Proceedings of the 1988 Winter Simulation Conference
M. Abrams, P. Haigh, and J. Comfort (eds.)

Stratified sampling in the simplex with applications to estimating statistical distributions

Russell C.H. Cheng and Teresa R. Davenport
School of Mathematics
University of Wales, Cardiff
Senghennydd Road, Cardiff CF2 4AG, U.K.

ABSTRACT

Many problems in statistics and operations research reduce to the evaluation of the distribution of a random variable, called the response, known to be a complicated function of a number, d say, of independent uniform variables. Monte Carlo estimation is often used for this purpose if the distribution is analytically intractable. Often the response possesses symmetry properties with respect to its arguments. It is then possible to restrict sampling to simplex regions of the sample space. This can be easily combined with stratified sampling to give variance reduction of order O(d) compared with normal stratified sampling. The theory of such methods is discussed and a simple stratified sampling scheme is applied to two examples giving a two to five fold reduction in the variance.

1. INTRODUCTION

Many of the most successful applications of variance reduction in Monte Carlo simulation have occurred in the context of estimating the distribution of a test statistic. A glance through the statistical literature shows this to be one of the areas where Monte Carlo simulation is most widely used, particularly for comparing the small sample distribution of a statistic with asymptotic, theoretically derived, results.

A typical scenario for such a study is as follows. A random sample is drawn from some specified distribution. A test statistic is then calculated from the random sample. This latter is often an involved calculation which consumes the bulk of the computing time. A crude way of estimating the distribution of the test statistic (but one which is often used) is to generate many independent test statistic values (from independent samples) and then estimate the distribution using the empirical distribution function. We shall consider how stratified sampling can be used to improve on this elementary method. The choice of stratified sampling rather than one of the other well-known variance reduction methods, like control variates or antithetic variates, has been based on the following factors.

(i) The distribution of a test statistic is precisely defined, and there is an arguably greater requirement to determine it more accurately, or at least to determine the accuracy of the Monte Carlo estimation more precisely, than in simulations of the operations research type where many more uncertain assumptions can influence the results. Thus variance reduction techniques which can introduce errors like bias, as in the case of control variates, are not so attractive as exact methods like stratification.

(ii) The technique should be applicable without the need to presume specialist features of the statistic under review for its implementation; ideally variance reduction should be guaranteed under very general conditions.

(iii) The technique should be easy to implement.

Cheng and Davenport (1988) discuss an elementary stratification scheme which meets the above three criteria. They discuss its implementation to problems of operations research type via the stratification of a so-called shadow response variable which is well correlated with the response of actual interest. This method is in principle applicable to the problem of concern to us here and indeed Cheng and Davenport mention one such application. We shall here adopt a much more direct approach. Our only restriction will be a symmetry condition on the response variable. This is not an inconsequential restriction; nevertheless many test statistics will satisfy it. Thus for example. those based on sums of squares of the observations, or connected with maximum likelihood estimation will usually satisfy the condition.

Section 2 outlines some of the ideas of stratified sampling to be used. Section 3 describes how they specialise under the symmetry condition. Example algorithms are discussed in Section 4. Two applications are described in Section 5.

2. STRATIFIED SAMPLING

We collect together some results concerning stratified sampling for later reference. Some, but not all, are well-known. Good reviews of stratified sampling include those by Rubinstein (1981), Wilson (1984) and Nelson and Schmeiser (1985).

Let $X_1, X_2, \ldots, X_d$ be a random sample from some distribution $F_X(.)$. Let $Y = Y(X_1, X_2, \ldots, X_d)$ be the test statistic of interest. We assume that each X_i is generated from an independent uniform U(0,1) variable, U_i, by the inverse distribution function transform

$$X_i = F_X^{-1}(U_i). \qquad (2.1)$$

Then Y can with no loss of generality be regarded as a function of the U_i. Thus writing $U = (U_1, U_2, \ldots, U_d)$, $I^d = \prod_{i=1}^{d} [0,1]$, so that I^d is the d-dimensional unit hypercube, we have

$$Y = Y(U) \qquad U \in I^d. \qquad (2.2)$$

We shall assume that it is the distribution of Y, and its major characteristics, like its mean, that are of interest to us.

Note that these characteristics can often be written as an integral. For example the mean of Y is

$$E(Y) = \int_{I_d} T(u)du \tag{2.3}$$

where T is identically equal to Y, i.e. T=Y. The cdf is

$$\Pr(T \leq y) = \int_{I_d} T(u;y)du \tag{2.4}$$

where

$$T(u;y) = \begin{cases} 1 & \text{if } Y(u) \leq y \\ 0 & \text{if } Y(u) > y. \end{cases} \tag{2.5}$$

Crude Monte Carlo sampling is thus equivalent to sampling a set of points $U_1, U_2, \ldots, U_N$ uniformly and independently distributed in I^d and then estimating the integral

$$\theta = \int_{I_d} T(u)du \tag{2.6}$$

by

$$\hat{\theta}_{crude} = N^{-1} \sum_{i=1}^{N} T(U_i). \tag{2.7}$$

The accuracy is well known to be

$$Var(\hat{\theta}_{crude}) = O(N^{-1}). \tag{2.8}$$

Consider now stratified sampling. We imagine I^d to be partitioned into N disjoint strata or *cells*, C_i, $i = 1, 2, \ldots, N$:

$$I^d = \bigcup_{i=1}^{N} C_i.$$

Let C_i have volume α_i. We shall consider the special case where we sample just one point U_i uniformly from each cell and use as the estimator

$$\hat{\theta}_{strat} = \sum_{i=1}^{N} \alpha_i \, T(U_i). \tag{2.9}$$

This is unbiased for θ.

We shall for convenience use the term STRAT1 for stratified sampling using one uniformly distributed point from each cell. We shall say the cells are *equiprobable* if $\alpha_i = N^{-1}$ for all i. We have the following result.

If the cells are equiprobable, then use of STRAT1 gives

$$Var(\hat{\theta}_{strat}) \leq Var(\hat{\theta}_{crude}). \tag{2.10}$$

This is a special case of the more general result described for example by Rubinstein (1981). The attraction of it is that is applies *simultaneously* to all functions T whose integrals are to be estimated. Thus the result guarantees, for example, that the entire cdf $F_Y(y)$ is estimated more accurately using stratification than if crude Monte Carlo is used.

A very obvious partition to use is that where the hypercube is divided in $N=k^d$ equal subcubes all with sides of length k^{-1}. We call this the *equal-subcubes partition of* I^d.

Cheng and Davenport show that provided T satisfies certain smoothness conditions (but which allow for discontinuities in T) then use of the equal-subcubes partition with STRAT1 gives

$$Var(\hat{\theta}_{strat}) = O\left[\frac{1}{N^{1+1/d}}\right]. \tag{2.11}$$

Thus, though better than crude Monte Carlo, STRAT1 will only be significantly better when d is small. It is perhaps worth comparing this result with the case where a fixed regular grid scheme is used (see for example, Niederreiter, 1978, or Ripley, 1987). The square of the so-called discrepancy

$$D_N^2 = O\left[\frac{1}{N^{2/d}}\right] \tag{2.12}$$

appears to be the most natural quantity to compare with (2.11); and it will be seen that STRAT1 always does better than the fixed regular grid.

We now consider the special case of estimating $F_Y(y)$ when the cells are not equiprobable. Let $F_Y(y) = p$, say. Using crude Monte Carlo we have that

$$Var(\hat{\theta}_{crude}) = p(1-p)/N. \tag{2.13}$$

We shall obtain a worst case bound on $Var(\hat{\theta}_{strat})$. Let

$$R(y) = \{u | u \in I^d, \ Y(u) \leq y\}.$$

Then the volume of R(y) is p. Let the area of $C_i \cap R$ be p_i. Then

$$Var(\hat{\theta}_{strat}) = \sum_{i=1}^{N} p_i(\alpha_i - p_i). \tag{2.14}$$

This is a concave function of the p_i, so its maximum can be found, subject to $\Sigma p_i = 1$, by the method of Lagrange multipliers. This gives after some manipulation

$$\max_{p_i} Var(\hat{\theta}_{strat}) \leq N^{-1}\Sigma\alpha_i^2 - N^{-1}(p-½)^2, \tag{2.15a}$$

or

$$\max_{p_i} Var(\hat{\theta}_{strat}) \leq ¼ \sum_{i=1}^{N} (\alpha_i - N^{-1})^2 + N^{-1}p(1-p). \tag{2.15b}$$

It should be emphasised that this last expression is a worst case situation. The actual variance will be less than this value. It shows that choice of unequal α_i can lead to a variance greater than that of crude Monte Carlo but it does place a limit on how much greater.

3. T SYMMETRIC IN ITS COMPONENTS

We consider the case where T(.) is symmetric in its components. That is, if

$$\pi(u) = (\pi(u_1), \pi(u_2), \ldots, \pi(u_d)) \qquad (3.1)$$

is any permutation of the components of u then

$$T(\pi(u)) = T(u). \qquad (3.2)$$

In this case the entire behaviour of T is encountered in the simplex

$$S = \{u | 0 \leq u_1 \leq u_2 \leq \ldots \leq u_d \leq 1, \ u \in I^d\}. \quad (3.3)$$

More precisely we have the following property:

Theorem 1 If T(.) is symmetric in its components then the distribution of T(U) is the same as the conditional distribution of T(U) given that U is uniformly distributed on the simplex S.

Proof We write

$$S(\pi) = \{u | 0 \leq \pi(u_1) \leq \pi(u_2) \leq \ldots \leq \pi(u_d) \leq 1, \ u \in I^d\}$$

to denote the simplex obtained from S by permuting its components to $\pi(u)$. There are altogether d! such simplices (including S); each has volume $(d!)^{-1}$; they are disjoint, so they partition I^d.

Now

$$\Pr(T \leq t) = \int_{T(u) \leq t} du$$

$$= \sum_{\pi} \int_{[T(u) \leq t] \cap S(\pi)} du$$

$$= \sum_{\pi} \int_{[T(\pi(u)) \leq t] \cap S(\pi)} du,$$

using (3.2). But

$$\int_{[T(\pi(u)) \leq t] \cap S(\pi)} du = \int_{[T(u) \leq t] \cap S} du .$$

Thus

$$\Pr(T \leq t) = d! \int_{[T(u) \leq t] \cap S} du$$

$$= \int_{[T(u) \leq t] \cap S} du \Big/ \int_S du$$

$$= \Pr(T \leq t | u \in S). \qquad \square\square\square$$

With crude Monte Carlo sampling the above Theorem does not help, as it is immaterial whether we sample uniformly over the entire cube I^d or merely over S. However with stratified sampling it can be used to greatly increase the effectiveness of the method.

Suppose that stratified sampling is used throughout I^d and that the partition used divides each simplex in the same way into N/d! equiprobable cells. For any cell, C, in S we can consider the corresponding cell (i.e. the one in the same position) in each of the other simplices. If T is symmetric in its components, the d! points falling in such a set of cells can be viewed as all having been sampled from the one cell in S. In this situation we can regard the method of sampling as taking d! uniformly distributed points from each cell of S. However from (2.11) we see that it would be better to further partition each cell of S into d! equiprobable cells and use stratified sampling in these cells; from (2.11) this reduces the variance by a factor $O[(d!)^{1/d}]$. We thus have

Theorem 2 Suppose use of STRAT1 with equiprobable cells gives $\mathrm{Var}(\hat{\theta}_{strat}) = O(N^{-1-1/d})$. Then for T symmetric in its components, application of STRAT1 to the simplex S using N equiprobable cells reduces $\mathrm{Var}(\hat{\theta}_{strat})$ by a factor $O[(d!)^{1/d}]$, compared with using STRAT1 on the cube I^d with N equiprobable cells. □□□

For d large $(d!)^{1/d} \approx d/e$, where e is the basis of natural logarithms.

4. TWO SAMPLING SCHEMES

In this section we describe two partiton methods of the simplex S. The first is a partition into equiprobable cells. The second is only approximately so, but would seem to be more convenient to use in practice. For this reason the first method will only be discussed briefly, but the second will be considered in more detail.

4.1 Equiprobable Partition of the Simplex S

The simplex S can be partitioned into k^d equiprobable subsimplices, all of volume $k^{-d}d!$; all are similar in shape from that of S. We shall not describe the case of general k, but the special case k=2.

Consider a point u ∈ S, so that $0 \leq u_1 \leq u_2 \leq \ldots \leq u_d \leq 1$. Let r be the subscript for which $u_r \leq \frac{1}{2} < u_{r+1}$. If we let

$$v_0 = \tfrac{1}{2}, \ v_i = u_i + \tfrac{1}{2}, \ i = 1, 2, \ldots, r, \text{ and } v_{r+1} = 1 \qquad \ldots (4.1)$$

then clearly

$$\tfrac{1}{2} \equiv v_0 \le v_1 \le \ldots \le v_r \le 1 \equiv v_{r+1}, \qquad (4.2)$$

so that we can think of $v_1, v_2, \ldots, v_r$ as dividing the interval $[\tfrac{1}{2},1]$ into $(r+1)$ subintervals:

$$[v_{j-1}, v_j] \qquad j = 1, 2, \ldots, r+1. \qquad (4.3)$$

Now the remaining u_i, $i = r+1, r+2, \ldots, d$, being all greater than ½, will fall into these subintervals. Let $a(j)$ be the number of u_i's that fall in $[v_{j-1}, v_j]$. Thus if

$$b(0) = r,\ b(j) = r + \sum_{i=1}^{j} a(i),\ j = 1, 2, \ldots, r+1$$

then

$$v_{j-1} \le u_{b(j-1)+1} \le \ldots \le u_{b(j)} \le v_j,$$
$$j = 1, 2, \ldots, r+1. \qquad (4.4)$$

We shall write $\mathbf{a} = (a(1), a(2), \ldots, a(r+1))$; for brevity we do not make explicit the dependence of $\mathbf{a}$ on r. Note also that for fixed r, different sets of $a(j)$ values are possible. Each $\mathbf{a}$ represents what is known as a different composition of $(d-r)$ objects into $(r+1)$ parts (see Nijenhuis and Wilf, 1978), and there are in all $\binom{d}{r}$ different compositions. In fact elementary considerations show that, for a given r and $\mathbf{a}$, then (4.1), (4.2) and (4.4) define a simplex of the same shape as S, but of volume $2^{-d}/d!$. Each different combination of r and $\mathbf{a}$ defines a different simplex and there is no overlap (apart from boundary points of probability zero). Summing over r shows there are 2^d subsimplices as required.

An algorithm for sampling one point from each simplex can be constructed from (4.2) and (4.4). For each r, we use the subroutine NEXCOM described by Nijenhuis and Wilf to generate each different $\mathbf{a}$ composition. For each composition we generate an ordered set of d uniforms on $(\tfrac{1}{2},1)$. We then leave the first $a(1)$ values unchanged, but subtract ½ from the $(a(1)+1)$th uniform; then leave the next $a(2)$ values unchanged but subtract ½ from the next. The process is repeated until all the uniforms have been considered. This set of processed uniforms constitutes our sample point from the subsimplex with the given r and $\mathbf{a}$ composition.

The main problem with the direct use of the algorithm is that the total number, k^d, of subsimplices increases too rapidly with d. Thus either some kind of transformation of the problem to one of lower dimension is necessary (see Cheng and Davenport, 1988) or else the subsimplices have to be regrouped into k^{d-b} cells each made up of k^b subsimplices with b suitably chosen. We do not consider these possibilities here.

4.2 Equal Subcubes Partition of S

The second partition does not produce equiprobable cells, but is the more obvious analogue of the equal-subcubes partition of I^d. Simply take the intersection of S with each subcube of the equal-subcubes partition as a cell of S. These cells can be classified as follows.

Divide the unit interval into k equal subintervals.

$$D_i = [(i-1)/k,\ i/k] \qquad i = 1, 2, \ldots, k. \quad (4.5)$$

For any point $\mathbf{u} \in S$ we count the number, a_j, say of components that fall in each subinterval D_j, and write

$$\mathbf{a} = (a_1, a_2, \ldots, a_k) \qquad (4.6)$$

All the points $\mathbf{u}$ with the same $\mathbf{a}$ are defined to be in the same cell, and every different $\mathbf{a}$ identifies a distinct cell which we can write as $C(\mathbf{a})$. Moreover, for a fixed $\mathbf{a}$, the a_j ranked components of $\mathbf{u}$ that fall in D_j sweep out a simplex n a_j-dimensional space of volume $k^{-a_j}/a_j!$. Thus the volume of $C(\mathbf{a})$ is

$$\mathrm{Vol}(\mathbf{a}) = k^{-d}/(a_1!\, a_2! \ldots a_k!). \qquad (4.7)$$

The different $\mathbf{a}$'s are, like the previous partition, the compositions - only now - of d objects divided into k groups. A convenient method of generating them is the NEXCOM subroutine previously mentioned.

An algorithm for generating one point from each cell $C(\mathbf{a})$ is readily constructed if we treat the a_j components of $\mathbf{u}$ falling in D_j as just a random sample of a_j uniform $U(k^{-1}(j-1), k^{-1}j)$ variables.

If we write $\mathbf{U}(\mathbf{a})$ as the vector of uniform variates formed in this way, then the estimate of θ, as given in (2.9), becomes

$$\hat{\theta}_{strat} = d!k^{-d} \sum_{\mathbf{a}} \left(\prod_{j=1}^{d} a_j! \right)^{-1} T[\mathbf{U}(\mathbf{a})]. \qquad (4.8)$$

The total number of cells is the number of compositions (see Nijenhuis and Wilf, 1978):

$$J(d,k) = \binom{d+k-1}{d}. \qquad (4.9)$$

For practical values of d and k this is spectacularly smaller than k^d (e.g. for $d = k = 10$, $J = 8008$, $k^d = 10^{10}$).

A weakness of this partition is that the cells are not equiprobable. However if d is much smaller than k then the partition is close to being equiprobable in the following sense.

Notice that in (2.15a) an upper bound for $\mathrm{Var}(\hat{\theta}_{strat})$ can be obtained by replacing every α_i by $\max \alpha_i = \alpha$, say. Now the largest cells of the partition are those which are full subcubes of volume k^{-d}. Because the sampling is restricted to points of S, this is equivalent to having $\alpha = k^{-d}d!$. Moreover, as $N = J(d,k)$, we can replace αN in (2.15b) by

$$\alpha N = (1 + k^{-1})(1 + 2k^{-1}) \ldots (1 + (d - 1)k^{-1})$$

to give

$$\text{Var}\left[\hat{\theta}_{strat}\right] \leq \frac{1}{4N}\left[\prod_{j=1}^{d-1} (1 + j/k) - 1\right] + \frac{p(1-p)}{N} .$$

Provided therefore that $d \ll k$, the right-hand side will not be much larger than $\text{Var}(\hat{\theta}_{crude})$ even in this, the very worst case. Overall this potential loss of efficiency needs to be weighed against the improvement of $O((d!)^{1/d})$ given in Theorem 2.

In use the weighting factors of (4.8) can be precalculated, or can be calculated recursively as each composition **a** is generated. Thus computationally the overhead associated with the use of (4.8) rather than (2.7) is small.

5. NUMERICAL EXAMPLES

5.1 A Statistical Estimation Problem

To illustrate the kind of variance reduction possible we consider the sampling scheme of Section 4.2 applied to a problem discussed by Thoman et al (1969) of estimating the bias of the maximum likelihood (ml) estimator of the shape parameter β of a Weibull variable W with cdf

$$F_W(w) = 1 - \exp[-(w/\gamma)^{\beta}]$$

when the scale parameter γ is unknown. Now the ml estimator $\hat{\beta}$ is not dependent on the order of the sampled observations. Suppose therefore that these are generated from uniform variables by the inverse distribution function transform method. We can then apply STRAT1 to these uniforms and moreover, do so using the simplex version of the equal subcubes partition described in Section 4.2. Table 1 gives the results of estimating $\theta = E(\hat{\beta})$ for sample size $d = 5$ with $k = 10$. This scheme uses 2002 points in one stratified run. To estimate the variability of the estimate of θ, the run was repeated 1000 times. For comparison the table also shows the means and variances of 1000 runs each of which estimated θ from 2002 independent samples (of size 5). It will be seen that the variance is reduced by a factor of over 10. Also tabulated are selected percentile estimates of the distribution of $\hat{\beta}$ and it will be seen that use of stratification leads to variance reduction for all percentiles, with a near five-fold reduction at the 99th percentile.

The computational overhead of using stratified sampling was less than 5% and so has been ignored.

Cheng (1984) discussed an antithetic technique for this problem which used antithetic sampling of certain control variables. A comparison of the results shows that the stratification method described here is more effective. However it would be possible to use stratified rather than antithetic sampling of the control variables, and it is hoped to report this elsewhere.

Table 1: Comparison of the Monte Carlo Estimates of the Mean and Selected Percentiles of the ml estimator, $\hat{\beta}$, of the shape parameter β in the Weibull distribution, using Random and Stratified Sampling

Number of replicates = 1000
$d = 5$, $k = 10$, $J = 2002$
True value of $\beta = 1$

	Mean of β	Percentile 1%	10%	50%	90%	99%
Replicate Means:						
Indep. Runs	1.4423	0.550	0.769	1.243	2.303	4.473
Strat. Runs	1.4420	0.551	0.766	1.240	2.306	4.533
Replicate Variances × 10^4						
Indep. Runs	3.4	2.0	1.0	2.2	25.0	751.
Strat. Runs	0.3	1.7	0.78	0.88	6.0	155.
Variance Ratio (≈ Efficiency)						
Indep/Strat	11.3	1.2	1.3	2.5	4.2	4.8

5.2 The Distribution of a Test Statistic

A number of goodness of fit test statistics have been proposed (see d'Agostino and Stephens, 1986, for a review) based on the differences between ordered observations. It is of some interest to tabulate the percentiles of such statistics under the null hypothesis. A well known example is Moran's statistic whose null distribution is that of

$$M = -\sum_{i=1}^{d+1} \log(U_{(i)} - U_{(i-1)}) \qquad (5.1)$$

where $U_{(1)} < U_{(2)} < \ldots U_{(d)}$ are an ordered sample of independent uniform $U(0,1)$ variates; and $U_{(0)} \equiv 0$, $U_{d+1} \equiv 1$. Though its asymptotic distribution is known, curiously there has been no attempt to tabulate its distribution for small samples until recently (see Cheng and Thornton, 1988). It will be of interest therefore to try Monte Carlo simulation in this case.

From (5.1) it is clear that M is symmetric in its components and so we can use the stratification scheme of Section 4.2 to estimate the small sample properties of M.

Table 2 gives the results of estimating the mean of M together with selected percentiles of its distribution for the case where $d+1 = 5$ with $k=18$. This uses 5985 points in one stratified run. To estimate the variability of estimates the run was repeated 500 times. For comparison Table 2 also shows the estimates and variances from 500 replicates each of which estimated the mean and the percentiles from 5985 independent runs.

As will be seen, stratification leads to variance reduction for all percentiles as well as for the mean even taking into account the computational overhead. The typical saving of 50% is perhaps not dramatic in this particular case. Two features of this problem are perhaps unusual in this respect. The statistic happens to be rather easy to calculate, so that the computational overhead of using stratified sampling is more noticeable than in general. Secondly, Moran's statistic is known to be sensitive to the smallest differences $U_{(i)} - U_{(i-1)}$ appearing in (5.1). Our method of sampling does not take especial advantage of this. A more effective application would probably be to use stratification directly on these differences.

These features should therefore not be allowed to obscure the main point, which is that the stratification scheme is extremely easy to apply and in general carries little computational overhead or danger of being variance increasing. The example does show that even where the stratification is not being applied in a subtle way, it still gives worthwhile variance reduction.

Table 2: Comparison of Monte Carlo Estimates of the Mean and Selected Percentiles of Moran's Statistic using Random and Stratified Sampling

Number of replicates = 500
d = 4, k = 18, J = 5985

	Mean of M	Percentile 1%	10%	50%	90%	99%
Replicate Means:						
Indep. Runs	10.415	8.23	8.69	10.05	12.62	15.74
Strat. Runs	10.416	8.23	8.69	10.05	12.62	15.71
Replicate Variances × 10^4						
Indep. Runs	4.27	1.51	2.35	6.16	30.9	257.0
Strat. Runs	1.30	0.79	0.86	2.93	17.3	151.7
Variance Ratio Indep/Strat	3.3	2.9	2.7	2.1	1.8	1.6
Efficiency Ratio (Variance Ratio × Labour Ratio)	2.5	1.5	2.1	1.6	1.4	1.3

REFERENCES

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 Thornton, K.M. (1988). Selected percentage points of the Moran Statistic. *Journal of Statistical Computation and Simulation*. To appear.

Cheng, R.C.H. and Davenport, T. (1988). The problem of dimensionality in stratified Monte Carlo sampling. *MATH Report 88-2*, UWIST, Cardiff, U.K.

D'Agostino, R.B. and Stephens, M.A. (1986). Goodness of fit techniques. *Marcel Dekker*, New York.

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.

Niederreiter, H. (1978). Quasi-Monte Carlo methods and pseudo-random numbers. *Bull. Amer. Math. Soc.*, 84, 957-1041.

Nijenhuis, A. and Wilf. H.S. (1978). Combinatorial algorithms. *Academic Press*, New York.

Ripley, B.D. (1987). Stochastic simulation. *Wiley*, New York.

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

Thoman, D.R., Bain, L.J. and Antle, C.E. (1969). Inferences on the parameters of the Weibull distribution. *Technometrics*, 11, 445-460.

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

AUTHORS' BIOGRAPHIES

RUSSELL C.H. CHENG obtained a B.A. 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 in 1972 and was appointed Reader in 1988. His main fields in 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.

School of Mathematics,
University of Wales College of Cardiff,
Senghennydd Road,
Cardiff CF2 4AG,
Great Britain.

TERESA R. DAVENPORT received her B.Sc.(Hons) degree in Statistics with Management Science Techniques in 1986 from the University of Wales. She became interested in discrete event simulation and was awarded a University of Wales Research Studentship to study variance reduction techniques. She is currently enrolled in the doctoral program of the School of Mathematics, University of Wales, Cardiff.