*Journal of the Korean Statistical Society (2003), 32: 1, pp 1–10*

# CONFIDENCE CURVES FOR A FUNCTION OF PARAMETERS IN NONLINEAR REGRESSION†

MYUNG-WOOK KAHNG[1]

ABSTRACT

We consider obtaining graphical summaries of uncertainty in estimates of parameters in nonlinear models. A nonlinear constrained optimization algorithm is developed for likelihood based confidence intervals for the functions of parameters in the model. The results are applied to the problem of finding significance levels in nonlinear models.

*AMS 2000 subject classifications.* Primary 62J02.
*Keywords.* Wald confidence interval, likelihood ratio confidence interval, efficient score, information matrix.

## 1. INTRODUCTION

The standard nonlinear regression model can be expressed as

$$y_i = f(\mathbf{x}_i, \boldsymbol{\theta}) + \epsilon_i, \ i = 1, 2, \ldots, n,$$

in which the $i^{th}$ response $y_i$ is related to the $q$-dimensional vector of known explanatory variable $\mathbf{x}_i$ through the known model function $f$, which depends on $p$-dimensional unknown parameter vector $\boldsymbol{\theta}$, and $\epsilon_i$ is error. We assume that $f$ is twice continuously differentiable in $\boldsymbol{\theta}$, and errors $\epsilon_i$ are independent, identically distributed normal random variables with mean 0 and variance $\sigma^2$. In matrix notation we will write

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon},$$

where $\mathbf{Y}$ is an $n$-dimensional vector with elements $y_1, y_2, \ldots, y_n$, $\mathbf{X}$ is an $n \times q$ matrix with rows $\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_n^T$, $\boldsymbol{\epsilon}$ is an $n$-dimensional vector with elements $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ and $\mathbf{f}(\mathbf{X}, \boldsymbol{\theta}) = (f(\mathbf{x}_1, \boldsymbol{\theta}), f(\mathbf{x}_2, \boldsymbol{\theta}), \ldots, f(\mathbf{x}_n, \boldsymbol{\theta}))^T$. The least squares

Received November 2001; accepted September 2002.
†This research was supported by the Sookmyung Women's University Research Grants, 2001.
[1]Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea

estimate of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, minimizes the residual sum of squares $S(\boldsymbol{\theta}) = \sum\{y_i - f(\mathbf{x}_i, \boldsymbol{\theta})\}^2$, and maximizes the likelihood function. Let the $p$-dimensional vector parameter $\boldsymbol{\theta}$ be partitioned in the form $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \theta_2)^T$ where $\theta_2$ is a single parameter which is of primary interest, and $\boldsymbol{\theta}_1$ is a $(p-1)$-dimensional nuisance vector. Define $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta}) = \partial \mathbf{f}/\partial \boldsymbol{\theta}^T$ to be the matrix of first derivatives and $\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\theta}})$. The problem of interest is to form confidence regions for $\theta_2$.

Probably the most common procedure for confidence region for the subset parameters $\theta_2$ is based on the large-sample normality of $\hat{\boldsymbol{\theta}}$ (Seber and Wild, 1989, pp. 25, 202). An approximate $100(1-\alpha)\%$ confidence region for $\theta_2$ is given by

$$\left\{ \theta_2 \;\middle|\; |\theta_2 - \hat{\theta}_2| \,/\sqrt{\hat{v}^{22} s^2} \;\leq\; t_\alpha(n-p) \right\}, \tag{1.1}$$

where $\hat{v}^{22}$ is the appropriate diagonal element of $(\hat{\mathbf{V}}^T\hat{\mathbf{V}})^{-1}$ and $s^2 = S(\hat{\boldsymbol{\theta}})/(n-p)$ is the usual unbiased estimator of $\sigma^2$. This confidence interval is called the Wald interval or the linearization interval.

An alternative procedure for defining confidence regions and confidence intervals is based on the likelihood ratio statistic. Let $S(\boldsymbol{\theta}) = S(\boldsymbol{\theta}_1, \theta_2)$. Standard asymptotic arguments suggest that the confidence region for $\theta_2$ is

$$\left\{ \theta_2 \;\middle|\; \sqrt{\{S(\hat{\boldsymbol{\theta}}_1(\theta_2), \theta_2) - S(\hat{\boldsymbol{\theta}})\}/s^2} \;\leq\; t_\alpha(n-p) \right\}, \tag{1.2}$$

where $\hat{\boldsymbol{\theta}}_1(\theta_2)$ denotes the $(p-1)$-dimensional vector valued function that minimizes $S(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$ for $\theta_2$ fixed. As suggested by Beale (1960, p. 53) and confirmed by Donaldson and Schnabel (1987), this regions may be preferable when intrinsic curvature (Bates and Watts, 1980 and 1981) is small.

Although Wald interval (1.1) and likelihood ratio confidence interval (1.2) are asymptotically the same for the nonlinear models and exactly the same for linear models, they can be quite different for small samples with nonlinear models. As noted by Donaldson and Schnabel (1987), confidence interval (1.2) has the coverage rate closer to the nominal value and thus is less affected by curvature than is Wald interval. However, it has computational disadvantages. Confidence intervals using likelihood method can be disjoint and unbounded, and this method is relatively computationally expensive.

## 2. CONFIDENCE CURVES FOR A SINGLE PARAMETER

Cook and Weisberg (1990) give the graphical alternative to likelihood and Wald confidence intervals for a component of the parameter vector $\boldsymbol{\theta}$. The reason

for using confidence curves is that likelihood intervals can have a different shape for each significance level $1-\alpha$. Wald intervals are always symmetric, and if we have a 95% interval, we can always get the 90% and 99% intervals in a simple way. With the likelihood intervals, this is not so; if we have the 95% interval, we cannot say much about the 90% or 99% interval. Consequently, a graphical summary that does not depend on level is desirable.

The confidence curve includes two curves – a likelihood confidence curve and Wald confidence curve. The likelihood confidence curves are the set of points defined by

$$\begin{cases} \sqrt{\{S(\hat{\boldsymbol{\theta}}_1(\theta_2),\theta_2) - S(\hat{\boldsymbol{\theta}})\}/s^2} & \text{on the horizontal axis,} \\ \theta_2 & \text{on the vertical axis.} \end{cases} \tag{2.1}$$

This is a modification of the graphical summary of the standard profile log-likelihood which has $\theta_2$ on the horizontal axis and $S(\hat{\boldsymbol{\theta}}_1(\theta_2),\theta_2)$ on the vertical axis. The plot (2.1) will be curves, with the amount of curvature giving information about the nonlinearity of the model. To this plot, two straight lines passing through $(0,\hat{\theta}_2)$ with slope $\pm se(\hat{\theta}_2)$ are added. These two straight lines represent the Wald interval. At any point on the horizontal axis of the confidence curve plot, the interval between the upper and lower curves provides a confidence interval for $\theta_2$ for some level of $1-\alpha$. The confidence level can be determined from the calibrating distribution for either the Wald or likelihood procedure, which, in the scale of the plot, is $t(\nu)$, where $s$ has $\nu$ degrees of freedom. If $t_\nu^{-1}(u)$ is the inverse of the $t$ cumulative distribution fuction with $\nu$ degrees of freedom evaluated at $u$, then the confidence level at a point along the horizontal axis is $1-2t_\nu^{-1}(\sqrt{\{S(\hat{\boldsymbol{\theta}}_1(\theta_2),\theta_2) - S(\hat{\boldsymbol{\theta}})\}/s^2}\,)$.

The Wald and the likelihood regions are tangent at the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$. If the likelihood is exactly quadratic, that is, if the log-likelihood is exactly normal, then the likelihood curves are same as the Wald curves (two straight lines). Non-normality is reflected in the likelihood confidence curves failing to be straight.

Computation of confidence curves requires a sequence of least squares fits to find $\hat{\boldsymbol{\theta}}_1(\theta_2)$ for selected values of $\theta_2$. Beginning at the maximum likelihood estimate $\hat{\theta}_2$, points on the curve are evaluated at $\hat{\theta}_2 + \Delta_2$. To determine the corresponding value of (2.1), we must find $\hat{\boldsymbol{\theta}}_1(\hat{\theta}_2+\Delta_2)$, which requires the iterative procedure that use the same routines that were used to find the full maximum likelihood estimate $\hat{\boldsymbol{\theta}}$. Once $\hat{\boldsymbol{\theta}}_1(\hat{\theta}_2+\Delta_2)$ has been computed, the next $\hat{\boldsymbol{\theta}}_1(\theta_2)$ can be found by taking $\hat{\theta}_2+\Delta_2$ as the starting point, and iterations as described

above. As Cook and Weisberg (1990) noted, we would like to choose the step size $\Delta_2$ as large as possible subject to: providing a visually satisfying trace of the curve, allowing reasonably accurate linear interpolation between points, and ensuring reliable algorithm for convergence.

To satisfy the above requirements they proposed the dynamic step size which depends on the shape of the confidence curve at the plotting point of $\theta_2$. With this dynamic step size, we have larger step sizes when the confidence curve is nearly linear, and smaller step sizes when the curve is nonlinear. Cook and Weisberg (1990) discussed that their methodology can obtain the confidence curve for the parameter subsets but it has limitation for a function of parameters. The method in the next section overcomes this limitaion.

## 3. CONFIDENCE CURVES FOR A FUNCTION OF PARAMETERS

### *3.1. Introduction*

Suppose that $c(\boldsymbol{\theta})$ is a continuous, twice differentiable and invertible function of $\boldsymbol{\theta}$ and is of particular interest. Confidence intervals for the function $c(\boldsymbol{\theta})$ can be obtained as in Section 2 for a special case $c(\boldsymbol{\theta}) = \theta_i,\ i = 1, 2, \ldots, p$, by using the large-sample normality of $c(\hat{\boldsymbol{\theta}})$ and the likelihood ratio statistic for testing $H_0 : c(\boldsymbol{\theta}) = \phi$.

By the large-sample normality, the maximum likelihood estimate of $c(\boldsymbol{\theta})$ is $c(\hat{\boldsymbol{\theta}})$ with standard error $se[c(\hat{\boldsymbol{\theta}})] = c'(\hat{\boldsymbol{\theta}})^T(\hat{\mathbf{V}}^T\hat{\mathbf{V}})^{-1}c'(\hat{\boldsymbol{\theta}})$ where $c'(\hat{\boldsymbol{\theta}}) = \partial c(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ evaluated at $\hat{\boldsymbol{\theta}}$. Using these, we can construct the Wald confidence interval for $c(\boldsymbol{\theta})$ corresponding to (1.1) having the form

$$\left\{ \phi \;\middle|\; |\phi - \hat{\phi}| \,/\, s\sqrt{c'(\hat{\boldsymbol{\theta}})^T(\hat{\mathbf{V}}^T\hat{\mathbf{V}})^{-1}c'(\hat{\boldsymbol{\theta}})} \;\leq\; t_\alpha(n-p) \right\}$$

where $\hat{\phi} = c(\hat{\boldsymbol{\theta}})$. The likelihood confidence interval corresponding to (1.2) is

$$\left\{ \phi \;\middle|\; \sqrt{\{S(\mathbf{m}_{\boldsymbol{\theta}}(\phi)) - S(\hat{\boldsymbol{\theta}})\}/s^2} \;\leq\; t_\alpha(n-p) \right\}$$

where $\mathbf{m}_{\boldsymbol{\theta}}(\phi)$ denotes the $p$-dimensional vector valued function that minimizes $S(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$ for given $c(\boldsymbol{\theta}) = \phi$.

Following Cook and Weisberg (1990) and generalizing Section 2, we will consider constructing confidence curves for the function of $c(\boldsymbol{\theta})$. In this plot, $\phi$ is on the vertical axis and the plotted point corresponding to $\phi$ on the horizontal axis is of the form $\sqrt{\{S(\mathbf{m}_{\boldsymbol{\theta}}(\phi)) - S(\hat{\boldsymbol{\theta}})\}/s^2}$. To this plot, two straight lines passing

through $(0, \hat{\phi})$ with slope $\pm se(\hat{\phi})$ are added. These straight lines represent the Wald interval.

Beginning at $\phi_0 = \hat{\phi} = c(\hat{\boldsymbol{\theta}})$, we need a series of values $\phi$ and $\mathbf{m}_{\boldsymbol{\theta}}(\phi)$ to construct the confidence curve. Thus computation of the confidence curve is a sequence of least squares fits that obtain the maximum likelihood estimates of $\boldsymbol{\theta}$ for selected values of $\phi$. When we construct a confidence curve for a component of the parameter vector, we compute the least squares estimate of $\boldsymbol{\theta}$ with one parameter as a fixed value in the model. But in the case for the function of $\boldsymbol{\theta}$, we are required to solve a constrained optimization by computing the least squares estimate $\boldsymbol{\theta}$ subject to the constraint that $c(\boldsymbol{\theta})$ is some fixed value $\phi$. This is called the nonlinear constrained optimization problem or nonlinear programming. This problem can be expressed as finding a local solution of $\mathbf{m}_{\boldsymbol{\theta}}(\phi)$:

$$\begin{cases} \text{minimize} & \mathrm{S}(\boldsymbol{\theta}), \\ \text{subject to} & \mathrm{c}(\boldsymbol{\theta}) = \phi. \end{cases}$$

We solve this problem using Lagrange multipliers in Section 3.2. Although we can sometimes compute a series of $\phi$ and $\mathbf{m}_{\boldsymbol{\theta}}(\phi)$ using the method of reparameterizing models with $c(\boldsymbol{\theta})$ as a single parameter, this is not always possible. Suppose that we have $\mathbf{m}_{\boldsymbol{\theta}}(\phi_0)$ and $S(\mathbf{m}_{\boldsymbol{\theta}}(\phi_0))$ for some value $\phi_0$ of $c(\boldsymbol{\theta})$, then the next point on the curve is evaluated at $\phi = \phi_0 + \Delta_\phi$, where $\Delta_\phi$ is a selected step size. We discuss the dynamic step size in Section 3.3.

### *3.2. Lagrange multipliers*

Suppose that initially we just wish to minimize an unconstrained function $S(\boldsymbol{\theta})$. Many minimization methods are based upon trying to locate a point $\boldsymbol{\theta}^*$ such that

$$S'(\boldsymbol{\theta}^*) = 0, \tag{3.1}$$

where $S'(\boldsymbol{\theta}) = \partial S(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$. In general, this point is referred to as a stationary point with respect to $\boldsymbol{\theta}$ of function $S(\boldsymbol{\theta})$. It is well known that (3.1) is a necessary condition for a minimum. A value of $\boldsymbol{\theta}$ for which $S'(\boldsymbol{\theta}) = 0$ does not necessarily minimize the function $S(\boldsymbol{\theta})$; it may maximize the function or be a saddle point, but under certain conditions it will definitely minimize $S(\boldsymbol{\theta})$ at least locally. One such condition arises if the function $S(\boldsymbol{\theta})$ is convex. The function $S(\boldsymbol{\theta})$ is said to be convex over a set $\boldsymbol{\Theta}_0$ if its matrix of second partial derivatives is positive semi-definite everywhere. In other words, $\mathbf{d}^T S''(\boldsymbol{\theta}_0)\mathbf{d} \leq 0$, for all vectors $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_0$ and $\mathbf{d} \neq 0$. If $S(\boldsymbol{\theta})$ is a convex function, any point where the derivative vanishes must be a minimum.

The problem becomes more difficult if we have to consider constraints. Suppose that we want to minimize the function $S(\boldsymbol{\theta})$ subject to the constraints

$$c(\boldsymbol{\theta}) = \phi. \tag{3.2}$$

If any point $\boldsymbol{\theta}$ satisfies the constraint (3.2), it is said to be a feasible point. We may be able to overcome the problem by using the constraints to solve for one variable in terms of others. If this can be done it reduces the dimension of the problem and the constraint will actually make things easier, but in general we cannot make this simplification.

There have been many contributions to the theory and application of constrained optimization. One approach is to solve the problem using Lagrange multipliers (Fletcher, 1987, p. 195). Since the function $c(\boldsymbol{\theta})$ must equal $\phi$, the problem is unaffected if we replace the object function $S(\boldsymbol{\theta})$ to be minimized by the Lagrangian function

$$\mathcal{L}(\boldsymbol{\theta}) = S(\boldsymbol{\theta}) - \lambda[c(\boldsymbol{\theta}) - \phi] \tag{3.3}$$

for any value of the quantity $\lambda$, which is called the Lagrange multiplier. The Lagrangian function $\mathcal{L}(\boldsymbol{\theta})$ is often written as $\mathcal{L}(\boldsymbol{\theta}, \lambda)$ to emphasize its dependence on both $\boldsymbol{\theta}$ and $\lambda$.

The Lagrange multiplier method is useful in that if we can find the value of $\lambda$ such that the vector $\boldsymbol{\theta}$ where $S(\boldsymbol{\theta})$ is minimized happens to satisfy the constraint $c(\boldsymbol{\theta}) = \phi$, then we will also have solved our original constrained problem. This is obvious since the Lagrangian function is reduced to our original object function $S(\boldsymbol{\theta})$ if the constraint is satisfied. Thus we try to find vector $\mathbf{m}_{\boldsymbol{\theta}}(\phi)$ and value $m_{\lambda}(\phi)$ which minimize (3.3) and also satisfy constraint (3.2). From the earlier discussion about minimization we know that a necessary condition for a minimum is $\partial\mathcal{L}/\partial\boldsymbol{\Psi} = 0$, where $\boldsymbol{\Psi} = (\boldsymbol{\theta}^T, \lambda)^T$. In other words, we need to solve the following system

$$\begin{cases} S'(\boldsymbol{\theta}) - \lambda c'(\boldsymbol{\theta}) = 0, \\ c(\boldsymbol{\theta}) = \phi. \end{cases} \tag{3.4}$$

We now have $p + 1$ equations for the $p + 1$ unknown $\boldsymbol{\theta}$ and $\lambda$ which we may possibly be able to solve.

A direct and efficient approach for solving (3.4) is to iterate on the basis of certain approximations of the function $\mathcal{L}(\boldsymbol{\theta}, \lambda)$. For this problem, the method can be explained as Newton's method applied to nonlinear equations. The Lagrangian function is defined in terms of variables $\boldsymbol{\theta}$ and $\lambda$, so a feature of the resulting method is that a sequence of approximations $\boldsymbol{\theta}^{(k)}$, $\lambda^{(k)}$ to both solution vector

$\mathbf{m}_{\boldsymbol{\theta}}(\phi_0)$ and the optimum Lagrange multiplier $m_\lambda(\phi)$ are generated. That is, given $\boldsymbol{\theta}^{(k)}$, $\lambda^{(k)}$ the new points $\boldsymbol{\theta}^{(k+1)}$, $\lambda^{(k+1)}$ are determined by setting $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \Delta\boldsymbol{\theta}$ and $\lambda^{(k+1)} = \lambda^{(k)} + \Delta\lambda$.

As usual a Taylor series for $\partial\mathcal{L}/\partial\Psi$ about $\boldsymbol{\theta}^{(k)}$, $\lambda^{(k)}$ gives

$$\mathcal{L}'(\boldsymbol{\theta}^{(k)} + \Delta\boldsymbol{\theta}, \lambda^{(k)} + \Delta\lambda) = \mathcal{L}'(\boldsymbol{\theta}^{(k)}, \lambda^{(k)}) + \mathcal{L}''(\boldsymbol{\theta}^{(k)}, \lambda^{(k)}) \begin{bmatrix} \Delta\boldsymbol{\theta} \\ \Delta\lambda \end{bmatrix} + \cdots$$

where

$$\mathcal{L}'(\boldsymbol{\theta}^{(k)}, \lambda^{(k)}) = \left.\frac{\partial\mathcal{L}}{\partial\Psi}\right|_{\Psi=(\boldsymbol{\theta}^{(k)},\lambda^{(k)})}, \quad \mathcal{L}''(\boldsymbol{\theta}^{(k)}, \lambda^{(k)}) = \left.\frac{\partial^2\mathcal{L}}{\partial\Psi\partial\Psi^T}\right|_{\Psi=(\boldsymbol{\theta}^{(k)},\lambda^{(k)})}.$$

Neglecting higher order terms and setting the left hand side equal to zero gives the iteration

$$\mathcal{L}''(\boldsymbol{\theta}^{(k)}, \lambda^{(k)}) \begin{bmatrix} \Delta\boldsymbol{\theta} \\ \Delta\lambda \end{bmatrix} = -\mathcal{L}'(\boldsymbol{\theta}^{(k)}, \lambda^{(k)}). \tag{3.5}$$

This can be solved to give corrections $\Delta\boldsymbol{\theta}$ and $\Delta\lambda$, and is Newton's method for the stationary point problem. Obtaining $\mathcal{L}'(\boldsymbol{\theta}^{(k)}, \lambda^{(k)})$, $\mathcal{L}''(\boldsymbol{\theta}^{(k)}, \lambda^{(k)})$ from (3.3) and substituting into (3.5) gives us the following system after some rearrangement:

$$\begin{bmatrix} \mathcal{L}''^{(k)} & -c'^{(k)} \\ -c'^{(k)} & 0 \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{\theta} \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} -S'^{(k)} + c^{(k)}\lambda^{(k)} \\ c^{(k)} \end{bmatrix},$$

where

$$\mathcal{L}'^{(k)} = \mathcal{L}'(\boldsymbol{\theta}^{(k)}) = \left.\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}}, \quad \mathcal{L}''^{(k)} = \mathcal{L}''(\boldsymbol{\theta}^{(k)}) = \left.\frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}},$$

$$S'^{(k)} = S'(\boldsymbol{\theta}^{(k)}), \quad S''^{(k)} = S''(\boldsymbol{\theta}^{(k)}), \quad c^{(k)} = c(\boldsymbol{\theta}^{(k)}), \quad c'^{(k)} = c'(\boldsymbol{\theta}^{(k)}).$$

It is more convenient to write $\lambda^{(k+1)} = \lambda^{(k)} + \Delta\lambda$ and $\boldsymbol{\Delta}^{(k)} = \Delta\boldsymbol{\theta}$, and to solve the equivalent system

$$\begin{bmatrix} \mathcal{L}''^{(k)} & -c'^{(k)} \\ -c'^{(k)} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta} \\ \lambda \end{bmatrix} = \begin{bmatrix} -S'^{(k)} \\ c^{(k)} \end{bmatrix} \tag{3.6}$$

to determine $\boldsymbol{\Delta}^{(k)}$ and $\lambda^{(k+1)}$. Then $\boldsymbol{\theta}^{(k+1)}$ is given by

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \boldsymbol{\Delta}^{(k+1)}. \tag{3.7}$$

This method requires initial approximations $\boldsymbol{\theta}^{(1)}$ and $\lambda^{(1)}$ and uses (3.6) and (3.7) to generate the iterative sequence $(\boldsymbol{\theta}^{(k)}, \lambda^{(k)})$.

First order necessary conditions (Fletcher, 1987, p. 200) for a feasible point $\boldsymbol{\theta}$ to solve this problem are that there exist Lagrange multipliers $\lambda$ such that

$$S'(\boldsymbol{\theta}) = \lambda c'(\boldsymbol{\theta}). \tag{3.8}$$

It is also possible to look at second order effects caused by feasible changes in the solution. Second order necessary conditions (Fletcher, 1987, pp. 207–212) are that the Lagrangian function has nonnegative curvature for all feasible directions, that is, $\mathbf{d}^T \mathcal{L}''(\boldsymbol{\theta})\mathbf{d} \geq \mathbf{0}$ for all $\mathbf{d} : c'(\boldsymbol{\theta})^T \mathbf{d} = \mathbf{0}$. More details of the theory of nonlinear programming can be found in Gill *et al.* (1981), Luenberger (1984) and Fletcher (1987).

### *3.3. Dynamic step size*

We consider, once again, the dynamic step size for the confidence curve for the function of parameters. A likelihood confidence curve consists of the set of two-dimensional points

$$\mathbf{P}(\phi) = \begin{bmatrix} P_1(\phi) \\ P_2(\phi) \end{bmatrix} = \begin{bmatrix} G^{1/2}(\phi)/s \\ \phi \end{bmatrix} \tag{3.9}$$

where $G(\phi) = S(\mathbf{m}_{\boldsymbol{\theta}}(\phi)) - S(\hat{\boldsymbol{\theta}})$. Suppose $\mathbf{P}(\phi_0)$ is a current plotting point, then a curvature of a plane curve (Stoker, 1969, p. 26) at $\mathbf{P}(\phi_0)$ is

$$C = \frac{|P_1' P_2'' + P_1'' P_2'|}{|{P_1'}^2 + {P_2'}^2|^{3/2}} = \frac{|P_1' 0 - P_1'' 1|}{(\|\mathbf{P}\|^2)^{3/2}} = \frac{|P_1''|}{se(\hat{\phi})\|\mathbf{P}'\|^3} \tag{3.10}$$

where $\mathbf{P}' = \mathbf{P}'(\phi_0)$ and $\mathbf{P}'' = \mathbf{P}''(\phi_0)$ are the first and second partial derivatives, $\partial\mathbf{P}/\partial\phi$ and $\partial^2\mathbf{P}/\partial\phi^2$ evaluated at $\phi_0$, respectively.

Cook and Weisberg suggest choosing step size inversely proportional to $C$ at $\mathbf{P}(\phi_0)$. We have the invariant step size as follows:

$$\Delta_\phi = \frac{1}{C} \frac{\tan(\eta)}{\|\mathbf{P}'\|}$$

and applying (3.10), we have

$$\Delta_\phi = \tan(\eta)\sqrt{2\, G_0\, (\hat{c}'^T \hat{\mathcal{L}}''^{-1} \hat{c}')}\, \frac{|\hat{\lambda}_0^2 + 2\, G_0\, (\hat{c}'^T \hat{\mathcal{L}}''^{-1} \hat{c}'))^{-1}|}{|\hat{\lambda}_0^2 - 2\, G_0\, ({c_0'}^T \mathcal{L}_0''^{-1} c_0')^{-1}|} \tag{3.11}$$

where $c' = c'(\hat{\boldsymbol{\theta}})$ and $\mathcal{L}'' = \mathcal{L}''(\hat{\boldsymbol{\theta}})$.

The first step size cannot be determined dynamically as (3.11) since at $\mathbf{P}(\hat{\phi})$, $G_0 = \hat{\phi} = 0$ and $\hat{\lambda}_0 = m_\lambda(\hat{\phi}) = 0$. Thus we have to set the first step size. A small fraction of $se(\hat{\phi})$, such as $0.5 \sim 5\%$ of $se(\hat{\phi})$ works well. Also we need to determine a user selected angle $\eta$; values between 5 and 10 degrees used by Cook and Weisberg (1990) seem to be reasonable in this case.

## 4. EXAMPLES

We now demonstrate the above methodology by using the data and the model taken from Clarke (1987). The data examines the weight of cut grass as a function of the weeks after commencement of grazing in a pasture for 13 cases. The proposed model is the Mitcherlitz equation, $f(x, \boldsymbol{\theta}) = \theta_3 + \theta_2 \exp(\theta_1 x)$.

Suppose that function of the parameters $\phi = \theta_3/\theta_2$ is of particular interest. To compute a confidence curve for $\phi$, we need to compute a series of least squares estimates of $\boldsymbol{\theta}$ subject to the constraint that $\phi$ is fixed. Using the procedures of Section 3 and dynamic step size (3.11), the confidence curve for $\phi$ is given in Figure 4.1.

The likelihood confidence curves are curved, indicating nonlinearity, and the 95% confidence interval for $\phi$ is $[-0.1474,\ 0.6637] = [\hat{\phi} - 0.5297,\ \hat{\phi} + 0.2814]$, which is very different from the symmetric Wald interval, $[0.009798,\ 0.7549] = [\hat{\phi} \pm 0.3725]$. This asymmetry of the likelihood confidence curves indicate that the linear approximation seems to be misleading.

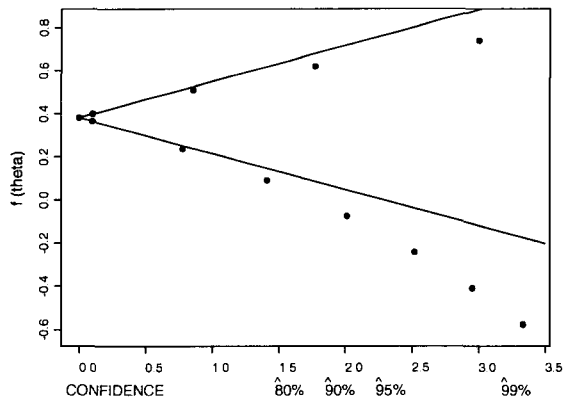


FIGURE 4.1 *Confidence curve for* $\theta_3/\theta_2$

## REFERENCES

BATES, D. M. AND WATTS, D. G. (1980). "Relative curvature measures of nonlinearity (with discussion)", *Journal of the Royal Statistical Society*, **B42**, 1–25.

BATES, D. M. AND WATTS, D. G. (1981). "Parameter transformations for improved approximate confidence regions in nonlinear least squares", *The Annals of Statistics*, **9**, 1152–1167.

BATES, D. M. AND WATTS, D. G. (1988). *Nonlinear Regression Analysis and Its Application*, John Wiley & Sons, New York.

BEALE, E. M. L. (1960). "Confidence regions in nonlinear estimation (with discussion)", *Journal of the Royal Statistical Society*, **B22**, 1152–1167.

CLARKE, G. P. Y. (1987). "Approximate confidence limits for a parameter function in nonlinear regression", *Journal of the American Statistical Association*, **82**, 221–230.

COOK, R. D. AND WEISBERG, S. (1990). "Confidence curves in nonlinear regression", *Journal of the American Statistical Association*, **85**, 544–551.

DONALDSON, J. R. AND SCHNABEL, R. B. (1987). "Computational experiences for confidence regions and confidence intervals for nonlinear least squares", *Technometrics*, **29**, 67–82.

FLETCHER, R. (1987). *Practical Methods of Optimization*, 2nd ed., Wiley, Chichester.

GILL, P. E., MURRAY, W. AND WRIGHT, M. H. (1981). *Practical Optimization*, Academic Press, New York.

LUENBERGER, D. G. (1984). *Linear and Nonlinear Programming*, 2nd ed., Addison-Wesley, Reading, Mass.

SEBER, G. A. F. AND WILD, C. J. (1989). *Nonlinear Regression*, John Wiley & Sons, New York.

STOKER, J. J. (1969). *Differential Geometry*, John Wiley & Sons, New York.