C. R. Acad. Sci. Paris, Ser. I 351 (2013) 311–316

Contents lists available at SciVerse ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics

Robustness of dual divergence estimators for models satisfying linear constraints

Robustesse des estimateurs par divergence duale pour des modèles satisfaisant des contraintes linéaires

Aida Toma [a,b]

[a] *Department of Applied Mathematics, Bucharest Academy of Economic Studies, Piața Romană 6, Bucharest, Romania*
[b] *"Gh. Mihoc–C. Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romanian Academy, Calea 13 Septembrie 13, Bucharest, Romania*

ARTICLE INFO

Article history:
Received 20 November 2011
Accepted after revision 12 February 2013
Available online 3 May 2013

Presented by the Editorial Board

ABSTRACT

We consider new classes of estimators and test statistics for models satisfying linear constraints with unknown parameter. These procedures are based on minimization of divergences through duality techniques. We prove that, for various divergences, the new approach provides robust estimation and test procedures, unlike the empirical likelihood method. We give general results using the influence function approach, which we exemplify in detail in the case of the Cressie–Read divergences. It is found that the Hellinger distance is one of the divergences that leads to robust procedures.

RÉSUMÉ

Nous considérons de nouvelles classes d'estimateurs et de procédures de test pour des modèles satisfaisant des contraintes linéaires à paramètre inconnu. Ces procédures sont basées sur la minimisation des divergences grâce à des techniques de dualité. Nous prouvons que, pour de nombreuses divergences, la nouvelle approche fournit des estimateurs et des tests robustes, contrairement à la méthode de vraisemblance empirique. Nous donnons des résultats généraux en utilisant l'approche par fonction d'influence, que nous illustrons en détail dans le cas des divergences de Cressie–Read. On remarque que la distance de Hellinger conduit à des procédures robustes.

Version française abrégée

Un modèle satisfaisant des contraintes linéaires est une famille $\mathcal{M}^1$ de mesures de probabilité, toutes définies sur le même espace mesurable $(\mathcal{X},\mathcal{B})$ telle que $\int g(x,\theta)\,\mathrm{d}Q(x)=0$, pour toute mesure $Q\in\mathcal{M}^1$. Le paramètre inconnu θ appartient à $\Theta\subset\mathbb{R}^d$. La fonction $g:=(g_1,\dots,g_l)^t$ est définie sur $\mathcal{X}\times\Theta$, chaque fonction g_i étant à valeurs réelles et les fonctions $g_1,\dots,g_l,\mathbf{1}_{\mathcal{X}}$ sont suposées être linéairement indépendantes. Nous notons par M^1 l'ensemble de toutes les mesures de probabilité sur $(\mathcal{X},\mathcal{B})$ et

$$\mathcal{M}_\theta^1:=\left\{Q\in M^1\colon\ \int g(x,\theta)\,\mathrm{d}Q(x)=0\right\}$$

E-mail address: aida_toma@yahoo.com.

1631-073X/$ – see front matter
http://dx.doi.org/10.1016/j.crma.2013.02.005

de sorte que $\mathcal{M}^1 = \bigcup_{\theta\in\Theta} \mathcal{M}^1_\theta$. Sur la base d'un échantillon i.i.d. de loi P_0 inconnue, nous voulons décider si P_0 appartient au modèle $\mathcal{M}^1$ et, si c'est le cas, trouver la valeur θ_0 pour laquelle $\int g(x,\theta)\,\mathrm{d}P_0(x) = 0$.

Pour ces modèles, Broniatowski and Keziou (2011) introduit de nouvelles procédures d'estimation et de tests par la minimisation des divergences. Ces procédures généralisent la méthode de vraisemblance empirique. Les estimateurs et les statistiques de test sont définis en utilisant une nouvelle représentation duale de divergences sous contraintes qui permet de dériver leurs distributions limites, à la fois sous l'hypothèse nulle et sous l'alternative, y compris en cas de mauvaise spécification ; ceci ne peut être atteint par l'approche classique de vraisemblance empirique. Dans le présent article, nous considérons ces classes d'estimateurs et de tests statistiques et étudions leurs propriétés de robustesse en utilisant l'approche par fonction d'influence. Lorsque la mesure P_0 appartient au modèle, les fonctions d'influence des estimateurs d'une classe donnée sont les mêmes pour toutes les divergences. Dans ce cas, le comportement des estimateurs par raport à la présence de valeurs aberrantes dans l'échantillon ne dépend pas du choix de la divergence. En revanche, lorsque $P_0 \notin \mathcal{M}_\theta$ ou lorsque $P_0 \notin \mathcal{M}$, la robustesse des estimateurs et des statistiques de test est dictée par le choix de la divergence. Dans ce cas, nous fournissons une étude détaillée pour la classe des divergences de Cressie–Read et montrons que, contrairement à KL_m (divergence qui est associée à l'approche de vraisemblance empirique classique), la divergence de Hellinger, ainsi que quelques autres appartenant à la famille des divergences de Cressie–Read, conduisent à des procédures robustes.

1. The model and existing results

The models satisfying linear constraints with unknown parameter and related estimation and testing methods have been investigated by many authors (see [11,12,4] and references therein). For such models, Broniatowski and Keziou [4] introduced new estimation and test procedures through divergence minimization. These procedures extend the empirical likelihood method and share common features with the generalized empirical likelihood approach. The estimators and test statistics are defined using a new dual representation of divergences under constraints that allows us to derive their limiting distributions, both under the null hypothesis and under alternative, including misspecification, which cannot be achieved through the classical empirical likelihood approach. The duality technique was previously implemented for estimation and testing in the case of continuous parametric models in [10,2,5,3] and for copula models in [1]. Corresponding robustness properties have been studied in [13,14,6]. In the present paper, we consider the classes of estimators and test statistics introduced by Broniatowski and Keziou in [4] in the context of models satisfying linear constraints and study their robustness properties through the influence function approach. We show that, depending on the used divergences, these statistical criteria may have robustness properties. We mention that the sensitivity of the classical empirical likelihood method in contaminated models has also been discussed by Glenn and Zhao [8], who proposed weighted empirical likelihood estimates with robustness properties superior to those of classical empirical likelihood estimates.

A model satisfying linear parametric constraints is a family $\mathcal{M}^1$ of probability measures (p.m.), all defined on the same measurable space $(\mathcal{X},\mathcal{B})$ such that $\int g(x,\theta)\,\mathrm{d}Q(x) = 0$, for all $Q \in \mathcal{M}^1$. The unknown parameter θ belongs to $\Theta \subset \mathbb{R}^d$. The function $g := (g_1,\dots,g_l)^t$ is defined on $\mathcal{X}\times\Theta$, each of the g_i's being real-valued and the functions $g_1,\dots,g_l,\mathbf{1}_\mathcal{X}$ being supposed to be linearly independent. Denote by M^1 the set of all probability measures on $(\mathcal{X},\mathcal{B})$ and

$$\mathcal{M}^1_\theta := \left\{ Q \in M^1 \colon \int g(x,\theta)\,\mathrm{d}Q(x) = 0 \right\} \tag{1}$$

so that $\mathcal{M}^1 = \bigcup_{\theta\in\Theta} \mathcal{M}^1_\theta$.

On the basis of an i.i.d. sample with unknown p.m., P_0 it is intending to decide if P_0 belongs to the model $\mathcal{M}^1$, and in the affirmative case to find the unknown true value θ_0 for which $\int g(x,\theta)\,\mathrm{d}P_0(x) = 0$.

Recently, Broniatowski and Keziou [4] developed an elegant methodology for this problem, by using dual representations of φ-divergences and projection methods in the sense of φ-divergences.

Let φ be a convex function defined on $\mathbb{R}$ and $[0,\infty]$ valued, such that $\varphi(1) = 0$ and let P be some p.m. For any signed finite measure Q absolutely continuous (a.c.) with respect to (w.r.t.) P, the φ-divergence between Q and P is defined by:

$$D_\varphi(Q,P) := \int \varphi\left(\frac{\mathrm{d}Q}{\mathrm{d}P}(x)\right)\mathrm{d}P(x). \tag{2}$$

When Q is not a.c. w.r.t. P, we set $D_\varphi(Q,P) = \infty$.

A well-known class of divergences between p.m.'s is the so called class of power divergences introduced by Cressie and Read [7] and defined by the convex functions:

$$x \in \mathbb{R}^*_+ \mapsto \varphi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma-1)} \tag{3}$$

for $\gamma \in \mathbb{R}\setminus\{0,1\}$ and $\varphi_0(x) := -\log x + x - 1$, $\varphi_1(x) := x\log x - x + 1$. The Kullback–Leibler divergence (KL) is associated with φ_1, the modified Kullback–Leibler (KL_m) with φ_0, the χ^2 divergence with φ_2, the modified χ^2 divergence (χ^2_m) with φ_{-1} and the Hellinger distance with $\varphi_{1/2}$. When φ_γ is not defined on $(-\infty,0)$ or when φ_γ is not convex, the definition of the corresponding power divergence function $Q \in M^1 \mapsto D_{\varphi_\gamma}(Q,P)$ can be extended on the whole set of signed finite measures by taking the following extension of φ_γ: $x \in \mathbb{R} \mapsto \varphi_\gamma(x)\mathbf{1}_{[0,\infty)}(x) + (+\infty)\mathbf{1}_{(-\infty,0)}(x)$.

The φ-divergence between some set Ω of signed finite measures and a p.m. P is defined by $D_\varphi(\Omega, P) = \inf_{Q\in\Omega} D_\varphi(Q, P)$. Assuming that $D_\varphi(\Omega, P)$ is finite, a measure $Q^* \in \Omega$ is called a φ-projection of P on Ω if $D_\varphi(Q^*, P) \leqslant D_\varphi(Q, P)$, for all $Q \in \Omega$.

In order to construct tests for the null hypothesis $\mathcal{H}_0 : P_0 \in \mathcal{M}^1$ against the alternative $\mathcal{H}_0 : P_0 \notin \mathcal{M}^1$, estimators of the φ-divergence between the model $\mathcal{M}^1$ and the law P_0 can be used. Let $X_1, \ldots, X_n$ be an i.i.d. sample with the distribution P_0. The "plug-in" estimator of the φ-divergence between the set $\mathcal{M}^1_\theta$ and the p.m. P_0 is defined by

$$\widehat{D}_\varphi\big(\mathcal{M}^1_\theta, P_0\big) = \inf_{Q\in\mathcal{M}^1_\theta} D_\varphi(Q, P_n) = \inf_{Q\in\mathcal{M}^1_\theta} \int \varphi\left(\frac{\mathrm{d}Q}{\mathrm{d}P_n}(x)\right)\mathrm{d}P_n(x) \tag{4}$$

where P_n is the empirical measure pertaining to the sample. If the projection of P_n on $\mathcal{M}^1_\theta$ exists, it is attained for a law a.c. w.r.t. P_n. Then it is natural to consider:

$$\mathcal{M}^{(n)}_\theta = \left\{Q \in M^1\colon\ Q \text{ a.c.w.r.t. } P_n \text{ and } \sum_{i=1}^n g(X_i, \theta) Q(X_i) = 0\right\} \tag{5}$$

for which the plug-in estimator (4) can be written as:

$$\widehat{D}_\varphi\big(\mathcal{M}^1_\theta, P_0\big) = \inf_{Q\in\mathcal{M}^{(n)}_\theta} \frac{1}{n}\sum_{i=1}^n \varphi\big(nQ(X_i)\big). \tag{6}$$

The infimum in this expression may be achieved on the frontier of the set $\mathcal{M}^{(n)}_\theta$, case in which the Lagrange method for characterizing the infimum and computing $\widehat{D}_\varphi(\mathcal{M}^1_\theta, P_0)$ is not valid.

In order to avoid this difficulty, Broniatowski and Keziou [4] have proposed to work on sets of signed finite measures and considered:

$$\mathcal{M}_\theta := \left\{Q \in M\colon \int \mathrm{d}Q = 1 \text{ and } \int g(x, \theta)\,\mathrm{d}Q(x) = 0\right\}, \tag{7}$$

where M denotes the set of all signed finite measures and $\mathcal{M} = \bigcup_\theta \mathcal{M}_\theta$. They showed that if Q^*_1, the projection of P_n on $\mathcal{M}^1_\theta$, is an interior point of $\mathcal{M}^1_\theta$, and Q^*, the projection of P_n on $\mathcal{M}_\theta$, is an interior point of $\mathcal{M}_\theta$, then both approaches based on signed finite measures, respectively on p.m.'s, coincide. In case Q^*_1 is a frontier point of $\mathcal{M}^1_\theta$, the estimator they proposed for θ_0 converges to θ_0. This justifies the substitution of $\mathcal{M}^1_\theta$ for $\mathcal{M}_\theta$.

In the following, we shortly recall the definitions of the estimators and of the test statistics proposed in [4].

For any p.m. P on $\mathcal{X}$ and any measurable real function f on $\mathcal{X}$, let Pf denote $\int f(x)\,\mathrm{d}P(x)$. For example, $P_0 g_j(\theta)$ will be used instead of $\int g_j(x, \theta)\,\mathrm{d}P_0(x)$. Denote $\bar{g}$ the function defined on $\mathcal{X} \times \Theta$ and $\mathbb{R}^{l+1}$ valued $\bar{g}(x, \theta) := (\mathbf{1}_\mathcal{X}(x), g_1(x, \theta), \ldots, g_l(x, \theta))^t$. For all $\theta \in \Theta$, denote also $\bar{g}(\theta)$, $g(\theta)$, $g_j(\theta)$ the functions defined respectively by $\bar{g}(\theta)(x) := \bar{g}(x, \theta)$, $g(\theta)(x) := g(x, \theta)$, $g_j(\theta)(x) := g_j(x, \theta)$ for all $j \in \{0, 1 \ldots, l\}$, where $g_0(x, \theta) = \mathbf{1}_\mathcal{X}(x)$.

Define $\psi(u) := u\varphi'^{-1}(u) - \varphi(\varphi'^{-1}(u))$ and for a given p.m. P,

$$\Lambda_\theta(P) := \left\{t \in \mathbb{R}^{l+1}\colon \int \left|\psi\left(t_0 + \sum_{j=1}^l t_j g_j(x, \theta)\right)\right| \mathrm{d}P(x) < \infty\right\}. \tag{8}$$

We will also use the notations Λ_θ for $\Lambda_\theta(P_0)$ and $\Lambda^{(n)}_\theta$ for $\Lambda_\theta(P_n)$.

Supposing that P_0 has a projection Q^*_θ on $\mathcal{M}_\theta$ with the same support as P_0 and that the function φ is strictly convex on its domain, then the φ-divergence $D_\varphi(\mathcal{M}_\theta, P_0)$ admits the dual representation:

$$D_\varphi(\mathcal{M}_\theta, P_0) = \sup_{t\in\Lambda_\theta} \int m(x, \theta, t)\,\mathrm{d}P_0(x) = \sup_{t\in\Lambda_\theta} P_0 m(\theta, t) \tag{9}$$

where $m(\theta, t)$ is the function defined on $\mathcal{X}$ by $m(\theta, t)(x) := m(x, \theta, t) = t_0 - \psi(t^t \bar{g}(x, \theta))$.

The supremum in (9) is unique and reached at $t_0 = c_0$, $t_1 = c_1, \ldots, t_l = c_l$, which are solutions of the system of equations:

$$\int \varphi'^{-1}\left(c_0 + \sum_{j=1}^l c_j g_j(x, \theta)\right)\mathrm{d}P_0(x) = 1 \tag{10}$$

$$\int g_j(x, \theta)\varphi'^{-1}\left(c_0 + \sum_{j=1}^l c_j g_j(x, \theta)\right)\mathrm{d}P_0(x) = 0, \quad j = 1, \ldots, l. \tag{11}$$

Denote c_θ this argsup:

$$c_\theta := \arg\sup_{t\in\Lambda_\theta} P_0 m(\theta, t). \tag{12}$$

Then, $D_\varphi(\mathcal{M}_\theta, P_0)$, c_θ, $D_\varphi(\mathcal{M}, P_0)$ and θ_0 can be estimated by:

$$\widehat{D}_\varphi(\mathcal{M}_\theta, P_0) := \sup_{t\in\Lambda_\theta^{(n)}} P_n m(\theta, t) \tag{13}$$

$$\widehat{c}_\theta := \arg\sup_{t\in\Lambda_\theta^{(n)}} P_n m(\theta, t) \tag{14}$$

$$\widehat{D}_\varphi(\mathcal{M}, P_0) := \inf_{\theta\in\Theta}\sup_{t\in\Lambda_\theta^{(n)}} P_n m(\theta, t) \tag{15}$$

$$\widehat{\theta}_\varphi := \arg\inf_{\theta\in\Theta}\sup_{t\in\Lambda_\theta^{(n)}} P_n m(\theta, t). \tag{16}$$

We refer to [4] for the complete study of the existence and of the asymptotic properties of the above estimators.

2. The influence functions of the estimators

The robustness measure that we use in the present paper is the influence function of the statistical functional corresponding to the estimator. Recall that a map T defined on a set of probability measures and parameter space valued is a statistical functional corresponding to an estimator $\widehat{\theta}_n$ of the parameter θ, if $\widehat{\theta}_n = T(P_n)$, where P_n is the empirical measure associated with the sample. The influence function of T at P_θ is defined by:

$$\mathrm{IF}(x; T, P_\theta) := \left.\frac{\partial T(\widetilde{P}_{\varepsilon x})}{\partial\varepsilon}\right|_{\varepsilon=0}, \tag{17}$$

where $\widetilde{P}_{\varepsilon x} := (1-\varepsilon)P_\theta + \varepsilon\delta_x$, δ_x being the Dirac measure putting all mass at x. Whenever the influence function is bounded with respect to x, the corresponding estimator is called robust (see [9] for details).

In the present section, we compute the influence functions of the estimators (13), (14), (15) and (16). We assume that the matrices $P_0\frac{\partial^2}{\partial^2 t}m(\theta, c_\theta)$ and $\{P_0\frac{\partial^2}{\partial t\partial\theta}m(\theta_0, c_{\theta_0})[P_0\frac{\partial^2}{\partial^2 t}m(\theta_0, c_{\theta_0})]^{-1}[P_0\frac{\partial^2}{\partial t\partial\theta}m(\theta_0, c_{\theta_0})]^t - P_0\frac{\partial^2}{\partial^2\theta}m(\theta_0, c_{\theta_0})\}$ are nonsingular and also assume that conditions allowing derivation under the integral sign are fulfilled.

For fixed θ, consider the estimator $\widehat{c}_\theta$ defined in (14). The statistical functional associated with $\widehat{c}_\theta$ is defined by:

$$T_\theta(P) := \arg\sup_{t\in\Lambda_\theta(P)} Pm(\theta, t). \tag{18}$$

Proposition 4.2. from [4] provides conditions for the existence of $T_\theta(P)$ for a given p.m. P. Since $\Lambda_\theta(P_n) = \Lambda_\theta^{(n)}$, we have $T_\theta(P_n) = \widehat{c}_\theta$. The functional T_θ satisfies $T_\theta(P_0) = c_\theta$, for each $\theta\in\Theta$.

Proposition 2.1. *For fixed θ, the influence function of the functional T_θ is given by*:

$$\mathrm{IF}(x; T_\theta, P_0) = P_0\left[\frac{g_i(\theta)g_j(\theta)}{\varphi''(\varphi'^{-1}(c_\theta^t\overline{g}(\theta)))}\right]_{i,j=0,l}^{-1}\big(1-\varphi'^{-1}\big(c_\theta^t\overline{g}(x,\theta)\big), -g_1(x,\theta)\varphi'^{-1}\big(c_\theta^t\overline{g}(x,\theta)\big), \ldots, -g_l(x,\theta)\varphi'^{-1}\big(c_\theta^t\overline{g}(x,\theta)\big)\big)^t. \tag{19}$$

As it is proved in [4], when P_0 belongs to the model $\mathcal{M}_{\theta_0}$, $c_{\theta_0} = (\varphi'(1), \underline{0}_l^t)^t$, where $\underline{0}_l$ is the null vector of size l. In this case, the formula of the influence function from Proposition 2.1 simplifies as:

$$\mathrm{IF}(x; T_{\theta_0}, P_0) = \varphi''(1)\big(0, \big(P_0 g(\theta_0)g(\theta_0)^t\big)^{-1}g(x,\theta_0)\big)^t. \tag{20}$$

We now consider the estimator $\widehat{D}_\varphi(\mathcal{M}_\theta, P_0)$ defined in (13). The corresponding statistical functional is defined by:

$$U_\theta(P) := \sup_{t\in\Lambda_\theta(P)} Pm(\theta, t) = Pm\big(\theta, T_\theta(P)\big). \tag{21}$$

Note that $U_\theta(P_0) = P_0 m(\theta, T_\theta(P_0)) = P_0 m(\theta, c_\theta) = D_\varphi(\mathcal{M}_\theta, P_0)$, for each $\theta\in\Theta$.

Proposition 2.2. *For fixed θ, the influence function of the functional U_θ is given by*:

$$\mathrm{IF}(x; U_\theta, P_0) = -D_\varphi(\mathcal{M}_\theta, P_0) + m(x, \theta, c_\theta). \tag{22}$$

When P_0 belongs to the model $\mathcal{M}_\theta$, it hold $m(x, \theta, c_\theta) = 0$ and $\mathrm{IF}(x; U_\theta, P_0) = 0$.

The statistical functional of a minimum empirical φ-divergence estimator defined by (16) is

$$T(P) = \arg\inf_{\theta\in\Theta} U_\theta(P) = \arg\inf_{\theta\in\Theta} Pm\big(\theta, T_\theta(P)\big). \tag{23}$$

Note that the target parameter is θ_0, equal to $T(P_0)$. Indeed, $T(P_0) = \arg\inf_{\theta\in\Theta} U_\theta(P_0) = \arg\inf_{\theta\in\Theta} D_\varphi(\mathcal{M}_\theta, P_0) = \theta_0$.

Proposition 2.3. *The influence function of the functional T corresponding to an estimator $\widehat{\theta}_\varphi$ is:*

$$\begin{aligned}\mathrm{IF}(x; T, P_0) = {} & \left\{P_0\frac{\partial^2}{\partial t\partial\theta}m(\theta_0, c_{\theta_0})\left[P_0\frac{\partial^2}{\partial^2 t}m(\theta_0, c_{\theta_0})\right]^{-1}\left[P_0\frac{\partial^2}{\partial t\partial\theta}m(\theta_0, c_{\theta_0})\right]^t - P_0\frac{\partial^2}{\partial^2\theta}m(\theta_0, c_{\theta_0})\right\}^{-1} \\ & \times\left\{P_0\frac{\partial^2}{\partial t\partial\theta}m(\theta_0, c_{\theta_0})\mathrm{IF}(x; T_{\theta_0}, P_0) + \frac{\partial}{\partial\theta}m(x, \theta_0, c_{\theta_0})\right\}.\end{aligned} \tag{24}$$

When P_0 belongs to the model $\mathcal{M}_{\theta_0}$,

$$\mathrm{IF}(x; T, P_0) = \left[P_0\frac{\partial}{\partial\theta}g(\theta_0)^t\big[P_0 g(\theta_0)g(\theta_0)^t\big]^{-1}P_0\frac{\partial}{\partial\theta}g(\theta_0)\right]^{-1}P_0\frac{\partial}{\partial\theta}g(\theta_0)^t\big[P_0 g(\theta_0)g(\theta_0)^t\big]^{-1}g(x, \theta_0). \tag{25}$$

The statistical functional associated with the estimator $\widehat{D}_\varphi(\mathcal{M}, P_0)$ defined in (15) is:

$$U(P) = \inf_{\theta\in\Theta} U_\theta(P) = \inf_{\theta\in\Theta} Pm\big(\theta, T_\theta(P)\big) = Pm\big(T(P), c\big(T(P), P\big)\big). \tag{26}$$

Note that $U(P_0) = \inf_{\theta\in\Theta} U_\theta(P_0) = \inf_{\theta\in\Theta} D_\varphi(\mathcal{M}_\theta, P_0) = D_\varphi(\mathcal{M}, P_0)$. When $P_0 \in \mathcal{M}$, it holds $U(P_0) = 0$.

Proposition 2.4. *The influence function of the functional U corresponding to $\widehat{D}_\varphi(\mathcal{M}, P_0)$ is:*

$$\mathrm{IF}(x; U, P_0) = -D_\varphi(\mathcal{M}, P_0) + m(x, \theta_0, c_{\theta_0}). \tag{27}$$

If P_0 belongs to the model $\mathcal{M}_{\theta_0}$, then $m(x, \theta_0, c_{\theta_0}) = 0$ and $\mathrm{IF}(x; U, P_0) = 0$.

3. The case of Cressie–Read divergences

When the measure P_0 belongs to the model, the influence functions of the estimators from a given class are the same for all divergences. Indeed, when using Cressie–Read divergences, it holds $\varphi''_\gamma(1) = 1$, which makes all the influence functions $\mathrm{IF}(x; T_\theta, P_0)$ in (20) to be the same. This is also the case of the influence functions $\mathrm{IF}(x; T, P_0)$ given by (25), which coincide, whatever the used divergence, which can be from the Cressie–Read class or not. On the other hand, $\mathrm{IF}(x; U_\theta, P_0)$ and $\mathrm{IF}(x; U, P_0)$ are identically zero. This entails that the behavior of the estimators with respect to the presence of outliers in the sample, as evaluated by the influence function, is not so influenced by the choice of the divergence.

When $P_0 \notin \mathcal{M}_\theta$ or when $P_0 \notin \mathcal{M}$, the robustness of the estimators defined by (13), (14), (15), (16) is dictated by the choice of the divergence.

For θ fixed, the influence function $\mathrm{IF}(x; T_\theta, P_0)$ in (19) is bounded w.r.t. x whenever $\frac{\partial}{\partial t}m(x, \theta, c_\theta)$ is bounded, while $\mathrm{IF}(x; U_\theta, P_0)$ given by (22) is bounded when $m(x, \theta, c_\theta)$ is bounded. In turn, $\mathrm{IF}(x; T, P_0)$ given by (24) is bounded when $\frac{\partial}{\partial t}m(x, \theta_0, c_{\theta_0})$ and $\frac{\partial}{\partial\theta}m(x, \theta_0, c_{\theta_0})$ are bounded, while $\mathrm{IF}(x; U, P_0)$ in (27) is bounded when $m(x, \theta_0, c_{\theta_0})$ is bounded.

For the Cressie–Read divergences, $m(x, \theta, c_\theta)$ is given by:

$$c_\theta^0 + \frac{1}{\gamma} - \frac{1}{\gamma}\big[(\gamma-1)c_\theta^t\overline{g}(x, \theta) + 1\big]^{\gamma/(\gamma-1)}, \quad \gamma \neq 0, 1 \tag{28}$$

$$c_\theta^0 + \log\big(1 - c_\theta^t\overline{g}(x, \theta)\big), \quad \gamma = 0 \tag{29}$$

$$c_\theta^0 - \exp\big(c_\theta^t\overline{g}(x, \theta)\big) + 1, \quad \gamma = 1 \tag{30}$$

$\frac{\partial}{\partial t}m(x, \theta, c_\theta)$ is:

$$\begin{aligned}&\Big(1 - \big[(\gamma-1)c_\theta^t\overline{g}(x, \theta) + 1\big]^{\frac{1}{\gamma-1}}, -g_1(x, \theta)\big[(\gamma-1)c_\theta^t\overline{g}(x, \theta) + 1\big]^{\frac{1}{\gamma-1}}, \ldots, \\ &\quad -g_l(x, \theta)\big[(\gamma-1)c_\theta^t\overline{g}(x, \theta) + 1\big]^{\frac{1}{\gamma-1}}\Big)^t, \quad \gamma \neq 0, 1\end{aligned} \tag{31}$$

$$\left(\frac{c_\theta^t\overline{g}(x, \theta)}{c_\theta^t\overline{g}(x, \theta) - 1}, \frac{g_1(x, \theta)}{c_\theta^t\overline{g}(x, \theta) - 1}, \ldots, \frac{g_l(x, \theta)}{c_\theta^t\overline{g}(x, \theta) - 1}\right)^t, \quad \gamma = 0 \tag{32}$$

$$\big(1 - \exp\big(c_\theta^t\overline{g}(x, \theta)\big), -g_1(x, \theta)\exp\big(c_\theta^t\overline{g}(x, \theta)\big), \ldots, -g_l(x, \theta)\exp\big(c_\theta^t\overline{g}(x, \theta)\big)\big)^t, \quad \gamma = 1 \tag{33}$$

and $\frac{\partial}{\partial\theta}m(x, \theta, c_\theta)$ expresses as:

$$-\left[c_\theta^t \frac{\partial}{\partial\theta}\bar{g}(x,\theta)\right]^t \left[(\gamma-1)c_\theta^t \bar{g}(x,\theta)+1\right]^{\frac{1}{\gamma-1}}, \quad \gamma \neq 0,1 \tag{34}$$

$$-\left[c_\theta^t \frac{\partial}{\partial\theta}\bar{g}(x,\theta)\right]^t \frac{1}{1-c_\theta^t \bar{g}(x,\theta)}, \quad \gamma = 0 \tag{35}$$

$$-\left[c_\theta^t \frac{\partial}{\partial\theta}\bar{g}(x,\theta)\right]^t \exp\left(c_\theta^t \bar{g}(x,\theta)\right), \quad \gamma = 1 \tag{36}$$

c_θ depending on divergence in each case.

In the following, we consider a model from [12] in order to exemplify the fact that the choice of the divergence determines the robustness properties of the estimators. This model is defined by $\mathcal{M} = \bigcup_{\theta\in\Theta}\mathcal{M}_\theta$ and satisfies linear constraints defined by the functions:

$$g_0(x) := 1, \qquad g_1(x) := x-\theta, \qquad g_2(x) := x^2 - 2\theta^2 - 1. \tag{37}$$

Let P_0 be the probability measure associated with the exponential distribution Exp(1). Note that $P_0 \notin \mathcal{M}$.

For a given divergence and a fixed θ, c_θ is given by (12), whenever there exists. In order to have the existence of c_θ, we need to have the integral $\int |\psi(t^t \bar{g}(x,\theta))|\,\mathrm{d}P_0$ finite for all $t := (t_0, t_1, t_2)$ in a neighborhood of c_θ.

For the KL_m divergence, $\psi(u) = -\log(1-u)$ and the domain of ψ is $(-\infty, 1)$. The integral $\int |\psi(t^t \bar{g}(x,\theta))|\,\mathrm{d}P_0$ is finite when the coefficient of the dominant term in $t^t\bar{g}(x,\theta)$ is negative, which implies that the coefficient of the dominant term in $c_\theta^t\bar{g}(x,\theta)$ is negative. Consequently, in this case, $m(x,\theta,c_\theta)$, given by (29), is unbounded w.r.t. x. This conclusion also holds for the Cressie–Read divergences associated with $\gamma < 0$ or to $\gamma > 1$.

When considering the KL divergence, $\psi(u) = \exp(u) - 1$, the domain of ψ is $\mathbb{R}$ and $\int |\psi(t^t \bar{g}(x,\theta))|\,\mathrm{d}P_0$ is finite when the coefficient of the dominant term in $t^t\bar{g}(x,\theta)$ is negative. This implies that the coefficient of the dominant term in $c_\theta^t\bar{g}(x,\theta)$ is negative and hence the corresponding $m(x,\theta,c_\theta)$ given by (30) is bounded w.r.t. x.

For the Hellinger divergence, $\psi(u) = \frac{2u}{2-u}$ and the domain of ψ is $(-\infty, 2)$. We impose the condition $t^t\bar{g}(x,\theta) = t_0 + t_1(x-\theta) + t_2(x^2 - 2\theta^2 - 1) < 2$, which is fulfilled when the coefficient of the dominant term is negative. In this case, the integral $\int |\psi(t^t \bar{g}(x,\theta))|\,\mathrm{d}P_0$ is finite and $m(x,\theta,c_\theta) = c_\theta^0 - \frac{2c_\theta^t\bar{g}(x,\theta)}{2-c_\theta^t\bar{g}(x,\theta)}$ is bounded w.r.t. x. This conclusion also holds for some other Cressie–Read divergences associated with $\gamma \in (0,1)$.

We conclude that, unlike KL_m and the divergences associated with $\gamma < 0$ or to $\gamma > 1$, the KL and the Hellinger divergences lead to robust test statistics $\widehat{D}_\varphi(\mathcal{M}_\theta, P_0)$ and $\widehat{D}_\varphi(\mathcal{M}, P_0)$.

Concerning the estimator $\widehat{\theta}_\varphi$, we remark that in the case of the KL divergence, both $\frac{\partial}{\partial t}m(x,\theta,c_\theta)$ and $\frac{\partial}{\partial\theta}m(x,\theta,c_\theta)$ are bounded, since the coefficient of the dominant term in $c_\theta^t\bar{g}(x,\theta)$ is negative, as explained above. These functions are also bounded when $\gamma \in (0,1)$ is chosen such that the degree of $[(\gamma-1)c_\theta^t\bar{g}(x,\theta)+1]^{1/(1-\gamma)}$ is greater than or equal to 2, for example when $\gamma = 1/2$. Moreover, depending on the values of the components of c_θ, it is possible that the functions $\frac{\partial}{\partial t}m(x,\theta,c_\theta)$ and $\frac{\partial}{\partial\theta}m(x,\theta,c_\theta)$ be bounded, even for the KL_m divergence. In all these situations $\mathrm{IF}(x; T, P_0)$ is bounded and the corresponding estimator $\widehat{\theta}_\varphi$ is robust.

The above study shows that the Cressie–Read class contains divergences which allow us to define robust statistical procedures for models satisfying linear constraints and that the Hellinger divergence is an example in this sense.

Acknowledgements

Financial support from the Romanian National Authority for Scientific Research, CNCS–UEFISCDI, Project number PN-II-RU-TE-2012-3-0007, is gratefully acknowledged.

References

[1] S. Bouzebda, A. Keziou, New estimates and tests of independence in semiparametric copula models, Kybernetika 46 (2010) 178–201.
[2] M. Broniatowski, A. Keziou, Minimization of ϕ-divergences on sets of signed measures, Stud. Sci. Math. Hung. 43 (2006) 403–442.
[3] M. Broniatowski, A. Keziou, Parametric estimation and tests through divergences and the duality technique, J. Multivar. Anal. 100 (2009) 16–31.
[4] M. Broniatowski, A. Keziou, Divergences and duality for estimation and test under moment condition models, J. Stat. Plan. Infer. 142 (2012) 2554–2573.
[5] M. Broniatowski, S. Leorato, An estimation method for Neyman chi-square divergence with application to test of hypothesis, J. Multivar. Anal. 97 (2006) 1409–1436.
[6] M. Broniatowski, A. Toma, I. Vajda, Decomposable pseudodistances and applications in statistical estimation, J. Stat. Plan. Infer. 142 (2012) 2574–2585.
[7] N. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests, J. R. Stat. Soc. Ser. B 46 (1984) 440–464.
[8] N.L. Glenn, Y. Zhao, Weighted empirical likelihood estimates and their robustness properties, Comput. Stat. Data An. 51 (2007) 5130–5141.
[9] F.R. Hampel, E. Ronchetti, P.J. Rousseeuw, W. Stahel, Robust Statistics: the Approach Based on Influence Functions, Wiley, New York, 1986.
[10] A. Keziou, Dual representation of ϕ-divergences and applications, C. R. Acad. Sci. Paris, Ser. I 336 (2003) 857–862.
[11] A.B. Owen, Empirical Likelihood, Chapman & Hall CRC, 2001.
[12] J. Qin, J. Lawless, Empirical likelihood and general estimating equations, Ann. Statist. 22 (1994) 300–325.
[13] A. Toma, M. Broniatowski, Dual divergence estimators and tests: Robustness results, J. Multivar. Anal. 102 (2011) 20–36.
[14] A. Toma, S. Leoni-Aubin, Robust tests based on dual divergence estimators and saddlepoint approximations, J. Multivar. Anal. 101 (2010) 1143–1155.