Computational Statistics and Data Analysis 83 (2015) 262–274

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

High finite-sample efficiency and robustness based on distance-constrained maximum likelihood

Ricardo A. Maronna [a,*], Victor J. Yohai [b]

[a] Department of Mathematics, School of Exact Sciences, National University of La Plata, Argentina

[b] Department of Mathematics, School of Exact and Natural Sciences, University of Buenos Aires and CONICET, Argentina

ARTICLE INFO

Article history:
Received 29 November 2013
Received in revised form 7 October 2014
Accepted 8 October 2014
Available online 22 October 2014

Keywords:
Finite-sample efficiency
Robust regression
Robust multivariate location and scatter
Kullback–Leibler divergence

ABSTRACT

Good robust estimators can be tuned to combine a high breakdown point and a specified asymptotic efficiency at a central model. This happens in regression with MM- and τ-estimators among others. However, the finite-sample efficiency of these estimators can be much lower than the asymptotic one. To overcome this drawback, an approach is proposed for parametric models, which is based on a distance between parameters. Given a robust estimator, the proposed one is obtained by maximizing the likelihood under the constraint that the distance is less than a given threshold. For the linear model with normal errors, simulations show that the proposed estimator attains a finite-sample efficiency close to one while improving the robustness of the initial estimator. The same approach also shows good results in the estimation of multivariate location and scatter.

1. Introduction

Since the seminal work of Huber (1964) and Hampel (1971), one of the main concerns of the research in robust statistics has been to derive statistical procedures that are simultaneously highly robust and highly efficient under the assumed model. The efficiency of an estimator is usually measured by the asymptotic efficiency, that is, by the ratio between the asymptotic variances of the maximum likelihood estimator (henceforth MLE) and of the robust estimator. However if the sample size n is not very large, this asymptotic efficiency may be quite different from the finite sample size one, defined as the ratio between the mean squared errors (MSE) of the MLE and of the robust estimator, for samples of size n. However, it is obvious that for practical purposes only the finite sample size efficiency matters.

Consider for example the case of a linear model with normal errors. In this case the MLE of the regression coefficients is the least squares estimator (LSE). It is well known that this estimator is very sensitive to outliers, and in particular its breakdown point is zero. To overcome this problem, several estimators combining high asymptotic breakdown point and high efficiency have been proposed. Yohai (1987) proposed MM-estimators, which have 50% breakdown point and asymptotic efficiency as close to one as desired. Yohai and Zamar (1988) proposed τ-estimates, which combine the same two properties as MM-estimators. Gervini and Yohai (2002) proposed regression estimators which simultaneously have 50% breakdown point and asymptotic efficiency equal to one.

However, as will be seen in Section 2.1, when n is not very large the finite sample efficiency of these estimators may be much smaller than the asymptotic one. On the other hand, a 50% breakdown point does not guarantee that the estimator is

* Correspondence to: Departamento de Matemática, Facultad de Ciencias Exactas, C.C. 172, La Plata 1900, Argentina.
E-mail address: rmaronna@retina.ar (R.A. Maronna).

http://dx.doi.org/10.1016/j.csda.2014.10.015
0167-9473/

highly robust. In fact, this only guarantees that given $\varepsilon < 0.5$ there exists $K(\varepsilon)$ such that if the data are contaminated with a fraction of outliers smaller than ε, the norm of the difference between the estimator and the true value is smaller than $K(\varepsilon)$. However $K(\varepsilon)$ may be very large, which makes the estimator unstable under outlier contamination of size ε.

Bondell and Stefanski (2013) proposed a regression estimator with maximum breakdown point and high finite-sample efficiency. However, as it will be seen in Section 2.1, the price for this efficiency is a serious loss of robustness.

An alternative approach to robust estimation is proposed by Olive and Hawkins (2010, 2011); see also Zhang et al. (2012). Their estimators are consistent and have high breakdown point, but since they are not equivariant, comparisons with them are difficult.

The purpose of this paper is to present estimators which have a high finite sample size efficiency and robustness even for small n. Besides, these estimators are highly robust using a robustness criterion better than the breakdown point, namely, the maximum MSE for a given contamination rate ε.

The procedure to define the proposed estimators is very general and may be applied to any parametric or semiparametric model. However in this paper the details are given only to estimate the regression coefficients in a linear model and the multivariate location and scatter of a random vector.

To define the proposed estimators we need an initial robust estimator, not necessarily with high finite sample efficiency. Then the estimators are defined by maximizing the likelihood function subject to the estimate being sufficiently close to the initial one. Doing so we can expect that the resulting estimator will have the maximum possible finite sample efficiency under the assumed model compatible with proximity to the initial robust estimator. This proximity guarantees the robustness of the new estimator.

The formulation of our proposal is as follows. Let D be a distance or discrepancy measure between densities. As a general notation, given a family of distributions with observation vector $\mathbf{z}$, parameter vector $\boldsymbol{\theta}$ and density $f(\mathbf{z}, \boldsymbol{\theta})$, put $d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = D(f(\mathbf{z}, \boldsymbol{\theta}_1), f(\mathbf{z}, \boldsymbol{\theta}_2))$. Let $\mathbf{z}_i$, $i = 1, \ldots, n$ be i.i.d. observations with distribution $f(\mathbf{z}, \boldsymbol{\theta})$, and let $\widehat{\boldsymbol{\theta}}_0$ be an initial robust estimator. Call $L(\mathbf{z}_1, \ldots, \mathbf{z}_n; \boldsymbol{\theta})$ the likelihood function. Then our proposal is to define an estimator $\widehat{\boldsymbol{\theta}}$ as

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\mathbf{z}_1, \ldots, \mathbf{z}_n; \boldsymbol{\theta}) \quad \text{with } d\left(\widehat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}\right) \leq \delta \tag{1}$$

where δ is an adequately chosen constant that may depend on n. We shall call this proposal "distance-constrained maximum likelihood' (DCML for short).

Several dissimilarity measures, such as the Hellinger distance, may be employed for this purpose. We shall employ as D the Kullback–Leibler (KL) divergence, because, as it will be seen, it yields easily manageable results. Therefore the d in (1) will be

$$d_{\mathrm{KL}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \int_{-\infty}^{\infty} \log\left(\frac{f(\mathbf{z}, \boldsymbol{\theta}_1)}{f(\mathbf{z}, \boldsymbol{\theta}_2)}\right) f(\mathbf{z}, \boldsymbol{\theta}_1)\, d\mathbf{z}.$$

In Sections 2 and 3 we apply this procedure to the linear model and to the estimation of multivariate location and scatter, respectively. In Section 4 the proposed estimators are applied to two data sets. Finally Section 5 summarizes the results.

2. Regression

Consider the family of distributions with $\mathbf{z} = (\mathbf{x}, y)$, with $\mathbf{x} \in R^p$ and $y \in R$, satisfying the model $y = \mathbf{x}'\boldsymbol{\beta} + \sigma u$, where $u \sim \mathrm{N}(0, 1)$ is independent of $\mathbf{x} \in R^p$. Here $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$. Let $\widehat{\boldsymbol{\theta}}_0 = \left(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0\right)$ be an initial robust estimator of regression and scale. We will actually consider σ as a nuisance parameter, and therefore we have

$$d_{\mathrm{KL}}\left(\boldsymbol{\beta}_0, \boldsymbol{\beta}\right) = \frac{1}{\sigma^2}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right)' \mathbf{C}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right) \tag{2}$$

with $\mathbf{C} = \mathrm{E}\mathbf{x}\mathbf{x}'$.

Here we replace σ with its estimator $\widehat{\sigma}_0$. The natural estimator of $\mathbf{C}$ would be $\widehat{\mathbf{C}} = n^{-1}\mathbf{X}'\mathbf{X}$, where $\mathbf{X}$ is the $n \times p$ matrix with rows $\mathbf{x}_i'$. Since it is not robust, we will employ a robust version thereof. Put for $\boldsymbol{\beta} \in R^p$, $r_i(\boldsymbol{\beta}) = y_i - \mathbf{x}'\boldsymbol{\beta}$, the residuals from $\boldsymbol{\beta}$. All "smooth" robust regression estimators, like S-estimators (Rousseeuw and Yohai, 1984), MM- and τ-estimators satisfy the estimating equations of an M-estimator, which can be written as weighted normal equations, namely

$$\sum_{i=1}^{n} W\left(\frac{r_i(\boldsymbol{\beta})}{\widehat{\sigma}_0}\right)\mathbf{x}_i r_i(\boldsymbol{\beta}) = \mathbf{0}, \tag{3}$$

where W is a "weight function". Then we define, as in Yohai et al. (1991)

$$\mathbf{C}_w = \frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i \mathbf{x}_i \mathbf{x}_i', \tag{4}$$