Proceedings of the 1983
Winter Simulation Conference
S. Roberts, J. Banks, B. Schmeiser (eds.)

VARIANCE REDUCTION: BASIC TRANSFORMATIONS

Barry L. Nelson
Bruce W. Schmeiser
School of Industrial Engineering
Purdue University
West Lafayette, IN 47907

A stochastic computer simulation model is a description of a system of interrelated random variables. Realizations of the system can be used to estimate parameters of interest. Variance reduction techniques (VRTs) transform simulation models into similar models that permit more precise estimation of the parameters. The basic types of transformations are defined and a simple example is given.

1. INTRODUCTION

A computer simulation model of a real or conceptual system is constructed by: 1) Characterizing those elements of the system that cannot be predicted with certainty (time between customer arrivals is distributed exponentially), 2) defining deterministic rules that explain how the system will react to actual values of the uncertain elements (customers will be served in order of arrival), and 3) deciding what performance measures will be recorded to estimate parameters of interest (mean of the customer waiting times to estimate expected waiting time). A model constructed in this way mimics the actual functioning of the system under study, or at least the essential elements of it. However, the only real restriction is that our model provides estimates of the parameters of interest. Unless we actually want to observe the simulated operation of the system, we will be satisfied with "good" estimates, no matter how we arrive at them.

Variance Reduction Techniques (VRTs) are transformations. They transform simulation models into related models that yield better estimates of the parameters of interest, where "better" means more precise. This gain in precision is often at the expense of the one-to-one correspondence between the model and the real or conceptual system. In this paper, we characterize the types of basic transformations that are combined to form VRTs. We do this by way of an illustration, although the definitions and properties of the basic transformations can be made mathematically and statistically rigorous (Nelson, 1983). We begin with some background and an abstract characterization of stochastic computer simulation models, then proceed to the example illustrating the basic transformations. A concluding section suggests how one might approach a model with the idea of applying VRTs.

2. BACKGROUND

Many VRTs used in computer simulation had their origins in Monte Carlo methods for estimating the value of mathematically intractable integrals (Hammersley and Handscomb, 1964), or in survey sampling methods for estimating characteristics of large populations (Cochran, 1977). One of the first extensions to systems simulation was Moy (1965). Kleijnen (1974) gives extensive descriptions of VRTs that are useful in simulation studies; Wilson (1983) provides a survey of the current state of variance reduction research. The most widely used VRTs are "common random numbers" and "antithetic variates" (Wilson, 1983), but "control variates" (Lavenberg and Welch, 1981) has received much recent interest. Probably the best practical guide on using VRTs is McGrath and Irving (1973, 1974). Current research efforts are in identifying classes of simulation models (stochastic networks, for instance) and establishing conditions that insure that a particular VRT will be effective. See, for instance, Sigal, Pritsker, and Solberg (1979), Kumamoto, Tanaka, Inoue, and Henley (1980), Carson (1983), Fishman (1983), Fox (1983), and Grant (1983). The author's work (Nelson, 1983) is in establishing a general mathematical-statistical framework for studying VRTs.

3. SIMULATION MODELS

For our purposes, a <u>simulation model</u> is a description of a system of random variables; the dimensions of this system may themselves be random variables. Given a source of randomness (usually U(0,1) random variables) realizations of the system can be generated that are consistent with the model. We partition these random variables according to how they are defined.

<u>Inputs</u> are random variables defined by known, possibly conditional, probability distributions. Examples might be service and interarrival times in a queuing simulation, or the quantity demanded per period in an inventory model. Another example is a service time distribution conditional on the number of customers in the system, but given this value the distribution is completely specified.

<u>Outputs</u> are defined by known, deterministic transformations of the inputs. These are the observations that we seek to generate. Their distributions are not known completely, but we have a rule or transformation that tells us how to generate realizations of them from realizations of the inputs. We can think of these transformations as representing the actual workings of our system.

<u>Statistics</u> are functions that aggregate outputs into point estimates of the parameters of interest. We often use some sort of sample mean as our statistic. Notice that we include the statistics as part of the description of the model, not as separate entities. When we talk about "variance reduction" we mean reducing the variance of the statistics.

Obviously we have glossed over a number of subtles in defining these sets of random variables. However, this loose characterization can be made rigorous, and it captures all the essential features of simulations that we need to discuss variance reduction. For further details see Nelson (1983).

VRTs are transformations that redefine these sets of random variables. To be effective, they must 1) increase the information available, and/or 2) make better use of the available information in the model about the parameters of interest. Here we use the term information in the sense of statistical information contained in a random variable about a particular parameter (see Barra, 1981). A basic principle of statistical information is that we cannot increase the information that a random variable contains by transforming it via a transformation that does not depend on the unknown parameter. However, there is a potential for loss of information when we transform inputs into outputs, and outputs into statistics. We may increase the information available by changing the definitions of the inputs, and/or by altering the allocation of our sampling effort toward random variables with greater information content. We may use the available information better by employing transformations that are statistically efficient.

In the next section we illustrate the basic ways we can redefine a model to accomplish these goals.

4. ILLUSTRATION OF TRANSFORMATIONS

In this section we define and illustrate the six basic types of transformations that form all VRTs. A transformation is not a VRT in and of itself, and application of a transformation does not guarantee a variance reduction. VRTs are formed by combinations of these basic transformations and prior knowledge (see section 5 below) about the specific problem at hand. Its effectiveness depends on the characteristics of the simulation model. The first two types of transformations are ways to redefine the inputs of a simulation model, the next two transform the outputs, and the final two change the statistics.

The example we use was originally suggested by Kahn (1956) to explain some basic VRTs. We will use it to illustrate the six types of transformations. Consider the problem of estimating the probability, p, that the sum of two fair dice is 3. Clearly p = 1/18, but suppose that we do not know this and want to estimate p by tossing dice. We will toss n pairs of dice (2n single dice), or let a computer program simulate these tosses. Let

$$X_i = \begin{cases} 1, & \text{if the sum of the ith toss is 3} \\ 0, & \text{otherwise} \qquad i = 1,2,\ldots,n \end{cases} \qquad (1)$$

Thus $p = \Pr(X_i = 1)$. As our statistic we take

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

for which

$$E(\overline{X}) = p \quad \text{and} \quad \mathrm{Var}(\overline{X}) \approx \frac{.052}{n}.$$

For convenience later we let

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

Our model is defined by the probabilities p_j that define the working of the dice (inputs), the transformation (1) that gives us our score (outputs), and the statistic $\overline{X}$. We now look at ways to transform this model. The names given to these classes of transformations are our own.

4.1 Distribution Replacement (DR)

<u>Redefine a marginal distribution without altering any statistical dependencies.</u>

Suppose that we redefine the working of our dice in the following way. Let

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

and

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

Thus the total 3 will now occur four times as often, on average. Notice that the individual die tosses are still independent. To compensate for the altered probabilities we let

$$\hat{x}_{dr} = \frac{1}{4}\,\overline{X}$$

which is an unbiased estimator of p, having variance

$$\mathrm{Var}(\hat{x}_{dr}) \approx \frac{.011}{n}$$

4.2 Dependence Induction (DI)

Redefine statistical dependencies without altering any marginal distributions.

On any particular pair of tosses, if we see the outcome (first, second), we were actually just as likely to have seen (7 - first, 7 - second). For instance, the events (2,1) and (5,6) have the same probability of occurrence. Now the well-known relation

$$\mathrm{Var}(X_i + X_j) = \mathrm{Var}(X_i) + \mathrm{Var}(X_j) + 2\mathrm{Cov}(X_i, X_j)$$

shows that negative covariance can decrease the variance of a sum. Thus, if we roll (first, second) on toss 2i - 1, we will just use (7-first, 7-second) for toss 2i. This causes

$$\mathrm{Cov}(X_{2i-1}, X_{2i}) = -p^2$$

and results in

$$\mathrm{Var}(\hat{x}_{di}) \approx \frac{.049}{n}$$

4.3 Sample Allocation (SA)

Redefine the allocation of sampling effort without altering the definition of individual observations.

We now approach the problem a bit differently. Suppose we realize $p = 2p_1p_2$ and we decide to use our 2n single die tosses to estimate p_1 and p_2. Let

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

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

and our statistic is

$$\hat{x}_{sa} = 2\overline{Y}\overline{Z}$$

The key point here is that the variance of $\hat{x}_{sa}$ depends on how we allocate our 2n tosses. In this case, the optimum allocation is to let m = n, and

$$\mathrm{Var}(\hat{x}_{sa}) \approx \frac{.031}{n} + \frac{.077}{n^2}$$

4.4 Equivalent Allocation (EA)

Redefine the observations sampled without altering the allocation of sampling effort.

We use the same approach as in illustrating SA, but now we score

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

This could be justified if we realize that $p_2 = p_4$. Again using the allocation m = n

$$\mathrm{Var}(\hat{x}_{ea}) \approx \frac{.022}{n} + \frac{.031}{n^2}$$

4.5 Auxiliary Information (AI)

Redefine the arguments of a statistic without altering its functional form.

We continue to work with the $\hat{x}_{sa}$ estimator. We might notice that we are actually not utilizing all the available information. Since $p_1 = p_2$, we can use the Y_i observations in $\overline{Z}$, and vice versa. Thus, both $\overline{Y}$ and $\overline{Z}$ are based on 2n observations, and

$$\mathrm{Var}(\hat{x}_{ai}) \approx \frac{.012}{n}$$

Of course $\hat{x}_{ai}$ is biased because $\overline{Y}$ and $\overline{Z}$ are dependent.

4.6 Equivalent Information (EI)

Redefine the functional form of a statistic without altering its arguments.

Recall our original estimator, $\overline{X}$. A class of unbiased statistics using equivalent information is

$$\hat{x}_{ei} = \sum_{i=1}^{n} w_i X_i$$

where $\sum w_i = 1$. For example, suppose we let

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

(Maybe we believe the dice "warm up.") In this case we have an estimator with greater variance

$$\mathrm{Var}(\hat{x}_{ei}) \approx \frac{.052}{n-2} + \frac{.130}{n^2}$$

Looking at the most used VRTs, we find that common random numbers and antithetic variates are based on DI transformations, while control variates combines AI and EI.

5. PRACTICAL GUIDE

How does one approach a large and complex simulation model and decide what kind of VRT to apply? Such a discussion is well beyond the scope of this paper. However, it is widely accepted that some sort of prior knowledge is required to effectively apply any VRT. By prior knowledge we mean any knowledge, either known with certainty or suspected, beyond what is needed to build the original simulation model. Some examples are:

1. Any analytic solutions to parts of the problem.
2. Any analytic models that approximate the system under study.
3. Previous experience with similar systems.
4. Statistical knowledge: correlations between random variables in the model or properties of their distributions.
5. Relative importance of particular random variables in the final estimate.

Prior knowledge makes effective use of the types of transformations outlined above possible.

REFERENCES

Barra, J.R. (1981), Mathematical Basis of Statistics, (L. Herbach, translation ed.), Academic Press, N.Y.

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

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

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

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

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

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

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

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

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

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

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

McGrath, E. and D. Irving (1973), "Techniques for Efficient Monte Carlo Simulation, Vol. III, Variance Reduction," ORLC Report SAI-72-509-LJ.

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

Nelson, B.L. (1983), "Variance Reduction: A Mathematical- Statistical Framework," unpublished Ph.D. dissertation, School of Industrial Engineering, Purdue University.

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

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