Electronic Journal of Statistics
Vol. 13 (2019) 4743–4793
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1650

Semiparametric density testing in the contamination model

Denys Pommeret and Pierre Vandekerkhove*

Aix Marseille Univ, CNRS, Centrale Marseille, I2M
e-mail: denys.pommeret@univ-amu.fr

Université Paris-Est Marne-la-Vallée, LAMA (UMR 8050) and UMI Georgia Tech - CNRS 2958, George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology
e-mail: pierre.vandek@univ-mlv.fr

In memory of Marie Duflo

Abstract: In this paper we investigate a semiparametric testing approach to answer if the parametric family allocated to the unknown density of a two-component mixture model with one known component is correct or not. Based on a semiparametric estimation of the Euclidean parameters of the model (free from the null assumption), our method compares pairwise the Fourier's type coefficients of the model estimated directly from the data with the ones obtained by plugging the estimated parameters into the mixture model. These comparisons are incorporated into a sum of square type statistic which order is controlled by a penalization rule. We prove under mild conditions that our test statistic is asymptotically χ_1^2-distributed and study its behavior, both numerically and theoretically, under different types of alternatives including contiguous nonparametric alternatives. We discuss the counterintuitive, from the practitioner point of view, lack of power of the maximum likelihood version of our test in a neighborhood of challenging non-identifiable situations. Several level and power studies are numerically conducted on models close to those considered in the literature, such as in McLachlan *et al.* [21], to validate the suitability of our approach. We also implement our testing procedure on the Carina galaxy real dataset which low luminosity mixes with the one of its companion Milky Way. Finally we discuss possible extensions of our work to a wider class of contamination models.

MSC 2010 subject classifications: Primary 62F03; secondary 62G10.
Keywords and phrases: Chi-squared test, contamination model, semiparametric estimator, maximum likelihood estimator, nonparametric contiguous alternative.

Received April 2019.

1. Introduction

Let us consider n independent and identically distributed random variables $(X_1, \dots, X_n)$ drawn from a two-component mixture model with probability den-

*First supporter of the project.

sity function g defined by:

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

where f_0 is a known probability density function, corresponding to a known signal, and where the unknown parameters of the model are the mixture proportion $p \in (0,1)$ and the probability density function $f \in \mathcal{F}$ (a given class of densities) associated to an unknown signal. Model (1) is widely used in statistics and is usually so-called the *contamination* model. This class of models is especially suitable for detection of differentially expressed genes under various conditions in microarray data analysis, see McLachlan *et al.* [21] or Dai and Charnigo [10]. In astronomy such a model has been used to model mixtures of X-ray sources, see Melchior and Goulding [23] and Patra and Sen [26]. Recently some applications have been also developed in selective Statistical Editing, see Di Zio and Guarnera [6], in biology to model trees diameters, see Podlaski and Roesch [27] or in kinetics to model plasma data, see Klingenber *et al.* [16].

Many techniques have been proposed to estimate the Euclidean and functional parameters p and f in model (1). The most popular methods for known finite order mixture models, such as the moment method, see Lindsay [19], the moment generating function based method, see Quandt and Ramsey [28], or the maximum likelihood method, see Lindsay [18], are largely used but suffer from the requirement of assigning a parametric form to the f density. Moreover, such parametric methods can be non-reliable if a model is close to a non-identifiable one, as for instance in the Gaussian case, taking in model (1) $f_0(\cdot) = f_{(0,1)}(\cdot)$ and $f(\cdot) = (f_{(0,1)}(\cdot) + f_{(\mu,1)}(\cdot))/2$ and writing the double representation

$$(1-p)f_{(0,1)}(x) + p\left(\frac{f_{(0,1)}(x) + f_{(\mu,1)}(x)}{2}\right) = (1-\frac{p}{2})f_{(0,1)}(x) + \frac{p}{2}f_{(\mu,1)}(x) \tag{2}$$

where $f_{(\mu,s)}$ denotes the Gaussian density function with mean μ and variance s. The above example illustrates Bordes *et al.* [8] (Proposition 2) which establishes an almost-everywhere identifiability result for the contamination model when f_0 is supposed to be zero-symmetric and f symmetric with repect to a location parameter μ. The sufficient identifiability constraint established by these authors is: $\mathrm{var}_{f_0} \neq \mathrm{var}_f + \mu\frac{2\pm k_0}{3k_0}$ for $k_0 \in \mathbb{N}^*$. Note that in the double representation (2) the previous condition is clearly not satisfied since we have $\mathrm{var}_{f_0} = \mathrm{var}_f + \mu\frac{2-k_0}{3k_0} = 1$ for $k_0 = 2$. This type of identifiability except on a set of parameters of measure zero, so called *generic identifiability*, is more widely investigated in Allman *et al.* [1].

As described in Section 4, slight modifications of the above model can easily trap parametric estimation near spurious modeling representations. Since then, some semiparametric approaches have been developed, such as the pioneer work by Bordes *et al.* [8], to relax that parametric modelling. These authors only restricted, for example, their study to the class of location-shift symmetric densities in order to make model (1) semiparametrically identifiable. More recently, different nonparametric approaches have been also considered, such as in Nguyen and Matias [22] where f_0 is a uniform distribution on $[0,1]$. In Ma and

Yao [20], where f_0 is only supposed to belong to a parametric family, a tail identifiability approach is used, considering symmetric distributions embedded in a nonparametric envelop. We also recommend the recent work by Al Mohamad and Boumahdaf [3] who consider situations where the unknown component f is defined through linear constraints. In Balabdaoui and Doss [5] a log-concave assumption is done on the family $\mathcal{F}$ to insure the identifiability of the model. In Patra and Sen [26] the identifiability and estimation problem is considered under tail conditions with very few shape constraints assumptions.

The goal of the present paper is to answer a very natural question, explicitly raised in McLachlan *et al.* [21] (Section 6) or Patra and Sen [26] (Section 9.2), which is basically "can we test if the unknown component of the contamination model belongs to a given class of parametric densities?", or more formally can we test

$$H_0 : f \in \mathcal{F} = \{f_\xi; \xi \in \Lambda\} \qquad \text{against} \qquad H_1 : f \notin \mathcal{F}, \tag{3}$$

where f_ξ is a probability density function parametrized by an Euclidean parameter ξ belonging to a parametric space Λ. For simplicity we will restrict ourselves to the case where f_ξ belongs to $\mathcal{S}$, the set of symmetric probability density functions with respect to a location parameter $\mu \in \mathbb{R}$, that is: there exists $\mu \in \mathbb{R}$, part of the parameter ξ, such that

$$f_\xi(x + \mu) = f_\xi(-x + \mu).$$

However we discuss in Section 10 how our approach can be generalized to any class of parametric densities provided that model (1) can be $\sqrt{n}$-estimated semiparametrically. This problem has been considered recently by Suesse *et al.* [32], who use a maximum likelihood estimate-based testing approach. In general the behavior of the maximum likelihood estimator is difficult to control or figure out, as illustrated in Section 7, under the alternative since the model is then misspecified. To get a consistent testing method under both H_0 and H_1, at the price of some shape restriction about H_1, we propose to use an $H_0 \cup H_1$ consistent semiparametric estimation approach in order to build a H_0-free statistic (do not forcing to fit into the parametric model). To the best of our knowledge this is the first time that an H_0-free semiparametric approach is used to test mixture models. The advantage of this new strategy will be demonstrated, both theoretically and numerically, on very counterintuitive examples in the close neighborhood of non-identifiable situations, see Fig. 1 and comments. For a general overview about semiparametric mixture models we recommend the recent surveys by Xiang *et al.* [38] or Gassiat [12]. Note that the test against a specific distribution, proposed in Bordes and Vandekerkhove [9] (Section 4.1), does not allow to test versus a complete class of probability density functions, which is our goal here. To point out the interest of the statistical community about the contamination problem testing, let us mention the very recent work by Arias-Castro and Huang [4] on the sparse variance contamination model testing and references therein.

The main idea of our test is based on the following data driven smooth test procedure developed by Ledwina [17], extending the idea of Neyman [25]. Writing $\mathcal{S}$ the set of symmetric probability density functions with respect to a location parameter $\mu \in \mathbb{R}$, the idea of our test consists in estimating the expansion coefficients of f in an orthogonal basis, first assuming $f \in \mathcal{S}$, and to compare these estimates to those obtained by assuming $f \in \mathcal{F}$. This approach has been used in Doukhan *et al.* [11], see also references therein, but the specificity of the two-component mixture model necessitates a special adaptation of the Neyman smooth test. In our case we develop a two rates procedure, one rate driven by the asymptotic normality of the test statistic and another one driven by the almost sure rate of convergence of the semiparametric estimators. As we will discuss along this paper, the approach of Suesse *et al.* [32], restricted to model (1), does not allow to investigate the asymptotic behavior of the test statistic under alternative assumptions (possibly contiguous) since the asymptotic behavior of the maximum likelihood estimator cannot be controlled properly under distribution misspecification. Another side of our nonparametric approach is that it can easily deal with situations where f_0 is only known through a training data. This situation is illustrated in Section 9 through a real dataset collecting the radial velocity of the Carina galaxy and its companion Milky Way.

The paper is organized as follows: in Section 2 we describe our two-step test methodology; in Section 3 we state the assumptions and asymptotic results under the null hypothesis; Section 4 is dedicated to the test behavior under the alternative; Section 5 is devoted to the study of our testing procedure under contiguous nonparametric alternatives (inspired from the parametric contiguous alternative concept); in Section 6 we discuss the choice of the reference measure when considering orthogonal bases for the unknown density decomposition; in Section 7 we conduct a power comparison between the semiparametric and maximum likelihood versions for our test, this section enlightens interestingly the fact that a maximum likelihood approach could force, in certain setups of the McLachlan *et al.* [21] (Section 6) Gaussian mixture model, to consider the number q of components defining f equal to 1 when in reality $q = 2$; Section 8 is dedicated to a simulation-based empirical and power levels study; in Section 9 we proceed with the application of our testing method to the datasets (breast cancer, colon cancer, HIV) previously studied in McLachlan *et al.* [21] and to the Galaxy dataset studied in Patra and Sen [26]. Finally in Section 10 we discuss further leads of research connected with the contamination model testing problem.

2. Testing problem

Let us consider an independent and identically distributed sample denoted $(X_1, \dots, X_n)$, drawn from a probability density function g defined in (1) with respect to a given reference measure ν. The problem addressed in this section deals with testing the unknown component f assuming the fact that f belongs to $\mathcal{S}$, the set of symmetric densities. More precisely, denoting

$$\mathcal{F} = \{f_\xi; \xi = (\mu, \theta) \in \Lambda\}$$

the set of symmetric densities with respect to ν, with mean μ and scale/shape parameter θ, where (μ, θ) is supposed to belong to a compact set Λ of $\mathbb{R} \times \Theta$, our goal is to test

$$H_0 : f \in \mathcal{F} \qquad \text{against} \qquad H_1 : f \in \mathcal{S} \backslash \mathcal{F}. \tag{4}$$

The symmetry condition, meanning that there exists $\mu \in \mathbb{R}$ such that $f(-x + \mu) = f(x + \mu)$, is a semiparametric identifiability and parameter picking condition proposed in Bordes *et al.* [8]. The scale/shape parameter θ corresponds to the variance in the Gaussian case but it can also be a vector as in the Generalized Gaussian distribution (GGD). In that latter case the scale parameter α along with the shape parameter β both belong to $\mathbb{R}^{+*}$ which leads to consider $\theta = (\alpha, \beta)^\top \in \mathbb{R}^{+*} \times \mathbb{R}^{+*} = \Theta$.

Our test procedure consists in estimating the expansion coefficients of the unknown density f in an orthogonal basis, first assuming $f \in \mathcal{S}$, and comparing in contrast these estimates to those obtained when f is supposed to strictly belong to a parametric sub-family $\mathcal{F}$ of $\mathcal{S}$. As intuitively expected, we will show how the study of the successive expansion coefficient differences helps in detecting possible departure from H_0 given the data. We will denote by $\mathcal{Q} = \{Q_k; k \in \mathbb{N}\}$, a ν-orthogonal basis satisfying $Q_0 = 1$ and such that

$$\int_{\mathbb{R}} Q_j(x) Q_k(x) \nu(dx) = q_k^2 \delta_{jk}, \tag{5}$$

with $\delta_{jk} = 1$ if $j = k$ and 0 otherwise, and where the normalizing factors $q_k^2 \geq 1$ will allow to control the variance of our estimators, as illustrated in Lemmas 1 and 3. We assume that $\mathcal{Q}$ is an $L^2(\mathbb{R}, \nu)$ Hilbert basis, which is satisfied if there exists $\kappa > 0$ such that $\int_{\mathbb{R}} e^{\kappa|x|} \nu(dx) < \infty$, and that the following integrability conditions are satisfied:

$$\int_{\mathbb{R}} f_0^2(x) \nu(dx) < \infty \qquad \text{and} \qquad \int_{\mathbb{R}} f^2(x) \nu(dx) < \infty.$$

Then, for all $x \in \mathbb{R}$, we have

$$g(x) = \sum_{k \geq 0} a_k Q_k(x) \qquad \text{with} \qquad a_k = \int_{\mathbb{R}} Q_k(x) g(x) \nu(dx) / q_k^2,$$

$$f_0(x) = \sum_{k \geq 0} b_k Q_k(x) \qquad \text{with} \qquad b_k = \int_{\mathbb{R}} Q_k(x) f_0(x) \nu(dx) / q_k^2,$$

$$f(x) = \sum_{k \geq 0} c_k Q_k(x) \qquad \text{with} \qquad c_k = \int_{\mathbb{R}} Q_k(x) f(x) \nu(dx) / q_k^2.$$

From (1) we have

$$a_k = (1 - p) b_k + p c_k.$$

Let us denote by Z a random variable with density $f_{\mu,\theta}$ and consider

$$\alpha_k(\mu,\theta) = \mathbb{E}(Q_k(Z))/q_k^2.$$

The null hypothesis can be reformulated as $c_k = \alpha_k(\mu,\theta)$, for all $k \geq 1$, or equivalently as

$$H_0: \quad \text{there exist } (\mu,\theta) \in \Lambda \text{ such that } a_k = (1-p)b_k + p\alpha_k(\mu,\theta), \text{ for all } k \geq 1. \tag{6}$$

Since the probability density function f_0 is known, the coefficients b_k are automatically known. For all $k \geq 1$, the coefficients a_k can be estimated empirically by:

$$a_{k,n} = \frac{1}{n}\sum_{i=1}^{n} \frac{Q_k(X_i)}{q_k^2}, \quad n \geq 1.$$

To obtain H_0-free estimators of the parameters (p,μ) and α_k's, the estimator of (p,μ) will be obtained without assuming the null hypothesis, that is using the semiparametric estimator $\bar{\vartheta}_n = (\bar{p}_n, \bar{\mu}_n)$ introduced in Bordes *et al.* [8] and studied more deeply in Bordes and Vandekerkhove [9]. Indeed, as numerically demonstrated in Section 7, the maximum likelihood estimator $(\widehat{p}_n, \widehat{\mu}_n, \widehat{\theta}_n)$ under the null assumption tends to provide the best H_0-fitted model when the semiparametric estimator of Bordes and Vandekerkhove [9] is not influenced by the constraint under the null and can provide very distant, both Euclidean and functional, estimations under H_1 (when the model is misspecified under the null assumption). In the same way, considering the relation (1), the estimator of θ is obtained by the H_0-free method of moments. The estimator of $\alpha_k(\mu,\theta)$ is obtained by using a standard plug-in approach, that is:

$$\alpha_{k,n} = \alpha_k(\bar{\mu}_n, \bar{\theta}_n).$$

To illustrate our general approach, let us consider the case where the parameter θ coincides with the variance parameter. This is the case for instance when $\mathcal{F}$ is equal to $\mathcal{G}$ the set of normal densities with mean μ and variance $\theta = s$. Then the method of moments yields

$$\bar{s}_n = \frac{\bar{M}_{2,n} - (1-\bar{p}_n)m_2}{\bar{p}_n} - (\bar{\mu}_n)^2, \tag{7}$$

where $\bar{M}_{2,n} = n^{-1}\sum_{i=1}^{n} X_i^2$, and $m_2 = \int_{\mathbb{R}} x^2 f_0(x)\nu(dx)$. Coming back now to generality and looking at the H_0 reformulation in (6), we can expect that the differences

$$R_{k,n} = a_{k,n} - \bar{p}_n(\alpha_{k,n} - b_k) - b_k, \quad \text{for all } k \geq 1,$$

will allow us to detect any possible departure from the null hypothesis. For simplicity matters and without loss of generality, since the b_k's are known constants, we assume from now on them to be equal to zero. For all $k \geq 1$, we define the

k-th order coefficient of our test statistic (incorporating the k-th order departure information from H_0)

$$T_{k,n} = nU_{k,n}^\top \widehat{D}_{k,n}^{-1} U_{k,n}, \tag{8}$$

where $U_{k,n} = (R_{1,n}, \dots, R_{k,n})$ and where $\widehat{D}_{k,n}$ is an estimator of

$$D_{k,n} = \mathrm{diag}(\mathrm{var}(R_{1,n}), \dots, \mathrm{var}(R_{k,n})),$$

normalizing the test statistic as in Munk *et al.* [24]. To avoid instability in the evaluation of $\widehat{D}_{k,n}^{-1}$, following Doukhan *et al.* [11], we add a trimming term $e(n)$ to every i-th, $i = 1, \dots, k$, diagonal element of $\widehat{D}_{k,n}$ as follows:

$$\widehat{D}_{k,n}[i] = \max(\widehat{\mathrm{var}}(R_{i,n}), e(n)), \quad 0 \le i \le k, \tag{9}$$

where $\widehat{\mathrm{var}}(R_{i,n})$ is a weakly consistent estimator of $\mathrm{var}(R_i)$ and $e(n) \to 0$, as $n \to +\infty$.

Following Ledwina [17] and Inglot *et al.* [15] we suggest a data driven procedure to select automatically the number of coefficients needed to answer the testing problem. We introduce the following penalized rule to pick parsimoniously (trade-off between H_0 departure detection and complexity of the procedure involved by index k) the "*best*" rank k for looking at $T_{k,n}$:

$$S_n = \min\big\{ \underset{1 \le k \le d(n)}{\mathrm{argmax}} \, (s(n)T_{k,n} - \beta_k pen(n)) \big\}, \tag{10}$$

where respectively $s(n) \to 0$ is a normalizing rate, $d(n) \to +\infty$, $pen(n)$ is a penalty term such that $pen(n) \to +\infty$, as $n \to +\infty$, and the β_k's are penalization factors. In practice we will consider $\beta_k = k$, $k \ge 1$, and $pen(n) = \log(n)$, $n \ge 1$. To match the asymptotic normality regime, under H_0, of the test statistic $T_{k,n}$ defined in (8), the normalizing factor $s(n)$ is usually taken equal to one, but in our case, due to the specificity of the semiparametric mixture estimation (possibly adapted to nonparametric contiguous alternatives), we chose:

$$s(n) = n^{\lambda - 1}, \quad \text{with} \quad 0 < \lambda < 1/2. \tag{11}$$

The above calibration is connected with the almost sure convergence rate of the estimators $\bar{p}_n$ and $\bar{\mu}_n$ which satisfy $|\bar{p}_n - p_0|^2 = o_{a.s.}(n^{-1/2+\alpha})$ and $|\bar{\mu}_n - \mu_0|^2 = o_{a.s.}(n^{-1/2+\alpha})$ for all $\alpha > 0$, see Theorem 3.1 in Bordes and Vandekerkhove [9]. Note that the selection rule in (10), adapted to the semiparametric framework, strongly differs from the BIC criterion used by Suesse *et al.* [32].

Remark 1. *It is important to notice at this point that we could have also investigated a test expressed like this:*

$$H_0: \text{ there exists } \xi = (\mu, \theta) \in \Lambda \text{ such that } F = F_\xi,$$

against its alternative, where F denotes the cumulative distribution function of f. For simplicity and without loss of generality, consider the Gaussian case with *known mean $\mu = 0$ (p still unknown) and variance parameter $s = \theta$ and write $F_s(\cdot) = F(\sqrt{s} \times \cdot)$. Write also $F_{(0,1)}$ the standard Gaussian cumulative distribution function. In such a perspective we could have used a strategy inspired from the simple hypothesis test of Bordes and Vandekerkhove [9] (Section 4.1). Since according to Theorem 3.2 in Bordes and Vandekerkhove [9] the semiparametric estimator $\widehat{F}_n$ of F satisfies a functional central limit theorem, one could consider s_n in (7) as a natural estimate of s under H_0 and evaluate the square of*

$$\sqrt{n}[\widehat{F}_{n,s_n} - F_{(0,1)}] = \sqrt{n}[\widehat{F}_{n,s_n} - F_{s_n}] + \sqrt{n}[F_{s_n} - F_{(0,1)}]$$

over a set of fixed values $(x_1, \dots, x_k)$, where $\widehat{F}_{n,s_n}(\cdot) = \widehat{F}_n(\sqrt{s_n} \times \cdot)$. By using the delta method, we can show that the second term of the above quantity is asymptotically normal, however the behavior of the first term looks much more difficult to analyze due to the random factor term s_n inside the semiparametric estimate $\widehat{F}_n$. In addition of this technical difficulty, it would also be more satisfactory to investigate a Kolmogorov type test based on $\sqrt{n}\sup_{x\in\mathbb{R}} |F_n(\sqrt{s_n}x) - F_{(0,1)}(x)|$, embracing the whole complexity of $F_{(0,1)}$, instead of a $\chi^2(k)$-type test based on the above expression evaluated over a k-grid. Again this is a very challenging problem. In that sense our approach allows to get a sort of asymptotic framework to capture the whole complexity of f through its (asymptotically unrestricted) decomposition in a base of orthogonal functions.

3. Assumptions and asymptotic behavior under H_0

To test consistently (4), based on the statistic $T(n) = T_{S_n,n}$, we will suppose the following conditions:

(A1) The coefficient order upper bound $d(n)$ involved in (10) satisfies $d(n) = O(\log(n)e(n))$, where $e(n)$ is the trimming term in (9).

(A2) For all $k \geq 1$, $\alpha_k(\cdot,\cdot)$ is a $\mathcal{C}^1$ function and there exists nonnegative constants M_1 and M_2 such that for all $(\mu,\theta) \in \Lambda$,

$$|\alpha_k(\mu,\theta)| \leq M_1 \quad \text{and} \quad \|\dot{\alpha_k}(\mu,\theta)\| \leq M_2,$$

where $\dot{\alpha_k}$ denotes the gradient $(\partial\alpha_k/\partial\mu, \partial\alpha_k/\partial\theta)^\top$ and $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^2$.

(A3) There exists a nonnegative constant M_3 such that for all $k \in \mathbb{N}^*$,

$$\frac{1}{k}\sum_{i=1}^{k} \mathrm{var}\left(\frac{Q_i(X_1)}{q_i^2}\right) \leq M_3.$$

Under these three conditions, which will be checked respectively in Lemma 1 and 3 for the Gaussian and the Lebesgue reference measure, we state the following theorem.

Theorem 2. *If assumptions* **(A1-3)** *hold, then under H_0:*

$$\mathbb{P}(S_n = 1) \longrightarrow 1, \quad \text{as} \quad n \to +\infty.$$

The above result states that under H_0 the infinite collection (6) of equality expansion conditions is asymptotically examined simply at rank $k = 1$, thanks to the penalization term involved in the test statistic $T_{S_n,n}$ where S_n is defined in (10). This basic result allows to deduce the following standard asymptotic distribution of the test statistic under the null, which convergence is also numerically illustrated in Appendix G.

Corollary 3. *If assumptions* **(A1-3)** *hold, we have under* H_0 *the convergence in law*

$$T(n) \xrightarrow{\mathcal{L}} Z \quad , \quad \text{as } n \to +\infty,$$

where Z *denotes a* χ^2*-distributed random variable with one degree of freedom.*

Remark 4. *Theorem 2 and Corollary 3 still hold if we replace in* $T(n)$ *the semiparametric estimators and their (asymptotic) variances by their maximum likelihood counterparts. The proofs of these two results are completely similar to the semiparametric case and rely on the asymptotic normality of the maximum likelihood estimator detailed in Appendix F. In this case the rate of the selection rule is the standard one, which namely is* $s(n) = 1$.

4. Asymptotic behavior under H_1

In the next proposition we study the behaviour of our test statistic under $H_1 : f \in \mathcal{S} \setminus \mathcal{F}$.

Proposition 1. *If* $f \in \mathcal{S} \setminus \mathcal{F}$*, then the test statistic* $T(n)$ *tends to* $+\infty$ *in probability with a drift-type behavior* $T_n \geq O_P(n^\lambda)$, $0 < \lambda < 1/2$*, as* $n \to +\infty$.

We would like to stress the fact that the identifiability conditions supposed when considering the class of symmetric densities $\mathcal{S}$, are crucial in the proof of Proposition 1. As mentioned in Bordes *et al.* [8], there exists various non identifiability cases for model (1). Let us remind the following one from Bordes and Vandekerkhove [9]: if we take f_0 to be φ and f to be $(\varphi(\cdot - \mu) + \varphi(\cdot + \mu))/2$, where φ is an even probability density function, then the parameters are not uniquely identifiable from g because we have the double writing

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

This example is very interesting since it clearly shows the difficulty of estimating model (1) when the probability density function of the unknown component is a mixture including exactly the same shape as the known component. In particular if φ is a given Gaussian distribution and we want to test if the 2nd component is Gaussian, we could possibly either reject or accept H_0 with our testing procedure depending on the convergence of our semiparametric estimators. Indeed the maximum likelihood estimator would converge towards the natural underlying Gaussian model and the semiparametric method could possibly converge towards both solutions. To avoid this concern, we recommend to check that the departures between the maximum likelihood estimator and the semiparametric one is not driven by a factor 2, *i.e* $\widehat{\mu}_n \approx 2\bar{\mu}_n$ and $\widehat{p}_n \approx \bar{p}_n/2$. To advise on this possible proximity, one could check if $\widehat{\mu}_n/2$ and $2\widehat{p}_n$ respectively belong to the 95% confidence intervals of μ and p derived from the asymptotic normality of $(\bar{p}_n, \bar{\mu}_n)$, see Bordes and Vandekerkhove [9]. Now if so, we suggest to initialize the semiparametric approach close the maximum likelihood estimator to force it to detect the possibly existing $\mathcal{F}$-component in model (1).

5. Contiguous alternatives

5.1. Contiguous contamination models

We consider in this section a *vanishing* convolution-class of nonparametric contiguous alternatives. More specifically, the null hypothesis consists here in considering that the observed sample $\mathbf{X}^n = (X_1, \dots, X_n)$ comes from

$$H_0: \quad X_i = (1 - U_i)Y_i + U_i Z_i, \quad i = 1, \dots, n,$$

where $(U_i)_{i\geq 1}$ and $(Y_i, Z_i)_{i\geq 1}$ are respectively independent and identically distributed sequences distributed according to a Bernoulli distribution with parameter p and $f_0 \otimes f_{\mu,\theta}$, where $f_{\mu,\theta}$ is the unknown density function with respect to the reference measure ν. For each $n \geq 1$, the contiguous alternative consists in the fact that the observed sample $\mathbf{X}^{(n)} = (X_1^n, \dots, X_n^n)$ comes from a *row independent* triangular array:

$$H_1^{(n)}: \quad X_i^n = (1 - U_i)Y_i + U_i Z_i^n, \quad i = 1, \dots, n, \tag{12}$$

where $Z_i^n = Z_i + \delta_n \varepsilon_i$, $(\varepsilon_i)_{i\geq 1}$ is an independent and identically distributed sequence of random variables, independent from the Z's and $\delta_n \to 0$ as $n \to +\infty$ (vanishing factor). We assume here that, $\forall i \geq 1$, $Z_i + \delta_n \varepsilon_i \notin \mathcal{S}$. In the Gaussian case this assumption is insured if the $\varepsilon' s$ are non Gaussian. It is also assumed that the $\mathbb{E}(e^{|\epsilon_1|}) < \infty$. This type of contiguous modeling looks natural to us as, in any experimental field, measurement errors could happen, represented above by the $\delta_n \varepsilon_i$'s, and additively impact the Z true underlying phenomenon. We also remind at this point that the distribution of the Y's is theoretically known by assumption.

The whole contiguous models collection will be denoted $H_1^* = \otimes_{n=1}^{\infty} H_1^{(n)}$. To emphasize the role of index n in the triangular array, we will denote all the estimators depending on $\mathbf{X}^{(n)}$ or any function depending on $G^{(n)}$, the cumulative distribution function of the $X_i^{(n)}$'s, with the extra superscript $^{(n)}$; for example, with this new notational rule, the estimator $\bar{p}_n(\mathbf{X}^{(n)})$ of p will be denoted $\bar{p}_n^{(n)}$. Similarly we will denote by $\widehat{g}_n^{(n)}$ the kernel density estimator of $g^{(n)}$ involved in the contiguous alternative setup, see Appendix D, defined by

$$\widehat{g}_n^{(n)}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i^n}{h_n}\right), \quad x \in \mathbb{R}, \tag{13}$$

where the bandwidth h_n satisfies $h_n \to 0$, $nh_n \to +\infty$ and K is a symmetric kernel density function detailed in Appendix E.1.

5.2. Fast contiguous alternatives

We first consider contiguous alternatives with a rate of convergence to the null hypothesis too fast to be detected by our testing procedure. Under the alternative $H_1^{(n)}$ we will consider the following assumptions:

(A4) The bandwidth setup is $h_n = n^{-1/4-\gamma}$, with $0 < \gamma < 1/12$.
(A5) The vanishing factor satisfies $\delta_n = n^{-\xi}$, with $\xi > \lambda + 2\gamma + 1/2$, where λ is given by (11).
(A6) There exists a nonnegative constant C such that for all $k \in \mathbb{N}$,

$$|\mathbb{E}\left(Q_k(X + \delta_n \varepsilon_1) - Q_k(X)\right)|/q_k^2 \leq C\delta_n,$$

where X is H_0 distributed.

Condition **(A6)** is checked in Lemmas 2-4 for the Gaussian and the Lebesgue reference measures. It is also satisfied for any reference measure with bounded support. For simplicity, we refer to condition **(A2-3)** under H_1^* in the proposition below. This means that both conditions are satisfied for all $n \geq 1$ replacing X_1 by X_1^n. Following the proof of these conditions in Appendix B under H_0 it is possible to establish explicit moment conditions on ε, adapted to the moments of Z, to insure **(A2-3)** under H_1^*. These conditions being technical and their proof being painful but straightforward, we do not detail them here.

Proposition 2. *If assumptions* **(A1-6)** *hold, then under* H_1^**:*

$$S_n \xrightarrow{P} 1 \quad \text{and} \quad T(n) \xrightarrow{\mathcal{L}} Z \quad , \quad \text{as } n \to +\infty,$$

where Z is a χ^2-distributed random variable with one degree of freedom.

5.3. Slow contiguous alternatives

In Assumption **(A5)** the convergence rate of δ_n to zero is too fast to distinguish the asymptotic null hypothesis when n tends to infinity. Contrarily, we now consider two convergence rates which are too slow to recover the asymptotic null distribution of the test statistic, despite the convergence of the contiguous alternative towards the null hypothesis. These convergence rates are given under the following assumptions:

(A7) $\mathbb{E}(\varepsilon) = 0$ and there exists $0 < \xi' < 1/4$ such that $\delta_n = n^{-\xi'}$.
(A8) $\mathbb{E}(\varepsilon) \neq 0$ and there exists $0 < \xi'' < 1/8$ such that $\delta_n = n^{-\xi''}$,

where ε denotes a generic random variable involved in the Z^n's definition. The rate in **(A7)** will control the mean deviation due to the perturbations ε and the rate given in **(A8)** will allow to control the variance of these perturbations when there is no mean deviation.

Proposition 3. *If assumption* **(A7)** *or* **(A8)** *holds, then under H_1^*:*

$$T(n) \xrightarrow{P} +\infty \quad , \quad as\ n \to +\infty.$$

Moreover, under **(A7)**, $S_n \xrightarrow{P} 1$ *when under* **(A8)**, $S_n \xrightarrow{P} 2$, *as $n \to +\infty$.*

Remark 5. *Le us notice that here is a gap between the rate $\xi > 1/2$ in* **(A5)** *and the rates $\xi' < 1/4$ and $\xi'' < 1/8$ in* **(A7-A8)**. *For now we are not able, given the various rates of convergence involved in our proofs, to establish the limiting distribution of $T(n)$ for a rate between $1/2$ and $1/4$.*

We provide, in Appendix L, a numerical illustration of our test sensitivity when applied to model (12) for a wide range of factors δ_n.

6. Choice of the reference measure and test construction

In order to run our test, we have to select now a reference measure ν and an *ad. hoc.* orthogonal family $\mathcal{Q} = \{Q_k,\ k \in \mathbb{N}\}$. The choice of the ν clearly relies on support of the X_i's. If the support is compact, one can choose a uniform distribution for ν and their associated Legendre polynomials. Since our numerical studies are dedicated to the Gaussian case, we illustrate here the choice of ν corresponding to two measures on the real line: the Gaussian and the Lebesgue one. The verification of conditions **(A2–3)** for these two measures is relegated in Appendix B.

Gaussian reference measure. In the present paper, we chose for ν the standard normal distribution to address the Gaussianity testing problem. This choice is adapted to any $\mathbb{R}$-supported probability distribution. The ν-orthogonal basis $\mathcal{Q}$ corresponds then to the collection of the $f_{(0,1)}$-orthogonal Hermite polynomials defined for all $k \geq 0$ by:

$$H_k(x) = k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m x^{k-2m}}{m!(k-2m)!2^m}, \quad x \in \mathbb{R}. \tag{14}$$

We have in particular $\|H_k\|^2 = k!$ and, for illustration purpose, the six first polynomials are:

$$H_0 = 1,\ H_1(x) = x,\ H_2(x) = x^2 - 1,\ H_3(x) = x^3 - 3x,$$
$$H_4(x) = x^4 - 6x^2 + 3,\ H_5(x) = x^5 - 10x^3 + 15x.$$

Lemma 1. *Let H_k be defined by (14) and let $Q_k(x) = H_k(x)$, for all $x \in \mathbb{R}$. Assume that we want to test $H_0 : f \in \mathcal{G}$, where $\mathcal{G}$ is the set of Gaussian densities. Then conditions* **(A2–3)** *are satisfied.*

Remark 6. *Lemma 1 can be extended to any non Gaussian null distribution f with known moments as discussed in Remark 7 in Appendix B*

Lemma 2. *Let H_k be defined by (14) and let $Q_k(x) = H_k(x)$, for all $x \in \mathbb{R}$. Then condition* **(A6)** *is satisfied.*

Lebesgue reference measure. Another possible choice for the reference measure ν is the Lebesgue measure over $\mathbb{R}$. In that case, we would rather consider the set of orthogonal Hermite functions defined by:

$$\mathcal{H}_k(x) = h_k(x)\exp(-x^2/2), \quad x \in \mathbb{R}, \tag{15}$$

where $h_k(x) = 2^{k/2}H_k(\sqrt{2}x)$, with H_k defined in (14). In that setup we have $\|\mathcal{H}_k\|^2 = k!2^k$.

Lemma 3. *Let $\mathcal{H}_k$ be defined by (15) and let $Q_k(x) = \mathcal{H}_k(x)$, for all $x \in \mathbb{R}$. Then conditions* **(A2–3)** *are satisfied.*

Lemma 4. *Let $\mathcal{H}_k$ be defined by (15) and let $Q_k(x) = \mathcal{H}_k(x)$, for all $x \in \mathbb{R}$. Then condition* **(A6)** *is satisfied.*

Test construction. The computation of the test statistic $T(n) = T_{S_n,n}$, see expressions (8) and (10), is grounded on the computation of the $\alpha_i(\mu, s)$ quantities. We detail here the expression of $R_{1,n}$ and $\text{var}(R_{1,n})$ when the reference measure is Gaussian associated with the Hermite polynomials. To overcome the complex dependence between the estimators $a_{1,n}$, $\bar{p}_n$, $\bar{\mu}_n$ and $\bar{s}_n$, we split the sample into four independent sub-samples of size n_1, n_2, n_3, n_4, with $n_1 + n_2 + n_3 + n_4 = n$. We use the first sample to estimate a_1, the second sample to estimate p, the third one to estimate μ, and the last one to estimate s. We get $\alpha_1(\mu, s) = \mu$ and $\alpha_{1,n} = \bar{\mu}_n$ which makes

$$R_{1,n} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i - \bar{p}_{n_2}\bar{\mu}_{n_3}, \quad \text{and}$$

$$\text{var}(R_{1,n}) = \frac{\text{var}(X)}{n_1} + \text{var}(\bar{p}_{n_2})\text{var}(\bar{\mu}_{n_3}) + \text{var}(\bar{p}_{n_2})\mathbb{E}(\bar{\mu}_{n_3})^2 + \mathbb{E}(\bar{p}_{n_2})^2\text{var}(\bar{\mu}_{n_3}).$$

We propose a consistent estimator of $\text{var}(R_{1,n})$:

$$V_{1,n} = S^2_{X,n_1} + v_{p,n_2}v_{\mu,n_3} + \bar{\mu}^2_{n_3}v_{p,n_2} + \bar{p}^2_{n_2}v_{\mu,n_3},$$

where S^2_{X,n_1} denotes the empirical variance based on $(X_1, \ldots, X_{n_1})$, and v_{p,n_2}, respectively v_{μ,n_3}, denotes the consistent estimator of $\text{var}(\bar{p}_{n_2})$, respectively $\text{var}(\bar{\mu}_{n_3})$), obtained from Bordes and Vandekerkhove [9]. The computation of the test statistic first requires the choice of $d(n)$, $e(n)$ and $s(n)$. A previous study showed us that the empirical levels and powers were overall weakly sensitive to $d(n)$ for $d(n)$ large enough. From that preliminary study we decided to set $d(n)$ equal to 10. The trimming $e(n)$ is calibrated equal to $(\log(n))^{-1}$. The normalization $s(n) = n^{\alpha-1}$ is setup close enough to $n^{-1/2}$, with α equal to 2/5, which seemed to provide good empirical levels.

Secondly, since the probability density functions considered in our set of simulation are $\mathbb{R}$-supported we use the standard Gaussian distribution for ν and its associated Hermite polynomials for $\mathcal{Q}$. All our simulations are based on 200 repetitions. Let us remind briefly that the empirical level is defined as the percentage of rejections under the null hypothesis and that the empirical power is the percentage of rejections under the alternative. Finally the asymptotic level is standardly fixed to 5%.

7. Semiparametric and maximum likelihood approaches comparison

In our testing procedure we estimate (p, μ) by using the semiparametric estimators proposed in Bordes and Vandekerkhove [9] instead of the maximum likelihood estimators. In the same way our estimation of θ, see expression (7), is H_0-free contrary to what would happen when using the maximum likelihood technique. Both approaches are asymptotically equivalent under the null hypothesis, see Remark 4, and all the simulations we did shown very similar empirical levels when comparing the semiparametric and maximum likelihood approaches under null models. However, under certain types of alternatives, the maximum likelihood approach can lead to very unexpected empirical powers. These behaviors are due to compensation phenomenon in models close, for example, to the non-identifiable one described in Section 4. To illustrate clearly this point we detail here the Gaussianity test in these cases. Write

$$g(x) = (1-p)f_{(0,1)}(x) + ph_{a,s}(x-\mu), \quad x \in \mathbb{R}, \tag{16}$$

where $h_{a,s}(x) = (f_{(0,s)}(x-a) + f_{(0,s)}(x+a))/2$, $a \neq 0$, $f_{(0,s)}$ being the Gaussian density, centered at zero, with variance s. When $\mu = \pm a$ and $s = 1$, we notice that (16) can be reformulated as

$$g(x) = (1 - \frac{p}{2})f_{(0,1)}(x) + \frac{p}{2}f_{(0,1)}(x - 2\mu), \quad x \in \mathbb{R}. \tag{17}$$

In this case there are two different parametrizations for (16): one that we call the *null parametrization*, coinciding with H_0 with null parameters $p_0 = p/2$, $\mu_0 = 2\mu$ and $s_0 = 1$, see the right hand side of (17). The other one is called the *alternative parametrization*, coinciding with H_1 with $p_1 = p$, $\mu_1 = \mu$ and $s_1 = \mu^2 + 1$, see the right hand side of (16). By construction the maximum likelihood estimator will favor the null parameters. We study now this phenomenon through a set of simulations where the parameters are $\mu = 4$, $s = 1$ and $p = 0.4$. For comparison, we use the same initial values for the both semiparametric and maximum likelihood algorithms, namely $(p, \mu, s) = (0.3, 6, 8.5)$, which is exactly between the null parametrization $(p, \mu, s) = (0.2, 8, 1)$, and the alternative parametrization $(p, \mu, s) = (0.4, 4, 17)$. It is of interest to study now the behavior of the semiparametric and maximum likelihood testing methods when the true model deviates smoothly from the null hypothesis in two ways:

i) the unknown component is a $h_{a,1}$ with $\mu \neq a$, *i.e*

$$g(x) = (1-p)f_{(0,1)}(x) + p \underbrace{\left(\frac{1}{2}f_{(0,1)}(x-a-\mu) + \frac{1}{2}f_{(0,1)}(x+a-\mu)\right)}_{\mu\text{-symmetric mixture detected by the semiparametric method}}$$
$$= \left((1-p)f_{(0,1)}(x) + \frac{p}{2}f_{(0,1)}(x+a-\mu)\right) + \frac{p}{2}f_{(0,1)}(x-a-\mu)$$
$$\approx \left(1-\frac{p}{2}\right)f_{(0,1)}(x) + \frac{p}{2}\underbrace{f_{(0,1)}(x-a-\mu)}_{(a+\mu)\text{-centered Gaussian attracting the maximum likelihood method}}, \quad \text{when } \mu \to a,$$

this case will be called the *mean deviation trap*, and ii) the unknown component is a $h_{a,s}$ with $\mu = a$ but $s \neq 1$, *i.e.*

$$g(x) = (1-p)f_{(0,1)}(x) + p \underbrace{\left(\frac{1}{2}f_{(0,s)}(x-2\mu) + \frac{1}{2}f_{(0,s)}(x)\right)}_{\mu\text{-symmetric mixture detected by the semiparametric method}}$$
$$= \left((1-p)f_{(0,1)}(x) + \frac{p}{2}f_{(0,s)}(x)\right) + \frac{p}{2}f_{(0,1)}(x-2\mu)$$
$$\approx \left(1-\frac{p}{2}\right)f_{(0,1)}(x)) + \frac{p}{2}\underbrace{f_{(0,s)}(x-2\mu)}_{(2\mu)\text{-centered Gaussian attracting the maximum likelihood method}}, \quad \text{when } s \to 1$$

this case will be called the *variance deviation trap*.

It is very important to point out now that the above phenomenons illustrate the risk of considering only one single Gaussian component ($q = 1$) in the generic mixture model defining f in McLachlan *et al.* [21] (Section 6) when actually two Gaussian components ($q = 2$) would be necessary to accurately fit the model.

Mean deviation trap. We consider deviations from the null model obtained by considering $\mu = 3, 2, 1$ and $s = 1$. Figure 1 shows the g probability density function under these respective alternatives. It can be observed that, if we try to visually detect a mixture of two Gaussian distributions, the probability density function of the left-side component moves clearly aside the Gaussian distribution family as μ moves largely away from $a = 4$, *i.e.* when $\mu = 1$, but we bet that many practitionners would probably vote "intuitively" for a mixture of two Gaussian distributions when $\mu = 3$ or 2. Figure 10 in Appendix H illustrates the difficulty of the maximum likelihood estimator to recognize the alternative model when the mean deviation is not distant enough (here $\mu = 3$ and $a = 4$). Based on a run of 200 repetitions, it is shown that the maximum likelihood estimation is trapped at the null parametrization which namely is $(p, \mu, s) = (0.2, 7, 1)$ when on the opposite, the semiparametric estimation detects the correct $(p, \mu, s) = (0.4, 3, 17)$ alternative parametrization. In Fig. 2 we display respectively the empirical power of our testing procedure based on the

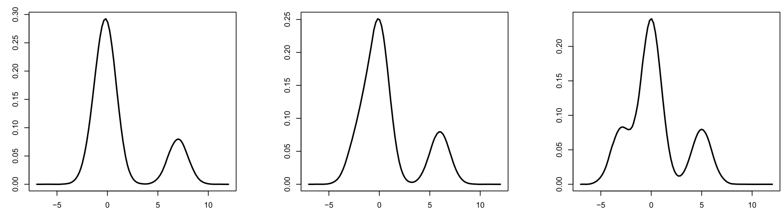

FIG 1. *The probability density function g in model (16) when $a = 4$, $s = 1$, and $\mu = 3, 2, 1$.*

maximum likelihood and the semiparametric approach for $\mu = 3, 2, 1$, $a = 4$, $s = 1$, and for $n = 1000, 2000, 5000$. As expected the maximum likelihood approach barely detects the alternative for small values of n when its semiparametric counterpart surpasses it with up to 10 times more correct decision results. The reason of this lack of power is due to the fact that our test focuses more on the moments of the second components than those of the first one and, as seen in Fig. 1, the second components looks pretty much Gaussian even for $\mu = 1$.

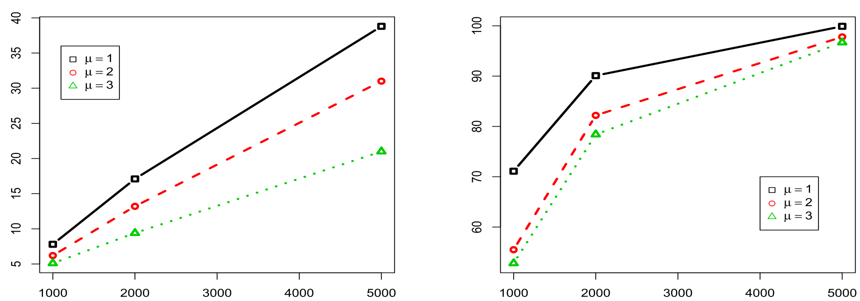

FIG 2. *Empirical powers obtained with the maximum likelihood approach (left) and semiparametric approach (right) under the mean deviation trap effect for $\mu = 3, 2, 1$ and $a = 4$.*

Variance deviation trap. We consider the variance deviations $s = 2, 3, 4$, fixing $\mu = a = 4$. Figure 3 shows the g probability density function under these alternatives. The corresponding empirical powers are displayed in Fig. 4. We

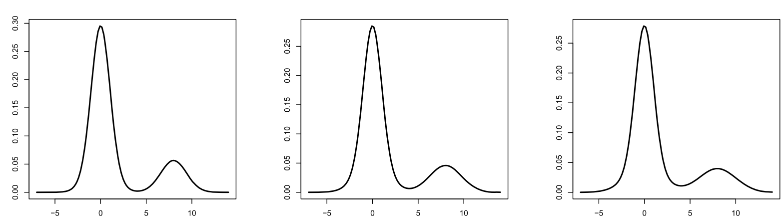

FIG 3. *The probability density function g in model (16) with $\mu = a = 4$ and $s = 2, 3, 4$.*

can observe that both powers associated with the maximum likelihood and semiparametric approach increase according to the variance deviation but it is worth to notice that the detection based on the maximum likelihood approach is again very poor compared to the semiparametric approach. As a conclusion,

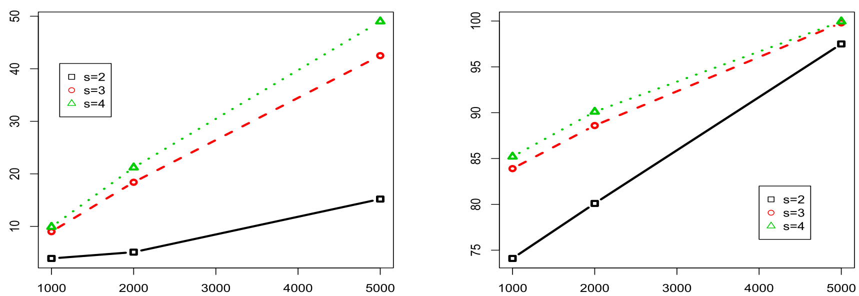

FIG 4. *Empirical powers obtained with the maximum likelihood approach (left) and semiparametric approach (right) under the variance deviation trap effetct for $s = 2, 3, 4 \neq 1$, with $\mu = a = 4$.*

this set of numerical experiments shows the clear interest, in terms of testing power, of considering the semiparametric versus the maximum likelihood approach especially in a close neighborhood of non-identifiable type (1) Gaussian models.

8. Simulations: empirical levels and powers

In all our simulation we consider the case where f_0 is a known Gaussian density, centered at zero and variance equal to one. The unknown density will be parameterized by its mean μ and its variance $\theta = s$. We chose as orthogonal basis the family of Hermite polynomials.

8.1. Empirical levels

McLachlan *et al.* [21] considered the two-component Gaussian version of the mixture model (1) through three datasets arising from the bioinformatics literature: the breast cancer data, with $n = 3226$, the colon cancer data, with $n = 2000$, and the HIV data, with $n = 7568$. The estimation of their associated parameters are respectively: $(\widehat{p}_n, \widehat{\mu}_n, \widehat{s}_n) = (0.36, 1.52, 0.99)$, $(0.58, 1.61, 2.08)$, and $(0.98, -0.15, 0.79)$. To make sure that our methodology will have reliable behaviors when applied on this collection of datasets, we investigate the empirical levels of our testing procedure across parameter values such as $n \in \{2000, 3000, 7500\}$ and $(p, \mu, s) = (1/3, 1.5, 1)$, $(0.5, 1.5, 2)$ and $(0.98, -0.15, 0.8)$ which are values in the range of the above targeted applications. For this purpose, for each value of n, p, μ and s, we compute the test statistic $T(n)$ based on the sample and compare it to the 5%-critical value of its approximated distribution under H_0 ($\chi^2(1)$ according to Corollary 3). Note that, for numerical simplicity, we initialize our parameter estimation step at the true value of the Euclidean parameter. The collection of empirical levels obtained for this set of simulated examples is reported in Fig. 11 of Appendix I. It appears that a significant number of observations is needed to get close to the theoretical level. This drawback can be balanced by the fact that today, as mentioned in the Introduction, genomic datasets usually contain thousands of genes which makes our methodology in practice suitable for a wide class of standard (from the sample size view point) microarray analysis problems.

8.2. Empirical powers

In this section we consider the Gaussian testing problem (3) with $\mathcal{F} = \mathcal{G}$ where $\mathcal{G} = \left\{ f_{(\mu,s)}; (\mu, s) \in \Lambda \subset \mathbb{R} \times \mathbb{R}^{+*} \right\}$ denotes the set of Gaussian densities with mean μ and variance s, compared to Student and Laplace alternatives. First a 1-shifted Student distribution $t(3)$, having a shape far enough from the Gaussian distribution, with a shift $\mu = 1$. Second a shifted Student $t(10)$, again with a shift equal to 1, but having a shape closer to the null Gaussian distribution. Third a Laplace distribution $\mathcal{L}(1, 1)$ with mean 1 and variance 2. The last alternative is a Laplace $\mathcal{L}(1, 2)$ with mean 1 and variance 8. The empirical powers for Student and Laplace alternatives are respectively summarized in Fig. 5 and 6. We also compare these empirical powers with those obtained by the maximum likelihood approach in Appendix K.

As expected, when comparing pairwise the Student alternatives, the power is greater for the $t(3)$ distribution compared to the $t(10)$ distribution. The $t(3)$ is very clearly detected by the test since the detection level is greater than 80% for all the cases and even close to 100% for $n = 7000$. Now, similarly to the mean and variance deviation trap setups investigated in Section 7, we can observe that the power is greater as p increases, which practically means that the Student component is enhanced in the model (remind that our test procedure is focused on the 2nd-component moments analysis). We display the mixture densities corresponding to this set of alternatives in Fig. 12 of Appendix J. For the first Student alternative, comparing $p = 1/2$ and $p = 0.98$, we can observe that a serious jump happens in terms of dissimilarity between the alternative model and the *best fitted* (same mean and variance) Gaussian null-model. For $p = 0.98$, the Student distribution strongly prevails and the test is automatically empowered. The second alternative is also detected, but with a lower power, let say between 40 % and 90%, due to the proximity of the Student $t(10)$ with the Gaussian $\mathcal{N}(0, 1)$.

In Fig. 12 of Appendix J we can see how close the null distribution and the $t(10)$ alternative are, especially for $p = 1/3$ and $p = 1/2$, and visually evaluate how challenging these testing problems really are.

The empirical powers for Laplace alternatives are given in Fig. 6. The power is larger with the alternative $\mathcal{L}(1, 2)$ than with the alternative $\mathcal{L}(1, 1)$. Indeed the $\mathcal{L}(1, 2)$ distribution has a stronger shape departure from the Gaussian than the $\mathcal{L}(1, 1)$, and the associated mixture densities inherit these characteristics as

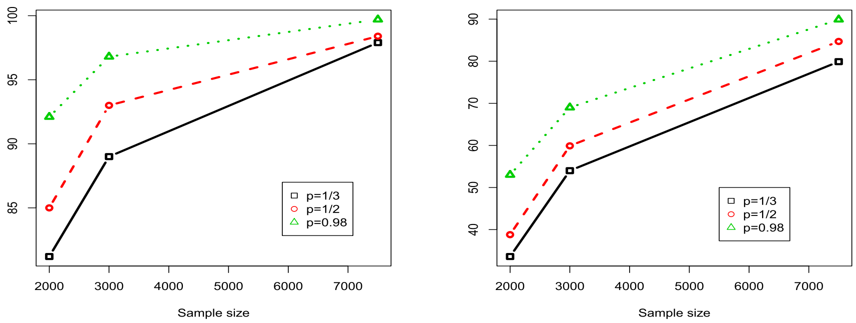

FIG 5. *Respectively left and right: the empirical powers when the alternative is a shifted Student $t(3)$, resp. a shifted Student $t(10)$, for parameter values $p = 1/3$ ($\square$), $p = 1/2$ ($\circ$) and $p = 0.98$ ($\triangle$) with sample sizes $n = 2000, 3000, 7500$.*

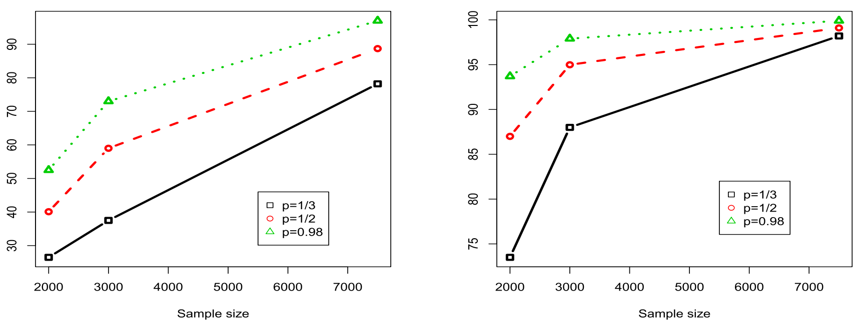

FIG 6. *Respectively left and right: the empirical powers when the alternative is a Laplace $\mathcal{L}(1,1)$, resp. a Laplace $\mathcal{L}(1,2)$, for parameter values $p = 1/3$ ($\square$), $p = 1/2$ ($\circ$) and $p = 0.98$ ($\triangle$) with sample sizes $n = 2000, 3000, 7500$.*

we can see in Fig. 12 of Appendix J. These alternatives are globally very well detected by our method and the power increases strongly when p gets closer to 1 (see Fig. 6 curve in green).

9. Real datasets

Microarray data We consider 3 datasets arising from the bioinformatics literature and studied in McLachlan *et al.* [21]. Figure 7 shows the non parametric kernel estimations of their probability density functions. Each of them deals with genes expressions modeled by the two-component mixture model (1) in which f was arbitrarily, for simplicity matters, considered as Gaussian (without any theoretical justification). The goal of this section is to answer if the classical Gaussian assumption was a posteriori correct or not.

<u>Breast cancer data</u>. We consider the breast cancer data studied in Hedenfalk *et al.* [14]. It consists in $n = 3226$ gene expressions in breast cancer tissues

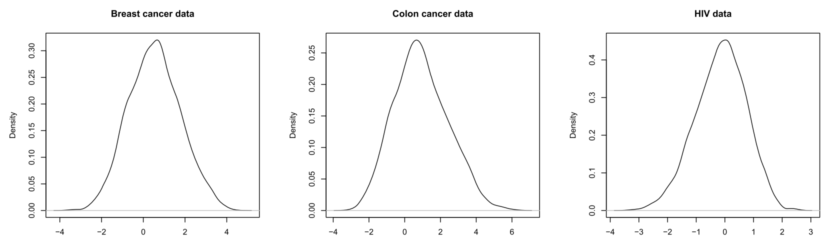

Fig 7. *Respectively the kernel density estimators of the breast data, colon data and* HIV *data distributions.*

from women with BRCA1 or BRCA2 gene mutations. The maximum likelihood parameter estimations under the Gaussian null model are $\widehat{p}_n = 0.36$, $\widehat{\mu}_n = 1.53$, $\widehat{s}_n = 0.98$. By the semiparametric method we obtain $\bar{p}_n = 0.41$, $\bar{\mu}_n = 1.35$ and $\bar{s} = 1.31$. It can be noticed here that nonparametric and maximum likelihood estimators give pretty similar results here which may corroborate the null hypothesis. Our test procedure provides a p-value equal to 0.82, with $S_n = 1$. As a consequence the normality of the second mixture component under H_0 cannot be rejected.

Colon cancer data. We consider the colon cancer data analysed in Alon *et al.* [2]. The samples comes from colon cancer tissues and normal colon tissues. It contains $n = 2000$ expressions of genes. The maximum likelihood estimations of the parameters are $\widehat{p}_n = 0.58$, $\widehat{\mu}_n = 1.61$, $\widehat{s}_n = 2.08$; The semiparametric method provides $\bar{p}_n = 0.72$, $\bar{\mu}_n = 1.28$ and $\bar{s} = 2.33$. By using our testing procedure we obtain a p-value less than 10^{-8} with $S_n = 4$. Here we clearly reject the normality under H_0. The rejection of the Gaussian mixture can be explained here by the fact that the nonparametric and the maximum likelihood estimators lead to notably different values especially on p.

HIV data. We consider the HIV dataset of vant' Wout *et al.* [36]. It contains expression levels of $n = 7680$ genes in CD4-T-cell lines, after infection with the HIV-1 virus. The maximum likelihood estimations of the parameters are $\widehat{p}_n = 0.98$, $\widehat{\mu}_n = -0.15$, $\widehat{s}_n = 0.79$. The semiparametric method provides $\bar{p}_n = 0.99$, $\bar{\mu}_n = 0.20$ and $\bar{s} = 0.80$. The p-value given by our testing procedure is equal to 0.64, associated with the decision $S_n = 1$. As a consequence the normality under H_0 cannot be rejected despite the fact that the maximum likelihood and semiparametric estimations of μ are quite different but both close to 0, meaning a strong overlap of the mixed distributions (see the almost symmetry of the third probability density function in Fig. 7).

Galaxy data We consider here the Carina dataset, see Walker *et al.* [35], studied previously in Patra and Sen [26]. Carina is a low luminosity galaxy companion of the Milky Way. The data collects $n = 1266$ measurements of the radial velocity of stars in Carina. This is a contamination model in the sense

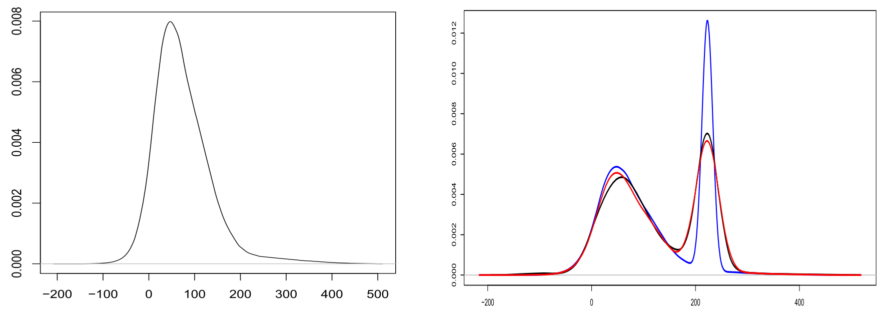

FIG 8. *Left side: estimated density of the Milky Way Radial velocities. Right side: in black, the plot of the Carina dataset nonparametric density estimate. In red, resp. in blue, the plot of the model (1) probability density function under $f = f_{(\mu,s)}$ and obtained by plugging $(\bar{p}_n, \bar{\mu}_n, \bar{s}_n)$, resp. $(\widetilde{p}_n, \widetilde{\mu}_n, \widetilde{s}_n)$, into (p, μ, s).*

that the measurements of stars in the Milky Way are mixed with some of Carina (overlapping). The Milky Way is largely observed, see Robin *et al.* [29]. Figure 8 shows the density f_0 of the radial velocity of Milky Way, estimated over $n' = 170,601$ observations. This density is clearly not zero-symmetric but in such a case it is enough to refer to the tail-oriented set of identifiability conditions of Proposition 3 i) in Bordes et al. (2006) to make the semiparametric estimation method still valid. Note also that the asymptotic results of Bordes and Vandekerkhove [9] still hold if the cumulative distribution function F_0 is replaced by a smooth empirical estimate $\widetilde{F}_{0,n'}$ based on a $n' = \varphi(n)$ sized training data provided with $n/n' \to 0$ as $n \to +\infty$. Unfortunately the study of the maximum likelihood estimate, see Section 5 of Appendix F, cannot be generalized straightforwardly since the non-parametric estimation of the Kullback distance, obtained by replacing f_0 by a kernel density estimate $\widehat{f}_{0,n}$ in the log-likelihood, is known to be very a delicate problem, see Berrett *et al.* [7] and references therein. Though, the fact that the unknown component of g under H_0 is supposed to have a parametric form should definitely help to control some technical tail issues specific to the Kullback estimation. We obtained for p and μ, respectively the proportion and the mean of the Carina radial velocity, the following estimations:

$$\bar{p}_n = 0.361 \quad \text{and} \quad \bar{\mu}_n = 222.60.$$

In their study, Patra and Sen (2016) obtained very similar values: $\widetilde{p} = 0.323$ and $\widetilde{\mu} = 222.9$. However, the estimation of the variance s appears to be highly sensitive to the estimation of p. Using the plug-in estimator given by (7) we get $\bar{s}_n = 453.93$. Note that the estimation given in Patra and Sen [26] was $\widetilde{s}_n = 56.4$ which looks far from the expected value given the data. To illustrate this remark, we compare in Fig. 8 the kernel density estimate of the observed data with the probability density of model (1), obtained by replacing (p, μ, s) by our estimates $(\bar{p}_n, \bar{\mu}_n, \bar{s}_n)$ and the Patra and Sen [26]'s estimates $(\widetilde{p}_n, \widetilde{\mu}_n, \widetilde{s}_n)$. We can observe that our estimation provides an excellent fitting when the variance estimated by Patra and Sen [26] appears to be way too small. Our test procedure yields a p-value equal to 0.75 with a test statistic $T_{S_n,n} = T_1 = 0.097$. As a consquence, there is no evidence here to reject the normality of the Carina radial velocity.

10. Discussion and perspectives

In this paper we proposed an H_0-free testing procedure to deal with the delicate problem of the contamination model parametrization. In our numerical study we focused our attention on the Gaussianity testing problem however it is very important to remind that our asymptotic results can be generalized to any suitable distribution (possibly non-symmetric). Indeed, if the unknown distribution of model (1) is embedded in a nonparametric envelop $\mathcal{S}$ provided with identifiability constraints and if there exists a corresponding semiparametric $\sqrt{n}$-consistent method, then the asymptotic results in Sections 3-4 extends straightforwardly. For this latter case, we recommend the recent work by Al Mohamad and Boumahdaf [3] who consider in model (1) an unknown component defined through linear constraints. In their paper, the authors derive an original consistent and asymptotically normally distributed semiparametric estimation method with asymptotic closed form variance expressions. Indeed, when considering null assumptions different from the Gaussian case, basically only the shape parameter estimation, usually deduced from moment equations, and the choice of the orthogonal basis described in Section 2 could possibly change, depending on the support of the tested distribution. Wavelet functions and Laguerre polynomials could respectively be used for probability density functions on the whole, respectively positive, real line, when Legendre, or cosine bases could be used for densities with compact support. Also, with a slight adaptation of our work, we could definitely test the unknown component of the contamination model considered in the recent work by Ma and Yao [20] where the first component density is only supposed to belong to a parametric family (the first component is not entirely known anymore). For each case, the use of the maximum likelihood or semiparametric approach could be again discussed. On the other hand, as it has been demonstrated in Section 7, see Figs. 2 and 4, the semiparametric testing approach shows better power performances than the maximum likelihood version especially in the neighborhood of the mean and variance deviation trap situations (up to 10 times more efficient for small sample sizes). We also proposed in Section 5 a vanishing convolution-class of nonparametric contiguous alternatives and studied theoretically their detectability under certain convergence rate conditions. In a futur work it would be very interesting to address the contiguous detection problem associated with the mean and variance deviation trap setups. This would namely consist in looking at the asymptotic behavior of our test when replacing respectively the parameters μ and s in the mean and variance deviation trap setups by sequences μ_n and s_n converging respectively towards a and 1 as n goes to infinity. The major technical difficulty here is that we are not able to establish yet optimal bounds of convergence for the semiparametric Euclidean estimator associated with a triangular array driven by the above asymptotic parametrization, see Remark 10 in Appendix E.2. Future work is also to consider a K-sample extension, $K \geq 2$, in the spirit of Wylupek [37], Ghattas *et al.* [13], or more recently Doukhan *et al.* [11]. More precisely, we could test the equality of K unknown components through K observed mixture models.

Appendix A: Proofs of the main results

Theorem 2. Let us prove that $\mathbb{P}(S_n \geq 2)$ vanishes as $n \to +\infty$. By definition of S_n in (10) and $\widehat{D}_{k,n}[\cdot]$ in (9) we have for all $\lambda \in]0, 1/2[$:

$$\begin{aligned}
&\mathbb{P}(S_n \geq 2) \\
&= \mathbb{P}\Big(\exists\, k \in \{2, \dots, d(n)\} : \\
&\qquad n^\lambda U_{k,n}^\top \widehat{D}_{k,n}^{-1} U_{k,n} - k \log(n) \geq n^\lambda U_{1,n}^\top \widehat{D}_{1,n}^{-1} U_{1,n} - \log(n)\Big) \\
&\leq \mathbb{P}\left(\exists\, k \in \{2, \dots, d(n)\} :\ n^\lambda U_{k,n}^\top \widehat{D}_{k,n}^{-1} U_{k,n} \geq (k-1)\log(n)\right) \\
&\leq \mathbb{P}\left(\exists\, k \in \{2, \dots, d(n)\} :\ \sum_{j=2}^{k} n^\lambda (R_{j,n})^2 \geq (k-1)\log(n) e(n)\right) \\
&\leq \mathbb{P}\left(\exists\, (j,k) \text{ with } 2 \leq j \leq k \leq d(n) :\ n^\lambda (R_{j,n})^2 \geq \log(n) e(n)\right) \\
&\leq \mathbb{P}\left(\sum_{j=2}^{d(n)} n^\lambda (R_{j,n})^2 \geq \log(n) e(n)\right).
\end{aligned} \tag{18}$$

It is important for us to keep the summation term up to $d(n)$ in the left hand side of the above inequality-type event in order to straightforwardly use the almost sure rate of convergence of the semiparametric Euclidean parameters, see (22)–(23). We decompose $R_{k,n}$ as follows:

$$R_{k,n} = (a_{k,n} - \mathbb{E}(a_{k,n})) - (\bar{p}_n \alpha_{k,n} - p_0 \alpha_k(\mu_0, \theta_0)), \quad 1 \leq k \leq d(n). \tag{19}$$

By using the inequality $(a+b)^2 \leq 2(a^2+b^2)$, for all $(a,b) \in \mathbb{R}^2$, we get

$$\begin{aligned}
&\mathbb{P}\left(\sum_{k=2}^{d(n)} n^\lambda (R_{k,n})^2 \geq \log(n) e(n)\right) \\
&\quad \leq \mathbb{P}\left(\sum_{k=2}^{d(n)} (a_{k,n} - \mathbb{E}(a_{k,n}))^2 \geq \frac{\log(n) e(n)}{4n^\lambda}\right) \\
&\quad + \mathbb{P}\left(\sum_{k=2}^{d(n)} (\bar{p}_n \alpha_{k,n} - p_0 \alpha_k(\mu_0, \theta_0))^2 \geq \frac{\log(n) e(n)}{4n^\lambda}\right).
\end{aligned} \tag{20}$$

We study now all the above quantities separately. By the Markov inequality, we first have

$$\begin{aligned}
\mathbb{P}&\left(\sum_{k=2}^{d(n)}(a_{k,n}-\mathbb{E}(a_{k,n}))^2 \geq \frac{\log(n)e(n)}{4n^\lambda}\right)\\
&\leq \frac{4n^\lambda}{\log(n)e(n)}\sum_{k=2}^{d(n)}\mathbb{E}\left((a_{k,n}-\mathbb{E}(a_{k,n}))^2\right)\\
&= \frac{4n^\lambda}{\log(n)e(n)}\sum_{k=2}^{d(n)}\frac{1}{n}\mathrm{var}\left(\frac{Q_k(X_1)}{q_k^2}\right)\\
&\leq \frac{4d(n)}{n^{1-\lambda}\log(n)e(n)}M_3, \qquad (21)
\end{aligned}$$

where the right hand side term goes to zero as $n\to+\infty$ since $d(n)/\log(n)e(n) = O(1)$ according to **(A1)** and (11).

Secondly, by decomposing $\bar{p}_n\alpha_{k,n}-p_0\alpha_k(\mu_0,\theta_0) = (\bar{p}_n-p_0)\alpha_{k,n}+p_0(\alpha_{k,n}-\alpha_k(\mu_0,\theta_0))$, we obtain the following majorization

$$\begin{aligned}
\mathbb{P}&\left(\sum_{k=2}^{d(n)}(\bar{p}_n\alpha_{k,n}-p_0\alpha_k(\mu_0,\theta_0))^2 \geq \frac{\log(n)e(n)}{4n^\lambda}\right)\\
&\leq \mathbb{P}\left(\sum_{k=2}^{d(n)}(\alpha_{k,n})^2(\bar{p}_n-p_0)^2 \geq \frac{\log(n)e(n)}{8n^\lambda}\right)\\
&\quad+\mathbb{P}\left(\sum_{k=2}^{d(n)}p_0^2(\alpha_{k,n}-\alpha_k(\theta_0,\mu_0))^2 \geq \frac{\log(n)e(n)}{8n^\lambda}\right).
\end{aligned}$$

Since the $\alpha_{k,n}$'s are bounded by M_1 according to **(A2)**, we have

$$\mathbb{P}\left(\sum_{k=2}^{d(n)}\alpha_{k,n}^2(\bar{p}_n-p_0)^2 \geq \frac{\log(n)e(n)}{8n^\lambda}\right) \leq \mathbb{P}\left((\bar{p}_n-p_0)^2 \geq \frac{\log(n)e(n)}{8n^\lambda M_1 d(n)}\right), \quad (22)$$

where the last right hand side term goes to zero as $n\to+\infty$ since $\lambda\in\,]0,1/2[$ and $|\bar{p}_n-p_0|^2 = o_{a.s.}(n^{-1/2+\alpha})$ for all $\alpha>0$, by Bordes and Vandekerkhove [9]. By denoting $\rho_0=(\mu_0,\theta_0)$ and $\bar{\rho}_n=(\bar{\mu}_n,\bar{\theta}_n)$, we also have $\|\bar{\rho}_n-\rho_0\|^2 = o_{a.s.}(n^{-1/2+\alpha})$, for all $\alpha>0$. Since the $\dot{\alpha}_{k,n}$'s are bounded by M_2 according to **(A2)**, using the *mean value* theorem we obtain:

$$\mathbb{P}\left(\sum_{k=2}^{d(n)}(\alpha_{k,n}-\alpha_k(\mu_0,\theta_0))^2 \geq \frac{\log(n)e(n)}{8n^\lambda}\right) \leq \mathbb{P}\left(\|\bar{\rho}_n-\rho_0\|^2 \geq \frac{\log(n)e(n)}{8n^\lambda M_2^2 d(n)}\right) \quad (23)$$

which right hand side goes to zero as $n\to+\infty$. Hence from (18) and the controls in probability (20–23), we obtain that $\mathbb{P}(S_n\geq 2)\to 0$ as $n\to+\infty$. □

Corollary 3. From Theorem 2, $T_{S_n,n}$ has the same limiting distribution as $T_{1,n} = nR_{1,n}^2/V_{1,n}$. Since the estimators $\bar{\theta}_n$ and $\bar{\mu}_n$ are independent and asymptotically Normally distributed towards the true values θ_0 and μ_0 we get, by using the delta method, the following convergence in distribution:

$$\sqrt{n}\alpha_1(\bar{\mu}_n, \bar{\theta}_n) \xrightarrow{\mathcal{L}} \mathcal{N}(\alpha_1(\mu_0, \theta_0), D(\mu_0, \theta_0)VD(\mu_0, \theta_0), \quad \text{as} \quad n \to +\infty,$$

where $D(\cdot, \cdot)$ is the gradient $\dot{\alpha}_1(\cdot, \cdot)$, and where V is the asymptotic variance of $(\sqrt{n}\bar{\mu}_n, \sqrt{n}\bar{\theta}_n)$. Combining this convergence in law with the following convergence in probability:

$$V_{1,n} \xrightarrow{P} \operatorname{var}(R_{1,n}) \quad \text{and} \quad \bar{p}_n \xrightarrow{P} p_0, \quad \text{as} \quad n \to +\infty,$$

along with the independence and the asymptotic normality of the first estimated coefficient $a_{1,n} = \sum_{i=1}^{n} Q_1(X_i)/nq_1^2$, we get, by using the Slutsky's Theorem, the following limiting distribution:

$$\sqrt{n}\frac{R_{1,n}}{\sqrt{V_{1,n}}} = \sqrt{\frac{n}{V_{1,n}}}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{Q_1(X_i)}{q_1^2} - \bar{p}_n\bar{\mu}_n\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \quad \text{as} \quad n \to +\infty,$$

which concludes the proof. □

Proposition 1. The advantage of considering the semiparametric approach in Bordes and Vandekerkhove [9] versus the maximum likelihood method is that under H_1 we keep the following consistency results in probability:

$$\bar{\vartheta}_n = (\bar{p}_n, \bar{\mu}_n) \xrightarrow{P} (p_0, \mu_0),\ \bar{\theta}_n \xrightarrow{P} \theta_0,\ R_i \xrightarrow{P} r_i = \mathbb{E}(Q_i(X)/q_i^2) - p_0\alpha_i(\mu_0, \theta_0),$$

as $n \to +\infty$, for $i \geq 1$, along with their associated asymptotic normality. As a consequence, by using the Slutsky's Theorem, the terms $\sqrt{n}(R_{i,n} - r_i)/\sqrt{\widehat{D}_{k,n}[i]}$, $1 \leq i \leq k$, are asymptotically normally distributed since $\widehat{D}_{k,n}[i]$ is a weakly consistent estimator of $\operatorname{var}(R_i)$. Now, Clearly by (1) (with $b_i = 0$), $\mathbb{E}(Q_i(X)) = p_0\mathbb{E}(Q_i(Y))$, where Y is a f-distributed random variable. Then we have the following equivalence

$$r_i = 0, \quad \text{for all} \ \ i \geq 1 \quad \Longleftrightarrow \quad \mathbb{E}(Q_i(Y)/q_i^2) = \alpha_i(\mu_0, \theta_0), \quad \text{for all} \ \ i \geq 1.$$

This condition implies that the expansion of the Ys' density matches with the expansion of the unknown density f with mean μ_0 and parameter θ_0, which is in contradiction with the semiparametric identifiability of model/setup H_1, see Bordes *et al.* [8]. Thus we can state that there exists an index j such that $r_j \neq 0$. For simplicity matters let us consider $j_0 = \min\{j \geq 1 :\ r_j \neq 0\}$. Since from (8), for every $k \geq 1$ fixed, we can decompose $T_{k,n}$ as follows:

$$\begin{aligned} s(n)T_{k,n} &= n^\lambda U_{k,n}^\top \widehat{D}_{k,n}^{-1} U_{k,n} \\ &= n^{\lambda-1} \sum_{\ell=1}^{k} \left(\sqrt{n} \left[\frac{R_{\ell,n} - r_\ell}{\sqrt{\widehat{D}_{k,n}[\ell]}} \right] \right)^2 + 2n^{\lambda-1/2} \sum_{\ell=1}^{k} \sqrt{n} \left[\frac{R_{\ell,n} - r_\ell}{\sqrt{\widehat{D}_{k,n}[\ell]}} \right] r_\ell \\ &\quad + n^\lambda \sum_{\ell=1}^{k} r_\ell^2, \end{aligned}$$

it comes that for all $k < j_0$, $T_{k,n} = O_p(n^{\lambda-1})$ since the r_ℓ's are all equal to zero for $1 \leq \ell \leq k$, when instead for the index j_0 we have $T_{j_0,n} \geq n^\lambda r_{j_0}^2 + O_p(n^{\lambda-1/2})$. It comes that for all $k < j_0$ we have

$$\mathbb{P}\left(s(n)T_{k,n} - \beta_k pen(n) < s(n)T_{j_0,n} - \beta_{j_0} pen(n)\right) \longrightarrow 1, \quad \text{as} \quad n \to +\infty.$$

This obviously shows, according to S_ns' definition (10), that $S_n \geq j_0$ with probability one as $n \to +\infty$. Now, since $T_{k,n}$ is a k-increasing sequence for every given $n \geq 1$, we have that $T_{S_n,n} \geq T_{j_0,n} \geq n^\lambda r_{j_0}^2 + O_p(n^{\lambda-1/2})$ which proves the wanted result. Note that the right hand side of the previous inequality shows clearly a drift of our test statistic in $O_P(n^\lambda)$, $0 < \lambda < 1/2$, under the alternative H_1. □

Proposition 2. Similarly to the proof of Theorem 2, we have

$$\mathbb{P}\left(S_n^{(n)} \geq 2\right) \leq \mathbb{P}\left(\sum_{k=2}^{d(n)} n^\lambda \left(R_{k,n}^{(n)}\right)^2 \geq \log(n)e(n)\right). \tag{24}$$

To prove that the right hand side term of the above probability goes to zero as $n \to +\infty$, we decompose $R_{k,n}^{(n)}$ as follows:

$$R_{k,n}^{(n)} = \left(a_{k,n}^{(n)} - \mathbb{E}(a_{k,n}^{(n)})\right) - \left(\bar{p}_n^{(n)} \alpha_{k,n}^{(n)} - p_0 \alpha_k(\mu_0, \theta_0)\right) + \psi_{k,n}, \tag{25}$$

with $\alpha_{k,n}^{(n)} = \alpha_k(\bar{\mu}_n^{(n)}, \bar{\theta}_n^{(n)})$, and

$$\psi_{k,n} = p_0 \mathbb{E}\left(Q_k(X_0 + \delta_n \varepsilon_1) - Q_k(X_0)\right)/q_k^2, \tag{26}$$

which denotes the expectation of the k-th difference between the $H_1^{(n)}$ and H_0-distribution type supported by the second component in the mixture model (1), X_0 being H_0 distributed. By **(A6)** there exists $c > 0$ such that

$$\psi_{k,n}^2 \leq c\delta_n^2. \tag{27}$$

We then have

$$\begin{aligned} &\mathbb{P}(S_n^{(n)} \geq 2) \\ &= \mathbb{P}\Bigg(n^\lambda \sum_{k=2}^{d(n)} \left((a_{k,n}^{(n)} - \mathbb{E}(a_{k,n}^{(n)})) - (\bar{p}_n^{(n)} \alpha_{k,n}^{(n)} - p_0 \alpha_k(\mu_0, \theta_0)) + \psi_{k,n} \right)^2 \end{aligned}$$

$$
\begin{aligned}
&\geq \log(n)e(n)\Bigg)\\
&\leq \mathbb{P}\Bigg(n^{\lambda}\sum_{k=2}^{d(n)}\Big(((a_{k,n}^{(n)}-\mathbb{E}(a_{k,n}^{(n)}))-(\bar{p}_n^{(n)}\alpha_{k,n}^{(n)}-p_0\alpha_k(\mu_0,\theta_0)))^2+\psi_{k,n}^2\Big)\\
&\qquad\geq \log(n)e(n)/2\Bigg)\\
&\leq \mathbb{P}\Bigg(n^{\lambda}\sum_{k=2}^{d(n)}\Big((a_{k,n}^{(n)}-\mathbb{E}(a_{k,n}^{(n)}))-(\bar{p}_n^{(n)}\alpha_{k,n}^{(n)}-p_0\alpha_k(\mu_0,\theta_0))\Big)^2\\
&\qquad\geq \log(n)e(n)/2-cn^{\lambda}d(n)\delta_n^2\Bigg)\\
&\leq \mathbb{P}\Bigg(\sum_{k=2}^{d(n)}\Big((a_{k,n}^{(n)}-\mathbb{E}(a_{k,n}^{(n)}))^2+(\bar{p}_n^{(n)}\alpha_{k,n}^{(n)}-p_0\alpha_k(\mu_0,\theta_0))^2\Big)\\
&\qquad\geq \log(n)e(n)/(4n^{\lambda})-cd(n)\delta_n^2/2\Bigg)\\
&\leq \mathbb{P}\Bigg(\sum_{k=2}^{d(n)}\Big(a_{k,n}^{(n)}-\mathbb{E}(a_{k,n}^{(n)})\Big)^2\geq C(k,n)/(8n^{\lambda})\Bigg)\\
&+\mathbb{P}\Bigg(\sum_{k=2}^{d(n)}\Big(\bar{p}_n^{(n)}\alpha_{k,n}^{(n)}-p_0\alpha_k(\mu_0,\theta_0)\Big)^2\geq C(k,n)/(8n^{\lambda})\Bigg)
\end{aligned}
$$

where $C(k,n)=\log(n)e(n)-2cd(n)n^{\lambda}\delta_n^2$. By **(A1)** we have $d(n)=O(\log(n)e(n))$, and $n^{\lambda}\delta_n^2\to 0$ as $n\to+\infty$ due to **(A5)**. It follows that

$$
C(k,n)=\log(n)e(n)+o(\log(n)e(n)). \tag{28}
$$

We study the two above probabilities separately. First we have, according to the Markov inequality and Condition **(A3)**, that

$$
\begin{aligned}
\mathbb{P}\left(\sum_{k=2}^{d(n)}\Big(a_{k,n}^{(n)}-\mathbb{E}(a_{k,n}^{(n)})\Big)^2\geq \frac{C(k,n)}{8n^{\lambda}}\right) &\leq \frac{8n^{\lambda}}{C(k,n)}\sum_{k=2}^{d(n)}\frac{1}{n}\mathrm{var}\left(\frac{Q_k(X_1^n)}{q_k^2}\right)\\
&\leq \frac{8d(n)}{n^{1-\lambda}C(k,n)}M_3,
\end{aligned}
$$

where the last right hand side term goes to zero as $n\to+\infty$ according to **(A1)**.

Secondly we have

$$
\mathbb{P}\left(\sum_{k=2}^{d(n)}\Big(\bar{p}_n^{(n)}\alpha_{k,n}^{(n)}-p_0\alpha_k(\mu_0,\theta_0)\Big)^2\geq \frac{C(k,n)}{8n^{\lambda}}\right)
$$

$$\leq \mathbb{P}\left(\sum_{k=2}^{d(n)}\left(\alpha_{k,n}^{(n)}\right)^2\left(\bar{p}_n^{(n)}-p_0\right)^2 \geq \frac{C(k,n)}{16n^{\lambda}}\right)$$
$$+\,\mathbb{P}\left(p_0^2\sum_{k=2}^{d(n)}\left(\alpha_{k,n}^{(n)}-\alpha_k(\mu_0,\theta_0)\right)^2 \geq \frac{C(k,n)}{16n^{\lambda}}\right).$$

By **(A2)** the α_k's are bounded by M_1 which leads to

$$\mathbb{P}\left(\sum_{k=2}^{d(n)}\left(\alpha_{k,n}^{(n)}\right)^2\left(\bar{p}_n^{(n)}-p_0\right)^2 \geq \frac{C(k,n)}{16n^{\lambda}}\right)$$
$$\leq \mathbb{P}\left(\sum_{k=2}^{d(n)}M_1^2\left(\bar{p}_n^{(n)}-p_0\right)^2 \geq \frac{C(k,n)}{16n^{\lambda}}\right)$$
$$\leq \mathbb{P}\left(\left(\bar{p}_n^{(n)}-p_0\right)^2 \geq \frac{C(k,n)}{16n^{\lambda}d(n)M_1^2}\right).$$

We next prove that the last right hand side term goes to zero as $n \to +\infty$. Combining **(A4)-(A5)** with (ii) of Theorem 8 in Appendix E.2 we have for all $\alpha > 0$ and $0 < \delta < 1/2$,

$$|\bar{p}_n^{(n)}-p_0| = O_{a.s.}\left(\left(n^{-1/2+\alpha}+\delta_n/h_n^2\right)^{1/2-\delta}\right)$$
$$= O_{a.s.}\left(\left(n^{-1/2+\alpha}+n^{-1/2+(1+2\gamma-\xi)}\right)^{1/2-\delta}\right).$$

Two cases have to be considered now.

First case: if $1+2\gamma-\xi \leq 0$, we obtain

$$|\bar{p}_n^{(n)}-p_0| = O_{a.s.}\left(\left(n^{-1/2+\alpha}\right)^{1/2-\delta}\right),$$

for all $\alpha > 0$ and $0 < \delta < 1/2$, and it follows that for all $\beta > 0$

$$|\bar{p}_n^{(n)}-p_0| = O_{a.s.}\left(n^{-1/4+\beta}\right).$$

Second case: if $1+2\gamma-\xi > 0$ we obtain

$$|\bar{p}_n^{(n)}-p_0| = O_{a.s.}\left(\left(n^{-1/2+1+2\gamma-\xi}\right)^{1/2-\delta}\right),$$

for all $0 < \delta < 1/2$. Writing $u = \xi - 1/2 - 2\gamma$, we have

$$|\bar{p}_n^{(n)} - p_0| = O_{a.s.}\left(n^{-u/2+\beta}\right),$$

for all $\beta > 0$. By **(A5)**, $u > \lambda$ and we obtain $|\bar{p}_n^{(n)} - p_0| = o_{a.s.}(n^{-\lambda/2})$.

Finally, in both cases we have

$$|\bar{p}_n^{(n)} - p_0|^2 = o_{a.s.}(n^{-\lambda}) \tag{29}$$

and the assertion that the right hand side term goes to zero follows from (28) and **(A1)**. Writing $\rho_0 = (\mu_0, \theta_0)$ and $\bar{\rho}_n^{(n)} = (\bar{\mu}_n^{(n)}, \bar{\theta}_n^{(n)})$, similarly **(A5)-(A6)** give $\|\bar{\rho}_n^{(n)} - \rho_0\|^2 = o_{a.s.}(n^{-\lambda})$. Since the $\dot{\alpha}_k$'s are bounded by M_2 according to **(A2)**, using the *mean value* Theorem, we obtain:

$$\mathbb{P}\left(p_0^2 \sum_{k=2}^{d(n)} \left(\alpha_{k,n}^{(n)} - \alpha_k(\mu_0, \theta_0,)\right)^2 \geq \frac{C(k,n)}{16n^\lambda}\right)$$
$$\leq \mathbb{P}\left(\|\bar{\rho}_n^{(n)} - \rho_0\|^2 \geq \frac{C(k,n)}{16n^\lambda d(n) M_2^2}\right),$$

which last term goes to zero as $n \to +\infty$ according to **(A1)**. Hence from (24), we obtain that $\mathbb{P}(S_n \geq 2) \to 0$ as $n \to +\infty$. Therefore, using the proofs of Corollary 3 we get the limiting distribution of the test statistic $T(n)$ under H_1^*. □

Proposition 3. Let us compute the close forms of the quantities $\psi_{1,n}$ and $\psi_{2,n}$ defined in (26). It first comes

$$\begin{aligned}\psi_{1,n} &= p_0 \mathbb{E}(Q_1(X_0 + \delta_n \varepsilon_1) - Q_1(X_0)) \\ &= p_0 \mathbb{E}(a_{1,1}(X_0 + \delta_n \varepsilon_1) + a_{1,0} - a_{1,1}(X_0) - a_{1,0}) \\ &= p_0 \delta_n \mathbb{E}(\varepsilon_1),\end{aligned}$$

and we have

$$\begin{aligned}R_{1,n}^{(n)} &= \left(a_{1,n}^{(n)} - \mathbb{E}(a_{1,n}^{(n)})\right) - \left(\bar{p}^{(n)} \alpha_{1,n}^{(n)} - p_0 \alpha(\mu_0, \theta_0)\right) + \psi_{1,n} \\ &= A - B + \psi_{1,n}.\end{aligned}$$

Combining Markov inequality and **(A3)** we obtain

$$\mathbb{P}(|a_{1,n}^{(n)} - \mathbb{E}(a_{1,n}^{(n)})| \geq 1/n) \leq \operatorname{var}\left(\frac{Q_1(X_1^n)}{q_1^2}\right) < M_3,$$

ensuring that $A = O_{a.s.}(1/n)$. Moreover

$$B = \alpha_{k,n}^{(n)}\left(\bar{p}^{(n)} - p_0\right) + p_0\left(\alpha_{k,n}^{(n)} - \alpha(\mu_0, \theta_0)\right) = B_1 + B_2.$$

From **(A2)** we have $|\alpha_{k,n}^{(n)}| \leq M1$ and from (29) we have $(\bar{p}^{(n)} - p_0) = o_{a.s.}(n^{-\lambda/2})$ which prove that $B_1 = o_{a.s.}(n^{-\lambda/2})$. In the same way, using **(A2)** we can show that $B_2 = o_{a.s.}(n^{-\lambda/2})$.

By **(A7)** it follows that almost surely $n^\lambda (R_{1,n}^{(n)})^2 \approx n^{\lambda-2\xi'} \to +\infty$, as $n \to +\infty$. By construction we have $T_{1,n}^{(n)} \geq n(R_{1,n}^{(n)})^2/(\widehat{\text{var}}(R_{1,n}^{(n)}) + e(n))$ which leads to the almost sure convergence

$$s(n)T_{1,n}^{(n)} - \log(n) \xrightarrow{a.s.} +\infty, \quad \text{as } n \to +\infty.$$

Under **(A8)** we obtain immediately that $\psi_{1,n} = 0$ and $R_{1,n} = o_{a.s.}(n^{-\lambda/2})$. Since $T_{1,n}^{(n)} \leq n(R_{1,n}^{(n)})^2/e(n)$, it follows that almost surely

$$s(n)T_{1,n}^{(n)} - \log(n) \xrightarrow{a.s.} -\infty, \quad \text{as } n \to +\infty.$$

We also have

$$\begin{aligned}
\psi_{2,n} &= p_0 \mathbb{E}(Q_2(X_0 + \delta_n \varepsilon_1) - Q_2(X_0)) \\
&= p_0(\mathbb{E}(a_{2,2}(X_0 + \delta_n \varepsilon_1)^2 + a_{2,1}(X_0 + \delta_n \varepsilon_1) + a_{2,0} \\
&\quad - a_{2,2}(X_0)^2 - a_{2,1}(X_0) - a_{2,0})) \\
&= 2p_0 a_{2,2} \delta_n^2 \mathbb{E}(\varepsilon_1^2).
\end{aligned}$$

From the above expressions and by definition of $R_{2,n}^{(n)}$ in (25) we can mimic the previous arguments to show that almost surely $R_{2,n}^{(n)} \approx \delta_n^2$ and that

$$\begin{aligned}
&s(n)T_{2,n}^{(n)} - 2\log(n) \\
&= s(n)\left(n(R_{1,n}^{(n)})^2 \widehat{D}_{1,n}^{-1} + n(R_{k,n}^{(n)})^2 \widehat{D}_{2,n}^{-1}\right) - 2\log(n) \\
&\geq n^\lambda \left((R_{1,n}^{(n)})^2/(e(n) + \widehat{\text{var}}(R_{1,n}^{(n)})) + (R_{2,n}^{(n)})^2/(e(n) + \widehat{\text{var}}(R_{2,n}^{(n)}))\right) - 2\log(n),
\end{aligned}$$

where the last right hand side term goes to infinity as $n \to +\infty$ which gives us the wanted result. □

Appendix B: Proofs of technical lemmas

Proof of Lemma 1. Since the parameters (μ, s) belong to a compact set we can fix: $s_0 < s < s_1$ and $|\mu| < \mu_1$. We consider for simplicity $k = 2\ell$, $\ell \geq 0$, in the k-th order Hermite polynomial expression (14) and notice that for all $(\mu, s) \in \Lambda$,

$$|\mathbb{E}(H_k(\sqrt{s}Z + \mu))| \leq k! \sum_{m=0}^{\ell} \frac{\mathbb{E}((\sqrt{s}Z + \mu)^{2(\ell-m)})}{m!(2(\ell-m))!2^m}, \tag{30}$$

where Z is a $\mathcal{N}(0,1)$ distributed random variable. Now since

$$\mathbb{E}((\sqrt{s}Z + \mu)^{2(\ell-m)}) = \sum_{j=0}^{2(\ell-m)} C_j^{2(\ell-m)} \sqrt{s}^j \mathbb{E}(Z^j)|\mu|^{2(\ell-m)-j}$$

$$
\begin{aligned}
&\leq \mathbb{E}(Z^{2(\ell-m)})(\sqrt{s}+|\mu|)^{2(\ell-m)} \\
&= \frac{2(\ell-m)!}{2^{(\ell-m)}(\ell-m)!}(\sqrt{s}+|\mu|)^{2(\ell-m)}, \qquad (31)
\end{aligned}
$$

including (31) in (30), we obtain:

$$
\begin{aligned}
|\mathbb{E}(H_k(\sqrt{s}Z+\mu))| &\leq k! \sum_{m=0}^{\ell} \frac{(\sqrt{s}+|\mu|)^{2(\ell-m)}}{m!(\ell-m)!2^{\ell}} \\
&= \frac{k!}{\ell!}\left(\frac{(\sqrt{s}+|\mu|)^2+1}{2}\right)^{\ell}. \qquad (32)
\end{aligned}
$$

Since $\alpha_k(\mu,s)=\mathbb{E}(H_k(\sqrt{s}Z+\mu))/q_k^2$, with $q_k^2=k!$, we deduce from (32) that for all $k\geq 0$ and for all $(\mu,s)\in\Lambda$,

$$
\begin{aligned}
|\alpha_k(\mu,s)| &\leq \frac{1}{\ell!}\left(\frac{(\sqrt{s}+|\mu|)^2+1}{2}\right)^{\ell} \\
&\leq \exp\left(\frac{(\sqrt{s}_1+\mu_1)^2+1}{2}\right),
\end{aligned}
$$

which proves the first part of **(A2)**.

For the second part of condition **(A2)**, we detail for simplicity the majorization of $\left|\frac{\partial}{\partial s}\mathbb{E}(H_k(sZ+\mu))\right|$, for $k=2\ell$ and $\ell\geq 1$, which is:

$$
\left|\frac{\partial}{\partial s}\mathbb{E}(H_k(\sqrt{s}Z+\mu))\right| \leq \frac{k!}{2\sqrt{s}} \sum_{m=0}^{\ell-1} \frac{2(\ell-m)\mathbb{E}\left(\left|Z(\sqrt{s}Z+\mu)^{2(\ell-m)-1}\right|\right)}{m!(2(\ell-m))!2^m}.
$$

Now since

$$
\begin{aligned}
\mathbb{E}\left(\left|Z(\sqrt{s}Z+\mu)^{2(\ell-m)-1}\right|\right) &\leq \sum_{j=0}^{2(\ell-m)-1} C_j^{2(\ell-m)-1}\sqrt{s}^j\mathbb{E}(Z^{j+1})|\mu|^{2(\ell-m)-1-j}, \\
&\leq \mathbb{E}(Z^{2(\ell-m)})(\sqrt{s}+|\mu|)^{2(\ell-m)-1},
\end{aligned}
$$

we obtain

$$
\begin{aligned}
\left|\frac{\partial}{\partial s}\mathbb{E}(H_k(\sqrt{s}Z+\mu))\right| &\leq \frac{k!}{2\sqrt{s}} \sum_{m=0}^{\ell-1} \frac{(s+|\mu|)^{2(\ell-m)-1}}{m!(\ell-m-1))!2^{\ell-1}} \\
&= \frac{k!}{2\sqrt{s}(\ell-1)!}\left(\frac{(\sqrt{s}+|\mu|)^2+1}{2}\right)^{\ell-1} \\
&\leq \frac{k!}{2\sqrt{s_0}}\exp\left(\frac{(\sqrt{s}_1+\mu_1)^2+1}{2}\right),
\end{aligned}
$$

which concludes the proof of **(A2)**.

We now consider condition **(A3)**. We have

$$\begin{aligned}\mathbb{E}(H_k^2(X_1)) &= (1-p)\mathbb{E}(H_k^2(Z)) + p\mathbb{E}(H_k^2(\sqrt{s}Z+\mu)) \\ &= (1-p)k! + p\mathbb{E}(H_k^2(\sqrt{s}Z+\mu)). \end{aligned} \tag{33}$$

Let us consider the last term of the above right-hand side equality, for $k = 2\ell$ and $\ell \geq 0$:

$$\mathbb{E}(H_k^2(\sqrt{s}Z+\mu)) = (k!)^2 \sum_{m,q=0}^{\ell} \frac{\mathbb{E}\left((\sqrt{s}Z+\mu)^{2(2\ell-(m+q)))}\right)}{m!q!(2(\ell-m))!(2(\ell-q))!2^{m+q}}.$$

By the Cauchy-Schwartz inequality, and the fact that for all $n \geq 1$, we have $\sqrt{2\pi}n^{n+1/2}e^{-n} \leq n! \leq en^{n+1/2}e^{-n}$, we obtain:

$$\begin{aligned}\mathbb{E}(H_k^2(\sqrt{s}Z+\mu)) &\leq (k!)^2 \left(\sum_{m=0}^{\ell} \frac{\sqrt{\mathbb{E}\left((\sqrt{s}Z+\mu)^{4(\ell-m)}\right)}}{m!(2(\ell-m))!2^m}\right)^2 \\ &= (k!)^2 \left(\frac{1}{\ell!2^\ell} + \sum_{m=0}^{\ell-1} \frac{\sqrt{\mathbb{E}\left((\sqrt{s}Z+\mu)^{4(\ell-m)}\right)}}{m!(2(\ell-m))!2^m}\right)^2 \\ &\leq (k!)^2 \left(\frac{1}{\ell!2^\ell} + \sum_{m=0}^{\ell-1} \frac{\sqrt{(4(\ell-m))!}(\sqrt{s}+|\mu|)^{2(\ell-m)}}{2^\ell(2(\ell-m)!)^{3/2}m!}\right)^2 \\ &\leq \frac{(k!)^2 e}{2^{2\ell+1/2}(2\pi)^3}\left(\frac{(2\pi)^3 2^{\ell+1/2}}{e\,\ell!} + \right. \\ &\qquad \left. \sum_{m=0}^{\ell-1} 2^{\ell-m}(\ell-m)^{-(\ell-m)-1}e^{\ell-m}(\sqrt{s}+|\mu|)^{2(\ell-m)}\right)^2 \\ &\leq \frac{(k!)^2 e}{2^{2\ell+1/2}(2\pi)^3}\left(\frac{(2\pi)^3 2^{\ell+1/2}}{e\,\ell!} + \sum_{u=1}^{\ell} \rho^u u^{-u-1}\right)^2, \end{aligned} \tag{34}$$

where $u = \ell - m$ and $\rho = 2e(\sqrt{s}+|\mu|)^2$. Clearly, $\rho \leq \rho_0 = 2e(\sqrt{s}_0 + \mu_0)^2$, and the series on the right hand side converges. Combining (33) and (34) we obtain

$$\begin{aligned}\mathrm{var}(Q_k(X_1)^2/q_k^2) &\leq \mathbb{E}(Q_k^2(X_1))/q_k^4 \\ &= (1-p)/(k!) + p\mathbb{E}(H_k^2(\sqrt{s}Z+\mu))/(k!)^2\end{aligned}$$

which gives us the wanted result. □

Proof of Lemma 3. The polynomials defined by (15) satisfy the following equations:

$$xh_k(x) = h_{k+1}(x)/2 + kh_{k-1}(x) \text{ and } h_k'(x) = 2kh_{k-1}(x), \text{ for all } x \in \mathbb{R}.$$

It is also well known, see for instance Szegö [33], that there exists a constant $C > 0$ such that, for all $x \in \mathbb{R}$:

$$|\mathcal{H}_k(x)| = \exp(-x^2/2)|h_k(x)| \le C\sqrt{k!2k}. \tag{35}$$

Since $\alpha_k(\mu, s) = \mathbb{E}(\mathcal{H}_k(sY + \mu))/q_k^2$, we deduce that for all $s > 0$, and $\mu \in \mathbb{R}$,

$$\alpha_k(\mu, s) \le C/\sqrt{k!2k},$$

which gives the first bound in **(A2)**. Moreover, we have

$$\begin{aligned}
\mathcal{H}_k'(x) &= \exp(-x^2/2)\left(-xh_k(x) + h_k'(x)\right) \\
&= \exp(-x^2/2)\left(-(h_{k+1}(x)/2 - kh_{k-1}(x)) + 2kh_{k-1}(x)\right) \\
&= -\mathcal{H}_{k+1}(x)/2 + k\mathcal{H}_{k-1}(x),
\end{aligned}$$

which leads to

$$\frac{\mathcal{H}_k'(x)}{q_k^2} = -\frac{\mathcal{H}_{k+1}(x)}{2^{k+1}k!} + \frac{\mathcal{H}_{k-1}(x)}{2^k(k-1)!}.$$

Combining this equality with (35) we obtain

$$\left|\frac{\mathcal{H}_k'(x)}{q_k^2}\right| \le C\left(\frac{\sqrt{k+1}}{\sqrt{2^{k+1}k!}} + \frac{1}{\sqrt{2^{k+1}(k-1)!}}\right). \tag{36}$$

Now since $\dot{\alpha_k}(\mu, s) = \mathbb{E}\left((s^{-1/2}\mathcal{H}_k(\sqrt{s}Y + \mu), \mathcal{H}_k(\sqrt{s}Y + \mu))\right)/2$ it follows that for all $s > 0$ and $\mu \in \mathbb{R}$:

$$\|\dot{\alpha_k}(\mu, s)\| \le (s^{-1/2}/2 + 1)C\left(\frac{\sqrt{k+1}}{\sqrt{(k)!}\sqrt{2^{k+1}}} + \frac{1}{2\sqrt{2^{k-1}(k-1)!}}\right),$$

which gives the second bound in **(A2)**.

Finally from (35) we obtain $\mathrm{var}(Q_k(X_1)/q_k^2) = \mathrm{var}(\mathcal{H}_k(X_1)/q_k^2) \le C^2/(k!2k)$, which directly insures **(A3)**. □

Remark 7. *Lemma 3 is very general. Lemma 1 can be extended to any null distribution f with known moments such that the series given in (30) is bounded. This is obviously the case for distributions with bounded support.*

Proof of Lemma 2. By the Taylor expansion formula, and noticing that Q_k is a polynomial of order k, we have

$$\begin{aligned}
|\mathbb{E}\left(Q_k(X_0 + \delta_n\varepsilon_1) - Q_k(X_0)\right)| &= |\sum_{j=1}^{k}\mathbb{E}\left((\delta_n\varepsilon_1)^j Q_k^{(j)}(X_0)/j!\right)| \\
&\le \sum_{j=1}^{k}\frac{\delta_n^j}{j!}\mathbb{E}\left(|\varepsilon_1|^j\right)\mathbb{E}\left(|Q_k^{(j)}(X_0)|\right),
\end{aligned}$$

where $Q_k^{(j)}$ denotes the j-th derivative of the Hermite polynomial Q_k. These polynomials, see for instance Szegö [33], satisfy $Q_k^{(1)} = kQ_{k-1}$, which implies that $Q_k^{(j)} = \frac{k!}{(k-j)!}Q_{k-j}$, for $j \leq k$. It follows that

$$\begin{aligned}
\mathbb{E}\left(|Q_k^{(j)}(X_0)|\right) &= \int_{\mathbb{R}} |Q_k^{(j)}(y)|g(y)\nu(dy) \\
&= \frac{k!}{(k-j)!}\int_{\mathbb{R}} |Q_{k-j}(y)|g(y)\nu(dy) \\
&\leq \frac{k!}{(k-j)!}\sqrt{\int_{\mathbb{R}} (Q_{k-j}(y))^2\nu(dy)\int_{\mathbb{R}} g^2(y)\nu(dy)} \\
&= \frac{k!}{(k-j)!}q_{k-j}G,
\end{aligned}$$

where $G = \sqrt{\int_{\mathbb{R}} g^2(y)\nu(dy)} < \infty$, since g belongs to $L^2(\nu)$. Now because $q_k^2 = k!$ and $\delta_n < 1$, we get

$$\begin{aligned}
|\mathbb{E}\left(Q_k(X_0 + \delta_n\varepsilon_1) - Q_k(X_0)\right)|/q_k^2 &\leq G\delta_n \sum_{j=1}^{k} \frac{1}{\sqrt{(k-j)!}j!}\mathbb{E}\left(|\varepsilon_1|^j\right) \\
&\leq G\delta_n \mathbb{E}\left(e^{|\varepsilon_1|}\right),
\end{aligned}$$

which gives the wanted result. □

Proof of Lemma 4. By the Mean Value Theorem there exists a random variable ξ lying between $X_0 + \delta_n\varepsilon_1$ and X_0, such that

$$\mathbb{E}\left(Q_k(X_0 + \delta_n\varepsilon_1) - Q_k(X_0)\right) = \mathbb{E}\left(\delta_n\varepsilon_1 Q_k'(\xi)\right).$$

From (36) we have $|Q_k'(x)|/q_k^2 < 2C$ for all $x \in \mathbb{R}$, and we get

$$\begin{aligned}
|\mathbb{E}\left(Q_k(X_0 + \delta_n\varepsilon_1) - Q_k(X_0)\right)|/q_k^2 &\leq \delta_n\mathbb{E}\left(|\varepsilon_1||Q_k'(\xi)|\right)/q_k^2 \\
&\leq 2C\delta_n\mathbb{E}\left(|\varepsilon_1|\right),
\end{aligned}$$

and the lemma follows. □

Appendix C: Contiguous alternative modelling

We study in this section the asymptotic behavior of the semiparametric estimator $(\bar{p}_n, \bar{\mu}_n)$ introduced in Bordes and Vandekerkhove [9] when their model is no longer fixed but depends on n through the following transformation:

$$g^{(n)}(x) = pf_0(x) + (1-p)f^{(n)}(x-\mu), \quad x \in \mathbb{R}, \tag{37}$$

where $(f^{(n)})_{n\geq 1}$ is a sequence of ν-pdfs converging towards the limiting pdf f. For simplicity, when $f^{(n)}$ is replaced by f in (37), the resulting model will be so-called the *asymptotic model*. In this framework, for each $n \geq 1$, we consider a sample $(X_1^n, \ldots, X_n^n)$ independent and identically drawn from the n-local probability density function $g^{(n)}n$. In addition, we suppose that for any $(n, m) \in \mathbb{N}^* \times \mathbb{N}$ such that $n \neq m$, we have $(X_1^n, \ldots, X_n^n)$ independent from $(X_1^m, \ldots, X_m^m)$. The sequence $(X_1^n, \ldots, X_n^n)_{n\geq 1}$ is commonly called a *row independent triangular-array*. To handle easily the asymptotic normality of the Bordes and Vandekerkhove [9] semiparametric estimator based on the "corrupted" sample $(X_1^n, \ldots, X_n^n)$, we consider the *coupling*:

$$\begin{cases} X_i^n &= (1-U_i)Y_i + U_i Z_i^n, \quad i = 1, \ldots, n \\ X_i &= (1-U_i)Y_i + U_i Z_i, \quad i \geq 1, \end{cases} \tag{38}$$

where $(U_i)_{i\geq 1}$ and $(Y_i, Z_i, \varepsilon_i)_{i\geq 1}$ are independent and identically distributed samples respectively drawn from a Bernoulli distribution with parameter p and a $f_0 \otimes f(\cdot - \mu) \otimes f_1$-distribution. The random variable $Z_i^n = Z_i + \delta_n \varepsilon_i$ is by construction distributed according to $f^{(n)}$. Note that we have the following stochastic bound:

$$|X_i^n - X_i| \leq \delta_n |\varepsilon_i|, \quad i = 1, \ldots, n. \tag{39}$$

Appendix D: Estimation method

The cumulative distribution function (cdf) $G^{(n)}$ associated with model (37) is defined by

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

where $G^{(n)}$, F_0 and $F^{(n)}$ are cdfs corresponding to the pdfs $g^{(n)}$, f_0 and $f^{(n)}$ respectively. Let us denote by ϑ the Euclidean part (p, μ) of the model parameters taking values in Γ. Assume that the asymptotic model is identifiable and denote by $\vartheta_0 = (p_0, \mu_0)$ the true value of its unknown parameter ϑ. A way to estimate consistently ϑ_0, based on the triangular array $(X_1^n, \ldots, X_n^n)$, is to follow step by step the Bordes and Vandekerkhove [9] procedure. Let us define

$$F^{(n)}(x) = \frac{1}{p}\left(G^{(n)}(x+\mu) - (1-p)F_0(x+\mu)\right), \quad x \in \mathbb{R}. \tag{40}$$

Because $F^{(n)}$ approximates the symmetric cdf F, we have $F^{(n)}(x) \approx 1 - F^{(n)}(-x)$, for all $x \in \mathbb{R}$. Let us introduce, for all $x \in \mathbb{R}$, the functions

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

and

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

We have, using (40) and the *almost*-symmetry of $F^{(n)}$,

$$H^{(n)}(x; \vartheta_0, G^{(n)}) = H_1^{(n)}(x; \vartheta_0, G^{(n)}) - H_2^{(n)}(x; \vartheta_0, G^{(n)}) \approx 0, \tag{41}$$

whereas we can expect that for all $\vartheta \neq \vartheta_0$ an *ad hoc* norm of the function $H^{(n)}$ will have a significant departure from zero. In Bordes *et al.* [8] 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 $d^{(n)}$ on Θ:

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

which will likely converge towards the contrast function

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

associated with the asymptotic model (1), see Bordes and Vandekerkhove [9].

Because $G^{(n)}$ is unknown it is natural to replace it by its empirical version $\widehat{G}_n^{(n)}$ obtained from the n-sample $(X_1^n, \ldots, X_n^n)$. However, because we aim to estimate ϑ by the minimum argument of the empirical version of $d^{(n)}$ using a differentiable optimization routine, we need to replace $G^{(n)}$ in $H^{(n)}$ by a regular version $\widetilde{G}_n^{(n)}$ of $\widehat{G}_n^{(n)}$. Therefore we obtain an emprical version $d_n^{(n)}$ of $d^{(n)}$ defined by

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

where

$$\widehat{G}_n^{(n)}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{X_i^n \leq x}, \quad x \in \mathbb{R},$$

and $\widetilde{G}_n^{(n)}(x) = \int_{-\infty}^{x} \widehat{g}_n^{(n)}(t)dt$ denotes the smoothed version of the empirical cdf $\widehat{G}_n^{(n)}$ since $\widehat{g}_n^{(n)}$ is a kernel density estimator of $g^{(n)}$ defined by (13). Note that additional conditions on the bandwidth h_n and the kernel function q will be specified afterward.

In the sequel, when the above quantities are considered without superscript (n) this will simply means that $G^{(n)}$ has been replaced by G and $\mathbf{X}^{(n)} = (X_1^n, \ldots, X_n^n)$ by $\mathbf{X}^n = (X_1, \ldots, X_n)$ accordingly in their respective analytical expressions. Note that these estimators are then exactly the ones considered in Bordes and Vandekerkhove [9] (Section 2). Finally we propose to estimate ϑ_0 by

$$\bar{\vartheta}_n^{(n)} = (\bar{p}_n^{(n)}, \bar{\mu}_n^{(n)}) = \arg\min_{\vartheta \in \Gamma} d_n^{(n)}(\vartheta).$$

Appendix E: Identifiability, consistency and asymptotic normality

E.1. General conditions and identifiability

In this section we give a set of conditions for which we obtain identifiability of the asymptotic model parameters, consistency and asymptotic normality of our estimators. 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 pdf f_0 we consider one assumption, for which the semiparametric identifiability of the model (1) parameters is obtained, see Bordes *et al.* [8] (Proposition 2, p. 736).

Identifiability condition (I). For all $n \geq 1$, let $(f_0, f) \in \mathcal{F}_3^2$, $\bar{f}_0 > 0$ and $(\mu_0, m)) \in \Phi_c^{(n)}$ 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 $\Xi = \{\mu; (\mu, m) \in \Phi_c\}$.

Kernel conditions (K).

1. The even kernel density function K is bounded, uniformly continuous, square integrable, of bounded variations and has second order moment.
2. The function K has first order derivative $K' \in L^1(\mathbb{R})$ and $K'(x) \to 0$ as $|x| \to +\infty$. In addition if γ is the square root of the continuity modulus of K, we have
$$\int_0^1 (\log(1/u))^{1/2}\, d\gamma(u) < \infty.$$

Approximation conditions (A). The even kernel density function K is bounded, twice differentiable with bounded first and second derivatives.

Bandwidth conditions (B).

1. $h_n \searrow 0$, $nh_n \to +\infty$ and $\sqrt{n}h_n^2 = o(1)$,
2. $nh_n/|\log h_n| \to +\infty$, $|\log h_n|/\log\log n \to +\infty$ and there exists a real number c such that $h_n \leq ch_{2n}$ for all $n \geq 1$,
3. $|\log h_n|/(nh_n^3) \to 0$.

Comments. The two first conditions in **(B)** are necessary to obtain the pointwize consistency of the $\widehat{g}_n$ sequence of kernel estimators towards g. The third condition allows to control the distance between the empirical cdf $\widehat{G}_n$ and its regularized version $\widetilde{G}_n$. By using Corollary 1 in Shorack and Wellner [30], see page 766, we obtain

$$\|\widetilde{G}_n - \widehat{G}_n\|_\infty = O_{a.s.}(h_n^2),$$

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

$$\|\widetilde{G}_n - G\|_\infty = O_{a.s.}\left(\left(\frac{\log\log n}{n}\right)^{-1/2}\right). \tag{42}$$

Lemma 5. *Suppose that the kernel function q satisfies Conditions* **(K)** *and* **(A)** *and that the bandwidth (h_n) satisfies Conditions* **(B)**, *then we have:*

$$
\begin{aligned}
&(i) \qquad \|\widetilde{G}_n^{(n)} - \widetilde{G}_n\|_\infty = O_{a.s.}\left(\delta_n/h_n\right),\\
&(ii) \qquad \|\widehat{g}_n^{(n)} - \widehat{g}_n\|_\infty = O_{a.s.}\left(\delta_n/h_n^2\right),\\
&(iii) \qquad \|(\widehat{g}_n^{(n)})' - (\widehat{g}_n)'\|_\infty = O_{a.s.}\left(\delta_n/h_n^3\right).
\end{aligned}
$$

Proof. Let us detail the proof of result (ii). For all $x \in \mathbb{R}$, the stochastic error between $\widehat{g}_n^{(n)}(x)$ and $\widehat{g}_n(x)$ is controlled as follows:

$$
\begin{aligned}
\left|\widehat{g}_n^{(n)}(x) - \widehat{g}_n(x)\right| &= \left|\frac{1}{nh_n}\sum_{i=1}^n \left(K\left(\frac{x - X_i^n}{h_n}\right) - K\left(\frac{x - X_i}{h_n}\right)\right)\right|, \quad x \in \mathbb{R}\\
&\leq \frac{1}{nh_n}\sum_{i=1}^n \left|K\left(\frac{x - X_i^n}{h_n}\right) - K\left(\frac{x - X_i}{h_n}\right)\right|\\
&\leq \frac{1}{nh_n^2}\sum_{i=1}^n ||K'||_\infty \,|X_i^n - X_i|\\
&\leq \frac{||K'||_\infty \delta_n}{h_n^2} \times \left(\frac{\sum_{i=1}^n |\varepsilon_i|}{n}\right),
\end{aligned} \tag{43}
$$

where the last inequality comes from (39). The above result shows that, according to the Strong Law of Large numbers, $\|\widehat{g}_n^{(n)} - \widehat{g}_n\|_\infty = O_{a.s.}(\delta_n/h_n^2)$. The proofs of (i) and (iii) are identic to the proof (ii). □

E.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 6. *Assume that Conditions* **(K)**, **(A)** *and* **(B)** *are satisfied and that Θ is a compact subset of $(0,1) \times \Phi_c$.*

(i) If K is bounded over $\mathbb{R}$ then $\sup_{\vartheta\in\Theta}\left|d_n^{(n)}(\vartheta) - d_n(\vartheta)\right| = O_{a.s.}(\delta_n/h_n)$.

(ii) If K' is bounded over $\mathbb{R}$ then $\left\|\dot{d}_n^{(n)}(\vartheta_0) - \dot{d}_n(\vartheta_0)\right\| = O_{a.s.}(\delta_n^2/h_n^3) + O_{a.s.}(\delta_n/h_n)$.

(iii) If K'' is bounded over $\mathbb{R}$ then $\sup_{\vartheta\in\Theta}\left\|\ddot{d}_n^{(n)}(\vartheta) - \ddot{d}_n(\vartheta)\right\| = O_{a.s.}(\delta_n/h_n^3)$.

Proof. For the proof of (i) let us write for all $\vartheta \in \Theta$:

$$
\begin{aligned}
\left|d_n^{(n)}(\vartheta) - d_n(\vartheta)\right| &= \left|\frac{1}{n}\sum_{i=1}^n \left(H^2(X_i^n;\vartheta,\widetilde{G}_n^{(n)}) - H^2(X_i;\vartheta,\widetilde{G}_n)\right)\right|\\
&\leq \frac{1}{n}\sum_{i=1}^n \left|H^2(X_i^n;\vartheta,\widetilde{G}_n^{(n)}) - H^2(X_i^n;\vartheta,\widetilde{G}_n)\right|
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{1}{n}\sum_{i=1}^{n}\left|H^2(X_i^n;\vartheta,\widetilde{G}_n) - H^2(X_i;\vartheta,\widetilde{G}_n)\right| \\
&\leq O_{a.s.}\left(\|\widetilde{G}_n^{(n)} - \widetilde{G}_n\|_\infty\right) \\
&+ O_{a.s.}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\widetilde{G}_n(X_i^n+\mu) - \widetilde{G}_n(X_i+\mu)\right|\right). \qquad (44)
\end{aligned}
$$

The second term in the right hand side of the above inequality can be handled by using the mean value theorem as follows:

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left|\widetilde{G}_n(X_i^n+\mu) - \widetilde{G}_n(X_i+\mu)\right| &\leq \frac{1}{n}\sum_{i=1}^{n}(\|\widetilde{g}_n - g\|_\infty + \|g\|_\infty)|X_i^n - X_i| \\
&\leq \delta_n(o_{a.s.}(1) + ||g||_\infty) \times \left(\frac{\sum_{i=1}^{n}|\varepsilon_i|}{n}\right),
\end{aligned}
$$

where according to Silverman [31], $\|\widetilde{g}_n - g\|_\infty = o_{a.s.}(1)$. Similarly to (43), using the Strong of Large Numbers on the $|\varepsilon_i|$'s, we get that this second term is a $O_{a.s.}(\delta_n)$. Since the first term in the right hand side of (44) is a $O_{a.s.}(\delta_n/h_n)$ according to Lemma 5 (i), we obtain the wanted result.

For the proof of result (ii), let proceed similarly to Bordes and Vandekerkhove [9] and investigate the partial derivative of $\dot{d}_n^{(n)}(\vartheta_0)$ with respect to μ (more complicated case). Consider for any cdf G, the generic expression $\mathcal{H}(x,\vartheta_0,G) := H(x;\vartheta_0,G)\frac{\partial H}{\partial\mu}(x;\vartheta_0,G)$, $x\in\mathbb{R}$. According to (2.4) in Bordes and Vandekerkhove [9], we have at point ϑ_0:

$$
\left|\frac{\partial d_n^{(n)}}{\partial\mu}(\vartheta_0) - \frac{\partial d_n}{\partial\mu}(\vartheta_0)\right| \leq \Delta_1(\mathbf{X}^{(n)},\widetilde{G}_n^{(n)},\widetilde{G}_n) + \Delta_2(\mathbf{X}^{(n)},\mathbf{X}^n,\widetilde{G}_n),
$$

$$
\begin{aligned}
\text{where}\quad \Delta_1(\mathbf{X}^{(n)},\widetilde{G}_n^{(n)},\widetilde{G}_n) &= \frac{2}{n}\sum_{i=1}^{n}\left|\mathcal{H}(X_i^n;\vartheta_0,\widetilde{G}_n^{(n)}) - \mathcal{H}(X_i^n;\vartheta_0,\widetilde{G}_n)\right|, \\
\Delta_2(\mathbf{X}^{(n)},\mathbf{X}^n,\widetilde{G}_n) &= \frac{2}{n}\sum_{i=1}^{n}\left|\mathcal{H}(X_i^n;\vartheta_0,\widetilde{G}_n) - \mathcal{H}(X_i;\vartheta_0,\widetilde{G}_n)\right|.
\end{aligned}
$$

For $\Delta_1(\mathbf{X}^{(n)},\widetilde{G}_n^{(n)},\widetilde{G}_n)$, since $H(\cdot;\vartheta_0,G)=0$ and $\frac{\partial H}{\partial\mu}(\cdot;\vartheta_0,G) = 2f(\cdot)$, we can write:

$$
\begin{aligned}
\Delta_1(\mathbf{X}^{(n)},\widetilde{G}_n^{(n)},\widetilde{G}_n) \leq{}& \frac{2}{n}\sum_{i=1}^{n}\left|H(X_i^n;\vartheta_0,\widetilde{G}_n^{(n)}) - H(X_i^n;\vartheta_0,\widetilde{G}_n)\right| \\
&\times\left|\frac{\partial H}{\partial\mu}(X_i^n;\vartheta_0,\widetilde{G}_n^{(n)}) - \frac{\partial H}{\partial\mu}(X_i^n;\vartheta_0,\widetilde{G}_n)\right| \\
&+ \frac{2}{n}\sum_{i=1}^{n}\left|H(X_i^n;\vartheta_0,\widetilde{G}_n) - H(X_i^n;\vartheta_0,G)\right|
\end{aligned}
$$

$$
\begin{aligned}
&\times \left|\frac{\partial H}{\partial \mu}(X_i^n;\vartheta_0,\widetilde{G}_n^{(n)}) - \frac{\partial H}{\partial \mu}(X_i^n;\vartheta_0,\widetilde{G}_n)\right| \\
&+ \frac{2}{n}\sum_{i=1}^n \left|\frac{\partial H}{\partial \mu}(X_i^n;\vartheta_0,\widetilde{G}_n) - 2f(X_i^n)\right| \\
&\times \left|H(X_i^n;\vartheta_0,\widetilde{G}_n^{(n)}) - H(X_i^n;\vartheta_0,\widetilde{G}_n)\right| \\
&+ \frac{4}{n}\sum_{i=1}^n |f(X_i^n)| \times \left|H(X_i^n;\vartheta_0,\widetilde{G}_n^{(n)}) - H(X_i^n;\vartheta_0,\widetilde{G}_n)\right| \\
&\le c_1\left(\|\widetilde{G}_n^{(n)} - \widetilde{G}_n\|_\infty + \|\widetilde{G}_n - G\|_\infty\right)\|\widetilde{g}_n^{(n)} - \widetilde{g}_n\|_\infty \\
&+ c_2\left(\|\widetilde{g}_n - g\|_\infty + \|f\|_\infty\right)\|\widetilde{G}_n^{(n)} - \widetilde{G}_n\|_\infty \\
&= O_{a.s.}\left(\frac{\delta_n^2}{h_n^3}\right) + O_{a.s.}\left(\frac{\delta_n}{h_n}\right).
\end{aligned}
$$

For $\Delta_2(\mathbf{X}^{(n)},\mathbf{X}^n,\widetilde{G}_n)$ let us notice first that for any $(x,x')\in\mathbb{R}^2$ we have:

$$
\begin{aligned}
&\left|H(x;\vartheta_0,\widetilde{G}_n) - H(x';\vartheta_0,\widetilde{G}_n)\right| \\
&\le \frac{1}{p_0}\left|(\widetilde{G}(x+\mu) - \widetilde{G}(x'-\mu)) + (\widetilde{G}(-x+\mu) - \widetilde{G}(-x'-\mu))\right| \\
&+ \frac{1-p_0}{p_0}\left|(F_0(x+\mu) - F_0(x'-\mu)) + (F_0(-x+\mu) - F_0(-x'-\mu))\right| \\
&\le \frac{2}{p_0}(\|\widetilde{g}_n - g\|_\infty + \|g\|_\infty)|x-x'| + \frac{2(1-p_0)}{p_0}\|f_0\|_\infty|x-x'|, \qquad (45)
\end{aligned}
$$

and

$$
\begin{aligned}
&\left|\frac{\partial H}{\partial \mu}(x;\vartheta_0,\widetilde{G}_n) - \frac{\partial H}{\partial \mu}(x';\vartheta_0,\widetilde{G}_n)\right| \\
&\le \frac{1}{p_0}\left|(\widetilde{g}_n(x+\mu) - \widetilde{g}_n(x'-\mu)) + (\widetilde{g}_n(-x+\mu) - \widetilde{g}_n(-x'-\mu))\right| \\
&+ \frac{1-p_0}{p_0}\left|(f_0(x+\mu) - f_0(x'-\mu)) + (f_0(-x+\mu) - f_0(-x'-\mu))\right| \\
&\le \frac{2}{p_0}(\|\widetilde{g}_n' - g'\|_\infty + \|g'\|_\infty)|x-x'| + \frac{2(1-p_0)}{p_0}\|f_0'\|_\infty|x-x'|. \qquad (46)
\end{aligned}
$$

Using (45) and (46) we obtain

$$
\begin{aligned}
\Delta_2(\mathbf{X}^{(n)},\mathbf{X}^n,\widetilde{G}_n) \le\ & \frac{2}{n}\sum_{i=1}^n \left|H(X_i^n;\vartheta_0,\widetilde{G}_n) - H(X_i;\vartheta_0,\widetilde{G}_n)\right| \\
&\times \left|\frac{\partial H}{\partial \mu}(X_i^n;\vartheta_0,\widetilde{G}_n) - \frac{\partial H}{\partial \mu}(X_i;\vartheta_0,\widetilde{G}_n)\right| \\
&+ \frac{2}{n}\sum_{i=1}^n \left|H(X_i;\vartheta_0,\widetilde{G}_n) - H(X_i;\vartheta_0,G)\right|
\end{aligned}
$$

$$
\begin{aligned}
&\times \left|\frac{\partial H}{\partial \mu}(X_i^n;\vartheta_0,\widetilde{G}_n) - \frac{\partial H}{\partial \mu}(X_i;\vartheta_0,\widetilde{G}_n)\right| \\
&+ \frac{2}{n}\sum_{i=1}^n \left|\frac{\partial H}{\partial \mu}(X_i;\vartheta_0,\widetilde{G}_n) - 2f(X_i)\right| \\
&\times \left|H(X_i^n;\vartheta_0,\widetilde{G}_n) - H(X_i;\vartheta_0,\widetilde{G}_n)\right| \\
&+ \frac{4}{n}\sum_{i=1}^n |f(X_i)| \times \left|H(X_i^n;\vartheta_0,\widetilde{G}_n) - H(X_i;\vartheta_0,\widetilde{G}_n)\right| \\
&= O_{a.s.}(\delta_n^2) + O_{a.s.}(\delta_n^2) + O_{a.s.}(\delta_n \|\widetilde{G}_n - G\|_\infty) + O_{a.s.}(\delta_n),
\end{aligned}
$$

which by (42) concludes the proof for (ii).

For the proof of result (iii) we use the following decomposition at any point $\vartheta \in \Gamma$:

$$
\begin{aligned}
&\left\|\ddot{d}_n^{(n)}(\vartheta) - \ddot{d}_n(\vartheta)\right\| \\
&\le \frac{2}{n}\sum_{k=1}^n \left\|H(X_i^{(n)};\vartheta,\widetilde{G}_n^{(n)})\ddot{H}(X_i^n;\vartheta,\widetilde{G}_n^{(n)}) - H(X_i;\vartheta,\widetilde{G}_n)\ddot{H}(X_i;\vartheta,\widetilde{G}_n)\right\| \\
&+ \frac{2}{n}\sum_{k=1}^n \left\|\dot{H}(X_i^{(n)};\vartheta,\widetilde{G}_n^{(n)})\dot{H}^T(X_i^{(n)};\vartheta,\widetilde{G}_n^{(n)}) - \dot{H}(X_i;\vartheta,\widetilde{G}_n)\dot{H}^T(X_i;\vartheta,\widetilde{G}_n)\right\| \\
&\le \sum_{j=1}^4 T_{j,1} + T_{j,2},
\end{aligned}
$$

where for $j = 1,\dots,4$, $T_{j,1}$ and $T_{j,2}$ are alternatively equal to

$$
\begin{aligned}
&\frac{2}{n}\sum_{k=1}^n |H(X_i^{(n)};\vartheta,\widetilde{G}_n^{(n)})| \left\|\ddot{H}(X_i^n;\vartheta,\widetilde{G}_n^{(n)}) - \ddot{H}(X_i^n;\vartheta,\widetilde{G}_n)\right\| = O_{a.s.}\left(\frac{\delta_n}{h_n^3}\right) \\
&\frac{2}{n}\sum_{k=1}^n |H(X_i^{(n)};\vartheta,\widetilde{G}_n^{(n)})| \left\|\ddot{H}(X_i^n;\vartheta,\widetilde{G}_n) - \ddot{H}(X_i;\vartheta,\widetilde{G}_n)\right\| = O_{a.s.}(\delta_n) \\
&\frac{2}{n}\sum_{k=1}^n \left\|\ddot{H}(X_i;\vartheta,\widetilde{G}_n)\right\| \left|H(X_i^n\vartheta,\widetilde{G}_n^{(n)}) - H(X_i^n;\vartheta,\widetilde{G}_n)\right| = O_{a.s.}\left(\frac{\delta_n}{h_n}\right) \\
&\frac{2}{n}\sum_{k=1}^n \left\|\ddot{H}(X_i;\vartheta,\widetilde{G}_n)\right\| \left|H(X_i^n;\vartheta,\widetilde{G}_n) - H(X_i;\vartheta,\widetilde{G}_n)\right| = O_{a.s.}(\delta_n) \\
&\frac{2}{n}\sum_{k=1}^n \left\|\dot{H}(X_i^n;\vartheta,\widetilde{G}_n^{(n)})\right\| \left\|\dot{H}(X_i^n;\vartheta,\widetilde{G}_n^{(n)}) - \dot{H}(X_i^n;\vartheta,\widetilde{G}_n)\right\| = O_{a.s.}\left(\frac{\delta_n}{h_n^2}\right) \\
&\frac{2}{n}\sum_{k=1}^n \left\|\dot{H}(X_i;\vartheta,\widetilde{G}_n^{(n)})\right\| \left\|\dot{H}(X_i^n;\vartheta,\widetilde{G}_n) - \dot{H}(X_i;\vartheta,\widetilde{G}_n)\right\| = O_{a.s.}\left(\frac{\delta_n}{h_n^2}\right) \\
&\frac{2}{n}\sum_{k=1}^n \left\|\dot{H}(X_i^n;\vartheta,\widetilde{G}_n)\right\| \left\|\dot{H}(X_i^n;\vartheta,\widetilde{G}_n^{(n)}) - \dot{H}(X_i^n;\vartheta,\widetilde{G}_n)\right\| = O_{a.s.}\left(\frac{\delta_n}{h_n^2}\right)
\end{aligned}
$$

$$\frac{2}{n}\sum_{k=1}^{n}\left\|\dot{H}(X_i;\vartheta,\widetilde{G}_n)\right\|\left\|\dot{H}(X_i^n;\vartheta,\widetilde{G}_n)-\dot{H}(X_i;\vartheta,\widetilde{G}_n)\right\| = O_{a.s.}(\delta_n).$$

The above results come from painful but straightforward calculations. To explain briefly how we get these rates we can basically say that the first factors after the sum sign are always $O_{a.s.}(1)$ due to Silverman [31] if they are $\widetilde{G}_n$ dependent and $O_{a.s.}(1+\delta_n/h_n^{1+k})$, where $k=0,1,2$ denotes the order of derivation of H, if they are $\widetilde{G}_n^{(n)}$ dependent. Next, due to the mean value theorem, Silverman [31] uniform consistency result on the kernel estimator and its derivatives and (39), the difference terms involving X_i^n and X_i based on $\widetilde{G}_n$ are all $O_{a.s.}(\delta_n)$. On the other hand due to approximation Lemma 6, the difference terms involving $\widetilde{G}_n^{(n)}$ and $\widetilde{G}_n$ located at the same argument value X_i^n are all $O_{a.s.}(\delta_n/h_n^{1+k})$ where $k=0,1,2$ denotes the order of derivation of H. □

Theorem 8. *(i) Suppose that Conditions* **(K)**, **(B)** *and* **(I)** *are satisfied,* Γ *is a compact subset of* $(0,1)\times\Phi_c$, G *is strictly increasing on* $\mathbb{R}$, F_0 *and* F *are twice continuously differentiable with second derivatives in* $L^1(\mathbb{R})$, *then we have* $\|\bar{\vartheta}_n-\vartheta_0\| = o_{a.s.}(n^{-1/4+\alpha})$ *for all* $\alpha>0$.

(ii) Suppose in addition that Condition **(A)** *is satisfied, then we have*

$$\|\bar{\vartheta}_n^{(n)}-\vartheta_0\| = O_{a.s.}\left(\left(n^{-1/2+\alpha}+\delta_n/h_n^2\right)^{1/2-\delta}\right),$$

for all $\alpha>0$ *and* $0<\delta<1/2$.

(iii) Under the conditions of (i), the estimator $\bar{\vartheta}_n=(\bar{p}_n,\bar{\mu}_n)$ *is asymptotically normally distributed:*

$$\sqrt{n}(\bar{p}_n-p_0,\bar{\mu}_n-\mu_0)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\Sigma),\quad as\quad n\to+\infty,$$

where $\Sigma=\mathcal{I}(\vartheta_0)^{-1}J(\theta_0)\mathcal{I}(\vartheta_0)^{-1}$, *with*

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

and $J(\theta_0)=\mathbb{V}(H(X_1,\vartheta_0,G)\dot{H}(X_1,\vartheta_0,G))$.

(iv) Under the conditions of (ii), and if

$$\sqrt{n}\left(\frac{\delta_n^2}{h_n^3}+\frac{\delta_n}{h_n}\right)\longrightarrow 0,\quad and\quad \frac{\delta_n}{h_n^3}\longrightarrow 0,\quad as\quad n\to+\infty, \tag{47}$$

the estimator $\bar{\vartheta}_n^{(n)}=(\bar{p}_n^{(n)},\bar{\mu}_n^{(n)})$ *associated with the triangular array* $(\mathbf{X}^{(n)})_{n\geq 1}$ *defined in (38) is asymptotically normally distributed:*

$$\sqrt{n}(\bar{p}_n^{(n)}-p_0,\bar{\mu}_n^{(n)}-\mu_0)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\Sigma),\quad as\quad n\to+\infty.$$

Proof. The proofs of (i) and (iii) are detailed in Bordes and Vandekerkhove [9]. For the proof of result (ii) it is enough to notice that

$$\sup_{\vartheta\in\Theta}|d_n^{(n)}-d|\leq \sup_{\vartheta\in\Theta}|d_n^{(n)}-d_n|+\sup_{\vartheta\in\Theta}|d_n-d|=O_{a.s.}(\delta_n/h_n+n^{-1/2+\alpha}),$$

with $\alpha>0$, and consider $\gamma_n=n^{-1/2+\alpha}+\delta_n/h_n$ along with $\eta_n=(n^{-1/2+\alpha}+\delta_n/h_n)^{1/2-\delta}$, with $\delta>0$ in the proof of Theorem 3.1 of Bordes and Vandekerkhove [9]. Doing so we insure that $\gamma_n=o(\eta_n^2)$ which concludes the proof of (ii).

For the proof of (iv) we consider the Taylor expansion of $\dot{d}_n^{(n)}$ around ϑ_0:

$$\ddot{d}_n^{(n)}(\vartheta{^*}_n^{(n)})\sqrt{n}(\bar{\vartheta}_n^{(n)}-\vartheta_0)=-\sqrt{n}\dot{d}_n^{(n)}(\vartheta_0)=-\sqrt{n}\dot{d}_n(\vartheta_0)+o_{a.s.}(1),$$

where $\vartheta{^*}_n^{(n)}$ lies in the line segment with extremities $\bar{\vartheta}_n^{(n)}$ and ϑ_0, and $o_{a.s.}(1)=-\sqrt{n}(\dot{d}_n^{(n)}(\vartheta_0)-\dot{d}_n(\vartheta_0))$ according to Lemma 6 if $\sqrt{n}(\delta^2/h_n^3+\delta/h_n)\to 0$ as $n\to+\infty$. Noticing now that:

$$\begin{aligned}\|\ddot{d}_n^{(n)}(\vartheta{^*}_n^{(n)})-\mathcal{I}(\vartheta_0)\| &\leq \|\ddot{d}_n^{(n)}(\vartheta{^*}_n^{(n)})-\ddot{d}_n(\vartheta_n^{*(n)})\|+\|\ddot{d}(\vartheta_n^{*(n)})-\mathcal{I}(\vartheta_0)\|\\ &\leq \sup_{\Theta}\|\ddot{d}_n^{(n)}-\ddot{d}_n\|+\|\ddot{d}(\vartheta_n^{*(n)})-\mathcal{I}(\vartheta_0)\|,\end{aligned}$$

where the first term in the right hand side is a $o_{a.s.}(1)$ if $\delta_n/h_n^3\to 0$ as $n\to+\infty$ according to Lemma 6 (iii) and the second term is also a $o_{a.s.}(1)$ according to (3.16) in the proof of Theorem 3.2 in Bordes and Vandekerkhove [9]. □

Remark 9. *Since the bandwidth rate recommended in Bordes and Vandekerkhove (2010, Remark 3.1) to satisfy Condition (B) is* $n^{-1/4-\gamma}$*, with* $\gamma\in(0,1/8)$ *we observe that for this range of rates condition (47) is satisfied if:*

$$\frac{\delta_n^2}{n^{-5/4-3\gamma}}+\frac{\delta_n}{n^{-3/4-\gamma}}\longrightarrow 0,\quad \text{and}\quad \frac{\delta_n}{n^{-3/4-3\gamma}}\longrightarrow 0,\quad \text{as}\quad n\to+\infty,$$

which leads to consider $\delta_n=n^{-3/4-\xi}$ *with* $\xi>3\gamma$*.*

Remark 10. *The conditions imposed in (47) do not look optimal to us but they provide for the first time, to the best of our knowledge, a framework for nonparametric contiguous alternatives in the parametric family testing problem. To improve these rates in the future we plan to carefully investigate the Donsker theorem associated with the empirical process* $\mathbb{G}_n=\sqrt{n}(\widehat{G}_n^{(n)}-G^{(n)})$*, where* $\widehat{G}_n^{(n)}$ *denotes the empirical cdf of a* $G^{(n)}$*-distributed generic triangular array* $(X_1^n,\dots,X_n^n)$*, where* $G^{(n)}$ *converges "smoothly enough" towards a given cdf* G *and revisit the uniform almost sure convergence results of the kernel density estimate and its derivatives in Silverman [31].*

Appendix F: Asymptotic behavior of the MLE

In this section we propose to derive the asymptotic covariance matrix involved in the Central Limit Theorem associated with the maximum likelihood estimator for the Gaussian case, that is when f belongs to $\mathcal{G}$ the set of normal densities $f_{(\mu,s)}$ with mean μ and variance $\theta = s$. Let us denote by $g_\phi(x) = (1-p)f_{(0,1)}(x) + pf_{(\mu,s)}(x)$ where $\phi = (\phi_1, \phi_2, \phi_3) = (p, \mu, s) \in (0,1) \times \Lambda$ and $\ell_\phi(x) = \ln(g_\phi(x))$. We now define the gradient of $\ell_\phi(x)$:

$$\dot{\ell}_\phi(x) = \left(\frac{\partial}{\partial \phi_1}\ell_\phi(x), \frac{\partial}{\partial \phi_2}\ell_\phi(x), \frac{\partial}{\partial \phi_3}\ell_\phi(x) \right)^\top.$$

For simplicity matters we denote $\dot{f}^{\phi_i}_{(\mu,s)}(x) := \frac{\partial}{\partial \phi_i} f_{(\mu,s)}(x)$, $i = 1,2,3$. We then obtain

$$\begin{aligned}
\frac{\partial}{\partial \phi_1}\ell_\phi(x) &= \frac{-f_{(0,1)}(x) + f_{(\mu,s)}(x)}{g_\phi(x)} \\
\frac{\partial}{\partial \phi_2}\ell_\phi(x) &= \frac{p\dot{f}^{\mu}_{(\mu,s)}(x)}{g_\phi(x)}, \quad \text{with} \quad \dot{f}^{\mu}_{(\mu,s)}(x) = \frac{x-\mu}{s} f_{(\mu,s)}(x) \\
\frac{\partial}{\partial \phi_3}\ell_\phi(x) &= \frac{p\dot{f}^{s}_{(\mu,s)}(x)}{g_\phi(x)}, \quad \text{with} \quad \dot{f}^{s}_{(\mu,s)}(x) = \left[-\frac{1}{2s} + \frac{(x-\mu)^2}{2s^2} \right] f_{(\mu,s)}(x).
\end{aligned}$$

The Hessian matrix of $\ell_\phi(x)$ is denoted $\ddot{\ell}_\phi(x) = \left(\frac{\partial^2}{\partial \phi_i \partial \phi_j}\ell_\phi(x) \right)_{1 \le i \le j \le 3}$ with:

$$\begin{aligned}
\frac{\partial^2}{\partial^2 \phi_1}\ell_\phi(x) &= -\frac{(-f_{(0,1)}(x) + f_{(\mu,s)})^2(x)}{g^2_\phi(x)} \\
\frac{\partial^2}{\partial^2 \phi_2}\ell_\phi(x) &= p\frac{\ddot{f}^{\mu}_{(\mu,s)}(x)}{g_\phi(x)} - \left(p\frac{\dot{f}^{\mu}_{(\mu,s)}(x)}{g_\phi(x)} \right)^2 \\
\frac{\partial^2}{\partial^2 \phi_3}\ell_\phi(x) &= p\frac{\ddot{f}^{s}_{(\mu,s)}(x)}{g_\phi(x)} - \left(p\frac{\dot{f}^{s}_{(\mu,s)}(x)}{g_\phi(x)} \right)^2,
\end{aligned}$$

and

$$\begin{aligned}
\ddot{f}^{\mu}_{(\mu,s)}(x) &= -\frac{1}{s} f_{(\mu,s)}(x) + \frac{x-\mu}{s}\dot{f}^{\mu}_{(\mu,s)}(x) = -\frac{1}{s} f_{(\mu,s)}(x) + \left(\frac{x-\mu}{s}\right)^2 f_{(\mu,s)}(x) \\
\ddot{f}^{s}_{(\mu,s)}(x) &= \left[\frac{1}{2s^2} - \frac{(x-\mu)^2}{s^3} \right] f_{(\mu,s)}(x) + \left[-\frac{1}{2s} + \frac{(x-\mu)^2}{2s^2} \right] \dot{f}^{s}_{(\mu,s)}(x)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial \phi_1 \partial \phi_2}\ell_\phi(x) &= \frac{\partial^2}{\partial \phi_2 \partial \phi_1}\ell_\phi(x) \\
&= \frac{\dot{f}^{\mu}_{(\mu,s)}(x)}{g_\phi(x)} - p\frac{(-f_{(0,1)}(x) + f_{(\mu,s)}(x))\dot{f}^{\mu}_{(\mu,s)}(x)}{g^2_\phi(x)} \\
\frac{\partial^2}{\partial \phi_1 \partial \phi_3}\ell_\phi(x) &= \frac{\partial^2}{\partial \phi_3 \partial \phi_1}\ell_\phi(x)
\end{aligned}$$

$$= \frac{\dot{f}^s_{(\mu,s)}(x)}{g_\phi(x)} - p\frac{(-f_{(0,1)}(x) + f_{(\mu,s)}(x))\dot{f}^s_{(\mu,s)}(x)}{g^2_\phi(x)}$$

$$\frac{\partial^2}{\partial\phi_2\partial\phi_3}\ell_\phi(x) = \frac{\partial^2}{\partial\phi_3\partial\phi_2}\ell_\phi(x) = \frac{p(x-\mu)}{s^2} \times \frac{\left[-f_{(\mu,s)}(x) + s\dot{f}^s_{(\mu,s)}(x)\right]}{g_\phi(x)} - \frac{p^2(x-\mu)}{s} \times \frac{f_{(\mu,s)}(x)\dot{f}^s_{(\mu,s)}(x)}{g^2_\phi(x)}.$$

Given the above expressions we can derive under standard conditions, see van der Vaart [34] page 63, the basic asymptotic normality of the MLE:

$$\sqrt{n}(\widehat{p}_n - p_0, \widehat{\mu}_n - \mu_0, \widehat{s}_n - s_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0_{\mathbb{R}^3}, A(\phi_0)^{-1}B(\phi_0)A(\phi_0)^{-1}),$$

as $n \to +\infty$, where

$$A(\phi_0) = \mathbb{E}\left(\ddot{\ell}_{\phi_0}(X_1)\right) \quad \text{and} \quad B(\phi_0) = \mathbb{E}\left(\dot{\ell}_{\phi_0}(X_1)\dot{\ell}^T_{\phi_0}(X_1)\right)$$

are respectively consistently estimated by

$$\widehat{A}_n = \frac{1}{n}\sum_{i=1}^n \ddot{\ell}_{\widehat{\phi}_n}(X_i) \quad \text{and} \quad \widehat{B}_n = \frac{1}{n}\sum_{i=1}^n \dot{\ell}_{\widehat{\phi}_n}(X_i)\dot{\ell}^\top_{\widehat{\phi}_n}(X_i).$$

Appendix G: Graphs illustrating the asymptotic convergence under the null

We illustrate the empirical distribution of $T(n)$ for $n = 1000, 2000, 3000$ under the null, with $(p, \mu, s) = (1/3, 1.5, 1)$ the first set of parameters considered in Section 8.1. Based on 300 replications, we obtain the empirical distribution functions displayed in Fig. 9.

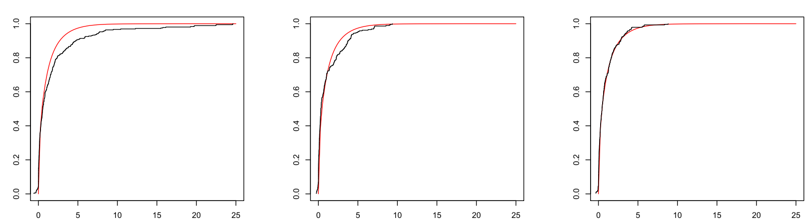

FIG 9. *In black: the empirical distribution functions of the test statistic $T(n)$ under H_0 with $(p, \mu, s) = (1/3, 1.5, 1)$ respectively for $n = 1000, 2000, 3000$. In red: the graph of the χ^2_1 asymptotic distribution function.*

- For $n = 1000$, a few values of the test statistic are very large leading to a heavy tailed empirical distribution function. We obtain for $T(n)$ a mean equal to 1.7 with a variance equal to 9.9. The Kolmogorov-Smirnov test rejects the χ^2_1 assumption with a p-value equal to 0.0009.

- For $n = 2000$, very few values exceed the expected values for a Chi-squared distribution. We obtain for $T(n)$ a mean equal to 1.5 and a variance equal to 3.3. The Kolmogorov-Smirnov test again rejects the χ_1^2 assumption, but here with a p-value equal to 0.08.
- Finally, for $n = 3000$, we obtain a mean for $T(n)$ equal to 1.2 and a variance equal to 2.2. The Kolmogorov-Smirnov test does not reject the χ_1^2 assumption, given a p-value equal to 0.39.

Appendix H: Graphs for maximum likelihood and semiparametric estimators comparison

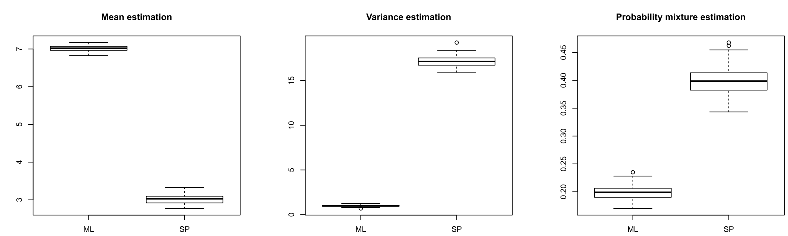

Fig 10. *Boxplot of the maximum likelihood and semiparametric estimators of m, s, p when $n = 1000$, under the mean deviation trap effect for $\mu = 3$ and $a = 4$, based on 200 repetitions.*

Appendix I: Empirical level graph

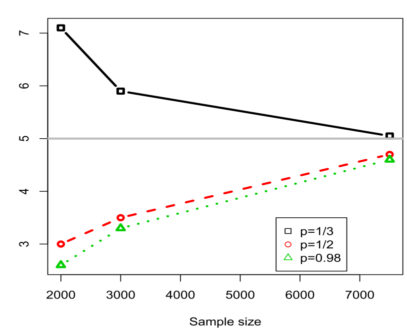

Fig 11. *Empirical levels for parameter values* $(p, \mu, s) = (1/3, 1.5, 1)$ *(□), $(p, \mu, s) = (0.5, 1.5, 2)$ (○) and $(p, \mu, s) = (0.98, -0.15, 0.8)$ (△) with sample sizes* $n = 2000, 3000, 7500$.

Appendix J: Graphs of the alternatives considered in Section 8.2

Row 1: 1-shifted Student $t(3)$ alternative distribution (plain) and a null-type Gaussian distribution with similar parameters $\mathcal{N}(1, 3)$ (dashed).

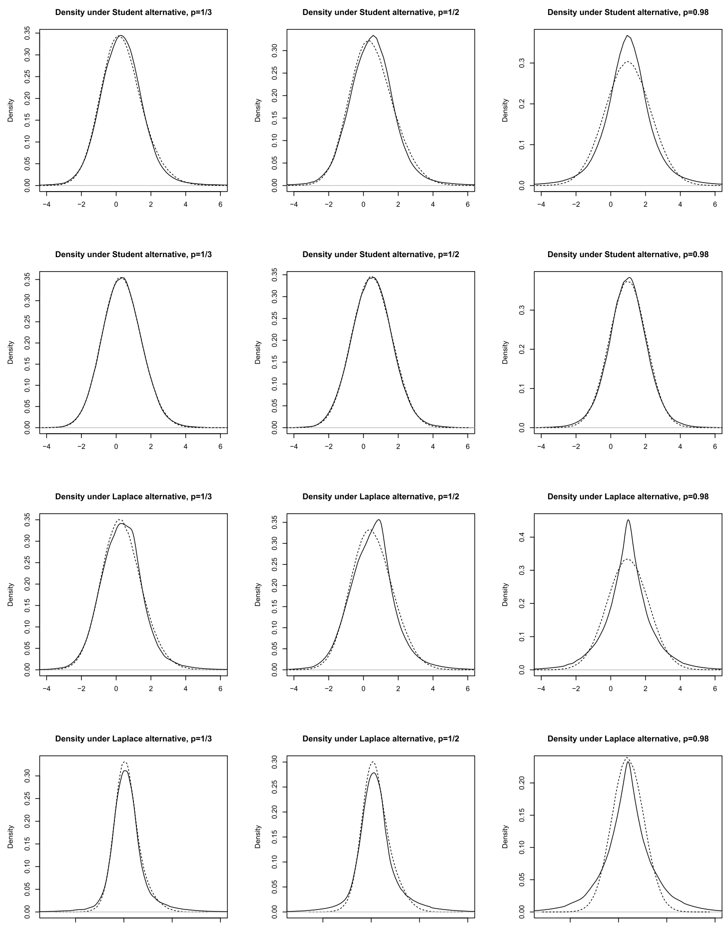

FIG 12. *Plot of the graph of g in model (1) under several conditions.*

Row 2: 1-shifted Student $t(10)$ alternative distribution (plain) and a null-type Gaussian distribution with similar parameters $\mathcal{N}(1, 1.25)$ (dashed).
Row 3: $\mathcal{L}(1, 1)$ Laplace distribution (plain) and a null-type Gaussian component with similar parameters $\mathcal{N}(1, 2)$ (dashed).
Row 4: $\mathcal{L}(1, 2)$ Laplace distribution (plain) and a null-type Gaussian component with similar parameters $\mathcal{N}(1, 8)$ (dashed).
Columns 1, 2, 3 correspond respectively to $p = 1/3,\ 1/2,\ 0.98$.

Appendix K: Empirical powers comparison

In contrast with the very separate behaviors noticed under the mean or variance deviation trap situations, see Section 7, there are most often situations where parametric and nonparametric approaches provide similar conclusions. For instance, for a mixture of a standard Gaussian and a Student distribution, the maximum likelihood and the nonparametric procedure give very similar powers. This behavioral closeness is demonstrated in Fig. 13 where the empirical powers obtained by both approaches under $p = 1/3$ and a shifted Student(3) alternative with $n = 2000, 3000, 7500$ are displayed.

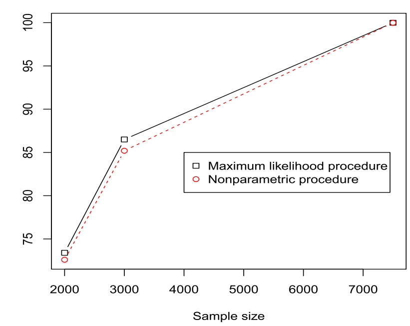

FIG 13. *Empirical powers (in percentage of rejections) when the alternative is a shifted Student* $t(3)$ *for parameter value* $p = 1/3$ *with sample sizes* $n = 2000, 3000, 7500$.

Appendix L: Sensitivity to contiguous alternatives

To illustrate the practical impact of dealing with Section 5.1 type contiguous models, we consider that in model (12) the $(U_i)_{i\geq 1}$ is an i.i.d. sequence of Bernoulli random variables with parameter $p = 1/3$ and $(Z_i)_{i\geq 1}$ is an i.i.d. sequence of Gaussian random variables with mean $\mu = 1.5$ and variance $s = 1$. We fixe $n = 3000$ and choose an exponential distribution with parameter $\lambda = 1/5$ for the sequence $(\varepsilon_i)_{i\geq 1}$. It is difficult to base our simulation on the theoretical rates given by **(A5), (A7), (A8)** since these rates are up to an unknown factor. However, to illustrate the effect of the deviation from H_0, we let the perturbation vary from $\delta = 0.05$ to 0.5 along a set of 8 different values. Figure 14 represents the rejection percentages with respect to this set of values. We can observe that the empirical power is quite sensitive with respect to the (*vanishing* by nature) perturbation factor δ, with about 50% of good decision recovery for $\delta = 0.2$ and more than 95% with $\delta = 0.4$.

Acknowledgement

The authors acknowledge the Office for Science and Technology of the Embassy of France in the United States, especially its antenna in Atlanta, for its valuable support to this work. The authors would like also to thank the Associated Editor along with the two referees for their remarks and insigthfull comments.

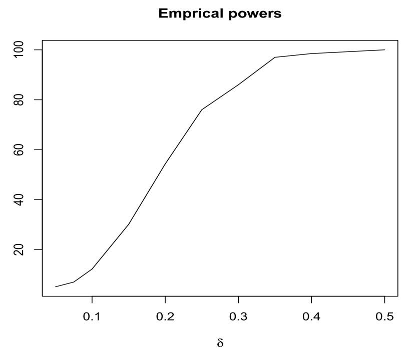

FIG 14. *Rejection percentages under model (12) Laplace type contiguous alternatives with* $\delta \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5\}$ *and* $n = 3000$.

References

[1] Allman, E. S., Matias, C. and Rhodes, J. A. (2009) Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.*, **37**, 3099–3132. MR2549554

[2] Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D. and Levine, A. J. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl Acad. Sci. USA*, **96**, 6745–6750.

[3] Al Mohamad, D. and Boumahdaf, A. (2018) Semiparametric two-component mixture models when one component is defined through linear constraints. *IEEE Trans. Information Theory*, **64**, 795–830. MR3762592

[4] Arias-Castro, E. and Huang, R. (2018) The sparse variance contamination model. *Preprint.* arXiv:807.10785v1.

[5] Balabdaoui, F. and Doss, C. R. (2018) Inference for a two-component mixture of symmetric distributions under log-concavity. *Bernoulli*, **24**, 1053–1071. MR3706787

[6] Di Zio, M. and Guarnera, U. (2013) A contamination model for selective editing. *J. Official Statist.*, **29**, 539–555.

[7] Berrett, T. B., Samworth, R. J., and Yuan, M. (2019) Efficient multivariate entropy estimation via k-nearest neighbour distances. *Ann. Statist.* **47**, 288–318. MR3909934

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

[9] Bordes, L. and Vandekerkhove, P. (2010) Semiparametric two-component mixture model when a component is known: an asymptotically normal estimator. *Math. Meth. Statist.*, **19**, 22–41. MR2682853

[10] Dai, H. and Charnigo, R. (2010) Contaminated normal modeling with application to microarray data analysis. *Can. J. Statist.* **38**, 315–332.

MR2730112

[11] Doukhan, P., Pommeret, D. and Reboul, L. (2015) Data driven smooth test of comparison for dependent sequences. *J. Multivar. Analys.*, **139**, 147–165. MR3349484

[12] Gassiat, E. (2018) Mixtures of nonparametric components and hidden Markov models. Handbook of Mixture Analysis (ed. G. Celeux, S. Fruhwirth-Schnatter, C. Robert, Chap. 12) *To appear*. MR3889699

[13] Ghattas, B., Pommeret, D., Reboul, L. and Yao, A. F. (2011) Data driven smooth test for paired populations. *J. Stat. Plan. Inference* **141**, 262–275. MR2719492

[14] Hedenfalk, I. *et al.* (2001) Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.*, **344**, 539–548.

[15] Inglot, T., Kallenberg, W. C. M. and Ledwina, T. (1997) Data driven smooth tests for composite hypotheses. *Ann. Statist.*, **25**, 1222–1250. MR1447749

[16] Klingenberg, C., Pirner, M. and Puppo, G. (2017) A consistent kinetic model for a two-component mixture with an application to plasma. *Kinet. Relat. Models*, **10**, 445–465. MR3579578

[17] Ledwina, T. (1994) Data-driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.* **89**, 1000–1005. MR1294744

[18] Lindsay, B. G. (1983) The geometry of mixture likelihoods: a general theory. *Ann. Statist.*, **11**, 86–94. MR0684866

[19] Lindsay, B. G. (1989) Moment matrices: applications in mixtures. *Ann. Statist.*, **17**, 722–740. MR0994263

[20] Ma, Y. and Yao, W. (2015) Flexible estimation of a semiparametric two-component mixture model with one parametric component. *Electr. J. Statist.*, **9**, 444–474. MR3326131

[21] McLachlan, G. J., Bean, R. W. and Ben-Tovim Jones, L. (2006) A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays. *Bioinformatics*, **22**, 1608–1615.

[22] Nguyen, V. H. and Matias, C. (2014) On efficient estimators of the proportion of true null hypotheses in a multiple testing setup. *Scan. J. Statist.*, **41**, 1167–1194. MR3277044

[23] Melchior, P. and Goulding, A. D. (2018) Filling the gaps: Gaussian mixture models from noisy, truncated or incomplete samples. *Astronomy and Computing*, **25**, 183–194.

[24] Munk, A., Stockis, J. P., Valeinis, J. and Giese, G. (2010) Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data. *Ann. Instit. Statist. Math.*, **63**, 939–959. MR2822962

[25] Neyman, J. (1937) Smooth test for goodness of fit. *Skandinavisk Aktuarietidskrift*, **20**, 149–199.

[26] Patra, R. K. and Sen, B. (2016) Estimation of a two-component mixture model with applications to multiple testing. *J. Roy. Statist. Soc., Series B*, **78**, 869–893. MR3534354

[27] Podlaski, R. and Roesch, F. A. (2014) Modelling diameter distributions of two-cohort forest stands with various proportions of dominant species:

A two-component mixture model approach. *Math. Biosci.*, **249**, 60–74. MR3173073

[28] Quandt, R. E. and Ramsey, J. B. (1978) Estimating mixtures of normal distributions and switching regressions (with comments). *J. Am. Statist. Ass.*, **73**, 730–752. MR0521324

[29] Robin, A. C., Reyl, C., Derrire, S. and Picaud, S. (2003) A synthetic view on structure and evolution of the Milky Way. *Astron. Astrophys.*, **409**, 523–540.

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

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

[32] Suesse, T., Rayner, J. C. W. and Thas, O. (2017) Assessing the fit of finite mixture distributions. *Aust. N. Z. J. Stat.*, **59**, 463–483. MR3760154

[33] Szegö, G. (1939) *Orthogonal Polynomials*. Colloquium Publications Volume XXIII. Amer. Math. Soc. MR0000077

[34] van der Vaart, A. W. (1998) *Asymptotic Statistics*. Cambridge University Press. MR1652247

[35] Walker, M. G., Mateo, M., Olszewski, E. W., Sen, B. and Woodroofe, M. (2009) Clean kinematic samples in drarf spheroidals: an algorithm for evaluating membership and estimating distribution parameters when contamination is present. *The Astronomical Journal*, **137**, 3109–3138.

[36] van't Wout, A. B. *et al.* (2003) Cellular gene expression upon human immunodeficiency virus type 1 infection of CD4+-T-cell lines. *J. Virol.*, **77**, 1392–1402.

[37] Wylupek, G. (2010) Data driven K-sample tests. *Technometrics*, **52**, 107–123. MR2654991

[38] Xiang, S., Yao, W. and Yang, G. (2019) An overview of semiparametric extensions of finite mixture models. *Statistical Sciences*, **34**, 391–404. MR4017520