Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®

Computational Statistics & Data Analysis 44 (2003) 273–295

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Implementing the Bianco and Yohai estimator for logistic regression

Christophe Croux[a], Gentiane Haesbroeck[b,*]

[a] *Department of Applied Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium*
[b] *Department of Mathematics, University of Liège (B37), Grande Traverse 12, B-4000 Liège, Belgium*

Received 31 July 2002; received in revised form 19 February 2003

Abstract

A fast and stable algorithm to compute a highly robust estimator for the logistic regression model is proposed. A criterium for the existence of this estimator at finite samples is derived and the problem of the selection of an appropriate loss function is discussed. It is shown that the loss function can be chosen such that the robust estimator exists if and only if the maximum likelihood estimator exists. The advantages of using a weighted version of this estimator are also considered. Simulations and an example give further support for the good performance of the implemented estimators.

Keywords: Robust estimation; Influence function; Logistic regression; Maximum likelihood

1. Introduction

Let Y_i, $1 \leqslant i \leqslant n$, be independent Bernoulli variables whose success probabilities depend on the values of p-dimensional explanatory variables $X_1, \ldots, X_n$ through the relation

$$\mathbb{P}(Y_i = 1 | X_i = x_i) = F(\alpha + \beta^t x_i),$$

where F is a strictly increasing cumulative distribution function. Taking $F(u) = 1/(1 + \exp(-u))$ results in the logit model, which is the model we will consider in this paper.

* Corresponding author. Tel.: +32-4-3669594; fax: +32-4-3669547.
E-mail addresses: christophe.croux@econ.kuleuven.ac.be (C. Croux), g.haesbroeck@ulg.ac.be (G. Haesbroeck).

0167-9473/03/$ - see front matter
doi:10.1016/S0167-9473(03)00042-2

To simplify notation, we will use $\gamma = (\alpha, \beta^t)^t$ and $z_i = (1, x_i^t)^t$ for all $1 \leqslant i \leqslant n$. An estimator for γ computed from the sample $X_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ is denoted $\hat{\gamma}_n$. The maximum likelihood (ML) estimator $\hat{\gamma}_n^{\mathrm{ML}}$ is defined as

$$\hat{\gamma}_n^{\mathrm{ML}} = \operatorname*{argmax}_{\gamma} \log L(\gamma; X_n) = \operatorname*{argmin}_{\gamma} \sum_{i=1}^{n} d(z_i^t \gamma; y_i), \tag{1.1}$$

where $\log L(\gamma; Z_n)$ is the conditional log-likelihood function calculated in γ and $d(z_i^t\gamma; y_i)$ is the deviance component given by

$$d(z_i^t\gamma; y_i) = -y_i \log F(z_i^t\gamma) - (1 - y_i) \log\{1 - F(z_i^t\gamma)\}.$$

The ML estimator is the most efficient estimator (asymptotically), but it may behave very poorly in presence of outliers. Therefore, robust alternatives need to be constructed.

In this paper, focus is on a generalization of (1.1) which consists of replacing the function d by another one. The estimator $\hat{\gamma}_n$ of interest is therefore defined by

$$\hat{\gamma}_n = \operatorname*{argmin}_{\gamma} \sum_{i=1}^{n} \varphi(z_i^t\gamma; y_i), \tag{1.2}$$

where φ is a positive and almost everywhere differentiable function. It needs to satisfy $\varphi(s; 0) = \varphi(-s; 1)$ for any score s, where a score value $s_i = z_i^t\gamma$ is obtained as a linear combination of a given parameter vector γ. Instead of working with φ, we will most of the time use the univariate function $\phi(s) = \varphi(s; 0)$. The value $\phi(s)$ corresponding to an observation with $y = 0$ gives the impact of a particular score s on the value of the objective function in (1.2). This function is assumed to be nondecreasing since a large positive score s should not be attributed to observations having a null y value and should therefore receive a larger weight in the function to minimize. We further require that $\lim_{s\to-\infty} \phi(s) = 0$, implying that a large negative score, what we expect for a value $y = 0$, is not contributing to the objective function in (1.2).

As (1.2) shows, the estimator $\hat{\gamma}_n$ belongs to the class of M-type estimators and the associated first-order condition is given by

$$\frac{1}{n} \sum_{i=1}^{n} \Psi(z_i^t\gamma; y_i) z_i = 0, \tag{1.3}$$

where $\Psi(s; 0) = \partial\varphi(s; 0)/\partial s$ and $\Psi(s; 1) = -\Psi(-s; 0)$. Notice that these first-order conditions may have multiple solutions and we need to use the value of the objective function in (1.2) to select the final estimator. Throughout the paper we will use the notation $\psi(s) = \Psi(s; 0) = \phi'(s)$.

Particular cases of (1.2) have already been considered in the literature. First, the ML estimator obviously belongs to this class of M-estimators with $\phi_{\mathrm{ML}}(s) = -\ln(1 - F(s))$. A more robust proposal is due to Pregibon (1982) who suggested an estimator defined by

$$\hat{\gamma}_n = \operatorname*{argmin}_{\gamma} \sum_{i=1}^{n} \lambda(d(z_i^t\gamma; y_i)),$$

where λ is a strictly increasing Huber's type function. This estimator was designed to give less weight to observations poorly accounted for by the model but did not