proportional to a_{ji} rather than a_{ij}. It is also given a chance of dying proportional to P_i. It is proportional rather than equal because in general

$$F_i=P_i+\sum_{j=1}^{n}a_{ji}$$

is not equal to one so that the a_{ji} must be divided by F_i as a normalizing factor. To keep things unbiased, the weight of the particle must be multiplied by the same F_i. The total weight of the dead particles divided by the number of particles tried is again an estimate of T. It is not actually necessary to carry along weighting factors as the particle can be allowed to split into r independent particples subject to the condition that if t_{ir} is the probability that the particle in the i'th state splits into r particles, then

$$\sum_{r=0}^{J} r t_{ir}=F_i.$$

This equivalence between the two problems immediately raises the question if one would not be able to use information from one of the problems to devise a more efficient sampling method for the other. To see that this is so, let us consider how we would proceed if we actually had a solution of eq 2, that is, if we know the ϕ_i. Instead of sampling from the P_i initially we would sample from $P_i\phi_i/T$. That this is properly normalized is obvious since $\Sigma P_i\ \phi_i=T$. This corresponds to sampling from an $f^*(x)$ instead of an $f(x)$ and means we have to multiply by the weighting factor f/f^* or T/ϕ_i. Similarly, the probability of going from i to j will be taken as $a^*_{ij}=a_{ij}\phi_j/i$, and the weighting factor is now ϕ_j/ϕ_i. The probability of dying will now be Q_i/ϕ_i. That these probabilities are properly normalized is guaranteed by the fact that ϕ_i is a solution of eq 2.

The thing to notice is that if we started in the i'th state and go to the n'th state through an inter-

mediate path involving the l'th and m'th state, the weight of the particle is independent of the intervening path for

$$\frac{T}{\phi_i}\times\frac{\phi_i}{\phi_e}\times\frac{\phi_e}{\phi_m}\times\frac{\phi_m}{\phi_n}=\frac{T}{\phi_n},$$

which is the same weight the particle would have had if it had started in the n'th state. Since $\phi_k=1$ the weight of the particle when it dies from the n'th state is $\frac{T}{\phi_m}\times\frac{\phi_m}{\phi_k}=\frac{T}{\phi_k}=T$, which is just the number we are trying to compute. In other words, even though the particles are doing a random walk, their final weight is not random, but deterministic, and we really haven't got a random process at all.

In actual practice one does not actually have a solution to the adjoint problem. However, it may be possible to get an approximate solution, and this approximate solution can be used to modify the sampling in the way described, to decrease the fluctuations in the estimate of the answer.

The close relation between the original and adjoint process also suggests a type of alternate sequential sampling, where the quantities corresponding to p_i are estimated by sampling from the adjoint process and then are used in sampling from the original process. It is not known to what extent this is computationally feasible.

The general importance of the adjoint process has been recognized by von Neumann and others.

We shall illustrate only one of many possibilities for the third type of sampling. Let $F(x, \alpha, \gamma)$ be the probability that a particle in the state (x, α, γ) in the slab will be transmitted without further collision. The function F can in general be determined exactly without difficulty. Then if $(x_r, \alpha_r, \gamma_r)$ represents the state of the particle after the r'th collision, we may give a score of $\sum_r F(x_r, \alpha_r, \gamma_r)$ to each particle, the sum being taken over the initial point and all collision points. The average score for a number of particles is then an estimate of the probability of transmission. The advantage of such an estimate is that even particles that are not transmitted give some information about the probability of transmission. This estimate can be improved by not actually picking the first point of collision but simply calling it x, integrating the $\sum_{r=1}^{n} F(x_r, \alpha_r, \gamma_r)$ over all x such that the particle stays within the slab, and then adding up the results for different n. This type of sampling seems to be most useful when there is no good system of importance sampling available or the number of collisions that one must follow is comparatively small.