Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: http://www.tandfonline.com/loi/lssp20

A perspective on variance reduction in dynamic simulation experiments

Barry L. Nelson

To cite this article: Barry L. Nelson (1987) A perspective on variance reduction in dynamic simulation experiments, Communications in Statistics - Simulation and Computation, 16:2, 385-426, DOI: 10.1080/03610918708812597

To link to this article: http://dx.doi.org/10.1080/03610918708812597

COMMUN. STATIST. -SIMULA., 16(2), 385-426 (1987)

A PERSPECTIVE ON VARIANCE REDUCTION IN DYNAMIC SIMULATION EXPERIMENTS

Barry L. Nelson

Department of Industrial and Systems Engineering
The Ohio State University
Columbus, OH 43210

Key Words and Phrases: antithetic variates; common random numbers; conditional expectations; control variates; generalized semi-Markov process; importance sampling; indirect estimators; Latin hypercube sampling; Monte Carlo; poststratifying the sample; Russian roulette; sampling; simulation; splitting; stratified sampling; swindle; systematic sampling; variance reduction.

ABSTRACT

A unifying perspective on variance reduction is presented that emphasizes broadly defined variance reduction strategies rather than specific variance reduction techniques (VRTs). The perspective is based on a new taxonomy of VRTs, which is reviewed in detail. The variance reduction problem is formulated as a constrained optimization problem, and results that guarantee the effectiveness of variance reduction strategies are summarized.

1. INTRODUCTION

Stochastic models are often used to describe real or conceptual systems in order to derive performance measures of interest. Analytic analysis methods are usually preferred, and there are rich classes of models for which analytic results are available (e.g. queueing models). However, when models are intractable we resort to simpler approximate models, numerical methods, or simulation. Each approach has its drawbacks: The validity of an approximation is often

established by tests on tractable models, but its performance on intractable models is uncertain. Good numerical methods converge to the performance measure of interest given infinite precision arithmetic and computing resources. However, their accuracy when arithmetic and computing resources are finite may be poor. Simulation experimentation is conceptually the simplest approach, because a simulation mimics the system of interest by sampling computer generated random numbers and exercising the model. Unfortunately, the variance of simulation estimators may be unacceptably large in practical problems.

In simulation, as in all sampling experiments, increased sampling usually reduces variance, but the cost of obtaining a large enough sample can be prohibitive. The availability of faster computers, instead of eliminating the problem, has spurred interest in using simulation to solve problems that were previously unmanageable. Optimization of stochastic models and incorporating stochastic models into real-time control systems via simulation are two current examples for which available computing budgets and available computer speed, respectively, are frequently inadequate.

Variance reduction techniques (VRTs) are techniques designed to reduce sampling error without a corresponding increase in computer effort. VRTs had their origins in survey sampling (Cochran, 1977) and Monte Carlo estimation (Hammersley and Handscomb, 1964). Unfortunately, techniques that work well for sampling from static populations or for evaluating definite integrals are often difficult to adapt to simulation of dynamic stochastic processes. Also, the lack of a unifying theory of variance reduction has been a hindrance. The result is that VRTs are infrequently used in practice, even though the author's experience indicates that large (one to two orders of magnitude) variance reductions are possible in practical problems.

Recently, Nelson and Schmeiser (1986a) proposed a taxonomy of VRTs that views VRTs as compositions of transformations from six elemental classes. The taxonomy is designed to serve as the basis for a unified theory of variance reduction, and ultimately as the basis for automated variance reduction. In the present survey paper, it facilitates formulating the variance reduction problem as a constrained optimization problem. Given this formulation, solutions to the

problem can be stated in terms of broadly defined variance reduction strategies rather than specific VRTs. Also, results that guarantee the effectiveness of these strategies, and the knowledge required to employ the results, can be conveniently organized. This is a fundamentally different and more unified presentation of variance reduction than previous surveys. It is **not** an exhaustive survey of specific VRTs, and we give no guidelines for implementation. Excellent surveys of this kind can be found in Kleijnen (1974), McGrath and Irving (1973), and a recent and highly recommended treatment in Wilson (1984). Textbook presentations of VRTs that have proven useful in simulation experiments are given by Bratley, Fox and Schrage (1983) and Law and Kelton (1982). See also Nelson (1985a), which contains an algorithm for selecting VRTs.

The paper is organized as follows: Section 2 reviews the Nelson and Schmeiser taxonomy of VRTs, and also presents the new formulation of the variance reduction problem. Given that background, sections 3 through 8 examine the classes of elemental transformations individually, giving broadly defined variance reduction strategies and summarizing theoretical results for the variance reduction problem. In section 9 we summarize and emphasize some of the important ideas in the previous sections. Finally, a concluding section raises issues relating to the automation of variance reduction and variance reduction in animation, both topics that we expect will receive significant future attention.

2. SIMULATION EXPERIMENTS AND VARIANCE REDUCTION

The basis for this survey is a taxonomy of variance reduction techniques that views VRTs as compositions of transformations from six elemental classes. A *transformation* maps one simulation experiment into another experiment that may have reduced variance. The six classes of transformations are based on a specific definition of simulation experiments, termed the *sample space definition* (Nelson and Schmeiser, 1986a). In this section we review the sample space definition, give an illustration, and then define the six classes of elemental transformations.

2.1 A Taxonomy of Variance Reduction Techniques

In the sample space definition, a simulation experiment has a fixed (Ω, θ). The *sample space of the input,* Ω, is a subset of R^∞ (the infinite dimensional

Euclidean space) and represents all possible realizations of the input (X); it is the range of the uncertain (uncontrollable) elements of the system. The *parameter of interest,* θ, is a vector of unknown real scalar constants. The purpose of performing a simulation experiment is to estimate θ.

The simulation *input,* X, is a matrix of scalar random variables modeled by F, the known cumulative probability distribution over Ω; notationally $X \sim F(x)$. The organization of the input into a matrix is for convenience: by convention, a column of X is an infinite sequence of identically distributed scalar inputs with the order of the elements in the column the same as the order in which realizations of the random variables will be generated in the simulation experiment. There is not necessarily a unique organization of X.

The *output,* Y, is a measurable function of X; notationally $Y = g(X)$. The output is the matrix of all essential random variables defined by functions of X, where *essential* means that any other random variable in the experiment that is a function of X can be calculated from Y provided no element of Y is deleted. The essential set restriction is necessary for theoretical reasons (see Nelson and Schmeiser, 1986a). Again, the output is organized into a matrix for convenience, using the same ordering convention as X.

The *sampling plan* R_* is a measurable function of X that constrains the number of elements in Y that will actually be realized; Y is countably infinite dimensional, since conceptually we can generate infinitely many inputs and corresponding outputs. The sampling plan allocates the available sampling effort by specifying the realized lengths of the columns of Y. Note that R_* constrains the outputs rather than the inputs, and implies a stopping rule for the simulation experiment.

The estimator of θ is the *statistic* Z, which is a measurable function of Y that aggregates (a subset of) the output into a vector of point estimators; notationally $Z = h(Y)$. A thorough discussion of the motivation for these definitions is given in Nelson and Schmeiser (1986a).

As an example, consider a simulation experiment performed on a model of an inventory system to estimate the expected stock level, one of the parameters of interest, θ. Suppose that demand for inventory is a random variable modeled by a specified probability distribution, implying that demand

is an input sequence in X. The stock currently on hand is an output sequence in Y. Since inventory position is given by the stock on hand plus the stock on order minus the backorders, inventory position is not an essential output, given the other three. The sampling plan R_* might specify the number of years of simulated operation during which we observe the stock level. If these observations of stock level are averaged to estimate the expected stock level, then that average is the statistic Z.

In this paper we will concentrate on scalar θ, or on only one element of multivariate θ. However, there are important unresolved issues in multivariate variance reduction. For scalar θ we can formulate the variance reduction problem considered here as

$$\textit{minimize } \mathrm{MSE}(Z,\theta) = \int_{\Omega} [h(g(x;R_*)) - \theta]^2 dF(x) \tag{1}$$

$$\textit{subject to: } \Omega, \theta \textit{ fixed} \tag{C.0}$$

which is minimization of the variance if $\mathrm{E}[Z] = \theta$. Of course, the analytic solution (unbiased, zero variance) for θ is the optimal solution to (1), but θ must be considered intractable if we are simulating. Thus, we seek instead a solution that yields an objective function value smaller than the original experiment does. Since constraint (C.0) fixes (Ω, θ), the *decision variables* are F, g, R_* and h, which can be changed by the experimenter.

Nelson and Schmeiser (1986a) define six mutually exclusive classes of transformations that exhaust the "decisions" that might reduce variance. A transformation maps a simulation experiment into another nonequivalent experiment by redefining F, g, R_* and/or h. A VRT is formed by composing members of these classes of transformations. See Nelson and Schmeiser (1986b) for the decomposition of seven well-known VRTs.

We now review the six classes of elemental transformations. A transformation is denoted by T, and if a transformation is defined as changing the definition of F, g, R_* or h alone, then the remaining components are unchanged. Each class of transformations redefines scalar random variables, where the "ij" subscript simply denotes scalar elements of matrices in the usual manner. Equality of distributions or functions is always on all but a set of

probability zero, and when distributions or functions are not equal it is on a set of positive probability. In addition, we are indifferent to one-to-one transformations of the outputs and equivalent essential sets; e.g. we treat Y_{ij} and $\ln Y_{ij}$ as equivalent. A prime (′) added to any symbol denotes a modified random variable, function or distribution, respectively.

1. *Distribution Replacement (DR):* $T \in$ DR if and only if T: $F(x) \to F'(x)$ such that $F' \neq F$, but

$$\frac{\Pr'(X_{ij} \in A \mid X \setminus X_{ij} \in B)}{\Pr'(X_{ij} \in A)} = \frac{\Pr(X_{ij} \in A \mid X \setminus X_{ij} \in B)}{\Pr(X_{ij} \in A)}$$

for all sets A and B with positive probability, and all ij. Transformations in DR redefine marginal distributions of elements of the input without altering statistical dependencies.

2. *Dependence Induction (DI):* $T \in$ DI if and only if T: $F(x) \to F'(x)$ such that $F' \neq F$, but $\Pr'(X_{ij} \in A) = \Pr(X_{ij} \in A)$ for all sets A with positive probability, and all ij. Transformations in DI redefine statistical dependencies among elements of the input without altering the marginal distributions.

3. *Equivalent Allocation (EA):* $T \in$ EA if and only if T: $g(X) \to g'(X)$ such that $g' \neq g$. Transformations in EA redefine the function from input to output without altering the allocation of sampling effort.

4. *Sample Allocation (SA):* $T \in$ SA if and only if T: $R_\bullet \to R'_\bullet$ such that $R'_\bullet \neq R_\bullet$. Transformations in SA redefine the allocation of sampling effort without altering the function from input to output.

5. *Equivalent Information (EI):* $T \in$ EI if and only if T: $h(Y_1) \to h'(Y_1)$ such that $h' \neq h$, but h and h' have identical arguments, $Y_1 \subset Y$. Transformations in EI redefine the function from output to statistic without altering the argument of the function.

6. *Auxiliary Information (AI):* $T \in$ AI if and only if T: $h(Y_1) \to h(Y_2)$ such that Y_1 and Y_2 are different subsets of Y. Transformations in AI redefine the argument of the statistic without altering the function, h.

Nelson and Schmeiser (1986a) prove that these classes of transformations are mutually exclusive, exhaustive, and nontrivial in the sense that each class contains transformations that reduce variance. In the present setting, they exhaust the potential solutions to (1), since under our formulation a solution corresponds to redefinition (transformation) of F, g, $R_\bullet$ and/or h.

There are infinitely many ways to transform a simulation experiment, but only a subset of them will reduce variance in any particular experiment. Well-known transformations that reduce variance in some experiments may be

ineffective or even increase variance in others. And worse yet, there is no theory to guide a search through the feasible transformations, because there is no general way to evaluate all potential VRTs except performing the experiment and estimating the variance. How, then, are VRTs ever applied? In practice, the experimenter depends on prior knowledge to select an appropriate VRT. *Prior knowledge* is any knowledge beyond what is necessary to construct the simulation experiment. There are several types of relevant prior knowledge: theoretical results (e.g., the Lehmann and Scheffé theorem; see EQUIVALENT INFORMATION below), experimenter knowledge (e.g., the sign of the correlation between two output random variables), definition of the model itself (e.g., mean of an input), or experimental results (e.g., pilot runs). From a practical standpoint, the available prior knowledge is another constraint on (1), even though we do not include it explicitly in the problem formulation since at present we have no taxonomy of prior knowledge.

2.2 Design and Analysis

Given (Ω, θ), we define a simulation experiment by specifying F, g, R_* and h, which in turn define X, Y and Z. We will call F, g and R_* the *design* of the experiment, since they define the outputs that will be generated, and h the *analysis*. Clearly these two aspects of the experiment are intimately related, since the desired analysis determines the appropriate design, and the feasible designs determine the appropriate analysis.

In all but trivial cases we cannot solve (1) as stated. In the following sections we examine further constrained versions of (1), beginning with the design fixed and only the analysis within our control, then progressively relaxing the constraints on the design and investigating other options. Seldom will an optimal solution to (1) be found even with additional constraints, but we will find improved solutions in the sense of reduced MSE. At the beginning of sections 3 through 8 we display the constraint that, along with (C.0), is enforced for the variance reduction problem (1).

Throughout the remainder of the paper, M_j and M_{ij} will denote the j^{th} column (sequence) and the ij^{th} element of a matrix M, respectively. Random variables are denoted by capital Roman letters, realizations by lower case

Roman letters, and other constants by lower case Greek letters. Finally, we use the term *crude experiment* to mean the original experiment prior to application of elemental transformations. Thus, "crude" is a relative, rather than absolute, term that depends on the experimenter.

3. EQUIVALENT INFORMATION (EI)

$$F, g, R_* \text{ and } Y_1 \text{ fixed} \tag{C.1}$$

Although the natural order in which to present the definition of simulation experiments is first the design (inputs, outputs, sampling plan) and then the analysis (statistics), to discuss variance reduction it is more natural to begin with the analysis. There are two reasons: First, variance reduction in the context of the analysis alone is that part of mathematical statistics dealing with optimality of estimators, and thus is well established. Secondly, if we know what conditions facilitate minimum variance estimation in the analysis, then we can use designs that produce those conditions, since simulation experiments often permit more flexibility in specifying the design than do classical sampling problems.

3.1 Variance Reduction Strategy

For the moment, we assume that we have a single output sequence $Y_1 = \{Y_{11}, Y_{21}, \ldots, Y_{I1}\}$ with unknown probability distribution, $p(.|\theta)$, which depends on θ (and possibly other parameters not currently of interest). The $\{Y_{i1}\}$ need not be independent or even identically distributed, although frequently they are. With F, g, R_* and a given output sequence Y_1 fixed as additional constraints, the solution to (1) is a transformation in EI that yields a statistic $h'(Y_1)$ with minimum MSE for estimating θ.

3.2 Results

To make determination of the best estimator possible, the constraint

$$F, g, R_* \text{ and } Y_1 \text{ fixed, and } \mathrm{E}[h'(Y)]=\theta \tag{C.1.1}$$

often replaces (C.1); i.e., h' is required to be *unbiased*. Under this constraint,

there may exist a uniformly minimum variance unbiased estimator (UMVUE) for all θ. When UMVUEs exist they are based on a sufficient statistic. For generality and for later discussion, we consider estimating any function of θ, say $q(\theta)$, where $q(.)$ may be the identity. The key theorem is:

Theorem (Lehmann and Scheffé). If $h(Y_1)$ is a complete, sufficient statistic for θ, and $s(Y_1)$ is unbiased for $q(\theta)$, then $h'(Y_1) = E[s(Y_1) | h(Y_1)]$ is a UMVUE of $q(\theta)$. If $\mathrm{Var}_\theta[h'(Y_1)] < \infty$ for all θ, then h' is the unique UMVUE for $q(\theta)$. (See Bickel and Doksum, 1977, page 122, for a proof.)

With the constraints (C.0) and (C.1.1), h' is the solution to (1). Unfortunately, the theorem is frequently of no practical value for finding h'. Also, h' and even s may not exist. For a discussion of minimum MSE estimators, see Kendall and Stuart (1979, pages 21-22).

Nelson and Schmeiser (1983) refer to VRTs increasing and/or making better use of the information available to estimate θ. Although we do not need a formal definition of statistical information here, it will be useful to consider an example based on the well-known Fisher information measure (Rao, 1973). Under certain regularity conditions, the minimum attainable variance for estimating $q(\theta)$ by any unbiased estimator $s(Y_1)$ of $q(\theta)$ is given by the Cramér-Rao lower bound $[dq(u)/du]^2/I_1(u)$ evaluated at $u = \theta$, where

$$I_1(u) = E_{Y_1}\left[\frac{\partial}{\partial u} \ln p(Y_1 | u)\right]^2 .$$

$I_1(\theta)$ is called the Fisher information measure. The bound shows that the minimum variance of any estimator of $q(\theta)$ is completely determined by the distribution of Y_1, $p(.|\theta)$, and the function of θ we estimate (in the present context $q(\theta) = \theta$, so $dq(\theta)/dy = 1$). The Fisher measure is typical of many concepts of statistical information. Thus, if we have a statistic whose variance achieves the lower bound, to achieve further reductions we must transform the experiment in a way that increases $I_1(\theta)$, and/or find an alternative representation of θ. Since the information depends on the distribution of Y_1, one way to increase the information is to change this distribution.

For example, consider the variance of the sample mean of a covariance stationary sequence Y_1.

$$\mathrm{Var}[I^{-1}\sum_{i=1}^{I} Y_{i1}] = I^{-1}\{\sigma^2 + 2\sum_{k=1}^{I-1}(1 - k/I)C_k\} \qquad (2)$$

where $C_k = \text{Cov}[Y_{i1}, Y_{j1}]$ when $|i - j| = k > 0$ and $\sigma^2 = \text{Var}[Y_{i1}]$. When the $\{Y_{i1}\}$ are independent, (2) reduces to the well-known expression σ^2/I. We consider only covariance stationary sequences because all results cited in this paper are based on (at least) that assumption. From (2) we see that the variance of a sample mean depends on the common variance of the scalar random variables, the covariance of pairs, and the number of observations. Changing these components affects the information about θ, and thus the variance of estimators of θ. There are variance reduction strategies, discussed in later sections, that concentrate on each of these components, and ones that use alternative representations of θ.

Although the asymptotic (large sample) properties of estimators have been widely studied, we do not discuss them here. Variance reduction is most relevant when obtaining even a small sample is costly. The reader is referred to Bickel and Doksum (1977), Kendall and Stuart (1979), Rao (1973) and references therein. However, if unbiasedness is not required, and if we cannot derive the variance or MSE directly, then the variance of the asymptotic distribution of candidate estimators can be used as a basis for comparison, and an asymptotically minimum variance estimator is usually preferred.

3.3 Examples

Consider the following example: Let $\{Y_{11}, Y_{21}\} \sim N(\theta, \theta, \sigma^2, \sigma^2, \rho)$, a bivariate normal distribution with common marginal mean and variance, and correlation coefficient ρ. Then the Cramér-Rao lower bound on the variance of any estimator of θ based on $\{Y_{11}, Y_{21}\}$ is $(1 + \rho)\sigma^2/2$. Note that the bound is lowered by decreasing σ^2 and ρ, which may be possible in simulation experiments. If $\rho = 0$ then $I_1(\theta) = \sigma^{-2} + \sigma^{-2}$, demonstrating that independent sources of information increase the total information additively. To affect the numerator of the lower bound we have to express θ as a function of another parameter (see AUXILIARY INFORMATION below).

When the output is a sequence of independently and identically distributed (i.i.d.) random variables, the sample mean is frequently the UMVUE of its expectation. When the outputs are not independent, knowledge of the joint distribution can lead to more efficient (smaller variance) estimators. Halfin (1982) derived optimal linear unbiased estimators for the mean of a stationary

stochastic process whose covariance function is a finite sum of decaying exponentials. For example, the covariance function of the queue length process of an M/M/s/c queue has this form. Of course, the covariance function of a simulation output sequence is seldom known, but Halfin noted (reassuringly) that the variance of the optimal estimator and the sample mean are the same asymptotically. His results illustrate the value of knowledge of the covariance structure of the outputs, and even knowledge less complete than Halfin assumes is often useful.

Bickel and Doksum (1977), Kendall and Stuart (1979), and Rao (1973) contain examples of UMVUEs and minimum MSE statistics for various estimation problems, a topic too large for consideration here. Many variance reduction strategies available for simulation experiments depend on augmenting Y_1 with additional outputs (auxiliary information), or redefining F, g and R_*, and then searching for a solution to (1). These strategies are addressed in sections 4 through 8.

4. AUXILIARY INFORMATION (AI)

$$F, g, R_*, \text{ and } h \text{ fixed} \qquad \text{(C.2)}$$

Auxiliary information (other simulation outputs not originally included in the analysis) has the potential to increase the available information about θ, making possible estimators with smaller variance than the crude estimator. Rarely can we simply augment or replace the output sequence Y_1 in the **same** function h and achieve a variance reduction, but it is possible. For instance, if we estimate θ by the sample mean of an i.i.d. sequence of random variables, Y_1 with $\text{Var}[Y_{i1}] = \sigma_1^2$, and there is another i.i.d. output sequence Y_2 with $\text{Var}[Y_{i2}] = \sigma_2^2 < \sigma_1^2$, then replacing Y_1 with the sequence Y_2 (of at least the same length) reduces variance. It is much more common to augment Y_1 with additional outputs Y_2, and to employ a different statistic h' to make use of the auxiliary information; i.e., using transformations from AI and equivalent information together. Thus, we relax constraint (C.2), replace it with (C.2.1) below, and investigate compositions of AI and EI.

$$F, g \text{ and } R_* \text{ fixed} \qquad \text{(C.2.1)}$$

After augmenting the original sequence with additional outputs, we again face the problem of selecting the best estimator. However, three broad (but not exhaustive) strategies are frequently employed to make use of auxiliary information. We will examine them individually because they indicate what kinds of auxiliary information and prior knowledge are useful for solving (1) by transformations in AI and EI. The three strategies are *control, conditional expectation,* and *indirect* strategies.

4.1 Control Strategies

Control estimators are statistics that attempt to correct the value of a crude estimator of θ using the discrepancy between the value of a second estimator and its known expectation. We characterize one class of control estimators as follows:

Suppose we have estimators $h(Y_1)$, $s(Y_2)$, and $h'(h, s)$ with the following properties:

(CV.1) $E[h] = \theta$, $E[s] \approx \alpha$ and α is known

(CV.2) $E[h'(h, s)] \approx \theta$

(CV.3) $h'(\theta, \alpha) = \theta$

(CV.4) $h'(h, \alpha) = h$

(CV.5) $|h'(h, \eta) - h|$ is nondecreasing in $|\eta - \alpha|$

Property (CV.1) establishes that h and s are unbiased estimators of their respective estimands θ and α; h is the crude estimator, θ is the parameter of interest and α is assumed known. Property (CV.2) establishes that h' is a useful estimator of θ (we leave the $\approx$ vague, but usually interpret it to mean that h' is at least a consistent estimator of θ). We make use of (CV.3) below. Property (CV.4) implies that no correction occurs if s is equal to its expectation, and (CV.5) establishes that the correction is greater as the discrepancy between s and α increases. We call estimators with properties (CV.1) through (CV.5) *control variate estimators (CV),* and we will call s the

control variate. When h and s are functions of outputs from the same simulated system the VRT is called *internal* CV, and when s is a function of outputs from a similar simulated system the technique is called *external* CV.

The most widely used form of control variate estimator is the univariate linear or regression CV, $h'(h, s) = h - \lambda(s - \alpha)$, although there are other potentially useful forms (Nelson, 1987). One attractive feature of the linear CV is that it readily generalizes to a regression on multiple control variates $s_k(Y_{k+1})$, $k = 1,2,..., q$,

$$h'(h, \{s_k\}) = h - \sum_{k=1}^{q} \lambda_k (s_k - \alpha_k)$$

and also generalizes to estimation of multivariate θ.

4.1.1 Results for Control Strategies

Using the first three terms in a Taylor series expansion, and invoking (CV.3), Nelson (1987) showed that to the order of the approximation the variance of h' — the general CV characterized by (CV.1) through (CV.5) — will be less than the variance of h if

$$\left[\frac{\partial h'}{\partial s}\right]^2 \mathrm{Var}[s] < -2\frac{\partial h'}{\partial s}\mathrm{Cov}[h, s]$$

where the partial derivatives are evaluated at (θ, α). Thus, for the CV to be effective the covariance between h and s must be large enough to counteract the variance introduced by incorporating s into the estimator h'. Auxiliary outputs that are not correlated with h are useless in this strategy. However, in simulation experiments we frequently have the capability to induce correlation when it is not inherent (see DEPENDENCE INDUCTION below).

For any constant λ the univariate linear CV is unbiased. The optimal choice of λ that minimizes the variance of h' is $\lambda^* = \mathrm{Cov}[h, s]/\mathrm{Var}[s]$; a similar result holds for the multivariate linear control. Unfortunately, estimating λ^* from Y_1 and Y_2 often causes h' to be biased.

Consider the special case when h and s_k, $k = 1,2,...,q$, are sample means of i.i.d. vectors $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i,q+1})$, $i = 1, 2,..., I$. Further, suppose that each

vector has a q+1-variate normal distribution with mean $\mu = \{\theta, \alpha_1, \ldots, \alpha_q\}$ and variance-covariance matrix Σ, where θ and Σ are unknown but the $\{\alpha_k\}$ are known. This assumption might be satisfied if $\mathbf{Y}_i$ results from averaging outputs in a single simulation run (independent replication). Under these assumptions the linear control is the conditional expectation (regression) of Y_{i1} on $\mathbf{Y}_i \setminus Y_{i1}$, and the problem of estimating θ and $\{\lambda_k\}$ can be viewed from the perspective of classical regression theory (Lavenberg and Welch, 1981). The key results are:

1. The CV estimator of θ is unbiased.

2. A valid confidence interval for θ can be constructed.

3. The ratio of the variance of the control variate estimator h' to the variance of h is $(1 - R_{YX}^2)(I - 2)/(I - q - 2)$, where R_{YX}^2 is the square of the multiple correlation coefficient of Y_{i1} and $\mathbf{Y}_i \setminus Y_{i1}$.

The last result shows that the decision to use more control variates involves a trade-off: Although $1 - R_{YX}^2$ is nonincreasing as control variates are added, the loss ratio $(I - 2)/(I - q - 2)$ is monotone increasing in q. Also, the variance reduction when the optimal $\{\lambda_k\}$ are known, which is $1 - R_{YX}^2$, cannot be achieved if we have to estimate the $\{\lambda_k\}$. Corresponding results for estimation of multivariate θ have been developed by Rubinstein and Markus (1985), Venkatraman (1983), and Venkatraman and Wilson (1986). Nozari, Arnold and Pegden (1984a) and Porta Nova and Wilson (1986) give similar results when the elements of θ are the parameters of a general linear model.

4.1.2 Examples of Control Strategies

There is a tremendous literature on control variate estimators. For a summary of work through 1974 see Kleijnen (1974). An interesting application in computer performance modeling appears in Lavenberg, Moeller and Welch (1982). The work of Wilson and Pritsker (1984ab) is particularly relevant in light of the results cited above: In the simulation of a network of q queues, Wilson and Pritsker used standardized sums of independent and identically distributed service time random variables at each queue as internal CVs. The standardization was done in such a way that the asymptotic distribution of the q control variates is q-variate normal with zero mean vector and variance-covariance matrix the q x q identity matrix. Thus, asymptotically, the

assumptions stated above are satisfied for these CVs. Wilson and Pritsker (1984b) summarizes a thorough experimental evaluation for four versions of a classical machine-repair system. They reported variance reductions in the range of 20% to 90% for estimators of machine utilization, repairman utilization, and expected repair time.

4.2 Poststratifying, A Different Control Strategy

An example of another control, but not control variate, VRT is *poststratifying the sample (PSTRAT).* PSTRAT is often viewed as a special case of stratified sampling since both VRTs have the same asymptotic variance, but our taxonomy shows how they are different. Like CV, PSTRAT uses auxiliary information to correct a crude estimator. Unlike CV, PSTRAT corrects for disproportionate sampling rather than for location. Stratified sampling, a sample allocation strategy, is discussed in the next section.

One source of variance in estimating θ is that the empirical distribution of the $\{Y_{i1}\}$ will almost surely not match the theoretical distribution. Of course the distribution of $\{Y_{i1}\}$ is unknown in general, so there is no way to measure how significant the deviation is. However, consider the following well-known relationship (Bickel and Doksum, 1977):

$$\theta = \mathrm{E}[Y_{i1}] = \int \mathrm{E}[Y_{i1} \mid Y_2 = y_2] dP(y_2) \tag{3}$$

where P is the distribution of Y_2, some auxiliary output sequence. If P is known, then one strategy is to correct for disproportionate sampling from P, indirectly correcting for disproportionate sampling of $\{Y_{i1}\}$.

For example, suppose we can pair outputs (Y_{i1}, Y_{i2}) so that different pairs are independent, and we can divide the range of Y_{i2} into n nonoverlapping intervals (strata), L_j, $j = 1,2,\ldots, n$. Then a special case of (3) is

$$\theta = \mathrm{E}[Y_{i1}] = \sum_{j=1}^{n} \mathrm{E}[Y_{i1} \mid Y_{i2} \in L_j] \Pr(Y_{i2} \in L_j) \tag{3'}$$

Let I_j be the number of $\{Y_{i2}\}$ falling in stratum j $(\Sigma I_j = I)$, and let $p_j = \Pr(Y_{i2} \in L_j)$ which is assumed known and constant over all i. Then if Y_{ij1} is the i^{th} observation of Y_1 when $Y_{i2} \in L_j$, the PSTRAT estimator is

$$h'(Y_1, Y_2) = \sum_{j=1}^{n} \sum_{i=1}^{I_j} \frac{p_j}{I_j} Y_{ij1}$$

Conditional on $I_j > 0$, the PSTRAT estimator is unbiased for θ if Y_{i1} is. Whereas a sample mean gives each observation weight $1/I$, the PSTRAT estimator gives weight p_j/I_j. PSTRAT checks the empirical distribution of $\{Y_{i2}\}$ over the strata and corrects for discrepancies. If the observations distribute themselves proportionately ($I_j = p_j I$) then $p_j/I_j = 1/I$. If a stratum is over or underrepresented probabilistically, then p_j/I_j is less or greater than $1/I$, respectively. The general relationship (3) is used to break up the range of Y_{i2} and thus exploit the (hopefully strong) relationship between Y_{i1} and Y_{i2}. The conditional expression (3) is exploited in several other variance reduction strategies.

4.2.1 Results for Poststratification

To terms of $O(I^{-2})$ the variance of the PSTRAT estimator is (Cochran, 1977)

$$I^{-1}\sum_{j=1}^{n} p_j \mathrm{Var}[Y_{i1} \mid Y_{i2} \in L_j] + I^{-2}\sum_{j=1}^{n}(1-p_j)\mathrm{Var}[Y_{i1} \mid Y_{i2} \in L_j]$$

The first term is the same as the variance of proportional stratified sampling (see below), so results that give the optimal strata boundaries L_j for stratified sampling can be used to determine nearly optimal boundaries for PSTRAT. Once we decide to use PSTRAT to solve the variance reduction problem (1), the only decisions that can be made to enhance its effectiveness are the selection of the stratification variable and the selection of the strata boundaries. Kleijnen (1974) derives confidence interval procedures for the PSTRAT estimator, and discusses procedures for dealing with empty ($I_j = 0$) strata.

4.2.2 Examples of Poststratification

Despite its simplicity, few examples of PSTRAT have appeared in the simulation literature. See Wilson and Pritsker (1984ab), who used their standardized control variates as poststratification variables as well.

4.3 Conditional Expectation Strategy

Conditional expectation estimators are based on the following well-known result

$$\mathrm{Var}\{\mathrm{E}[Y_{i1} \mid Y_{j2}]\} = \mathrm{Var}[Y_{i1}] - E\{\mathrm{Var}[Y_{i1} \mid Y_{j2}]\} \tag{4}$$

which, based on (3), suggests that if we can find an output sequence $\{Y_{j2}\}$ for which the conditional expectation $E[Y_{i1} | Y_{j2} = y_{j2}]$ can be calculated for all y_{j2}, then we might want to replace $\{Y_{i1}\}$ by $\{E[Y_{i1} | Y_{j2}]\}$ in h; this is the *conditional expectation (CE)* estimator h'.

Dramatic variance reductions from CE have been reported in Monte Carlo estimation of location and dispersion; Simon (1976) gives a survey. However, these applications are characterized by the existence of an independent sequence $\{Y_{j2}\}$ on which to condition. In simulation experiments the $\{Y_{j2}\}$ are usually dependent outputs from a stochastic process, implying that (4) alone is not enough to ensure a variance reduction.

4.3.1 Results for Conditional Expectation Strategies

If h is a sample mean the CE estimator h' that averages $\{E[Y_{i1} | Y_{j2}]\}$ is unbiased for θ if Y_{i1} is unbiased. If both $\{Y_{i1}\}$ and $\{E[Y_{i1} | Y_{j2}]\}$ are sequences of i.i.d. random variables of the same length, then the CE estimator will have no larger variance than the sample mean of $\{Y_{i1}\}$ (since from (4) each summand has no larger variance).

Unfortunately, when the sequences are not independent there are no readily verifiable conditions that guarantee the CE estimator will have smaller variance. Suppose we have covariance stationary sequences $\{Y_{i1}\}$ $i=1, 2,..., I_1$ and $\{E[Y_{i1} | Y_{j2}]\}$ $j=1, 2,..., I_2$. Then we know that $\text{Var}\{E[Y_{i1} | Y_{j2}]\} \leq \text{Var}[Y_{i1}]$ for any combination of i and j. The expression (2) for the variance of a sample mean suggests that it would be useful to have a condition that ensures $\Sigma_k \text{Cov}\{E[Y_{i1} | Y_{j2}], E[Y_{i1} | Y_{j+k,2}]\} < \Sigma_k \text{Cov}[Y_{i1}, Y_{i+k,1}]$ as well. However, CE is often used when $I_2 \gg I_1$, so that one may expect a variance reduction based only on the relative number of observations.

4.3.2 Examples of Conditional Expectation Strategies

In simulation experiments on models of dynamic systems there have been successful applications of CE that computed expectations of rare events conditional on the system state (Y_{j2} in our development) at particular points in time. Carter and Ignall (1975) estimated the expected response time of fire fighting equipment to serious fires by conditioning on the disposition of

equipment at randomly selected times. Lavenberg and Welch (1979) estimated the expected delay for jobs at various devices in a multiprogrammed computer system in a similar manner. That these applications yielded variance reductions depended more on the increased sample size they produced, and on the intuition of the researchers, than on (4).

On the other hand, Ross and Schechner (1985) were able to prove that their CE strategy is effective in certain cases. Looking at both discrete and continuous time Markov chains, they estimated parameters related to the time until the process enters a fixed subset of states, ψ. At each transition of the process, they computed the conditional probability that the next transition would take the process into ψ, and based their estimators on these "observed hazards." An interesting side note is that the new estimator of convolutions of random variables reported by Ross and Schechner (1985, p. 233) is suggested by an earlier conditioning argument used in the simulation of stochastic networks (Burt and Garman, 1971). In Burt and Garman's work the time to traverse a path through the network is a convolution. Stochastic network simulation is an application in which CE strategies have been extremely effective (see for instance Sigal, Pritsker and Solberg, 1980, Fishman 1985ab).

4.4 Indirect Strategies

Suppose that we can write $\theta = q(\gamma)$ and there is an output sequence $Y_2 = \{Y_{12}, \ldots, Y_{I2}\}$ such that $Y_2 \sim p(Y_2|\gamma)$; the parameter γ could be multivariate, but we assume it is a scalar here. Then it may better to estimate γ directly by a statistic, say $t(Y_2)$, and estimate θ indirectly by $h(Y_2) = q[t(Y_2)]$. We call such strategies *indirect estimators (INDIR).*

4.4.1 Results for Indirect Strategies

From the Cramer-Rao lower bound (section 3.2), it appears that INDIR will be effective if the information about γ is greater than the information about θ, and/or $[dq(u)/du]^2$ evaluated at $u = \gamma$ is less than one. Unfortunately it is difficult to say anything more definitive. INDIR strategies are problem specific because of the need to express θ as $q(\gamma)$, so it is difficult to predict whether they will be effective in new situations. Also, there is no general way to select the

optimal $q(\gamma)$ from among several alternative representations of θ. Finally, INDIR estimators may be biased unless $q(\gamma)$ is linear in γ. However, if effective functional relationships can be found for very broad classes of models (all queueing networks, for example), then INDIR could be a powerful strategy.

4.4.2 Examples of Indirect Strategies

Because simulation experiments are performed on models of real or conceptual systems, functional relationships among parameters of interest are sometimes known from physical properties of the system. Law (1975) and Carson and Law (1977, 1980) exploited functional relationships among the following steady-state parameters of GI/G/s queues: expected customer sojourn time in the system, expected number of customers in the queue, expected number of customers in the system, and the expected amount of work in the system. They considered indirect estimators of these quantities as functions of a direct estimator of γ, the expected customer delay in the queue, and showed analytically that the asymptotic variance of the INDIRs is smaller than the corresponding variance of the direct estimator. Their results were derived using regenerative analysis (Bratley, Fox and Schrage, 1983), and suggest that, in queueing simulation, statistics based on time spent in the system may have smaller variance than statistics based on the number in the system. However, when Cooper (1981, pages 293-295) compared two estimators of server utilization in an M/G/1/1 queue, one based on number of customers lost and the other based on total busy time, he found that the variance depends on the ratio of the arrival rate to the service rate, and that neither estimator is uniformly superior over all values of this ratio. Thus, the results for GI/G/s queues do not generalize to all types of queues.

Similarly, Minh and Sorli (1983) expressed the expected customer delay in queue (γ) for the GI/G/1 queue as a function of the expected duration of server idle periods (η), and suggested estimating η directly and γ indirectly. Again using regenerative analysis, they proved that for the M/G/1 queue the asymptotic variance of the INDIR is smaller than the corresponding variance of the direct estimator of γ, and that the variance decreases further as the traffic intensity approaches one (the performance of estimators in queueing simulation often deteriorates in heavy traffic). This suggests combining the Minh and Sorli INDIR for γ with the Carson and Law INDIRs for the other quantities.

5. SAMPLE ALLOCATION (SA)

F and g fixed (C.3)

VRTs whose composition includes transformations in SA depend on a simple principle: For any reasonable estimator, the variance of the estimator (and the bias if it is consistent) decreases as the number of observations on which is it based increases. Mathematically, if Z_I is a statistic based on a sequence Y_1 of length I, then $\text{Var}[Z_{I'}] < \text{Var}[Z_I]$ if $I' > I$. If Z_I is a sample mean and Y_1 consists of i.i.d. random variables with common variance σ^2, then we have the well-known result $\text{Var}[Z_I] = \sigma^2/I$. More generally (2) gives the variance as a function of I. The most basic variance reduction strategy is to increase the length of the output sequence, but this involves additional computer effort and the rate of reduction is slow ($O(I^{-1})$). However, when the parameter of interest is a function of several parameters, distributing a fixed amount of sampling effort intelligently among several output sequences can lead to substantial variance reduction.

5.1 Variance Reduction Strategy

Consider a statistic h that is a function of n output sequences, Y_j, $j = 1,2,...,n$. Let the length of the j^{th} sequence be I_j. Then the variance of $Z = h(Y_1, \ldots, Y_n)$ is a function of the $\{I_j\}$. As a particular example, suppose that the Y_j are individually composed of i.i.d. random variables with variances σ_j^2, the sequences are mutually independent, and h is a linear combination of the sample means $I_j^{-1}\Sigma Y_{ij}$, $j = 1,2,...,n$. Then

$$\text{Var}[Z] = \sum_{j=1}^{n} c_j^2 \sigma_j^2 I_j^{-1} \tag{5}$$

where the $\{c_j\}$ are constants depending on the linear combination. When the elements of each sequence Y_j are not independent the same result applies with σ_j^2/I_j replaced by (2). If the sequences themselves are dependent, the expression is more complicated.

As a general strategy, we want to allocate sampling effort to sequences that contribute the most to the variance of the estimator. Which sequences contribute the most depends not only on the distribution of the outputs, but also

on the estimator used. In (5), $\{\sigma_j\}$ and $\{c_j\}$ quantify the respective contributions.

One sample allocation strategy is to again use (3') to express θ as a function of conditional expectations. However, instead of the $\{I_j\}$ being random outputs as in PSTRAT, they will be fixed as part of the sampling plan, R_*. The problem then becomes estimating $E[Y_{i1} | Y_{i2} \in L_j]$ for $j = 1, 2, \ldots, n$ individually, and allocating the I available observations of (Y_{i1}, Y_{i2}) in a way that minimizes the variance of the combined estimator of θ. This VRT is called *stratified sampling (STRAT).*

A strategy related to STRAT is *splitting (SPLT),* which is particularly useful when Y_{i2} and Y_{i1} are realized in sequence; i.e. a realization of Y_{i2} is generated, then a realization of Y_{i1} follows, for $i = 1, 2, \ldots$. SPLT allocates sampling effort to estimate $E[Y_{i1} | Y_{i2} \in L_{j^*}]$ for some particular (or in general several) j^*. However, rather than directly allocating I_{j^*} observations to the sequence $\{Y_{i2}\}$, that portion of the total sample I remains unspecified. Instead, each time the event $[Y_{i2} \in L_{j^*}]$ occurs naturally, m observations of $\{Y_{i1}\}$ — conditional on the current Y_{i2} — are generated. Thus, SPLT is a more dynamic version of STRAT. This complicated sampling plan is represented in R_* by only allowing realizations of Y_{i1} to occur in multiples of m when $[Y_{i2} \in L_{j^*}]$. SPLT has been called a special case of importance sampling, but we define importance sampling in a way that separates the two (see DISTRIBUTION REPLACEMENT below).

5.2 Results

Given the strata, $\{L_j\}$, the *optimal allocation* for STRAT is (Cochran, 1977)

$$I_j^* = I \frac{p_j \sigma_j}{\sum p_i \sigma_i}$$

where $\sigma_j^2 = \mathrm{Var}[Y_{i1} | Y_{i2} \in L_j]$. Notice that the optimal allocation depends on the variance within the stratum (σ_j^2) and the likelihood of an observation coming from the stratum (p_j). Of course, σ_j^2 is seldom known, so *proportional allocation* ($I_j = p_j I$), which only takes into account the likelihood, is often used.

The variance under proportional allocation is the same as the first term in the variance of PSTRAT, and is smaller than the variance of a sample mean based on simple random sampling (Rubinstein, 1981). Kahn (1956) showed that the optimal choice of m for SPLT also depends on $\sigma_{j^*}^2$, the variance within stratum L_{j^*}.

The preceding results assumed the strata $\{L_j\}$ were given. However, intelligent selection of the strata is part of the solution to the variance reduction problem (1) via STRAT. In other words, we must also determine the way the sequences in Y will be partitioned. The optimal strata boundaries under optimal and proportional allocation schemes are known when the stratification variable is Y_{i1} itself (Dalenius, 1950, Sethi, 1963). When a secondary stratification variable Y_{i2} is used — usually the only feasible option in simulation experiments — it is necessary to assume a functional relationship between Y_{i1} and Y_{i2} to derive optimal boundaries (Dalenius and Gurney, 1951). The optimal boundaries depend not only on the variance within the strata, but also the variation between them.

5.3 Examples

The survey sampling literature abounds with examples of STRAT, often combined with other sampling techniques. A good general reference is Cochran (1977). Unfortunately, as Kleijnen (1974) points out, even when a stratification variable can be identified it is difficult to control the sample allocation in dynamic stochastic simulation. Good candidates for stratification variables in dynamic simulation are initial conditions that are selected randomly at the beginning of independent simulation replications. Bratley, Fox and Schrage (1983) give an example in the simulation of a bank where the expected number of customers served each day (total arrivals minus those that balk) is of interest. The number of tellers at work each day is a random variable. By conditioning on the number tellers at work, they allocate (rather than sample) the days when 1, 2 or 3 tellers are available in a way that reduces variance.

Kioussis and Miller (1983) use SPLT to estimate the probability of system failure, a rare event, in a fault tolerant computer system. They condition on the number of active faults (Y_{i2} in our development) to replicate situations when system failure is likely. See also Hopmans and Kleijnen (1979).

6. EQUIVALENT ALLOCATION (EA)

$$F, R_* \text{ fixed} \tag{C.4}$$

The function g defines how inputs sampled from a static probability distribution F are transformed into outputs that describe a dynamic stochastic process. The logic of the system of interest is embodied in g. VRTs are infrequently composed of transformations in EA because changing g can mean simulating a fundamentally different stochastic process. However, if g **itself** can be modeled, then the effect of such a radical change might be known in advance. Transformations in EA, by redefining g, can also provide additional outputs necessary for strategies employing auxiliary information; we discuss this idea later.

6.1 Variance Reduction Strategies

A fundamental property of statistical information is that information about an unknown parameter, θ, contained in a random variable, X, may be decreased, but not increased, by a function of the random variable, $Y = g(X)$, provided the function does not depend on θ. Of course, the information in Y may be more useful in the sense that we know how to estimate θ from Y but not from X. With F fixed, solving the variance reduction problem (1) is facilitated by a function g' that preserves all of the information about θ in X and still permits estimation of θ from Y. The function g is usually implicit (represented by computer code), complicating the search for g'.

Since g represents the logic of the system of interest, one approach is to model g itself as a stochastic process. A stochastic process with sufficient generality to model a wide range of simulation experiments, yet having enough structure to permit generic results to be derived and applied, is needed. One promising candidate is the generalized semi-Markov process (GSMP). A complete treatment of GSMPs is beyond the scope of this paper (see for instance Glynn, 1983 and Whitt, 1980, and references therein), but we briefly discuss their structure.

A GSMP models the dynamic changes in a discrete-event simulation by transitions from one of a countable set of *states* to another. A state transition is

caused by one of a countable set of *events*. An event is triggered by the expiration of its *clock*, and there is one clock associated with each pending event. Critical specifications for a GSMP are the state transition probabilities and the probability distributions of the new clock settings after the occurrence of an event. Clearly this characterization describes many discrete-event simulations, but it also has sufficient structure to permit useful results to be derived. The reader is encouraged to see Whitt (1980) for conditions ensuring the existence of invariant (steady-state) probability distributions for GSMPs. Several researchers are using the GSMP framework to study simulation output analysis, and significant results for variance reduction should be forthcoming. One EA strategy suggested by the GSMP framework is to simulate an embedded stochastic process made obvious by the GSMP representation, rather than the original process.

6.2 Examples

We consider a simple example of EA alone, which is estimating the integral

$$\theta = \int_0^1 g(x)dx$$

where $0 \le g(x) \le 1$ in the region of integration. In "hit-or-miss" Monte Carlo, we sample inputs $\{X_{1j}, X_{2j}, \ldots, X_{Ij}\} \sim$ i.i.d. $U(0,1)$ for $j = 1, 2$, where $U(0,1)$ denotes the uniform distribution on the interval (0,1), then estimate θ by

$$Z = I^{-1}\sum_{i=1}^{I} 1_g(X_{i1}, X_{i2})$$

where 1_g is an indicator such that $1_g(X_{i1}, X_{i2}) = 1$ if $X_{i2} \le g(X_{i1})$, and 0 otherwise. The summation counts the number of "hits", where a hit occurs if a point generated from a uniform distribution on the unit square is under the curve $g(x)$. The ratio of hits to trials estimates the fraction of the unit square under $g(x)$, its area in this example, and $\mathrm{Var}[Z] = \theta(1-\theta)/I$. An example of EA is replacing the output function 1_g with $g(X_{i1})$, which always reduces $\mathrm{Var}[Z]$. While the new output function is derived from knowledge of the conditional expectation, the transformation is EA since additional information is created that could not have been obtained from the original outputs.

For an example of simulating an embedded process, see Fox and Glynn (1986). We encounter the situation when EA is employed to provide auxiliary information in section 8 below.

7. DEPENDENCE INDUCTION (DI)

marginals of F fixed (C.5)

We have seen that functional and statistical relationships between outputs and parameters of interest can be exploited to reduce variance. For example, linear correlation of sufficient magnitude facilitates control strategies. Also, expressions such as (2) show that dependence between outputs is a component of the statistic's variance. Transformations in DI, by redefining the dependence between inputs, can change both the marginal distributions and dependence structure of the outputs. However, variance reduction strategies employing DI usually induce dependence between outputs without changing the marginal distributions of the outputs, either. DI is a well-known variance reduction strategy for Monte Carlo estimation, and is the most widely used strategy in simulation. Research has concentrated on three areas: 1) strategies for inducing the desired dependence among inputs, 2) strategies for preserving the dependence in the outputs, and 3) strategies for design and analysis under induced dependence. In this section we concentrate on strategies where DI reduces variance directly, but DI can facilitate any strategy that exploits dependence.

Because much of the work on dependence induction is intimately tied to the methods used to generate realizations of X, we introduce some additional notation to represent that aspect. From a practical perspective, variate generation can almost always be described as follows: Let $U_j = \{U_{1j}, U_{2j}, \ldots, U_{mj}\}$ be a vector of random variables, independent and with identical marginal distributions that are $U(0,1)$. Realizations of $X_j = \{X_{1j}, X_{2j}, \ldots, X_{nj}\}$ are generated from realizations of U_j via a function e_j; that is $X_j = e_j(U_j)$. The function e_j need not map a single U_{ij} into a single X_{ij}. To generate a realization of a scalar X_{ij} with cdf F_{ij}, the most commonly used function is the *inverse cdf* $e_j(U_{ij}) = F_{ij}^{-1}(U_{ij})$, where F_{ij}^{-1} is the inverse function of F_{ij}. The inverse cdf has the attractive property of being monotone in U_{ij}, which is useful for inducing dependence (see below).

The sample space definition does not encompass variate generation; it stresses the resulting distribution F and not the methods used to generate realizations of X or methods used to induce dependence. Thus, it is insensitive to whether the $\{U_{ij}\}$ are random or pseudorandom (deterministically generated, but appearing random). However, a useful extension of the sample space definition is to view Ω as being induced by the functions e_j from a more basic sample space, $(0,1)^\infty$, the infinite dimensional unit hypercube that is the sample space of the $U(0,1)$ variates. However, because variance reduction concerns the population variance of Z, and the variance is determined by F rather than the particular variate generation algorithm, the sample space definition does not specify a variate generation scheme.

While the class DI is defined in terms of F, as a practical matter dependence is often induced by making the $\{U_{ij}\}$ functionally dependent. This is accomplished by sampling one m-dimensional U_j randomly and letting $U_i = d_i(U_j), i \neq j$, for some function d_i. The function d_i is chosen to have the property that the marginal distribution of U_i is still m-dimensional uniform.

There is some disagreement regarding appropriate analysis when the basic sample space is $(0,1)^\infty$, but the uniform variates are actually pseudorandom. In this case, a single "seed" ξ_j and generator completely determine the j^{th} sequence U_j. To facilitate dependence induction, simulation experiments may employ multiple random number sequences, requiring the selection of multiple starting seeds. The seeds $\{\xi_j\}$, and thus the sequences $\{U_j\}$, are under the complete control of the experimenter. To legitimately treat the the simulation output as a random sample, Mihram (1974) suggested that some or all of the seeds must be selected randomly from the set of possible seeds. However, this view is not universally held, and the debate will not be settled here. We raise the issue to point out that the sample space definition is not affected by it, but that it may be fundamental to the statistical analysis. For further discussion, see Kleijnen (1985), Mihram (1974), Schruben and Margolin (1978), and Wilson (1984).

7.1 Variance Reduction Strategies

We begin by considering a simplified version of the variance reduction problem (1), which is to minimize $\mathrm{Var}[Z]$ when

$$Z = I^{-1}\sum_{j=1}^{I} g_j(X_j) \tag{6}$$

where $X_j = \{X_{1j}, X_{2j}, \ldots, X_{nj}\}$, $X_j = e_j(U_j)$, and $Y_{j1} = g_j(X_j)$ is a scalar valued output. Assume initially that the $\{U_j\}$ are independent vectors of finite length m, implying that the vectors $\{X_j\}$ (and corresponding $\{Y_{j1}\}$) are also independent, although not necessarily identically distributed. The problem then becomes selecting a joint distribution for $\{X_1, \ldots, X_I\}$ such that the scalar marginal distributions are preserved and the variance of Z is minimized. In other words, we search for the optimal DI transformation. In the special case where the $\{Y_{j1}\}$ form a covariance stationary sequence, the variance of Z is given by (2). We induce dependence among the inputs in hopes of realizing a favorable covariance structure in the outputs. Of course, we would like to specify the joint distribution of the outputs directly, since we cannot guarantee that dependence induced between inputs (e.g. negatively correlated) will be reflected in the outputs unless we know quite a bit about g (e.g. monotone). Unfortunately, we do not know the distribution of Y, and g is usually implicit in the simulation code.

7.2 Results

The results we cite are sometimes called "antithetic variates theorem" results, although we have couched our problem in terms that also include "common random numbers" results. Given the sample space definition of simulation experiments, the distinction is irrelevant. The definition of the experiment includes **all** parameters to be estimated, whether they are absolute system parameters of interest, the relative difference between the parameters of two or more systems, or the parameters of a statistical model (a case not covered by (6)). For example, if $g_j = g$ for all j and the X_j are identically distributed, then (6) estimates an absolute parameter. On the other hand, if g_{2j-1} is the output of one system, and g_{2j} is the negative of the output of a second system, then (6) estimates a difference. From a practical perspective it is often useful to think of antithetic variate strategies as generating compensating (negative) dependence for estimating an absolute measure, and common random number strategies as guaranteeing homogeneous experimental conditions (positive dependence) for estimating relative performance measures. But again,

in our definition there is only one experiment, no matter how many conceptual systems are involved. See Nelson and Schmeiser (1986b) for more traditional definitions of *antithetic variates* and *common random numbers.*

The results cited below guarantee, in various situations, that the optimal solution to (1) when Z is given by (6) can be achieved by inducing functional dependence among the $\{U_j\}$ in a way that leaves their marginal distributions, and the marginal distributions of the $\{X_j\}$, unchanged. These are remarkable results, because they show that optimal dependence induction can be accomplished at the variate generation level. Unfortunately, some of the results guarantee only the existence of an optimal set of functions $\{d_i\}$, and are no direct help in finding them. Figure 1 organizes some references to important theoretical results, categorizing them in terms of the number of outputs, I, the dimension of U_j, m, and the dimension of X_j, n, in the experiment.

Hammersley and Mauldon (1956), Handscomb (1958), and Wilson (1979, 1983) provide existence theorems, first for bounded, and later for unbounded g_j. Whitt (1976) also gives an existence theorem, but for a more general problem which includes (6) as a special case. Fishman and Huang (1983) and Roach and Wright (1977) find optimal functions $\{d_i\}$ among certain restricted classes of functions. Snijders (1984) derives the optimal functions when the $\{X_j\}$ are identically distributed, binary random variables. Rubinstein, Samorodnitsky and Shaked (1985) show that a particular simple function d_j is optimal given that X_j is generated via the inverse cdf and g_j has certain monotonicity properties. McKay, Beckman and Conover (1979) investigate an effective class of induction schemes for multidimensional problems that they call *Latin hypercube sampling;* also see Stein (1985) for extensions of McKay et al. Finally, Granovsky (1983) provides an existence theorem that extends Whitt (1976).

The results of Fishman and Huang (1983) and Roach and Wright (1977) are special cases of what are known in sampling theory as *systematic sampling (SYS)* plans (Madow and Madow, 1944). The distinction between SYS and strategies based on sample allocation (SA) has been unclear in the past, but in light of our taxonomy the distinction is obvious. Antithetic variates, common random numbers, Latin hypercube sampling, and SYS induce dependence (functionally) among randomly sampled inputs, and realize variance reductions

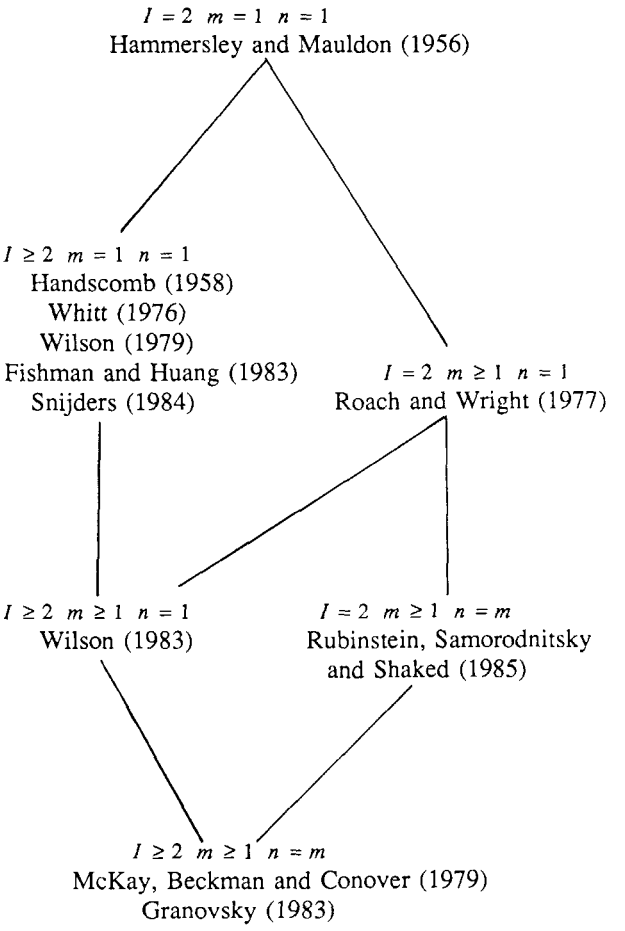

FIG. 1. Cases of the DI Problem (6).

because of favorable dependence among the outputs. STRAT and other SA strategies realize variance reductions by deterministically allocating sampling effort in the outputs where it does the most good. Although we have never seen an application, there is no inherent reason why DI and SA strategies cannot be used together.

The most common way to apply DI transformations is to hold the marginal distributions of the inputs fixed and induce dependence via the variate

generation algorithm. The references in Figure 1 illustrate this approach for estimators like (6). Monotonicity is the key property for proving the effectiveness of DI strategies. Using the inverse cdf guarantees a monotone mapping of U_{ij} into X_{ij}. If the $\{g_j\}$ are *concordant,* meaning that with respect to each component of X_j they are monotone in the same direction, then it is often possible to prove that a DI strategy reduces variance (see Rubinstein et al. and McKay et al., for example).

Assuming that some known dependence can be induced, by whatever means, how can it best be used? The correlation induction strategies of Schruben (1979) and Schruben and Margolin (1978) are one answer. They present experimental designs for estimating the parameters $\theta = (\theta_1, \ldots, \theta_p)$ of a general linear model

$$Y_i = \sum_{j=1}^{p} \theta_j w_{ij} + \varepsilon_i$$

where w_{ij} is the setting of the i^{th} experimental factor at the j^{th} design point, and ε_i is a random error term. Assuming that a particular covariance matrix for the responses (outputs) $\{Y_i\}$ can be induced, these designs reduce the determinant of the covariance matrix for ordinary and generalized least squares estimators of θ as compared to independent sampling and other dependence induction strategies. This is a distinctly different approach, because it specifies an optimal experimental design under induced dependence, rather than specifying how the dependence is induced. However, Schruben and Margolin's assumptions that lead to the known correlation structure are somewhat controversial. See Kleijnen (1985), Nozari, Arnold and Pegden (1984b) and Tew and Wilson (1985) for up-to-date treatments.

7.3 Examples

Techniques for generating bivariate inputs with extremal distributions (maximum or minimum possible covariance with given marginal distributions) are based on the inverse cdf (see Whitt, 1976). DI is more difficult when the variate generation algorithm does not monotonically transform a fixed length vector U_j into X_j. Variate generation methods for dependence induction are outside the scope of this paper, however, Bratley, Fox and Schrage (1983) is a

good general reference, and Cheng (1985), Fishman and Moore (1984), and Schmeiser and Kachitvichyanukul (1986) give some variate generation algorithms that facilitate dependence induction.

Cooley and Houck (1982) used the Schruben and Margolin designs in response surface methodology (RSM) for optimization problems when the objective function is evaluated via simulation. RSM requires fitting low order polynomial models to the system response under different configurations (the decision variables) of the simulated system. Dependence induction reduces the variance of estimators of the coefficients of these models. Cooley and Houck extended the Schruben and Margolin methodology to second order models, and demonstrated the technique by finding the optimal reorder point and reorder quantity to minimize annual cost for an inventory system. Although they were successful, follow-up articles by Safizadeh and Thornton (1982), Cooley and Houck (1983), and Safizadeh (1983) further debate the use of these experimental designs.

8. DISTRIBUTION REPLACEMENT (DR)

conditionals of F preserved (C.6)

In Monte Carlo evaluation of integrals, it is sometimes theoretically possible to design a zero variance experiment using *importance sampling (IS)*. The decomposition of IS includes a transformation from DR. The idea, as it is conventionally portrayed, is to bias sampling toward outputs that contribute the most to the variance of the statistic, and then correct for this bias after sampling. Unlike VRTs based on sample allocation that deterministically allocate sampling effort to the outputs, IS does not require the facility to control R_*. However, like transformations in equivalent allocation, it is often difficult to predict the effect of DR transformations in a dynamic simulation. Changing the marginal distributions of $F(x)$ changes the marginal distributions of the outputs, and possibly their expectation, which is central if the statistic is a sample mean. The bias correction that is easily computed in Monte Carlo problems is more difficult to derive when the outputs are dependent. Computing and using the bias correction requires other transformations, usually from equivalent allocation, equivalent information and auxiliary information. The spectacular potential of DR strategies in Monte Carlo estimation has not been realized to date in dynamic simulation.

8.1 Variance Reduction Strategy

Loosely speaking, the variance of a random variable is determined by the possible deviations from its mean and the probability it assumes those values. DR strategies preserve the range of the outputs but alter the probabilities. Ideally, we would like to work with the marginal distributions of the outputs directly, but these distributions are unknown. Thus, as in dependence induction strategies, DR strategies work indirectly through the inputs. Usually DR strategies replace the marginal distributions of **independent** inputs in dynamic simulation. Independence facilitates computing the bias correction as a running product as the inputs are generated (see below). This is particularly important in simulation because the number of inputs that will be realized may not be known a priori.

Let Z be a statistic, and let $X_1 = \{X_{11}, \ldots, X_{n1}\}$ be a sequence of independent, but not necessarily identically distributed, inputs (here we break with our usual convention of organizing identically distributed sequences in columns of X). Let the density or mass function of X_{i1} be denoted by $f_i(x_{i1})$, $i = 1, 2, \ldots, n$, which we assume exists. Now consider a different sequence of marginal distributions $\{f'_i\}$ with the same support the original sequence has. Then if Z is an unbiased estimator of θ, so is the new estimator

$$Z' = Z \prod_{i=1}^{n} \frac{f_i(x_{i1})}{f'_i(x_{i1})} = Z \prod_{i=1}^{n} Y_{i3} \tag{7}$$

where $X_1 \sim \prod f'_i$; this property holds even if Z is not a function of X_1. Here a new output sequence, $\{Y_{i3}\}$, is formed (EA transformation), and a new statistic (EI transformation) results from combining this output with Z (AI transformation). The variance reduction strategy is to select new distributions $\{f'_i\}$ that reduce the variance of Z' as compared to Z. When the elements of X_1 are not independent then the IS estimator is

$$Z' = Z \frac{f(x_{11}, x_{21}, \ldots, x_{n1})}{f'(x_{11}, x_{21}, \ldots, x_{n1})} \tag{7'}$$

where f is the joint distribution of the sequence X_1, making the selection of the new distribution even more difficult.

Expressions (7) and (7') are more general formulations of the IS-type estimator than is found in the literature. Standard presentations assume Z is a sample mean of i.i.d. outputs $\{Y_{i1}\}$, and the bias correction $\{Y_{i3}\}$ is applied to each Y_{i1} individually; e.g. $Z'' = n^{-1}\Sigma Y_{i1}Y_{i3}$. In this case Z'' is easier to work with than Z', but both are composed of the same classes of transformations (Nelson and Schmeiser, 1986b).

A DR strategy more suited to simulation than IS is *Russian roulette (RR).* Like SPLT, RR is useful when outputs Y_{i2} and Y_{i1} are realized in sequence for $i = 1, 2,....$ RR alters the probability that Y_{i1} will be realized (which may have been 1 originally) whenever the event $[Y_{i2} \in L_{j\#}]$ occurs; $L_{j\#}$ is some stratum in the range of Y_{i2} as defined in section 5. RR is often combined with SPLT, but they work differently. SPLT allocates sampling effort to interesting regions of Y_1 (an SA strategy), while RR biases sampling away from uninteresting regions by altering the probability of selection (a DR strategy).

8.2 Results

An old, but as yet unsurpassed, presentation of standard IS results is Kahn (1956). While the existence of optimal DR strategies that reduce variance to zero is theoretically interesting, it is not practically useful in Monte Carlo, much less in simulation. However, these results indicate that the new input distribution $\prod f_i'$ should be roughly proportional to $|E[Z|x_1]| \prod f_i(x_i)$. Of course, $E[Z|x_1]$ is unknown in general. However, if we restrict attention to a particular parametric family of distributions, say $f'_i(x_i;\alpha)$, then it may be possible to estimate the optimal parameter α for that family in a particular problem. Marshall (1956) presents a two-stage sampling procedure for estimating the optimal α; see also Kleijnen (1974).

8.3 Examples

Importance sampling applications abound in the Monte Carlo literature, particularly in particle transport problems; see for instance Carter and Cashwell (1975) and Hammersley and Handscomb (1964). Jeruchim (1984) describes an example in dynamic simulation. He considered simulating a "two-hop" satellite communication system consisting of a transmitter, a transponder, and a receiver

in that order. In terms of the sample space definition, inputs are the signal sequence (binary), and transmitter and transponder noise (individually independent, and identically normally distributed). The output is the signal produced by the receiver. The parameter of interest is the probability of an error, for example the probability that a "1" is received when a "0" is sent. Since the output from each device depends on all signals currently in its memory, the outputs are correlated.

Jeruchim replaced the marginal distributions of the noise components with normal distributions having the same means but altered variances, and computed the bias correction as a product as each noise component was generated. The independence of the inputs made IS feasible.

Kioussis and Miller (1983) applied RR combined with SPLT in the fault tolerant computer simulation mentioned earlier.

9. RECAPITULATION

In the preceding sections we saw that improved, and sometimes optimal, solutions to further constrained versions of the variance reduction problem (1) have been developed. A solution to (1) is a VRT composed of transformations from the six elemental classes. Unfortunately, the number of potential variance reducing transformations is infinite, so an exhaustive search is not possible. In practice, feasible transformations are identified from the available prior knowledge, which varies from experimenter to experimenter. Prior knowledge establishes conditions needed to invoke results ensuring a variance reduction strategy will work. We have summarized these results, but it is prior knowledge that determines when they are applicable.

Based on the preceding development, the kinds of prior knowledge that are useful include:

1. Covariance structure of the outputs. In particular, if the outputs are not all independent, knowledge about the covariance structure can facilitate more efficient estimators through direct use of the structure or control strategies.

2. Distribution of an input or output. Knowing something (moment, range, distribution) of a random variable that will nevertheless be sampled facilitates control strategies (correction based on the knowledge), sample allocation

strategies (allocation based on relative likelihood or variance), or distribution replacement strategies.

3. Functional relationships. When parameters of interest are functionally related, indirect strategies are facilitated. Any knowledge about conditional relationships between random variables can be exploited, through conditional expectations strategies that use the relationship directly, or sample allocation strategies that improve the estimation of a conditional expectation.

4. Properties of g. If g can be modeled, then it may be possible to replace it with g' that may not necessarily mimic the system of interest, but does facilitate efficient estimation. Information about how g maps inputs into outputs (e.g. monotonely) facilitates dependence induction and distribution replacement strategies.

5. Variate generation via the inverse cdf. As a practical matter, the inverse cdf approach is needed for dependence induction strategies.

Some classes of transformations are particularly useful for "setting-up" others. The dependence induction class facilitates other strategies that exploit dependence, such as control strategies and the experimental designs of Schruben and Margolin. The option to induce correlation where it is not inherently present is usually not available in classical sampling experiments, but is always possible (if not useful) in simulation experiments. Equivalent allocation transformations are most frequently used to generate auxiliary outputs that facilitate other strategies. The bias correction for IS is an example. We conjecture that distribution replacement transformations could facilitate sample allocation strategies, since optimal STRAT schemes are available for stratification variables with certain distributions (Dalenius and Gurney, 1951). However, we know of no such application.

This list does not exhaust the kinds of useful prior knowledge by any means. An algorithm that attempts to deduce the available prior knowledge an experimenter has, organize it, and use it to select candidate VRTs is given in Nelson (1985a). Ultimately a taxonomy of prior knowledge, similar to the taxonomy of variance reduction presented here, is needed to facilitate the identification and application of prior knowledge in simulation experiments.

The preceding development points out many open areas of immediate research interest. We review just three:

1. Conditions that ensure a variance reduction when employing conditional expectations (CE) with correlated outputs.

2. Models of the simulation itself as a stochastic process.

3. A comprehensive framework for producing and validating the conditions necessary to employ the Schruben and Margolin designs.

10. THE FUTURE

Instead of surveying VRTs, we have tried to present broadly defined variance reduction strategies using particular VRTs as illustrations. One difficulty in surveying VRTs is that there is no standard definition of any VRT. A benefit of our taxonomy is that a variance reduction strategy can be unambiguously defined in terms of the classes of transformations it employs. Such a structured framework is necessary if variance reduction is ever to be automated, and automation is the only hope for widespread application. By automation we mean that a variance reduction generator, incorporated into a standard simulation programming package, will work interactively with the experimenter to develop and implement an effective VRT in general simulation experiments. An automated VRT generator might work in the following way (Nelson, 1985b):

Step 1. Translate the simulation experiment, expressed in a general purpose simulation language, into a standard form such as the sample space definition.

Step 2. Interactively determine the available prior knowledge, based on a taxonomy of prior knowledge, by querying the model, by querying the experimenter, or by performing experiments.

Step 3. Match the prior knowledge with classes of transformations that employ it, thereby generating candidate variance reduction strategies.

Step 4. Evaluate candidate strategies by invoking known results or performing experiments.

Step 5. Implement the resulting VRT.

Step 6. Repeat steps 2 - 5 if necessary as the experiment progresses.

Research in this area is currently underway.

Historically, VRTs have been applied to estimators of unknown system performance parameters. The increasing use of animation — driving a graphical representation of the system of interest in real time — should spur further interest in variance reduction. Animation necessarily implies that only a brief

realization of system behavior can be observed. To make reliable decisions, representative samples, or samples that exhibit the most critical behavior or range of behavior, are essential. Some existing VRTs may be adaptable, but new ideas will undoubtedly be required. Incorporating animation into simulation design and analysis promises to be an exciting new research area.

ACKNOWLEDGEMENTS

This research was partially supported by a University Seed Grant from the Office of Research and Graduate Studies, The Ohio State University. The author benefited from discussions with Bruce Schmeiser of Purdue University and the reviews of J.P.C. Kleijnen and two referees.

BIBLIOGRAPHY

Bickel, P.J. and K.A. Doksum (1977), *Mathematical Statistics, Basic Ideas and Selected Topics,* Holden-Day, San Francisco.

Bratley, P., B.L. Fox and L.E. Schrage (1983), *A Guide to Simulation,* Springer-Verlag, NY.

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

Carson, J.S. and A.M. Law (1977), "Conservation Equations and Variance Reduction in Queueing Simulations," *1977 Winter Simulation Conference Proceedings* (H. Highland, R. Sargent, J. Schmidt, eds.), 187-189.

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

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

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, TN.

Cheng, R.C.H. (1985), "Generation of Multivariate Normal Samples with Given Sample Mean and Covariance Matrix," *J. Statist. Comput. Simul.,* 21, 39-49.

Cochran, W.G. (1977), *Sampling Techniques,* Wiley, NY.

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

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

Cooper, R.B. (1981), *Introduction to Queueing Theory,* North Holland, NY.

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, 133-148.

Fishman, G.S. (1985a), "Estimating Network Characteristics in Stochastic Activity Networks," *Mgmt. Sci.,* 31, 5, 579-593.

Fishman, G.S. (1985b), "Estimating Critical Path and Arc Probabilities in Stochastic Activity Networks," *Naval Res. Logist. Quart.,* 32, 2, 249-261.

Fishman, G.S. and B.D. Huang (1983), "Antithetic Variates Revisited," *Commun. ACM,* 26, 964-971.

Fishman, G.S. and L.R. Moore III (1984), "Sampling from a Discrete Distribution while Preserving Monotonicity," *American Statistician,* 38, 3, 219-223.

Fox, B.L. and P.W. Glynn (1986), "Discrete-time Conversion for Simulating Semi-Markov Processes," *Opns. Res. Lett.,* 5, 4, 191-196.

Glynn, P.W. (1983), "On the Role of Generalized Semi-Markov Processes," *1983 Winter Simulation Conference Proceedings* (S. Roberts, J. Banks, B. Schmeiser, eds.), 39-42.

Granovsky, B.L. (1983), "Optimal Variance Reduction Theorem in Simulation by the Monte Carlo Method," Sitzungsberichte der Oesterreichischen Akademie der Wissenschaften, Abt. II, *Mathematische, Physikalische und Technische Wissenschaften,* 192, 8-10, 329-335.

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

Hammersley, J.M. and D.C. 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.

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

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

Jeruchim, M.C. (1984), "On the Application of Importance Sampling to the Simulation of Digital Satellite and Multihop Links," *IEEE Trans. Commun.,* COM-32, 10, 1088-1092.

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

Kendall, M. and A. Stuart (1979), *The Advanced Theory of Statistics, Vol. 2, Inference and Relationships,* Charles Griffin and Co. Ltd., London.

Kioussis, L.C. and D.R. Miller (1983), "An Importance Sampling Scheme for Simulating the Degradation and Failure of Complex Systems During Finite Missions," *1983 Winter Simulation Conference Proceedings* (J. Banks, S. Roberts, B. Schmeiser, eds.), 631-639.

Kleijnen, J.P.C. (1974), *Statistical Techniques in Simulation, Part I,* Marcel Dekker, NY.

Kleijnen, J.P.C. (1985), "Analyzing Simulation Experiments with Common Random Numbers," Research Memorandum, Department of Economics, Tilburg University, The Netherlands.

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

Lavenberg, S.S. and P.D. Welch (1979), "Using Conditional Expectation to Reduce Variance in Discrete Event Simulation," *1979 Winter Simulation Conference Proceedings* (H. Highland, R. Shannon, M. Spiegel, eds.), 291-294.

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

Law, A.M. (1975), "Efficient Estimators for Simulated Queueing Systems," *Mgmt. Sci.,* 22, 1, 30-41.

Law, A.M. and W.D. Kelton (1982), *Simulation Modeling and Analysis,* McGraw-Hill, NY.

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

Marshall, A.W. (1956), "The Use of Multi-Stage Sampling Schemes in Monte Carlo Computations," in *Symposium of Monte Carlo Methods* (H. Meyer, ed.), Wiley, NY.

McGrath, E.J. and D.C. Irving (1973), "Techniques for Efficient Monte Carlo Simulation, Vol. III, Variance Reduction," ORNL Report SAI-72-509-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,* 21, 2, 239-245.

Mihram, G.A. (1974), "Blocking in Simular Experimental Designs," *J. Statist. Comput. Simul.,* 3, 1, 29-32.

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

Nelson, B.L. (1985a), "A Decomposition Approach to Variance Reduction," *1985 Winter Simulation Conference Proceedings* (G. Blais, D. Gantz, S. Solomon, eds.), 23-32.

Nelson, B.L. (1985b), "An Illustration of the Sample Space Definition of Simulation and Variance Reduction," *Trans. Society for Computer Simulation,* 2, 3, 237-247.

Nelson, B.L. (1987), "On Control Variate Estimators," *Computers & Operations Research,* in press.

Nelson, B.L. and B. Schmeiser (1983), "Variance Reduction: Basic Transformations," *1983 Winter Simulation Conference Proceedings* (S. Roberts, J. Banks, B. Schmeiser, eds.), 255-258.

Nelson, B.L. and B. Schmeiser (1986a), "A Simulation Framework for Variance Reduction," Working Paper Series No. 1986-011, Department of Industrial and Systems Engineering, The Ohio State University.

Nelson, B.L. and B. Schmeiser (1986b), "Decomposition of Some Well-Known Variance Reduction Techniques," *J. Statist. Comput. Simul.,* 23, 3, 183-209.

Nozari, A., S.F. Arnold and C.D. Pegden (1984a), "Control Variates for Multipopulation Simulation Experiments," *IIE Trans.,* 16, 2, 159-169.

Nozari, A., S.F. Arnold and C.D. Pegden (1984b), "Statistical Analysis Under Schruben and Margolin Correlation Induction Strategy," Technical Report, School of Industrial Engineering, Univ. of Oklahoma.

Porta Nova, A.M.O. and J.R. Wilson (1986), "Using Control Variates to Estimate Multiresponse Simulation Metamodels," *1986 Winter Simulation Conference Proceedings* (J. Wilson, J. Henrikson, S. Roberts, eds.), 326-334.

Rao, C.R. (1973), *Linear Statistical Inference and Its Applications,* Wiley, NY.

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

Ross, S.M. and Z. Schechner (1985), "Using Simulation to Estimate First Passage Distribution," *Mgmt. Sci.,* 31, 2, 224-233.

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

Rubinstein, R.Y. and R. Markus (1985), "Efficiency of Multivariate Control Variables in Monte Carlo Simulation," *Opns. Res.,* 33, 3, 661-677.

Rubinstein, R.Y., G. Samorodnitsky and M. Shaked (1985), "Antithetic Variates, Multivariate Dependence and Simulation of Stochastic Systems," *Mgmt. Sci.,* 31, 1, 66-77.

Safizadeh, M.H. and B.M. Thornton (1982), "An Alternative Variance-Reduction Strategy for RSM Simulation Studies," *Decision Sci.,* 13, 322-330.

Safizadeh, M.H. (1983), "More on Variance-Reduction Strategies for RSM Simulation Studies," *Decision Sci.,* 14, 138-139.

Schmeiser, B. and V. Kachitvichyanukul (1986), "Correlation Induction Without the Inverse Transform," *1986 Winter Simulation Conference Proceedings* (J. Wilson, J. Henrikson, S. Roberts, eds.), 266-274.

Schruben, L.W. (1979), "Designing Correlation Induction Strategies for Simulation Experiments," in *Current Issues in Computer Simulation* (N.R. Adam, A. Dogramaci, eds.), Academic Press, NY.

Schruben, L.W. and B.H. Margolin (1978), "Pseudorandom Number Assignment in Statistically Designed Simulation and Distribution Sampling Experiments," *J. Amer. Statist. Assoc.,* 73, 363, 504-525.

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

Sigal, C.E., A.A.B. Pritsker and J.J. Solberg (1980), "The Use of Cutsets in Monte Carlo Analysis of Stochastic Networks," *Math. Comput. Simul.* XXI, 376-384.

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

Snijders, T.A.B. (1984), "Antithetic Variates for Monte Carlo Estimation of Probabilities," *Statistica Neerlandica,* 38, 2, 55-73.

Stein, M. (1985), "The Use of Latin Hypercube Sampling for Variance Reduction in Simulations with Many Random Parameters," Research Report RC 11166 (#50265), IBM T.J. Watson Research Center, Yorktown Heights, NY.

Tew, J.D. and J.R. Wilson (1985), "Validation of Correlation Induction Strategies for Simulation Experiments," *1985 Winter Simulation Conference Proceedings* (G. Blais, D. Gantz, S. Solomon, eds.), 190-195.

Venkatraman, S. (1983), "Application of the Control Variate Technique to Multiple Simulation Output Analysis," unpublished M.S. thesis, Department of Mechanical Engineering, University of Texas, Austin, Texas.

Venkatraman, S. and J.R. Wilson (1986), "The Efficiency of Control Variates in Multiresponse Simulation," *Opns. Res. Lett.,* 5, 1, 37-42.

Whitt, W. (1976), "Bivariate Distributions with Given Marginals," *Ann. Statist.,* 4, 1280-1289.

Whitt, W. (1980), "Continuity of Generalized Semi-Markov Processes," *Math. Opns. Res.,* 5, 4, 494-501.

Wilson, J.R. (1979), "Proof of the Antithetic Variate Theorem for Unbounded Functions," *Proc. Camb. Phil. Soc.*, 86, 477-479.

Wilson, J.R. (1983), "Antithetic Sampling with Multivariate Inputs," *Am. J. Math. Mgmt. Sci.*, 3, 2, 121-144.

Wilson, J.R. (1984), "Variance Reduction Techniques for Digital Simulation," *Am. J. Math. Mgmt. Sci.*, 4, 3 & 4, 277-312.

Wilson, J.R. and A.A.B. Pritsker (1984a), "Variance Reduction in Queueing Simulation Using Generalized Concomitant Variables," *J. Statist. Comput. Simul.*, 19, 129-153.

Wilson, J.R. and A.A.B. Pritsker (1984b), "Experimental Evaluation of Variance Reduction Techniques for Queueing Simulation Using Generalized Concomitant Variables," *Mgmt. Sci.*, 30, 12, 1459-1472.

Received by Editorial Board member November, 1985; Revised January, 1986.

Recommended by Jack P. C. Kleijnen, Catholic University Brabant, Tilburg, The Netherlands.

Refereed by Ardaven Nozari, Salomon Brothers, Inc, New York, NY and James R. Wilson, Purdue University, West Lafayette, IN.