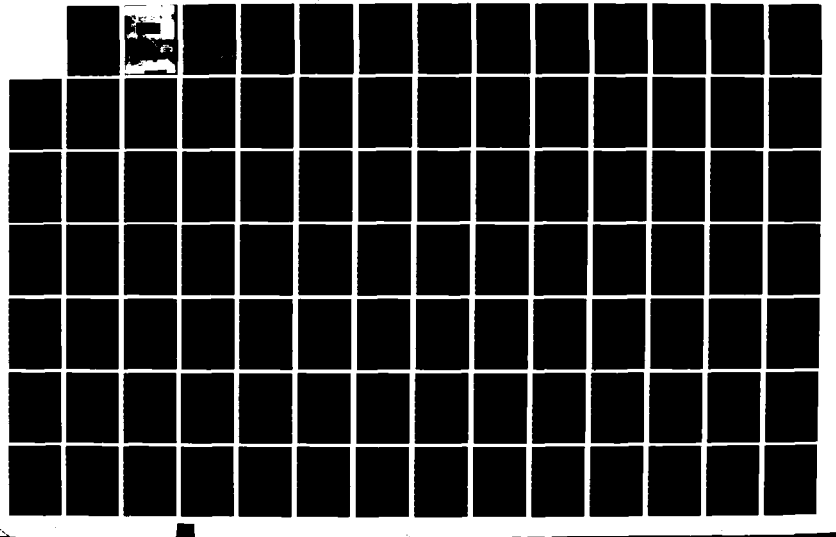
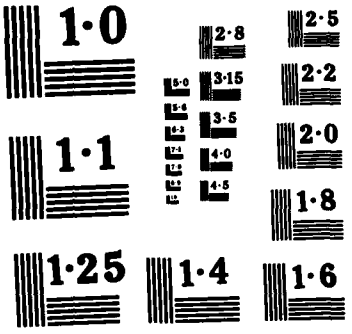
AD-A158 146 VARIANCE REDUCTION IN SIMULATION EXPERIMENTS: A MATHEMATICAL-STATISTICAL FRAMEWORK(U) PURDUE UNIV LAFAYETTE IN SCHOOL OF INDUSTRIAL ENGINEERING 1/2

UNCLASSIFIED B L NELSON DEC 83 N00014-79-C-0832 F/G 12/1 NL





NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

Variance Reduction in Simulation Experiments: A Mathematical-Statistical Framework

Barry L. Nelson

December 1983

2

Variance Reduction in Simulation Experiments: A Mathematical-Statistical Framework

Barry L. Nelson

December 1983

Research funded under contract N00014-79-C-0832 for the Office of Naval Research. Principal investigator: Bruce W. Schmeiser.

To Sharon and LeRoy Nelson,
and Mary Smith

ACKNOWLEDGMENTS

I would like to thank my major professor Dr. Bruce W. Schmeiser for his ideas, advice, enthusiasm, and appreciation of "truth and beauty." I also thank committee members Drs. Louis J. Cote, Stephen D. Roberts, and James J. Solberg for permitting me to attempt an unusual and somewhat risky topic and being concerned enough to make sure I did not get into too much trouble. To my wife Jeanne, for allowing me to spend our first year of marriage in the office, I give special thanks and condolences.

I would like to acknowledge the support provided by a fellowship from the General Electric Corporation, a David Ross Grant from the Purdue Research Foundation, a teaching assistantship from the School of Industrial Engineering, a grant from the Office of Naval Research (N00014-79-C-0832), and an occasional dose of reality from Jack W. Posey, who always kept everything in perspective. Finally, I would like to acknowledge myself for typing Mr. Dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS

$E[X]$, $Var[X]$ - expected value, variance of X

$Cov[X,Y]$ - covariance of X and Y

$\bar{}$ - denotes sample mean

σ^2 - variance

S - statistical space

Ω, Θ, Γ - sample spaces

$\mathfrak{X}$, $\mathfrak{Y}$, $\mathfrak{Z}$ - event sets

A, B - sigma algebras

X, Y, Z - input, output, and statistic random variables

$[X|Y]$ - concatenation of X and Y

θ - vector of parameters of interest

$E(f_\omega,(g;R_\Psi),h;\Omega,\theta)$ - simulation experiment

$E^S(\Omega,\theta)$ - experiment set

P_θ , $\theta \in \Theta$ - family of probability measures

$I(\theta)$ - measure of information for estimation of θ

$l(Z,\theta)$ - loss function for estimation of θ by Z

$L(\theta,x)$ - likelihood function of θ given X

f - probability distribution

$r(x)$ - ratio of conditional density functions

g, h - transformations of random variables

T - transformation of simulation experiment

$*$ - denotes "fixed" when attached as subscript or superscript

R - feasible region

R_* - sampling plan

$I = [I_1, \ldots, I_{\ell_*}]$ - lengths of output sequences

n - number of observations

$\underline{0}$ - zero vector

$U(0,1)$ - uniform distribution on $(0,1)$

$\overset{c}{=}$ $\overset{e}{=}$ $\overset{d}{=}$ $\overset{s}{=}$ - equivalence relations

e - measure of effort

α, γ - constants

i, j, k, ℓ, m - indices

p, p_j - probabilities

L_j - strata

δ_c - indicator function on condition c

t, Δt - time, time increment

U, V, W - sets of random variables

$\{a_i\}$, $\{b_i\}$, $\{c_i\}$ - arbitrary sequences

ABSTRACT

Nelson, Barry Lee. Ph.D., Purdue University, December 1983. Variance Reduction in Simulation Experiments: A Mathematical - Statistical Framework. Major Professor: Dr. Bruce W. Schmeiser.

With the expanding use of computer simulation to model and solve industrial engineering problems, there has been increasing interest in the development of efficient simulation techniques. When the concern is for statistical efficiency of results that are random variables, such approaches are usually called variance reduction techniques (VRTs).

Many of the fundamental ideas in simulation, and particularly techniques for efficient simulation, had their origins in the Monte Carlo estimation literature. The theory of sampling is another closely related field that predates the development of simulation. Although there has been significant research interest in variance reduction, there have been few attempts to structure and define the discipline.

VRTs are transformations. They transform simulation experiments into related experiments that yield better estimates of some parameters of interest, where better usually means more precise. This research identifies and defines the components from which all variance reduction techniques are built. Given a general mathematical-statistical definition of simulation experiments, these components or classes of transformations are shown to be useful, to be mutually exclusive, and to

generate all possible VRTs via composition. Benefits of the research include: 1) the facility to unambiguously define new or existing VRTs, eliminating confusion that currently exists in the literature, 2) the facility to decompose VRTs into combinations of transformations, making the relationships between VRTs clear, 3) the development of a theoretical foundation for analytical treatment of VRTs, and 4) the development of a setting for proposing new VRTs and research questions. In addition, increased understanding of the area should promote more and better application of variance reduction in practice.

INTRODUCTION

With the expanding use of computer simulation to model and solve industrial engineering problems, there has been increasing interest in the development of efficient simulation techniques. By efficient techniques is meant approaches that produce accurate answers with reasonable computing cost and analyst effort. When the concern is for statistical efficiency of results that are random variables, such approaches are usually called variance reduction techniques (VRTs).

Definitions

In this research, simulation will refer to digital computer models of stochastic systems. Often these models are characterized by explicit accounting of the passage of time, although this is not a requirement. "By simulation is meant the technique of setting up a stochastic model of a real situation, and then performing sampling experiments upon the model." (Harling, 1958) The experiment is done to obtain performance measures for the system, but since the models are driven by stochastic inputs, the measures are only estimates of the true performance of the system.

Many of the fundamental ideas in simulation, and particularly techniques for efficient simulation, had their origins in the Monte

events.'" (Carter and Ignall, 1975, p. 608)

"The idea [of virtual measures for a particular inventory problem] is surely suggestive of the splitting and enrichment techniques." (Carter and Ignall, 1975, p. 614)

"In the slab problem considered above, the process of splitting, accompanied by Russian Roulette, may be thought of as an example of importance sampling where the transport kernel is modified." (Carter and Cashwell, 1975, p. 17)

"Latin Hypercube Sampling is a variant of stratification especially appropriate for multivariate problems with restricted sampling budgets." (Swain, 1981, p. 40)

"It [Latin Hypercube Sampling] is an extension of quota sampling, and it is a first cousin to the 'random balance' design...and to the highly fractionalized factorial designs... and to lattice sampling." (McKay, Beckman and Conover, 1979, p. 243)

"It is interesting to note that the exponential transform is a form of quota sampling." (Kahn, 1950b, p. 62)

"Another instance of control variate sampling is the use of the same sequence of random numbers in two different phases of the simulation. The idea is to introduce a positive correlation for quantities that are to be subtracted." (Gentle, 1975, p. 7)

"The idea behind stratified sampling is essentially the same as that of importance sampling." (Gentle, 1975, p. 7)

"This last technique [stratified sampling] is sort of a combination of Importance Sampling and Systematic Sampling." (Kahn, 1956, p. 155)

"Indeed the family of [antithetic variate] sampling plans...is the family of systematic sampling plans." (Roach and Wright, 1974, p. 8)

"Systematic sampling plans form a computationally feasible subset of the family of antithetic sampling plans originally described by Hammersley, Handscomb, and Mauldin." (Roach and Wright, 1974, p. 32)

"The method of stratified sampling is closely related to that of using 'control variables.'" (Hartley, 1977, p. 23)

"It is possible to view virtual measures as, in the words of a referee, 'a re-packaging of conditional Monte Carlo for the estimation of rare

Biased Estimators (Rubinstein, 1981)
Weighted Uniform Sampling (Powell and Swann, 1966)

Random Quadrature Method (Rubinstein, 1981)
Use of Orthonormal Functions (Hammersley and Handscomb, 1964)

Many researchers have conjectured that relationships exist between various VRTs. Most of these conjectures are in fact true, but can seem contradictory without a unifying theory that makes relationships and differences apparent. Below are several examples of statements that have appeared in the variance reduction literature.

"The antithetic variate technique is a particular case of this situation (regression method)" (Hammersley and Handscomb, 1964, p. 66)

"...so the antithetic variate method is equivalent to using the control variate t'=1/2t - 1/2t'', whose expectation is zero." (Hammersley and Morton, 1956, p. 449)

Control variates can be "a special form of the use of the same [common] random numbers..." (Kleijnen, 1974, p. 205)

Common random numbers is a special case of control variates (paraphrased). (Gentle, 1975, p. 8)

Selective Sampling (Brenner, 1963)
Fixed Sequence Principle (Ehrenfeld and Ben-Tuvia, 1962)

Sequential Sampling (McGrath and Irving, 1973a)

Poststratified Sampling (Wilson, 1979b)
Stratification after Sampling (Kleijnen, 1974)

Importance Sampling
Partition of Region (Rubinstein, 1981)
Correction Sampling (Hartley, 1977)
Multi-Stage Sampling (Marshall, 1956)
Method of the Essential Sample (Kohlas, 1982)
Sampling with Probability Proportional to Size (Moy, 1965)

Transformations (McGrath and Irving, 1973a)

Expected Values (McGrath and Irving, 1973a)
Conditional Expectations (Law and Kelton, 1982)
Conditioned Sampling (Garman, 1972)
Statistical Estimation (McGrath and Irving, 1973a)
Virtual Measures (Carter and Ignall, 1975)
Prior Information (Pritsker and Pegden, 1979)
Reducing the Dimensionality (Rubinstein, 1981)
Strict Conditional Monte Carlo (Fox, 1983)
Extended Conditional Monte Carlo (Fox, 1983)

Indirect Estimation (Law and Kelton, 1982)

Adjoint Formulations (McGrath and Irving, 1973a)
Mathematical Analog (Kahn, 1950)

Conditional Monte Carlo (Hammersley and Handscomb, 1964)
Conditional Sampling (Hartley, 1977)
Classic Conditional Monte Carlo (Fox, 1983)

Correlation Induction Strategies (Schruben, 1979)

Control Variables
Control Variates
Control Variate Sampling (Swain, 1981)
Concomitant Control Variables (Lavenberg and Welch, 1981)
Internal Controls (Iglehart, 1979)
Concomitant Information (Ehrenfeld and Ben-Tuvia, 1962)
Regression Sampling (Kleijnen, 1974)
Extraction of the Regular Part (Shreider, 1966)
Comparison Method (Kohlas, 1982)

Regression Methods (Hammersley and Handscomb, 1964)
Regression on Concomitant Variables (Gentle, 1975)

Common Random Numbers
Correlated Sampling (Law and Kelton, 1982)
Using the same random numbers
Correlation of Samples (Kahn and Marshall, 1953)

History Reanalysis (McGrath and Irving, 1973a)

Systematic Sampling
Simple Stratified Sampling
Dagger Sampling (Kumamoto, Tanaka and Inoue, 1980a)
Sequential Destruction Method (Easton and Wong, 1980)
Systematic Source Sampling (Carter and Cashwell, 1975)

Quasi-random Numbers (Hammersley and Handscomb, 1964)
Latin Hypercube Sampling (McKay, Beckman and Conover, 1979)

Stratified Sampling
Stratification of Random Numbers
Quota Sampling (Kahn, 1954)
Adaptive Stratified Sampling
Critical Value Stratified Sampling (Surkis, Gordon and Hynes, 1975)
Representative Sampling (Delanius, 1950)
Bowley-sampling (Delanius, 1950)
Neyman-sampling (Delanius, 1950)
Proportional Sampling (Ehrenfeld and Ben-Tuvia, 1962)
Group Sampling (Shreider, 1966)

any knowledge, either known with certainty or suspected, beyond what is needed to draw samples from the experiment. A transformation is a modification of a problem situation so that a variance reduction might be achieved. Combined with the necessary prior knowledge, transformations produce VRTs. In a real application, VRTs are implemented as algorithms. The theoretical framework developed here defines six classes of transformations and shows how they are composed into VRTs. The format used to define VRTs should be illustrative for developing algorithms.

Existing Variance Reduction Techniques

The number of VRTs and their variations is staggering. The following is a list of VRTs found in the simulation, Monte Carlo, and sampling literature. Multiple names for the same or very similar techniques are grouped together, and references are given for names not in common use.

Antithetic Variates
Antithetic Sampling
Antithetic Transformation (Halton, 1979)
Antithetic Variate Sampling Plans (Roach and Wright, 1977)
Supplemental Variables (Mize and Cox, 1968)
Antithetic Control Variables (Cheng, 1981)
Complementary Random Numbers (Hiller and Lieberman, 1974)
Complementary Antithetic Variates (George, 1977)
Correlation Selection (Ermakov and Zolotukhin, 1960)
Use of Dependent Variables (Shreider, 1966)
Symmetrization of the Integral (Shreider, 1966)
Basic Antithetic Variates (Roach and Wright, 1977)
Antithetical Variables (Kohlas, 1982)
Compensation Methods (Kohlas, 1982)
Randomization Sampling (Deutsch and Schmeiser, 1980)

it is possible (conceptually) to represent simulation experiments similarly by thinking of the simulation as some input distributions and output transformations from which one can sample. It may not be possible to write an explicit expression for the integral in all cases.

The terms crude Monte Carlo or crude sampling are used to describe the following technique for estimating (2.1)

0. Formulate (2.1) as in (2.2)
1. Sample n values of X, $(X_1, X_2, \ldots, X_n)$
2. Estimate θ by

$$Z = \frac{1}{n} \sum_{i=1}^{n} g_2(X_i)$$

Z is an unbiased estimator of θ with variance

$$\mathrm{Var}[Z] = \frac{\sigma^2}{n}$$

where

$$\sigma^2 = E[g_2(X)-\theta]^2$$

Clearly the variance of Z can be reduced by increasing n.

VRTs usually attempt to attain an estimator with smaller variance for n observations, or the same variance with fewer observations. It is generally agreed that prior knowledge is required to achieve a variance reduction. For the purposes of this research prior knowledge will mean

LITERATURE REVIEW

This chapter reviews the names of and relationships between VRTs as well as several key survey papers. The intention is not to explain individual VRTs in detail here; see Chapter 5.

Definitions

Simulation, Monte Carlo estimation, and sampling are defined as in Chapter 1. Recall that the value of any integral can be expressed as the expected value of a random variable. For example, consider the simple scalar integral

$$\theta = \int_A g_1(x)dx \tag{2.1}$$

where $g_1(x)$ is a real valued function of elements in A, a subset of the real line. Now if $f(x)$ is a probability density function on A, and $f(x) = 0$ only if $g_1(x) = 0$, then

$$\theta = \int_A \frac{g_1(x)}{f(x)} f(x)dx = \int_A g_2(x)f(x)dx \tag{2.2}$$

and $E[g_2(X)] = \theta$, where X is a random variable with density function $f(x)$. Monte Carlo estimation problems are formulated as integrals, and

concerning the definitions and relationships between VRTs is presented. Chapter 3 develops a general mathematical-statistical definition of simulation experiments, which is necessary to define the classes of transformations and establish their properties (Chapter 4). Chapter 5 reviews five well-known VRTs in light of the results just presented. The final chapter contains concluding comments and directions for future research.

The current research identifies and defines the components from which all variance reduction techniques are built. Given a general mathematical-statistical definition of simulation experiments, these components or classes of transformations are shown to be useful, to be mutually exclusive, and to generate all possible VRTs via composition. The scope of the results is not limited to simulation since Monte Carlo and sampling theory problems are special cases of the general simulation experiment.

Benefits of the research include: 1) the facility to unambiguously define new or existing VRTs, eliminating the confusion that currently exists in the literature, 2) the facility to decompose VRTs into combinations of transformations, making the relationships between VRTs clear, 3) the development of a theoretical foundation for analytical treatment of VRTs, and 4) the development of a setting for proposing new VRTs and research questions. In addition, increased understanding of the area should promote more and better application of variance reduction in practice.

Additional results of the research are a graphical scheme for describing VRTs (which is applied to five of the most common techniques), and an extensive bibliography of variance reduction literature.

Organization of the Dissertation

Chapter 2 of the dissertation contains a literature review emphasizing previous attempts to develop a unifying framework. In addition, a brief examination of the confusion that presently exists

the sum of the variance and the square of the bias. However, other measures could be proposed. For the purposes of this research, the term variance reduction will mean the following more general goal:

$$\text{minimize } E_Z[l(Z, \theta)]$$

where

$$l(\theta, \theta) = 0$$

and

$$l(Z, \theta) > l(Z', \theta) \quad \text{iff } |Z - \theta| > |Z' - \theta|$$

where $|\ |$ is a metric. This general loss function includes both variance and MSE, as well as others. The particular loss function is application dependent.

Research

Although there has been significant research interest in variance reduction, there have been few attempts to structure and define the discipline. The primary exception is McGrath and Irving (1973a). They classify variance reduction techniques according to whether they modify the sampling process, make use of analytic equivalences, or are simply specialized techniques. This classification fails to show which techniques are related to or are particular cases of others, or provide insight into the underlying theory of variance reduction. In addition, the catchall category of "specialized techniques" is not satisfying.

considered over repeated realizations; i.e. its expected squared deviation from its own expectation. To account for the greater effort usually involved with achieving greater precision, measures that incorporate "effort" have been proposed. Most take the following form (Hammersley and Handscomb, 1964):

$$\frac{e_1 \sigma_1^2}{e_2 \sigma_2^2}$$

where σ_1^2 and σ_2^2 are the variances of estimators 1 and 2, respectively, and e_1 and e_2 are some measures of the effort involved in using estimators 1 and 2 (computer or analyst time, for instance).

Contrast the idea of increased precision (reduced variance) with increased accuracy. Accuracy refers to the absolute deviation of the value of the estimator from the quantity to be estimated. In some situations this quantity can be bounded. A similar measure is the expected value of the difference between the estimator and the parameter, called the bias. It is clear that any arbitrary constant has optimal precision, but it will probably be biased (unless one is so lucky as to select the value to be estimated). It is also clear that variance reduction in the context of numerical or quasi-Monte Carlo integration procedures has little meaning, while accuracy does. If unbiased estimators are employed, then precision is the only quantity to worry about. However, some VRTs trade variance of the estimator for bias, and this may be quite acceptable.

There are many possible solutions to this problem of definition. One could talk about mean squared error (MSE) reduction, since MSE is

potential error, and is often employed to construct confidence intervals for the unknown quantity being estimated. The smaller the variance of the estimator, the more certain one is that the estimate is not misleading. VRTs are usually considered to be methods for achieving a given level of precision at reduced cost, or greater precision at the same estimation cost. Precision is a quantity inversely proportional to the variance.

VRTs, sometimes called Monte Carlo swindles, have long been applied to Monte Carlo and sampling problems. It is possible to represent an integral as the expected value of a random variable, to sample from the random variable, and to use the sample average as an estimator of the integral. Increasing the number of observations will decrease the variance of the estimator, but an excessive number of observations may be required to achieve acceptable precision in the absence of variance reduction techniques. Similarly, to obtain an estimate with acceptable precision from a simple random sample of a large and diverse population an unreasonably large sample may be required.

More recently variance reduction techniques--many direct analogues of Monte Carlo and sampling methods--have been applied to computer simulation experiments. Increasing the length or number of simulation runs will improve the precision of the estimators, but not without cost. "Sometimes a variance reduction technique, properly applied, can make the difference between an impossibly expensive simulation and a frugal, useful one" (Law and Kelton, 1982).

In the simulation and Monte Carlo literature, <u>variance reduction</u> refers to any attempt to decrease the variance of an estimator

Carlo estimation literature. Monte Carlo estimation refers to the use of probabilistic models to evaluate mathematically intractable integrals. These problems may not be inherently stochastic. The difference between Monte Carlo and standard statistical estimation problems is that

> In the standard statistical-estimation problem both the probability distribution and the parameter to be estimated are assumed to be fixed; typically, given a sample of n values from the distribution, the best (or minimum variance) estimate of the parameter is to be found. In Monte Carlo calculations only ξ, the answer, is really fixed and the problem is to sample from that distribution which produces the minimum (or a substantially smaller) variance estimate of this number, for fixed cost. (Kahn and Marshall, 1953)

The theory of sampling is another closely related field that predates the development of simulation. Sampling refers to selecting a subset of the members of some population to discover or estimate some characteristic of the whole population. The measures derived from a sample are in general subject to random variation. "The purpose of sampling theory is to make sampling more efficient. It attempts to develop methods of sample selection and of estimation that provide, at the lowest possible cost, estimates that are precise enough for our purpose." (Cochran, 1977)

Variance Reduction

In sampling, Monte Carlo estimation, and computer simulation problems, one is often as interested in how far from the actual value the estimate of a quantity may be as in the value of the estimate itself. The variance of the estimator is a common measure of the

"These involve importance sampling, including the special case of the exponential transform, splitting, and Russian Roulette." (Steinberg, 1963, p. 142)

"...indeed the CV [control variate] technique is sometimes called regression sampling." (Law and Kelton, 1982, p. 360)

"The idea of this technique [stratified sampling] is similar to the idea of importance sampling..." (Rubinstein, 198., p. 131)

Control Variates is a form of correlated sampling (paraphrased). (Kahn and Marshall, 1953, p. 269)

"Use of quasi-random numbers is essentially an instance of systematic sampling." (Gentle, 1975, p. 8)

"A special case of the regression method is the use of antithetic sampling." (Gentle, 1975, p. 4)

"The antithetic-variate method is a variation of the regression sampling method introduced earlier." (Moy, 1965, p. 18)

"...in general, quota sampling may be described as stratified sampling

with a more or less nonrandom selection of units within strata." (Cochran, 1963, p. 137)

Surveys and Frameworks

A number of survey papers and book chapters on variance reduction have appeared in the literature. Most have concentrated on developing formulation, theory, and application of a particular VRT, and occasionally they have advanced ideas about the relationships between techniques or their taxonomy. In this section some of these works are reviewed, with particular emphasis on those that propose a classification scheme.

An important early survey paper is Kahn (1956), which appeared in the classic collection of papers from a symposium on Monte Carlo methods held at the University of Florida in 1954. Kahn uses the simple problem of repeatedly tossing two dice to estimate the probability that the sum is three to illustrate six VRTs that he feels are useful in Monte Carlo and simulation studies. Although the problem is easily solved analytically, the example makes the idea behind each approach clear. He continues the exposition using an integral formulation of the Monte Carlo estimation problem to describe each VRT from a mathematical point of view. He uses the integral problem because "it is the application in which the ideas are most clearly defined (p.147)." The development is sophisticated, considering a general multidimensional integral and deriving expressions for the variance of each modified estimator. When possible, potential applications of each VRT are given and optimal implementation strategies derived. Kahn does not attempt to define a

set of underlying concepts or develop a framework. However, in an earlier paper (Kahn, 1950) he does identify four general techniques for reducing variance in the context of neutron transport problems; they are: integration by random sampling, using a mathematical analog, quota sampling, and statistical estimation. There is considerable overlap between the categories, and many existing VRTs are not covered.

Steinberg (1963) proposes two principle classes of VRTs, those designed to reduce the theoretical variance of each sample "history", and those that reduce the variance of a set of sample "histories." A "history" is essentially an observation. This breakdown differentiates between techniques that change the individual observations and those that reduce variance through a cumulative effect. It ignores VRTs that change the statistic (function of the observations) used.

Probably the most cited reference in all the variance reduction literature is Hammersley and Handscomb (1964). Although many researchers consider Monte Carlo Methods to be concerned only with Monte Carlo problems, the authors are also interested in simulation and devote a chapter (Chapter 4) to the subject. In fact, their definition of Monte Carlo methods is quite general:

> Monte Carlo methods comprise that branch of experimental mathematics which is concerned with experiments on random numbers. (p. 2)

The text presents a brief history and overview of Monte Carlo methods and problems, develops the basic techniques (Chapter 5), and demonstrates their application to problems in areas such as solution of linear operator equations, radiation shielding and nuclear reaction criticality.

Like Kahn, Hammersley and Handscomb develop variance reduction techniques from an integral representation of the problem. They propose no general framework, but stress that two basic concepts underlying several VRTs are: 1) sampling from an advantageous distribution, not necessarily the one that naturally appears in the problem, and 2) replacing an estimate by an exact value when possible. VRTs are described, then illustrated using a simple integral for which an analytic solution is known. A major feature of the book is a list of original or early variance reduction and Monte Carlo references.

Another extensive list of variance reduction references is found in Kleijnen (1974). Kleijnen is particularly interested in VRTs that can be used in a wide range of simulation studies. As a result, he does not discuss some VRTs that appear in the Monte Carlo literature and includes one (Selective Sampling) that is unique to simulation. A description and "critical appraisal" of six VRTs is given, with extensions, limitations, and corrections presented. Also, the combined use of two well known VRTs (antithetic variates and common random numbers), and the resulting dangers, are discussed.

While Kleijnen does not propose a specific variance reduction framework, he opens his chapter with the following comments:

> Some VRT's change the original sampling process completely.... Other VRT's use the same sampling process as in crude sampling, but after the sampling has ended, they do not use the sample average $\bar{x}$ but a more sophisticated estimator.... Some VRT's modify the sampling process in a very subtle way....

While this is clearly true, it fails to completely describe all the possibilities. However, a slight modification of Kleijnen's ideas is

exhaustive: A VRT can change the inputs to the simulation, it can change the simulation model, or it can change the simulation outputs. This framework has merit, but it groups techniques that exploit different knowledge or problem characteristics. These concepts deal more with when than what a VRT does.

Another survey paper is Gentle (1975). Gentle is interested in VRTs that are robust to deviations of the simulation or Monte Carlo model assumptions from reality, and deviations of the distributions produced by the random variate generators from the desired distributions. He describes nine distinct techniques and combinations of some of them. Like Kleijnen he warns that "simultaneous use of variance-reducing techniques may not be effective when different methods achieve reduction in conflicting ways...." (p. 9)

McGrath and Irving (1973a) is the first real attempt to define the concepts underlying variance reduction methods and develop a framework based on them. The overall purpose of the paper is to provide analysts with some understanding of variance reduction techniques and a useful guide for selecting a VRT for a particular problem (p. 5). In this context they identify the following concepts that variance reduction techniques employ to increase the efficiency of simulation: 1) Modify the simulation procedure, 2) Utilize approximate or analytic information, and 3) Study the system within a different context or abstract representation. Based on these concepts the authors propose the following categories:

1. Modification of the Sampling Process

2. Use of Analytical Equivalence

3. Specialized Techniques

> The techniques which modify the sampling process effectively alter the probability distributions of 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 [including the combination of two or more techniques in the other categories]. (p. 27)

McGrath and Irving are able to classify sixteen VRTs using the above scheme, but they comment that many of these techniques are related and it is quite difficult to arrive at a completely distinct classification structure. This is indeed the case. For instance, the authors place systematic sampling and antithetic variates in categories 1 and 2, respectively, yet it was noted above that Roach and Wright (1974) claim one is a subset of the other. Also, there are several problems with the "Specialized Techniques" category. Since combinations of other VRTs are contained here, it means that techniques related only by being combinations of others are grouped together. Also, there is an implicit assumption that all of the VRTs classified in categories 1 and 2 are fundamental, or else they would be in group 3. However, it can be shown that some of these VRTs are built up from still more fundamental concepts (see Chapter 5). Finally, the category is a catchall for any VRT that does not fall in either 1 or 2, which is not desirable.

Like many of the other authors, McGrath and Irving stress an integral formulation of the general Monte Carlo or simulation problem, claiming that it is completely general. To demonstrate the generality they formulate a network queuing problem in this way. For most VRTs covered in the paper, a theoretical development, comparison with crude sampling, and example application are given. Also, a concluding section of the report provides a systematic procedure for selecting and applying several of the more important VRTs.

Kohlas (1978) claims that the most widely known VRTs can be divided into two groups: correlation methods and the methods of essential sample. Correlation methods are further divided into comparison and compensation methods; both involve manipulating or taking advantage of correlation between observations to increase statistical efficiency. Essential sample methods attempt to concentrate sampling in regions that will make "significant contributions" to the estimate.

Chapter 4 of Rubinstein (1981) is essentially an updated version of Hammersley and Handscomb's Chapter 5. Rubinstein reviews most of the same techniques the previous authors did, occasionally adding theorems concerning conditions that insure a variance reduction, an expression for the theoretical variance of an estimator, or an implementation algorithm. Also, several less well-known VRTs, some developed since the publication of Monte Carlo Methods, are explained. No general concepts or framework is presented, but a significant list of references is included.

A recent paper by Wilson (1983a) proposes another framework for classifying VRTs. He has two categories:

1. Importance Methods
2. Correlation Methods

The importance methods achieve improved efficiency by concentrating the sampling in regions of the input domain that make the greatest contribution to the integral. The correlation methods are further divided into those that induce favorable correlation between blocks of simulation runs, and those that exploit "inherent" correlation between output variables within each run (p. 1). The author does not claim that this categorization includes all existing VRTs, but he is able to fit eight well-known methods into the scheme. The primary purpose of the paper is to survey recent research in variance reduction and comment on its potential benefit in simulation studies.

Although this chapter did not review any sampling literature, the interested reader is referred to Cochran (1977).

SIMULATION

To provide a context for the discussion of variance reduction and the results presented in the next chapter, a rigorous and detailed definition of simulation experiments is developed in this chapter. As a preliminary, some notation is established and the concept of statistical "information" is discussed.

Notation

Descriptions will generally be in terms of matrices, columns of matrices, and elements of matrices. Letters, Greek or English, without subscripts will denote matrices, letters with single subscripts will denote columns, and doubly subscripted letters are elements, using the usual row-column convention. For instance, X_{ik} is the i^{th} element of column vector X_k, which is the k^{th} column in the matrix X.

A letter with subscripts in parentheses indicates a set of variables with indices in a fixed set. $X_{(ab)}$ would be used to designate all the elements X_{ik} in X with subscripts in index set (ab), a set that would have to be defined. The ( ) is a mapping from a single index to a set of indices.

Random variables will be denoted by capital English letters, and realizations of these random variables by lower case letters. For

example, $y_{i\ell}$ is a realization of random variable $Y_{i\ell}$. Any notation that is counter to the above conventions will be specifically defined when used.

Information

The information about an unknown parameter contained in a random variable refers to the extent to which uncertainty about the value of the parameter is reduced by observing the random variable. The term information will always be used in this statistical sense. It should not be confused with knowledge of the system to be modeled, or knowledge of certain mathematical or statistical relationships. Knowledge relates to what is understood or recognized by the experimenter, while information is an abstract statistical quantity. A specific measure of information is not needed, but any measure that is used or proposed must satisfy the requirements given below.

Consider a statistical space $S = (\Omega, A, P_\theta, \theta \in \Theta)$, where Ω is a sample space, A a σ-algebra of subsets of Ω (the events), and P_θ is a family of probability measures on Ω indexed by a parameter θ. Although this development is quite general, Ω and Θ can be restricted to subsets of some finite dimensional Euclidean spaces without loss of generality. Define a random variable Z to be a measurable mapping from (Ω, A) to a measurable space (Φ, B) that does not depend on θ. Then Z induces a statistical space $S_Z = (\Phi, B, P_Z)$; Z may be a statistic used to estimate θ, the identity mapping, or in general any arbitrary measurable function. Consider the amount of information available to estimate θ, and call $I(\theta)$ a measure of this information only if it satisfies the

following properties.[1]

(1) The amount of information available with respect to an unknown parameter θ is defined for a statistical space S independently of any estimation procedure used or inference desired.

(2) The amount of information contained in Z equals the amount of information contained in the statistical space S_Z induced by Z.

(3) The amount of information contained in Z is less than or equal to the amount of information contained in the statistical space S on which Z is defined.

(4) A sufficient statistic[2] Z contains all of the information included in the statistical space on which Z is defined. If there is a unique value of Z (a.s.) corresponding to each possible value of the parameter θ, then Z contains the maximum possible information; if Z has the same distribution for all values of the parameter (a.s.) then it provides no information.

(5) The information given by two statistically independent functions defined on the statistical space S is the sum of the information given by each of them. If the two functions are not independent, then there may be a cumulative effect resulting in the total information being greater or smaller than the sum of the individual informations.

(6) The efficiency (as measured by the variance of the estimator) with which θ can be estimated is nondecreasing in the amount of information contained in the statistical space on which the estimator is

1. Barra, J.R. (1981), Mathematical Basis of Statistics, (L. Herbach, translation ed.), Academic Press, N.Y.
2. Bickel, P.J. and K.A. Doksum (1977), Mathematical Statistics: Basic Ideas and Selected Topics, Holden-Day, San Francisco.

based; greater information implies the potential for more efficient estimation.

A well-known measure of the information in a real valued vector or general multidimensional random variable, X, about a real valued parameter θ, is the Fisher information number[3]; if θ is a scalar then

$$I(\theta) = E\left[\left[\frac{d\log L(\theta,X)}{d\theta}\right]^2\right]$$

where L is the likelihood function of θ given X. Under certain regularity conditions[4] the minimum variance attainable when estimating θ from X is a function of $I(\theta)$. The Fisher information measure extends in a straightforward way to a vector θ. This measure also satisfies the six properties stated above. Clearly the value of $I(\theta)$ is independent of any function of X since it depends only on the probability distribution of X. The logarithmic form of the measure implies that (4) and (5) will hold because of the product form of the distribution of independent random variables and the factorization result for sufficient statistics. That (3) holds is well-known, and the Cramer-Rao lower bound[5] relates the variance of an estimator to $I(\theta)$.

As an illustration of some of these properties consider the following example: Let the sample space be R^2, the two dimensional Euclidean space, with the probability measure being the bivariate normal distribution with identical marginals denoted by $N(\mu,\mu,\sigma^2,\sigma^2,\rho)$. A random sample (X_1,X_2) has Fisher information measure

3. Fisher, R.A. (1925), "Theory of Statistical Estimation," Proc. Camb. Phil. Soc., 22, 700-725.
4. Rao, C.R. (1973), Linear Statistical Inference and Its Applications, Wiley, N.Y.
5. Bickel, P.J. and K.A. Doksum (1977), Mathematical Statistics: Basic Ideas and Selected Topics, Holden-Day, San Francisco.

$$I(\mu) = \frac{2}{(1+\rho)\sigma^2}$$

relative to μ. Note that in this example $I(\mu)$ does not depend on μ, which is not true in general. There are two cases to examine: If $\rho = 0$, so that X_1 and X_2 are independent $N(\mu, \sigma^2)$ random variables, then

$$I(\mu) = \frac{2}{\sigma^2} = \frac{1}{\sigma^2} + \frac{1}{\sigma^2}$$

The total information is the sum of the information contained in the independent sources. Note that the amount of information is increased if σ^2 is decreased. If, on the other hand, $\rho \neq 0$ there is a cumulative effect leading to more ($\rho < 0$) or less ($\rho > 0$) information than in the independent case. For this example the conditions of the Cramer-Rao lower bound are satisfied, implying that the minimum variance attainable for an unbiased estimate of μ from (X_1, X_2) is $\frac{1}{I(\mu)}$.

Measures of information other than the kind considered here have appeared in the literature. Kullback[6] discusses a measure of the amount of information available for discriminating between two hypotheses about the probability measure on the statistical space S. If one considers the information for discriminating between θ and $\theta + \Delta\theta$, then the Kullback and Fisher measures are closely related . Shannon's measure of information[7] is used in communication theory to quantify the amount of uncertainty or entropy present in a message source. The more uncertainty (freedom of choice) there is in composing a message, the

6. Kullback, S. (1959), Information Theory and Statistics, Wiley, N.Y.
7. Shannon, C.E. and W. Weaver (1963), The Mathematical Theory of Communication, The Univ. of Illinois Press, Urbana.

more information the message itself contains. This measure is quite different from those of Fisher and Kullback, although some parallels can be drawn.[8]

The concept of statistical information is important in the discussion of variance reduction since, as will be shown, variance reduction techniques achieve their results by increasing the amount of information available and/or making more efficient use of the available information.

Simulation Experiments

In this section a definition of simulation experiments is given. Later, Monte Carlo and sampling problems are shown to be special cases of this definition. The definition proves useful for discussion of variance reduction by showing how the random variables in a simulation are defined, where statistical information about the parameters of interest is, and where information may be increased or lost. As will be shown, VRTs transform the random variables in a simulation experiment to increase and/or make better use of information, so this perspective captures the essential features of variance reduction. The development below is rather abstract. Later in the chapter examples are given that show how some of the constructs usually encountered in simulations (such as time and initial conditions) are contained in this definition.

Before beginning, a concept is introduced that will be useful later. Consider three sets of random variables U, V, and W. Let V =

8. Schutzenberger, M.P. (1956) "On Some Measures of Information Used In Statistics," in Information Theory, (C. Cherry, ed.), Academic Press, N.Y.

variables. For instance, in the Monte Carlo problem (2.2) the outputs are ordered $Y_1,\ldots,Y_n$ corresponding to when the observations are generated. Trying to relate successive rows in X and Y to the passage of time is tempting, but is a limited viewpoint true only in the simplest experiments. However, the order <u>within</u> columns does correspond to the order in which realizations within that column will be generated. Since time is such a common construct in simulation, it is worthwhile to consider briefly how it is incorporated into the definition given above. There are two cases: time advanced in fixed increments and at random event times.

When time is advanced in known, fixed increments, Δt, the time increment is part of the definition of an output transformation. The clock time after i increments is a transformation of the previous clock time t_{i-1}, namely,

$$t_i = g(t_{i-1}, \Delta t) = t_{i-1} + \Delta t$$

Thus, the clock time is an output, and would occupy a column in Y (if it is essential).

In "discrete event simulation" (see for instance, Fishman, 1978, Pritsker and Pegden, 1979, Law and Kelton, 1982, Banks and Carson, 1983, or Bratley, Fox, and Schrage, 1983) clock time advances in discrete, random steps between the occurrence of events. Usually the probability distribution of the interval is not explicitly known and advances are generated by a complicated transformation of other random variables. For example, in a large queuing network with many servers, service centers, and customer types, the interaction of various service times,

$$Z = \left[\overline{W}\right]$$

and $\theta = [\theta_w]$ is the parameter of interest. Replacing A, S, Q, β, W, $\overline{W}$ with appropriately subscripted elements of X, Y and Z is straightforward, and g is given by equation (3.4), which is written in a form like (3.2). The feasible region for $I = [I_w, I_q]$ is

$$R = \{(n,n): \ n = 1,2,\ldots\}$$

since generation of a realization of W requires a corresponding Q. If the experiment requires a sample of 500 waiting times, then $R_* = \{500, 500\}$.

As a second example, consider the general Monte Carlo estimation problem of Chapter 2, specifically equation (2.2). In that example, θ, f, g, X, and Z correspond to the similar terms defined above. Now let

$$h(Y) = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

where $Y_i = g_2(X_i)$. The function h is defined by its functional form (a summation of terms divided by the number of terms in the sum) and the particular argument, Y_i. Note that $I = n$ can be any arbitrary positive integer, and the arguments of h all form a single Y column.

Time

Many simulation experiments explicitly account for the passage of time, and all simulations have some underlying ordering of their random

the system be denoted by A, a random variable distributed exponentially with some known, fixed rate. Let the service time be S, distributed exponentially with parameter β, where β is a function of Q, the number of customers in the queue at the beginning of service. Time in the queue for the i^{th} customer, W_i, is given by the well-known relation:

$$W_i = \max(0, W_{i-1} + S_{i-1} - A_i) \tag{3.4}$$

An estimator of θ_w is

$$\overline{W} = \frac{1}{n} \sum_{i=1}^{n} W_i$$

where n is the number of customers. Based on the definitions above, the input, output, and statistic matrices are:

$$X = \begin{bmatrix} A_1 & S_1 \\ . & . \\ . & . \\ . & . \\ A_n & S_n \end{bmatrix}$$

$$Y = \begin{bmatrix} W_1 & Q_1 \\ . & . \\ . & . \\ . & . \\ W_n & Q_n \end{bmatrix}$$

Definition: E and E' are everywhere equivalent ($E \overset{e}{=} E'$) if and only if

$$X \overset{d}{=} X'$$

$$g(x) = c(g'(x)) \quad R_* = R'_*$$

and

$$h(y) = h'(y') \quad \forall \text{ realizations } y, y' \text{ generated from common } x$$

Clearly, everywhere equivalence implies strong equivalence.

Considering all experiments based on some given Ω and θ, the definitions of equivalence partition the simulation experiments into equivalence classes. A VRT transforms an experiment into another experiment with the same context, but hopefully one that is not d-equivalent and in fact has statistics with better estimation properties. Usually d and s-equivalence are conditions that cannot be verified. However, a VRT should yield an experiment that is not e-equivalent to the original one. Since e-equivalence is the finest partition, characterizing the ways in which simulation experiments are transformed into other, non e-equivalent experiments characterizes the ways to transform them into non s and d-equivalent experiments as well.

Examples and Common Constructs

Consider a simulation model of a single server, first-come-first-served queuing system that is used to estimate θ_w, the expected time a customer spends waiting for service. Let the time between arrivals to

realizations.

Beyond this trivial distinction, definitions of three types of equivalences between experiments will be given. To be equivalent in any sense used here, experiments must be based on the same sample space Ω and have the same parameters of interest, θ. Let E and E' be two simulation experiments with common context (Ω,θ).

Definition: E and E' are equivalent in distribution ($E \overset{d}{=} E'$) if and only if

$$Z \overset{d}{=} Z'$$

Since the ultimate goal is to estimate θ via the statistics Z, if two experiments are equivalent in distribution they have the same statistical properties. However, the distribution of Z is not generally known, while the following may be:

Definition: E and E' are strongly equivalent ($E \overset{s}{=} E'$) if and only if

$$X \overset{d}{=} X'$$

and

$$h(g(x)) = h'(g'(x)) \quad R_* = R'_* \quad \forall \text{ realizations } x \text{ of } X$$

Clearly, strong equivalence implies equivalence in distribution.

Definition: g and g' are equivalent except for coding ($g \overset{c}{=} g'$) if there exists a one-to-one mapping c such that $g(x) = c(g'(x))$.

The next definition of equivalence is at the level on which random variables in the experiment are actually defined.

The preceding definitions imply that f_ω, $(g;R_*)$ and h are unambiguously defined (up to equivalent essential sets) for a given simulation experiment.

Equivalence of Experiments

In the preceding section a definition of simulation experiments was presented with the idea that, given a particular experiment, it is possible to partition the random variables into matrices X, Y and Z, and to (at least implicitly) identify f_ω, $(g;R_*)$, and h. It has already been noted that the identification is not unique, since for a given experiment it is possible to reorder the column indices of the variables and to have different values of ℓ^* and k^* without changing the experiment. Distinguishing between experiments that are identical except for the order of their subscripts or values of ℓ^* and k^* is not necessary. Specifically, consider an arbitrary sequence $\{a_i\}$ $i = 1,2,\ldots,I_a$. Also consider two subsequences $\{b_i\}$ and $\{c_i\}$ with the property that

$$\{b_i\} \cap \{c_i\} = \phi$$

and

$$\{b_i\} \cup \{c_i\} = \{a_i\}$$

for all values of I_a. Then for the purposes of this research the two representations of the sequence are equivalent. Identify sequences such as $\{f_{ik}\}$ and $\{g_{i\ell}\}$ with $\{a_i\}$. Order in the sequence and subsequences is only constrained by the correspondence to order of generation of

transformations of elements in X, the relative lengths of the different sequences being subject to some restrictions. The vector Z is defined by functions of the columns of Y that contain information about θ.

Axioms of Simulation

A simulation experiment is composed of the sets of random variables X, Y, and Z, which are defined by f_ω, $(g;R_*)$, and h, respectively. However, the existence of such definitions does not in itself constitute a simulation. There are two necessary axioms that must be acknowledged before X, Y, and Z represent a simulation experiment:

Axiom 1 (Existence of Information): The random variables X, Y, and Z have probability distributions that depend on θ.

Axiom 2 (Existence of Realizations): Given a source of randomness, realizations of X, Y, and Z can be (perhaps iteratively) generated.

Axiom 1 guarantees that estimation of θ from realizations of X, Y, and Z is not fruitless. Axiom 2 is more subtle; it establishes that the dependencies between and within X and Y are such that R is not empty. This restriction limits the potential transformations of an experiment.

Definition: A simulation experiment, denoted by $E(f_\omega,(g;R_*),h;\Omega,\theta)$, is specified by a context (Ω,θ), a probability distribution f_ω on Ω, output transformation and sampling plan $(g;R_*)$, and statistics function h, subject to Axioms 1 and 2. Where there will be no confusion, the shortened notation E will be used.

$$Z = h(Y) = \{h_m(Y_{(m)})\} \quad m = 1,2,\ldots,m^* \tag{3.3}$$

For each m, h_m is a transformation from a set of output sequences $Y_{(m)}$ to Z_m. These functions do not depend on θ or I_m. The dimension m^* is known, finite, and equals the dimension of θ.

The statistical space induced by h is $S_z = (\Gamma, \mathcal{Z}, f_h)$, where again f_h is usually not known, or if known is not used to generate realizations of Z.

Notationally, let:

$$h = \left[h_1 \quad . \quad . \quad . \quad h_{m^*} \right]$$

and

$$Z = \left[Z_1 \quad . \quad . \quad . \quad Z_{m^*} \right].$$

The m^{th} element in Z is defined by the function h_m and its argument $Y_{(m)}$. Thus, the functional forms of h_m and h'_m might both involve a summation of squared terms even though $(m) \neq (m)'$. Each h_m is defined for arbitrary length of the sequences $Y_{(m)}$.

The statistics aggregate output random variables. Note that each element θ_m in θ has a corresponding estimator Z_m in Z. Often h_m can be thought of as being computed in stages, defining certain intermediate random variables from Y and then combining them into a final estimator.

To summarize, X is the matrix of all random variables in the experiment whose probability distributions are known, in the sense stated above. Elements of Y are obtained by application of sequences of

defined on X may not be unique. A useful restriction that can be imposed without loss of generality is: A column $Y_\ell \subset Y$ will represent 1) an argument, considered as a whole, for one or more statistics, Z (see below), and/or 2) a sequence $\beta_{i\ell}$ that parameterizes f_{ik} for one or more k, and/or 3) an output needed to specify the sampling plan, R_*. Thus, columns are delimited by the purpose they serve. Note that 1) - 3) define the essential random variables; random variables that are either arguments for the statistics, necessary for generation of realizations of X, or determine when the desired sample has been obtained. Information about θ may be lost, but not gained, in the transformation from S_x to S_y (see properties of information above). The matrix Y is a minimal set containing all the useful information about θ.

An alternative representation of (3.1) is

$$Y_{i\ell} = g_{i\ell}([X|Y]_{(i\ell)}) \tag{3.2}$$

where $[X|Y]$ is the matrix obtained by including all columns of X and Y. This notation may seem more natural in the simulation context, since realizations of outputs are often generated iteratively from transformations of Xs and previous Ys. However, since all elements of Y are ultimately functions of X, (3.1) is completely general and has the advantage that, given a particular experiment, $X_{(i\ell)}$ is unique.

Definition: Z, the row vector of statistics, has real valued, scalar random variables as elements that are estimators of θ. The statistics are specified by h, where

from a set of elements $X_{(i\ell)}$ to $Y_{i\ell}$. These transformations do not depend on θ. The region R is an ℓ^* dimensional feasible region that accounts for the interrelationship between the individual $g_{i\ell}$s; R is determined by all possible realizations of X. The sampling plan R_* is specified so that $I \in R_*$ with probability one when realizations are generated. The column dimension ℓ^* is finite and does not depend on I.

The statistical space induced by g is $S_y = (\Phi, \Psi, f_g)$. The probability distribution f_g is usually not known, or if known is not used to generate realizations of Y. It is, however, naturally parameterized by θ so that estimation of θ is possible from realizations of Y.

Notationally, let:

$$g = \begin{bmatrix} g_{i1} & \cdots & g_{i\ell^*} \end{bmatrix}$$

and

$$Y = \begin{bmatrix} Y_{11} & \cdots & Y_{1\ell^*} \\ Y_{21} & \cdots & Y_{2\ell^*} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \end{bmatrix}$$

The row index i corresponds to the order in which realizations are generated in an output sequence. The number of sequences, ℓ^*, is not uniquely determined since some sequences could be divided or merged and still satisfy the definition just given. Also, the essential set

all that is required is that some joint or marginal distribution including X_{ik} be known. f is the vector of all marginal distributions of f_ω. A special case of an element of X is a constant selected by the experimenter from a distribution known to him (implicitly or explicitly).

The lengths of the sequences $\{X_{ik}\}$ are infinite, but the number of realizations is determined by the frequency of sampling from $\{f_{ik}\}$. The index i corresponds to the order of sampling. The number of input sequences, k^*, is finite and known. However, k^* is not uniquely determined since some sequences could be divided or merged and still satisfy the definition just given.

In later discussions it will be necessary to work with conditional distributions of f_ω. The distribution of $X_{(ik)} \subset X$ given $X_{(j\ell)} \subset X$ is denoted

$$f_{(ik)|(j\ell)}(x_{(ik)}|x_{(j\ell)})$$

where the parameter β has been suppressed. A shorthand version is $f_{(ik)|(j\ell)}$.

Definition: Y, the matrix of outputs, has real valued, scalar random variables as elements and is an essential set of all random variables defined on X. The outputs are specified by $(g;R_*)$, where

$$Y = g(X) = \{g_{i\ell}(X_{(i\ell)})\} \quad i = 1,2,\ldots,I_\ell \quad \ell = 1,2,\ldots,\ell^* \qquad (3.1)$$

$$I = [I_1\ I_2\ \ldots\ I_{\ell^*}] \in R_* \subset R$$

For each ℓ, $\{g_{i\ell}\}$ is a potentially infinite sequence of transformations

multivariate) random variable with unknown distribution (see below).

The associated statistical space is $S_x = (\Omega, \mathcal{F}, f_\omega)$, where Ω is the sample space for X, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and f_ω is the probability distribution on Ω.

All of the information in the simulation experiment for estimation of θ is contained in the statistical space S_x. However, the probability distribution f_ω is not "naturally parameterized" by θ, which means that θ is some unknown or complicated function of parameters of f_ω, making estimation of θ directly from realizations of X difficult. Thus, transformation of S_x into a space whose induced probability distribution is conveniently parameterized by θ is desirable. Notationally, let:

$$f = \begin{bmatrix} f_{i1} & f_{i2} & \cdot & \cdot & \cdot & f_{ik^*} \end{bmatrix}$$

and

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdot & \cdot & \cdot & X_{1k^*} \\ X_{21} & X_{22} & \cdot & \cdot & \cdot & X_{2k^*} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

The $\{f_{ik}\} = \{f_{ik}(x_{ik}|\beta_{(ik)})\}$ are k^* sequences of scalar marginal distributions of f_ω. For fixed k, they are identical for all i expect possibly for the value of $\beta_{(ik)} \subseteq \beta$. Each element X_{ik} has marginal distribution f_{ik}, but statistical dependencies can exist within or between column elements. Note that f_{ik} may only be known implicitly;

carried out on a computer, these assumptions are not restrictive.

Context of the Experiment

Definition: θ, the row vector of parameters of interest, has unknown, but fixed, real scalar constants as elements, and has dimension m^*. The purpose of performing a simulation experiment is to estimate θ.

Note that θ is fixed, but other elements of the simulation experiment are not (see below). Although θ is just a vector of constants, it has a context given by Ω, a sample space sampled from to estimate θ.

Definition: Ω, the sample space of the inputs, is a subset of some infinite dimensional Euclidean space; Ω is the intersection of the set of all sample spaces that can be sampled from according to known probability distributions (see below), and the set of all sample spaces whose sampling distributions contain information about θ.

Note that Ω is a fixed space that will be sampled from according to a known distribution. It is possible that some subsets of Ω have probability zero. There will be other sample spaces in the simulation experiment induced by transformations of Ω.

Definition of the Experiment

Definition: X, the matrix of inputs, has real valued, scalar random variables as elements and known multivariate probability distribution. $X \quad f_\omega(x|\beta)$. Here, known distribution means that f_ω is specified by an analytic or tabular expression with parameter β, a real (possibly

$V(u)$ and $W = W(v)$, where u and v are realizations of U and V, respectively.

Definition: A subset $V' \subseteq V$ is an essential set defined on U if 1) for each element V_0 of V there exists a known transformation from elements in V' to V_0, and the transformation does not depend on the probability distribution of V, and 2) for any element $V_0 \in V'$, no such set of transformations exists for $V' - V_0$.

Thus, W can be defined as a transformation of V' alone, but deletion of any element of V' means W may not be defined. Note that the essential subset may not be unique. For example, suppose U is a scalar, and $V_1 = U$, while $V_2 = U^2$. Let $W = V_1 + V_2$. Then $V' = \{V_1\}$, unless it is known that $U \geq 0$, in which case V' could be either $\{V_1\}$ or $\{V_2\}$.

As a second example, let $U = \{U_1, U_2\}$, and $V = \{V_1, V_2\}$ where

$$V_1 = U_1^2 \quad \text{and} \quad V_2 = \frac{U_2}{U_1^2}$$

Then if $W = \frac{V_2}{2}$, $V' = V$.

As a less abstract example, consider the simulation of a service system and three random variables associated with each customer: waiting time, service time, and total delay. As essential subset is any two of the three, since given two the third can be derived by simple arithmetic.

In the development that follows, all sample spaces are subsets of Euclidean spaces of some dimension, and all random variables defined on them are real valued. Also, all transformations are Lebesgue measurable and integrable. Since the concern is for experiments that can be

interarrival times, and queuing disciplines determine when the next event will occur. Clearly, the time step is a member of Y. The clock time may be in Y if it is essential.

Initial Conditions

Two long standing issues in the simulation of systems in steady state are the specification of initial "startup" conditions and the estimation of steady state parameters from outputs that may be contaminated by the chosen conditions.[9] The estimation problem is a question of what statistic to choose. Initial conditions are often constants, chosen for convenience ("empty and idle") or because they are expected to be consistent with the steady state distribution. They may also be selected randomly from a known distribution. In either case they would be classified as inputs in the simulation experiment.

As an example, consider a simulation that generates outputs described by

$$Y_{i1} = \alpha Y_{i-1,1} + X_{i2}$$

an autoregressive process of order one, where X_{i2} $i = 1,2,\ldots$ are identically distributed with some known distribution. If the outputs are to begin with Y_{11}, then an initial value for Y_{01} must be given. Let that value be X_{11}, either a constant or a random variable with a known distribution. Then the matrices of inputs and outputs look like:

9. Wilson, J.R. and A.A.B. Pritsker (1978), "A Survey of Research on the Simulation Startup Problem," Simulation, 31, 55-58.

$$X = \begin{bmatrix} X_{11} & X_{12} \\ & X_{22} \\ & \cdot \\ & \cdot \\ & \cdot \end{bmatrix}$$

$$Y = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

and

$$Y_{i1} = g([X|Y]_{(i1)}) = \delta_{i>1}(\alpha Y_{i-1,1}) + \delta_{i=1}(\alpha X_{11}) + X_{i2}$$

where δ_c is an indicator function that equals 1 or 0 if condition c is true or false, respectively.

Aside: This case should not be confused with using a time series algorithm to generate random variables with a known distribution; if the distribution is known then the random variables are inputs, no matter what method we use to generate realizations. For example, if $Y_{11}, Y_{21}, \ldots, Y_{n1}$ have a known multivariate normal distribution and are generated as such, then they become $X_{11}, X_{21}, \ldots, X_{n1}$.

Stopping Rules

In the examples given earlier, the sampling plan R_* was specified by a number, say n, of observations of a particular output. This was easily denoted by a single element of R. Two other cases are possible.

It may be that R_* is a region, rather than a point, in which case sampling stops when I is first contained in R_*. For instance, suppose that in a network queuing simulation the run is terminated when station 1 has serviced 50, and/or station 2 has serviced 60 customers. If $I = [I_1, I_2]$, then

$$R_* = \{I_1, I_2: \ I_1 \geq 50 \ \text{ or } \ I_2 \geq 60 \ \text{ and } \ I \in R\}$$

The second, more interesting case is stopping rules based on satisfying a condition other than a count. A simulation run that terminates when the clock time is 480 minutes, or one that stops when a resource is depleted are examples. Situations such as these can always be represented by an output variable whose realization indicates that the stopping condition has been met. For convenience denote this output by Y_{ℓ_*}, and write

$$R_* = \{I: \ I_{\ell_*} = 1 \ \text{ and } \ I \in R\}$$

Sequential Procedures

In simulation experiments, as well as in general statistical experiments, sequential procedures may be employed to estimate a parameter of interest. Such procedures are characterized by the

selection of an initial sample, analysis of the results, and a decision to continue sampling or stop based on the results of the analysis.

Considering such procedures in light of the definition of simulation experiments, modelling sequential sampling might at first appear to require some sort of "feedback" or control structure from the statistics (h(Y)) to the output transformations (g(X)). However, it is our conscious intention to make a distinction between the "design" and the "analysis" aspects of simulation experiments. Sequential procedures affect the sampling plan, R_*, and thus are a part of the design aspect (as are the inputs). In fact, sequential procedures are simply a kind of stopping rule, as discussed earlier. The statistics are the analysis part of the experiment, and will always be functions of a fixed pool of data. While this is not the only possible perspective, it seems justified since restrictions on sampling are embodied in f_ω and g.

Joint Distributions of the Inputs

As stated earlier, the elements in X can have any feasible joint distribution. Often the elements in a column are independent, identically distributed random variables indexed by the order in which realizations are sampled. When encountering statistical dependencies, however, two types are common: Identically distributed multivariate vectors where each element in the vector has a different marginal distribution, and identically distributed scalar random variables that are pairwise (or in general m-tuplewise) correlated. The first type would be represented by columns of scalars where corresponding (same row index) elements have known joint distribution. The second case is

denoted by a single column where the row indices indicate the correlation structure. Remember, also, that if the distribution is not completely specified, then the unknown parameters for marginal distributions f_{ik} are given by one or more columns in Y.

Confidence Intervals

In this research the assumed goal of the simulation experiment is to derive point estimates of unknown, real parameters. However, outputs from simulation experiments are often used to construct confidence intervals on these parameters. Variance reduction and confidence interval construction are related because the properties of the interval are generally a function of the properties of the point estimator(s). Thus, while attention is restricted to point estimates and their variance the research is relevant to confidence interval construction.

Types of Statistics

The statistics defined by a simulation experiment can be separated into two types based on the outputs that are their arguments. Observational statistics are based on individual, discrete outputs without relation to when the output was generated. Time-persistent outputs have values defined over time, and require not only the value but also the time period over which it persisted ("time" can be any index). In the definition of simulation the distinction is irrelevant, since both types of outputs are represented by scalar random variables. Also, it does not matter whether the outputs came from replication

"runs" (usually implying independent replications) or a single steady state "run"; columns in Y can represent either kind of output. Of course, the particular estimators used will depend on both these factors.

Extension to Realizations

The definition of simulation experiments given above can be easily extended to describe how realizations of the experiment are generated. A basic source of randomness--usually scalar, independent, identically distributed U(0,1) random variables--is transformed into random variables X distributed according to f_ω. The transformations g and h are employed to yield a realization of Y and Z from a realization of X. Of course, simulations are usually implemented as computer algorithms. Note again that, when considering variance reduction, the interest is in how random variables are defined and not how realizations are actually generated, although the method of generation will often affect what can be achieved in practice.

Monte Carlo and Sampling Experiments

The general Monte Carlo integration problem fits easily into the simulation characterization. Since the problem is that of evaluating a known integral, and since the problem can be made stochastic by introducing a probability distribution into the integral, the defining f distributions, g transformations, and h functions are easily identified and can be expressed in closed form (see the earlier example). However,

even the solution of problems like (2.1) by numerical or quasi-Monte Carlo methods (Hammersley and Handscomb, 1964) is covered by the definition. In those situations, the points at which the integrand is evaluated (the inputs) are known constants. Thus, there is zero variance; independent realizations of the experiment will all yield the same estimate. However, the accuracy of the method (difference between the estimate and the true answer) will not be zero in general.

That survey sampling problems can be characterized as above is less obvious. However, in the usual case of probability sampling (Cochran, 1977) the relationship can be demonstrated. In probability sampling a set of distinct samples from a (usually finite) population that the sampling procedure is capable of selecting is defined, and each possible sample is assigned a probability of selection. One of the samples is selected with likelihood given by this probability, and an estimate of whatever quantity is of interest is made from the responses given by the elements in the sample. In terms of statistical spaces, the triple $(\Omega, \mathcal{F}, f_\omega)$ corresponds to the possible samples, the events, and the probabilities assigned to each possible sample. The most common sampling distribution is simple random sampling, where each possible sample is equally likely to be selected. Sampling with or without replacement are two procedures for generating such samples, depending on whether the same element can appear in a sample more than once. The g transformations that induce the space $(\Phi, \mathcal{F}, f_g)$ are more implicit, representing how responses from the sample are obtained and the allocation of sampling effort. The space S_z is as before, representing the estimators employed.

Consider, for example, an experiment to determine the expected lifetime of a type of light bulb. The procedure might be to take a sample of size n from a lot of bulbs in such a way that each possible size n sample is equally likely, burn the bulbs, record the time until burnout, and estimate the expected life by the average of these values. The population of light bulbs and the sampling procedure define X, the sample size n and the method of establishing lifetimes defines Y, and Z is defined by the estimation rule (simple average) and the outputs, Y.

Analytic Solutions

Consider a given simulation experiment, E. Given sufficient insight, it may be that the value of θ can be deduced analytically. Such a solution procedure is not outside the scope of this research. Think of a continuum between the original stochastic experment and an analytic determination of θ, specified by the precision of the statistics Z in the experiment. Then VRTs transform a given experiment into another one in this continuum. An analytic solution, of course, is the limiting case having infinite precision.

CLASSES OF TRANSFORMATIONS

Given the definition of simulation experiments in the previous chapter, this chapter develops a framework for VRTs based on six classes of transformations of simulation experiments. The six classes are defined in the next section. After defining the classes it will be shown that they generate all possible VRTs via composition, that they are disjoint classes, and that they are useful for achieving a variance reduction. Uniqueness of the particular partitioning of the transformations is not claimed. Trying to relate each class of transformations directly to a well-known VRT is tempting, but misses the point. A transformation redefines an experiment in a way that may be favorable; it is not necessarily a VRT nor does its use imply a variance reduction. The last two sections discuss subclasses of transformations within the original six and explain the relationship between the six classes of transformations and information for estimation of θ.

Definitions of the Transformations

Recall that VRTs attempt to increase information and/or make better use of information for estimation of θ in a simulation experiment. Three of the classes affect the amount of information, while the

remaining three concern the use of information. The six classes of transformations, along with some additional refinements that will be discussed later, are:

Amount of Information

- Distribution Replacement (DR)
- Dependence Induction (DI)
  - (independent case)
  - (dependent case)
- Sample Allocation (SA)
  - (series)
  - (replication)

Use of Information

- Equivalent Allocation (EA)
- Equivalent Information (EI)
- Auxiliary Information (AI)
  - (about θ)
  - (about Z)

Definition: The experiment set, $E^S(\Omega,\theta)$, is

$$E^S(\Omega,\theta) = \bigcup E(f_\omega,(g;R_*),h;\Omega,\theta)$$

where the union is over all $(f_\omega,(g;R_*),h)$ for a fixed context (Ω,θ) such that Axioms 1 and 2 are satisfied.

The definitions below establish classes of transformations with domain and range $E^S(\Omega,\theta)$ for fixed (Ω,θ). These transformations map a simulation experiment into another non e-equivalent (but possibly s or d-equivalent) experiment in the same experiment set. A transformation will be denoted by T, possibly with subscripts. If a transformation is defined as altering the definition of f_ω, g, or h alone, then the

remaining components are unchanged. Note that any experiment set has six classes of transformations associated with it.

In this chapter, if two distributions are not equal then they are not equal on a region of positive probability.

Transformation of the Inputs

Distribution Replacement (DR): $T_1 \in$ DR if and only if

$$T_1: \quad f_\omega(x) \rightarrow f'_\omega(x)$$

such that

$$f'_\omega(x) \neq f_\omega(x) \quad \text{for some } x$$

and

$$f'_{ik|}(c) = f_{ik|}(c) \frac{f'_{ik}}{f_{ik}} \quad \forall\, ik$$

where $f_{ik|}(c)$ is the probability distribution of X_{ik} given $X - X_{ik}$.

Dependence Induction (DI): $T_2 \in$ DI if and only if

$$T_2: \quad f_\omega(x) \rightarrow f'_\omega(x)$$

such that

$$f'_\omega(x) \neq f_\omega(x) \quad \text{for some } x$$

and

$$f'_{ik} = f_{ik} \quad \forall\, ik$$

Transformation of the Outputs

Equivalent Allocation (EA): $T_3 \in EA$ if and only if

$$T_3: \; g(x) \rightarrow g'(x)$$

such that

$$g'(x) \overset{c}{\neq} g(x) \quad \text{for some } x$$

and

$$R'_* = R_*$$

Sample Allocation (SA): $T_4 \in SA$ if and only if

$$T_4: \; g(x) \rightarrow g'(x)$$

such that

$$R'_* \neq R_*$$

and

$$\{g'_{il}(x)\} \overset{c}{=} \{g_{il}(x)\} \quad \forall\, il,\, x$$

(Recall that $\overset{c}{=}$ means equal except for coding; see Chapter 3)

ansformation of the Statistics

Equivalent Information (EI): $T_5 \in$ EI if and only if

$$T_5: \quad h(y) \rightarrow h'(y)$$

ch that

$$h'(y) \neq h(y) \quad \text{for some } y$$

d

$$(m)' = (m) \quad \forall\ m$$

Auxiliary Information (AI): $T_6 \in$ AI if and only if

$$T_6: \quad h(y) \rightarrow h'(y)$$

ch that

$$(m)' \neq (m) \quad \text{for some } m$$

d

$$h'_m(y) = h_m(y) \quad \forall\ m, y$$

Notice that the transformations for each set X, Y, and Z are rallel, and each one changes the definition of scalar random variables the sets.

)erties

Before proceeding to the main results in the next section, three ıas establishing properties of the classes DR and DI are proved. In se proofs and those in the remainder of the chapter, all probability tributions are assumed to be integrable, and all integrals are over entire domain of the variable of integration. If the distribution liscrete, then the integral would be replaced by a summation over all sible values in the domain and the proof would proceed as given. In e of the proofs there are ratios of distributions that could have o denominators for some values of the function arguments. For all h values of the arguments the ratio is also multiplied by zero. se situations will be left undefined and attention restricted to ues of the arguments where the denominators are nonzero.

ma 1: For all $T \in DR$, $T: f_\omega \rightarrow f'_\omega$

$$f_{ik|}(c) = f'_{ik|}(c) \frac{f_{ik}}{f'_{ik}} \quad \forall\, ik$$

of: by the definition of the class DR and simple algebra.

qed

ma 2: $T \in DR$, $T: f_\omega \rightarrow f'_\omega$ if and only if

$$w_1 = \frac{1}{2n} \quad w_n = \frac{3}{2n} \quad w_i = \frac{1}{n} \quad i = 2,3,\ldots,n-1$$

it is believed the dice "warm up.") In this case the estimator
ater variance

$$\mathrm{Var}[Z_{ei}] = \frac{.052}{n-2} + \frac{.130}{n^2}$$

, going from Z_{ei} back to Z is an example of an effective use of

qed

sses of Transformations

his section briefly discusses some interesting subclasses within
ix classes of transformations. These refinements are useful in
ce, and could be the subject of future research.

hen a transformation from the class DI is applied to a simulation
ment, it is most often to induce statistical dependence between
that were originally independent. See Chapter 5 for two examples.
dependence is usually induced within columns of identically
buted input sequences in X, but may also be across columns.

n the class SA it is often of practical importance to distinguish
n those transformations that yield a different number of (usually
ndent) replications, and those that alter the length of an output
ce in a single "run." Stated another way, some change the largest
of an output sequence, while others result in additional or fewer
ations of a sequence.

could be justified if the fact that $p_2 = p_4$ is known. Again using llocation m = n

$$\mathrm{Var}[Z_{ea}] = \frac{.022}{n} + \frac{.031}{n^2}$$

iary Information (AI)

Continuing to work with the (original) Z_{sa} estimator, notice that all the available information is utilized. Since $p_1 = p_2$, use the bservations in $\bar{Y}_3$, and vice versa. Thus, both $\bar{Y}_2$ and $\bar{Y}_3$ are based observations, and

$$\mathrm{Var}[Z_{ai}] = \frac{.012}{n}$$

urse Z_{ai} is biased because $\bar{Y}_2$ and $\bar{Y}_3$ are dependent. However, it is consistent.

alent Information (EI)

Recall the original estimator, Z. A statistic using equivalent mation is

$$Z_{ei} = \sum_{i=1}^{n} w_i Y_{i1}$$

$\Sigma w_i = 1$. For example, suppose

$$Y_{i2} = \begin{cases} 1, \text{ if the ith single toss is 1} \\ 0, \text{ otherwise} \qquad i = 1,2,\ldots,m \end{cases}$$

$$Y_{i3} = \begin{cases} 1, \text{ if the ith single toss is 2} \\ 0, \text{ otherwise} \qquad i = m+1,\ldots,2n \end{cases}$$

let the statistic be

$$Z_{sa} = 2\overline{Y}_2\overline{Y}_3$$

key point here is that the variance of Z_{sa} depends on how the 2n ss are allocated. In this case, the optimum allocation is to let m and

$$Var[Z_{sa}] = \frac{.031}{n} + \frac{.077}{n^2}$$

valent Allocation (EA)

Use the same approach as in illustrating SA, but now score

$$Y_{i3} = \begin{cases} \frac{1}{2}, \text{ if the ith single toss is 2} \\ \frac{1}{2}, \text{ if the ith single toss is 4} \\ 0, \text{ otherwise} \qquad i = m+1,\ldots,2n \end{cases}$$

r the altered probabilities (using EI) let

$$Z_{dr} = \frac{1}{4} Z$$

ich is an unbiased estimator of p, having variance

$$\text{Var}[Z_{dr}] \approx \frac{.011}{n}$$

pendence Induction (DI)

On any particular pair of tosses, the outcome (first, second) is ıst as likely as the outcome (7 - first, 7 - second). For instance, ıe events (2,1) and (5,6) have the same probability of occurrence. ıus, if (first, second) is rolled on toss 2i - 1, use (7-first, ·second) for toss 2i. This causes

$$\text{Cov}[Y_{2i-1,1}, Y_{2i,1}] = -p^2$$

ıd results in

$$\text{Var}[Z_{di}] \approx \frac{.049}{n}$$

ımple Allocation (SA)

Now approach the original problem a bit differently. Since $p = p_1 p_2$ use the 2n single die tosses to estimate p_1 and p_2. Let

$$Y_{i1} = \begin{cases} 1, \text{ if the sum of the ith toss is } 3 \\ 0, \text{ otherwise} \end{cases} \qquad i = 1,2,\ldots,n \tag{4.4}$$

Thus $p = Pr(Y_{i1} = 1)$. As the statistic take

$$Z = \frac{1}{n} \sum_{i=1}^{n} Y_{i1}$$

for which

$$E[Z] = p \quad \text{and} \quad Var[Z] = \frac{.052}{n}$$

For convenience later define

$$p_j = Pr(\text{toss of a single die} = j)$$

The experiment is defined by the probabilities p_j that define the working of the dice (inputs), the transformation (4.4) that gives the score (outputs), the sampling plan $R_* = \{2n\}$, and the statistic Z.

Distribution Replacement (DR)

Suppose that the working of the dice is redefined in the following way. Let

$$p_1 = p_2 = \frac{1}{3}$$

and

$$p_3 = p_4 = p_5 = p_6 = \frac{1}{12}$$

Thus the total 3 occurs four times as often, on average. To compensate

the inputs. Thus, {DR, DI} are necessary to generate those transformations that redefine the inputs. Similarly, the classes {EA, SA} and {EI, AI} are necessary. The result is then an immediate consequence of Theorem 2, which shows that these pairs of classes are all disjoint.

qed

Theorem 3: Under the loss function

$$l(Z,\theta) = (Z - \theta)^2$$

where θ is a scalar, there exists for each of the six classes a simulation experiment E that can be transformed by a transformation whose composition includes a member of that class into some E' such that

$$E_Z[l(Z',\theta)] < E_Z[l(Z,\theta)]$$

where Z and Z' are both consistent estimators of θ.

Proof: The proof is by example. The example used was originally suggested by Kahn (1956) to explain some basic VRTs. Consider the problem of estimating the probability, p, that the sum of two fair dice is 3. Clearly $p = 1/18$, but suppose that this is not known and p will be estimated by tossing dice. Toss n pairs of dice (2n single dice), or have a computer program simulate these tosses, and let

$$f'_{ik|}(c) \neq f_{ik|}(c) \quad \text{for some ik by lemma 3.}$$

which implies that $f'_{ik} \neq f_{ik}$ for some ik by the definition of DR. Thus, $T_1 \notin$ DI by the definition of DI.

Next consider $T_2 \in$ DI. Then T_2: $f_\omega \rightarrow f'_\omega$ such that

$$f'_\omega \neq f_\omega \quad \text{and} \quad f'_{ik} = f_{ik} \quad \forall\, ik$$

Suppose that

$$f'_{ik|}(c) = f_{ik|}(c) \frac{f'_{ik}}{f_{ik}} = f_{ik|}(c) \quad \forall\, ik \qquad (4.3)$$

This implies that

$$f'_\omega = f_\omega$$

by lemma 3, which is a contradiction. Thus, there exists some ik such that (4.3) does not hold. This in turn implies that $T_2 \notin$ DR.

qed

corollary 2.1: Theorem 1 does not hold if any class is eliminated from the six classes.

proof: All six classes of transformations are needed to prove Theorem 1, unless some classes contain elements that have the same effect as members of other classes. However, by definition {EA, SA} and {EI, AI} do not transform the inputs, so no composition of them will transform

$$T_7: \quad h_m(Y_{(m)},) \rightarrow h'_m(Y_{(m)},) \quad \forall\ m$$

Thus, $T_7 \in EI$ by definition of EI.

Finally, let

$$T = T_0 \circ T_1 \circ T_2 \circ T_3 \circ T_4 \circ T_5 \circ T_6 \circ T_7$$

Then, by construction

$$T: \quad E \rightarrow E'$$

and T is a composition of transformations in the six classes DR, DI, EA, SA, EI, and AI.

qed

Theorem 2: The six classes of transformations are disjoint.

Proof: For any given experiment set, the classes EA, SA, EI, and AI are mutually disjoint by their definition and the definition of simulation experiments. Also, they are clearly disjoint from DR and DI. What remains to be shown is that DR and DI do not overlap.

Consider $T_1 \in DR$. Then $T_1: \quad f_\omega \rightarrow f'_\omega$ such that

$$f'_\omega \neq f_\omega \quad \text{and} \quad f'_{ik|}(c) = f_{ik|}(c) \frac{f'_{ik}}{f_{ik}} \quad \forall\ ik$$

Now since $f'_\omega \neq f_\omega$, then

$$\{g_{i\ell}\} \quad i = 1,2,\ldots \quad \ell = 1,2,\ldots,\ell^*$$

and

$$\{g'_{i\ell}\} \quad i = 1,2,\ldots \quad \ell = 1,2,\ldots,\ell^*,$$

Let T_4 be the transformation such that

$$T_4: \ \{g_{i\ell}\} \rightarrow c\{g'_{i\ell}\}$$

$$i = 1,2,\ldots \quad \ell = 1,2,\ldots\ell^*,$$

Thus, $T_4 \in$ EA by definition of EA and the fact that $\underline{0}$ is always achievable. Now let T_5 be the transformation such that

$$T_5: \ \underline{0} \rightarrow R'_*$$

Thus, $T_5 \in$ SA by definition of SA under the representation g'.

Next consider h(Y) and h'(Y'), where $Y \overset{c}{=} Y'$. Since different coding is irrelevant, without loss of generality let

$$h' \leftarrow h' \circ c$$

where $\circ$ denotes composition. Thus, h(Y) and h'(Y) are being compared. Let T_6 be the transformation such that

$$T_6: \ (m) \rightarrow (m)' \quad \forall \ m$$

Thus, $T_6 \in$ AI by definition of AI. Now let T_7 be the transformation such that

Proof: The proof is by construction of T. Assume a definition of E and E' given by $(f_\omega,(g;R_*),h)$ and $(f'_\omega,(g';R'_*),h')$, respectively. Let T_0 be the transformation such that

$$T_0: \quad R_* \rightarrow \underline{0}$$

where $\underline{0}$ is a vector of all zeroes of the same dimension as R_*. Thus, $T_0 \in$ SA since $\underline{0}$ is always an achievable sample allocation.

Next consider $f_\omega(x)$ and $f'_\omega(x)$. Both have the same support, Ω. Let

$$\amalg f_{ik} \quad \text{and} \quad \amalg f'_{ik}$$

be the product of all scalar marginal distributions of f_ω and f'_ω, respectively. Let T_1 be the transformation such that

$$T_1: \quad f_\omega \rightarrow \amalg f_{ik}$$

Thus, $T_1 \in$ DI by definition of DI. Now let T_2 be the transformation such that

$$T_2: \quad \amalg f_{ik} \rightarrow \amalg f'_{ik}$$

Thus, $T_2 \in$ DR by definition of DR. Now let T_3 be the transformation such that

$$T_3: \quad \amalg f'_{ik} \rightarrow f'_\omega$$

Thus, $T_3 \in$ DI by definition of DI.

Next consider $(g(x);\underline{0})$ and $(g'(x);R'_*)$ defined on the same x. Consider the representations

$$f_{1,2|3,4,\ldots,n-i+1} = f_{1|2,3,\ldots,n-i+1}\, f_{2|3,4,\ldots,n-i+1}$$

$$= f_{2|1,3,4,\ldots,n+i-1}\, f_{1|3,4,\ldots,n-i+1}$$

The proof proceeds exactly as above to show that the (n-i-1)st order scalar conditionals are determined. By induction, this shows that all scalar conditionals (including the first order scalar marginal distributions) are determined. And since

$$f_{1,2,\ldots,n} = f_1\, f_{2|1}\, f_{3|1,2} \cdots f_{n|1,2,\ldots,n-1}$$

then $f_{1,2,\ldots,n}$ is determined.

qed

Main Results

In this section the main results of this research are proved; namely that for a given experiment set $E^S(\Omega,\theta)$ the associated six classes of transformations generate all possible VRTs via composition, they are disjoint classes, and they are useful. Remember that these transformations are from any e-equivalent class of experiments to any other e-equivalent class in $E^S(\Omega,\theta)$. In the results that follow, E and E' are both elements of $E^S(\Omega,\theta)$.

Theorem 1: Given $E \overset{e}{\not\sim} E'$, there exists a transformation T: E -> E', and T is a composition of members of the six classes of transformations.

$$\frac{f_{2|3,4,\ldots,n}}{f_{1|3,4,\ldots,n}} = \frac{f_{2|1,3,\ldots,n}}{f_{1|2,3,\ldots,n}} = r(x) \qquad (4.2)$$

for all $x = (x_1,\ldots,x_n)$. It is easy to show that if the denominator of (4.2) is zero for some x, then the numerator is as well; in fact so is $f_{1,2,\ldots,n}$. Such values of x are not of interest and the value of $r(x)$ will be left undefined there, restricting attention to values of x where this is not the case.

Now for any fixed value of $(x_3,x_4,\ldots,x_n)$ equation (4.2) gives

$$1 = \int f_{2|3,4,\ldots,n} dx_2 = f_{1|3,4,\ldots,n} \int r(x)dx_2$$

which implies that

$$f_{1|3,4,\ldots,n} = \frac{1}{\int r(x)dx_2}$$

By the argument above, $r(x) > 0$ in the regions of interest, so the density exists. Since $r(x)$ is given, $f_{1|3,4,\ldots,n}$ is determined. Using a similar argument, it can be shown that all (n-2)nd order scalar conditionals

$$f_{j|1,\ldots,j-1,j+1,\ldots,n} \qquad j = 1,2,\ldots,n$$

are determined by the (n-1)st order scalar conditionals.

The remainder of the proof proceeds by induction. Assume that the (n-i)th order scalar conditionals are determined for some $i = 2,3,\ldots,n$. For example, $f_{1|2,3,\ldots,n-i+1}$ and $f_{2|1,3,4,\ldots,n-i+1}$. Then write

The condition is not necessary if, for instance, $f_{(c)|ik}$ is constant over those regions where f_{ik} and f'_{ik} differ.

The next result characterizes multivariate distributions in terms of conditional distributions. The notation

$$f_{1,2,\ldots,n} = f_{1,2,\ldots,n}(x_1,\ldots,x_n)$$

will be used for the n variate density and/or mass function, and the usual notation for conditional distributions will also be used. For example

$$f_{1|2} = f_{1|2}(x_1|x_2)$$

lemma 3: The distribution $f_{1,2,\ldots,n}$ is determined (up to a set of Lebesgue measure zero) by

$$f_{1|2,3,\ldots,n}\ ,\ f_{2|1,3,4,\ldots,n}\ ,\ldots,\ f_{n|1,2,\ldots,n-1}$$

the (n-1)st order scalar conditional distributions.

proof: From elementary properties of conditional distributions

$$f_{1,2|3,4,\ldots,n} = f_{1|2,3,\ldots,n}\, f_{2|3,4,\ldots,n}$$

$$= f_{2|1,3,\ldots,n}\, f_{1|3,4,\ldots,n}$$

After some rearrangement

$$f'_{(c)|ik} = f_{(c)|ik} \frac{f'_{(c)}}{f_{(c)}} \quad \forall\ ik \tag{4.1}$$

proof: Let $T \in DR$. For any fixed ik

$$\frac{f'_{\omega}}{f'_{(c)}} = f'_{ik|(c)} = f_{ik|(c)} \frac{f'_{ik}}{f_{ik}} \quad \text{(since } T \in DR\text{)}$$

$$= \frac{f_{\omega}}{f_{(c)}} \frac{f'_{ik}}{f_{ik}} = f_{(c)|ik} \frac{f'_{ik}}{f_{(c)}}$$

$$\rightarrow \quad \frac{f'_{\omega}}{f'_{ik}} = f_{(c)|ik} \frac{f'_{(c)}}{f_{(c)}}$$

$$\rightarrow \quad f'_{(c)|ik} = f_{(c)|ik} \frac{f'_{(c)}}{f_{(c)}}$$

A parallel argument shows that (4.1) implies the definition of DR.

qed

Lemmas 1 and 2 are alternative definitions of the class DR. A sufficient condition for there to exist a transformation in DR that transforms f_{ik} into $f'_{ik} \neq f_{ik}$ is that X_{ik} and $X - X_{ik}$ are independent.

Transformations in the AI class are characterized by altering the argument set of the function h. In practice, members of AI most often recruit additional outputs, thus providing more information for estimation of θ. Restricting attention to those transformations in AI that augment the original argument set, it is possible to identify two important subclasses: those that yield more information about θ (AI.θ), and those that recruit outputs containing information about Z (AI.Z).

In Chapter 3 the concept of information about a parameter contained in a random variable was discussed. The idea can be extended quite naturally to information contained in one random variable about realizations of another. In particular, if the two random variables are independent, then looking at the realization of one reveals nothing about the other. However, if the two are statistically dependent, then the realization of one may reveal something about the likelihood of the particular realization of the other. Thus, uncertainty about the likelihood of the realization of a statistic may be reduced, and the estimate may be modified based on this knowledge. In Chapter 5 a VRT (control variates) that makes use of this type of information is discussed. Note that AI.θ and AI.Z need not be disjoint, but that members of AI - AI.θ - AI.Z will never be effective for reducing variance.

The Transformations and Information

Throughout the development of this research, the concept of statistical information and its usefulness in the discussion of variance reduction has been stressed. The six classes of transformations are a

particularly useful breakdown of the possible transformations because of the close association of each class with either the idea of increasing the available information or making better use of information. As noted briefly above, DR, DI, and SA transformations can increase information in the simulation experiment. This is because f_ω determines the information content of the initial sample (X) and R_* controls how the sampling effort is allocated. On the other hand, g and h determine the information loss via transformation of the intial sample. Transformations in the EA, EI, and AI classes can reduce the loss.

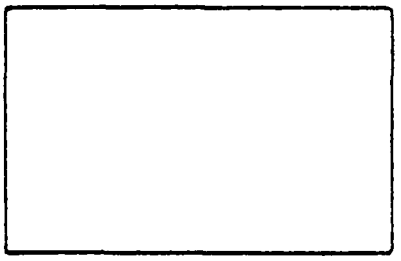
It should be noted that effective members of the EI class of transformations have been extensively studied in the classical statistical literature on optimality of estimators. Concepts such as sufficiency and minimum variance estimators are results of this work. Estimators using auxiliary information have also been studied in the same manner.

FIVE WELL-KNOWN VRTS

In this chapter five of the more well-known VRTs are reviewed in light of the framework developed in Chapters 3 and 4. The VRTs considered are antithetic variates, common random numbers, control variates, stratified sampling, and use of conditional expectations. For each a brief review of the literature is presented, along with a description of the VRT and graphical display of how the VRT is composed of members of the six classes of transformations. The purpose is not to propose a precise definition of these five VRTs, since the same names are used for several variations. Rather, an attempt is made to present the most widely accepted version of each technique. First, a symbol set is given to be used for graphical presentation.

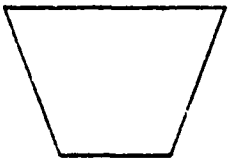
Symbol Set

Only three basic symbols are needed (see Figure 5.1). A rectangle will enclose a definition of an input, output, or statistic in the simulation experiment. A circle denotes a class of transformations, and a trapezoid some prior knowledge used to make the application of the transformation reasonable. The progression is from left to right, proceeding from a definition of some input, output, or statistic to a new definition via a transformation.

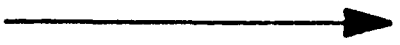


definition of random variables

class of transformations



prior knowledge



progression

Figure 5.1 Symbol Set for VRTs

Recall that VRTs are composed of members of the six classes of transformations combined with prior knowledge; given a specific problem they can be implemented as algorithms. The presentation in this chapter is on the VRT level.

Antithetic Variates (AV)

Antithetic Variates is a VRT that has been extensively studied in the context of Monte Carlo estimation. The technique was invented by Hammersley and Morton (1956), with further developments by Hammersley and Mauldon (1956), Morton (1956), Halton and Handscomb (1957), and Handscomb (1958). In its broadest sense, "we use the term antithetic variates to describe any set of estimators which mutually compensate each other's variations." (Hammersley and Handscomb, 1964, p. 61) Statistical results such as

$$\mathrm{Var}(Y_i \pm Y_j) = \mathrm{Var}(Y_i) + \mathrm{Var}(Y_j) \pm 2\mathrm{Cov}(Y_i, Y_j) \qquad (5.1)$$

make clear the advantage of random variables being correlated. Correlation may be inherent in the outputs of a simulation experiment. However, AV attempts to force a correlation structure onto the observations while preserving their marginal distributions. The correlation is usually accomplished by making them analytically dependent. When an estimator consists of a sum of n random observations, the antithetic-variates theorem (Hammersley and Mauldon, 1956, Handscomb, 1958, Wilson, 1979, 1982a, 1983c) shows that under fairly general conditions the greatest lower bound of the variance of the estimator can be approached arbitrarily closely by generating all n

observations from a deterministic transformation of one random observation. Texts that discuss AV in both the Monte Carlo and simulation contexts include: Tocher (1963), Hammersley and Handscomb (1964), Shreider (1964, 1966), Mize and Cox (1968), Meier, Newell and Pazer (1969), Fishman (1973, 1978), Gaver and Thompson (1973), Hillier and Lieberman (1974), Kleijnen (1974), Carter and Cashwell(1975), Yakowitz (1977), Pritsker and Pegden (1979), Rubinstein (1981), Kohlas (1982), Law and Kelton (1982), and Payne (1982). Research into the application of AV in the simulation of stochastic networks has been done by Burt, Gaver, and Perlas (1970), Burt and Garman (1971a, 1971b), Sullivan, Hayya, and Schual (1982), Carson (1983), Grant (1983, 1980), and Grant and Solberg (1983); Kumamoto, Tanaka, Inoue and Henley (1980a) investigate Monte Carlo evaluation of fault trees. Moy (1965, 1971), Page (1965), Gaver (1969), and Mitchell (1973) use AV for simulating queuing systems. George (1977) looks at simulating replacement processes with AV, while Fishman (1981, 1982a, 1982b, 1983a) deals with simulation of Markov chains and processes. Issues relating to the generation of antithetic observations in various contexts are discussed in Fishman (1972, 1974), Franta (1975), and Cheng (1981, 1982, 1983a, 1983b). Combining AV and other VRTs is examined in Fishman (1974), Gentle (1975), Kleijnen (1974, 1975); Schruben (1978, 1979), Schruben and Margolin (1978), and Cooley and Houck (1982) incorporate AV and CRN (see below) into the design of simulation experiments. Roach and Wright (1977) correctly state that systematic sampling (SYS) plans are a subset of general AV sampling plans; see Madow and Madow (1944) for a reference on SYS. Other papers of interest are Deutsch and Schmeiser (1980),

Fishman (1979, 1968), McGrath and Irving (1973a, 1974), Simon (1976), Lavenberg and Welch (1978), Halton (1979), Rubinstein and Samordnitsky (1982), Rubinstein, Samordnitsky, and Shaked (1982), and Wilson (1982, 1983a, 1983b, 1983d).

Consider estimating θ_1 using a simulation experiment defining

$$Y_{i1} = g_{i1}(X_{(i1)}) \quad i = 1,2,\ldots,I_1 = 2n$$

where $E(Y_{i1}) = \theta_1$, with statistic

$$Z_1 = \frac{1}{I_1} \sum_{i=1}^{I_1} Y_{i1}$$

Further, suppose that

$$X_{(i1)} \sim \text{i.i.d. } f_{(1)}(X_{(i1)}). \tag{5.2}$$

The usual AV transformation is to redefine the joint distribution of

$$(X_{(2i-1,1)}, X_{(2i,1)}) \quad i=1,2,\ldots,n$$

such that they still have the marginals given in (5.2), but the pairs are negatively correlated in some way. When $X_{(i1)}$ is a scalar, or if AV is only used on a scalar component of $X_{(i1)}$, the correlation is most often induced by generating realizations via the inverse cumulative distribution function (cdf) of $X_{(i1)}$ in the following manner:

$$X_{(2i-1,1)} = F^{-1}{}_{(1)}(U_i)$$

$$X_{(2i,1)} = F^{-1}{}_{(1)}(1-U_i)$$

where $U_i \sim U(0,1)$ $i = 1,2,\ldots,n$. When $X_{(i1)}$ is multivariate, there are a variety of approaches and objectives. The reason for redefining the inputs to be dependent is to cause

$$Cov(Y_{2i-1,1}, Y_{2i,1}) < 0$$

reducing the variance of Z_1 via (5.1). Figure 5.2 shows how AV employs the DI class of transformations.

Common Random Numbers (CRN)

Common random numbers is often called "correlated sampling" (CS). There is some confusion because CRN is both a method for generating correlated samples and a VRT that exploits induced correlation. "The name of the technique stems from the possibility in some situations of using the same stream of basic U(0,1) random variables to drive each of the alternative models through time..." (Law and Kelton, 1982, p.350). Here, the term CRN is used in the sense of CS, meaning that correlation is induced (by whatever means) between certain inputs to obtain positively correlated outputs, and the interest is in estimating a parameter that can be expressed as a difference.

CRN has the distinction of being "...the only VRT that is as a rule used by practitioners of simulation" (Kleijnen, 1974, p. 206). Papers and books discussing various aspects of CRN include Kahn and Marshall (1953), Jessop (1956), Conway (1963), Fishman (1968, 1974), Ignall (1972), McGrath and Irving (1973a, 1974), Becker (1974), Kleijnen (1974,

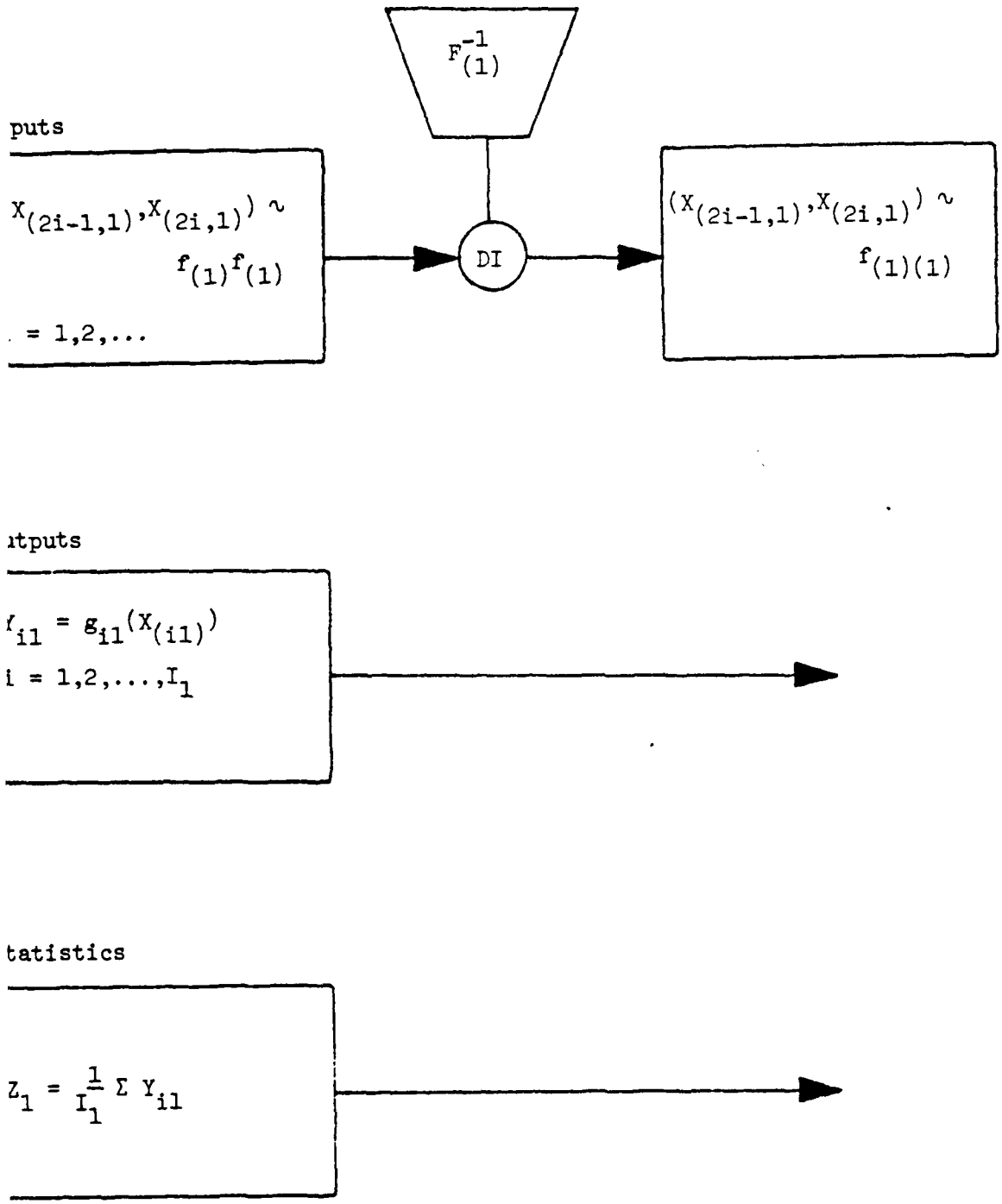


Figure 5.2 Antithetic Variates (AV)

75), Lavenberg and Welch (1978), Heidelberger and Iglehart (1978), itsker and Pegden (1979), Rubinstein (1981), Gal, Rubinstein, and Ziv 981), Wilson (1982a, 1982b, 1983a, 1983b), Law and Kelton (1982), hlas (1982), Banks and Carson (1983), and Bratley, Fox and Schrage 983). Gentle (1975) calls the technique control variates. Mihram 974), Heikes, Montgomery, and Rardin (1976), Schruben (1978, 1979), hruben and Margolin (1978), and Cooley and Houck (1982) investigate corporating CRN into the design of the simulation experiment as a ndom block effect. The last three papers consider incorporating AV th CRN, as do Fishman (1974) and Kleijnen (1974, 1975). Wright and msey (1979) give a simple example of an inventory simulation where sing CRN gives counterintuitive results.

Consider estimating

$$\theta_1 = \alpha_2 - \alpha_3$$

sing a simulation experiment defining

$$Y_{i\ell} = g_{i\ell}(X_{(i\ell)}) \quad i = 1,2,\ldots,I_\ell \quad \ell = 2,3$$

here $E(Y_{i\ell}) = \alpha_\ell$, with statistic

$$Z_1 = \frac{1}{I_2}\sum_{i=1}^{I_2} Y_{i2} - \frac{1}{I_3}\sum_{i=1}^{I_3} Y_{i3}$$

$$= \bar{Y}_2 - \bar{Y}_3$$

or convenience assume that $I_2 = I_3$. The basis for CRN is the well-

ı relation

$$\mathrm{Var}(\bar{Y}_2 - \bar{Y}_3) = \mathrm{Var}(\bar{Y}_2) + \mathrm{Var}(\bar{Y}_3) - 2\mathrm{Cov}(\bar{Y}_2, \bar{Y}_3)$$

ıdefining the joint distribution of $(X_{(i2)}, X_{(i3)})$ $i = 1,2,\ldots,I_1$ that they are positively correlated--without redefining their ınal distributions--it is hoped that $\mathrm{Cov}(\bar{Y}_2, \bar{Y}_3) > 0$, reducing the ance of Z_1. Several of the references cited discuss conditions and ods that insure a favorable covariance term. See Figure 5.3 for a ıical presentation of this VRT.

rol Variates (CV)

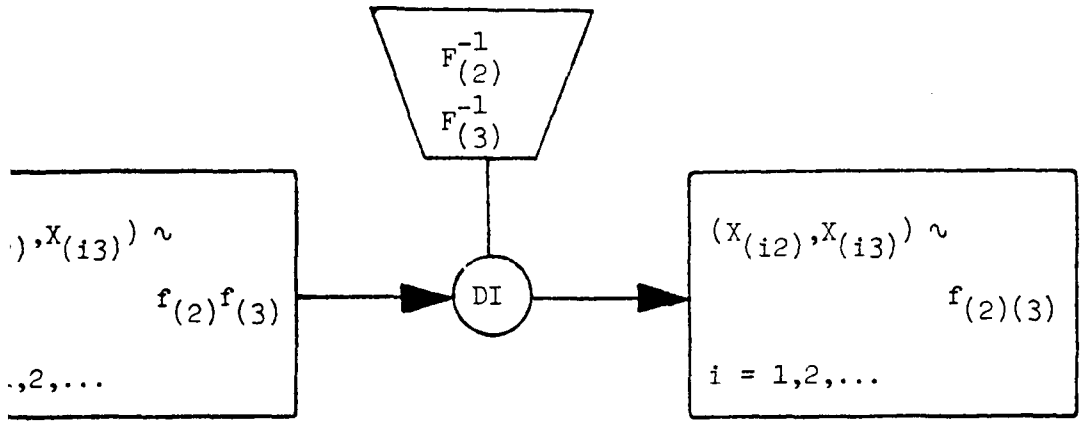
The term control variates has a variety of meanings. Here, it will used to describe a class of statistics that attempt to correct the e of an estimator based on the discrepancy between the value of a nd estimator and the known value of its expectation. For example, $Y_{(1)}$ and $Y_{(2)}$ be sets of output random variables in a simulation riment, and s_1 and s_2 be scalar valued functions such that

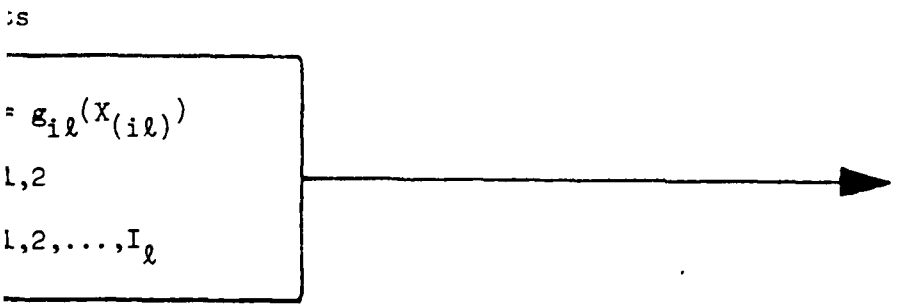
$$E[s_1(Y_{(1)})] = \theta_1 \qquad E[s_2(Y_{(2)})] = \alpha$$

e θ_1 and α are real scalars; θ_1 is the parameter of interest and α nown. The two most common CV estimators are the linear control

$$Z_c = s_1(Y_{(1)}) - b(s_2(Y_{(2)}) - \alpha) \tag{5.3}$$

e b is a constant, and the ratio estimator





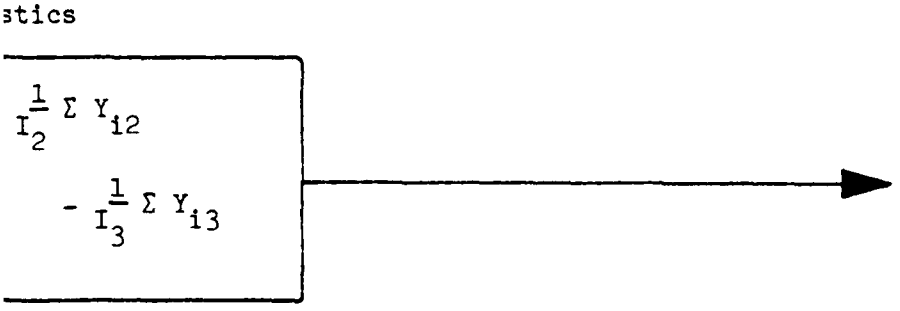


Figure 5.3 Common Random Numbers (CRN)

$$Z_c = \frac{s_1(Y_{(1)})}{s_2(Y_{(2)})}\,\alpha \qquad (5.4)$$

ction s_2 is the control variate. Both (5.3) and (5.4) extend
ly to multiple control variates, but that extension will not be
red here. Also not discussed is the determination of the
er b, except to cite several references. In the simulation
ure, a distinction is made between "internal" control variates
variables that are part of the same real or conceptual system)
ternal" control variates (random variables that are part of a
real or conceptual system). This distinction is not relevant
ince both types are simply functions of outputs in the simulation
ent. However, external control variates employ a transformation
e class DI, while internal CV does not.

xtbooks providing general descriptions of CV for Monte Carlo
ion are Hammersley and Handscomb (1964), Shreider (1964, 1966),
z (1977), and Rubinstein (1981). Concentrating more specifically
ulation are Tocher (1963), Fishman (1973, 1978), Gaver and
n (1973), Kleijnen (1974), Pritsker and Pegden (1979), Law and
(1982), and Bratley, Fox, and Schrage (1983); see Cochran (1977)
cussion of CV in survey sampling. Use of CV in the simulation of
tic networks is investigated in Burt, Gaver and Perlas (1970),
d Garman (1971), Grant (1980) and Grant and Solberg (1983), while
o, Tanaka, and Inoue (1977) apply CV to fault tree analysis.
rg, Moeller, and Welch (1977a, 1977b, 1978, 1982), Taaffe and
1983), Wilson (1979b, 1983e, 1983f), Wilson and Pritsker (1982)
th CV in the simulation of queuing systems. The selection or

ın of the CV multipliers under various assumptions for linear like (5.3) has received considerable attention; see for Cheng (1978), Hopmans and Kleijnen (1980) and Koehler (1981). ınd Beale (1983) discuss verification of the hypothesis that the ıhip with the CV is indeed linear. Olkin (1958), Matern (1962), 977), and Isaki (1983) treat CV in the survey sampling context, (1956), Swain (1981, 1982), and Swain and Schmeiser (1983) ıte on Monte Carlo problems. Iglehart and Lewis (1979) consider the general context of regenerative simulations. Initial ıons of CV in simulations used a single, linear control. In an t survey paper, Lavenberg and Welch (1981) summarize results on ıltiple linear control variates. Rubinstein and Markus (1982) ıese results to estimation of multiple parameters with multiple , and Nozari, Arnold, and Pegden (1983) extend them to ılation simulation experiments. Other papers of interest are: i and Ben Tuvia (1962), Moy (1965, 1971), Gaver (1969), Gaver ler (1971), McGrath and Irving (1973a, 1974), Gentle (1975), g and Welch (1978), Cheng and Feast (1980) and Wilson (1982b, 983b).

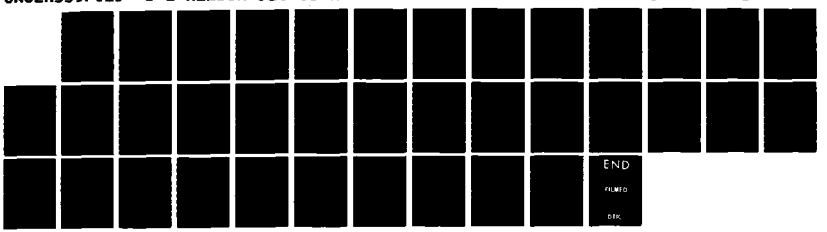
ore describing the general characterization of CV an expression variance of a function of two random variables will be derived. nd Q_2 be two real, scalar valued random variables with finite Let

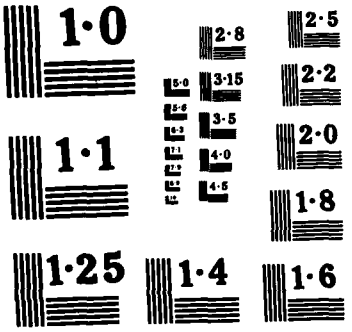
$$E[Q_i] = \gamma_i \quad i = 1,2$$

a function $h(Q_1,Q_2)$ that is analytic at (γ_1,γ_2) for all values

AD-A158 146 VARIANCE REDUCTION IN SIMULATION EXPERIMENTS: A MATHEMATICAL-STATISTICAL FRAMEWORK(U) PURDUE UNIV LAFAYETTE IN SCHOOL OF INDUSTRIAL ENGINEERING 2/2

UNCLASSIFIED B L NELSON DEC 83 N00014-79-C-0832 F/G 12/1 NL





of Q_1 and Q_2. Then to two terms, the Taylor series expansion of h about this point is

$$h(Q_1,Q_2) = h(\gamma_1,\gamma_2) + \sum_{i=1}^{2} \frac{\partial h(\gamma_1,\gamma_2)}{\partial Q_i} (Q_i - \gamma_i) + R_2 \qquad (5.5)$$

where R_2 is a remainder term given by the next term in the Taylor expansion with h evaluated at an unknown point between γ_i and Q_i $i = 1,2$ for Q_1 and Q_2, respectively. Ignore this term for the moment. Using $E[Q_i] = \gamma_i$, (5.5) implies that

$$E[h(Q_1,Q_2)] \approx h(\gamma_1,\gamma_2)$$

and

$$E[h^2(Q_1,Q_2)] \approx h^2(\gamma_1,\gamma_2) +$$

$$\sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\partial h(\gamma_1,\gamma_2)}{\partial Q_i} \frac{\partial h(\gamma_1,\gamma_2)}{\partial Q_j} \mathrm{Cov}[Q_i,Q_j]$$

Combining these two gives

$$\mathrm{Var}[h(Q_1,Q_2)] \approx \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\partial h(\gamma_1,\gamma_2)}{\partial Q_i} \frac{\partial h(\gamma_1,\gamma_2)}{\partial Q_j} \mathrm{Cov}[Q_i,Q_j] \qquad (5.6)$$

Interestingly, to make (5.6) an equality only requires adding the term $\mathrm{Var}(R_2)$ to the right hand side.

Now, returning to the description given at the beginning of this section, the CV estimator will be of the form

$$Z_c = h(s_1(Y_{(1)}), s_2(Y_{(2)}))$$

with the restriction that $h(\theta_1, \alpha) = \theta_1$. Assume that h is analytic. It is clear that (5.3) and (5.4) are of this general form. Several authors have noted that these two estimators are in the same general class, including Kleijnen (1974) and Olkin (1983).

Result:

$$\mathrm{Var}(Z_c) \approx \left[\frac{\partial h}{\partial s_1}\right]^2 \mathrm{Var}(s_1) + \left[\frac{\partial h}{\partial s_2}\right]^2 \mathrm{Var}(s_2) + 2\,\frac{\partial h}{\partial s_1}\,\frac{\partial h}{\partial s_2}\,\mathrm{Cov}[s_1, s_2]$$

where all the partial derivatives are evaluated at (θ_1, α).

Proof: In (5.6), identify s_i with Q_i, θ_1 with γ_1, and α with γ_2.

qed

Note that for the linear control, $R_2 = 0$. The result shows that nonzero covariance between the estimator of the parameter of interest and the control variate is usually necessary for CV to be effective. The estimator s_2 contains information about s_1, in the sense that uncertainty about the expected value of s_1 is reduced by knowledge of s_2. The covariance term represents this information. See Figure 5.4 for a description of how CV combines transformations from the EI and AI classes.

Stratified Sampling (STRAT)

Books discussing stratified sampling in Monte Carlo problems include Hammersley and Handscomb (1964), Shreider (1964, 1966) and Rubinstein (1981). In the context of survey sampling, see Cochran

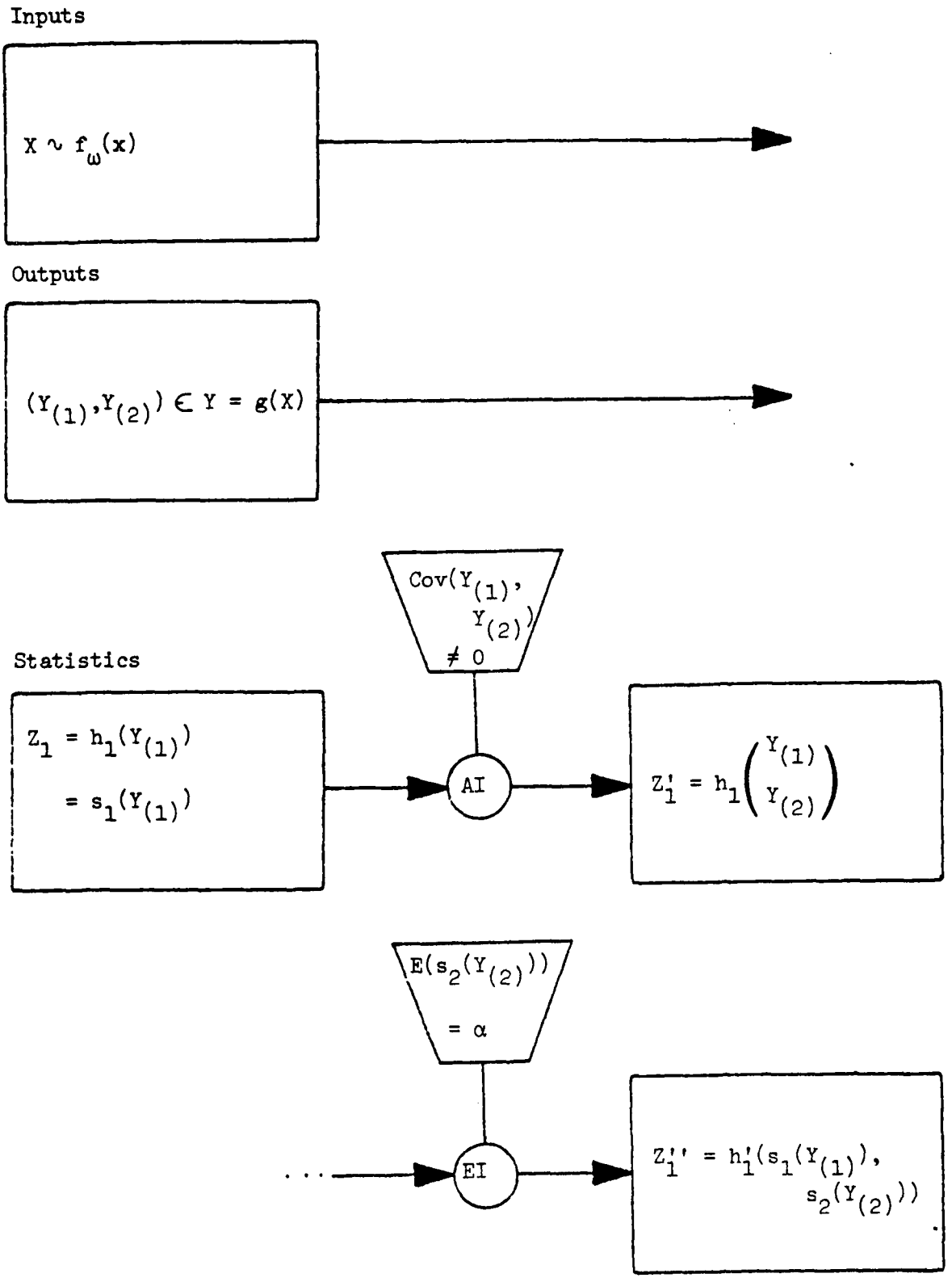


Figure 5.4 Control Variates (CV)

(1977). Some books or chapters dealing with stratified sampling specifically in systems simulation are Moy (1971), Kleijnen (1974), Hillier and Lieberman (1974), Pritsker and Pegden (1979), Payne (1982), and Bratley, Fox and Schrage (1983). Methods for setting stata boundaries are examined in Delanius (1950), Delanius and Gurney (1951), Sethi (1963), and Singh (1975a, 1975b, 1977). Papers of general interest in simulation, Monte Carlo, and survey sampling contexts include Ehrenfeld and Ben-Tuvia (1962), Moy (1965), Sardnal (1968), Burt and Garman (1971a), Bayes (1972), McGrath and Irving (1973a, 1974), Gentle (1975), Hartley (1977), and Wilson (1979b, 1982b, 1983a, 1983b). Kahn (1950a, 1950b) and Steinberg (1963) discuss it under the name "quota sampling." For interesting application papers see Surkis, Gordon, and Hynes (1975), Gordon and Hynes (1978), and Diegert and Diegert (1981). DeGroot and Starr (1969) look at the problem from a Bayesian viewpoint; the stratum means and proportion of the population in each stratum have prior distributions.

Consider estimating θ_1 when it is possible to sample I_1 observations of Y_{i1}, where $E(Y_{i1}) = \theta_1$, $i = 1,2,\ldots,I_1$. Thus, the crude estimator of θ_1 might be

$$Z_1 = h_1(Y_{(1)})$$

$$= \frac{1}{I_1} \sum_{i=1}^{I_1} Y_{i1} .$$

Now suppose Y_{il} can be expressed as a transformation of $(X_{(i1)}, X_{ik})$ for

some fixed column index k of X. For convenience write

$$Y_{i1} = g_{i1}(X_{ik}) \tag{5.7}$$

suppressing the $X_{(i1)}$. Assume that X_{ik} are i.i.d. random variables for all i, and that the range of X_{ik} can be divided into n nonoverlapping, exhaustive intervals (strata). Denote these strata by L_j , j = 2,...,n+1. An equivalent way to view (5.7) is

$$Y_{ij} = g_{ij}(X_k) \quad j = 2,\ldots,n+1 \quad i = 1,\ldots,I_j$$

such that Y_{ij} is the i^{th} observation of Y_1 when $X_{mk} \in L_j$, and

$$\sum_{j=2}^{n+1} I_j = I_1$$

Now, if $p_j = P(X_{mk} \in L_j)$ is known for all j, and if the values of I_j , j = 2,...,n+1 can be fixed arbitrarily, then the STRAT estimator

$$Z'_1 = h'_1(Y_{(1)})$$

$$= \sum_{j=2}^{n+1} p_j \frac{1}{I'_j} \sum_{i=1}^{I'_j} Y_{ij}$$

may have smaller variance that Z_1, depending on the allocation I'_j. Allocation strategies will not be discussed here (see for instance, Cochran, 1977), but if <u>proportional allocation</u> ($I_j = I_1 p_j$) is used then

$$\mathrm{Var}(Z'_1) \leq \mathrm{Var}(Z_1)$$

(Rubinstein, 1981). Note that if I_j is not altered then the VRT is known as poststratified sampling. See Figure 5.5.

Conditional Expectations (CE)

Conditional expectations is often called "conditional Monte Carlo" (CMC). However, CMC is a sampling technique originally developed by Trotter and Tukey (1956) to "use a family of transformations to convert given samples into samples conditioned on a given characteristic (p. 64)." The original CMC was not inherently a variance reduction technique, although when used as one it is most closely akin to importance sampling (Dubi and Horowitz, 1979). Here, the term conditional Monte Carlo is reserved for the original sampling technique. Other references include Arnold, Bucher, Trotter, and Tukey (1956), Hammersley (1956), Wendel (1957), Hammersley and Handscomb (1964), Granovsky (1981), Rubinstein (1981), and Wilson (1983b).

The use of conditional expectations (CE) will be described as the term is used by Law and Kelton (1982). Fishman (1973) and Pritsker and Pegden (1979) refer to it as use of "prior information"; Carter and Ignall (1970, 1975) use the term "virtual measures." Brown, Solomon, and Stephens (1979) use CE for estimating the expected number of renewals in $[0,t]$ for a renewal process, while Andrews, Bickel, Hampel, Huber, Rogers and Tukey (1972) employ conditioning in Monte Carlo estimation of location parameters. Lavenberg and Welch (1979) surveys applications of CE. Other papers of interest are Kahn (1950, 1956), Kahn and Marshall

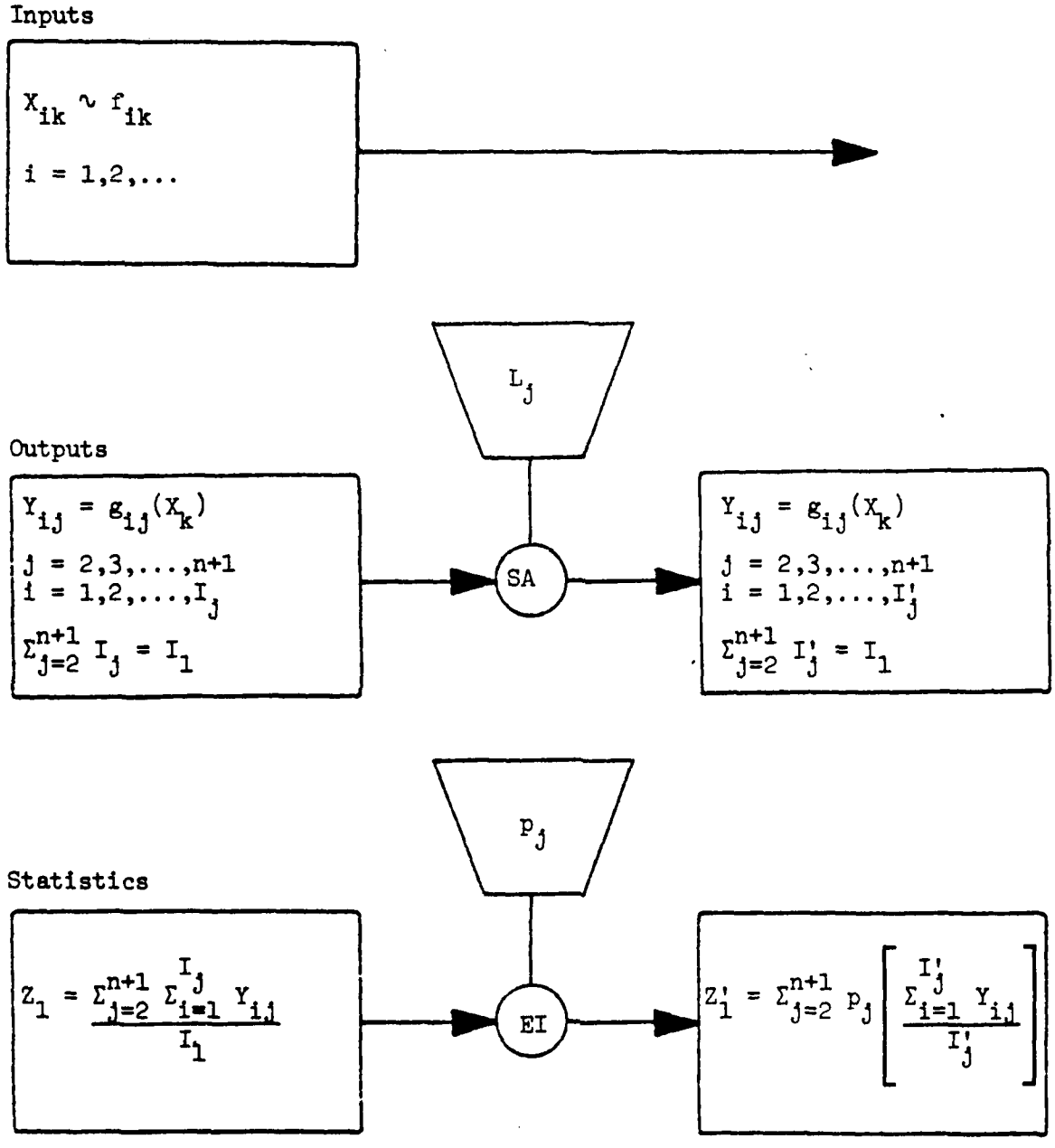


Figure 5.5 Stratified Sampling (STRAT)

(1953), McGrath and Irving (1973a, 1974), Gross (1973), Simon (1976), Lavenberg and Welch (1978), and Wilson (1982, 1983a, 1983b). The latter paper by Wilson compares CMC and CE. Fox (1983) establishes conditions that guarantee effectiveness of CE when based on correlated outputs; see also Bratley, Fox and Schrage (1983).

Consider estimating θ_1 using a simulation experiment defining

$$Y_{i1} = g_{i1}(X_{(i1)}) \quad i = 1,2,\ldots,I_1$$

where $E(Y_{i1}) = \theta_1$, with statistic

$$Z_1 = h_1(Y_{(1)}) = \frac{1}{I_1} \sum_{i=1}^{I_1} Y_{i1} \tag{5.8}$$

However, suppose there is another output random variable Y_{i2} $i = 1,2,\ldots,I_2$ and

$$E[Y_1|Y_{i2} = y_{i2}]$$

can be calculated for all realizations y_{i2} of Y_{i2}. Here Y_1 is generic for any of Y_{i1}. Based on the well-known relation

$$Var[E[Y_1|Y_{i2}]] = Var[Y_1] - E[Var[Y_1|Y_{i2}]]$$

use the estimator

$$h'_1(Y_{(1)}') = \frac{1}{I_2} \sum_{i=1}^{I_2} E[Y_1|Y_{i2}] \tag{5.9}$$

The estimator (5.9) is unbiased for θ_1, and if the Y_{i2} are independent then it has no greater variance than (5.8) if $I_1 = I_2$. See Fox (1983)

for more general conditions. However, CE is often employed when $I_2 > I_1$, such as when Y_{i1} are results of "rare events." Clearly the estimator (5.9) may be based on a vector of outputs, not just a scalar Y_{i2}. Note that Y_{i1} has not been redefined, but rather other outputs in the simulation experiment are used. CE is shown in Figure 5.6.

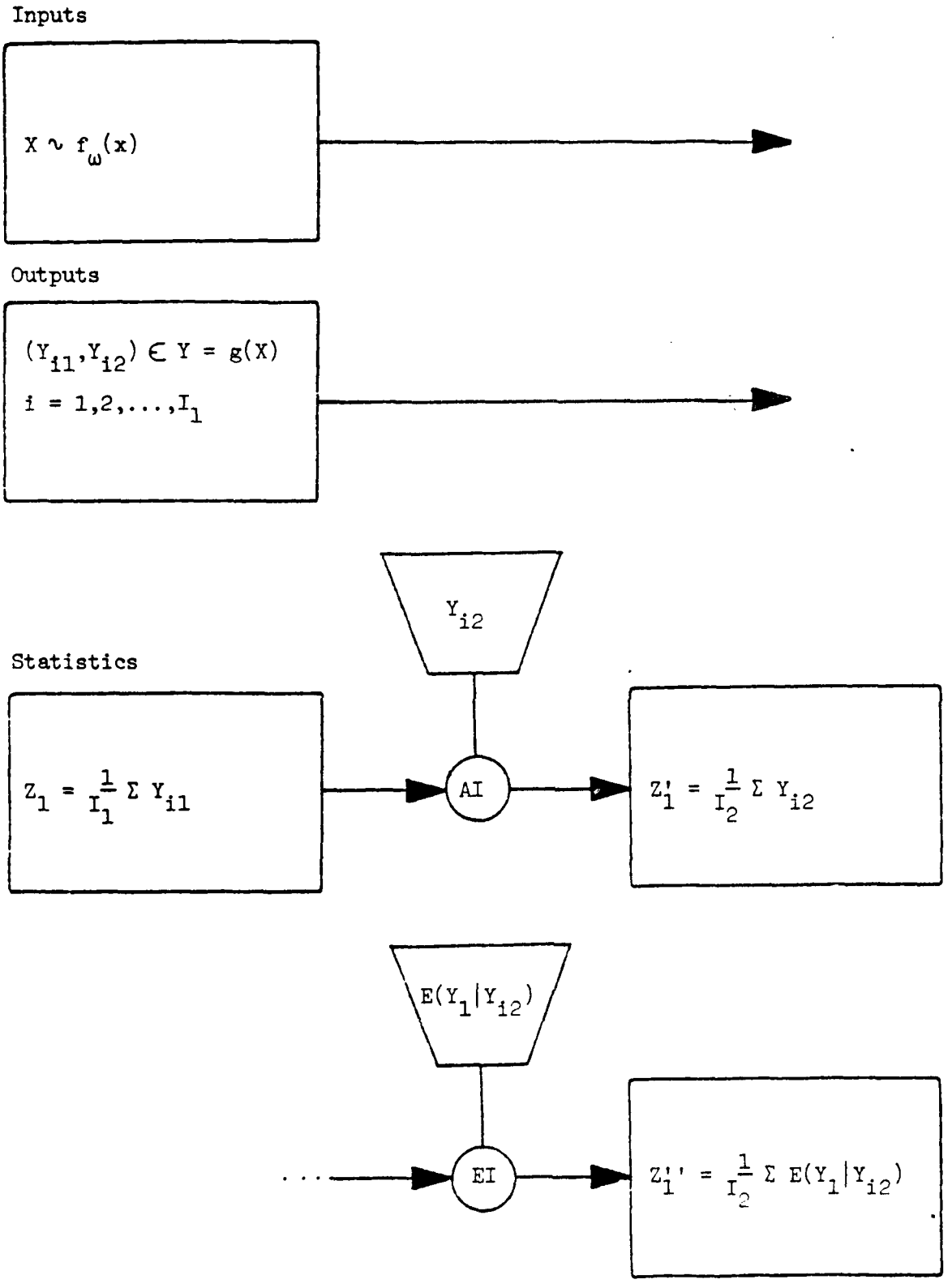


Figure 5.6 Conditional Expectations (CE)

CONCLUSIONS

Although the research presented here may seem remote from actual problems, it should have ramifications on practical simulation experiments. Conversations with practitioners indicate that, with the exception of the most simple applications of AV and CRN, variance reduction techniques are seldom employed. This is due partly to a lack of knowledge and understanding; to the casual student of simulation, variance reduction appears to be a collection of special purpose techniques that need to be rederived for each application. The existence of an underlying theory and a small number of elemental transformations provides structure not previously available. This structure should facilitate coherent instruction in variance reduction and also provide common ground for reporting applications that take place. The distinction between transformations, VRTs, and algorithms is central: 1) The six classes generate all of variance reduction, in a sense providing a checklist of possible approaches to take based on what prior knowledge is available. 2) Graphical presentation of VRTs (Chapter 5) provides a clear description of general VRTs, what knowledge is commonly needed to "make them go," and yet does not relate them to a particular application. 3) Algorithms, the problem specific part, do not seem as ad hoc when they are just examples of general methods and even

more basic transformations. The definition of simulation experiments (Chapter 3) not only provides the structure needed to prove the results of Chapter 4, but also structures the thinking of the experimenter, helping him to recognize prior knowledge that can be exploited.

The research presented here depends heavily on the usefulness and validity of the definition of simulation experiments; this definition is consistent with the general idea of an experiment in probability and statistics, and appears to be broad enough to encompass Monte Carlo experiments and survey sampling. We suspect that the characterization covers any "sampling experiment," but do not know how to prove such a conjecture. The terms and definitions used are well-known statistical objects: sample spaces, events, probability distributions, and transformations. The formal definition permits investigation of issues such as experimental design, restrictions on sampling, efficient estimation techniques, and the trade off between variance and bias. An unexpected bonus is that numerical techniques and analytic solutions are special cases. Finally, the axioms are few and reasonable: the experiment is relevant to the estimation problem, and the experiment can be performed.

Other approaches could have been taken. The definition ignores issues of model validity, implementation of computer algorithms, and numerical limitations of the computer. It is often useful (as a modelling perspective) to view a computer simulation as a stochastic process, but our definition does not do so. In variance reduction we are concerned with statistical properties of estimators in a sampling experiment, and we capture that aspect.

ıvenberg, S., T. Moeller and P. Welch (1977a), "The Application of Control Variables to the Simulation of Closed Queuing Networks," 1977 Winter Simulation Conference Proceedings, 152-155.

ıvenberg, S., T. Moeller and P. Welch (1977b), "Control Variables Applied to the Simulation of Queuing Models of Computer Systems," IBM Research Report RC 6630.

ıw, A.M. (1975), "Efficient Estimators for Simulated Queuing Systems," Mgt. Sci., 22, 1, 30-41.

ıw, A.M. and W.D. Kelton (1982), Simulation Modeling and Analysis, McGraw-Hill Book Co., N.Y.

ɔcks, M.O. (1979), "Evaluating the KTI Monte Carlo Method for Systems Reliability Calculations," IEEE Trans. on Reliability, R-28, 5, 368-372.

ıdow, W.G. and L.H. Madow (1944), "On the Theory of Systematic Sampling," Ann. Math. Stat., 15, 1-24.

ırshall, A.W. (1956), "The Use of Multi-Stage Sampling Scheme in Monte Carlo Computations," in Symposium on Monte Carlo Methods (H.A. Meyer, ed.), Wiley, N.Y.

ıtern, B. (1962), "On the Use of Information on Supplementary Variables in Estimating a Distribution," Rev. Int. Stat. Inst., 30, 2, 121-135.

ızumdar, M. (1975), "Importance Sampling in Reliability Estimation," Reliability and Fault Tree Analysis, SIAM, Philadelphia, 153-163.

cGrath, E.J. and D.C. Irving (1974), "Application of Variance Reduction to Large Scale Simulation Problems," Comput. Ops. Res., 1, 283-311.

cGrath, E.J. and D.C. Irving (1973a), "Techniques for Efficient Monte Carlo Simulation, Vol. III, Variance Reduction," ORNL Report. SAI-72-509-LJ.

Koehler, K.J. (1981), "An Improvement of a Monte Carlo Technique Using Asymptotic Moments with an Application to the Likelihood Ratio Statistic," Commun. Statist.-Simula. Computa., B10, 4, 343-357.

Kohlas, J. (1982), Stochastic Methods of Operations Research, Cambridge University Press, N.Y.

Koop, J.C. (1971), "Short Communications on Splitting a Systematic Sample for Variance Reduction," Ann. Math. Stat., 42, 3, 1084-1087.

Kumamoto, H., K. Tanaka and K. Inoue (1977), "Efficient Evaluation of System Reliability by Monte Carlo Method," IEEE Trans. on Reliability, R-26, 5, 311-315.

Kumamoto, H., K. Tanaka, K. Inoue and E. Henley (1980a), "Dagger Sampling Monte Carlo for System Unavailability Evaluation," IEEE Trans. on Reliability, R-29, 2, 122-125.

Kumamoto, H., K. Tanaka, K. Inoue and E. Henley (1980b), "State-Transition Monte Carlo for Evaluating Large, Repairable Systems," IEEE Trans. on Reliability, R-29, 5, 376-380.

Lavenberg, S. and P. Welch (1981), "A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations," Mgt. Sci., 27,3, 322-335.

Lavenberg, S. and P. Welch (1979), "Using Conditional Expectation to Reduce Variance in Discrete Event Simulation," 1979 Winter Simulation Conference Proceedings, 291-294.

Lavenberg, S. and P. Welch (1978), "Variance Reduction Techniques," 1978 Winter Simulation Conference Proceedings, 167-170.

Lavenberg, S., T. Moeller and P. Welch (1982), "Statistical Results on Control Variables with Application to Queuing Network Simulation," Op. Res., 30, 1, 182-202.

Lavenberg, S., T. Moeller and P. Welch (1978), "Statistical Results on Multiple Control Variables with Applications to Variance Reduction in Queuing Network Simulation," IBM Research Report RC 7423.

Innis, G. and E. Rexstad (1983), "Simulation Model Simplification Techniques," Simulation, 41, 1, 7-15.

Isaki, C.T. (1983), "Variance Estimation Using Auxiliary Information," JASA, 78, 381, 117-123.

Jessop, W.N. (1956), "Monte Carlo Methods and Industrial Problems," Appl. Statist., 5, 3, 158-165.

Kahn, H. (1956), "Use of Different Monte Carlo Sampling Techniques," in Symposium on Monte Carlo Methods, (H. Meyer ed.), Wiley, N.Y., 146-190.

Kahn, H. (1950a), "Random Sampling (Monte Carlo) Techniques in Neutron Attenuation Problems - II," Nucleonics, 6, 6, 60-65.

Kahn, H. (1950b), "Random Sampling (Monte Carlo) Techniques in Neutron Attenuation Problems - I," Nucleonics, 6, 5, 27-33, 37.

Kahn, H. and A. Marshall (1953), "Methods of Reducing Sample Size in Monte Carlo Computations," Op. Res., 1, 5, 263-278.

Kennedy, W. and J. Gentle (1980), Statistical Computing, Marcel Dekker, N.Y.

Kleijnen, J. (1978), "Analysis of Simulation With Common Random Numbers: A Note on Heikes, et. al.," Simuletter, 11, 2, 7-10, 13.

Kleijnen, J. (1975), "Antithetic Variates, Common Random Numbers and Optimal Computer Time Allocation in Simulation," Mgt. Sci., 21, 1176-1185.

Kleijnen, J. (1974), Statistical Techniques in Simulation, Part I, Marcel Dekker, N.Y.

Kloek, T. and H.K. van Dijk (1978), "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," Econometrica, 46, 1, 1-19.

Hammersley, J.M. and K. Morton (1954), "Poor Man's Monte Carlo," J. Roy. Statist. Soc., Series B, 16, 1, 23-38.

Handscomb, D.C. (1958), "Proof of the Antithetic-Variates Theorem for n > 2," Proc. Camb. Phil. Soc., 54, 2, 300-301.

Harling, J. (1958), "Simulation Techniques in Operations Research - A Review," Op. Res., 6, 3, 307-319.

Hartley, H.O. (1977), "Solution of Statistical Distribution Problems by Monte Carlo Methods," in Statistical Methods for Digital Computers, Vol. III, (Enslein, Ralston, Wilf eds.), Wiley, N.Y., 16-34.

Heidelberger, P. and D. Iglehart (1978), "Comparing Stochastic Systems Using Regenerative Simulation With Common Random Numbers," Technical Report No. 47, Dept. of Operations Research, Stanford Univ.

Heikes, R., D. Montgomery and R. Rardin (1976), "Using Common Random Numbers in Simulation Experiments -- An Approach to Statistical Analysis," Simulation, 25, 81-85.

Henley, E.J. and H. Kumamoto (1981), Reliability Engineering and Risk Assessment, Prentice-Hall, Englewood Cliffs, N.J.

Hillier, F.S. and G.J. Lieberman (1974), Operations Research, Holden-Day, Inc., San Francisco.

Hopmans, A.C.M. and J.P.C. Kleijnen (1980), "Regression Estimation in Simulation," Report FEW-89, Dept of Economics, Katholieke Hogeschool, Tilberg (Netherlands).

Hopmans, A.C.M. and J.P.C. Kleijnen (1979), "Importance Sampling in System Simulation: A Practical Failure?" Math. Computers in Simulation XXI, 209-220.

Iglehart, D. and P. Lewis (1979), "Regenerative Simulation with Internal. Controls," JACM, 26, 2, 271-282.

Ignall, E.J. (1972), "On Experimental Designs for Computer Simulation Experiments," Mgt. Sci., 18, 7, 384-388.

Granovsky, B.L. (1981), "Optimal Formulae of the Conditional Monte Carlo," SIAM J. Alg. Disc. Meth., 2, 3, 289-294.

Grant, F.H. (1983), "A Note on Efficiency of the Antithetic Variate Method for Simulating Stochastic Networks," Mgt. Sci., 29, 3, 381-384.

Grant, F.H. (1980), "Variance Reduction Techniques in Stochastic Shortest Route Analysis," unpublished Ph.D. thesis, School of Industrial Engineering, Purdue Univ.

Grant, F.H. and J.J. Solberg (1983), "Variance Reduction Techniques in Stochastic Shortest Route Analysis: Application, Procedures and Results," Math. Computers in Simulation XXV, 366-375.

Gross, A.M. (1973), "A Monte Carlo Swindle for Estimators of Location," Appl. Statist., 22, 347-353.

Halfin, S. (1982), "Linear Estimators for a Class of Stationary Queuing Processes," Op. Res., 30, 3, 515-529.

Halton, J.H. (1979), "An Explicit Formulation of the Generalized Antithetic Transformation for Monte Carlo Integration," Computer Sciences Technical Report 348, Computer Sciences Dept., Univ. of WI-Madison.

Halton, J.H. and D.C. Handscomb (1957), "A Method for Increasing the Efficiency of Monte Carlo Integration," JACM, 4, 3, 329-340.

Hammersley, J.M. (1956), "Conditional Monte Carlo," JACM, 3, 73-76.

Hammersley, J.M. and D. Handscomb (1964), Monte Carlo Methods, Chapman and Hall, London.

Hammersley, J.M. and J. Mauldon (1956), "General Principles of Antithetic Variates," Proc. Camb. Phil. Soc., 52, 476-481.

Hammersley, J.M. and K. Morton (1956), "A New Monte Carlo Technique: Antithetic Variates," Proc. Camb. Phil. Soc., 52, 449-475.

Fishman, G.S. and B. Huang (1980), "Antithetic Variates Revisited," Technical Report No. 80-4, Curriculum in Operations Research and Systems Analysis, Univ. of North Carolina at Chapel Hill.

Fox, B.L. (1983a), "When Conditional Monte Carlo Works," Departement d'informatique et de recherche operationnelle, Univ. de Montreal.

Fox, B.L. (1983b), "Counterparts of Variance Reduction Techniques for Quasi-Monte Carlo Integration," Departement d'informatique et de recherche operationnelle, Univ. de Montreal.

Franta, W.R. (1975), "A Note on Random Variate Generators and Antithetic Sampling," INFOR, 13, 1, 112-117.

Gallaher, L.J. (1973), "Algorithm 440: A Multidimensional Monte Carlo Quadrature with Adaptive Stratified Sampling," CACM, 16, 1.

Garman, M.B. (1972), "More Conditioned Sampling in the Simulation of Stochastic Networks," Mgt. Sci., 19, 1, 90-95.

Gaver, D.P., Jr. (1969), "Statistical Methods for Improving Simulation Efficiency," Management Sciences Report No. 169, Graduate School of Industrial Administration, Carnegie-Mellon Univ.

Gaver, D.F. and G.S. Shedler (1971), "Control Variable Methods in the Simulation of a Model of Multiprogrammed Computer System," Naval Res. Logistics Quart., 18, 435-450.

Gaver, D.P. and G.L. Thompson (1973), Programming and Probability Models in Operations Research, Brooks/Cole Pub. Co., Monterey, C.A.

Gentle, J. (1975), "A Survey of Variance Reduction and Robust Procedures in Monte Carlo Analysis," Presented before the Fall Joint National ORSA/TIMS Meetings, Las Vegas.

George, L.L. (1977), "Variance Reduction for a Replacement Process," Simulation, 29, 3, 65-74.

Gordon, G.R. and J. Surkis (1978), "A General Simulation Methodology for Optical Bar Code Evaluation," J. Statist. Comput. Simul., 8, 2, 85-101.

Fishman, G.S. (1983b), "Variance Reduction in the Simulation of Stochastic Activity Networks," Technical Report No. UNC/ORSA/TR-82/7, Curriculum in Operations Research and Systems Analysis, Univ. of North Carolina at Chapel Hill.

Fishman, G.S. (1982a), "Superefficient Simulation of Markov Chains and Semi-Markov Processes," Technical Report No. UNC/ORSA/TR-82/5, Curriculum in Operations Research and Systems Analysis, Univ. of North Carolina at Chapel Hill.

Fishman, G.S. (1982b), "Generating Parallel Correlated Transition Frequencies for a Markov Chain," Technical Report No. UNC/ORSA/TR-82/8, Curriculum in Operations Research and Systems Analysis, Univ. of North Carolina at Chapel Hill.

Fishman, G.S. (1981), "Accelerated Convergence in the Simulation of Countably Infinite State Markov Chains," Technical Report No. 81-4, Curriculum in Operations Research and Systems Analysis, Univ. of North Carolina at Chapel Hill.

Fishman, G.S. (1979), "Variance Reduction for Population Growth Simulation Models," Op. Res., 27, 5, 997-1010.

Fishman, G.S. (1978), Principles of Discrete Event Simulation, Wiley, N.Y.

Fishman, G.S. (1974), "Correlated Simulation Experiments," Simulation, 23, 6.

Fishman, G.S. (1973a), "Variance Reduction for Normal Variates in Monte Carlo Studies," J. Statist. Comput. Simul., 2, 4, 365-374.

Fishman, G.S. (1973b), Concepts and Methods in Discrete Event Digital Simulation, Wiley, N.Y.

Fishman, G.S. (1972), "Variance Reduction in Simulation Studies," J. Statist. Comput. Simul., 1, 173-182.

Fishman, G.S. (1968), "The Allocation of Computer Time in Comparing Simulation Experiments," Op. Res., 16, 2, 280-295.

Dalenius, T. (1950), "The Problem of Optimum Stratification," Skandinavisk Aktuarietidskrift, 33, 203-213.

Dalenius, T. and M. Gurney (1951), "The Problem of Optimum Stratification. II," Skandinavisk Aktuarietidskrift, 34, 113-148.

DeGroot, M.H. and N. Starr (1969), "Optimal Two-Stage Stratified Sampling," Ann. Math. Stat., 40, 2, 575-582.

Deutsch, S.J. and B.W. Schmeiser (1980), "Estimation of the Sum of Differences Distribution," Commun. Statist.-Simula. Computa., B9, 6, 563-587.

Diegert, C. and K.V. Diegert (1981), "How to Cheat a Monte Carlo Estimation of a System's Reliability," Sandia National Laboratories, Alburquerque, NM.

Dubi, A. and Y.S. Horowitz (1979), "The Interpretation of Conditional Monte Carlo as a Form of Importance Sampling," SIAM J. Appl. Math., 36, 1, 115-122.

Easterfield, T.E. (1961), "A Short Cut in a Class of Simulation Problems," Opl. Res. Quart., 12, 4, 221-225.

Easton, M.C. and C.K. Wong (1980), "Sequential Destruction Method for Monte Carlo Evaluation of System Reliability," IEEE Trans. on Reliability, R-29, 1, 27-32.

Ehrenfeld, S. and S. Ben-Tuvia (1962), "The Efficiency of Statistical Simulation Procedures," Technometrics, 4, 2, 257-275.

Ermakov, S.M. and V.G. Zolotukhin (1960), "Polynomial Approximations and the Monte-Carlo Method," Theory of Prob. and Its Applications, 5, 4, 428-431.

Fieller, E.C. and H.O. Hartley (1954), "Sampling with Control Variables," Biometrika, 41, 494-501.

Fishman, G.S. (1983a), "Accelerated Accuracy in the Simulation of Markov Chains," Op. Res., 31, 3, 466-487.

Cheng, R.C.H. (1983a), "Generation of Multivariate Normal Samples with Given Sample Mean and Covariance Matrix," Math Report 83-2, Dept. of Mathematics, UWIST, Cardiff, Britain.

Cheng, R.C.H. (1983b), "Generation of Inverse Gaussian Variates with Given Sample Mean and Dispersion," Math Report 83-1, Dept. of Mathematics, UWIST, Cardiff, Britain.

Cheng, R.C.H. (1982), "The Use of Antithetic Variates in Computer Simulations," J. Opl. Res. Soc., 33, 229-237.

Cheng, R.C.H. (1981), "The Use of Antithetic Control Variates in Computer Simulations," 1981 Winter Simulation Conference Proceedings, 313-318.

Cheng, R.C.H. (1978), "Analysis of Simulation Experiments Under Normality Assumptions," J. Opl. Res. Soc., 29, 5, 493-497.

Cheng, R.C.H. and G.M. Feast (1980), "Control Variables with Known Mean and Variance," J. Opl. Res. Soc., 31, 1, 51-56.

Clark, C.E. (1961), "Importance Sampling in Monte Carlo Analysis," Op. Res., 9, 5, 603-620.

Cochran, W.G. (1977), Sampling Techniques, Wiley, N.Y.

Conway, R.W. (1963), "Some Tactical Problems in Digital Simulation," Mgt. Sci., 10, 1, 47-61.

Cook, T.M. and R.H. Jennings (1979), "Estimating a Project's Completion Time Distribution Using Intelligent Simulation Methods," J. Opl. Res. Soc., 30, 1103-1108.

Cooley, B.J. and E.C. Houck (1983), "On an Alternative Variance-Reduction Strategy for RSM Simulation Studies," Decision Sciences, 14, 134-137.

Cooley, B.J. and E.C. Houck (1982), "A Variance-Reduction Strategy of RSM Simulation Studies," Decision Sciences, 13, 303-321.

Brenner, M.E. (1963), "Selective Sampling -- A Technique for Reducing Sample Size in Simulation of Decision-Making Problems," Journal of Industrial Engineering, 14, 6, 291-296.

Brewer, K.R.W. (1963), "A Model of Systematic Sampling with Unequal Probabilities," Austral. J. Statist., 5, 13, 5-15.

Brown, M., H. Solomon and M.A. Stephens (1979), "Variance Reduction in Monte Carlo Simulation," Technical Report No. 274, Dept. of Statistics, Stanford Univ.

Burt, J.M. and M.B. Garman (1971a), "Monte Carlo Techniques for Stochastic PERT Network Analysis," INFOR, 9, 3, 248-262.

Burt, J.M., D.P. Gaver and M. Perlas (1970), "Simple Stochastic Networks: Some Problems and Procedures," Naval Res. Logistics Quart., 17, 4, 439-459.

Burt, J.M. and M.B. Garman (1971b), "Conditional Monte Carlo: A Simulation Technique for Stochastic Network Analysis," Mgt. Sci., 18, 3, 207-217.

Carson, J.S. (1983), "On the Efficiency of Antithetic Sampling and Conditional Monte Carlo in Simulations of Stochastic Activity Networks," Technical Report, School of Industrial and Systems Engineering, Georgia Inst. of Technology.

Carson, J.S. and A.M. Law (1980), "Conservation Equations and Variance Reduction in Queuing Simulations," Op. Res., 28, 3, 535-546.

Carter, G. and E. Ignall (1975), "Virtual Measures: A Variance Reduction Technique for Simulation," Mgt. Sci., 21, 6, 607-616.

Carter, G. and E. Ignall (1970), "A Simulation Model of Fire Department Operations: Design and Preliminary Results," IEEE - Systems Science and Cybernetics, 6, 4, 282-293.

Carter, L.L. and E.D. Cashwell (1975), Particle-Transport Simulation with the Monte Carlo Method, ERDA Critical Review Series, USERDA Technical Information Center, Oak Ridge, Tennessee.

BIBLIOGRAPHY

Adkins, G.W. (1982), "Variance Reduction in Estimating the Mean Flow Time in Open Queuing Network Simulations," 1982 Winter Simulation Conference Proceedings, 450-454.

Andrews, D.F., P.J. Bickel, F.R. Hampel, P.J. Huber, W.H. Rogers, and J.W. Tukey (1972), Robust Estimates of Location: Survey and Advances, Princeton Univ. Press, Princeton, NJ.

Arnold, H.J., B.D. Bucher, H.F. Trotter and J.W. Tukey (1956), "Monte Carlo Techniques in a Complex Problem About Normal Samples," in Symposium on Monte Carlo Methods, (H. Meyer, ed.), Wiley, N.Y., 80-88.

Ashford, R.W. and E.M.L. Beale (1983), "A Cross-Estimation Technique for Using Control Variables in Stochastic Simulations," Scicon Limited, Great Britain.

Banks, J. and J.S. Carson (1983), Discrete-Event Simulation, Prentice-Hall, Englewood Cliffs, N.J.

Bayes, A.J. (1972), "A Minimum Variance Sampling Technique for Simulation Models," JACM, 19, 4, 734-741.

Becker, P. (1974), "Finding the Better of Two Similar Designs by Monte Carlo Techniques," IEEE Trans. on Reliability, R-23, 4, 242-246.

Birge, J.R. and R.L. Smith (1982), "New Monte Carlo Procedures for Numerical Integration," Technical Report 82-8, Dept. of Industrial and Operations Engineering, Univ. of Michigan.

Bratley, P., B.L. Fox and L.W. Schrage (1983), A Guide to Simulation, Springer-Verlag, N.Y.

BIBLIOGRAPHY

useful, for instance. The implications of the various equivalence relations for simulation experiments also appears to be an interesting area for future research.

Recent research in the area of variance reduction has emphasized the application of VRTs in specific, but hopefully broad classes of problems. Attempts have been made to specify rather weak conditions under which a variance reduction is guaranteed. We hope that our research will accelerate the effort, making it possible to derive even more general conditions for even broader classes of problems. The ultimate achievement, which is probably not possible, would be necessary and sufficient conditions under which application of each class of transformations would achieve a variance reduction.

As mentioned above, the definition of simulation is quite broad in scope. Certainly the inclusion of survey sampling should be studied more thoroughly. By having such a general model of sampling experiments, defining the differences between particular cases such as "systems simulation," Monte Carlo, and survey sampling is possible. For instance, McGrath and Irving (1973a) stated that simulation experiments can be viewed as Monte Carlo estimation problems like (2.2). The definition reveals why this is sometimes difficult. In systems simulation the integral might be of the form:

$$\theta = \int_A g(x) f(x|\beta) dx$$

where β is a function of $g(x)$.

To attack the problem of determining necessary and sufficient conditions for a transformation to be effective, the theory of statistical information might be a key tool. An investigation of the types and value of different kinds of auxiliary information is in order. The result in Chapter 5 demonstrates that linear correlation can be

McGrath, E.J. and D.C. Irving (1973b), "Demonstration of Improved Monte Carlo Simulation Techniques for APAIR Antisubmarine Warfare Program," ORNL Report SAI-73-533-LJ.

McKay, M.D., R.J. Beckman and W.J. Conover (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," Technometrics, 2, 2, 239-245.

Meier, R.C., W.T. Newell and H.L. Pazer (1969), Simulation in Business and Economics, Prentice-Hall, Englewood Cliffs, N.J.

Mihram, G.A. (1974), "Blocking in Simular Experimental Designs," J. Statist. Comput. Simul., 3, 4, 29-32.

Mihram, G.A. (1972), Simulation: Statistical Foundations and Methodology, Academic Press, N.Y.

Minh, D.L. and R.M. Sorli (1983), "Simulating the GI/G/1 Queue in Heavy Traffic," Op. Res., 31, 5, 966-971.

Mitchell, B. (1973), "Variance Reduction by Antithetic Variates in GI/G/1 Queuing Simulations," Op. Res., 21, 4, 988-997.

Mize, J.H. and J.G. Cox (1968), Essentials of Simulation, Prentice-Hall, Englewood Cliffs, N.J.

Morton, K.W. (1957), "A Generalization of the Antithetic Variate Technique for Evaluating Integrals," J. of Math. and Physics, 36, 3, 289-293.

Moy, W. (1971), "Variance Reduction," in Computer-Simulation Experiments with Models of Economic Systems, (T.H. Naylor, ed.), Wiley, N.Y.

Moy, W.A. (1965), "Sampling Techniques for Increasing the Efficiency of Simulations of Queuing Systems," Ph.D. dissertation, Northwestern University, University Microfilms, International, Ann Arbor, Michigan.

Naylor, T., J. Balintfy, D. Burdick and K. Chu (1966), Computer Simulation Techniques, Wiley, N.Y.

Naylor, T.H. (ed.) (1971), Computer Simulation Experiments with Models of Economic Systems, Wiley, N.Y.

Nozari, A., S.F. Arnold and C.D. Pegden (1983), "Control Variates for Multipopulation Simulation Experiments," to appear in IIE Transactions.

Olkin, I. (1958), "Multivariate Ratio Estimation for Finite Populations," Biometrika, 45, 154-165.

Page, E.S. (1965), "On Monte Carlo Methods in Congestion Problems: II. Simulation of Queuing Systems," Op. Res., 13, 2, 300-305.

Payne, J.A. (1982), Introduction to Simulation, McGraw-Hill Book Co., New York.

Powell, M.J.D. and J. Swann (1966), "Weighted Uniform Sampling -- A Monte Carlo Technique for Reducing Variance," J. Inst. Maths. Applics., 2, 3, 228-236.

Pritsker, A.A.B. and C.D. Pegden (1979), Introduction to Simulation and SLAM, Wiley, N.Y.

Roach, W. and R. Wright (1977), "Optimal Antithetic Sampling Plans," J. Statist. Comput. Simul., 5, 2, 99-114.

Roach, W. and R. Wright (1974), "Optimally Efficient Antithetic Sampling Plans for Model Sampling," Statistical Computing Section of American Statistical Association Meetings, St. Louis.

Ross, S.M. and Z. Schechner (1983), "Using Simulation to Estimate First Passage Distributions," ORC 83-1, Operations Research Center, Univ. of California, Berkley.

Rubinstein, R.Y. and G. Samorodnitsky (1982), "Antithetic and Common Random Numbers in Simulation of Stochastic Systems," Monograph Series No. 308, Operations Research, Statistics and Economics, Technion, Israel Inst. of Technology.

Rubinstein, R.Y., G. Samorodnitsky and M. Shaked (1982), "Antithetic Variates, Multivariate Dependence and Simulation of Complex Stochastic Systems," Technical Report, Technion, Israel Inst. of Technology.

Rubinstein, R.Y. (1981), Simulation and the Monte Carlo Method, Wiley, N.Y.

Rubinstein, R.Y., S. Gal and A. Ziv (1981), "On the Optimality and Efficiency of Common Random Numbers, " IBM Israel Scientific Center.

Rubinstein, R.Y. and R. Markus (1981), "Efficiency of Multivariate Control Variates in Monte Carlo Simulation," National Research Institute for Mathematical Sciences, (NRIMS/W/68/5), Internal Report No. I 339.

Sarndal, C.E. (1968), "The Use of a Stratification Variable in Estimation by Proportional Stratified Sampling, " JASA, 63, 324, 1310-1320.

Schruben, L. (1979), "*Designing* Correlation Induction Strategies for Simulation Experiments," in Current Issues in Simulation, Academic Press, N.Y.

Schruben, L. (1978), "Designing Correlation Induction Strategies for Simulation Experiments," Technical Report No. 373, School of Operations Research and Industrial Engineering, Cornell Univ.

Schruben, L. and B. Margolin (1978), "Pseudorandom Number Assignment in Statistically Designed Simulation and Distribution Sampling Experiments," JASA, 73, 363, 504-525.

Seila, A.F. (1982), "Stratified Estimation in Regenerative Simulations," Univ. of Georgia.

Sethi, V.K. (1963), "A Note on Optimum Stratification of Populations for Estimating the Population Mean," Austral. J. Statist., 5, 13, 20-33.

Shreider, Yu A. (1966), The Monte Carlo Method: The Method of Statistical Trials, Pergamon Press, N.Y.

Shreider, Yu A. (1964), Method of Statistical Testing: Monte Carlo Method, Elsevier Pub. Co., N.Y.

Sigal, C.E. (1977), "The Stochastic Shortest Route Problem," unpublished Ph.D. thesis, School of Industrial Engineering, Purdue Univ.

Sigal, C.E., A.A.B. Pritsker and J.J. Solberg (1979), "The Use of Cutsets in Monte Carlo Analysis of Stochastic Networks," Math. Computers in Simulation XXI, 376-384.

Simon, G. (1976), "Computer Simulation Swindles, with Applications to Estimates of Location and Dispersion," Appl. Statist., 25, 3, 266-274.

Singh, R. (1977), "A Note on Optimum Stratification for Equal Allocation with Ratio and Regression Methods of Estimation," Austral. J. Statist., 19, 2, 96-104.

Singh, R. (1975a), "On Optimum Stratification for Proportional Allocation," Sankhya: The Indian Journal of Statistics, Series C, 37, 109-115.

Singh, R. (1975b), "An Alternative Method of Stratification on the Auxiliary Variable," Sankhya: The Indian Journal of Statistics, Series C, 37, 100-108.

Sobol', I.M. (1974), The Monte Carlo Method, Popular Lectures in Mathematics, The Univ. of Chicago Press, Chicago.

Spanier, J. (1970), "An Analytic Approach to Variance Reduction," SIAM J. Appl. Math. 18, 1, 172-190.

Spanier, J. and E.M. Gelbard (1969), Monte Carlo Principles and Neutron Transport Problems, Addison-Wesley Pub. Co., Reading, MA.

Steinberg, H.A. (1963), "Generalized Quota Sampling," Nuc. Sci. and Engr., 15, 2, 142-145.

Sullivan, R., J. Hayya and R. Schaul (1982), "Efficiency of the Antithetic Variate Method for Simulating Stochastic Networks," Mgt. Sci., 28, 5, 563-572.

Surkis, J, G. Gordon and J. Hynes (1975), "A New Variance Reduction Technique for Evaluating System Reliability Problems by Simulation" J. Statist. Comput. Simul., 4, 3, 215-224.

Swain, J. (1982), "Monte Carlo Estimation of the Sampling Distribution of Nonlinear Parameter Estimators," unpublished Ph.D. thesis, School of Industrial Engineering, Purdue Univ.

Swain, J. (1981), "Monte Carlo Estimation of Sampling Distributions Arising in Nonlinear Statistical Modeling," Ph.D. Thesis Proposal, School of Industrial Engineering, Purdue Univ.

Swain, J. and B.W. Schmeiser (1983), "Monte Carlo Estimation of the Sampling Distribution of Nonlinear Model Parameter Estimators," Technical Report C-83-1, Georgia Inst. of Technology.

Taaffe, M.R. and S.A. Horn (1983), "External Control Variance Reduction for Nonstationary Simulation," to appear in 1983 Winter Simulation Conference Proceedings.

Tocher, K.D. (1963), The Art of Simulation, D. Van Nostrand Co., Princeton, N.J.

Trotter, H.F. and J.W. Tukey (1956), "Conditional Monte Carlo for Normal Samples," in Symposium for Monte Carlo Methods, (H. Meyer, ed.), 64-79, Wiley, N.Y.

Wendel, J.G. (1957), "Groups and Conditional Monte Carlo," Ann. Math. Stat., 28, 4, 1048-1052.

Wilson, J.R. (1983a), "Variance Reduction: The Current State," Math. Computers in Simulation XXV, 55-59.

Wilson, J.R. (1983b), "Variance Reduction Techniques in Stochastic Simulation," to appear in Encyclopedia of Systems and Control, (M. Singh, ed.), Pergamon Press, Oxford.

Wilson, J.R. (1983c), "Antithetic Sampling with Multivariate Inputs," American J. Math. and Mgt. Sci., 3, 2.

Wilson, J.R. (1983d), "Optimal Univariate Rotation Sampling," Dept. of Mechanical Engineering Technical Report, Univ. of Texas.

Wilson, J.R. (1983e), "Variance Reduction in Queuing Simulation Using Generalized Concomitant Variables," to appear in J. Statist. Comput. Simul.

Wilson, J.R. (1983f), "Experimental Evaluation of Variance Reduction Techniques for Queuing Simulations Using Generalized Concomitant Variables," to appear in Mgt. Sci.

Wilson, J.R. (1982a), "Antithetic Sampling with Multivariate Inputs," Dept. of Mechanical Engineering Technical Report, Univ. of Texas.

Wilson, J.R. (1982b), "Variance Reduction Techniques," 1982 Winter Simulation Conference Proceedings, 604-612.

Wilson, J.R. (1979a), "Proof of the Antithetic Variate Theorem for Unbounded Functions, " Math. Proc. Camb. Phil. Society, 86, 477.

Wilson, J.R. (1979b), "Variance Reduction Techniques for the Simulation of Queuing Networks, " Unpublished Ph.D. Dissertation, School of Industrial Engineering, Purdue Univ.

Wilson, J.R. and A.A.B. Pritsker (1982), "Variance Reduction in Queuing Simulations Using Generalized Concomitant Variables, I: Principles and Methods," Dept. of Mechanical Engineering Technical Report, Univ. of Texas.

Wright, R. and T. Ramsey (1979), "On the Effectiveness of Common Random Numbers," Mgt. Sci., 25, 7, 649-656.

Yakowitz, S.J. (1977), Computational Probability and Simulation, Addison-Wesley, London.

Yakowitz, S., J.E. Krimmel and F. Szidarovoszky (1978), "Weighted Monte Carlo Integration," SIAM J. Numer. Anal., 15, 6, 1289-1300.

VITA

VITA

Barry Lee Nelson was born on December 21, 1956 in Vincennes, Indiana. He attended public school in Columbus, Indiana and graduated from Columbus East High School in 1975. He received a B.A. (Summa Cum Laude, Phi Beta Kappa) from Depauw University, Greencastle, Indiana in mathematics and computer science in 1979. At Purdue University, West Lafayette, Indiana, he earned M.S. and Ph.D. degrees in operations research from the School of Industrial Engineering in 1981 and 1983, respectively. His Master's thesis, "Combined Graphical-Simulation Analysis of a USAF Staffing Problem," won the 1981 Institute of Industrial Engineers Graduate Research Award. The title of his Ph.D. dissertation was "Variance Reduction in Simulation Experiments: A Mathematical - Statistical Framework." It was completed under the direction of Dr. Bruce W. Schmeiser. In the dissertation he derived definitions of mutually exclusive classes of transformations that exhaust the set of all variance reduction techniques under composition. In December 1983 Barry accepted a position as an Assistant Professor in the Department of Industrial and Systems Engineering at The Ohio State University.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. A158146	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Variance Reduction in Simulation Experiments: A Mathematical-Statistical Framework		5. TYPE OF REPORT & PERIOD COVERED Ph.D. Dissertation 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Barry L. Nelson		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0832
9. PERFORMING ORGANIZATION NAME AND ADDRESS School of Industrial Engineering Purdue University West Lafayette, Indiana 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N042-477
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research -- Code 411 800 North Quincy Street Arlington, VA 22217		12. REPORT DATE December 1983 13. NUMBER OF PAGES 128
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release: Distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Simulation, Variance Reduction, Monte Carlo, Sampling, Swindle, Estimation		

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

With the expanding use of computer simulation to model and solve industrial engineering problems, there has been increasing interest in the development of efficient simulation techniques. When the concern if for statistical efficiency of results that are random variables, such approaches are usually called variance reduction techniques (VRTs).

Many of the fundamental ideas in simulation, and particularly techniques for efficient simulation, had their origins in the Monte Carlo estimation

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

literature. The theory of sampling is another closely related field that predates the development of simulation. Although there has been significant research interest in variance reduction, there have been few attempts to structure and define the discipline.

VRTs are transformations. They transform simulation experiments into related experiments that yield better estimates of some parameters of interest, where better usually means more precise. This research identifies and defines the components from which all variance reduction techniques are built. Given a general mathematical-statistical definition of simulation experiments, these components or classes of transformations are shown to be useful, to be mutually exclusive, and to generate all possible VRTs via composition. Benefits of the research include: 1) the facility to decompose VRTs into combinations of transformations, making the relationships between VRTs clear, 2) the facility to unambiguously define new or existing VRTs, eliminating confusion that currently exists in the literature, 3) the development of a theoretical foundation for analytical treatment of VRTs, and 4) the development of a setting for proposing new VRTs and research questions. In addition, increased understanding of the area should promote more and better application of variance reduction in practice.

END

FILMED

9-85

DTIC