Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 2531–2541
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Algorithms for bounded-influence estimation☆

Ruggero Bellio*

Department of Statistics, University of Udine, Via Treppo 18, 33100 Udine, Italy

Received 21 November 2005; received in revised form 15 September 2006; accepted 15 September 2006
Available online 10 October 2006

Abstract

Bounded-influence estimation is a well developed and useful theory. It provides fairly efficient estimators which are robust to outliers and local model departures. However, its use has been limited thus far, mainly because of computational difficulties. A careful implementation in modern statistical software can effectively overcome the numerical problems of bounded-influence estimators. The proposed approach is based on general methods for solving estimating equations, together with suitable methods developed in the statistical literature, such as the delta algorithm and nested iterations. The focus is on Mallows estimation in generalized linear models and on optimal bias-robust estimation in models for independent data, such as regression models with asymmetrically distributed errors.

Keywords: Estimating equation; Numerical algorithm; Optimal bias-robust estimator; Regression model; Robustness

1. Introduction

The theory of bounded-influence estimation provides robust estimators with fairly good efficiency. The theory was originally proposed by Hampel (1974), and further developed by Hampel and co-workers (Hampel et al., 1986). See also Carroll and Ruppert (1988, Chapter 6) and Ronchetti (1997).

Bounded-influence estimation is a potentially useful tool in applied work, for example as a diagnostic method in maximum likelihood estimation; see the discussion given in Carroll and Ruppert (1988, Chapter 6). Nonetheless, its use in applications has been rather limited thus far. A major obstacle are the serious numerical difficulties which hamper the computation of bounded-influence estimates. These are for instance mentioned in Stefanski et al. (1986), Carroll and Ruppert (1988, Section 6.3), Dupuis and Field (1998) and Mills et al. (2002). As a result, the implementation of these methods within commonly used statistical software has been restricted to a few special cases.

The numerical difficulties involved when obtaining bounded-influence estimates are partly due to the fact that these are typically obtained by solving suitable estimating equations, rather than from the optimization of an objective function. As known from numerical analysis, the resolution of nonlinear equations is generally more difficult than function optimization. Furthermore, the equations can be very complicated, making the problem burdensome.

The main point of this paper is to show how the availability of increasing computing power and modern statistical software enables one to successfully handle the numerical difficulties posed by bounded-influence estimation. In partic-

☆ This work was supported by *Ministero dell'Università e della Ricerca*.
* Tel.: +39 0432 249574; fax: +39 0432 249595.
E-mail address: ruggero.bellio@dss.uniud.it.

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.09.013

ular, Mallows estimation for generalized linear models and optimal bias-robust estimation for models with independent data are considered.

For generalized linear models, the delta algorithm (Jørgensen, 1984) is found to be quite effective for Mallows estimation. Its implementation is not too difficult, while the gain in computational stability seems remarkable with respect to competing algorithms. In the case of optimal bias-robust estimation, a modification to commonly used algorithms is proposed. The key idea is to use nested iterations (Smyth, 1996), which improve the convergence properties of numerical algorithms.

The paper is organized as follows. Section 2 presents a brief review of the most common methods for solving numerical equations. Section 3 explores the application of these methods to the estimator defined in Cantoni and Ronchetti (2001) for generalized linear models. Section 4 deals with the computation of the optimal bias-robust estimator in models for independent data. The algorithm is then applied to a regression problem with skew-normal errors. The final discussion follows in Section 5. Some details on the software used for the computations are given in Appendix A.

2. Numerical methods for solving nonlinear equations

In broad generality, given the data y for a model with a p-dimensional parameter θ, the aim is to solve the nonlinear equation

$$g(\theta; y) = 0. \tag{1}$$

Note that g has the same dimension of θ, so that (1) is actually a convenient way for denoting the set of equations $g_j(\theta_1, \ldots, \theta_p; y) = 0$, $j = 1, \ldots, p$. It is not assumed that g is the gradient of an objective function, as this is not the case for many estimating functions in robust statistics.

In the following, the attention will be restricted to a few selected methods for solving nonlinear equations, which are among the most commonly used by practitioners. Details can be found in specialized texts, such as Fletcher (1980), Gill et al. (1981), Dennis and Schnabel (1983), or, for more statistically oriented readers, Lange (1999) and Monahan (2001).

(i) *Newton–Raphson algorithm*: The basic method for solving (1) is the Newton–Raphson algorithm. If $\tilde{\theta}_k$ is the value of θ at the kth iteration and $J(\theta; y)$ the Jacobian of $-g(\beta; y)$, the update at step $k + 1$ can be written as

$$\tilde{\theta}_{k+1} = \tilde{\theta}_k + b_k, \tag{2}$$

where b_k is the Newton direction, $b_k = J(\tilde{\theta}_k; y)^{-1} g(\tilde{\theta}_k; y)$. If the computation of J is too demanding, finite-difference approximations or other approximations to J are usually employed (Dennis and Schnabel, 1983, Chapter 8). A variation of the method, specifically developed for solving estimating equations, is given by the Fisher-scoring algorithm. It replaces $J(\theta; y)$ with its (approximate) expected value.

(ii) *Levenberg–Marquandt algorithm*: Least-squares algorithms try to solve (1) by finding solutions to the minimization problem

$$\min_{\theta} F(\theta; y) = \sum_{j=1}^{p} g_j(\theta; y)^2, \tag{3}$$

using a suitable method for function optimization.

Least-squares algorithms are popular among statisticians, for the main reason that code for function optimization is much more widely available than code for nonlinear equations. However, some care is generally required, as optimization algorithms may converge to local minima of $F(\theta; y)$ that do not solve (1); see for example Fletcher (1980, p. 100).

A popular least-squares method is the Levenberg–Marquandt algorithm (e.g., Fletcher, 1980, Chapter 5), which tries to minimize (3) by taking an update that uses only g and J,

$$\tilde{\theta}_{k+1} = \tilde{\theta}_k + \{J(\tilde{\theta}_k; y)^{\mathrm{T}} J(\tilde{\theta}_k; y) + \alpha_k I_p\}^{-1} J(\tilde{\theta}_k; y)^{\mathrm{T}} g(\tilde{\theta}_k; y),$$

where I_p is the p-dimensional identity matrix and α_k is a suitable scalar value.

(iii) *Powell's hybrid method*: Powell's hybrid method is a modification of the Newton–Raphson algorithm. It checks whether a Newton step (2) reduces the value of $F(\tilde{\theta}_k; y)$ in (3), otherwise suitable modifications of the update direction