Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 188 (2008) 460–484
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Stochastics and Statistics

Dealing with the multiplicity of solutions of the ℓ_1 and ℓ_∞ regression models

Enrique Castillo [a,b,*], Roberto Mínguez [b], Carmen Castillo [c], Antonio S. Cofiño [a]

[a] *Department of Applied Mathematics and Computational Sciences, University of Cantabria, Spain*
[b] *Department of Applied Mathematics, University of Castilla-La Mancha, Spain*
[c] *Department of Civil Engineering, University of Castilla-La Mancha, Spain*

Received 25 June 2006; accepted 16 April 2007
Available online 21 April 2007

Abstract

It is well known that the least absolute value (ℓ_∞) and the least sum of absolute deviations (ℓ_1) algorithms produce estimators that are not necessarily unique. In this paper it is shown how the set of all solutions of the ℓ_1 and ℓ_∞ regression problems for moderately large sample sizes can be obtained. In addition, if the multiplicity of solutions wants to be avoided, two new methods giving the same optimal ℓ_1 and ℓ_∞ values, but supplying unique solutions, are proposed. The idea consists of using two steps: in the first step the optimal values of the ℓ_1 and ℓ_∞ errors are calculated, and in the second step, in case of non-uniqueness of solutions, one of the multiple solutions is selected according to a different criterion. For the ℓ_∞ the procedure is used sequentially but removing, in each iteration, the data points with maximum absolute residual and adding the corresponding constraints for keeping these residuals, and this process is repeated until no change in the solution is obtained. In this way not only the maximum absolute residual values are minimized in the modified method, but also the maximum absolute residual values of the remaining points sequentially, until no further improvement is possible. In the ℓ_1 case a least squares criterion is used but restricted to the ℓ_1 residual condition. Thus, in the modified ℓ_1 method not only the ℓ_1 residual is minimized, but also the sum of squared residuals subject to the ℓ_1 residual. The methods are illustrated by their application to some well known examples and their performances are tested by some simulations, which show that the lack of uniqueness problem cannot be corrected for some experimental designs by increasing the sample size.

Keywords: Least squares; ℓ_1-norm; ℓ_∞-norm; Parameter estimation; Set of all solutions; Asymptotic properties

1. Introduction

Consider the regression model

$$y = f(\boldsymbol{x};\boldsymbol{\beta}) + \varepsilon, \tag{1}$$

* Corresponding author. Tel.: +34 942201722; fax: +34 942201703.
E-mail address: castie@unican.es (E. Castillo).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2007.04.020

where y is the response variable, $\boldsymbol{x}$ is the vector of p predictor variables, $\boldsymbol{\beta}$ is the parameter vector, ε is the error term, and we assume that the Jacobian determinant $|J| = |\partial f(\boldsymbol{x};\boldsymbol{\beta})/\partial\boldsymbol{\beta}|$ is non-null. If $f(x;\boldsymbol{\beta}) = \boldsymbol{X}\boldsymbol{\beta}$ with $\boldsymbol{X}$ the design matrix, we have the linear regression model.

Several estimation methods are used for estimating the parameters of linear and non-linear regression models of the form (1). Among them, the least squares (ℓ_2), the least sum of absolute deviations (ℓ_1) and the least absolute value (ℓ_∞) are the most common (see, for example, Laplace, 1789; Meketon, 1986; Belsley et al., 1980; Bloomfield and Steiger, 1980; Dodge, 1987, 1992, 1997; Dodge and Falconer, 2002; Rao and Toutenburg, 1999 or Chatterjee et al., 2000 for recent papers).

The most popular method for estimating the regression parameters $\boldsymbol{\beta}$ of the models in (1) is the ℓ_2 method, where the sum of squared distances between observed and predicted values is minimized, that is,

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \quad Z_{\ell_2} = \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta}))^2. \tag{2}$$

Though ℓ_1 and ℓ_∞ methods had initially a great success, they were obscured by the appearance of the ℓ_2 method. Later they recovered some prestige (see Edgeworth, 1887, 1888), when it was discovered that they correspond to maximum likelihood estimators for the uniform and double exponential residuals, respectively, and gave iterative methods for finding the solution, but soon they returned to obscurity mainly due to their associated computational complexities.

A posterior prestige recovery of these methods and a more frequent use of them took place because of the important contribution of Mosteller et al. (1950), who discovered the possibility of stating these problems as linear programming problems. Since then, the advances of mathematical programming were applied to these regression problems and many new results appeared. Recently, Portnoy and Koenker (1997) have shown the interesting result that there are algorithms that make them competitive with the ℓ_2 method, and even superior for some sample sizes.

One important property of the ℓ_1 and ℓ_∞ methods is that they are less sensitive to extreme errors (outliers) than the ℓ_2 method, as already pointed out by Bowditch (see Eisenhart, 1961). Some interesting sensitivity measures are given, for example, in Chatterjee and Hadi (1988), and a recent sensitivity analysis of the three regression methods has been presented by Castillo et al. (2004).

All three regression models can be seen as particular cases of the weighted regression model

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \quad Z = \sum_{i=1}^{n} w_i |y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta})|, \tag{3}$$

where the weights are $w_i = |y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta})|$ for the least squares, $w_i = 1$ if $|y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta})| = \max_j |y_j - f(\boldsymbol{x}_j;\boldsymbol{\beta})|$ and $w_i = 0$, otherwise, for the ℓ_∞ method, and $w_i = 1$ for the ℓ_1 method. Thus, the least squares method gives much more weight to large residuals, the ℓ_∞ method gives only weight to the maximum residual, and the ℓ_1 method gives equal weight to all residuals. This immediately suggests when each method should be used in a particular application.

1.1. The ℓ_1 regression method

In the ℓ_1 regression problem, the sum of absolute residuals is minimized, i.e.

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \quad Z_{\ell_1} = \sum_{i=1}^{n} |y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta})|. \tag{4}$$

The ℓ_1 method is very old. The first idea seems to be attributed to Boscovich in 1760 (see Stigler, 1984, 1986), who added to (4) the condition of the regression line to pass through the centroid of the mass of points. However, the first written solution to this problem is due to Laplace (1789).

One century later, Edgeworth (1887, 1888) removed the constraint and stated the unconstrained problem. So, we must give the credit for the first proposal of the ℓ_1 method to Edgeworth. He also proposed a numerical method for solving problem (4). However, as indicated by Hawley and Gallagher (1994), his method cycles when the data have some special degeneracies. Other efficient methods for solving this problem were given