Working Paper Series, N. 1, February 2005

On Empirical Likelihood in Semiparametric Two-Sample Density Ratio Models

Amor KEZIOU
L.S.T.A.
University Paris 6
France.

Samuela LEONI-AUBIN
Department of Statistical Sciences
University of Padua
Italy

Abstract: We consider estimation and test problems for some semiparametric two-sample density ratio models. The profile empirical likelihood (EL) poses an irregularity problem under the null hypothesis that the laws of the two samples are equal. We show that a "dual" form of the profile EL is well defined even under the null hypothesis. A statistical test, based on the dual form of the EL ratio statistic (ELRS), is then proposed. We give an interpretation for the dual form of the ELRS through ϕ-divergences and "duality" technique. The asymptotic properties of the test statistic are presented both under the null and the alternative hypotheses, and an approximation to the power function is deduced.

Keywords: Biased two samples problem, Logistic regression, Case-control data, Test of homogeneity, Empirical likelihood, ϕ-divergences, Duality.

Department of Statistical Sciences
University of Padua
Italy

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Final version (2004-02-07)

This Paper is partly financed by Progetto Ateneo CPDA023415

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Samuela Leoni-Aubin
tel: +39 049 827 4168
fragola@stat.unipd.it

On Empirical Likelihood in Semiparametric Two-Sample Density Ratio Models

Amor KEZIOU
L.S.T.A.
University Paris 6
France.

Samuela LEONI-AUBIN
Department of Statistical Sciences
University of Padua
Italy

Abstract: We consider estimation and test problems for some semiparametric two-sample density ratio models. The profile empirical likelihood (EL) poses an irregularity problem under the null hypothesis that the laws of the two samples are equal. We show that a "dual" form of the profile EL is well defined even under the null hypothesis. A statistical test, based on the dual form of the EL ratio statistic (ELRS), is then proposed. We give an interpretation for the dual form of the ELRS through ϕ-divergences and "duality" technique. The asymptotic properties of the test statistic are presented both under the null and the alternative hypotheses, and an approximation to the power function is deduced.

Keywords: Biased two samples problem, Logistic regression, Case-control data, Test of homogeneity, Empirical likelihood, ϕ-divergences, Duality.

1 Introduction and notations

In this paper, we consider the following problems : two-sample test for comparing two populations and estimation of the parameters for some semiparametric density ratio models. We dispose of two samples: $X_1, \ldots, X_{n_0}$ with distribution P and $Y_1, \ldots, Y_{n_1}$ with distribution Q. We consider the following semiparametric density ratio model

$$\frac{dQ}{dP}(x) := \exp\left\{\alpha_T + \beta_T^T r(x)\right\}, \tag{1.1}$$

where $\theta_T^T := (\alpha_T, \beta_T^T)$ is the true unknown value of the parameter which we suppose to belong to some open set $\Theta \subset \mathbb{R}^{1+d}$. For simplicity, we sometimes write $m(\theta, x)$ instead of $\exp\left\{\alpha + \beta^T r(x)\right\}$. $r(\cdot)$ is a known function with values in $\mathbb{R}^d$. It often takes the form $r(x) = (x, x^2, \ldots, x^d)^T$, and the model (1.1) is sometimes called "log-linear model" in this case. The supports of the two laws Q and P may be known or unknown, discrete or continuous. We now give some statistical examples and motivations for model (1.1).

1.1 Comparison of two populations

In applications, we often come across with the problem of comparing two laws. The use of the well known $t-$test requires to assume that both samples are normally distributed with unknown means and common known or unknown variance. The $t-$test enjoys several optimal properties, for example it is the uniformly most powerful unbiased test (see e.g. Lehmann (1986)). If both Q and P are normally distributed with equal variance

$$Q = \mathcal{N}(\mu_1, \sigma^2) \text{ and } P = \mathcal{N}(\mu_2, \sigma^2),$$

then, the ratio $\frac{dQ}{dP}$ takes the form

$$\frac{dQ}{dP}(x) = \exp\left\{\alpha + \beta x\right\}, \text{ where } \alpha = \frac{\mu_1^2 - \mu_2^2}{2\sigma^2} \text{ and } \beta = \frac{\mu_2 - \mu_1}{\sigma^2}. \tag{1.2}$$

It follows that testing the hypothesis $\mathcal{H}_0 : Q = P$ is equivalent to testing the parametric hypothesis $\mathcal{H}_0 : \beta = 0$. We underline that $\beta = 0$ implies $\alpha = 0$.
Kay and Little (1987) and Fokianos (2002) observed that there are cases in which the choice

$$\frac{dQ}{dP}(x) = \exp\left\{\alpha + \beta r(x)\right\}, \tag{1.3}$$

where $r(.)$ is an arbitrary but known function, is more appropriate.

When the two samples $X_1, \ldots, X_{n_0}$ and $Y_1, \ldots, Y_{n_1}$ are independent, Fokianos *et al.* (2001) present a statistical test, for the null hypothesis $\mathcal{H}_0 : Q = P$ or equivalently $\mathcal{H}_0 : \beta_T = 0$, where the test statistic is based on a "constrained" empirical likelihood estimate of the parameter β_T (see Qin (1998)) and an empirical estimate of the limit variance.
In the case when the semiparametric assumption (1.1) fails, the test commonly used is the non parametric Wilcoxon rank-sum (see e.g Randles and Wolfe (1979) and Hollander and Wolfe (1999)). We expect it not to be powerful, since it does not use the model (1.1).

For the model (1.1), the empirical likelihood ratio statistic is not well defined under the null hypothesis $\mathcal{H}_0 : Q = P$ (see Section 1.3 below). This problem has been observed also by Zou *et al.* (2002) in the context of a semiparametric mixture models with known weights (see Zou *et al.* (2002) Theorem 1). We propose to use, instead of the empirical likelihood ratio statistic, its "dual" form (see (2.11)) (to perform a test of the null hypothesis $\mathcal{H}_0 : Q = P$) which is well defined regardless of the null hypothesis. Simulation results, presented in Section 4 below, show that the observed level of the test based on the statistic (2.11) converges (to the nominal level) better than the observed level of the test proposed by Fokianos *et al.* (2001). Using ϕ-divergences and "duality" technique, we give an interpretation for the statistic (2.11), the dual form of the empirical likelihood ratio statistic; see (2.21). This interpretation allows us to give the asymptotic law of the proposed test statistic under the alternative hypothesis. We apply this result to give an approximation to the power function in a similar way to Morales and Pardo (2001) who gave some approximations to power functions of ϕ-divergences tests in parametric models. Duality technique has been used by Broniatowski (2003) in order to estimate the Kullback-Leibler divergence without making use of any partitioning nor smoothing. It has been used also by Keziou (2003) and Broniatowski and Keziou (2003) in order to estimate ϕ-divergences between probability measures (without smoothing), and to introduce a new class of estimates and test statistics for discrete or continuous parametric models extending maximum likelihood approach; the use of the duality technique in the context of ϕ-divergences allows also to study the asymptotic properties of the test statistics (including the likelihood ratio one) both under the null and the alternative hypotheses. Recall that a ϕ-divergence between two probability measures Q and P, when Q is absolutely continuous with respect to P, is defined by

$$\phi(Q,P) := \int \varphi\left(dQ/dP\right)\ dP, \tag{1.4}$$

where φ is a real nonnegative convex function satisfying $\varphi(1)=0$. Note that $\phi(Q,P)$ is nonnegative, $\phi(Q,P)=0$ when $Q=P$. Further, if φ is strictly convex on a neighborhood of one, then $\phi(Q,P)=0$ if and only if $Q=P$; we refer to Liese and Vajda (1987) for a systematic theory of ϕ-divergences.

1.2 Logistic model and multiplicative-intercept risk model

Consider the logistic model which has been widely used in statistical applications for the analysis of binary data (see e.g. Agresti (1990), Hosmer and Lemeshow (1999) and Hosmer and Lemeshow (2000)). Suppose that y is a binary response variable and that x is the associate covariate vector. The logistic model has the form

$$\Pr(y=1|x) = \frac{\exp(\gamma+\beta^T x)}{1+\exp(\gamma+\beta^T x)}, \quad \gamma \in \mathbb{R}, \quad \beta \in \mathbb{R}^d. \tag{1.5}$$

Note that the marginal density of x, noted $f(x)$, is left completely unspecified. One of the major reasons of the logistic regression model has seen such a wide use, especially in epidemiologic research, is the ease of obtaining adjusted odds ratios from the estimated slope coefficients when sampling is performed conditional on the outcome variables, as in a case-control study. In a case-control study the binary outcome variable is fixed by stratification. In this type of study design, two random samples of sizes n_0 and n_1 are chosen from the two strata defined by the outcome variable, i.e, from the subsets of the population with $y=0$ and $y=1$, respectively. Assume that $x_1,\ldots,x_{n_0}$ are the observed covariates from the control group and let $x_{n_0+1},\ldots,x_n$ ($n=n_0+n_1$) be those from the case group. We aim to estimate the parameters γ and β using the two samples $X_1,\ldots,X_{n_0}$ and $X_{n_0+1},\ldots,X_n$. We show that the logistic model (1.5) writes in the form of the model (1.1). So, let f denote the density function of the covariates x, and put

$$\pi = \Pr(y=1) = \int \Pr(y=1|x) f(x)\ dx,$$

and assume that

$$f_i(x) = f(x|y=i) = dF(x|y=i)/dx \qquad i=0,1$$

exist and represent the conditional density function of x given $y=i$. It is not difficult to manipulate the case-control likelihood function to obtain a logistic regression model in which the dependent variable is the outcome variable of interest to the investigator. The key step in this development is an application of the Bayes theorem, that yields

$$f_1(x) = \frac{\exp(\gamma + \beta^T x)}{(1 + \exp(\gamma + \beta^T x))\pi} f(x) \text{ and } f_0(x) = \frac{\exp(\gamma + \beta^T x)}{(1 + \exp(\gamma + \beta^T x))(1 - \pi)} f(x).$$

So,

$$\frac{f_1(x)}{f_0(x)} = \frac{1-\pi}{\pi} \exp(\gamma + \beta^T x) = \exp\left\{\gamma + \log\left(\frac{1-\pi}{\pi}\right) + \beta^T x\right\} =: \exp(\alpha + \beta^T x),$$

where $\alpha := \gamma + \log\left(\frac{1-\pi}{\pi}\right)$. Thus, model (1.5) is equivalent to the following two-sample semiparametric model

$$\begin{aligned} x_1, \ldots, x_{n_0} &\sim f(x|y=0) = f_0(x), \\ x_{n_0+1}, \ldots, x_n &\sim f(x|y=1) = f_1(x) = \exp(\alpha + \beta^T x) f_0(x). \end{aligned} \tag{1.6}$$

More generally, we can consider the multiplicative-intercept risk model, i.e,

$$\Pr(y = 1|x) = \frac{\exp(\gamma + \beta^T r(x))}{1 + \exp(\gamma + \beta^T r(x))}, \quad \gamma \in \mathbb{R}, \quad \beta \in \mathbb{R}^d,$$

where $r(.)$ is a given function. In this case, as above, we obtain the following two-sample semiparametric model

$$\begin{aligned} x_1, \ldots, x_{n_0} &\sim f_0(x) \\ x_{n_0+1}, \ldots, x_n &\sim f_1(x) = \exp(\alpha + \beta^T r(x)) f_0(x). \end{aligned} \tag{1.7}$$

Models (1.6) and (1.7) are particular cases of models (1.1) by taking $\frac{dP}{dx} = f_0$ and $\frac{dQ}{dx} = f_1$. For models (1.1), when the two samples $X_1, \ldots, X_{n_0}$ and $Y_1, \ldots, Y_{n_1}$ are independent, Qin (1998) presents an estimation procedure of θ_T based on the empirical likelihood approach (see Owen (1988), Owen (1990) and Owen (2001)), using the likelihood of the independent variables $X_1, \ldots, X_{n_0}$, $Y_1, \ldots, Y_{n_1}$. However, an important special case of the case-control study is the matched (or paired) study. In this design, subjects are stratified on the basis of variables believed to be associated with the outcome (an example of stratification variable is the age for each of the individuals in the survey). Within each stratum, samples of cases ($y = 1$) and controls ($y = 0$) are chosen; the most common matched design includes one case and one control per stratum and is thus referred as 1-1 matched study.
The rest of the paper is organized as follows: we end this Section recalling the estimation method proposed by Qin (1998). In Section 2, we show that the irregularity problem of the profile empirical likelihood can be adjusted in the context of model (1.1). We next give a regularized version of the profile empirical likelihood using duality techniques. A statistical test, for the null hypothesis $\mathcal{H}_0 : Q = P$, is then proposed. An other point of view at the test statistic is given using ϕ-divergences and "duality" technique. In Section 3, we study the asymptotic behavior of the proposed test statistic under the null and the alternative hypotheses with independent samples, and we give an approximation to the power function which leads to

an approximation to the sample sizes n_0 and n_1 guaranteeing a desired power for a given alternative. In Section 4, we present simulation results. Concluding remarks and possible developments are presented in Section 5. All proofs are in Section 6. In the sequel, we sometimes write Pf instead of $\int f(x)\, dP(x)$ for any function f and any measure P.

1.3 The profile empirical likelihood (EL) and its irregularity under the null hypothesis $\mathcal{H}_0 : Q = P$

In the present setting, the estimation method proposed by Qin (1998), which is based on the empirical likelihood approach (see Owen (1988), Owen (1990) and Owen (2001)), can be summarized as follows. For any $\theta \in \Theta$, the empirical likelihood of the two samples $X_1, \dots, X_{n_0}$ and $Y_1, \dots, Y_{n_1}$, if they are independent, is

$$L(\theta) := \prod_{i=1}^{n_0} p(X_i) \prod_{j=1}^{n_1} q(Y_j).$$

For simplicity, denote $(t_1, \dots, t_n)$ the combined sample $(X_1, \dots, X_{n_0}, Y_1, \dots, Y_{n_1})$, where $n := n_0 + n_1$. Since $q(x) = m(\theta, x)p(x)$, then $L(\theta)$ writes

$$L(\theta) = \prod_{i=1}^{n} p(t_i) \prod_{i=n_0+1}^{n} m(\theta, t_i).$$

For convenience we write p_i instead of $p(t_i)$. Hence, the log-likelihood writes

$$l(\theta, p) := \sum_{i=1}^{n} \log p_i + \sum_{i=n_0+1}^{n} \log\left[m(\theta, t_i)\right].$$

The profile log-likelihood (in θ) is then

$$l(\theta) := \sup_{p \in \mathcal{C}_\theta} l(\theta, p), \tag{1.8}$$

where p is constrained to the set

$$\mathcal{C}_\theta := \left\{ p \in \mathbb{R}_+^n \text{ such that } \sum_{i=1}^{n} p_i = 1 \text{ and } \sum_{i=1}^{n} p_i \left[m(\theta, t_i) - 1\right] = 0 \right\}.$$

The EL estimate of θ_T, proposed by Qin (1998), is then

$$\tilde{\theta} := \arg\sup_{\theta \in \Theta} l(\theta). \tag{1.9}$$

Qin (1998) has proved that the estimate $\tilde{\theta}$ is optimal (in the sense of Godambe (1960)), in the class of all estimates obtained by unbiased estimating functions, when $m(\theta, x)$ takes the form $\exp\{\alpha + \beta^T r(x)\}$ and α is unknown (see Qin (1998) Theorem 3).

For a given $\theta \in \Theta$, the profile log-likelihood $l(\theta)$ is well defined (and finite) if and only if

$$\text{there exists } p \in \mathcal{C}_\theta \text{ such that } l(\theta, p) < \infty. \tag{1.10}$$

This condition means that 0 is inside the convex hull generated by the points $[m(\theta, t_1) - 1], \ldots, [m(\theta, t_n) - 1]$, i.e,

$$\min_{1 \leq i \leq n} [m(\theta, t_i) - 1] < 0 < \max_{1 \leq i \leq n} [m(\theta, t_i) - 1]. \tag{1.11}$$

So, when $\beta_T \neq 0$ and if P is not degenerate, using similar arguments to those in Zou *et al.* (2002) Theorem 1, we can show that there exists a neighborhood of θ_T, say $N(\theta_T)$, such that for all $\theta \in N(\theta_T)$, the assumption (1.10) holds as $n_0 \to \infty$. Hence, $\theta \in N(\theta_T) \mapsto l(\theta)$ is well defined for n_0 sufficiently large. However, when $Q = P$ (i.e, when $\beta_T = 0$), then obviously the set $\mathcal{C}_\theta$ is empty for all $\theta = (\alpha, \beta^T)^T \in \Theta$ with $\alpha \neq 0$ and $\beta = 0$. So, when $Q = P$ (i.e, when $\theta_T = 0$), there exists no neighborhood $N(\theta_T)$ of θ_T such that the profile empirical log-likelihood function $\theta \mapsto l(\theta)$ is well defined on all $N(\theta_T)$. Consequently the estimate $\tilde{\theta}$ is not well defined also in this case. In the following Section, we will show, using some arguments of duality theory, that this problem can be adjusted in the context of the model (1.1).

2 Adjustment of the profile empirical likelihood

If the assumption (1.10) holds, then $l(\theta)$ is finite, and the unique "optimal solution" (i.e, the value of p which yields the supremum in (1.8)), as an explicit expression of (1.8) can be derived by a Lagrange multiplier argument and the Khun-Tucher Theorem (see e.g Rockafellar (1970) Section 28). In fact, the "dual" problem associated to the "primal" problem (i.e, the optimization problem (1.8)) writes as follows

$$\begin{aligned} \inf_{\lambda_0, \lambda_1 \in \mathbb{R}} \quad & \left\{ \lambda_0 - n - \sum_{i=1}^{n} \log\left(\lambda_0 + \lambda_1 \left[m(\theta, t_i) - 1\right]\right) \right. \\ & \left. + \sum_{i=n_0+1}^{n} \log\left[m(\theta, t_i)\right] \right\}. \end{aligned} \tag{2.1}$$

So, by the Khun-Tucher Theorem, under condition (1.10), the infimum in (2.1) is attained, and the following equality

$$\begin{aligned} \sup_{p \in \mathcal{C}_\theta} l(\theta, p) = \inf_{\lambda_0, \lambda_1 \in \mathbb{R}} \quad & \left\{ \lambda_0 - n - \sum_{i=1}^{n} \log\left(\lambda_0 + \lambda_1 \left[m(\theta, t_i) - 1\right]\right) \right. \\ & \left. + \sum_{i=n_0+1}^{n} \log\left[m(\theta, t_i)\right] \right\} \end{aligned} \tag{2.2}$$

holds. The "dual" optimal solution, say $(\overline{\lambda}_0, \overline{\lambda}_1)$, (i.e, the argument infimum in (2.1)) can be derived by differentiation. Hence, we obtain $\overline{\lambda}_0 = n$ and $\overline{\lambda}_1$ is the solution (in λ_1) of the equality

$$\sum_{i=1}^{n} \frac{m(\theta, t_i) - 1}{n + \lambda_1 [m(\theta, t_i) - 1]} = 0.$$

Finally, under condition (1.10), the equality

$$l(\theta) := \sup_{p \in \mathcal{C}_\theta} l(\theta, p) = \inf_{\lambda \in \mathbb{R}} \left\{ -\sum_{i=1}^{n} \log \left[n \left(1 + \lambda \left[m(\theta, t_i) - 1 \right] \right) \right] + \sum_{i=n_0+1}^{n} \log \left[m(\theta, t_i) \right] \right\} \tag{2.3}$$

holds with finite values, and the unique optimal solution $(\overline{p}_1, \ldots, \overline{p}_n)$ exists and it is given by

$$\overline{p}_i = \frac{1}{n} \frac{1}{1 + \overline{\lambda}[m(\theta, t_i) - 1]} \quad \text{for all } i = 1, \ldots, n, \tag{2.4}$$

where $\overline{\lambda}$ is the unique dual optimal solution in (2.3). It is the solution (in λ) of the equation

$$-\sum_{i=1}^{n} \frac{1}{n} \frac{m(\theta, t_i) - 1}{1 + \lambda \left[m(\theta, t_i) - 1 \right]} = 0. \tag{2.5}$$

Hence, the EL estimate $\tilde{\theta}$ of θ_T writes

$$\tilde{\theta} := \arg \sup_{\theta \in \Theta} \inf_{\lambda \in \mathbb{R}} \left\{ -\sum_{i=1}^{n} \log \left[n \left(1 + \lambda \left[m(\theta, t_i) - 1 \right] \right) \right] + \sum_{i=n_0+1}^{n} \log \left[m(\theta, t_i) \right] \right\}. \tag{2.6}$$

By differentiation with respect to α and λ, we can see by simple calculus that the Lagrange multiplier $\overline{\lambda}$ in (2.6) has the explicit solution $\overline{\lambda}(\tilde{\theta}) = \frac{n_1}{n}$ which does not depend on the data. Hence, the value of the log-likelihood (2.3) in $\tilde{\theta}$ is

$$l(\tilde{\theta}) = -n \log n - \sum_{i=1}^{n} \log \left(1 + \frac{n_1}{n} \left[m(\tilde{\theta}, t_i) - 1 \right] \right) + \sum_{i=n_0+1}^{n} \log \left[m(\tilde{\theta}, t_i) \right], \tag{2.7}$$

and the EL estimate $\tilde{\theta}$ can be written as

$$\tilde{\theta} = \arg \sup_{\theta \in \Theta} \left\{ -n \log n - \sum_{i=1}^{n} \log \left(1 + \frac{n_1}{n} \left[m(\theta, t_i) - 1 \right] \right) + \sum_{i=n_0+1}^{n} \log \left[m(\theta, t_i) \right] \right\}. \tag{2.8}$$

Under the null hypothesis $\mathcal{H}_0 : Q = P$, i.e, when $\beta_T = 0$, the profile log-likelihood $l(\theta)$ is not defined for some θ (see Section 1.3). So, in view of (2.7) and (2.8), we propose to consider, instead of $l(\theta)$, the "dual form":

$$l_d(\theta) := -n \log n - \sum_{i=1}^{n} \log \left(1 + \frac{n_1}{n} \left[m(\theta, t_i) - 1 \right] \right) + \sum_{i=n_0+1}^{n} \log \left[m(\theta, t_i) \right], \tag{2.9}$$

which is well defined for all $\theta \in \Theta$ regardless of the null hypothesis $\mathcal{H}_0 : Q = P$, and to redefine the EL estimate as

$$\hat{\theta} := \arg\sup_{\theta\in\Theta} l_d(\theta), \tag{2.10}$$

Note that, under condition (1.10), we have $\hat{\theta} = \tilde{\theta}$ and $l_d(\hat{\theta}) = l(\tilde{\theta})$. Now, we give an interpretation to the "dual form"

$$S_n := 2W_d(\hat{\theta}) := 2\left[\sup_{\theta\in\Theta} l_d(\theta) + n\log n\right] \tag{2.11}$$

of the empirical likelihood ratio statistic

$$2W(\tilde{\theta}) := 2\left[\sup_{\theta\in\Theta} l(\theta) + n\log n\right] \tag{2.12}$$

(associated to the null hypothesis $\mathcal{H}_0 : Q = P$). First, denote $\rho_n := n_1/n_0$, $a_n := n\rho_n(1+\rho_n)^{-2}$, and let Q_{n_1} and P_{n_0} to be, respectively, the empirical measures associated to the samples $Y_1, \ldots, Y_{n_1}$ and $X_1, \ldots, X_{n_0}$, namely

$$Q_{n_1} := \frac{1}{n_1}\sum_{i=1}^{n_1} \delta_{Y_i} \quad \text{and} \quad P_{n_0} := \frac{1}{n_0}\sum_{i=1}^{n_0} \delta_{X_i},$$

with δ_x denotes the Dirac measure at point x, for all x. By simple calculus, we can show that the statistic (2.11) writes as follows:

$$S_n = 2a_n \sup_{\theta\in\theta}\left\{\int f_{\rho_n}(\theta,x)\, dQ_{n_1}(x) - \int g_{\rho_n}(\theta,x)\, dP_{n_0}(x)\right\}, \tag{2.13}$$

where

$$f_{\rho_n}(\theta,x) := (1+\rho_n)\log\left[m(\theta,x)\right] - (1+\rho_n)\log\left[1+\rho_n m(\theta,x)\right] + (1+\rho_n)\log\left(1+\rho_n\right) \tag{2.14}$$

and

$$g_{\rho_n}(\theta,x) := \frac{1+\rho_n}{\rho_n}\log\left[1+\rho_n m(\theta,x)\right] - \frac{1+\rho_n}{\rho_n}\log\left(1+\rho_n\right). \tag{2.15}$$

In (2.13), the sequence a_n is a normalizing term and the second term can be seen as an empirical estimate of

$$\sup_{\theta\in\Theta}\left\{\int f_\rho(\theta,x)\, dQ(x) - \int g_\rho(\theta,x)\, dP(x)\right\}, \tag{2.16}$$

where $\rho := \lim_{n\to\infty}\rho_n$ (which we suppose to be positive),

$$f_\rho(\theta,x) := (1+\rho)\log\left[m(\theta,x)\right] - (1+\rho)\log\left[1+\rho m(\theta,x)\right] + (1+\rho)\log\left(1+\rho\right) \tag{2.17}$$

and

$$g_\rho(\theta,x) := \frac{1+\rho}{\rho}\log\left[1+\rho m(\theta,x)\right] - \frac{1+\rho}{\rho}\log\left(1+\rho\right). \tag{2.18}$$

On the other hand, using the so-called "dual representation of ϕ-divergences" (see Theorem 2.1 in Keziou (2003) or Theorem 4.4 in Broniatowski and Keziou (2004)) and choosing the class of functions $\mathcal{F} := \left\{x \mapsto \varphi_\rho^{\star\prime}\left(m(\theta, x)\right); \theta \in \Theta\right\}$, we can prove the equality

$$\sup_{\theta\in\Theta}\left\{\int f_\rho(\theta,x)\ dQ(x) - \int g_\rho(\theta,x)\ dP(x)\right\} = \int \varphi_\rho^\star\left(\frac{dQ}{dP}\right)\ dP =: \phi^\star(Q,P), \tag{2.19}$$

where $\varphi_\rho^\star$ is the nonnegative real strictly convex function defined on $\mathbb{R}_+$ by

$$\varphi_\rho^\star(x) := (1+\rho)\left[x\log x - \frac{1+\rho x}{\rho}\log(1+\rho x) + \frac{1}{\rho}\log(1+\rho) + x\log(1+\rho)\right], \tag{2.20}$$

which is a member of the class of $\phi-$divergences (1.4). In other words, by (2.13), (2.16) and (2.19), $a_n^{-1}W_d\left(\hat\theta\right)$ can be seen as an empirical estimate (which we denote $\hat\phi^\star(Q,P)$) of $\phi^\star(Q,P)$, the $\phi^\star$-divergence between Q and P, i.e, $\hat\phi^\star(Q,P) := (2a_n)^{-1}S_n$. Since $\phi^\star(Q,P)$ is nonnegative and takes value 0 only when $Q = P$, it is reasonable to perform a test that rejects the null hypothesis $\mathcal{H}_0 : Q = P$ when the statistic

$$S_n = 2a_n\hat\phi^\star(Q,P) = 2a_n\sup_{\theta\in\theta}\left\{\int f_{\rho_n}(\theta,x)\ dQ_{n_1}(x) - \int g_{\rho_n}(\theta,x)\ dP_{n_0}(x)\right\} \tag{2.21}$$

(see (2.11) and (2.13)) takes large values.

The estimate $\hat\theta$ of θ_T (see (2.10)), writes

$$\hat\theta = \arg\sup_{\theta\in\Theta}\left\{\int f_{\rho_n}(\theta,x)\ dQ_{n_1}(x) - \int g_{\rho_n}(\theta,x)\ dP_{n_0}(x)\right\}. \tag{2.22}$$

On the other hand, by Theorem 2.1 in Keziou (2003) or Theorem 4.4 in Broniatowski and Keziou (2004), we can prove that the supremum in (2.16) is unique and reached at $\theta = \theta_T$. This indicates that the estimate $\hat\theta$ of θ_T may converge (as M-estimate) to θ_T even when the samples are paired.

3 Asymptotic behavior of the estimate and test statistic under the null and the alternative hypotheses, and approximation of the power function

In this Section, for independent samples, we give the asymptotic properties of the estimate $\hat\theta$ (of the parameter θ_T) and the test statistic (2.21) both under the null and the alternative hypotheses. As an application, we obtain an approximation to the power function for a given alternative. In all the sequel, $f'(\theta,x)$ and $f''(\theta,x)$ denote respectively the gradient and the Hessian of f at the point θ, for all x and any function f. $|.|$ denotes the Euclidean norm. Let $\rho_{n_1} := n_1/n$ and $\rho_{n_0} := n_0/n$, and assume that $\rho_{n_1} \to \rho_1 > 0$ and $\rho_{n_0} \to \rho_0 > 0$ when $n = n_0 + n_1 \to \infty$. Denote also $l_{\phi^\star}(\theta) := a_n\left[Q_{n_1}f_{\rho_n}(\theta) - P_{n_0}g_{\rho_n}(\theta)\right]$. In all the sequel, for simplicity, we write f and g instead of f_ρ and g_ρ defined in (2.17) and (2.18). We give our results under the following assumptions

(A.1) There exists a neighborhood $N(\theta_T)$ of θ_T such that the third order partial derivative functions $\left\{x \mapsto (\partial^3/\partial\theta_i\partial\theta_j\partial\theta_k)f(\theta,x);\ \theta \in N(\theta_T)\right\}$ (resp. $\left\{x \mapsto (\partial^3/\partial\theta_i\partial\theta_j\partial\theta_k)g(\theta,x);\ \theta \in N(\theta_T)\right\}$) are dominated by some function Q-integrable (resp. some function P-integrable);

(A.2) The integrals $Q\left|f'(\theta_T)\right|^2$, $P\left|g'(\theta_T)\right|^2$, $Q\left|f''(\theta_T)\right|$ and $P\left|g''(\theta_T)\right|$ are finite, and the matrix $[Qf''(\theta_T) - Pg''(\theta_T)]$ is non singular.

Theorem 3.1 *Assume that assumptions (A.1-2) hold.*

(a) *Let $B(\theta_T, n^{-1/3}) := \left\{\theta \in \Theta;\ |\theta - \theta_T| \le n^{-1/3}\right\}$. Then as $n \to \infty$, with probability one, $l_{\phi^\star}(\theta)$ attains its maximum value at some point $\hat\theta$ in the interior of the ball $B(\theta_T, n^{-1/3})$, and the estimate $\hat\theta$ satisfies $l'_{\phi^\star}(\hat\theta) = 0$.*

(b) *$\sqrt{n}\left(\hat\theta - \theta_T\right)$ converges in distribution to a centered multivariate normal variable with covariance matrix*

$$\begin{aligned} LCM &= \left[-Qf''(\theta_T) + Pg''(\theta_T)\right]^{-1} \cdot \left[\rho_1^{-1}\left(Qf'(\theta_T)f'(\theta_T)^T - \right.\right. \\ &\quad \left. -Qf'(\theta_T)Qf'(\theta_T)^T\right) + \rho_0^{-1}\left(Pg'(\theta_T)g'(\theta_T)^T - \right. \\ &\quad \left.\left. -\ Pg'(\theta_T)Pg'(\theta_T)^T\right)\right] \cdot \left[-Qf''(\theta_T) + Pg''(\theta_T)\right]^{-1}. \end{aligned} \tag{3.1}$$

If $Q = P$, then the limit covariance matrix is

$$LCM = \frac{(1+\rho)^2}{\rho}\begin{bmatrix} 1 & Pr^T \\ Pr & P(rr^T) \end{bmatrix}^{-1}. \tag{3.2}$$

(c) *Under the null hypothesis $\mathcal{H}_0 : Q = P$, the statistic S_n converges in distribution to a χ^2 variable with d degrees of freedom.*

In order to give the asymptotic properties of the test statistic S_n under the alternative hypothesis $\mathcal{H}_1 : Q \neq P$, we need the following additional assumption pertaining to the function f and g defined in (2.17) and (2.18).

(A.3) The integrals $Q\left(f(\theta_T)^2\right)$ and $P\left(g(\theta_T)^2\right)$ are finite.

Theorem 3.2 *Assume that assumptions (A.1-3) hold. Then, under the alternative hypothesis $\mathcal{H}_1 : Q \neq P$, we have*

$$\sqrt{a_n}\left[(2a_n)^{-1}S_n - \phi^\star(Q,P)\right] \tag{3.3}$$

converges in distribution to a centered normal variable with variance

$$\sigma^2(\theta_T) = \rho_0\left[Q(f^2) - (Qf)^2\right] + \rho_1\left[P(g^2) - (Pg)^2\right]. \tag{3.4}$$

Remark 3.1 *Using Theorem 3.1 part (c), we propose to reject the null hypothesis $\mathcal{H}_0 : Q = P$ if $S_n > \chi^2_\epsilon(d)$, where $\chi^2_\epsilon(d)$ is the $(1-\epsilon)$-quantile of the χ^2 distribution with d degrees of freedom. This leads to a test asymptotically of level ϵ. The* *asymptotic result in Theorem 3.2 allows to give an approximation to the power function for a given alternative: for a given* $\beta_T \neq 0$*, we obtain for the power function* $\beta(\theta_T) := P_{\theta_T}\left\{S_n > \chi^2_\epsilon(d)\right\}$ *the following approximation*

$$\beta(\theta_T) \approx 1 - F_{\mathcal{N}}\left(\frac{\sqrt{a_n}}{\hat{\sigma}(\theta_T)}\left[(2a_n)^{-1}\chi^2_\epsilon(d) - H_n(\theta_T)\right]\right), \tag{3.5}$$

where $F_{\mathcal{N}}(.)$ *is the cumulative distribution function of a normal variable with mean zero and variance one,*

$$\hat{\sigma}(\theta_T)^2 := \rho_{n_0}\left[Q_{n_1}(f(\theta_T)^2) - (Q_{n_1}f(\theta_T))^2\right] + \rho_{n_1}\left[P_{n_0}(g(\theta_T)^2) - (P_{n_0}g(\theta_T))^2\right],$$

and $H_n(\theta_T) := Q_{n_1}f(\theta_T) - P_{n_0}g(\theta_T)$*. Note also that the power* $\beta(\theta_T)$*, by the asymptotic result in Theorem 3.2, tends to one, as* $n \to \infty$*, under the alternative hypothesis* $\mathcal{H}_1 : Q \neq P$*.*

4 Simulation results

In this Section, we present some simulation results concerning the testing problem of the null hypothesis of homogeneity (see Example 1 below). Various examples of the choices of $m(\theta, x)$ can be founded in the papers by Qin (1998), Kay and Little (1987) and Cox and Ferry (1991). In all examples, we consider the nominal level 5%; it is represented, in all figures, by an horizontal dotted line. The value of β_T corresponding to $\mathcal{H}_0$ in all cases is represented by a vertical dotted line. The power is plotted as a function of β; note that for any test, the power associated to the value of β corresponding to the null hypothesis $\mathcal{H}_0$ is the observed level of the test. Example 2 concerns the power approximation discussed in Remark 3.1.

4.1 Example 1- Comparison of two populations

We compare the power function of the ELR test (ELRT), defined in (2.21), with the power function of the two-sample t-test, Wilcoxon rank-sum test and Fokianos *et al.* (2001) test. We recall that Fokianos *et al.* (2001) test statistic is based on a constrained empirical likelihood estimate of the parameter (see Qin (1998)) and an empirical estimate of its limit variance. Three cases are considered. In the first case, we have $X \sim \mathcal{N}(\beta, 1)$, $Y \sim \mathcal{N}(0, 1)$ and $m(\theta, x) = \exp\{\alpha + \beta x\}$. In the second case, we have two lognormal populations, $X \sim LN(\beta, 1)$, $Y \sim LN(0, 1)$ and $m(\theta, x) = \exp\{\alpha + \beta \log x\}$. In the third case, we have two gamma populations $X \sim Ga(3 + \beta, 1)$, $Y \sim Ga(3, 1)$ and $m(\theta, x) = \exp\{\alpha + \beta \log x\}$. The power function is plotted for sample sizes $n_0 = n_1 = 50$. Each power entry was obtained from 1000 independent runs. Under normal and variance equality assumptions, we observe (see Fig. 1) that the four tests are very similar. The fact that our test displays more power than the t-test in the cases of lognormal (see Fig. 2) and gamma populations (see Fig. 3) shows that a departure from the classical normal and variance equality assumptions can considerably weaken the t-test. Note that the ELRT is not dominated by the t-test in the present normal example with equal variances. Apparently, the Wilcoxon rank-sum test has less power than the test

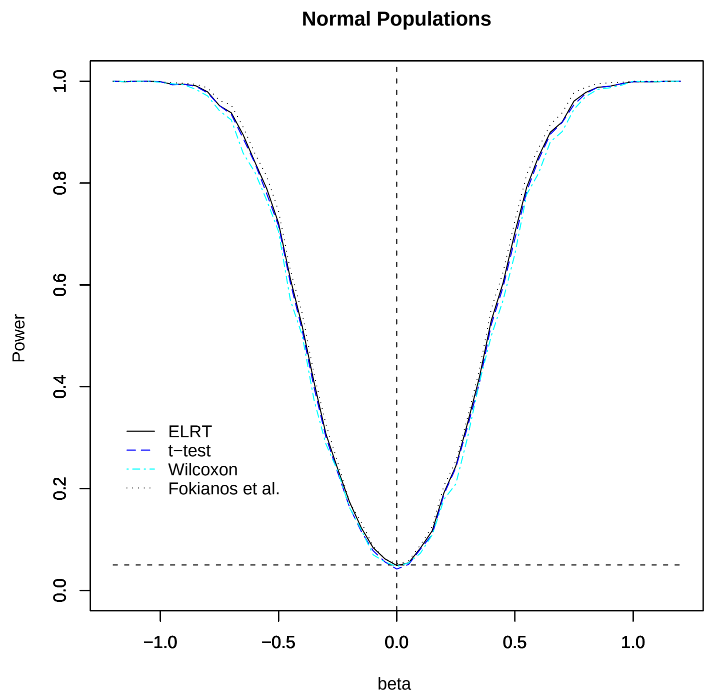

Figure 1: Example 1.a - Two normal populations.

provided here in all the three cases considered. Finally, note that in the gamma case (see Fig. 3) the observed level of the test proposed by Fokianos *et al.* (2001) is far from the nominal level 5%. We conclude that the ELRT (2.21) is more convenient.

4.2 Example 2- Power approximation

In the context of the model $m(\theta, x) = \exp\{\alpha + \beta x\}$ we consider the problem of testing

$$\mathcal{H}_0 : Q = P \text{ versus } \mathcal{H}_1 : Q \neq P$$

or equivalently

$$\mathcal{H}_0 : \beta_T = 0 \text{ versus } \mathcal{H}_1 : \beta_T \neq 0$$

based on the test statistic S_n (2.21). In this example, we consider $X \sim \mathcal{N}(\beta, 1)$, $Y \sim \mathcal{N}(0, 1)$. We study numerically the accuracy of the power approximation given in Remark 3.1. We recall that the approximation (given in Remark 3.1) for the power function $\beta(\theta_T) = P_{\theta_T}\left(S_n \geq \chi^2_{0.05}(1)\right)$ is

$$approx(\theta_T) = 1 - F_{\mathcal{N}}\left(\frac{\sqrt{a_n}}{\hat{\sigma}(\theta_T)}\left[(2a_n)^{-1}\chi^2_{0.05}(1) - H_n(\theta_T)\right]\right), \tag{4.1}$$

where $F_{\mathcal{N}}(.)$ is the cumulative distribution function of a normal variable with mean zero and variance one,

$$\hat{\sigma}(\theta_T)^2 := \rho_{n_0}\left[Q_{n_1}(f(\theta_T)^2) - (Q_{n_1}f(\theta_T))^2\right] + \rho_{n_1}\left[P_{n_0}(g(\theta_T)^2) - (P_{n_0}g(\theta_T))^2\right],$$

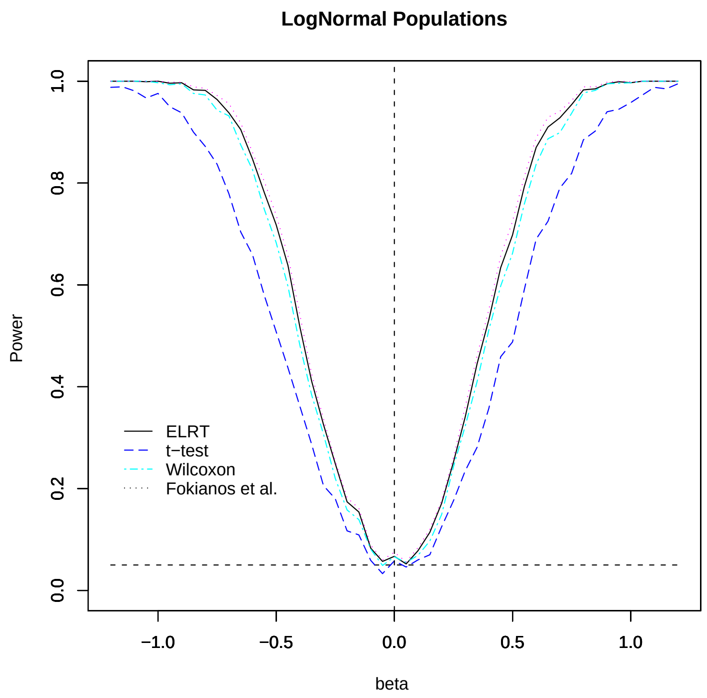

Figure 2: Example 1.b - Two lognormal populations.

and $H_n(\theta_T) := Q_{n_1}f(\theta_T) - P_{n_0}g(\theta_T)$. The power $\beta(\theta_T)$ is plotted for sample sizes $n_0 = n_1 = 30$, $n_0 = n_1 = 50$ and $n_0 = n_1 = 100$, and for different values of β. Each power entry was obtained from 1000 independent runs. The approximation (4.1) is plotted as a function of β by a dotted line; H_n and $\hat{\sigma}$, in (4.1), are calculated (from 1000 simulations) with sample sizes $n_0 = n_1 = 30$, $n_0 = n_1 = 50$ and $n_0 = n_1 = 100$. We observe (see Fig. 4) that the approximation is accurate for alternatives which are not very "near" to the null hypothesis even for moderate sample sizes.

5 Concluding remarks and possible developments

We have addressed the problems of estimation and test of homogeneity in semi-parametric two-sample density ratio models. The profile EL poses an irregularity problem under the null hypothesis $\mathcal{H}_0$ that the two laws of the two samples are equal. We have showed that the dual form of the profile EL is well defined even under the null hypothesis, then we have proposed a test of homogeneity based on the dual form of the EL ratio statistic. We have showed, using the dual representation of ϕ-divergences, that the test statistic can be seen as an estimate of the particular divergence $\phi^\star$ between the two laws, and that the EL estimate $\hat{\theta}$ of θ_T can be seen as the dual optimal solution in the dual representation of the $\phi^\star$-divergence. The advantage of this interpretation is twice:

- It permits to obtain the limit law of the test statistic under the alternative

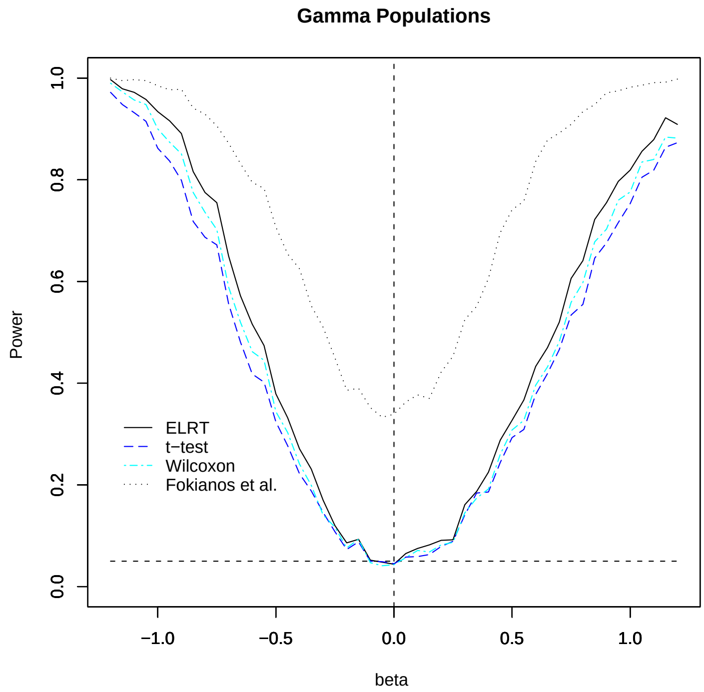

Figure 3: Example 1.c - Two gamma populations.

hypothesis which we use to give an approximation to the power function.

- It suggests to generalize the test and the estimate of the parameter to a class of tests and to a class of estimates using other divergences, and it would be interesting in this case to give how to choose the divergence which leads to an "optimal" (in some sense) estimate or test in term of efficiency and robustness.

In the important case of paired samples, the asymptotic results presented in Section 4 hold with some modifications. The method can be generalized to corresponding problems involving more than two samples. Simple and composite tests on the parameter and approximations to the corresponding power functions can be obtained in a similar way. It would be worthwhile also to involve the problem of Bartlett correctability of the test statistic S_n. These developments will be reported in future communications.

6 Proofs

Proof of Theorem 3.1. (a) We prove this part using some similar arguments to those in Qin and Lawless (1994) and Zou *et al.* (2002). Simple calculus give

$$Qf'(\theta_T) - Pg'(\theta_T) = 0 \tag{6.1}$$

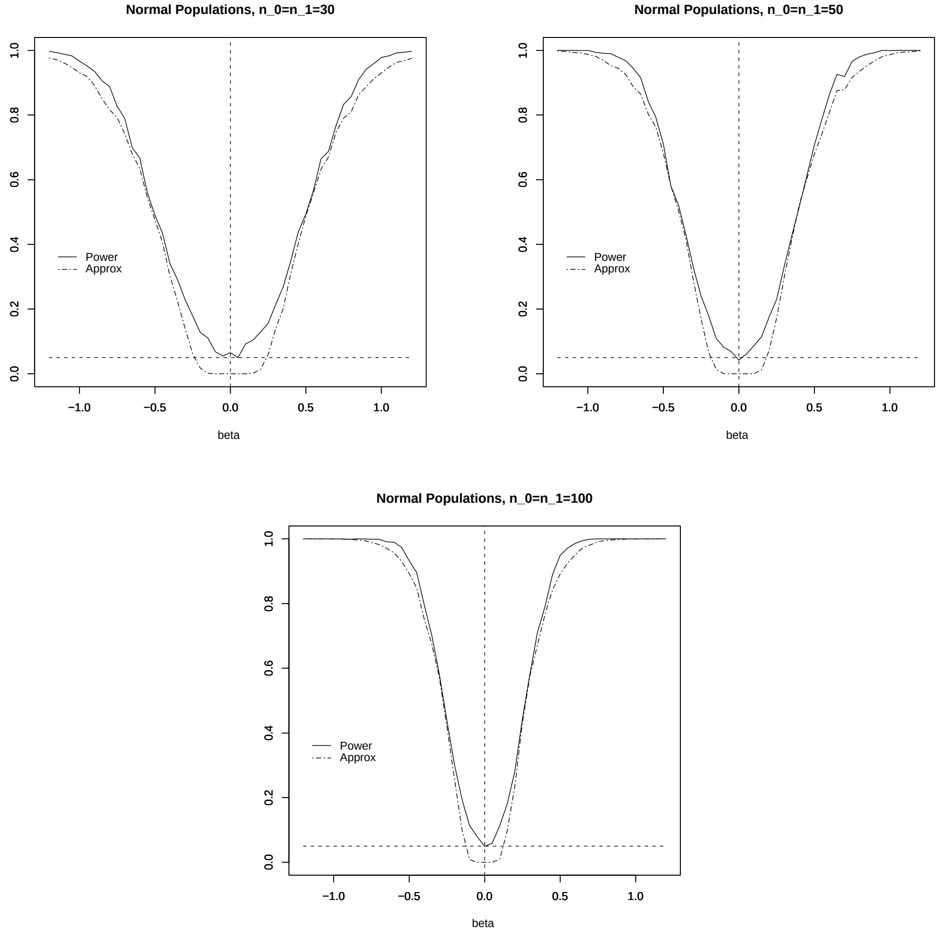

Figure 4: Example 2- Approximation of power function.

and

$$Qf''(\theta_T) - Pg''(\theta_T) = -P\left(m'(\theta_T)m'(\theta_T)^T\varphi_\rho^{\star\prime\prime}(m(\theta_T))\right) =: -V_1(\theta_T). \tag{6.2}$$

Observe that the matrix (6.2) is symmetric. Let $U_n(\theta_T) := Q_{n_1}f'(\theta_T) - P_{n_0}g'(\theta_T)$ and use (6.1) and condition (A.2) in connection with the Central Limit Theorem to see that

$$\sqrt{n}U_n(\theta_T) \to \mathcal{N}\left(0, V_2(\theta_T)\right), \tag{6.3}$$

with $V_2(\theta_T) := \rho_1^{-1}\left[Q\left(f'f'^T\right) - Qf'Qf'^T\right] + \rho_0^{-1}\left[P\left(g'g'^T\right) - Pg'Pg'^T\right]$. Also, let $V_n(\theta_T) := Q_{n_1}f''(\theta_T) - P_{n_0}g''(\theta_T)$ and use (A.2) and (6.2) in connection with the Law of Large Numbers to conclude that

$$V_n(\theta_T) \to -V_1(\theta_T) \quad (a.s). \tag{6.4}$$

Now for $\theta = \theta_T + un^{-1/3}$ with $|u| \leq 1$ consider a Taylor expansion of $l_{\phi^\star}(\theta)$ around θ_T, and use (A.1) and the fact that $l'_{\phi^\star}(\theta) = n\rho_n(1+\rho_n)^{-2}U_n(\theta)$ with $\rho_n \to \rho > 0$, to see that (a.s)

$$l_{\phi^\star}(\theta) - l_{\phi^\star}(\theta_T) = n^{2/3}\rho(1+\rho)^{-2}u^T U_n + 2^{-1}n^{1/3}\rho(1+\rho)^{-2}u^T V_n u + O(1) \quad (6.5)$$

uniformly on u with $|u| \leq 1$. Now, using (6.4) and the fact that $U_n = O\left(n^{-1/2}(\log\log n)^{1/2}\right)$ (a.s) to conclude that

$$l_{\phi^\star}(\theta) - l_{\phi^\star}(\theta_T) = O\left(n^{1/6}(\log\log n)^{1/2}\right) - 2^{-1}\rho(1+\rho)^{-2}u^T V_1 u n^{1/3} + O(1) \ \ (a.s).$$

Hence, uniformly on the surface of the ball $B(\theta_T, n^{-1/3})$ (i.e, uniformly on u with $|u| = 1$), we have

$$l_{\phi^\star}(\theta) - l_{\phi^\star}(\theta_T) \leq O\left(n^{1/6}(\log\log n)^{1/2}\right) - 2^{-1}\rho(1+\rho)^{-2}cn^{1/3} + O(1) \ \ (a.s), \quad (6.6)$$

where c is the smallest eigenvalue of the matrix V_1. Note that c is positive since the matrix V_1 defined in (6.2) is positive definite (it is symmetric, and non singular by assumption). In view of (6.6), by the continuity of $\theta \mapsto l_{\phi^\star}(\theta)$, it holds that as $n \to \infty$, with probability one, $l_{\phi^\star}(\theta)$ attains its maximum value at some point $\hat{\theta}$ in the interior of the ball $B(\theta_T, n^{-1/3})$, and therefore the estimate $\hat{\theta}$ satisfies $l'_{\phi^\star}(\hat{\theta}) = 0$ and $\hat{\theta} - \theta_T = O(n^{-1/3})$.
(b) Using the fact that $l'_{\phi^\star}(\hat{\theta}) = 0$ and a Taylor expansion of $l'_{\phi^\star}(\hat{\theta})$ around θ_T, we obtain

$$0 = a_n^{-1} l'_{\phi^\star}(\hat{\theta}) = a_n^{-1} l'_{\phi^\star}(\theta_T) + a_n^{-1} l''_{\phi^\star}(\theta_T)\left(\hat{\theta} - \theta_T\right) + o(n^{-1/2}).$$

Hence,

$$\sqrt{n}\left(\hat{\theta} - \theta_T\right) = -V_n^{-1}(\theta_T)\sqrt{n}U_n(\theta_T) + o(1) \quad (6.7)$$

where U_n and V_n are defined as in the proof of part (a). Using (6.3) and (6.4), by application of Slutsky Theorem, we may conclude then $\sqrt{n}\left(\hat{\theta} - \theta_T\right) \to \mathcal{N}(0, LCM)$ where LCM is given by (3.1). When $Q = P$, simple calculus leads to (3.2).

(c) First, recall that $Q = P$ implies that $\theta_T = 0$. Hence, from (6.7) using the convergence (6.4), we get

$$\hat{\theta} = V_1^{-1}(0)U_n(0) + o(n^{-1/2}), \quad (6.8)$$

where $V_1(0) = P\left[(1, r^T)^T(1, r^T)\right]$, $U_n(0) = \left(0, W_n(0)^T\right)^T$ and $W_n(0) := Q_{n_1}(\partial/\partial\beta)f(0) - P_{n_0}(\partial/\partial\beta)g(0)$. On the other hand, a Taylor expansion of $2l_{\phi^\star}(\hat{\theta})$ in $\hat{\theta}$ around $\theta_T = 0$, using the fact that $l_{\phi^\star}(0) = 0$, gives

$$\begin{aligned} 2l_{\phi^\star}(\hat{\theta}) &= 2l'_{\phi^\star}(0)^T\hat{\theta} + \hat{\theta}^T l''_{\phi^\star}(0)\hat{\theta} + o(1) \\ &= 2a_n U_n(0)^T\hat{\theta} + a_n\hat{\theta}^T V_n(0)\hat{\theta} + o(1) \\ &= 2a_n U_n(0)^T\hat{\theta} - a_n\hat{\theta}^T V_1(0)\hat{\theta} + o(1). \end{aligned}$$

Combining this with (6.8) to conclude that

$$2l_{\phi^\star}(\hat{\theta}) = a_n W_n(0)^T {V_P}^{-1} W_n(0) + o(1) \text{ where } V_P := P(rr^T) - (Pr)(Pr)^T.$$

It follows that $2l_{\phi^\star}(\hat\theta)$ converges in distribution to a χ^2 variable with d degrees of freedom, since $\sqrt{a_n}W_n(0) \to \mathcal{N}(0, V_P)$ in distribution.

Proof of Theorem 3.2. First, observe that when $Q \neq P$, then $\beta_T \neq 0$ and $\theta_T = (\alpha_T, \beta_T^T)^T \neq 0$. Furthermore,

$$\phi^\star(Q,P) = \int \varphi_\rho^\star(dQ/dP)\; dP = \int \varphi_\rho^\star(m(\theta_T))\; dP = Qf(\theta_T) - Pg(\theta_T) \tag{6.9}$$

which is finite (by assumption (A.3)) and positive. A Taylor expansion, of $(2a_n)^{-1}S_n = a_n^{-1}l_{\phi^\star}(\hat\theta)$ in $\hat\theta$ around θ_T, gives

$$(2a_n)^{-1}S_n = Q_{n_1}f(\theta_T) - P_{n_0}g(\theta_T) + o(n^{-1/2}). \tag{6.10}$$

Combining this with (6.9) to conclude that

$$\begin{aligned}\sqrt{a_n}\left[(2a_n)^{-1}S_n - \phi^\star(Q,P)\right] = \sqrt{a_n}\left[Q_{n_1}f(\theta_T) - Qf(\theta_T)\right] - \\ \sqrt{a_n}\left[P_{n_0}g(\theta_T) - Pf(\theta_T)\right] + o(1)\end{aligned}$$

which converges in distribution to a centered normal variable with variance

$$\sigma^2(\theta_T) = \rho_0\left[Q(f^2) - (Qf)^2\right] + \rho_1\left[P(g^2) - (Pg)^2\right].$$

References

Agresti, A. (1990). *Categorical data analysis*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York. A Wiley-Interscience Publication.

Broniatowski, M. (2003). Estimation of the Kullback-Leibler divergence. *Math. Methods Statist.*, **12**(4), 391–409 (2004).

Broniatowski, M. and Keziou, A. (2003). Parametric estimation and testing through divergences. *Preprint 2004-1, L.S.T.A - Université Paris 6.*

Broniatowski, M. and Keziou, A. (2004). On the mimimization of ϕ-divergences on sets of signed measures. *Submitted to S.T.U.D.I.A.*

Cox, T. F. and Ferry, G. (1991). Robust logistic discrimination. *Biometrika*, **78**(4), 841–849.

Fokianos, K. (2002). Box-cox transformation for semiparametric comparison of two samples, *to appear in. Shoresh Conference Proceedings.*

Fokianos, K., Kedem, B., Qin, J., and Short, D. A. (2001). A semiparametric approach to the one-way layout. *Technometrics*, **43**(1), 56–65.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.*, **31**, 1208–1211.

Hollander, M. and Wolfe, D. A. (1999). *Nonparametric statistical methods*. John Wiley & Sons Inc., New York, second edition.

Hosmer, Jr., D. W. and Lemeshow, S. (1999). *Applied survival analysis*. Wiley Series in Probability and Statistics: Texts and References Section. John Wiley & Sons Inc., New York. Regression modeling of time to event data, A Wiley-Interscience Publication.

Hosmer, Jr., D. W. and Lemeshow, S. (2000). *Applied logistic regression*. Wiley Series in Probability and Statistics: Texts and References Section. John Wiley & Sons Inc., New York.

Kay, R. and Little, S. (1987). Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika*, **74**(3), 495–501.

Keziou, A. (2003). Dual representation of ϕ-divergences and applications. *C. R. Math. Acad. Sci. Paris*, **336**(10), 857–862.

Lehmann, E. L. (1986). *Testing statistical hypotheses*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, second edition.

Liese, F. and Vajda, I. (1987). *Convex statistical distances*, volume 95. BSB B. G. Teubner Verlagsgesellschaft, Leipzig.

Morales, D. and Pardo, L. (2001). Some approximations to power functions of ϕ-divergences tests in paramtric models. *Test*, **10**(2), 249–269.

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. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika*, **85**(3), 619–630.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**(1), 300–325.

Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics*. John Wiley & Sons, New York-Chichester-Brisbane. Wiley Series in Probability and Mathematical Statistics.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton University Press, Princeton, N.J.

Zou, F., Fine, J. P., and Yandell, B. S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika*, **89**(1), 61–75.

Acknowledgements

The authors thank Professor Michel Broniatowski for his helpful discussions and suggestions leading to improvement of this paper.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

Department of Statistical Sciences
University of Padua
Italy