Available online at www.sciencedirect.com

ScienceDirect

Journal of Statistical Planning and Inference 138 (2008) 915–928

journal of statistical planning and inference

www.elsevier.com/locate/jspi

On empirical likelihood for semiparametric two-sample density ratio models

Amor Keziou[a,*], Samuela Leoni-Aubin[a,b]

[a]*Laboratoire de Mathématiques (UMR 6056) CNRS, Université de Reims and LSTA, Université Paris 6, UFR Sciences, Moulin de la Housse, B.P. 1039, 51687 Reims, France*

[b]*Université de Technologie de Compiègne, Centre de Recherche de Royallieu, Rue Personne de Roberval - BP 20529, 60205 Compiègne, France*

Received 18 February 2005; received in revised form 4 October 2006; accepted 28 February 2007

Available online 22 April 2007

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 techniques. The asymptotic properties of the test statistic are presented both under the null and the alternative hypotheses, and approximation of the power function of the test is deduced.

Keywords: Biased two samples problem; Logistic regression; Case-control data; Test of homogeneity; Power function; Empirical likelihood; ϕ-Divergences; Duality

1. Introduction and notation

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:

$$\frac{\mathrm{d}Q}{\mathrm{d}P}(x) := \exp\{\alpha_T + \beta_T^{\mathrm{T}} r(x)\}, \tag{1.1}$$

where $\theta_T^{\mathrm{T}} := (\alpha_T, \beta_T^{\mathrm{T}})$ is the true unknown value of the parameter which we suppose to belong to some open set $\Theta \subset \mathbb{R}^{1+d}$ and $r(\cdot)$ is a known function with values in $\mathbb{R}^d$. The two densities are assumed unknown but are related, however, through a tilt (or distortion) which determines the difference between them. For simplicity, we sometimes write $m(\theta, x)$ instead of $\exp\{\alpha + \beta^{\mathrm{T}} r(x)\}$. The supports of the two laws Q and P may be known or unknown, discrete or continuous.

The density ratio model has attracted much attention recently, because it relaxes several conventional assumptions in the context of multi-sample problems and because fitting can be easily implemented in standard software.

* Corresponding author.

E-mail addresses: amor.keziou@univ-reims.fr (A. Keziou), samuela.leoni-aubin@utc.fr (S. Leoni-Aubin).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2007.02.009

For an application of the density ratio model to meteorological data, see Fokianos et al. (1998). For further applications of model (1.1), see Fokianos et al. (2001) and Qin et al. (2002).

It is useful to say that expression (1.1) can be viewed as a biased sampling model with weights depending on parameters. Vardi (1982, 1985) and Gill et al. (1988) have discussed inference in biased sampling models with known weight functions. Gilbert (2000) and Gilbert et al. (1999) considered weight functions depending on an unknown finite dimensional parameter.

We now give some statistical examples and motivations for model (1.1).

1.1. Logistic model

Model (1.1) can be viewed as a generalization of logistic regression (taking $r(x) = x$, see Qin, 1998). This kind of model has been widely used in statistical applications for the analysis of binary data (see e.g., Agresti, 1990; Hosmer and Lemeshow, 1999, 2000). One of the major reasons that 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 1, respectively.

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 (EL) approach (see Owen, 1988, 1990, 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. In this important case, the asymptotic results presented in Section 3 hold with some modifications.

1.2. 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

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

then, the ratio $\mathrm{d}Q/\mathrm{d}P$ takes the form

$$\frac{\mathrm{d}Q}{\mathrm{d}P}(x) = \exp\{\alpha + \beta x\} \quad \text{where } \alpha = \frac{\mu_1^2 - \mu_2^2}{2\sigma^2} \text{ and } \beta = \frac{\mu_2 - \mu_1}{\sigma^2}.$$

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 (2003) observed that there are cases in which the choice

$$\frac{\mathrm{d}Q}{\mathrm{d}P}(x) = \exp\{\alpha + \beta r(x)\},$$

where $r(\cdot)$ is an arbitrary but known function, is more appropriate. For example, if the two distributions are lognormal or gamma, $r(x) = \log(x)$ is the right choice. We underline that there are no normal cases in which we have $r(x) = x$ (for example, when we consider the ratio of two exponential densities), so this approach generalizes the classical normal-based one-way analysis of variance in the sense that it obviates the need for a completely specified parametric model, see Fokianos et al. (2001).

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" EL 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 nonparametric Wilcoxon rank-sum (see e.g., Randles and Wolfe, 1979; 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 EL ratio statistic (ELRS) is not well defined under the null hypothesis $\mathcal{H}_0 : Q = P$ (see Section 1.3 below). This problem has also been observed 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 ELRS, its "dual" form (see (2.7)) (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 (2.7) converges (to the nominal level) better than the observed level of the test proposed by Fokianos et al. (2001). Using ϕ-divergences and duality techniques, we give an interpretation for the statistic (2.7), the dual form of the ELRS; see (2.13). This interpretation allows us to give the asymptotic law of the proposed test statistic under the alternative hypothesis. We apply this result to give approximation of the power function of the test in a similar way to Morales and Pardo (2001) who gave some approximations of the 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 or 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

$$\phi(Q, P) := \int \varphi(\mathrm{d}Q/\mathrm{d}P)\,\mathrm{d}P, \tag{1.2}$$

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.

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 EL can be adjusted in the context of model (1.1). We next give a regularized version of the profile EL using duality techniques. A statistical test, for the null hypothesis $\mathcal{H}_0 : Q = P$, is then proposed. Another point of view at the test statistic is given using ϕ-divergences and duality techniques. 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 approximation to the power function which leads to 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. In the sequel, we sometimes write Pf instead of $\int f(x)\,\mathrm{d}P(x)$ for any function f and any measure P.

1.3. The profile 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 EL approach (see Owen, 1988, 1990, 2001), can be summarized as follows. For any $\theta \in \Theta$, the EL 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 by $(t_1, \dots, t_n)$ the combined sample $(X_1, \dots, X_{n_0}, Y_1, \dots, Y_{n_1})$, where $n := n_0 + n_1$. Hence, the log-likelihood can be written as

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

where $p_i := p(t_i)$. The profile log-likelihood (in θ) is then

$$l(\theta) := \sup_{p \in \mathcal{C}_\theta} l(\theta, p), \tag{1.3}$$

where p is constrained to the set

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

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

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

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 $\exp\{\alpha + \beta^{\mathrm{T}} r(x)\}$ 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

$$\text{there exists } p \in \mathcal{C}_\theta \text{ such that } |l(\theta, p)| < \infty. \tag{1.4}$$

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.,

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

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 (1.4) 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 $\theta = (\alpha, \beta^{\mathrm{T}})^{\mathrm{T}} \in \Theta$ 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 EL

If the assumption (1.4) holds, then $l(\theta)$ is finite, and the unique "optimal solution", say $\overline{p}$ (i.e., the value of p which yields the supremum in (1.3)), as an explicit expression of (1.3) can be derived by a Lagrange multiplier argument and the Kuhn–Tucker Theorem (see e.g., Rockafellar, 1970, Section 28). In fact, the "dual" problem associated to the "primal" problem (i.e., the optimization problem (1.3)) can be written as follows:

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

where $\overline{\log}(.)$ is the function defined on $\mathbb{R}$ by $\overline{\log}(x) = \log x$ if $x > 0$ and $\overline{\log}(x) = -\infty$ elsewhere. So, by the Kuhn–Tucker Theorem, under condition (1.4), the infimum in (2.1) is attained, and the following equality

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

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. Furthermore, $(\overline{\lambda}_0, \overline{\lambda}_1)$ and $\overline{p}$ satisfy

$$\overline{p}_i = \frac{1}{-\overline{\lambda}_0 - \overline{\lambda}_1 [m(\theta, t_i) - 1]}, \quad \forall i = 1, \ldots, n.$$

Hence, using the fact that $\overline{p}$ satisfies the constraints, 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 (1.4), the equality

$$l(\theta) := \sup_{p \in \mathcal{C}_\theta} l(\theta, p) = \inf_{\lambda \in \mathbb{R}} \left\{ -\sum_{i=1}^{n} \overline{\log}[n(1 + \lambda[m(\theta, t_i) - 1])] + \sum_{i=n_0+1}^{n} \log[m(\theta, t_i)] \right\} \tag{2.2}$$

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

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

where $\overline{\lambda}$ is the unique dual optimal solution in (2.2). Hence, the EL estimate $\tilde{\theta}$ of θ_T can be written as follows:

$$\tilde{\theta} = \arg \sup_{\theta \in \Theta} \inf_{\lambda \in \mathbb{R}} \left\{ -\sum_{i=1}^{n} \overline{\log}[n(1 + \lambda[m(\theta, t_i) - 1])] + \sum_{i=n_0+1}^{n} \log[m(\theta, t_i)] \right\}. \tag{2.3}$$

By differentiation with respect to α and λ, we can see by simple calculus that the Lagrange multiplier $\overline{\lambda}$ in (2.3) has the explicit solution $\overline{\lambda}(\tilde{\theta}) = n_1/n$ which does not depend on the data. Hence, the value of the log-likelihood (2.2) in $\tilde{\theta}$ is

$$l(\tilde{\theta}) = -n \log n - \sum_{i=1}^{n} \log\left(1 + \frac{n_1}{n}[m(\tilde{\theta}, t_i) - 1]\right) + \sum_{i=n_0+1}^{n} \log[m(\tilde{\theta}, t_i)], \tag{2.4}$$

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

$$\tilde{\theta} = \arg \sup_{\theta \in \Theta} \left\{ -n \log n - \sum_{i=1}^{n} \log\left(1 + \frac{n_1}{n}[m(\theta, t_i) - 1]\right) + \sum_{i=n_0+1}^{n} \log[m(\theta, t_i)] \right\}. \tag{2.5}$$

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 (2.4) and (2.5), we propose to consider, instead of $l(\theta)$, the dual form

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

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 $\tilde{\theta}$ as follows:

$$\widehat{\theta} := \arg \sup_{\theta \in \Theta} l_d(\theta). \tag{2.6}$$

Note that, under condition (1.4), we have $\widehat{\theta} = \tilde{\theta}$ and $l_d(\widehat{\theta}) = l(\tilde{\theta})$. Now, we give an interpretation to the dual form

$$S_n := 2W_d(\widehat{\theta}) := 2\left[\sup_{\theta \in \Theta} l_d(\theta) + n \log n\right] \tag{2.7}$$

of the EL ratio test statistic (ELRTS)

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

(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

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

with $\delta_x(\cdot)$ denoting the Dirac measure at point x, for all x. By simple calculus, we can show that the statistic (2.7) can be written as

$$S_n = 2a_n \sup_{\theta\in\theta}\left\{\int f_{\rho_n}(\theta, x)\,\mathrm{d}Q_{n_1}(x) - \int g_{\rho_n}(\theta, x)\,\mathrm{d}P_{n_0}(x)\right\}, \tag{2.8}$$

where

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

and

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

In (2.8), the sequence a_n is a normalizing term and the second term can be seen as an empirical estimate of

$$\sup_{\theta\in\Theta}\left\{\int f_\rho(\theta, x)\,\mathrm{d}Q(x) - \int g_\rho(\theta, x)\,\mathrm{d}P(x)\right\}, \tag{2.9}$$

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

$$f_\rho(\theta, x) := (1+\rho)\log[m(\theta, x)] - (1+\rho)\log[1+\rho m(\theta, x)] + (1+\rho)\log(1+\rho) \tag{2.10}$$

and

$$g_\rho(\theta, x) := \frac{1+\rho}{\rho}\log[1+\rho m(\theta, x)] - \frac{1+\rho}{\rho}\log(1+\rho). \tag{2.11}$$

On the other hand, using the so-called "dual representation of ϕ-divergences" (see Keziou, 2003, Theorem 2.1; Broniatowski and Keziou, 2006, Theorem 4.4) and choosing the class of functions

$$\mathcal{F} := \{x \mapsto \varphi_\rho^{\star\prime}(m(\theta, x)); \theta \in \Theta\},$$

we can prove the equality

$$\sup_{\theta\in\Theta}\left\{\int f_\rho(\theta, x)\,\mathrm{d}Q(x) - \int g_\rho(\theta, x)\,\mathrm{d}P(x)\right\} = \int \varphi_\rho^\star\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\right)\mathrm{d}P, \tag{2.12}$$

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

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

which is a member of the class of ϕ-divergences (1.2). We denote by $\phi^\star(Q, P)$ this divergence. In other words, by (2.8), (2.9) and (2.12), $a_n^{-1}W_d(\widehat{\theta})$ can be seen as an empirical estimate (which we denote by $\widehat{\phi}^\star(Q, P)$) of $\phi^\star(Q, P)$, the $\phi^\star$-divergence between Q and P, i.e., $\widehat{\phi}^\star(Q, P) := (2a_n)^{-1}S_n$. 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

$$S_n = 2a_n\widehat{\phi}^\star(Q, P) = 2a_n \sup_{\theta\in\theta}\left\{\int f_{\rho_n}(\theta, x)\,\mathrm{d}Q_{n_1}(x) - \int g_{\rho_n}(\theta, x)\,\mathrm{d}P_{n_0}(x)\right\}, \tag{2.13}$$

see (2.7) and (2.8), takes large values.

The estimate $\widehat{\theta}$ of θ_T (see (2.6)) can be written then as follows:

$$\widehat{\theta} = \arg\sup_{\theta\in\Theta}\left\{\int f_{\rho_n}(\theta,x)\,\mathrm{d}Q_{n_1}(x) - \int g_{\rho_n}(\theta,x)\,\mathrm{d}P_{n_0}(x)\right\}.$$

On the other hand, by Keziou (2003, Theorem 2.1) and Broniatowski and Keziou (2006, Theorem 4.4), we can prove that the supremum in (2.9) is unique and reached at $\theta=\theta_T$. This indicates that the estimate $\widehat{\theta}$ 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 $\widehat{\theta}$ (of the parameter θ_T) and the test statistic (2.13) both under the null and the alternative hypotheses. As an application, we obtain approximation of 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 $n\to\infty$. Denote also

$$l_{\phi^\star}(\theta) := a_n[Q_{n_1}f_{\rho_n}(\theta) - P_{n_0}g_{\rho_n}(\theta)].$$

In all the sequel, for simplicity, we write f and g instead of f_ρ and g_ρ defined in (2.10) and (2.11).

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 $\{x\mapsto(\partial^3/\partial\theta_i\partial\theta_j\partial\theta_k)f(\theta,x);\ \theta\in N(\theta_T)\}$ (resp. $\{x\mapsto(\partial^3/\partial\theta_i\partial\theta_j\partial\theta_k)g(\theta,x);\ \theta\in N(\theta_T)\}$) are dominated by some function Q-integrable (resp. some function P-integrable).

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

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

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

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

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

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

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

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

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

$$Qf'(\theta_T) - Pg'(\theta_T) = 0 \tag{3.3}$$

and

$$Qf''(\theta_T) - Pg''(\theta_T) = -P(m'(\theta_T)m'(\theta_T)^{\mathrm{T}}\varphi_\rho^{\star\,\prime\prime}(m(\theta_T))) =: -V_1(\theta_T). \tag{3.4}$$

Observe that the matrix $V_1(\theta_T)$ is symmetric and positive since the second derivative $\varphi_\rho^{\star\prime\prime}$ is nonnegative by the convexity of $\varphi_\rho^\star$. Let $U_n(\theta_T) := Q_{n_1} f'(\theta_T) - P_{n_0} g'(\theta_T)$ and use (3.3) and condition (A.2) in connection with the Central Limit Theorem to see that

$$\sqrt{n} U_n(\theta_T) \to \mathcal{N}(0, V_2(\theta_T)), \tag{3.5}$$

with $V_2(\theta_T) := \rho_1^{-1}[Q(f' f'^{\mathrm{T}}) - Qf' Qf'^{\mathrm{T}}] + \rho_0^{-1}[P(g' g'^{\mathrm{T}}) - Pg' Pg'^{\mathrm{T}}]$. Also, let $V_n(\theta_T) := Q_{n_1} f''(\theta_T) - P_{n_0} g''(\theta_T)$ and use (A.2) and (3.4) in connection with the Law of Large Numbers to conclude that

$$V_n(\theta_T) \to -V_1(\theta_T) \quad \text{(a.s.)}. \tag{3.6}$$

Now, for $\theta = \theta_T + un^{-1/3}$ with $|u| \leqslant 1$ 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.)

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

uniformly on u with $|u| \leqslant 1$. Now, use (3.6) and the fact that $U_n = \mathrm{O}(n^{-1/2}(\log\log n)^{1/2})$ (a.s.) to conclude that

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

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

$$l_{\phi^\star}(\theta) - l_{\phi^\star}(\theta_T) \leqslant \mathrm{O}(n^{1/6}(\log\log n)^{1/2}) - 2^{-1}\rho(1+\rho)^{-2} c n^{1/3} + \mathrm{O}(1) \quad \text{(a.s.)}, \tag{3.7}$$

where c is the smallest eigenvalue of the matrix V_1. Note that c is positive since the matrix V_1 defined in (3.4) is positive definite (it is symmetric, positive and nonsingular by assumption A.2). In view of (3.7), 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 $\widehat{\theta}$ in the interior of the ball $B(\theta_T, n^{-1/3})$, and therefore the estimate $\widehat{\theta}$ satisfies $l'_{\phi^\star}(\widehat{\theta}) = 0$ and $\widehat{\theta} - \theta_T = \mathrm{O}(n^{-1/3})$.

(b) Using the fact that $l'_{\phi^\star}(\widehat{\theta}) = 0$ and a Taylor expansion of $l'_{\phi^\star}(\widehat{\theta})$ around θ_T, we obtain

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

Hence,

$$\sqrt{n}(\widehat{\theta} - \theta_T) = -V_n^{-1}(\theta_T)\sqrt{n} U_n(\theta_T) + \mathrm{o}_p(1), \tag{3.8}$$

where U_n and V_n are defined as in the proof of part (a). Using (3.5) and (3.6), by application of Slutsky Theorem, we may conclude then $\sqrt{n}(\widehat{\theta} - \theta_T) \to \mathcal{N}(0, \mathrm{CM})$ where CM 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 (3.8) and using the convergence (3.6), we get

$$\widehat{\theta} = V_1^{-1}(0) U_n(0) + \mathrm{o}_p(n^{-1/2}), \tag{3.9}$$

where $V_1(0) = P[(1, r^{\mathrm{T}})^{\mathrm{T}}(1, r^{\mathrm{T}})]$, $U_n(0) = (0, W_n(0)^{\mathrm{T}})^{\mathrm{T}}$ 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 $2l_{\phi^\star}(\widehat{\theta})$ around $\theta_T = 0$, using the fact that $l_{\phi^\star}(0) = 0$, gives

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

Combine this with (3.9) to conclude that

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

It follows that $2l_{\phi^\star}(\widehat{\theta})$ 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. □

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 (2.10) and (2.11).

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

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

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

converges in distribution to a centered normal random variable with variance

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

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

$$\phi^\star(Q, P) = \int \varphi_\rho^\star(\mathrm{d}Q/\mathrm{d}P)\,\mathrm{d}P = \int \varphi_\rho^\star(m(\theta_T))\,\mathrm{d}P = Qf(\theta_T) - Pg(\theta_T) \tag{3.10}$$

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

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

Combine this with (3.10) to conclude that

$$\sqrt{a_n}[(2a_n)^{-1}S_n - \phi^\star(Q, P)] = \sqrt{a_n}[Q_{n_1}f(\theta_T) - Qf(\theta_T)] - \sqrt{a_n}[P_{n_0}g(\theta_T) - Pg(\theta_T)] + \mathrm{o}_p(1)$$

which converges in distribution to a centered normal variable with variance

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

Remark 3.1. Using Theorem 3.1 part (c), we propose to reject the null hypothesis $\mathcal{H}_0 : Q = P$ if $S_n > \chi^2_\varepsilon(d)$, where $\chi^2_\varepsilon(d)$ is the $(1-\varepsilon)$-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 approximation of the power function for a given alternative: for a given $\beta_T \neq 0$, we obtain for the power function $\theta_T \mapsto P_{\theta_T}\{S_n > \chi^2_\varepsilon(d)\}$ the following approximation:

$$P_{\theta_T}\{S_n > \chi^2_\varepsilon(d)\} \approx 1 - F_{\mathcal{N}}\left(\frac{\sqrt{a_n}}{\widehat{\sigma}(\theta_T)}[(2a_n)^{-1}\chi^2_\varepsilon(d) - H_n(\theta_T)]\right),$$

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

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

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 found 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 a 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_T; note that for any test, the power associated to the value of β_T 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.

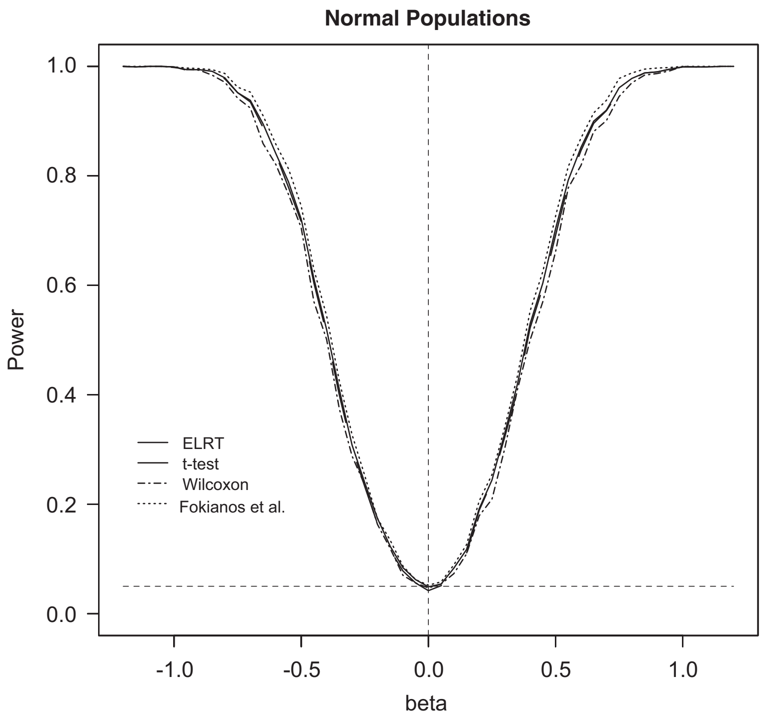

Fig. 1. Example 1a—two normal populations.

4.1. *Example 1—comparison of two populations*

We compare the power function of the ELRT, defined in (2.13), 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 EL 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 $X \sim \mathrm{Ga}(3+\beta, 1)$, $Y \sim \mathrm{Ga}(3, 1)$ 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 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 (2.13) is more convenient.

4.2. *Example 2—approximation of the power function*

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

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

or equivalently

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

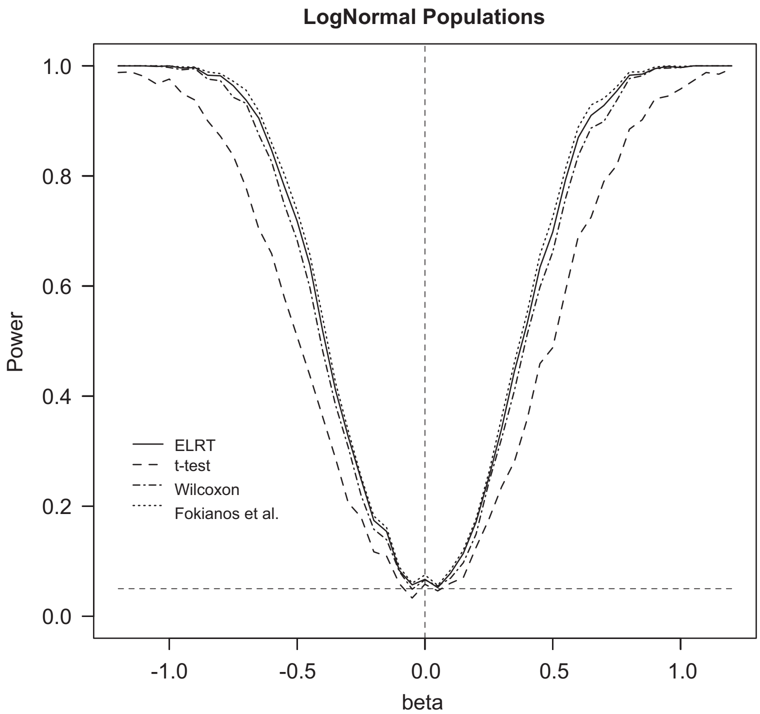

Fig. 2. Example 1b—two lognormal populations.

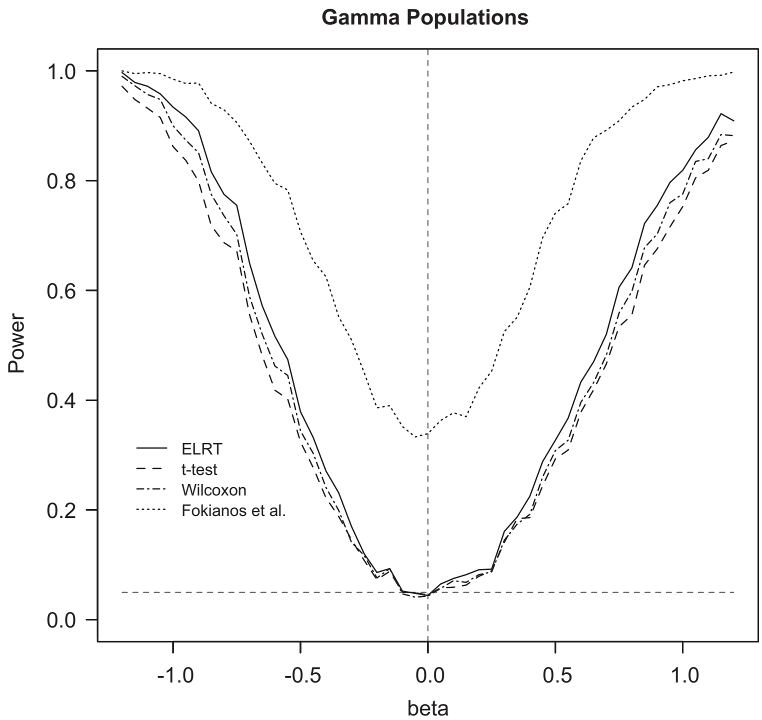

Fig. 3. Example 1c—two gamma populations.

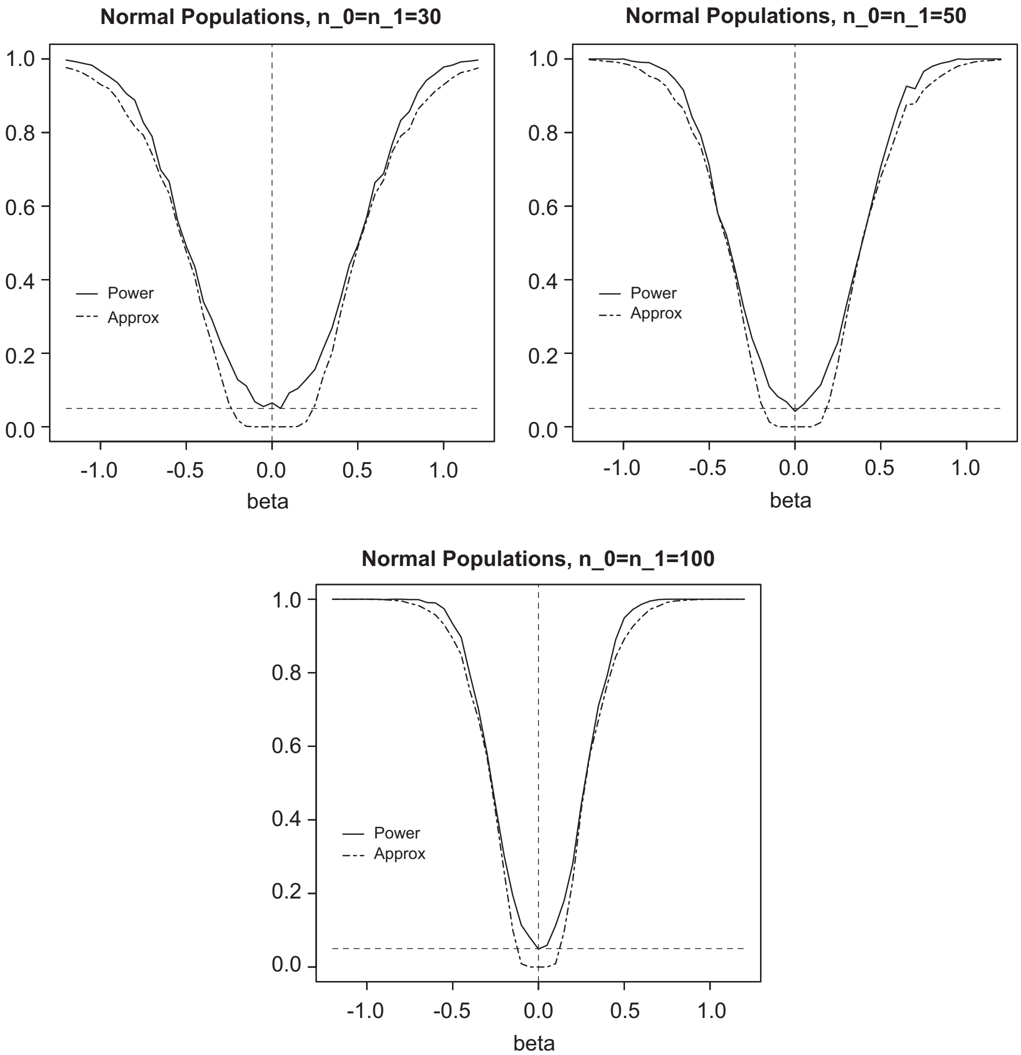

Fig. 4. Example 2—approximation of the power function.

based on the test statistic S_n (2.13). In this example, we consider $X \sim \mathcal{N}(\beta, 1)$ and $Y \sim \mathcal{N}(0, 1)$. We study numerically the accuracy of the power approximation given in Remark 3.1. We recall that the approximation of the power function $\theta_T \mapsto P_{\theta_T}(S_n \geqslant \chi^2_{0.05}(1))$ is

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

where $F_{\mathcal{N}}(.)$ is the cumulative distribution function of a standard normal variable,

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

and $H_n(\theta_T) := Q_{n_1}f(\theta_T) - P_{n_0}g(\theta_T)$. The power function 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_T. Each power entry was obtained from 1000 independent runs. The approximation (4.1) is plotted as a function of β_T by a dotted line; H_n and $\widehat{\sigma}$, 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 semiparametric 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 proved, 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 $\widehat{\theta}$ 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 hypothesis which we use to give approximation of the power function of the test.
- 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 terms of efficiency and robustness.

In the important case of paired samples, the asymptotic results presented in Section 3 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 of 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.

Acknowledgments

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

References

Agresti, A., 1990. Categorical data analysis. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley, New York (a Wiley-Interscience publication).
Broniatowski, M., 2003. Estimation of the Kullback–Leibler divergence. Math. Methods Statist. 12 (4), 391–409.
Broniatowski, M., Keziou, A., 2003. Parametric estimation and testing through divergences. Preprint 2004-1, L.S.T.A, Université Paris 6.
Broniatowski, M., Keziou, A., 2006. Minimization of ϕ-divergences on sets of signed measures. Studia Sci. Math. Hungar. 43 (4), 403–442.
Cox, T.F., Ferry, G., 1991. Robust logistic discrimination. Biometrika 78 (4), 841–849.
Fokianos, K., 2003. Box–Cox transformation for semiparametric comparison of two samples. In: Haitovsky, H.R.L.Y., Ritov, Y. (Eds.), Foundations of Statistical Inference. Physica, Heidelberg, pp. 131–140.
Fokianos, K., Kedem, B., Qin, J., Haferman, J., Short, D., 1998. On combining instruments. J. Appl. Meteorology 37.
Fokianos, K., Kedem, B., Qin, J., Short, D.A., 2001. A semiparametric approach to the one-way layout. Technometrics 43 (1), 56–65.
Gilbert, P.B., 2000. Large sample theory of maximum likelihood estimates in semiparametric biased sampling models. Ann. Statist. 28 (1), 151–194.
Gilbert, P.B., Lele, S., Vardi, Y., 1999. Maximum likelihood estimation in semiparametric selection bias models with application to aids vaccine trials. Biometrika 86 (1), 27–43.
Gill, R., Vardi, Y., Wellner, J., 1988. Large sample theory of empirical distributions in biased sampling models. Ann. Statist. 16 (3), 1069–1112.
Godambe, V.P., 1960. An optimum property of regular maximum likelihood estimation. Ann. Math. Statist. 31, 1208–1211.
Hollander, M., Wolfe, D.A., 1999. Nonparametric Statistical Methods. second ed. Wiley, New York.
Hosmer Jr., D.W., Lemeshow, S., 1999. Applied survival analysis. Wiley Series in Probability and Statistics: Texts and References Section. Wiley, New York. Regression modeling of time to event data. A Wiley-Interscience Publication.
Hosmer Jr., D.W., Lemeshow, S., 2000. Applied logistic regression. Wiley Series in Probability and Statistics: Texts and References Section. Wiley, New York.
Kay, R., 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. second ed. Wiley, New York.
Liese, F., Vajda, I., 1987. Convex Statistical Distances, vol. 95. BSB B.G. Teubner Verlagsgesellschaft, Leipzig.
Morales, D., Pardo, L., 2001. Some approximations to power functions of ϕ-divergences tests in parametric 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 & Hall, New York.
Qin, J., 1998. Inferences for case-control and semiparametric two-sample density ratio models. Biometrika 85 (3), 619–630.
Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Ann. Statist. 22 (1), 300–325.
Qin, J., Berwick, M., Ashbolt, R., Dwyer, T., 2002. Quantifying the change of melanoma incidence by breslow thickness. Biometrics 58 (3), 665–670.
Randles, R.H., Wolfe, D.A., 1979. Introduction to the theory of nonparametric statistics. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, Chichester, Brisbane.
Rockafellar, R.T., 1970. Convex Analysis. Princeton University Press, Princeton, NJ.
Vardi, Y., 1982. Nonparametric estimation in the presence of length bias. Ann. Statist. 10 (2), 616–620.
Vardi, Y., 1985. Empirical distributions in selection bias models. Ann. Statist. 13 (1), 178–203.
Zou, F., Fine, J.P., Yandell, B.S., 2002. On empirical likelihood for a semiparametric mixture model. Biometrika 89 (1), 61–75.