Proceedings of the 1994 Winter Simulation Conference
ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila

SELECTING INPUT MODELS

Russell C. H. Cheng

Institute of Mathematics and Statistics
The University of Kent at Canterbury
Canterbury, Kent CT2 7NF
England

ABSTRACT

Discrete-event simulation almost invariably makes uses of random quantities drawn from given probability distributions to model chance fluctuations. This advanced tutorial discusses how to choose appropriate distributions. The two main points addressed are (a) the factors which should be considered in selecting input distributions, and (b) how the effect of errors in making an inappropriate or inexact choice will influence the accuracy of final simulation results.

1 INTRODUCTION

I should mention at the outset that I have given two introductory WSC tutorials (Cheng 1992, 1993) on input distribution selection and on variate generation. Those tutorials concentrated on the elementary aspects of distribution fitting and variate generation. The question of distribution selection was only touched upon. The methodologies of how to fit distributions and how to generate variates are well developed and are relatively easy to understand. Once these aspects have been dealt with then the most interesting problem, from the point of view of the practitioner, is how to select or choose candidate distributions.

In this advanced tutorial I take up the story where the introductory tutorials left off, focusing on two particular aspects of input modelling. Firstly, there are the wider issues underlying distribution selection. In particular I shall try to consider rather broader classes of distributions than usually considered. The other aspect that will be considered is the effect on the accuracy of the final simulation output when errors are made in the choice or fitting of input distributions. This is a topic that has received relatively little attention in the literature. The accuracy of the final results of a simulation is clearly influenced by the accuracy of the input variate streams. The tutorial will try to quantify this more precisely. It is a question that deserves a lot more attention not least because of its practical importance.

The references drawn on in this tutorial are understandingly more widely scattered than for the introductory tutorials. Many interesting ideas have been proposed through this Conference in previous years. Good basic references which I have found useful are Law and Kelton (1991, 2nd Ed.), Lewis and Orav (1989) and Bratley, Fox and Schrage(1983). Since the aspects to be discussed are mainly statistical, the books by Devroye (1986) and Ripley (1987) might also be consulted as they focus more on statistical aspects.

2 THE SIMULATION MODEL

We wish to highlight the use of empirical data in simulations. We shall assume that the simulation requires k streams of variates. For each stream we have available a sample of empirical data:

$$\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{il_i}) \; i = 1, 2, ..., k. \qquad (1)$$

Each sample is assumed to have a joint distribution with probability increment $DF_i(\mathbf{x}_i, \theta)$ $(i = 1, 2, ...k)$. We assume that the distribution depends on a vector $\theta = (\theta_1, \theta_2, ..., \theta_p)$ of p parameters which are unknown. If the sample is a random sample then the probability increment becomes a product of l_i copies of a univariate increment: $DF_i(\mathbf{x}_i, \theta) = \prod_{j=1}^{l_i} DF_i(x_{ij}, \theta)$, but in our general discussion we do not necessarily have to assume this.

The variates used in one simulation run will be denoted by

$$\mathbf{x}_i^* = (x_{i1}^*, x_{i2}^*, ..., x_{im_i}^*) \; i = 1, 2, ..., k. \qquad (2)$$

It will be convenient to regard these as having been obtained by transformation

$$\mathbf{x}_i^* = \phi_i(\mathbf{u}_i, \theta) \qquad (3)$$

of a corresponding set of independent uniform $U(0,1)$ variates:

$$\mathbf{u}_i = (u_{i1}, u_{i2}, ..., u_{im_i}) \; i = 1, 2, ..., k. \quad (4)$$

Note that we assume the uniforms to be independent, but the components of the $\mathbf{x}_i^*$ are not necessarily so. The output of interest from the simulation run, y, can then be thought of as being a function of the $\mathbf{u}_i$, albeit a complicated one:

$$y = y(\mathbf{U}, \theta), \quad (5)$$

where

$$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k). \quad (6)$$

We regard the objective as being to estimate the expected value of y, and note that this is a function of θ only:

$$\eta(\theta) = E(y, \theta) = y(\mathbf{U}, \theta)\mathbf{dU}. \quad (7)$$

As we have already remarked, we wish to consider how the empirical data can be used in the simulation and how its use affects inferences about $\eta(\theta)$. There are three main possibilities. We can either directly use these data as the input variable streams, or we can use some resampled version of these samples as occurs for instance in 'jackknife' methods, or we can fit parametric models which exactly reproduce, or at least accurately approximate, the $F_i(\mathbf{x}_i, \theta)$ and then sample variates from these fitted distributions for use in the simulation.

We shall consider asymptotic properties which obtain when $l_i \to \infty$, all i. To simplify the notation we shall assume that $l_i = \alpha_i l$ where the proportions α_i are assumed fixed with $\sum \alpha_i = 1$. This allows us to give asymptotic results in the form $l \to \infty$ rather than consider each l_i separately. Obvious variations to the results apply if the $l_i \to \infty$ at different rates.

In principle the m_i can also be thought of as being variable. However it is more convenient to think of the length of a simulation run as being fixed so that the m_i are thus also fixed. We consider the overall simulation experiment as being made up of n runs. The responses or outputs from these runs will be written as:

$$y_j(\mathbf{U}_j, \theta) = \eta(\theta) + e_j(\mathbf{U}_j, \theta), \; j = 1, 2, ..., n. \quad (8)$$

Note that these outputs depend on the parameter vector θ. The 'error' variable e_j is the random difference between the jth simulation run output and $\eta(\theta)$. We shall assume $E(e_j) = 0$ and $Var(e_j) = \sigma^2$ for $j = 1, 2, ..., n$. Thus, assuming that θ is fixed for the moment, we have

$$E[y_j(\mathbf{U}_j, \theta)] = \eta(\theta), \quad (9)$$

and the mean of the outputs

$$\bar{y} = \bar{y}(\mathbf{U}_1, \mathbf{U}_2, ..., \mathbf{U}_n, \theta) = \sum_{j=1}^{n} y_j(\mathbf{U}_j, \theta)/n, \quad (10)$$

is an unbiased estimator of $\eta(\theta)$ with

$$Var[\bar{y}] = \sigma^2/n. \quad (11)$$

3 FITTING MODELS BY MAXIMUM LIKELIHOOD

We gather together some well-known facts involving *maximum likelihood* (ml) estimation that we shall use in what follows (see for example Efron and Tibshirani, 1993). In a number of situations we will suppose that some particular distribution has been selected to be an input model that is dependent on the vector θ of unknown parameters. The likelihood is simply the joint distribution probability element evaluated at the observed values, and then treated as a function of θ. It is usually easier to work with its logarithm (log-likelihood):

$$L(\theta) = \sum_{i=1}^{k} \log Df_i(\mathbf{x}_i, \theta). \quad (12)$$

Let θ_0 denote the unknown true parameter value. Its maximum likelihood estimate, $\hat{\theta}$, is the value of θ which maximizes the loglikelihood. The derivative of $L(\theta)$, $L'(\theta) = \partial L(\theta)/\partial\theta$, is called the score function. When the loglikelihood is maximized at an interior point of the parameter space then $\hat{\theta}$ satisfies $L'(\theta) = \mathbf{0}$. The observed and expected information are defined as

$$I(\theta) = -\partial^2 L(\theta)/\partial\theta^2 \; and \; i(\theta) = E[I(\theta)]. \quad (13)$$

Under general regularity conditions the ml estimate has the important property that its distribution is asymptotically normal as the sample size $l \to \infty$, i.e.

$$\hat{\theta} \sim N(\theta_0, \mathbf{i}(\theta_0)^{-1}) \; l \to \infty.$$

In practice, as θ_0 is unknown, use is made of one of the asymptotically equivalent versions: $\hat{\theta} \sim N(\hat{\theta}, \mathbf{i}(\hat{\theta})^{-1})$ or $\hat{\theta} \sim N(\hat{\theta}, \mathbf{I}(\hat{\theta})^{-1})$. A typical application of this result is its use in the construction of confidence intervals which contain the unknown true parameters with prescribed degree of confidence.

4 SELECTING INPUT MODELS

4.1 Univariate Models

Input models or distributions are the probability distributions of random variables used to drive the sim-

ulation. We begin by considering univariate models. In discrete event simulation relatively few theoretical distributions are used in practice. The main continuous distributions are the uniform, normal, exponential, gamma, lognormal, Weibull, beta of the first and second kinds, the triangular and the inverse Gaussian. The main discrete distributions are the discrete uniform, Bernoulli, binomial, geometric and negative binomial. A good review of these distributions is given for example by Law and Kelton (1991). Many statistical packages, whether they are specifically targetted for the simulation community or not, contain fitting routines for many of these. For example Palisade Corporation's 'BestFit' allows fitting of a whole range of distributions, together with ranking based on goodness of fit criteria so that the user has guidance on which distribution will be satisfactory or best.

There are two provisos. Firstly, despite the apparent choice, there are limitations. Only the most widely known models tend to be considered and these tend to be the ones implemented in packages. For example, despite its clear analytic tractability and its useful provenance, the inverse Gaussian distribution is often not listed in simulation texts as a possible model, and not always included as a possible candidate in packages. Such omissions are clearly just a matter of taste.

The second proviso is more serious. Use of standard distributions tends to limit the amount of control one has over the shape of the fitted model. Most of the models listed above contain two parameters which in some sense control the location and scale of the variable, though often one or other will influence the shape of the distribution as well. Thus, for example, the gamma model with density

$$f(x,\beta,\gamma) = \Gamma^{-1}(\beta)\gamma^{-\beta}x^{\beta-1}e^{-x/\gamma},\ x>0, \qquad (14)$$

has mean $\beta\gamma$ and variance $\beta\gamma^2$. The shape of the distribution is determined exclusively by β. Obviously we cannot separately select the location, scale and shape with only two parameters.

With increasingly wider applications of computer simulation has come the need to consider more flexible distributions. One natural extension is to consider known families of more generality than the above listed distributions. UniFit II (Vincent and Law 1992), in addition to the above models, includes the Pearson Type V and VI families. Swain, Venkatraman and Wilson (1988) consider use of the Johnson system of distributions. This extends the number of parameters to four in most cases.

An even more comprehensive solution is to use methods which extend the parameterization in an unrestricted way. Hora (1983), Avramidis and Wilson (1989) discuss such techniques. Flanigan-Wagner and Wilson (1993) address this issue and discuss an interesting approach using what they call the Bézier distribution. The advantage of such a technique is that it brings into play the power and intuitive graphic properties of interactive Bézier curve fitting. The disadvantage of such an approach is that it is difficult then to assess the statistical properties of the method, such as significance of the goodness of fit.

Despite advances in such flexible methods, there is nevertheless still scope for considering simple extensions of well known distributions. The reasons are:

(i) Standard distributions are still widely used, so simple variations of them which extend their scope will be convenient and may therefore gain more ready acceptance.

(ii) Simple extensions often have some practical interpretation making their use more meaningful than more elaborate models.

(iii) It will be easier to generate variate values from simple extensions.

We now discuss some extensions of this kind.

4.2 Three Parameter Models

There are several ways that an additional parameter can be incorporated into a standard distribution to give more flexibility for fitting purposes.

One method is to incorporate a power transform. For example the gamma model (14) becomes:

$$f(x,\alpha,\beta,\gamma) = \mid \alpha \mid \Gamma^{-1}(\beta)\gamma^{-\beta}x^{\alpha\beta-1}e^{-x^{\alpha}/\gamma},\ x>0.$$

The statistical properties of this generalized gamma model were considered by Stacy (1962). Variates are easily generated using $Y = X^{1/\alpha}$, where X has the gamma density (14). Another example occurs in certain generalizations of the inverse Gaussian model considered by Jϕrgensen (1982).

Another method is to include a shifted threshold parameter. For example the gamma model (14) becomes:

$$f(x,\alpha,\beta,\gamma) = \Gamma^{-1}(\beta)\gamma^{-\beta}(x-\alpha)^{\beta-1}e^{-(x-\alpha)/\gamma},\quad x>\alpha.$$

There is an extensive literature on the problems of fitting this type of distribution, see for example Smith (1985) and Cheng and Iles (1990) and the references therein. The interesting thing that occurs with this type of model is that it contains a *non-degenerate* two parameter embedded model obtained as certain of the parameters tend to zero. In the gamma model this

occurs if we substitute β and γ by μ and σ using the reparameterization:

$$\beta = [(\mu - \alpha)/\sigma]^2 \quad and \quad \gamma = \sigma^2/(\mu - \alpha).$$

Now if we let $\alpha \to -\infty$, keeping μ and σ constant, we get the normal model with density:

$$f(x, \mu, \sigma) = (2\pi\sigma^2)^{-1/2} \exp[-(x - \mu)^2/(2\sigma^2)]$$

in the limit. This possibility must be allowed for if numerical instability is not to occur in the fitting process.

A third method is to use *tilting*. Tilted distributions are useful in importance sampling methods. Nakayama (1992) describes efficient methods for generating certain exponentially tilted variates.

A final method is to treat one of the parameters as being random with a distribution of its own. An extra parameter can then be introduced via this mixing distribution. Scattered examples of such models occur in the literature. An interesting example is the generalization of the two parameter Weibull to a three parameter Burr distribution (see for example Dubey, 1968). For the gamma model we proceed as follows. Suppose that given Y, X has the conditional gamma distribution with density:

$$f(x, \beta, \gamma/Y) = \Gamma^{-1}(\beta)(Y/\gamma)^{\beta} x^{\beta-1} e^{-xY/\gamma}, \; x > 0.$$

Now let Y be a random variable in its own right, with gamma density

$$g(y, \alpha) = \Gamma^{-1}(\alpha^{-1})\alpha^{-1/\alpha} y^{\frac{1}{\alpha}-1} e^{-y/\alpha}, \; y > 0.$$

Then the unconditional distribution of X is the Beta distribution

$$f(x, \alpha, \beta, \gamma) = \frac{(\alpha\beta/\gamma)^{\beta} x^{\beta-1}}{B(\beta, 1/\alpha)(1 + \alpha\beta x/\gamma)^{\beta+\frac{1}{\alpha}}}.$$

Cheng, Evans and Traylor (1993) review a number of such models. Again the problem of embedded models occurs and this must be allowed for in the fitting process. Generation of variates from this type of mixture model is easy. We generate Y from the mixing distribution, then, given this value of Y we generate from the conditional parent X distribution.

The effect of the mixing parameter is to spread the distribution so that it generally has thicker tails. When extreme values are expected more often than occurs with standard distributions then such models may be useful, especially as the mixing parameter may have a sensible interpretation in the context of the application.

4.3 Multivariate and Correlated Variables

We end this section with some brief comments on the fitting of correlated variables and their generation. These topics are increasingly being studied. However, except for certain families of models, an established methodology has yet to be properly developed. I think there are several reasons why more work is needed before methods of satisfying generality might be thought to exist. Firstly there is the sheer range of processes of practical interest when we extend from the univariate and independent case to situations with multivariate and correlated structures. Fitting methods then tend to depend on graphical and interactive techniques which are too complex or ill defined to be amenable to tractable analysis of their statistical properties. Methods and models like those described by Flanigan-Wagner and Wilson (1993) and by Melamed, Hill and Goldsman (1992) perhaps fall into this category. Finally there is the problem of devising methods of variate generation which possess sufficient generality to have wide applicability.

Areas which have received particular attention where progress has been made include the generation of correlated time-series. Examples of sophisticated generators of this kind include those proposed by Melamed, Hill and Goldsman (1992), by Chen and Schmeiser (1992) and by Song and Hsiao (1993). Such models are complex. More accessible flexible methods should perhaps be based on more general theoretical models. The recent work on characterizing mixture models for time series done by Jalali and Pemberton (1994) might prove a useful basis for such developments.

One notable area where the methodology is more clear is the use of Markov chains to fit and generate correlated series. A good example of use of such a model is given by Keezer, Fenic and Nelson (1992) who also cite various other previous applications of such modelling. A good reference concerning the fitting of Markov chains, and indeed the fitting of general stochastic processes is the book by Basawa and Rao (1980).

Finally it should be mentioned that the modelling of spatial processes, especially with reference to image processing, has received huge interest in recent years. Here the methodology is well established and rich. A good introductory reference is given by Ripley (1988). However as the topic falls rather outside our discrete-event simulation remit, we shall not pursue it further here.

5 THE EFFECT OF ESTIMATING INPUT DISTRIBUTIONS

5.1 Bias and Variance

We consider three ways of using empirical data in a simulation:

1. Generate variates for use in the simulation by sampling directly from the empirical cdf. This is what has come to be called the 'bootstrap' method in statistics.

2. Carry out some form of smoothing of the empirical cdf, including extension to allow sampling outside the range of observed values. Sample variates from this smoothed distribution for use in the simulation. A number of 'smoothed bootstrap' methods fall into this category.

3. Fit a theoretical parametric model to the data and sample variates from the fitted model for use in the simulation. This is what is fashionably called the 'parametric' bootstrap.

There are two potential sources of error when using an input model that has been fitted to empirical data:
(a) The bias error that occurs through fitting an incorrect model.
(b) The variance error arising from the variability of estimators of parameters even if the model fitted is the correct one.

The bootstrap technique is attractive in that it does not suffer from bias error. It does not however remove variance error as this is a consequence of the finiteness of the empirical data, i.e. the finiteness of l. The attraction of the bootstrap method is that it allows this variability to be accurately gauged by increasing n. The classic technique (see for example Efron and Tibshirani, 1993) translates into the simulation context as follows: Group the n runs into B blocks of r runs each. For the bth block calculate the sample variance, s_b^2, from the y_j of that block . The average of the s_b^2, S^2 say, of the B blocks estimates the variance of the mean of the y_j in each block. The bootstrap claim is that as $B \rightarrow \infty$ this accurately estimates the variance of the *population* block mean.

The weakness of the standard bootstrap is that if the sample size of the empirical data is small then the bootstrap observations take rather a restricted set of values. Moreover the tail behaviour is severely curtailed by the restricted range of empirical values. Barton and Schruben (1993) suggest smoothed versions of the bootstrap using a two step strategy. In the first step a standard bootstrap resample is obtained (alternatively a uniform resample is obtained - this latter suggestion is in effect the method suggested by Rubin, 1981, for a Bayesian resample). In the second step the empirical cdf of this bootstrap sample is smoothed and then used in the simulation to generate variates.

Given that sampling of the smoothed empirical cdf takes place in the second step, and this is in effect a bootstrap strategy, it is not clear that the first step is necessary or even helpful to carryout. In most cases I suspect that the first stage can be omitted altogether. The method becomes the standard smoothed bootstrap method (see Efron 1982). Banks (1989) shows that in many situations this smoothing leads to improved estimator characteristics over the standard bootstrap technique.

An attractive variation is the method suggested by Bratley, Fox and Schrage (1987) for adding an exponential tail to the smoothed empirical cdf. The effect of this method on the bias error has been theoretically analysed for certain queues in interesting work by Shanker and Kelton (1994). They develop a methodology for testing alternative input models which should therefore help to reduce this bias error.

The analysis of bootstrap techniques usually becomes bogged down in rather intractable algebra. Even where progress is possible the resulting formulas tend to be complicated and not easy to interpret. The third technique, the parametric bootstrap, is probably the most widely used method in practice at present. It suffers from bias error more noticeably than the other two methods. However assuming that we have been careful in our choice of model then analysis of the effect of variance error is more readily carried out, and this is what we turn to in the next sub-section.

5.2 Parametric Bootstrap

We assess the effect of estimating θ_0 on the estimation of $\eta(\theta)$. When θ_0 is estimated then the expression (8) for the observations should be written as:

$$y_j(\mathbf{U}_j, \hat{\theta}) = \eta(\hat{\theta}) + e_j(\mathbf{U}_j, \hat{\theta}) \; j = 1, 2, ..., n \qquad (15)$$

where both $\hat{\theta}$ and $\mathbf{U}_j$ are random. We can calculate the variance of the estimate of the response using:

$$\begin{aligned} &Var[\textstyle\sum_{j=1}^{n} y_j(\mathbf{U}_j, \hat{\theta})/n] = \\ &\hat{\theta} Var\{\mathbf{u}_j E[\textstyle\sum_{j=1}^{n} y_j(\mathbf{U}_j, \hat{\theta})/n \mid \hat{\theta}]\} \\ &+\hat{\theta} E\{\mathbf{u}_j Var[\textstyle\sum_{j=1}^{n} y_j(\mathbf{U}_j, \hat{\theta})/n \mid \hat{\theta}]\}. \end{aligned} \qquad (16)$$

As $\hat{\theta}$ and the $\mathbf{U}_j$ are mutually independent, and as $E[e_j(\mathbf{U}_j, \hat{\theta}) \mid \hat{\theta}] = 0$, the first term on the right hand

side reduces to:

$$\hat{\theta}Var\{\mathbf{u}_j E[\sum_{j=1}^{n} y_j(\mathbf{U}_j,\hat{\theta})/n \mid \hat{\theta}]\} = \hat{\theta}Var[\eta(\hat{\theta})]. \quad (17)$$

If we expand $\eta(\hat{\theta})$ as a Taylor series about θ_0:

$$\eta(\hat{\theta}) = \eta(\theta_0) + \eta'(\theta_0)^T \cdot (\hat{\theta} - \theta_0) + O[(\hat{\theta} - \theta_0)^2],$$

where

$$\eta'(\theta_0) = \partial\eta(\theta)/\partial\theta \mid_{\theta_0},$$

and evaluate the variance of this, we have to first order:

$$\hat{\theta}Var[\eta(\hat{\theta})] = \eta'(\theta_0)^T \mathbf{V}(\theta_0)\eta'(\theta_0), \quad (18)$$

where $\mathbf{V}(\theta_0) = \mathbf{i}(\theta_0)^{-1}$ is the Variance-Covariance matrix of $\hat{\theta}$ as previously defined. The second member of (16) involves terms of the form $\mathbf{u}_j Var[y_j(\mathbf{U}_j,\hat{\theta}) \mid \hat{\theta}]$. These can be evaluated by writing $y_j(\mathbf{U}_j,\hat{\theta})$ as $\eta(\hat{\theta}) + e_j(\mathbf{U}_j,\hat{\theta})$ and expanding $e_j(\mathbf{U}_j,\hat{\theta})$ as a Taylor series:

$$e_j(\mathbf{U}_j,\hat{\theta}) = e_j(\mathbf{U}_j,\theta_0) + e_j'(\mathbf{U}_j,\theta_0)^T \cdot (\hat{\theta} - \theta) + O[(\hat{\theta} - \theta_0)^2],$$

where

$$e_j'(\mathbf{U}_j,\theta_0) = \partial e_j(\mathbf{U}_j,\theta)/\partial\theta \mid_{\theta_0}$$

We have to first order

$$\begin{aligned}\mathbf{u}_j Var[y_j(\mathbf{U}_j,\hat{\theta}) \mid \hat{\theta}] &= \mathbf{u}_j Var[e_j(\mathbf{U}_j,\theta_0)] \\ &+ 2\mathbf{u}_j Cov[e_j(\mathbf{U}_j,\theta_0),\ e_j'(\mathbf{U}_j,\theta_0)^T \cdot (\hat{\theta} - \theta_0)] \\ &+ (\hat{\theta} - \theta_0)^T W(\theta_0)(\hat{\theta} - \theta_0).\end{aligned} \quad (19)$$

where $W(\theta_0)$ is the variance-covariance matrix of $e_j'(\mathbf{U}_j,\theta_0)$. Now

$$\mathbf{u}_j Var[e_j(\mathbf{U}_j,\theta_0)] = \sigma^2,$$

$$\hat{\theta}E\{\mathbf{u}_j Cov[e_j(\mathbf{U}_j,\theta_0),\ e_j'(\mathbf{U}_j,\theta_0)^T \cdot (\hat{\theta} - \theta_0)] \mid \hat{\theta}\} = O(l^{-1}),$$

and

$$\hat{\theta}E[(\hat{\theta} - \theta_0)^T W(\theta_0)(\hat{\theta} - \theta_0) \mid \hat{\theta}] = O(l^{-1}).$$

Combining these results we find to first order that (16) becomes:

$$\begin{aligned}Var[\textstyle\sum_{j=1}^{n} y_j(\mathbf{U}_j,\hat{\theta})/n] &= \eta'(\theta_0)^T \mathbf{V}(\theta_0)\eta'(\theta_0) \\ &+ \sigma^2/n \\ &= O(l^{-1}) + O(n^{-1}).\end{aligned} \quad (20)$$

This result shows that to first order the variability resulting from estimating parameters from empirical data can be separated from that arising from the simulation experiment. The result should be interpreted with a little care. Second order terms, viz. those of order $O(l^{-2})$, $O(n^{-2})$ and $O(l^{-1}n^{-1})$, have been omitted. This means that if, say, we have a large number of runs, i.e. n is large compared to l, then the term of order $O(l^{-1})$ will dominate. However the next most important contribution would not be the $O(n^{-1})$ term, but the $O(l^{-2})$ term, which is not shown.

An obvious, but important consequence of the result is that there is little point in making the simulation over exact compared with the quality of the empirical data. In the limit, as $n \to \infty$, $Var(\bar{y}) \downarrow Var[g(\hat{\theta})]$.

The overall variance can be estimated using the formula (20). The variance σ^2 can be estimated from the sample variance of the observed responses (15). The gradient vector $\mathbf{g}(\theta_0) = \eta'(\theta_0)$ can be estimated by making runs in sets of $(k+1)$ with $\theta_1 = \hat{\theta}$ used in the first run but replaced by $\theta_s = \hat{\theta} + \delta\mathbf{e}_{s-1}$, for $s = 2,3,...,k+1$ where $\mathbf{e}_{s-1}$ is the k dimensional vector with zero entries except for unity in the $(s-1)$th component. The small displacement, δ, has to be appropriately chosen. The same uniforms should be used in all the runs of a given set:

$$\begin{aligned}y_j(\mathbf{U}_j,\theta_s) &= \eta(\theta_s) + e_j(\mathbf{U}_j,\theta_s) \\ j &= 1,2,...,n \\ s &= 1,2,...,k+1\end{aligned}$$

The ith component, $g_i(\theta_0)$, of the gradient vector $\mathbf{g}(\theta_0)$ is estimated by:

$$\widehat{g_i(\theta_0)} = \textstyle\sum_{j=1}^{n}[y_j(\mathbf{U}_j,\theta_{i+1}) - y_j(\mathbf{U}_j,\theta_1)]/(n\delta), \quad i = 1,2,..,k.$$

Confidence intervals for $\eta(\theta_0)$ should be based on (20) and not on (11) alone.

REFERENCES

Avramidis, A. and Wilson, J.R. 1989. A flexible method for estimating inverse distributions in simulation experiments. In *Proceedings of the 1989 Winter Simulation Conference* (ed. E.A. McNair, K.J. Musselman and P. Heidelberger), IEEE Piscataway, New Jersey, 428-436.

Banks, D.L. 1989. Improving the Bayesian bootstrap. Unpublished paper. Dept of Pure Mathematics and Mathematical Statistics, Cambridge University.

Barton, R.R. and Schruben, L.W. 1993 Uniform and bootstrap resampling of empirical distributions. In *Proceedings of the 1993 Winter Simulation Conference* (ed. G.W. Evans, M. Mollaghasemi, E.C.

Russell and W.E. Biles), IEEE Piscataway, New Jersey, 503-508.

Basawa, I.V. and Pakasa Rao, B.L.S. 1980. *Statistical Inference for Stochastic Processes.* London: Academic Press.

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

Chen, H. and Schmeiser, B.W. 1992. Simulation of Poisson processes with trigonometric rates. In *Proceedings of the 1992 Winter Simulation Conference* (ed. J.J. Swain, D. Goldsman, R.C. Crain and J.R. Wilson), IEEE, Piscataway, New Jersey, 609-617.

Cheng, R.C.H. 1992. Distribution fitting and random number and random variate generation. In *Proceedings of the 1992 Winter Simulation Conference* (ed. J.J. Swain, D. Goldsman, R.C. Crain and J.R. Wilson), IEEE Piscataway, New Jersey, 74-81.

Cheng, R.C.H. 1993. Selecting input distributions and random variate generation. In *Proceedings of the 1993 Winter Simulation Conference* (ed. G.W. Evans, M. Mollaghasemi, E.C. Russell and W.E. Biles), IEEE Piscataway, New Jersey, .

Cheng, R.C.H. and Iles, T.C. 1990. Embedded models in three-parameter distributions and their estimation. *J. R. Statist. Soc. B*, 52, 135-149.

Cheng, R.C.H., Evans, B.E. and Traylor, L. 1993. Fitting three-parameter mixture models with flexible tail behaviour. MATHS Report 93-2, School of Mathematics, Univ. of Wales, Cardiff.

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

Dubey, S.D. 1968. A compound Weibull distribution. *Naval Res. Logistics Quarterly*, 15, 179-188.

Efron, B. 1982. *The jackknife, the bootstrap and other resampling plans.* Volume 38 of CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM.

Efron, B. and Tibshirani, R.J. 1993 *An Introduction to the Bootstrap.* New York and London: Chapman and Hall.

Flanigan-Wagner, M.A. and Wilson, J.R. 1993. Using univariate Bézier distributions to model simulation input processes. In *Proceedings of the 1993 Winter Simulation Conference* (ed. G.W. Evans, M. Mollaghasemi, E.C. Russell and W.E. Biles), IEEE Piscataway, New Jersey, 365-373.

Hora, S.C. 1983. Estimation of the inverse function for random variate generation. Communications of the ACM, 26 (8) 590-594.

Jalali, A. and Pemberton, J. 1994. Mixture models for time series. To appear in J. of Applied Probability.

Jϕrgensen, B. 1982. Identifiability problems in Hadwiger fertility graduation, *Scand. Actuarial J.*, 103-109.

Keezer, W.S., Fenic, A.P. and Nelson, B.L. 1992. Representation of user transaction processing behavior with a state transition matrix. In *Proceedings of the 1992 Winter Simulation Conference* (ed. J.J. Swain, D. Goldsman, R.C. Crain and J.R. Wilson), IEEE, Piscataway, New Jersey, 1223-1231.

Law, A.M. and Kelton, W.D. 1991. *Simulation Modeling and Analysis 2nd Edition.* 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: Wadsworth and Brooks/Cole.

Melamed, B., Hill, J.R. and Goldsman, D. 1992. The TES methodology: Modeling empirical stationary time series. In *Proceedings of the 1992 Winter Simulation Conference* (ed. J.J. Swain, D. Goldsman, R.C. Crain and J.R. Wilson), IEEE, Piscataway, New Jersey, 135-144.

Nakayama, M.K. 1992. Efficient methods for generating some exponentially tilted random variates. In *Proceedings of the 1992 Winter Simulation Conference* (ed. J.J. Swain, D. Goldsman, R.C. Crain and J.R. Wilson), IEEE Piscataway, New Jersey, 603-608.

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

Ripley, B.D. 1988. *Statistical Inference for Spatial Stochastic Processes.* Cambridge: CUP.

Rubin, D.B. 1981. The Bayesian bootstrap. *Ann. Statist.* 9, 130-134.

Shanker, A. and Kelton, W. D. 1994. Measuring output error due to input error in simulation: analysis of fitted vs. mixed empirical distributions for queues. To appear.

Smith, R.L. 1985. Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, 72, 67-90.

Song, W.T. and Hsiao, L.-C. 1993. Generation of autocorrelated random variables with a specified marginal distribution. In *Proceedings of the 1993 Winter Simulation Conference* (ed. G.W. Evans, M. Mollaghasemi, E.C. Russell and W.E. Biles), IEEE Piscataway, New Jersey, 374-377.

Stacy, E.W. 1962. A generalization of the gamma distribution, *Ann. math. Statist.*, 33, 1187-1192.

Swain, J.J., Venkatraman, S. and Wilson, J.R. 1988. Distribution selection and validation. J. of Statist. Comput. and Simul., 29, 271-297.

Vincent, S.G. and Law, A.M. 1992. UniFit II: Total support for simulation input modeling. In *Proceedings of the 1992 Winter Simulation Conference* (ed. J.J. Swain, D. Goldsman, R.C. Crain and J.R. Wil-

son), IEEE Piscataway, New Jersey, 371-376.

AUTHOR BIOGRAPHY

RUSSELL C. H. CHENG is Professor of Operational Research in the Institute of Mathematics and Statistics at the University of Kent at Canterbury. He has an M.A. and the Diploma in Mathematical Statistics from Cambridge University, England. He obtained his Ph.D. from Bath University. He is General Secretary of the U.K. Simulation Society, a Fellow of the Royal Statistical Society, Member of the Operational Research Society and was a former President of the Cardiff Branch of the Mathematical Association. His research interests include: variance reduction methods and parametric estimation methods. He is an Associate Editor for *Management Science* and for the *ACM Transactions on Modeling and Computer Simulation.*