MONTE CARLO

Herman Kahn
Irwin Bunn

P-1165

July 30, 1957

The RAND Corporation
1700 MAIN ST • SANTA MONICA • CALIFORNIA

CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION CFSTI
DOCUMENT MANAGEMENT BRANCH 410.11

LIMITATIONS IN REPRODUCTION QUALITY

ACCESSION # AD 606529

☒ 1. WE REGRET THAT LEGIBILITY OF THIS DOCUMENT IS IN PART UNSATISFACTORY. REPRODUCTION HAS BEEN MADE FROM BEST AVAILABLE COPY.

☐ 2. A PORTION OF THE ORIGINAL DOCUMENT CONTAINS FINE DETAIL WHICH MAY MAKE READING OF PHOTOCOPY DIFFICULT.

☐ 3. THE ORIGINAL DOCUMENT CONTAINS COLOR, BUT DISTRIBUTION COPIES ARE AVAILABLE IN BLACK-AND-WHITE REPRODUCTION ONLY.

☐ 4. THE INITIAL DISTRIBUTION COPIES CONTAIN COLOR WHICH WILL BE SHOWN IN BLACK-AND-WHITE WHEN IT IS NECESSARY TO REPRINT.

☐ 5. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED, DOCUMENT WILL BE AVAILABLE IN MICROFICHE ONLY.

☐ 6. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED DOCUMENT WILL NOT BE AVAILABLE.

☐ 7. DOCUMENT IS AVAILABLE IN MICROFICHE ONLY.

☐ 8. DOCUMENT AVAILABLE ON LOAN FROM CFSTI (TT DOCUMENTS ONLY).

☐ 9.

PROCESSOR: [illegible]

TS..-107-10 64

PREFACE

This is a draft of a report which is being circulated for information and comment. We hope to make it a chapter of a book titled Military Planning In An Uncertain World, and would appreciate any comments, criticism, ideas, and examples that readers may have. This draft began as a transcript of an informal talk and, despite some rewriting, it probably still suffers (like many such talks) from being "fashionable." We are aware that it has a number of other weaknesses and assume there are still others of which we are not aware. We hope to give it a thoughtful and leisurely review but are deferring this until we get some outside criticism.

A table of contents is given on the next page to show the relation of this chapter to the rest of the book. The chapter may not be quite self-contained as a paper, as it occasionally refers to other chapters; but we trust this will be understood or overlooked.

A more complete introduction and list of acknowledgements are given in RM-1829-1.

MILITARY PLANNING IN AN UNCERTAIN WORLD

CONTENTS OF BOOK

---

[1] Has already appeared as RM-1829-1

[2] This document

[3] Has already appeared as P-1166

[4] Has already appeared as P-1167

[5] Has already appeared as RM-1937

INTRODUCTION*

A problem doesn't have to be very complicated before it gets too difficult for even a modern high-speed computer to do in a straightforward fashion. One of the powerful techniques available that will often make a seemingly intractable problem tractable, if not easy, and one that is particularly well suited to the electronic computer, is the so-called Monte Carlo method. This fashionable name has been given to any technique which uses sampling to estimate the answer to a precise mathematical problem. A game of chance is devised with the property that the average of the scores of a large number of plays of the game is the number being estimated. While the game may be played by using gambling devices such as a roulette wheel, dice, or coins, usually the simplest and most practical such device from the viewpoint of the computer is a table of random numbers. Such a table presents a strange appearance to the layman. It is nothing but a haphazard collection of the digits 0 to 9, but extreme care is taken to see that the collection is uniformly haphazard and random. By using this table it is possible to simulate the play of any mechanical gambling device.

The game of chance can be a direct analogue of the problem being studied or it can be an artificial invention. The only property that it must possess is that its average score is the answer to the desired problem.

---

*This paper was largely drawn from another paper written by one of us, "Use of Different Monte Carlo Sampling Techniques", published by Wiley and Sons Co., "Symposium on Monte Carlo Methods" 1956. The reader is referred to that paper for a more technical discussion. We are indebted to Wiley and Sons Co. for allowing us to do this.

It is not necessary to use the first game that comes to mind since it is often possible to render this estimation more efficient by making changes in the game that do not disturb its expected score. The new game may be more efficient because it is less erratic or perhaps because it is cheaper to play.

Therefore, when using the Monte Carlo technique to solve a problem, one directs attention to three main topics:

(1) choosing the probability process (picking the game of chance)

(2) generating sample values of the random variables on a given computing machine (playing the game)

(3) designing and using variance reducing techniques (modifying the game to be more efficient).

The first two of these topics were discussed in the example on the attrition of bombers in Chapter 2, pages 46-50. The reader is referred to these pages if he wishes an elementary discussion of a typical Monte Carlo problem.

This chapter will discuss only the last topic, how to increase the efficiency of a Monte Carlo calculation by proper experimental design. It is, of course, true that in actual problems one cannot isolate variance reduction from the first two topics. The methods that can be used to reduce variance are often sharply dependent upon the probability model and in some cases on the techniques used to generate values of the random variables. Also, the greatest gains in variance reduction are often made by exploiting specific details of the problem, rather than by routine application of general principles. However, there do seem to be some general ideas on reducing variance which can be used in many problems. Seven techniques seem to be most useful. They are:

1. Importance Sampling
2. Russian Roulette and Splitting
3. Use of Expected Values (combination of analytic and probabilistic methods)
4. Correlation and Regression
5. Systematic Sampling
6. Stratified Sampling
7. Specialized Techniques

While all of these techniques can be used in standard statistical sampling problems, the first three seem to have found particular and specialized usefulness in Monte Carlo applications as differentiated from the usual applications in ordinary sampling. This is mainly because of the fact that in a Monte Carlo problem the experimenter has complete control of his sampling procedure. If, for example, he were to want a green-eyed pig with curly hair and six toes and if this event had a non-zero probability, then the Monte Carlo experimenter, unlike the agriculturist, could immediately produce the animal.

In order to illustrate the general nature of the techniques, we will apply them to a very simple example--so simple, in fact, that the reader will have to exercise his imagination in order to see that there is a problem.

Consider the problem of calculating the probability of obtaining a total of three when two ordinary dice are tossed. Each die is of the standard sort with six faces labeled from one to six and constructed so that each face has the same probability (1/6) of being on top. This problem can, of course, be solved analytically. Any particular combination of the dice has a probability equal to 1/6 times 1/6 of occurring. Since there are two combinations

which make three (one-two and two-one), the probability of getting a three in a random toss of the dice is 2/36 or 1/18.

In doing the problem by Monte Carlo one could simply toss the dice[1] N times, count the number (n) of successes (threes) and then estimate the probability (p) of success by

$$\hat{p} = \frac{n}{N} \tag{1}$$

Typically, $\hat{p}$ differs from p; that is, the estimate has a statistical error. This statistical error is usually measured by the standard deviation σ.[2] In this case,

$$\sigma = \sqrt{\frac{p(1-p)}{N}} \tag{2}$$

The percent standard deviation or standard error is then given by

$$\frac{100\sigma}{p} = 100\sqrt{\frac{1-p}{Np}} \tag{3}$$

As is intuitively clear, and as is shown by formula (3), this error goes down (though not very rapidly) as the number of trials is increased. In what follows, we will illustrate other ways than increasing N by which the error can be decreased. These are very important in practice, for to

---

[1] Usually one would not toss physical dice, but simulate the tosses with the aid of a table of random numbers.

[2] In Monte Carlo problems the error has statistical properties which can usually be described in the following manner. The probability that the absolute value of the error will be larger than $m\sigma$ is given by the following table:

m	Prob.
.67	.50
1.00	.32
2.00	.05
3.00	.003
4.00	.0001

σ is called the standard deviation and in our case is given by equation (2) above. A more complete explanation of statistical errors can be found in almost any elementary book on statistics.

diminish a standard deviation ten fold by simply increasing N requires that N be increased one hundred fold. We can often get improvements of the same order by very simple changes in the sampling procedure.

1. IMPORTANCE SAMPLING

If by some method we can increase the effective value of p, equation (3) shows that even though the same N is used, the percent error will be reduced. This increase in the effective value of p can be obtained very easily. We could, for example, bias the dice so that the probability for either a one or a two would be twice as great as usual, that is 1/3 rather than 1/6. This could be done with physical dice by "loading" them, or with mathematically simulated dice by using a biased table of random numbers. If a one and two each had a probability 1/3 of occurring, then the probability of getting a three, instead of being 1/18, would be four times as great or 2/9. The percent error is then cut by slightly more than a factor of two. Of course, equation (1) can no longer be used to estimate p but

$$p = \frac{1}{4}\frac{n}{N} \qquad (4)$$

must be used instead. The 1/4 in equation (4) is called a weighting factor. By using it, the distortion introduced by the biased sampling is removed. It is vital to appreciate that even if p were difficult to compute (as we are in effect pretending) the weighting factor is still easy to compute.

This illustrates the general idea of Importance Sampling--which is to draw samples from a distribution other than the one suggested by the problem and then to carry along an appropriate weighting factor, which, when multiplied into the final results, corrects for having used the wrong distribution. The biasing is done in such a way that the probability of the sample's being drawn from an "interesting"[3] region is increased; the probability that

---

[3] The words "interesting" and "uninteresting" refer to the amount of effort or interest the sensible calculator would show in the region.

it comes from an "uninteresting" region is correspondingly decreased. The reader should verify for himself that it would be good to carry the bias illustrated in the example to its natural limit; the probability of getting a one or a two could be increased by a factor of three, making the probability of obtaining one of these numbers 1/2 and making the probability of obtaining any other number zero.

This natural limit is not the ultimate limit. For example, if we tossed the dice one at a time, then we might want to bias the second die differently from the first one. In particular, if we were willing to let the biasing of the second die depend on the outcome of the first throw, we might consider the following scheme.

1. Increase the probability of getting a one or a two on the first die by a factor of three. This means, of course, that there will be a zero probability of getting any other numbers.
2. If the first die comes up one, increase the probability of the second die coming up two by a factor of six; if the first die comes up two, increase the probability of the second die coming up one by a factor of six.

If this scheme is followed every toss of the dice will yield a three so that the number of successes (n) will be equal to the number of trials (N). The weighting factor will be 1/3 times 1/6 or 1/18 and the estimate will be

$$p = \frac{n}{18N} \qquad (5)$$

$$= \frac{1}{18}$$

which is exactly equal to p. We have devised a sampling procedure which has

zero variance. It is always possible to design an Importance Sampling scheme that has zero variance, but the computation needed to compute what the proper biasing procedure may be is exactly that needed to solve the original problem and in some cases even more.

Examples[4]

One of the major problems on which Monte Carlo is used is to calculate the probability that nuclear particles will penetrate shields. In such a problem, the particle starts at one side of the shield, and has collisions of different types with the atoms of the shield, finally being either reflected backwards, absorbed in the shield, or transmitted. The calculation can be done by Monte Carlo by simulating the particle histories with the aid of random numbers. As this idea of Importance Sampling suggests, the simulation should not be faithful. For example, the following types of random events (typically) increase the probability of penetration and should be emphasized at the expense of equally, or even more probable, but less "important," ones:

a. collisions resulting in a forward direction of motion

b. collisions that result in small energy losses

c. long forward jumps and short backward jumps (the so-called exponential transformation)

d. survival vs. absorption (if carried to the limit, this can be looked on as an application of Use of Expected Values).

---

[4]Some of the examples discussed at the end of each section may be hard for the non-professional to follow. He will find that if he skip them, the later text starts again from the beginning.

The calculator is, of course, confronted with the problem of how far to go in altering the sampling. In deciding this he must use a combination of judgment and calculation which cannot be entered into here.[5]

Another important application of Monte Carlo is in the design and analysis of reactors. Here again we are studying the various ways in which nuclear particles--particularly neutrons--behave in matter. It is found that those which wander away from the center of the reactor will not contribute much of interest to the process. On the contrary, these neutrons which wander back toward the fissionable material are the ones which contribute most to the answer. The sampling must then be designed to sample more frequently among the second type of neutrons and less frequently among the first type of neutrons.

Monte Carlo is also applied to Operations Research problems. In a typical problem of this type we might try to calculate the vulnerability of a piece of equipment or of an airfield to some offensive weapon. In such cases, one often Monte Carlos, for example, the error of the missile which is doing the destruction. The distribution of errors is determined by a parameter called the Circular Probable Error (CEP). The CEP is the radius of the circle about the aiming point into which the missile will fall 50% of the time. Since we are interested here in destruction, if we Monte Carlo from a distribution defined from a smaller CEP than that which obtains in the real world, we will find that more of the interesting processes (hits) happen in the simulated experiment than would happen naturally. The less interesting processes (misses) are then discriminated against.

---

[5] See previous reference or RAND report RM-1237 "Applications of Monte Carlo."

For another example in the same field, consider queuing problems. In these problems one is often interested in the mean and variance of the waiting time. One then wishes to bias the sampling to emphasize long waits. This could be done by sampling from new distributions that simulate increased traffic, increased servicing time, or increased servicing requirements.

It is worth noting that any set of samples obtained with the use of Importance Sampling is less effective in estimating certain auxiliary quantities than a set that has been obtained in a straightforward fashion. For instance in the shielding problem, a sampling design that leads to an accurate estimate of the probability of penetration will be very poor for estimating the probability of reflection; in the reactor problem the suggested sampling would not be good at estimating leakage; in the vulnerability problem we will lose information about light damage and the location of misses; and finally in the queuing problem we will not get a good estimate of the idle time of the servicing facilities.

It is in fact usually (but not always) true that to design an efficient Monte Carlo calculation one must direct attention to those things he is really interested in and ignore other aspects of the problem. It may even be better to do more than one calculation than to compromise the goals of any particular design. This can be a serious disadvantage if the computation is lengthy, as the decision about what aspects to concentrate on must be made early and therefore may easily be made wrongly.

2. RUSSIAN ROULETTE AND SPLITTING[6]

Let us assume that the dice are tossed one at a time and that the cost of the problem is measured by the total number of individual tosses. Now it is immediately clear that if the first die is tossed and if it happens to come up three or greater, it will be impossible to get a total of three, no matter how the second die comes up. Under these circumstances, there is no point in making the second toss and we can simply record a zero for the experiment. This makes it unnecessary to toss the second die 2/3 of the time. Therefore on the average we will do 1/3 fewer tosses in an experiment.

In more complicated examples where the sampling is done in stages, it is often possible to examine the sample at each stage and clarify it as being in some sense "interesting" or "uninteresting." The sensible calculator is willing to spend more than an average amount of work on the "interesting" ones and less on the "uninteresting" ones. This can be done by splitting the "interesting" samples into independent branches, thus getting more of them, and by killing off some percent (in the above example 100%) of the "uninteresting" ones. The first process is Splitting and the second Russian Roulette.

The "killing off" is usually done by a supplementary game of chance. If the supplementary game is lost the sample is killed; if it is won the sample is counted with an extra weight to make up for the fact that it ran some risk of being killed. The game has a certain similarity to the Russian game of chance played with revolvers and forehands--whence the name.

The idea of Russian Roulette and Splitting is similar in spirit to the

---

[6] Both the idea and the names are due to J. von Neumann and S. Ulam.

sequential sampling schemes of quality control, though quite different in detail. It was first thought of in connection with particle diffusion problems. Particles that get into interesting regions are split into n independent sub-particles, each with one n'th of the weight of the original particle. Particles that get into uninteresting regions are, in effect, amalgamated into a smaller number of heavier particles. In this way the calculator achieves his goal of allocating his effort sensibly.

Most of the examples mentioned in Importance Sampling also could be used to illustrate the use of Russian Roulette and Splitting. However, there are differences. In Importance Sampling the samples are forced into the regions and prevented from entering "uninteresting" regions. Sometimes, though, this is hard to arrange--particularly when the sampling distribution is not given explicitly, but a complicated process for getting sample values is given instead. In these circumstances we may have very good estimates of the relative "importance" of different regions and still not be able to arrange for the proper biasing. We can then use Russian Roulette and Splitting which, in effect, does the same thing ad hoc; that is, we wait to see what region is entered and then decide what the size of the sample should be.

3. USE OF EXPECTED VALUES

If the sampling is being done in two stages, then even if we aren't clever enough to calculate the combinatorics of the whole problem, we still might be clever enough to notice that there is no point in tossing the second die; that is, once the first die is tossed, it is trivially easy to calculate the probability of obtaining a total of three. For example, when the first die comes up one, the only way we can get the three total is for the second die to come up two. This event obviously has a probability of 1/6. Similarly if the first die comes up two, the only way to get three is for the second die to be one. This event also has a probability of 1/6. Finally, all the other possibilities for the first die (three to six) have a zero probability of giving three. If we record the probabilities rather than toss the second die, then it is a fact that the average of these probabilities is an estimate of p.[7] This method simultaneously reduces the number of tosses we need by a factor of two and decreases the variance, so that the tosses we do make are more effective.

The possibility of using expected values to great advantage occurs frequently in practice. The illustration is not artificial. In many probabilistic problems, much of the variance or fluctuation is introduced by a part which can be calculated analytically, while another part which is hard to calculate analytically may, in fact, not introduce much fluctuation. In these cases the sensible calculator combines analytic and probabilistic methods--calculating analytically that which is easy and Monte Carloing that

---

[7] It is probably worth while to point out that the Use of Expected Values is quite different from using an expected-value model (see Chapter 1). The former is exact and the latter is an approximation.

which is hard.

Examples

Expected values can be used in all the problems mentioned under Importance Sampling. In the shielding problem, for example, instead of counting the particles that succeed in piercing the shield, we can calculate the probability after each collision that the particle would go directly through the shield. An unbiased and improved estimate of the penetration is obtained by summing these probabilities.

Another application of the same idea is in the calculation of multiple scattering effects in nuclear cross section measurements. In this calculation one simulates a scattering experiment in which the main interest is in the number of particles that enter a detector. This probability may be very small, say of the order 10^{-4} or 10^{-6}, so that if one sampled in a straightforward fashion, there would be practically no counts. It is, however, very simple to calculate the probability that the particle will enter the detector from any point in the system. If these probabilities are recorded, rather than the number of particles counted, one usually finds that the percent variance is enormously reduced, factors of 10^{-3} being typical. The factor is larger in this case than in the previous (shielding) example because in this problem many samples tend to get into a region where the particle has a relatively high (though absolutely low) probability of getting into the detector; in the previous problem this is not true unless Importance Sampling or Russian Roulette and Splitting is also used.

This idea can be applied to more general problems than particle diffusion. For example, in vulnerability studies, it is often true that the problem being studied involves several stages and that in the final stage

there is given a probability p of achieving or not achieving a kill. It should be obvious to the reader at this point that it would be wrong to sample from this binomial distribution; rather, the p's themselves should be recorded and their average used for the estimate.

A similar situation occurs in the particle diffusion problems mentioned previously. Usually there is a non-zero probability that the particles will be absorbed after they have had a collision. This binomial process could be sampled, but it is generally more accurate not to sample, but simply to weight the particle with the probability of survival, multiplying all the probabilities together after it has had all its collisions. This increases the average length (number of collisions) of a history and therefore its cost, but this effect is usually dwarfed by the decrease in variance.

Another way to use expected values is to integrate a sample over the initial conditions. For example, in particle diffusion problems, it is not much work to translate, rotate, or reflect histories (thus getting new ones) and then average these translated, rotated, or reflected histories over their a priori probabilities.

The three techniques discussed above can be most effective in realistic applications. The authors are familiar with cases in which each technique has, by itself, decreased the effective variance by factors of the order of 10^4 to 10^6. In most cases this means changing the problem from one which cannot be done because it would be too expensive or lengthy to one which is easily done on modern computing machines or even by hand computers. However, these large reductions were all associated with physics or engineering problems. In Operation Research applications the reductions while important are much more modest. They tend to fall in the range 2-100. This is still

of value if the problem is at all lengthy.

The techniques that we are now going to discuss are, in general, not as effective as the three already mentioned. However, they often are very easy to use and may yield worthwhile improvements.

4. CORRELATION AND REGRESSION

Correlation or regression is used when we are going to make comparisons or calculate differences and wish to eliminate irrelevant fluctuations which do not affect the comparison or difference.

In order to illustrate this technique, it will be necessary to change the problem slightly. Assume, for instance, that the proprietor of one of the gaming establishments in Las Vegas wants to change the rules in force at his dice tables. Under current rules, if a player tosses a 2, 3, or 12, on the first throw of dice, the player loses. If he tosses a 7 or 11, he wins, and if he tosses a 4, 5, 6, 8, 9, or 10, he will win or lose, depending on whether or not that number or a 7 comes up first in his subsequent throws.

Now let the rule change being considered be the interchange of the roles of 3 and 4 and assume that, unlike most of the proprietors in Las Vegas, the one we are considering is unsophisticated and wants to determine by sampling what the change in his revenue will be. The obvious way to do this is to run two sets of experiments, one with the old rules and one with the new rules, and then compare the two experimentally-determined revenues. Under these circumstances, one is subtracting two relatively large, fluctuating quantities to determine a small quantity. In general, this yields a process with a large percent error.

There is a better way to do this problem. Instead of running two independent games, the proprietor could run only one game and apply both sets of rules simultaneously to this game. In fact, he can choose to estimate the difference in revenue directly rather than the revenue that would be achieved under each set of rules.

This clearly amounts to playing the following game:

1. Whenever a 3 comes up, continue to toss the dice until either a 3 or a 7 comes up. In the first happenstance, record a minus two, since under the old rules the customer would have lost a dollar, but under the new rules he wins one; in the second happenstance, record a zero because under both sets of rules the customer loses.
2. Follow similar process if a 4 comes up.
3. If a number other than 3 or 4 comes up terminate the play then and there and record a zero. (Because of this rule, the effects of chance fluctuations in the proportionate number of times that the numbers 2, 5, 6, . . ., 11, 12 come up are eliminated from the comparison.)

The specific game that is played is quite different from the two games that are being compared. In this case there are three sources of savings; first only one set of games is played, and second the number and kinds of chance fluctuations that can affect the results are greatly reduced and lastly the stop rule makes the average game shorter. It is in fact generally true that if we wish to compare two or more situations, we can, by combining this comparison into a single problem, reduce the work substantially. As in the example, only one problem, rather than several, has to be done, and the direct estimate of the difference can usually be made more accurately than estimates of separate individual quantities.

This is a substantial virtue of the Monte Carlo method. In many complicated problems we are not actually interested in absolute values but only in comparisons. We may wish, for example, to know if Strategy A is better than Strategy B, or if Engineering Design A is better than Engineering Design B.

We might, in fact, not even believe the absolute values because the idealizations are so rough, but do believe in the existence of the qualitative features implied by differences in the calculated performances. Monte Carlo can then be used to estimate the thing that we actually desire to know and that we believe, and we can thus bypass the estimate of less important quantities. Usually however, we can obtain these less important quantities also, but at some extra cost.

Correlated Sampling can also often be used to test the accuracy of an approximate theory. If the approximate theory happens to be an exact treatment of an idealized situation, and if the idealized situation happens to be "structurally" similar to the unidealized situation, then it is often possible to design very efficient sampling schemes to calculate the difference between the idealized and unidealized situations. The answer to the problem posed by the unidealized situation can then be obtained by adding together the results of the approximate analytic calculation and the Monte Carlo difference calculation.

Comparing Different Bombing Strategies

If a strategic or tactical bombing campaign is studied by Monte Carlo, one usually has to estimate the effects of the following sequence of random events.

1. Number of aircraft that abort
2. Number of aircraft shot down by area defense on the way into the target
3. Number of aircraft that stray through navigational errors
4. Number of aircraft shot down by local defense at the target
5. Weather conditions over target (affects reconnaissance and CEP)

6. Place where bombs land
7. Damage done
8. Number of aircraft shot down by area defense on the way out of the target area
9. Number of aircraft that don't get back for miscellany of minor reasons.

In comparing different bombing strategies it is often effective to use correlation to cut down the sample size required to get significant information. If the correlation is done by using the same random numbers the computer cannot use a single list of random numbers in sequence in the two problems, for they would soon get out of step. He can either throw away the excess random numbers or, what is sometimes better, save them for use on later strikes. For example, if a larger number of targets were attacked on the first strike of one of the strategies, the extra random numbers that were used to determine the weather on these excess targets can be saved. If in a later strike an excess number of targets is attacked under the other strategy, the previously saved random numbers can then be used on these targets. Correlation can thus be achieved by using the same random numbers whenever the two strategies give rise to the same type of contingencies--even if they are on different strikes with different planes and targets.

Because the point is sometimes misunderstood, we would like to emphasize that when the results of the same type of contingencies are being correlated, the contingencies do not necessarily have the same detailed character. If for example in Strategy A, n_1 aircraft come up to the area defenses and in Strategy B, n_2 aircraft are used, the problems can still be correlated by using the same random numbers in computing the number that survive. This is

true as long as the picking is so arranged that the degrees of success are monotone functions of the uniform random numbers, so that fluctuations in the values of the uniform random numbers affect the two situations in the same qualitative way.

If the different strategies are such that a definite type of event is all-important to the comparison, then correlation by weighting may be better than by using the same random numbers. For example, if the effect of different types of defensive armament is being studied, the same kill probabilities could be used for the enemy fighters in the sampling, and weighting factors carried along to account for the differences being studied. The correlation may be higher if this is done, because exactly the same number of bombers is shot down each time, so all of the subsequent history is the same. If the correlating were done by using the same random numbers, different numbers of aircraft would be shot down and the actual progress of the two strategic campaigns might be quite different. It would still be possible to obtain correlation by using the same random numbers for the same contingencies, but it is unlikely that the correlation would be as high. Weighting will, of course, not work well if by its use one is forced to use probability density functions for the sampling which will themselves introduce a lot of variance because the sampling is then not good Importance Sampling for all the cases being considered.

Another case where weighting might be preferable to using the same random numbers would be when two different reconnaissance devices were being compared. The possible weather situations can then be classified according to the following criteria:

1. Both devices work

2. One works and the other does not

3. Neither works

Only situation 2 makes a difference between the two devices so that in the sampling only it should be allowed to occur. If 1 and 3 occur, the sample would give zero for the estimate, so they need not be calculated; only the fraction r of time they occur is needed. Account of this is automatically taken by the weighting factors. If instead of weighting factors the random numbers were used to do the correlating, then (1-r) of the time the sample would be calculating zero and be wasted. If instead of being an all or nothing situation the devices have different probabilities of working as the weather changes, then the appropriate modification must be made in the sampling. This last is as much an example of Importance Sampling as of Correlation.

5. SYSTEMATIC SAMPLING

If we are doing a multi-stage sampling problem, it often turns out to be very easy to do the first stage systematically. For example, in our problem, if we are going to toss the dice one at a time, then there is really no point in actually tossing the first die. If, for example, we were planning on getting 600 samples, we would expect on the average that each die would come up one about 100 times, two another 100, and so on. It is easy to show that we do not bias the results if we assume that the first 100 tosses of the first die actually do come up one, the second 100 tosses of this die come up two, etc. and so only toss the second die. The main advantage in doing this is that we have eliminated the error caused by fluctuations in the proportions of ones, twos, etc. which would result if the first toss was random.

In practice, however, doing the first stage of the sampling systematically does not usually lead to substantial improvements in efficiency. Generally, in fact, it will only reduce the number of samples required by relatively few percent--say 5 to 30. However, it ordinarily does not cost anything to apply this technique, so that there is no point in not using it. Also, it is sometimes interesting thus to estimate how the expected score depends on what happens at the first stage. About the only time we may not be able to use it conveniently is when we do not know in advance how big a sample we will want, and even then it may not be impractical.

The main application of Systematic Sampling is in those multi-stage problems where it is trivial to calculate the distribution of events at the first stage. In that case, the sampling should be done systematically.

6. STRATIFIED SAMPLING

This technique is a sort of combination of Importance Sampling and Systematic Sampling. For example, if we were only a little bit sophisticated and were doing the systematic sampling described above, we would soon notice that there is no point in considering the 400 tosses in which we had assigned the values three to six for the first toss of the die, since under these circumstances, we can never get a total of three. Therefore, we might systematically divide the sample into halves rather than sixths. In the first half we would say that the first die came up one, and in the second half that the first die came up two.

In theory, this method could be as powerful as Importance Sampling. In actual practice, the fact that you have to sample systematically turns out to decrease sharply the number of places in which it can be used conveniently. However, at those places in a calculation where it can be used, it is usually better than Importance Sampling and in any case never worse. Therefore, whenever the costs of the two techniques are comparable, Stratified Sampling is preferable to Importance Sampling.

The last remark on the applications of Systematic Sampling also applies to Stratified Sampling: it is usually useful when it is trivial to calculate the distribution of events at the first stage. There is only the additional fact that in more general problems one must have some idea of the relative importance where importance is defined differently from the systematic sampling case.

7. SPECIALIZED TECHNIQUES

If one were throwing the dice one at a time in a naive fashion, it would be a mistake to separate throws into disjoint pairs and look just at these pairs to see if they had produced any threes. For example, one can look not only at dice one and two and then at three and four, but also dice two and three and also four and five.[8] One will then have doubled to a first approximation the total effective number of throws for the same amount of work. Actually this doubling is not as effective as one might first think because there is a high correlation between successive throws and, therefore, the fluctuation will not be decreased as much as if all the pairs had been independent. However, this technique should be used because it is essentially free.

For another example, assume that one happened to have a machine which could throw, say, six dice at a time. One way to use this machine would be to simply group the six dice into three disjoint pairs which would correspond to the naive estimate. A much better way would be to consider all possible pairs of dice that could be obtained from the six. This is not as difficult as might seem at first glance. One would first calculate the total number of possible pairs that one could have ($6 \times 5/2 = 15$) and then by superficial observation find the number of pairs which could yield a total of three (i.e., if there are m 2's and n 1's, there are $m \times n$ pairs that will yield 3's; the estimate will then be $m \times n / 15$ for that sample).

The reader will undoubtedly think of many other specialized estimates that depend upon the fact that one is specifically calculating the probability

---

[8] We are indebted to J. F. Mcbacker for suggesting this idea.

of getting a three on a throw of two dice and which could not be generalized to other problems. This lack of generality is no reason for not investing at least a modest effort in looking for such methods and using them.

* * *

It is probably clear to the reader that the problems faced by the Monte Carlo experimenter in trying to cut down his statistical fluctuations are in some respects similar to those that are faced in almost any application of sampling. Therefore, much of the literature of statistics is relevant to the problems we have been considering. In fact, a fairly complete discussion of the fifth and sixth techniques, and to a slightly lesser extent, the preceding two, can be found in many statistics textbooks; only the first two do not seem to be applicable to ordinary statistical practice and have therefore not been discussed. For this reason it is very valuable to have professional statistical help in designing these calculations. However, if one has to choose between a person who is mainly interested in statistics and one who is mainly interested in the problem itself, experience has shown that, in this field at least, the latter is preferable. This last remark is not intended as a slur on statisticians, but simply to amplify a comment made earlier, that, "the greatest gains in variance reduction are often made by exploiting specific details of the problem, rather than by the routine application of general principles."