Confidence regions and intervals in nonlinear regression*

MIRTA BENŠIĆ†

Abstract. *This lecture presents some methods which we can apply in searching for confidence regions and intervals for true values of regression parameters. The nonlinear regression models with independent and identically distributed errors and L_p norm estimators are discussed.*

Key words: *nonlinear regression, confidence regions, confidence intervals, L_p norm*

Sažetak. **Područja i intervali povjerenja nelinearne regresije**. *U ovom predavanju opisane su neke metode koje možemo primijeniti za traženje područja i intervala povjerenja pravih vrijednosti regresijskih parametara. Pri tome razmatrani su nelinearni regresijski modeli s nezavisno i jednoliko distribuiranim greškama u L_p normi.*

Ključne riječi: *nelinearna regresija, područja povjerenja, intervali povjerenja, L_p norma*

1. Introduction

This lecture presents a brief review of methods which we apply in searching for confidence regions and confidence intervals for true values of the regression parameters in nonlinear models.

As it is known, the usual least squares estimator for the regression parameters is not always the best choice for an estimator if the additive errors are not normal [10]. Namely, this estimator is known to be sensitive to departures from normality in the residual distributions. As alternatives to the least squares estimator in these cases, members of the class of L_p norm estimators have been proposed. In this context we discuss the ways for computing the confidence intervals and regions if the L_p-norm estimator is used.

*The lecture presented at the MATHEMATICAL COLLOQUIUM in Osijek organized by Croatian Mathematical Society - Division Osijek, January 31, 1997.

†Faculty of Economics, University of Osijek, Gajev trg 7, HR-31 000 Osijek, Croatia, e-mail: bensic@oliver.efos.hr

2. The model

Let us suppose that we have response variables y_i ($i = 1, \ldots, n$), observed with unknown errors e_i ($i = 1, \ldots, n$) and we want to fit them to m fixed predictor variables $x_{i1}, \ldots, x_{im}$ ($\mathbf{x}_i = [x_{i1}, \ldots, x_{im}]^\tau$,) $i = 1, \ldots, n$ using a function $f(\mathbf{x}_i; \boldsymbol{\theta})$. Here $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_k]^\tau$ is the vector of k unknown parameters and we suppose the function f is nonlinear in its parameters.

When the errors e_i are additive random variables, the response variables can be shown by

$$\mathbf{y} = F(\dot{\boldsymbol{\theta}}) + \mathbf{e},$$

$$\mathbf{y} = [y_1, \ldots, y_n]^\tau, \quad F(\boldsymbol{\theta}) = [f(\mathbf{x}_1, \boldsymbol{\theta}), \ldots, f(\mathbf{x}_n, \boldsymbol{\theta})]^\tau$$

$$\mathbf{e} = [e_1, \ldots, e_n]^\tau.$$

Here $\dot{\boldsymbol{\theta}}$ denotes a true, but unknown value of the vector parameter $\boldsymbol{\theta}$. Moreover, let us suppose that the errors e_i (i=1,...,n) are independent and identically distributed random variables.

The L_p-norm estimator of the parameter $\boldsymbol{\theta}$ is the value $\hat{\boldsymbol{\theta}}^{(p)} = [\hat{\theta}_1^{(p)}, \ldots, \hat{\theta}_k^{(p)}]$ that minimizes the sum of the p-th exponent of the absolute value of residuals, $p \in [1, \infty)$. Thus, if we denote

$$|r_i(\boldsymbol{\theta})| = |y_i - f(\mathbf{x}_i; \boldsymbol{\theta})|, \quad i = 1, \ldots, n$$

$$S_p(\boldsymbol{\theta}) = \sum_{i=1}^{n} |r_i(\boldsymbol{\theta})|^p$$

$\hat{\boldsymbol{\theta}}^{(p)}$ is a vector which satisfies

$$S_p(\hat{\boldsymbol{\theta}}^{(p)}) = \min_{\boldsymbol{\theta} \in \Theta} S_p(\boldsymbol{\theta})$$

if it exists. Here, Θ is the set of all possible values for the vector parameter $\boldsymbol{\theta}$, $\Theta \subseteq \mathbf{R}^k$. Being special cases, this class of estimators contains the minimum absolute deviations estimator (MAD, $p = 1$) and the least squares estimator (LSE, $p = 2$).

As we suppose, $\mathbf{y}$ is a random vector which depends on m deterministic predictor variables. It means that the vector $\hat{\boldsymbol{\theta}}^{(p)}$ will also be random and for the specific realizations of $\mathbf{y}$ we have different values of $\hat{\boldsymbol{\theta}}^{(p)}$. Which is the true one? We can answer this question only through the confidence regions, i.e. using the properties of $\hat{\boldsymbol{\theta}}^{(p)}$ as a random vector, we can indicate with some specific confidence level $1 - \alpha$ ($\alpha \in (0, 1)$) in what region about $\hat{\boldsymbol{\theta}}^{(p)}$ we might reasonably expect $\dot{\boldsymbol{\theta}}$ to be. Such regions are known as $100(1 - \alpha)\%$ confidence regions.

A joint confidence region for all parameters $\dot{\theta}_1, \ldots, \dot{\theta}_k$ is defined using a function

$$CR_\alpha : Y \rightarrow \text{a region in } \mathbf{R}^k$$

that satisfies

$$P\{\dot{\boldsymbol{\theta}} \in CR_\alpha(y)\} = 1 - \alpha.$$

A confidence interval for an individual parameter $\dot{\theta}_j$ is defined using a function

$$CI_{j,\alpha} : Y \to \text{an interval in } \mathbf{R}$$

that satisfies

$$P\{\dot{\theta}_j \in CI_{j,\alpha}(y)\} = 1 - \alpha$$

(see, for example, [3]).

3. Least squares estimator

If $p = 2$, the L_p norm estimator is in fact the LSE. There are currently many results regarding this case for computing confidence regions in approximate sense (for large samples) or in exact sense (for small samples). These results mostly regard models with normal, independent and identically distributed errors (i.i.d. errors). Thus, in this section we will suppose that the errors are normal i.i.d.

The least expensive computational procedure for computing confidence regions if LSE is applied arose from the linearization approach and suggests $(1-\alpha)\%$ confidence regions for $\dot{\boldsymbol{\theta}}$:[1]

$$\{\boldsymbol{\theta} : (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\tau \hat{V}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \le kF_{k,n-k,1-\alpha}\},$$

where

$$\hat{V} = s^2[(J(\hat{\boldsymbol{\theta}}))^\tau J(\hat{\boldsymbol{\theta}})]^{-1},$$

$$s^2 = \frac{S_2(\hat{\boldsymbol{\theta}})}{n-k},$$

$J(\hat{\boldsymbol{\theta}})$ is the Jacobian matrix of $F(\boldsymbol{\theta})$ at $\hat{\boldsymbol{\theta}}$. This approach also suggests the $(1-\alpha)\%$ confidence interval for θ_j, $j = 1, \ldots, k$:

$$\{\boldsymbol{\theta} : |\theta_j - \hat{\theta}_j| \le \hat{V}_{jj}^{1/2} t_{n-k,1-\alpha/2}\},$$

where V_{jj} is the $(j,j)^{th}$ element of $\hat{V}$.

If the contours of $S_2(\boldsymbol{\theta})$ are approximately elliptical (exactly elliptical in the linear case), these approximations will be adequate but this method can be very poor if the contours of $S_2(\boldsymbol{\theta})$ are not close to ellipses.

There are two other methods for large samples which are more consistent with the Bates and Watts curvature measures ([1]). Thus, the likelihood approach suggests the $(1-\alpha)\%$ confidence region for $\dot{\boldsymbol{\theta}}$:

$$\{\boldsymbol{\theta} : S_2(\boldsymbol{\theta}) \le S_2(\hat{\boldsymbol{\theta}})[1 + \frac{k}{n-k} F_{k,n-k,1-\alpha}]\}$$

and the lack-of-fit approach suggests:

$$\{\boldsymbol{\theta} : \frac{R(\boldsymbol{\theta})^\tau P(\boldsymbol{\theta}) R(\boldsymbol{\theta})}{R(\boldsymbol{\theta})^\tau (I - P(\boldsymbol{\theta})) R(\boldsymbol{\theta})} \le \frac{k}{n-k} F_{k,n-k,1-\alpha}\}$$

[1] Here $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(2)}$.

$$P(\boldsymbol{\theta}) = J(\boldsymbol{\theta})(J(\boldsymbol{\theta})^\tau J(\boldsymbol{\theta}))^{-1} J(\boldsymbol{\theta})^\tau$$
$$R(\boldsymbol{\theta}) = Y - F(\boldsymbol{\theta}).$$

As we can see, these two methods are computationally very expensive, requiring the evaluation of a sufficient number of points to produce a contour. The lack-of-fit approach gives in fact the exact regions which are not dependent on $\hat{\boldsymbol{\theta}}$. The assumption that the errors are normal is the only reason for puting this method in the LSE case. Namely, if the errors are normal, LSE is better than the other estimators from the class of L_p norm estimators.

When the sample size n is small, Duncan ([4]) suggested the jackknife procedure for computing the confidence region. The procedure is as follows:

1. Let $\hat{\boldsymbol{\theta}}_{(i)}$ be the least squares estimate of $\boldsymbol{\theta}$ when the i^{th} case is deleted from the sample.

2. Calculate the pseudo-values as the vectors

$$\mathbf{T}_i = n\hat{\boldsymbol{\theta}} - (n-1)\hat{\boldsymbol{\theta}}_{(i)}$$

The sample mean and variance of $\mathbf{T}_i$ are given by

$$\bar{\mathbf{T}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{T}_i, \quad \bar{\mathbf{T}} = [\bar{T}_1, \ldots \bar{T}_k]^\tau$$

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n} (\mathbf{T}_i - \bar{\mathbf{T}})(\mathbf{T}_i - \bar{\mathbf{T}})^\tau \;\; \mathbf{S} = [\mathbf{S}_{ij}] \; i,j = 1, \ldots, k.$$

3. A $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta}$ is

$$\{\boldsymbol{\theta} | (\mathbf{T} - \boldsymbol{\theta})^\tau S^{-1} (\mathbf{T}_i - \boldsymbol{\theta}) \leq \frac{k}{n-k} F_{k,n-k,1-\alpha}\}.$$

A $100(1-\alpha)\%$ confidence interval for θ_i can be constructed as[2]

$$\bar{T}_i \pm \sqrt{\frac{k}{n-k} F_{k,n-k,1-\alpha} \mathbf{S}_{ii}}.$$

4. L_p norm estimation

The confidence intervals for regression parameters in nonlinear models that have been suggested to this day were computed using the results on asymptotic distribution of the L_p norm estimator in the linear models and the fact that the errors are additive (see [8], [9],[7]). These intervals are only asymptotical.

Thus, if the L_1 norm estimator is applied, a $100(1-\alpha)\%$ confidence interval for θ_j is given by

$$\theta_j^{(1)} \pm z_{\alpha/2} \sqrt{\omega_1^2 (J(\boldsymbol{\theta}^{(1)})^\tau J(\boldsymbol{\theta}^{(1)})_{jj}}$$

[2] As we can see this procedure requires $n+1$ nonlinear estimations. Fox et al. ([5], [6]) describes a linear jackknife procedure which is not so computationaly expensive as this.

where

$$\omega_1^2 = \frac{1}{[2f(m)]^2}$$

and $f(m)$ is the ordinate of the error distribution at the median m. ω_1^2 can be estimated by the Cox and Hinkley estimator ([2], [6]).

If the L_p norm is applied, $p \in (1, \infty)$ then a $100(1-\alpha)\%$ confidence interval for θ_j is given by

$$\theta_j^{(p)} \pm z_{\alpha/2}\sqrt{\omega_p^2 (J(\boldsymbol{\theta}^{(p)})^\tau J(\boldsymbol{\theta}^{(p)})_{jj}}$$

where

$$\omega_p^2 = \frac{E[|e_i|^{2p-p}]}{[(p-1)E(|e_i|^{p-2})]^2}.$$

Here are ω_p^2 for some symmetric distributions:

- The uniform distribution on $[-b, b]$:

$$\omega_p^2 = \frac{b^2}{2p-1} = \frac{3\sigma^2}{2p-1}$$

- The normal distributions $\mathcal{N}(0, \sigma^2)$:

$$\omega_p^2 = \frac{2\sqrt{\pi}\sigma^2\Gamma(p-\frac{1}{2})}{(p-1)^2\Gamma^2(\frac{p-1}{2})}$$

- The symmetric Laplace distribution (density function: $f(x) = \frac{1}{2b}e^{-|x|/b}$):

$$\omega_p^2 = \frac{\sigma^2\Gamma(2p-1)}{2(p-1)^2\Gamma^2(p-1)}$$

Note that $\omega_p^2 = \sigma^2$ when $p = 2$ for all the distributions.

References

[1] D. M. Bates, D. G. Watts, *Relative curvature measures of nonlinearity*, J.R. Statist. Soc., Ser. B **42**(1980), 1–25.

[2] D. R. Cox, D. V. Hinkley, *Theoretical statistics*, Chapman and Hall, London, 1974.

[3] J. R. Donaldson, R. B. Schnabel, *Computational expirience with confidence regions and confidence intervals for nonlinear least squares*, Technometrics **29**(1987), 67–82.

[4] G. T. Duncan, *An empirical study of jackknife - constructed confidence regions in nonlinear regression*, Technometrics **20**(1978), 123–129.

[5] T. Fox, D. V. Hinkley, K. Larntz, *Jackknife in nonlinear regression*, Technometrics **22**(1980), 29–33.

[6] R. Gonnin, A. H. Money, *Nonlinear L_p-norm estimation*, Marcel Dekker, Inc., New York and Basel, 1989.

[7] R. Gonnin, A. H. Money, *Nonlinear L_p-norm estimation: Part I - On the choice of the exponent, p, where the errors are additive*, Commun. Statist. - Theor. Meth. **14**(1985), 827–840.

[8] R. I. Jenrich, *Asymptotic properties of non-linear least squares estimators*, Ann. Math. Statist. **40**(1969), 633–643.

[9] H. Nyquist, *The optimal L_p norm estimator in linear regression models*, Commun. Statist. - Theor. Meth. **12**(1983), 2511–2524.

[10] V. A. Sposito, M. L. Hand, B. Skarpness, *On the efficiency of using the sample kurtosis in selecting optimal L_p norm estimators*, Commun. Statist. - Simula. Computa. **12**(1983), 265–272.