OAK RIDGE NATIONAL LABORATORY

operated by
UNION CARBIDE CORPORATION

for the
U.S. ATOMIC ENERGY COMMISSION

ORNL-RSIC-38
(Vol. III)

TECHNIQUES for EFFICIENT MONTE CARLO SIMULATION

Volume III

VARIANCE REDUCTION

E. J. McGrath
D. C. Irving

RADIATION SHIELDING INFORMATION CENTER

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

ORNL-RSIC-38 (Vol. III)

Contract No. W-7405-eng-26

(SAI-72-590-LJ)

TECHNIQUES FOR EFFICIENT MONTE CARLO SIMULATION

Volume III

VARIANCE REDUCTION

E. J. McGrath
D. C. Irving

APRIL 1975

Reprinted December 1974

Prepared for the

Office of Naval Research (Code 462)
Department of the Navy
Arlington, Virginia 22217

by Science Applications, Incorporated

NOTE:

This work partially supported by
DEFENSE NUCLEAR AGENCY

Reprinted by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37830
operated by
UNION CARBIDE CORPORATION
for the
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION

SAI-72-590-LJ

TECHNIQUES FOR EFFICIENT MONTE CARLO SIMULATION

VOLUME III

VARIANCE REDUCTION

Scientific Officer, Office of Naval Research (Code 462)

J. R. Simpson

Project Principal Investigator

E. J. McGrath

Co-author

D. C. Irving

ABSTRACT

Many Monte Carlo simulation problems lend themselves readily to the application of variance reduction techniques. These techniques can result in great improvements in simulation efficiency. This document describes the basic concepts of variance reduction (Part I), and a methodology for application of variance reduction techniques is presented in Part II. Appendices include the basic analytical expressions for application of variance reduction schemes as well as an abstracted bibliography.

The techniques considered here include importance sampling, Russian roulette and splitting, systematic sampling, stratified sampling, expected values, statistical estimation, correlated sampling, history reanalysis, control variates, antithetic variates, regression, sequential sampling, adjoint formulation, transformations, orthonormal and conditional Monte Carlo. Emphasis has been placed on presentation of the material for application by the general user. This has been accomplished by presenting a step by step procedure for selection and application of the appropriate technique(s) for a given problem.

CONTENTS

PART I

PART II

FIGURES

TABLES

PART I

BASIC CONCEPTS OF VARIANCE REDUCTION

EXECUTIVE SUMMARY

Monte Carlo simulation is one of the most powerful and commonly used techniques for analyzing complex physical problems. Applications can be found in many diverse areas from radiation transport to river basin modeling. Important Navy applications include analysis of antisubmarine warfare exercises and operations, prediction of aircraft or sensor performance, tactical analysis, and matrix game solutions where random processes are considered to be of particular importance. The range of applications has been broadening and the size, complexity, and computational effort required have been increasing. However, such developments are expected and desirable since increased realism is concomitant with more complex and extensive problem descriptions.

In recognition of such trends, the requirements for improved simulation techniques are becoming more pressing. Unfortunately, methods for achieving greater efficiency are frequently overlooked in developing simulations. This can generally be attributed to one or more of the following reasons:

- Analysts usually seek advanced computer systems to perform more complex simulation studies by exploiting increased speed and/or storage capabilities. This is often achieved at a considerably increased expense.

- Many efficient simulation methods have evolved for specialized applications. For example, some of the most impressive Monte Carlo techniques have been developed in radiation transport, a discipline that does not overlap into areas where even a small number of simulation analysts are working.

- Known techniques are not developed to the point where they can be easily understood or applied by even a small fraction of the analysts who are performing simulation studies or developing simulation models.

- Appendix B, "MIRAN - A Machine Independent Package For Generating Uniform Random Numbers," describes a uniform random number generator that can be used on any machine that does not have a reliable generator or on several different machines where identical random numbers are to be generated for comparison and cross checking.

Before proceeding it must be recognized that a "good" uniform random number generator is generally assumed to be available to the user. This is often not the case, although most computers today have uniform random number generators included as part of the system software. Unfortunately, many of the uniform random number generators in current use do not adequately approximate randomness to be sufficient for all Monte Carlo calculations. To alleviate this difficulty, a machine independent package for generating uniform random numbers is provided (Appendix B).

VARIANCE REDUCTION

1. INTRODUCTION

A useful feature of Monte Carlo simulation is that the analyst has the flexibility to dictate his simulation conditions and sampling plans to a much greater extent than does an experimenter in a real world environment. This extra latitude provides an excellent opportunity for optimal design of simulations to obtain estimates with minimal sampling size. This will effectively reduce the time and effort involved in computation as the number of trials necessary to achieve a given accuracy is thereby reduced. In view of the large number of situations where simulation results can be substantially improved, it is fair to say that no simulation problem has been justly treated until the possibility of applying variance reduction techniques has been seriously considered.

The procedures which are available in the design of a Monte Carlo simulation for minimizing the required sample size are generally called variance reduction techniques. The intent here is to provide the analyst with an understanding of and an appreciation for several variance reduction techniques and to provide a useful guide for selecting and using the most appropriate technique for his particular problem.

It is difficult to provide a complete perspective on variance reduction techniques. This is primarily due to the fact that there are an infinite number of ways Monte Carlo simulation can be accomplished for a given problem and each could conceivably be used to calculate the simulation objective although with greatly different efficiencies. However, it appears fair to say that the approach to improving simulation efficiency was not seriously considered until the work on the atomic bomb during the Second World War.(14) This work initially involved the use of "straightforward" Monte Carlo simulation for nuclear particle transport, but early in these investigations Von Neumann and Ulam(18) applied certain variance reduction techniques. A systematic development of these techniques was presented by Harris and Kahn about 1948.(19) Although comprehensive, this detailed work is difficult to apply to general problems. Subsequent application and development of variance reduction techniques has been almost exclusively carried out within the radiation transport community. This has resulted in limited application in other areas where Monte Carlo simulation is used. It is the purpose of this document to provide a mechanism to aid in a wider application of variance reduction. This has been attempted by presenting the material in two parts.

Part I, BASIC CONCEPTS OF VARIANCE REDUCTION, presents the fundamental principles and relationships among several variance reduction techniques. Part I is intended to provide the reader with a background and an understanding of variance reduction. It is recommended that the user who is not familiar with the basic concepts review Part I before attempting to implement variance reduction.

Part II of this volume, APPLICATION OF VARIANCE REDUCTION TECHNIQUES, comes as close as currently practical to being a step-by-step procedure for application of variance reduction. However, the reader should have an understanding of the basic principles involved. In most cases considerable ingenuity and insight will also be necessary. The approach here has been to present a convenient characterization of the various methods considered for purposes of selection. This is followed by a summary of guidellnes on how to actually apply each method.

This volume also includes other information useful in applying variance reduction techniques to Monte Carlo problems. Appendix A presents a summary of the pertinent analytical formulations and Appendix B is an abstracted bibliography of useful references.

2. CHARACTERIZATION OF VARIANCE REDUCTION TECHNIQUES

In this section the general characteristics of variance reduction techniques will be introduced. In Section 3 each method will be discussed in detail.

2.1 CLASSIFICATION OF TECHNIQUES

As the name implies, variance reduction is concerned with increasing the accuracy of Monte Carlo estimates of parameters. A simulation using one or more reduction techniques can be contrasted with what may be considered the crude (sometimes called direct or straightforward) Monte Carlo approach where an attempt is made to create true-to-life or actual models of the process. In crude sampling, flows through the model and sampling probability distributions are chosen to reflect the real situation as exactly as possible. On the other hand, variance reduction techniques attempt to increase the effectiveness of the Monte Carlo method by:

- Modifying the simulation procedure
- Utilization of approximate or analytical information
- Studying the system within a different context or abstract representation

Based on these approaches a general classification of several known variance reduction schemes is presented in Table 2.1. Many of the techniques presented in Table 2.1 are related and it is difficult to arrive at a completely distinct classification. However, the manner in which they are presented here is useful for subsequent discussions.

Modifying the sampling process is usually achieved by using more effective sampling techniques or altering the sampling distributions. As an example consider the problem of estimating the probability of an early failure in a piece of electronic equipment, and suppose that the failure distribution for this equipment is exponential with a very long mean time between failures (MTBF). In a crude Monte Carlo evaluation the ratio of the number of early failure to the total number of simulated failures is very small. Thus, in

TABLE 2.1

Classification of Variance Reduction Techniques

- MODIFICATION OF THE SAMPLING PROCESS
  - Importance Sampling
  - Russian Roulette and Splitting
  - Systematic Sampling
  - Stratified Sampling
- USE OF ANALYTICAL EQUIVALENCE
  - Expected Values
  - Statistical Estimation
  - Correlated Sampling
  - History Reanalysis
  - Control Variates
  - Antithetic Variates
  - Régression
- SPECIALIZED TECHNIQUES
  - Sequential Sampling
  - Adjoint Formulation
  - Transformations
  - Orthonormal Functions
  - Conditional Monte Carlo

order to generate confidence in an estimate for the probability of early failure, one must simulate a very large number of failures. The number of simulated events required can be substantially reduced, however, if the failure distribution in the simulation is suitably modified. In particular, if an exponential distribution with a short MTBF is substituted for the actual failure distribution, more early failures will be observed, and thus a more accurate answer can be derived with less simulation effort. This procedure is referred to as importance sampling. Of course, the modifications introduced in the sampling distribution must be accounted for when determining the desired estimate since the failure processes, (actual and modified) are not the same.

The above example, simulating events of very low probability, illustrates one area where variance reduction techniques are always beneficial, if not an absolute necessity. If the occurrence of an event in a process is on the order of one in a thousand, then one would expect an event to occur only once in every thousand direct simulations of the process. Since the accuracy in measuring an event is related to the number of times it occurs, the crude simulation has to be run many thousands of times before much accuracy is achieved. The common variance reduction procedure in these cases involves altering the simulation in a known way so that the rare events can be observed more frequently.

Other forms of variance reduction are based on the fact that analytic procedures are usually preferable to simulation. Thus, reverting to simulation implies the problem does not have a readily available analytic solution. However, in many cases segments of the process may be amenable to determining a closed form solution. In other cases, the overall process or segments of the process may be closely correlated to a simpler, approximate process with known analytic solutions. In both situations substantial improvement can be realized by taking advantage of this knowledge. This class of techniques is described by the term "use of analytical equivalence".

As a simple example of the use of analytical equivalence, consider again a piece of electronic equipment. Suppose this time, however, that the failure distribution of the equipment is not exponential, but assume that the exponential distribution may serve as a first approximation to it. The correlation approach to variance reduction involves investigating the failure properties of this equipment by taking advantage of this knowledge and simulating the difference between the actual and the approximate exponential failure rate instead of simulating the actual process. The properties of the actual process can then be inferred using the analytic properties of the exponential distribution and the results from the simulation on the difference between the actual and exponential distribution. This approach is called control variates.

In addition to sampling modification and analytical equivalence, there are certain specialized techniques that can be used to achieve variance reduction. These procedures may include the application of one or more of the above techniques in its implementation. One powerful procedure is called sequential Monte Carlo. In order to effectively employ variance reduction in a simulation, some knowledge about the process and the answers to be generated must exist. One way to gain this information is through a direct simulation of the process. Results from this simulation can then be used to define variance reduction techniques which will refine and improve the efficiency of a second simulation. In complex problems, several iterations may be called for.

Another procedure which often proves valuable in developing variance reduction procedures is to consider the process from various viewpoints. In many flow processes, for example, hints for effective importance functions can be gained by considering the process in reverse or looking at the mathematical adjoint of the problem under study. However, as with many of the specialized techniques described in Table 2.1, it is not adequately developed for general application.

Generally variance reduction techniques can be aimed at reducing the variance of the estimate of only one parameter or aspect of the process being simulated. Using variance reduction techniques on one parameter can reduce the effectiveness of the simulation to estimate other parameters. It is very important, therefore,to first determine all of the results which will be desired from the simulation before searching for a technique to apply to a given situation.

If several quantities (parameters) are to beestimated by the simulation, the selection of a variance reduction technique has to be considered from the standpoint of all of these parameters. In many circumstances it may be beneficial to create a different Monte Carlo method to estimate each parameter. The goal for each simulator would be efficient measurement of a specific parameter.

Each of the techniques or procedures introduced in Table 2.1 will be discussed in detail in subsequent sections.

2.2 VARIANCE REDUCTION AND KNOWLEDGE OF THE PROCESS TO BE SIMULATED

As the discussion of the previous section suggests, variance reduction can be viewed as a means to use known, usually qualitative, information about the process in an explicit and quantitative manner. In fact, if nothing is known about the process to be simulated, variance reduction cannot be directly achieved. (However, sequential sampling may be used to generate the required knowledge.) The other extreme from no knowledge is complete knowledge, and in this case a zero variance simulation can be devised. Put very simply, variance reduction techniques cannot give the user something for nothing; it is merely a way of not wasting information. Therefore, the more that is known about the problem, the more effective variance reduction can be and the more powerful are the techniques that can be employed. Hence, it is always important to clearly define as much as is possible what is known about a problem.

Knowledge of a process to be simulated can be qualitative and/or quantitative. Both are useful. It is important to use all the information available, and in fact it may be useful to do limited crude simulations of the process to gain some knowledge, especially if a little data might lead to extensive insight. Selection of a variance reduction technique(s) for a particular simulation is thus peculiar to that simulation, and general procedures are difficult to establish. However, the mental exercise and the initial groundwork that must be established in order to select or evalute the usefulness of applying these techniques is almost always worth the effort. Searching for a technique forces the simulation designer into asking the basic questions of: (1) "What answers are to be generated from the simulation," and (2) what is known about the behavior of the process"?

Problem definition is thus of paramount importance. Before considering variance reduction techniques it is important to characterize aspects of the problem which might indicate which might be fruitfully applied. To evaluate the usefulness of these methods for a particular problem it is necessary to:

- List all of the parameters to be estimated from the simulation.
- Determine all the available knowledge on the internal workings of the process to be simulated.

In fact,clearly delineating such information is the basis for the approach presented in Part II, APPLICATION OF VARIANCE REDUCTION TECHNIQUES.

2.3 INTEGRAL REPRESENTATION

In principle a Monte Carlo procedure can be interpreted as a method for evaluating an integral, or more graphically, the area under a curve. Since integrals can also be evaluated by analytic or numerical methods, reverting to Monte Carlo simulation implies either a very complex integration or, more generally, an inability to represent the problem in integral form. Knowledge that the Monte Carlo procedure does have an integral representation and determining the explicit form of that integral is fundamental to understanding and developing variance reduction techniques.

An intuitive justification for the integral representation can be given by considering how the Monte Carlo method works. The model of the process, or simulation, is exercised numerous times. Conclusions about the process are drawn by averaging the individual outcomes. From a probabilistic viewpoint, averaging is a means for estimating particular types of integrals known as expectations or expected values.

Symbolically, suppose $g(X_1,\ldots,X_n)$ is the outcome or result obtained from a simulation. The X_i values represent a particular outcome* from each of the random processes affecting the characteristic of the system being estimated. To simplify the presentation, let $\vec{x}$ represent the vector $(x_1,\ldots,x_n)$. If $f(\vec{x})$ denotes the probability density function of $\vec{x}$ (i.e., joint probability density function of $x_1,\ldots,x_n$), then the objective of the Monte Carlo simulation is to estimate the integral

$$I = E[g(\vec{x})] = \int g(\vec{x})f(\vec{x})d\vec{x} \quad . \tag{2.1}$$

A crude application of Monte Carlo would obtain an estimate I by selecting a random sample $\vec{X}_1,\ldots,\vec{X}_N$ from $f(\vec{x})$ and compute the sample mean using

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} g(\vec{X}_i) \tag{2.2}$$

The law of large numbers ensures the convergence of $\hat{I}$ to I in most cases.(14)

It is, of course, true that $\hat{I}$ is a random variable and that the expected value of $\hat{I}$ equals I. That is,

$$E[\hat{I}] = I \tag{2.3}$$

It is said that $\hat{I}$ is an unbiased estimator for I when (2.3) holds. This is important to keep in mind when estimators for variance reduction are constructed since variance reduction can lead to biased estimators unless care is taken.

*Using general notation, X represents a particular outcome of the random variable x.

An estimate of the error in the estimator $\hat{I}$ is given by the sample variance S^2, where

$$S^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} [g(\vec{X}_i) - \hat{I}]^2 = \frac{N}{N-1} \left\{ \frac{1}{N} \sum_{i=1}^{N} g^2(\vec{X}_i) - \hat{I}^2 \right\} \qquad (2.4)$$

S^2 is commonly used as an estimate for σ^2, the population variance, which is defined as

$$\sigma^2 = E[\{g(\vec{x}) - I\}^2] \qquad (2.5)$$

S^2 is also used as a basis for evaluating the effectiveness of Monte Carlo simulations. A basic measure for such effectiveness if $E[(\hat{I}-I)^2]$. It is easy to see that[*]

$$E[(\hat{I}-I)^2] = \sigma^2/N \qquad (2.6)$$

Note that as $N \to \infty$, $E[(\hat{I}-I)^2] \to 0$.[**]

Now, since σ^2 is estimated using S^2, an estimate for $E[(\hat{I}-I)^2]$ is constructed using

$$s^2 = \frac{S^2}{N} = \frac{1}{(N-1)} \left\{ \frac{1}{N} \sum_{i=1}^{N} g^2(\vec{X}_i) - \hat{I}^2 \right\} \qquad (2.7)$$

The estimator $s^{\cdot}$ is often used as an absolute measure of the accuracy of a simulation.

[*] It is assumed that a simulation will consist of N statistically independent histories.

[**] Since $E[(\hat{I}-I)^2] \to 0$ as $N \to \infty$, then $\hat{I}$ is said to be a consistent estimator for I.

Use of the integral representation provides a convenient mechanism to develop and apply variance reduction in simulation, and if possible, such a representation should be established. As a trivial example of how this might be accomplished consider the queueing system shown in Fig. 2.1. Here t indicates time. Further it is assumed that $f_1(t), \ldots, f_7(t)$ are probability density functions for the time required to go through the corresponding box. p_{11} and p_{12} are respective probabilities for going along the paths indicated. Similarly for p_{21} and p_{22}.

It is easy to see that the average time to pass through the system is given by

$$I = \int_0^\infty t[f_1(t) + p_{11}f_2(t) + p_{12}f_3(t) + f_4(t) + f_5(t) + p_{21}f_6(t) + f_7(t)]dt$$

$$= \int_0^\infty t\, f(t)dt$$

which has the same qualitative form as Eq. 2.1. Such integral representation can greatly simplify the application of variance reduction techniques and will be used as a basis for the discussion presented later.

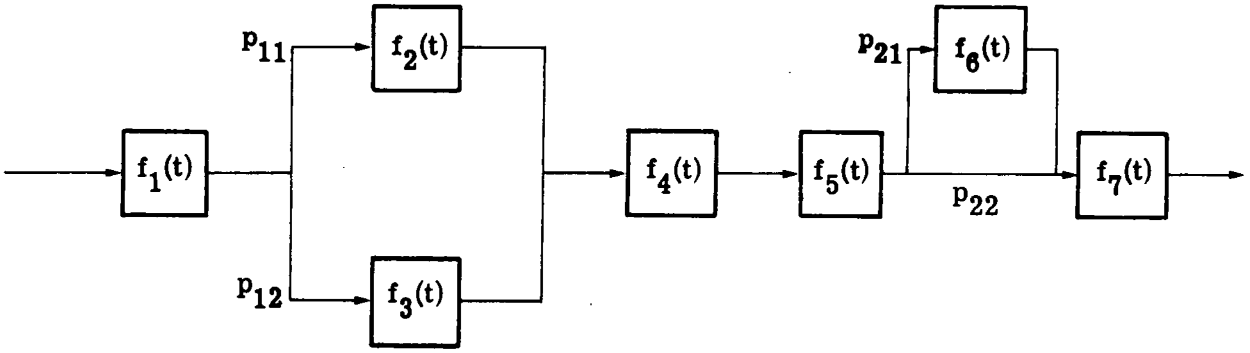


Fig. 2.1. Schematic of a Simple Queueing System

2.4 EFFICIENCY OF VARIANCE REDUCTION

This section presents the basic ideas and practical expressions for estimating the efficiency of variance reduction techniques.

2.4.1 General Concepts

The measure introduced in the previous discussion that will be used to evaluate the effectiveness of a simulation was $E[(\hat{I}-I)^2]$. This is estimated using s^2 defined by (2.7). That is,

$$s^2 = \frac{1}{(N-1)} \left\{ \frac{1}{N} \sum_{i=1}^{N} g^2(\vec{X}_i) - \hat{I}^2 \right\} \qquad (2.8)$$

s^2 is an estimate for the variance of $\hat{I}$. It can be shown that

$$E[s^2] = E[(\hat{I} - I)^2] = \sigma^2/N \qquad (2.9)$$

where σ^2 is the variance of $g(\vec{x})$ and N is the sample size or the number of histories.

It can be seen from (2.9) that, as the number of histories, N, increases, the closer $\hat{I}$ will come to I.

Another way to consider this is in terms of intervals of uncertainty. For example it is known from basic statistics[(14)] that, with high probability the estimate I will fall between $I - k\sigma/\sqrt{N}$ and $I + k\sigma/\sqrt{N}$ where k is some constant. Thus for a fixed k, the convergence of the estimate is related to the number of histories , N, and the variance of $g(\vec{x})$.

Two approaches can be taken to increase the accuracy of the estimator, $\hat{I}$. One is to increase the number of histories. The other is to reduce the variance (σ) associated with each observation. The disadvantage of increasing the number of iterations (i.e., the size of N) is obvious. For example, to reduce the interval of uncertainty by a factor of two, thus doubling the accuracy, four times as many histories would be required (for a fourfold increase in computing time). Eventually it becomes prohibitively expensive to gain further accuracy by increasing the number of histories. Therefore, achieving variance reduction which reduces the variance associated with each history, σ, is highly desirable for improvements in the answers.

To evaluate the efficiency gained in the use of variance reduction techniques it is clearly desirable to have a quantitative measure. This can readily be established based on the ideas introduced above. Suppose two simulation method exist for estimating the same parameter I. Let the variance per history associated with the first simulation method be σ_1^2 and that associated with the second be σ_2^2. It is desired that the result be known within an uncertainty of ϵ (i.e., the estimate $\hat{I}$ fall in the interval $I-\epsilon$ to $I+\epsilon$). For this to happen with high probability will require $N_1 = k^2\sigma_1^2/\epsilon^2$ histories for the first method. For the second method, it will require $N_2 = k^2\sigma_2^2/\epsilon^2$ histories. In general the two methods will require different amounts of computational effort to generate each history. Let the computer time taken per history by the first method be t_1 seconds and by the second t_2 seconds. Then the total time required for the first method to achieve the desired accuracy would be $k^2\sigma_1^2t_1/\epsilon^2$. Total time for the second method would be $k^2\sigma_2^2t_2/\epsilon^2$. The relative efficiency of the two simulation methods is given by the ratio of the computing times required. Thus,

$$\text{efficiency} = \epsilon = \frac{t_1\sigma_1^2}{t_2\sigma_2^2} \tag{2.10}$$

which is the relative time advantage gained by using the second method.

In most applications a variance reduction method is being compared to crude sampling. That is, t_1 and σ_1^2 would be that obtained when crude sampling is used, while t_2 and σ_2^2 refer to the computation using the variance reduction method.

2.4.2 Estimation of Variance Reduction Efficiency

The difficulty in using definition (2.10) for efficiency is that σ_1^2 and σ_2^2 are rarely known. However, it is reasonable to replace them by their estimators and get an estimator for ϵ,

$$\hat{\epsilon} = \frac{t_1 S_1^2}{t_2 S_2^2} \tag{2.11}$$

where

$$S_1^2 = \frac{N_1}{N_1-1}\left[\frac{1}{N_1}\sum_{i=1}^{N_1} g^2(\vec{X}_i) - \hat{I}_1^2\right] \tag{2.12}$$

with

$$\hat{I}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} g(\vec{X}_i) \tag{2.13}$$

and $\vec{X}_1, \ldots, \vec{X}_{N_1}$ being a random sample obtained with crude Monte Carlo. Also,

$$S_2^2 = \frac{N_2}{N_2-1}\left[\frac{1}{N_2}\sum_{i=1}^{N_2} g^2(\vec{X}_i') - \hat{I}_2^2\right] \tag{2.14}$$

with

$$\hat{I}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} g(\vec{X}_i') \tag{2.15}$$

and $X'_1, \ldots, X'_{N_2}$ being a random sample obtained using variance reduction.

It is important to recognize that $\hat{\epsilon}$ is a random variable and in practical application will be subject to random variations. In fact, as S_1^2 and S_2^2 are second order quantities, they will be subject to much larger variation than first order parameters such as $\hat{I}$.

Note that the use of (2.12) and (2.14) assumes that independent random histories were available. However, the application of many variance reduction techniques will not produce histories that are statistically independent. This is particularly true when stratified, systematic sampling, or Russian Roulette and splitting are used. .In some cases correlated sampling and history reanalysis will also produce samples that are not independent.

In cases where a truly random sample is not available (or suspected to be not available), it is convenient to use a batching process to estimate the sample variance. The general guidelines to follow in application of batching are:

1. Obtain a sample, say $g(\vec{X}_1), \ldots, g(\vec{X}_N)$ consisting of N histories (which may or may not be independent).
2. Group the histories into batches such that the batches are independent and equivalent. For example, it may be possible to arrange the histories so that the sample contained within any batch will be independent from the samples in any other batch. However, the samples within a batch may be correlated with each other. In the case of stratified sampling, each batch must consist of the same number of samples from the same strata. (Typically, the number of batches, N_B, should be between 10 and 50.)
3. Construct an average in each batch for the parameters being estimated. That is, if $g(\vec{X}_1), \ldots, g(\vec{X}_{N_1})$ are contained in batch 1, then set

$$\hat{I}_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} g(\vec{X}_i) \qquad (2.16)$$

where it is assumed that there are N_1 sample in each batch.

4. Construct a final estimate for I using

$$\hat{I} = \frac{1}{N_B} \sum_{i=1}^{N_B} \hat{I}_i \tag{2.17}$$

5. Obtain an estimate for σ^2 from

$$S^2 = \frac{1}{(N_B-1)} \sum_{i=1}^{N_B} (\hat{I}_i - \hat{I})^2 = \frac{N_B}{N_B-1} \left[\frac{1}{N_B} \sum_{i=1}^{N_B} \hat{I}_i^2 - \hat{I}^2 \right] \tag{2.18}$$

In essence, each batch is being considered as a separate small simulation run. Parameters are estimated as the average of the estimates obtained in each batch. The sample variance of the different batch estimates provides a basis for estimating the variance of the final average. This technique is completely general; it will work in all cases no matter what combination of variance reduction techniques are being used nor what kind of parameter is being estimated. Batching may not provide the best estimate in all cases; usually a better estimator can be constructed for any particular techniques being used. However, there frequently are easily-missed subtleties in ensuring that an estimator is based on independent and equivalent samples. It is generally best to avoid the analysis required to generate an estimator valid for the particular methods employed - and also avoid the pitfall of constructing an erroneous estimator - by using batching to calculate variances.

2.4.3 Estimation of Confidence Intervals

In some applications, it is of interest to calculate confidence intervals for estimated parameters when variance reduction is used. Under the usual assumptions,(14) the confidence interval of size α can be obtained from the following expression:

$$P\left[\hat{I} - \frac{TS}{\sqrt{N}} \leq I \leq \hat{I} + \frac{TS}{\sqrt{N}}\right] \cong \frac{1}{\sqrt{2\pi}} \int_{-T}^{T} e^{-t^2/2}\, dt = \alpha \qquad (2.19)$$

where α may be obtained from Table 2.2. The value of S may be obtained using (2.4) or (2.18). Then, the interval $I - \frac{TS}{\sqrt{N}}$; $I + \frac{TS}{\sqrt{N}}$ is said to be a 100 α% confidence interval for the estimate of I.

2.5 THE PITFALLS OF OVERBIASING AND UNDERBIASING

The goal of variance reduction is improved efficiency, that is, making the best use of computing time to simulate events which are most significant to the final answer. In modifying the sampling to bring this about, it is possible to overshoot the mark and produce a sampling scheme that is so strongly biased as to be less efficient than crude sampling. This is termed 'overbiasing' or 'oversampling'. The opposite term, 'underbiasing' or 'undersampling', is used to apply to the crude or slightly modified sampling scheme when the result depends heavily on infrequent events and not enough observations occurred for good statistics.

It is a general characteristic of both overbiased and underbiased situations that most of the time the answers generated are too small. This produces an apparently consistent bias in the results which can be more troublesome than poor confidence intervals in the result. Furthermore, variance estimates are also generally small so that the confidence intervals calculated in the simulation will tend to indicate that the results are much more accurate than they really are. This generates a false sense of security and faith in results which are actually consistently bad.

As an extremely simplified example, consider a simulation in which there are basically two classes of events. One type of event (X_1) occurs frequently ($f(X_1) = .9999$) but contributes only a small amount $(g(X_1) = .01)$ to the final result while the other type (event X_2) is rare $(f(X_2) = .0001)$ but

TABLE 2.2

Table of the Standard Cumulative Normal Distribution

$$F(y) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{y} e^{-z^2/2}\, dz \qquad y = 0.00\,(0.01)\,4.99^{*}$$

y	·00	·01	·02	·03	·04	·05	·06	·07	·08	·09
·0	·5000	·5040	·5080	·5120	·5160	·5199	·5239	·5279	·5319	·5359
·1	·5398	·5438	·5478	·5517	·5557	·5596	·5636	·5675	·5714	·5753
·2	·5793	·5832	·5871	·5910	·5948	·5987	·6026	·6064	·6103	·6141
·3	·6179	·6217	·6255	·6293	·6331	·6368	·6406	·6443	·6480	·6517
·4	·6554	·6591	·6628	·6664	·6700	·6736	·6772	·6808	·6844	·6879
·5	·6915	·6950	·6985	·7019	·7054	·7088	·7123	·7157	·7190	·7224
·6	·7257	·7291	·7324	·7357	·7389	·7422	·7454	·7486	·7517	·7549
·7	·7580	·7611	·7642	·7673	·7703	·7734	·7764	·7794	·7823	·7852
·8	·7881	·7910	·7939	·7967	·7995	·8023	·8051	·8078	·8106	·8133
·9	·8159	·8186	·8212	·8238	·8264	·8289	·8315	·8340	·8365	·8389
1·0	·8413	·8438	·8461	·8485	·8508	·8531	·8554	·8577	·8599	·8621
1·1	·8643	·8665	·8686	·8708	·8729	·8749	·8770	·8790	·8810	·8830
1·2	·8849	·8869	·8888	·8907	·8925	·8944	·8962	·8980	·8997	·90147
1·3	·90320	·90490	·90658	·90824	·90988	·91149	·91309	·91466	·91621	·91774
1·4	·91924	·92073	·92220	·92364	·92507	·92647	·92785	·92922	·93056	·93189
1·5	·93319	·93448	·93574	·93699	·93822	·93943	·94062	·94179	·94295	·94408
1·6	·94520	·94630	·94738	·94845	·94950	·95053	·95154	·95254	·95352	·95449
1·7	·95543	·95637	·95728	·95818	·95907	·95994	·96080	·96164	·96246	·96327
1·8	·96407	·96485	·96562	·96638	·96712	·96784	·96856	·96926	·96995	·97062
1·9	·97128	·97193	·97257	·97320	·97381	·97441	·97500	·97558	·97615	·97670
2·0	·97725	·97778	·97831	·97882	·97932	·97982	·98030	·98077	·98124	·98169
2·1	·98214	·98257	·98300	·98341	·98382	·98422	·98461	·98500	·98537	·98574
2·2	·98610	·98645	·98679	·98713	·98745	·98778	·98809	·98840	·98870	·98899
2·3	·98928	·98956	·98983	$\cdot 9^{2}0097$	$\cdot 9^{2}0358$	$\cdot 9^{2}0613$	$\cdot 9^{2}0863$	$\cdot 9^{2}1106$	$\cdot 9^{2}1344$	$\cdot 9^{2}1576$
2·4	$\cdot 9^{2}1802$	$\cdot 9^{2}2024$	$\cdot 9^{2}2240$	$\cdot 9^{2}2451$	$\cdot 9^{2}2656$	$\cdot 9^{2}2857$	$\cdot 9^{2}3053$	$\cdot 9^{2}3244$	$\cdot 9^{2}3431$	$\cdot 9^{2}3613$
2·5	$\cdot 9^{2}3790$	$\cdot 9^{2}3963$	$\cdot 9^{2}4132$	$\cdot 9^{2}4297$	$\cdot 9^{2}4457$	$\cdot 9^{2}4614$	$\cdot 9^{2}4766$	$\cdot 9^{2}4915$	$\cdot 9^{2}5060$	$\cdot 9^{2}5201$
2·6	$\cdot 9^{2}5339$	$\cdot 9^{2}5473$	$\cdot 9^{2}5604$	$\cdot 9^{2}5731$	$\cdot 9^{2}5855$	$\cdot 9^{2}5975$	$\cdot 9^{2}6093$	$\cdot 9^{2}6207$	$\cdot 9^{2}6319$	$\cdot 9^{2}6427$
2·7	$\cdot 9^{2}6533$	$\cdot 9^{2}6636$	$\cdot 9^{2}6736$	$\cdot 9^{2}6833$	$\cdot 9^{2}6928$	$\cdot 9^{2}7020$	$\cdot 9^{2}7110$	$\cdot 9^{2}7197$	$\cdot 9^{2}7282$	$\cdot 9^{2}7365$
2·8	$\cdot 9^{2}7445$	$\cdot 9^{2}7523$	$\cdot 9^{2}7599$	$\cdot 9^{2}7673$	$\cdot 9^{2}7744$	$\cdot 9^{2}7814$	$\cdot 9^{2}7882$	$\cdot 9^{2}7948$	$\cdot 9^{2}8012$	$\cdot 9^{2}8074$
2·9	$\cdot 9^{2}8134$	$\cdot 9^{2}8193$	$\cdot 9^{2}8250$	$\cdot 9^{2}8305$	$\cdot 9^{2}8359$	$\cdot 9^{2}8411$	$\cdot 9^{2}8462$	$\cdot 9^{2}8511$	$\cdot 9^{2}8559$	$\cdot 9^{2}8605$
3·0	$\cdot 9^{2}8650$	$\cdot 9^{2}8694$	$\cdot 9^{2}8736$	$\cdot 9^{2}8777$	$\cdot 9^{2}8817$	$\cdot 9^{2}8856$	$\cdot 9^{2}8893$	$\cdot 9^{2}8930$	$\cdot 9^{2}8965$	$\cdot 9^{2}8999$
3·1	$\cdot 9^{3}0324$	$\cdot 9^{3}0646$	$\cdot 9^{3}0957$	$\cdot 9^{3}1260$	$\cdot 9^{3}1553$	$\cdot 9^{3}1836$	$\cdot 9^{3}2112$	$\cdot 9^{3}2378$	$\cdot 9^{3}2636$	$\cdot 9^{3}2886$
3·2	$\cdot 9^{3}3129$	$\cdot 9^{3}3363$	$\cdot 9^{3}3590$	$\cdot 9^{3}3810$	$\cdot 9^{3}4024$	$\cdot 9^{3}4230$	$\cdot 9^{3}4429$	$\cdot 9^{3}4623$	$\cdot 9^{3}4810$	$\cdot 9^{3}4991$
3·3	$\cdot 9^{3}5166$	$\cdot 9^{3}5335$	$\cdot 9^{3}5499$	$\cdot 9^{3}5658$	$\cdot 9^{3}5811$	$\cdot 9^{3}5959$	$\cdot 9^{3}6103$	$\cdot 9^{3}6242$	$\cdot 9^{3}6376$	$\cdot 9^{3}6505$
3·4	$\cdot 9^{3}6631$	$\cdot 9^{3}6752$	$\cdot 9^{3}6869$	$\cdot 9^{3}6982$	$\cdot 9^{3}7091$	$\cdot 9^{3}7197$	$\cdot 9^{3}7299$	$\cdot 9^{3}7398$	$\cdot 9^{3}7493$	$\cdot 9^{3}7585$
3·5	$\cdot 9^{3}7674$	$\cdot 9^{3}7759$	$\cdot 9^{3}7842$	$\cdot 9^{3}7922$	$\cdot 9^{3}7999$	$\cdot 9^{3}8074$	$\cdot 9^{3}8146$	$\cdot 9^{3}8215$	$\cdot 9^{3}8282$	$\cdot 9^{3}8347$
3·6	$\cdot 9^{3}8409$	$\cdot 9^{3}8469$	$\cdot 9^{3}8527$	$\cdot 9^{3}8583$	$\cdot 9^{3}8637$	$\cdot 9^{3}8689$	$\cdot 9^{3}8739$	$\cdot 9^{3}8787$	$\cdot 9^{3}8834$	$\cdot 9^{3}8879$
3·7	$\cdot 9^{3}8922$	$\cdot 9^{3}8964$	$\cdot 9^{4}0039$	$\cdot 9^{4}0426$	$\cdot 9^{4}0799$	$\cdot 9^{4}1158$	$\cdot 9^{4}1504$	$\cdot 9^{4}1838$	$\cdot 9^{4}2159$	$\cdot 9^{4}2468$
3·8	$\cdot 9^{4}2765$	$\cdot 9^{4}3052$	$\cdot 9^{4}3327$	$\cdot 9^{4}3593$	$\cdot 9^{4}3848$	$\cdot 9^{4}4094$	$\cdot 9^{4}4331$	$\cdot 9^{4}4558$	$\cdot 9^{4}4777$	$\cdot 9^{4}4988$
3·9	$\cdot 9^{4}5190$	$\cdot 9^{4}5385$	$\cdot 9^{4}5573$	$\cdot 9^{4}5753$	$\cdot 9^{4}5926$	$\cdot 9^{4}6092$	$\cdot 9^{4}6253$	$\cdot 9^{4}6406$	$\cdot 9^{4}6554$	$\cdot 9^{4}6696$
4·0	$\cdot 9^{4}6833$	$\cdot 9^{4}6964$	$\cdot 9^{4}7090$	$\cdot 9^{4}7211$	$\cdot 9^{4}7327$	$\cdot 9^{4}7439$	$\cdot 9^{4}7546$	$\cdot 9^{4}7649$	$\cdot 9^{4}7748$	$\cdot 9^{4}7843$
4·1	$\cdot 9^{4}7934$	$\cdot 9^{4}8022$	$\cdot 9^{4}8106$	$\cdot 9^{4}8186$	$\cdot 9^{4}8263$	$\cdot 9^{4}8338$	$\cdot 9^{4}8409$	$\cdot 9^{4}8477$	$\cdot 9^{4}8542$	$\cdot 9^{4}8605$
4·2	$\cdot 9^{4}8665$	$\cdot 9^{4}8723$	$\cdot 9^{4}8778$	$\cdot 9^{4}8832$	$\cdot 9^{4}8882$	$\cdot 9^{4}8931$	$\cdot 9^{4}8978$	$\cdot 9^{5}0226$	$\cdot 9^{5}0655$	$\cdot 9^{5}1066$
4·3	$\cdot 9^{5}1460$	$\cdot 9^{5}1837$	$\cdot 9^{5}2199$	$\cdot 9^{5}2545$	$\cdot 9^{5}2876$	$\cdot 9^{5}3193$	$\cdot 9^{5}3497$	$\cdot 9^{5}3788$	$\cdot 9^{5}4066$	$\cdot 9^{5}4332$
4·4	$\cdot 9^{5}4587$	$\cdot 9^{5}4831$	$\cdot 9^{5}5065$	$\cdot 9^{5}5288$	$\cdot 9^{5}5502$	$\cdot 9^{5}5706$	$\cdot 9^{5}5902$	$\cdot 9^{5}6089$	$\cdot 9^{5}6268$	$\cdot 9^{5}6439$
4·5	$\cdot 9^{5}6602$	$\cdot 9^{5}6759$	$\cdot 9^{5}6908$	$\cdot 9^{5}7051$	$\cdot 9^{5}7187$	$\cdot 9^{5}7318$	$\cdot 9^{5}7442$	$\cdot 9^{5}7561$	$\cdot 9^{5}7675$	$\cdot 9^{5}7784$
4·6	$\cdot 9^{3}7888$	$\cdot 9^{5}7987$	$\cdot 9^{5}8081$	$\cdot 9^{5}8172$	$\cdot 9^{5}8258$	$\cdot 9^{5}8340$	$\cdot 9^{5}8419$	$\cdot 9^{5}8494$	$\cdot 9^{5}8566$	$\cdot 9^{5}8634$
4·7	$\cdot 9^{5}8699$	$\cdot 9^{5}8761$	$\cdot 9^{5}8821$	$\cdot 9^{5}8877$	$\cdot 9^{5}8931$	$\cdot 9^{5}8983$	$\cdot 9^{6}0320$	$\cdot 9^{6}0789$	$\cdot 9^{6}1235$	$\cdot 9^{6}1661$
4·8	$\cdot 9^{6}2067$	$\cdot 9^{6}2453$	$\cdot 9^{6}2822$	$\cdot 9^{6}3173$	$\cdot 9^{6}3508$	$\cdot 9^{6}3827$	$\cdot 9^{6}4131$	$\cdot 9^{6}4420$	$\cdot 9^{6}4696$	$\cdot 9^{6}4958$
4·9	$\cdot 9^{6}5208$	$\cdot 9^{6}5446$	$\cdot 9^{6}5673$	$\cdot 9^{6}5889$	$\cdot 9^{6}6094$	$\cdot 9^{6}6289$	$\cdot 9^{6}6475$	$\cdot 9^{6}6652$	$\cdot 9^{6}6821$	$\cdot 9^{6}6981$

*Note. $\dfrac{1}{\sqrt{2\pi} - 1} \displaystyle\int_{-T}^{T} e^{-z^{2}/2}\, dt = 2\,F\,(T) - 1 = \alpha$

makes a large contribution ($g(X_2) = 100$) when it does occur. In this example, the integral I being estimated has the correct value:

$$\begin{aligned} I &= \sum_i (\text{probability of event } i)*(\text{value of event } i) \\ &= f(X_1)g(X_1) + f(X_2)g(X_2) \\ &= .02 \end{aligned} \tag{2.20}$$

However, using crude Monte Carlo with a moderate (several hundred to a thousand) number of histories, event X_2 would very probably never occur and the 'underbiased' answer would be recorded as

$$\hat{I}_u = g(X_1) = .01 \tag{2.21}$$

If it was realized that X_2 events made such a heavy contribution to the result, one natural response would be to modify the simulation so that X_2 events occurred frequently (see the discussion of importance sampling in Section 3.1.1 for an explanation of the formulas used in this example). If this modification was carried to excess, say new probabilities of $f^*(X_2) = .9999$ and $f^*(X_1) = .0001$ were used, then X_1 events would not occur in a run of moderate size and the 'overbiased' estimate would turn out to be

$$I_0 = g(X_2) \cdot \frac{f(X_2)}{f^*(X_2)} = 100 \cdot \frac{.0001}{.9999} \approx .01 \ . \tag{2.22}$$

The proper modification for this example is to let X_1 and X_2 events occur with equal probability, $f^*(X_1) = f^*(X_2) = .5$. Then, the contribution from each history is

$$g(X_1)\frac{f(X_1)}{f^*(X_1)} = .01 \cdot \frac{.9999}{.5} \approx .02 = g(X_2)\frac{f(X_2)}{f^*(X_2)} = 100 \cdot \frac{.0001}{.5} \tag{2.23}$$

and the final estimate from a small sample would be

$$\hat{I} = .02 \tag{2.24}$$

The above example is somewhat extreme but illustrates the general nature of most simulations where variance reduction is needed. The underlying distribution is highly skewed with the large majority of cases making little or no contribution to the final answer while a small number of cases can make large contributions. In both the 'overbiased' and 'underbiased' example, the final estimates were smaller than the correct value and this is also a general characteristic of such cases. In the example, if a set of 100 histories was simulated using crude Monte Carlo, then most likely there would be no X_2 events observed and the (incorrect) estimate would be .01. Once in every 100 sets of 100 histories, a single X_2 event would be simulated. For that set of histories the estimate would be

$$\hat{I}'_u = 1/100[99 \cdot .01 + 1 \cdot 100] \approx 1.01 \quad , \tag{2.25}$$

a number very much larger than the correct value. (Notice that this makes the estimation average out correctly in the long run.) Unfortunately, at this stage the human factor enters the problem. Most users confronted with several similar runs giving values of .01 and one run giving 1.01 will decide that the 1.01 estimate was the result of some input mistake or computer error, and throw out that run.

In this example the variance estimates produced would be zero for all runs except the one in a hundred which had a mean value of 1.01. For this case the relative standard deviation would be almost 100%, a sure sign of insufficient sampling.

Therefore caution is recommended in simulations where most histories contribute a small bit to the answer but a few histories contribute a large value, and complete faith should not be placed in estimates of variance especially when the results are smaller than expected or if the possibility of overbiasing or underbiasing is suspected.

3. VARIANCE REDUCTION TECHNIQUES

To provide a reasonable presentation of variance reduction, it is imperative that some organization be given to relate the various techniques. To this end the techniques or approaches for achieving variance reduction were grouped in the following three classes which were introduced in the previous section.

- Modification of the sampling process
- Use of analytical equivalence
- Specialized techniques

A summary of the specific variance reduction techniques in each of these classes was presented in Table 2.1.

The techniques which modify the sampling process effectively alter the probability distributions of the random variables so that the more significant events are observed more often. The use of analytical equivalence exploits analytical expressions and expected values to explain or approximate the majority of the phenomena, thus leaving only the most interesting portions of the process to be simulated. Specialized approaches encompass the more sophisticated techniques for achieving variance reduction.

In this section of the report, the techniques presented in each of these three classes is discussed in detail.

3.1 MODIFICATION OF THE SAMPLING PROCESS

Variance reduction techniques in this class include:

- Importance Sampling
- Russian Roulette and Splitting
- Systematic Sampling
- Stratified Sampling

These have several common characteristics in that they all reduce the variance of the estimate by sampling from a probability distribution different from the true physical distribution. In this way more of the interesting events will be observed, i.e., more of the events that contribute to the result being estimated will be observed and less computing time will be spent on events of no importance to the results. These techniques also preserve the actual physical process of the system in the simulation mode. Only the probabilities are altered; the flow of events remains essentially the same.

3.1.1 Importance Sampling(3,4,5,7,12,13,14,16,17,18,19,20,22,26,28,29,34,35,36,37)

3.1.1.1 General Concepts

Under this scheme the sampling distributions which would be used in the direct simulation are replaced with ones which force the sampling into more interesting, or important regions. For instance, in tossing a pair of dice, if one is interested in the occurrence of a three, one could weight or bias each die toward the numbers one and two. The biasing of the sampling distributions is done in a known fashion so that this information can be used to alter the computation of the results so as to unbias the answers.

Mathematically the importance sampling idea can be illustrated by considering a Monte Carlo estimate of a parameter I where

$$I = E[g(x)] = \int g(x)f(x)dx \quad . \tag{3.1}$$

The direct or straightforward Monte Carlo procedure would be as follows:

- Select a random sample $X_1, \ldots, X_N$ from the distribution with density $f(x)$
- Estimate $\hat{I}$ using

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} g(X_i) \quad . \tag{3.2}$$

As indicated in Section 2.3, the sample variance for this estimate is given by

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} [g(X_i) - \hat{I}]^2 = \frac{N}{N-1} \left[\frac{1}{N} \sum_{i=1}^{N} g^2(X_i) - \hat{I}^2 \right] \tag{3.3}$$

Now suppose the sampling was not from f(x), but rather from a distribution f*(x). Then it is clear from (3.1) that I may be expressed as

$$I = \int \frac{g(x)f(x)}{f^*(x)} f^*(x)dx \tag{3.4}$$

where it is assumed f*(x) does not go to zero when g(x)f(x) is not zero.

Now, if a sampling procedure were set up which selected a random sample $X_1, \ldots, X_N$ from f*(x), then the new estimator for I would be given by

$$\hat{I}_1 = \frac{1}{N} \sum_{i=1}^{N} \frac{g(X_i)f(X_i)}{f^*(X_i)} \tag{3.5}$$

Thus, when X_i is selected from f*(x), the sample is weighted by $\frac{f(X_i)}{f^*(X_i)}$ in the final result. Also, the sample variance is given by

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left[\frac{g(X_i)f(X_i)}{f^*(X_i)} - \hat{I}_1 \right]^2 = \frac{N}{N-1} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[\frac{g(X_i)f(X_i)}{f^*(X_i)} \right]^2 - \hat{I}_1^2 \right\} \tag{3.6}$$

It is of interest to consider the expected value of the square of the difference between $\hat{I}_1$ and I. That is,

$$E[(\hat{I}_1 - I)^2] = E\left[\left(\frac{1}{N}\sum_{i=1}^{N}\frac{g(X_i)f(X_i)}{f^*(X_i)} - I\right)^2\right]$$

$$= \frac{1}{N}\left\{\int\left[\frac{g(x)f(x)}{f^*(x)}\right]^2 f^*(x)dx - I^2\right\} \qquad (3.7)$$

Now it is seen that if[+]

$$f^*(x) = \frac{g(x)f(x)}{I} \qquad (3.8)$$

then $E[(I_1 - I)^2] = 0$, a desirable situation. But this implies the ridiculous condition that I is known. (This is the extreme situation indicated previously in that if the answer is known, a sampling scheme can be developed with expected zero variance.) However, (3.8) does suggest that if something close to the form $\frac{g(x)f(x)}{I}$ can be conveniently selected for $f^*(x)$ then a large improvement in the simulation should be possible. For example, consider Fig. 3.1. which qualitatively shows $f(x)$ and $\frac{g(x)f(x)}{I}$. A reasonable sampling function $f^*(x)$ which approximates $\frac{g(x)f(x)}{I}$ is indicated. $f^*(x)$ is called the importance sampling function since it tends to emphasize the areas where the expression $\frac{f(x)g(x)}{I}$ is most important. $f^*(x)$ could be something as simple and easy to work with as an exponential or normal distribution. The aim of importance sampling can, therefore, be to concentrate the distribution of sample points in the parts of the interval which are most important. This demonstrates again the utilization of knowledge of the process to accomplish variance reduction. It is desirable of course that $f^*(x)$ be easy to work with (i.e., integrable) which is usually a conflicting requirement to having $f^*(x)$ as close to $\frac{g(x)f(x)}{I}$ as possible.

[+] Note that if $g(x)$ ever changes sign, a zero variance sampling function is not so easily obtained since $f^*(x)$ must be non-negative to be a density function.

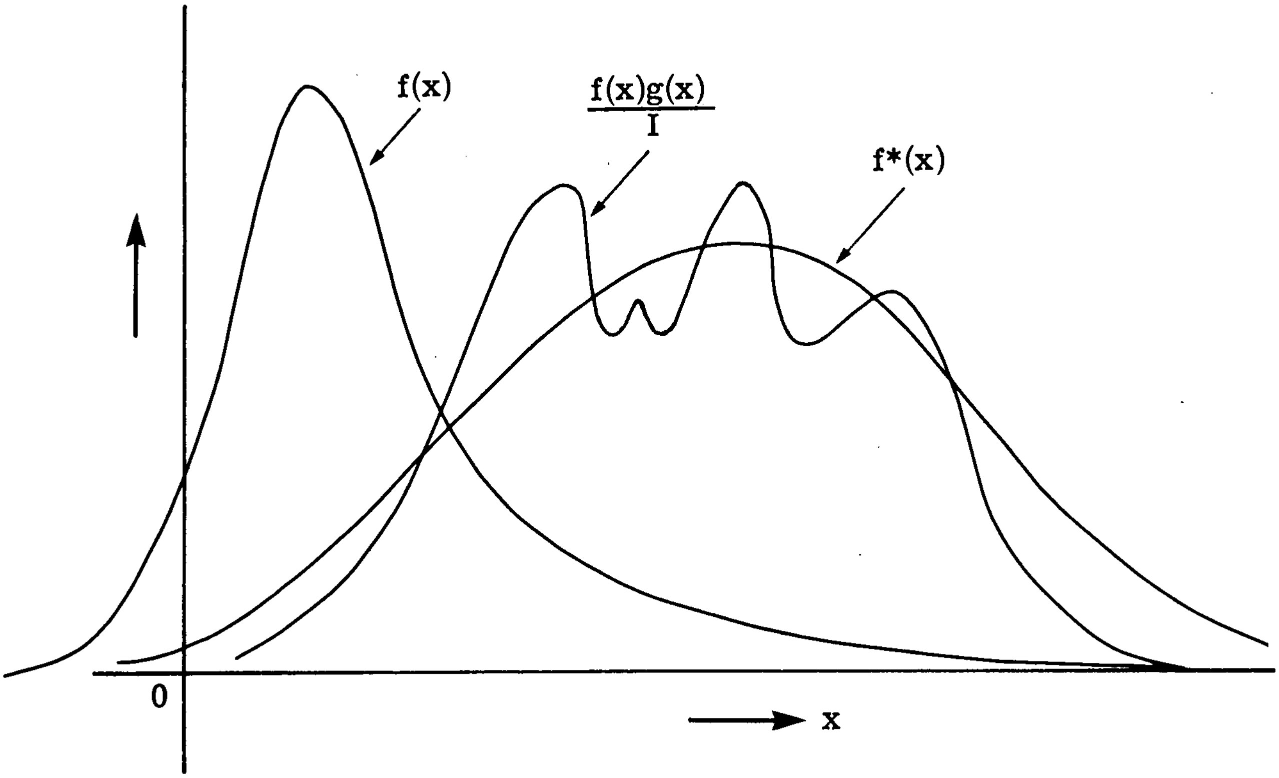


Fig. 3.1. Illustration of the Importance Sampling Concept.

3.1.1.2 Comparison of Importance Sampling with Straightforward Sampling

Unless carefully implemented, importance sampling has the potential of giving worse results than crude sampling. This can be seen by a comparison of the expected values of the sampling variances in the two situations. That is, from (3.3) and (3.6),

$$E[S^2 - S_1^2] = E[S^2] - E[S_1^2] = \int g^2(x) \left[1 - \frac{f(x)}{f^*(x)}\right] f(x)dx \qquad (3.9)$$

There is no assurance (3.9) will be positive. Therefore, in selection of f*(x), a worse result could be obtained from the selection of f*(x) over f(x) as the sampling distribution. This can be avoided, however, by careful selection of the importance function f*(x).

3.1.1.3 Extensions of Importance Sampling Concepts

One extension of interest in variance reduction is in applications involving two or more variables. To see how an importance function can be developed in the general situation consider the integral

$$I = \int_{\vec{x}} g(\vec{x})f(\vec{x})d\vec{x} = \int_{\vec{x}} \frac{g(\vec{x})f(\vec{x})}{f^*(\vec{x})} f^*(\vec{x})d\vec{x} \qquad (3.10)$$

Now, if a random sample $\vec{X}_1, \ldots, \vec{X}_N$ is obtained from the importance function $f^*(\vec{x})$, then the estimator for I is

$$\hat{I}_1 = \frac{1}{N} \sum_{i=1}^{N} \frac{g(\vec{X}_i)f(\vec{X}_i)}{f^*(\vec{X}_i)} \qquad (3.11)$$

The sample variance is of the same form as (3.6).

As in (3.7), consider

$$E[(\hat{I}_1 - I)^2] = E\left[\left\{\frac{1}{N}\sum_{i=1}^{N} \frac{g(\vec{X}_i)f(\vec{X}_i)}{f^*(\vec{X}_i)} - I\right\}^2\right] = \frac{1}{N}\left\{\int_{\vec{x}} \left(\frac{g(\vec{x})f(\vec{x})}{f^*(\vec{x})}\right)^2 f^*(\vec{x})d\vec{x} - I^2\right\}$$

$$(3.12)$$

As in (3.7), the "best" (i.e., when $E[(I_1 - I)^2] = 0$) importance function to select is

$$f^*(\vec{x}) = \frac{g(\vec{x})f(\vec{x})}{I} \qquad (3.13)$$

The arguments for selecting $f^*(\vec{x})$ is, therefore, identical to those used for selection of $f^*(x)$. However, in practice it is generally difficult to develop $f^*(\vec{x})$ due to the multidimensional aspects. An alternate approach is to try to select some sort of conditional importance function. For example, suppose $\vec{x} = (x, y)$. Then an importance sampling function for x, say $f^*(x)$ can be developed as follows:

$$I = \int_{x,y} g(x,y)f(x,y)dxdy = \int_{x,y} g(x,y)f(x)f(y\,|x)dxdy$$

$$= \int_{x,y} \frac{g(x,y)f(x)}{f^*(x)} f^*(x)f(y|x)dxdy \quad . \tag{3.14}$$

Now, if $X_1,\ldots,X_N$ is selected from $f^*(x)$ and $Y_1,\ldots,Y_N$ selected from $f(y\,|X_i)f^*(X_i)$, then the estimator for I is

$$\hat{I}_1 = \frac{1}{N}\sum_{i=1}^{N} \frac{g(X_i,Y_i)f(X_i)}{f^*(X_i)} \quad . \tag{3.15}$$

The sample variance in this case is

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{g(X_i,Y_i)f(X_i)}{f^*(X_i)} - \hat{I}_1\right]^2$$

$$= \frac{1}{N-1}\left\{ \frac{1}{N}\sum_{i=1}^{N}\left[\frac{g(X_i,Y_i)f(X_i)}{f^*(X_i)}\right]^2 - \hat{I}_1^2\right\} \tag{3.16}$$

In a manner similar to that used to arrive at (3.12) it is easy to see that

$$E[(\hat{I}_1-I)^2] = \frac{1}{N}\left\{\int_{x,y}\left[\frac{g(x,y)f(x)}{f^*(x)}\right]^2 f^*(x,y)dxdy - I^2\right\}$$

$$= \frac{1}{N}\left\{\int_x \frac{f^2(x)}{f^*(x)}\int_y g^2(x,y)f(y|x)dydx - I^2\right\} \tag{3.17}$$

But

$$E[g^2(x,y)\,|x] = \int_y g^2(x,y)f(y|x)dy \tag{3.18}$$

so the "best" importance function is

$$f^*(x) = \frac{f(x)\left\{E[g^2(x,y)|x]\right\}^{1/2}}{\int\left\{E[g^2(x,y)|x]\right\}^{1/2} f(x)dx} \tag{3.19}$$

which reduces (3.17) to $\frac{1}{N}\left\{[\int\left\{E[g^2(x,y)|x]\right\}^{1/2} f(x)dx]^2\right\}$.

In the general multidimensional case, it follows that the importance function for x should be

$$f^*(x) = \frac{f(x)\left\{E[g^2(x,\vec{y})|x]\right\}^{1/2}}{\int\left\{E[g^2(x,\vec{y})|x]\right\}^{1/2} f(x)dx} \tag{3.20}$$

where $\vec{y}$ refers to all the random variables except x.

The estimator for I and the sample variance are given respectively by expressions similar to (3.15) and (3.16).

The selection of the "best" importance function implies of course that the answer being sought is known. Thus, it is clear that the arbitrary selection of the best importance function would be a matter of luck. However, an understanding of the above formulations can lead to guidance to selecting an importance function. For example consider (3.20). In this case it may be possible to obtain an estimate for $E[g^2(x,y)|x]$ by performing a simulation for fixed values of x and selecting an approximate form for the results. This and many other variations become readily evident when serious considerations of importance sampling are undertaken. General guidelines for achieving such benefits are outlined in Part II of this document.

3.1.1.4 General Areas of Applicability for Importance Sampling

Application of the importance sampling technique can be very useful in simulations which are attempting to estimate very low probability events. One of the major areas to which this method has been applied is in nuclear physics in calculating probabilities concerning nuclear particle behavior. Examples are estimating the probability of penetrating a shield or barrier or analysis of the wandering of particles within nuclear reactors. Application of these techniques can also prove fruitful in problems which are more oriented towards operations research. For example, in vulnerability studies of weapons, the number of critical hits on a target can be increased by reducing the circular error probability (CEP) of the weapon from that normally expected. Another application is in queueing problems where improvements in estimates for the waiting time can be achieved by increasing the arrival rate or increasing the service time.

The effectiveness of importance sampling techniques are, of course, directly related to the ability to select good importance sampling distributions. This, in turn, is related to what might be called a priori or beforehand knowledge of the process being simulated. In essence, if the answers to the questions being sought are approximately or qualitatively known, then very good importance functions can be determined. In less favorable situations, the use of importance sampling might involve an iterative simulation procedure. For example, results from an initial simulation might be used to generate importance sampling distributions in a second simulation. Such iterations could proceed through several stages.

It is also worth noting that in importance sampling, as is the case for most variance reduction procedures, the samples obtained from the resulting simulation may be less effective for estimating certain quantities than crude sampling. Since the importance functions are selected to increase the effectiveness of estimating specific quantities or parameters, the estimation of other parameters, not involved in this selection, can be greatly impaired by this procedure.

3.1.2 Russian Roulette and Splitting(12, 14, 16, 19, 20, 36)

3.1.2.1 General Concepts

This technique can be very effective in problems which are characterized by a series of events. Examples are random walk, random movement of a submarine on maneuvers, subsystems in series, etc.

Generally, simulation of a series process of this type can be structured such that during the simulation it can be examined at various stages. At one or more of these stages it may be possible to establish whether or not the process is in an interesting or uninteresting state. (Interesting means likely to contribute to the desired result.) If the state of a given stage is not of interest, then one might want to restrict further investigation; that is, kill off the process with a known probability (Russian Roulette). If, howevr, the process is in an interesting state, one may want to conduct additional investigations; that is, increase the number of simulations starting from that desirable situation (splitting).

This technique can also be particularly useful for simulations involving a large number of discrete situations. For example, consider a queueing system in which a large number of individuals are being tracked. Then at a certain stage in the problem, one of these individuals can be selected and removed from the system with probability p. If this individual is not removed from the system he is allowed to continue through the system with a weight $(1-p)^{-1} = 1/q$. This can be repeated with more individuals (with the same or different values of p) until the number of individuals being tracked is reduced to a desired size.

Conversely the number of individuals being tracked in the system can be increased by splitting. For example, suppose an individual has an assigned weight w, then he can be replaced by n individuals each having a weight $w' = w/n$. The n individuals can then proceed independently through the system, except that the weight assigned at the splitting must be maintained.

It should be evident from the above descriptions that Russian Roulette and splitting techniques can be useful when simulating events of low probability and thus its application can prove beneficial in many of the same situations where importance sampling may be indicated. Indeed, there is a great resemblance between the two methods in that both force the simulation into interesting areas by modification of the sampling distributions. The difference between the two is the method of choosing the important areas. Russian Roulette and splitting is an "after-the-fact" or passive approach which uses a straightforward simulation but limits or increases the sampling as a function of the events which occur during the simulation. Importance sampling, on the other hand, attempts to force the paths into the more interesting areas by a prior alteration of the underlying random process.

3.1.2.2 Application to a Two-Stage Problem

To illustrate some of the more fundamental aspects of Russian Roulette and splitting,consider the two-stage process in Fig. 3.2. Let X denote the random observations from the first stage, and Y denote the observations from the second. Suppose the parameter to be estimated is

$$I = E[g(x,y)] \tag{3.21}$$

Crude sampling would generate pairs of values $X_1, Y_1; \ldots ; X_N, Y_N$ and estimate I using

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} g(X_i, Y_i) \tag{3.22}$$

Suppose, however, it can be determined from the characteristics of the problem that certain values of X would probably lead to more interesting results than others. On this basis then, Russian Roulette and splitting would be implemented by dividing the first stage into the following two mutually exclusive sets of states:

R_1: The set of states where Russian Roulette is used and the simulation is terminated with probability $p = 1 - q$. If the simulation continues, the estimated parameter is weighted by $1/q$.

R_2: The set of states where splitting is employed by breaking each simulation reaching a point in R_2 into n simulations to be continued from this point in the process. The weight assigned to each new simulation is $1/n$ of the weight of the original simulation.

This procedure would be then repeated for N starting situations as shown in Fig. 3.3. It is clear the sampling process has been modified and thus the estimator must be adjusted accordingly. In this case the estimator becomes:

$$\hat{I} = \frac{1}{N} \left\{ \sum_{X_i \epsilon R_1} \frac{g(X_i, Y_i)}{q} + \sum_{X_i \epsilon R_2} \sum_{j=1}^{n} \frac{g(X_i, Y_j)}{n} \right\} \qquad (3.23)$$

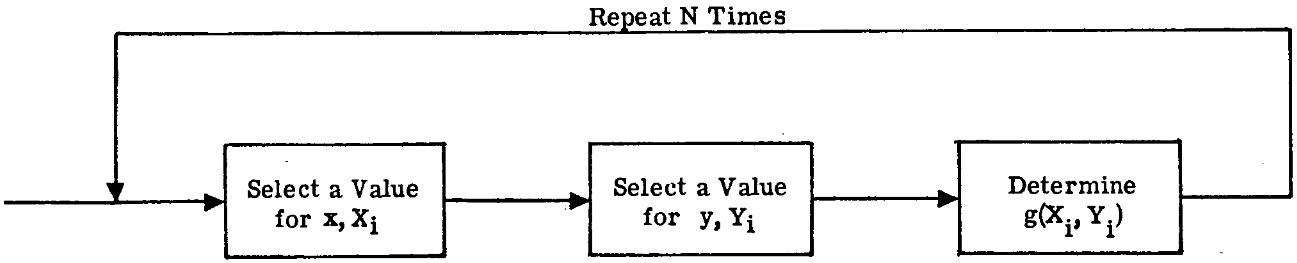
It can easily be shown that $\hat{I}$ is an unbiased estimator for I.

Estimation of the sample variance in this case is easy to accomplish. Defining I_i (i.e., $I_i = 0$, $g(X_i, Y_i)/q$, or $\sum_{j=1}^{n} g(X_i, Y_j)$ as the contribution to the estimator from history i, and since

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} I_i \qquad (3.24)$$

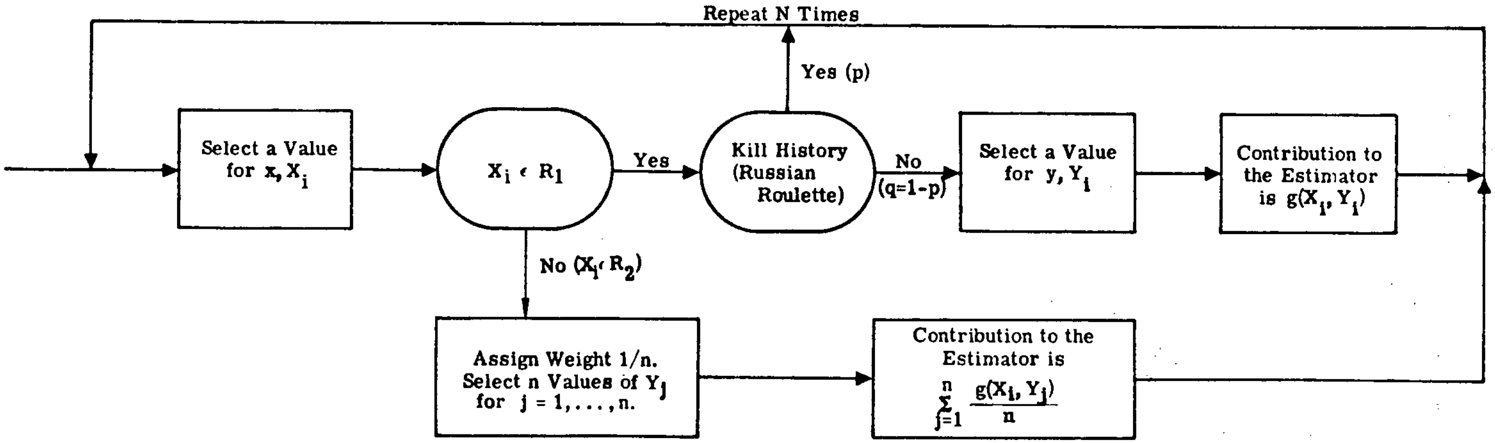
then the sample variance is estimated using

$$S^2 = \frac{N}{N-1} \left\{ \frac{1}{N} \sum_{i=1}^{N} I_i^2 - \hat{I}^2 \right\} \qquad (3.25)$$



$$\hat{I} = 1/N \sum_{i=1}^{N} g(X_i, Y_i)$$

Fig. 3.2. Crude Sampling Procedure for a Two-Stage Problem


Repeat N Times
Yes (p)
Select a Value for x, X_i
X_i ε R_1
Yes
Kill History (Russian Roulette)
No (q=1-p)
Select a Value for y, Y_i
Contribution to the Estimator is g(X_i, Y_i)
No (X_i ε R_2)
Assign Weight 1/n. Select n Values of Y_j for j = 1, . . . , n.
Contribution to the Estimator is Σ_{j=1}^{n} g(X_i, Y_j)/n

$$\hat{I} = 1/N\left[\sum_{X_i \epsilon R_1} \frac{g(X_i, Y_i)}{q} + \sum_{X_i \epsilon R_2} \sum_{j=1}^{n} \frac{g(X_i, Y_j)}{n}\right]$$

Fig. 3.3. Russian Roulette and Splitting for a Two-Stage Problem

Alternately, batching as described in Section 2.4.2 could be used, although (3.25) is recommended.

3.1.2.3 Application to a Three-Stage Problem

Although the basic concepts of Russian Roulette and splitting are as simple as presented above, they can be applied to rather large multistage problems. To illustrate this, consider the three-stage problem shown in Fig. 3.4 where it is shown within the context of a crude sampling. Assume Russian Roulette and splitting is applied between the first and second stages. The procedure may be accomplished as follows (see also Fig. 3.5).

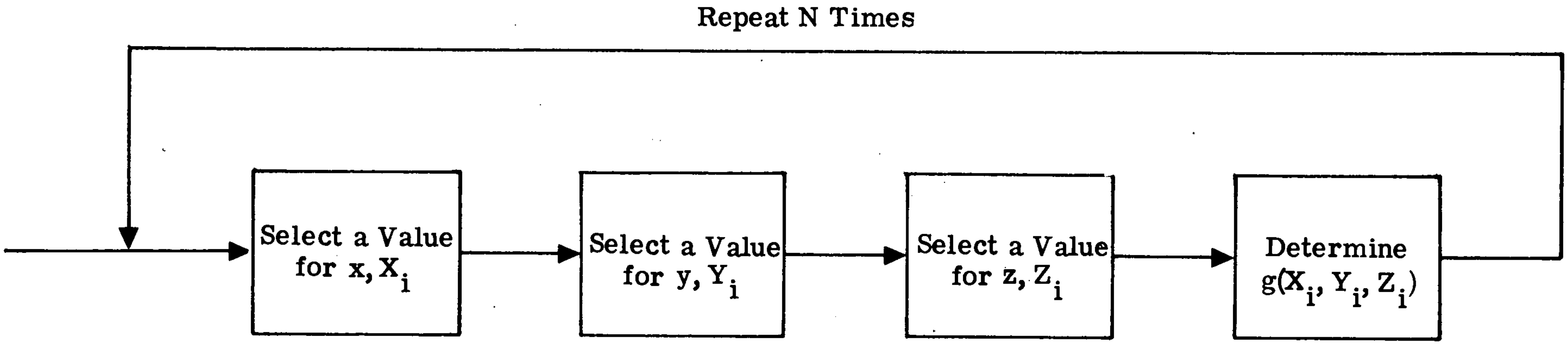
1. First generate a value for x, X_i. If $X_i \epsilon R_{11}$, the history is terminated with probability $p_1 = 1-q_1$(i.e., Russian Roulette). If the history is killed, there is no contribution to the estimator.

2. If the history is not killed, a value for y, Y_i, is selected. The history now has a weight $1/q_1$. If $Y_i \epsilon R_{21}$, the history is terminated (Russian Roulette) with probability $p_2 = 1-q_2$. If the history is not terminated here, a value of z, Z_i is generated. The weight of the history is then $(q_1q_2)^{-1}$ and the contribution to the estimate for I is

$$\frac{f(X_i, Y_i, Z_i)}{q_1q_2}$$

3. If the history is not killed on X (with weight $1/q_1$) and $Y_i \epsilon R_{22}$ then the history is split into n_2 histories. Next, n_2 values for Z; $Z_{i1}, \ldots, Z_{in_2}$ are generated and assigned weights of this history to the estimator is

$$\sum_{j=1}^{n_2} \frac{g(X_i, Y_i, Z_{ij})}{n_2q_1}$$


Repeat N Times
Select a Value for x, X_i
Select a Value for y, Y_i
Select a Value for z, Z_i
Determine $g(X_i, Y_i, Z_i)$

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} g(X_i, Y_i, Z_i)$$

Fig. 3.4. Crude Sampling in a Three Stage Problem

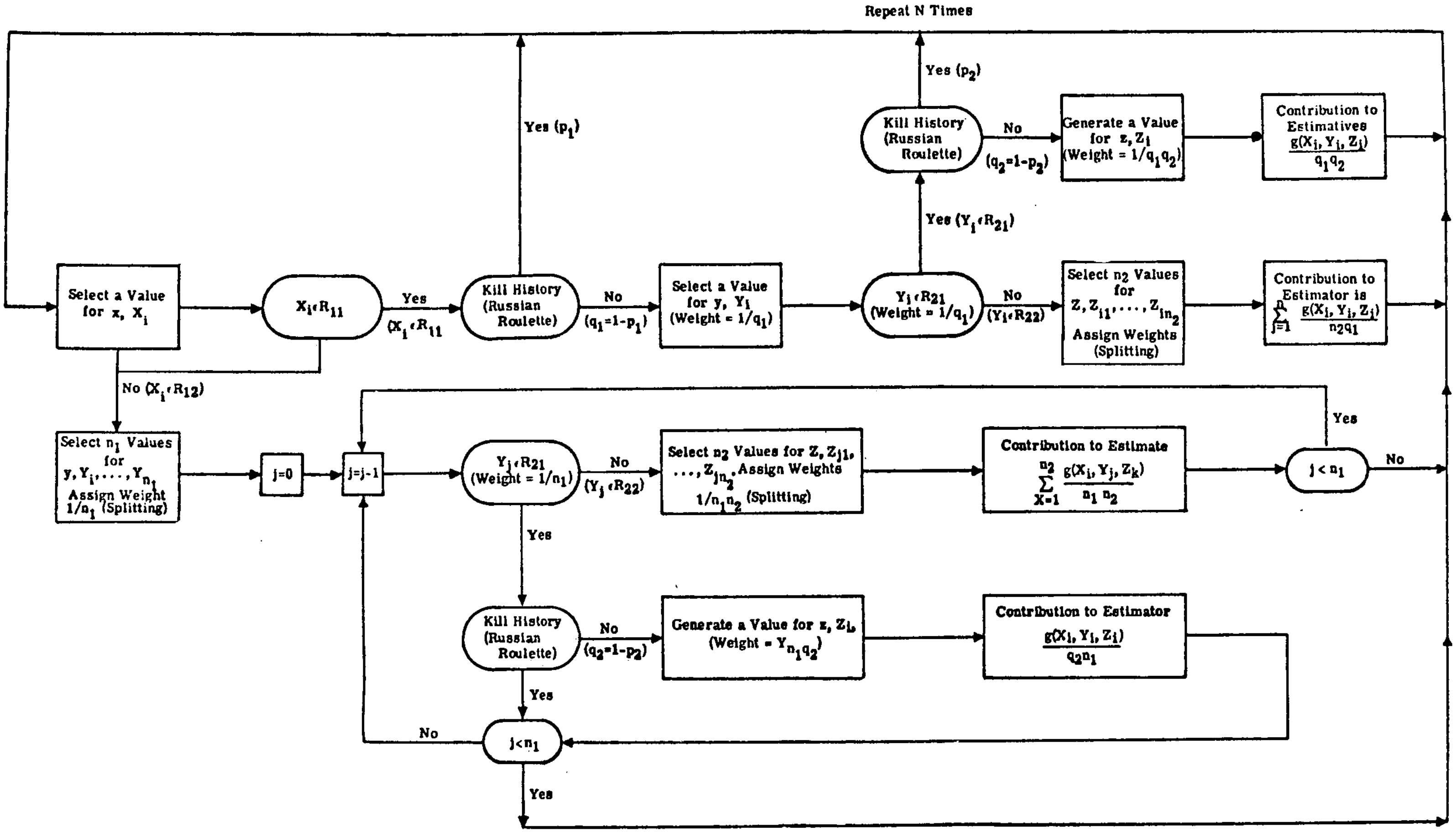

Repeat N Times
Yes (p₁)
Yes (p₂)
Kill History (Russian Roulette)
No
(q₂=1-p₂)
Generate a Value for z, Zi (Weight = 1/q₁q₂)
Contribution to Estimatives g(Xi, Yi, Zi) / q₁q₂
Yes (Yi ∈ R21)
Select a Value for x, Xi
Xi ∈ R11
Yes
(Xi ∈ R11
Kill History (Russian Roulette)
No
(q₁=1-p₁)
Select a Value for y, Yi (Weight = 1/q₁)
Yi ∈ R21 (Weight = 1/q₁)
No
(Yi ∈ R22)
Select n2 Values for Z, Zi1, ..., Zin2 Assign Weights (Splitting)
Contribution to Estimator is Σ_{j=1}^{n} g(Xi, Yi, Zj) / n2q1
No (Xi ∈ R12)
Select n1 Values for y, Yi, ..., Yn1 Assign Weight 1/n1 (Splitting)
j=0
j=j-1
Yj ∈ R21 (Weight = 1/n1)
No
(Yj ∈ R22)
Select n2 Values for Z, Zj1, ..., Zjn2, Assign Weights 1/n1n2 (Splitting)
Contribution to Estimate Σ_{X=1}^{n2} g(Xi, Yj, Zk) / n1 n2
Yes
j < n1
No
Yes
Kill History (Russian Roulette)
No
(q₂=1-p₂)
Generate a Value for z, Zi, (Weight = Yn1q2)
Contribution to Estimator g(Xi, Yi, Zi) / q2n1
Yes
No
j<n1
Yes

Fig. 3.5. Russian Roulette and Splitting for a Three-Stage Process

4. If $X_i \epsilon R_{12}$, then the history (with weight 1) is split into n_1 values $Y_1, \ldots, Y_{n_1}$ and assigned weights $1/n_1$.

5. Now, each Y_j(weight $= 1/n_1$), $j = 1, \ldots, n_1$ is considered in turn. If $Y_j \epsilon R_{21}$, the history is killed with probability $p_2 = 1 - q_2$. If killed, there is no contribution to the estimator. If the history is not killed here a value for Z, say Z_j, is selected. The weight of Z_j is $1/(n_1 q_2)$. The contribution to the estimator in this situation is now given by

$$\sum_{Y_j \epsilon R_{21}} \frac{g(X_i, Y_j, Z_j)}{n_1 q_2}$$

6. If Y_j (weight $1/n_1$) ϵR_{22}, this history is split into n_2 histories. Then n_2 values of Z are selected $Z_{j1}, \ldots, Z_{jn_2}$ and a weight of $\frac{1}{n_1 n_2}$ is assigned to each. The contribution to the estimator along this path is

$$\sum_{Y_j \epsilon R_{22}} \sum_{k=1}^{n_2} \frac{g(X_i, Y_j, Z_{jk})}{n_1 n_2}$$

This procedure is repeated N times as indicated in Fig. 3.5. For each X_i selected then the contribution to the estimator is, for $X_i \epsilon R_{11}$,

$$\hat{I}_i = \sum_{Y_i \epsilon R_{21}} \frac{g(X_i, Y_i, Z_i)}{q_1 q_2} + \sum_{Y_i \epsilon R_{22}} \sum_{j=1}^{n_2} \frac{g(X_i, Y_i, Z_j)}{n_2 q_1} \qquad (3.26)$$

and if $X_i \epsilon R_{12}$

$$\hat{I}_i = \sum_{Y_j \epsilon R_{21}} \frac{g(X_i, Y_j, Z_j)}{n_1 q_2} + \sum_{Y_j \epsilon R_{22}} \sum_{k=1}^{n_2} \frac{g(X_i, Y_j, Z_{jk})}{n_1 n_2} \qquad (3.27)$$

Assuming the entire process is repeated N times (i.e.,: N starting values for x, $X_1, \ldots, X_n$ are selected) the estimator is

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N}\hat{I}_i = \frac{1}{N}\left[\sum_{X_i \epsilon R_{11}}\left\{\sum_{Y_i \epsilon R_{21}} \frac{g(X_i, Y_i, Z_i)}{q_1 q_2} + \sum_{Y_i \epsilon R_{22}}\sum_{j=1}^{n_2}\frac{g(X_i, Y_i, Z_j)}{n_2 q_1}\right\}\right.$$

$$\left. + \sum_{X_i \epsilon R_{12}}\left\{\sum_{Y_j \epsilon R_{22}}\sum_{k=1}^{n_2}\frac{g(X_i, Y_j, Z_{jk})}{n_1 n_2} + \sum_{Y_j \epsilon R_{21}}\frac{g(X_i, Y_j, Z_j)}{n_1 q_2}\right\}\right] \qquad (3.28)$$

The procedure, although rather complex to write down as formal expressions can be seen to be rather straightforward.

As in the two-stage case, the best estimation to use for the sample variance is

$$S^2 = \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}\hat{I}_i^2 - \hat{I}^2\right] \qquad (3.29)$$

S^2 can then be used to compare the efficiency of Russian Roulette and Splitting to the crude sampling.

3.1.2.4 Weight Standards for General Application

For a general application of Russian Roulette and splitting, it is best to introduce the concept of weight standards. Let us presume that the problem has been broken up into several regions, $R_1, R_2, \ldots, R_N$. (These 'regions' do not necessarily denote geometrical volumes, but rather ranges of the random variables that describe a state in the system being studied.) For each region, there will be a high weight, w_{Hi}, a low weight, w_{Li}, and an average weight, w_{Ai}. Now, whenever a history enters region i, the current weight, w, of the history is compared to the weight standards as follows:

1. If $w < w_{Li}$, Russian Roulette is implemented as follows:

- With probability $1 - \frac{w}{w_{Ai}}$, the history will be killed.
- With probability $\frac{w}{w_{Ai}}$, the history will survive with a new weight of w_{Ai}. (Note that the expected weight surviving from this process is w, which it must be to conserve weights).

2. If $w > w_{Hi}$, splitting is implemented as follows:

- Find n such that $w - nw_{Ai} < w_{Ai}$.
- Create n histories which start from this point with a weight w_{Ai}.
- With probability $\frac{w-nw_{Ai}}{w_{Ai}}$, create one more daughter history to start from this point with a weight w_{Ai}. (This procedure conserves the expected weight while making all histories start from this point with the same weight, w_{Ai}.)

3. If $w_{Li} < w < w_{Hi}$, do nothing to the history.

The underlying assumption in the above procedure is that each region describes a volume of approximately constant importance. The importance varies from region to region in a manner inversely proportional to the average weight, w_A. Thus, histories moving into a region of higher importance (lower weight) will be split while those moving into a region of lower importance (higher weight) will suffer Russian Roulette. For maximum efficiency in allocating computer time, all histories in a region of constant importance should have the same weight. The use of a fixed average weight standard, rather than fixed splitting parameter, n, or fixed Russian Roulette probability, p, ensures this in a multiregion setting.

The high and low weight standards, w_H and w_L, are only used to define upper and lower limits for triggering the Russian Roulette and splitting processes. If Russian Roulette and splitting are the only variance reduction techniques being employed and the history weights are not otherwise being varied, it is probably best to set $w_H = w_A = w_L$. On the other hand, if there are other techniques in use which are changing the history weights, it is best to put a spread between w_H and w_L within which the weight is allowed to vary. If the spread between w_H and w_L is too small, there will be a loss of efficiency due to computing time spent in the bookkeeping involved with frequent Russian Roulette and splitting actions. Conversely, if the spread is too large, there will be a loss of efficiency as equal amounts of computing time are expended on histories with unequal weight.

To estimate expected values and variances, the contribution from a single original history is computed using

$$\hat{I}_i = \sum_j g(\vec{X}_{ij})w(\vec{X}_{ij}) \qquad (3.30)$$

where the summation runs over all contributions from split histories, j, which originated from the same initial history, i. Then the final estimate of I for N initial histories is

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} \hat{I}_i \qquad (3.31)$$

and the sample variance is given by (3.29).

3.1.2.5 Selection of Criteria for Russian Roulette and Splitting

One difficulty in the application of Russian Roulette and splitting is in the selection of values for the parameters being used, either weight standards or Russian Roulette kill probability and number of splittings. The ideal approach would be to select these parameters to minimize the variance in the estimate as was done with importance sampling; however, this is generally not practical. Consequently, intuitive information along with practical limitations (e.g., computer storage) and simplifications must be resorted to. For example, if it is 'felt' that a certain range of Y is twice as important than the remainder of the range of Y, then a splitting with $n = 2$ of histories inside the important range or a Russian Roulette kill factor of .5 outside the range would be not unreasonable. A clue to the optimum standards to be selected is given by the results of analysis for importance sampling (3.20) or stratified sampling (see Section 3.1.4). In both cases the resulting weights will be proportional to $E[g^2(x)]^{-1/2}$. Thus the weight standards in a given region should be inversely proportional to the root mean square average of the 'pay-off' or result function i.e., weight standards should be high in regions of low value and low in regions of high value.

3.1.3 Systematic Sampling(7, 12, 14, 20, 23, 24, 36)

3.1.3.1 General Concept

Systematic sampling is a procedure that modifies selection from the sample space in a somewhat structured manner. This serves to reduce the random variation that is introduced into the results when crude Monte Carlo sampling is used. An important characteristic of systematic sampling is that if used it will always result in a reduction in variance from the that obtained using crude sampling. Also, the method rarely involves any significant effort to implement. Unfortunately, the improvement is generally less than impressive although as a general rule it should be used whenever the opportunity arises.

Its potential application can generally be associated with initial or starting conditions in a problem. For example, systematic sampling could be applied to the distribution of interarrival times of individuals entering a queueing system, the initial position of a submarine in simulation of an ASW exercise, etc. Generally, any Monte Carlo problem which has a probability distribution to characterize the initial conditions can be considered as a candidate for application of systematic sampling.

Two methods commonly used for systematic sampling will be described below. As will be seen, systematic sampling is similar to stratified sampling to be described next. Stratified sampling can be considered an optional form of systematic sampling.

In each of the methods to be presented below, the usual form of the integral,

$$I = \int_{-\infty}^{\infty} g(x)f(x)dx \qquad (3.32)$$

will be considered.

3.1.3.2 Method I for Systematic Sampling

In the first method for applying systematic sampling, assume the range of the density function f(x) is broken up into N equal regions each having an area 1/N (N should typically vary between 5 and 50). This scheme is shown in Fig. 3.6 for both the probability density f(x) and the cumulative distribution function $F(X) = \int_{-\infty}^{X} f(x)dx$

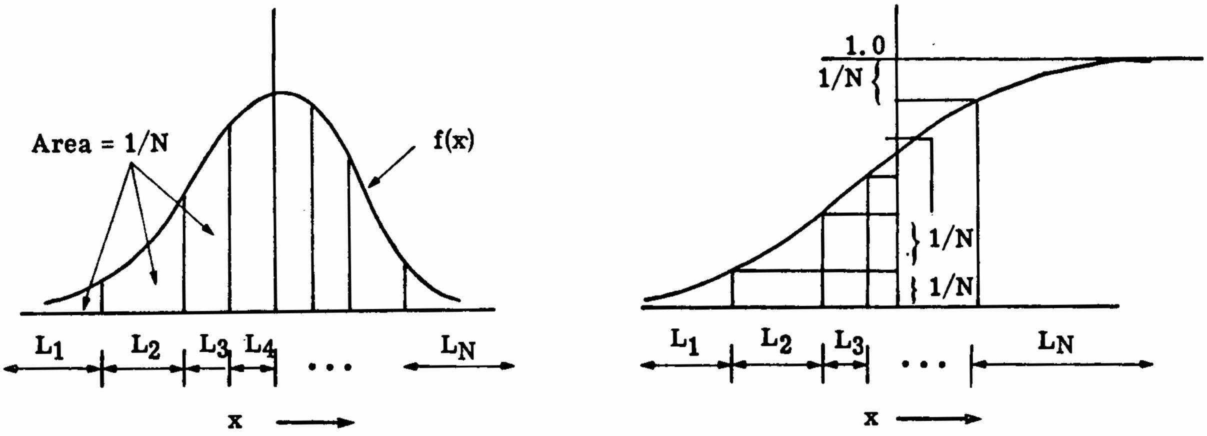


Fig. 3.6. Interval characterization for systematic sampling

It is clear that

$$\frac{1}{N} = \int_{X \epsilon L_j} f(x)dx \quad ; \quad j = 1, \ldots, N \tag{3.33}$$

Now, assume a sequence of random numbers, $R_1, \ldots, R_n$ is selected from the uniform distribution on the interval (0, 1). This form of systematic sampling will then generate the following sequence of numbers

$$R_{ij} = \frac{R_i}{N} + \frac{(j-1)}{N} \quad ; \quad \begin{matrix} i = 1, \ldots, n \\ j = 1, \ldots, N \end{matrix} \tag{3.34}$$

For each value of i, this procedure effectively assigns a value of R_{ij} to each interval j.

The next step is to determine X_{ij} from

$$R_{ij} = \int_{\infty}^{X_{ij}} f(x)dx \quad ; \quad \begin{matrix} i = 1, \ldots, n \\ j = 1, \ldots, N \end{matrix} \tag{3.35}$$

The estimator for I is

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{n} \hat{I}_i = \frac{1}{nN} \sum_{i=1}^{n} \sum_{j=1}^{N} g(X_{ij}) \tag{3.36}$$

where

$$\hat{I}_i = \frac{1}{N} \sum_{j=1}^{N} g(X_{ij}) \quad ; \quad i = 1, \ldots, N \tag{3.37}$$

is the contribution from the ith batch of histories.

The sample variance is computed using

$$S^2 = \frac{n}{n-1}\left\{1/n\sum_{i=1}^{n}\hat{I}_i^2 - \hat{I}^2\right\} \tag{3.38}$$

3.1.3.3 Method II for Systematic Sampling

A second and generally better method to perform systematic sampling is to allocate N independent samples to each interval defined in Fig. 3.6 rather than scale each random number R_i into N intervals. This can be accomplished by selecting R_{ij}; $i = 1, \ldots, n$; $j = 1, \ldots, N$ random numbers from a uniform distribution U(0,1). Then, n random numbers are allocated to each of the N intervals using

$$R'_{ij} = \frac{j - R_{ij}}{N} \quad ; \quad \begin{array}{l} i = 1, \ldots, n \\ j = 1, \ldots, N \end{array} \tag{3.39}$$

The values of X_{ij} are then determined from

$$R'_{ij} = \int_{-\infty}^{X_{ij}} f(x)dx \tag{3.40}$$

The estimators for the sample mean and variance in this case are given by 3.36 and 3.38 respectively.

Of the two methods described above, the second will always give the better answer in the sense of smaller variance. However, Method II requires that a larger number of random samples be selected from U(0,1). Generally it is recommended that Method II be used in spite of the slightly additional computation effort required. In both cases, the efficiency of systematic sampling compared to crude sampling is approximately proportional to N^2.

3.1.4 Stratified Sampling(1, 6, 7, 11, 12, 14, 15, 20, 24, 27, 32, 33)

3.1.4.1 General Concept

Stratified sampling (sometimes called quota sampling) is similar to systematic sampling with additional considerations directed toward structuring the sampling process so that regions of large variance will receive more samples. In this sense, therefore, stratified sampling seeks to combine the systematic and importance sampling schemes. Alternately, stratified sampling can be viewed as a special case of systematic sampling where optimal distribution of samples is attempted.

Generally, all the problem characteristics that serve to define the applicability of systematic sampling apply to stratified sampling. However, if additional information on which portions of the sampling distribution tend to contribute more to the total variance is available, additional reduction in the variance can be achieved using stratified sampling.

Assume the sampling range for f(x) is broken up into N regions of length $L_1, \ldots, L_N$. In this case, however, assume L_j is selected according to some specified P_j where

$$P_j = \int_{X \epsilon L_j} f(x)dx \quad ; \quad j = 1, \ldots, N \tag{3.41}$$

Schematically this structure is similar to that in Fig. 3.6. In fact, if $P_j = 1/N$,then this structuring would be the same as systematic sampling. A general rule to follow for selecting the P_j is to select them such that the variation in g(x)f(x) is the same in each of the intervals.

Once the intervals $L_1, \ldots, L_N$ are selected, the next requirement is to define the number of samples to assign to each interval.

More specifically, let n_j be the number of samples assigned to interval L_j where,

$$\sum_{j=1}^{N} n_j = n \tag{3.42}$$

The n_j samples can be assigned to interval L_j as follows:

Select $R_{1j}, \ldots, R_{n_j j}$ from $U(0,1)$. Then, $X_{ij} \epsilon L_j$ are determined by

$$R_{ij} P_j + \sum_{\ell=1}^{j-1} P_\ell = \int_{-\infty}^{X_{ij}} f(x)dx \quad ; \quad i = 1, \ldots, n_j \tag{3.43}$$

An unbiased estimate for I is

$$\hat{I} = \sum_{j=1}^{N} \frac{P_j}{n_j} \left[\sum_{i=1}^{n_j} g(X_{ij}) \right] = \sum_{j=1}^{N} P_j \hat{I}_j \tag{3.44}$$

where

$$\hat{I}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} g(X_{ij}) \tag{3.45}$$

To see that (3.44) is unbiased consider

$$E(\hat{I}_j) = I_j = \int_{x \epsilon L_j} \frac{f(x)}{P_j} g(x)dx \tag{3.46}$$

from which it follows that

$$I = \sum_{j=1}^{N} P_j I_j \tag{3.47}$$

To select the n_j, consider

$$E[(\hat{I} - I)^2] = E\left[\left(\sum_{j=1}^{N} P_j\hat{I}_j - I\right)^2\right] = E\left[\left\{\sum_{j=1}^{N} P_j(\hat{I}_j - I_j)\right\}^2\right]$$

$$= \sum_{j=1}^{N} \frac{P_j^2\sigma_j^2}{n_j} \qquad (3.48)$$

where

$$\sigma_j^2 = \int_{X_j \in L_j} \frac{f(x)}{P_j}[g(x) - I_j]^2 dx = n_j E[(\hat{I}_j - I_j)^2] \qquad (3.49)$$

is the variance in the interval L_j.

Now, if the n_j are selected to minimize (3.48) subject to (3.42), then it can be shown (24) that n_j should be selected to satisfy

$$n_j \simeq \frac{nP_j\sigma_j}{\sum_{j=1}^{N} P_j\sigma_j} \qquad (3.50)$$

Thus, the sample size in each interval should be selected to be proportional to the fraction of the variance in each interval. The obvious difficulty is, of course, that the σ_j^2 are not known. However, they can be estimated using

$$S_j^2 = \frac{1}{n_j'-1}\sum_{i=1}^{n_j'} [g(X_{ij}) - \hat{I}_j]^2 = \frac{n_j'}{n_j'-1}\left[\frac{1}{n_j}\sum_{i=1}^{n_j'} g^2(X_{ij}) - \hat{I}_j^2\right] \qquad (3.51)$$

where n_j' samples are arbitrarily selected in each interval, An iterative scheme can be structured to estimate n_j as the sampling is carried out.

The sample variance using stratified sampling may be estimated using

$$S^2 = \sum_{j=1}^{N} \frac{P_j^2}{n_j - 1} \sum_{i=1}^{n_j} [g(X_{ij}) - \hat{I}_j]^2 = \sum_{j=1}^{N} \frac{n_j P_j^2}{n_j - 1} \left[\frac{1}{n_j} \sum_{i=1}^{n_j} g^2(X_{ij}) - \hat{I}_j^2 \right] \qquad (3.52)$$

or a batching procedure (see Section 2.4.2) could be used.

As in the case of systematic sampling, the efficiency of stratified sampling in comparison with crude sampling is $\simeq N^2$.

3.2 ANALYTICAL EQUIVALENCE TECHNIQUES

This group of variance reduction techniques is based on using prior knowledge of the processes involved to form analytical or approximate solutions to the problem being simulated. This is another means to utilize information about the process and is also based on the fact that it is generally beneficial to use analytical solutions to parts of the problem whenever sufficient prior knowledge allows. This may mean that a related process is solved exactly using analytical or other low variance techniques and that the difference between the exact and related processes is derived by Monte Carlo techniques. All of the techniques discussed below are based on this concept and many are very closely related in the principles and ideas involved.

3.2.1 Expected Value(18, 19, 20, 35, 36)

This technique is based on the fact that analytic determinations are usually preferred to results gained through simulation. Thus any portion of a process which can be analytically determined should be replaced by its analytical representation in the model whenever that can be done without losing an essential element from the simulation. The name "expected value" refers to the basic notion that Monte Carlo simulation of any parameter is equivalent to estimating its expected value, i.e., evaluating an integral. Thus any portion of the simulation which can be evaluated analytically can be replaced by its expected value, and this is likely to improve the efficiency of the simulation.

To demonstrate the application of the expected value technique, consider the two-dimensional integration.

$$I = \int\int f(x, y)g(x, y)dxdy \quad . \tag{3.53}$$

This could, for example, be a two-stage problem such as that described under Russian Roulette and splitting (Section 3.1.2) and shown schematically in Fig. 3.7.

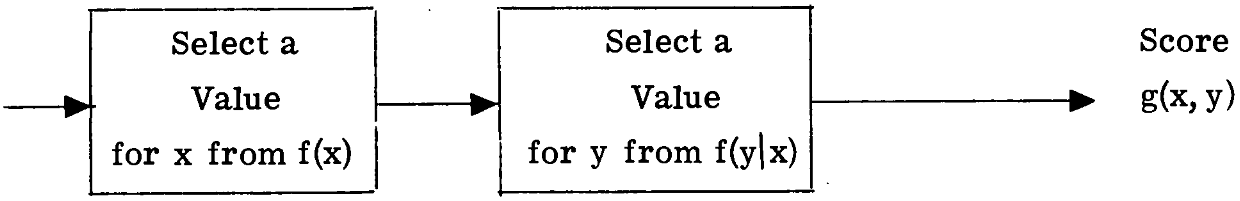


Fig. 3.7. Crude simulation of a two-step process

where first a random sample X is selected from the density function f(x) and then a random sample Y is selected from the conditional distribution f(y|X). Now, if this is repeated N times, the crude Monte Carlo estimator for I is

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} g(X_i, Y_i) \quad . \tag{3.54}$$

Assume, however, that it is possible to compute the expected value of the conditional probability in the second stage given the result of the first stage, X_i. That is, suppose $E[g(y|x)]$ is known analytically.

Then the simulation could be performed by simply generating N values for X , $X_i, \ldots, X_N$ and using the expected value estimator given by

$$\hat{I}_E = \frac{1}{N} \sum_{i=1}^{N} E[g(y|X_i)] \tag{3.55}$$

This is an unbiased estimator for I since $E[\hat{I}_E] = I$.

The sample variance is given by

$$S^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (E[g(y|X_i)] - \hat{I}_E)^2 = \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N} E^2[g(y\ X_i)] - \hat{I}_E^2\right] \qquad (3.56)$$

It is easy to show that this approach will always give results that are better than the straightforward Monte Carlo procedure.

The trivial nature of the above description should not be interpreted as indicating limited potential for this technique. Indeed, its application often results in a difficult simulation becoming an easy one. Furthermore, it can find application in a vast number of problems. For example, in computing the average time spent in a queueing system, the simulation of the server(s) can be replaced by the mean service time. In radiation transport, the stochastic absorption of particles is almost always replaced by a weighting system involving the expected absorption percentage.

It is not always possible, of course, to calculate the expected value of a process in the simulation - if all expected values could be calculated analytically there would be no need for simulation. Even if the expected value can be calculated, it may not be possible to replace the process by its expected value. The entire distribution involved in the process may be important in the simulation, or in other words, the second and higher moments may be important to the final answer and not just the first moment or expected value. In a few cases, replacing the stochastic process can actually reduce the efficiency. This may be true whenever the stochastic process is one of the decision points where the simulation may be terminated. Replacing the termination decision by its expected value involves assigning a weight to the history and modifying that weight to allow for the expected percentage of terminations at each decision point. When the survival probability is small, this can lead to computing time being wasted in simulating a history which may have a vanishingly small weight after passing a few decision points.

Several significant aspects must be considered before the expected value techniques can be implemented. Generally, these are:

1. Identify those parts of the overall simulation for which the expected value can be determined efficiently.
2. For each such process identified in 1., a determination must be made as to whether the random nature of the process is an essential element of the overall simulation or whether it may be replaced by a deterministic process without loss of desired realism in the model, i. e., does the fact that the stochastic process results in a range of outcomes rather than a single expected value affect the final answers of the simulation? Or, put in different terms, replacing the random process by its expected value preserves the first moment of the distribution but alters all the higher order moments. If these higher order moments are important to the overall answer (e. g., in determining a probability distribution) then the stochastic process cannot be replaced by its expected value. On the other hand, if the higher order moments do not contribute to the final result, then replacement by the expected value can be considered. For a particular physical system, the determination of which stochastic elements are essential may depend on the particular parameters being estimated.
3. Finally, it must be determined if the replacement of the random process by its expected value will increase the efficiency. This is generally true, but not always. If the process in question is a branch point where the history may go in either of two (or more) directions, then replacing the stochastic event by its expected value requires splitting the history with each part going in one of the directions and carrying the probability of that branch as a weight. Should enough of these events be encountered the number of split histories which must be computed can easily expand beyond a reasonable bound. Alternatively, one of the branches of the decision can be to terminate the history; in this case the history is not split but continues from the branch point with a weight representing the survival probability. This can easily lead to histories with very low weights which usually represents a loss in efficiency in the calculation. Again, this determination is likely to depend on the particular parameters of interest in the calculation.

Once the decision has been made to replace the stochastic process by its expected value, the implementation depends on the role of the process in the overall simulation. Specifically,

1. If the process is one of selection of a random variable, then the process becomes merely a deterministic setting of the variable to its expected value and the simulation proceeds as before with no change in estimators.

2. If the process represents a decision between terminating or not terminating the history, then the history continues but with a reduced weight representing the probability of survival. That is,

$$w_{new} = w_{old} \cdot p_s \tag{3.57}$$

where p_s is the probability of survival (non-termination) at the decision point and w_{old} and w_{new} are the weights of the history before and after the replaced random process.

For any parameter being calculated, an estimate for each history can be made by summing the contributions from that history. That is,

$$\hat{I}_i = \sum_j w_{ij}\, g(X_{ij}) \tag{3.58}$$

where w_{ij} is the weight of the ith history at the time of the jth contribution to the final result. Then the final estimate and the sample variance are given by

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} \hat{I}_i \tag{3.59}$$

and

$$S^2 = \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N} \hat{I}_i^2 - \hat{I}^2\right] \tag{3.60}$$

If the contributions to a parameter from a history would have come from the terminations in the process which was replaced by its expected value, then the loss of weight at each such step is the proper estimate for the expected terminations. In this case we set

$$I_i = \sum_j (w_{old, ij} - w_{new, ij}) \cdot g(X_{ij}) = \sum_j w_{old, ij}(1-p_S) \cdot g(X_{ij}) \tag{3.61}$$

where j denotes the jth occurrence of the replaced event in the ith history. The estimators for I and S^2 remain as in (3.59) and (3.60) above.

3. If the process represents a decision between two or more branch points, then the history must be split and followed from that point on as two separate histories, each taking a different branch and carrying a weight equal to the probability of that branch. Parameters are estimated by summing weighted contributions from all daughter histories resulting from an original parent history, using formulas identical to (3.58), (3.59), and (3.60).

In cases 2 and 3 above, histories may develop weights which are very small. As this may entail spending a good deal of computing time calculating histories that can make only a trivial contribution to the result, the efficiency may be very low. To remedy this, Russian Roulette (see Section 3.1.2) can be used to eliminate those histories whose weights become too small.

Figure 3.9 shows a schematic flow of a multistage simulation when a branch process that is a possible termination point for a history is replaced by its expected value. This may be contrasted to Fig. 3.8 which shows the crude Monte Carlo approach to the same simulation and Fig. 3.10 which shows statistical estimation used on the same problem.

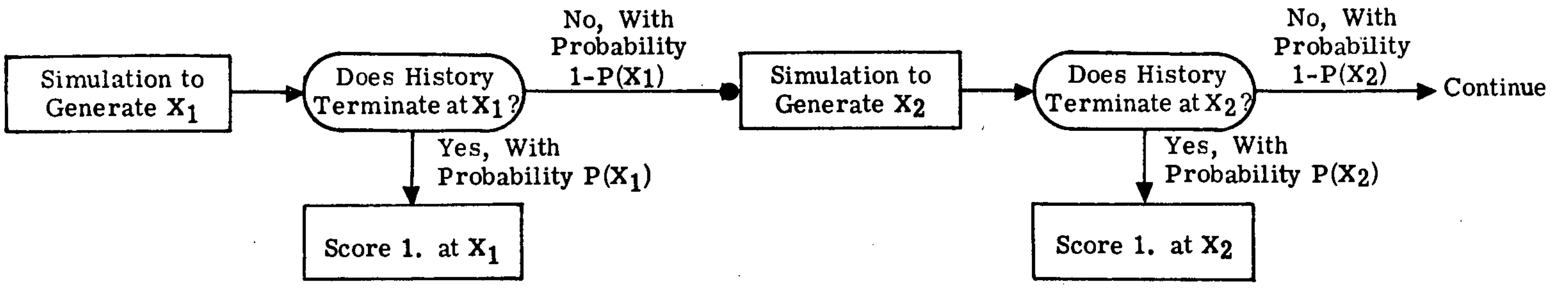

Simulation to Generate X_1
Does History Terminate at X_1?
No, With Probability $1-P(X_1)$
Yes, With Probability $P(X_1)$
Score 1. at X_1
Simulation to Generate X_2
Does History Terminate at X_2?
No, With Probability $1-P(X_2)$
Continue
Yes, With Probability $P(X_2)$
Score 1. at X_2

Fig. 3.8. Multistage Simulation with Crude Monte Carlo Approach

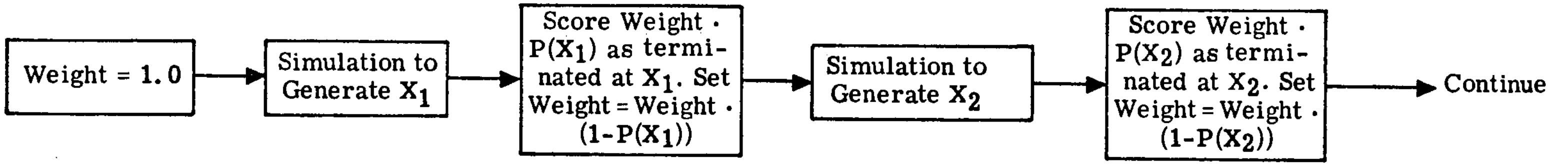

Weight = 1.0
Simulation to Generate X_1
Score Weight · $P(X_1)$ as terminated at X_1. Set Weight = Weight · $(1-P(X_1))$
Simulation to Generate X_2
Score Weight · $P(X_2)$ as terminated at X_2. Set Weight = Weight · $(1-P(X_2))$
Continue

Fig. 3.9. Multistage Simulation with Expected Value Technique

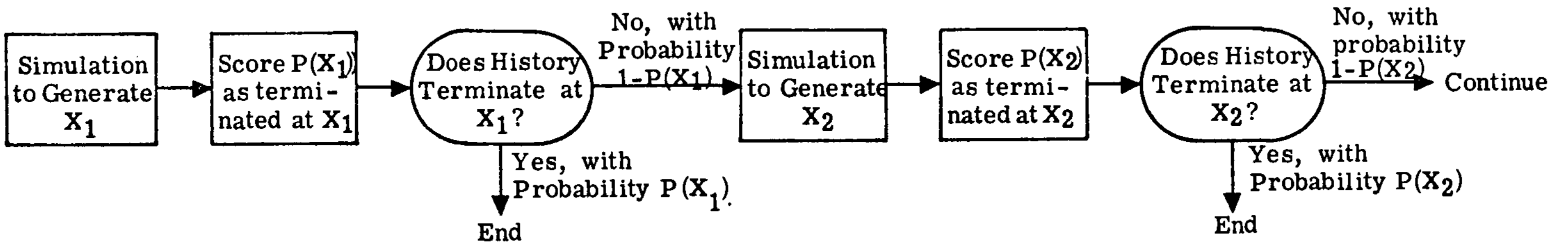

Simulation to Generate X_1
Score $P(X_1)$ as terminated at X_1
Does History Terminate at X_1?
No, with Probability $1-P(X_1)$
Yes, with Probability $P(X_1)$
End
Simulation to Generate X_2
Score $P(X_2)$ as terminated at X_2
Does History Terminate at X_2?
No, with probability $1-P(X_2)$
Continue
Yes, with Probability $P(X_2)$
End

Fig. 3.10. Multistage Simulation with Statistical Estimation

3.2.2 Statistical Estimation(12, 18, 19, 20, 34, 35, 36)

It is not essential, and frequently not efficient, for a simulation of a physical process to be carried out to the natural termination of the process in estimating final outcomes. It is always proper to stop the simulation at any point and to calculate through analytic or numerical means the expectation of reaching any final outcome. Indeed, the sooner the simulation is stopped and the more analytic calculations are done, the lower the variance will be. Obviously, however, the sooner the simulation is stopped the more complex and difficult the analytic calculations become and the point is quickly reached where the overall efficiency is less despite the gain in variance reduction. At the last step in the simulated process, the probability of reaching the various final outcomes needs to be determined in order to do the simulation. Thus, it is generally advantageous to use analytic expectations for the final step. Whether the analytic calculations should be carried beyond the final step will depend on the particular process and results desired, but generally it is less efficient to use analytic expectations for more than the last step.

If the process being simulated is a once-through process, i.e., the final step can be reached only once each history, then the use of expected outcomes is equivalent to the expected value technique. If the process is iterative or repetitive with many passes through a branch point where a final outcome is possible, there are two ways of using the analytic computations. One is by the expected value technique as outlined in the previous section. The other is called statistical estimation and should be used whenever the expected value technique would be inefficient. In statistical estimation the stochastic process is not removed from the simulation, but the expected value, rather than the result of the simulation, is then used in the estimation.

Consider a simulation consisting of many repetitive steps in which one step is a random choice between arriving at some final outcome, Y_f , or continuing through the simulation process with some other value of y .

Let the probability of Y_f at this step be $P(Y_f | \vec{X})$ where $\vec{X}$ denotes all the other random variables determined at earlier steps in the process. In crude Monte Carlo, a random number, R, would be generated at this step, and if $R < P(Y_f | \vec{X})$, then the history would be terminated with a score of 1. If $R > P(Y_f | \vec{X})$, the history would continue with no score being made. After N histories the estimate for the probability of reaching Y_f would be

$$\hat{P}_c(Y_f) = \frac{n}{N} \tag{3.62}$$

where n is the number of histories which terminated at Y_f. In statistical estimation, no change is made in the simulation process, i.e., a random number, R, is drawn and tested to see if the history continues or is terminated. However, the estimation or scoring technique is changed. Every time the particular step is encountered, a contribution of $P(Y_f | \vec{X})$ is added to the estimate, regardless of what the actual outcome of the simulation was. Then the final estimate is given by

$$\hat{P}_{SE}(Y_f) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j} P(Y_f | X_{ij}) \tag{3.63}$$

where the j summation runs over all occurances of the (possibly-) final step in the course of the i^{th} simulation. An estimate of the variance may be calculated from

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(\hat{P}_i - \hat{P}_{SE}(Y_f)\right)^2 = \frac{N}{N-1} \left[\frac{1}{N} \sum_{i=1}^{N} \hat{P}_i^2 - \hat{P}_{SE}(Y_f) \right] \tag{3.64}$$

where

$$P_i = \sum_{j} P(Y_f | \vec{X}_{ij}) \tag{3.65}$$

is the estimate resulting from the i^{th} history.

The schematic flow for the statistical estimation technique in a multi-stage process is shown in Fig. 3.10 where it can be contrasted with the use of crude Monte Carlo (Fig. 3.8) and expected value technique (Fig. 3.9) for the same process.

If the calculations do not get too complicated, the statistical estimation procedure can be extended to using the probability that the simulation will reach the desired end in one or two more stages. If the analytic calculations of such expected values are difficult computationally, then statistical estimation may be less efficient than crude estimators. In employing statistical estimation, the actual simulations which reach the desired end point must be neglected to avoid double counting and only the 'statistically estimated' results used.

The use of statistical estimation will always improve the variance but it can be particularly useful if the probability of reaching the desired end point is small at all intermediate stages. It becomes not just useful but essential when the probability of the end point becomes vanishingly small. In such a case no actual simulations would reach the desired end point and the crude Monte Carlo estimator would give a zero result. If there were many intermediate stages which could, with very low probability, reach the desired end point, then statistical estimation might calculate the desired result with good accuracy.

3.2.3 Correlated Sampling[8,9,12,14,16,18,19,20,34,36]

3.2.3.1 General Concept

Correlated sampling can be one of the most powerful variance reduction techniques due to the wide applicability of the technique as well as to the large efficiency gains which can be obtained. Frequently the primary objective of a simulation study is to determine the effect of a small change in the system. A crude sampling approach would make two independent runs, with and without the change in the system being modeled, and subtract the results obtained. Unfortunately the difference being calculated is often small compared to the two separate results while the variance of the difference will be the sum of the variances in the two runs. Thus the relative uncertainty in the difference is generally very large and it can easily happen that the effect being calculated is smaller than its statistical uncertainty. In such cases the use of correlated sampling can be essential to obtaining a statistically significant result. If, instead of being independent, the two simulations use the same random numbers at comparable stages in the computation, the results can be highly correlated. The effect of this correlation is to reduce the variance of the difference in the two results while not changing the variance in either individual result. As a consequence the effect of the difference in the system will be known to a much greater accuracy than it would be otherwise. Another way of viewing correlated sampling through random number control is to realize that the use of the same random numbers will generate identical histories in those parts of the two systems which are the same. Thus any difference in the results will be due directly to the difference in the systems and not to random noise from the unchanged, but stochastic, elements in the rest of the stimulation. This obviously leads to a gain in efficiency compared to the uncorrelated case.

There are several types of situations where the use of correlated sampling is indicated. These include:

- The effect of a small change in the system is to be calculated.
- The difference in a parameter in two or more similar cases is of more interest than its absolute value in any one case.
- A parametric study of several problems is to be performed. This has greatest potential when the problems are relatively similar in nature.
- The answer to one of several similar problems is known accurately. The answers to the unknown problems can often be estimated from the known result.

3.2.3.2 Analytical Formulation

To provide insight into the concept of correlated sampling, consider the following integrals which characterize different (but hopefully similar or related) problems:

$$I_1 = \int f_1(x)g_1(x)dx \tag{3.66}$$

and

$$I_2 = \int f_2(y)g_2(y)dy \tag{3.67}$$

of primary interest is the difference

$$\Delta = I_1 - I_2 \quad . \tag{3.68}$$

The obvious crude approach is to select N values of X from $f_1(x)$, say $X_1, \ldots, X_N$ and N values of Y from $f_2(y)$, say $Y_1, \ldots, Y_N$ and compute

$$\hat{\Delta} = \hat{I}_1 - \hat{I}_2 = \frac{1}{N}\sum_{i=1}^{N}[g_1(X_i) - g_2(Y_i)] = \frac{1}{N}\sum_{i=1}^{N} g_1(X_i) - \frac{1}{N}\sum_{i=1}^{N} g_2(Y_i) \tag{3.69}$$

The variance in $\hat{\Delta}$ is

$$\sigma^2(\hat{\Delta}) = \sigma_1^2(\hat{I}_1) + \sigma_2^2(\hat{I}_2) - 2 \text{ cov } (\hat{I}_1, \hat{I}_2) \tag{3.70}$$

where

$$\sigma_1^2(\hat{I}_1) = E[(\hat{I}_1 - I_1)^2] \tag{3.71}$$

$$\sigma_2^2(\hat{I}_2) = E[(\hat{I}_2 - I_2)^2] \tag{3.72}$$

and

$$\mathrm{cov}(\hat{I}_1,\hat{I}_2) = E[(\hat{I}_1 - I_1)(\hat{I}_2 - I_2)] = E[(\hat{I}_1\hat{I}_2)] - I_1I_2 \tag{3.73}$$

Now if $\hat{I}_1$ and $\hat{I}_2$ are statistically independent (i.e., no correlation) then

$$\mathrm{cov}\ (\hat{I}_1,\hat{I}_2) = 0 \tag{3.74}$$

and

$$\sigma^2(\hat{\Delta}) = \sigma_1^2(\hat{I}_1) + \sigma_2^2(\hat{I}_2) \tag{3.75}$$

However, if the random variables $\hat{I}_1$ and $\hat{I}_2$ are positively correlated then

$$\mathrm{cov}\ (\hat{I}_1,\hat{I}_2) \geq 0 \tag{3.76}$$

and the variance in the correlated case will be less than that realized with no correlation.

3.2.3.3 Implementation of Correlated Sampling

The key to reducing the variance of the estimate of Δ in (3.69) is to ensure positive correlation between the estimators I_1 and I_2. This can be achieved in several ways although the easiest to implement is to obtain correlated samples through random number control. Specifically, this can be accomplished by using as many of the same random numbers as possible in paired situations in the two simulations. One way this might be accomplished is by using the same sequence of pseudo-random numbers in the two simulations. For example, in the above problem the same sequence of uniform random numbers, $R_1, \ldots, R_N$ from $U(0,1)$ could be used to generate the two sequences $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_N$ by using

$$R_i = \int_{-\infty}^{X_i} f_1(x)dx = \int_{-\infty}^{Y_i} f_2(y)dy \tag{3.77}$$

Clearly the random variables X_i and Y_i are positively correlated since they both used the same R_i . In fact, if f_1 is very similar to f_2 , the random sequences will be very highly correlated.

As another example, consider a multistage problem where many of the events which occur at various stages are not subject to the differences in problem structure. Then, identical random numbers should be used at those stages which are not impacted by the problem differences to produce some positive correlation between the two simulations and to eliminate statistical noise from parts of the system which are unchanged.

It is difficult to be specific as to how random number control should be applied in a general problem. As a general rule, however, to achieve the maximum correlation, the same random numbers should be used whenever the similarities in problem structure will permit this to occur.

Use of the same sequence of random numbers in two separate runs means that the histories generated will be identical up to the point where the difference in the system first comes into play. This complete correlation will obviously eliminate all variance in the difference due to the first, common part of the simulation. In addition, it is possible to save computational time by doing the first simulation and storing the knowledge of the state of the system at the first point in the history where the difference in the two systems affects the simulation. The second simulation could then start at this point rather than recomputing the identical first part of the history. However, this frequently requires more programming effort to implement than is justified.

If it is possible to return to the same sequence of random choices after the calculations concerned with the perturbation, then obviously all the variance in the simulation will be associated with the perturbation, with maximum effectiveness. However, this is generally not possible. Usually the perturbation forces a difference in decision and the two histories proceed in divergent directions following the perturbation. At the completion of one history and the start of the next, it is then necessary to re-synchronize the random number sequences to begin the next histories identically.

In order to estimate the variance in Δ obtained through the use of correlation, it is necessary only to view Δ as if it were being directly simulated and to calculate the sample variance of the difference as

$$S^2 = \frac{N}{N-1} \left[\frac{1}{N} \sum_{i=1}^{N} \hat{\Delta}_i^{\,2} - \hat{\Delta}^2 \right] \tag{3.78}$$

where

$$\hat{\Delta}_i = g_1(X_i) - g_2(Y_i) \tag{3.79}$$

3.2.4 History Reanalysis(18)

3.2.4.1 General Concept

History reanalysis is essentially a form of correlated sampling except that one does not actually run a second simulation using the same random numbers as in the first. Instead, the detailed results of the first simulation are reanalyzed to calculate an answer for the second process. In this case the first process is treated as an altered or 'biased' modification of the second process. In addition to the reduced variance obtained by the correlation, history reanalysis reduces the computational time involved by not actually performing the second simulation. This can often lead to quite high efficiencies for this technique.

Since history reanalysis is a form of correlated sampling, it will apply to the same types of problems indicated in Section 3.2.3.1. However, there is an additional constraint that the differences in the systems being simulated must be expressible as a difference in a probability distribution or in the scoring function.

3.2.4.2 Analytical Formulation

For the purposes here, it is assumed that there are two problems of interest which involve estimating I_1 and I_2 as given by 3.66 and 3.67. It is assumed that a random sample $X_1, \ldots, X_N$ has been obtained from $f_1(x)$. The estimator for I_1 is as usual

$$\hat{I}_1 = \frac{1}{N} \sum_{i=1}^{N} g_1(X_i) \tag{3.80}$$

Since

$$I_2 = \int f_2(x) g_2(x) dx = \int f_1(x) \frac{g_2(x) f_2(x)}{f_1(x)} dx \ , \tag{3.81}$$

an estimate for I_2 can be obtained using

$$I_2 = \frac{1}{N} \sum_{i=1}^{N} \frac{g_2(X_i) f_2(X_i)}{f_1(X_i)} \tag{3.82}$$

where $f_1(X_i) \neq 0$ is implied whenever $g_2(X_i) f_2(X_i) \neq 0$. This is of course very reminiscent of the formulas for importance sampling (see Eq. 3.5). The sample variance for I_2 is

$$S_2^2 = \frac{N}{N-1} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[\frac{g_2(X_i) f_2(X_i)}{f_1(X_i)} \right]^2 - \hat{I}_2^2 \right\} \tag{3.83}$$

which may be used in efficiency calculations. However, to properly calculate the effect of the correlation, it is necessary to estimate the variance of the difference directly. That is, if

$$\hat{\Delta}_i = \frac{g_2(X_i) f_2(X_i)}{f_1(X_i)} - g_1(X_i) \tag{3.84}$$

is the difference in the i^{th} history and

$$\hat{\Delta} = \frac{1}{N} \sum_{i=1}^{N} \hat{\Delta}_i \tag{3.85}$$

is the average difference, then the sample variance is

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N}(\hat{\Delta}_i - \hat{\Delta})^2\right] = \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_i^2 - \hat{\Delta}^2\right] \tag{3.86}$$

Alternatively, batching (Section 2.4.2) can be applied to the differences.

3.2.4.3 Further Considerations

The equations in the preceding section show that by simulating I_1 and biasing the results appropriately, an estimate for I_2 can be readily generated. This can obviously be generalized to the case of three or more similar problems. The time saving gained by not making several separate simulations is obvious. In addition, there will be all the advantages of high correlation due to the use of a common set of random numbers.

However, the use of history reanalysis is not universally beneficial and may sometimes be less efficient than independent simulations. The simularity of the equations for history reanalysis to those of importance sampling has already been noted. It should then be clear that the random sample $X_1, \ldots, X_N$ which has been chosen from $f_1(x)$ is not likely to be the optimum choice, in the sense of 'importance', for the simulation of $g_2(x)f_2(x)$. Thus, the variance of $\hat{I}_2$ is likely to be greater than that which would be obtained from a direct simulation of I_2. Hopefully, the gain in efficiency effected by the correlation and reduced computation will more than offset this variance increase but this will not be true in all cases. Obviously the more similar the two cases are, the more optimum the selection will be for computing I_2. Thus, history reanalysis works best for the problems which are most similar which are the cases where variance reduction is most necessary.

There is an important class of problems where history reanalysis is trivially accomplished. This occurs when

$$g_2(x)f_2(x) = \begin{cases} g_1(x)f_1(x) & , \text{ in some region A} \\ 0 & , \text{ elsewhere} \end{cases} \tag{3.87}$$

An example of this is a simulation that is run for a fixed real-time interval, T_1, and it is desired to know the results of a case that was limited to a shorter time interval, T_2. Then history reanalysis consists of making a single simulation with the longer time limit, T_1, scoring for the first case all events, and scoring for the second case only those events for which time is less than T_2.

Several extensions readily come to mind. Most significantly, parametric studies to determine the impact of several forms of a sampling distribution can be readily performed. This capability is often overlooked in simulation studies resulting in considerable unnecessary expense.

3.2.5 Control Variates(11, 14, 20, 24, 34, 36)

3.2.5.1 General Concept

In many situations where analytic models are difficult or impossible to develop, there exist simplications or approximations to the problem having analytic or closed form solutions. In these situations, the analytic information can be beneficially exploited to reduce variance by what is referred to as control variates. With this technique, instead of estimating a parameter directly, the difference between the problem of interest and some analytical model is simulated. The variance reduction, or increase in accuracy in estimating the parameters of interest, is directly related to the degree of correlation between the analytic and the true process. Application of this technique is again very general and should prove very useful when analytical representations of simplified models for the system exist.

The control variate method has several of the features similar to those of the correlation technique and indeed in some instances is addressed within the context of correlation. However, the manner in which this technique is applied is somewhat distinctive and, therefore, will be treated separately here.

3.2.5.2 Analytical Formulation

Again consider the integral

$$I = \int_{-\infty}^{+\infty} g(x)f(x)dx \tag{3.88}$$

Assume that it is possible to determine a function $h(x)$ whose expected value is known (or analytically determinable) and which closely approximates $g(x)$. Qualitatively such a situation is shown in Fig. 3.11. Let

$$\theta = \int_{-\infty}^{+\infty} h(x)f(x)dx \tag{3.89}$$

and assume that θ is known

Then I can be expressed as

$$I = \int_{-\infty}^{+\infty} h(x)f(x)dx + \int_{-\infty}^{+\infty} [g(x) - h(x)]f(x)dx$$

$$= \theta + \int_{-\infty}^{+\infty} [g(x) - h(x)]f(x)dx = \theta + I_1 \tag{3.90}$$

The function $h(x)$ is called the control variate for $g(x)$ and may be some approximation (or guess) to $g(x)$.

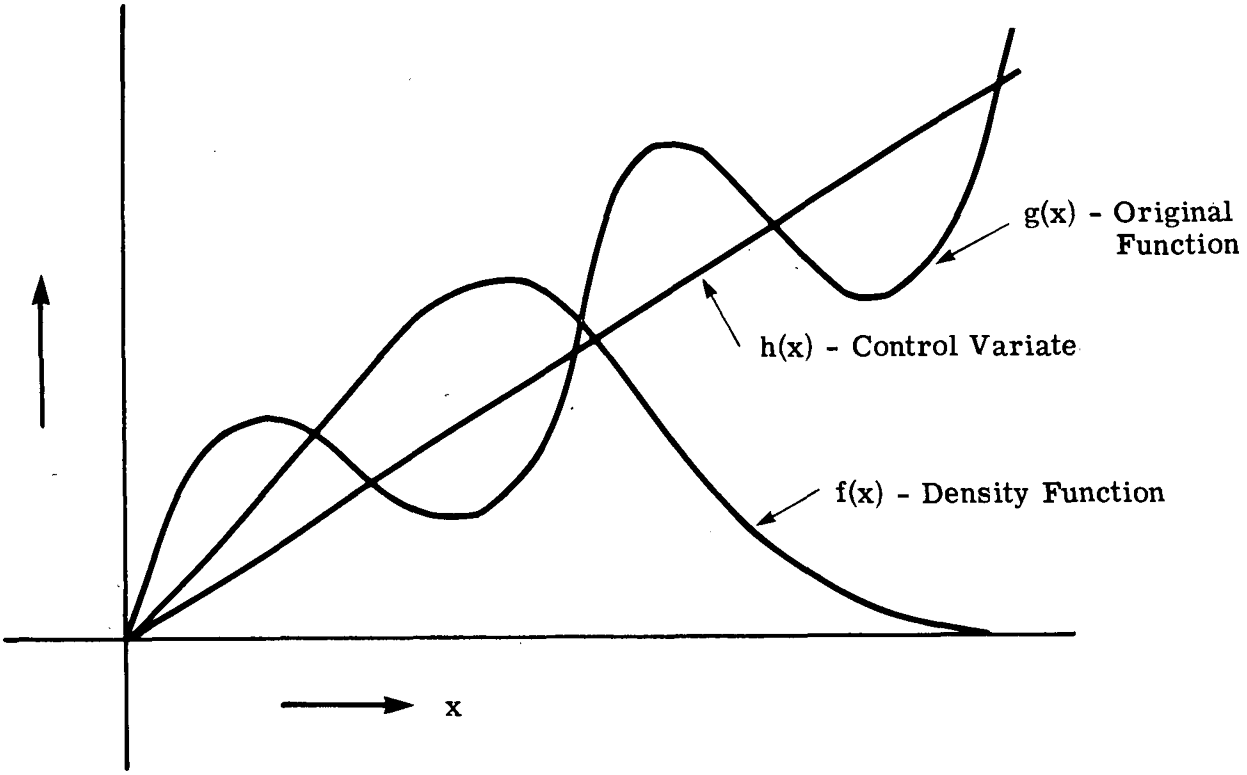


Fig. 3.11 Illustration for Control Variates

Now, since h(x) has been selected so that the first integration can be completed, simulation is required only on the second term,

$$I_1 = \int_{-\infty}^{\infty} [g(x) - h(x)]f(x)dx \tag{3.91}$$

If crude sampling is used to simulate I_1, then a random sample would be obtained by selecting $X_1, \ldots, X_N$ from f(x) and using

$$I = \theta + 1/N \sum_{i=1}^{N} g(X_i) - 1/N \sum_{i=1}^{N} h(X_i) = \theta + 1/N \sum_{i=1}^{N} \hat{\Delta}_i \tag{3.92}$$

where

$$\hat{\Delta}_i = g(X_i) - h(X_i) \tag{3.93}$$

An estimate for the sample variance for purposes of efficiency calculations is given by

$$s^2 = \frac{N}{N-1} \left\{ 1/N \sum_{i=1}^{N} \hat{\Delta}_i^2 - \hat{\Delta}^2 \right\} \tag{3.94}$$

where

$$\hat{\Delta} = 1/N \sum_{i=1}^{N} \Delta_i \tag{3.95}$$

The use of control variates is but another manifestation of the use of information about the problem to reduce the variance. In this case a knowledge of the approximate behavior of the system was used to advantage. Its effectiveness is greatly dependent, however, on how good h(x) can be selected to approximate g(x).

It is worthwhile to note that if an approximate shape for g(x) is not known, it is often possible to obtain an approximation by simply selecting a few values of x and plotting the results. A straight line fit to the results or some other simple formulation may significantly improve the efficiency of the simulation.

3.2.6 Antithetic Variates(9, 11, 12, 14, 28, 34, 36)

3.2.6.1 General Concept

The concept of antithetic variates is somewhat related to that for control variates except that, rather than seeking a function that is similar to the function being estimated, a function is sought which is negatively correlated.

The estimation process is then structured to exploit this negative correlation to reduce the variance in the estimator. The basic idea can be used to develop very sophisticated and powerful methods. Two methods will be presented below.

3.2.6.2 Method I for Antithetic Variates

The use of antithetic variates can be introduced very simply as follows: consider again the parameter I to be estimated where

$$I = \int_{-\infty}^{+\infty} g(x)f(x)dx \tag{3.96}$$

Assume an unbiased estimator $\hat{I}_1$ for I exists. For example, if crude sampling is used

$$\hat{I}_1 = 1/N \sum_{i=1}^{N} g(X_i). \tag{3.97}$$

Suppose a second unbiased estimator, $\hat{I}_2$ for I also exists.

A third unbiased estimator $\hat{\theta}$ for I can be constructed using

$$\hat{\theta} = 1/2(\hat{I}_1 + \hat{I}_2) \tag{3.98}$$

and

$$E[\hat{\theta}] = I$$

The variance in the estimator $\hat{\theta}$ is given by

$$\sigma^2(\hat{\theta}) = 1/4\, \sigma^2(\hat{I}_1) + 1/4\, \sigma^2(\hat{I}_2) + 1/2 \operatorname{cov} (\hat{I}_1\hat{I}_2) \tag{3.99}$$

Now, if $\hat{I}_1$ and $\hat{I}_2$ are selected such that they are negatively correlated, then

$$\operatorname{cov} (\hat{I}_1,\hat{I}_2) \leq 0 \tag{3.100}$$

If $\mathrm{cov}(\hat{I}_1, \hat{I}_2)$ is sufficiently large (negatively), then

$$\sigma^2(\hat{\theta}) < \sigma^2(\hat{I}_1) \tag{3.101}$$

and

$$\sigma^2(\hat{\theta}) < \sigma^2(\hat{I}_2) \tag{3.102}$$

Thus, the combined estimator $\hat{\theta}$ of $\hat{I}_1$ and $\hat{I}_2$ will have a smaller variance than either $\hat{I}_1$ or $\hat{I}_2$.

The estimator, $\hat{I}_2$, is called the antithetic variate since it is an estimator that compensates for the variation in $\hat{I}_1$. This is, of course, the concept of negative correlation.

There is a convenient manner in which an antithetic variate can be obtained. This is as follows:

Consider the estimator $\hat{I}_1$ to be derived from crude sampling. To accomplish this a set of random numbers $R_1, \ldots, R_N$ will be generated from $U(0,1)$ and the corresponding values of X, say $X_1, \ldots, X_N$ can be obtained from

$$R_i = \int_{-\infty}^{X_i} f(x)dx \quad ; \quad i = 1, \ldots, N \tag{3.103}$$

It is clear that $\{X_i\}$ are from the distribution $f(x)$. Now consider generation of another set of values of X, $X_1', \ldots, X_N'$ using

$$1 - R_i = \int_{-\infty}^{X_i'} f(x)dx \quad ; \quad i = 1, \ldots, N \tag{3.104}$$

Again $X_1', \ldots, X_N'$ will be from the distribution $f(x)$. The pairs of values of X_i and X_i' are, of course, correlated since the same random number

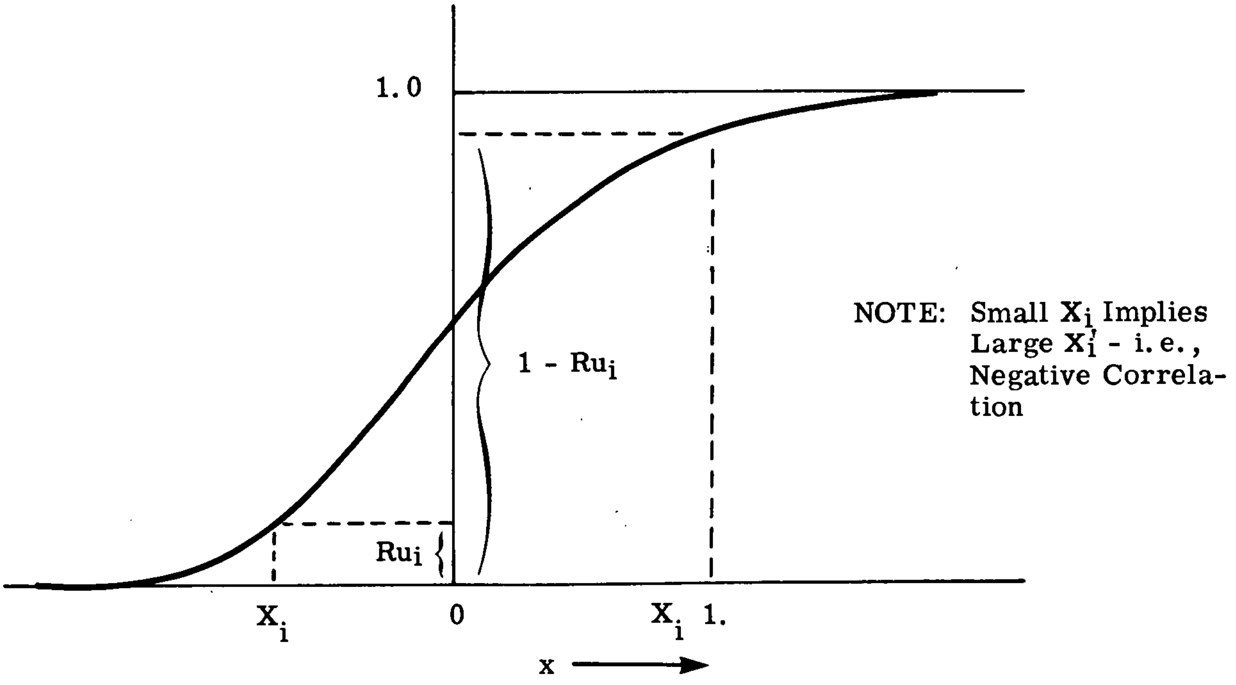
R_i was used to generate both values of X. Furthermore, these values of X_i and X'_i are neglatively correlated. That is, small X_i corresponds to large X'_i. This is shown conceptually in Fig. 3.12.

Defining

$$\theta_i = 1/2[g(X_i) + g(X'_i)] \tag{3.105}$$

Then the estimator for I using the antithetic variates is

$$\theta = 1/N\sum_{i=1}^{N} \theta_i = 1/2N\sum_{i=1}^{N} [g(X_i) + g(X'_i)] \tag{3.106}$$



NOTE: Small X_i Implies Large X'_i - i.e., Negative Correlation

Fig. 3.12 Schematic Showing a Method to Generate Antithetic Variates

The sample variance is determined from

$$S^2(\hat{\theta}) = \frac{1}{(N-1)} \sum_{i=1}^{N} (\theta_i - \hat{\theta})^2 = \frac{N}{N-1} \left\{ \frac{1}{N} \sum_{i=1}^{N} \theta_i^2 - \hat{\theta}^2 \right\} \qquad (3.107)$$

3.2.6.3 Method II for Antithetic Variates

A second approach to antithetic variates that has proven very successful is to use a combination of stratified sampling along with antithetic variates. Consder a case with 2 strata as shown in Fig. 3.13. Assume the range

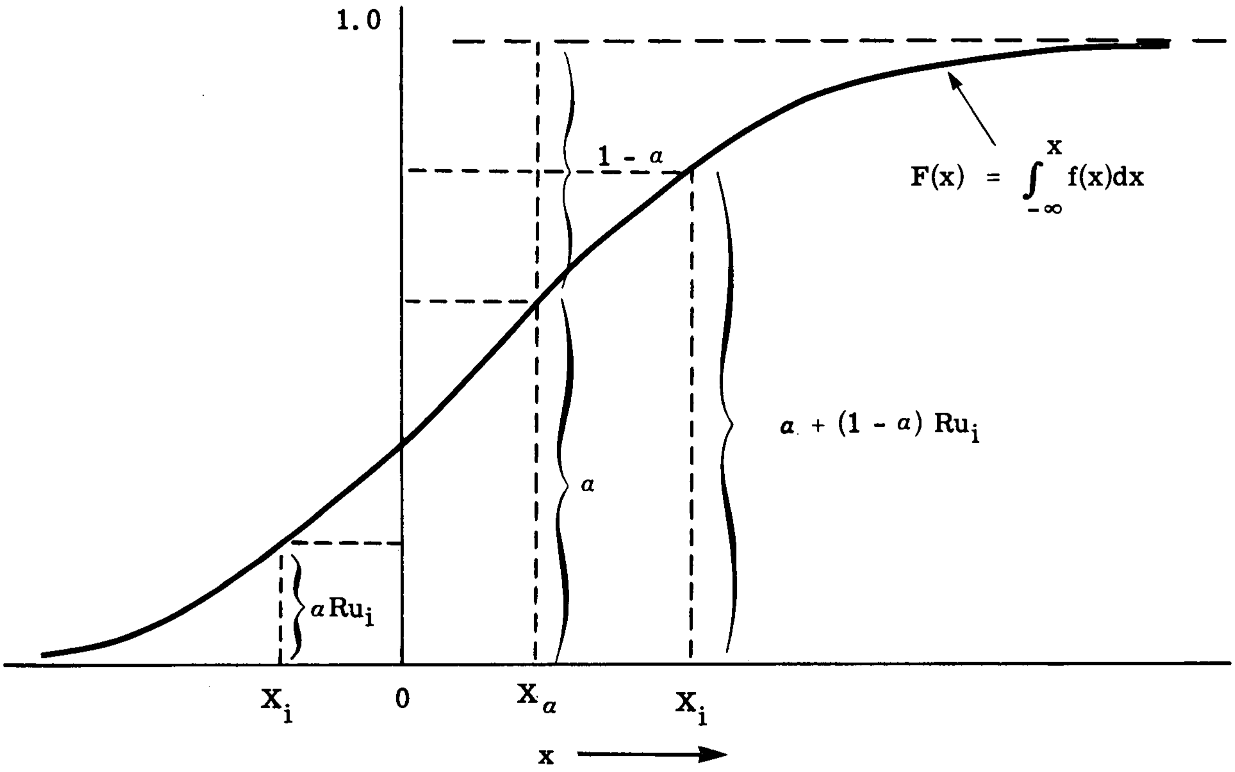


Fig. 3.13 Method II for Application of Antithetic Variates

of $f(x)$ is broken up by X_α into the ranges $-\infty < x < X_\alpha$ and $X_\alpha < x < \infty$. Now, suppose a random number is selected from U(0, 1). Then select X_i from

$$\alpha R_i = \int_{-\infty}^{X_i} f(x)dx \tag{3.108}$$

and select X_i' from

$$\alpha + (1-\alpha)R_i = \int_{-\infty}^{X_i'} f(x)dx \ . \tag{3.109}$$

Clearly X_i and X_i' are distributed according to $f(x)$ within their appropriate ranges. Also, X_i and X_i' are negatively correlated since small X_i implies large X_i' and vice versa. Now define

$$\hat{\theta}_i = \alpha g(X_i) + (1-\alpha)g(X_i') \tag{3.110}$$

An unbiased estimator for I is

$$\hat{\theta} = 1/N \sum_{i=1}^{N} \hat{\theta}_i = 1/N \sum_{i=1}^{N} [\alpha g(X_i) + (1-\alpha)g(X_i')] \tag{3.111}$$

and the sample variance is

$$S^2 = \frac{N}{N-1} \left\{ 1/N \sum_{i=1}^{N} \hat{\theta}_i^2 - \hat{\theta}^2 \right\} \tag{3.112}$$

If $\alpha = 1/2$, then Eq. 3.111 reduces to Eq. 3.106.

The difficulty in the use of this second approach is in the selection of α. A general rule is to select α such that

$$g(X_\alpha) = \alpha g(X_L) + (1-\alpha)g(X_U) \tag{3.113}$$

where X_U and X_L are the upper and lower limits of the range of $f(x)$.

An alternate approach is to utilize a trial and error method to test various values of α and estimate the improvement realized in the efficiency.

It is important to recognize that the choice of α will not impact the simulation in the sense that the estimator will still be unbiased. However, it may result in some loss of efficiency if a poor value is selected.

3.2.7 Regression(7, 11, 14)

3.2.7.1 General Concepts

Regression techniques have found limited application in Monte Carlo simulation in spite of the seemingly important advantages that

- They can be applied to a wide variety of Monte Carlo simulations
- They will produce unbiased estimates

- They can be applied in a situation where correlation is known to exist and will take advantage of such correlation

- If applied to a situation where no correlation exists, nothing is lost except the additional computational effort involved

Its use appears to be rather limited due to the effort involved in formulation of the appropriate estimators and the difficulty encountered when attempts are made to view a practical simulation problem within the context of known formulations of the regression method.

3.2.7.2 Analytical Formulation

To formalize the regression method, assume a set of integrals $I_1, \ldots, I_p$ are to be estimated. Assume a set of estimates $\hat{\theta}_1, \ldots, \hat{\theta}_n$ $(n \geq p)$ are available satisfying the condition that

$$E[\hat{\theta}_j] = a_{j1}I_1 + \ldots + a_{jp}I_p \quad , \quad j = 1, \ldots, n \qquad (3.114)$$

where the matrix

$$\vec{A} = \begin{pmatrix} a_{11} \cdots a_{1p} \\ a_{21} \cdots a_{2p} \\ \cdot \\ \cdot \\ \cdot \\ a_{n1} \cdots a_{np} \end{pmatrix} \qquad (3.115)$$

is known. It is assumed that a sample is available consisting of N independent sets of simulated values for $\dot{\theta}_j$, namely $\hat{\theta}_{1j}, \ldots \hat{\theta}_{Nj}$; $j = 1, \ldots, n$. Then

$$\hat{\theta}_j = \frac{1}{N} \sum_{i=1}^{N} \theta_{ij} \quad ; \; j = 1, \ldots, n \qquad (3.116)$$

and the column matrix

$$\vec{\hat{\theta}} = \begin{pmatrix} \hat{\theta}_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \hat{\theta}_n \end{pmatrix} \tag{3.117}$$

can be readily constructed.

Now, an estimate for the matrix $\vec{I}$ is desired where $\vec{I}$ is defined as

$$\vec{I} = \begin{pmatrix} I_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ I_p \end{pmatrix} \tag{3.118}$$

It will be recalled from elementary statistics that the minimum variance unbiased estimator for $\vec{I}$ is given by

$$\vec{\hat{I}} = (\vec{A}^T \vec{V}^{-1} \vec{A})^{-1} \vec{A}^T \vec{V}^{-1} \vec{\hat{\theta}} \tag{3.119}$$

where

$$\vec{V} = \begin{pmatrix} v_{11} \cdots v_{1n} \\ v_{21} \cdots v_{2n} \\ \cdot \\ \cdot \\ \cdot \\ v_{n1} \cdots v_{nn} \end{pmatrix} \tag{3.120}$$

is the covariance matrix for $\hat{\theta}_1, \ldots, \hat{\theta}_n$ and A^T is the transpose of A

That is

$$v_{ij} = E[\{\hat{\theta}_i - E(\hat{\theta}_i)\}\{\hat{\theta}_j - E(\hat{\theta}_j)\}] \qquad \begin{array}{l} i = 1,\ldots,n \\ j = 1,\ldots,n \end{array} \qquad (3.121)$$

Unfortunately v_{ij} is usually not known. However, an estimate for $\vec{V}$, can be obtained using

$$\hat{v}_{ij} = \sum_{k=1}^{N} (\theta_{ki} - \hat{\theta}_i)\ (\theta_{kj} - \hat{\theta}_j) \ ; \qquad \begin{array}{l} i = 1,\ldots,n \\ j = 1,\ldots,n. \end{array} \qquad (3.122)$$

where $\hat{\theta}_i$; $i = 1,\ldots,n$ are obtained from Eq. 3.116.

and

$$\vec{\hat{V}} = \begin{pmatrix} \hat{v}_{11}\cdots\hat{v}_{1n} \\ \hat{v}_{21}\cdots\hat{v}_{2n} \\ \cdot \\ \cdot \\ \cdot \\ \hat{v}_{2n}\cdots\hat{v}_{nn} \end{pmatrix} \qquad (3.123)$$

The new estimator is therefore

$$\hat{I}* = (\vec{A}^T \vec{\hat{V}}^{-1} \vec{A})^{-1} \vec{A}^T \vec{\hat{V}}^{-1} \hat{\theta} \qquad (3.124)$$

This is still unbiased since

$$E[\hat{I}*] = I \qquad (3.125)$$

It is recommended that batching be used to obtain an estimate for the variance σ^2.

As it is formulated above, the regression technique is very easy to apply. All that is required is to obtain $\vec{V}$ and $\vec{\theta}$ from the sample values and use Eq. 3.124 to obtain an unbiased estimator for I.

In spite of its relatively simple formulation which is based on some elementary statistical concepts, the method is difficult to apply in practice primarily because it is generally difficult to formulate the estimators $\hat{\theta}_1, \ldots, \hat{\theta}_n$. It has evidently been applied in only trivial situations and realization of its full potential must await additional development and experience. Clearly, one characteristic a problem should have before attempting to apply this method is a linear combination of the estimator and parameters to be estimated as indicated by Eq. 3.114.

3.3 SPECIALIZED TECHNIQUES

In the foregoing sections several very useful and well developed Monte Carlo techniques were presented and discussed. There are, however, a large number of additional procedures that might warrant consideration in situations where some of the preceding techniques proved ineffective. These are either not well developed (e.g., orthonormal) or they may be extremely specialized and have therefore found application in very specific problems (e.g., the adjoint method). It must be recognized, however, that the application of these specialized techniques may be necessary to achieve a reasonable answer in very difficult problems but should be resorted to after this becomes abundantly clear. These specialized techniques are however fertile fields for further research into variance reduction.

3.3.1 Sequential Sampling(14, 19, 25, 30, 34)

Occasionally there is little or no a priori information concerning the expected results of the simulation or perhaps what knowledge there is strictly qualitative with no quantitative values on which to base a choice of an importance function or Russian Roulette or splitting standards. In such a case it may be possible to use sequential sampling. This is not a specific variance reduction technique but rather a general approach to the use of other techniques. In sequential sampling an initial run is made with little or no variance reduction used. Then the results of this first run are analyzed to calculate an importance function or used to estimate Russian Roulette standards, stratified sampling parameters, etc. A second run is made using a variance reduction technique with the parameters estimated in the first run. Now these results can be analyzed in conjunction with the first set of histories, to improve the estimation of the sampling parameters. A third run can then be made using the improved sampling parameters and this 'self-learning' process can be carried out through an indefinite number of stages with the efficiency of the sampling improving at each stage. Despite the simplicity and intuitive appeal of such an approach, little or no work on sequential sampling has been performed. (There has been some development of 'self-learning' techniques applied to stratified sampling, and preliminary work is in progress in some other areas.) Consequently little can be said regarding implementation techniques, trade-offs of computation required to estimate sampling parameters versus the efficiency gain from improved sampling, or possible pit-falls (e.g., can an initial choice of 'underbiased' or 'overbiased' parameters lead to estimation of parameters that are even more underbiased or overbiased with the sequential process feeding on itself destructively?)

3.3.2 Orthonormal Functions(14, 19, 25, 30, 34)

The use of orthonormal functions in general Monte Carlo simulations has received little attention, although it does have potential for greatly improving simulation efficiency when it is applied to problems having a large number of dimensions.

Basically, the approach is to first define a set of orthogonal functions over a region of multiple integration. Next, a sampling scheme must be structured that will permit efficient sampling over this region from a joint probability density function. In general the procedures to accomplish these tasks are not well developed and will not be further discussed here. This does not imply, however, that the potential gains that can be realized with this technique are not worth the effort but only that no general guidelines or problem approach can be presented to provide reasonable assurance that the effort would be fruitful.

3.3.3 Adjoint Method(15, 17, 21)

In formulating the mathematical equations for the simulation of a process, it frequently is the case that there is another set of equations, "inverted" or "adjoint" with respect to the first, that is mathematically equivalent in the sense that a solution to one set of equations will also give a solution to the second set. This second set, the adjoint equations, may not represent any real process but can be simulated anyway. Depending on the nature of the problem and the result being calculated, it may be easier or more efficient to simulate the adjoint equations than to simulate the direct process. It may also be possible to split the problem into two parts, one of which is best simulated directly and one of which is best simulated by the adjoint process.

There is a close interrelationship between the adjoint solution and the importance of sampling in the direct simulation. This leads to interesting possibilities such as using approximate methods or a simplified process to calculate the adjoint, then using the adjoint solution to generate importance sampling for a full simulation of the direct equations. Another alternative is a form of sequential calculation where direct and adjoint solutions alternate with each solution serving as the importance function for sampling the next solution.

As with many of the more powerful variance reduction techniques, the adjoint has been exploited very successfully in the area of radiation transport. This was possible due to the formulation of the radiation transport problem in a precise (although difficult to solve) linear integral equation where an adjoint formulation could be easily established.

Unfortunately, in most Monte Carlo simulations such a compact formulation is not generally available and furthermore would be difficult to develop. However, the concept of the adjoint offers some intriguing possibilities. For example, rather than tracing an individual history through the system in a natural manner, (i. e., from start to finish) it is possible to trace the individual from a final exit point to the starting position. As an example, should an adjoint formulation be developed with respect to antisubmarine warfare application it would not simulate in a forward manner to determine events that result in a submarine kill, but would rather start from a submarine kill and trace backward through the simulation to determine what sequence of events could have led up to the kill. Many other applications could be envisioned which could exploit the use of the adjoint or backward formulation. This technique must however await further development and use before it becomes a generally applicable method such as is found in importance sampling or correlation.

3.3.4 Transformations(4, 12, 34)

The transformation method is essentially a special form of importance sampling. It differs from other types of importance sampling in that a priori information about the process is formulated in a parametric, closed-form representation which is then used to alter the sampling procedure by the transformation. For example if an approximate, parametric representation of the importance function is known, then a transformation can generate an altered process where the important areas have greater probability and the unimportant areas have low probability.

This method has been largely employed in radiation transport calculations where the functions of interest frequently have an approximately exponential form. There an exponential transform generates an altered process with a greatly reduced variance.

3.3.5 Conditional Monte Carlo(2, 10, 14, 34)

If the particular problem being investigated is very complex in that it deals with a complicated sample space or the probability density function is difficult to select from, it may be possible to embed the given sample space in a much larger space in which the desired density function appears as a conditional probability. The larger space and its accompanying density are chosen to be much simpler in definition although they involve more variables. Simulation of the large problem can be much simpler than the original complex problem, and, despite the added computation required to calculate the conditional probabilities, the gain in efficiency can be quite high. Furthermore, the added degrees of freedom gained by the added variables and the choice of a space and density function in which to embed the original problem can be utilized to secure additional variance reduction.

Despite the potential power of conditional Monte Carlo for solving complex simulation problems, it has seen very little use. In large part this is due to the creative leap needed to view the problem in a larger context and to design the larger space in which to embed the problem. In addition, while the theoretical basis for this technique has been developed, very little in the way of practical examples or applications has been produced, and the method is still not well understood.

PART II

APPLICATION OF VARIANCE REDUCTION TECHNIQUES

4. SELECTION OF VARIANCE REDUCTION TECHNIQUES

Unless one is very familiar with the concepts of variance reduction, the selection of a promising approach for a particular problem can cause considerable difficulty due to the large number of possibilities available. This section of the report will be directed toward aiding the analyst in selection and implementation of an appropriate variance reduction technique or techniques. This is accomplished by way of a systematic procedure to:

- Define the problem information that can be used as a basis to select an appropriate technique or techniques.
- Select the specific technique or techniques that should be considered for a given problem.
- Provide basic guidelines to implement the selected procedure.

Each of these aspects are described in Sections 4.1 through 4.3 respectively.

There are several approaches to use the information of this part with that of Part I. The first, and probably the most effective, is to review briefly the material in Part I and then proceed to defining the available information on the problem, selecting the appropriate technique and proceeding to its implementation. Alternately, the required information for selection of the particular variance reduction technique could first be defined and the procedure selected prior to reviewing the material in Part I.

4.1 DEFINITION OF PROBLEM INFORMATION

The usefulness of variance reduction techniques is ultimately determined by how effectively known information about the problem is utilized. Problem definition is thus of paramount importance. Before considering variance reduction techniques, it is essential to characterize the aspects of the problem that might indicate which techniques could be fruitfully applied. To evaluate the usefulness of these methods for a particular problem, it is helpful to know the information defined in Table 4.1. Such information is not strictly required as certain approaches (such as sequential sampling) can generate useful information, but this is not generally accomplished without cost. Thus prior information is highly desirable. The more that is known, the better the ultimate results will be. This information will be used in conjunction with the characteristics of the techniques described in the next section.

Consider item 1 in Table 4.1. Here it is required to clearly define which parameters are to be estimated. This could include mean values, variances or probabilities. Furthermore, if it is known that the problem is such that sensitivity or perturbation studies are to be performed, it is important that this be recognized at the outset. Additional information of significance includes the sequential nature of a problem, as well as identifying any input conditions that are random variables.

Under the second item in Table 4.1, the significance of integral formulations for parameters to be estimated is pointed out. The importance of integral formulations cannot be over-emphasized since it is this basic integral structure which is used to understand almost every variance reduction method. In addition to the analytical forms, the knowledge of various expected values in the problem and the availability of simplified analytical expressions which are positively or negatively correlated with the parameters whose expected values are being estimated can provide key information as to the variance reduction approach to be finally implemented.

Next, identifying intermediate events or parameters which assume significance relative to their importance, unimportance or insensitivity to the problem outcomes can provide valuable information. A key ingredient for improving the efficiency of multistage simulations is identification of variables or outcomes in the problem that will probably lead to either important or unimportant outcomes of the final events. Finally, identifying those final outcomes

TABLE 4.1

Recommended Problem Information to be Defined Prior to Selecting and Implementing Variance Reduction Techniques

1. Define nature of the problem relative to

   - expected values (means, variances, probabilities, etc.) to be estimated
   - sensitivities, perturbations or variations of parameters of interest
   - possible mathematical formulations (e.g., integral equations, expected values, etc.)
   - any sequential characteristics such as independent paths, outcomes dependent on intermediate steps, etc.
   - input conditions which are random variables to be sampled

2. Identify portions of the problem or parameters to be estimated that can be

   - expressed in an analytical form such as single or multidimensional integrals, differential and/or integral equations
   - solved analytically, such as expected values, variances, probabilities, etc.
   - represented by approximate, simplified positively correlated analytical expressions
   - represented by approximate, simplified negatively correlated analytical representations
   - established as relatively not important to final outcomes compared to other aspects of the problem

3. Identify variables in the problem which

   - are very important to the expected outcome
   - are not expected to significantly impact the results
   - over their range of variation have relatively little effect on the problem
   - are strongly correlated with other variables

4. Locate final events or outcomes of the problem which

   - have very small probabilities
   - have very large probabilities
   - have outcomes relatively insensitive to problem parameters
   - have known probabilities of occurrence from intermediate stages in the problem
   - are linear combinations of other events or random variables
   - have known correlation with other events or outcomes

which occur with large or small probabilities, are insensitive to problem parameters, have correlation with other events, or the final events which have known probabilities of occurrence from intermediate stages will also prove to be very useful in effectively reducing the variance.

It should also be recognized that, in general, variance reduction techniques are aimed at reducing the variance of only one parameter or aspect of the process being simulated. Using variance reduction techniques designed for one parameter will usually reduce the effectiveness of the simulation to estimate other parameters. It is very important, therefore, to determine all of the results which will be desired from the simulation before searching for a technique to apply to a given situation. If several quantities are to be estimated by the simulation, the selection of a variance reduction technique has to be considered from the standpoint of all of these parameters. In many circumstances it may be beneficial (or even necessary) to implement a different variance reduction technique for each parameter. This might be accomplished in the extreme case by developing a different simulator for each parameter of interest.

4.2 SELECTION OF VARIANCE REDUCTION TECHNIQUE(S)

A comprehensive summary of the variance reduction techniques considered in Section 3 is shown in Table 4.2. Here, each alternative is described briefly along with the suggested criteria for application. In addition, advantages, disadvantages, and typical applications are noted. As will be seen many of these techniques are interrelated, although their method of application may differ substantially.

Also shown for each technique is the section numbers of this report in which details of the approach can be found. The first section noted refers to the material in Part I and the second references Part II. As may be seen from a brief review of Table 4.2, there is substantial variation in the criteria

TABLE 4.2

Summary of Variance Reduction Techniques and Characteristics

Variance Reduction Technique or Procedure[1]	Description of Technique	Suggested Criteria for Application	Advantages[2]	Disadvantages	Comments on Typical Applications[3]
Importance Sampling (Secs. 3.1.1, 4.3.1)	Structuring the sampling such that the more important regions of interest are sampled more frequently.	. Certain events are known to be important . Parameters to be estimated are based on events with very low probability of occurrence . Certain events are not of interest . Regions of importance can be defined at various stages in the simulation	● can give great improvement with limited effort ● well developed procedures ● can often be easily implemented ● may be only way to get a reasonable answer	● can give worse results if not properly applied ● may require considerable a priori knowledge ● may require additional computation time to implement and to correct for biased sampling	Most commonly used method. Applications in PERT, reliability, fault TREE analysis, queueing and radiation transport. Applicable to almost any problem having criteria indicated
Russian Roulette and Splitting (Secs. 3.1.2, 4.3.2)	Use of probabilities to limit or kill samples in an uninteresting portion of the simulation (Russian Roulette) and to increase the number of samples in interesting regions using probabilities (splitting).	● process is sequential in nature ● low probability events are involved ● importance of all outcomes are qualitatively known ● often useful where importance and/or expected values are used	. Very little computation . Convenient way to accomplish crude importance sampling	. Requires a priori knowledge of importance of intermediate steps . There may be counteracting pitfalls if used with importance sampling . May not be effective if used alone . Splitting may be difficult to implement	Classical application has been in particle transport. Can be used in search problems, network analysis, traffic flow, random walk, etc. when low probability events or unimportant events are involved
Systematic Sampling (Secs. 3.1.3, 4.3.3)	Sampling is accomplished by selecting random numbers and systematically distributing them over the sample space.	. Random variables are to be generated at the beginning of the problem (source or initial conditions) . Relative importance of ranges of random variables being sampled is unknown	● easy to implement ● little additional computation required ● no risk in application ● always get variance reductions	● often gives marginal improvement ● effective only in a limited number of situations	Application to many problems beyond direct integral evaluation has been limited. Could be useful in any problem where random inputs are of interest. Examples are in queueing, reliability. etc.

TABLE 4.2 (Continued)

Variance Reduction Technique or Procedure[1]	Description of Technique	Suggested Criteria for Application	Advantages[2]	Disadvantages	Comments on Typical Applications[3]
Stratified or Quota Sampling (Secs. 3.1.4. 4.3.4)	Sampling is distributed over the sample space by dividing the distribution into sections and sampling from each section. More important areas of sampling are emphasized.	● certain ranges of the variables are more important than others. ● random variables are to be generated at the beginning of the problem (source or initial conditions)	● easy to implement ● can give great improvement ● gives better results than systematic if properly applied ● can be optimally applied	● requires a priori knowledge of importance ● can lead to worse results if not properly applied ● sometimes difficult to define sampling ranges	Known applications are primarily in queueing. Any problem with initial random condition should find this technique useful. Includes reliability, network analysis, etc. (1, 2, 4, 14, 15, 17, 18, 27, 28).
Expected Value (Secs. 3.1.5, 4.3.5)	Include known expected values in the process to replace stochastic portions of a simulation.	. Analytical results, probabilities, or expected values of portions of the problem are known. . Stochastic nature (i.e., higher moments) of portion whose expected value is known is not essential to overall simulation.	. Almost always gives improvements . Trivial to implement in most useful cases	● may require prohibitive analysis/computation	Has been applied to PERT and radiation transport. Its use in other areas has been limited. However, could be useful in almost any area where known results are available for parts of the problem
Statistical Estimation (Secs. 3.1.6, 4.3.6)	Include known expected value or probability in estimator but not in simulation model itself.	. Analytic results, probabilities or expected values of portions of the problem are known. . Stochastic nature is essential to overall simulation. . Process is sequential in nature. . Probability of final outcomes is known, but small, at all intermediate stages.	. Always gives improvement . Can get great improvement . Often only way to get an answer	. May require prohibitive analysis or computation	Primary application has been in radiation transport. Potential exists for use on almost all simulations.
Correlated Sampling (Secs. 3.1.7, 4.3.7)	Some or part of the random numbers are used in successive trials to provide correlation among the outcome in successive trials.	● sensitivity studies are of interest ● effect of parameter variations to be determined ● small perturbations are of interest	● easy to comprehend ● applicable to many practical situations ● little risk involved in application ● often only way to detect small effects	● may be difficult to implement ● sensitive to specific applications	Can be used in almost any problem where sensitivities are of interest. Examples are network design, tactics evaluation, reliability, queueing system design, etc.

TABLE 4.2 (Continued)

Variance Reduction Technique or Procedure(1)	Description of Technique	Suggested Criteria for Application	Advantages(2)	Disadvantages	Comments on Typical Applications(3)
History Reanalysis (Secs. 3.1.8, 4.3.8)	Takes results of one simulation and reinterprets as results of second simulation weighting answers to remove bias.	·Results are needed for two or more similar problems ·Difference in problems may be written as a difference in probability distributions	.Results are highly correlated, thus reducing variance in difference .Saves computation time for second simulation	.Difference between problems may be more than a difference in probability distribution .Variance of second problem may be prohibitively high in history reanalysis	.Potentially many applications but, so far, mainly used in radiation transport.
Control Variates (Secs. 3.1.9, 4.3.9)	Using simplified or approximate representations for undetermined parts in the simulation. These representations are positively correlated in the sense that they follow the trend of the true expressions.	● portions of the problem can be approximated with an analytical representation ● a simple low variance approximate simulation is known	● simple calculations usually required ● provides considerable flexibility ● can give great improvement	● requires some understanding of the process ● simple analytic approximation may not exist	Usefulness to queueing problems has been demonstrated. Its potential is great for use in almost all areas where some approximate representation of the random variable whose expected value is to be determined is known
Antithetic Variates (Secs. 3.1.10, 4.3.10)	Using simplified or approximate expressions for undetermined parts of the problem except that an inverse or negative correlation is maintained in the sense they follow the opposite trend of the true expressions.	● analytical or approximate representations are known and correlate in a negative or inverse fashion with the problem variables	● simple calculations usually required ● provides considerable flexibility ● can give great improvement	● requires some understanding of the process ● usually easier to use than control variates	Has been applied in queueing problems to improve efficiency. Could be used in a variety of operations research problems in reliability queueing, sensitivity analysis, etc.
Regression Method (Secs. 3.1.11, 4.3.11)	Used with linear combinations of parameters being estimated in a simulation model. An approximate minimum variance estimate is included which exploits any inherent correlation among the parameters to reduce the variance.	● estimated parameters are linear combinations of random variables ● correlation among variables comprising the estimator are known to exist	● little, if any, bias results ● exploits correlation ● little to lose if there is no correlation ● can be applied with little or no information about the problem	● increased computation required ● limited to linear combinations of random variables in its usual form	Should be applicable in almost any problem (reliability, queueing, etc.) where the outcome is a linear combination of statistical estimators.

TABLE 4. 2 (Continued)

Variance Reduction Technique or Procedure(1)	Description of Technique	Suggested Criteria for Application	Advantages(2)	Disadvantages	Comments on Typical Applications(3)
Sequential Sampling (Sec. 3.4.1)	Development of estimates for parameters to be used in other variance reduction techniques by using information generated from previous sampling	● little or no information is available about the process ● other variance reduction techniques (e.g. importance, control variates, etc.) are to be implemented	● can almost always be applied ● requires little or no a priori knowledge ● can be used in conjunction with other techniques	● may require considerable calculations initially ● efficiency may be very low ● may not be useful in certain problems (such as low probabilities) ● not well developed	Evaluation of multiple integrals, estimation of distribution parameters and has found use in lifetime studies
Orthonormal Functions (Sec. 3.4.2)	The orthogonal characteristics of certain orthonormal functions are exploited in formulating a sampling scheme with reduced variance.	● many-dimensional variables in problem ● multiple integral formulations can be determined	● offers great possibilities for problems with multidimensional integrals	● not well developed for general application ● generally difficult to efficiently implement ● may require considerable calculations ● requires construction of orthonormal functions ● involves sampling from joint random variables	Useful when considering multiple integrals. Has been applied in nuclear physics. However, it may be useful in complex systems problems where a multiple integral formulation is possible
Adjoint Method (Sec. 3.4.3)	The problem is simulated in the backward or reverse sense which is often described by a mathematical adjoint formulation.	● reverse process can be formulated ● starting positions for final outcomes can be defined ● mathematical adjoint formulation is known	● can give indication of what importance sampling is required ● very powerful method when it can be developed ● useful for certain problems involving system input sensitivity studies	● not well developed for most all applications ● difficult to implement ● adjoint concept difficult to formulate for most problems	Has been used almost exclusively in radiation transport. Could be useful in queueing, inventory, etc. if problem can be formulated as integrals or differential equations

TABLE 4.2 (Continued)

Variance Reduction Technique or Procedure[1]	Description of Technique	Suggested Criteria for Application	Advantages[2]	Disadvantages	Comments on Typical Applications[3]
Transformation (Sec. 3.4.4.)	Use a transform or modification of the process being simulated to give a problem that can be simulated with a reduced variance.	● approximations are known for parts of the problem ● an analytical formulation for the problem is known ● transformed equations can be easily simulated	● permits parametric representation of knowledge about the system ● can lead to substantial improvements	● not well developed for most all applications ● may require complete analytical statement of the problem ● appears to be limited in scope of application	Has been used almost exclusively in radiation transport problems. Has potential for queueing, inventory, etc. (1, 22, 25, 26)
Conditional Monte Carlo (Sec. 3.4.5)	Imbeds problem in larger process using conditional probabilities in larger process.	. Process is extremely complicated but can be embedded in a simpler simulation.	. When applicable can give substantial improvements . Some algorithms do exist for network applications	● not well developed ● sometimes difficult to recognize applicability	Improving the efficiency of estimating the distribution of the function of times to pass through a network. Network application of interest are stochastic PERT, CPM, etc.

[1] Numbers in brackets refer to section where the technique or approach is discussed in detail.

[2] Improvements are always noted relative to straightforward sampling except where indicated.

[3] Numbers in brackets refer to references.

to be used for selecting various techniques. This indicates, of course, the importance of problem definition and the value of known information prior to selecting an approach.

The results of the information requirements defined as noted in Table 4.1 can readily be used in conjunction with Table 4.2 to define a recommended variance reduction approach. For example, if at a certain stage in the problem it is known that a certain range of variables would not be of interest to the final outcomes relative to a second range of the variables, the application of importance sampling or Russian Roulette and splitting is suggested by Table 4.2. The next step would be to proceed to the sections indicated.

A list of references which describe one or more of the various aspects of each of these techniques is included in the corresponding section indicated in Part I.

4.3 PROCEDURES FOR IMPLEMENTATION OF THE SELECTED VARIANCE REDUCTION TECHNIQUES

This section presents general guidelines to implement the more important variance reduction techniques. For convenience, the order in which the methods were presented in Part I will be followed here. It is recommended that the material presented here be used in conjunction with that presented in the corresponding section of Part I. Specifically, the implementation guidelines are presented in the following subsections:

4.3.1 Importance Sampling
4.3.2 Russian Roulette and Splitting
4.3.3 Systematic Sampling
4.3.4 Stratified Sampling
4.3.5 Expected Value
4.3.6 Statistical Estimation
4.3.7 Correlated Sampling
4.3.8 History Reanalysis
4.3.9 Control Variates
4.3.10 Antithetic Variates
4.3.11 Regression

No procedures are presented for implementation of the specialized techniques (sequential sampling, orthonormal functions, adjoint method, transformations and conditional Monte Carlo) presented in Section 3.4 since these are not considered to be well enough developed for general application.

It should be mentioned that the material presented here is intended as a basic guide to provide general procedures for implementation of the variance reduction technique selected from Table 4.2. In many cases, it is difficult to provide anything more than a rather general description of the steps to be implemented. However, where possible specific formulae or recipes which have general applicability were included. The specific analytical formulae of interest are also summarized in Appendix A.

4.3.1 Importance Sampling

Importance sampling is the term for modifying the sampling procedure in a manner that will tend to emphasize the more important aspects of the problem. The results must be corrected to account for this modification.

Importance sampling is, in many cases, necessary for obtaining a reasonable answer and, in other cases, can give outstanding improvements in efficiency. This is particularly true when very small probability events can contribute significantly to the outcome of the problem.

'One danger with the application of importance sampling is that it can lead to results worse than that obtained using straightforward sampling. Such a situation can occur when the importance function is not carefully selected. Furthermore, the method requires a fairly good understanding of the problem.

4.3.1.1 Implementation Guidelines for Importance Sampling

The general considerations that should be followed in application of importance sampling are as follows:

1. Attempt to identify one random variable x for importance sampling and its density function f(x). Express if possible the expected value being estimated as

$$I=\int g(x)f(x)dx \qquad (4.1)$$

2. Determine the functional form of g(x). This may be known analytically in trivial cases. In complex simulations, it may be possible to input selected values of x (not necessarily from f(x)) and actually obtain an estimate for the form of g(x).

3. Plot the shape of f(x)g(x) and select an importance function f*(x) that is "similar" in form to g(x)f(x). A sketch of the basic ideas involved is shown in Fig. 4.1.

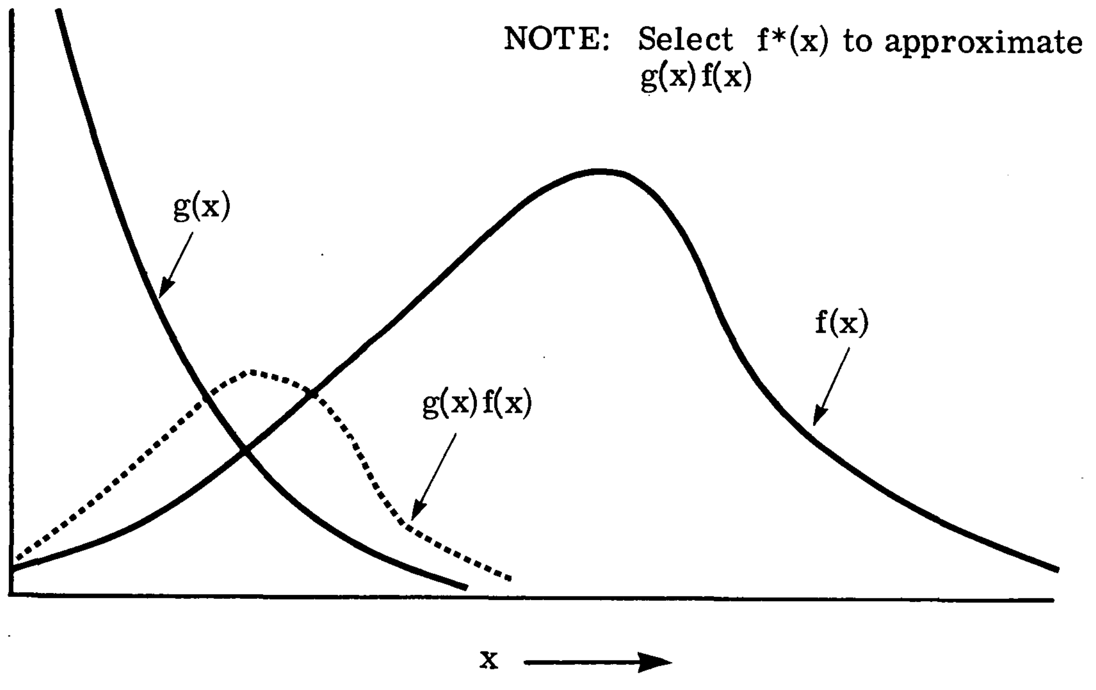


Fig. 4.1. Qualitative Description of Importance Sampling

4. The new estimator for I is

$$\hat{I}_1 = \frac{1}{N} \sum_{i=1}^{N} \frac{g(X_i)f(X_i)}{f^*(X_i)} \tag{4.2}$$

where $X_1, \ldots, X_n$ is a random sample from $f^*(x)$.

5. The estimator for the sample variance is given by

$$S^2 = \frac{N}{N-1}\left[\frac{1}{N} \sum_{i=1}^{N} \left\{\frac{g(X_i)f(X_i)}{f^*(X_i)}\right\}^2 - \hat{I}_1^2\right] \tag{4.3}$$

6. Obtain a random sample $X_1, \ldots, X_N$ using crude Monte Carlo from $f(x)$ and estimate I_1 and S_1^2.

7. Obtain an estimate for the efficiency of importance sampling using

$$\hat{\epsilon} = \frac{tS^2}{t_1 S_1^2} \tag{4.4}$$

where t_1 and t are the times required to obtain N samples with and without importance sampling respectively.

It should be noted that $\hat{\epsilon}$ is a random variable and is subject to uncertainty which will depend on the sample size N. Thus, it is usually a good practice to make N as large as reasonably possible to obtain a good estimate for ϵ. In the event several random variables are involved in the problem, the suggested procedure is:

1. Identify one random variable x for importance sampling and express the estimator as

$$I = \int f(x)\left[\int g(x,\vec{y})f(\vec{y}|x)d\vec{y}\right]dx \tag{4.5}$$

where the vector $\vec{y}$ is all the remaining random variables.

2. For a range of values of x, estimate $E[g^2(x,\vec{y}|x)]$ where $\vec{y}$ are selected from their corresponding probability distributions. If $\vec{y}_1,\dots,\vec{y}_n$ are random samples of $\vec{y}$, then

$$\hat{g}^2(x,y|x) = \frac{1}{n}\sum_{i=1}^{n} g^2(x,\vec{y}_i|x) \qquad (4.6)$$

is used to estimate $E[g^2(x,\vec{y}|x)]$.

3. Select $f^*(x)$ to approximate

$f(x)E[\hat{g}^2(x,y|x)]^{1/2}$.

(This can sometimes be accomplished graphically.)

4. Estimators with importance sampling are now respectively

$$\hat{I}_1 = \frac{1}{N}\sum_{i=1}^{N}\frac{g(X_i,\vec{Y}_i)f(X_i)}{f^*(X_i)} \qquad (4.7)$$

and

$$S_1^2 = \frac{N}{N-1}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{g(X_i,\vec{Y}_i)f(X_i)}{f^*(X_i)}\right]^2 - \hat{I}_1^2\right\} \qquad (4.8)$$

The efficiency is computed in the same manner as in the previous case

4.3.2 Russian Roulette and Splitting

Russian Roulette and splitting is a powerful technique that is easiest to apply when the problem is characterized by a series of events. Examples are found in problems in queueing, series subsystems, radiation transport, random walk, etc.

This technique is in essence a simplified form of importance sampling. One potential difficulty with Russian Roulette and splitting is the possibility that it may lead to a large number of histories being traced through the system at one time. While Russian Roulette is generally easy to implement, incorporating splitting, (especially in an existing program), may be more difficult due to the need to store problem conditions and later 'restart' in mid-history. Following a 'split', the program must continue with the simulation of one of the histories until it terminates, and the program must then go back to the point of the split and restore the program conditions at that time so that the next 'daughter' simulation can proceed.

The general steps that can be followed for Russian Roulette and splitting are:

1. Identify stages in the problem for which the possible conditions at those stages can be divided into regions $R_1, R_2, \ldots, R_N$ such that all the points in any one region have roughly the same importance.

2. Choose average weight standards, w_{A_i}, $i = 1, N$, for each region that are inversely proportional to that region's importance. The mean weight standard at any stage should be roughly the average weight expected to reach that stage from the previous stage.

3. If no other variance reduction techniques are being employed, set high and low weight standards, w_{H_i} and w_{L_i}, equal to the average, w_{A_i}. If there are other variance reduction techniques in use which are causing weight changes, then w_{H_i} and w_{L_i} should be spaced sufficiently far above and below w_{A_i} so that there is no unnecessary Russian Roulette and splitting but also so that there is not a wide variation of weights among histories of roughly the same importance.

4. Whenever a history arrives at a particular stage in region R_i with a weight w, carry out the following manipulations:

   a. If $w < w_{L_i}$, play Russian Roulette:

      i. Kill (terminate) the history with probability $1 - \frac{w}{w_{A_i}}$, or

ii. Let the history survive (continue) carrying a new weight w_{A_i} with probability w/w_{A_i}.

b. If $w_{L_i} < w < w_{H_i}$, continue the history with weight w.

c. If $w > w_{H_i}$, carry out splitting:

i. Determine n such that $0 \leq w - n\, w_{A_i} < w_{A_i}$

ii. Split the history into n 'daughter' histories which start at this point with weight w_{A_i}.

iii. With probability $\frac{w - nw_{A_i}}{w_{A_i}}$, create one more daughter history with weight w_{A_i}.

5. In scoring, accumulate the outcomes from all daughter histories which originated from the same initial or parent history. That is, form estimates

$$\hat{I}_j = \sum_{\ell,\text{ daughter of } j} g(\vec{X}_\ell) w_\ell \tag{4.9}$$

6. Form the final estimate by averaging estimates from N starting histories

$$\hat{I} = \frac{1}{N} \sum_{j=1}^{N} \hat{I}_j \tag{4.10}$$

and calculate the sample variance:

$$S^2 = \frac{N}{N-1} \left[\frac{1}{N} \sum_{j=1}^{N} \hat{I}_j^2 - \hat{I}^2 \right] \tag{4.11}$$

4.3.3 Systematic Sampling

There are two ways to implement systematic sampling. Both are presented below although it is generally recommended that Method II be used. The application of systematic sampling can be generally most effective when initial conditions for a problem are selected from a probability distribution, although other applications can be identified. In any event it is convenient to consider the usual integral form

$$I = \int_{-\infty}^{+\infty} g(x)f(x)dx \qquad (4.12)$$

Method I

In this method, systematic sampling is implemented as follows:

1. The cumulative distribution for f(x) is determined as indicated in Fig. 4.2. The range (0,1) is divided into N intervals, each of width 1/N as indicated. N should vary between 5 and 50.

2. A random sample $R_1, \ldots, R_n$ of size n is selected from the uniform distribution U(0,1).

3. Using the sequence, $R_1, \ldots, R_n$, n numbers are allocated to each interval using.

$$R_{ij} = \frac{j - R_i}{N} \; ; \quad \begin{array}{l} i = 1, \ldots, n \\ j = 1, \ldots, N \end{array} \qquad (4.13)$$

4. Determine the values of X_{ij} from

$$R_{ij} = \int_{-\infty}^{X_{ij}} f(x)dx \; ; \quad \begin{array}{l} i = 1, \ldots, n \\ j = 1, \ldots, N \end{array} \qquad (4.14)$$

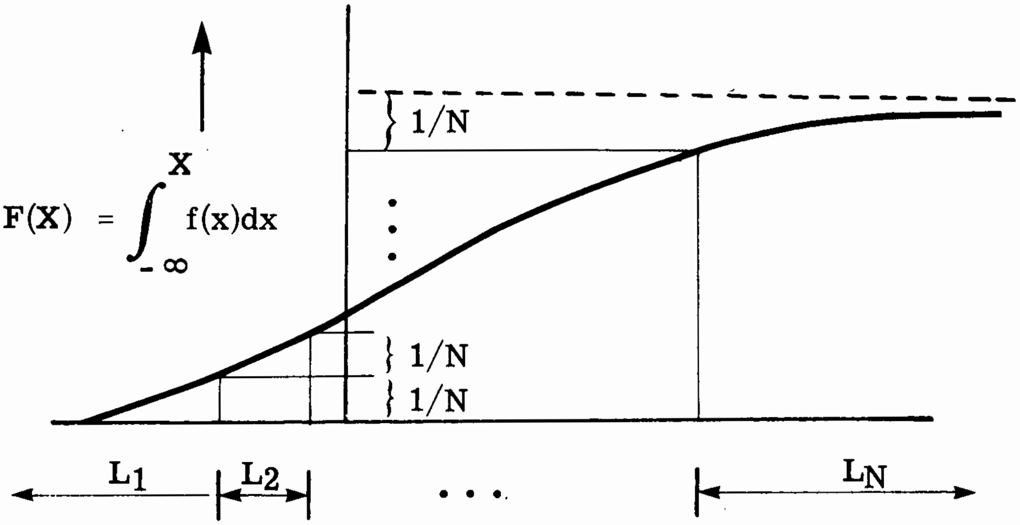


Fig. 4.2. Systematic Sampling

5. Once the values for X_{ij} are obtained, the estimator used for I is

$$\hat{I} = \frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N} g(X_{ij}) = \frac{1}{n}\sum_{i=1}^{n}\hat{I}_i \tag{4.15}$$

where

$$\hat{I}_i = \frac{1}{N}\sum_{j=1}^{N} g(X_{ij}) \tag{4.16}$$

6. Estimate the sample variance using

$$S^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}\hat{I}_i^2 - \hat{I}^2\right] \tag{4.17}$$

Method II

In this method, the sampling is structured as follows:

1. The cumulative distribution for f(x) is determined as indicated in Fig. 4.2. The range (0, 1) is divided into N intervals, each of width 1/N. N should vary between 5 and 50.

2. n sets of N random numbers each, $R_{i1}, \ldots, R_{iN}; \ldots ; R_{n1}, \ldots, R_{nN}$ are selected from U(0, 1).

3. n random numbers are allocated to each interval according to

$$R'_{ij} = \frac{j-R_{ij}}{N} \quad ; \quad \begin{matrix} i = 1, \ldots, n \\ j = 1, \ldots, N \end{matrix} \tag{4.18}$$

4. The values of X_{ij} are determined from

$$R'_{ij} = \int_{-\infty}^{X_{ij}} f(x)dx \tag{4.19}$$

5. The estimator for the integral I is then obtained with

$$\hat{I} = \frac{1}{nN} \sum_{i=1}^{n} \sum_{j=1}^{N} g(X_{ij}) = \frac{1}{n} \sum_{i=1}^{n} \hat{I}_i \qquad (4.20)$$

where

$$\hat{I}_i = \frac{1}{N} \sum_{j=1}^{N} g(X_{ij}) \qquad (4.21)$$

6. The estimate for the sample variance is obtained using

$$S^2 = \frac{n}{n-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{I}_i^2 - \hat{I}^2 \right\} \qquad (4.22)$$

It should be noted that the difference between Method II and Method I is that the random numbers are generated independently in each of the N intervals. This requires more effort than Method I, although Method II will generally give better results.

4.3.4 Stratified Sampling

This variance reduction technique, sometimes called quota sampling, is similar to systematic sampling in that specific numbers of samples are generated in each of several intervals spanning the sample space. In systematic sampling the number of cases in each interval is determined from the 'natural' proportions of the process being simulated. In stratified sampling, on the other hand, the number of samples in each interval is chosen to optimize the accuracy of the simulation.

Stratified sampling can be implemented in the following steps:

1. Break the range of the random variable being simulated into N intervals of length $L_1, \ldots, L_N$ as indicated in Fig. 4.3. Typically N should be between 5 and 50. Each L_j is selected so the variation in $g(x)f(x)$ is approximately the same.

2. Determine P_j, the probability that x will be in the interval L_j, from

$$P_j = \int_{x \epsilon L_j} f(x)dx \qquad j = 1, \ldots, n \tag{4.23}$$

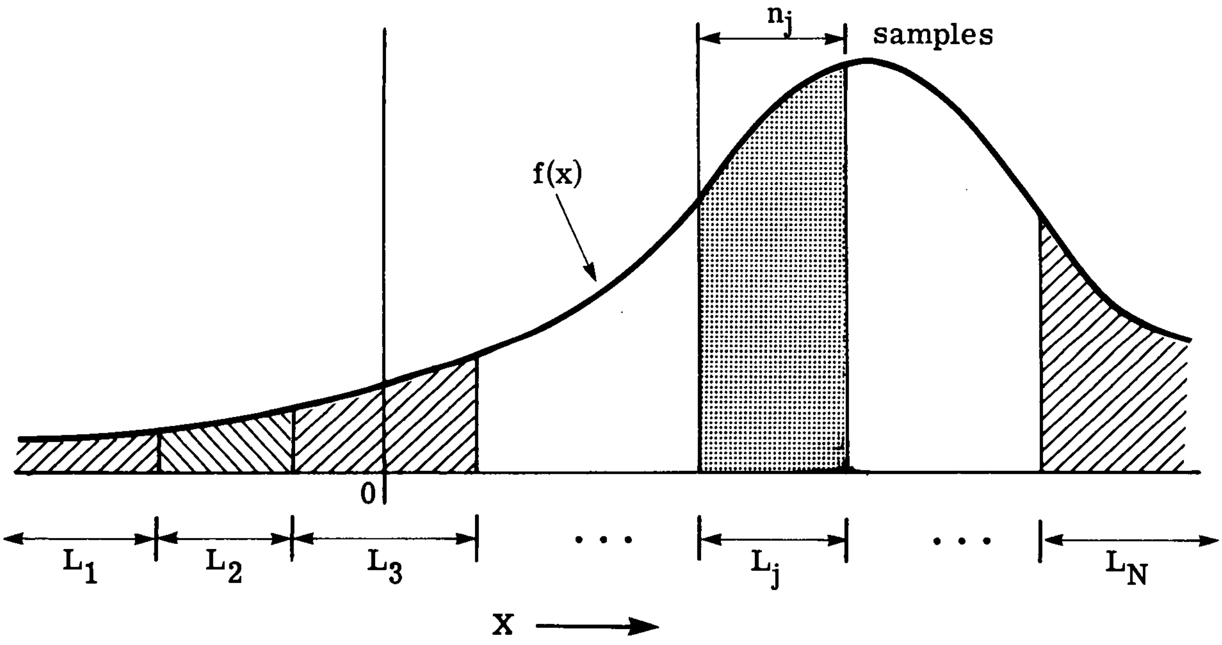


Fig. 4.3. Illustration of Systematic Sampling

3. Arbitrarily assign n_j' ; $j = 1,\ldots,N$ as the number of samples to select from each interval where $\Sigma n_j' = n$, the total number of samples desired. Select R_{ij}' ; $i = 1,\ldots,n_j'$; $j = 1,\ldots,N$ from $U(0,1)$.

4. Determine X_{ij}' from

$$R_{ij}' P_j + \sum_{\ell=1}^{j-1} P_\ell = \int_{-\infty}^{X_{ij}'} f(x)dx \tag{4.24}$$

and determine

$$S_j'^2 = \frac{n_j'}{n_j'-1}\left[\frac{1}{n_j'}\sum_{i=1}^{n_j'} g^2(X_{ij}') - \hat{I}_j^2\right] \tag{4.25}$$

where

$$I_j' = \frac{1}{n_j'}\sum_{i=1}^{n_j'} g(X_{ij}') \tag{4.26}$$

5. Determine n_j using

$$\frac{n_j}{n} \simeq \frac{P_j S_j'}{\sum_{j=1}^{n} P_j S_j'} \tag{4.27}$$

where n is the total sample size to be selected $\left(\text{i.e., } \sum_{j=1}^{N} n_j = n\right)$.

6. Select R_{ij} ; $i = n_j' + 1, \ldots, n_j$; $j = 1, \ldots, N$ from $U(0,1)$ and determine X_{ij} ; $i = n_j' + 1, \ldots, n_j$; $j = 1, \ldots, N$ from

$$R_{ij}P_j + \sum_{\ell=1}^{j-1} P_\ell = \int_{-\infty}^{X_{ij}} f(x)dx \tag{4.28}$$

7. Estimate I and σ^2 using

$$\hat{I} = \sum_{j=1}^{N} P_j \hat{I}_j \tag{4.29}$$

$$S^2 = \sum_{j=1}^{N} \frac{n_j P_j^2}{n_j - 1} \left[\frac{1}{n_j} \sum_{i=1}^{n_j} g^2(X_{ij}) - \hat{I}_j^2 \right] \tag{4.30}$$

where

$$\hat{I}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} g(X_{ij}) \tag{4.31}$$

The efficiency of stratified sampling can now be estimated using S^2 as determined in the last step.

If the

$$\sigma_j^2 = \int_{X \epsilon L_j} \frac{f(x)}{P_j} [g(x) - I_j]^2 dx \tag{4.32}$$

are known or can be estimated from a priori knowledge of the system being simulated, then steps 3 and 4 can be omitted and n_j can be determined directly from

$$n_j \approx \frac{nP_j\sigma_j}{\sum_{j=1}^{N} P_j\sigma_j} \qquad (4.33)$$

Alternatively, steps 3 through 5 can be performed iteratively to determine a best set of values for $n_1, \ldots, n_N$

4.3.5 Expected Value

In any simulation consisting of several stages, it may be that the expected value of some of the stages is either known or can be determined analytically. In such cases the possibility of achieving variance reduction through replacing one of the random stages by its expected value should be investigated. The steps which must be taken to determine if replacement by an expected value is feasible are:

1. Identify the stochastic processes in the overall simulation for which the expected value can be calculated efficiently.
2. For each stochastic process identified in 1., determine if the random element in the process is an essential part of the simulation model. If the fact that the process randomly takes on a range of values affects the rest of the simulation, then the process cannot be replaced by its mean value. If, on the other hand, only the first moment of the random distribution, and not any higher moments, affects the rest of the simulation, then it is possible to replace the random process by its first moment or expected value. For any given physical system, the determination of which stochastic elements are essential usually depends on the particular parameters being estimated.
3. If a random process can be replaced by its expected value without loss of realism, that will always reduce the variance. However, it may not improve efficiency as it may cause excessive computation. If the process in question is a branch point where the history may go in either of two (or more) directions, then replacing the stochastic event by its expected value requires splitting the history with each part going in one of the

directions and carrying the probability of that branch as a weight. Should enough of these events be encountered the number of split histories which must be computed can easily expand beyond a reasonable bound. Alternatively, one of the branches of the decision can be to terminate the history; in this case the history is not split but continues from the branch point with a weight representing the survival probability. This can easily lead to histories with very low weights which usually represents a loss in efficiency in the calculation. Again, this determination is likely to depend on the particular parameters of interest in the calculation and it is impossible to give general guidelines.

Figure 4.4 shows, in abbreviated form, the considerations used in choosing between expected value, statistical estimation, and crude Monte Carlo techniques for the simulation of a random process.

Once the decision has been made to replace a stochastic process by its expected value, the implementation depends on the role of the process in the overall simulation. Specifically,

1. If the process is one of selection of a random variable, then the process becomes merely a deterministic setting of the variable to its expected value and the simulation proceeds as before with no change in estimators, that is, if y is to be selected from f(y), then set

$$y = E[f(y)] \tag{4.34}$$

and continue the simulation.

2. If the process represents a decision between terminating or not terminating the history, then the history continues but with a reduced weight representing the probability of survival. That is,

$$w_{new} = w_{old} \cdot p_s \tag{4.35}$$

where p_s is the probability of survival (non-termination) at the decision point and w_{old} and w_{new} are the weights of the history before and after the replaced random process.

For any parameter being calculated, an estimate for each history can be made by summing the contributions from that history. That is,

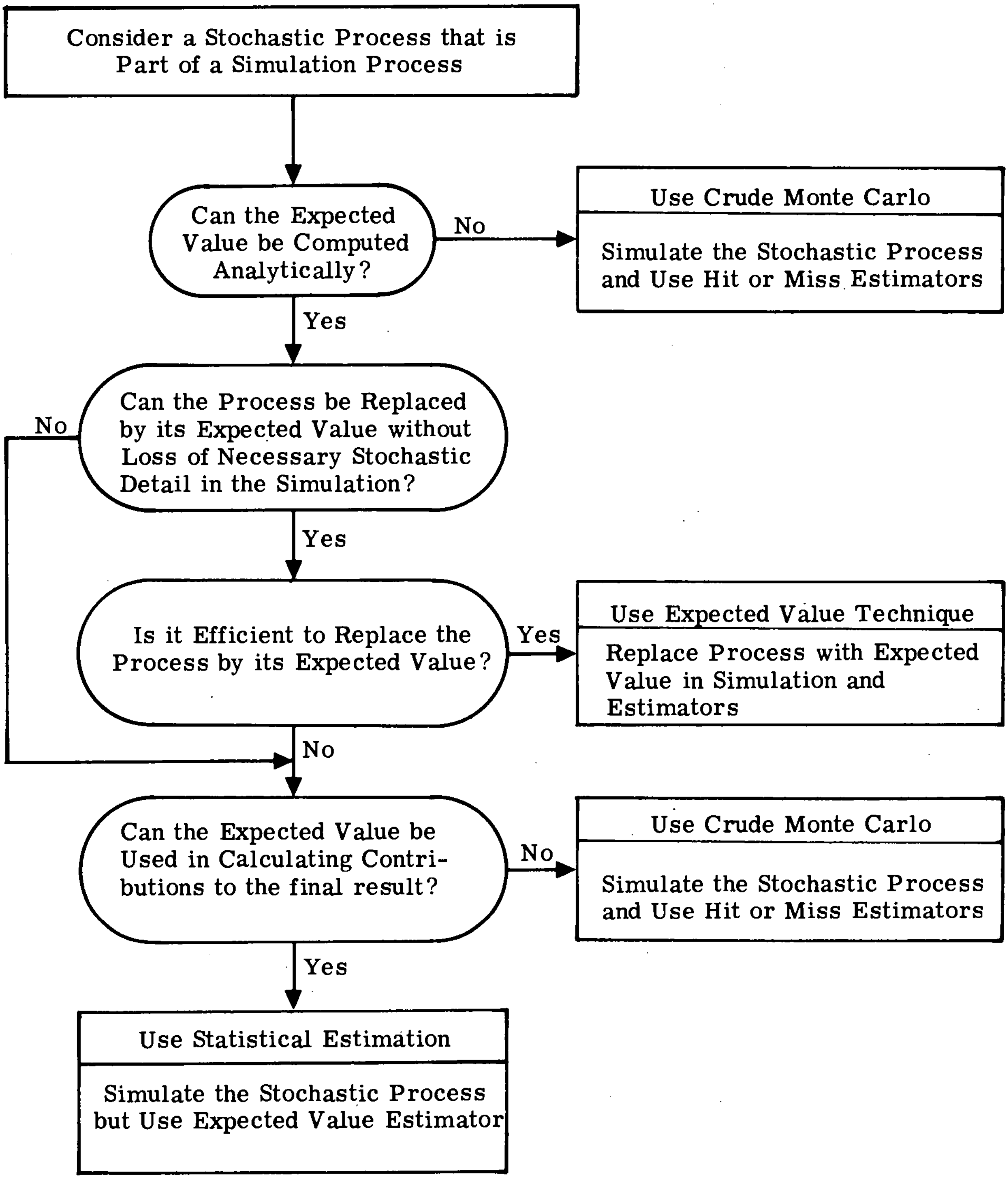

Consider a Stochastic Process that is Part of a Simulation Process
Can the Expected Value be Computed Analytically?
No
Use Crude Monte Carlo
Simulate the Stochastic Process and Use Hit or Miss Estimators
Yes
Can the Process be Replaced by its Expected Value without Loss of Necessary Stochastic Detail in the Simulation?
No
Yes
Is it Efficient to Replace the Process by its Expected Value?
Yes
Use Expected Value Technique
Replace Process with Expected Value in Simulation and Estimators
No
Can the Expected Value be Used in Calculating Contributions to the final result?
No
Use Crude Monte Carlo
Simulate the Stochastic Process and Use Hit or Miss Estimators
Yes
Use Statistical Estimation
Simulate the Stochastic Process but Use Expected Value Estimator

Fig. 4.4 Problem Characteristics and the Choice of Crude Monte Carlo, Expected Value, and Statistical Estimation Techniques

$$\hat{I}_i = \sum_j w_{ij}\, g(X_{ij}) \qquad (4.36)$$

where w_{ij} is the weight of the i^{th} history at the time of the j^{th} contribution to the final result. Then the final estimate and the sample variance are given by

$$\hat{I} = 1/N \sum_{i=1}^{N} \hat{I}_i \qquad (4.37)$$

and

$$S^2 = \frac{N}{N-1} \left\{ 1/N \sum_{i=1}^{N} \hat{I}_i^2 - \hat{I}^2 \right\} \qquad (4.38)$$

If the contributions to a parameter from a history would have come from the terminations in the process which was replaced by its expected value, then the loss of weight at each such step is the proper estimate for the expected terminations. In this case we get

$$\hat{I}_i = \sum_j (w_{old,\, ij} - w_{new,\, ij}) \cdot g(X_{ij}) = \sum_j w_{old,\, ij}\, (1-p_s) \cdot g(X_{ij}) \qquad (4.39)$$

where j denotes the j^{th} occurrence of the replaced event in the i^{th} history. The estimates for $\hat{I}$ and S^2 remain as in (4.37) and (4.38) above.

3. If the process represents a decision between two or more branch points, then the history must be split and followed from that point on as two separate histories, each taking a different branch and carrying a weight equal to the probability of that branch. Parameters are estimated by summing weighted contributions from all daughter histories resulting from an

original parent history, using formulas identical to (4. 36), (4. 37) and (4. 38). In cases 2 and 3 above, histories may develop weights which are very small. As this may entail spending a good deal of computing time calculating histories that can make only a trivial contribution to the result, the efficiency may be very low. To remedy this, Russian Roulette (see Section 4. 1. 2) can be used to eliminate those histories where weights become too small.

4.3.6 Statistical Estimation

It is not essential, and frequently not efficient, for a simulation of a physical process to be carried out to the natural termination of the process in estimating final outcomes. It is always proper to stop the simulation at any point and to calculate through analytic or numerical means the expectation of reaching any final outcome. Indeed, the sooner the simulation is stopped and the more analytic calculations are done, the lower the variance will be. Obviously, however, the sooner the simulation is stopped, the more complex and difficult the analytic calculations become and a point is quickly reached where the overall efficiency is less despite the gain in variance reduction. At the last step in the simulated process, the probability of reaching the various final outcomes needs to be determined in order to do the simulation. Thus, it is generally advantageous to use analytic expectations for the final step. Whether the analytic calculations should be carried beyond the final step will depend on the particular process and results desired, but generally it is less efficient to use analytic expectations for more than the last step.

If the process being simulated is a once-through process, i.e., the final step can be reached only once each history, then the use of expected outcomes is equivalent to the expected value technique. If the process is iterative or repetitive with many passes through a branch point where a final outcome is possible, there are two ways of using the analytic computations. One is by the expected value technique as outlined in the previous section. The other is called statistical estimation and should be used whenever the expected value technique would be inefficient. In certain cases where the probability of the desired final outcome is extremely small, statistical estimation may be the only way to obtain an answer. Figure 4.4 shows, in abbreviated form, the considerations used in choosing between statistical estimation, expected value, and crude Monte Carlo techniques for the simulation of a stochastic process.

Statistical estimation is implemented as follows:

1. Identify the stochastic processes in the simulation which have the desired final outcome as one possible alternative.

2. Each time such a process is encountered in simulating a history, a contribution of

$$g(\vec{X}, Y_f) f(Y_f | \vec{X}) \tag{4.40}$$

is scored, where $g(\vec{x}, y)$ is the function being integrated by the simulation, Y_f is the desired outcome of the particular process at hand, $\vec{X}$ denotes the current state of all the other variables in the system, and $f(Y_f | \vec{X})$ is the conditional probability of obtaining outcome Y_f given $\vec{X}$ as the status of the system.

3. Do not modify the simulation itself, but continue to model the stochastic process by drawing a random number and probabilistically selecting an outcome, i.e., select a Y from $f(y|\vec{x})$

4. Do not mix statistical estimation with crude Monte Carlo, i.e., if the outcome of Step 3 turns out to be Y_f, no additional scoring is made. The contribution at this step remains $g(\vec{X}_j, Y_f) f(Y_f | \vec{X}_j)$.

5. Form an estimate for the entire history by summing the contributions

$$\hat{I}_i = \sum_j g(\vec{X}_{ij}, Y_f) f(Y_f | \vec{X}_{ij}) \ . \tag{4.41}$$

where j runs over all occurrences of the particular process being estimated in the i^{th} history.

6. The final estimate is averaged over all histories

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} \hat{I}_i \tag{4.42}$$

and the sample variance is

$$S^2 = \frac{N}{N-1} \left[\frac{1}{N} \sum_{i=1}^{N} \hat{I}_i^2 - \hat{I}^2 \right] . \tag{4.43}$$

4.3.7 Correlated Sampling

Correlated sampling can be one of the most useful variance reduction techniques due to the wide applicability of the technique as well as to the large efficiency gains which can be realized.

There are several types of situations where the use of correlated sampling is indicated. These include:

- The effect of a small change in the system is to be calculated.
- The difference in a parameter in two or more similar cases is of more interest than absolute values in any one case.
- A parametric study of several problems is to be performed. This has greatest potential when the problems are relatively similar in nature.
- The answer to one of several similar problems is known accurately. The answers to the unknown problems can often be estimated from the known result.

The aim of correlated sampling is to produce a high positive correlation between two similar simulations so that the variance of the difference in results is considerably smaller than it would be if the two simulations were statistically independent. Unfortunately, there is no general procedure that can be implemented in correlated sampling. However, the following procedures can convey some notion of the methods useful in producing correlation. Let us begin by considering two similar simulations involving only a single variable, i.e., it is desired to estimate

$$\Delta = I_1 - I_2 \tag{4.44}$$

where

$$I_1 = \int_{-\infty}^{+\infty} g_1(x) f_1(x) dx \tag{4.45}$$

and

$$I_2 = \int_{-\infty}^{+\infty} g_2(y) f_2(y) dy \tag{4.46}$$

Then the implementation of correlated sampling proceeds as follows:

1. Generate a random sample $X_1, \ldots, X_N$ from $f_1(x)$ and a sample $Y_1, \ldots, Y_N$ from $f_2(y)$ using

$$R_i = \int_{-\infty}^{X_i} f_1(x) dx = \int_{-\infty}^{Y_i} f_2(y) dy \; ; \quad i = 1, \ldots, N \tag{4.47}$$

   where R_i ; $i = 1, \ldots, N$ is a random sample from U(0, 1).

2. Estimate Δ using

$$\hat{\Delta} = \frac{1}{N} \sum_{i=1}^{N} [g_1(X_i) - g_2(Y_i)] = \frac{1}{N} \sum_{i=1}^{N} \Delta_i \tag{4.48}$$

   where

$$\Delta_i = g_1(X_i) - g_2(Y_i) \tag{4.49}$$

   Estimate the sample variance using

$$S^2 = \frac{N}{N-1} \left\{ \frac{1}{N} \sum_{i=1}^{N} \Delta_i^2 - \hat{\Delta}^2 \right\} \tag{4.50}$$

   (Batching may also be used.)

If $f_1(x)$ is similar to $f_2(y)$, the random samples $X_1, \ldots, X_N$ and $Y_1, \ldots., Y_N$ will be highly correlated. If $g_1(x)$ is also similar to $g_2(y)$ then the estimates will also be highly correlated. This will greatly reduce the variance in Δ, as the history values, Δ_i, will reflect almost totally the real differences in $g_1(x)f_1(x)$ and $g_2(y)f_2(y)$ and not random "noise" due to a difference in random numbers used.

In the more general case the simulation involves a sequence of random variables $\vec{x} = x_1, x_2, \ldots, x_k$ and the integrals being estimated are

$$I_1 = \int g(\vec{x}) f(\vec{x}) d\vec{x} = \iiint \ldots \int g(x_1, x_2, \ldots, x_k) f(x_1) f(x_2 | x_1) \ldots .$$
$$f(x_k | x_1 x_2, \ldots, x_{k-1}) dx_1 dx_2, \ldots, dx_k \tag{4.51}$$

and similarly for I_2. The procedure now is as follows:

1. Identify, to the maximum extent possible, where identical random numbers can be used on both problems. Clearly, when parameters are changed between the problems, this may not always be possible. However, it may be possible to use the same uniform random numbers throughout. (In sequential or multistage problems it may be possible to precompute once the portions of the simulation which will be identical in the two cases and then use these computations in the two simulations, thereby reducing the computational effort required.)

2. For each history i, generate a random sample $R_{i1}, \ldots, R_{ik}$ from the unit uniform distribution U(0, 1). Solve for $X_{i1}, \ldots, X_{ik}$ using

$$R_{ij} = \int_{-\infty}^{X_{ij}} f_1(x_j | X_{i1}, X_{i2}, \ldots, X_{i(j-1)}) dx_j \tag{4.52}$$

and for $Y_{i1}, \ldots, Y_{ik}$ using

$$R_{ij} = \int_{-\infty}^{Y_{ij}} f_2(y_j | Y_{i1}, Y_{i2}, \ldots Y_{i(j-1)}) dy_j \tag{4.53}$$

3. Form an estimate for each history

$$\hat{\Delta}_i = g_1(X_{i1}, X_{i2}, \ldots, X_{ik}) - g_2(Y_{i1}, Y_{i2}), \ldots, Y_{ik}) \tag{4.54}$$
$$= \hat{I}_{1i} - \hat{I}_{2i}$$

4. Calculate the final estimate by

$$\hat{\Delta} = \frac{1}{N} \sum_{i=1}^{N} \hat{\Delta}_i$$

and the sample variance

$$S^2 = \frac{N}{N-1} \left[1/N \sum_{i=1}^{N} \hat{\Delta}_i^2 - \hat{\Delta}^2 \right]$$

(Alternatively, batching can be used to calculate the variance.)

In most practical problems one does not want to develop a completely new simulator to estimate the difference in parameters but is rather faced with the problem of using an existing simulator program which was designed to solve a single case. Thus two separate runs must be made, but the correlation generated in step 2 can be retained if the basic random sample $R_{11}, R_{12}, \ldots, R_{1k}, \quad R_{21}, \ldots, R_{Nk}$ is generated in both programs. Here the property possessed by the congruential uniform random number generator of always producing the same sequence of numbers when given the same starting value becomes very useful. It is then only necessary to ensure that the two separate runs start with the same random value and they will continue to generate the same sequence $R_{11}, \ldots, R_{Nk}$. However, this is not quite enough for most simulations. It is usually the case that k, the number of random variables in a history, is itself a random variable and can vary from one simulation to the other due to the difference in the problem solved. Thus, for the maximum correlation the random number generator should be set at the start of each history to a value that is common in both runs, i.e., force the values $R_{11}, R_{21}, R_{31}, \ldots, R_{N1}$ to be the same in both runs and all the rest of the random sequence will be identical in the two cases. If step 1 identified portions of the simulation which could be identical in the two cases, it would be desirable to force common starting values on the random number generator at the start of each such portion of the simulation and not just at the start of the history.

To generate values $R_{11}, R_{21}, R_{31}, \ldots, R_{N1}$ which are themselves random numbers but are consistent in the two runs, a second random number generator is used which does nothing but generate starting values for the main random number generator used in the simulation. As this second generator is used only once each history, it is unaffected by the difference in the two cases and will generate identical starting values in both runs. (Note that one should use the binary integer produced by the second generator and not the floating point random number as a starting value for the main random number generator.)

Having made two separate runs which are correlated, the problem then is to compute the difference estimates. To estimate the variance produced by the correlation one must save the estimates, $\hat{I}_{1i}$ and $\hat{I}_{2i}$, produced each history (or each batch, if batching is used), and then in a separate calculation obtain the estimated difference and the sample variance from 4.54, 4.55, and 4.56.

4.3.8 History Reanalysis

History reanalysis involves generating a series of histories for one case and then reanalyzing them to generate an estimate for a similar case. This combines the advantages of a saving in computer time with correlated sampling (Section 4.3.7) since only one simulation was run to get two results which are correlated due to the use of identical random numbers.

Basically, the types of problems to which history reanalysis can be useful are a subset of those where correlated sampling is useful. That is, when differences in similar problems are to be addressed or when sensitivity analyses are to be performed (see Section 4.3.7). In addition, it is necessary that the difference in the cases studied be expressable as a difference in the random distributions used or in the pay-off function, $g(x)$, and not be a difference in deterministic elements of the simulation. It is commonly a sensitivity analysis where history reanalysis is likely to be most effective.

As in the case of correlated sampling, there is no general procedure that can be followed in history reanalysis. However, the following procedure illustrates the general principles used to derive the results for one problem (I_2) from another problem (I_1) where, as usual,

$$I_1 = \int_{-\infty}^{+\infty} g_1(x)f_1(x)dx \tag{4.57}$$

and

$$I_2 = \int_{-\infty}^{+\infty} g_2(x)f_2(x)dx\ . \tag{4.58}$$

1. Generate a random sample $X_1, \ldots, X_N$ from $f_1(x)$.

2. Obtain an estimate for I_1 from

$$\hat{I}_1 = \frac{1}{N}\sum_{i=1}^{N} g_1(X_i) \tag{4.59}$$

and for σ_1^2 using

$$S_1^2 = \frac{N}{N-1}\left\{\frac{1}{N}\sum_{i=1}^{N} g_1^2(X_i) - \hat{I}_1^2\right\} \tag{4.60}$$

3. Obtain an estimate for I_2 from

$$\hat{I}_2 = \frac{1}{N}\sum_{i=1}^{N} g_2(X_i)\,\frac{f_2(X_i)}{f_1(X_i)} \tag{4.60}$$

and for σ_2^2 from

$$S_2^2 = \frac{N}{N-1}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[g_2(X_i)\,\frac{f_2(X_i)}{f_1(X_i)}\right]^2 - \hat{I}_2^2\right\} \tag{4.61}$$

The above procedure could clearly be used in the analysis of several other integrals and also for the differences in integrals (as was the case used for correlated sampling).

In problems of a sequential (or multistage) nature there may be several points at which reanalysis of the original problem is performed in a manner similar to that described above. Care must be taken however to avoid potential difficulties where branching decision, etc. are based on the outcome of prior events in the problem. These must be appropriately accounted for, but the general procedure outlined above can be useful.

For proper use of history reanalysis, $f_1(x)$ cannot be zero for any point x where $f_2(x)$ is not also zero. The converse, however, is not true. In fact there is a large set of cases where history reanalysis is most useful where $g_1(x)$ is the same as $g_2(x)$ and $f_2(x) \begin{cases} =f_1(x) & \text{for } x \in R_2 \\ =0 & \text{for } x \notin R_2 \end{cases}$. In this case the "weights" used in calculating I_2, $\frac{f_2(X_i)}{f_1(X_i)}$, are either 1 or 0 and we have as a replacement for 4.61

$$\hat{I}_2 = \frac{1}{N_2}\sum_{X_i \in R_2} g_1(X_i) \tag{4.62}$$

and as a sample variance

$$S^2 = \frac{N_2}{N_2-1}\left[\frac{1}{N_2}\sum_{X_i \in R_2} g_1^2(X_i) - \hat{I}_2^2\right]$$

where N_2 is the number of histories for which $X_i \epsilon R_2$. As an example of this kind of case consider a simulation of an antisubmarine mission where the problem is limited by the total mission time. It is desired to calculate kill probabilities for a range of mission times. The simulation is run for the longest time of interest, and the histories can then be reanalyzed to determine kill probabilities for shorter times by simply ignoring the kills which occur after the time in question.

One worry in history reanalysis is that $f_1(x)$ may be too different from $f_2(x)$ to do a reasonable job of estimating I_2. The result may be that 4.61 will prove to be an 'overbiased' or 'underbiased' estimation. It is recommended that users be aware of the considerations mentioned in Section 2.5 whenever using history reanalysis.

4.3.9 Control Variates

In the calculation of an integral

$$I = \int g(x)\ f(x)\ dx, \qquad (4.63)$$

if an approximate function, $h(x) \approx g(x)$, can be found such that $\theta = \int h(x)\ f(x)\ dx$ is known or can easily be determined analytically, then the control variate technique should be used.

In this case the integral I may be written as

$$I = \int_{-\infty}^{+\infty} h(x)f(x)dx + \int_{-\infty}^{+\infty} [(g(x) - h(x)]f(x)dx \qquad (4.64)$$

$$= \theta + \int_{-\infty}^{+\infty} [g(x) - h(x)]f(x)dx = \theta + I_1$$

Then, the simulation is not performed on I directly, but rather on the expected difference between g(x) and h(x), I_1.

The procedure to follow in implementation of control variates is straightforward. Namely,

1. Express the parameter or parameters to be estimated in integral form as indicated above.
2. For each expected value, I, attempt to obtain an approximating function h(x) whose expected value, θ, is known.
3. Structure the simulation such that the difference between h(x) and g(x) given by

$$I_1 = \int_{-\infty}^{+\infty} [g(x) - h(x)] f(x)\, dx \tag{4.65}$$

   is simulated.

4. Generate a random sample $X_1, \ldots, X_N$ from f(x) and estimate I_1 using

$$\hat{I}_1 = \frac{1}{N} \sum_{i=1}^{N} [g(X_i) - h(X_i)] \tag{4.66}$$

   whose sample variance is given by

$$S^2 = \frac{N}{N-1} \left\{ \frac{1}{N} \sum_{i=1}^{N} [g(X_i) - h(X_i)]^2 - \hat{I}_1^2 \right\} \tag{4.67}$$

Frequently, the real process being simulated will give clues as to potential approximating functions. However in many cases an approximate value for g(x) will not be available. This can sometimes be achieved by the use of a sequential sampling procedure in which a few simulations are performed to obtain an approximate representation to g(x). Clearly, the better the approximation for g(x) that can be obtained, the better the results will be.

The extension of the control variate concept to multiple dimensional integrals is clearly evident and is accompanied with the usual complications associated with such extensions.

4.3.10 Antithetic Variates

When two estimators for a parameter of interest are known, then it is possible to combine them to form a third estimator. If the two original estimators are negatively correlated, then the combined estimator can have a variance which is smaller than the variance of either of the original estimators. The usual method for achieving negative correlation is to manipulate the random number generation. Although there are many different ways this can be achieved, the following formulation (which uses a variation of stratified sampling) is very easy to implement.

1. Express, as usual, the parameter (or parameters) to be estimated in integral form as

$$I = \int_{-\infty}^{+\infty} g(x) f(x) dx \tag{4.68}$$

2. Select a value for the parameter $\alpha (0 < \alpha < 1)$ and select X_i and X_1' for i = 1, . . . , N from

$$\alpha R_i = \int_{-\infty}^{X_i} f(x) dx \tag{4.69}$$

and

$$1 - \alpha R_i = \int_{-\infty}^{X_i'} f(x) dx. \tag{4.70}$$

where R_i; i = 1, . . . , N is a random sample from U(0, 1).

3. Construct the unbiased estimator $\hat{\theta}$ using

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} \hat{\theta}_i \tag{4.71}$$

where

$$\hat{\theta}_i = \alpha g(X_i) + (1 - \alpha) g(X_i') \; ; \; i = 1, \ldots, N \tag{4.72}$$

with the sample variance

$$S^2 = \frac{N}{N-1} \left\{ \frac{1}{N} \sum_{i=1}^{N} \hat{\theta}_i^2 - \hat{\theta}^2 \right\} \tag{4.73}$$

Selection of an appropriate value for α is not always clear. One use of antithetic variates uses $\alpha = 1/2$. Another approach is to perform several simulations for various values of α and estimate the efficiency as a function of α.

4.3.11 Regression

The application of regression techniques to reduce variance in simulations can be associated with problems in which a set of integrals $I_1, \ldots, I_p$ are to be estimated from a set of estimators $\hat{\theta}_1, \ldots, \hat{\theta}_n (n \geq p)$ satisfying

$$E[\vec{I}] = \begin{pmatrix} E[I_1] \\ E[I_2] \\ \cdot \\ \cdot \\ \cdot \\ E[I_p] \end{pmatrix} = \vec{A} \; E[\hat{\theta}] = \vec{A} \begin{pmatrix} E[\hat{\theta}_1] \\ \cdot \\ \cdot \\ \cdot \\ E[\hat{\theta}_n] \end{pmatrix} \tag{4.74}$$

where $\vec{A}$ is a known n x p matrix of the form

$$\vec{A} = \begin{pmatrix} a_{11} \cdots a_{1p} \\ \cdot \\ \cdot \\ \cdot \\ a_{n1} \cdots a_{np} \end{pmatrix} \tag{4.75}$$

Based on the concept of minimum variance unbiased estimators, the following procedure may be used to obtain an estimate for I using regression.

1. Perform a simulation N times to obtain N values for each $\theta_1, \ldots, \theta_n$. Define these values as

$$\theta_{ki} \; ; \; \begin{array}{l} k = 1, \ldots, N \\ i = 1, \ldots, n \end{array}$$

2. Obtain the sample means

$$\hat{\theta}_i = \frac{1}{N} \sum_{k=1}^{N} \theta_{ki} \; ; \; i = 1, \ldots, n \tag{4.76}$$

and construct the matrix

$$\vec{\hat{\theta}} = \begin{pmatrix} \hat{\theta}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\theta}_n \end{pmatrix} \tag{4.77}$$

3. Estimate the covariance matrix

$$\vec{\hat{V}} = \begin{pmatrix} \hat{v}_{11} \cdots \hat{v}_{1n} \\ \hat{v}_{21} \cdots \hat{v}_{2n} \\ \cdot \qquad \cdot \\ \cdot \qquad \cdot \\ \cdot \qquad \cdot \\ \hat{v}_{n1} \cdots \hat{v}_{nn} \end{pmatrix} \tag{4.78}$$

where

$$\hat{v}_{ij} = \sum_{i=1}^{N} (\theta_{ki} - \hat{\theta}_i)(\theta_{kj} - \hat{\theta}_j) \; ; \; \begin{array}{l} i = 1, \ldots, n \\ j = 1, \ldots, n \end{array} \tag{4.79}$$

4. The unbiased estimator for $\vec{I}$ is obtained from

$$\vec{\hat{I}} = (\vec{A}^T \vec{\hat{V}}^{-1} \vec{A})^{-1} \vec{A}^T \vec{\hat{V}}^{-1} \vec{\hat{\theta}} \tag{4.80}$$

($\vec{A}^T$ is the transpose of $\vec{A}$.)

It is recommended that an estimate for the sample variance be obtained by batching.

APPENDIX A

SUMMARY OF ANALYTICAL EXPRESSIONS FOR APPLICATION OF VARIANCE REDUCTION TECHNIQUES

APPENDIX A

SUMMARY OF ANALYTICAL EXPRESSIONS FOR APPLICATION OF VARIANCE REDUCTION TECHNIQUES

A convenient summary of the basic expressions used in implementing the more important variance reduction techniques is presented in Table A1. For the most part the table is self explanatory. However, it should be noted that all possibilities are not considered. For example, the results of applying Russian Roulette and splitting is shown for a two-stage problem only.

Also it should be noted that the specialized techniques which were introduced in Part I were not included here.

TABLE A-1

Summary of Analytical Expressions for Application of Variance Reduction

Simulation Technique	Parameter(s) to be Estimated	Random Sample	Estimator, I	Estimator S^2 For σ^2	Comments
Straightforward Sampling	$I = \int_{-\infty}^{+\infty} g(x)f(x)dx$	$X_1, \ldots, X_N$ from $f(x)$	$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} g(X_i)$	$\frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N} g^2(X_i) - \hat{I}^2\right]$	• Also called crude Monte Carlo • (used as a basis for efficiency measures)
Importance Sampling (Single Variable)	$I = \int_{-\infty}^{+\infty} g(x)f(x)dx$	$X_1, \ldots, X_N$ from $f^*(x)$	$\hat{I} = \frac{1}{N}\sum_{i=1}^{N}\frac{g(X_i)f(X_i)}{f^*(X_i)}$	$\frac{N}{N-1}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{g(X_i)f(X_i)}{f^*(X_i)}\right]^2 - \hat{I}^2\right\}$	• $f^*(x)$ is the importance function
(Multiple Variables)	$\int_{-\infty}^{+\infty} f(x)\int_{-\infty}^{+\infty} g(x,\vec{y})f(\vec{y} \mid x)d\vec{y}dx$	$X_1, \ldots, X_N$ from $f^*(x)$ $\vec{Y}_1, \ldots, \vec{Y}_N$ from $f(\vec{y} \mid x)$	$\frac{1}{N}\sum_{i=1}^{N}\frac{g(X_i,\vec{Y}_i)f(X_i)}{f^*(X_i)}$	$\frac{N}{N-1}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{g(X_i,\vec{Y}_i)f(X_i)}{f^*(X_i)}\right]^2 - \hat{I}^2\right\}$	
Russian Roulette and Splitting (Two-stage)	$I = \int\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy$	X_i from $f(x)$ if $X_i \epsilon R_1, Y_i$ from $f(y \mid x)$. If $X_i \epsilon R_2$, $Y_{i1}, \ldots, Y_{in}$ a random sample from $f(y \mid x)$	$\hat{I} = \frac{1}{N}\left[\sum_{X_i \epsilon R_1}\frac{g(X_i,Y_i)}{q} + \sum_{X_i \epsilon R_2}\sum_{j=1}^{n}\frac{g(X_i,Y_j)}{n}\right]$	$\frac{N}{N-1}\left\{\frac{1}{N}\left[\sum_{X_i \epsilon R_2}\frac{g^2(X_i,Y_i)}{q^2} + \sum_{X_i \epsilon R_2}\sum_{j=1}^{n}\frac{g^2(X_i,Y_j)}{n^2}\right] - \hat{I}^2\right\}$	• R_1 = region of x for Russian Roulette with probability p = 1-q • R_2 = region of x for splitting • can be extended to multistage
Systematic Sampling (Method II)	$I = \int^{+\infty} g(x)f(x)dx$	R_{ij} ; $i = 1, \ldots, n,$ $j = 1, \ldots, N$ from $U(0,1)$ X_{ij} from $\frac{i - R_{ij}}{N} = \int_{-\infty}^{X_{ij}} f(x)dx$	$\hat{I} = \frac{1}{N}\sum_{j=1}^{N}\hat{I}_j$ where $\hat{I}_j = \frac{1}{n}\sum_{i=1}^{n} g(X_{ij})$	$\frac{N}{N-1}\left\{\frac{1}{N}\sum_{j=1}^{N}\hat{I}_j^2 - \hat{I}^2\right\}$	• N = number of samples in each bin • n = number of bins
Stratified Sampling	$I = \int_{-\infty}^{+\infty} g(x)f(x)dx$	R_{ij} ; $i = 1, \ldots, N;$ $j = 1, \ldots, n$ from $U(0,1)$ X_{ij} from $R_{ij}p_j + \sum_{\ell=1}^{j-1} p_\ell = \int_{-\infty}^{X_{ij}} f(x)dx$	$\hat{I} = \sum_{j=1}^{n}\frac{P_j}{N_j}\sum_{i=1}^{N_j} g(X_{ij})$	$\sum_{j=1}^{n}\frac{N_jP_j^2}{N_j-1}\left[\frac{1}{N_j}\sum_{j=1}^{N_j} g^2(X_{ij}) - \hat{I}_j^2\right]$	• N_j = number of samples in interval L_j $N_1 + \ldots + N_n = N$ • N = number of intervals • $P_j = \int_{X \epsilon L_j} f(x)dx$ • P_j is selected so variation in $g(x)f(x)$ is the same in each interval
Expected Values (Two-stage)	$I = \int_{-\infty}^{+\infty}\int g(x,y)f(x,y)dxdy$	$X_1, \ldots, X_N$ from $f(x)$	$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} E[g(y \mid X_i)]$	$\frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N} E^2[g(y \mid X_i)] - \hat{I}^2\right]$	• $E[g(y \mid X_i)]$ known • Can be extended to multistage

TABLE A-1 (contd)

Simulation Technique	Parameter(s) to be be Estimated	Random Sample	Estimator, I	Estimator S^2 For σ^2	Comments
Statistical Estimation	$I = \iint_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy$	$X_1, \ldots, X_N$ and $Y_1, \ldots, Y_N$ from $f(x,y)$	$\hat{I} = \frac{1}{N} \Sigma [g(y\|X_i)]$	$\frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N} E^2[g(y\|X_i)] - \hat{I}_i^2\right]$	● Useful when process is repetitive and values of y, which were selected in the random sample but not used in the estimator, will be needed in generation of next value of x.
Correlated Sampling	$\Delta = \int_{-\infty}^{+\infty} g_1(x)f_1(x)dx$ $- \int_{-\infty}^{+\infty} g_2(y)f_2(y)dy$	$R_1, \ldots, R_N$ from U(0, 1) X_i and Y_i from $R_i = \int_{-\infty}^{X_i} f_1(x)dx$ $R_i = \int_{-\infty}^{Y_i} f_2(y)dy$	$\hat{\Delta} = \frac{1}{N}\sum_{i=1}^{N}[g_1(X_i)$ $- g_1(Y_i)]$	$\frac{N}{N-1}\left\{\frac{1}{N}\sum_{i=1}^{N}[g_1(X_i)-g_2(Y_i)]^2 - I^2\right\}$	● Many variations in correlation techniques are available.
History Reanalysis	$I_1 = \int_{-\infty}^{+\infty} g_1(x)f_1(x)dx$ $I_2 = \int_{-\infty}^{+\infty} g_2(x)f_2(x)dx$	$X_1, \ldots, X$ from $f_1(x)$	$\hat{I}_1 = \frac{1}{N}\sum_{i=1}^{N} g_1(X_i)$ $\hat{I}_2 = \frac{1}{N}\sum_{i=1}^{N}\frac{g_2(X_i)f_2(X_i)}{f_1(X_i)}$	$S_1^2 = \frac{N}{N-1}\left\{\frac{1}{N}\sum_{i=1}^{N} g_1^2(X_i) - \hat{I}_1^2\right\}$ $S_2^2 = \frac{N}{N-1}\left\{\frac{1}{N}\left[\sum_{i=1}^{N}\frac{g_2(X_i)f_i(X_i)}{f_1(X_i)}\right]^2 - \hat{I}_2^2\right\}$	● History to estimate I_1 is available ● May be applied to several I_2 ● Similar to importance sampling and correlation
Control Variates	$I = \int_{-\infty}^{+\infty} g(x)f(x)dx$ $= \theta + \int_{-\infty}^{+\infty}[g(x)-h(x)]f(x)dx$ $= \theta + I_1$	$X_1, \ldots, X_N$ from $f(x)$	$\hat{I}_1 = \frac{1}{N}\sum_{i=1}^{N}[g(X_i)-h(X_i)]$	$\frac{N}{N-1}\left\{\frac{1}{N}\sum_{i=1}^{N}[g(X_i)-h(X_i)]^2 - \hat{I}_1^2\right\}$	● $\theta = \int_{-\infty}^{+\infty} h(x)f(x)dx$ ● h(x) = approximation to g(x) with known expected value ● Estimate for I is $\int_{-\infty}^{+\infty} g(x)f(x)dx = \theta + \hat{I}$

TABLE A-1 (contd)

Simulation Technique	Parameter(s) to be Estimated	Random Sample	Estimator, I	Estimator S^2 For σ^2	Comments
Antithetic Variates	$\int_{-\infty}^{+\infty} g(x)f(x)dx$	R_i ; $i = 1,\ldots,N$ from $U(0,1)$. Determine X_i and X_i' from $\alpha R_i = \int_{-\infty}^{X_i} f(x)dx$ and $1-\alpha R_i = \int_{-\infty}^{X_i} f(x)dx$	$\frac{1}{N} \sum_{i=1}^{N} [\alpha g(X_i)+(1-\alpha)g(X_i)]$	$\frac{N}{N-1}\left\{ \frac{1}{N} \sum_{i=1}^{N} [\alpha g(X_i)+(1-\alpha)g(X_i)]^2 - \hat{I}^2 \right\}$	• $0 < \alpha < 1$ • Select α as that value that minimizes $\hat{S}^2$
Regression	$\vec{I} = \begin{pmatrix} I_1 \\ \vdots \\ I_p \end{pmatrix}$ (a vector of integrals)	θ_{ij} ; $i = 1,\ldots,N$ $j = 1,\ldots,n$	$(\vec{A}^T \vec{V}^{-1} \vec{A})^{-1} \vec{A}^T \vec{V}^{-1} \theta$ $\vec{I} = \vec{A} \cdot E[\vec{\hat{\theta}}]$ where $\vec{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_n \end{pmatrix}$; $\vec{A} = \begin{pmatrix} a_{11} \cdots a_{1n} \\ \vdots \quad\quad \vdots \\ a_{p1} \cdots a_{pn} \end{pmatrix}$ ($\vec{A}^T$ is the transpose of $\vec{A}$.) $\hat{\theta}_j = \frac{1}{N} \sum_{i=1}^{N} \theta_{ij}$. $\hat{\vec{V}} = \begin{pmatrix} \hat{v}_{11} \cdots \hat{v}_{1n} \\ \hat{v}_{n1} \cdots \hat{v}_{nn} \end{pmatrix}$ where $v_{ij} = \sum_{k=1}^{N} (\theta_{ki} - \hat{\theta}_i)(\theta_{kj} - \hat{\theta}_j)$	Use Batching	

APPENDIX B

REFERENCES AND ABSTRACTED BIBLIOGRAPHY

1. Bracken, J., McCormick, G.P., "Selected Applications of Non-Linear Programming," John Wiley & Sons, New York, 1968.

A book that presents several selected optimization problems. Of particular interest here is the application of optimization methods to selection of optimal strata for sampling in the sense of minimum variance.

2. Burt, J.M. and M.B. Garman, "Conditional Monte Carlo: A Simulation Technique for Stochastic Network Analysis," Management Science, 18, No. 3, 207-217, Nov. 1971.

This paper is concerned with a simulation procedure for estimating the distribution functions of the time to complete stochastic networks. The procedure, called conditional Monte Carlo, is shown to be substantially more efficient (in terms of the computational effort required) than traditional simulation methods. The efficiency of conditional Monte Carlo and its use in conjunction with other Monte Carlo methods is illustrated for the Wheatstone bridge network. The applicability of the procedure to larger networks, as well as other stochastic systems, is discussed, and a general method is given for its implementation.

3. Clark, C.E., "Importance Sampling in Monte Carlo Analyses," Operations Research, 603-620, Sept-Oct. 1961.

Some Monte Carlo analyses require hundreds of hours of high speed computer time. Many problems of current interest cannot be handled because the computer time required would be too great. Statistical sampling procedures have been developed that greatly reduce the required computer time. Importance sampling is one of these. This paper is an elementary description of importance sampling as used in Monte Carlo analyses.

4. Clark, F.H., "The Exponential Transform as an Importance Sampling Device - A Review," Oak Ridge National Laboratory (AEC) ORNL-RSIC-14, 1-50, January 1966.

The exponential transform is reviewed, with emphasis on its use as a guide to effective importance sampling in the solution of the Boltzmann equation by Monte Carlo methods. Contributions of various workers are assembled, along with numerical results. Special consideration is

given to approximate forms and to effective practical methods. Problems related to negative effective cross sections, tracking across discontinuities, directional biasing in inhomogeneous media, and high variance in back-scattered components are specifically treated.

5. Conveyou, R.R., V.R. Cain and K.J. Yost, "Adjoint and Importance in Monte Carlo Application," Nuclear Science and Engineering, 27, 219-234, 1967.

The use of the Monte Carlo method for the study of deep penetration of radiation into and through shields entails the use of sophisticated methods of variance reduction to make such calculations economical or even feasible. This paper presents an exposition of the most useful methods of variance reduction. The exposition is unified by consistent exploitation of adjoint formulations to estimate expected values, as in previous work, and further to evaluate the variance of the resulting estimates.

The connection between adjoint formulations and the choice of biasing schemes is also investigated. In particular, it is shown that the value function (the solution of the integral equation of the adjoint formulation) is always a good choice for importance function biasing; a sharp upper bound, independent of the particular problem is found for the resulting variance. Predicted (analytic) and experimental (Monte Carlo) results are also given for a simple one-dimensional problem.

6. DeGrott, M.H. and N. Starr, "Optimal Two-Stage Stratified Sampling," The Annals of Math. Statistics, 40, No. 2, 575-582, 1969.

This paper develops effective approximations to the optimal sampling for situations where the total number of available observations is large, and, therefore the optimal number of observations that should be obtained at the first stage will also be large in a two strata population where the sampling is accomplished in two stages. The techniques can be extended to multistrata problems provided the observations at each strata have a normal distribution.

7. Ehrenfeld, S. and S. Ben-Tuvia, "The Efficiency of Statistical Simulation Procedures," Technometrics, 4, No. 2, 257-275, May, 1962.

Various methods for improving the efficiency of statistical simulation of complex systems are described and illustrated for simple queueing situations. The paper proposes that the efficiency and effectiveness of statistical simulations can be increased through the adaptation of experimental design principles which exploit any qualitative knowledge surrounding the problem under study. Some techniques explored are stratified sampling, sequential sampling, importance sampling and the use of concomitant information.

8. Evans, D.H., "Applied Multiplex Sampling," Technometrics, Vol. 5, No. 3, August 1963.

Multiplex sampling is an extension of standard Monte Carlo methods for estimating characteristics of the distribution of a response when the response is a function of several variables, each of which comes from a known distribution. The extension is required when each variable is available in a variety of distributions. Depending on the number of variables there are many possible different combinations each of which, in general, will give a different distribution to the response. If characteristics of the response are to be estimated for many or all of these combinations, there will be a plethora of Monte Carlos to be performed if usual procedures are followed. This in turn can require a great number of observations of the response; if these are difficult to obtain, e.g., if they must be determined experimentally, the carrying out of such a program can easily prove impracticable. Multiplex sampling is a method for estimating the characteristics for all the different distributions for the response by using a relatively small number of observations. This is accomplished by sampling from an efficient sample space and then using weighted sampling formulas. The functional form for the probability density function describing this sample space is derived in a companion paper; here we assume this form and consider the practical aspects.

9. Fishman, G.S., "The Allocation of Computer Time in Comparing Simulation Experiments," Operations Research, 16, 280-295, March-April, 1968.

This paper investigates the problem of efficiently allocating computer time between two simulation experiments when the objective is to make a statistical comparison of means. For a given level of accuracy the results show that significantly less computer time is required when the sample sizes are determined according to a certain rule than when the sample sizes are equal. A graphical analysis suggests that small errors in estimating the population parameters of the allocation rule do not significantly affect the efficient allocation of time. The influence that the degree of autocorrelation has on the time allocation is also investigated; results show that small differences in the autocorrelation functions are important when each process is highly autocorrelated. Positively correlated samples for the two experiments are examined and incorporated into the efficient allocation rule. It is shown that their use leads to a saving in computer time. A two-stage procedure is described wherein initial estimates of the population parameters are computed which permit the experimenterto estimate how many more observations to collect on each experiment. The procedure is simple and straightforward to implement and should be of practical value. When the computer time requirements turn out to be prohibitive, we suggest using negatively correlated replications on each experiment. This may be accomplished by using antithetic variates. The two-stage procedure also applies in this case.

10. Garman, M.B., "More on Conditioned Sampling in the Simulation of Stochastic Networks," Management Science, Vol. 17, No. 1, September 1972.

The technique of conditioned sampling has been shown to improve simulation efficiency in the estimation of stochastic activity network duration. This paper describes a method for generalizing the conditioned sampling approach from its current use of product-form estimators to the use of product/convolution-form estimators. Estimators of the latter type are constructed and demonstrated to require fewer samples per realization (hence increased estimation accuracy) in almost all networks. An algorithm for estimator construction is presented and proven to apply to any given activity network. It is also shown that the derived product-convolution-form estimators may require a precedence structure within the sampling sequence which creates their corresponding realizations.

11. Gaver, D.P. Jr., "Statistical Methods for Improving Simulation Efficiency," Carnegie-Mellon Universtiy, Pittsburgh, Pa., August 1969. AD694445

The paper presents a variety of statistical devices for improving the effectiveness of computer simulations of random processes. The methods are illustrated by examples from a queueing problem that is inadequately treated by analytical approaches.

12. Goertzel., G. and M.H. Kalos, "Monte Carlo Methods in Transport Problems," Progress in Nuclear Science, Series I, Volume II, Pergamon Press, p. 315-369.

The article is devoted to the discussion of the applications of the Monte Carlo method in the field of nuclear energy. An account of the theory is given, including preliminary material on random and pseudorandom numbers and on choosing from probability distributions. The target

game and the transport game are described in detail, with the emphasis put on generality. The final section deals with specific applications to some shielding and reactor core calculations.

13. Hague, J. F., "Variance Reduction in the Monte Carlo Method for Determining the Volume of Multidimensional Non Analytic Solids," Nuclear Instruments and Methods, 47, 194-200, 1967.

A Monte Carlo method for finding the volume of any definable object located within a unit cube is considered. The method, which does not require the surface of the solid to be described by an explicit function, is developed into suitable program form and is tested for "cylinders, spheres and pyramids in 2, 4 and 6 dimensions. Variance reduction factors, over straightforward Monte Carlo, of up to 30 for a 6-dimensional "cylinder," and 3 for a 6-dimensional "pyramid" are obtained. An example is given of the application of the method to high energy particle physics.

14. Hammersley, J.M. and D.C. Handscomb, Monte Carlo Methods, Methuen & Co. Ltd. London, 1964.

One of the most useful references available today on Monte Carlo, it presents the general Monte Carlo concepts and methods, techniques, for generation of random numbers and applications to problems in solution of linear equations, reactor shielding, statistical mechanics flow in random media (percolation processes) and multivariable systems.

15. Hartley, H. O.. and J. Rao, "Variance Estimation in Linear Models Applied to Stratification Problems," Biometrics, 23, 380, 1967.

It is well known that for sampling from finite populations with numerous strata the allocation of one unit per stratum often results in highest efficiency. On the other hand, it will not in general be possible to obtain unbiased estimates of the variance of the stratified estimator. Various solutions (including the so-called collapsing of strata into pairs) have been tried but most of these are afflicted by an unknown bias. The present approach uses a linear model which will usually result in a considerable reduction of the bias in variance estimation. The problem is reduced to the following general problem in variance estimation for linear models. Given a familiar linear model $y = X \quad + e$, where the residual vector e consists of n independent elements with mean vector 0 and the variance vector ^{2}. If X is assumed to represent an x by k matrix the total number of unknown parameters is k + n and these are clearly not estimable. However, if at least k linear restrictions are assumed to hold between the elements of E^2 the problem becomes estimable. For specific linear restrictions unbiased estimators are derived. Specifically, the application to the above stratification problem is discussed.

16. Householder, A.S. (Ed), "Monte Carlo Methods," National Bureau of Standards Applied Mathematics Series 12, June 1951.

Proceedings of a Symposium held June 29, 30 and July 1, 1949 on Monte Carlo Methods. Papers included several Monte Carlo applications and random number generation.

17. Irving, D.C., "The Adjoint Boltzmann Equation and Its Simulation by Monte Carlo," ORNL-TM-2879, May 18, 1970, Nucl. Eng. Des, 15(3), 273-293.

The Boltzmann equation for neutron transport is discussed in both integro-differential and integral form. The 'value' or 'importance' equation is derived and shown to be equivalent, in the integral form, to the adjoint of the collision density. However, the value is also equivalent to the adjoint of the flux when the adjoint operation is carried out on the integro-differential equations. Possible ways of simulating both the forward and adjoint equations by Monte Carlo are discussed. Because the value equation is a 'flux-like' equation, direct simulation of it proves to be unwieldy. Instead, a 'collision density' for adjoint particles, equal to the value or adjoint flux times the total cross section, is introduced. The equation for this adjuncton collision density may be simulated by the same routines as were used for the forward calculation and only the cross sections need to be changed. The extension of this to problems involving multiplying media is also included.

18. Kahn, H. and A.W. Marshall, "Methods of Reducing Sample Size in Monte Carlo Computations," Operations Research, 1, 263-278, 1953,

This paper deals with the problem of increasing the efficiency of Monte Carlo calculations. The methods of doing so permit one to reduce the sample size required to produce estimates of a fixed level of accuracy or, alternatively, to increase the accuracy of the estimates for a fixed cost of computation. Few theorems are known with regard to optimal sampling schemes, but several helpful ideas of very general applicability are available for use in desiging Monte Carlo sampling schemes. Three of these ideas are discussed and illustrated in simple cases. These ideas are (1) correlation of samples, (2) importance, and (3) statistical estimation.

19. Kahn, H., "Modification of The Monte Carlo Method," The Rand Corporation Publication.

The theory behind several useful variance reduction methods such as importance sampling, sequential sampling, correlation, Russian Roulette and splitting.

20. Kahn, H., "Applications of Monte Carlo," The Rand Corporation, Santa Monica, Calif. AECU-3259, 1-250, 19 April 1954.

A classic report that provides a comprehensive and detailed survey of random number generation and variance reduction techniques. Several examples pertaining to the area of radiation transport are presented to demonstrate the applicability of variance reduction.

21. Kalos, M.H., "Monte Carlo Integration of the Adjoint Gamma-Ray Transport Equation, Nuclear Science and Engineering; 33, 284-290 (1968).

The adjoint transport problem for gamma radiation is formulated and prescriptions for its Monte Carlo solution are given. Emphasis is put upon requirements for calculation of effects in shielding against fallout and the differential effect of source position. Results are given for two situations: a detector three feet above a uniform infinite source of 1.25-MeV photons, and another detector placed in an open pit with a similar source.

22. Karcher, R.H., "Static Fault Tree Analysis by Monte Carlo With Some Results," Homes & Narver, Inc., Sept. 1967.

In this paper a Monte Carlo method for evaluation of fault trees is presented along with some results. Of particular interest is the application of importance sampling for improvement of the sampling efficiency.

23. Koop, J.C., "Short Communications on Splitting a Systematic Sample for Variance Estimation," The Annals of Math. Statistics 42, No. 3, 1084-1087, 1971.

Variance estimation in systematic sampling by splitting the sample into equal halves can lead to very serious bias. The expression for this bias relative to the true variance is given in terms of intraclass correlation coefficients. The danger of serious bias is still present when successive pairs of units are treated as "independent" replicates; an expression for this relative bias is also given.

24. McGrath, E.J., "Fundamentals for Operations Research," West Coast University, 1970.

A graduate level text book which includes a chapter on Monte Carlo simulation. Variance reduction techniques considered include systematic and stratified sampling, importance sampling and use of control variates.

25. Moshman, J., "The Application of Sequential Estimation to Computer Simulation and Monte Carlo Procedures," J. Assoc. Computing Mach., 5, 343-352, 1968.

This paper considers a number of sequential techniques for estimating the parameters of Gaussian and binomial populations. Some techniques will be exact ones; others will have symptotic validity. In every case it is possible by proper programming, and possibly some preliminary analysis, to have the computer evaluate the sample obtained thus far and determine whether additional samples are required to obtain some specified precision. In some cases, the evaluation is made after each sample unit; in other cases, evaluation takes place at certain intervals.

26. Nagel, P.M., "A Monte Carlo Method to Compute Fault Tree Probabil Probabilities," System Safety Symposium, Seattle, Wash., June 8-9, 1965.

This report presents a discussion of the application of Monte Carlo methods to the fault tree and demonstrates a methodology to reduce a large simulation into a smaller simulation for application of importance sampling. A small fault tree example is analyzed to demonstrate the technique.

27. Nilsson, G., "Optimal Stratification According to the Method of Least Sequences," Skandmavisk. Aklurarietidskuft, 1967, p. 128-136.

A method is presented that selects the optimal set of points of stratification in the sense of minimum variance.

28. Page, E.S., "On Monte Carlo Methods in Congestion Problems: II. Simulation of Queueing Systems," Operations Research, 13, 300-305, March 1965.

In this paper the application of the antithetic variate technique to reduce variance is shown to possess advantages in a simple queueing system and its application to more complex situations is proposed.

29. Pugh, E. L., "Some Examples of Stochastic Distortion, a Monte Carlo Technique," SP-1584, System Development Corp., Santa Monica, Calif. March 6, 1964.

The effects of importance sampling on the variance of a Monte Carlo estimation of tail probabilities is presented for both the exponential and the gamma distributions. Also presented is the effect of the distortion on the required sample size for a desired accuracy-confidence statement.

30. Pugh, E. L., "A Gradient Technique of Adaptive Monte Carlo," SP-1921/000/01, System Development Corp., Santa Monica, Calif. Sept. 8, 1965.

A technique of Monte Carlo estimation is presented which is "adaptive" in the sense that its efficiency increases as the sampling proceeds. It is based on sequentially estimating the gradient of the variance and following the path of steepest descent. The technique is applied to a problem of estimating the survival probability of a repairable machine.

31. Relles, Daniel A., "Variance Reduction Techniques for Monte Carlo Samples from Students Distribution," Technometrics, Vol. 12, No. 3, August 1970.

A Monte Carlo design is presented for estimating the variance and cumulative distribution function of translation and scale invariant statistics based on independent Student random variables. One obvious application is studying estimates of the location parameter from a symmetric, possibly long-tailed distribution. The method itself amounts to suppressing some of the variability in the sampled objects by integrating these objects over appropriate regions of the underlying probability space. Indications are that, in cases of interest, the variability is thereby considerably reduced, as is illustrated in an application concerning trimmed and Winsorized means.

32. Sarndal, C. E., "The Use of Stratification Variables in Estimation by Proportional Stratified Sampling," Amer. Statistical Assoc. J., 63, 1310-1320, 1968.

This paper deals with proportional stratified sampling in the situation where the estimation variable X is difficult and expensive to observe, while the possible erroneous stratification variable Y is easy and inexpensive to get at. The usually biased estimate

$$I_a = \sum_{i=1}^{k} P_i y_i$$

is compared with the unbiased estimate

$$I_b = \sum_{i=1}^{k} P_i x_i ,$$

where the P_i are stratum weights and y_i and x_i are means of the units sampled from the i:th stratum. The two estimates are similar in that they utilize information from only those population units that make up the sample. While I_a is the more inexpensive estimate, I_b is usually preferable if one judges by the size of the mean square error, which, among other things, depends on the number of strata and the location of the stratum boundaries. In particular, the properties of I_a and I_b are discussed in connection with an example involving the bivariate normal distribution.

33. Serfling, R. J., "Approximately Optimal Stratification," Amer. Statistical Assoc. J., 63, 1298-1309, 1968.

The cum f method of Dalenius and Hodges for approximately optimal construction of strata is utilized to approximate the variance of the stratified estimate, for estimation of the population mean of a random variable Y by the technique of stratified random sampling. The approximation provides a basis for choosing optimally, for fixed cost, the number of strata to be constructed and the total sample size to be used. It also facilitates other purposes, such as the comparison of optimal stratification with optimal simple random sampling. The study is carried out for the situations of stratification on the estimation variable and of stratification on a covariable closely associated with the estimation variable.

34. Shreider, Yu. A., "The Monte Carlo Method," Pergamon Press, 1966.

A general Monte Carlo reference that addresses the general principles, application of simulation to evaluation of definite integrals, neutron physics, servicing processes, communications theory and generation of random variables. A limited amount of material is presented on the formal aspects of variance reduction.

35. Spanier, J., "An Analytic Approach to Variance Reduction," SIAM J. Appl. Math., 18, No. 1, 172-190, January 1970.

This paper presents a study of the variance of the weight of particles actually transmitted through slabs of various dimensions. Similar

techniques may be used to study estimates of transmission which are collision types, i.e., estimators which record, on every collision, the probability of direct transmission on the next flight.

36. Spanier, J. and E.M. Gelbard, "Monte Carlo Principles and Neutron Transport Problems," 1-234, Addison-Wesley Publishing Co., Reading, Mass. 1969.

A comprehensive reference presenting fundamentals of Monte Carlo, discrete and continuous random walk processes, standard variance reduction techniques and several applications to radiation transport problems.

37. Van Slyke, R.M., "Monte Carlo Methods and the Pert Problem," Operations Research, 11, 839-860, 1963.

In this paper the results of a Monte Carlo simulation of PERT networks are given. First the concept of using Monte Carlo methods to give solutions to PERT problems under less restrictive assumptions is discussed. Results are given for the accuracy obtainable, for the computer time required and devices for reducing computational effort. Finally, a "criticality" index is defined for each activity. This index is simply the probability that the activity will be on the critical path. The ramifications and uses of this parameter, which are not available using current techniques, are developed.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

Science Applications, Inc.
P. O. Box 2351, 1200 Prospect Street
La Jolla, California 92037

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

Techniques for Efficient Monte Carlo Simulations
Volume III
Variance Reduction

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*

Elgie J. McGrath, David C. Irving

6. REPORT DATE March 1973	7a. TOTAL NO. OF PAGES 153	7b. NO. OF REFS 37

8a. CONTRACT OR GRANT NO.

N00014-72-C-0293

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

SAI-72-590-LJ

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Office of Naval Research (Code 462)
Department of the Navy
Arlington, Virginia 22217

13. ABSTRACT

Many Monte Carlo simulation problems lend themselves readily to the application of variance reduction techniques. These techniques can result in great improvements in simulation efficiency. This document describes the basic concepts of variance reduction (Part I), and a methodology for application of variance reduction techniques is presented in Part II. Appendices include the basic analytical expressions for application of variance reduction schemes as well as an abstracted bibliography.

The techniques considered here include importance sampling, Russian roulette and splitting, systematic sampling, stratified sampling, expected values, statistical estimation, correlated sampling, history reanalysis, control variates, antithetic variates, regression, sequential sampling, adjoint formulation, transformations, orthonormal and conditional Monte Carlo. Emphasis has been placed on presentation of the material for application by the general user. This has been accomplished by presenting a step by step procedure for selection and application of the appropriate technique(s) for a given problem.

ORNL-RSIC-38
Vol. III

INTERNAL DISTRIBUTION

1. L. S. Abbott
2. R. G. Alsmiller, Jr.
3. C. E. Clifford
4. R. R. Coveyou
5. F. C. Maienschein
6. F. R. Mynatt
7. R. W. Peelle
8. F. G. Perey
9. H. Postma
10. M. W. Rosenthal
11. D. Steiner
12. D. B. Trauger
13. D. K. Trubey
14. G. E. Whitesides
15. A. Zucker
16. H. Feshbach (Consultant)
17. P. F. Fox (Consultant)
18. W. W. Havens, Jr. (Consultant)
19. A. F. Henry (Consultant)

20-320. RSIC
321-322. Central Research Library
323. ORNL — Y-12 Technical Library
Document Reference Section
324. Laboratory Records Department
325. Laboratory Records, ORNL, R.C.
326. ORNL Patent Office

EXTERNAL DISTRIBUTION

327. P. B. Hemmig, Division of Reactor Research and Development, ERDA, Washington, D.C. 20545
328. D. C. Irving, Science Applications, Inc., Box 2351, La Jolla, California 92037
329. E. J. McGrath, Science Applications, Inc., Box 2351, La Jolla, California 92037
330. Capt. R. G. Powell, Defense Nuclear Agency, Washington, D.C. 20305
331. L. K. Price, Division of Controlled Thermonuclear Research, ERDA, Washington, D.C. 20545
332. Burt Zolatar, Electric Power Research Insitute, 3412 Hillview Avenue, Palo Alto, California 94304
333. Directorate of Licensing, NRC, Washington, D.C.
334. Directorate of Regulatory Standards, NRC, Washington, D.C. Attn: Director

335-336. Technical Information Center, ERDA, Oak Ridge, Tennessee
337. Research and Technical Support Division, ORO, Oak Ridge, Tennessee
338. Reactor Division, ORO, Oak Ridge, Tennessee

☆U. S. GOVERNMENT PRINTING OFFICE: 1974-748-189/182