Semiparametric two-component mixture model with a known component: an asymptotically normal estimator

Laurent Bordes and Pierre Vandekerkhove

Laboratoire de Mathématiques Appliquées

UMR CNRS 5142

Université de Pau et des Pays de l'Adour

France

E-mail: laurent.bordes@univ-pau.fr

Laboratoire d'Analyse et de Mathématiques

Appliquées - UMR CNRS 8050

Université de Marne-la-Vallée

France

E-mail: pierre.vandek@univ-mlv.fr

Phone: +33 (0) 1 60 95 75 25

Fax: +33 (0) 1 60 95 75 45

Abstract: In this paper we consider a two-component mixture model one component of which has a known distribution while the other is only known to be symmetric. The mixture proportion is also an unknown parameter of the model. This mixture model class has proved to be useful to analyze gene expression data coming from microarray analysis. In this paper is proposed a general estimation method leading to a joint central limit result for all the estimators. Applications to basic testing problems related to this class of models are proposed, and the corresponding inference procedures are illustrated through some simulation studies.

AMS 2000 subject classifications: Primary 62G05, 62G20; secondary 62E10..

Keywords and phrases: Semiparametric, two-component mixture model, functional delta method, asymptotic normality, chi-square tests..

1. Introduction

Let us consider n independent and identically distributed random variables $X_1, \ldots, X_n$ coming from the two-component mixture model with density function (df) g defined by

$$g(x) = (1-p)f_0(x) + pf(x-\mu), \quad \forall x \in \mathbb{R}, \tag{1.1}$$

where f_0 is a known df and where the unknown parameters are the mixture proportion $p \in (0,1)$, the non-null location parameter μ and the df $f \in \mathcal{F}$ (the set of even df). This class of models extends classical two-component mixture models in the sense that one component is supposed to be symmetric only, without assuming that it belongs to a known parametric family. In the parametric setup this model is sometimes referred as contamination model (see Naylor and Smith, 1983, for application to chemistry, see Pal and Sengupta, 2000, for application to reliability, see also McLachlan and Peel, 2000, for further applications). This class of models is especially suitable for gene expression data coming from microarray analysis. An application to two bovine gestation mode comparison is performed in Bordes, Delmas and Vandekerkhove (2006). Connexions between model (1.1) and false discovery rate is also extensively discussed in Efron (2007). In Robin, Bar-Hen, Daudin and Pierre (2007) a convergent algorithm is proposed for model (1.1) without assuming that the nonparametric component is symmetric.

Nonparametric estimation of finite mixture models with learning data had been extensively studied in the eighties (see e.g. Hall and Titterington, 1985, Titterington, Smith and Makov, 1985). Because finite mixture models were reputed nonparametrically nonidentifiable very few authors tried to work on nonparametric finite mixture model without learning data. It is worth to point out the work of Hettmansperger and Thomas (2000) and later Cruz-Medina and Hettmansperger (2004), that considered ways to estimate the mixing proportions in a finite mixture distribution without making parametric assumptions about the component distributions. Note that other types of semiparametric mixture models are also discussed in Lindsay and Lesperance (1995).

Recently new classes of semiparametric mixture models has been considered. Qin (1999) investigates a real-valued two-component mixture model for which the log-likelihood ratio of the unknown components is an affine function. Hall and Zhou (2003) consider a $\mathbb{R}^p$-valued two-component mixture model for which the component distributions have independent components. This model is extended into a more general model in Hall, Neeman, Pakyari and Elmore (2005). Bordes, Mottelet and Vandekerkhove (2006) and Hunter, Wang and Hettmansperger (2007) consider real-valued finite mixture models the components of which are symmetric and equal up to a shift parameter. In Bordes, Delmas and Vandekerkhove (2006) model (1.1) is under consideration but the estimation method based on minimum L^q-distance estimation was shown to be numerically instable and did not allow to derive a central limit theorem. For all these models one of the crucial issue is to derive the identifiability of the model parameters when there is no learning data. For the above models estimation methods are generally linked to the model structure (invertible nonlinear system in Hall, Neeman, Pakyari and Elmore, 2005; symmetry of the unknown component distribution in Bordes, Mottelet and Vandekerkhove, 2006, Bordes, Delmas and Vandekerkhove 2006, and Hunter, Wang and Hettmansperger, 2007) but general stochastic EM-algorithm such as the one developed in Bordes, Chauveau and Vandekerkhove (2007) can be adapted to estimate all the above mentioned semi- or non-parametric mixture models. However obtaining the asymptotic behavior of these estimators remains an open question.

There are very few results concerning central limit theory for the above mentioned semiparametric mixture models. In Bordes, Mottelet, and Vandekerkhove (2006), the authors only prove that their estimators are $n^{-1/4+\alpha}$ a.s. consistent for all $\alpha > 0$, whereas in Hunter, Wang and Hettmansperger (2007) the authors prove under abstract technical conditions, that the Euclidean part of their estimators is asymptotically normally distributed. In Hall and Zhou (2003) and Hall, Neeman, Pakyari and Elmore (2005), the rate $O_P(n^{-1/2})$ is obtained for the whole parameters but none of the above papers propose a joint central limit theory for the whole parameters with consistent estimators of the asymptotic covariance function. Recently in Sugakova (2009), Maiboroda and Sugakova (2009), the authors consider a generalized estimating equations (GEE) method to estimate the Euclidean parameters of model (1.1). They prove, under mild conditions adapted to the GEE approach, the consistency and asymptotic normality of their estimators. The later is the main goal of this paper. In addition we stress that these results are certainly the preamble to omnibus tests construction in order to check that one model component belongs to a parametric family. The paper is organized in the following way. Section 2 is devoted to the estimation method whereas in Section 3 are gathered the large sample results and the estimators of various asymptotic expressions. In Section 4 we apply our large sample results to simple hypothesis testing which is illustrated by a Monte Carlo study.

2. Estimation method

Suppose that we observe n independent and identically distributed random variables $X_1, \ldots, X_n$ with cumulative distribution function (cdf) G defined by model (1.1), that is

$$G(x) = (1-p)F_0(x) + pF(x-\mu), \quad \forall x \in \mathbb{R}, \tag{2.1}$$

where G, F_0 and F are cumulative distribution functions (cdf) corresponding to df g, f_0 and f respectively. Let us denote by ϑ the Euclidean part (p, μ) of the model parameters taking values in Θ. We say that the parameters of model (1.1) are semiparametrically identifiable on $\Theta \times \mathcal{F}$ if for $\vartheta = (p, \mu) \in \Theta$, $\vartheta' = (p', \mu') \in \Theta$ and $(f, f') \in \mathcal{F}^2$

$$(1-p)f_0(\cdot) + pf(\cdot - \mu) = (1-p')f_0(\cdot) + p'f'(\cdot - \mu') \quad \lambda - \text{a.e.},$$

we have $\vartheta = \vartheta'$ and $f = f'$ λ-a.e. on $\mathbb{R}$ where λ is the Lebesgue measure on $\mathbb{R}$.

Assume that model (1.1) is identifiable, then we have

$$F(x) = \frac{1}{p}\left(G(x+\mu) - (1-p)F_0(x+\mu)\right), \quad \forall x \in \mathbb{R}. \tag{2.2}$$

Because F is the cdf of a symmetric distribution with respect to 0, we have $F(x) = 1 - F(-x)$, for all $x \in \mathbb{R}$. We denote by $\vartheta_0 = (p_0, \mu_0)$ the true value of the unknown parameter ϑ. Let us introduce, for all $x \in \mathbb{R}$, the functions

$$H_1(x; \vartheta, G) = \frac{1}{p}G(x+\mu) - \frac{1-p}{p}F_0(x+\mu),$$

and

$$H_2(x; \vartheta, G) = 1 - \frac{1}{p}G(-x+\mu) + \frac{1-p}{p}F_0(-x+\mu).$$

We have, using (2.2) and the symmetry of F,

$$H(x; \vartheta_0, G) \equiv H_1(x; \vartheta_0, G) - H_2(x; \vartheta_0, G) = 0 \quad \forall x \in \mathbb{R}, \tag{2.3}$$

whereas we can expect that for all $\vartheta \neq \vartheta_0$ an *ad hoc* norm of the function H will be strictly positive. In Bordes, Mottelet and Vandekerkhove (2006) the authors considered the $L^2_G(\mathbb{R})$-norm that proved to be interesting from both theoretical and numerical point of view. Considering such a norm leads to the following function d on Θ:

$$d(\vartheta) = \int_{\mathbb{R}} H^2(x; \vartheta, G)dG(x),$$

where obviously $d(\vartheta) \geq 0$ for all $\vartheta \in \Theta$ and $d(\vartheta_0) = 0$. Because G is unknown it is natural to replace it by its empirical version $\hat{G}_n$ obtained from the n-sample. However, because we aim to estimate ϑ by the minimum argument of the empirical version of d using a differentiable optimization routine, we need to replace G in H by a regular version $\tilde{G}_n$ of $\hat{G}_n$. Therefore we obtain an emprical version d_n of d defined by

$$d_n(\vartheta) = \int_{\mathbb{R}} H^2(x; \vartheta, \tilde{G}_n)d\hat{G}_n(x) = \frac{1}{n}\sum_{i=1}^{n} H^2(X_i; \vartheta, \tilde{G}_n) \tag{2.4}$$

where

$$\hat{G}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \le x}, \quad \forall x \in \mathbb{R},$$

and $\tilde{G}_n(x) = \int_{-\infty}^{x} \hat{g}_n(t)dt$ denotes the smoothed version of the empirical cdf $\hat{G}_n$ since $\hat{g}_n$ is a kernel density estimator of g defined by

$$\hat{g}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} q\left(\frac{x - X_i}{h_n}\right), \quad \forall x \in \mathbb{R},$$

with $h_n \to 0$, $nh_n \to +\infty$ and q is a symmetric kernel density function. For example we can choose $q(x) = (1-|x|)\mathbf{1}_{-1\le x\le 1}$. Note that additional conditions on the bandwidth h_n and the kernel function q will be specified afterward. Finally we propose to estimate ϑ_0 by

$$\hat{\vartheta}_n = (\hat{p}_n, \hat{\mu}_n) = \arg\min_{\vartheta\in\Theta} d_n(\vartheta).$$

Using the relation (2.2) we can estimate F by:

$$\hat{F}_n(x) = \frac{1}{\hat{p}_n}\left(\hat{G}_n(x + \hat{\mu}_n) - (1 - \hat{p}_n)F_0(x + \hat{\mu}_n)\right), \quad \forall x \in \mathbb{R}.$$

Note that generally $\hat{F}_n$ is not a legitimate cdf since it is generally not nondecreasing on the whole real line. However, the Glivenko-Cantelli strong consistency result obtained in Section 3 shows that it is not a serious drawback whenever the sample size is large enough.

Again by formula (2.2) a natural estimator of the df f is defined by

$$\tilde{f}_n(x) = \frac{1}{\hat{p}_n}\left(\hat{g}_n(x + \hat{\mu}_n) - (1 - \hat{p}_n)f_0(x + \hat{\mu}_n)\right), \quad \forall x \in \mathbb{R}.$$

Because generally $\tilde{f}_n$ will not be a density, it can be modified into a legitimate df estimator $\hat{f}_n$ defined by

$$\hat{f}_n = \frac{1}{s_n}\tilde{f}_n \mathbf{1}_{\tilde{f}_n \ge 0},$$

where $s_n = \int_{\mathbb{R}} \tilde{f}_n(x)\mathbf{1}_{\tilde{f}_n(x)\ge 0}dx$.

The advantage of choosing the above $L^2_G(\mathbb{R})$-norm is that it leads to an explicit empirical function d_n since replacing G by $\hat{G}_n$ transforms the integral sign into a simple sum. However other choices are possible. In Bordes, Delmas and Vandekerkhove (2006), $L^q(\mathbb{R})$ $(1 \leq q < +\infty)$ distances between H_1 and H_2 are discussed. These authors show also that it is possible to reduce the Euclidean parameter to one of the two parameters p and μ using the first order moment of G that can be estimated directly from the n-sample. However such a reduction of parameters can lead to serious numerical instability when the product $p\mu$ is small, which is frequent, e.g., for applications to microarray experiments.

3. Identifiability, consistency and asymptotic normality

3.1. General conditions and identifiability

In this section we give a set of conditions for which we obtain identifiability of the model parameters, consistency and asymptotic normality of our estimators. We denote by $\vartheta_0 = (p_0, \mu_0)$ the true value the unknown Euclidean parameter $\vartheta = (p, \mu)$ of model (1.1). Let us denote by m_0 and m the second-order moments of f_0 and f respectively. We introduce the set

$$\Phi = \mathbb{R}^* \times]0, +\infty[\setminus \cup_{k \in \mathbb{N}^*} \Phi_k$$

where

$$\Phi_k = \left\{ (\mu, m) \in \mathbb{R}^* \times]0, +\infty[; m = m_0 + \mu^2 \frac{k \pm 2}{3k} \right\}.$$

Let us define $\mathcal{F}_q = \{f \in \mathcal{F}; \int_{\mathbb{R}} |x|^q f(x)dx < +\infty\}$ for $q \geq 1$. Denoting by $\bar{f}_0$ the Fourier transform of the df f_0 we consider one assumption, for which the semi-parametric identifiability of the model (1.1) parameters is obtained, see Bordes, Delmas and Vandekerkhove (2006, Proposition 2, p. 736).

Identifiability condition (I). Let $(f_0, f) \in \mathcal{F}_3^2$, $\bar{f}_0 > 0$ and $(\mu_0, m) \in \Phi_c$ where Φ_c a compact subset of Φ. We have $\vartheta_0 = (p_0, \mu_0) \in \Theta$ where Θ is a compact subset of $(0, 1) \times \Xi$ where $\Xi = \{\mu; (\mu, m) \in \Phi_c\}$.

Comments and remarks. Note that condition (I) is fulfilled if f_0 is the density function of a centered gaussian (not necessarily normalized).

As it is mentioned in Bordes, Delmas and Vandekerkhove (2006), there exists various non-identifiability cases for model (1.1). Let us focus our attention on the following one:

$$(1-p)\varphi(x) + pf(x-\mu) = (1-\frac{p}{2})\varphi(x) + \frac{p}{2}\varphi(x-2\mu), \quad \forall x \in \mathbb{R}, \quad (3.1)$$

where a is a real number, φ is an even df, $p \in (0,1)$ and $f(x) = (\varphi(x-a) + \varphi(x+a))/2$. This example is particularly interesting since it clearly shows the danger of estimating model (1.1) when the df of the unknown component has exactly the same shape as the known df.

An other very important point which needs to be explained is the fact that the value $\mu = 0$ must be rejected from the parametric space. In fact it is easy to check that for all $p \in (0,1)$, $\vartheta = (p,0)$ is always a solution of $d(\vartheta) = 0$. Indeed for all $p \in (0,1)$ and all cdf G satisfying (2.1), we have

$$\forall x \in \mathbb{R}, \quad H_1(x;(p,0),G) = F(x), \quad H_2(x;(p,0),G) = 1 - F(-x) = F(x).$$

It follows that $\mu = 0$ does not match the contrast property of d. Even if $\mu = 0$ is a forbidden value by Condition (I), when the true value of μ is close to zero the empirical contrast may fail in finding a good estimate of the location parameter. Note however that in microarray experiments a value of μ that is close to 0 may be balanced by a large sample size. The same remark holds for $p = 0$.

Kernel conditions (K).

(i) The even kernel density function q is bounded, uniformly continuous, square integrable, of bounded variations and has second order moment.

(ii) The function q has first order derivative $q' \in L^1(\mathbb{R})$ and $q'(x) \to 0$ as $|x| \to +\infty$. In addition if γ is the square root of the continuity modulus of q, we have

$$\int_0^1 (\log(1/u))^{1/2}\, d\gamma(u) < \infty.$$

Comments. More general conditions on the kernel function q may be founded, e.g., in Silverman (1978) and in Giné and Guillou (2002). From a practical point of view triangular, gaussian, cauchy or many other standard kernels satisfy these conditions.

Bandwidth conditions (B).

(i) $h_n \searrow 0$, $nh_n \to +\infty$ and $\sqrt{n}h_n^2 = o(1)$,

(ii) $nh_n/|\log h_n| \to +\infty$, $|\log h_n|/\log\log n \to +\infty$ and there exists a real number c such that $h_n \leq ch_{2n}$ for all $n \geq 1$,

(iii) $|\log h_n|/(nh_n^3) \to 0$.

Comments. The two first conditions in (B) (i) are necessary to obtain the pointwize consistency of g kernel estimators. The third condition allows to control the distance between the empirical cdf $\hat{G}_n$ and its regularized version $\tilde{G}_n$. By using Corollary 1 in Shorack and Wellner (1986, p. 766) we obtain

$$\left\|\tilde{G}_n - \hat{G}_n\right\|_\infty = O_{a.s.}(h_n^2) \tag{3.2}$$

which by (i) and the law of iterated logarithm, leads to

$$\left\|\tilde{G}_n - G\right\|_\infty = O_{a.s.}\left(\left(\frac{\log\log n}{n}\right)^{-1/2}\right). \tag{3.3}$$

Conditions (ii) and (iii) allows to obtain the following result.

Lemma 3.1. *Suppose that the kernel function q satisfies Conditions (K) and that the bandwidth (h_n) satisfies Conditions (B).*

(i) If g and g' are uniformly continuous on $\mathbb{R}$, then

$$\|\hat{g}_n - g\|_\infty = o_{a.s.}(1) \quad \textit{and} \quad \|\hat{g}_n' - g'\|_\infty = o_{a.s.}(1).$$

(ii) If g is Lipschitz on $\mathbb{R}$, then

$$\|\hat{g}_n - g\|_\infty = O_{a.s.}\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/2}\right) + O(h_n).$$

Proof. Result (i) is given in Silverman (1978, Theorems A and C). For (ii) we have

$$\|\hat{g}_n - g\|_\infty \leq \|\hat{g}_n - E(\hat{g}_n)\|_\infty + \|E(\hat{g}_n) - g\|_\infty.$$

From Giné and Guillou (2002) we have

$$\|\hat{g}_n - E(\hat{g}_n)\|_\infty = O_{a.s.}\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/2}\right),$$

whereas the $O(h_n)$ term holds because g is Lipschitz on $\mathbb{R}$. □

Remark 3.1. *The convergence rate given by Giné and Guillou (2002) for the multivariate case was given in Silverman (1978) under stronger conditions on the bandwidth. Note that we put conditions in order to obtain simultaneously (i) and (ii) of Lemma 3.1. However each of the two results do not require all the assumptions made in (K) and (B) to hold. Finally it is worth to note that for example the bandwidth rate $n^{-1/4-\delta}$, with $\delta \in (0, 1/8)$, is convenient since it meets all the conditions that are given in (B).*

3.2. Consistency and preliminary convergence rate

We denote for simplicity by $\dot{h}(\vartheta)$ and $\ddot{h}(\vartheta)$ the gradient vector and hessian matrix of any real function h (when it makes sense) with respect to argument $\vartheta \in \mathbb{R}^2$.

Lemma 3.2. *Assume that Condition (I) is satisfied and that Θ is a compact subset of $(0, 1) \times \Phi_c$.*

(i) The function d is continuous on Θ.

(ii) If G is strictly increasing then d is a contrast function, i.e. for all $\vartheta \in \Phi_c$, $d(\vartheta) \geq 0$ and $d(\vartheta) = 0$ if and only if $\vartheta = \vartheta_0$.

(iii) *If* F_0 *and* F *are Lipschitz on* $\mathbb{R}$, *then* d *is Lipschitz on* Θ *and* $\sup_{\vartheta\in\Phi_c}|d_n(\vartheta)-d(\vartheta)| = o_{a.s.}(n^{-1/2+\alpha})$, *for all* $\alpha > 0$.

(iv) *If supp*$(g) = \mathbb{R}$ *then*

$$\ddot{d}(\vartheta_0) = 2\int_{\mathbb{R}} \dot{H}(x;\vartheta_0,G)\dot{H}^T(x;\vartheta_0,G)dG(x) > 0.$$

Proof. Let us show (i). The function $\vartheta \mapsto H(x;\vartheta,G)$ being bounded and continuous at any point $\vartheta \in \Theta$ for all fixed $x \in \mathbb{R}$, the wanted result is a direct consequence of the Lebesgue dominated convergence Theorem.

Let us show (ii). If $\vartheta = \vartheta_0$ then $d(\vartheta) = 0$. To prove the reciprocal let us remark that $d(\vartheta) = 0$ implies $H_1(\cdot;\vartheta,G) = H_2(\cdot;\vartheta,G)$ which leads, because G is strictly increasing on $\mathbb{R}$, to

$$g(x+\mu) - (1-p)f_0(x+\mu) = g(-x+\mu) - (1-p)f_0(-x+\mu), \quad \forall \; x \in \mathbb{R}.$$

Using (1.1) with $\vartheta = \vartheta_0$ we obtain for all $x \in \mathbb{R}$

$$\begin{aligned}&(1-p_0)f_0(x+\mu) + p_0(x+\mu-\mu_0) - (1-p)f_0(x+\mu)\\ =\;& (1-p_0)f_0(-x+\mu) + p_0(-x+\mu-\mu_0) - (1-p)f_0(-x+\mu).\end{aligned}$$

Assume now that $\vartheta \neq \vartheta_0$. Considering the Fourier transform of the above equality we obtain

$$(p-p_0)\sin(t\mu)\bar{f}_0(t) = p_0\sin(t(\mu-\mu_0))\bar{f}(t), \quad \forall \; t \in \mathbb{R}. \tag{3.4}$$

Since $\bar{f}_0(t) > 0$, it comes that $t(\mu-\mu_0) \in \pi\mathbb{Z} \Rightarrow t\mu \in \pi\mathbb{Z}$, which proves that it exists $k_0 \in \mathbb{Z}$ such that $|\mu|/|\mu-\mu_0| = k_0$. Taking in addition the first and third order derivatives of (3.4) at $t = 0$ we obtain

$$p_0\mu_0 = p\mu \quad \text{and} \quad (p-p_0)\mu(3m_0+\mu^2) + p_0(\mu-\mu_0)(3m+(\mu-\mu_0)^2) = 0. \tag{3.5}$$

The above results imply that

$$m = m_0 + \mu_0^2\frac{k_0 \pm 2}{3k_0}, \tag{3.6}$$

and then $(\mu, m) \in \cup_{k\in\mathbb{N}^*}\Phi_k$, which in turn implies that $(\mu, m) \notin \Phi_c$. It follows that $\vartheta = \vartheta_0$.

Let us show (iii). Let ϑ and ϑ' be two points in Θ. We have

$$\begin{aligned} & |d(\vartheta) - d(\vartheta')| \\ \leq & \int_{\mathbb{R}} |H(x;\vartheta,G) + H(x;\vartheta',G)| \times |H(x;\vartheta,G) - H(x;\vartheta',G)|\, dG(x) \\ \leq & \; c\int_{\mathbb{R}} |H(x;\vartheta,G) - H(x;\vartheta',G)|\, dG(x), \end{aligned} \tag{3.7}$$

where c is a constant coming from the boundedness of $(x, \vartheta) \mapsto H(x; \vartheta, G)$ on $\mathbb{R} \times \Theta$. Moreover, using the compacity of Θ and properties of cdf it is easy to show that there exist constants α, β and γ such that

$$\begin{aligned} & |H(x;\vartheta,G) - H(x;\vartheta',G)| \\ \leq & \; \alpha\left(|F(x-\mu) - F(x-\mu')| + |F(x+\mu) - F(x+\mu')|\right) \\ & +\beta\left(|F_0(x-\mu) - F_0(x-\mu')| + |F_0(x+\mu) - F_0(x+\mu')|\right) + \gamma|p-p'|. \end{aligned}$$

Using (3.7) and the above inequality we obtain that there exists a constant c' such that $|d(\vartheta) - d(\vartheta')| \leq c'\|\vartheta - \vartheta'\|_2$, where $\|\cdot\|_2$ denote the Euclidean norm on $\mathbb{R}^2$.

Considering for all $k \geq 0$ the random variable $Z_k(\vartheta) = H^2(X_k; \vartheta, G)$, we have for all $\vartheta \in \Theta$

$$\begin{aligned} |d_n(\vartheta) - d(\vartheta)| \;\; \leq \;\; & \left|\frac{1}{n}\sum_{k=1}^{n}\Big(H^2(X_k;\vartheta,\tilde{G}_n) - H^2(X_k;\vartheta,G)\Big)\right| \\ & + \left|\frac{1}{n}\sum_{k=1}^{n}(Z_k(\vartheta) - E(Z_k(\vartheta)))\right| \\ \leq \;\; & O_{a.s.}(||\tilde{G}_n - G||_\infty) + \sup_{\vartheta\in\Theta}\left|\frac{1}{n}\sum_{k=1}^{n}(Z_k(\vartheta) - E(Z_k(\vartheta)))\right|. \end{aligned}$$

Noticing that $||\tilde{G}_n - G||_\infty$ is $O_{a.s.}(\sqrt{n^{-1}\log\log n})$ (see Shorack and Wellner, 1986, p. 766) and that the last term of the right hand side is the supremum of an empirical process indexed by a class of Lipschitz bounded function, which is known to be a $o_{a.s.}(n^{-1/2+\alpha})$ for $\alpha > 0$ (see Bordes, Mottelet and Vandekerkhove, 2006, for the details), we get the wanted result.

Let us show (iv). First we have

$$\begin{aligned}\ddot{d}(\vartheta_0) &= 2\int_{\mathbb{R}} \left(\ddot{H}(x;\vartheta_0,G)H(x;\vartheta_0,G) + \dot{H}(x;\vartheta_0,G)\dot{H}^T(x;\vartheta_0,G)\right) dG(x) \\ &= 2\int_{\mathbb{R}} \dot{H}(x;\vartheta_0,G)\dot{H}^T(x;\vartheta_0,G)dG(x)\end{aligned}$$

because $H(\cdot;\vartheta_0,G) \equiv 0$ on $\mathbb{R}$. Let v be a vector in $\mathbb{R}^2$, we have

$$v^T\ddot{d}(\vartheta_0)v = 2\int_{\mathbb{R}} \left(v^T\dot{H}(x;\vartheta_0,G)\right)^2 dG(x) \geq 0.$$

It follows that $\ddot{d}(\vartheta_0)$ is a positive 2×2 real valued matrix. Let us show that it is also definite. Let $v \in \mathbb{R}^2$ be a non null column vector such that $v^T\ddot{d}(\vartheta_0)v = 0$, then $v^T\dot{H}(\cdot;\vartheta_0,G) = 0$ λ-everywhere on $\mathbb{R}$. It is straightforward to show that

$$\frac{\partial H}{\partial \mu}(x;\vartheta_0,G) = 2f(x), \tag{3.8}$$

and

$$\frac{\partial H}{\partial p}(x;\vartheta_0,G) = \frac{1}{p_0}\left[F_0(x+\mu_0) + F_0(\mu_0 - x) - 1\right]. \tag{3.9}$$

Therefore, using the above derivatives with the condition $v^T\dot{H}(\cdot;\vartheta_0,G) = 0$ we obtain the proportionality of the functions f and $F_0(\cdot+\mu_0) + F_0(\mu_0 - \cdot) - 1$. Because f_0 is an even function we obtain

$$f(x) = \frac{F_0(x+\mu_0) - F_0(x-\mu_0)}{\int_{\mathbb{R}}\left(F_0(x+\mu_0) - F_0(x-\mu_0)\right)dx} = \frac{1}{2\mu_0}\left(F_0(x+\mu_0) - F_0(x-\mu_0)\right).$$

Finally, computing the second order moment of f, we obtain by the integration by part formula

$$\begin{aligned}m = \int_{\mathbb{R}} x^2 f(x)ds &= \frac{1}{2\mu_0}\int_{\mathbb{R}} x^2\left(F_0(x+\mu_0) - F_0(x-\mu_0)\right)dx \\ &= \frac{2\mu_0 m_0}{2\mu_0} = m_0,\end{aligned}$$

which is not possible, by Condition (I), since m is different from $m_0+\mu_0^2(k\pm 2)/3k$ for all $k \in \mathbb{N}^*$ (hence from m_0 taking $k = 2$). □

Theorem 3.1.

(i) *Suppose that Condition (I) is satisfied,* Θ *is a compact subset of* $(0,1)\times\Phi_c$*,* G *is strictly increasing on* $\mathbb{R}$*, and that* F_0 *and* F *are Lipschitz on* $\mathbb{R}$*. Then the estimator* $\hat{\vartheta}_n$ *converges almost surely to* ϑ_0*.*

(ii) *If in addition* F_0 *and* F *are twice continuously differentiable with second derivatives in* $L^1(\mathbb{R})$*, then we have* $|\hat{\vartheta}_n - \vartheta_0| = o_{a.s.}(n^{-1/4+\alpha})$ *for all* $\alpha > 0$*.*

Proof. Let us show (i). This proof follows entirely the proof of i) in Theorem 1 given in Bordes *et al.* (2006b) by using (i)–(iii) of Lemma 3.2.

Let us show (ii). By Lemma 3.2 (iv) there exists $\alpha > 0$ such that for all $v \in \mathbb{R}^2$, $v^T \ddot{d}(\mu_0) v > \alpha \|v\|_2^2$. By a two order Taylor expansion of d at ϑ_0, we can find $\eta > 0$ such that for all v satisfying $\|v\|_2 < \eta$ and $\vartheta_0 + v \in \overset{\circ}{\Theta}$, we have

$$d(\vartheta_0 + v) \geq \frac{\alpha}{4}\|v\|_2^2. \tag{3.10}$$

Let us consider $B_0(\eta_n)$ the open ball centered at ϑ_0 with radius $\eta_n > 0$. Following the proof of Theorem 3.3 in Bordes, Mottelet and Vandekerkhove (2006) we show that for all $\vartheta \in \Theta \setminus B_0(\eta_n)$, we have the following events inclusion

$$\begin{aligned} &\limsup_n \left\{\hat{\vartheta}_n \notin B_0(\eta_n)\right\} \\ \subseteq\ &\limsup_n \left\{\inf_{\vartheta \in \Theta \setminus B_0(\eta_n)} d(\vartheta) < \gamma_n\right\} \cup \limsup_n \left\{\gamma_n \leq 2\sup_{\vartheta\in\Theta} |d_n(\vartheta) - d(\vartheta|\right\}. \end{aligned}$$

for any arbitrary sequence γ_n. Choosing now $\gamma_n = n^{-1/2+\alpha}$, and $\eta_n = n^{-1/4+\beta/2}$, with $0 < \alpha < \beta$ taken arbitrarily small, it follows from (3.10) and the uniform almost sure rate of convergence of d_n towards d given in Lemma 3.2 (iii), that

$$P\left(\limsup_n \left\{\inf_{\vartheta \in \Theta \setminus B_0(\eta_n)} d(\vartheta) < \gamma_n\right\}\right) = 0,$$

and

$$P\left(\limsup_n \left\{\gamma_n \leq 2\sup_{\vartheta\in\Theta} |d_n(\vartheta) - d(\vartheta)|\right\}\right) = 0.$$

In conclusion $\hat{\vartheta}_n$ converges almost surely towards ϑ_0 at rate $n^{-1/4+\delta}$, with $\delta > 0$ chosen arbitrarily small. $\square$

3.3. Asymptotic normality

For simplicity, we postpone in Appendix the definition of the 3×3 matrices L, J and Σ, and their empirical estimates $\hat{L}$, $\hat{J}$ and $\hat{\sigma}$, involved in Theorem 3.2 which establishes the joint asymptotic normality of $(\hat{\vartheta}_n, \hat{F}_n(\cdot))$.

Theorem 3.2. *Suppose that Conditions (I), (K) and (B) are satisfied, and that Θ is a compact subset of $(0,1) \times \Phi_c$, G is strictly increasing on $\mathbb{R}$, and that F_0 and F are twice continuously differentiable with second derivatives in $L^1(\mathbb{R})$. Then, denoting by $D(\mathbb{R})$ the space of cadlag function on $\mathbb{R}$,*

$$\sqrt{n}\left(\hat{\mu}_n - \mu_0, \hat{p}_n - p_0, \hat{F}_n(\cdot) - F(\cdot)\right)^T \rightsquigarrow \mathcal{G} \text{ in } \mathbb{R}^2 \times D(\mathbb{R}), \tag{3.11}$$

where $\mathcal{G} = (\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3)^T$ is a Gaussian process with correlation function

$$\Gamma(x, y; \vartheta_0) = L(x; \vartheta_0))J^{-1}(\vartheta_0)\Sigma(x, y)J^{-1}(\vartheta_0)(L(x; \vartheta_0))^T$$

such that for

$$\hat{\Gamma}(x, y) = \hat{L}(x)\hat{J}^{-1}\hat{\Sigma}(x, y)\hat{J}^{-1}(\hat{L}(y))^T,$$

we have

$$\sup_{(x,y)\in\mathbb{R}^2} \|\hat{\Gamma}(x, y) - \Gamma(x, y; \vartheta_0)\|_\infty \xrightarrow{a.s.} 0. \tag{3.12}$$

Proof. By a Taylor expansion of $\dot{d}_n$ around ϑ_0, we have

$$\ddot{d}_n(\vartheta_n^*)\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) = -\sqrt{n}\dot{d}_n(\vartheta_0), \tag{3.13}$$

where ϑ_n^* lies in the line segment with extremities $\hat{\vartheta}_n$ and ϑ_0.

Step 1. Let us prove that

$$\dot{d}_n(\vartheta_0) = \frac{2}{n}\sum_{i=1}^{n} H(X_i; \vartheta_0, \hat{G}_n)\dot{H}(X_i; \vartheta_0, G) + o_{a.s.}(n^{-1/2}). \tag{3.14}$$

We only investigate the partial derivative of $d_n(\theta_0)$ with respect to μ, the partial derivative with respect to p being easier to study. According to expression (2.4) we have

$$\frac{\partial d_n}{\partial \mu}(\vartheta_0) = \frac{2}{n}\sum_{i=1}^{n} H(X_i;\vartheta_0,\tilde{G}_n)\frac{\partial H}{\partial \mu}(X_i;\vartheta_0,\tilde{G}_n). \tag{3.15}$$

Let us consider now the following decomposition

$$\frac{\partial d_n}{\partial \mu}(\vartheta_0) - \frac{2}{n}\sum_{i=1}^{n} H(X_i;\vartheta_0,\hat{G}_n)\frac{\partial H}{\partial \mu}(X_i;\vartheta_0,G) = L_n^{(1)} + L_n^{(2)}$$

where

$$L_n^{(1)} = \frac{2}{n}\sum_{i=1}^{n} H(X_i;\vartheta_0,\tilde{G}_n)\left(\frac{\partial H}{\partial \mu}(X_i;\vartheta_0,\tilde{G}_n) - \frac{\partial H}{\partial \mu}(X_i;\vartheta_0,G)\right)$$

and

$$L_n^{(2)} = \frac{2}{n}\sum_{i=1}^{n} \frac{\partial H}{\partial \mu}(X_i;\vartheta_0,G)\left(H(X_i;\vartheta_0,\tilde{G}_n) - H(X_i;\vartheta_0,\hat{G}_n)\right),$$

with $\frac{\partial H}{\partial \mu}(x;\vartheta_0,G) = 2f(x)$ by (3.8).

Notice first that if f and f_0 are bounded on $\mathbb{R}$, then $x \mapsto \frac{\partial H}{\partial \mu}(x;\vartheta_0,G)$ is bounded on $\mathbb{R}$. Because $H(\cdot;\vartheta_0,G)=0$ we have

$$\begin{aligned}
&|L_n^{(1)}| \\
\leq\ & \frac{2}{n}\sum_{i=1}^{n}\left|H(X_i;\vartheta_0,\tilde{G}_n) - H(X_i;\vartheta_0,G)\right|\left|\frac{\partial H}{\partial \mu}(X_i;\vartheta_0,\tilde{G}_n) - \frac{\partial H}{\partial \mu}(X_i;\vartheta_0,G)\right| \\
\leq\ & 2\left\|H(\cdot;\vartheta_0,\tilde{G}_n) - H(\cdot;\vartheta_0,G)\right\|_\infty \times \left\|\frac{\partial H}{\partial \mu}(\cdot;\vartheta_0,\tilde{G}_n) - \frac{\partial H}{\partial \mu}(\cdot;\vartheta_0,G)\right\|_\infty \\
\leq\ & c\|\tilde{G}_n - G\|_\infty \times \|\tilde{g}_n - g\|_\infty = o_{a.s.}(n^{-1/2})
\end{aligned}$$

since c is a constant, $\|\tilde{G}_n - G\|_\infty = O_{a.s.}(\sqrt{n^{-1}\log\log n})$ by (3.2) and (3.3), and $\|\tilde{g}_n - g\|_\infty = O_{a.s.}((|\log h_n|/(nh_n))^{1/2}) + O(h_n)$ by Lemma 3.1 (ii).
Let us now consider the $L_n^{(2)}$ term. We have

$$\begin{aligned}
|L_n^{(2)}| &\leq \frac{2}{n}\sum_{i=1}^{n}\left|\frac{\partial H}{\partial \mu}(X_i;\vartheta_0,G)\right|\left|H(X_i;\vartheta_0,\tilde{G}_n) - H(X_i;\vartheta_0,\hat{G}_n)\right| \\
&\leq 2\left\|\frac{\partial H}{\partial \mu}(\cdot;\vartheta_0,G)\right\|_\infty \times \left\|H(\cdot;\vartheta_0,\tilde{G}_n) - H(\cdot;\vartheta_0,\hat{G}_n)\right\|_\infty \\
&\leq c\|\tilde{G}_n - \hat{G}_n\|_\infty = O_{a.s.}(h_n^2),
\end{aligned}$$

which in turn gives the wanted result because by Condition (B) we have $\sqrt{n}h_n^2 = o(1)$. This finishes the proof of the first step.

Step 2. We need to prove that

$$\ddot{d}_n(\vartheta_n^*) \xrightarrow{a.s.} \mathcal{I}(\vartheta_0), \tag{3.16}$$

where $\mathcal{I}(\vartheta_0) = \int_{\mathbb{R}} \dot{H}(x;\vartheta_0,G)\dot{H}^T(x;\vartheta_0,G)dG(x) > 0$.

In order to prove statement (3.16) let us remark that

$$\begin{aligned} \ddot{d}_n(\vartheta_n^*) &= \frac{2}{n}\sum_{k=1}^n H(X_k;\vartheta_n^*,\tilde{G}_n)\ddot{H}(X_k;\vartheta_n^*,\tilde{G}_n) \\ &\quad + \frac{2}{n}\sum_{k=1}^n \dot{H}(X_k;\vartheta_n^*,\tilde{G}_n)\dot{H}^T(X_k;\vartheta_n^*,\tilde{G}_n) \\ &= T_n^{(1)} + T_n^{(2)} + T_n^{(3)}, \end{aligned}$$

where

$$\begin{aligned} T_n^{(1)} &= \frac{2}{n}\sum_{k=1}^n \Big(H(X_k;\vartheta_n^*,\tilde{G}_n) - H(X_k;\vartheta_0,G)\Big)\ddot{H}(X_k;\vartheta_n^*,\tilde{G}_n), \\ T_n^{(2)} &= \frac{2}{n}\sum_{k=1}^n \dot{H}(X_k;\vartheta_n^*,\tilde{G}_n)\dot{H}^T(X_k;\vartheta_n^*,\tilde{G}_n) \\ &\quad - \frac{2}{n}\sum_{k=1}^n \dot{H}(X_k;\vartheta_0,G)\dot{H}^T(X_k;\vartheta_0,G), \\ T_n^{(3)} &= \frac{2}{n}\sum_{k=1}^n \dot{H}(X_k;\vartheta_0,G)\dot{H}^T(X_k;\vartheta_0,G). \end{aligned}$$

Because the df f and f_0 are bounded it easy to show that the strong law of large numbers holds for $T_n^{(3)}$ and therefore $T_n^{(3)} \to \mathcal{I}(\vartheta_0)$ almost surely. It remains to show that $T_n^{(1)}$ and $T_n^{(2)}$ converge almost surely to 0. For $T_n^{(1)}$ let us remark that

$$T_n^{(1)} \leq \Big\|H(\cdot;\vartheta_n^*,\tilde{G}_n) - H(\cdot;\vartheta_0,G)\Big\|_\infty \times \Big\|\ddot{H}(\cdot;\vartheta_n^*,\tilde{G}_n)\Big\|_\infty. \tag{3.17}$$

Because $\ddot{H} = \ddot{H}_1 - \ddot{H}_2$ and H_1 and H_2 are very similar, we only prove that the supremum norm of each term in $\ddot{H}_1(\cdot;\vartheta_n^*,\tilde{G}_n)$ matrix is bounded. We only handle the more complicated term in $\ddot{H}_1(\cdot;\vartheta_n^*,\tilde{G}_n)$ which is the second order derivative with respect to μ. We have

$$\frac{\partial^2 H_1}{\partial \mu^2}(x;\vartheta_n^*,\tilde{G}_n) = \frac{1}{\hat{p}_n}\tilde{g}_n'(x+\hat{\mu}_n) - \frac{1-\hat{p}_n}{\hat{p}_n}f_0'(x+\hat{\mu}_n),$$

where $\tilde{g}_n' = \tilde{G}_n''$ is an estimator of g'. Because f' and f_0' are bounded we have

$$\left\|\frac{\partial^2 H_1}{\partial \mu^2}(\cdot;\vartheta_n^*,\tilde{G}_n)\right\|_\infty \leq O_{a.s.}\left(\|\tilde{g}_n' - g'\|_\infty\right) + O_{a.s.}(1).$$

According to Silverman (1978), dealing with the uniform consistency of kernel estimators of a density and its derivatives, we have

$$\|\tilde{g}_n - g\|_\infty = o_{a.s.}(1) \quad \text{and} \quad \|\tilde{g}_n' - g'\|_\infty = o_{a.s.}(1).$$

Finally we obtain

$$\left\|\ddot{H}(\cdot;\vartheta_n^*,\tilde{G}_n)\right\|_\infty = o_{a.s.}(1) + O_{a.s.}(1) = O_{a.s.}(1).$$

Painfull but straitghforward calculations lead to

$$\left\|H(\cdot;\vartheta_n^*,\tilde{G}_n) - H(\cdot;\vartheta_0,G)\right\|_\infty \leq O_{a.s.}\left(\|\hat{\vartheta}_n - \vartheta_0\|_2\right) + O_{a.s.}\left(\|\tilde{G}_n - G\|_\infty\right).$$

According to Corollary 1 in Shorack and Wellner (1986, p. 766) and the Law of the Iterated Logarithm for the empirical cdf, we have

$$\|\tilde{G}_n - G\|_\infty = O_{a.s}(\sqrt{n^{-1}\log\log(n)}),$$

hence by (3.17) and the above results we have

$$T_n^{(1)} = O_{a.s.}\left(\|\vartheta_n^* - \vartheta_0\|_2 + \sqrt{n^{-1}\log\log(n)}\right). \tag{3.18}$$

Finally by Theorem 3.1 we have that $\|\vartheta_n^* - \vartheta_0\|_2 = o_{a.s.}(n^{-1/4+\alpha})$ for any $\alpha > 0$, then $T_n^{(1)}$ converges almost surely to 0.

The same kind of calculations allow to prove that $T_n^{(2)} \to 0$ almost surely, which concludes the proof of statement (3.16).

Step 3. Using the fact that at the $\sqrt{n}$-rate $\hat{G}_n$ and $\tilde{G}_n$ are interchangeable, and using the properties of F and F_0 we obtain the following uniform (with respect to $x \in \mathbb{R}$) approximation

$$\begin{aligned} & \sqrt{n}\left(\hat{F}_n(x) - F(x)\right) \qquad (3.19) \\ = \quad & \sqrt{n}(\hat{p}_n - p_0)\left(\frac{F_0(x+\mu_0) - G(x+\mu_0)}{p_0^2}\right) \\ & + \sqrt{n}\frac{1}{p_0}\left(\hat{G}_n(x+\mu_0) - G(x+\mu_0)\right) + \sqrt{n}(\hat{\mu}_n - \mu_0) f(x) + o_{a.s.}(1). \end{aligned}$$

Step 4. Let us prove (3.11). By (3.14), (3.16) and (3.19) we obtain

$$\sqrt{n}\begin{pmatrix} \hat{p}_n - p_0 \\ \hat{\mu}_n - \mu_0 \\ \hat{F}_n(\cdot) - F(\cdot) \end{pmatrix} = L(\cdot;\vartheta_0)J^{-1}(\vartheta_0)\sqrt{n}\begin{pmatrix} \mathcal{U}_1(\hat{G}_n) \\ \mathcal{U}_2(\hat{G}_n) \\ \mathcal{U}_3(\hat{G}_n) \end{pmatrix} + o_{a.s.}(1), \qquad (3.20)$$

where for a cdf V

$$\begin{aligned} \mathcal{U}_1(V) &= 2\int_{\mathbb{R}} H(x;\vartheta_0, V) h_1(x) dV(x) \\ \mathcal{U}_2(V) &= 2\int_{\mathbb{R}} H(x;\vartheta_0, V) h_2(x) dV(x) \\ \mathcal{U}_3(V) &= V(\cdot + \mu_0) - G(\cdot + \mu_0). \end{aligned}$$

Considering $\mathcal{U} = (\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3)^T$ as a function from $BV_1(\mathbb{R})$ to $\mathbb{R}^2 \times D(\mathbb{R})$ where $BV_1(\mathbb{R})$ is the space of functions with variations bounded by 1 on $\mathbb{R}$, it is easy to see that $\mathcal{U}$ is Hadamard differentiable (see e.g. van der Vaart, 1998) on $BV_1(\mathbb{R})$ with derivative

$$K \mapsto \begin{pmatrix} \frac{2}{p_0}\int_{\mathbb{R}} h_1(y)\left(K(\mu_0+y) + K(\mu_0-y)\right) dG(y) \\ \frac{2}{p_0}\int_{\mathbb{R}} h_2(y)\left(K(\mu_0+y) + K(\mu_0-y)\right) dG(y) \\ K(\cdot + \mu_0) \end{pmatrix}.$$

It follows by the δ-method theorem (see e.g. van der Vaart, 1998) that

$$\sqrt{n}\begin{pmatrix} \mathcal{U}_1(\hat{G}_n) \\ \mathcal{U}_2(\hat{G}_n) \\ \mathcal{U}_3(\hat{G}_n) \end{pmatrix} = \begin{pmatrix} \frac{2}{p_0}\int_{\mathbb{R}} h_1(y)\left(\mathbb{G}_n(\mu_0+y) + \mathbb{G}_n(\mu_0-y)\right) dG(y) \\ \frac{2}{p_0}\int_{\mathbb{R}} h_2(y)\left(\mathbb{G}_n(\mu_0+y) + \mathbb{G}_n(\mu_0-y)\right) dG(y) \\ \mathbb{G}_n(\cdot + \mu_0) \end{pmatrix} + o_P(1),$$

where $\mathbb{G}_n = \sqrt{n}(\hat{G}_n - G)$. By the Donsker theorem $\mathbb{G}_n \rightsquigarrow \mathcal{B}$ where $\mathcal{B}$ is a gaussian process with correlation function ρ. Then the following weak convergence holds in $\mathbb{R}^2 \times D(\mathbb{R})$

$$\sqrt{n}\begin{pmatrix} \mathcal{U}_1(\hat{G}_n) \\ \mathcal{U}_2(\hat{G}_n) \\ \mathcal{U}_3(\hat{G}_n) \end{pmatrix} \rightsquigarrow \mathcal{H} = \begin{pmatrix} \frac{2}{p_0}\int_{\mathbb{R}} h_1(y)\left(\mathcal{B}(\mu_0+y)+\mathcal{B}(\mu_0-y)\right)dG(y) \\ \frac{2}{p_0}\int_{\mathbb{R}} h_2(y)\left(\mathcal{B}(\mu_0+y)+\mathcal{B}(\mu_0-y)\right)dG(y) \\ \mathcal{B}(\cdot+\mu_0) \end{pmatrix},$$

where $\mathcal{H}$ is a gaussian process as a linear form on $\mathcal{B}$ the correlation function of which is defined by $\Sigma(x,y) = E\left[\mathcal{H}(x)\mathcal{H}^T(y)\right]$. This with (3.20) lead to (3.11).

Step 5. It remains to prove (3.12). For this purpose it is sufficient to prove the convergence in probability of $\hat{J}_{ij}$ to $J(\vartheta_0)$, that

$$\max_{i=1,2,3} \|\hat{h}_i - h_i(\cdot;\vartheta_0)\|_\infty \xrightarrow{a.s.} 0,$$

(which gives the strong uniform convergence of $\hat{L}$ to $L(\cdot;\vartheta_0)$) and that

$$\sup_{(x,y)\in\mathbb{R}^2} |\hat{\sigma}_{ij}(x,y) - \sigma_{ij}(x,y)| \xrightarrow{a.s.} 0,$$

for all $1 \le i,j \le 3$. Because the proof is quite repetitive, we only consider one of the more difficult terms, other terms can be handled in the same way. First

$$\sup_{x\in\mathbb{R}} |\hat{h}_1(x) - h_1(x)| \xrightarrow{a.s.} 0,$$

by the strong consistency result of Theorem 3.1 and the Lipschitz property of F_0. Using repetively the telescoping rule we show that

$$\begin{aligned} & \sup_{x\in\mathbb{R}} |\tilde{f}_n(x) - f(x)| \\ \le\ & \frac{1}{\hat{p}_n}\left(\|\hat{g}-g\|_\infty + \|g(\cdot+\hat{\mu}_n) - g(\cdot+\mu_0)\|_\infty + \|f_0(\cdot+\hat{\mu}_n) - f_0(\cdot+\mu_0)\|_\infty\right) \\ & + \frac{\|g\|_\infty + \|f_0\|_\infty}{p_0\hat{p}_n}|\hat{p}_n - p_0|. \end{aligned}$$

The above inequality with the strong consistency result of Theorem 3.1 and the Lipschitz properties of f_0 and g lead to

$$\sup_{x\in\mathbb{R}} |\tilde{f}_n(x) - f(x)| \xrightarrow{a.s.} 0.$$

Now recall that $s_n = \int_{\mathbb{R}} \tilde{f}_n(x)\mathbf{1}_{\tilde{f}_n(x)}dx \xrightarrow{a.s.} 1$ (see Bordes, Mottelet and Vandekerkhove, 2006), then we can write

$$|\tilde{f}_n(x) - \hat{f}_n(x)| \leq \|\tilde{f}_n\|_\infty \frac{|1-s_n|}{s_n} + |\tilde{f}_n(x)|\mathbf{1}_{\tilde{f}_n(x)<0},$$

where because f is bounded and $\|\tilde{f}_n - f\|_\infty \xrightarrow{a.s.} 0$, the right hand side of the above inequality converges strongly to 0. It follows that

$$\sup_{x\in\mathbb{R}} |\hat{h}_2(x) - h_2(x)| \xrightarrow{a.s.} 0.$$

By similar arguments we prove the uniform strong convergence of $\hat{\rho}$, $\hat{\ell}$ and $\hat{k}$ to ρ, ℓ and k respectively. Now let us show that $\hat{\sigma}_{12}$ converges strongly to σ_{12}. Let us consider, for example, the convergence of $\hat{\sigma}_{12}$. From previous results it is straightforward to obtain

$$\hat{\sigma}_{12} = \frac{8}{n(n-1)p_0^2} \sum_{i=2}^{n} \sum_{j=1}^{i-1} h_1(X_i)h_2(X_j)k(X_i, X_j) + o_{a.s.}(1),$$

where by the strong law of large numbers the right hand side term is a 2-order U-statistic, with kernel function $(u, v) \mapsto h(u, v) = h_1(u)h_2(v)k(u, v)$, converging strongly to $\sigma_{12} = 4E(h(X_1, X_2))/p_0^2$. □

4. Applications to testing and simulations

The aim of this section is not to develop some sophisticated testing procedures, the aim is rather to show how to apply our CLT in order to build, quite directly, some basic tests adapted to few hypotheses relative to the semiparametric mixture model (1.1). However, more general testing procedures could be developed. For example it should be possible to test the hypothesis that the nonparametric component belongs to a parametric family, but such a test is beyond the scope of the paper and will be developed elsewhere for a larger class of semiparametric mixture models. First we propose a chi-square test for a simple hypothesis for the three parameters of model (1.1) next the same type of test is proposed to check the hypothesis that the nonparametric component of the mixture model has a symmetric distribution.

4.1. Testing some simple hypothesis

We propose in this section to consider the following simple hypothesis testing problem:

$$\mathcal{H}_0 : (p, \mu, F) = (p_\star, \mu_\star, F_\star) \qquad \text{versus} \qquad \mathcal{H}_1 : (p, \mu, F) \neq (p_\star, \mu_\star, F_\star),$$

where $(p_\star, \mu_\star) \in \Theta$ and $F_\star$ is a known cdf function. Because under $\mathcal{H}_0$ the joint asymptotic behavior of $\sqrt{n}(\hat{p}_n - p_\star, \hat{\mu}_n - p_\star, \hat{F}_n - F_\star)$ is known, it is possible to base a testing procedure on the asymptotic distribution. Such a procedure requires to choose a discrepancy measure between the estimates and the estimators, and of course, several choices are possible. The test we propose is based on the frequently used chi-square measure, leading to chi-square type tests. Let us fix k real numbers $s_1 < \cdots < s_k$ such that $0 < F_\star(s_1) < \cdots < F_\star(s_k) < 1$. Because Moore (1971) and Ruymgaart (1975) proved that under smooth conditions chi-square statistics can also be constructed using random cells boundaries it is generally possible to choose equidistributed cells to increase the probability that there are enough data in each interval $]s_{\ell-1}, s_\ell]$ for $\ell = 1, \ldots, k$ with $s_0 = -\infty$.

We consider now the random vector W_n defined by

$$W_n = \sqrt{n} \begin{pmatrix} \hat{p}_n - p_\star \\ \hat{\mu}_n - \mu_\star \\ \hat{F}_n(s_1) - F_\star(s_1) \\ \vdots \\ \hat{F}_n(s_k) - F_\star(s_k) \end{pmatrix}.$$

According to Theorem 3.2 we have

$$W_n \rightsquigarrow \mathcal{N}(0, V),$$

where V is the $(k+2)\times(k+2)$ correlation matrix with entries defined by

$$v_{ij} = \Gamma_{ij}, \quad \text{for } 1 \leq i \leq j \leq 2,$$

$$v_{1j} = \Gamma_{13}(s_{j-2}) \text{ and } v_{2j} = \Gamma_{23}(s_{j-2}), \quad \text{for } 3 \leq j \leq k+2,$$

and

$$v_{ij} = \Gamma_{33}(s_{i-2}, s_{j-2}), \quad \text{for } 3 \leq i \leq j \leq k+2.$$

Defining $\hat{V}$ as V but using $\hat{\Gamma}$ instead of Γ (see the beginning of section 3), we obtain that

$$W_n^T \hat{V}^{-1} W_n \rightsquigarrow \chi^2_{k+2},$$

where χ^2_m is the chi-square distribution with m degrees of freedom. We reject $\mathcal{H}_0$ at the level $\alpha \in (0,1)$ if $W_n^T \hat{V}^{-1} W_n > \chi^2_{k+2,1-\alpha}$ where $\chi^2_{k+2,1-\alpha}$ is the quantile of order $1-\alpha$ of the χ^2_{k+2} distribution.

4.2. Testing the symmetry of the nonparametric component

Testing the symmetry of the nonparametric component is equivalent to test the hypothesis

$$\mathcal{H}_0 : F(x) + F(-x) = 1, \qquad \text{for all } x \in \mathbb{R}^+,$$

versus

$$\mathcal{H}_1 : \text{there exists } x \in \mathbb{R}^+ \text{ such that } F(x) + F(-x) \neq 1.$$

It is therefore necessary to compare on $\mathbb{R}^+$ the random maps $x \mapsto \hat{F}_n(x) + \hat{F}_n(-x)$ to $x \mapsto F(x)+F(-x)$ which is constant equal to 1 under $\mathcal{H}_0$. When $p = 1$ in model (1.1), there are several ways to test $\mathcal{H}_0$. Some test statistics are based on ranks (see e.g. Shorack and Wellner, 1986) while others are based on empirical cdf (see e.g. Schuster and Barker, 1987). Again, for simplicity, we choose a chi-square measure to test $\mathcal{H}_0$. Note that combination of maximum deviation measure with bootstrapped critical value proposed Schuster and Barker (1987) and studied by Arcones and Giné (1991) could certainly be used here. Let us consider k positive real numbers $0 < s_1 < \cdots < s_k$ satisfying $F(s_1) < \cdots < F(s_k)$ under $\mathcal{H}_0$. We consider the following discrepancy measure

$$Z_n = \sqrt{n} \begin{pmatrix} \hat{F}_n(-s_1) + \hat{F}_n(s_1) - 1 \\ \vdots \\ \hat{F}_n(-s_k) + \hat{F}_n(s_k) - 1 \end{pmatrix}.$$

Then, under $\mathcal{H}_0$ we have $Z_n = AY_n$ where A is the $k \times 2k$ matrix defined by $A = (I_k, I_k)$ with I_k is the identity matrix of order k, and Y_n is the $2k$-dimensional random vector defined by

$$Y_n = \sqrt{n} \begin{pmatrix} \hat{F}_n(s_1) - F(s_1) \\ \vdots \\ \hat{F}_n(s_k) - F(s_k) \\ \hat{F}_n(-s_1) - F(-s_1) \\ \vdots \\ \hat{F}_n(-s_k) - F(-s_k) \end{pmatrix}.$$

According to Theorem 3.2 we have

$$Y_n \rightsquigarrow \mathcal{N}(0, \Lambda),$$

where Λ is the $2k \times 2k$ correlation matrix with entries λ_{ij} defined by

$$\lambda_{ij} = \Gamma_{33}(s_i^*, s_j^*), \quad \text{for } 1 \leq i, j \leq 2k,$$

where $s_i^* = s_i$ for $1 \leq i \leq k$ and $s_i^* = -s_i$ for $k+1 \leq i \leq 2k$.

Defining $\hat{\Lambda}$ as Λ but using $\hat{\Gamma}$ instead of Γ, we obtain that

$$Z_n^T (A\hat{\Lambda}A^T)^{-1} Z_n \rightsquigarrow \chi_k^2.$$

We reject $\mathcal{H}_0$ at the level $\alpha \in (0,1)$ if $Z_n^T (A\hat{\Lambda}A^T)^{-1} Z_n > \chi^2_{k,1-\alpha}$.

4.3. Simulation study

Let Φ be the cdf of a $\mathcal{N}(0,1)$ distribution. In this section data are simulated from model (2.1) with $p = 0.7$, $\mu = 3$, $F_0 = \Phi$ and $F = \Phi(\cdot/0.5)$. We performed

Table 1
Mean (Stand. Dev.) of 200 estimates of p, μ and $F(0.5)$.

Sample size	$p = 0.7$	$\mu = 3$	$F(0.5) = 0.8413$
100	0.7106 (0.0498)	2.9912 (0.0757)	0.8415 (0.0378)
400	0.7048 (0.0277)	2.9959 (0.0355)	0.8390 (0.0177)
1000	0.7018 (0.0167)	2.9977 (0.0225)	0.8409 (0.0107)

the contrast (2.4) minimization using the `optim` function (quasi-Newton BFGS method) of `R` software. We used the triangular kernel q defined by $q(x) = (1 - |x|)\mathbf{1}_{|x|<1}$ and the bandwidth is provided by the function `density` of `R` software. Table 1 shows the good behavior of our estimators even for moderate sample size. The standard deviations, within parentheses, are computed from the 200 estimates of each parameter and are quite small.

Another important question is the quality of the asymptotic variance estimators. Let us recall that these estimators involve both U-statistics and estimation of the density f. It is therefore important to check that these estimators have sufficiently good properties to make our central limit theorem useful in practice. Figure 1 shows that these good properties are satisfied even when the sample size is moderate. This is especially true for the functional parameter F the estimator of which has very good behavior even if its variance estimation requires to estimate the unknown density function f.

With the same data as those we used to obtain Figure 1 we calculated the power of some basic tests based on direct application of the central limit theorem (see Figure 2). Indeed, for various values of n (100, 400 and 1000) we calculate the power as a function of:

(first row) $p^\star \in (0,1)$, $\mathcal{H}_0 : p = p^\star$ vs. $\mathcal{H}_1 : p = 0.7$.

(second row) $\mu^\star \in (2,4)$, $\mathcal{H}_0 : \mu = \mu^\star$ vs. $\mathcal{H}_1 : \mu = 3$.

(third row) $s^\star \in (0.1,2)$, $\mathcal{H}_0 : F(0.5) = \Phi(0.5/s^\star)$ vs. $\mathcal{H}_1 : F(0.5) = \Phi(0.5/\sigma)$ with $\sigma = 0.5$.

All the graphs show that the power of the various tests increases with the sample size. Because these test are constructed at the 95% level we can see that when

$\mathcal{H}_0$ and $\mathcal{H}_1$ are identical the 5% rejection rate is well satisfied.

To finish this section let us show that chi-square tests proposed in Sections 4.1 and 4.2 have the expected asymptotically free chi-square distributions. Figure 3 shows the level plot of the power as a function of $p^\star$ and $\mu^\star$ in testing $\mathcal{H}_0 : (p, \mu) = (p^\star, \mu^\star)$ versus $\mathcal{H}_1 : (p, \mu) = (0.7, 3)$. The sample size is quite large and as a consequence the power is quickly close to one whenever (p, μ) moves away from $(0.7, 3)$.

In Figure 4 we compare the asymptotic chi-square cdf of the symmetry tests we proposed in Section 4.2 with the empirical cdf obtained from 200 tests produced under the null hypothesis. The test is based on the comparison of $F(-x)+F(x)$ and 1. For one value of x (first row) the asymptotic distribution is a chi-square distribution with 1 degree of freedom, whereas for two (resp. three) values of x (second row) (resp. third row) the asymptotic law is a chi-square distribution with 2 (resp. 3) degrees of freedom. The asymptotic distribution is generally well reached even if from time to time the test may appear a little bit conservative.

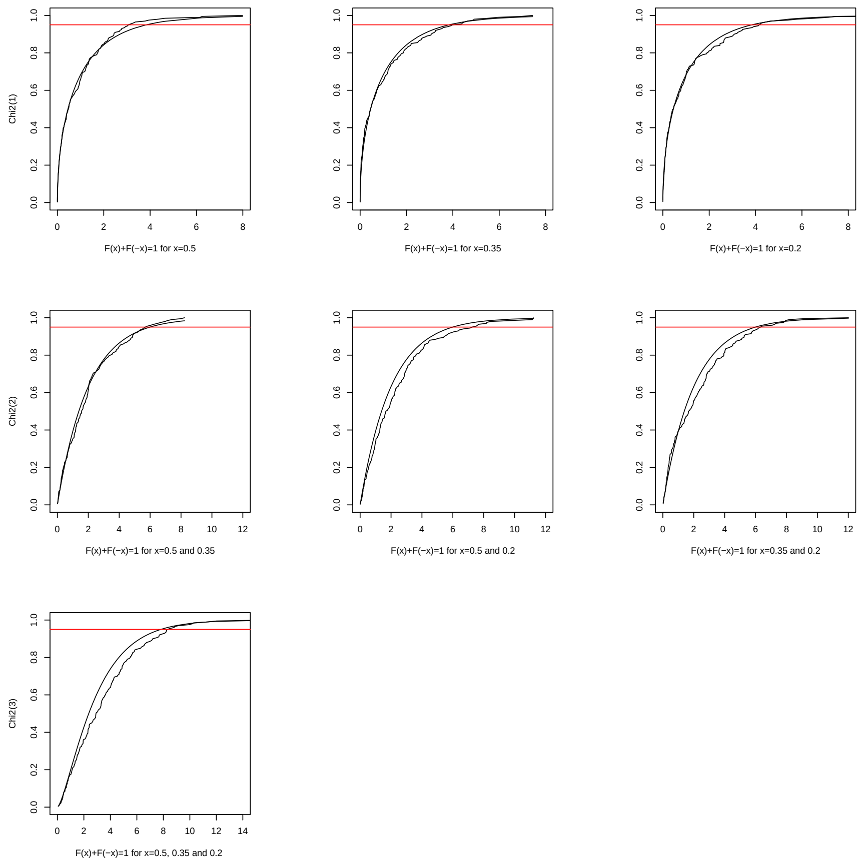

FIG 4. *Empirical cdf of 200 simulated values of the Chi-square symmetry test for* $n = 1000$. *First row: testing at one point, second row: testing at two points, and third row: testing at three points. The horizontal lines correspond to the 95% level.*

Appendix A: Matrices of Theorem 3.2

We define the 3×3 real valued matrix $\Sigma(x,y)$ by its components $\sigma_{ij}(x,y)$ $(1 \leq i,j \leq 3)$ where for $(x,y) \in \mathbb{R}^2$

$$\begin{aligned}
\sigma_{11}(x,y) &\equiv \sigma_{11} = \frac{4}{p_0^2} \int_{\mathbb{R}^2} h_1(u)h_1(v)k(u,v)dG(u)dG(v),\\
\sigma_{22}(x,y) &\equiv \sigma_{22} = \frac{4}{p_0^2} \int_{\mathbb{R}^2} h_2(u)h_2(v)k(u,v)dG(u)dG(v),\\
\sigma_{12}(x,y) &\equiv \sigma_{21} = \frac{4}{p_0^2} \int_{\mathbb{R}^2} h_1(u)h_2(v)k(u,v)dG(u)dG(v),\\
\sigma_{13}(x,y) &\equiv \sigma_{13}(y) = \frac{2}{p_0} \int_{\mathbb{R}} h_1(u)\ell(y,u)dG(u),\\
\sigma_{23}(x,y) &\equiv \sigma_{23}(y) = \frac{2}{p_0} \int_{\mathbb{R}} h_2(u)\ell(y,u)dG(u),\\
\sigma_{31}(x,y) &= \sigma_{13}(x),\\
\sigma_{32}(x,y) &= \sigma_{23}(x),\\
\sigma_{33}(x,y) &= \rho(x+\mu_0, y+\mu_0),
\end{aligned}$$

with

$$\begin{aligned}
h_1(x) &= \frac{1}{p_0}\left(F_0(\mu_0+x) + F_0(\mu_0-x) - 1\right),\\
h_2(x) &= 2f(x),\\
\rho(x,y) &= G(x \wedge y)(1 - G(x \vee y)),\\
\ell(x,y) &= \rho(\mu_0+x, \mu_0+y) + \rho(\mu_0+x, \mu_0-y),\\
k(x,y) &= \ell(x,y) + \ell(-x,y).
\end{aligned}$$

Note that $(h_1, h_2)^T$ is equal to $\dot{H}(\cdot; \vartheta_0, G)$ by (3.8) and (3.9). Let $J(\vartheta_0) = (J_{ij}(\vartheta_0))_{1\leq i,j\leq 3}$ be the 3×3 real valued matrix with entries

$$\begin{aligned}
J_{11}(\vartheta_0) &= -2\int_{\mathbb{R}} h_1^2(u)dG(u), \\
J_{12}(\vartheta_0) &= J_{21}(\vartheta_0) = -2\int_{\mathbb{R}} h_1(u)h_2(u)dG(u), \\
J_{22}(\vartheta_0) &= -2\int_{\mathbb{R}} h_2^2(u)dG(u), \\
J_{13}(\vartheta_0) &= J_{23}(\vartheta_0) = J_{31}(\vartheta_0) = J_{32}(\vartheta_0) = 0, \\
J_{33}(\vartheta_0) &= 1.
\end{aligned}$$

We also define the 3×3 real valued matrix $L(x; \vartheta_0)$ by

$$L(x; \vartheta_0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h_3(x) & f(x) & 1/p_0 \end{pmatrix},$$

where

$$h_3(x) = \frac{F_0(x + \mu_0) - G(x + \mu_0)}{p_0^2}.$$

For each of the above quantities we can define natural estimators. Let us define

$$\begin{aligned}
\hat{h}_1(x) &= \frac{1}{\hat{p}_n}\left(F_0(\hat{\mu}_n + x) + F_0(\hat{\mu}_n - x) - 1\right), \\
\hat{h}_2(x) &= 2\hat{f}_n(x), \\
\hat{h}_3(x) &= \frac{F_0(x + \hat{\mu}_n) - \hat{G}_n(x + \hat{\mu}_n)}{\hat{p}_n^2}, \\
\hat{\rho}(x, y) &= \hat{G}_n(x \wedge y)(1 - \hat{G}_n(x \vee y)), \\
\hat{\ell}(x, y) &= \hat{\rho}(\hat{\mu}_n + x, \hat{\mu}_n + y) + \hat{\rho}(\mu_n + x, \hat{\mu}_n - y), \\
\hat{k}(x, y) &= \hat{\ell}(x, y) + \hat{\ell}(-x, y).
\end{aligned}$$

Then we can estimate $\Sigma(x,y)$ by $\hat{\Sigma}(x,y)$ where

$$\begin{aligned}
\hat{\sigma}_{11}(x,y) &\equiv \hat{\sigma}_{11} = \frac{8}{n(n-1)\hat{p}_n^2} \sum_{1\leq i<j\leq n} \hat{h}_1(X_i)\hat{h}_1(X_j)\hat{k}(X_i,X_j),\\
\hat{\sigma}_{22}(x,y) &\equiv \hat{\sigma}_{22} = \frac{8}{n(n-1)\hat{p}_n^2} \sum_{1\leq i<j\leq n} \hat{h}_2(X_i)\hat{h}_2(X_j)\hat{k}(X_i,X_j),\\
\hat{\sigma}_{12}(x,y) &\equiv \hat{\sigma}_{21} = \frac{4}{n(n-1)\hat{p}_n^2} \sum_{1\leq i\neq j\leq n} \hat{h}_1(X_i)\hat{h}_2(X_j)\hat{k}(X_i,X_j),\\
\hat{\sigma}_{13}(x,y) &\equiv \hat{\sigma}_{13}(y) = \frac{2}{n\hat{p}_n} \sum_{i=1}^{n} \hat{h}_1(X_i)\hat{\ell}(y,X_i),\\
\hat{\sigma}_{23}(x,y) &\equiv \hat{\sigma}_{23}(y) = \frac{2}{n\hat{p}_n} \sum_{i=1}^{n} \hat{h}_2(X_i)\hat{\ell}(y,X_i),\\
\hat{\sigma}_{31}(x,y) &= \hat{\sigma}_{13}(x),\\
\hat{\sigma}_{32}(x,y) &= \hat{\sigma}_{23}(x),\\
\hat{\sigma}_{33}(x,y) &= \hat{\rho}(x+\hat{\mu}_n, y+\hat{\mu}_n),
\end{aligned}$$

and $\hat{J} = (\hat{J}_{ij})_{1\leq i,j\leq 3}$ with

$$\begin{aligned}
\hat{J}_{11} &= -\frac{2}{n}\sum_{i=1}^{n} \hat{h}_1^2(X_i), \quad \hat{J}_{12} = \hat{J}_{21} = -\frac{2}{n}\sum_{i=1}^{n} \hat{h}_1(X_i)\hat{h}_2(X_i),\\
\hat{J}_{22} &= -\frac{2}{n}\sum_{i=1}^{n} \hat{h}_2^2(X_i),\\
\hat{J}_{13} &= \hat{J}_{23} = \hat{J}_{31} = \hat{J}_{32} = 0 \quad \text{and} \quad \hat{J}_{33} = 1.
\end{aligned}$$

The $L(x;\vartheta_0)$ matrix is estimated by $\hat{L}(x)$ the third line of which being therefore $(\hat{h}_3(x), \hat{h}_2(x)/2, 1/\hat{p}_n)$.

References

[1] Arcones, M. A. and Giné, E. (1991). Some bootstrap tests of symmetry for univariate continuous distributions. *Ann. Statist.* **19** 1496–1511. MR1126334

[2] Bordes, L., Mottelet, S. and Vandekerkhove, P. (2006a). Semiparametric estimation of a two-component mixture model. *Ann. Statist.* **34** 1204–1232. MR2278356

[4] Bordes, L., Delmas, C. and Vandekerkhove, P. (2006b). Semiparametric estimation of a two-component mixture model when a component is known. *Scand. J. Statist.* **33** 733–752. MR2300913

[4] Bordes, L., Chauveau, D. and Vandekerkhove, P. (2007). A stochastic EM algorithm for a semiparametric mixture model. *Comput. Statist. Data Anal.* **51** 5429–5443.

[5] Cruz-Medina, I. R. and Hettmansperger, T. P. (2004). Nonparametric estimation in semiparametric univariate mixture models. *Journal of Statistical Computation and Simulation* **74** 513–524. MR2073229

[6] Efron, B. (2007). Size, power and false discovery rate. *Ann. Statist.* **35** 1351–1377.

[7] Giné, E. and Guillou A. (2002). Rates of Strong Uniform Consistency for Multivariate Kernel density estimators. *Ann. Inst. H. Poincaré* **38** 503–522. MRMR1955344

[8] Hall, P. and Zhou, X. H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.* **31** 201–224. MR1962504

[9] Hall, P., Neeman, A., Pakyari, R. and Elmore, R. (2005). Nonparametric inference in multivariate mixtures. *Biometrika* **92** 667–678. MR2202653

[10] Hettmansperger, T. P. and Thomas, H. (2000). Almost nonparametric inference for repeated measures in mixture models. *J. Royal Statist. Soc. Ser. B* **62** 811–825. MR1796294

[11] Hunter, D. R., Wang, S. and Hettmansperger, T. P. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist.* **35** 224–251.

[12] Lindsay, B. G. and Lesperance, M. L. (1995). A review of semiparametric mixture models. *J. Statist. Plann. Inf.* **47** 29–39. MRMR1360957

[13] McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models.* Wiley, New York. MR1789474

[14] Maiboroda, R. and Sugakova, O. (2009). Generalized estimating equations for symmetric distributions observed with admixture. *Preprint*.

[15] Moore, D.S. (1971). A chi-square statistic with random cell boundaries. *Ann. Math. Statist.* **42** 147–156. MR0275601

[16] Naylor, J. C. and Smith, A. F. M. (1983). A contamination model in clinical chemistry: an illustration of a method for the efficient computation of posterior distributions. *The Statistician* **32** 82–87.

[17] Pal, C. and Sengupta, A. (2000). Optimal tests for no contamination in reliability models. *Lifetime Data Anal.* **6** 281–290. MR1786803

[18] Ruymgaart, F. H. (1975). A note on chi-square statistics with random cell boundaries. *Ann. Statist.* **3** 965–968. MR0378183

[19] Robin, S. A., Bar-Hen, A., Daudin, J-J. and Pierre, L. (2007). Semi-parametric approach for mixture models: Application to local false discovery rate estimation. *Comput. Statist. Data Anal.* In press.

[20] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York. MR0838963

[21] Schuster, E. F. and Barker, R. C. (1987). Using the bootstrap in testing symmetry versus asymmetry. *Comm. Statist. Simulation Comput.* **16** 69–84. MR0883122

[22] Silverman, B. W. (1978). Weak and strong consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177–184. MR0471166

[23] Sougakova, O. (2009) Estimation of location parameter by observations with admixture. *Theorija Imovirnosti ta Matematychna Statystyka*, **80**, to appear.

[24] Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Wiley, Chichester. MR0838090

[25] van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, New York. MR1652247

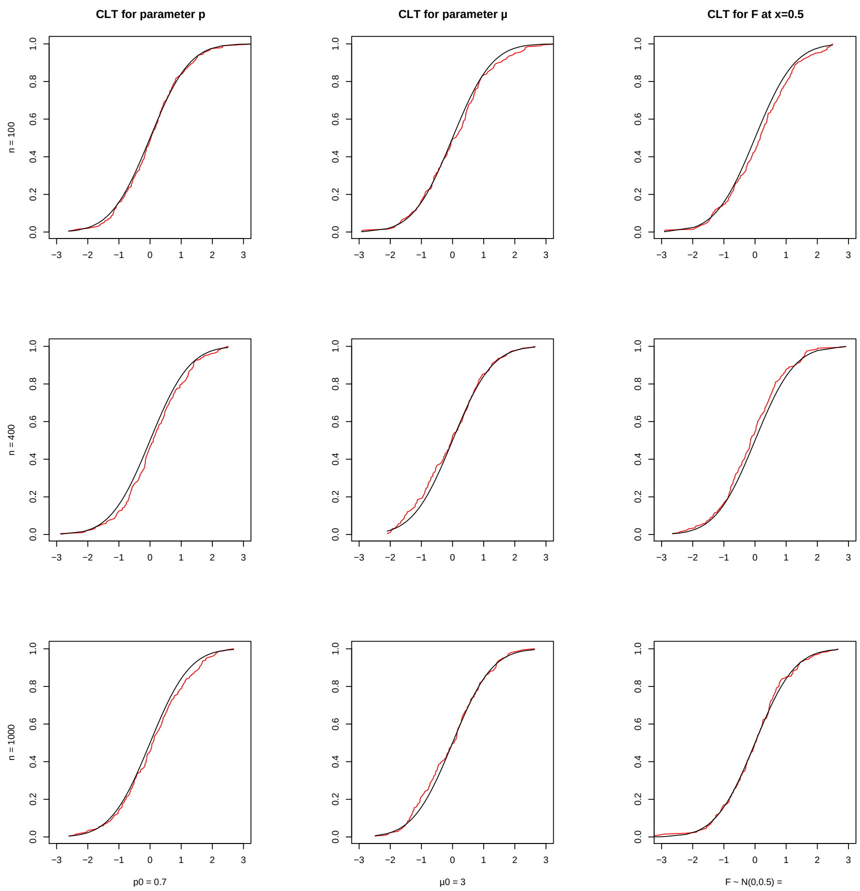

FIG 1. *Comparison of the empirical cdf of 200 estimates of p (first column), μ (second column) and $F(0.5)$ (third column) with the cdf of a $\mathcal{N}(0,1)$ for sample sizes equal to 100 (first row), 400 (second row) and 1000 (third row). Each estimate is centered on the parameter and reduced using the estimated standard deviation.*

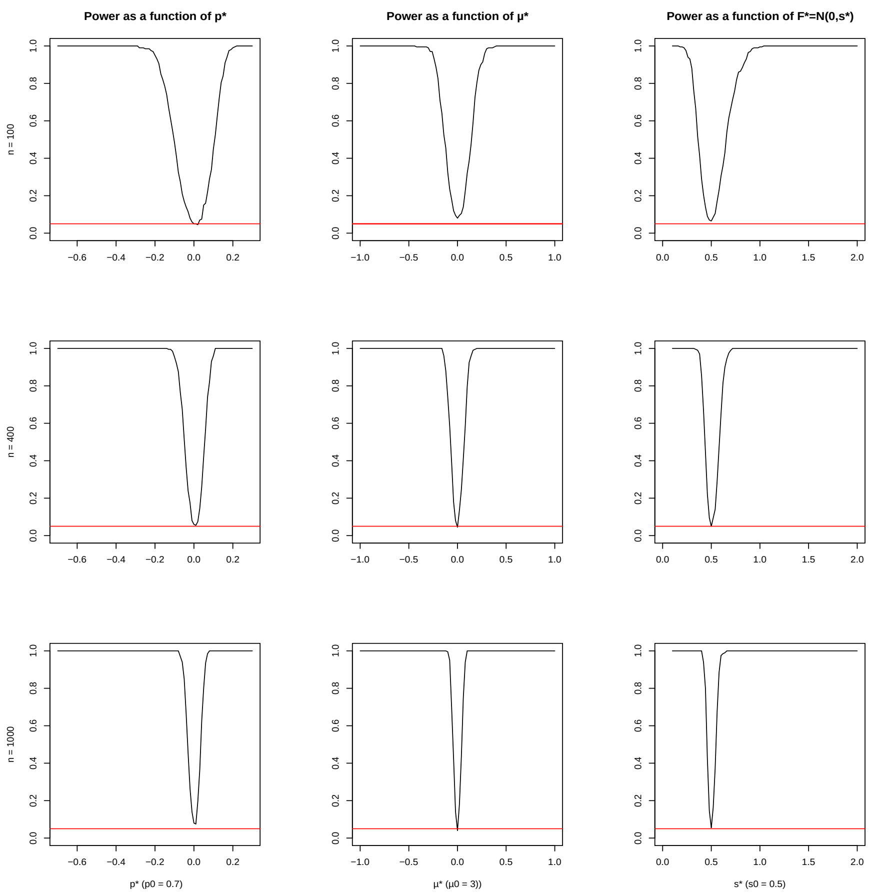

FIG 2. *Power calculations for* $n = 100$ *(first row),* 400 *(second row) and* 1000 *(third row) for testing* $p = p^\star$ *(first column),* $\mu = \mu^\star$ *(second column) and* $F(0.5) = F^\star(0.5) \equiv \Phi(0.5/s^\star)$*. Under* $\mathcal{H}_1$ *we have* $(p, \mu, F(0.5) = (0.7, 3, \Phi(1))$*. The horizontal lines correspond to the 5% level.*

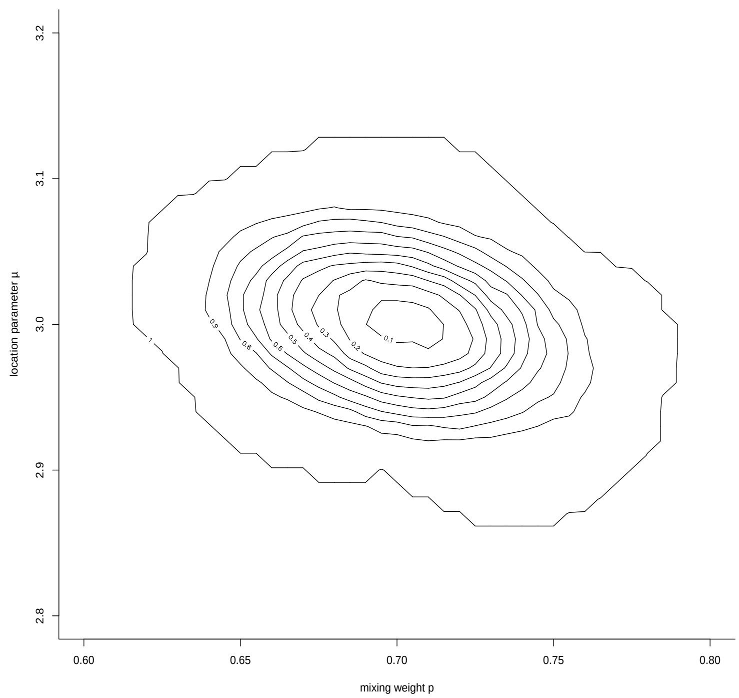

FIG 3. *Power estimation based on 200 estimates of* (p, μ). *The null hypothesis is that* $(p, \mu) \in [0.6, 0.8] \times [2.8, 3.2]$ *against* $(p, \mu) = (0.7, 3)$. *The sample size is* $n = 1000$.