Proceedings of the 2005 Winter Simulation Conference
M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.

BOOTSTRAPPING SIMULTANEOUS CONFIDENCE BANDS

Russell C.H. Cheng

School of Mathematics
University of Southampton
Southampton, SO17 1NQ, U.K.

ABSTRACT

Once a regression has been fitted to data, it is usually necessary to add confidence intervals to indicate the accuracy of the fitted regression line. This can easily be done for individual explanatory variable values. However sometimes confidence limits are needed simultaneously for the whole range of explanatory variable values of interest. In other words the problem is to construct a confidence band within which the entire unknown true regression line lies with given confidence. This article discusses computer intensive methods for doing this. The advantage of such methods compared with classical asymptotic methods is that accurate coverages can be obtained quite easily using bootstrap resampling.

1 INTRODUCTION

We consider a typical and quite general situation where a regression metamodel is used to represent the output from a simulation study. Barton (1998) gives a good introduction to the metamodelling approach. We suppose that the simulation study comprises a number of independent runs, *n* say, of the simulation model and that the observed output of interest, denoted by *y*, is random but dependent on an independent variable *x*. A convenient representation of this situation is by means of a statistical metamodel. Our approach is much in the spirit of that adopted by Krzanowski (1998). The observations, which we write as a vector $\mathbf{y} = (y_1, y_2, ..., y_n)$, can be written as

$$y_j = \eta(x_j, \boldsymbol{\theta}) + \varepsilon_j, \quad j = 1,2,...,n . \tag{1}$$

Here $\eta(x, \boldsymbol{\theta})$ denotes a deterministic function, called the *regression function*, and $\boldsymbol{\theta}$ is a vector of *p* coefficients, or parameters, on which the function depends. The functional form of $\eta(x, \boldsymbol{\theta})$ is assumed known, but the values of entries in the parameter $\boldsymbol{\theta}$ are assumed unknown. The ε_j represent independent random errors or perturbations, with mean zero. i.e.

$$E(\varepsilon) = 0 .$$

A typical more explicit assumption is that the errors are normally distributed and have constant variance, i.e.

$$\varepsilon \sim N(0, \sigma^2) \tag{2}$$

but we shall not need to assume this necessarily holds.

The condition $E(\varepsilon) = 0$ means that the expected value of *y* at a given *x* is simply

$$E(y|x) = \eta(x, \boldsymbol{\theta}) . \tag{3}$$

We shall assume that the determination of $E(y|x)$ for a given range $a \le x \le b$ is the primary objective of the simulation.

This question is meaningful if we suppose that there is a true, but unknown, value for $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0$ say. The first problem therefore is to estimate this value from the observations **y**.

By far the best general way estimating the parameters $\boldsymbol{\theta}$ is the method of maximum likelihood (ML). The method and the properties of the resulting estimator, $\hat{\boldsymbol{\theta}}$, are well known and are summarised in the next section.

The fitted *y*,

$$\hat{y}(x) = \eta(x, \hat{\boldsymbol{\theta}}), \ a \le x \le b , \tag{4}$$

is the obvious estimate of $E(y|x)$. The main question then is 'How accurate is the estimate $\hat{y}(x) = \eta(x, \hat{\boldsymbol{\theta}})$ of $E(y|x)$?' The classical answer to this question is to use asymptotic theory and the so-called *delta method* to calculate, for any given value of *x*, a confidence interval for $\eta(x, \boldsymbol{\theta}_0)$. The method for doing this is summarized in the next section also.

A more interesting problem is to find a *simultaneous* confidence interval. For given confidence level α, we wish to find a confidence band B with lower and upper limits $y_L(x), y_U(x),\ a \le x \le b$ such that with confidence level $100(1-\alpha)\%$

$$y_L(x) \le \eta(x,\boldsymbol{\theta}_0) \le y_U(x), \text{ for all } a \le x \le b\ .$$

Thus, with given confidence, $\eta(x,\boldsymbol{\theta}_0)$ lies within the given limits *simultaneously* for all $a \le x \le b$. This problem is also well-known, with established solutions. See for example Miller (1981). When $\eta(x,\boldsymbol{\theta})$ is a sufficiently smooth function of $\boldsymbol{\theta}$, in the neighbourhood of $\boldsymbol{\theta}_0$, to admit a Taylor series expansion in $\boldsymbol{\theta}-\boldsymbol{\theta}_0$, then this can be combined with asymptotic theory to give a simple explicit solution to this problem. We outline this method also in the next section.

The confidence band calculations use approximations, particularly the Taylor series approximation for $\eta(x,\boldsymbol{\theta}_0)$, that can result, especially for small n, in the actual coverage being rather less than the nominal stated confidence level. We show how resampling methods can provide a method for overcoming this last difficulty. This is the main purpose of this paper. The resampling methods are discussed in Section 3. A numerical example involving the modelling of a tuberculosis notification rates is discussed in Section 4.

2 MAXIMUM LIKELIHOOD ESTIMATION

2.1 The Maximum Likelihood Method

Suppose $\mathbf{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ is a set of observations where the ith observation, Y_i, is a random variable drawn from the continuous distribution with pdf $f_i(y,\ \boldsymbol{\theta})$ (i = 1, 2, …, n). The subscript i indicates that the distributions of the y_i can all be different. As an example, in the regression situation where the errors are normally distributed then $Y_i \sim N(\eta(x_i,\boldsymbol{\theta}),\sigma^2)$, i = 1, 2, ..., n. The pdf of Y_i is then

$$f_i(y,\boldsymbol{\theta}) = \frac{\exp\{-[y-\eta(x_i,\boldsymbol{\theta})]^2/(2\sigma^2)\}}{\sqrt{2\pi\sigma^2}}\ .$$

Thus $\mathbf{Y}$ is *not* a random sample in this case, because the observations are not all identically distributed. However ML estimation still works in this case.

We now describe the method. Suppose that $\mathbf{y} = \{y_1, y_2, \ldots, y_n\}$ is a sampled value of $\mathbf{Y} = \{Y_1, Y_2, \ldots, Y_n\}$. Then we write down the joint distribution of $\mathbf{Y}$ evaluated at the sampled value $\mathbf{y}$ as:

$$Lik(\boldsymbol{\theta},\mathbf{y}) = f_1(y_1,\boldsymbol{\theta})f_2(y_2,\boldsymbol{\theta})\ldots f_n(y_n,\boldsymbol{\theta})\ .$$

This expression, *treated as a function of* $\boldsymbol{\theta}$**,** is called the *likelihood* (of the sampled value $\mathbf{y}$)**.** The logarithm:

$$L(\boldsymbol{\theta},\mathbf{y}) = \log\{Lik(\boldsymbol{\theta},\mathbf{y})\} = \sum_{j=1}^{n}\log f_j(y_j,\boldsymbol{\theta})$$

is called the *loglikelihood*. The ML estimate, $\hat{\boldsymbol{\theta}}$, is that value of $\boldsymbol{\theta}$ which maximizes the loglikelihood.

The MLE is illustrated in Figure 5 in the one parameter case. In some cases the maximum can be obtained explicitly as the solution of the vector equation

$$\frac{\partial L(\boldsymbol{\theta},\mathbf{y})}{\partial\boldsymbol{\theta}} = \mathbf{0}$$

which identifies the stationary points of the likelihood. The maximum is often obtained at such a stationary point. This equation is called the *likelihood equation*. The MLE illustrated in Figure 1 corresponds to a stationary point.

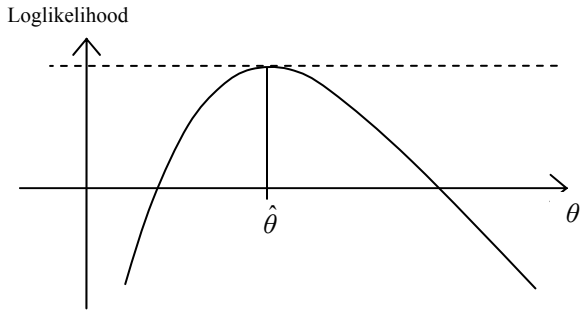


Figure 1: The Maximum Likelihood Estimator $\hat{\boldsymbol{\theta}}$

In certain situations, and this includes some well known standard ones, the likelihood equations can be solved to give the ML estimators explicitly. This is preferable when it can be done. However in general the likelihood equations are not very tractable. Then a much more practical approach is to obtain the maximum using a numerical search method. In the examples given below we have used the *Nelder-Mead* method. This is a flexible method that seems very robust in practice.

2.2 Accuracy of ML Estimators

An important property of the MLE, $\hat{\boldsymbol{\theta}}$, is that its asymptotic probability distribution is known to be normal under very general conditions. In fact it is known that, as the sample size $n \to \infty$,

$$\hat{\boldsymbol{\theta}} \sim N\{\boldsymbol{\theta}_0, \mathbf{V}(\boldsymbol{\theta}_0)\} \qquad (5)$$

where $\boldsymbol{\theta}_0$ is the unknown true parameter value and the variance has the form

$$\mathbf{V}(\boldsymbol{\theta}_0) = [\mathbf{I}(\boldsymbol{\theta}_0)]^{-1},$$

where

$$\mathbf{I}(\boldsymbol{\theta}) = E\left(-\partial^2 \mathbf{L} / \partial \boldsymbol{\theta}^2\right)$$

is called the *information matrix*. Thus the asymptotic variance of $\hat{\boldsymbol{\theta}}$ is the inverse of the information matrix evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Its value cannot be computed precisely as it depends on the unknown $\boldsymbol{\theta}_0$, but it can be approximated by

$$\mathbf{V}(\hat{\boldsymbol{\theta}}) = \left[\mathbf{I}(\hat{\boldsymbol{\theta}})\right]^{-1}.$$

The expectation in the definition of $\mathbf{I}(\boldsymbol{\theta})$ is with respect to the joint distribution of $\mathbf{Y}$ and this expectation can be hard to evaluate. In practice the approximation

$$\mathbf{V}(\hat{\boldsymbol{\theta}}) \cong [\, -\partial^2 \mathbf{L}(\boldsymbol{\theta}, \mathbf{y}) / \partial \boldsymbol{\theta}^2 \big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}]^{-1},$$

where we replace the information matrix by its sample analogue, the latter being called the *observed information*, is quite adequate. Practical experience indicates that it tends to give a better indication of the actual variability of the MLE. Thus the working version of (5) is

$$\hat{\boldsymbol{\theta}} \sim N\{\boldsymbol{\theta}_0, \mathbf{V}(\hat{\boldsymbol{\theta}})\} \tag{6}$$

The *Hessian* of second derivatives of the loglikelihood, $\partial^2 \mathbf{L}(\boldsymbol{\theta}, \mathbf{y}) / \partial \boldsymbol{\theta}^2$, that appears in the expression for $\mathbf{V}(\hat{\boldsymbol{\theta}})$, measures the *rate of change of the derivative* of the loglikelihood. This is essentially the *curvature* of the loglikelihood. Thus it will be seen that the variance is simply the inverse of the *magnitude* of this curvature at the stationary point.

Though easier to calculate than the information, the observed information $-\partial^2 \mathbf{L}(\boldsymbol{\theta}, \mathbf{y}) / \partial \boldsymbol{\theta}^2$ can still be very messy to evaluate analytically. Again it is usually much easier to calculate this numerically using a finite-difference formula for the second derivatives. The expression is a matrix of course, and the variance-covariance matrix of the MLE is the negative of its *inverse*. A numerical procedure is needed for this inversion.

The result (6) provides readily calculated confidence intervals. For example a $(1-\alpha)100\%$ confidence interval for the coefficient θ_1 is

$$\hat{\theta}_1 \pm z_{\alpha/2} \sqrt{V_{11}(\hat{\boldsymbol{\theta}})}$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

We are interested not in $\boldsymbol{\theta}$ directly, but in the regression function $\eta(x, \boldsymbol{\theta})$. For fixed x, this is simply a given function of $\boldsymbol{\theta}$, $g(\boldsymbol{\theta})$ say. The general *invariant* property of the MLE then means that the MLE of $g(\boldsymbol{\theta}_0)$ is

$$\hat{g} = g(\hat{\boldsymbol{\theta}}).$$

The so-called delta-method provides an approximate confidence interval for the unknown $g(\boldsymbol{\theta}_0)$ provided $g(\boldsymbol{\theta})$ possesses a Taylor series expansion about $\boldsymbol{\theta}_0$; that is

$$g(\hat{\boldsymbol{\theta}}) = g(\boldsymbol{\theta}_0) + \left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}_0}^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + r(\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right|) \tag{7}$$

where $r(t)$ is a remainder term that is of order t^{-2} in probability, i.e. $r(t) = O_p(t^{-2})$. Then we have that

$$\begin{aligned} Var[g(\hat{\boldsymbol{\theta}})] &\cong Var[\left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}_0}^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)] \\ &= E\left[\left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}_0}^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}_0}\right] \\ &= \left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}_0}^T \mathbf{V}(\hat{\boldsymbol{\theta}}) \left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}_0} \\ &= \left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)_{\hat{\boldsymbol{\theta}}}^T \mathbf{V}(\hat{\boldsymbol{\theta}}) \left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)_{\hat{\boldsymbol{\theta}}}. \end{aligned}$$

An approximate $(1-\alpha)100\%$ confidence interval for $g(\boldsymbol{\theta})$ is thus

$$g(\hat{\boldsymbol{\theta}}) \pm z(\alpha/2) \sqrt{(\partial g / \partial \boldsymbol{\theta})\big|_{\hat{\boldsymbol{\theta}}}^T \mathbf{V}(\hat{\boldsymbol{\theta}}) (\partial g / \partial \boldsymbol{\theta})\big|_{\hat{\boldsymbol{\theta}}}} \tag{8}$$

where $z(\alpha)$ is the upper α quantile of the standard normal distribution. In this formula the first derivative of $g(\boldsymbol{\theta})$ is required. As with the evaluation of the information matrix, it can be obtained numerically using a finite-difference calculation.

2.3 Confidence Bands

A conservative confidence band can be constructed for the entire regression curve $\eta(x,\boldsymbol{\theta}_0)$, $a \le x \le b$ in a similar, but more subtle way, to that used in calculating individual confidence intervals.

From (7) it follows that

$$C = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \mathbf{V}^{-1}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{9}$$

is asymptotically distributed as a chi-squared variate with p degrees of freedom. Thus

$$\Pr\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \mathbf{V}^{-1}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \le \chi_p^2(\alpha)\} = 1-\alpha,$$

where $\chi_p^2(\alpha)$ is the upper α chi-squared quantile with p degrees of freedom. To invert this probability and change it into a confidence region, we replace $\boldsymbol{\theta}_0$ by $\boldsymbol{\theta}$ and take all $\boldsymbol{\theta}$ satisfying

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{V}^{-1}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \le \chi_p^2(\alpha). \tag{10}$$

We denote this ellipsoidal region as

$$R(\alpha) = \{\boldsymbol{\theta} : (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{V}^{-1}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \le \chi_p^2(\alpha)\}. \tag{11}$$

Clearly we are $(1-\alpha)100\%$ confident that $\boldsymbol{\theta}_0 \in R(\alpha)$. Thus $R(\alpha)$ is a $(1-\alpha)100\%$ confidence region for $\boldsymbol{\theta}_0$.

Now, for each x satisfying $a \le x \le b$ we can also find the values $\boldsymbol{\theta}_L(x)$ and $\boldsymbol{\theta}_U(x)$ that respectively minimize and maximize $\eta(x,\boldsymbol{\theta})$, i.e.

$$\begin{aligned} \boldsymbol{\theta}_L(x) &= \arg[\min_{\boldsymbol{\theta} \in R(\alpha)} \eta(x,\boldsymbol{\theta})], \\ \boldsymbol{\theta}_U(x) &= \arg[\max_{\boldsymbol{\theta} \in R(\alpha)} \eta(x,\boldsymbol{\theta})]. \end{aligned} \tag{12}$$

If we set

$$\left.\begin{aligned} y_L(x) &= \eta(x,\boldsymbol{\theta}_L(x)) \\ y_U(x) &= \eta(x,\boldsymbol{\theta}_U(x)) \end{aligned}\right\}, \quad a \le x \le b \tag{13}$$

we must have

$$y_L(x) \le \eta(x,\boldsymbol{\theta}) \le y_U(x), \quad a \le x \le b, \ \boldsymbol{\theta} \in R(\alpha), \tag{14}$$

and, as we are $(1-\alpha)100\%$ confident that $\boldsymbol{\theta}_0 \in R(\alpha)$, it follows that $y_L(x)$ and $y_U(x)$, $a \le x \le b$, as given in (11) is a $(1-\alpha)100\%$ confidence band for $\eta(x,\boldsymbol{\theta}_0)$, $a \le x \le b$.

The band is a conservative one because it is possible that there are $\boldsymbol{\theta} \notin R(\alpha)$ for which

$$y_L(x) \le \eta(x,\boldsymbol{\theta}) \le y_U(x), \quad a \le x \le b.$$

An explicit approximate formula for $y_L(x)$ and $y_U(x)$, $a \le x \le b$, can be obtained using the delta technique if we use the linear approximation (7) for $\eta(x,\boldsymbol{\theta})$ and minimize and maximize this linear approximation subject to (10). The problem is thus, for each x to

$$\min_{\boldsymbol{\theta}}/\max \eta(x,\boldsymbol{\theta}) = \eta(x,\hat{\boldsymbol{\theta}}) + \left(\frac{\partial \eta(x,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\hat{\boldsymbol{\theta}}}^T (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

subject to

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{V}^{-1}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \le \chi_p^2(\alpha).$$

This is easily solved by the method of Lagrange multipliers and yields the solution

$$y_L(x), y_U(x) = \eta(x,\hat{\boldsymbol{\theta}}) \mp h(x,\hat{\boldsymbol{\theta}}), \quad a \le x \le b$$

where

$$h(x,\hat{\boldsymbol{\theta}}) = \sqrt{\chi_p^2(\alpha)\left(\frac{\partial \eta(x,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\hat{\boldsymbol{\theta}}}^T \mathbf{V}(\hat{\boldsymbol{\theta}})\left(\frac{\partial \eta(x,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\hat{\boldsymbol{\theta}}}}. \tag{15}$$

The form of (15) is exactly the same as for the individual confidence interval (8) except that $\sqrt{\chi_p^2(\alpha)}$ replaces $z(\alpha/2)$ in the formula.

3 BOOTSTRAP CONFIDENCE BANDS

We consider two ways that the formula (15) might be made more exact. The first avoids the use of the linear approximation (7) for $\eta(x,\boldsymbol{\theta})$ but retains the asymptotic normality assumption for $\hat{\boldsymbol{\theta}}$. Resampling is not strictly necessary but can be used in this method. The second avoids the linear approximation and also uses bootstrap resampling to better estimate the distribution of $\hat{\boldsymbol{\theta}}$. Good accounts of bootstrapping are given in Davison and Hinkley (1997) and Hjorth (1994). A more general reference is Chernick (1999)

3.1 Exact Calculation of $\eta(x,\boldsymbol{\theta})$

The first adjustment that we consider is where we avoid using the linear approximation (7) for $\eta(x,\boldsymbol{\theta})$, but simply calculate *numerically* the minimum and maximum of

$\eta(x,\boldsymbol{\theta})$ subject (10). We consider the most likely situation (which is actually that tacitly assumed in the case where we use a linear approximation) that $\eta(x,\boldsymbol{\theta})$ is relatively slowly varying so that the minimum and maximum points lie on $\partial[R(\alpha)]$, the boundary of $R(\alpha)$. We therefore only need look for the minimum or maximum, using a direct search method like the Nelder-Mead, restricting the search to points on $\partial[R(\alpha)]$.

An even cruder but quite adequate approach, at least when p is small, is to calculate the values of $\eta(x,\boldsymbol{\theta})$ for a sufficiently large selection of points $\boldsymbol{\theta}_i \in \partial[R(\alpha)]$, and then take as minimum and maximum of $\eta(x,\boldsymbol{\theta})$ the smallest and largest of the $\eta(x,\boldsymbol{\theta}_i)$ amongst those points evaluated. We only use these optimized values in the calculation of $y_L(x)$ and $y_U(x)$ in (13). The behaviour of $\eta(x,\boldsymbol{\theta})$ is quadratic in the difference between the selected $\boldsymbol{\theta}_i$ and the optimum $\boldsymbol{\theta}^*$, so this apparently crude calculation should be adequate provided we take sufficient points on $\partial[R(\alpha)]$.

Points randomly and uniformly distributed on $\partial[R(\alpha)]$ are easily obtained on noting that

$$\partial[R(\alpha)] = \{\boldsymbol{\theta} : (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{V}^{-1}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \chi_p^2(\alpha)\}$$

is simply a level surface of the density of the variable $\mathbf{W} = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ that is (approximately) multivariate normally distributed $MVN[\mathbf{0}, \mathbf{V}(\hat{\boldsymbol{\theta}})]$. Therefore $\mathbf{W}$ can be written as

$$\mathbf{W} = \mathbf{LZ}$$

where $\mathbf{L}$ is the lower triangular matrix of the Cholesky decomposition of $\mathbf{V}$:

$$\mathbf{V} = \mathbf{L}\mathbf{L}^T .$$

We can therefore form a point on $\partial[R(\alpha)]$ by taking $\mathbf{W}$ with components

$$W_i = \chi_p^2(\alpha) z_i^2 / \sum_{j=1}^{p} z_j^2$$

where the $z_i, i = 1,2,...,p$ are p independent $N(0,1)$ variables. (See section 5.5.2 in Banks, 1998)

The projection of a sample of 500 $\mathbf{W}$'s each of dimension 3 onto a two dimensional plane is illustrated in Figure 2.

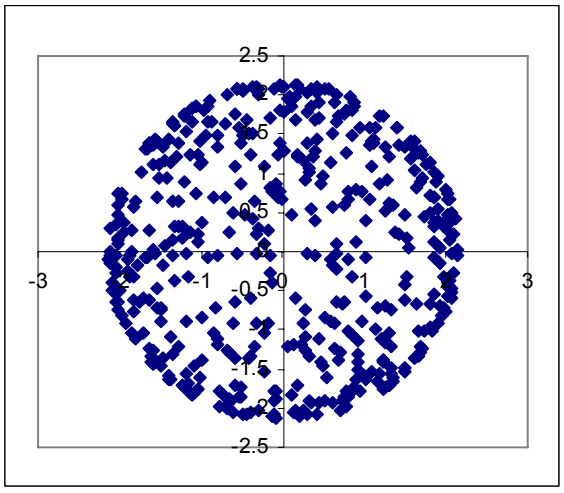


Figure 2: 500 Randomly Generated 3-Dimensional Points on the Surface $\partial[R(\alpha)]$ Projected on 2-Space.

3.2 Bootstrap Resampling of $\hat{\boldsymbol{\theta}}$

Bootstrap resampling can be used as an alternative to avoid the assumption of asymptotic normality of the distribution of $\hat{\boldsymbol{\theta}}$. We consider just the case of parametric bootstrapping. A **y*,** bootstrap sample of **y,** can be obtained with the components of **y*** generated from

$$y_j^* = \eta(x_j, \hat{\boldsymbol{\theta}}) + \varepsilon_j^*, \ j = 1,2,...,n ,$$

where, for simplicity we assume that the $\varepsilon_j^*, \ j = 1,2,...,n$ are independent $N(0, \hat{\sigma}^2)$ variates. For each such sample we carry out ML estimation of $\boldsymbol{\theta}$ to obtain the bootstrap MLE $\hat{\boldsymbol{\theta}}^*$. This therefore yields a bootstrap value of C defined in (9) as

$$C^* = (\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})^T \mathbf{V}^{-1}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}) .$$

We can (easily) replicate the bootstrapping, B times say, to produce a bootstrap sample of C values:

$$\mathbf{C}^* = \{C^{(1)*}, C^{(2)*}, ..., C^{(B)*}\} .$$

A simple estimate of the upper α quantile, which we write as $C^*(\alpha)$, can be obtained from the empirical distribution function of the sample $\mathbf{C}^*$ (or a smoothed version of it). This quantile estimator, $C^*(\alpha)$, can then be used instead of $\chi_p^2(\alpha)$ in the definition of $\partial R(\alpha)$ used to calculate the limits $y_L(x)$ and $y_U(x), a \le x \le b$, of (13).

In the next Section, we illustrate the sampling methods just described.

4 A NUMERICAL EXAMPLE

Though we have discussed the calculation of confidence bands in a simulation setting, the method does not depend on whether the data is real or simulated.

The following is a set of data giving the number of notifications of pulmonary TB (per 100,000) in Morocco in four selected years 1980, 1986, 1993, 2000, grouped by age.

Table 1: Morocco Pulmonary TB notifications per 100,000

Av. Age	Year	1980	1986	1993	2000
2	0-4	1.26	2.78	0.63	0.34
7	5-9	3.53	4.10	1.31	0.91
12	10-14	11.98	13.14	9.86	6.53
19.5	15-24	90.82	97.12	75.85	59.46
29.5	25-34	83.45	116.62	104.00	80.85
39.5	35-44	55.98	67.28	79.33	82.66
54.5	45-64	66.32	78.53	69.10	67.27
75	65+	39.42	55.35	60.76	73.20

We fitted the model

$$y_j = (\theta_2 + \theta_4 x_j + \theta_6 x_j^2)\frac{\exp[\theta_5(x_j - \theta_3)]}{1+\exp[\theta_5(x_j - \theta_3)]} + \varepsilon_j,$$

where $\varepsilon_j \sim N(0, \theta_1^2)$. The model can be criticized on various grounds. We treated all 32 observations as being independent and as have the same variance. Neither assumption is fully justified. However the example is for illustration only so we will not investigate its adequacy any further. The model is highly nonlinear and the form has been carefully selected to give a meaningful parameterization. The ML estimates for the parameters are given in Table 2.

We then calculated 90% confidence intervals and confidence bands using the asymptotic methods of Section 2 and the bootstrapping methods of Section 3.

The resulting bands are shown in Figure 3. It will be seen that the bands are broadly similar. The confidence bands are wider than the confidence intervals. The more accurate version of the bootstrap confidence bands are, as expected, somewhat wider than the bands using the linear approximation for $\eta(x, \boldsymbol{\theta})$. Also the bootstrap versions arguably have the better characteristics. For example the bootstrap bands do not go negative, a weakness of the bands constructed from asymptotic normality theory.

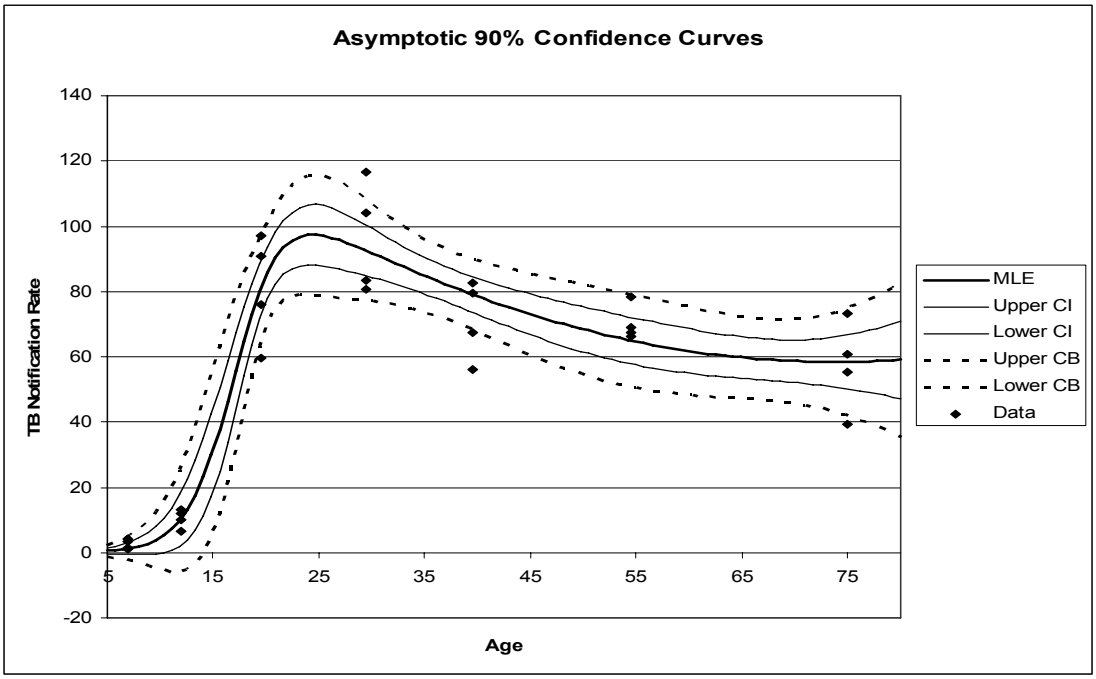
Table 2: MLE's for the Morocco TB Model

Parameter	MLE
θ_1	10.17
θ_2	153.79
θ_3	17.26
θ_4	-2.58
θ_5	0.455
θ_6	0.01746

5 CONCLUSIONS

The resampling and bootstrap methods for constructing confidence bands are easy to apply. The spreadsheet implementations are available and can be downloaded from the author's personal Web page at www.maths.soton.ac.uk/staff/Cheng/Teaching/GTPBootstrap and can be easily modified to handle particular problems.

The bootstrap versions are based on bootstrap percentile confidence intervals. It is generally accepted that use of bootstrap studentized intervals can give more accurate coverage. A detailed discussion of different possible confidence intervals is given in Davison and Hinkley (1997). This has not been attempted in this paper. It is hoped that this work will be carried out shortly and that simulation comparisons will be made in which actual coverage probabilities are determined.



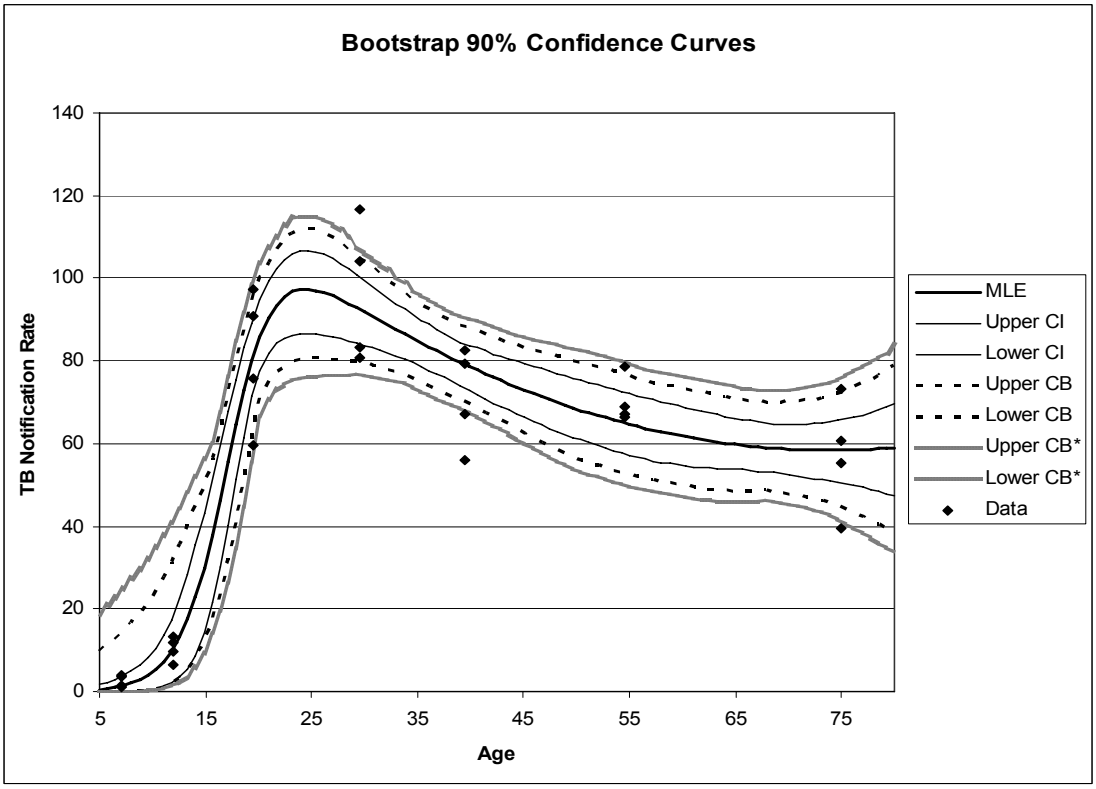


Figure 3: Asymptotic and Bootstrap Confidence Curves for the Moroccan TB Data.

REFERENCES

Banks, J. 1998. *Handbook of simulation*. New York: Wiley.

Barton, R.R., 1998. Simulation metamodels. *Proc. of the 1998 Winter Simulation Conference*, D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds., IEEE, Piscataway, NJ. 167-174.

Chernick, M.R. 1999. *Bootstrap methods, a practitioner's guide*. New York: Wiley.

Davison, A.C. and Hinkley, D.V. 1997. *Bootstrap methods and their application*. Cambridge: Cambridge University Press.

Hjorth, J.S.U. 1994. *Computer intensive statistical methods*. London: Chapman & Hall.

Krzanowski, W.J. 1998. *An introduction to statistical modelling*. London: Arnold.

Law, A.M. and Kelton, W.D. 1991. *Simulation modeling and analysis, 2nd Ed.*, New York: McGraw-Hill.

Miller, R.G. 1981. *Simultaneous Statistical Inference*. New York: Springer.

AUTHOR BIOGRAPHY

RUSSELL C. H. CHENG is Professor and Head of Operational Research 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 Royal Statistical Society and the British Computer Society, Member of the Operational Research Society. His research interests include: variance reduction methods and parametric estimation methods. He was a Joint Editor of the *IMA Journal of Management Mathematics*. His email and web addresses are R.C.H.Cheng@maths.soton.ac.uk and www.maths.soton.ac.uk/staff/Cheng