Proceedings of the 1986 Winter Simulation Conference
J. Wilson, J. Henriksen, S. Roberts (eds.)

VARIANCE REDUCTION METHODS

Russell C.H. Cheng
Department of Mathematics
University of Wales Institute of Science and Technology
Colum Drive, Cardiff, CF1 3EU, U.K.

ABSTRACT

A computer simulation model is unusual in that the random error is under the total control of the experimenter. Variance reduction methods aim to take advantage of this to improve experimental accuracy. The fundamental ideas behind the most important of these methods will be described and illustrated with simple examples.

1. INTRODUCTION

The purpose of a scientific experiment is to understand how a system, whether physical, biological, organisational, or whatever, can be expected to behave. Such an experiment is subject to chance errors and one of the aims of statistical methodology is to attempt to reduce the effects of such errors as far as is possible. Simulation experiments differ from this kind of experiment in one great respect: in a simulation experiment the random error is under the total control of the experimenter. By being selective as to how this error is introduced and controlled it is possible to obtain more information from a simulation than might at first appear. There exists a number of different techniques for doing this and they are referred to as *variance reduction* (VR) methods. The aim of this tutorial is to describe the fundamental ideas behind the most important of these and to illustrate their application with simple examples.

It is not possible, nor desirable, in a short survey to attempt to give a comprehensive picture; however to avoid merely offering a set of unrelated recipes, an attempt will be made to indicate the relations between some of the more well-known methods, and thus clarify some of the basic issues involved in the use of VR Techniques.

The literature on VR techniques is a fast growing one. We list some selected references which provide a wider reference base and which give more details of the methods themselves and their applications than can be given here. Law and Kelton (1982) and Morgan (1984) provide good coverage at the introductory level. Interesting detailed applications are given by Fishman (1978). A fuller account at a more technical level is provided by Bratley, Fox and Schrage (1983). Two good advanced articles which pay particular attention to fundamental aspects are Wilson (1984) and Nelson and Schmeiser (1985). In addition to these, more specific references will be given for the individual VR methods as we come to them.

2. SOME PRELIMINARY IDEAS

Before looking at specific VR methods, it is necessary to set out some preliminary ideas. In particular we (i) identify the key features of a simulation that we will manipulate and (ii) recall a number of statistical formulas which will be used repeatedly.

Historically Hammersley and Handscomb (1964) introduced many of the ideas of VR through the Monte Carlo evaluation of an integral and this lead is still often followed. However the specialised format tends to obscure how certain of the VR methods can be extended to more general problems; in particular to discrete-event simulations. We shall not use this approach but will take a viewpoint more directed towards these latter problems.

The following problem will be used to introduce the basic ideas. For clarity it is kept very simple, and indeed there would be no need to tackle such a problem by simulation methods at all as its properties can be obtained easily by direct analysis. However it has sufficient structure to illustrate all the VR methods to be considered.

Suppose we wish to investigate the number of men, Y, assigned each day to a long running project. The assignment is made daily, each day being independent of other days. The number of men, X, available for assignment also varies daily and, from past records, X is known to be a random variable with probability distribution: $P(X=1) = 0.3$, $P(X=2) = 0.45$, $P(X=3) = 0.25$. Each day Y is determined from X in the following way: If only one worker (i.e. X=1) is available (s)he works on the project with probability 0.4. If two (i.e. X=2) are available, then the first is assigned with probability 0.71 and the second is assigned only if the first is, and then only with probability 0.72. If three (X=3) are available then one is definitely assigned to the project, one is definitely not, and the third is assigned with probability 0.8.

A Monte Carlo simulation of the daily assignment process is easily set up as follows:

(i) Generate a uniform U(0,1) variable, U_1, to determine whether the number of

men available on a given day, X, is 1, 2 or 3.

(ii) If one man is available use a second U(0,1) variable, U_2 to determine if he works on the project.

If two men are available use a second U(0,1) variable, U_2 to determine if the first is assigned, and if so, use a third U_3 to determine if the second is assigned.

If three men are available, allocate one man and use a second U(0,1) variable, U_2, to determine if a second man is assigned or not that day.

(iii) Output the number of men, Y, assigned to the project on that day.

The above can be regarded as a typical *simulation run* which, in an overall simulation, will be repeated (replicated) n times, say. Each individual run, in general, has the following features

(A1) A stream of U(0,1) *input variables* possibly variable in number (in the example, two or three) is used in the simulation to produce

(A2) A *response variable* Y (in the example, the number of men allocated to the project).

It is the distribution of Y, or characterizing properties, like its mean E(Y) and variance var(Y) that is of major interest. More generally we may be interested in some quantity g(Y) associated with Y. (Strictly speaking, g(Y) is known as a *functional* of Y.) Our objective is to carry out the n simulation runs in such a way that we obtain as good an estimate of g(Y) as is possible from the results of the runs.

Note that we think of a single run as producing just one observation of the response Y. We need to be a little careful with this interpretation in certain discrete-event simulations where in a single run we may measure a number of identically distributed random variates, like waiting times W_j j=1, 2, ..., m, which are not independent; and where we are interested in the distribution of a typical W. This situation does not quite fall into our above format. If however we limit interest to E(W) (or Var(W)) then an estimate of this such as

$$Y = \frac{1}{m} \sum W_j \qquad (2.1)$$

can be thought of as the response of interest obtained from a single run. In steady-state simulations replication can be achieved by simply extending the length of a run, so confusion can arise over what constitutes a single run unless it is clearly defined.

It is possible to discuss VR methods using only assumptions A1 and A2. However it turns out to be much more satisfactory if we introduce one additional component to our view of a basic simulation run:

(A3) In addition to a response, Y, the run produces a *shadow response*, X. In the worker-assignment example we shall take X to be the number of workers available on a given day.

Simulation and regression literature sees various uses of concomitant, control, auxiliary, intermediate variables which are related to the response but are not of primary interest in themselves. Our definition and use of a shadow response is not unlike that of the well-known control variable, but as will become clear, is more general purpose. Unlike the response, which is usually prescribed and over which there is no choice, the shadow response, X can be selected by the simulator as he or she sees fit. The major property that it should have is

(C1) X is (highly) correlated with Y.

In addition it should ideally satisfy either or both the following conditions:

(C2) The distribution of X is known.

(C3) The conditional distribution of the input stream U_i, i=1, 2, ..., M given X is known.

For clarity we shall usually think of Y, X as univariate (i.e. scalar) quantities and of a single input stream. However all our remarks generalise to the multivariate context. This is an important practical point as almost all simulations of real practical interest are multivariate.

We now come to the key step in our view of VR methods. What we do is to strip away the simulation context of the simulation run and regard each run as producing a single observation, (X, Y), from a bivariate distribution. The simulation can then be thought of as being essentially a regression problem in which Y is the dependent variable and X is the independent variable. The joint distribution will not be known; in particular the marginal distribution of Y is not known, and it is this and properties of it that are to be estimated.

In the worker-assignment example, because of its simple nature, we can calculate the exact joint distribution of X and Y. This is given in Table 1 together with associated means and variances and covariances. When we come to use the example we shall assume that certain of this information is not known or only partially known, as would be the case in real problems.

The value of the regression viewpoint is that we can study all VR methods by means of two fundamental formulas. The first shows how a quantity of interest, g(Y), can be decomposed by conditioning on the value of the shadow response, X:

Table 1. Joint distribution of the daily number of men available (X) and assigned (Y)

		X = 1	X = 2	X = 3	P(Y=j)
Y	2	.0	.09	.20	.29
	1	.12	.23	.05	.40
	0	.18	.13	.0	.31
	P(X=i)	.3	.45	.25	

i	1	2	3
E(Y|X=i)	.4	.91	1.8
Var(Y|X=i)	.24	.48	.16

E(X) = 1.95 Var(X) = .5475
E(Y) = .98 Var(Y) = .5996
cov(X,Y) = .379

E(Var(Y|X)) = .3285 Var(E(Y|X)) = .2711

First fundamental formula of VR (Estimator decomposition)

$$g(Y) = \sum_i g(Y|X = i)\ P(X = i) \qquad (2.2)$$

where g(Y|X = i) is the quantity of interest calculated under the condition X = i.

A compact way of writing (2.2) is

$$g(Y) = E(g(Y|X)) \qquad (2.2\ bis)$$

where the outer expectation is evaluated with respect to the distribution of X.

Two particular cases of special interest are

$$E(Y) = \sum_i E(Y|X = i)\ P(X = i) \qquad (2.3)$$

and the probability that Y equals some particular value y

$$P(Y=y) = \sum_i P(Y=y|X=i)\ P(X=i). \qquad (2.4)$$

In an actual simulation we cannot use these formulas directly as quantities like P(Y=y|X=i) will not be known. However we shall find that estimates can also be decomposed and studied using analogous formulas.

The second fundamental formula gives us a decomposition of the variance in a similar way:

Second fundamental formula of VR (Variance decomposition)

$$Var(Y) = E(Var(Y|X)) + Var(E(Y|X)) \qquad (2.5)$$

This formula will be used repeatedly to examine the variance reduction achieved with the different VR methods. The formula shows that the variance can be decomposed into two parts. For any given value of X, Y possesses a residual variance and its average value is the first term on the right in (2.5). However in addition there is a contribution due to the variability in X and this results in the second term on the right in (2.5). For instance if we apply (2.5) directly to the worker-assignment example: E(Var(Y|X)) = 0.3285, Var(E(Y|X)) = 0.2711 so that Var(Y) = 0.5996.

A more general version of the formula is occasionally useful:

$$cov(Y,Z) = E(cov(Y,Z|X)) + cov(E(Y|X), E(Z|X)) \qquad (2.6)$$

The above formulas are important in their own right in stochastic modelling and a clear introduction to them is given in Ross (1980). Their use in VR is discussed by Bratley, Fox and Schrage (1983).

We turn now to the VR methods themselves.

3. VARIANCE REDUCTION METHODS

The notation in this section follows that of section 2. In particular Y is the response of interest, X is the shadow response.

Each VR method will be described in general terms, then illustrated using the worker-assignment example introduced in section 2, followed by brief comments concerning their properties.

In the example, X and Y will always be taken to be respectively the number of men available and actually assigned each day. The objective of the simulation will be to estimate $E(Y) = \theta$.

We begin with the basic simulation method which we can use for comparisons. It does not incorporate any VR as such, but because of its simplicity is by far the most used method.

3.1 Crude Monte Carlo

The basic simulation technique is often called crude Monte Carlo. We make n independent runs giving the responses Y_i i=1, 2, ..., n. The shadow response values are not used at all, and θ is estimated by

$$\hat{\theta}_{crude} = \frac{1}{n} \sum Y_i \ (=\bar{Y}). \qquad (3.1)$$

This estimator is unbiased and its variance is

$$\mathrm{Var}(\hat{\theta}_{crude}) = \frac{1}{n}\,\mathrm{Var}(Y)\ . \qquad (3.2)$$

Example. From Table 1 we have

$$\mathrm{Var}(\hat{\theta}_{crude}) = \frac{1}{n} \times 0.5996\ . \qquad (3.3)$$

Comments. The method has the merit of total simplicity. The number of replications n can be increased till sufficient accuracy is achieved. The method suffers from the basic problem of all methods using increased replication to achieve improved accuracy: namely that the standard deviation decreases as $n^{-\frac{1}{2}}$ only. Often this implies that n be unduly, even hopelessly, large before sufficient accuracy is achieved. Wilson (1982) gives an example showing the huge number of runs needed to estimate the average waiting time in a single server queue.

As an initial exercise on the use of the fundamental formula (2.3) note that if N_i is the number of times X=i is observed and Y_{ij}, j=1, 2, ..., N_i are the corresponding values of Y (when X=i), then we can write:

$$\hat{\theta}_{crude} = \sum_i \bar{Y}_i (N_i/n) \qquad (3.4)$$

where

$$\bar{Y}_i = N_i^{-1} \sum_j Y_{ij}\,. \qquad (3.5)$$

Clearly $\bar{Y}_i$ and N_i/n respectively estimate $E(Y|X=i)$ and $P(X=i)$, so $\hat{\theta}_{crude}$ can be interpreted as being simply (2.3) with $E(Y|X=i)$ and $P(X=i)$ replaced by these estimates.

3.2 Stratified Sampling

The shadow response, X, has to satisfy C2 and C3 for stratified sampling to be possible. Knowledge of the marginal distribution of X is used to ensure that the sampled proportions of the different values of X in the n runs are exactly equal to their probabilities of occurrence. Thus if n_i is the number of times X=i is sampled, we ensure that

$$n_i = nP(X=i). \qquad (3.6)$$

Writing Y_{ij} j = 1, 2, ..., n_i for the n_i values of Y observed when X=i and $\bar{Y}_i$ for their mean we can use the estimate

$$\hat{\theta}_{strat} = \sum_i \bar{Y}_i P(X=i). \qquad (3.7)$$

This corresponds to estimating $E(Y|X=i)$ by $\bar{Y}_i$ in (2.3), together with use of the correct value for P(X=i). The estimator is unbiased, moreover

$$\begin{aligned}\mathrm{Var}(\hat{\theta}_{strat}) &= \sum_i P^2(X=i)\ \mathrm{Var}(\bar{Y}_i)\\ &= \frac{1}{n}\sum_i \mathrm{Var}(Y|X=i)\ P(X=i)\\ &= \frac{1}{n}\,E(\mathrm{Var}(Y|X))\\ &= \frac{1}{n}\,[\mathrm{Var}(Y) - \mathrm{Var}(E(Y|X))]. \end{aligned} \qquad (3.8)$$

This shows that stratification can only reduce the variance.

Example. From Table 1, $E(\mathrm{Var}(Y|X)) = .3285$ so that

$$\mathrm{Var}(\hat{\theta}_{strat}) = \frac{1}{n}\,.3285\,. \qquad (3.9)$$

Comments. This method seems rather underused in practice. The key steps are the choice of shadow response, X, and the need, C3, in each run, to generate the random stream of uniforms A1, *given* the value of X. The difficulty of this last requirement is perhaps the main discouragement to the use of the method. There is one useful class of shadow responses which is amenable to stratification. This comprises those which are sums of independently and identically distributed random variables from the gamma, normal or inverse Gaussian families or from the binomial or Poisson families. Thus, for example, if in a queueing simulation we use the sum of a set of exponential customer inter-arrival times as a shadow response, it is possible to generate a set of exponential random variables conditional on their sum being equal to some preassigned (stratified) value. Cheng (1983, 84) discusses details of such conditional sampling albeit in the context of antithetic variates (see §3.6) where such similar sampling is also required.

There are two forms of stratification possible. The above is known as *pre-stratification* because the n_i are fixed before the simulation runs. The second form is called *post-stratification* and is easier to apply because condition C2 is needed but C3 is not. We make n runs as in crude Monte Carlo, but then use

$$\hat{\theta}_{post} = \sum \bar{Y}_i\ P(X=i)\,. \qquad (3.10)$$

This looks the same as $\hat{\theta}_{strat}$. The difference is that here $\bar{Y}_i$ is the mean of a random number, N_i of observations (if N_i=0 set $\bar{Y}_i$=0, say).

Fieller and Hartley (1954) discuss the post-stratification method when estimating P(Y=j) as in equation (2.4) and give a detailed formula for the variance of the estimator in this case. Using pre-stratification yields an estimator with the same variance of

$$\mathrm{Var}(\hat{P}(Y=j)) = \frac{1}{n}[P(Y=j) - \sum_i \frac{P^2(Y=j,X=i)}{P(x=i)}], \tag{3.11}$$

to order n^{-1}. However the post-stratification estimator includes additional positive terms of order n^{-2} showing that the pre-stratified estimator is preferable if it can be applied.

The advantage of the post-stratification method is of course that it can be applied without requiring condition C3.

3.3 Importance Sampling

This is in effect the stratified sampling method, only taken a step further. The shadow response X needs to satisfy conditions C2 and C3. Now in stratified sampling the number of times each different value of X is sampled is in strict proportion to its probability of occurrence. Thus we set n_i, the number of times X=i equal to nP(X=i). There is no need to use this precise value. Whatever value we use for n_i we can still use

$$\hat{\theta}_{imp} = \sum_i \bar{Y}_i \, P(X=i) \tag{3.12}$$

and this still corresponds to estimating θ from the formula (2.3) with $E(Y|X=i)$ replaced by $\bar{Y}_i$.

Now, (3.12) can be viewed as

$$\hat{\theta}_{imp} = \sum_i (\bar{Y}_i r_i) \, P(Z=i) \tag{3.13}$$

where $r_i = P(X=i)/P(Z=i)$ and Z is a variable with probability mass function

$$P(Z=i) = n_i/n \tag{3.14}$$

Thus we can view (3.13) as the situation where we have stratified according to the distribution of Z and have compensated by weighting the $\bar{Y}_i$ by r_i.

How should we choose the values of n_i? An obvious way is to choose them to minimise $\mathrm{Var}(\hat{\theta}_{imp})$. From (3.12) we have

$$\mathrm{Var}(\hat{\theta}_{imp}) = \sum_i \frac{P^2(X=i)}{n_i} \mathrm{Var}(Y|X=i), \tag{3.15}$$

and minimizing this with respect to n_i subject to $\sum n_i = n$ gives

$$n_i = k \, P(X=i) \sqrt{\mathrm{Var}(Y|X=i)} \tag{3.16}$$

where k is a constant of proportionality. This optimal choice is called *importance sampling* because it assigns most weight to values of X which occur with high probability and for which also the conditional variance of Y is large; such combinations being the most important in their contribution to E(Y). Unfortunately, though P(X=i) is assumed known it is unusual to know $\mathrm{Var}(Y|X=i)$. If however we can guess at its relative magnitude this might lead to worthwhile improvement.

Example. If we insert the values for P(X=i) and $\mathrm{Var}(Y|X=i)$ given in Table 1 into (3.16) we find that n_0, n_1, n_2 should be in the proportions 17 : 72 : 11, giving

$$\mathrm{Var}(\hat{\theta}_{imp}) = \frac{1}{n} \times 0.31219 \tag{3.17}$$

As expected this is less than $\mathrm{Var}(\hat{\theta}_{strat})$.

Comments. Just as we can post as well as pre-stratify so we can do the same with importance sampling. The details are entirely analogous and are omitted here. Importance sampling is usually presented for a continuous X without any stratification applied to the possible values of X. In this case the method reduces to the post-sampling form (see Morgan 1984).

Importance sampling can be unsatisfactory, even dangerous, to use. The optimal weighting, assuming it can be found at all, is only optimal for estimating θ and not necessarily any other quantity. A well-chosen X which makes $\mathrm{Var}(Y|X=i)$ fairly constant, as i varies, renders the method effectively the same as stratified sampling; and even when this is not the case, as in the example, the improvement over stratified sampling may often not be particularly worthwhile.

3.4 Conditional Monte Carlo

This is the complementary method to stratified sampling in the sense that it removes variability in the estimation of E(Y|X) whilst stratification removes variability in the estimation of P(X=i). As in crude Monte Carlo, n runs are made, except that in the runs only sufficient of the simulation need be carried out to enable the shadow response values X_j, j=1, ..., n to be obtained. From each run, we then calculate $E(Y|X_j)$, and take the estimator as

$$\hat{\theta}_{cond} = \frac{1}{n} \sum_j E(Y|X_j) \, . \tag{3.18}$$

Notice that this can be interpreted in terms of the fundamental formula (2.3) as

$$\hat{\theta}_{cond} = \sum_i E(Y|X=i) \, \hat{P}(X=i) \, . \tag{3.19}$$

The estimator is unbiased, moreover

$$\mathrm{Var}(\hat{\theta}_{cond}) = \frac{1}{n} \mathrm{Var}(E(Y|X)) = \frac{1}{n} [\mathrm{Var}(Y) - E(\mathrm{Var}(Y|X))] \tag{3.20}$$

so that conditional Monte Carlo can only reduce the variance.

Example. Suppose that the distribution of X is not known but that, given X=i, the distribution of Y is known. This implies that the quantities E(Y|X=i) are known and are as given in Table 1. Thus

$$\hat{\theta}_{cond} = 0.4 \times \hat{P}(X=1) + 0.9111 \times \hat{P}(X=2) + 1.8 \times \hat{P}(X=3)$$

and

$$Var(\hat{\theta}_{cond}) = \frac{1}{n} Var(E(Y|X)) = \frac{1}{n} \times 0.2711$$

Comments. The main difficulty with conditional Monte Carlo is finding a shadow response for which $E(Y|X=i)$ is known. A useful list of examples is given by Bratley, Fox and Schrage (1983). One case is where Y is observed at each epoch of some particular event. Balking and other queue joining properties fall into this category. If X respresents the state of the system, then it may be that $E(Y|X)$ is known at each event epoch. Note that it will usually still be necessary to calculate Y as its value may be needed to continue the simulation.

3.5 Control Variates

In importance sampling, the effect of variability of $\bar{Y}_i$ in (3.12) is reduced by rescaling it to $\bar{Y}_i r_i$ as in (3.13). In control variates a similar effect is achieved by a *relocation* or *differencing* adjustment. The idea is that if X is correlated with Y (for example when a large value of X tends to give a large value of Y and a small value of X tends to give a small value of Y) then one should compensate for this in estimating θ. Only the mean of X, E(X), need be known and not necessarily its entire distribution. The estimate of θ is then taken as

$$\hat{\theta}_{cont} = \bar{Y} - c[\bar{X} - E(X)] \qquad (3.21)$$

which is in effect the equivalent of the modified fundamental formula:

$$E(Y) = \sum_i [E(Y|X=i) - c(i - E(X))] P(X=i) \qquad (3.22)$$

with $E(Y|X=i)$ and $P(X=i)$ replaced by their estimates. The constant, c, is chosen to minimize $Var(\hat{\theta}_{cont})$, which can be written as

$$Var(\hat{\theta}_{cont}) = \frac{1}{n} (Var(Y) - 2c\, cov(X,Y) + c^2 Var(X)) \qquad (3.23)$$

or using (2.5), as:

$$Var(\hat{\theta}_{cont}) = \frac{1}{n}[E(Var(Y|X)) + Var(E(Y|X) - cX)] \qquad (3.24)$$

From (3.23), the best choice of c is:

$$c = cov(X,Y)/Var(X) \qquad (3.25)$$

giving

$$Var(\hat{\theta}_{cont}) = \frac{1}{n} (1-\rho^2) Var(Y) \qquad (3.26)$$

where $\rho = correl(X,Y)$. From (3.24) we see that unless $E(Y|X)$ is linear in X even this best choice of c will not reduce the term $Var(E(Y|X) - cX)$ to zero, showing that in general the control variates method will not give as much VR as stratification.

Usually c cannot be set to its optimal value anyway as cov(X,Y) will not be known. To achieve variance reduction we need $c(cVar(X) - 2cov(X,Y)) < 0$, i.e. we need c to have the same sign as cov(X,Y) and

$$|c| < 2|cov(X,Y)|/Var(X) \qquad (3.27)$$

In practice we would hope that $|\rho| > \frac{1}{2}$, say, before applying the method and if this is so then for the case $\rho > 0$:

$$\frac{2cov(X,Y)}{Var(X)} = 2\rho\sqrt{\frac{VarY}{VarX}} > \sqrt{\frac{VarY}{VarX}} \qquad (3.28)$$

A reasonably safe choice is therefore to take c=1 provided X has been scaled, if necessary, so that $Var(X) \approx Var(Y)$ and so that $\rho > 0$.

Example. From Table 1 we have $Var(X) = 0.5475$ and $cov(X,Y) = 0.379$. The optimal choice is thus

$$c = 0.379/0.5475 = 0.692$$

giving

$$Var(\hat{\theta}_{cont}) = \frac{1}{n} \times 0.3376 \qquad (3.29)$$

which, as expected, is greater than $Var(\hat{\theta}_{strat})$. If we had used c=1, then

$$Var(\hat{\theta}_{cont}) = \frac{1}{n}(.5996 + .5475 - 2 \times .379) = \frac{1}{n} \times 0.3891 \qquad (3.30)$$

Comments. We have discussed control variates assuming c is a preset value. In practice one can *estimate* the best value of c from the bivariate observations themselves. The situation is closely akin to regression analysis. A very detailed account is given by Lavenberg and Welch (1981); a compact summary is given by Wilson (1984). It is worth noting that if Y and X are both time or event average quantities (like average queue length, service times) then a normal distribution can be assumed leading to very complete distributional results.

3.6 Antithetic Variates (AV)

Suppose the aim is to estimate $E(Y) = \theta$. We use the same idea of compensation as was used in the control variates method. The shadow response X is assumed to satisfy C2 and C3 and is correlated with Y so that a large or small value of X tends, say, to give respectively a large or small value of Y. Then if in one simulation run we obtain a larger than expected value of X, we try to compensate, not, as in control variates, by adjusting the estimate, but by ensuring that X in a second or *antithetic* run is smaller than expected. The procedure is thus to make runs in pairs in which the shadow responses

X, X′ of the two runs are negatively correlated; this in turn will induce a negative correlation between the responses Y and Y′. The average

$$\bar{Y} = \tfrac{1}{2}(Y+Y') \tag{3.31}$$

which estimates θ will then have variance

$$\mathrm{Var}(\bar{Y}) = \tfrac{1}{2}(1+\rho)\ \mathrm{Var}(Y) \tag{3.32}$$

where ρ is the correlation between Y and Y′. If ρ is negative, then Var(Ȳ) will be less than if we had just made two independent runs (when ρ=0).

The pairs of antithetic runs are obtained as follows. Firstly, a pair of antithetic shadow responses is generated. A convenient method is to use the inverse transform method. If

$$X = F_X^{-1}(U) \tag{3.33}$$

where F_X is the cdf of X and U is a uniform U(0,1) variable, then U′ = 1−U is also U(0,1) and is negatively correlated with U. Thus

$$X' = F_X^{-1}(U') \tag{3.34}$$

is negatively correlated with X. The variates X and X′ are then used in separate runs to obtain the antithetic response pair Y and Y′.

As in stratified sampling, condition C3 is needed to be able to generate an appropriate stochastic input stream from X in order to make the simulation. The comments made in the discussion of stratification in connection with this form of conditional sampling apply here also. Some applications of this form of sampling are given by Cheng (1983, 84).

Example. If antithetic sampling is used for X only, then (3.33) and (3.34) reduce to:

$$X = \begin{cases} 1 \text{ if } 0 < U < .3 \\ 2 \text{ if } .3 < U < .75 \\ 3 \text{ if } .75 < U < 1 \end{cases} \quad X' = \begin{cases} 1 \text{ if } .7 < U < 1 \\ 2 \text{ if } .25 < U < .7 \\ 3 \text{ if } 0 < U < .25 \end{cases}$$

A direct calculation then shows that the correlation between Y and Y′ is ρ = −0.3868, so that (3.32) gives

$$\mathrm{Var}(\bar{Y}) = \tfrac{1}{2} \times 0.3677 .$$

Comments. The formula (2.6) can be used to give some idea of when the antithetic variate technique is likely to be successful. We have

$$\begin{aligned} \mathrm{cov}(Y,Y') &= E[\mathrm{cov}(Y,Y'|X,X')] \\ &\quad + \mathrm{cov}[E(Y|X,X'),\ E(Y'|X,X')] \\ &= 0 + \mathrm{cov}[E(Y|X),\ E(Y|X')] \end{aligned} \tag{3.35}$$

Now X′ decreases if X increases, so that a sufficient condition for cov(Y,Y′) to be negative is if E(Y|X) is monotone in X. If however the behaviour of E(Y|X) is quadratic in X (first decreasing and then increasing as X increases) then cov(Y,Y′) might easily be positive and the antithetic method be counter-productive.

In this connexion, the traditional approach to antithetic variates is to make the entire input sequence of the second run antithetic to the first, i.e. if $U_1, U_2, \ldots$ is the sequence used in the first run then $1-U_1, 1-U_2, \ldots$ is the sequence used in the second. From the above, we see this is a reasonable procedure if (i) Y is monotone in each U_i and (ii) the objective is solely to estimate E(Y). For example, if Y is an average queue length and each U_i is used with the inverse distribution method to generate either a service time or an inter-arrival time then the first condition is likely to be met. If however Var(Y) is of interest as well as E(Y), then this can be regarded as the situation where the response is $Z = (Y-E(Y))^2$ and its mean E(Z) is to be estimated. Equation (3.35) then indicates that the antithetic technique applied to X will likely yield a positive correlation between Z and Z′ and so be counterproductive.

The technique of antithetic variates can be thought of as a special case of systematic sampling schemes where the uniform input streams of several runs are correlated. Fishman and Huang (1983) give details of such methods. In their limiting form, as the number of correlated runs becomes large, these methods become not dissimilar to stratification.

3.7 Common Random Numbers (CRN)

This is the most commonly used and perhaps the most successful VR method. It is used to compare two similar systems. As far as is possible the two systems are subjected to the same simulation conditions, so that any difference between the two responses Y and Y′ can be ascribed to systematic differences rather than differences between random errors used in the two simulation runs. In other words we try to induce (in contrast to antithetic variates) a *positive* correlation between Y and Y′ so that for their difference:

$$\mathrm{Var}(Y-Y') = \mathrm{Var}(Y) + \mathrm{Var}(Y') - 2\mathrm{cov}(Y,Y') \tag{3.36}$$

will be less than if two independent runs had been made.

This positive correlation can be induced via a shadow response in a manner akin to the antithetic variates method. Suppose that a shadow response has distribution F_1 and F_2 in the two systems. Then we use the inverse distribution transform method with the *same* random number, U, to generate the shadow responses X, X′ used in the two runs:

$$X = F_1^{-1}(U)\ ,\ X' = F_2^{-1}(U) \tag{3.37}$$

This gives

$$\begin{aligned} \mathrm{cov}(Y,Y') &= E(\mathrm{cov}(Y,Y'|X,X')) \\ &\quad + \mathrm{cov}(E(Y|X),\ E(Y'|X')) \\ &= 0 + \mathrm{cov}(E(Y|X),\ E(Y'|X')) \end{aligned}$$

This shows that providing $E(Y|X)$ and $E(Y'|X')$ remain in step in terms of monotonicity relative to U (i.e. one is increasing or decreasing when the other is respectively increasing or decreasing) then $cov(Y,Y')$ will be positive.

Notice that this is a less stringent condition, particularly when similar systems are being compared, than that discussed for the antithetic variates method, so that CRN is often much more likely to be successful.

A convenient way to view CRN is to think of the random number U used to generate X and X' as being itself the shadow variable, and because the above weak monotonicity condition is often satisfied by all the variates U_1, U_2, ... of the input sequence, we can then apply the method simply by using, as far as is possible, the *same* input sequence for both runs. This is the way the method is usually described.

Example. Suppose that the assignment probabilities are modified (in a way which need not concern us) so that the resulting joint distribution of Y and X is as given in Table 2, and we wish to estimate the change in the average number of personnel assigned each day. Let Y' be the number assigned daily in the new arrangement. Then $Var(Y-Y') = .5996 + .5256 = 1.1252$ if Y and Y' are obtained independently. If we apply CRN to the generation of X as in (3.37) then $cov(E(Y|X), E(Y'|X')) = .2486$ which reduces the variance to:

$$Var(Y-Y') = 1.1252 - 0.4972 = 0.6280$$

Table 2. Modified joint distribution of daily number of men available (X) and assigned (Y)

		X		
		1	2	3
Y	2	.0	.11	.22
	1	.18	.25	.03
	0	.12	.09	.0

4. DISCUSSION

We have not given much details of the minutiae often required for the application of VR methods in practical cases but instead have concentrated on describing the motivation behind the methods. Moreover our emphasis has been on VR in discrete-event simulations.

There is another broad area where Monte Carlo simulations are used, namely *distribution sampling experiments*, and our discussion covers this situation as well. Typically we are interested in properties of a statistic calculated from a random sample. For instance what is the bias of a certain estimator of a certain parameter from such-and-such a distribution? Here the statistic is the response and the input sequence comprises the uniforms used to generate the random sample.

In either type of simulation the choice of a good shadow response is often crucial if worthwhile VR is to be achieved. Kleijnen (1974) reviews some useful types; see also Cheng and Feast (1980) and Wilson and Pritsker (1984).

It will be evident from the description that care is often required to apply VR methods effectively. We have tried to show the relationship between the different VR methods and to show when they may be expected to be successful. A description of VR using a unified framework is invaluable for this and the work of Nelson and Schmeiser (1985) seems particularly interesting in this regard. One simple consequence is that it more clearly shows when VR methods can be used in combination and when they cannot. For example use of conditional Monte Carlo and stratification is a particularly effective combination. Use of AV and CRN is an intriguing case; Law and Kelton (1982) point to the problems involved and Schruben and Margolin (1978) give a particularly interesting account in the context of design of experiments.

Finally, it should be noted that the VR methods described here by no means exhaust the list. However those discussed should at least provide a base from which other more sophisticated or elaborate methods can be studied.

REFERENCES

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

Cheng, R.C.H. (1983). Random samples with known sample statistics: with application to variance reduction. *Proceedings of the 1983 Winter Simulation Conference.* Institute of Electrical and Electronic Engineers, Washington, D.C., 395-400.

Cheng, R.C.H. (1984). Generation of inverse Gaussian variates with given sample mean and dispersion. *Applied Statistics*, 33, 309-316.

Cheng, R.C.H. and Feast, G.M. (1980). Control variables with known mean and variance. *Journal of the Operational Research Society*, 31, 51-56.

Fieller, E.C. and Hartley, H.O. (1954). Sampling with control variables, *Biometrika*, 41, 494-501.

Fishman, G.S. (1978). *Principles of Discrete Event Simulation*, Wiley, New York.

Fishman, G.S. and Huang, B.D. (1983). Antithetic variates revisited. *Communications of the ACM*, 26, 964-971.

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

Kleijnen, J.P.C. (1974). *Statistical Techniques in Simulation, Part 1*. Marcel Dekker, New York.

Lavenberg, S.S. and Welch, P.D. (1981). A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Management Science*, 27, 322-335.

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

Morgan, B.J.T. (1984). *Elements of Simulation*, Chapman and Hall, London.

Nelson, B.L. and Schmeiser, B.W. (1985). Decomposition of some well-known variance reduction techniques. *Research Memorandum 84-6*, School of Industrial Engineering, Purdue University, West Lafayette, Indiana.

Ross, S.M. (1980). *Introduction to Probability Models*, 2nd ed, Academic Press, New York.

Schruben, L.W. and Margolin, B.H. (1978). Pseudorandom number assignment in statistically designed simulation and distribution sampling experiments. *Journal of the American Statistical Association*, 73, 504-520.

Wilson, J.R. (1982). Variance reduction techniques. *Proceedings of the 1982 Winter Simulation Conference*. Institute of Electrical and Electronic Engineers, Washington D.C. 605-612.

Wilson, J.R. (1984). Variance reduction techniques for digital simulation. *American Journal of Mathematical and Management Sciences*, 4, 277-312.

Wilson, J.R. and Pritsker, A.A.B. (1984). Variance reduction in queueing simulation using generalized concomitant variables. *Journal of Statistical Computation and Simulation*, 19, 129-153.

AUTHOR'S BIOGRAPHY

RUSSELL C.H. CHENG obtained a BA from Cambridge University, England, in 1968, and the Diploma in Mathematical Statistics in 1969. He obtained his Ph.D. in 1972 from Bath University working on computer simulation models of industrial chemical plants in association with ICI. He joined the Mathematics Department of the University of Wales Institute of Science and Technology (known and pronounced invariably as UWIST) in 1972 and was appointed Senior Lecturer in 1980. His main fields of interest include: computer generation of random variates, variance reduction methods, parametric estimation methods, applications of Markov decision processes to industrial processes and more recently: ship simulation; and he has published a number of articles in all these fields.

Russell C.H. Cheng
Department of Mathematics
UWIST
Colum Drive
Cardiff CF1 3EU
U.K.