



Estimation and tests for models satisfying linear constraints with unknown parameter

Michel Broniatowski, Amor Keziou

► To cite this version:

Michel Broniatowski, Amor Keziou. Estimation and tests for models satisfying linear constraints with unknown parameter. 2008. hal-00340760

HAL Id: hal-00340760

<https://hal.science/hal-00340760>

Preprint submitted on 21 Nov 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ESTIMATION AND TESTS FOR MODELS SATISFYING LINEAR CONSTRAINTS WITH UNKNOWN PARAMETER

MICHEL BRONIATOWSKI* AND AMOR KEZIOU**

ABSTRACT. We introduce estimation and test procedures through divergence minimization for models satisfying linear constraints with unknown parameter. Several statistical examples and motivations are given. These procedures extend the empirical likelihood (EL) method and share common features with generalized empirical likelihood (GEL). We treat the problems of existence and characterization of the divergence projections of probability measures on sets of signed finite measures. Our approach allows for a study of the estimates under misspecification. The asymptotic behavior of the proposed estimates are studied using the dual representation of the divergences and the explicit forms of the divergence projections. We discuss the problem of the choice of the divergence under various respects. Also we handle efficiency and robustness properties of minimum divergence estimates. A simulation study shows that the Hellinger divergence enjoys good efficiency and robustness properties.

Key words: Empirical likelihood; Generalized Empirical likelihood; Minimum divergence; Efficiency; Robustness; Duality; Divergence projection.

MSC (2000) Classification: 62G05; 62G10; 62G15; 62G20; 62G35.

JEL Classification: C12; C13; C14.

1. INTRODUCTION AND NOTATION

A model satisfying partly specified linear parametric constraints is a family of distributions \mathcal{M}^1 all defined on a same measurable space $(\mathcal{X}, \mathcal{B})$, such that, for all Q in \mathcal{M}^1 , the following condition holds

$$\int g(x, \theta) dQ(x) = 0.$$

The unspecified parameter θ belongs to Θ , an open set in \mathbb{R}^d . The function $g := (g_1, \dots, g_l)^T$ is defined on $\mathcal{X} \times \Theta$ with values in \mathbb{R}^l , each of the g_i 's being real valued and the functions $g_1, \dots, g_l, \mathbf{1}_{\mathcal{X}}$ are assumed linearly independent. So \mathcal{M}^1 is defined through l -linear constraints indexed by some d -dimensional parameter θ . Denote M^1 the collection of all probability measures on $(\mathcal{X}, \mathcal{B})$, and

$$\mathcal{M}_{\theta}^1 := \left\{ Q \in M^1 \text{ such that } \int g(x, \theta) dQ(x) = 0 \right\}$$

so that

$$(1.1) \quad \mathcal{M}^1 = \bigcup_{\theta \in \Theta} \mathcal{M}_{\theta}^1.$$

Assume now that we have at hand a sample X_1, \dots, X_n of independent random variables (r.v.'s) with common unknown distribution P_0 . When P_0 belongs to the model (1.1), we denote θ_0 the value of the parameter θ such that \mathcal{M}_{θ_0} contains P_0 . Obviously, we assume that θ_0 is unique.

The scope of this paper is to propose new answers for the classical following problems

Problem 1: Does P_0 belong to the model \mathcal{M}^1 ?

Date: October 2008.

Problem 2: When P_0 is in the model, which is the value θ_0 of the parameter for which $\int g(x, \theta_0) dP_0(x) = 0$? Also can we perform simple and composite tests for θ_0 ? Can we construct confidence areas for θ_0 ? Can we give more efficient estimates for the distribution function than the usual empirical cumulative distribution function (c.d.f.)?

We present some examples and motivations for the model (1.1) and Problems 1 and 2.

1.1. Statistical examples and motivations.

Example 1.1. Suppose that P_0 is the distribution of a pair of random variables (X, Y) on a product space $\mathcal{X} \times \mathcal{Y}$ with known marginal distributions P_1 and P_2 . Bickel et al. (1991) study efficient estimation of $\theta = \int h(x, y) dP_0(x, y)$ for specified function h . This problem can be handled in the present context when the spaces \mathcal{X} and \mathcal{Y} are discrete and finite. Denote $\mathcal{X} = \{x_1, \dots, x_k\}$ and $\mathcal{Y} = \{y_1, \dots, y_r\}$. Consider an i.i.d. bivariate sample $(X_i, Y_i), 1 \leq i \leq n$ of the bivariate random variable (X, Y) . The space \mathcal{M}_θ in this case is the set of all p.m.'s Q on $\mathcal{X} \times \mathcal{Y}$ satisfying $\int g(x, y, \theta) dQ(x, y) = 0$ where $g = (g_1^{(1)}, \dots, g_k^{(1)}, g_1^{(2)}, \dots, g_r^{(2)}, g_1)^T$, $g_i^{(1)}(x, y, \theta) = \mathbb{1}_{\{x_i\} \times \mathcal{Y}}(x, y) - P_1(x_i)$, $g_j^{(2)}(x, y, \theta) = \mathbb{1}_{\mathcal{X} \times \{y_j\}}(x, y) - P_2(y_j)$ for all $(i, j) \in \{1, \dots, k\} \times \{1, \dots, r\}$, and $g_1(x, y, \theta) = h(x, y) - \theta$. Problem 1 turns to be the test for " P_0 belongs to $\bigcup_{\theta \in \Theta} \mathcal{M}_\theta$ ", while Problem 2 pertains to the estimation and tests for specific values of θ . Motivation and references for this problem are given in Bickel et al. (1991).

Example 1.2. (Generalized linear models). Let Y be a random variable and X a l -dimensional random vector. Y and X are linked through

$$Y = m(X, \theta_0) + \varepsilon$$

in which $m(\cdot, \cdot)$ is some specified real valued function and θ_0 , the parameter of interest, belongs to some open set $\Theta \subset \mathbb{R}^d$. ε is a measurement error. Denote P_0 the law of the vector variable (X, Y) and suppose that the true value θ_0 satisfies the orthogonality condition

$$\int x(y - m(x, \theta_0)) dP_0(x, y) = 0.$$

Consider an i.i.d. sample $(X_i, Y_i), 1 \leq i \leq n$ of r.v.'s with same distribution as (X, Y) . The existence of some θ_0 for which the above condition holds is given as the solution of Problem 1, while Problem 2 aims to provide its explicit value; here \mathcal{M}_θ^1 is the set of all p.m.'s Q on \mathbb{R}^{l+1} satisfying $\int g(x, y, \theta) dQ(x, y) = 0$ with $g(x, y, \theta) = x(y - m(x, \theta))$.

Qin and Lawless (1994) introduce various interesting examples when (1.1) applies. In their example 1, they consider the existence and estimation of the expectation θ of some r.v. X when $E(X^2) = m(\theta)$ for some known function $m(\cdot)$. Another example is when a bivariate sample (X_i, Y_i) of i.i.d. r.v.'s is observed, the expectation of X_i is known and we intend to estimate $E(Y_i)$. Haberman (1984) and Sheehy (1987) consider estimation of $F(x)$ based on i.i.d. sample X_1, \dots, X_n with distribution function F when it is known that $\int T(x) dF(x) = a$, for some specified function $T(\cdot)$. For this problem, the function $g(x, \theta)$ in the model (1.1) is equal to $T(x) - \theta$ where $\theta = a$ is known. This example with a unknown is treated in details in Section 3 of the present paper. We refer to Owen (2001) for more statistical examples when model (1.1) applies.

Another motivation for our work stems from confidence region (C.R.) estimation techniques. The empirical likelihood method provides such estimation (see Owen (1990)). We will extend this approach providing a wide range of such C.R.'s, each one depending upon a specific criterion, one of those leading to Owen's C.R.

An important estimator of θ_0 is the generalized method of moments (GMM) estimator of Hansen (1982). The empirical likelihood approach developed by Owen (1988) and Owen (1990) has been adapted in the present setting by Qin and Lawless (1994) and Imbens (1997) introducing the empirical likelihood estimator (EL). The recent literature in econometrics focusses on such models, the paper by Newey and Smith (2004) provides an exhaustive list of works dealing with the statistical properties of GMM and generalized empirical likelihood (GEL) estimators.

Our interest also lays in the behavior of the estimates under misspecification. In the context of tests of hypothesis, the statistics to be considered is some estimate of some divergence between the unknown distributions of the data and the model. We are also motivated by the behavior of those statistics under misspecification, i.e., when the model is not appropriated to the data. Such questions have not been addressed until now for those problems in the general context of divergences. Schennach (2007) consider the asymptotic properties of the empirical likelihood estimate under misspecification. As a by product, we will prove that our proposal leads to consistent test procedures; furthermore, the asymptotic behavior of the statistics, under \mathcal{H}_1 , provides the fundamental tool in order to achieve Bahadur efficiency calculations (see Nikitin (1995)).

An important result due to Newey and Smith (2004) states that EL estimate enjoys optimality properties in term of efficiency when bias corrected among all GEL and GMM estimators. Also Corcoran (1998) and Baggerly (1998) proved that in a class of minimum discrepancy statistics, EL ratio is the only that is Bartlett correctable. However, these results do not consider the optimality properties of the tests for *Problems 1* and *2*. Also, in connection with estimation problem, they do not consider the properties of EL estimate with respect to robustness. So, the question regarding divergence-based methods remains open at least in these two instances.

The approach which we develop is based on *minimum discrepancy estimates*, which have common features with minimum distance techniques, using merely divergences. We present wide sets of estimates, simple and composite tests and confidence regions for the parameter θ_0 as well as various test statistics for *Problem 1*, all depending on the choice of the divergence. Simulations show that the approach based on Hellinger divergence enjoys good robustness and efficiency properties when handling *Problem 2*. As presented in Section 5, empirical likelihood methods appear to be a special case of the present approach.

1.2. Minimum divergence estimates. We first set some general definition and notation. Let P be some probability measure (p.m.). Denote $M^1(P)$ the subset of all p.m.'s which are absolutely continuous (a.c.) with respect to P . Denote M the space of all signed finite measures on $(\mathcal{X}, \mathcal{B})$ and $M(P)$ the subset of all signed finite measures a.c. w.r.t. P . Let φ be a convex function from $[-\infty, +\infty]$ onto $[0, +\infty]$ with $\varphi(1) = 0$. For any signed finite measure Q in $M(P)$, the ϕ -divergence between Q and the p.m. P is defined through

$$(1.2) \quad \phi(Q, P) := \int \varphi\left(\frac{dQ}{dP}\right) dP.$$

When Q is not a.c. w.r.t. P , we set $\phi(Q, P) = +\infty$. This definition extends Rüschemdorf (1984)'s one which applies for ϕ -divergences between p.m.'s; it also differs from Csiszár (1963)'s one, which requires a common dominating σ -finite measure, noted λ , for Q and P . Since we will consider subsets of $M^1(P)$ and subsets of $M(P)$, it is more adequate for our sake to use the definition (1.2). Also note that all the just mentioned definitions of ϕ -divergences coincide on the set of all p.m.'s a.c. w.r.t. P and dominated by λ .

For all p.m. P , the mappings $Q \in M \rightarrow \phi(Q, P)$ are convex and take nonnegative values. When $Q = P$ then $\phi(Q, P) = 0$. Further, if the function $x \rightarrow \varphi(x)$ is strictly convex on neighborhood of $x = 1$, then the following basic property holds

$$(1.3) \quad \phi(Q, P) = 0 \text{ if and only if } Q = P.$$

All these properties are presented in Csiszár (1963), Csiszár (1967) and Liese and Vajda (1987) Chapter 1, for ϕ -divergences defined on the set of all p.m.'s M^1 . When the ϕ -divergences are defined on M , then the same arguments as developed on M^1 hold.

When defined on M^1 , the Kullback-Leibler (KL), modified Kullback-Leibler (KL_m), χ^2 , modified χ^2 (χ_m^2), Hellinger (H), and L^1 divergences are respectively associated to the convex functions $\varphi(x) = x \log x - x + 1$, $\varphi(x) = -\log x + x - 1$, $\varphi(x) = \frac{1}{2}(x-1)^2$, $\varphi(x) = \frac{1}{2}(x-1)^2/x$, $\varphi(x) = 2(\sqrt{x}-1)^2$ and $\varphi(x) = |x-1|$. All those divergences except the L^1 one, belong to the class of power divergences introduced in Cressie and Read (1984) (see also Liese and Vajda (1987) Chapter 2). They are defined through the class of convex functions

$$(1.4) \quad x \in \mathbb{R}_+^* \mapsto \varphi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)}$$

if $\gamma \in \mathbb{R} \setminus \{0, 1\}$ and by $\varphi_0(x) := -\log x + x - 1$ and $\varphi_1(x) := x \log x - x + 1$. For all $\gamma \in \mathbb{R}$, $\varphi_\gamma(0) := \lim_{x \downarrow 0} \varphi_\gamma(x)$ and $\varphi_\gamma(+\infty) := \lim_{x \uparrow +\infty} \varphi_\gamma(x)$. So, the KL -divergence is associated to φ_1 , the KL_m to φ_0 , the χ^2 to φ_2 , the χ_m^2 to φ_{-1} and the Hellinger distance to $\varphi_{1/2}$. For all $\gamma \in \mathbb{R}$, we sometimes denote ϕ_γ the divergence associated to the convex function φ_γ . We define the derivative of φ_γ at 0 by $\varphi'_\gamma(0) := \lim_{x \downarrow 0} \varphi'_\gamma(x)$. We extend the definition of the power divergences functions $Q \in M^1 \rightarrow \phi_\gamma(Q, P)$ onto the whole set of signed finite measures M as follows. When the function $x \rightarrow \varphi_\gamma(x)$ is not defined on $(-\infty, 0[$ or when φ_γ is defined on \mathbb{R} but is not a convex function we extend the definition of φ_γ through

$$(1.5) \quad x \in [-\infty, +\infty] \mapsto \varphi_\gamma(x) \mathbb{1}_{[0, +\infty]}(x) + (\varphi'_\gamma(0)x + \varphi_\gamma(0)) \mathbb{1}_{[-\infty, 0[}(x).$$

For any convex function φ , define the *domain* of φ through

$$(1.6) \quad D_\varphi = \{x \in [-\infty, +\infty] \text{ such that } \varphi(x) < +\infty\}.$$

Since φ is convex, D_φ is an interval which may be open or not, bounded or unbounded. Hence, write $D_\varphi := (a, b)$ in which a and b may be finite or infinite. In this paper, we will only consider φ functions defined on $[-\infty, +\infty]$ with values in $[0, +\infty]$ such that $a < 1 < b$, and which satisfy $\varphi(1) = 0$, are strictly convex and are \mathcal{C}^2 on the interior of its domain D_φ ; we define $\varphi(a)$, $\varphi'(a)$, $\varphi''(a)$, $\varphi(b)$, $\varphi'(b)$ and $\varphi''(b)$ respectively by $\varphi(a) := \lim_{x \downarrow a} \varphi(x)$, $\varphi'(a) := \lim_{x \downarrow a} \varphi'(x)$, $\varphi''(a) := \lim_{x \downarrow a} \varphi''(x)$, $\varphi(b) := \lim_{x \uparrow b} \varphi(x)$, $\varphi'(b) := \lim_{x \uparrow b} \varphi'(x)$ and $\varphi''(b) := \lim_{x \uparrow b} \varphi''(x)$. These quantities may be finite or infinite. All the functions φ_γ (see (1.5)) satisfy these conditions.

Definition 1.1. Let Ω be some subset in M . The ϕ -divergence between the set Ω and a p.m. P , noted $\phi(\Omega, P)$, is

$$\phi(\Omega, P) := \inf_{Q \in \Omega} \phi(Q, P).$$

Definition 1.2. Assume that $\phi(\Omega, P)$ is finite. A measure $Q^* \in \Omega$ such that

$$\phi(Q^*, P) \leq \phi(Q, P) \text{ for all } Q \in \Omega$$

is called a ϕ -projection of P onto Ω . This projection may not exist, or may be not defined uniquely.

We will make use of the concept of ϕ -divergences in order to perform estimation and tests for the model (1.1). So, let X_1, \dots, X_n denote an i.i.d. sample of r.v.'s with common distribution P_0 . Let

P_n be the empirical measure pertaining to this sample, namely

$$P_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

in which δ_x is the Dirac measure at point x . When P_0 and all $Q \in \mathcal{M}^1$ share the same discrete finite support S , then the ϕ -divergence $\phi(Q, P_0)$ can be written as

$$(1.7) \quad \phi(Q, P_0) = \sum_{j \in S} \varphi \left(\frac{Q(j)}{P_0(j)} \right) P_0(j).$$

In this case, $\phi(Q, P_0)$ can be estimated simply through the plug-in of P_n in (1.7), as follows

$$(1.8) \quad \hat{\phi}(Q, P_0) := \sum_{j \in S} \varphi \left(\frac{Q(j)}{P_n(j)} \right) P_n(j).$$

In the same way, for any θ in Θ , $\phi(\mathcal{M}_\theta^1, P_0)$ is estimated by

$$(1.9) \quad \hat{\phi}(\mathcal{M}_\theta^1, P_0) := \inf_{Q \in \mathcal{M}_\theta^1} \sum_{j \in S} \varphi \left(\frac{Q(j)}{P_n(j)} \right) P_n(j),$$

and $\phi(\mathcal{M}^1, P_0) = \inf_{\theta \in \Theta} \phi(\mathcal{M}_\theta^1, P_0)$ can be estimated by

$$(1.10) \quad \hat{\phi}(\mathcal{M}^1, P_0) := \inf_{\theta \in \Theta} \inf_{Q \in \mathcal{M}_\theta^1} \sum_{j \in S} \varphi \left(\frac{Q(j)}{P_n(j)} \right) P_n(j).$$

By uniqueness of $\inf_{\theta \in \Theta} \phi(\mathcal{M}_\theta^1, P_0)$ and since this infimum is reached at $\theta = \theta_0$, we estimate θ_0 through

$$(1.11) \quad \hat{\theta}_\phi := \arg \inf_{\theta \in \Theta} \inf_{Q \in \mathcal{M}_\theta^1} \sum_{j \in S} \varphi \left(\frac{Q(j)}{P_n(j)} \right) P_n(j).$$

The infimum in (1.9) (i.e., the projection of P_n on \mathcal{M}_θ^1) may be achieved on the frontier of \mathcal{M}_θ^1 . In this case the Lagrange method is not valid. Hence, we endow our statistical approach in the global context of signed finite measures with total mass 1 satisfying the linear constraints.

$$(1.12) \quad \mathcal{M}_\theta := \left\{ Q \in M \text{ such that } \int dQ = 1 \text{ and } \int g(x, \theta) dQ(x) = 0 \right\}$$

and

$$(1.13) \quad \mathcal{M} := \bigcup_{\theta \in \Theta} \mathcal{M}_\theta,$$

sets of signed finite measures that replace \mathcal{M}_θ^1 and \mathcal{M}^1 .

As above, we estimate $\phi(\mathcal{M}_\theta, P_0)$, $\phi(\mathcal{M}, P_0)$ and θ_0 respectively by

$$(1.14) \quad \hat{\phi}(\mathcal{M}_\theta, P_0) := \inf_{Q \in \mathcal{M}_\theta} \sum_{j \in S} \varphi \left(\frac{Q(j)}{P_n(j)} \right) P_n(j),$$

$$(1.15) \quad \hat{\phi}(\mathcal{M}, P_0) := \inf_{\theta \in \Theta} \inf_{Q \in \mathcal{M}_\theta} \sum_{j \in S} \varphi \left(\frac{Q(j)}{P_n(j)} \right) P_n(j),$$

and

$$(1.16) \quad \hat{\theta}_\phi := \arg \inf_{\theta \in \Theta} \inf_{Q \in \mathcal{M}_\theta} \sum_{j \in S} \varphi \left(\frac{Q(j)}{P_n(j)} \right) P_n(j).$$

Enhancing \mathcal{M}^1 to \mathcal{M} is motivated by the following arguments

- For all θ in Θ , denote Q_1^* and Q^* respectively the projection of P_n on \mathcal{M}_θ^1 and on \mathcal{M}_θ , as defined in (1.9) and in (1.14). If Q_1^* is an interior point of \mathcal{M}_θ^1 , then, by Proposition 2.5 below, it coincides with Q^* , the projection of P_n on \mathcal{M}_θ , i.e., $Q_1^* = Q^*$. Therefore, in this case, both approaches coincide.
- It may occur that for some θ in Θ , Q_1^* , the projection of P_n on \mathcal{M}_θ^1 , is a frontier point of \mathcal{M}_θ^1 , which makes a real difficulty for the estimation procedure. We will prove in Theorem 3.4 that $\hat{\theta}_\phi$, defined in (1.16) and which replaces (1.11), converges to θ_0 . This validates the substitution of the sets \mathcal{M}_θ^1 by the sets \mathcal{M}_θ . In the context of a test problem, we will prove that the asymptotic distributions of the test statistics pertaining to Problem 1 and 2 are unaffected by this change.

This modification motivates the above extensions in the definitions of the φ functions on $[-\infty, +\infty]$ and of the ϕ -divergences on the whole space of finite signed measures M .

In the case when Q and P_0 share different discrete finite support or share same or different discrete infinite or continuous support, then formula (1.8) is not defined, due to lack of absolute continuity of Q with respect to P_n . Indeed

$$(1.17) \quad \hat{\phi}(Q, P_0) := \phi(Q, P_n) = +\infty.$$

The plug-in estimate of $\phi(\mathcal{M}_\theta, P_0)$ is

$$(1.18) \quad \hat{\phi}(\mathcal{M}_\theta, P_0) := \inf_{Q \in \mathcal{M}_\theta} \phi(Q, P_n) = \inf_{Q \in \mathcal{M}_\theta} \int \varphi\left(\frac{dQ}{dP_n}(x)\right) dP_n(x).$$

If the infimum exists, then it is clear that it is reached at a signed finite measure (or probability measure) which is a.c. w.r.t. P_n . So, define the sets

$$(1.19) \quad \mathcal{M}_\theta^{(n)} := \left\{ Q \in M \text{ such that } Q \ll P_n, \sum_{i=1}^n Q(X_i) = 1 \text{ and } \sum_{i=1}^n Q(X_i)g(X_i, \theta) = 0 \right\},$$

which may be seen as subsets of \mathbb{R}^n . Then, the plug-in estimate (1.18) of $\phi(\mathcal{M}_\theta, P_0)$ can be written as

$$(1.20) \quad \hat{\phi}(\mathcal{M}_\theta, P_0) = \inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i)).$$

In the same way, $\phi(\mathcal{M}, P_0) := \inf_{\theta \in \Theta} \inf_{Q \in \mathcal{M}_\theta} \phi(Q, P_0)$ can be estimated by

$$(1.21) \quad \hat{\phi}(\mathcal{M}, P_0) = \inf_{\theta \in \Theta} \inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i)).$$

By uniqueness of $\inf_{\theta \in \Theta} \phi(\mathcal{M}_\theta, P_0)$ and since this infimum is reached at $\theta = \theta_0$, we estimate θ_0 through

$$(1.22) \quad \hat{\theta}_\phi = \arg \inf_{\theta \in \Theta} \inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i)).$$

Note that, when P_0 and all $Q \in \mathcal{M}^1$ share the same discrete finite support, then the estimates (1.22), (1.21) and (1.20) coincide respectively with (1.16), (1.15) and (1.14). Hence, in the sequel, we study the estimates $\hat{\phi}(\mathcal{M}_\theta, P_0)$, $\hat{\phi}(\mathcal{M}, P_0)$ and $\hat{\theta}_\phi$ as defined in (1.20), (1.21) and (1.22), respectively. We propose to call the estimates $\hat{\theta}_\phi$ defined in (1.22) “Minimum Empirical ϕ -Divergences Estimates” (ME ϕ DE’s). As will be noticed later on, the empirical likelihood paradigm (see Owen (1988) and Owen (1990)), which is based on this plug-in approach, enters as a special case of the statistical issues related to estimation and tests based on ϕ -divergences with $\varphi(x) = \varphi_0(x) = -\log x + x - 1$, namely on KL_m -divergence. The empirical log-likelihood ratio

for the model (1.12), in the context of ϕ -divergences, can be written as $-\widehat{nKL_m}(\mathcal{M}_\theta, P_0)$. In the case of a single functional, for example when $g(x, \theta) = x - \theta$ with x and θ belong to \mathbb{R} , Owen (1988) shows that $2n\widehat{KL_m}(\mathcal{M}_\theta, P_0)$ has an asymptotic $\chi^2_{(1)}$ distribution when P_0 belongs to \mathcal{M}_θ . (see Owen (1988) Theorem 1). This result is a nonparametric version of Wilks's theorem (see Wilks (1938)). In the multivariate case, the same result holds (see Owen (1990) Theorem 1). When we want to extend the arguments used in Owen (1988) and Owen (1990) in order to study the limiting behavior of the statistics $\widehat{\phi}(\mathcal{M}_\theta, P_0)$, when $P_0 \notin \mathcal{M}_\theta$ (for example, when $\theta_0 \neq \theta$), most limiting arguments become untractable. We propose to use the so-called “dual representation of ϕ -divergences” (see Keziou (2003)), a device which is well known for the Kullback-Leibler divergence in the context of large deviations, and which has been used in parametric statistics in Keziou (2003) and Broniatowski and Keziou (2009). The estimates then turn to be M-estimates whose limiting distributions are obtained through classical methods. On the other hand, the obtention of the limit distributions of the statistics $\widehat{\phi}(\mathcal{M}_\theta, P_0)$ when $P_0 \notin \mathcal{M}_\theta$, requires the study of the existence and the characterization of the projection of the p.m. P_0 on the sets \mathcal{M}_θ .

This paper is organized as follows : In Section 3, we study the asymptotic behavior of the proposed estimates (1.20), (1.21) and (1.22) giving solutions to *Problem 2*. We then address *Problem 1*, namely : does there exist some θ_0 in Θ for which P_0 belongs to \mathcal{M}_{θ_0} ? In Section 4, extending the result by Qin and Lawless (1994), we give new estimates for the distribution function using the ϕ -projections of P_n on the model \mathcal{M} . We show that the new estimates of the distribution function are generally more efficient than the empirical cumulative distribution function. Section 5 illustrates the concept of empirical likelihood in the context of ϕ -divergences techniques. In Section 6, we focus on robustness and efficiency of the ME ϕ D estimates. A simulation study aims at emphasizing the specific advantage of the choice of the Hellinger divergence in relation with robustness and efficiency considerations. All proofs are in Section 7.

2. ESTIMATION FOR MODELS SATISFYING LINEAR CONSTRAINTS

At this point, we must introduce some notational convention for sake of brevity and clearness. For any p.m. P on \mathcal{X} and any measurable real function f on \mathcal{X} , Pf denotes $\int f(x) dP(x)$. For example, $P_0 g_j(\theta)$ will be used instead of $\int g_j(\theta, x) dP_0(x)$. Hence, we are led to define the following functions : denote \overline{g} the function defined on $\mathcal{X} \times \Theta$ with values in \mathbb{R}^{l+1} by

$$\begin{aligned} \overline{g} &: \mathcal{X} \times \Theta \rightarrow \mathbb{R}^{(l+1)} \\ (x, \theta) &\mapsto \overline{g}(x, \theta) := (\mathbf{1}_{\mathcal{X}}(x), g_1(x, \theta), \dots, g_l(x, \theta))^T, \end{aligned}$$

and for all $\theta \in \Theta$, denote also $\overline{g}(\theta)$, $g(\theta)$, $g_j(\theta)$ the functions defined respectively by

$$\begin{aligned} \overline{g}(\theta) &: \mathcal{X} \rightarrow \mathbb{R}^{l+1} \\ x &\mapsto \overline{g}(x, \theta) := (g_0(x, \theta), g_1(x, \theta), \dots, g_l(x, \theta))^T, \text{ where } g_0(x, \theta) := \mathbf{1}_{\mathcal{X}}(x), \\ g(\theta) &: \mathcal{X} \rightarrow \mathbb{R}^l \\ x &\mapsto g(x, \theta) := (g_1(x, \theta), \dots, g_l(x, \theta))^T \\ \text{and} \\ g_j(\theta) &: \mathcal{X} \rightarrow \mathbb{R} \\ x &\mapsto g_j(x, \theta), \text{ for all } j \in \{0, 1, \dots, l\}. \end{aligned}$$

We now turn back to the setting defined in the Introduction and consider model (1.12). For fixed θ in Θ , define the class of functions

$$\mathcal{F}_\theta := \{g_0(\theta), g_1(\theta), \dots, g_l(\theta)\},$$

and consider the set of finite signed measures \mathcal{M}_θ defined by $(l+1)$ linear constraints as defined in (1.12)

$$\mathcal{M}_\theta := \left\{ Q \in M_{\mathcal{F}_\theta} \text{ such that } \int dQ(x) = 1 \text{ and } \int g(x, \theta) dQ(x) = 0 \right\}.$$

We present explicit tractable conditions for the estimates (1.20), (1.21) and (1.22) to be well defined. This will be done in Propositions 2.1, Remark 2.1, Proposition 2.2 and Remark 2.3 below. First, we present sufficient conditions which assess the existence of the infimum in (1.20), noted \widehat{Q}_θ^* , the projection of P_n on \mathcal{M}_θ . We also provide conditions under which the Lagrange method can be used to characterize \widehat{Q}_θ^* . The Fenchel-Legendre transform of φ will be denoted φ^* , i.e.,

$$(2.1) \quad t \in \mathbb{R} \mapsto \varphi^*(t) := \sup_{x \in \mathbb{R}} \{tx - \varphi(x)\}.$$

Define

$$(2.2) \quad \mathcal{D}_\phi^{(n)} := \left\{ Q \in M \text{ such that } Q \ll P_n \text{ and } \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i)) < \infty \right\},$$

i.e., the *domain* of the function

$$(Q(X_1), \dots, Q(X_n))^T \in \mathbb{R}^n \mapsto \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i)).$$

We have

Proposition 2.1. *Assume that there exists some measure R in the interior of $\mathcal{D}_\phi^{(n)}$ and in $\mathcal{M}_\theta^{(n)}$ such that for all Q in $\partial \mathcal{D}_\phi^{(n)}$, the frontier of $\mathcal{D}_\phi^{(n)}$, we have*

$$(2.3) \quad \frac{1}{n} \sum_{i=1}^n \varphi(nR(X_i)) < \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i)).$$

Then the following holds

(i) *there exists an unique \widehat{Q}_θ^* in $\mathcal{M}_\theta^{(n)}$ such that*

$$(2.4) \quad \inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i)) = \frac{1}{n} \sum_{i=1}^n \varphi(n\widehat{Q}_\theta^*(X_i))$$

(ii) *\widehat{Q}_θ^* is an interior point of $\mathcal{D}_\phi^{(n)}$ and satisfies for all $i = 1, \dots, n$*

$$(2.5) \quad \widehat{Q}_\theta^*(X_i) = \frac{1}{n} \overleftarrow{\varphi}' \left(\sum_{j=0}^l \widehat{c}_j g_j(X_i, \theta) \right),$$

where $(\widehat{c}_0, \widehat{c}_1, \dots, \widehat{c}_l)^T := \widehat{c}_\theta$ is solution of the system of equations

$$(2.6) \quad \begin{cases} \int \overleftarrow{\varphi}' \left(\widehat{c}_0 + \sum_{i=1}^l \widehat{c}_i g_i(x, \theta) \right) dP_n(x) &= 1 \\ \int g_j(x, \theta) \overleftarrow{\varphi}' \left(\widehat{c}_0 + \sum_{i=1}^l \widehat{c}_i g_i(x, \theta) \right) dP_n(x) &= 0, \quad j = 1, \dots, l. \end{cases}$$

Example 2.1. *For the χ^2 -divergence, we have $\mathcal{D}_{\chi^2}^{(n)} = \mathbb{R}^n$. Hence condition (2.3) holds whenever $\mathcal{M}_\theta^{(n)}$ is not void. Therefore, the above Proposition holds always independently upon the distribution P_0 . More generally, the above Proposition holds for any ϕ -divergence which is associated to φ function satisfying $D_\varphi = \mathbb{R}$. (See (1.6) for the definition of D_φ).*

Example 2.2. In the case of the modified Kullback-Leibler divergence, which turns to coincide with the empirical likelihood technique (see Section 5), we have $\mathcal{D}_{KL_m}^{(n)} =]0, +\infty[)^n$. For α in Θ , define the assertion

$$(2.7) \quad \begin{aligned} & \text{there exists } q = (q_1, \dots, q_n) \text{ in } \mathbb{R}^n \text{ with } 0 < q_i < 1 \text{ for all } i = 1, \dots, n \\ & \text{and } \sum_{i=1}^n q_i g_j(X_i, \alpha) = 0 \text{ for all } j = 1, \dots, l. \end{aligned}$$

A sufficient condition, in order to assess that condition (2.3) in the above Proposition holds, is when (2.7) holds for $\alpha = \theta$. In the case when $g(x, \theta) = x - \theta$, this is precisely what is checked in (Owen (1990)), p. 100, when θ is an interior point of the convex hull of (X_1, \dots, X_n) .

Example 2.3. For the modified χ^2 -divergence, we have $\mathcal{D}_{\chi_m^2}^{(n)} =]0, \infty[)^n$, and therefore, condition (2.7) for $\alpha = \theta$ is sufficient for the condition (2.3) to hold. So, conditions which assess the existence of the projection \widehat{Q}_θ^* are the same for the modified χ^2 -divergence and the KL_m -divergence.

Remark 2.1. If there exists some $Q_0 \in \mathcal{M}_\theta^{(n)}$ such that

$$(2.8) \quad a < \inf_i nQ_0(X_i) \leq \sup_i nQ_0(X_i) < b,$$

then applying Corollary 2.6 in Borwein and Lewis (1991), we get

$$\inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i)) = \sup_{t \in \mathbb{R}^{(l+1)}} \left\{ t_0 - \int \psi(t^T \bar{g}(x, \theta)) dP_n(x) \right\}$$

with dual attainment. Furthermore, if

$$\varphi'(a) < \inf_i \widehat{c}_\theta^T \bar{g}(X_i, \theta) \leq \sup_i \widehat{c}_\theta^T \bar{g}(X_i, \theta) < \varphi'(b),$$

with \widehat{c}_θ a dual optimal, then the unique projection \widehat{Q}_θ^* of P_n on $\mathcal{M}_\theta^{(n)}$ is given by (2.5).

We will make use of the dual representation of ϕ -divergences (see Broniatowski and Keziou (2006) theorem 4.4). So, define

$$(2.9) \quad \mathcal{C}_\theta := \{t \in \mathbb{R}^{l+1} \text{ such that } t^T \bar{g}(\cdot, \theta) \text{ belongs to } \text{Im } \varphi' (P_0 - a.s.)\},$$

and

$$(2.10) \quad \mathcal{C}_\theta^{(n)} := \{t \in \mathbb{R}^{l+1} \text{ such } t^T \bar{g}(X_i, \theta) \text{ belongs to } \text{Im } \varphi' \text{ for all } i = 1, \dots, n\}.$$

We omit the subscript θ when unnecessary. Note that both \mathcal{C}_θ and $\mathcal{C}_\theta^{(n)}$ depend upon the function φ but, for simplicity, we omit the subscript φ .

If P_0 admits a projection Q_θ^* on \mathcal{M}_θ with the same support as P_0 , using the second part in Corollary 3.5 in Broniatowski and Keziou (2006), there exist constants c_0, \dots, c_l , obviously depending on θ , such that

$$\varphi' \left(\frac{dQ_\theta^*}{dP_0}(x) \right) = c_0 + \sum_{j=1}^l c_j g_j(x, \theta), \text{ for all } x (P_0 - a.s.).$$

Since Q_θ^* belongs to \mathcal{M}_θ , the real numbers c_0, c_1, \dots, c_l are solutions of

$$(2.11) \quad \begin{cases} \int \varphi'^{-1} \left(c_0 + \sum_{j=1}^l c_j g_j(x, \theta) \right) dP_0(x) & = 1 \\ \int g_j(x, \theta) \varphi'^{-1} \left(c_0 + \sum_{j=1}^l c_j g_j(x, \theta) \right) dP_0(x) & = 0, j = 1, \dots, l. \end{cases}$$

Since $Q \mapsto \phi(Q, P_0)$ is strictly convex, the projection Q_θ^* of P_0 on the convex set \mathcal{M}_θ is unique. This implies, by Broniatowski and Keziou (2006) Corollary part 1, that the solution

$$c_\theta := (c_0, c_1, \dots, c_l)^T$$

of the system (2.11) is unique provided that the functions $g_i(\theta)$ are linearly independent. Further, using the dual representation of ϕ -divergences (see Broniatowski and Keziou (2006) Theorem 4.4), we get

$$\phi(\mathcal{M}_\theta, P_0) := \phi(Q_\theta^*, P_0) = \sup_{f \in \mathcal{F}} \left\{ \int f dQ_\theta^* - \int \varphi^*(f) dP_0 \right\},$$

and the sup is unique and is reached at $f = \varphi'(dQ_\theta^*/dP_0) = c_0 + \sum_{j=1}^l c_j g_j(\cdot, \theta)$, if it belongs to \mathcal{F} . This motivates the choice of the class \mathcal{F} through

$$\mathcal{F} := \{x \mapsto t^T \bar{g}(x, \theta) \text{ for } t \text{ in } \mathcal{C}_\theta\}.$$

It is the smallest class of functions that contains $\varphi'(dQ_\theta^*/dP_0)$ and which does not presume any knowledge on Q_θ^* . We thus obtain

$$\phi(\mathcal{M}_\theta, P_0) = \sup_{t \in \mathcal{C}_\theta} \int m(x, \theta, t) dP_0(x),$$

where $m(\theta, t)$ is the function defined on \mathcal{X} by

$$x \in \mathcal{X} \mapsto m(x, \theta, t) := t_0 - \varphi^*(t^T \bar{g}(x, \theta)) = t_0 - (t^T \bar{g}(x, \theta)) \varphi'^{-1}(t^T \bar{g}(x, \theta)) + \varphi(\varphi'^{-1}(t^T \bar{g}(x, \theta))).$$

With the above notation, we state

$$(2.12) \quad \phi(\mathcal{M}_\theta, P_0) = \sup_{t \in \mathcal{C}_\theta} P_0 m(\theta, t).$$

So, a natural estimate of $\phi(\mathcal{M}_\theta, P_0)$ is

$$(2.13) \quad \sup_{t \in \mathcal{C}_\theta^{(n)}} P_n m(\theta, t)$$

which coincides with the estimate defined in (1.20). Hence, we can write

$$(2.14) \quad \hat{\phi}(\mathcal{M}_\theta, P_0) = \sup_{t \in \mathcal{C}_\theta^{(n)}} P_n m(\theta, t).$$

which transforms the constrained optimization in (1.20) into the above unconstrained one.

On the other hand, the sup in (2.12) is reached at $t_0 = c_0, \dots, t_l = c_l$ which are solutions of the system of equations (2.11), i.e.,

$$(2.15) \quad c_\theta = \arg \sup_{t \in \mathcal{C}_\theta} P_0 m(\theta, t).$$

So, a natural estimate of c_θ in (2.15) is therefore defined through

$$(2.16) \quad \arg \sup_{t \in \mathcal{C}_\theta^{(n)}} P_n m(\theta, t).$$

This coincides with \hat{c}_θ , the solution of the system of equations (2.6). So, we can write

$$(2.17) \quad \hat{c}_\theta = \arg \sup_{t \in \mathcal{C}_\theta^{(n)}} P_n m(\theta, t).$$

Using (2.14), we obtain the following representations for the estimates $\widehat{\phi}(\mathcal{M}, P_0)$ in (1.21) and $\widehat{\theta}_\phi$ in (1.22)

$$(2.18) \quad \widehat{\phi}(\mathcal{M}, P_0) = \inf_{\theta \in \Theta} \sup_{t \in \mathcal{C}_\theta^{(n)}} P_n m(\theta, t)$$

and

$$(2.19) \quad \widehat{\theta}_\phi = \arg \inf_{\theta \in \Theta} \sup_{t \in \mathcal{C}_\theta^{(n)}} P_n m(\theta, t),$$

respectively.

Formula (2.12) also has the following basic interest : Consider the function

$$(2.20) \quad t \in \mathcal{C}_\theta \mapsto P_0 m(\theta, t),$$

In order for integral (2.20) to be properly defined, we assume that

$$(2.21) \quad \int |g_i(x, \theta)| dP_0(x) < \infty, \quad \text{for all } i \in \{1, \dots, l\}.$$

The domain of the function (2.20) is

$$(2.22) \quad \mathcal{D}_\phi(\theta) := \{t \in \mathcal{C}_\theta \text{ such that } P_0 m(\theta, t) > -\infty\}.$$

The function $t \mapsto P_0 m(\theta, t)$ is strictly concave on the convex set $\mathcal{D}_\phi(\theta)$. Whenever it has a maximum t^* , then it is unique, and if it belongs to the interior of $\mathcal{D}_\phi(\theta)$, then it satisfies the first order condition. Therefore t^* satisfies system (2.11). In turn, this implies that the measure Q^* defined through $dQ^* := \varphi'^{-1} \left(t^{*T} \overline{g}(\theta) \right) dP_0$ is the projection of P_0 on Ω , by Theorem 3.4 part 1 in Broniatowski and Keziou (2006). This implies that Q^* and P_0 share the same support. We summarize the above arguments as follows

Proposition 2.2. *Assume that (2.21) holds and that*

- (i) *there exists some s in the interior of $\mathcal{D}_\phi(\theta)$ such that for all t in $\partial \mathcal{D}_\phi(\theta)$, the frontier of $\mathcal{D}_\phi(\theta)$, it holds $P_0 m(\theta, t) < P_0 m(\theta, s)$;*
- (ii) *for all t in the interior of $\mathcal{D}_\phi(\theta)$, there exists a neighborhood $V(t)$ of t , such that the classes of functions $\left\{ x \rightarrow \frac{\partial}{\partial r_i} m(x, \theta, r), \quad r \in V(t) \right\}$ are dominated (P_0 -a.s.) by some P_0 -integrable function $x \rightarrow H(x, \theta)$.*

Then P_0 admits an unique projection Q_θ^* on \mathcal{M}_θ having the same support as P_0 and

$$(2.23) \quad dQ_\theta^* = \varphi'^{-1} (c_\theta^T \overline{g}(\theta)) dP_0,$$

where c_θ is the unique solution of the system of equations (2.11).

Remark 2.2. *In the case of KL-divergence, comparing this Proposition with Theorem 3.3 in Csiszár (1975), we observe that the dual formula (2.12) provides weaker conditions on the class of functions $\{\overline{g}(\theta), \theta \in \Theta\}$ than the geometric approach.*

Remark 2.3. *The result of Borwein and Lewis (1991), with some additional conditions, provides more practical tools for obtaining the results in Proposition 2.2. Assume that the functions $g_j(\theta)$ belongs to the space $L_p(\mathcal{X}, P_0)$ with $1 \leq p \leq \infty$ and that the following “constraint qualification” holds*

$$(2.24) \quad \text{there exists some } Q_0 \text{ in } \mathcal{M}_\theta \text{ such that : } a < \inf \frac{dQ_0}{dP_0} \leq \sup \frac{dQ_0}{dP_0} < b,$$

with (a, b) is the domain D_φ of the divergence function φ and \mathcal{M}_θ is the set of all signed measures Q a.c. w.r.t. P_0 , satisfying the linear constraints and such that $\frac{dQ}{dP_0}$ belong to $L_q(\mathcal{X}, P_0)$, ($1 \leq q \leq \infty$ and $1/p + 1/q = 1$). In this case, applying Corollary 2 in Borwein and Lewis (1991), we obtain

$$\phi(\mathcal{M}_\theta, P_0) = \sup_{t \in \mathbb{R}^{(l+1)}} \left\{ t_0 - \int \varphi^*(t^T \bar{g}(x, \theta)) dP_0(x) \right\}$$

(with dual attainment). Furthermore, if for a dual optimal c_θ , it holds

$$\lim_{y \downarrow -\infty} \frac{\varphi(y)}{y} < \inf_x c_\theta^T \bar{g}(x, \theta) \leq \sup_x c_\theta^T \bar{g}(x, \theta) < \lim_{y \uparrow +\infty} \frac{\varphi(y)}{y} \text{ for all } x \text{ (} P_0 \text{ a.s.)},$$

then the unique projection Q_θ^* of P_0 on \mathcal{M}_θ is given by

$$(2.25) \quad dQ_\theta^* = \varphi^{*'}(c_\theta^T \bar{g}(\theta)) dP_0.$$

Note that if φ^* is strictly convex, then c_θ is unique and

$$\sup_{t \in \mathbb{R}^{(l+1)}} \left\{ t_0 - \int \varphi^*(t^T \bar{g}(x, \theta)) dP_0(x) \right\} = \sup_{t \in \mathcal{C}_\theta} \left\{ t_0 - \int \varphi^*(t^T \bar{g}(x, \theta)) dP_0(x) \right\}, \text{ and}$$

$$\varphi^{*'}(c_\theta^T \bar{g}(\theta)) = \varphi'^{-1}(c_\theta^T \bar{g}(\theta)).$$

Léonard (2001a) and Léonard (2001b) gives, under minimal conditions, duality theorems of minimum ϕ -divergences and characterization of projections under linear constraints, which generalize the results given by Borwein and Lewis (1991) and Borwein and Lewis (1993). These results are used recently by Bertail (2004) and Bertail (2006) in empirical likelihood.

3. ASYMPTOTIC PROPERTIES AND STATISTICAL TESTS

In the sequel, we assume that the conditions in Proposition 2.1 (or Remark (2.1)) and in Proposition 2.2 (or Remark (2.3)) hold. This allows to use the representations (2.14), (2.18) and (2.19) in order to study the asymptotic behavior of the proposed estimates (1.20), (1.21) and (1.22). All the results in the present Section are obtained through classical methods of parametric statistics; see e.g. van der Vaart (1998) and Sen and Singer (1993). We first consider the case when θ is fixed, and we study the asymptotic behavior of the estimate $\hat{\phi}(\mathcal{M}_\theta, P_0)$ (see (1.20)) of $\phi(\mathcal{M}_\theta, P_0) := \inf_{Q \in \mathcal{M}_\theta} \phi(Q, P_0)$ both when $P_0 \in \mathcal{M}_\theta$ and when $P_0 \notin \mathcal{M}_\theta$. This is done in the first Subsection. In the second Subsection, we study the asymptotic behavior of the EM ϕ D estimates $\hat{\theta}_\phi$ and the estimates $\hat{\phi}(\mathcal{M}, P_0)$ both in the two cases when P_0 belongs to \mathcal{M} and when P_0 does not belong to \mathcal{M} . The solution of *Problem 1* is given in Subsection 3.3 while *Problem 2* is treated in Subsections 3.1, 3.2, 3.3 and 3.4.

3.1. Asymptotic properties of the estimates for a given $\theta \in \Theta$. First we state consistency.

Consistency. We state both weak and strong consistency of the estimates \hat{c}_θ and $\hat{\phi}(\mathcal{M}_\theta, P_0)$ using their representations (2.17) and (2.14), respectively. Denote $\|\cdot\|$ the Euclidian norm defined on \mathbb{R}^d or on \mathbb{R}^{l+1} . In order to state consistency, we need to define

$$T_\theta := \{t \in \mathcal{C}_\theta \text{ such that } P_0 m(\theta, t) > -\infty\},$$

and denote T_θ^c the complementary of the set T_θ in the set \mathcal{C}_θ , namely

$$T_\theta^c := \{t \in \mathcal{C}_\theta \text{ such that } P_0 m(\theta, t) = -\infty\}.$$

Note that, by Proposition 2.2, the set T_θ contains c_θ .

We will consider the following condition

$$(C.1) \quad \sup_{t \in T_\theta} |P_n m(\theta, t) - P_0 m(\theta, t)| \text{ converges to 0 a.s. (resp. in probability);}$$

(C.2) there exists $M < 0$ and $n_0 > 0$, such that, for all $n > n_0$, it holds $\sup_{t \in T_\theta^c} P_n m(\theta, t) \leq M$ a.s. (resp. in probability).

The condition (C.2) makes sense, since for all $t \in T_\theta^c$ we have $P_0 m(\theta, t) = -\infty$.

Since the function $t \in T_\theta \mapsto P_0 m(\theta, t)$ is strictly concave, the maximum c_θ is isolated, that is

$$(3.1) \quad \text{for any positive } \epsilon, \text{ we have } \sup_{\{t \in C_\theta : \|t - c_\theta\| \geq \epsilon\}} P_0 m(\theta, t) < P_0 m(\theta, c_\theta).$$

Proposition 3.1. *Assume that conditions (C.1) and (C.2) hold. Then*

- (i) *the estimates $\hat{\phi}(\mathcal{M}_\theta, P_0)$ converge to $\phi(\mathcal{M}_\theta, P_0)$ a.s. (resp. in probability).*
- (ii) *the estimates \hat{c}_θ converge to c_θ a.s. (resp. in probability).*

Asymptotic distributions. Denote $m'(\theta, t)$ the $(l+1)$ -dimensional vector with entries $\frac{\partial}{\partial t_i} m(\theta, t)$, $m''(\theta, t)$ the $(l+1) \times (l+1)$ -matrix with entries $\frac{\partial^2}{\partial t_i \partial t_j} m(\theta, t)$, $\underline{0}_l := (0, \dots, 0)^T \in \mathbb{R}^l$, $\underline{0}_d := (0, \dots, 0)^T \in \mathbb{R}^d$, \underline{c} the $(l+1)$ -vector defined by $\underline{c} := (0, \underline{0}_l^T)^T$, and $P_0 g(\theta) g(\theta)^T$ the $l \times l$ -matrix defined by

$$P_0 g(\theta) g(\theta)^T := [P_0 g_i(\theta) g_j(\theta)]_{i,j=1,\dots,l}.$$

We will consider the following assumptions

- (A.1) \hat{c}_θ converges in probability to c_θ ;
- (A.2) the function $t \mapsto m(x, \theta, t)$ is \mathcal{C}^3 on a neighborhood $V(c_\theta)$ of c_θ for all x (P_0 -a.s.), and all partial derivatives of order 3 of the function $\{t \mapsto m(x, \theta, t), t \in V(c_\theta)\}$ are dominated by some P_0 -integrable function $x \mapsto H(x)$;
- (A.3) $P_0 (\|m'(\theta, c_\theta)\|^2)$ is finite, and the matrix $P_0 m''(\theta, c_\theta)$ exists and is invertible.

Theorem 3.2. *Assume that assumptions (A.1-3) hold. Then*

- (1) $\sqrt{n}(\hat{c}_\theta - c_\theta)$ converges to a centered normal multivariate variable with covariance matrix

$$(3.2) \quad V = [-P_0 m''(\theta, c_\theta)]^{-1} [P_0 m'(\theta, c_\theta) m'(\theta, c_\theta)^T] [-P_0 m''(\theta, c_\theta)]^{-1}.$$

In the special case, when P_0 belongs to \mathcal{M}_θ , then $c_\theta = \underline{c}$ and

$$(3.3) \quad V = \varphi''(1)^2 \begin{bmatrix} 0 & \underline{0}_l^T \\ \underline{0}_l & [P_0 g(\theta) g(\theta)^T]^{-1} \end{bmatrix}.$$

- (2) *If P_0 belongs to \mathcal{M}_θ , then the statistics*

$$\frac{2n}{\varphi''(1)} \hat{\phi}(\mathcal{M}_\theta, P_0)$$

converge in distribution to a χ^2 variable with l degrees of freedom.

- (3) *If P_0 does not belong to \mathcal{M}_θ , then*

$$\sqrt{n} \left(\hat{\phi}(\mathcal{M}_\theta, P_0) - \phi(\mathcal{M}_\theta, P_0) \right)$$

converges to a centered normal variable with variance

$$\sigma^2 := P_0 m(\theta, c_\theta)^2 - (P_0 m(\theta, c_\theta))^2.$$

Remark 3.1. (a) *When specialized to the modified Kullback-Leibler divergence, Theorem 3.2 part (2) gives the limiting distribution of the empirical log-likelihood ratio $2n \widehat{KL}_m(\mathcal{M}_\theta, P_0)$*

which is the result in Owen (1990) Theorem 1. Part (3) gives its limiting distribution when P_0 does not belong to \mathcal{M}_θ .

- (b) Nonparametric confidence regions (CR_ϕ) for θ_0 of asymptotic level $(1-\epsilon)$ can be constructed using the statistics

$$\frac{2n}{\varphi''(1)} \widehat{\phi}(\mathcal{M}_\theta, P_0),$$

through

$$CR_\phi := \left\{ \theta \in \Theta \text{ such that } \frac{2n}{\varphi''(1)} \widehat{\phi}(\mathcal{M}_\theta, P_0) \leq q_{(1-\epsilon)} \right\},$$

where $(1-\epsilon)$ is the $(1-\epsilon)$ -quantile of a $\chi^2(l)$ distribution. It would be interesting to obtain the divergence leading to optimal confidence regions in the sense of Neyman (1937) (see Takagi (1998)), or the optimal divergence leading to confidence regions with small length (volume, area or diameter) and covering the true value θ_0 with large enough probability.

3.2. Asymptotic properties of the estimates $\widehat{\theta}_\phi$ and $\widehat{\phi}(\mathcal{M}, P_0)$.

Consistency. We assume that when P_0 does not belong to the model \mathcal{M} , the minimum, say θ^* , of the function $\theta \in \Theta \mapsto \inf_{Q \in \mathcal{M}_\theta} \phi(Q, P_0)$ exists and is unique. Hence P_0 admits a projection on \mathcal{M} which we denote Q_{θ^*} . Obviously when P_0 belongs to the model \mathcal{M} , then $\theta^* = \theta_0$ and $Q_{\theta^*} = P_0$. We will consider the following conditions

(C.3) $\sup_{\{\theta \in \Theta, t \in T_\theta\}} |P_n m(\theta, t) - P_0 m(\theta, t)|$ tends to 0 a.s. (resp. in probability);

(C.4) there exists a neighborhood $V(c_{\theta^*})$ of c_{θ^*} such that

- (a) for any positive ϵ , there exists some positive η such that for all $t \in V(c_{\theta^*})$ and all $\theta \in \Theta$ satisfying $\|\theta - \theta^*\| \geq \epsilon$, it holds $P_0 m(\theta^*, t) < P_0 m(\theta, t) - \eta$;
- (b) there exists some function H such that for all t in $V(c_{\theta^*})$, we have $|m(t, \theta_0)| \leq H(x)$ (P_0 -a.s.) with $P_0 H < \infty$;

(C.5) there exists $M < \infty$ and $n_0 > 0$ such that for all $n \geq n_0$, we have

$$(3.4) \quad \sup_{\theta \in \Theta} \sup_{t \in T_\theta^c} P_n m(\theta, t) \leq M \text{ a.s. (resp. in probability).}$$

Proposition 3.3. Assume that conditions (C.3-5) hold. Then

- (i) the estimates $\widehat{\phi}(\mathcal{M}, P_0)$ converge to $\phi(\mathcal{M}, P_0)$ a.s. (resp. in probability).
- (ii) $\sup_{\theta \in \Theta} \|\widehat{c}_\theta - c_\theta\|$ converge to 0 a.s. (resp. in probability).
- (iii) The ME ϕ D estimates $\widehat{\theta}_\phi$ converge to θ^* a.s. (resp. in probability).

Asymptotic distributions. When $P_0 \in \mathcal{M}$, then by assumption, there exists unique $\theta_0 \in \Theta$ such that $P_0 \in \mathcal{M}_{\theta_0}$. Hence $\theta^* = \theta_0$ and $c_{\theta^*} = c_{\theta_0} = \underline{c}$. We state the limit distributions of the estimates $\widehat{\theta}_\phi$ and $\widehat{c}_{\widehat{\theta}_\phi}$ when $P_0 \in \mathcal{M}$ and when $P_0 \notin \mathcal{M}$. We will make use of the following assumptions

(A.4) Both estimates $\widehat{\theta}_\phi$ and $\widehat{c}_{\widehat{\theta}_\phi}$ converge in probability respectively to θ^* and c_{θ^*} ;

(A.5) the function $(\theta, t) \mapsto m(x, \theta, t)$ is \mathcal{C}^3 on some neighborhood $V(\theta^*, c_{\theta^*})$ for all x (P_0 -a.s.), and the partial derivatives of order 3 of the functions

$\{(\theta, t) \mapsto m(x, \theta, t), (\theta, t) \in V(\theta^*, c_{\theta^*})\}$ are dominated by some P_0 -integrable function $H(x)$;

(A.6) $P_0 \left(\left\| \frac{\partial}{\partial t} m(\theta^*, c_{\theta^*}) \right\|^2 \right)$ and $P_0 \left(\left\| \frac{\partial}{\partial \theta} m(\theta^*, c_{\theta^*}) \right\|^2 \right)$ are finite, and the matrix

$$S := \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix},$$

with $S_{11} := P_0 \frac{\partial^2}{\partial t^2} m(\theta^*, c_{\theta^*})$, $S_{12} = S_{21}^T := P_0 \frac{\partial^2}{\partial t \partial \theta} m(\theta^*, c_{\theta^*})$ and $S_{22} := P_0 \frac{\partial^2}{\partial \theta^2} m(\theta^*, c_{\theta^*})$, exists and is invertible.

Theorem 3.4. *Let P_0 belongs to \mathcal{M} and assumptions (A.4-6) hold. Then, both $\sqrt{n}(\widehat{\theta}_\phi - \theta_0)$ and $\sqrt{n}(\widehat{c}_{\widehat{\theta}_\phi} - c)$ converge in distribution to a centered multivariate normal variable with covariance matrix, respectively*

$$(3.5) \quad V = \left\{ \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right] [P_0 (g(\theta_0)g(\theta_0)^T)]^{-1} \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right]^T \right\}^{-1},$$

and

$$\begin{aligned} U &= \varphi''(1)^2 \begin{bmatrix} 0 & \underline{0}_l^T \\ \underline{0}_l & [P_0 g(\theta_0)g(\theta_0)^T]^{-1} \end{bmatrix} - \varphi''(1)^2 \begin{bmatrix} 0 & \underline{0}_l^T \\ \underline{0}_l & [P_0 g(\theta_0)g(\theta_0)^T]^{-1} \end{bmatrix} \times \\ &\times \begin{bmatrix} \underline{0}_d, P_0 \frac{\partial}{\partial \theta} g(\theta_0) \end{bmatrix}^T V \begin{bmatrix} \underline{0}_d, P_0 \frac{\partial}{\partial \theta} g(\theta_0) \end{bmatrix} \begin{bmatrix} 0 & \underline{0}_l^T \\ \underline{0}_l & [P_0 g(\theta_0)g(\theta_0)^T]^{-1} \end{bmatrix}, \end{aligned}$$

and the estimates $\widehat{\theta}_\phi$ and $\widehat{c}_{\widehat{\theta}_\phi}$ are asymptotically uncorrelated.

Remark 3.2. *When specialized to the modified Kullback-Leibler divergence, the estimate $\widehat{\theta}_{KL_m}$ is the empirical likelihood estimate (ELE) (noted $\tilde{\theta}$ in Qin and Lawless (1994)), and the above result gives the limiting distribution of $\sqrt{n}(\widehat{\theta}_{KL_m} - \theta_0)$ which coincides with the result in Theorem 1 in Qin and Lawless (1994). Note also that all ME ϕ DE's including ELE have the same limiting distribution with the same variance when P_0 belongs to \mathcal{M} . Hence they are all equally first order efficient.*

Theorem 3.5. *Assume that P_0 does not belong to \mathcal{M} and that assumptions (A.4-6) hold. Then*

$$\sqrt{n} \begin{pmatrix} \widehat{c}_{\widehat{\theta}_\phi} - c_{\theta^*} \\ \widehat{\theta}_\phi - \theta^* \end{pmatrix}$$

converges in distribution to a centered multivariate normal variable with covariance matrix

$$W = S^{-1}MS^{-1}$$

where

$$M := P_0 \left(\begin{pmatrix} \frac{\partial}{\partial \theta} m(\theta^*, c_{\theta^*}) \\ \frac{\partial}{\partial \theta} m(\theta^*, c_{\theta^*}) \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial \theta} m(\theta^*, c_{\theta^*}) \\ \frac{\partial}{\partial \theta} m(\theta^*, c_{\theta^*}) \end{pmatrix}^T \right).$$

θ^* and c_{θ^*} are characterized by

$$\theta^* := \arg \inf_{\theta \in \Theta} \phi(\mathcal{M}_\theta, P_0),$$

$$dQ_{\theta^*}^* = \varphi'^{-1}(c_{\theta^*}^T \bar{g}(\theta)) dP_0 \quad \text{and} \quad Q_{\theta^*}^* \in \mathcal{M}_{\theta^*}.$$

3.3. Tests of model. In order to test the hypothesis $\mathcal{H}_0 : P_0$ belongs to \mathcal{M} against the alternative $\mathcal{H}_1 : P_0$ does not belong to \mathcal{M} , we can use the estimates $\widehat{\phi}(\mathcal{M}, P_0)$ of $\phi(\mathcal{M}, P_0)$, the ϕ -divergences between the model \mathcal{M} and the distribution P_0 . Since $\phi(\mathcal{M}, P_0)$ is nonnegative and take value 0 only when P_0 belongs to \mathcal{M} (provided that P_0 admits a projection on \mathcal{M}), we reject the hypothesis \mathcal{H}_0 when the estimates take large values. In the following Corollary, we give the asymptotic law of the estimates $\widehat{\phi}(\mathcal{M}, P_0)$ both under \mathcal{H}_0 and under \mathcal{H}_1 .

Corollary 3.6.

- (i) *Assume that the assumptions of Theorem 3.4 hold and that $l > d$. Then, under \mathcal{H}_0 , the statistics*

$$\frac{2n}{\varphi''(1)} \widehat{\phi}(\mathcal{M}, P_0)$$

converge in distribution to a χ^2 variable with $(l - d)$ degrees of freedom.

(ii) Assume that the assumptions of Theorem 3.5 hold. Then, under \mathcal{H}_1 , we have :

$$(3.6) \quad \sqrt{n} \left(\widehat{\phi}(\mathcal{M}, P_0) - \phi(\mathcal{M}, P_0) \right)$$

converges to centered normal variable with variance

$$\sigma^2 = P_0 m(\theta^*, c_{\theta^*})^2 - (P_0 m(\theta^*, c_{\theta^*}))^2$$

where θ^* and c_{θ^*} satisfy

$$\theta^* := \arg \inf_{\theta \in \Theta} \phi(\mathcal{M}_\theta, P_0),$$

$$\varphi' \left(\frac{dQ_{\theta^*}^*}{dP_0}(x) \right) = c_{\theta^*}^T \bar{g}(x, \theta^*) \quad \text{and} \quad Q_{\theta^*}^* \in \mathcal{M}_{\theta^*}.$$

Remark 3.3. This Theorem allows to perform tests of model of asymptotic level α ; the critical regions are

$$(3.7) \quad C_\phi := \left\{ \frac{2n}{\varphi''(1)} \widehat{\phi}(\mathcal{M}, P_0) > q_{(1-\alpha)} \right\},$$

where $q_{(1-\alpha)}$ is the $(1-\alpha)$ -quantile of the χ^2 distribution with $(l-d)$ degrees of freedom. Also these tests are all asymptotically powerful, since the estimates $\widehat{\phi}(\mathcal{M}, P_0)$ are n -consistent estimates of $\phi(\mathcal{M}, P_0) = 0$ under \mathcal{H}_0 and \sqrt{n} -consistent estimates of $\phi(\mathcal{M}, P_0)$ under \mathcal{H}_1 .

We assume now that the p.m. P_0 belongs to \mathcal{M} . We will perform simple and composite tests on the parameter θ_0 taking into account of the information $P_0 \in \mathcal{M}$.

3.4. Simple tests on the parameter. Let

$$(3.8) \quad \mathcal{H}_0 : \theta_0 = \theta_1 \quad \text{versus} \quad \mathcal{H}_1 : \theta_0 \in \Theta \setminus \{\theta_1\},$$

where θ_1 is a given known value. We can use the following statistics to perform tests pertaining to (3.8)

$$S_n^\phi := \widehat{\phi}(\mathcal{M}_{\theta_1}, P_0) - \inf_{\theta \in \Theta} \widehat{\phi}(\mathcal{M}_\theta, P_0).$$

Since

$$\phi(\mathcal{M}_{\theta_1}, P_0) - \inf_{\theta \in \Theta} \phi(\mathcal{M}_\theta, P_0) = \phi(\mathcal{M}_{\theta_1}, P_0)$$

are nonnegative and take value 0 only when $\theta_0 = \theta_1$, we reject the hypothesis \mathcal{H}_0 when the statistics S_n^ϕ take large values.

We give the limit distributions of the statistics S_n^ϕ in the following Corollary which we can prove using some algebra and arguments used in the proof of Theorem 3.4 and Theorem 3.5.

Corollary 3.7.

(i) Assume that assumptions of Theorem 3.4 hold. Then under \mathcal{H}_0 , the statistics

$$\frac{2n}{\varphi''(1)} S_n^\phi$$

converge in distribution to χ^2 variable with d degrees of freedom.

(ii) Assume that assumptions of Theorem 3.4 hold. Then under \mathcal{H}_1 ,

$$\sqrt{n} (S_n^\phi - \phi(\mathcal{M}_{\theta_1}, P_0))$$

converges to a centered normal variable with variance

$$\sigma^2 = P_0 m(\theta_1, c_{\theta_1})^2 - (P_0 m(\theta_1, c_{\theta_1}))^2.$$

Remark 3.4. When specialized to the KL_m -divergence, the statistic $2nS_n^{KL_m}$ is the empirical likelihood ratio statistic (see Qin and Lawless (1994) Theorem 2).

3.5. Composite tests on the parameter. Let

$$(3.9) \quad h : \mathbb{R}^d \rightarrow \mathbb{R}^k$$

be some function such that the $(d \times k)$ -matrix $H(\theta) := \frac{\partial}{\partial \theta} h(\theta)$ exists, is continuous and has rank k with $0 < k < d$. Let us define the composite null hypothesis

$$(3.10) \quad \Theta_0 := \{\theta \in \Theta \text{ such that } h(\theta) = 0\}.$$

We consider the composite test

$$(3.11) \quad \mathcal{H}_0 : \theta_0 \in \Theta_0 \text{ versus } \mathcal{H}_1 : \theta_0 \in \Theta \setminus \Theta_0,$$

i.e., the test

$$(3.12) \quad \mathcal{H}_0 : P_0 \in \bigcup_{\theta \in \Theta_0} \mathcal{M}_\theta \text{ versus } \mathcal{H}_1 : P_0 \in \bigcup_{\theta \in \Theta \setminus \Theta_0} \mathcal{M}_\theta.$$

This test is equivalent to the following one

$$(3.13) \quad \mathcal{H}_0 : \theta_0 \in f(B_0) \text{ versus } \mathcal{H}_1 : \theta_0 \notin f(B_0),$$

where $f : \mathbb{R}^{(d-k)} \rightarrow \mathbb{R}^d$ is a function such that the matrix $G(\beta) := \frac{\partial}{\partial \beta} g(\beta)$ exists and has rank $(d - k)$, and $B_0 := \{\beta \in \mathbb{R}^{(d-k)} \text{ such that } f(\beta) \in \Theta_0\}$. Therefore $\theta_0 \in \Theta_0$ is an equivalent statement for $\theta_0 = f(\beta_0), \beta_0 \in B_0$.

The following statistics are used to perform tests pertaining to (3.13) :

$$T_n^\phi := \inf_{\beta \in B_0} \hat{\phi}(\mathcal{M}_{f(\beta)}, P_0) - \inf_{\theta \in \Theta} \hat{\phi}(\mathcal{M}_\theta, P_0).$$

Since

$$\inf_{\beta \in B_0} \phi(\mathcal{M}_{f(\beta)}, P_0) - \inf_{\theta \in \Theta} \phi(\mathcal{M}_\theta, P_0) = \inf_{\beta \in B_0} \phi(\mathcal{M}_{f(\beta)}, P_0)$$

are nonnegative and take value 0 only when \mathcal{H}_0 holds, we reject the hypothesis \mathcal{H}_0 when the statistics T_n^ϕ take large values.

We give the limit distributions of the statistics T_n^ϕ in the following Corollary.

Corollary 3.8.

- (i) Assume that assumptions of Theorem 3.4 hold. Under \mathcal{H}_0 , the statistics T_n^ϕ converge in distribution to a χ^2 variable with $(d - k)$ degrees of freedom.
- (ii) Assume that there exists $\beta^* \in B_0$, such that $\beta^* = \arg \inf_{\beta \in B_0} \phi(\mathcal{M}_{f(\beta)}, P_0)$. If the assumptions of Theorem 3.5 hold for $\theta^* = f(\beta^*)$, then

$$\sqrt{n} (T_n^\phi - \phi(\mathcal{M}_{\theta^*}, P_0))$$

converges to a centered normal variable with variance

$$\sigma^2 = P_0 m(\theta^*, c_{\theta^*})^2 - (P_0 m(\theta^*, c_{\theta^*}))^2.$$

4. ESTIMATES OF THE DISTRIBUTION FUNCTION THROUGH PROJECTED DISTRIBUTIONS

In this Subsection, the measurable space $(\mathcal{X}, \mathcal{B})$ is $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. For all ϕ -divergence, by (1.21), we have

$$\widehat{\phi}(\mathcal{M}, P_0) = \phi(\mathcal{M}, P_n) = \phi\left(\widehat{Q_{\theta_\phi}^*}, P_n\right).$$

Proposition 2.4 above provides the description of $\widehat{Q_{\theta_\phi}^*}$.

So, for all ϕ -divergence, we estimate the distribution function F using $\widehat{Q_{\theta_\phi}^*}$ the ϕ -projection of P_n on \mathcal{M} , through

$$\begin{aligned} \widehat{F}_n(x) &:= \sum_{i=1}^n \widehat{Q_{\theta_\phi}^*}(X_i) \mathbb{1}_{(-\infty, x]}(X_i) \\ (4.1) \quad &= \frac{1}{n} \sum_{i=1}^n \varphi' \left(\widehat{c_{\theta_\phi}^T} \bar{g}(X_i, \theta_\phi) \right) \mathbb{1}_{(-\infty, x]}(X_i). \end{aligned}$$

Remark 4.1. When the estimating equation

$$(4.2) \quad \frac{1}{n} \sum_{i=1}^n g(X_i, \theta) = \underline{0}_d$$

admits a solution $\widetilde{\theta}_n$, then P_n belongs to \mathcal{M} . If the solution is unique then $\widehat{\theta}_\phi = \widetilde{\theta}_n$. Hence by Proposition 2.1

$$\text{for all } i \in \{1, 2, \dots, n\}, \text{ we have } \widehat{Q_{\theta_\phi}^*}(X_i) = \frac{1}{n},$$

and $\widehat{F}_n(x)$, in this case, is the empirical cumulative distribution function, i.e.,

$$\widehat{F}_n(x) = F_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, x]}(X_i).$$

So, the main interest is in the case where (4.2) does not admit a solution, that is in general when $l > d$.

Remark 4.2. The ϕ -projections $\widehat{Q_{\theta_\phi}^*}$ of P_n on \mathcal{M} may be signed measures. For all ϕ -divergence satisfying $D_\phi = \mathbb{R}_+^*$, the ϕ -projection $\widehat{Q_{\theta_\phi}^*}$ is a p.m. if it exists. (for example, KL_m , KL , Hellinger, and χ_m^2 divergences all provide p.m.'s).

We give the limit law of the estimates \widehat{F}_n of the distribution function F in the following Theorem. We will see that the estimate $\widehat{F}_n(x)$ is generally more efficient than the empirical cumulative distribution function $F_n(x)$.

Theorem 4.1. Under the assumptions of Theorem 3.4, $\sqrt{n} \left(\widehat{F}_n(x) - F(x) \right)$ converges in distribution to a centered normal variable with variance

$$(4.3) \quad W(x) = F(x) (1 - F(x)) - [P_0(g(\theta_0) \mathbb{1}_{(-\infty, x]})]^T \Gamma [P_0(g(\theta_0) \mathbb{1}_{(-\infty, x]})],$$

with

$$\begin{aligned} \Gamma &= [P_0 g(\theta_0) g(\theta_0)^T]^{-1} - [P_0 g(\theta_0) g(\theta_0)^T]^{-1} \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right]^T V \times \\ &\quad \times \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right] [P_0 g(\theta_0) g(\theta_0)^T]^{-1}, \end{aligned}$$

and

$$V = \left\{ \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right] [P_0 (g(\theta_0)g(\theta_0)^T)]^{-1} \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right]^T \right\}^{-1}.$$

5. EMPIRICAL LIKELIHOOD AND RELATED METHODS

In the present setting, the empirical likelihood (EL) approach for the estimation of the parameter θ_0 can be summarized as follows. For any θ in Θ , define the *profile likelihood ratio* of the sample $X := (X_1, \dots, X_n)$ through

$$L_n(\theta) := \sup \left\{ \prod_{i=1}^n nQ(X_i) \text{ where } Q(X_i) \geq 0, \sum_{i=1}^n Q(X_i) = 1, \sum_{i=1}^n g(X_i, \theta)Q(X_i) = 0 \right\}.$$

The estimate of θ_0 through empirical likelihood (EL) approach is then defined by

$$(5.1) \quad \hat{\theta}_{EL} := \arg \sup_{\theta \in \Theta} L_n(\theta).$$

The paper by Qin and Lawless (1994) introduces $\hat{\theta}_{EL}$ and presents its properties. In this Section, we show that $\hat{\theta}_{EL}$ belongs to the family of ME ϕ D estimates for the specific choice $\varphi(x) = -\log x + x - 1$. We also discuss the problem of the existence of the solution of (5.1) for all n .

When $\varphi(x) = -\log x + x - 1$, formula (1.22) clearly coincides with $\hat{\theta}_{EL}$. For test of hypotheses given by $\mathcal{H}_0 : P_0 \in \mathcal{M}_\theta$ against $\mathcal{H}_1 : P_0 \notin \mathcal{M}_\theta$ or for construction of nonparametric confidence regions for θ_0 , the statistic $2n\widehat{KL}_m(\mathcal{M}_\theta, P_0)$ coincides with the empirical log-likelihood ratio introduced in Owen (1988), Owen (1990) and Qin and Lawless (1994). We state the results of Section 3 in the present context. We will see that the approach of empirical likelihood by divergence minimization, using the dual representation of the KL_m -divergence and the explicit form of the KL_m -projection of P_0 , yields to the limit distribution of the statistic $2n\widehat{KL}_m(\mathcal{M}_\theta, P_0)$ under \mathcal{H}_1 , which can not be achieved using the approach in Owen (1990) and Qin and Lawless (1994). Consider

$$\hat{\theta}_{KL_m} = \arg \inf_{\theta \in \Theta} \widehat{KL}_m(\mathcal{M}_\theta, P_0)$$

where

$$(5.2) \quad \widehat{KL}_m(\mathcal{M}_\theta, P_0) = \sup_{t \in \mathcal{C}_\theta} P_n m(\theta, t)$$

with $\varphi(x) = \varphi_0(x) = -\log x + x - 1$. The explicit form of $m(\theta, t)$ in this case is

$$(5.3) \quad \begin{aligned} x \mapsto m(x, \theta, t) &= t_0 - (t^T \bar{g}(x, \theta)) \frac{1}{1 - t^T \bar{g}(x, \theta)} + \log(1 - t^T \bar{g}(x, \theta)) + \frac{1}{1 - t^T \bar{g}(x, \theta)} - 1. \\ &= t_0 + \log(1 - t^T \bar{g}(x, \theta)). \end{aligned}$$

For fixed $\theta \in \Theta$, the sup in (5.2), which we have noted \hat{c}_θ , satisfies the following system

$$(5.4) \quad \begin{cases} \int \frac{1}{1 - c_0 - \sum_{j=1}^l c_j g_j(x, \theta)} dP_n(x) &= 1 \\ \int \frac{g_j(x, \theta)}{1 - c_0 - \sum_{j=1}^l c_j g_j(x, \theta)} dP_n(x) &= 0, \text{ for all } j = 1, \dots, l \end{cases}$$

a system of $(l + 1)$ equations and $(l + 1)$ variables. The projection \widehat{Q}_θ^* is then obtained using Proposition 2.1 part (ii). We have for all $i \in \{1, \dots, n\}$

$$\frac{1}{\widehat{Q}_\theta^*(X_i)} = n \left(1 - c_0 - \sum_{j=1}^l c_j g_j(X_i, \theta) \right)$$

which, multiplying by $\widehat{Q}_\theta^*(X_i)$ and summing upon i yields $c_0 = 0$. Therefore the system (5.4) reduces to the system (3.3) in Qin and Lawless (1994) replacing c_1, \dots, c_l by $-t_1, \dots, -t_l$. Simplify (5.3) plugging $t_0 = 0$. Notice that $2n\widehat{KL}_m(\mathcal{M}_\theta, P_0) = l_E(\theta_0)$ in the notation of Qin and Lawless (1994), and that the function of $t = (0, -\tau_1, \dots, -\tau_l)$ defined by

$$t \mapsto P_n m(\theta, t)$$

coincide with the function

$$\tau \rightarrow P_n \log(1 + \tau^T g(\cdot, \theta))$$

used in Qin and Lawless (1994). The interest in formula (5.2) lays in the obtention of the limit distributions of $2n\widehat{KL}_m(\mathcal{M}_\theta, P_0)$ under \mathcal{H}_1 . By Theorem 3.2, we have

$$\sqrt{n} \left(\widehat{KL}_m(\mathcal{M}_\theta, P_0) - KL_m(\mathcal{M}_\theta, P_0) \right)$$

converges to a normal distribution variable, which proves consistency of the test; this results cannot be obtained by the Qin and Lawless (1994)'s approach.

The choice of φ depends on some a priori knowledge on θ_0 . Hopefully, some divergences do not have such an inconvenient. We now clarify this point. For fixed θ in Θ , let $\mathcal{M}_\theta^{(n)}$ and $\mathcal{D}_\phi^{(n)}$ be defined respectively as in (1.19) and in (2.2). Assume that $\mathcal{M}_\theta^{(n)} \cap \mathcal{D}_\phi^{(n)}$ is not void. Then P_n has a projection \widehat{Q}_θ^* on $\mathcal{M}_\theta^{(n)}$ and $\phi(\widehat{Q}_\theta^*, P_n)$ is finite. The estimation of θ_0 is achieved minimizing $\widehat{\phi}(\mathcal{M}_\theta, P_0)$ on the sets

$$\Theta_n^\phi := \left\{ \theta \in \Theta \text{ such that } \mathcal{M}_\theta^{(n)} \cap \mathcal{D}_\phi^{(n)} \text{ is not void} \right\}.$$

Clearly the description of Θ_n^ϕ depends on the divergence ϕ . Consider the following example, with $n = 2$, $X = (X_1, X_2)$ and $g(x, \theta) = x - \theta$. Then

$$\mathcal{M}_\theta = \left\{ (q_1, q_2)^T \text{ such that } q_1 + q_2 = 1 \text{ and } q_1(X_1 - \theta) + q_2(X_2 - \theta) = 0 \right\}$$

and

$$\mathcal{D}_\phi^{(2)} = \left\{ (q_1, q_2) \text{ such that } \frac{1}{2} \sum_{i=1}^2 \varphi(2q_i) < \infty \right\}.$$

When $\phi = KL_m$, then $\mathcal{D}_{KL_m}^{(2)} = \mathbb{R}_+^* \times \mathbb{R}_+^*$. So, according to the value of θ , $\mathcal{M}_\theta^{(n)} \cap \mathcal{D}_{KL_m}^{(n)}$ may be void and therefore $\Theta_n^{KL_m}$ has a complex structure. At the opposite, for example when $\phi = \chi^2$, then $\mathcal{D}_{\chi^2}^{(2)} = \mathbb{R}^2$. Hence $\mathcal{M}_\theta^{(n)} \cap \mathcal{D}_{\chi^2}^{(n)} = \mathcal{M}_\theta^{(n)}$ which is not void for all θ and hence $\Theta_n^{\chi^2} = \Theta$.

On the other hand, we have for any ϕ -divergence

$$\begin{aligned} \widehat{\theta}_\phi &:= \arg \inf_{\theta \in \Theta} \inf_{Q \in \mathcal{M}_\theta^{(n)}} \widehat{\phi}(Q, P_0) \\ &= \arg \inf_{\theta \in \Theta} \inf_{Q \in \mathcal{M}_\theta^{(n)} \cap \mathcal{D}_\phi^{(n)}} \widehat{\phi}(Q, P_0). \end{aligned}$$

When $\mathcal{D}_\phi^{(n)} \neq \mathbb{R}^n$, the infimum in θ above should be taken upon Θ_n^ϕ which might be quite cumbersome. Owen (2001) indeed mentions such a difficulty.

In relation to this problem, Qin and Lawless (1994) bring some asymptotic arguments in the case of the empirical likelihood. They show that there exists a sequence of neighborhoods

$$V_n(\theta_0) := \left\{ \theta \text{ such that } \|\theta - \theta_0\| \leq n^{-1/3} \right\}$$

on which, with probability one as n tends to infinity, $L_n(\theta)$ has a maximum. This turns out, in the context of ϕ -divergences, to write that the mapping

$$\theta \mapsto \inf_{Q \in \mathcal{M}_\theta^{(n)}} KL_m(Q, P_n)$$

has a minimum when θ belongs to $V_n(\theta_0)$. This interesting result does not solve the problem for fixed n , as θ_0 is unknown. For such problem, the use of ϕ -divergences, satisfying $\mathcal{D}_\phi^{(n)} = \mathbb{R}^n$ (for example χ^2 -divergence), might give information about θ_0 and localizes it through ϕ -divergence confidence regions (CR_ϕ 's).

The choice of the divergence ϕ also depends upon some knowledge on the support of the unknown p.m. P_0 . When P_0 has a projection on \mathcal{M} with same support as P_0 , Proposition 2.2 yields its description and its explicit calculation. A necessary condition for this is that \mathcal{C}_θ , as defined in (2.9), has non void interior in $\mathbb{R}^{(l+1)}$. Consider the case of the empirical likelihood, that is when $\varphi(x) = -\log x + x - 1$; then $\text{Im } \varphi' =]-\infty, 1[$. Consider $g(x, \theta) = x - \theta$, i.e., a constraint on the mean. Assume that the support of P_0 is unbounded. Then

$$\mathcal{C}_\theta = \{t \in \mathbb{R}^2 \text{ such that for all } x (P_0 - a.s.) , t_0 + t_1(x - \theta) \in]-\infty, 1[\}.$$

Therefore, $t_1 = 0$ and $\mathcal{C}_\theta =]-\infty, 1[\times\{0\}$ which implies that the interior of \mathcal{C}_θ is void. This results indicates that the support of Q^* is not the same as the support of P_0 . Hence in this case we cannot use the dual representation of $KL_m(\mathcal{M}_\theta, P_0)$. The arguments used in Section 3 for the obtention of limiting distributions cannot be used, if the support of P_0 is unbounded, in order to obtain the limiting distribution of the estimates $\widehat{KL_m}(\mathcal{M}_\theta, P_0)$ under \mathcal{H}_1 (i.e., when P_0 does not belong to \mathcal{M}_θ). We thus cannot conclude in this case that the tests pertaining to θ_0 are consistent.

6. ROBUSTNESS AND EFFICIENCY OF ME ϕ D ESTIMATES AND SIMULATION RESULTS

Lindsay (1994) introduced a general instrument for the study of the asymptotic properties of parametric estimates by minimum ϕ -divergences, called Residual Adjustment Function (RAF). We first recall its definition. Let $\{P_\theta ; \theta \in \Theta\}$ be some parametric model defined on a finite set \mathcal{X} . Let X_1, \dots, X_n a sample with distribution P_{θ_0} . A minimum ϕ -divergence estimate (M ϕ DE) (called also minimum disparity estimator) of θ_0 is given by

$$(6.1) \quad \tilde{\theta}_\phi := \arg \inf_{\theta \in \Theta} \sum_{x \in \mathcal{X}} \varphi \left(\frac{P_\theta(x)}{P_n(x)} \right) P_n(x),$$

where $P_n(x)$ is the proportion of the sample point that take value x . When the parametric model $\{P_\theta : \theta \in \Theta\}$ is regular, then $\tilde{\theta}_\phi$ is solution of the equation

$$(6.2) \quad \sum_{x \in \mathcal{X}} \varphi' \left(\frac{P_\theta(x)}{P_n(x)} \right) \dot{P}_\theta(x) = 0,$$

which can be written as

$$(6.3) \quad \sum_{x \in \mathcal{X}} A_\varphi(\delta(x)) \dot{P}_\theta(x) = 0.$$

In this display, $A_\varphi(u) := \varphi' \left(\frac{1}{u+1} \right)$ depends only upon the divergence function φ and

$$\delta(x) := \frac{P_n(x)}{P_\theta(x)} - 1$$

is the ‘‘Pearson Residual’’ at x which belongs to $] -1, +\infty[$. The function $A_\varphi(\cdot)$ is the RAF.

The points x for which $\delta(x)$ is close to -1 are called “inliers”, whereas points x such that $\delta(x)$ is large are called “outliers”. Efficiency properties are linked with the behavior of $A_\varphi(\cdot)$ in the neighborhood of 0 (see Lindsay (1994) Proposition 3 and Basu and Lindsay (1994)) : the smaller the value of $|A''_\varphi(0)|$, the more second efficient the estimate $\tilde{\theta}_\phi$ in the sense of Rao (1961).

It is easy to verify that the RAF's of the power divergences ϕ_γ , defined by the divergence functions in (1.4), have the form

$$(6.4) \quad A_\gamma(\delta) = \frac{(\delta + 1)^{1-\gamma} - 1}{(\gamma - 1)}.$$

In particular, the $M\phi_\gamma$ DE of (6.2) with the RAF in (6.4) corresponds to the maximum likelihood when $\gamma = 0$, minimum Hellinger distance when $\gamma = 0.5$, minimum χ^2 divergence when $\gamma = 2$, minimum modified χ^2 divergence when $\gamma = -1$ and minimum KL divergence when $\gamma = 1$.

From (6.4), we see that $A''_\gamma(0) = \gamma$. Hence for the maximum likelihood estimate, we have $|A''_\gamma(0)| = |A''_0(0)| = 0$ which is the smallest value of $|A''_\gamma(0)|$, $\gamma \in \mathbb{R}$. Therefore, according to Proposition 3 in Lindsay (1994), the maximum likelihood estimate is the most second-order efficient estimate (in the sense of Rao (1961)) among all minimum power divergences estimates.

Robustness features of $\tilde{\theta}_\phi$ against inliers and outliers are related to the variations of $A_\varphi(u)$ or $\varphi(x)$ when u or x close to -1 and $+\infty$, respectively as seen through the following heuristic arguments. Let ϕ_1 and ϕ_2 two divergences associated to the functions φ_1 and φ_2 . If

$$\lim_{x \downarrow 0} \frac{\varphi_1(x)}{\varphi_2(x)} = +\infty,$$

then the estimating equation (6.2) corresponding to φ_1 is not as stable as that corresponding to φ_2 , and hence the $ME\phi_2$ DE is more robust than $ME\phi_1$ DE against outliers. If

$$\lim_{x \uparrow +\infty} \frac{\varphi_1(x)}{\varphi_2(x)} = +\infty,$$

then the estimating equation (6.2) corresponding to φ_1 is not as stable as that corresponding to φ_2 , and hence the $ME\phi_2$ DE is more robust than $ME\phi_1$ DE against inliers.

In all cases, the divergence associated to the divergence function having the smallest variations on its domain leads to the most robust estimate against both outliers and inliers.

It is shown also in Jiménez and Shao (2001) that no minimum power divergence estimate (including the maximum likelihood one) is better than the minimum Hellinger divergence in terms of both second-order efficiency and robustness.

In the examples below, we compare by simulations the efficiency and robustness properties of some $ME\phi$ DE's for some models satisfying linear constraints. We will see that the minimum empirical Hellinger divergence estimate represents a suitable compromise between efficiency and robustness. A theoretical study of efficiency and robustness properties of $ME\phi$ DE's is necessary and should involve second-order efficiency versus robustness since all $ME\phi$ DE's are all equally first-order efficient (see Remark 3.2 and Theorem 3.4).

Numerical Results. We consider for illustration the same model as in Qin and Lawless (1994) Section 5 Example 1. The model \mathcal{M}_θ (see 1.12) here is the set of all signed finite measures Q

satisfying

$$(6.5) \quad \int dQ = 1 \quad \text{and} \quad \int g(x, \theta) dQ(x) = 0,$$

with $g(x, \theta) = ((x - \theta), (x^2 - 2\theta^2 - 1))^T$ and θ , the parameter of interest, belongs to \mathbb{R} .

In Examples 1.a and 1.b below, we compare the efficiency property of various estimates : we generate 1000 pseudorandom samples of sizes 25, 50, 75 and 100 from a normal distribution with mean θ_0 and variance $\theta_0^2 + 1$ (i.e., $P_0 = \mathcal{N}(\theta_0, \theta_0^2 + 1)$) for two values of θ_0 : $\theta_0 = 0$ in Example 1.a and $\theta_0 = 1$ in Example 1.b. Note that P_0 satisfies (6.5).

For each sample, we consider various estimates of θ_0 : the sample mean estimate (SME), the parametric ML estimate (MLE) based on the normal distribution $\mathcal{N}(\theta, \theta^2 + 1)$ and ME ϕ D estimates $\hat{\theta}_\phi$ associated to the divergences : $\phi = \chi_m^2$, H , KL , χ^2 and KL_m -divergence (which coincides with the MEL one, i.e., ME KL_m E=MELE).

For all divergence ϕ considered, in order to calculate the ME ϕ DE $\hat{\theta}_\phi$, we first calculate $\hat{\phi}(\mathcal{M}_\theta, P_0)$ for all given θ (using the representation (2.14)) by Newton's method, and then minimize it to obtain $\hat{\theta}_\phi$.

The results of Theorem 3.4 show that for all ϕ -divergence

$$\sqrt{n}(\hat{\theta}_\phi - \theta_0) \rightarrow \mathcal{N}(0, V)$$

where V is independent of the divergence ϕ ; it is given in Theorem 3.4. For the present model, following Qin and Lawless (1994), V writes

$$(6.6) \quad V = \text{Var}(X) - \Delta^{-1} [m'(\theta_0)\text{Var}(X) + \theta_0 m(\theta_0) - E(X^3)]^2$$

where $\Delta = E[m'(\theta_0)(X - \theta_0) + m(\theta_0) - X^2]^2$ and $m(\theta) := 2\theta^2 + 1$. Thus $V \leq \text{Var}(X)$ which is the variance of $\sqrt{n}(\bar{X}_n - \theta_0)$ with $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$, the sample mean estimate (SME) of θ_0 . So, EM ϕ D estimates are all asymptotically at least as efficient as \bar{X}_n .

6.1. Example 1.a. In this example the true value of the parameter is $\theta_0 = 0$.

n	ME χ_m^2 DE		ME KL_m DE=MELE		ME H DE		ME KL DE	
	mean	var	mean	var	mean	var	mean	var
25	0.0089	0.0314	0.0086	0.0315	0.0084	0.0315	0.0082	0.0314
50	-0.0116	0.0209	-0.0118	0.0210	-0.0119	0.0210	-0.0120	0.0210
75	-0.0025	0.0171	-0.0024	0.0170	-0.0023	0.0170	-0.0022	0.0169
100	-0.0172	0.0112	-0.0174	0.0111	-0.0174	0.0111	-0.0175	0.0112
n	ME χ^2 DE		PMLE		SME			
	mean	var	mean	var	mean	var		
25	0.0077	0.0313	0.0026	0.0318	0.0081	0.0394		
50	-0.0125	0.0212	-0.0063	0.0196	-0.0040	0.0200		
75	-0.0019	0.0167	-0.0011	0.0170	0.0013	0.0164		
100	-0.0177	0.0112	-0.0158	0.0108	-0.0149	0.0102		

TABLE 1. Estimated mean and variance of the estimates of θ_0 in Example 1.a.

We can see from Table 1 that all the estimates converge in a satisfactory way. The estimated variances are almost the same for all estimates. This is not surprising since the limit variance of all estimates in this Example (when $\theta_0 = 0$) is close to $V(X)$.

6.2. Example 1.b. In this example the true value of the parameter is $\theta_0 = 1$.

n	ME χ_m^2 DE		MEKL $_m$ DE=MELE		MEHDE		MEKLDE	
	mean	var	mean	var	mean	var	mean	var
25	0.9394	0.0310	0.9387	0.0312	0.9385	0.0313	0.9378	0.0316
50	0.9994	0.0186	0.9967	0.0186	0.9954	0.0186	0.9941	0.0187
75	1.0009	0.0156	0.9988	0.0154	0.9975	0.0154	0.9966	0.0153
100	0.9984	0.0113	0.9959	0.0112	0.9945	0.0112	0.99315	0.0112

n	ME χ^2 DE		PMLE		SME	
	mean	var	mean	var	mean	var
25	0.9350	0.0322	0.9540	0.0325	1.0033	0.0810
50	0.9909	0.0190	1.0036	0.0174	1.0021	0.0407
75	0.9940	0.0152	1.0003	0.0149	0.9912	0.0288
100	0.9900	0.0113	0.9970	0.0107	0.9851	0.0262

TABLE 2. Estimated mean and variance of the estimates of θ_0 in Example 1.b.

We can see from Table 2 and Figure 1 that the estimated bias of E ϕ DE's are all smaller than the SME one for moderate and large sample sizes. Furthermore, from Figure 2, we observe that the estimated variances of E ϕ DE's are all less than the SME one. They lie between that of the sample mean and that of the parametric maximum likelihood estimate. We observe also that the estimated variances of the MELE and MEHDE are equal and are the smallest among the variances of all ME ϕ DE's considered. It should be emphasized that even for small sample sizes, the MSE of the SM is larger than any of ME ϕ DE's.

In Examples 2.a and 2.b below, we compare robustness property of the estimates considered above for contaminated data : we consider the same model \mathcal{M}_θ as in (6.5).

6.3. Example 2.a. In this Example, we generate 1000 pseudo-random samples of sizes 25, 50, 75 and 100 from a distribution

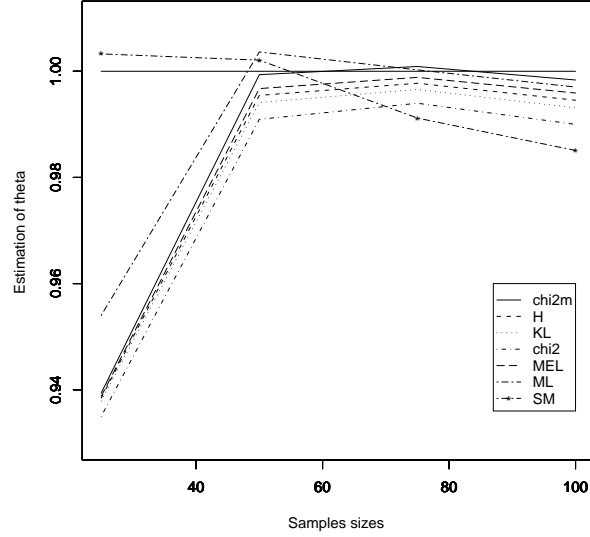
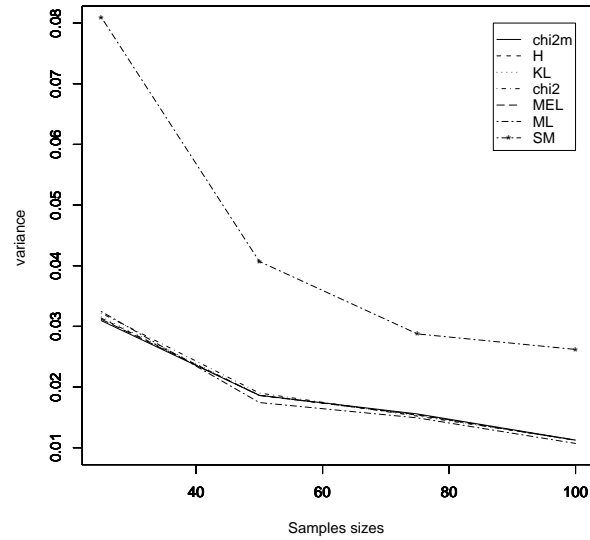
$$\widetilde{P}_0 = (1 - \epsilon)P_0 + \epsilon\delta_5$$

where $P_0 = \mathcal{N}(\theta_0, \theta_0^2 + 1)$, $\epsilon = 0.15$ and $\theta_0 = 2$. We consider the same estimates as in the above examples.

In this Example, we can see from Table 3 and Figure 3 that the ME χ^2 D estimate is the most robust and ME χ_m^2 estimate is the least robust. We observe also that the MELE which is the MEKL $_m$ DE is less robust than the MEKLDE and that the MEHD estimate is more robust than MEL one.

6.4. Example 2.b. In this Example, we generate 1000 pseudo-random samples of sizes 50, 100, 150 and 200 from a distribution $P_0 = \mathcal{N}(\theta_0, \theta_0^2 + 1)$ with $\theta_0 = 2$ and we cancel the observations in the interval $[4, 5]$. We consider the same estimates as in the above examples.

In this example, in contrast with Example 2.b, we observe that the ME χ_m^2 DE is the most robust, ME χ^2 DE is the least robust and MEKLDE is less robust than MEKL $_m$ DE (=MELE). Generally,

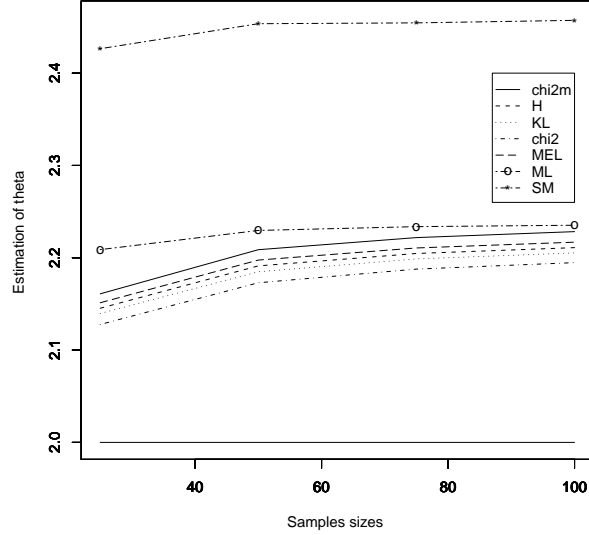

 FIGURE 1. Estimated mean of the estimates of θ_0 in Example 1.b.

 FIGURE 2. Estimated variance of the estimates of θ_0 in Example 1.b.

if a $\text{ME}\phi\text{DE}$ is more robust than its adjoint¹ (i.e., $\text{ME}\phi^\sim\text{DE}$) against “outliers”, then it is less

¹For all divergence ϕ associated to a convex function φ , its adjoint, noted ϕ^\sim , is the divergence associated to the convex function, noted φ^\sim , defined by : $\varphi^\sim(x) = x\varphi(1/x)$, for all x .

n	$ME\chi_m^2 DE$		$MEKL_m DE=MELE$		$MEHDE$		$MEKLDE$	
	mean	var	mean	var	mean	var	mean	var
25	2.1609	0.0654	2.1513	0.0653	2.1453	0.0653	2.1396	0.0652
50	2.2087	0.0303	2.1975	0.0304	2.1912	0.0307	2.1848	0.0309
75	2.2218	0.0214	2.2106	0.0213	2.2046	0.0213	2.1987	0.0215
100	2.2283	0.0151	2.2169	0.0149	2.2110	0.0148	2.2052	0.0149

n	$ME\chi^2 DE$		PMLE		SME		
	mean	var	mean	var	mean	var	
25	2.1278	0.0646	2.2088	0.0581	2.4265	0.2178	
50	2.1729	0.0316	2.2296	0.0280	2.4535	0.1076	
75	2.1877	0.0219	2.2337	0.0197	2.4545	0.0721	
100	2.1947	0.0151	2.2352	0.0139	2.4572	0.0543	

TABLE 3. Estimated mean and variance of the estimates of θ_0 in Example 2.a.FIGURE 3. Estimated mean of the estimates of θ_0 in Example 2.a.

robust then its adjoint against “inliers” (see Examples 2.a and 2.b). The Hellinger divergence has not this disadvantage since it is self-adjoint (i.e., $H = H^\sim$).

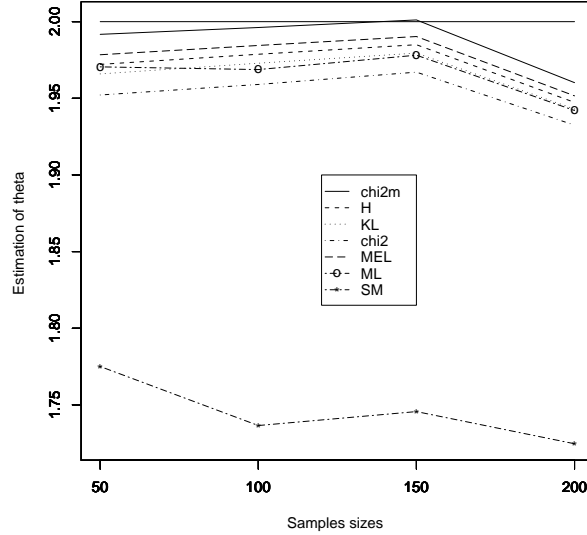
7. PROOFS

7.1. **Proof of Proposition 2.1.** Proof of part (i). The function

$$(Q(X_1), \dots, Q(X_n))^T \in \mathbb{R}^n \mapsto \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i))$$

n	$ME\chi_m^2DE$		$MEKL_mDE=MELE$		$MEHDE$		$MEKLDE$	
	mean	var	mean	var	mean	var	mean	var
50	1.9917	0.0451	1.9784	0.0431	1.9721	0.0426	1.9659	0.0423
100	1.9962	0.0362	1.9844	0.0346	1.9787	0.0341	1.9729	0.0336
150	2.0011	0.0150	1.9903	0.0142	1.9849	0.0139	1.9795	0.0137
200	1.9602	0.0162	1.9516	0.0158	1.9473	0.0157	1.9430	0.0156

n	$ME\chi^2DE$		PMLE		SME		
	mean	var	mean	var	mean	var	
50	1.9522	0.0428	1.9705	0.0358	1.7750	0.1039	
100	1.9590	0.0329	1.9687	0.0298	1.7365	0.0576	
150	1.9671	0.0135	1.9781	0.0121	1.7456	0.0283	
200	1.9325	0.0155	1.9420	0.0146	1.7247	0.0317	

 TABLE 4. Estimated mean and variance of the estimates of θ_0 in Example 2.b.

 FIGURE 4. Estimated mean of the estimates of θ_0 in Example 2.b.

is continuous and nonnegative on $\mathcal{D}_\phi^{(n)}$. Furthermore, the set $\mathcal{M}_\theta^{(n)}$ is closed in \mathbb{R}^n . Hence, by condition (2.3), the infimum of the function

$$(Q(X_1), \dots, Q(X_n))^T \in \mathbb{R}^n \mapsto \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i))$$

on the set $\mathcal{D}_\phi^{(n)} \cap \mathcal{M}_\theta^{(n)}$ exists as an interior point of $\mathcal{D}_\phi^{(n)}$. Since the above function is strictly convex and the set $\mathcal{D}_\phi^{(n)} \cap \mathcal{M}_\theta^{(n)}$ is convex, then this infimum is unique. It is noted \widehat{Q}_θ^* . This concludes the proof of part (i).

Proof of part (ii). Since $(Q(X_1), \dots, Q(X_n))^T \in \mathbb{R}^n \mapsto \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i))$ is \mathcal{C}^1 on the interior

of $\mathcal{D}_\phi^{(n)}$, and since \widehat{Q}_θ^* is in the interior of $\mathcal{D}_\phi^{(n)}$, we can use the Lagrange method. This yields the explicit form (2.5) of the projection \widehat{Q}_θ^* in which \widehat{c}_0 is the Lagrange multiplier associated to the constraint $\sum_{i=1}^n Q(X_i) = 1$ and \widehat{c}_j to the constraint $\sum_{i=1}^l Q(X_i)g_j(X_i, \theta) = 0$, for all $j = 1, \dots, l$. This concludes the proof of Proposition 2.1.

7.2. Proof of Proposition 3.1. Define the estimates

$$\tilde{c}_\theta = \arg \inf_{t \in T_\theta} P_n m(\theta, t) \quad \text{and} \quad \tilde{\phi}(\mathcal{M}_\theta, P_0) = \sup_{t \in T_\theta} P_n m(\theta, t).$$

By condition (C.2), for all n sufficiently large, we have

$$\widehat{c}_\theta = \tilde{c}_\theta \quad \text{and} \quad \widehat{\phi}(\mathcal{M}_\theta, P_0) = \tilde{\phi}(\mathcal{M}_\theta, P_0).$$

We prove that $\tilde{\phi}(\mathcal{M}_\theta, P_0)$ and \tilde{c}_θ converge to $\phi(\mathcal{M}_\theta, P_0)$ and c_θ respectively. Since c_θ is isolated, then consistency of \tilde{c}_θ holds as a consequence of Theorem 5.7 in van der Vaart (1998). For the estimate $\tilde{\phi}(\mathcal{M}_\theta, P_0)$, we have

$$\left| \tilde{\phi}(\mathcal{M}_\theta, P_0) - \phi(\mathcal{M}_\theta, P_0) \right| = |P_n m(\theta, \tilde{c}_\theta) - P_0 m(\theta, c_\theta)| := |A|,$$

which implies

$$P_n m(\theta, c_\theta) - P_0 m(\theta, c_\theta) < A < P_n m(\theta, \tilde{c}_\theta) - P_0 m(\theta, \tilde{c}_\theta).$$

Both the RHS and the LHS terms in the above display go to 0, under condition (C.1). This implies that A tends to 0, which concludes the proof of Proposition 3.1.

7.3. Proof of Theorem 3.2. . Proof of part (1). Some calculus yield

$$(7.1) P_0 m'(\theta, c_\theta) = P_0 \left(1 - \overleftarrow{\varphi}'(c_\theta^T g(\theta)), -g_1(\theta) \overleftarrow{\varphi}'(c_\theta^T g(\theta)), \dots, -g_l(\theta) \overleftarrow{\varphi}'(c_\theta^T g(\theta)) \right)^T = \underline{0}_l^T.$$

and

$$(7.2) \quad P_0 m''(\theta, c_\theta) = P_0 \left[-\frac{g_i g_j}{\overleftarrow{\varphi}''(\overleftarrow{\varphi}'(c_\theta^T g(\theta)))} \right]_{i,j=0,\dots,l},$$

which implies that the matrix $P_0 m''(\theta, c_\theta)$ is symmetric. Under assumption (A.2), by Taylor expansion, there exists $t_n \in \mathbb{R}^{l+1}$ inside the segment that links c_θ and \widehat{c}_θ with

$$(7.3) \quad \begin{aligned} 0 &= P_n m'(\theta, \widehat{c}_\theta) \\ &= P_n m'(\theta, c_\theta) + (P_n m''(\theta, c_\theta))^T (\widehat{c}_\theta - c_\theta) \\ &\quad + \frac{1}{2} (\widehat{c}_\theta - c_\theta)^T P_n m'''(\theta, t_n) (\widehat{c}_\theta - c_\theta), \end{aligned}$$

in which, $P_n m'''(\theta, t_n)$ is a $(l+1)$ -vector whose entries are $(l+1) \times (l+1)$ -matrices. By (A.2), we have for the sup-norm of vectors and matrices

$$\|P_n m'''(\theta, t_n)\| := \left\| \frac{1}{n} \sum_{i=1}^n m'''(X_i, \theta, t_n) \right\| \leq \frac{1}{n} \sum_{i=1}^n |H(X_i)|.$$

By the Law of Large Numbers (LLN), $P_n m'''(\theta, t_n) = O_P(1)$. So using (A.1), we can write the last term in the right hand side of (7.3) as $o_P(1) (\widehat{c}_\theta - c_\theta)$. On the other hand by (A.3), $P_n m''(\theta, c_\theta) := \frac{1}{n} \sum_{i=1}^n m''(X_i, \theta, c_\theta)$ converges to the matrix $P_0 m''(\theta, c_\theta)$. Write $P_n m''(\theta, c_\theta)$ as $P_0 m''(\theta, c_\theta) + o_P(1)$ to obtain from (7.3)

$$(7.4) \quad -P_n m'(\theta, c_\theta) = (P_0 m''(\theta, c_\theta) + o_P(1)) (\widehat{c}_\theta - c_\theta).$$

Under (A.3), by the Central Limit Theorem, we have $\sqrt{n} P_n m'(\theta, c_\theta) = O_P(1)$, which by (7.4) implies that $\sqrt{n} (\widehat{c}_\theta - c_\theta) = O_P(1)$. Hence, from (7.4), we get

$$(7.5) \quad \sqrt{n} (\widehat{c}_\theta - c_\theta) = [-P_0 m''(\theta, c_\theta)]^{-1} \sqrt{n} P_n m'(\theta, c_\theta) + o_P(1).$$

Under (A.3), the Central Limit Theorem concludes the proof of part 1. In the case when P_0 belongs to \mathcal{M}_θ , then $c_\theta^T = (\varphi'(1), \underline{0}^T) := \underline{c}$ and calculation yields

$$P_0 m'(\theta, \underline{c}) m'(\theta, \underline{c})^T = \begin{pmatrix} 0 & \underline{0}_l^T \\ \underline{0}_l & P_0 g(\theta) g(\theta)^T \end{pmatrix} \quad \text{and} \quad -\varphi''(1) P_0 m''(\theta, \underline{c}) = \begin{pmatrix} 1 & \underline{0}_l^T \\ \underline{0}_l & P_0 g(\theta) g(\theta)^T \end{pmatrix}.$$

A simple calculation yields (3.3).

Proof of part (2). By Taylor expansion, there exists \bar{t}_n inside the segment that links c_θ and \hat{c}_θ with

$$\begin{aligned} \hat{\phi}_n(\mathcal{M}_\theta, P_0) &= P_n m(\theta, \hat{c}_\theta) \\ &= P_n m(\theta, c_\theta) + (P_n m'(\theta, c_\theta))^T (\hat{c}_\theta - c_\theta) \\ &\quad + \frac{1}{2} (\hat{c}_\theta - c_\theta)^T [P_n m''(\theta, c_\theta)] (\hat{c}_\theta - c_\theta) \\ &\quad + \frac{1}{3!} \sum_{1 \leq i, j, k \leq d} (\hat{c}_\theta - c_\theta)_i (\hat{c}_\theta - c_\theta)_j \times \\ &\quad \times (\hat{c}_\theta - c_\theta)_k P_n \frac{\partial^3}{\partial t_i \partial t_j \partial t_k} m(\theta, \bar{t}_n). \end{aligned} \tag{7.6}$$

When P_0 belongs to \mathcal{M}_θ , then $c_\theta^T = \underline{c}$. Hence $P_n m(\theta, c_\theta) = P_n m(\theta, \underline{c}) = P_n 0 = 0$. Furthermore, by part (1) in Theorem 3.2, it holds $\sqrt{n}(\hat{c}_\theta - c_\theta) = O_P(1)$. Hence, by (A.1), (A.2) and (A.3), we get

$$\begin{aligned} \hat{\phi}_n(\mathcal{M}_\theta, P_0) &= (P_n m'(\theta, c_\theta))^T (\hat{c}_\theta - c_\theta) + \\ &\quad \frac{1}{2} (\hat{c}_\theta - c_\theta)^T [P_0 m''(\theta, c_\theta)] (\hat{c}_\theta - c_\theta) + o_P(1/n), \end{aligned}$$

which by (7.5), implies

$$\begin{aligned} \hat{\phi}_n(\mathcal{M}_\theta, P_0) &= [P_n m'(\theta, c_\theta)]^T [-P_0 m''(\theta, c_\theta)]^{-1} [P_n m'(\theta, c_\theta)] + \\ &\quad \frac{1}{2} [P_n m'(\theta, c_\theta)]^T [P_0 m''(\theta, c_\theta)]^{-1} [P_n m'(\theta, c_\theta)] + o_P(1/n) \\ &= \frac{1}{2} [P_n m'(\theta, c_\theta)]^T [-P_0 m''(\theta, c_\theta)]^{-1} [P_n m'(\theta, c)] + o_P(1/n). \end{aligned}$$

This yields to

$$(7.7) \quad \frac{2n}{\varphi''(1)} \hat{\phi}_n(\mathcal{M}_\theta, P_0) = [\sqrt{n} P_n m'(\theta, c_\theta)]^T [-\varphi''(1) P_0 m''(\theta, c_\theta)]^{-1} [\sqrt{n} P_n m'(\theta, c_\theta)] + o_P(1).$$

Note that when P_0 belongs to \mathcal{M}_θ , then $c_\theta^T = \underline{c}$ and calculation yields

$$P_0 m'(\theta, \underline{c}) m'(\theta, \underline{c})^T = \begin{pmatrix} 0 & \underline{0}_l^T \\ \underline{0}_l & P_0 g(\theta) g(\theta)^T \end{pmatrix} \quad \text{and} \quad -\varphi''(1) P_0 m''(\theta, \underline{c}) = \begin{pmatrix} 1 & \underline{0}_l^T \\ \underline{0}_l & P_0 g(\theta) g(\theta)^T \end{pmatrix}.$$

Combining this with (7.7), we conclude the proof of part (2).

Proof of part (3). Since $(\hat{c}_\theta - c_\theta) = O_P(1/\sqrt{n})$ and $P_n m'(\theta, c_\theta) = P_0 m'(\theta, c_\theta) + o_P(1) = 0 + o_P(1) = o_P(1)$, then, using (7.6), we obtain

$$\begin{aligned} \sqrt{n} \left(\hat{\phi}_n(\mathcal{M}_\theta, P_0) - \phi(\mathcal{M}_\theta, P_0) \right) &= \sqrt{n} \left(\hat{\phi}_n(\mathcal{M}_\theta, P_0) - P_0 m(\theta, c_\theta) \right) \\ &= \sqrt{n} (P_n m(\theta, c_\theta) - P_0 m(\theta, c_\theta)) + o_P(1), \end{aligned}$$

and the Central Limit Theorem yields to the conclusion of the proof of Theorem 3.2.

7.4. Proof of Proposition 3.3. Define the estimates

$$\begin{aligned}\tilde{\theta}_\phi &:= \arg \inf_{\theta \in \Theta} \sup_{t \in T_\theta} P_n m(\theta, t), \\ \tilde{\phi}(\mathcal{M}, P_0) &:= \inf_{\theta \in \Theta} \sup_{t \in T_\theta} P_n m(\theta, t)\end{aligned}$$

and for all $\theta \in \Theta$,

$$\tilde{c}_\theta := \arg \sup_{t \in T_\theta} P_n m(\theta, t).$$

By condition (C.5), for all n sufficiently large, it holds

$$\hat{\theta}_\phi = \tilde{\theta}_\phi \quad \text{and} \quad \hat{\phi}(\mathcal{M}, P_0) = \tilde{\phi}(\mathcal{M}, P_0).$$

We prove that $\tilde{\theta}_\phi$ and $\tilde{\phi}(\mathcal{M}, P_0)$ are consistent. First, we prove the consistency of $\tilde{\phi}(\mathcal{M}, P_0)$. We have

$$\left| \tilde{\phi}(\mathcal{M}, P_0) - \phi(\mathcal{M}, P_0) \right| = \left| P_n m(\tilde{\theta}_\phi, \tilde{c}_{\tilde{\theta}_\phi}) - P_0 m(\theta^*, c_{\theta^*}) \right| =: |A|.$$

This implies

$$P_n m(\tilde{\theta}_\phi, c_{\theta^*}) - P_0 m(\tilde{\theta}_\phi, c_{\theta^*}) \leq A \leq P_n m(\theta^*, \tilde{c}_{\tilde{\theta}_\phi}) - P_0 m(\theta^*, \tilde{c}_{\tilde{\theta}_\phi}).$$

By condition (C.3), both the RHS and LHS terms in the above display go to 0. This implies that A tends to 0 which concludes the proof of part (i).

Proof of part (ii). Since for sufficiently large n , by condition (C.5), we have $\hat{c}_\theta = \tilde{c}_\theta$ for all $\theta \in \Theta$, the convergence of $\sup_{\theta \in \Theta} \|\tilde{c}_\theta - c_\theta\|$ to 0 implies (ii). We prove now that $\sup_{\theta \in \Theta} \|\tilde{c}_\theta - c_\theta\|$ tends to 0. By the very definition of \tilde{c}_θ and condition (C.3), we have

$$\begin{aligned}P_n m(\theta, \tilde{c}_\theta) &\geq P_n m(\theta, c_\theta) \\ (7.8) \quad &\geq P_0 m(\theta, c_\theta) - o_P(1),\end{aligned}$$

where $o_P(1)$ does not depends upon θ (due to condition (C.3)). Hence, we have for all $\theta \in \Theta$,

$$(7.9) \quad P_0 m(\theta, c_\theta) - P_0 m(\theta, \tilde{c}_\theta) \leq P_n m(\theta, \tilde{c}_\theta) - P_0 m(\theta, \tilde{c}_\theta) + o_P(1).$$

The term in the RHS of the above display is less than

$$\sup_{\theta \in \Theta, t \in T_\theta} |P_n m(\theta, t) - P_0 m(\theta, t)| + o_P(1)$$

which by (C.3), tends to 0. Let $\epsilon > 0$ be such that $\sup_{\theta \in \Theta} \|\tilde{c}_\theta - c_\theta\| > \epsilon$. There exists some $a_n \in \Theta$ such that $\|\tilde{c}_{a_n} - c_{a_n}\| > \epsilon$. Together with the strict concavity of the function $t \in T_\theta \rightarrow P_0 m(\theta, t)$ for all $\theta \in \Theta$, there exists $\eta > 0$ such that

$$P_0 m(a_n, c_{a_n}) - P_0 m(a_n, \tilde{c}_{a_n}) > \eta.$$

We then conclude that

$$P \left\{ \sup_{\theta \in \Theta} \|\tilde{c}_\theta - c_\theta\| > \epsilon \right\} \leq P \{ P_0 m(a_n, c_{a_n}) - P_0 m(a_n, \tilde{c}_{a_n}) > \eta \},$$

and the RHS term tends to 0 by (7.9). This concludes the proof part (ii).

Proof of part (iii). We prove that $\tilde{\theta}_\phi$ converges to θ^* . By the very definition of $\tilde{\theta}_\phi$, condition (C.4.b) and part (ii), we obtain

$$\begin{aligned}P_n m(\tilde{\theta}_\phi, \tilde{c}_{\tilde{\theta}_\phi}) &\leq P_n m(\theta^*, \tilde{c}_{\theta^*}) \\ &\leq P_0 m(\theta^*, \tilde{c}_{\tilde{\theta}_\phi}) - o_P(1),\end{aligned}$$

from which

$$\begin{aligned}
 P_0 m(\tilde{\theta}_\phi, \widetilde{c_{\tilde{\theta}_\phi}}) - P_0 m(\theta^*, \widetilde{c_{\tilde{\theta}_\phi}}) &\leq P_0 m(\tilde{\theta}_\phi, \widetilde{c_{\tilde{\theta}_\phi}}) - P_n m(\tilde{\theta}_\phi, \widetilde{c_{\tilde{\theta}_\phi}}) + o_P(1) \\
 (7.10) \qquad \qquad \qquad &\leq \sup_{\{\theta \in \Theta, t \in T_\theta\}} |P_n m(\theta, t) - P_0 m(\theta, t)| + o_P(1).
 \end{aligned}$$

Further, by part (ii) and condition (C.4.a), for any positive ϵ , there exists $\eta > 0$ such that

$$P \left\{ \left\| \tilde{\theta}_\phi - \theta^* \right\| > \epsilon \right\} \leq P \left\{ P_0 m(\tilde{\theta}_\phi, \widetilde{c_{\tilde{\theta}_\phi}}) - P_0 m(\theta^*, \widetilde{c_{\tilde{\theta}_\phi}}) > \eta \right\}.$$

The RHS term, under condition (C.3), tends to 0 by (7.10). This concludes the proof of Proposition 3.3.

7.5. Proof of Theorem 3.4. Since $P_0 \in \mathcal{M}$, then $c_\theta = \underline{c}$. Some calculus yield

$$\begin{aligned}
 (7.11) \qquad \frac{\partial}{\partial t} m(\theta_0, \underline{c}) &= [0, -g_1(\theta_0), \dots, -g_l(\theta_0)]^T = -[0, g(\theta_0)^T]^T, \\
 \frac{\partial}{\partial \theta} m(\theta, t) &= - \sum_{j=0}^l t_i \overleftarrow{\varphi'}(t^T g(\theta)) \frac{\partial}{\partial \theta} g_j(\theta), \quad \frac{\partial}{\partial \theta} m(\theta_0, \underline{c}) = \underline{0}_d, \\
 \frac{\partial^2}{\partial \theta \partial t} m(\theta_0, \underline{c}) &= \left[\underline{0}_d, -\frac{\partial}{\partial \theta} g_1(\theta_0), \dots, -\frac{\partial}{\partial \theta} g_l(\theta_0) \right] = - \left[\underline{0}_d, \frac{\partial}{\partial \theta} g(\theta) \right] \\
 \frac{\partial^2}{\partial t \partial \theta} m(\theta_0, \underline{c}) &= \left[\frac{\partial^2}{\partial \theta \partial t} m(\theta_0, \underline{c}) \right]^T, \quad \frac{\partial^2}{\partial \theta^2} m(\theta_0, \underline{c}) = [\underline{0}_d, \dots, \underline{0}_d],
 \end{aligned}$$

and

$$\frac{\partial^2}{\partial t^2} m(\theta_0, \underline{c}) = \frac{1}{\varphi''(1)} [-g_i(\theta_0) g_j(\theta_0)]_{i,j=0,1,\dots,l} := \frac{-1}{\varphi''(1)} (\overline{g}(\theta_0) \overline{g}(\theta_0)^T).$$

Integrating w.r.t. P_0 , we obtain

$$(7.12) \qquad P_0 \frac{\partial}{\partial t} m(\theta_0, \underline{c}) = \underline{0}_l, \quad P_0 \frac{\partial}{\partial \theta} m(\theta_0, \underline{c}) = \underline{0}_d, \quad P_0 \frac{\partial^2}{\partial \theta^2} m(\theta_0, \underline{c}) = [\underline{0}_d, \dots, \underline{0}_d],$$

$$(7.13) \qquad P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta_0, \underline{c}) = - \left[\underline{0}_d, P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right],$$

$$(7.14) \qquad P_0 \frac{\partial^2}{\partial t \partial \theta} m(\theta_0, \underline{c}) = \left[P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta_0, \underline{c}) \right]^T = - \left[\underline{0}_d, P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right]^T,$$

and

$$\begin{aligned}
 (7.15) \qquad P_0 \frac{\partial^2}{\partial t^2} m(\theta_0, \underline{c}) &= \frac{-1}{\varphi''(1)} [P_0 g_i(\theta_0) g_j(\theta_0)]_{i,j=0,1,\dots,l} \\
 &= \frac{-1}{\varphi''(1)} \begin{pmatrix} 1 & \underline{0}_l^T \\ \underline{0}_l & P_0 g(\theta_0) g(\theta_0)^T \end{pmatrix}
 \end{aligned}$$

By the very definition of $\widehat{\theta}_\phi$ and $\widehat{c_{\widehat{\theta}_\phi}}$, they both obey

$$\begin{cases} P_n \frac{\partial}{\partial t} m(\theta, t) &= 0 \\ P_n \frac{\partial}{\partial \theta} m(\theta, t(\theta)) &= 0, \end{cases}$$

i.e.,

$$\begin{cases} P_n \frac{\partial}{\partial t} m(\widehat{\theta}_\phi, \widehat{c_{\widehat{\theta}_\phi}}) &= 0 \\ P_n \frac{\partial}{\partial \theta} m(\widehat{\theta}_\phi, \widehat{c_{\widehat{\theta}_\phi}}) + P_n \frac{\partial}{\partial t} m(\widehat{\theta}_\phi, \widehat{c_{\widehat{\theta}_\phi}}) \frac{\partial}{\partial \theta} \widehat{c_{\widehat{\theta}_\phi}} &= 0. \end{cases}$$

The second term in the left hand side of the second equation is equal to 0, due to the first equation. Hence $\widehat{c_{\theta_\phi}}$ and $\widehat{\theta_\phi}$ are solutions of the somehow simpler system

$$\begin{cases} P_n \frac{\partial}{\partial t} m(\widehat{\theta_\phi}, \widehat{c_{\theta_\phi}}) &= 0 \text{ (E1)} \\ P_n \frac{\partial}{\partial \theta} m(\widehat{\theta_\phi}, \widehat{c_{\theta_\phi}}) &= 0 \text{ (E2)}. \end{cases}$$

Use a Taylor expansion in (E1); there exists $(\widetilde{\theta}_n, \widetilde{t}_n)$ inside the segment that links $(\widehat{\theta_\phi}, \widehat{c_{\theta_\phi}})$ and $(\theta_0, \underline{c})$ such that

$$\begin{aligned} 0 &= P_n \frac{\partial}{\partial t} m(\theta_0, \underline{c}) + \left[\left(P_n \frac{\partial^2}{\partial t^2} m(\theta_0, \underline{c}) \right)^T, \left(P_n \frac{\partial^2}{\partial \theta \partial t} m(\theta_0, \underline{c}) \right)^T \right] a_n \\ &\quad + \frac{1}{2} a_n^T A_n a_n, \end{aligned} \tag{7.16}$$

with

$$a_n := \left((\widehat{c_{\theta_\phi}} - \underline{c})^T, (\widehat{\theta_\phi} - \theta_0)^T \right)^T \tag{7.17}$$

and

$$A_n := \begin{pmatrix} P_n \frac{\partial^3}{\partial t^3} m(\widetilde{\theta}, \widetilde{c}_n) & P_n \frac{\partial^3}{\partial t \partial \theta \partial t} m(\widetilde{\theta}, \widetilde{c}_n) \\ P_n \frac{\partial^3}{\partial \theta \partial t^2} m(\widetilde{\theta}, \widetilde{c}_n) & P_n \frac{\partial^3}{\partial \theta^2 \partial t} m(\widetilde{\theta}, \widetilde{c}_n) \end{pmatrix}. \tag{7.18}$$

By (A.5), the LLN implies that $A_n = O_P(1)$. So using (A.4), we can write the last term in right hand side of (7.16) as $o_P(1)a_n$. On the other hand by (A.6), we can write also $\left[\left(P_n \frac{\partial^2}{\partial t^2} m(\theta_0, \underline{c}) \right)^T, \left(P_n \frac{\partial^2}{\partial \theta \partial t} m(\theta_0, \underline{c}) \right)^T \right]$ as $\left[P_0 \frac{\partial^2}{\partial t^2} m(\theta_0, \underline{c}), \left(P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta_0, \underline{c}) \right)^T \right] + o_P(1)$ to obtain from (7.16)

$$-P_n \frac{\partial}{\partial t} m(\theta_0, \underline{c}) = \left[P_0 \frac{\partial^2}{\partial t^2} m(\theta_0, \underline{c}) + o_P(1), \left(P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta_0, \underline{c}) \right)^T + o_P(1) \right] a_n. \tag{7.19}$$

In the same way, using a Taylor expansion in (E2), there exists $(\bar{\theta}_n, \bar{t}_n)$ inside the segment that links $(\widehat{\theta_\phi}, \widehat{c_{\theta_\phi}})$ and $(\theta_0, \underline{c})$ such that

$$\begin{aligned} 0 &= P_n \frac{\partial}{\partial \theta} m(\theta_0, \underline{c}) + \left[\left(P_n \frac{\partial^2}{\partial t \partial \theta} m(\theta_0, \underline{c}) \right)^T, \left(P_n \frac{\partial^2}{\partial \theta^2} m(\theta_0, \underline{c}) \right)^T \right] a_n \\ &\quad + \frac{1}{2} a_n^T B_n a_n, \end{aligned} \tag{7.20}$$

with

$$B_n := \begin{bmatrix} P_n \frac{\partial^3}{\partial t^2 \partial \theta} m(\bar{\theta}_n, \bar{t}_n) & P_n \frac{\partial^3}{\partial t \partial \theta^2} m(\bar{\theta}_n, \bar{t}_n) \\ P_n \frac{\partial^3}{\partial \theta \partial t \partial \theta} m(\bar{\theta}_n, \bar{t}_n) & P_n \frac{\partial^3}{\partial \theta^3} m(\bar{\theta}_n, \bar{t}_n) \end{bmatrix}.$$

As in (7.19), we obtain

$$-P_n \frac{\partial}{\partial \theta} m(\theta_0, \underline{c}) = \left[\left(P_0 \frac{\partial^2}{\partial t \partial \theta} m(\theta_0, \underline{c}) \right)^T + o_P(1), P_0 \frac{\partial^2}{\partial \theta^2} m(\theta_0, \underline{c}) + o_P(1) \right] a_n. \tag{7.21}$$

From (7.19) and (7.21), we get

$$(7.22) \quad \begin{aligned} \sqrt{n}a_n &= \sqrt{n} \begin{pmatrix} P_0 \frac{\partial^2}{\partial t^2} m(\theta_0, c(\theta_0)) & \left(P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta_0, \underline{c}) \right)^T \\ \left(P_0 \frac{\partial^2}{\partial t \partial \theta} m(\theta_0, c(\theta_0)) \right)^T & P_0 \frac{\partial^2}{\partial \theta^2} m(\theta_0, c(\theta_0)) \end{pmatrix}^{-1} \times \\ &\times \begin{pmatrix} -P_n \frac{\partial}{\partial t} m(\theta_0, \underline{c}) \\ -P_n \frac{\partial}{\partial \theta} m(\theta_0, \underline{c}) \end{pmatrix} + o_P(1). \end{aligned}$$

Denote S the $(l+1+d) \times (l+1+d)$ -matrix defined by

$$(7.23) \quad S := \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} := \begin{pmatrix} P_0 \frac{\partial^2}{\partial t^2} m(\theta_0, c(\theta_0)) & \left(P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta_0, \underline{c}) \right)^T \\ \left(P_0 \frac{\partial^2}{\partial t \partial \theta} m(\theta_0, c(\theta_0)) \right)^T & P_0 \frac{\partial^2}{\partial \theta^2} m(\theta_0, c(\theta_0)) \end{pmatrix}.$$

We have

$$(7.24) \quad S_{11} = \frac{-1}{\varphi''(1)} \begin{pmatrix} 1 & \underline{0}_l^T \\ \underline{0}_l & P_0 g(\theta_0) g(\theta_0)^T \end{pmatrix}$$

$$(7.25) \quad S_{12} = - \left[\underline{0}_d, P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right]^T, \quad S_{21} = - \left[\underline{0}_d, P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right] \quad \text{and}$$

$$(7.26) \quad S_{22} = P_0 \frac{\partial^2}{\partial \theta^2} m(\theta_0, \underline{c}) = [\underline{0}_d, \dots, \underline{0}_d].$$

The inverse matrix S^{-1} of the matrix S writes

$$(7.27) \quad S^{-1} = \begin{pmatrix} S_{11}^{-1} + S_{11}^{-1} S_{12} S_{22.1}^{-1} S_{21} S_{11}^{-1} & -S_{11}^{-1} S_{12} S_{22.1}^{-1} \\ -S_{22.1}^{-1} S_{21} S_{11}^{-1} & S_{22.1}^{-1} \end{pmatrix},$$

where

$$(7.28) \quad \begin{aligned} S_{22.1} &= -S_{21} S_{11}^{-1} S_{12} \\ &= \left[\underline{0}_d, P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right] [\varphi''(1)] \begin{bmatrix} 1 & \underline{0}_l^T \\ \underline{0}_l & [P_0 g(\theta_0) g(\theta_0)^T]^{-1} \end{bmatrix} \left[\underline{0}_d, P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right]^T \\ &= \varphi''(1) \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right] [P_0 g(\theta_0) g(\theta_0)^T]^{-1} \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right]^T. \end{aligned}$$

From (7.22), using (7.23) and (7.27), we can write

$$(7.29) \quad \begin{aligned} \sqrt{n} \begin{pmatrix} \widehat{c_{\theta_\phi}} - \underline{c} \\ \widehat{\theta_\phi} - \theta_0 \end{pmatrix} &= \begin{pmatrix} S_{11}^{-1} + S_{11}^{-1} S_{12} S_{22.1}^{-1} S_{21} S_{11}^{-1} & -S_{11}^{-1} S_{12} S_{22.1}^{-1} \\ -S_{22.1}^{-1} S_{21} S_{11}^{-1} & S_{22.1}^{-1} \end{pmatrix} \times \\ &\times \sqrt{n} \begin{pmatrix} -P_n \frac{\partial}{\partial t} m(\theta_0, \underline{c}) \\ \underline{0}_d \end{pmatrix} + o_P(1). \end{aligned}$$

Note that

$$(7.30) \quad \sqrt{n} \begin{pmatrix} -P_n \frac{\partial}{\partial t} m(\theta_0, \underline{c}) \\ \underline{0}_d \end{pmatrix},$$

under assumption (A.6), by the Central Limit Theorem, converges in distribution to a centered multivariate normal variable with covariance matrix

$$(7.31) \quad M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

where

$$(7.32) \quad M_{11} = \begin{pmatrix} 0 & \underline{0}_l^T \\ \underline{0}_l & P_0 g(\theta_0) g(\theta_0)^T \end{pmatrix}, \quad M_{12} = \begin{pmatrix} \underline{0}_d^T \\ \vdots \\ \underline{0}_d^T \end{pmatrix}, \quad M_{21} = \begin{pmatrix} 0 & \underline{0}_l^T \\ \vdots & \vdots \\ 0 & \underline{0}_l^T \end{pmatrix} \quad \text{and} \quad M_{22} = \begin{pmatrix} \underline{0}_d^T \\ \vdots \\ \underline{0}_d^T \end{pmatrix}.$$

Hence, from (7.30), we deduce that

$$(7.33) \quad \sqrt{n} \begin{pmatrix} \widehat{c_{\hat{\theta}_\phi}} - \underline{c} \\ \widehat{\theta}_\phi - \theta_0 \end{pmatrix}$$

converges in distribution to a centered multivariate normal variable with covariance matrix

$$(7.34) \quad C = S^{-1} M [S^{-1}]^T := \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},$$

and using (7.32) and some algebra, we get

$$(7.35) \quad \begin{aligned} C_{11} &= \varphi''(1)^2 \begin{bmatrix} 0 & \underline{0}_l^T \\ \underline{0}_l & [P_0 g(\theta_0) g(\theta_0)^T]^{-1} \end{bmatrix} - \varphi''(1)^2 \begin{bmatrix} 0 & \underline{0}_l^T \\ \underline{0}_l & [P_0 g(\theta_0) g(\theta_0)^T]^{-1} \end{bmatrix} \times \\ &\times \begin{bmatrix} \underline{0}, P_0 \frac{\partial}{\partial \theta} g(\theta_0) \end{bmatrix}^T [C_{22}] \begin{bmatrix} \underline{0}, P_0 \frac{\partial}{\partial \theta} g(\theta_0) \end{bmatrix} \begin{bmatrix} 0 & \underline{0}_l^T \\ \underline{0}_l & [P_0 g(\theta_0) g(\theta_0)^T]^{-1} \end{bmatrix}, \end{aligned}$$

$$(7.36) \quad C_{12} = [\underline{0}_l, \dots, \underline{0}_l], \quad C_{21} = [\underline{0}_d, \dots, \underline{0}_d]$$

and

$$(7.37) \quad C_{22} = \left\{ \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right] [P_0 (g(\theta_0) g(\theta_0)^T)]^{-1} \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right]^T \right\}^{-1}.$$

From (7.33), we deduce that C_{11} and C_{22} are respectively the limit covariance matrix of $\sqrt{n} (\widehat{c_{\hat{\theta}_\phi}} - \underline{c})$ and $\sqrt{n} (\widehat{\theta}_\phi - \theta_0)$, i.e., $U = C_{11}$ and $V = C_{22}$. (7.36) implies that $\sqrt{n} (\widehat{c_{\hat{\theta}_\phi}} - \underline{c})$ and $\sqrt{n} (\widehat{\theta}_\phi - \theta_0)$ are asymptotically uncorrelated. This concludes the Proof of Theorem 3.4.

7.6. Proof of Theorem 3.5. Under assumptions (A.4-6), as in the proof of Theorem 3.4, we obtain

$$\sqrt{n} \begin{pmatrix} \widehat{c_{\hat{\theta}_\phi}} - c_{\theta^*} \\ \widehat{\theta}_\phi - \theta^* \end{pmatrix} = \sqrt{n} S^{-1} \begin{pmatrix} -P_n \frac{\partial}{\partial l} m(\theta^*, c_{\theta^*}) \\ -P_n \frac{\partial}{\partial \theta} m(\theta^*, c_{\theta^*}) \end{pmatrix} + o_P(1),$$

and the CLT concludes the proof.

7.7. Proof of Theorem 4.1. Using Taylor expansion at $(\underline{c}, \theta_0)$, we get

$$(7.38) \quad \begin{aligned} \widehat{F}_n(x) &:= \sum_{i=1}^n \widehat{Q_{\hat{\theta}_\phi}^*} \mathbb{1}_{(-\infty, x]}(X_i) := \frac{1}{n} \sum_{i=1}^n \overleftarrow{\varphi'} \left(\widehat{c_{\hat{\theta}_\phi}}^T \overline{g}(X_i, \widehat{\theta}_\phi) \right) \mathbb{1}_{(-\infty, x]}(X_i) \\ &= F_n(x) + \frac{1}{n} \left[\sum_{i=1}^n \overline{g}(X_i, \theta_0) \mathbb{1}_{(-\infty, x]}(X_i) \right]^T \frac{1}{\varphi''(1)} (\widehat{c_{\hat{\theta}_\phi}} - \underline{c}) + o_P(\delta_n), \end{aligned}$$

where $\delta_n := \|\widehat{c_{\theta_\phi}} - \underline{c}\| + \|\widehat{\theta_\phi} - \theta_0\|$, which by Theorem 3.4, is equal to $O_P(1/\sqrt{n})$. Hence, (7.38) yields

$$\begin{aligned} \sqrt{n} \left(\widehat{F}_n(x) - F(x) \right) &= \sqrt{n} (F_n(x) - F(x)) + \\ &\quad + \frac{1}{\varphi''(1)} [P_0 (\bar{g}(\theta_0) \mathbf{1}_{(-\infty, x]})]^T \sqrt{n} (\widehat{c_{\theta_\phi}} - \underline{c}) + o_P(1). \end{aligned} \quad (7.39)$$

On the other hand, from (7.29), we get

$$\sqrt{n} (\widehat{c_{\theta_\phi}} - \underline{c}) = H \sqrt{n} \left(-P_n \frac{\partial}{\partial \theta} m(\theta_0, \underline{c}) \right) + o_P(1) \quad (7.40)$$

with

$$H = S_{11}^{-1} + S_{11}^{-1} S_{12} S_{22.1}^{-1} S_{21} S_{11}^{-1}. \quad (7.41)$$

We will use $f(\cdot)$ to denote the function $\mathbf{1}_{(-\infty, x]}(\cdot) - F(x)$, for all $x \in \mathbb{R}$. Substituting (7.40) in (7.39), we get

$$\begin{aligned} \sqrt{n} (\widehat{F}_n(x) - F(x)) &= \sqrt{n} P_n f + \frac{1}{\varphi''(1)} [P_0 (\bar{g}(\theta_0) \mathbf{1}_{(-\infty, x]})]^T H \times \\ &\quad \times \sqrt{n} \left(-P_n \frac{\partial}{\partial \theta} m(\theta_0, \underline{c}) \right) + o_P(1). \end{aligned} \quad (7.42)$$

By the Multivariate Central Limit Theorem, the vector

$\sqrt{n} \left(P_n f, -[P_n \frac{\partial}{\partial \theta} m(\theta_0, \underline{c})]^T \right)^T$ converges in distribution to a centered multivariate normal variable which implies that $\sqrt{n} (\widehat{F}_n(x) - F(x))$ is asymptotically centered normal variable. We calculate now its limit variance, noted $W(x)$.

$$\begin{aligned} W(x) &= F(x)(1 - F(x)) + \frac{1}{\varphi''(1)^2} [P_0 (\bar{g}(\theta_0) \mathbf{1}_{(-\infty, x]})]^T U [P_0 (\bar{g}(\theta_0) \mathbf{1}_{(-\infty, x]})] + \\ &\quad + 2 \frac{1}{\varphi''(1)} [P_0 (\bar{g}(\theta_0) \mathbf{1}_{(-\infty, x]})]^T H \left[P_0 \left(-\frac{\partial}{\partial \theta} m(\theta_0, \underline{c}) \mathbf{1}_{(-\infty, x]} \right) \right]. \end{aligned} \quad (7.43)$$

Use the explicit forms of $\frac{\partial}{\partial \theta} m(\theta_0, \underline{c})$, the matrices U and V and some algebra to obtain

$$W(x) = F(x)(1 - F(x)) - [P_0 (g(\theta_0) \mathbf{1}_{(-\infty, x]})]^T \Gamma [P_0 (g(\theta_0) \mathbf{1}_{(-\infty, x]})],$$

with

$$\begin{aligned} \Gamma &= [P_0 g(\theta_0) g(\theta_0)^T]^{-1} - [P_0 g(\theta_0) g(\theta_0)^T]^{-1} \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right]^T V \times \\ &\quad \times \left[P_0 \frac{\partial}{\partial \theta} g(\theta_0) \right] [P_0 g(\theta_0) g(\theta_0)^T]^{-1}. \end{aligned}$$

This concludes the proof of Theorem 4.1.

REFERENCES

- Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika*, **85**(3), 535–547.
- Basu, A. and Lindsay, B. G. (1994). Minimum disparity estimation for continuous models: efficiency, distributions and robustness. *Ann. Inst. Statist. Math.*, **46**(4), 683–705.

- Bertail, P. (2004). Empirical likelihood in nonparametric and semiparametric models. In *Parametric and semiparametric models with applications to reliability, survival analysis, and quality of life*, Stat. Ind. Technol., pages 291–306. Birkhäuser Boston, Boston, MA.
- Bertail, P. (2006). Empirical likelihood in some semiparametric models *Bernoulli*, **12**(2), 299–331.
- Bickel, P. J., Ritov, Y., and Wellner, J. A. (1991). Efficient estimation of linear functionals of a probability measure P with known marginal distributions. *Ann. Statist.*, **19**(3), 1316–1346.
- Borwein, J. M. and Lewis, A. S. (1991). Duality relationships for entropy-like minimization problems. *SIAM J. Control Optim.*, **29**(2), 325–338.
- Borwein, J. M. and Lewis, A. S. (1993). Partially-finite programming in L_1 and the existence of maximum entropy estimates. *SIAM J. Optim.*, **3**(2), 248–267.
- Broniatowski, M. and Keziou, A. (2006). Minimization of ϕ -divergences on sets of signed measures. *Studia Sci. Math. Hungar.*, **43**(4), 403–442.
- Broniatowski, M. and Keziou, A. (2009). Parametric estimation and testing through divergences. *Journal of Multivariate Analysis*, **43**, 16–36.
- Corcoran, S. (1998). Bertlett adjustment of empirical discrepancy statistics. *Biometrika*, **85**, 967–972.
- Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B*, **46**(3), 440–464.
- Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **8**, 85–108.
- Csiszár, I. (1967). On topology properties of f -divergences. *Studia Sci. Math. Hungar.*, **2**, 329–339.
- Csiszár, I. (1975). I -divergence geometry of probability distributions and minimization problems. *Ann. Probability*, **3**, 146–158.
- Haberman, S. J. (1984). Adjustment by minimum discriminant information. *Ann. Statist.*, **12**(3), 971–988.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**(4), 1029–1054.
- Imbens, G. W. (1997). One-step estimators for over-identified generalized method of moments models. *Rev. Econom. Stud.*, **64**(3), 359–383.
- Jiménez, R. and Shao, Y. (2001). On robustness and efficiency of minimum divergence estimators. *Test*, **10**(2), 241–248.
- Keziou, A. (2003). Dual representation of ϕ -divergences and applications. *C. R. Math. Acad. Sci. Paris*, **336**(10), 857–862.
- Léonard, C. (2001a). Convex conjugates of integral functionals. *Acta Math. Hungar.*, **93**(4), 253–280.
- Léonard, C. (2001b). Minimizers of energy functionals. *Acta Math. Hungar.*, **93**(4), 281–325.
- Liese, F. and Vajda, I. (1987). *Convex statistical distances*, volume 95. BSB B. G. Teubner Verlagsgesellschaft, Leipzig.
- Lindsay, B. G. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *Ann. Statist.*, **22**(2), 1081–1114.
- Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*.
- Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Phil. Trans. Roy. Soc. Ser.*, **A**(236), 333–380.
- Nikitin, Y. (1995). *Asymptotic efficiency of nonparametric tests*. Cambridge University Press, Cambridge.
- Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**(1), 90–120.
- Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**(2), 237–249.

- Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall, New York.
- Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**(1), 300–325.
- Rao, C. R. (1961). Asymptotic efficiency and limiting information. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pages 531–545. Univ. California Press, Berkeley, Calif.
- Rüschendorf, L. (1984). On the minimum discrimination information theorem. *Statist. Decisions*, (suppl. 1), 263–283. Recent results in estimation theory and related topics.
- Schennach, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *Ann. Statist.*, **35**(2), 634–672.
- Sen, P. K. and Singer, J. M. (1993). *Large sample methods in statistics*. Chapman & Hall, New York.
- Sheehy, A. (1987). Kullback-Leibler constrained estimation of probability measures. *Report, Dept. Statistics, Stanford Univ.*
- Takagi, Y. (1998). A new criterion of confidence set estimation: improvement of the Neyman shortness. *J. Statist. Plann. Inference*, **69**(2), 329–338.
- van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.
- Wilks, S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, **9**, 60–62.

*LSTA-UNIVERSITÉ PARIS 6. E-MAIL: MICHEL.BRONIATOWSKI@UPMC.FR

**LABORATOIRE DE MATHÉMATIQUES (FRE 3111) CNRS, UNIVERSITÉ DE REIMS AND LSTA-UNIVERSITÉ PARIS 6., E-MAIL: AMOR.KEZIOU@UPMC.FR