Journal of Multivariate Analysis 102 (2011) 20–36

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Dual divergence estimators and tests: Robustness results

Aida Toma [a,b,*], Michel Broniatowski [c]

[a] Mathematics Department, Academy of Economic Studies, Piaţa Romană 6, Bucharest, Romania

[b] "Gheorghe Mihoc – Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Calea 13 Septembrie 13, Bucharest, Romania

[c] LSTA, Université Paris 6, 175 Rue du Chevaleret, 75013 Paris, France

ARTICLE INFO

Article history:
Received 17 November 2007
Available online 1 August 2010

AMS subject classifications:
62F03
62F05
62F10
62F35

Keywords:
Location model
Minimum divergence estimators
Robust estimation
Robust test
Scale model

ABSTRACT

The class of dual ϕ-divergence estimators (introduced in Broniatowski and Keziou (2009) [5]) is explored with respect to robustness through the influence function approach. For scale and location models, this class is investigated in terms of robustness and asymptotic relative efficiency. Some hypothesis tests based on dual divergence criteria are proposed and their robustness properties are studied. The empirical performances of these estimators and tests are illustrated by Monte Carlo simulation for both non-contaminated and contaminated data.

1. Introduction

Minimum divergence estimators and related methods have received considerable attention in statistical inference because of their ability to reconcile efficiency and robustness. Among others, Beran [3], Tamura and Boos [24], Simpson [22,23] and Toma [25] proposed families of parametric estimators minimizing the Hellinger distance between a nonparametric estimator of the observations density and the model. They showed that those estimators are both asymptotically efficient and robust. Generalizing earlier work based on the Hellinger distance, Lindsay [18], Basu and Lindsay [2], and Morales et al. [19] have investigated minimum divergence estimators, for both discrete and continuous models. Some families of estimators based on approximate divergence criteria have also been considered; see Basu et al. [1].

Broniatowski and Keziou [5] have introduced a new minimum divergence estimation method based on a dual representation of the divergence between probability measures. Their estimators are defined in a unified way for both continuous and discrete models. They do not require any prior smoothing and include the classical maximum likelihood estimators as a benchmark. A special case for the Kullback–Leibler divergence is presented in Broniatowski [4]. The present paper presents robustness studies for the classes of estimators generated by the minimum dual ϕ-divergence method, as well as for some tests based on corresponding estimators of the divergence criterion.

We give general results that allow us to identify robust estimators in the class of dual ϕ-divergence estimators. We apply this study to the Cressie–Read divergences and state explicit robustness results for scale models and location models. Gains in robustness are often paid for by some loss in efficiency. This is discussed for some scale and location models. Our

* Corresponding author at: Mathematics Department, Academy of Economic Studies, Piaţa Romană 6, Bucharest, Romania.
E-mail addresses: aida_toma@yahoo.com (A. Toma), michel.broniatowski@upmc.fr (M. Broniatowski).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2010.07.010

main remarks are as follows. All the relevant information pertaining to the model and the true value of the parameter to be estimated should be used in order to define, when possible, robust and fairly efficient procedures. Some models allow for such procedures. The example provided by the scale normal model shows that the choice of a good estimation criterion is heavily dependent on the acceptable loss in efficiency, for achieving a compromise with the robustness requirement. When sampling under the model is overspread (typically for Cauchy and logistic models), not surprisingly the maximum likelihood estimator is both efficient and robust and therefore should be preferred (see Section 3.2).

On the other hand, these estimation results constitute the premises for constructing some robust tests. The purpose of robust testing is twofold. First, the level of a test should be stable under small arbitrary departures from the null hypothesis (i.e. robustness of validity). Second, the test should have a good power under small arbitrary departures from specified alternatives (i.e. robustness of efficiency). To control the test stability against outliers in the aforementioned senses, we compute the asymptotic level of the test under a sequence of contaminated null distributions, as well as the asymptotic power of the test under a sequence of contaminated alternatives. These quantities are seen to be controlled by the influence function of the test statistic. In this way, the robustness of the test is a consequence of the robustness of the test statistic based on a dual ϕ-divergence estimator. In many cases, this requirement is met when the dual ϕ-divergence estimator itself is robust.

The paper is organized as follows. In Section 2 we present the classes of estimators generated by the minimum dual ϕ-divergence method. In Section 3, for these estimators, we compute the influence functions and give the Fisher consistency. We particularize this study for the Cressie–Read divergences and state robustness results for scale models and location models. Basic examples (normal scale with known location, Cauchy or logistic location under known scale) are handled in order to shed light on robustness properties of the estimates. Section 4 is devoted to hypothesis testing. We give general convergence results for contaminated observations and use them to compute the asymptotic level and the asymptotic power for the tests that we propose. In Section 5, the performances of the estimators and tests are illustrated by Monte Carlo simulation studies. In Section 6 we briefly present a proposal for the adaptive choice of tuning parameters.

All the proofs are included in the technical report [27].

2. Minimum divergence estimators

2.1. Minimum divergence estimators

Let φ be a non-negative convex function defined from $(0, \infty)$ onto $[0, \infty]$ and satisfying $\varphi(1) = 0$. Also extend φ to 0, defining $\varphi(0) = \lim_{x\downarrow 0} \varphi(x)$. Let $(\mathcal{X}, \mathcal{B})$ be a measurable space and P be a probability measure (p.m.) defined on $(\mathcal{X}, \mathcal{B})$. Following Rüschendorf [21], for any p.m. Q absolutely continuous (a.c.) w.r.t. P, the ϕ-divergence between Q and P is defined by

$$\phi(Q, P) := \int \varphi\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\right) \mathrm{d}P. \tag{1}$$

When Q is not a.c. w.r.t. P, we set $\phi(Q, P) = \infty$. We refer the reader to Liese and Vajda [16] for an overview on the origin of the concept of divergence in statistics.

A commonly used family of divergences is the so-called "power divergences", introduced by Cressie and Read [9] and defined by the class of functions

$$x \in \mathbb{R}_+^* \mapsto \varphi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)} \tag{2}$$

for $\gamma \in \mathbb{R} \setminus \{0, 1\}$ and $\varphi_0(x) := -\log x + x - 1$, $\varphi_1(x) := x \log x - x + 1$ with $\varphi_\gamma(0) = \lim_{x\downarrow 0} \varphi_\gamma(x)$, $\varphi_\gamma(\infty) = \lim_{x\to\infty} \varphi_\gamma(x)$, for any $\gamma \in \mathbb{R}$. The Kullback–Leibler divergence (KL) is associated with φ_1, the modified Kullback–Leibler divergence (KL_m) with φ_0, the χ^2 divergence with φ_2, the modified χ^2 divergence (χ_m^2) with φ_{-1} and the Hellinger distance with $\varphi_{1/2}$.

Let $\{P_\theta : \theta \in \Theta\}$ be some identifiable parametric model with Θ a subset of $\mathbb{R}^d$. Consider the problem of estimation of the unknown true value of the parameter θ_0 on the basis of an i.i.d. sample $X_1, \ldots, X_n$ with p.m. P_{θ_0}.

When all p.m. P_θ share the same finite support S which is independent of the parameter θ, the ϕ-divergence between P_θ and P_{θ_0} has the form

$$\phi(P_\theta, P_{\theta_0}) = \sum_{j\in S} \varphi\left(\frac{P_\theta(j)}{P_{\theta_0}(j)}\right) P_{\theta_0}(j).$$

For this case, Liese and Vajda [17], Lindsay [18] and Morales et al. [19] investigated the so-called "minimum ϕ-divergence estimators" (minimum disparity estimators in [18]) of the parameter θ_0 defined by

$$\widehat{\theta}_n := \arg\inf_{\theta\in\Theta} \phi(P_\theta, P_n), \tag{3}$$