

CONDITIONAL MONTE CARLO FOR NORMAL SAMPLES⁺

Hale F. Trotter and John W. Tukey

Princeton University

The techniques presented here represent what sometimes happens to lazy people who start doing a computation before they quite know what they are going to do. Contrary to experience and high moral principles, this time it worked out all right. We solved our problem (cp. "Monte Carlo techniques in a complex problem about normal samples" pp. 80 ff., below) and we were able to extract some relatively general techniques from what we were driven to do in attempting to solve a particular problem.

2. We have not tried to express our results in the greatest possible generality, suspecting that there may be enough difficulty in coming to grips with them in a relatively special case -- however, we shall say a word or two about greater generality at the end of the account. In very rough terms we seek to use a family of transformations to convert given samples into samples conditioned on a given characteristic. We find that we can do this legitimately, and can even reuse the same samples when conditioning to another value. This seems unlikely and of doubtful legitimacy, but adequate arguments can be given for its validity and wisdom.

MONTE CARLO IN GENERAL

3. We shall begin by talking about Monte Carlo in general. One of our neighbors in the audience was asking us about the previous paper -- asking whether it was an example of Monte Carlo, or of synthetic (experimental) sampling, or whether Monte Carlo had taken synthetic sampling over. Our answer was that we thought that the last paper was experimental sampling -- that we should not like to call it Monte Carlo unless the sampling was a bit trickier. Leaving aside the important ways of using ratio and regression estimates (and covariance in general), the way in which sampling is made trickier is by the use of weights. (In the simplest situations, the weights are, of course, inversely proportional to the factor by which the probability of occurrence has been distorted.)

⁺Prepared in connection with research sponsored by the Office of Naval Research.

4. As long as we stick to the simpler situations of Monte Carlo -- and we will talk (both here and in the later paper) only about sampling from a very simple distribution -- the one essential point is that any sample can come from any distribution. The particular distribution with which we are concerned is reflected, sometimes dimly, in the weights to be attached to the samples. A table of random samples from the unit normal distribution can be used as a table of random samples from any distribution (over real numbers) whatever -- if we use the correct weights. The whole trick is to generate samples and weights which match them in such a way as to get efficient results in a manageable manner. We must, therefore, study processes which provide us pairs, say (y, w) where y is the number, sample, neutron path, or what not in whose properties we are interested, and w is the corresponding weight.

5. A process generating pairs (y, w) , w real, is said to produce weighted samples from a given distribution if two conditions are satisfied:

- (a) the pairs (y, w) are samples from some joint distribution, and
- (b) for all reasonable real-valued functions $\phi(y)$ we have the equality:

$$\text{ave} [w \phi(y) \mid \text{process}] = \text{ave} [\phi(z) \mid \text{distribution}]$$

6. Almost all Monte Carlo techniques are devoted to finding suitable processes yielding weighted samples, and then using the results of the process to estimate the average of one (or more) interesting function(s) of z , where z has a prescribed distribution. As noted above, the values of z (and y) may be numbers, vectors, samples of numbers, samples of vectors, neutron paths or what have you. The two conditions above are of very different nature and importance. Condition (b) states that we get the right answer on the average. Condition (a) says something about the convergence of means toward this average -- saying that means behave like (exactly like!) means of random samples. We are interested in learning ways to make new processes out of old ones, retaining condition (b) as we do this. In some of our heuristic proofs, we shall go through certain intermediate processes which satisfy (b) but not (a). This failure of (a) will be unimportant, since the necessary

convergence can be established by arguments which we shall omit. Notice that condition (a) is always fulfilled in two limiting situations

- (1) the x are a sample, and $w = w(x)$ depends on x alone
- (2) the x are a sample, w is independent of x , the w are a sample (in particular, are constant).

7. If we have N weighted samples $(y_1, w_1), (y_2, w_2), \dots, (y_N, w_N)$ from some distribution and are interested in $\text{ave} [\phi(z) | \text{distribution}]$ we can estimate the average either by

$$\frac{\sum w_i \phi(y_i)}{N} \quad \text{or by} \quad \frac{\sum w_i \phi(y_i)}{\sum w_i}$$

Experience shows that the former is almost always better than the latter (as well as being unbiased).

8. The equivalence for large samples of these two estimates follows from 5(b) by taking $\phi(z) = 1$, whence

$$\text{ave} [w | \text{process}] = 1.$$

Initially we didn't realize how important it was to distinguish these two estimates -- but experience showed the advantage of dividing by N , the average total weight, rather than by the realized total weight, $\sum w_i$. Indeed, it is possible to use

$$\frac{\sum w_i \phi(y_i)}{\lambda N + (1 - \lambda) \sum w_i}$$

for any real λ , and there is some ground for anticipating that λ 's somewhat larger than unity will often be best. (We don't know of any practical experience with such more general estimates.)

MONTE CARLO FOR STATISTICIANS

9. Today we are interested in Monte Carlo to solve problems which come up in statistics. There are many places where the statistician can properly say "I think I know how to analyze the data in principle, but how can I find the critical numbers?" For this reason, statisticians are going to use more and more Monte Carlo as time goes on. Because their main interest is in critical numbers (critical values), their Monte Carlo problems tend to be simpler than those of some others. As statisticians we are likely to need the solution of some such relation as

$$\text{Prob (something } \geq h) = 5\%$$

or

$$\text{Prob (something } \geq h) = 1\%$$

for h , where "something" is likely to be a rather complex function of a number, sample of numbers, etc. which has a given distribution. It is not quite obvious that such problems are asymptotically of the form

$$\text{find } \text{ave} [\phi(z) \mid \text{distribution}]$$

but if we give a couple of wisely chosen values to h , and for each put

$$\phi(z) = \begin{cases} 0, & \text{something } < h, \\ 1, & \text{something } \geq h, \end{cases}$$

we can show the asymptotic equivalence.

10. In dealing with such functions $\phi(z)$, where $\phi^2 = \phi$, we find it relatively easy to specify what are good Monte Carlo processes. The statistician often works with the variance, defined by

$$\text{var } (u) = \text{ave } (u^2) - [\text{ave } (u)]^2$$

as his measure of nonconstancy. We have

$$\begin{aligned} \text{var } \left[\frac{\sum w_i \phi(y_i)}{N} \right] &= \frac{1}{N} \text{var } [w \phi(y)] \\ &= \frac{1}{N} \left[\text{ave } [w^2 \phi(y)] - [\text{ave } [w \phi(y)]]^2 \right] \end{aligned}$$

where we have made use of 5(a) and have written $w^2 \phi(y)$ for $w^2 \phi^2(y)$ as we may since $\phi^2 = \phi$. If now we observe that $\text{ave} [w \phi(y)] = \text{ave} [\phi(z) | \text{distribution}]$ is exactly what we are trying to find and is therefore constant, we see that minimizing the variance of estimate is merely minimizing $\text{ave} [w^2 \phi(y)]$. If we call those y 's for which $\phi(y) = 1$ the exceptional set, then we are to try to make w^2 small for exceptional y 's. In most cases we do very well with the rule of thumb that the largest w for an exceptional y should be kept as small as possible. (The w 's for non-exceptional y 's are multiplied by $\phi(y) = 0$ and are of no account.) This helps to simplify Monte Carlo for statisticians. (Indeed, we give an important part of the picture if we say that small weights in the exceptional sets must be reflected in many samples in the exceptional set, while having many samples actually involved in the computation tends to reduce the effect of sampling fluctuations. All this is true, but it is only part of the story.)

CONDITIONAL MONTE CARLO

11. All of the group here today should be quite clear that the only good Monte Carlos are dead Monte Carlos -- the Monte Carlos we don't have to do. In other words, good Monte Carlo ducks chance processes as much as possible. In particular, if the last step of the process we are studying is a probability of reaction, it is wasteful, often very wasteful, to force a "yes" or "no" with a random number instead of accepting the numerical value of the probability of reaction and averaging this numerical value. The use of conditional Monte Carlo is, in a certain sense, the converse of this process, since it leads to integration over the first coordinate instead of over the last. As a way of reducing the sway of chance processes, it is almost always useful and sometimes exceedingly useful.

12. We are concerned with Monte Carlo for the statistical problems associated with samples from a normal distribution. Such problems involving "plus," "minus," "times," and "divided by" have a reasonably good chance of being solved by analytic means and avoiding Monte Carlo. It is the problems which involve "max," "min," "or" and "and" which have given rise to difficulty. These are the connectives which will drive the statistician to Monte Carlo. And for these connectives, conditional distributions are particularly advantageous. For example, if "or" means "either or both," we have:

$$\text{Prob} [A(z) \geq a \text{ or } B(z) \geq b] = \int_{-\infty}^a \text{Prob} [B(z) \geq b | A(z) = t] dF_A(t) + \text{Prob} [A(z) \geq a].$$

where $\text{Prob} [B(z) \geq b | A(z) = t]$ signifies the probability that $B(z) \geq b$ under the condition that $A(z) = t$, and $F_A(t) = \text{Prob} [A(z) \leq t]$ is the cumulative distribution of $A(z)$.

This example is not unrelated to our general class of problems and the form of the answer is particularly helpful. Often we can calculate the last term analytically, and if it is larger than the integrated term we have notably reduced the size of the probability to be evaluated by Monte Carlo. This automatically improves the sampling situation.

13. We remark that, in many problems, the natural conditions are conditions of scale, so that we may choose $A(z)$ to be homogeneous of the first degree in the elements of the sample. Indeed the two most prominent choices are the range (highest minus lowest) of the sample and alternatively, the square root of the sum of deviations. Both of these are invariant under translation, a property that we shall also find convenient.

NORMAL SAMPLES

14. We shall need some notation, and shall begin by writing $x = (x_1, x_2, \dots, x_m)$ for the sample and its elements. Multiplication by a real number proceeds coordinate-wise, so that $\alpha x = (\alpha x_1, \alpha x_2, \dots, \alpha x_m)$.

15. Some abbreviations for hypotheses are useful. We shall use:

NID (0, 1) = randomly normally and independently distributed with mean zero and unit variance,

NID' (0, 1) = same modulo translations,

ONID (0, 1) = like NID (0, 1) except that entries in each sample are rearranged to run from - to +,

ONID' (0, 1) = ordered NID' (0, 1).

Here by "modulo translations" we mean that (x_1, x_2, \dots, x_m) and $(x_1 + h, x_2 + h, \dots, x_m + h)$ are identified (treated as equivalent) for all values of h . (This may, if we wish, be described for special purposes as going from (x_1, x_2, \dots, x_m) to $(x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_m - \bar{x})$, where $\bar{x} = \sum x_j$, or in terms of any other special representation.) Many problems have the corresponding invariance under translation.

16. If we really accepted the idea that a sample from one distribution is a sample from any distribution (if appropriately weighted) then we should not be surprised at the next two results stated below. It requires only a look at the probability densities to determine the factor by which the weights are altered. The case $w(x) = 1$ will account for most of the uses of these results, but there is no difficulty in including the general case.

17. If x is NID $(0, 1)$ with weight $w(x)$, and if α is either fixed, or distributed independently of x , then $y = \alpha x$ is NID $(0, 1)$ with weight

$$w(\alpha, x) = w(x) \alpha^m e^{-\frac{1}{2}(\alpha^2 - 1)\sum x_i^2}$$

18. If x is NID' $(0, 1)$ with weight $w(x)$, and if α is either fixed, or distributed independently of x , then $y = \alpha x$ is NID' $(0, 1)$ with weight

$$w(\alpha, x) = w(x) \alpha^{m-1} e^{-\frac{1}{2}(\alpha^2 - 1) S(x)}$$

$$\text{where } S(x) = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{1}{m} (\sum x_i)^2.$$

19. In both of the last statements, the dependence of $w(x)$ on x need not be definite, the real condition is as in 5(a) that (x, w) be a sample from a joint distribution with the cases $w = 1$ as a sure function of x and $w = 1$ as special cases. Exactly similar results apply when NID or NID' is replaced by ONID or ONID'.

SKULLDUGGERY

20. We come now to the central results -- results which at first glance will seem unintuitive, improper and obviously wrong. All these discomforts are associated with the appearance of a new, additional weight function which has nothing at all to do with the way in which our samples were drawn. It is clear that there should be no place for such an additional weight function, yet it has a place, and in the next section we shall try to explain what this is.

21. We now introduce an arbitrary additional weight function $\mu(\alpha)$ with

(a) $\mu(\alpha) \geq 0$ and

(b) $\int \mu(\alpha) d\alpha/\alpha = 1,$

where the integration is over nonnegative α , and a homogeneous real-valued function $A(x)$ which is

(a) almost never zero and

(b) positively homogeneous of the first degree in the components of the sample, so that $A(\alpha x) = \alpha A(x)$.

We shall be interested in obtaining, from arbitrary samples, weighted normal samples conditioned on $A(y) = R$ a given constant. In view of our build-up it should be clear that we are going to do this by placing $y = \alpha x$ and choosing α separately for each x so that $A(y) = A(\alpha x) = \alpha A(x)$ has the correct value. Choosing α is easy, but do there exist any weights that will serve and what are they?

22. If $y = \alpha x$ is, for each fixed α , a sample from the given distribution with weight $w(\alpha, x)$, then $y = \alpha x$, with $\alpha = \alpha(x, R)$ varying from one x to another so that $A(\alpha x) = R$, R fixed, in a sample from the same distribution conditioned on $A(y) = R$ with weight

$$\frac{1}{K} \mu(\alpha) w(\alpha, x) = \frac{1}{K(R)} \mu[R/A(x)] w[R/A(x), x]$$

where

$$K = R \frac{d \text{ Prob } [A(x) < R]}{dR}$$

23. We shall try to explain the possibility of this result before we apply it.

PARTIAL EXPLANATION

24. Our explanation will take the form of an analogy. We discuss the case where we are dealing with (u, v) , v real, which is a sample from a given distribution with weight $w = w(u, v)$, and with its translations $(u, v + \gamma)$ which, for γ fixed, are samples from the same given distribution with weight $w(\gamma, u, v)$. We shall be concerned with this given distribution conditioned on $v = 0$. This version has two advantages. On the one hand, if we take u to be a scalar we can draw illuminating pictures (and we urge the reader to draw them). On the other, if we take u to be a vector, we can derive the result of 21. So this form is unusually convenient.

25. In principle, and heuristically, we are quite clear about how to find the distribution of (u, v) conditioned on $v = 0$. (Our language will now refer to the case where u is a scalar, although the arguments will be of general applicability.) We have only to draw the lines $v = \pm \epsilon$, to take a very large sample of pairs (u, v) , and to record the set of u 's among them for which $-\epsilon < v < \epsilon$. Then we repeat this with a smaller value for ϵ and an even larger sample. In the limit, the sets of u 's we obtain converge to the conditional distribution of u with $v = 0$. (The appearance of weights makes no essential change in this procedure.) We shall follow this process with some inversions of limiting processes. In particular, we shall fix ϵ until further notice.

26. For each fixed γ , $(u, v + \gamma)$ is a sample from the desired distribution with weight $w(\gamma, u, v)$. If we give γ not 1 single fixed value for each and every (u, v) , but instead 2 fixed values, 3 fixed values, ... for each and every (u, v) what we obtain will be a weighted sample for the desired distribution in a sense which is an inessential generalization of 5. The generalization lies in the fact that 5(a) fails through the $(u + v + \gamma)$'s appearing in 2's, 3's, ... with fixed spacings along the v -axis, all else being as in a sample. This generalization does not affect

anything material. As far as 5(b) is concerned, if the values of γ are γ_p , $p = 1, 2, \dots, P$ we have

$$\frac{\sum_i \sum_p w(\gamma_p, u_i, v_i) \phi(u_i, v_i + \gamma_p)}{NP} =$$

$$\frac{1}{P} \sum_p \frac{\sum_i w(\gamma_p, u_i, v_i) \phi(u_i, v_i + \gamma_p)}{N}$$

and hence the average of the left hand side is the mean of the averages of the P terms on the right. But each of these P averages is known to have the desired value, so the average of the left hand side also has the same value, and 5(b) holds.

27. If we can give γ 1 value, or 2 values, or 137 values or 5,937,263 values, there is nothing heuristic which prevents us from giving γ a distribution, say $\nu(\gamma)d\gamma$. Now every pair (u, v) has many translates $(u, v + \gamma)$ which fall in the magic strip $-\epsilon \leq v \leq +\epsilon$. Neglecting terms of higher order there will be a number proportional to

$$\int_{-v-\epsilon}^{-v+\epsilon} \nu(\gamma) d\gamma \sim 2\epsilon \nu(-v)$$

each with weight

$$w(\gamma, u, v) \sim w(-v, u, v)$$

and the total weight corresponding to (u_i, v_i) is proportional to the product

$$2\epsilon \nu(-v_i) w(-v_i, u_i, v_i).$$

Going now to the limit with ϵ , we find that $(u_i, 0)$ with weight

$$\frac{1}{K} \nu(-v_i) w(-v_i, u_i, v_i)$$

is behaving as a weighted sample from the desired distribution conditioned on $v = 0$.

28. In this analog, we can see how the arbitrary distribution $\nu(\gamma)d\gamma$ enters. It is the result of a perfectly legitimate choice, and arises quite naturally. Indeed, we see that it need not be constant, for if we replace $\nu(\gamma)$ by $\nu(\gamma, u) \geq 0$ with $\int \nu(\gamma, u) d\gamma = 1$ for every u , we shall be able to carry through the same argument.

29. To obtain the form given in 22, we have only (a) to determine K and (b) to place

$$v = \log A(x) - \log R,$$

$$u = (A(x))^{-1} x.$$

Thus there is reason to 22 after all.

FIRST APPLICATIONS

30. We can now go back and combine 22 with 17 and 18. The results are interesting and powerful.

31. If x is NID(0, 1) with weight $w(x)$, then $y = \alpha x$, with $A(y) = A(\alpha x) = R$ determining $\alpha = \alpha(x, R)$, is NID(0, 1) conditional on $A(y) = R$ with weight

$$\frac{1}{K} \mu(\alpha) w(x) \alpha^m e^{-\frac{1}{2}(\alpha^2 - 1) \sum x_i^2}$$

32. If x is NID'(0, 1) with weight $w(x)$ and $A(z)$ is invariant under translation, then $y = \alpha x$ with $A(y) = R$ is NID'(0, 1) conditional on $A(y) = R$ with weight

$$\frac{1}{K} \mu(\alpha) w(x) \alpha^{m-1} e^{-\frac{1}{2}(\alpha^2 - 1) S(x)}$$

33. Similar statements also apply with ONID replacing NID and with ONID' replacing NID'.

34. These results mean, in particular, that if we start out with 1,000 samples of m from a normal distribution, we can rescale each sample so that its range is $\pi = 3.14159\dots$,

and provided we use the right weights, we can use the 1,000 rescaled samples as samples from a unit normal distribution conditioned by "range = π ". And so on.

35. In particular, the following choice for μ is admissible for any $k > 0$:

$$\mu(\alpha) = \frac{2[kR/\sqrt{2}]^{m-1} e^{-\frac{1}{2}k^2 R^2} \alpha^{1-m} e^{\frac{1}{2}k^2(\alpha^2 - 1) A^2(x)}}{\Gamma\left[\frac{m-1}{2}\right]}$$

$$= \frac{2[kR/\sqrt{2}]^{m-1} e^{-\frac{1}{2}k^2 R^2} \alpha^{1-m} e^{\frac{1}{2}k^2(\alpha^2 - 1) R^2/\alpha^2}}{\Gamma\left[\frac{m-1}{2}\right]}$$

Notice that although the first form looks unusable since $A(x)$ depends on x_j , the substitution $A(x) = R/\alpha$ eliminates this dependence, as we see in the second form.

36. The use of 35 in 32 gives for the weight function

$$\frac{2w(x)}{k\Gamma\left[\frac{m-1}{2}\right]} \left[kR/\sqrt{2}\right]^{m-1} e^{-\frac{1}{2}k^2 R^2} e^{-\frac{1}{2}(\alpha^2 - 1) [S(x) - k^2 A^2(x)]}$$

Here we begin to reach a specifically applicable formula -- one we shall need in the later paper.

REUSE OF SAMPLES

37. We have a method for starting with an unconditioned set of samples and converting all of them to meet the condition $A(y) = R$. We could convert the very same set of samples to meet $A(y) = R_1$, where $R_1 \neq R$. And to meet $A(y) = R_2$. And so on. For each of these values of R taken separately we

obtain a weighted sample from the corresponding conditional distribution. Each by itself is obviously all right. But when we use conditional distributions we usually need to use them for a number of values of the condition, summing or integrating over these values. If this is what we are up to, is it proper to reuse the original set of samples without change? The answer turns out to be that it is proper, but the result may not be as precise as if we drew a new sample for each condition. In terms of precision for a given amount of effort, however, we may be very much better off to reuse the single set.

38. If we return to our example (in 12) of a possible use for conditional distributions, we have

$$\text{Prob} [A(z) \geq a \text{ or } B(z) \geq b] = \int_{-\infty}^a \text{Prob}[B(z) \geq b | A(z) = t] dF_A(t) + \text{Prob}[A(z) \geq a]$$

and if we know the distribution of $A(z)$ then we have only (a) to find, or at least to estimate, the conditional probability that $B(z) \geq b$ conditional on $A(z)$ equalling suitable values, and (b) to evaluate the integral. One natural approach is to choose a moderate number of values for the condition, obtain a single set of samples, adjust each and every one of them to meet each and every condition, thereby obtaining estimates of each of the corresponding conditional probabilities, and then use some reasonable quadrature formula of numerical integration to approximately estimate the value of the integral. (As we shall see in the later paper, this works quite well.)

39. Another approach is to use each sample not just for a few conditioning values, but rather to use it for every possible conditioning value and integrate out the result. The estimate for the integral will be the integral of our conditional estimates. Each conditional estimate can be written as the mean over the finite set of samples of certain expressions. If we invert the operations of integration and finite summation, as we always may, we obtain an estimate for the integral as the mean over the finite set of samples of certain integrals. If $\phi(z)$ is always either 0 or 1, these integrals need be extended only over values of α with $\phi(\alpha x) = 1$. Assuming that this condition describes manageable sets of α 's, we obtain the results which follow.

40. If $y = \alpha x$, for each fixed α , is a sample from a given distribution with weight $w(\alpha, x)$, and if we wish to estimate

$$\text{ave } [\phi(z)] = \text{prob } [g(z) = 1]$$

then we may use

$$\frac{1}{N} \sum_{\substack{\int \\ \phi(\alpha x)=1}} \mu(\alpha) w(\alpha, x) \frac{d\alpha}{\alpha}$$

where the summation is over the N samples x and the integration is over the α for which $\phi(\alpha x) = 1$ (over the α for which αx is exceptional).

41. If x is NID (0, 1) with weight $w(x)$, and we wish to estimate

$$\text{ave } [\phi(z)] = \text{prob } [g(z) = 1]$$

we may use

$$\frac{1}{N} \sum_{\substack{\int \\ g(\alpha x)=1}} w(x) \frac{\mu(\alpha)}{\alpha} \alpha^m e^{-\frac{1}{2}(\alpha^2 - 1) \sum x_i^2} d\alpha$$

as the estimate, where $\mu(\alpha)$ is as in 21.

42. If x is NID' (0, 1) with weight $w(x)$, and $\phi(x)$ is invariant under translation, then we may estimate

$$\text{ave } [\phi(z)] = \text{prob } [g(z) = 1]$$

by means of

$$\frac{1}{N} \sum_{\substack{\int \\ g(\alpha x)=1}} w(x) \frac{\mu(\alpha)}{\alpha} \alpha^{m-1} e^{-\frac{1}{2}(\alpha^2 - 1) S(x)} d\alpha$$

where $S(x) = \sum (x_i - \bar{x})^2$ and $\mu(\alpha)$ is as in 21.

43. Under the hypotheses of 42 we may use

$$\frac{1}{N} \sum \frac{2w(x)}{\Gamma\left[\frac{m-1}{2}\right]} \left[kA(x)/\sqrt{2} \right]^{m-1} e^{-\frac{1}{2}[S(x) - k^2 A^2(x)]} \int_{g(\alpha x)=1}^{\infty} \alpha^{m-2} e^{-\frac{1}{2}\alpha^2 S(x)} d\alpha$$

as an estimate.

44. Clearly 41, 42, and 43 derive from 40 by using 31, 32, and 36. Entirely similar statements hold when NID or NID' is replaced by ONID or ONID'. The form of these estimates -- as the means of suitable functions of the separate samples -- implies a variance of estimate proportional to $\frac{1}{N}$.

Thus either the variance is infinite (which does not occur for reasonable cases) or it converges to zero. This provides a proof that this technique of infinite reuse really works. (We have not made any empirical tests to see how well it works -- however, tables of incomplete normal moment functions are all that is needed to make 43 operate.)

TROTTER'S TECHNIQUE

45. The particular problem which led to the development of these techniques involved the special case where $A(x) = \text{range } [x_i]$. A special technique applicable to this case was developed and proved unusually useful. For convenience we shall say that (x_1, x_2, \dots, x_k) are UID $[0, 1]$ if they are uniformly and independently distributed on the interval $[0, 1]$ -- that is to say if they are independent random numbers (not random digits) in the usual sense of these words. We write OUID $[0, 1]$ for the ordered case. With these definitions we have the following results.

46. If a, b, \dots, j are $m - 2$ quantities OUID $[0, 1]$, then $y = Rx = (0, Ra, Rb, \dots, Rj, R)$ is a sample from ONID' $(0, 1)$ conditioned by range $[y_i] = R$ with weight

$$(2\pi)^{-(m-1)/2} R^{\frac{m-2}{2}} \frac{e^{-\frac{1}{2}SR^2}}{f(R)}$$

where $f(R)$ is the density function of the distribution of the unit normal range of m , and $S = (a^2 + b^2 + \dots + j^2 + 1) - \frac{1}{m}(a + b + \dots + j + 1)^2$.

47. Under the hypotheses of 42 and 46, we may take as an estimate

$$\frac{1}{N} \sum (2\pi S)^{-(m-1)/2} \int_{g(ux/\sqrt{S})=1}^{\infty} u^{m-2} e^{-\frac{u^2}{2}} du.$$

MORALS AND EXTENSIONS

48. If there are any morals to this development, we feel that these are the two most important ones:

- (a) there are many ways to introduce good "swindles" into Monte Carlo, and only a few have been discovered -- when in trouble look for a new one,
- (b) so long as we deal with linear problems and have choices, it is likely to be worthwhile to take our choices several different ways at once.

To (a) this account offers the example of the whole method. To (b) it offers two examples which may be less clear to the reader. On the one hand, the use of many values of α (and the selection of only the relevant one) lies back of the procedure of getting a conditional sample. On the other hand, the use of the same original sample as a contributor to estimates for several or many values of the condition is the other novel feature. Both of these may properly be thought of as examples of taking several choices simultaneously where the naive approach would be to take only one choice. Linearity is a great thing.

49. As for extensions, there are likely to be many. We have discussed only the case where the family of transformations are one-dimensional, but extensions to several dimensions undoubtedly exist. We have covered simple additive and multiplicative transformations, but the transformation

$$(x_1, x_2, x_3, x_4, x_5, x_6) \rightarrow (\alpha x_1, \alpha x_2, x_3, x_4, x_5, x_6)$$

might be useful in obtaining critical values of

$$\frac{|x_1 - x_2|}{|x_3 - x_4| + |x_5 - x_6|}$$

We leave this, and other possibilities, to the reader.

Symposium on Monte Carlo Methods

Held at the University of Florida

Conducted by the Statistical Laboratory

Sponsored by Wright Air Development Center
of the Air Research and Development Command

March 16 and 17, 1954

Edited by
HERBERT A. MEYER



New York
London

John Wiley & Sons, Inc.
Chapman & Hall, Limited

511.5
F636

COPYRIGHT © 1956

BY

JOHN WILEY & SONS, INC.

All Rights Reserved

*This book or any part thereof must not
be reproduced in any form without
the written permission of the publisher.*

COPYRIGHT, CANADA, 1956, INTERNATIONAL COPYRIGHT, 1956
JOHN WILEY & SONS, INC., PROPRIETOR

All Foreign Rights Reserved

Reproduction in whole or in part forbidden.

SECOND PRINTING, NOVEMBER, 1958

Library of Congress Catalog Card Number: 56-9825

PRINTED IN THE UNITED STATES OF AMERICA

TABLE OF CONTENTS

I. PAPERS IN ORDER OF PRESENTATION

A. W. Marshall	1. An Introductory Note	1
Olga Taussky and John Todd	2. Generation of Pseudo Random Numbers	15
N. Metropolis	3. Phase Shifts - Middle Squares - Wave Equation	29
G. E. Albert	4. A General Theory of Stochastic Estimates of the Neumann Series for the Solution of Certain Fredholm Integral Equations	37
Theodore S. Motzkin	5. Neighbor Sets for Random Walks and Difference Equations	47
Nancy M. Dismuke	6. Monte Carlo Computations	52
S. Ulam	7. Applications of Monte Carlo Methods to Tactical Games	63
Hale F. Trotter and John W. Tukey	8. Conditional Monte Carlo for Normal Samples	64
Harvey J. Arnold, Bradley D. Bucher, Hale F. Trotter and John W. Tukey	9. Monte Carlo Techniques in a Complex Problem about Normal Samples	80
Martin J. Berger	10. An Application of the Monte Carlo Method to a Problem in Gamma Ray Diffusion	89
L. A. Beach and R. B. Theus	11. Stochastic Calculations of Gamma Ray Diffusion	103
A. W. Marshall	12. Application of Multiple Stage Sampling Procedures to Monte Carlo Problems	123
John E. Walsh	13. Questionable Usefulness of Variance for Measuring Estimate Accuracy in Monte Carlo Importance Sampling Problems	141

CONTENTS		
C. W. Vickery	14. Experimental Determination of Eigenvalues and Dynamic Influence Coefficients for Complex Structures such as Airplanes	145
Herman Kahn	15. Use of Different Monte Carlo Sampling Techniques	146
J. H. Curtiss	16. A Theoretical Comparison of the Efficiencies of Two Classical Methods and a Monte Carlo Method for Computing One Component of the Solution of a Set of Linear Algebraic Equations	191
E. J. Lytle, Jr.	17. A Description of the Generation and Testing of a Set of Random Normal Deviates	<u>234</u>
James W. Butler	18. Machine Sampling from Given Probability Distributions	249
John E. Walsh	19. A Monte Carlo Technique for Obtaining Tests and Confidence Intervals for Insurance Mortality Rates	265
Alwin Walther	20. Experiments and Models for the Monte Carlo Method	278
II. BIBLIOGRAPHY		283