ANALYSIS OF SIMULATION OUTPUT BY RESAMPLING

R.C.H. CHENG

Faculty of Mathematical Studies
University of Southampton
Southampton SO17 1BJ

Abstract: Bootstrap resampling is an extremely practical and effective way of studying the distributional properties of simulation output when this is subject to random variation. The purpose of this paper is to show that many problems of sensitivity analysis (SA) and validation can be very simply handled by this method. We consider classical non-parametric version of the bootstrap and show that it can be applied in a natural way in simulation studies without becoming immersed in complicated statistical methodology. The main message of the paper is that the bootstrap method should be the statistical method of first resort in SA and validation studies.

Keywords: Bootstrapping, sensitivity analysis, validation, metamodels

1 INTRODUCTION

Bootstrapping is a well-understood method for studying the distribution of statistical variables by resampling [Davison and Hinkley 1997, Chernick 1999]. This latter reference, with its 84 *page* bibliography, indicates the huge interest in the topic. Thus the common initial reaction of many non-experts that it is a suspect method where one tries to get 'something for nothing', is quite unfounded. It is true that for finite sample sizes an approximation is involved; however as has been amply demonstrated both theoretically and in practice, the method is often especially effective when sample sizes are small, when other well-used and much relied on methods have properties that are dubious [Hall, 1992].

The purpose of this article is to point out the generality of application of bootstrapping by considering, in a unified way, how it can be applied to a range of problems involving sensitivity analysis (SA) and validation of simulation output.

To structure the discussion we shall think of a simulation experiment as being made up of a number of *runs*. Each simulation run produces an output that is a vector function:

$$\mathbf{y} = \{\mathbf{y}(t), 0<t<T\}.$$

In discrete event simulation this output is not deterministic but is a *stochastic process*. It may be of interest to study this stochastic process in its own right. However we shall refine our problem to the study of statistics

$$\mathbf{q} = \mathbf{q}(\mathbf{y})$$

calculated from **y.** A typical example is where **q** is an average value

$$\mathbf{q} = \frac{1}{T}\Xi\mathbf{y}(t)dt.$$

Technically **q** is thus a *vector functional* of **y**, but we need not concern ourselves with this; all we need to note is that **q** is a *random variable* so that its study reduces to considering its distributional properties.

In the next section we consider in more detail the ssetting typically encountered in a simulation model. Section 3 describes the bootstrap method in general terms, whilst in Section 4 it is applied to two basic problems in sensitivity analysis and in validation. Section 5 discusses two, more advanced, problems covering (i) local behaviour of the statistic of interest, and (ii) optimisation of performance measures. Section 6 contains some concluding remarks.

2 STRUCTURE OF SIMULATION OUTPUT

We follow the simulation setting adopted in [Cheng 1995]. However our formulation here is somewhat more general, allowing the simulation output to depend on a larger set of possible factors.

We suppose that there are n simulation runs yielding the observations:

$$\mathbf{q}_j = \mathbf{q}(\mathbf{u}_j, \mathbf{v}_j, \boldsymbol{\theta}(\mathbf{w}), \mathbf{x}_j)\ j = 1, 2, ..., n$$

In our formulation we have allowed for **q** to depend on a number of quantities that we now explain.

Firstly $\mathbf{u}_j$ denotes the stream of uniformly distributed U(0,1) random numbers used in the jth run. Typically the uniforms are not used directly, but are transformed into random variables drawn from distributions other than the uniform. However such transforms are allowed for in supposing that **q** is a function of the **u** so the transforms do not have to be explicitly displayed.

Next comes **v.** This represents a sequence of inputs that are random, but that has been generated independently of the simulation model. A typical instance is where **v** comprises sampled real observations taken from some real system, separate from the system being modelled, but on which **q** depends. Such a sampled real process is sometimes called a *trace*. An example is where **v** is a set of observed true daily temperatures in the simulation of daily gas demand over a given period of time. It could be that **v** is itself a stochastic process, i.e. $\mathbf{v} = \{\mathbf{v}(t), 0<t<T\}$. We shall not consider this general form here but simply think of **v** as being just a sample of observations.

In addition there may be further quantities which may affect **q**. These are denoted by **x** and **θ.** There is no essential difference between **x** and **θ** in the way that they influence **q**. They are simply variables on which **q** depends. However we make a distinction in supposing that **x** comprises those quantities whose values are known exactly in the simulation run; **x** thus includes *decision variables*, i.e. quantities that can be selected by the simulator. However we also include in **x** parameters whose values are known, and that might affect **q,** but which are *not* selectable by the simulator. The reason for adopting this convention is that it then allows us to assume **θ** to be those parameter values, whose values are *not* known, and so have to be estimated. We will therefore assume that in addition to **v**, there exists input data **w** that is used exclusively to estimate **θ.** The **θ** can in principle be allowed to depend on the run j, but there is no loss of generality in assuming that it is the same **θ** in all runs.

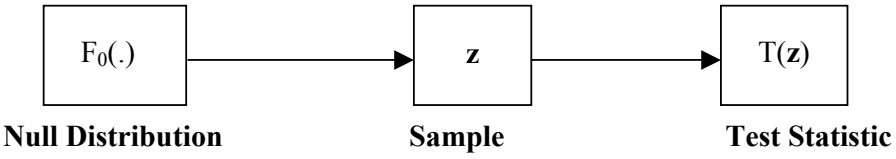
A simple example is a simulation model of a multiserver queue, where **y**(t) is the queue length over a given period, **v** might be a previously observed set of interarrival times, **x** might be the (scalar) number of servers and **θ** the service rates of servers. Here we treat **x** as being selectable by the simulator so that it is a design variable, **θ** may not be known and have to be estimated from available sampled real service times, **w.**

Our study focuses on identifying the distribution of **q**. Before discussing this we describe the basic method that we shall use for doing this.

3. BOOTSTRAP METHOD

In classical statistical inference, many, if not most, problems reduce to studying the distributional properties of a test statistic calculated from a random sample of observations. Mathematical statistics focuses on obtaining such distributions, or, where this is not possible, on obtaining the asymptotic distribution.

Figure 1: Basic Sampling Process



An alternative strategy is possible with the increased power of computers. Bootstrapping is a computer intensive method for determining the distribution of a (test) statistic in a very direct and intuitive way. It is a *resampling* method that operates by sampling from the original data used to calculate the statistic There are many accounts of this, see for example [Efron and Tibshirani, 1993].

The initial way the problem arises is illustrated in Figure 1. This depicts the steps of drawing a sample $\mathbf{z} = (z_1, z_2, ..., z_n)$ of size n from a distribution $F_0(.)$, and then calculating the statistic of interest T from **z**.

If we could repeat this a large number of times, B say, this would give a large sample $\{T_1, T_2,..., T_B\}$ of test statistics, and the sample distribution of these estimates, namely the *empirical distribution function* (EDF) itself estimates the distribution of T.

From now on we shall use the notation $G\{\mathbf{z}\}$ or $G\{z_j\}$, whichever is more convenient, to denote the EDF formed from a sample **z** or from a sample with typical observation z_j.

The fundamental theorem of sampling says that: *If **z** is a random sample of size n, and if n tends to infinity, then G{z} will tend with probability one to $F_0(.)$, the underlying cumulative distribution function (CDF), of the distribution from which the sample is drawn.* Applying this in our case shows that as B tends to infinity then $G\{T_j\}$ will tend to the CDF .of T.

Unfortunately to apply this result requires repeating the basic process of Figure 1 many times. This is often too expensive to do as the step of obtaining **z** from $F_0(.)$ may involve a complicated and time-consuming experimental procedure.

The bootstrap method is based on the idea of replacing $F_0(.)$ by the best estimate we have for it. This is simply $G\{\mathbf{z}\}$**.** Thus we mimic the basic process depicted in Figure 1 but instead of sampling from $F_0(.)$ we sample from $G\{\mathbf{z}\}$. This is exactly the same as sampling with replacement from **z.** We carry out this process B times to get B *bootstrap samples* $\mathbf{z}_1^*$, $\mathbf{z}_2^*$, ..., $\mathbf{z}_B^*$**.** From each of these bootstrap samples we calculate a corresponding *bootstrap test statistic value* $T_j^* = T(\mathbf{z}_j^*,)$ i = 1,2,...,B. The process is depicted in Figure 2

The EDF of the bootstrap sample of T_i^*s, namely $G\{T_i^*\}$, now estimates the distribution of T. This is depicted in Figure 3, where, without loss of generality we have assumed the sample to be reordered so that $T_1^* < T_2^* <...< T_B^*$,

Figure 2: Bootstrap Process:

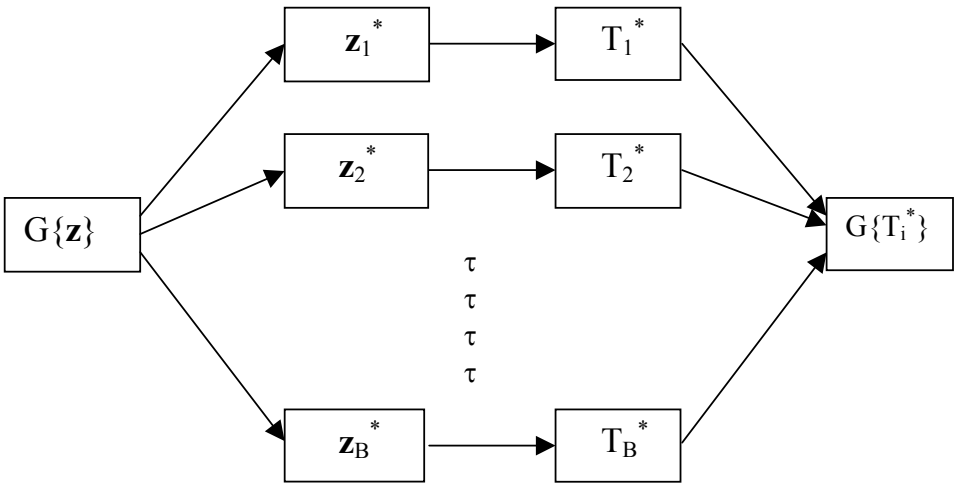
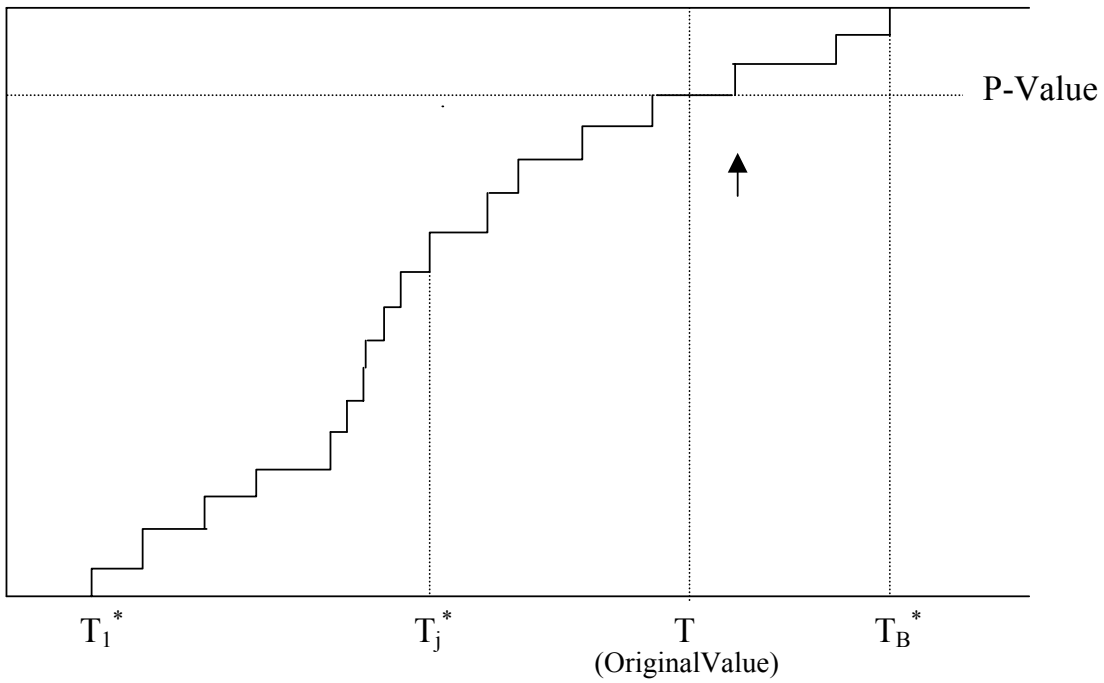


Figure 3: Empirical Distribution Function of the T_i^*



Also included in the Figure is the test statistic value T calculated from the original sample. In the present situation, if the bootstrap process has worked properly, then both T and the bootstrap T_i^* values will have been drawn from near identical distributions. If B is large, T should typically be an unexceptional value, with p-value that is not extreme (as depicted in the Figure).

As will be seen later, there will be situations where T and T_i^* only have near identical distributions under a null hypothesis. The procedure of examining the p-value of T thus provides a simple statistical test of the null hypothesis.

In practice typical values for B are 500 or 1,000.

The version we have described of the bootstrap is the non parametric case where resampling of the original observations is carried out. There is an alternative strategy known as the *parametric bootstrap.* This can be applied where $F_0(.)$ is known apart from parameters $\boldsymbol{\varphi}$ on which it depends**,** i.e.

$$F_0(.) = F_0(.,\boldsymbol{\varphi}).$$

The parameters $\boldsymbol{\varphi}$ can be estimated from **z**. We then have a fitted distribution

$$\hat{F}_0(.) = F_0(., \hat{\varphi}).$$

The parametric bootstrap replaces $G\{\mathbf{z}\}$ in Figure 2 by this fitted distribution.

For more details of different bootstrap methods see [Davison and Hinkley, 1997].

4. ELEMENTARY SA AND VALIDATION

We focus now on the use of bootstrap in simulation experiments. This is a subject which has only received limited study. For some initial discussion of bootstrapping as applied to simulation see [Barton and Schruben, 1993], [Kim et al, 1993], [Shiue et al 1993].

We consider first an elementary problem of Sensitivity Analysis (SA). Suppose that **x** is held

fixed. The basic question of sensitivity analysis is then:

(A) What is the distribution of the random variable

$$\mathbf{q} = \mathbf{q}(\mathbf{u}, \mathbf{v}, \boldsymbol{\theta}(\mathbf{w}), \mathbf{x}),$$

taking into account the probabilistic variation due to **u**, **v** and **w**? A discussion of this problem is given in [Cheng and Holland, 1995a and 1995b].

If we have enough input data we can answer (A) by making n runs

$$\mathbf{q}_j = \mathbf{q}(\mathbf{u}_j, \mathbf{v}_j, \boldsymbol{\theta}(\mathbf{w}_j), \mathbf{x}), \quad j = 1, 2, ..., n$$

If n is sufficiently large then $G\{\mathbf{q}_j\}$ estimates the distribution of **q**.

Here we suppose that we have access to enough input data to be able to use independent sets $\mathbf{v}_j$ (each having the same joint distribution) in each run. Likewise we suppose the $\mathbf{w}_j$ are separate samples, each drawn from the same joint distribution. The $\mathbf{q}_j$ are thus independently and identically distributed and they incorporate the variation due to variation in **v** and **w**.

This formulation is likely to be unsatisfactory on two grounds. Firstly, if, as we have supposed, $\boldsymbol{\theta}$ is unknown and we use **w** simply to estimate $\boldsymbol{\theta}$, then it would be perverse to divide up **w** just to provide separate estimates of $\boldsymbol{\theta}$ in the different runs in order to gauge how the variability in the estimate of $\boldsymbol{\theta}$ influences the distribution of **q.** We would in practice almost certainly use the whole of **w** to estimate $\boldsymbol{\theta}$. The simulation outputs are thus of the form:

$$\mathbf{q}_j = \mathbf{q}(\mathbf{u}_j, \mathbf{v}_j, \boldsymbol{\theta}(\mathbf{w}), \mathbf{x}) \quad j = 1, 2, ..., n$$

and the distribution of **q** as estimated by $G\{\mathbf{q}_i\}$ is conditional on $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathbf{w})$. The same problem arises with **v**. If there are insufficient observations to divide up **v** into parts $\mathbf{v}_j$ for the different runs then we will have to use the same **v** in each run giving:

$$\mathbf{q}_j = \mathbf{q}(\mathbf{u}_j, \mathbf{v}, \boldsymbol{\theta}(\mathbf{w}), \mathbf{x}) \quad j = 1, 2, ..., n$$

and the distribution of **q** is conditional on both $\mathbf{V}=\mathbf{v}$ and $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathbf{w})$.

We can overcome both problems by using bootstrapping on **v** and **w**. In terms of Figure 1, **u,** **v** and **w** play the role of **z**; whilst **q** plays the role of T. The only difference is that we do not bother to bootstrap **u,** as this is easy to generate. This gives

$$\mathbf{q}_j^* = \mathbf{q}(\mathbf{u}_j, \mathbf{v}_j^*, \boldsymbol{\theta}(\mathbf{w}_j^*), \mathbf{x}) \quad j = 1, 2, ..., n$$

The distribution of **q** is now estimated by the bootstrap EDF $G\{\mathbf{q}_j^*\}$.

Next we consider a basic validation problem:

(B) Does the output of the simulation model accurately represent the output of the system being modelled?

A discussion of validation involving trace-driven simulation is described in [Kleijnen et al., 2000]. We adopt a more general approach here.

We suppose that we wish to compare the distribution of the simulation output **q** with its counterpart obtained either from an existing or a hypothetical system. Suppose that $\{\mathbf{q}_j\}$ is a sample of simulation outputs and $\{\mathbf{q}_j^0\}$ is a sample of outputs obtained from an existing system. A natural way to compare the distributions is to use a statistic that measures the difference between $G\{\mathbf{q}_j\}$ and $G\{\mathbf{q}_j^0\}$. A particularly effective goodness-of-fit statistic (less subjective than the usual chi-squared goodness-of-fit, and more powerful than Kolmogorov-Smirnov) is the two sample Cramer - von Mises statistic [cf Anderson, 1962]:

$$W^2 = \Xi[G\{\mathbf{q}_j^0\} \; ? \; G\{\mathbf{q}_j\}]^2 dG\{\mathbf{q}_j\}.$$

The distribution of W^2 is not easy to obtain, especially when parameters have to be estimated.

We can use a bootstrap technique in this case as follows. We obtain B pairs of bootstrap samples $\{\mathbf{q}_j^*\}_{1i}$, $\{\mathbf{q}_j^*\}_{2i}$, i = 1,2,...,B, each sample of the pair being obtained by resampling with replacement from $\{\mathbf{q}_j\}$, with $\{\mathbf{q}_j^*\}_{1i}$ the same sample size as $\{\mathbf{q}_j\}$, and $\{\mathbf{q}_j^*\}_{2i}$ the same sample size as $\{\mathbf{q}_j^0\}$. Denote the EDFs of the samples of each pair by $G_{1i}\{\mathbf{q}_j^*\}$ and $G_{2i}\{\mathbf{q}_j^*\}$ i = 1,2,...,B. We can now calculate B bootstrap Cramer - von Mises statistics from each pair:

$$W_i^{2*} = \Xi[G_{2i}\{\mathbf{q}_j^*\} \; ? \; G_{1i}\{\mathbf{q}_j^*\}]^2 dG_{1i}\{\mathbf{q}_j^*\}.$$

Under the null assumption that $G\{\mathbf{q}_j\}$ and $G\{\mathbf{q}_j^0\}$ have the same distribution it follows that each

W_i^{2*} is just a bootstrap version of W^2. Thus a critical value can be obtained from corresponding p-value of $G\{W_i^{2*}\}$ and the null hypothesis that $G\{\mathbf{q}_j\}$ and $G\{\mathbf{q}_j^0\}$ have the same distribution can be tested simply by checking if the original W^2 exceeds this p-value or not.

Figure 3 illustrates this situation if we take **z** to be **q** and T to be W^2.

5. ADVANCED PROBLEMS IN SA

The situation is more complicated if we allow **q** to depend on **x** as well as on **v** and **θ.** We consider two problems: (i) how strong is this dependence? and (ii) how does this dependence vary as the position of the design point **x** varies?. In either case we shall treat the problem as one of *regression analysis*. We use the *metamodel* approach [Kleijnen and Van Groenendaal, 1992] where we assume that the output from the jth run takes the form

$$\mathbf{q}_j = \eta(\mathbf{x}_j, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}_j$$

Here $\boldsymbol{\eta}(\mathbf{x}_j, \boldsymbol{\beta})$ is the *regression function* or *metamodel.* It is dependent on **x,** the values of the decision variables used in the run. We assume it is also dependent on **β,** a set of unknown parameter values that have to be estimated from the observed output values $\mathbf{q}_j$. The term $\boldsymbol{\varepsilon}_j$ is a random 'error'; this has a distribution that is either of some assumed form, or that will have to be estimated.

In terms of the metamodel, the question concerning the strength of dependence on particular components of **x** can be readily formulated in statistical terms if we restrict ourselves to considering just the local behaviour of **q** for values of decision variables in the neighborhood of some specific operating point $\mathbf{x}_0$. Mathematically we are simply asking the question:

C: Which derivatives $\partial\boldsymbol{\eta}(\mathbf{x}_0)/\partial x_i$ (evaluated at $\mathbf{x}_0$) are large?

Though readily formulated the problem can be hard to solve fully. For example if the number of decision variables, s say, is large, then even the problem of identifying the subset of k largest derivatives can be a difficult problem [Cheng and Holland, 1995a, b].

The second question, concerning how the dependence of **q** varies with **x**, encompasses a number of interesting problems.

(i) A basic problem is where the precise form of η is not known and we suppose that

$$\eta(\mathbf{x}, \beta) = \sum_{k=1}^{K} \beta_k \eta_k(\mathbf{x})$$

is the linear combination of basis functions $\eta_k(\mathbf{x})$, k = 1,2,...,K. The problem is then a statistical one where we have to estimate the coefficients β_k. It may also be necessary to decide on how many terms are needed. This is equivalent to deciding how many of the β_k are non zero and which are sufficiently small to be treated as being zero.

Uncertainty in the number of terms to take can be handled by bootstrapping. We simply take K to be a statistic of interest and to be one of the components of **q**. Thus we have a problem of type (A).

(ii) A more difficult version of the problem arises where it is thought that η is one of a number of possible competing functions, η_k, k = 1,2,...,K, and we take

$$\eta(\mathbf{x}, \beta, \theta) = \sum_{k=1}^{K} \beta_k \eta_k(\mathbf{x}, \theta_k)$$

subject to $\sum_{k=1}^{K} \beta_k = 1$. If one of the η_k is the correct model then this can be obtained by estimating the β_k, when the correct β_k will be close to unity and the others near zero. The complication here is that the η_k each are functions involving unknown parameters θ_k. This turns out to be a much more difficult statistical problem than the previous case as it does not satisfy the usual statistical regularity conditions. [Cheng and Traylor, 1995] have discussed this problem from a statistical standpoint. (See also McLachlan and Basford, 1988].)

Though there are complications, again bootstrapping can be used to establish distributional properties of K in this as in the previous case.

(iii) Estimation of η usually requires that we give an indication of how accurate the estimated function $\hat{\eta}$ is. We may be interested in calculating confidence intervals at particular design points **x**. A typical confidence interval statement at $\mathbf{x}_0$ takes the form: 'The unknown true

value of $\boldsymbol{\eta}(\mathbf{x}_0)$ lies in the interval $\hat{\boldsymbol{\eta}}(\mathbf{x}_0)\pm\boldsymbol{\delta}(\mathbf{x}_0)$ with confidence α'. Alternatively it may be of interest to calculate a confidence band which takes the form: 'The unknown true value of $\boldsymbol{\eta}(\mathbf{x})$ lies in the interval $\hat{\boldsymbol{\eta}}(\mathbf{x})\pm\boldsymbol{\delta}(\mathbf{x})$, for all $\mathbf{a} \le \mathbf{x} \le \mathbf{b}$, with confidence α'. In either case we can use bootstrap instead of more traditional confidence interval calculations.

The first case is really just a type (A) problem if we treat $\hat{\boldsymbol{\eta}}(\mathbf{x}_0)$ as being the **q** of interest.

The second case is more interesting [cf Hall and Pittelkow, 1990]. It can be tackled by bootstrapping if we generate B bootstrap fitted curves $\hat{\boldsymbol{\eta}}_i^*(\mathbf{x})$, for all $\mathbf{a} \le \mathbf{x} \le \mathbf{b}$, i=1,2,...,B. We now select a proportion (1- α) of the B curves, the selection being random, with each curve equally likely to be selected. Suppose, with no loss of generality we label the selected curves as i=1,2,3, ..., S=(1-α)B. Then the *envelope* created by these selected curves, with lower and upper limits:

$\boldsymbol{\eta}_L(\mathbf{x}) = \min_{1\le i\le S} \hat{\boldsymbol{\eta}}_i^*(\mathbf{x})$, for all $\mathbf{a} \le \mathbf{x} \le \mathbf{b}$

and

$\boldsymbol{\eta}_U(\mathbf{x}) = \max_{1\le i\le S} \hat{\boldsymbol{\eta}}_i^*(\mathbf{x})$, for all $\mathbf{a} \le \mathbf{x} \le \mathbf{b}$

is a band that includes at least (1-α) of the fitted curves in their entirety; and so for large B, it will be a conservative (1-α) confidence band for the unknown true curve.

(iv) Variations of the previous problem occur when we consider system optimisation. Suppose that **q** is a scalar performance index. We may then be interested in optimal operating points, that is, values of **x** for which $\eta(\mathbf{x})$ is a minimum or a maximum. Again we may be interested in constructing a confidence interval either for the *optimum value* or a confidence region for the *optimum point*.

In the former instance construction of a confidence interval for the optimal value is readily treated as being a type (A) problem.

Construction of a confidence region for the optimum point can be handled in a similar way to that used for constructing a confidence band. We generate B bootstrap fitted curves $\hat{\boldsymbol{\eta}}_i^*(\mathbf{x})$, for all $\mathbf{a} \le \mathbf{x} \le \mathbf{b}$, i=1,2,...,B. For each we find the minimum point $\mathbf{x}_i^{(min)}$, say. Now select a proportion (1- α) of the points, the selection being random, with each minimum point equally likely to be selected. Suppose, with no loss of generality we label the selected points as i=1,2,3, ..., S=(1-α)B. Then the *convex hull* of these selected points contains (1-α) of these points at least. Assuming smooth behaviour of η, for large B this convex hull will be a (conservative) (1-α) confidence region for the optimum design point.

The problem of obtaining optimum conditions can be posed in a non parametric setting. Here we take the observations as having the form

$$q_j = \eta(\mathbf{x}_j) + \varepsilon_j$$

without supposing $\eta(\mathbf{x})$ has any specific form other than that it has a minimum. Optimisation must therefore be carried out in a non-parametric context. The biggest bugbear is the problem of dimensionality when s, the number of components of **x**, is large. The solution space is then large, and it can be extremely expensive to locate a minimum especially in the presence of measurement uncertainty.

6 CONCLUSIONS

The above examples and methodology hopefully provide an indication of the simplicity and general applicability of the use of bootstrap resampling in simulation experiments. Subject to weak, very generally satisfied conditions, the method can be used to handle the widely encountered situation where we have an output **q** that is a function of a sample **z** and the problem of interest is the determination of the distribution of **q.**

Bootstrap resampling seems a natural accompaniment to simulation experiments. All the sampling routines for the required random variables are bound to be already in place for the simulation itself. It is thus a simple extension to use these for the resampling procedures of the bootstrap. The methodology has much to recommend it and will surely gain much wider acceptance and use as its power and simplicity become recognised.

REFERENCES

Anderson, T.W. 1962, "On the distribution of the two-sample Cramer-von Mises criterion." *Annals of Mathematical Statistics*, **33,** 1148-1159.

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.

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." SAMO95 International Symposium, Sept. 1995, Belgirate, Italy.

Cheng, R.C.H. and Traylor, L. 1995. "Non-regular maximum likelihood problems (with Discussion)", *J.R. Statist. Soc*. *B*, **57**, 3-44.

Cheng, R.C.H. 1995, "Bootstrap methods in computer simulation experiments" In *Proceedings of 1995 Winter Simulation Conference, WSC' 95* 171-177.

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

Hall, P.and Pittelkow, Y.E. 1990, "Simultaneous bootstrap confidence bands in regression." *J. Statist. Comput. Simul.* **37**, 99-113.

Hall, P. 1992, "*The Bootstrap and Edgeworth Expansion.*" Springer-Verlag, New York.

Kim, Y.B., Willemain, T.R., Haddock, J. and Runger, G.C. 1993, "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.

Kleijnen, J.P.C. and Van Groenendaal, W. 1992, "*Simulation: A Statistical Perspective"* Wiley, Chichester, UK.

Kleijnen, J.P.C., Cheng, R.C.H. and Bettonvil, B. 2000**,** "Validation of trace-driven simulation models: more on bootstrap tests" In *Proceedings of the 2000 Winter Simulation Conference*, eds. J.A. Joines, R.R. Barton, K. Kang and P.A. Fishwick.

McLachlan, G.J. and Basford, K.E. 1988, "*Mixture Models: Inference and Applications to Clustering"* Marcel Dekker, New York.

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

BIOGRAPHY

RUSSELL C. H. CHENG is Professor and Head of the Operational Research Group in the Faculty of Mathematical Studies at the University of Southampton. 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 a former Chairman of the U.K. Simulation Society, a Fellow of the Institute of Mathematics, a Fellow of the Royal Statistical Society, and Member of the Operational Research Society. His research interests include simulation experimental design and analysis, and parametric estimation methods. He is Joint Editor of the *IMA Journal on Management Mathematics* and an Editorial Panel member of *Simulation News Europe.*

Email: R.C.H.Cheng@maths.soton.ac.uk
Homepage:
http://www.maths.soton.ac.uk/staff/Cheng/