Proceedings of the 1992 Winter Simulation Conference
ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson

DISTRIBUTION FITTING AND RANDOM NUMBER AND VARIATE GENERATION

Russell C.H. Cheng

University of Wales College of Cardiff
Senghennydd Road,
Cardiff, U.K.

ABSTRACT

Discrete—event simulation programs invariably use random variates to model chance fluctuations. The basic requirement is to appropriately select probability distributions, such as the exponential or normal, and then sample variate values from these distributions to represent quantities, such as arrival and service times, needed to carry out the simulation itself. This tutorial discusses the basic techniques and the points that need to be borne in mind for both selecting distributions and generating variate values from these, once chosen.

1. INTRODUCTION

Random variate generation is a basic, and hence important, element of a simulation program. However, because of its specific nature it forms a rather narrow part of simulation. Providing the user has an understanding of elementary ideas about random variables and their distributions, then the standard methods available for their generation are easily explained and understood. Moreover variate generation routines have become increasingly available in computer library subroutines for an ever—widening list of distributions. It is all too easily possible, especially at the introductory level, for a user to conduct simulations with only a sketchy knowledge of the processes or methods used for generating random variates. With these points in mind, I shall try to discuss two aspects.

Firstly, in accordance with the intention of the tutorial, I shall cover basic methods of variate generation. However there are many excellent texts with much more space to fully describe the range of methods available. The aim here will be to provide a brief guide, highlighting points of particular generality or usefulness, which will hopefully provide an easy introduction to more comprehensive accounts.

The second aim is much more difficult but arguably more important and interesting for the practitioner. Stochastic simulation is just one particular kind of statistical study. And the core of all statistical studies is the examination of random variables of interest and their distributions. Rather tritely, these are the only two fundamental questions we need or ***can*** ask: What are the quantities of interest in the simulation? What are their distributions? A consequence of this is that a user needs to have a clear idea of certain basic properties of random variables if a simulation is to be built on other than a shaky foundation. In particular the choice of what input distributions to use is very important as far as the modelling process is concerned.

Once input distributions have been selected, the mechanics of generating variates from them is actually rather simple. In fact, as far as discrete event simulation is concerned relatively few specific distributions are seriously used in practice. The main continuous theoretical distributions are the uniform, normal, exponential, gamma, lognormal, Weibull, beta of the 1st and 2nd kind and possibly the triangular. I would personally add the inverse Gaussian distribution making it the round ten. This last is undeservedly obscure. It is easy to use, and is interesting both theoretically and practically with useful structure and characteristics not present in the others. The main discrete distributions are the discrete uniform, Bernoulli, binomial, Poisson, geometric and negative binomial.

As far as generation of variates from these distributions is concerned it would be possible simply to give cookbook recipes for each. However this would not be very insightful and it is not done here. But the user should be aware that practically

this is all variate generation boils down to.

Section 2 recalls some basic definitions concerning random variables and their distributions. It is assumed these ideas are familiar.

Section 3 discusses the issues involved in choosing appropriate input distributions. There are two main aspects. How to select a distribution by fitting a theoretical distribution to actual data, and how to assess if the fit is a good one.

Section 4 describes the basic methods of variate generation, and Section 5 discusses some specific examples drawn from the above list.

There are many good accounts of random variate generation. For this tutorial I found Law and Kelton (1991, 2nd Ed), Lewis and Orav (1989) and Morgan (1984) most apposite; in particular the order of presentation of material in Law and Kelton seemed just right and I have used essentially the same order; but with my own comments, of course.

Finally, I have confined discussion specifically to variate generation in the context of discrete event simulation. Variate generation in the context of simulation in statistics is actually a much richer, more demanding and interesting area. Lewis and Orav place some emphasis on this and I would certainly recommend also Ripley (1987) and Devroye (1986). There are some truly general variate methods in this latter reference which I have found particularly useful in statistical studies of non—standard distributions.

2. RANDOM VARIATES AND THEIR DISTRIBUTION

This section is simply a stricture on the absolute minimum about random variates and probability distributions that someone needs to know if their use in simulations is to be meaningful.

Formally a random variable, X is a function defined on a sample space (Law and Kelton §4.2). In a simulation experiment this sample space comprises all the possible outcomes and a random variable is simply some quantity of interest associated with these outcomes. For example in a single server queue, X might be the total number of customers served, the length of time it takes to serve the first customer, the time of arrival of the second customer, and so on. Because chance assigns different probabilities to these outcomes, this means that X takes on different values with different probabilities. The cumulative distribution function (cdf), F(x), is a very convenient way of defining these probabilities:

$$F(x) = P(X \leq x) \quad -\infty < x < \infty \tag{2.1}$$

where $P(X \leq x)$ means the probability that, as a result of the experiment, X takes on a value less than or equal to x. The main properties are that: $0 \leq F(x) \leq 1$; it is nondecreasing as x increases; F(x) tends to 0 as $x \to -\infty$ and to 1 as $x \to \infty$. If X can take a continuous range of values (as is the case for example when X is an arrival time) then it is termed a continuous random variable. Then F(x) can be written as

$$F(x) = \int_{-\infty}^{x} f(y)dy \tag{2.2}$$

where f(x) is called the probability density function (pdf).

The value of f at x, f(x), is always positive and it gives, in a precise way, the relative chance of X taking a value close to x. The larger f is, the greater this chance. The shape of f(x) is thus important in determining the behaviour of X. For instance the value of f in the 'tails' of the distribution, viz. when x is large in magnitude (negative or positive) determines how often extreme values of X will occur.

The equation

$$p = F(x) \quad -\infty < x < \infty \tag{2.3}$$

defines the distribution in the sense that it tells us the probability or p—value corresponding to different x. However it can be regarded in its inverse form as determining x, call it x_p, for a given p—value; x_p is called the pth quantile or percentile or percentage point. The equation then becomes

$$x_p = F^{-1}(p) \quad 0 < p < 1 \tag{2.4}$$

where F^{-1} is the inverse of F. An example of this inversion can be seen with the exponential distribution. This has cdf $F(x) = 1 - e^{-\alpha x}$, $x > 0$, and for this case (2.4) takes the simple form

$$x_p = -\alpha^{-1} \log(1-p). \tag{2.5}$$

We shall see in the next section, that, as in the exponential case, variate generation from the

distribution with cdf F(x) is easy when F^{-1} can be written in closed form.

If X can take only a fixed set of prescribed values $x_o, x_1, x_2, \ldots$, (for example when X is the number of customers served in a given time) then it is called a discrete random variable and F(x) becomes piecewise constant. The analogue of (2.2) is

$$F(x) = \sum_{x_i \leq x} p_i \qquad (2.6)$$

where p_i = probability that X equals x_i. The quantile x_p is not uniquely determinable for all p values in this case. If $p = \sum_{i=o}^{j} p_i$ for some j then a range of x values satisfies (2.6). However if we restrict choice to the prescribed values and set

$$x_p = x_i \text{ where } F(x_{i-1}) < p \leq F(x_i), \qquad (2.7)$$

with $F(x_{-1}) = 0$, then this fixes x_p uniquely given $0 < p < 1$ so that (2.7) may be regarded as the analogue of (2.4). An elementary example is the Bernoulli random variable

X=1 with probability θ, X = 0 otherwise.

Here

$$\begin{aligned} x_p &= 0 \quad \text{if} \quad 0 < p \leq 1-\theta \\ &= 1 \quad \text{if} \quad 1-\theta < p < 1 \end{aligned} \qquad (2.8)$$

The above terminology refers to underlying population distributions. A corresponding terminology exists for a random sample of observations. If the sample is ordered $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$, then the empirical distribution function is defined as

$$\begin{aligned} F_n(x) &= 0 \quad x < x_{(1)} \\ &= \frac{i}{n} \quad x_{(i)} \leq x < x_{(i+1)} \quad 1 \leq i \leq n-1 \\ &= 1 \quad x_{(n)} \leq x \end{aligned} \qquad (2.9)$$

and this is the analogue of (2.6). The analogue of (2.7) is

$$x_p = x_{(i)} \quad \text{where } \frac{i-1}{n} < p \leq \frac{i}{n}. \qquad (2.10)$$

3. CHOOSING INPUT DISTRIBUTIONS

3.1 Input Distributions

When a simulation run is carried out, random variates will be generated from various distributions to represent quantities like interarrival times between customers and service times of customers. Selecting appropriate distributions to use is a very important step in building the simulation model. It is certainly much more difficult than the actual business of generating random variates once the distribution has been chosen.

The approach will depend on whether information or data exists about the input variables of interest or not.

If information does exist, for example if there are records of times taken to serve customers, then there are three possibilities. If there is enough such information then the data themselves can be used directly. Secondly the bootstrap approach can be taken whereby the data are used to form an empirical distribution function and variates can then be generated from this empirical distribution treating it as being the population distribution. Thirdly a theoretical distribution can be fitted to the data using some standard statistical estimation technique. The theoretical distribution is known up to some unknown parameters and the fitting process comes up with estimates of these parameters based on these data.

The first approach is very useful for comparison purposes. For example in a study of gas demand and how best to control supply, the investigators found it useful to use the actual day to day temperature variations for a year when there had been known to be gas supply difficulties. The third method is attractive on a number of counts. Use of a theoretical distribution with known properties adds 'structure' to the overall model. This is often useful in giving insight into the behaviour of the system by showing how it depends on specific characteristics of the input distributions. There is one danger in using a theoretical distribution. It may be that the actual distribution has characteristics that are subtly different from those which can be represented by the chosen theoretical

distribution. For instance the probability of extreme events occurring is determined by the 'tail' behaviour of the distribution. If a theoretical distribution is used whose tail behaviour is not appropriate then the occurrence of extreme events will not be correctly modelled in the simulation.

Even if an appropriate theoretical distribution is chosen there remains the problem of assessing the accuracy of the estimated parameters. For instance if the parameter α of the exponential distribution of (2.5) had been estimated then some account should be taken of this in the simulation. One possibility is to carry out simulations at both the upper and lower limits of a confidence interval for the unknown true α—value.

A similar approach can be adopted in the situation when no data are available. Theoretical input distributions should be used with sufficient flexibility to model the range of behaviour that might occur. Simulations can then be run with the parameters set at different values to reflect this range of behaviour.

We now look at the two main aspects of fitting distributions in more detail: how to estimate parameters and assess the accuracy of these estimates and how to check if the fitted distribution is a good fit to the data.

3.2 Fitting Distributions

Suppose we have selected a candidate distribution to represent a required input distribution. For example we might think the gamma distribution is an appropriate model of service times in some queueing model. Its density is $f(x) = x^{\alpha-1} e^{-x/\beta}/(\Gamma(\alpha)\beta^{\alpha})$. The distribution is specified apart from the values of the parameters α and β. We wish to estimate α and β using n observed sample service times: $x_1, x_2, \ldots, x_n$. We denote the unknown parameters by a vector θ. By far and away the most powerful general method for doing this is the method of maximum likelihood (ML). The likelihood is simply the product of the pdf's evaluated at $x_1, x_2, \ldots, x_n$, viewed as a function of θ. The maximum likelihood estimates, $\hat{\theta}$, are the values of the parameters which maximize the likelihood. Sometimes it is possible to obtain closed form expressions for maximum likelihood estimators, for example this occurs with the mean and variance of the normal distribution. In general a numerical technique has to be used (see Law and Kelton for a discussion). I find that, in general, a search technique which works by directly evaluating the likelihood at selected parameter values is as convenient a method as any. This avoids having to evaluate derivatives.

A very important feature of the maximum likelihood method is the fact that the distribution of the estimators, under fairly general conditions, approaches normality as the size of the sample used in the estimation becomes large. This allows confidence intervals to be found in which the unknown true parameter value lies with prescribed degree of confidence. There are three basic ways of doing this calculation based on the Wald statistic, score statistic or the likelihood ratio. The last method appears to be the most accurate. It additionally best matches the numerical search technique for finding the ML estimates. It is based on the result that

$$2(L(\hat{\theta}) - L(\theta^*))$$

is approximately chi—squared with p degrees of freedom. Here p is the number of parameters fitted, θ^* is the unknown true parameter value and

$$L(\theta) = \sum_{i=1}^{n} \log f(x_i, \theta)$$

is the log of the likelihood. The import of the result is that if we evaluate L at a θ value and compare it with $L(\hat{\theta})$, then $L(\theta)$ is unusually small compared with a chi—squared variate with p degrees of freedom and will indicate a θ value unlikely to be the true one. The easiest way is to implement this result for each parameter separately. Consider θ_1 and denote by $L^*(\theta_1)$ the value of the likelihood maximised with respect to the other parameters assuming θ_1 fixed. If we denote by $\chi_1^2(\alpha)$ the upper α quantile of the chi—squared distribution with one degree of freedom and carry out a search on θ_1 for the two values, θ_{1L} and θ_{1U} say, satisfying

$$L(\hat{\theta}) - L^*(\theta_1) = \tfrac{1}{2}\chi_1^2(\alpha),$$

then these two points give the limits of a α 100% confidence interval for θ_1.

The attraction of this method is that as derivatives are not involved it is easy to modify the original ML estimation process to do this calculation also.

This uncertainty in the parameter value can be taken into account by carrying out simulations at both θ_{1L} and θ_{1U}.

3.3 Goodness of Fit Tests

If there is any doubt that a selected input distribution is not appropriate, because its characteristics do not properly represent or display all those of the sampled observations, then it is important to carry out tests to see if the fit is a good one or not.

There are a whole battery of so—called goodness of fit tests that can be used to assess how well the fitted distribution represents the sample. Arguably the best known is the chi—squared goodness of fit test. This works by comparing sampled and fitted frequencies and is in effect a comparison of density functions. It is easy to carry out and has the merit of allowing the fact that parameters have been fitted to be taken into account in the test process. It has a disadvantage in that it requires the observations to be grouped into a number of ranges, and there is a certain subjectiveness in this.

Various good tests are available which compare the empirical distribution function with that of the fitted cdf. The best known is the Kolmogorov Smirnov test which looks at the maximum difference between the two distribution functions. It has two obvious weaknesses. Firstly, because both distribution functions are bound to start at zero and end at unity, the test tends to be insensitive to differences in the distributions in the tails. Secondly it is sensitive to whether parameters have been fitted or not. Stephens (1986) shows how with several commonly occurring distributions this can be allowed for in the test.

I would also recommend the Anderson—Darling test which also looks at the difference between the empirical and fitted distributions, but in an aggregated way with especial weight given to the tails. In practice this is a very sensitive test and it is not really much more difficult to carry out than the others.

Finally it is worth considering graphical methods which give a clear indication if a fit is not good and if not, why not. The so—called quantile—quantile (Q—Q) plot or the probability—probability (P—P) plot, as their names imply, compare like quantile and like probability values. When equal this gives straight line plots. Differences show up as deviations from the straight line. Good discussions of these plotting methods are given by Law and Kelton and by Lewis and Orav.

The former authors give helpful suggestions as to the general shape of different distributions and in what applications they are likely to be of use.

4. RANDOM VARIATE GENERATION

4.1 Random Numbers

Before discussing random variate generation in general we look at the special case of random numbers. Like many mathematical terms with an apparently non specific name, the term random numbers is used in a precise sense; it always means a uniform random variate from the continuous uniform distribution U(0,1) i.e. with pdf $f(x) = 1$ for $0 < x < 1$, $f(x) = 0$ otherwise. Its pre—eminent place lies in the fact that it provides a convenient starting point, both theoretically and practically, for generating variates from other distributions. Random numbers will be denoted by U or U_1, U_2.

In practice on a computer, instead of producing true random numbers one has instead what are deterministic, often called pseudo random number, algorithms which produce a stream of numbers with the appearance of randomness. The theory and the methods of testing such generators is one of the most difficult in simulation. At the introductory level it would be inappropriate to attempt any kind of serious technical discussion. However some brief remarks are in order.

Most computing systems provide a generator. But there have been instances of examples whose performance has not been satisfactory. (See Ripley 1987, Lewis and Orav 1988 for instances where the deterministic nature of pseudo random numbers can resurface in disconcerting ways.) Some important features to look out for and check are discussed by Ripley (1989) and include closeness to being uniformly distributed, especially independence in a reasonable number of dimensions, repeatability, speed and long period.

The most commonly occurring generators seem to be of linear congruential or of shift register type together with shuffling to increase apparent randomness.

My main suggestion to a (first time) user of a

random number generator is to be aware of the requirements, to make sure that they use a generator that has been documented and if possible to ascertain that at least some of the characteristics mentioned above have been investigated and found satisfactory.

If there is any doubt it is better to avoid the library function and incorporate a routine one has selected oneself. Ripley (1989), Lewis and Orav (1988), Bratley, Fox and Schrage (1983) give simple examples with tested provenance which are easy to code.

All methods of generating random variates from a given distribution use random numbers as their starting point.

At the elementary level there are really only three methods that can claim to be general. The inverse transform, the composition and the acceptance/rejection method. Each will be described briefly.

4.2 The Inverse Transform Method

This method can be applied to both discrete and continuous variables. Equations (2.7) and (2.4) are formulas for the pth quantile x_p in terms of the probability value p. If p is actually generated as a U(0,1) variable U say, then x_p becomes a random variable, X say. Its cdf is

$$\begin{aligned} P(X \leq x) &= P(F^{-1}(U) \leq x) \quad \text{(from 2.4)} \\ &= P(U \leq F(x)) \\ &= F(x). \end{aligned} \tag{4.2}$$

Thus X generated in this way has precisely the distribution we want. In the example of the exponential distribution this gives a closed formula

$$X = -\alpha^{-1} \log(1-U). \tag{4.2}$$

In fact (1—U) can be replaced by U itself as both have the same distribution, U(0,1).

The result works for the discrete case using (2.7) instead of (2.4). The Bernoulli case (2.8) reduces to the obviously sensible

$$\begin{aligned} X &= 0 \quad \text{if } 0 < U < 1 - \theta \\ &= 1 \quad \text{if } 1 - \theta < U < 1. \end{aligned} \tag{4.3}$$

4.3 The Composition Method

Sometimes it is possible to write the cdf of interest as

$$F(x) = \sum_{j=1}^{J} p_j F_j(x) \tag{4.4}$$

where $\{p_j\}$ is a discrete probability distribution and the $F_j(x)$ are cdf's. F(x) is called a mixture distribution. The mixture may arise naturally from the context of the problem. Alternatively it may simply be an ingenious but contrived decomposition. Either way it is possible to generate X with distribution F(x) by selecting the jth cdf of the set F_j with probability p_j, generating a variate X_j with distribution F_j, and setting $X = X_j$.

The main way the method is implemented is to try and make p_1 as large as possible and at the same time to have F_1 a distribution that is easy to generate from ; p_1 large then ensures that F_1 is chosen frequently. A typical example is the Marsaglia and Bray (1964) 'convenient' method for generating normal variates, where over 86% of the time a linear combination of three U(0,1) variables is taken. The remainder of the time more elaborate cdf's have to be used.

The composition method can be used for discrete distributions rather neatly. See Law and Kelton for an example.

4.4 Acceptance—Rejection

This is mainly used for continuous random variables. Suppose f(x) is the pdf of the distribution we want to generate variates from. Suppose g(x) is the pdf of another distribution for which we already have a method of generating variates from, and that moreover we can find a scaling factor K so that

$$e(x) = Kg(x) > f(x). \tag{4.5}$$

Thus e(x) is an envelope whose graph lies completely above the graph of f(x). If we generate X so it has pdf g(x) and take Y = Ue(X) where U is U(0,1) and independent of X, then (X,Y) is a point under the graph of e(x) and in fact it turns

out to be uniformly distributed under this graph. Imagine a large number of points generated in this way. They will be scattered uniformly under e(x). If we now insert the graph of f(x) and discard points that lie above it then the remainder will be uniformly distributed under f(x). Thus the number of points with a given x value will be proportional to f(x). In practice X's with the right distribution can be obtained one at a time: points (X,Y) are generated until one is found satisfying $Y < f(X)$; then X is the next accepted variate value. A formal proof is given in Law and Kelton Appendix 8A.

The efficiency of acceptance rejection methods is dependent on the scaling factor K of (4.5) which gives the average number of pairs (X,Y) needed for each X accepted.

As an example consider the gamma distribution with

$$f(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-x} \quad x > 0.$$

A large number of acceptance rejection methods have been suggested for this distribution. I often use the method of Fishman (1976, & G3 1978) valid for $\alpha > 1$, provided α is less than 3 say, as it is one of the simplest methods. It uses a negative exponential envelope. The scaling factor K is proportional to $\alpha^{\frac{1}{2}}$ so the method becomes increasingly inefficient as α increases. The method, GB, using a log logistic envelope which I suggested (Cheng, 1977) has $K \leq 1.47$ for all $\alpha \geq 1$, and is satisfactory for larger α.

4.5 Discrete Distributions

The above methods can be applied to discrete distributions reasonably satisfactorily. For example the inverse transform method works well if the distribution function is known numerically. In this latter case because of the way the method then works it is sometimes called the 'table—look—up' method.

There is a further general approach introduced by Walker (1977) and improved by Kronmal and Peterson (1979) called the alias method which is powerful and quick. A detailed description is given by Law and Kelton. The method has a modified acceptance technique where a uniform variate is tested. Depending on the outcome one of two variate values is returned (hence the name alias). As a variate is always returned, no rejection occurs. The method is expensive to set up and applies only to discrete distributions with a finite range, but is otherwise very effective.

5. EXAMPLES

Though the basic methods are reasonably useful two reservations need to be made. Firstly even if, for the given distribution of interest, one of the basic methods can be used, it may be that some special property peculiar to that distribution can be used to produce a more effective method. Indeed there are many relationships between different cases so that using certain distributions as basic building blocks a fairly comprehensive set of distributions can be covered.

A typical example of a special method is if independent pairs of U(0,1) variates are simply added:

$$X = U_1 + U_2.$$

The resulting variate X has the triangular distribution.

Convolution sums are a convenient way of generating variates from certain other distributions. Erlang variates can be obtained as sums of exponentials, chi—squared variates as sums of squares of independent standard normals.

Others can be obtained as ratios. For example Y_1, Y_2 are independent gamma variates with parameters α_1 and α_2 then $X = Y_1/(Y_1 + Y_2)$ is a beta variate with parameters α_1 and α_2.

We end with the celebrated example of the Box Muller method for generating standard normal variates. There is a simple way of viewing the method geometrically.

Suppose we imagine ourselves to be at the origin of the xy plane and take one step in the x direction and one in the y direction. If each step is standard normal in length z_1, z_2 say, then the overall distance D travelled from the origin satisfies $D^2 = z_1^2 + z_2^2$. This makes D^2 an exponential variable with parameter 2. Moreover any direction is equally likely. In other words the random angle so obtained has the uniform distribution $V = 2\pi U_1$. The Box Muller method simply reverses the

process. Take a point at distance D from the origin, using the inverse transform method to generate D say, at a random angle V (from the x axis). Then the projections of this point onto the x and y axis are:

$$Z_1 = D \cos 2\pi U_1 \, , \, Z_2 = D \sin 2\pi U_1$$

and these will be independent N(0,1) variables.

REFERENCES

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

Cheng, R.C.H. 1977. The generation of gamma variables with non—integral shape parameter, *Applied Statistics*, 26, 71—75.

Devroye, L. 1986. *Non—Uniform Random Variate Generation.* New York: Springer—Verlag.

Fishman, G.S. 1976. Sampling from the gamma distribution on a computer. *Communications of the ACM*, 19, 407—409.

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

Kronmal, R.A. and Peterson, A.V. Jr. 1979. On the alias method for generating random variables from a discrete distribution, *American Statistician*, 33, 214—218.

Law, A.M. and Kelton, W.D. 1991. *Simulation Modeling and Analysis. Second Ed.* New York: McGraw—Hill.

Lewis, P.A.W. and Orav, E.J. 1989. *Simulation Methodology for Statisticians, Operations Analysts, and Engineers, Vol. 1*, Pacific Grove California: Wadsworth and Brooks/Cole.

Marsaglia, G. and Bray, T.A. 1964. A convenient method for generating normal variables. *SIAM Review*, 6, 260—264.

Morgan, B.J.T. 1984. *The Elements of Simulation.* London: Chapman and Hall.

Ripley, B.D. 1987. *Stochastic Simulation.* New York, Wiley.

Ripley, B.D. 1989. Thoughts on pseudorandom number generators. *Journal of Computational and Applied Mathematics*, 31, 153—163.

Stephens, M.A. 1986. *EDF Statistics. Chapter 4 of Goodness—of—Fit Techniques* (Ed. R.B. d'Agostino and M.A. Stephens): New York: Marcel Dekker.

Walker, A.J. 1977. An efficient method of generating discrete random variables with general distributions. *ACM Transactions on Mathematical Software*, 3, 253—256.

RUSSELL C.H. CHENG is Reader in Statistics at the University of Wales, College of Cardiff. He 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 assocation 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 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. He has published a number of articles in all these fields.