*Proceedings of the 1995 Winter Simulation Conference*
*ed. C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman*

# BOOTSTRAP METHODS IN COMPUTER SIMULATION EXPERIMENTS

Russell C. H. Cheng

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

## ABSTRACT

We critically review the work that has been done in applying basic, smoothed and parametric bootstrap methods to simulation experiments. We develop a framework to classify bootstrap methods in this context and use it to compare various bootstrap schemes. Most bootstrap methods are hard to analyse theoretically. An exception is the parametric case for which a detailed analysis can be carried out. An interesting result in this case is that, whereas in standard statistical experiments bootstrap samples give only information about the variance of a statistic and not its mean, this turns out not to be so in simulation experiments. Thus parametric bootstrap samples can be advantageously included in estimates of the responses of interest.

## 1 INTRODUCTION

Suppose we wish to assess the variability of a statistic of interest calculated from a random sample. Bootstrapping is a way to do this by numerical computation. It works by (re)sampling, with replacement, from the original sample, then calculating the statistic for each these bootstrap (re)samples. Under certain general conditions, the sample variability of these bootstrap statistics turns out to have nearly the same variability as that of the original statistic, and so can be used to estimate it.

Bootstrap methods have been proposed in many areas of statistical inference. We review the work that has been done in applying basic, smoothed and parametric bootstrap methods in the special case of simulation experiments. This seems a particularly appropriate use, as simulation experiments also use sampling techniques to reconstruct the statistical distributions of quantities of interest. We develop a framework which can be used to classify and consider bootstrap methods in this context.

It is important to realise that, in simulation experiments, bootstrapping can be used in two quite distinct ways.

Firstly, it can be used to assess variability within the simulation experiment itself. Here, input variates, generated from given probability distributions are used in simulation runs to produce response outputs of interest. If we equate this simulation output with the observations obtained in a statistical experiment, then bootstrap sampling of this simulation output can be used to assess its variability. We shall call this the *known input distribution* case.

Secondly, bootstrapping can be used to assess that variability of the simulation output stemming from use of finite-sized real data to estimate or construct the input probability distributions of the simulation model. Here the bootstrap methodology is different from the first case because there are two sources of statistical variation: that due to the variability of the observed true data and that due to the variability in generating input variates used in the simulation experiment itself. We shall call this the *unknown input distribution* case.

We shall consider both cases. In particular we discuss the statistical properties of the bootstrap technique for the second case, showing that it differs significantly from that of the first.

Nelson (1990) points out that bootstrap methods are computationally expensive, thus ruling them out for use in complex simulations. However this not a problem in the known input distribution case, as the bootstrap resampling does not involve any more simulation runs; these runs being the expensive part of the overall process.

We also consider a method of parametric bootstrapping, in the unknown input distribution case, which can be regarded as an integral part of the original simulation scheme. This method of bootstrapping yields information about the mean of the response of interest as well as its variability. This improved efficiency should make it worth considering even in complex simulations.

There is a third possible use of bootstrapping: to validate stochastic input models. Here classical non-parametric bootstrapping cannot be used. However if parametric bootstrapping is used then this provides a ready means of carrying out the validation. Space prevents discussion of this application here. We hope to describe the technique and its extension to goodness-of-fit and to sensitivity analysis elsewhere.

## 2 THE SIMULATION MODEL

We assume a framework that highlights the dependence of the simulation on input variates. We shall assume that the simulation requires $s$ streams of variates. The variates used in one simulation run will be denoted by

$$\xi_i = (\xi_{i1}, \xi_{i2}, ..., \xi_{il_i}) \; i = 1, 2, ..., s. \tag{1}$$

We shall assume that the total number of variates used is $l$, with the $i$th stream using $l_i \simeq \alpha_i l$, where $\alpha_i$ is a fixed proportion of the total. This allows us to think of $l$ as being a generalised run length, and to consider asymptotic results, where the run length becomes large, by letting $l \to \infty$. When referring to all the variates we use the notation:

$$\Xi = (\xi_1, \xi_2, ..., \xi_s).$$

The output of interest from the simulation run, $y$, can be regarded as a function of the $\xi_i$:

$$y = y(\Xi). \tag{2}$$

We consider two possible formulations. The first is where each sample is drawn from a joint distribution with probability increment $DF_i(\xi_i)$, $i = 1, 2, ..., s$. This is the *non-parametric* case. The second is where each sample is assumed to have a joint distribution with probability increment $DF_i(\xi_i, \theta)$, $i = 1, 2, ..., s$ which depends on a vector $\theta = (\theta_1, \theta_2, ..., \theta_p)$ of $p$ parameters. We call this the *parametric* case. The input distributions are known, in the non-parametric case once the $F_i$ are specified, and in the parametric case once both the $F_i$ and the values of the parameters have been specified.

In what follows we shall include $\theta$ in the notation so that we can consider the parametric case explicitly. The non-parametric case is recovered from the parametric case, simply by ignoring the dependence on $\theta$.

It will be convenient to regard the input variates as having been obtained by the transformation

$$\xi_i = \xi_i(\mathbf{u}_i, \theta) \tag{3}$$

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

$$\mathbf{u}_i = (u_{i1}, u_{i2}, ..., u_{il_i}) \; i = 1, 2, ..., s. \tag{4}$$

We assume the uniforms to be independent, but the components of the $\xi_i$ are not necessarily so. The output (2), can then be regarded as a function of the $\mathbf{u}_i$:

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

where

$$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_s). \tag{6}$$

We assume that the purpose of the simulation experiment is to estimate the expected value of $y$, which is a function of $\theta$ only:

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

We consider the overall simulation experiment as being made up of $r$ runs. The responses or outputs from these runs will be written as:

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

These outputs depend on the parameter vector $\theta$. The "error" variable $e_j$ is the random difference between the $j$th simulation run output and $\eta(\theta)$. We shall assume $E(e_j) = 0$ and $Var(e_j) = \tau^2/l$ for $j = 1, 2, ..., r$. Note that we emphasize the dependence of $Var(e_j)$ on the run length $l$.

Assuming that $\theta$ is fixed for the moment, we have

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

and the mean of the outputs

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

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

$$Var[\bar{y}] = \tau^2/rl. \tag{11}$$

Note that this is the total variance of the response only when the input distributions are completely known. In this case there is no problem in estimating this variance. We can use the sample variance of the $y_j$'s, which will be denoted by $s^2$.

The more interesting situation is when the $F_i$ are not completely known. We assume, in this case, that there is available a sample of empirical data for each stream:

$$\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{in_i}) \; i = 1, 2, ..., s. \tag{12}$$

As in the case of the $l_i$, we assume that the total number of observations is $n$, and that $n_i = \beta_i n$, is a fixed proportion, $\beta_i$, of the total. This allows us to consider asymptotic results as sample size becomes large simply by letting $n \to \infty$. We write

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_s). \quad (13)$$

We shall consider how use of these values to estimate $F_i$ in the non-parametric case, or $F_i$ and $\theta$ in the parametric case, affects the variability of the response, and in particular how then to estimate $Var[\bar{y}]$.

## 3 A POSSIBLE CONFUSION IN TERMINOLOGY

There is the possibility of confusion in terminology when considering bootstrapping in simulation experiments. In classical statistical experiments, bootstrapping is a second stage technique. In the first stage a statistic, which we can view as being the response of interest, is calculated from data obtained from a statistical experiment. Bootstrapping is then used, as an alternative to standard statistical analysis, to assess the variability of the response. Placed within this framework, the simulation runs of a simulation experiment constitute the first stage; the simulation output being the response of interest. It just happens that this first stage is a bootstrap method in its own right, because the runs are based on sampling of input variates, and this is a bootstrapping technique. If we therefore apply bootstrapping in the second stage to assess the variability of the simulation response then this could be viewed as a double use of the bootstrapping technique.

Cheng (1994) discusses how the simulation experiment itself can be regarded as a bootstrap method. To avoid confusion we shall not make explicit use of this interpretation but reserve bootstrapping terminology only for the second stage process of estimating the variability of the response output.

## 4 THE STANDARD BOOTSTRAP

The standard bootstrapping technique works as follows. Let

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_s) \quad (14)$$

be a (multivariate) sample of actual data, the $\mathbf{x}_i$ having the form (12). From this a statistic (response) of interest, $t(\mathbf{X})$, is calculated. We wish to estimate the variability of $t(\mathbf{X})$. The bootstrap method for doing this is as follows. Form $b$ bootstrap samples. Each bootstrap sample is obtained by sampling, with replacement, $l_i$ values from the original sample $\mathbf{x}_i$, $i = 1, 2, ..., s$

$$\mathbf{X}_k^* = (\mathbf{x}_{k1}^*, \mathbf{x}_{k2}^*, ..., \mathbf{x}_{ks}^*) \quad k = 1, 2, ..., b. \quad (15)$$

Calculate the bootstrap response, $t^*(\mathbf{X}_k^*)$, for each bootstrap sample, and obtain the sample variance of these bootstrap responses. This sample variance estimates the variance of $t(\mathbf{X})$. The process is illustrated in Figure 1.

$$\begin{array}{ccccccc}
\mathbf{X} & & \longrightarrow & & t(\mathbf{X}) & & \\
 & \searrow & & & & & \\
 & & \mathbf{X}_1^* & \to & t(\mathbf{X}_1^*) & & \\
 & \searrow & & \cdots & & \searrow & \\
 & & \mathbf{X}_k^* & \to & t(\mathbf{X}_k^*) & \to & s_b^2 \\
 & \searrow & & \cdots & & \nearrow & \\
 & & \mathbf{X}_b^* & \to & t(\mathbf{X}_b^*) & &
\end{array}$$

Figure 1: Basic Bootstrap Method

### 4.1 Known Input Distribution Case

In the case where the input distributions are known, we can view the outputs $y_j$, $j = 1, 2, ..., r$ as being the observed sample. Thus we simply let $\mathbf{X} = \mathbf{y} = \mathbf{x}_1$ (as $s = 1$ in this case), and $t(\mathbf{X}) = \bar{y}$. All the techniques of standard bootstrapping are now available in this situation; see Efron and Tibshirani (1993) for an introduction; see also Hinkley (1988); DiCiccio and Romano (1988).

Applications involving the construction of bootstrap confidence intervals using this structure are discussed by Shiue et al. (1993), who investigate their properties through simulation studies.

An interesting variation of this case is also considered by Shiue et al., where the $y_j$, $j = 1, 2, ..., r$ are taken to be bivariate output from regenerative cycles in the operation of a queue.

Other variations are possible. Kim et al. (1993b) consider the case where the output of the simulation is a time-series, obtained from a single run. Whereas Shiue et al. ensure independence of the $y_j$ by using regenerative cycles, Kim et al. achieve this by clipping the time-series into binary form and then taking consecutive runs of zeroes and ones as being the (now independent) $y_j$. The $y_j$ can then be bootstrap (re)sampled. Additional methods for handling such autocorrelated timeseries are considered by Kim et al. (1993a). A good review of these and other nonparametric methods, as they apply to simulation, is given by Yücesan (1994).

### 4.2 Unknown Input Distribution Case

Observe that we can apply the basic bootstrap method in a simulation experiment where the raw data $\mathbf{X}$ is unaltered, and is to be used directly as the input to produce the response, $y(\mathbf{X}) = \mathbf{t}(\mathbf{X})$. This is represented by the top line in Figure 1: $\mathbf{X} \rightarrow t(\mathbf{X})$. The other lines, $\mathbf{X}_k^* \rightarrow t(\mathbf{X}_k^*)$, $k = 1, 2, ..., b$, represent the $b$ simulation runs using bootstrap samples $\mathbf{X}_k^*$ obtained directly by resampling the original data, $\mathbf{X}$, and then using the resulting bootstrap as input to the simulation which then produces the response $y(\mathbf{X}_k^*) = t(\mathbf{X}_k^*)$. This is the simplest example of the non-parametric bootstrap applied to estimate the variation due to using finite sample-size data to construct input distributions; $\mathbf{X}$ and $\Xi$ being one and the same in this case.

## 5 SMOOTHED BOOTSTRAPPING

In the standard bootstrap, the resampled observations are restricted to the values observed in the original sample. When the original sample size is small, the tail behaviour may be particularly unrepresentative. To try to overcome such difficulties, methods have been suggested for smoothing the empirical distribution function (EDF). The standard method of smoothing is depicted in Figure 2. The only difference from the scheme of Figure 1 is that a smoothed EDF, $\tilde{F}_i$, is first constructed from each sample $\mathbf{x}_i$, for $i = 1, ..., s$. We write $\tilde{\mathbf{F}} = (\tilde{F}_1, ..., \tilde{F}_s$.) These are then sampled to give bootstrap samples, $\mathbf{X}_k^*$, of the same size as $\mathbf{X}$, for use as inputs to the simulations.

$$
\begin{array}{ccccccc}
\mathbf{X} & \longrightarrow & & y(\mathbf{X}) & & & \\
\downarrow & & & & & & \\
\tilde{\mathbf{F}}(.) & & & & & & \\
& \searrow & & & & & \\
& & \mathbf{X}_1^* & \rightarrow & y^*(\mathbf{X}_1^*) & & \\
& \searrow & \cdots & & & \searrow & \\
& & \mathbf{X}_k^* & \rightarrow & y^*(\mathbf{X}_k^*) & \rightarrow & s_b^2 \\
& \searrow & \cdots & & & \nearrow & \\
& & \mathbf{X}_b^* & \rightarrow & y^*(\mathbf{X}_b^*) & &
\end{array}
$$

Figure 2: Basic Smoothed Bootstrap Method

### 5.1 Known Input Distribution Case

This is basically similar to the basic bootstrap case depicted in Figure 2, except that smoothing is done prior to resampling. We again view the outputs $y_j$, $j = 1, 2, ..., r$ as being the observed sample, and let $\mathbf{X} = \mathbf{y} = \mathbf{x}_1$, and $t(\mathbf{X}) = \bar{y}$. The techniques of standard smoothed bootstrapping are now available in this situation; see Efron (1982); Silverman and Young (1987); Banks (1989); Young (1990).

### 5.2 Unknown Input Distribution Case

This is an interesting case. A smoothing method tailored for the simulation context has been given by Barton and Schruben (1993), and a slight generalization of it is depicted in Figure 3.

The empirical real data, $\mathbf{X}$, is as in (14). The original simulation experiment is depicted in Figure 3 by the top line: $\mathbf{X} \rightarrow \tilde{\mathbf{F}}(.) \rightarrow \Xi_j \rightarrow y_j(\Xi_j) \rightarrow \bar{y}$. Here, as before, $\tilde{\mathbf{F}} = (\tilde{F}_1, ..., \tilde{F}_s)$ denotes smoothed EDF's constructed from each sample $\mathbf{x}_i$, $i = 1, ..., s$. The experiment comprises $r$ runs. In the $j$th run, the input variates, $\xi_{ij}$, $i = 1, ..., s$, used in the run are sampled from these smoothed EDFs; $y_j(\Xi_j)$, $j = 1, ..., s$ are the outputs from the runs, and $\bar{y}$ their mean. In the bootstrap versions a bootstrap sample, $\mathbf{X}_k^*$, is formed from $\mathbf{X}$. The bootstrap simulations: $\mathbf{X}_k^* \rightarrow \tilde{\mathbf{F}}_k^*(.) \rightarrow \Xi_{jk}^* \rightarrow y(\Xi_{jk}^*) \rightarrow \bar{y}_k$ precisely mimic the original simulation experiment.

$$
\begin{array}{lcccccccccc}
\mathbf{X} \rightarrow \tilde{\mathbf{F}}(.) \rightarrow \Xi_j \rightarrow y_j(\Xi_j) \rightarrow \bar{y} \quad j = 1, ..., r & & & & \\
\searrow & & & & \\
\quad \mathbf{X}_1^* \rightarrow \tilde{\mathbf{F}}_1^*(.) \rightarrow \Xi_{j1}^* \rightarrow y(\Xi_{j1}^*) \rightarrow \bar{y}_1 & & & & \\
\searrow \qquad\qquad\qquad\qquad\qquad\qquad\qquad \searrow & & & & \\
\quad \mathbf{X}_k^* \rightarrow \tilde{\mathbf{F}}_k^*(.) \rightarrow \Xi_{jk}^* \rightarrow y(\Xi_{jk}^*) \rightarrow \bar{y}_k \rightarrow s_b^2 & & & & \\
\searrow \qquad\qquad\qquad\qquad\qquad\qquad\qquad \nearrow & & & & \\
\quad \mathbf{X}_b^* \rightarrow \tilde{\mathbf{F}}_b^*(.) \rightarrow \Xi_{jb}^* \rightarrow y(\Xi_{jb}^*) \rightarrow \bar{y}_b & & & &
\end{array}
$$

Figure 3: Smoothed Simulation Bootstrap Method

Barton and Schruben (1993) also consider an alternative where a uniform resample is obtained - this suggestion is in effect the method suggested by Rubin, 1981, for a Bayesian resample.

A refinement to this method has been suggested by Bratley, Fox and Schrage (1987) who suggest adding an exponential tail to the smoothed empirical cdf. The bias error introduced by this tail has been theoretically analysed for certain queues by Shanker and Kelton (1994).

There is an important difference between the schemes of Figures 2 and 3. An additional element of random error is introduced with the smoothed bootstrap of Figure 3 through the sampling step, $\tilde{\mathbf{F}}(.) \rightarrow \Xi_j$ in the original simulation and hence in the bootstrap version, $\tilde{\mathbf{F}}_k^*(.) \rightarrow \Xi_{jk}^*$. Analysis as to how this additional variation must be taken into account seems difficult however. Barton and Schruben (1993) investigate this point through some simulation studies.

## 6 THE PARAMETRIC BOOTSTRAP

The intractability of the multinomial distribution of bootstrap samples, has meant that most authors have resorted to simulation studies to report their properties. We consider now the parametric bootstrap. This is a method that has received less attention in the statistical literature. However in the simulation context it seems a worthy contender, as most simulations use parametric input models. Moreover it proves possible to give a fairly detailed analysis of the properties of the method. The case where the parameters are known is rather elementary. We consider therefore only the case of inknown distributions.

### 6.1 Unknown Input Distribution Case

Suppose the original samples, $\mathbf{x}_i$, (14), are assumed to depend on a set of unknown parameters but are otherwise supposed known. The $\mathbf{x}_i$ are used to fit the parameters; the simulation then uses input variates sampled from the fitted distributions. The method is depicted in Figure 4. It will be seen that the structure is like that of the smoothed bootstrap method except that fitted distributions, rather than smoothed distributions, are used to generate the input variates.

(This scheme is not the only possible. One might obtain bootstrap resamples, $\mathbf{X}_k^*$, directly from $\mathbf{X}$, rather than sample from the fitted distribution, $\mathbf{F}(.,\hat{\theta})$. But we run into the difficulty of a scarcely tractable multinomial resampling distribution. We do not consider this version further here.)

$$\mathbf{X}$$
$$\downarrow$$
$$\hat{\theta}$$
$$\downarrow$$
$$\mathbf{F}(.,\hat{\theta}) \to \Xi_j \to y_j(\Xi_j) \to \bar{y} \quad j = 1,...,r$$
$$\searrow$$
$$\mathbf{X}_1^* \to \hat{\theta}_1^*, \mathbf{F}(.,\hat{\theta}_1^*) \to \Xi_{j1}^* \to y_{j1}(\Xi_{j1}^*) \to \bar{y}_1$$
$$\searrow \qquad \searrow$$
$$\mathbf{X}_k^* \to \hat{\theta}_k^*, \mathbf{F}(.,\hat{\theta}_k^*) \to \Xi_{jk}^* \to y_{jk}(\Xi_{jk}^*) \to \bar{y}_k \to s_b^2$$
$$\searrow \qquad \nearrow$$
$$\mathbf{X}_b^* \to \hat{\theta}_b^*, \mathbf{F}(.,\hat{\theta}_b^*) \to \Xi_{jb}^* \to y_{jb}(\Xi_{jb}^*) \to \bar{y}_b$$

Figure 4: Parametric Simulation Bootstrap Method

As in the smoothed case of Figure 3 the additional variation in $y$ introduced through sampling of input variates must be taken into account. It proves much easier to carry out the analysis in the scheme depicted in Figure 4. We concentrate on this case in the remainder of the paper.

## 7 VARIANCE ERROR

We consider how variability of the simulation output is affected by the use of sampled data in the parametric scheme of Figure 4. Specifically, in this case, we wish to assess the effect on the variance of $\bar{y}$ of estimating $\theta_0$, the unknown true parameter value.

Consider first the basic simulation experiment depicted in Figure 4 by the top line: $\mathbf{X} \to \hat{\theta} \to \mathbf{F}(.,\hat{\theta}) \to \Xi_j \to y_j(\Xi_j) \to \bar{y}$, $j = 1,...,r$. When $\theta_0$ is estimated then the expression (8) for the observations should be written as:

$$y_j(\Xi_j) \equiv y_j(\mathbf{U}_j,\hat{\theta}) = \eta(\hat{\theta}) + e_j(\mathbf{U}_j,\hat{\theta}) \qquad j = 1,2,...,r \tag{16}$$

where both $\hat{\theta}$ and $\mathbf{U}_j$ are random. The variance of the estimate of the expected response can be written as

$$\begin{aligned} Var[\textstyle\sum_{j=1}^r y_j(\mathbf{U}_j,\hat{\theta})/r] = \\ \underset{\hat{\theta}}{Var}\{\underset{\mathbf{U}_j}{E}[\textstyle\sum_{j=1}^r y_j(\mathbf{U}_j,\hat{\theta})/r \mid \hat{\theta}]\} \\ +\underset{\hat{\theta}}{E}\{\underset{\mathbf{U}_j}{Var}[\textstyle\sum_{j=1}^r y_j(\mathbf{U}_j,\hat{\theta})/r \mid \hat{\theta}]\}. \end{aligned} \tag{17}$$

Cheng (1994) shows that, to first order, (17) reduces to

$$\begin{aligned} Var[\textstyle\sum_{j=1}^r y_j(\mathbf{U}_j,\hat{\theta})/r] &= \eta'(\theta_0)^T\mathbf{V}(\theta_0)\eta'(\theta_0) \\ &\quad +\tau^2/rl \\ &= \sigma^2/n + \tau^2/rl. \end{aligned} \tag{18}$$

where $\mathbf{V}(\theta_0)$ is the covariance matrix of the estimates of the $\theta_i$, and

$$\eta'(\theta_0) = \partial\eta(\theta)/\partial\theta \mid_{\theta_0} . \tag{19}$$

The first term on the right in (18) is simply the variability resulting from estimating parameters from empirical data, whilst the second term is the variability arising from the simulation experiment.

Confidence intervals for $\eta(\theta_0)$ can be constructed by directly estimating the variance terms in (18). The difficult term is $\sigma^2/n$. Cheng (1994) discusses how the gradients (19) can be estimated, by perturbing the parameter estimates one at a time. This perturbation method is expensive if the number of parameters is large. A conservative estimate is obtained by a simultaneous perturbation. Such a method has been considered by Cheng and Holland (1995b). The bootstrap method of estimating $\sigma^2/n$ is given in Figure 4. This shows the way the bootstrap experiment is constructed: $\mathbf{F}(.,\hat{\theta}) \to \mathbf{X}_k^* \to \hat{\theta}_k^*, \mathbf{F}(.,\hat{\theta}_k^*) \to \Xi_{jk}^* \to y_{jk}(\Xi_{jk}^*) \to \bar{y}_k$. The fitted cdf $\mathbf{F}(.,\hat{\theta})$ is used in each bootstrap experiment to generate a bootstrap sample,

$\mathbf{X}_k^*$, of the same size as the original $\mathbf{X}$. The bootstrap sample is then used to estimate $\theta$, producing a bootstrap estimated cdf, $\mathbf{F}(.,\hat{\theta}_k^*)$, which is then used to generate input variates for the bootstrap simulation. Each bootstrap experiment comprises $r^*$ runs, which may be different from the number of runs, $r$, in the basic experiment. We let the length of the bootstrap runs be $l^*$, which may be different from $l$, the length of the basic runs. The observations can be written as

$$y_{jk}(\Xi_{jk}^*) \equiv y_{jk}(\mathbf{U}_{jk},\hat{\theta}_k) = \eta(\hat{\theta}_k)+e_{jk}(\mathbf{U}_{jk},\hat{\theta}_k) \quad j=1,2,...,r^* \tag{20}$$

Each bootstrap experiment yields a bootstrap mean

$$\bar{y}_k = \sum_{j=1}^{r^*} y_{jk}/r^*.$$

We denote the average of these means by

$$\breve{y} = b^{-1}\sum_{k=1}^{b} \bar{y}_k.$$

Cheng and Holland (1995a) show that, to first order, this has conditional variance:

$$Var[\breve{y} \mid \hat{\theta}] = (\sigma^2/n + \tau^2/r^*l^*)/b. \tag{21}$$

This quantity is estimated by $s_b^2$, the sample variance of the bootstrap means, $\bar{y}_k$, $k = 1,...,b$.

The simulation variance, $\tau^2/r^*l^*$, can be estimated from any of the $b$ bootstrap experiments, or from the original experiment (which estimates $\tau^2/rl$). Thus we can obtain an estimate of $\sigma^2/n$.

An interesting and important feature of the scheme in Figure 4 is that we have *two possible estimators of* $\eta(\theta_0)$*:* $\bar{y}$ *and* $\breve{y}$. In standard statistical applications, the variance of $\bar{y}$ cannot be reduced by combining with the bootstrap $\bar{y}_k$'s as $\bar{y}$ is already sufficient for $\theta$. But the simulation experiment stage injects a variability into the estimation of $\eta(\theta_0)$ that is not present in the standard statistical case.

Cheng and Holland (1995a) show that the unconditional variance of $\breve{y}$ is to first order:

$$Var[\breve{y}] = \sigma^2/n + (\sigma^2/n + \tau^2/r^*l^*)/b. \tag{22}$$

and that the variance of the estimator:

$$\tilde{y} = \alpha\bar{y} + (1-\alpha)\breve{y}$$

is minimized when

$$\alpha = \alpha_{\min} \equiv \frac{\sigma^2/n + \tau^2/r^*l^*}{b\tau^2/rl + \sigma^2/n + \tau^2/r^*l^*}.$$

The minimized value

$$Var_{\min}(\tilde{y}) = \sigma^2/n + \alpha_{\min}\tau^2/rl \tag{23}$$

is smaller than either $Var(\bar{y})$ or $Var(\breve{y})$.

These results allow us optimally to allocate computing time, if this is scarce, between the original experimental runs and the bootstrap runs. We deduce, from the form of (23), that it is best to allocate only sufficient time to achieve a prescribed accuracy in estimating $\sigma^2/n$ from the bootstrap runs. The remainder of the effort should be concentrated in the original experiment to keep the variability of the estimate of the response obtained from these runs as small as possible.

## 8 CONCLUSIONS

It should be realised that bootstrapping methods are an alternative to, rather than a replacement for, more standard statistical procedures. Nevertheless in simulation experiments, as in standard statistical experiments, there are situations where bootstrapping seems an attractive option. In particular the parametric bootstrap is especially amenable to analysis and seems to have the potential to provide a useful tool for estimating how the finite sample size of real-data affects the variability of the output of simulation experiments.

## REFERENCES

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.

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

Cheng, R.C.H. 1994. Selecting input models. In *Proceedings of the 1994 Winter Simulation Conference* (ed. J.D. Tew, S. Manivannan, D.A. Sadowski and A.F. Seila), IEEE Piscataway, New Jersey, 184-191.

Cheng, R.C.H. and Holland, W. 1995a. The effect of input parameters on the variability of simulation output. In *Proceedings of UKSS'95, The Second Conference of the U.K. Simulation Society* (ed. R.C.H. Cheng and R.J. Pooley), UK Simulation Society, EUCS Reprographics, Edinburgh Univ., 29-36.

Cheng, R.C.H. and Holland, W. 1995b. The sensitivity of computer simulation experiments to errors in input data. Accepted for SAMO95 International Symposium, Sept. 1995, Belgirate, Italy.

DiCiccio, T.J. and Romano, J.P. 1988. A review of bootstrap confidence intervals. *J. R. Statist. Soc.* B, **50**, 338-354.

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.

Hinkley D.J. 1988. Bootstrap methods. *J. R. Statist. Soc.* B, **50**, 321-337.

Kim, Y. B., Haddock, J. and Willemain, T. R. 1993a. The binary bootstrap: inference with correlated data. *Commun. Statist.-Simula.*, **22**, 205-216.

Kim, Y.B., Willemain, T.R., Haddock, J. and Runger, G.C. 1993b. The threshold bootstrap: a new approach to simulation output analysis. 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, 498-502.

Nelson, B. 1990. Control-variate remedies, *Opns Res.* **38,** 974-992.

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

Shiue, W.-K., Xu, C.-W. and Rea, C.B. 1993. Bootstrap confidence intervals for simulation outputs. *J. Statist. Comput. Simul.*, **45**, 249-255.

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.

Silverman, B.W. and Young, G.A. 1987. The bootstrap: to smooth or not to smooth? *Biometrika*, **74**, 469-479.

Young, G.A. 1990. Alternative smoothed bootstraps. *J.R. Statist. Soc.* B, **52**, 477-484.

Yücesan, E. 1994. Nonparametric statistics in simulation analysis: a tutorial. (ed. J.D. Tew, S. Manivannan, D.A. Sadowski and A.F. Seila), IEEE Piscataway, New Jersey, 99-105.

## 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.*