Semiparametric false discovery rate model Gaussianity test

Denys Pommeret, Pierre Vandekerkhove

Semiparametric false discovery rate model Gaussianity test

Denys Pommeret and Pierre Vandekerkhove

Institut Mathématique de Marseille
Campus de Luminy,
13288 Marseille Cedex 9, France
e-mail: denys.pommeret@univ-amu.fr

Université Paris-Est Marne-la-Vallée
LAMA (UMR 8050), UPEMLV
F-77454, Marne-la-Vallée, France
and
UMI Georgia Tech - CNRS 2958,
Georgia Institute of Technology, USA
e-mail: pierre.vandekerkhove@u-pem.fr

Abstract: In this paper we investigate a semiparametric testing approach to answer if the Gaussian assumption made by McLachlan *et al.* (2006) on the unknown component of their false discovery type mixture model was a posteriori correct or not. Based on a semiparametric estimation of the Euclidean parameters of the model (free from the Gaussian assumption), our method compares pairwise the Hermite coefficients of the model estimated directly from the data with the ones obtained by plugging the estimated parameters into the Gaussian version of the false discovery 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 $\chi^2(1)$-distributed and study its behavior under different types of alternatives, including contiguous nonparametric alternatives. Several level and power studies are numerically conducted on models close to those considered in McLachlan *et al.* (2006) to validate the suitability of our approach. We also discuss the lack of power of the maximum likelihood version of our test in a neighborhood of certain non identifiable situations and implement our testing procedure on the three microarray real datasets analyzed in McLachlan *et al.* (2006) and comment our results. Finally we discuss possible extension of this work to more general models.

MSC 2010 subject classifications: Primary 62F03, 28C20; secondary 33C45.
Keywords and phrases: Asymptotic normality, Chi-squared test, False Discovery Rate, maximum likelihood estimator, nonparametric contiguous alternative, semiparametric estimator, two-component mixture model..

1. Introduction

Let us consider n independent and identically distributed (iid) random variables $(T_1, \ldots, T_n)$ coming from the two-component mixture model with probability density function (pdf) g defined by

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

where f_0 is a known pdf and where the unknown parameters are the mixture proportion $p \in (0,1)$ and the pdf $f \in \mathcal{F}$ (a given class of densities). This class of models is especially suitable for detection of differentially expressed genes under various conditions in microarray data analysis. For this purpose a test statistic is built for each gene. Under the null hypothesis, corresponding to a lack of difference of expression under the various conditions, this statistic is supposed to have a known distribution (Student, Fisher, etc.). We then observe thousands of genes, corresponding in practice to thousands of statistical tests (the T_i's). The sample generated in this way comes from a mixture of distributions: the known distribution f_0 (genes under the null hypothesis) and an unknown alternative distribution corresponding to f. Using Bayes Theorem in (1.1), the posterior probability that the ith gene is not differentially expressed, or equivalently f_0-distributed, is then given by

$$\tau_0(T_i) = \frac{pf_0(T_i)}{g(T_i)}, \quad i = 1, \ldots, n. \tag{1.2}$$

In that framework, the above gene-specific posterior probabilities provide a good tool for statistical inference about differential expression. The posterior probability $\tau_0(\cdot)$ has been termed the *local false discovery rate* (local FDR) by Efron and Tibshirani (2002). It quantifies the gene-specific evidence for each gene. As noted by Efron (2004), the use of this quantity can be viewed as an empirical Bayes version of the Benjamini-Hochberg (1995) methodology, using densities rather than tail areas. For convenience and as a tribute to Efron and Tibshirani (2002), model (1.1) will be so called the *false discovery mixture model*. It can be seen from (1.2) that in order to use this posterior probability of non-differential expression in practice, we need to be able to estimate p, the mixture density g and the null density f_0, or equivalently, the ratio of densities f_0/g. Efron *et al.* (2001) have developed a simple empirical Bayes approach to this problem with minimal assumptions. This problem has been studied since under more specific assumptions, including works by Newton *et al.* (2004), Lönnstedt and Speed (2002), Pan *et al.* (2003), Zhao and Pan (2003), Broët *et al.* (2004), Do *et al.* (2005) and *et al.* (2006), among many others. In McLachlan *et al.* (2006), similarly to Efron (2004), the authors suggest to transform the observed value of the test statistic to a Z-score given by

$$Z_i = \Phi^{-1}(1 - P_i), \quad i = 1, \ldots, n,$$

where $P_i = 1 - F_0(T_i) + F_0(-T_i)$ denotes the p-value corresponding to the original test statistic T_i and Φ denotes the cumulative distribution function (cdf) of the $\mathcal{N}(0,1)$ distribution. If F_0 is the true null cdf, then the null distribution of the new test statistic Z_i is exactly standard normal. The main advantage of such a transformation is that it provides a parametric version of model (1.1) that is easy to fit by using a standard expectation-maximization (EM) algorithm. In fact the common law of the Z_i's is to be represented by a normal mixture

model, which is model (1.1) where

$$f(x) = \sum_{k=1}^{q} p_k f_{(\mu_k, s_k)}(x), \quad x \in \mathbb{R}, \tag{1.3}$$

and $f_{(\mu,s)}$ denotes the normal pdf with mean μ and variance s and $\sum_{k=1}^{q} p_k = 1$. The meaning of the above representation is that f can basically be approximated with arbitrary accuracy by taking q sufficiently large in the normal mixture representation (1.3). In McLachlan *et al.* (2006) expression (15), the authors consider empirically that in the datasets they had to analyze, it was sufficient to consider $q = 1$ in (1.3).

The aim of the present paper is to answer more precisely if the normality assumption for f is realistic or not. The solution we propose here is to consider the estimation and testing methodology suggested in the semiparametric setup by Bordes and Vandekerkhove (2010) adapted to the following specific model

$$g(x) = (1-p) f_{(0,1)}(x) + p f(x-\mu), \quad x \in \mathbb{R}, \tag{1.4}$$

where $f \in \mathcal{S}^\star$ (the set of zero-symmetric pdf's). For a general overview about semiparametric approaches in missing data models we recommend the reading of Gassiat (2018). Note that the test against a specific distribution, proposed in Bordes and Vandekerkhove (2010, Section 4.1), does not allow to test versus a complete class of pdf's which is our goal here. To the best of our knowledge only the MLE-based recent paper by Suesse *et al.* (2017) also addresses the componentwize goodness of fit testing problem for mixture models. The main idea of our test is based on the Neyman (1934) smooth test procedure which consists in estimating the expansion coefficients of f in an orthogonal basis, first assuming $f \in \mathcal{S}$ (the set of symmetric pdf's with respect to a location parameter $\mu \in \mathbb{R}$), and to compare this estimates to those obtained by assuming $f \in \mathcal{G}$. This approach has been used in Doukhan *et al.* (2015) (see also references therein), but the specificity of the two component mixture model necessitates a particular adaptation of the Neyman smooth tests. 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 our paper, the MLE-based approach of Suesse *et al.*, restricted to model (1.4), does not allow to investigate the asymptotic behavior of the test statistic under alternative assumptions (possibly contiguous) since the asymptotic behavior of the MLE cannot be controlled properly under Gaussianity misspecification.

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 divergence under the alternative; Section 5 is devoted to the study of our testing procedure under contiguous nonparametric alternatives; in Section 6 we discuss the choice of the reference measure when considering orthogonal bases for the unknown density decomposition; Section 7 is dedicated to a simulation-based empirical and

power levels study; in Section 8 we proceed with the application of our testing method to the datasets (breast cancer, colon cancer, HIV) previously studied in McLachlan *et al.* (2006). Finally in Section 9 we discuss further leads of research connected with the FDR Gaussianity test problem.

2. Testing problem

Let consider an iid sample denoted for simplicity $(X_1, \ldots, X_n)$, playing the same role as the sample $(T_1, \ldots, T_n)$ presented in the very beginning of the Introduction section, drawn from a pdf g defined in (1.1) with respect to a given reference measure ν. In that model we suppose that the pdf with respect to ν, denoted generically latter on ν-pdf, f_0 is supposed to be known with a variance s_{f_0} fixed to be equal to 1, when the ν-pdf f, the mixture proportion $p \in (0,1)$ are both unknown. The problem addressed in this paper deals with the normality testing of the unknown component f assuming the fact that f belongs to $\mathcal{S}$, the set of zero-symmetric densities provided with identifiability conditions in Bordes and Vandekerkhove (2010, p. 25). More precisely, denoting $\mathcal{G} = \{f_{(\mu,s)}; (\mu, s) \in \Lambda\}$ the set of ν-normal densities, with mean μ and variance s where (μ, s) is supposed to belong to a compact set Λ of $\mathbb{R}^{+*} \times \mathbb{R}$, our goal is to test

$$H_0 : f \in \mathcal{G} \quad \text{vs} \quad H_1 : f \in \mathcal{S} \backslash \mathcal{G}. \tag{2.1}$$

Our test procedure is based on the Neyman's one (1937) and consists in estimating the expansion coefficients of the unknown ν-pdf 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 belong strictly to the sub-parametric family $\mathcal{G}$. 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 $\mathcal{Q} = \{Q_k; k \in \mathbb{N}\}$, an ν-orthogonal basis satisfying $Q_0 = 1$ and such that

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

with $\delta_{jk} = 1$ if $j = k$ and 0 otherwise. We assume that $\mathcal{Q}$ is an $L^2(\mathbb{R}, \nu)$ Hilbert basis, which is satisfied if there exists $\theta > 0$ such that $\int_{\mathbb{R}} e^{\theta|x|} \nu(dx) < \infty$, and that the following integrability conditions are satisfied:

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

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

$$g(x) = \sum_{k\geq 0} a_k Q_k(x) \quad \text{with} \quad a_k := \int_{\mathbb{R}} Q_k(x) g(x) \nu(dx)/q_k^2,$$
$$f_0(x) = \sum_{k\geq 0} b_k Q_k(x) \quad \text{with} \quad b_k := \int_{\mathbb{R}} Q_k(x) f_0(x) \nu(dx)/q_k^2,$$
$$f(x) = \sum_{k\geq 0} c_k Q_k(x) \quad \text{with} \quad c_k := \int_{\mathbb{R}} Q_k(x) f(x) \nu(dx)/q_k^2.$$

From (1.1) we have

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

Let us denote by Z a $\mathcal{N}(0,1)$ random variable and consider

$$\alpha_k(\mu, s) := \mathbb{E}(Q_k(\sqrt{s}Z + \mu))/q_k^2.$$

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

$$H_0 : \ a_k = (1-p)b_k + p\alpha_k(\mu, s), \qquad \text{for all } k \geq 1. \tag{2.2}$$

Since the distribution of 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 avoid possible compensation phenomenon under H_1 between the estimation of $\vartheta := (p, \mu)$ and the estimation of the 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.* (2006) and studied more deeply in Bordes and Vandekerkhove (2010). Indeed the Maximum Likelihood Estimator (MLE) $(\widehat{p}_n, \widehat{\mu}_n, \widehat{s}_n)$ under the Gaussian assumption tends to provide the best Gaussian fitted model when the semiparametric estimator of Bordes and Vandekerkhove (2010) is not affected by this constraint and can provide very distant, euclidean and functional, estimations under H_1 (when the model is misspecified under the Gaussianity assumption). In the same way, given $s_{f_0} = 1$ and considering the relation (1.4), the estimator of s is obtained by the H_0-free semiparametric plug-in moment method:

$$\bar{s}_n := \frac{\bar{M}_{2,n} - (1-\bar{p}_n)}{\bar{p}_n} - \left(\frac{\bar{M}_{1,n}}{\bar{p}_n}\right)^2, \tag{2.3}$$

where $\bar{M}_{1,n} := n^{-1}\sum_{i=1}^{n} X_i$ and $\bar{M}_{2,n} := n^{-1}\sum_{i=1}^{n} X_i^2$. The estimator of $\alpha_k(\mu, s)$ is obtained by using a standard plug-in approach, that is:

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

Now looking at the H_0 reformulation in (2.2) we 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)

$$T_{k,n} := nU_{k,n}^{\top}\widehat{D}_{k,n}^{-1}U_{k,n}, \tag{2.4}$$

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

$$D_{k,n} := diag(\mathbb{V}(R_{1,n}), \ldots, \mathbb{V}(R_{k,n})),$$

normalizing the test statistic as in Munk *et al.* (2009). To avoid instability in the evaluation of $\widehat{D}_{k,n}^{-1}$, following Doukhan *et al.* (2015), we add a trimming term $e(n)$ to every i-th, $i = 1, \ldots, k$, diagonal element of $\widehat{D}_{k,n}$ as follows:

$$\widehat{D}_{k,n}[i] := \max(\widehat{\mathbb{V}}(R_{i,n}), e(n)), \quad 0 \leq i \leq k, \tag{2.5}$$

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

Following Ledwina (1994) and Kallenberg and Ledwina (1995), 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 parcimouniously (trade-off between H_0 departure detection and complexity of the procedure involved by index k) the "*best*" rank k for looking at $T_{k,n}$:

$$S_n := \min\Big\{ \underset{1\leq k\leq d(n)}{\operatorname{argmax}} \left(s(n)T_{k,n} - \beta_k pen(n)\right)\Big\}, \tag{2.6}$$

where $s(n) \to 0$ is a normalizing rate, $d(n) \to +\infty$ as $n \to +\infty$, $pen(n)$ 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 $\beta_k = k$, $k \geq 1$, and $pen(n) = \log(n)$, $n \geq 1$. To match the asymptotic normality regime, under H_0, of the test statistic $T_{k,n}$ defined in (2.4), 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:

$$s(n) = n^{\lambda-1}, \text{ with } \lambda \in (0, 1/2). \tag{2.7}$$

The above calibration is connected with the *a.s.* convergence rate of the estimators $\bar{p}_n$ and $\bar{\mu}_n$ (see Theorem 3.1 in Bordes and Vandekerkhove, 2010). Note that the selection rule in (2.6), adapted to the semiparametric framework, strongly differs from the BIC criterion used by Suesse *et al.* (2017, p. 9).

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

$$H_0 : \exists s \in \mathbb{R}^{+*} \text{ s.t. } F_s = F_{(0,1)} \quad vs \quad H_1 : \nexists s \in \mathbb{R}^{+*} \text{ s.t. } F_s = F_{(0,1)},$$

where respectively, $F_s(\cdot) := F(\sqrt{s}\times\cdot)$ *and* $F_{(0,1)}$ *are the cdf of* $f \in \mathcal{S}$ *and* $f_{(0,1)}$*. In such a perspective we could have used a strategy inspired from the simple hypothesis test of Bordes and Vandekerkhove (2010, Section 4.1). Since according to Theorem 3.2 in Bordes and Vandekerkhove (2010) the semiparametric estimator* $\widehat{F}_n$ *of* F *satisfies a functional central limit theorem, one could consider* s_n *in (2.3) 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 (2.1), 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 (2.6) satisfies $d(n) = O(\log(n)e(n))$, where $e(n)$ is the trimming term in (2.5).
- **(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 $(\mu,s) \in \Lambda \subset \mathbb{R}\times\mathbb{R}^{+*}$,

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

  where $\dot{\alpha_k}$ denotes the gradient $(\partial\alpha_k/\partial\mu, \partial\alpha_k/\partial s)^T$ and $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^2$.
- **(A3)** There exists a nonnegative constant M_3 such that for all $(k,i) \in \mathbb{N}^*\times\mathbb{N}^*$,

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

Under these three conditions we state the following theorem.

Theorem 2. *If assumptions* **(A1-3)** *hold, then, under* H_0, S_n *converges in Probability towards 1 as* $n \to +\infty$.

Corollary 3. *Under* **(A1)-(A3)**, *the test statistic* $T(n)$ *converges in law towards a* χ^2*-distribution with one degree of freedom as* $n \to +\infty$.

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 MLE detailed the supplementary material Section 14. In this case the rate of the selection rule is the standard one, which is namely* $s(n) = 1$.

4. Asymptotic behavior under H_1

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

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

We would like to stress out the fact that the identifiability conditions supposed when considering the class of densities $\mathcal{S}$, see definition in Section 2, are crucial in the proof of Proposition 1. As mentioned in Bordes, Delmas and Vandekerkhove (2006), there exists various non identifiability cases for model (1.4). Let us remind the following one from Bordes and Vandekerkhove (2010):

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

where φ is an even pdf, $p \in (0,1)$ and $f(x) = (\varphi(x-\mu) + \varphi(x+\mu))/2$. This example is very interesting since it clearly shows the danger of estimating model (1.4) when the pdf of the unknown component has exactly the same shape as the known pdf. In particular if φ is a given Gaussian distribution we could possibly either reject or accept H_0 with our testing procedure depending on the convergence of our semiparametric estimators. Indeed the MLE would converge towards the natural underlying Gaussian model and the semiparametric method could possibly converge towards both solutions. To avoid this very well identified concern, we recommend to check if the departures between the MLE estimator and the semiparametric one is not driven by a factor 2, *i.e* $\widehat{\mu}_n \approx 2\bar{\mu}_n$ and $\widehat{p}_n \approx \bar{p}_n/2$. If so, we suggest to initialize the semiparametric approach close the MLE estimator to force it to detect the possibly existing Gaussian f-component in model (1.1).

5. Contiguous alternatives

5.1. Detected contiguous alternatives

We consider in this section a *vanishing* convolution-class of nonparametric contiguous alternatives. More specifically, the null hypothesis consists here in con-

sidering that the observed sample $\mathbf{X}^n := (X_1, \ldots, X_n)$ comes from

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

where $(U_i)_{i\geq 1}$ is an iid sequence of $\mathcal{B}(p)$ random variable and $(Y_i, Z_i)_{i\geq 1}$ is an iid sequence of random variables distributed according to $f_0 \otimes f(\cdot - \mu)$, where f is a $\mathcal{N}(0, s)$-df. On the other hand, for each $n \geq 1$, the contiguous alternative consists in the fact that the observed sample $\mathbf{X}^{(n)} := (X_1^n, \ldots, X_n^n)$ comes from a *row independent* triangular array:

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

where $Z_i^n := Z_i + \delta_n \varepsilon_i$, $(\varepsilon_i)_{i\geq 1}$ is an iid sequence of non Gaussian random variables, independent from the Z's and $\delta_n \to 0$ as $n \to +\infty$ (vanishing factor). 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 cdf 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 the supplementary material Section 11, defined by

$$\widehat{g}_n^{(n)}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i^n}{h_n}\right), \quad \forall x \in \mathbb{R}, \tag{5.1}$$

where the bandwidth h_n satisfies $h_n \to 0$, $nh_n \to +\infty$ and K is a symmetric kernel density function detailed in Section (11). We will denote also by $\mathbb{E}^{(n)}$ and $\mathbb{P}^{(n)}$ the expectation and probability distribution under the alternative $H_1^{(n)}$ and consider the following assumptions:

(A4) The bandwidth setup is $h_n = n^{-1/4-\gamma}$ with $\gamma \in (0, 1/12)$.
(A5) The vanishing factor satisfies $\delta_n = n^{-3/4-\xi}$, with $3\gamma < \xi < 2\gamma + 1/2$.

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 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* **(A1-5)** *hold, then, under* H_1^*, S_n *converges in Probability towards 1 and* $T(n)$ *converges in law towards a* χ^2*-distribution with one degree of freedom, as* $n \to +\infty$.

5.2. Undetected contiguous alternatives

Combining Assumptions **(A4)** and **(A5)**, we clearly have $0 < \xi < 2/3$ and then there exists $\widetilde{\xi} = 3/4 + \xi \in (3/4, 17/12)$ such that $\delta_n = n^{-\widetilde{\xi}}$. The convergence rate

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

(A6) $\mathbb{E}(\varepsilon) = 0$ and there exists $0 < \xi' < 1/4$ such that $\delta_n = n^{-\xi'}$.
(A7) $\mathbb{E}(\varepsilon) \neq 0$ and there exists $0 < \xi'' < 1/8$ such that $\delta_n = n^{-\xi''}$,

where ε denotes a generic non Gaussian random variable involved in the above definition of the Z^n's. The rate in **(A6)** will control the mean deviation due to the non Gaussian perturbations ε and the rate given in **(A7)** will allow to control the variance of these perturbations when there is no mean deviation.

Proposition 3. *If assumptions* **(A6)** *or* **(A7)** *holds, then, under* H_1^*, $T(n)$ *converges a.s. towards* $+\infty$. *Moreover, under* **(A6)** S_n *converges a.s. towards 1, and under* **(A7)** S_n *converges a.s. towards 2.*

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 $\mathcal{Q} = \{Q_k,\ k \in \mathbb{N}\}$. Since the test procedure is dedicated to Gaussianity, we consider here two measures on the real line, the Gaussian one and the Lebesgue one. The verification of conditions **(A2–3)** for these two measures is relegated in Appendix.

Gaussian reference measure. In practice, in the present paper, we chose for ν the standard normal distribution. The set $\mathcal{Q}$ is constructed from the $f_{(0,1)}$-orthogonal Hermite polynomials defined for all $k \geq 0$ by:

$$H_k(x) = k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m x^{k-2m}}{m!(k-2m)!2^m}, \quad x \in \mathbb{R}. \tag{6.1}$$

We have $\|H_k\|^2 = k!$, and 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 (6.1) and let* $Q_k(x) = H_k(x)$, *for all* $x \in \mathbb{R}$. *Then conditions* **(A2–3)** *are satisfied.*

Lebesgue reference measure. Another ν reference measure simple choice could have been the Lebesgue measure over $\mathbb{R}$. In that case, we could have consider the set of orthogonal Hermite functions defined by:

$$\mathcal{H}_k(x) = h_k(x) \exp(-x^2/2), \quad x \in \mathbb{R}, \tag{6.2}$$

where $h_k(x) = 2^{k/2} H_k(\sqrt{2}x)$, with H_k defined in (6.1). We have $\|\mathcal{H}_k\|^2 = k!2^k$.

Lemma 2. *Let $\mathcal{H}_k$ be defined by (6.2) and let $Q_k(x) = \mathcal{H}_k(x)$, for all $x \in \mathbb{R}$. Then conditions* **(A2–3)** *are satisfied.*

Test construction. The computation of the test statistic $T(n) = T_{S_n,n}$, see expressions (2.4) and (2.6), is grounded on the computation of the $\alpha_i(\mu, s)$'s quantities. We detail here the expression of $R_{1,n}$ and $\mathbb{V}(R_{1,n})$ when the reference measure is Gaussian associated with 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

$$\begin{aligned} R_{1,n} &= n_1^{-1} \sum_{i=1}^{n_1} X_i - \bar{p}_{n_2} \bar{\mu}_{n_3}, \quad \text{and} \\ \mathbb{V}(R_{1,n}) &= \mathbb{V}(X)/n_1 + \mathbb{V}(\bar{p}_{n_2})\mathbb{V}(\bar{\mu}_{n_3}) + \mathbb{V}(\bar{p}_{n_2})\mathbb{E}(\bar{\mu}_{n_3})^2 + \mathbb{E}(\bar{p}_{n_2})^2 \mathbb{V}(\bar{\mu}_{n_3}). \end{aligned}$$

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

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

where S^2_{X,n_1} denotes the empirical variance based on $(X_1, \ldots, X_{n_1})$, and v_{p,n_2} (resp. v_{μ,n_3}) denotes the consistent estimator of $\mathbb{V}(\bar{p}_{n_2})$ (resp. $\mathbb{V}(\bar{\mu}_{n_3})$) obtained from Bordes and Vandekerkhove (2010, p. 40).

7. Simulation study

The computation of the test statistic first requires the choice of $d(n)$, $e(n)$ and $s(n)$. A previous study showed that the empirical levels and powers were overall weakly sensitive to $d(n)$ for $d(n)$ large enough. From this preliminary study we decided to set $d(n)$ equal to 10. The trimming $e(n)$ was calibrated equal to $(\log\log(n))^{-1}$. The normalization $s(n) = n^{\alpha-1}$ was chosen close enough to $n^{-1/2}$, with α equal to $2/5$, which seemed to provide good empirical levels.

Secondly, since the pdfs considered in our set of simulation are supported by $\mathbb{R}$ we used the standard Gaussian distribution for ν and its associated Hermite polynomials for $\mathcal{Q}$ (see Appendix). 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 was standardly fixed to 5%.

7.1. Comparing semiparametric and maximum likelihood approaches

In our testing procedure we estimate p, μ by the semiparametric (SP) estimators proposed in Bordes and Vandekerkhove (2010) instead of the maximum likelihood (ML) estimator. In the same way our estimation of s, see expression (2.3),

is H_0-free contrary to the ML 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 SP and ML approaches under null models. However, under certain types of alternatives, the ML 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 notice that

$$g(x) = (1-p)f_{(0,1)}(x) + ph_{a,s}(x-\mu), \quad x \in \mathbb{R}, \tag{7.1}$$

where $h_{a,s}(x) := (f_{(0,s)}(x-a) + f_{(0,s)}(x+a))/2$, $a \neq 0$, turns to satisfy, when $a = \mu$ and $s = 1$, the following rewritting

$$g(x) = (1-\frac{p}{2})f_{(0,1)}(x) + \frac{p}{2}f_{(0,1)}(x-2\mu), \quad x \in \mathbb{R}. \tag{7.2}$$

In this case there are two different parametrizations for (7.1): one that we call the *null parametrization*, coinciding with H_0 with null parameters $p_0 = p/2, \mu_0 = 2\mu$ and $s_0 = 1$ (right hand side of (7.2)). The other one is called the *alternative parametrization*, coinciding with H_1 with $p_1 = p, \mu_1 = \mu$ and $s_1 = \mu^2 + 1$ (right hand side of (7.1)). By construction the ML will favor the null parameters. We study now this phenomenon through a set of simulations where the parameters are $\mu = 4,\ s = 1$ and $p = 0.4$. The corresponding densities of h and g are respectively displayed in Fig. 1. For comparison, we used the same initial values in SP and ML 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 $(p, \mu, s) = (0.4, 4, 17)$.

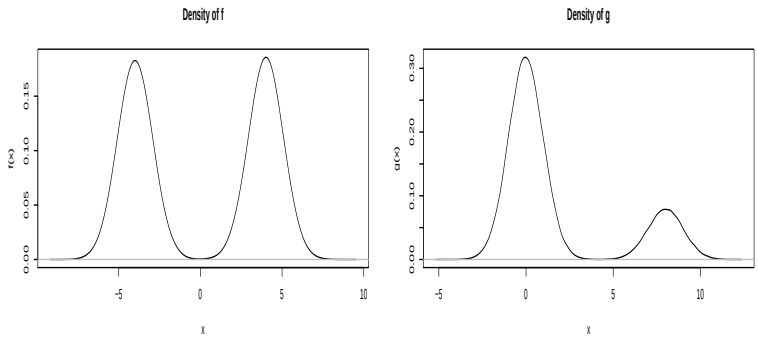

Fig 1: h pdf (left) and g pdf (right) corresponding to model (7.1) with $\mu = 4,\ s = 1$ and $p = 0.4$.

It is of interest to study the behavior of the SP and ML 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*

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

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

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

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

Mean deviation trap. We consider deviations from the null model obtained by considering $a = 3, 2, 1$ and $s = 1$. Fig. 2 shows the g pdf under these respective alternatives. It can be observed that, if we try to visually detect a mixture of two Gaussian distributions, the pdf of the left-side component moves clearly aside the Gaussian distribution family as a moves away from $\mu = 4$.

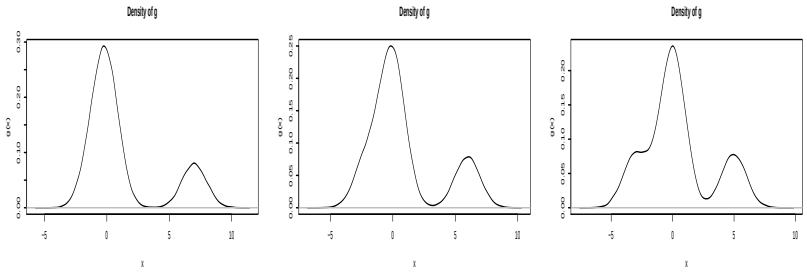

Fig 2: g pdf in model (7.1) when $a = 3, 2, 1$, $\mu = 4$ and $s = 1$.

Fig. 3 illustrates the difficulty of the ML estimator to recognize the alternative model when the mean deviation is not distant enough (here $a = 3$). Based on

a run of 200 repetitions, it is shown that the ML estimation is trapped at the null parametrization which namely is $(p, \mu, s) = (0.2, 7, 1)$ when on the opposite, the SP estimation detects the correct $(p, \mu, s) = (0.4, 3, 17)$ alternative parametrization.

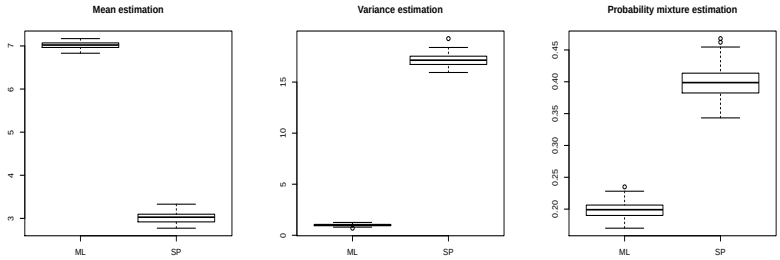

Fig 3: Boxplot for ML and SP estimators of m, s, p when $n = 1000$, under mean deviation $a = 3$, based on 200 repetitions.

In Fig. 4 we display respectively the empirical power of our testing procedure based on the ML and the SP approach for $a = 3, 2, 1$ and for $n = 1000, 2000, 5000$. As expected the ML approach barely detects the alternative for small values of n. 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. 2, the second components looks pretty much Gaussian even for $a = 1$.

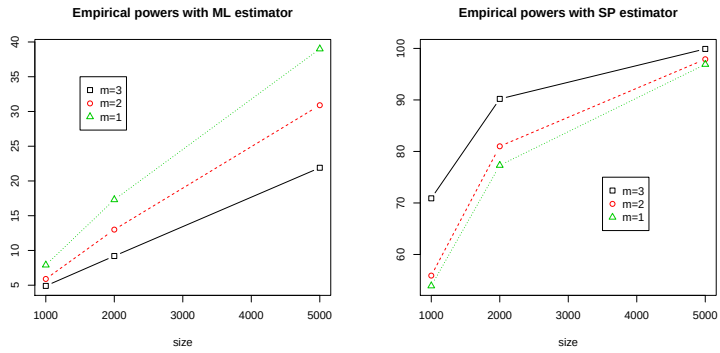

Fig 4: Empirical powers obtained with the ML approach (left) and SP approach (right) under *mean deviation* when $a = 3, 2, 1$.

Variance deviation trap. We consider the variance deviations $s = 2, 3, 4$. Fig. 5 shows the g pdf under these alternatives.

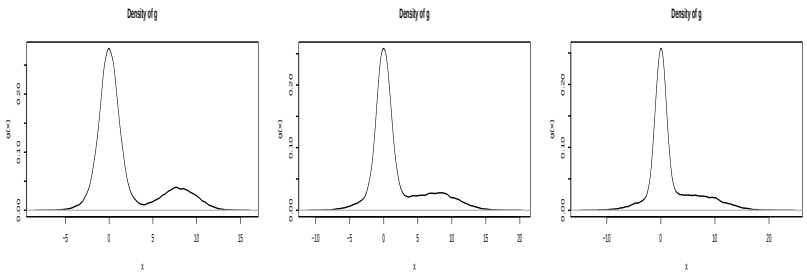

Fig 5: g pdf in model (7.1) with $a = \mu = 4$ and $s = 2, 3, 4$.

Empirical powers are displayed in Fig. 6. We can observe that both powers associated with the ML and SP approach increase according to the variance deviation but it is worth to notice that the detection based on the ML approach is again very poor compared to the SP approach.

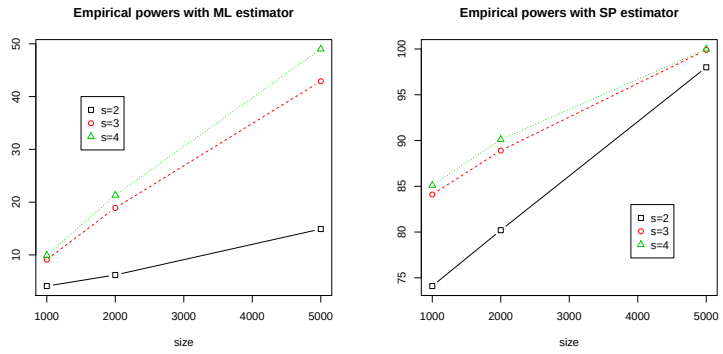

Fig 6: Empirical powers obtained with the ML approach (left) and SP approach (right) under variance deviation when $s = 2, 3, 4 \neq 1$.

As a conclusion, this set of numerical experiments shows the clear interest, in terms of testing power, of considering the SP *vs* ML approach especially in a close neighborhood of non-identifiable type (1.4) Gaussian models.

7.2. Empirical levels

McLachlan *et al.* (2006) considered the two-component mixture model (1.4) 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 *empirical level* of the test is defined as the percentage of rejection of the null hypothesis over 100 repetitions of the test statistic. The collection of empirical levels obtained for this set of simulated examples is reported in Fig. 7. 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.

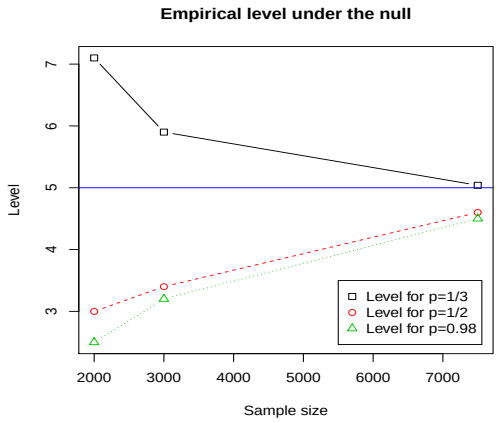

Fig 7: Empirical levels for parameter values $(p, \mu, s) = (1/3, 1.5, 1)$ ($\square$), $(p, \mu, s) = (0.5, 1.5, 2)$ ($\circ$) and $(p, \mu, s) = (0.98, -0.15, 0.8)$ ($\triangle$) with sample sizes $n = 2000, 3000, 7500$.

7.3. Empirical powers

We consider Student and Laplace distributions as 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. 8 and 9.

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 7.1, 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 the Appendix section Fig. 11. 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. 11 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.

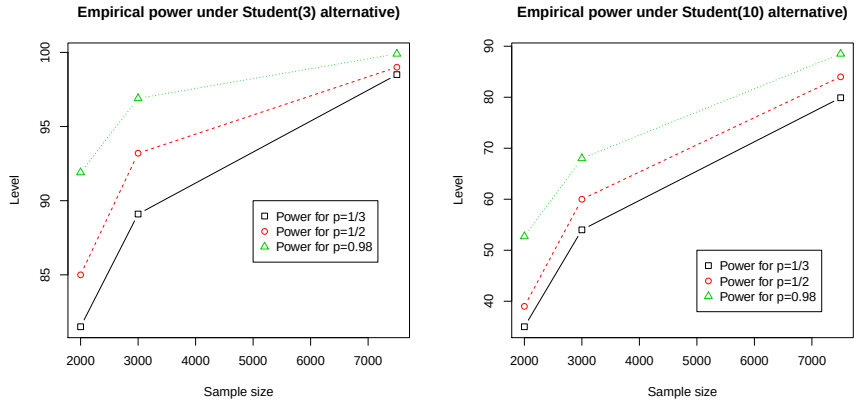

Fig 8: Empirical powers with alternative a shifted Student $t(3)$ (left) and a shifted Student $t(10)$ for parameter values $p = 1/3$ ($\square$), $p = 1/2$ ($\circ$) and $p = 0.98$ ($\triangle$) with sample sizes $n = 2000, 3000, 7500$.

The empirical powers for Laplace alternatives are given in Fig. 9. 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 we can see in Fig. 11. These alternatives are globally very well detected by our

method and the power increases strongly when p gets closer to 1 (see Fig. 9 curve in green).

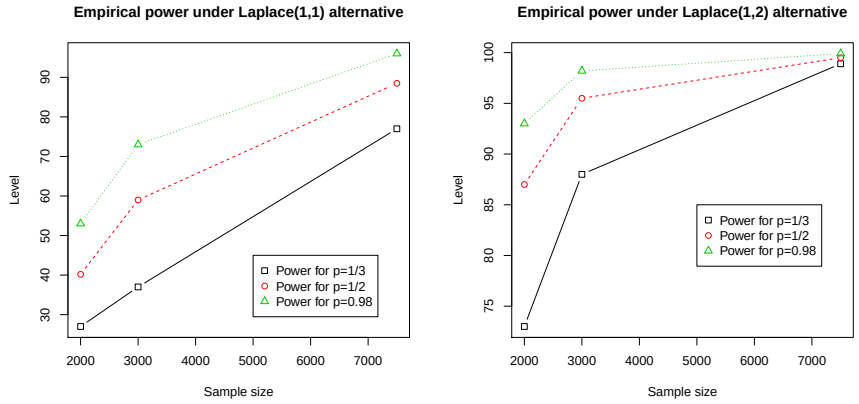

Fig 9: Empirical powers with alternative a Laplace $\mathcal{L}(1,1)$ (left) and a Laplace $S\mathcal{L}(1,2)$ (right) for parameter values $p = 1/3$ ($\square$), $p = 1/2$ ($\circ$) and $p = 0.98$ ($\triangle$) with sample sizes $n = 2000, 3000, 7500$.

8. Real data sets

We consider 3 datasets arising from the bioinformatics literature and studied in McLachlan *et al.* (2006). Fig. 10 shows the non parametric kernel estimations of their pdf's. Each of them deals with genes expressions modeled by the two-component mixture model (1.4) 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, made for instance by MacLachlan *et al.* (2006), was a posteriori correct or not. Let us remark that the questioning about parametric modeling of the f component in the false discovery problem (1.1) is mainly the cause of all the literature about semiparametric mixture models with one known component. For a good screening of these methods (which do not impose any parametric structure on f) we recommend Bordes *et al.* (2006) which suppose the zero-symmetry of f, Guan *et al.* (2008) in the compact support case, Ma and Yao (2015) in the symmetric case with tails conditions, or Yang *et al.* (2013) for parametric and semiparametric mixture approaches comparisons in the local false discovery rate problem.

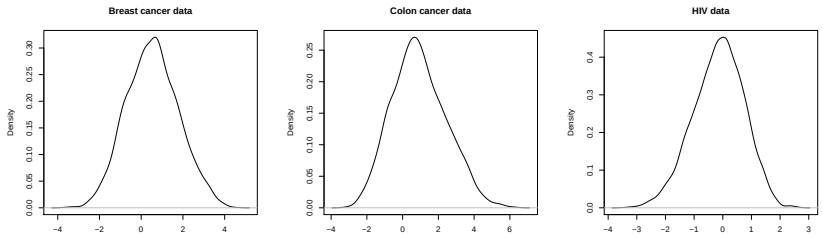

Fig 10: Respectively the kernel estimators of the breast data, colon data and HIV data distributions.

Breast cancer data. We consider the breast cancer data studied in Hedenfalk *et al.* (2001). It consists in $n = 3226$ gene expressions in breast cancer tissues 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.* (1999). 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.

HIV data. We consider the HIV dataset of vant' Wout *et al.* (2003). 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 ML and SP 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 pdf in Fig. 10).

9. Discussion and perspectives

In this paper we focused our work on the Gaussianity testing problem of the unknown component in the FDR model (1.4). However, it is very important to notice that the theoretical results we developed here can be extended to any symmetric distribution, or any non-symmetric distribution provided with relevant identifiability constraints. For this latter case, we recommend the recent work by Al Mohammad and Boumahdaf (2018) who consider in model (1.4) 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, resp. positive, real line, when Legendre, or cosine bases could be used for densities with compact support. For each case, the use of the ML or SP approach could be again discussed. On the other hand, as it has been demonstrated in Section 7.1, the SP testing approach shows better power performances than the ML version especially in the neighborhood of the *mean and variance deviation* pivotal situations. 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* setup. This would namely consist in looking at the asymptotic behavior of our test when replacing respectively the parameters a and s in the mean and variance deviation trap setups by sequences a_n and s_n converging respectively towards μ 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 SP Euclidean estimator associated with a triangular array driven by the above asymptotic parametrization, see Remark 7 in the supplementary material Section. Future work is also to consider a K-sample extension, $K \geq 2$, in the spirit of Wylupeck (2010), Ghattas *et al.* (2011), or more recently Doukhan *et al.* (2015). More precisely, we could test the equality of K unknown components through K observed mixture models.

Acknowlegement. 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.

10. Appendix: proofs of the main results

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 (6.1) and notice that for all $(\mu, s) \in \Lambda$,

$$|\mathbb{E}(H_k(\sqrt{s}Z+\mu))| \leq k! \sum_{m=0}^{\ell} \frac{\mathbb{E}((\sqrt{s}X+\mu)^{2(\ell-m)})}{m!(2(\ell-m))!2^m}. \tag{10.1}$$

Now since

$$\begin{aligned}\mathbb{E}((\sqrt{s}Z+\mu)^{2(\ell-m)}) &= \sum_{j=0}^{2(\ell-m)} C_j^{2(\ell-m)} \sqrt{s}^j \mathbb{E}(Z^j)|\mu|^{2(\ell-m)-j} \\ &\leq \mathbb{E}(Z^{2(\ell-m)})(\sqrt{s}+|\mu|)^{2(\ell-m)} \\ &= \frac{2(\ell-m)!}{2^{(\ell-m)}(\ell-m)!}(\sqrt{s}+|\mu|)^{2(\ell-m)}, \end{aligned} \tag{10.2}$$

including (10.2) in (10.1), we obtain:

$$\begin{aligned}|\mathbb{E}(H_k(\sqrt{s}Z+\mu))| &\leq k! \sum_{m=0}^{\ell} \frac{(\sqrt{s}+|\mu|)^{2(\ell-m)}}{m!(\ell-m)!2^\ell} \\ &= \frac{k!}{\ell!}\left(\frac{(\sqrt{s}+|\mu|)^2+1}{2}\right)^\ell. \end{aligned} \tag{10.3}$$

Since $\alpha_k(\mu,s) = \mathbb{E}(H_k(\sqrt{s}Z+\mu))/q_k^2$, with $q_k^2 = k!$, we deduce from (10.3) 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 prove 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 $k = 2\ell$, $\ell \geq 1$:

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

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

$$\begin{aligned}\mathbb{E}(H_k^2(X_1)) &= (1-p)\mathbb{E}(H_k^2(Z))+p\mathbb{E}(H_k^2(\sqrt{s}Z+\mu))\\ &= (1-p)k!+p\mathbb{E}(H_k^2(\sqrt{s}Z+\mu)).\end{aligned}\tag{10.4}$$

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:

$$\begin{aligned}\mathbb{E}(H_k^2(\sqrt{s}Z+\mu)) &\leq (k!)^2\left(\sum_{m=0}^{\ell}\frac{\sqrt{\mathbb{E}\left((\sqrt{s}Z+\mu)^{4(\ell-m)}\right)}}{m!(2(\ell-m))!2^m}\right)^2\\ &= (k!)^2\left(\frac{1}{\ell!2^\ell}+\sum_{m=0}^{\ell-1}\frac{\sqrt{\mathbb{E}\left((\sqrt{s}Z+\mu)^{4(\ell-m)}\right)}}{m!(2(\ell-m))!2^m}\right)^2\\ &\leq (k!)^2\left(\frac{1}{\ell!2^\ell}+\sum_{m=0}^{\ell-1}\frac{\sqrt{(4(\ell-m))!}(\sqrt{s}+|\mu|)^{2(\ell-m)}}{2^\ell(2(\ell-m)!)^{3/2}m!}\right)^2\\ &\leq \frac{(k!)^2e}{2^{2\ell+1/2}(2\pi)^3}\left(\frac{(2\pi)^3 2^{\ell+1/2}}{e\,\ell!}+\right.\\ &\qquad\left.\sum_{m=0}^{\ell-1}2^{\ell-m}(\ell-m)^{-(\ell-m)-1}e^{\ell-m}(\sqrt{s}+|\mu|)^{2(\ell-m)}\right)^2\\ &\leq \frac{(k!)^2e}{2^{2\ell+1/2}(2\pi)^3}\left(\frac{(2\pi)^3 2^{\ell+1/2}}{e\,\ell!}+\sum_{u=1}^{\ell}\rho^u u^{-u-1}\right)^2,\end{aligned}\tag{10.5}$$

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 (10.4) and (10.5) we

obtain

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

and we get the wanted result. □

Lemma 2. The polynomials defined by (6.2) satisfy the following relations:

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

It is also well known (see for instance Szegö, 1939) that there exists a constant $C > 0$ such that, for all $x \in \mathbb{R}$:

$$|\mathcal{H}(x)| = \exp(-x^2/2)|h_k(x)| \leq C\sqrt{k!2k}. \tag{10.6}$$

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

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

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 (10.6) we obtain

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

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 $\mu \in \mathbb{R}$:

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

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

Finally from (10.6) we obtain $\mathbb{V}(Q_k(X)/q_k^2) = \mathbb{V}(\mathcal{H}_k(X_1)/q_k^2) \leq C^2/(k!2k)$, which directly insures **(A3)**. □

Theorem 2. Let us prove that $\mathbb{P}_0(S_n \geq 2)$ vanishes as $n \to +\infty$. By definition of S_n in (2.6) and $\widehat{D}_{k,n}[\cdot]$ in (2.5) we have for all $\lambda \in (0, 1/2)$:

$$\begin{aligned}
&\mathbb{P}_0(S_n \geq 2)\\
&= \mathbb{P}_0\left(\exists k, 2 \leq k \leq d(n) :\ n^\lambda U_{k,n}^\top \widehat{D}_{k,n}^{-1} U_{k,n} - k\log(n) \geq n^\lambda U_{1,n}^\top \widehat{D}_{1,n}^{-1} U_{1,n} - \log(n)\right)\\
&\leq \mathbb{P}_0\left(\exists k, 2 \leq k \leq d(n) :\ n^\lambda U_{k,n}^\top \widehat{D}_{k,n}^{-1} U_{k,n} \geq (k-1)\log(n)\right)\\
&\leq \mathbb{P}_0\left(\exists k, 2 \leq k \leq d(n) :\ \sum_{j=2}^{k} n^\lambda (R_{j,n})^2 \geq (k-1)\log(n) e(n)\right)\\
&\leq \mathbb{P}_0\left(\exists (j,k), 2 \leq j \leq k \leq d(n) :\ n^\lambda (R_{j,n})^2 \geq \log(n) e(n)\right)\\
&\leq \mathbb{P}_0\left(\sum_{j=2}^{d(n)} n^\lambda (R_{j,n})^2 \geq \log(n) e(n)\right).
\end{aligned}$$

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 (10.8–10.9). We decompose $R_{k,n}$ as follows:

$$R_{k,n} = (a_{k,n} - \mathbb{E}_0(a_{k,n})) - (\bar{p}_n \alpha_{k,n} - p_0 \alpha_k(\mu_0, s_0)), \quad 1 \leq k \leq d(n).$$

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

$$\begin{aligned}
&\mathbb{P}_0\left(\sum_{k=2}^{d(n)} n^\lambda (R_{k,n})^2 \geq \log(n) e(n)\right)\\
&\qquad \leq \mathbb{P}_0\left(\sum_{k=2}^{d(n)} (a_{k,n} - \mathbb{E}_0(a_{k,n}))^2 \geq \frac{\log(n) e(n)}{4n^\lambda}\right)\\
&\qquad\quad + \mathbb{P}_0\left(\sum_{k=2}^{d(n)} (\bar{p}_n \alpha_{k,n} - p_0 \alpha_k(\mu_0, s_0))^2 \geq \frac{\log(n) e(n)}{4n^\lambda}\right).
\end{aligned}$$

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

first have

$$\mathbb{P}_0\left(\sum_{k=2}^{d(n)}(a_{k,n}-\mathbb{E}_0(a_{k,n}))^2\geq\frac{\log(n)e(n)}{4n^\lambda}\right)$$
$$\leq\frac{4n^\lambda}{\log(n)e(n)}\sum_{k=2}^{d(n)}\mathbb{E}_0\left((a_{k,n}-\mathbb{E}_0(a_{k,n}))^2\right)$$
$$=\frac{4n^\lambda}{\log(n)e(n)}\sum_{k=2}^{d(n)}\frac{1}{n}\mathbb{V}\left(\frac{Q_k(X_1)}{q_k^2}\right)$$
$$\leq\frac{4d(n)}{n^{1-\lambda}\log(n)e(n)}M_3, \tag{10.7}$$

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 (2.7).

Secondly, by decomposing

$$\bar{p}_n\alpha_{k,n}-p_0\alpha_k(\mu_0,s_0)=(\bar{p}_n-p)\alpha_{k,n}+p_0(\alpha_{k,n}-\alpha_k(\mu_0,s_0)),$$

we obtain the following majorization

$$\mathbb{P}_0\left(\sum_{k=2}^{d(n)}(\bar{p}_n\alpha_{k,n}-p_0\alpha_k(\mu_0,s_0))^2\geq\frac{\log(n)e(n)}{4n^\lambda}\right)$$
$$\leq\mathbb{P}_0\left(\sum_{k=2}^{d(n)}(\alpha_{k,n})^2(\bar{p}_n-p_0)^2\geq\frac{\log(n)e(n)}{8n^\lambda}\right)$$
$$+\mathbb{P}_0\left(\sum_{k=2}^{d(n)}p_0^2(\alpha_{k,n}-\alpha_k(s_0,\mu_0))^2\geq\frac{\log(n)e(n)}{8n^\lambda}\right).$$

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

$$\mathbb{P}_0\left(\sum_{k=2}^{d(n)}\alpha_{k,n}^2(\bar{p}_n-p_0)^2\geq\frac{\log(n)e(n)}{8n^\lambda}\right) \tag{10.8}$$
$$\leq\mathbb{P}_0\left((\bar{p}_n-p_0)^2\geq\frac{\log(n)e(n)}{8n^\lambda M_1 d(n)}\right),$$

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 (2010). By denoting $\theta_0=(\mu_0,s_0)$ and $\bar{\theta}_n=(\bar{\mu}_n,\bar{s}_n)$, we also have $\|\bar{\theta}_n-\theta_0\|^2=o_{a.s.}(n^{-1/2+\alpha})$, for all $\alpha>0$. Since the $\dot{\alpha}_{k,n}$'s are bounded by $M2$ according to

(A2), using the *mean value* theorem we obtain:

$$\mathbb{P}_0\left(\sum_{k=2}^{d(n)}(\alpha_{k,n}-\alpha_k(s_0,\mu_0))^2 \geq \frac{\log(n)e(n)}{8n^\lambda}\right) \leq \mathbb{P}_0\left(\|\bar\theta-\theta_0\|^2 \geq \frac{\log(n)e(n)}{8n^\lambda M_2^2 d(n)}\right), \tag{10.9}$$

which last term goes to zero as $n\to+\infty$. Hence from (10.13) and the controls in probability (10.7–10.9), 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 $\bar s_n$ and $\bar\mu_n$ are independent and asymptotically Normally distributed towards the true values s_0 and μ_0, we get by using the delta method:

$$\sqrt{n}\alpha_1(\bar s_n,\bar\mu_n)\xrightarrow{\mathcal{L}}\mathcal{N}(\alpha_1(s_0,\mu_0),v_1d_1^2(s_0)+v_2d_2^2(\mu_0)),\quad \text{as } n\to+\infty,$$

where (d_1,d_2) is the gradient $\dot\alpha_1(\cdot,\cdot)$, and where v_1 and v_2 are respectively the asymptotic variance $\sqrt{n}\bar s_n$ and $\sqrt{n}\bar\mu_n$. Combining this convergence in law with the following convergence in probability:

$$V_{1,n}\xrightarrow{\mathbb{P}}\mathbb{V}(R_{1,n})\quad\text{and}\quad \bar p_n\xrightarrow{\mathbb{P}}p_0,\quad \text{as } n\to+\infty,$$

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:

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

which concludes the proof. □

Proposition 1. The advantage of considering the semiparametric approach in Bordes and Vandekerkhove (2010) versus the ML method is that under H_1 we keep the following consistency results:

$$\bar\vartheta_n=(\bar p_n,\bar\mu_n)\xrightarrow{\mathbb{P}_1}(p_0,\mu_0),\ \bar s_n\xrightarrow{\mathbb{P}_1}s_0,\ R_i\xrightarrow{\mathbb{P}_1}r_i=\mathbb{E}(Q_i(X)/q_i^2)-p_0\alpha_i(\mu_0,s_0),$$

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 $\mathbb{V}(R_i)$. Now, Clearly by (1.1) (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

$$r_i=0,\quad\forall\, i\geq 1\quad\Longleftrightarrow\quad\mathbb{E}(Q_i(Y)/q_i^2)=\alpha_i(\mu_0,s_0),\quad\forall\, i\geq 1.$$

This condition implies that the expansion of the Y density matches with the expansion of a Gaussian density with mean μ_0 and variance s_0, which is in contradiction with the semiparametric identifiability of model/setup H_1, see Bordes *et al.* (2006). 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 (2.4), for every $k \geq 1$ fixed, we can decompose $T_{k,n}$ as follows:

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

it comes that for all $k < j_0$, $T_{k,n} = O_P(n^{s-1})$ 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 $T_{j_0,n} \geq n^{\lambda} r_{j_0}^2 + O_P(n^{s-1/2})$. It comes that for all $k < j_0$ we have

$$\mathbb{P}_1 \left(T_{k,n} - \beta_k pen(n) < T_{j_0,n} - \beta_{j_0} pen(n) \right) \longrightarrow 1, \quad \text{as} \quad n \to +\infty.$$

This obviously shows, according to S_ns' definition (2.6), 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 $T_{S_n,n} \geq T_{j_0,n} \geq n^{\lambda} r_{j_0}^2 + O_P(n^{s-1/2})$ 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

$$\mathbb{P}^{(n)} \left(S_n^{(n)} \geq 2 \right) \leq \mathbb{P}^{(n)} \left(\sum_{k=2}^{d(n)} n^s \left(R_{k,n}^{(n)} \right)^2 \geq \log(n) e(n) \right). \tag{10.10}$$

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:

$$R_{k,n}^{(n)} = a_{k,n}^{(n)} - \mathbb{E}^{(n)}(a_{k,n}^{(n)}) - \left(\bar{p}_n^{(n)} \alpha_{k,n}^{(n)} - p_0 \alpha_k(\mu_0, s_0) \right) + \psi_{k,n}, \tag{10.11}$$

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

$$\psi_{k,n} := p_0 \mathbb{E} \left(Q_k(s_0 Y_1 + \mu_0 + \delta_n \varepsilon_1) - Q_k(s_0 Y_1 + \mu_0) \right) / q_k^2, \tag{10.12}$$

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

(1.4), Y_1 being $\mathcal{N}(0,1)$ distributed. We then have

$$\begin{aligned}
&\mathbb{P}^{(n)}(S_n^{(n)} \geq 2) \\
&\leq \mathbb{P}^{(n)}\left(\sum_{k=2}^{d(n)}\left(a_{k,n}^{(n)} - \mathbb{E}^{(n)}(a_{k,n}^{(n)}) - \bar{p}_n^{(n)}\alpha_{k,n}^{(n)} - p_0\alpha_k(\mu_0,s_0)\right)^2 \geq \frac{C(k,n)}{2n^s}\right) \\
&\leq \mathbb{P}^{(n)}\left(\sum_{k=2}^{d(n)}\left((a_{k,n}^{(n)} - \mathbb{E}^{(n)}(a_{k,n}^{(n)}))^2 + (\bar{p}_n^{(n)}\alpha_{k,n}^{(n)} - p_0\alpha_k(\mu_0,s_0))^2\right) \geq \frac{C(k,n)}{4n^s}\right) \\
&\leq \mathbb{P}^{(n)}\left(\sum_{k=2}^{d(n)}\left(a_{k,n}^{(n)} - \mathbb{E}^{(n)}(a_{k,n}^{(n)})\right)^2 \geq \frac{C(k,n)}{8n^s}\right) \\
&\quad + \mathbb{P}^{(n)}\left(\sum_{k=2}^{d(n)}\left(\bar{p}_n^{(n)}\alpha_{k,n}^{(n)} - p_0\alpha_k(\mu_0,s_0)\right)^2 \geq \frac{C(k,n)}{8n^s}\right),
\end{aligned}$$

where $C(k,n) := \log(n)e(n) - 2n^s\psi_{k,n}^2$. By **(A2)** we have

$$C(k,n) \geq \log(n)e(n) - 2n^\lambda M_2^2\delta_n^2\mathbb{E}(|\varepsilon_1|)^2, \tag{10.13}$$

which shows that $C(k,n) = O(\log(n)e(n))$ since $n^\lambda\delta_n^2 \to 0$ as $n \to +\infty$ due to **(A5)** (key point of the proof). We study the two above probabilities separately. First we have, according to the Markov inequality and Condition **(A3)**, that

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

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

$$\begin{aligned}
&\mathbb{P}^{(n)}\left(\sum_{k=2}^{d(n)}\left(\bar{p}_n^{(n)}\alpha_{k,n}^{(n)} - p_0\alpha_k(\mu_0,s_0)\right)^2 \geq \frac{C(k,n)}{8n^\lambda}\right) \\
&\quad \leq \mathbb{P}^{(n)}\left(\sum_{k=2}^{d(n)}\left(\alpha_{k,n}^{(n)}\right)^2\left(\bar{p}_n^{(n)} - p_0\right)^2 \geq \frac{C(k,n)}{16n^\lambda}\right) \\
&\quad + \mathbb{P}^{(n)}\left(p_0^2\sum_{k=2}^{d(n)}\left(\alpha_{k,n}^{(n)} - \alpha_k(s_0,\mu_0)\right)^2 \geq \frac{C(k,n)}{16n^\lambda}\right).
\end{aligned}$$

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

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

where the last right hand side term goes to zero as $n\to+\infty$ since $|\bar{p}_n^{(n)}-p_0|^2=o_{a.s}(n^{-\lambda})$. In fact, since $\lambda\in(0,1/2)$ can be parametrized as $\lambda=1/2-\alpha$ for all $0<\alpha<1/2$, according to Theorem 5 ii) and Remark 6 in the supplementary material Section, the last assertion is true. Writing $\theta_0=(\mu_0,s_0)$ and $\bar{\theta}_n^{(n)}=(\bar{\mu}_n^{(n)},\bar{s}_n^{(n)})$, similarly we have $\|\bar{\theta}_n^{(n)}-\theta_0\|^2=o_{a.s}(n^{-\lambda})$. Since the $\dot{\alpha}_k$'s are bounded by M_2 according to **(A2)**, using the *mean value* Theorem, we obtain:

$$\begin{aligned}\mathbb{P}^{(n)}&\left(p_0^2\sum_{k=2}^{d(n)}\left(\alpha_{k,n}^{(n)}-\alpha_k(s_0,\mu_0)\right)^2\geq\frac{C(k,n)}{16n^\lambda}\right)\\&\leq\mathbb{P}^{(n)}\left(\|\bar{\theta}_n^{(n)}-\theta_0\|^2\geq\frac{C(k,n)}{16n^\lambda d(n)M_2^2}\right),\end{aligned}$$

which last term goes to zero as $n\to+\infty$ according to **(A1)**. Hence from (10.10), we obtain that $\mathbb{P}^{(n)}(S_n\geq 2)\to 0$ 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 (10.12). It first comes

$$\begin{aligned}\psi_{1,n}&=p_0\mathbb{E}^{(n)}(Q_1(s_0Y_1+\mu_0+\delta_n\varepsilon_1)-Q_1(s_0Y_1+\mu_0))\\&=p_0\mathbb{E}^{(n)}(a_{1,1}(s_0Y_1+\mu_0+\delta_n\varepsilon_1)+a_{1,0}-a_{1,1}(s_0Y_1+\mu_0)-a_{1,0})\\&=p_0\delta_n\mathbb{E}^{(n)}(\varepsilon_1),\end{aligned}$$

and we have

$$\begin{aligned}R_{1,n}^{(n)}&=\left(a_{1,n}^{(n)}-\mathbb{E}^{(n)}(a_{1,n}^{(n)})\right)-\left(\bar{p}^{(n)}\alpha_{k,n}^{(n)}-p_0\alpha(\mu_0,s_0)\right)+\psi_{1,n}\\&=A-B+\psi_{1,n}.\end{aligned}$$

As seen previously, $A=O_{a.s.}(n)$ and

$$\begin{aligned}B&=\alpha_{k,n}^{(n)}\left(\bar{p}^{(n)}-p_0\right)+p_0\left(\alpha_{k,n}^{(n)}-\alpha(\mu_0,s_0)\right)\\&=B_1+B_2.\end{aligned}$$

Again, we have seen that $B_1 = o_{a.s.}(n^{-\lambda/2})$ and $B_2 = o_{a.s.}(n^{-\lambda/2})$. By **(A6)** it follows that a.s. $n^\lambda (R_{1,n}^{(n)})^2 \approx n^{\lambda - 2\xi'} \to +\infty$, as $n \to +\infty$. By construction we have $T_{1,n}^{(n)} \geq n(R_{1,n}^{(n)})^2/(\widehat{\mathbb{V}}(R_{1,n}^{(n)}) + e(n))$ which leads to

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

Under **(A7)** 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

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

We also have

$$\begin{aligned}
\psi_{2,n} &= p_0 \mathbb{E}^{(n)}(Q_2(s_0 Y_1 + \mu_0 + \delta_n \varepsilon_1) - Q_2(s_0 Y_1 + \mu_0)) \\
&= p_0 (\mathbb{E}^{(n)}(a_{2,2}(s_0 Y_1 + \mu_0 + \delta_n \varepsilon_1)^2 + a_{2,1}(s_0 Y_1 + \mu_0 + \delta_n \varepsilon_1) + a_{2,0} \\
&\quad - a_{2,2}(s_0 Y_1 + \mu_0)^2 - a_{2,1}(s_0 Y_1 + \mu_0) - a_{2,0})) \\
&= 2 p_0 a_{2,2} \delta_n^2 \mathbb{E}(\varepsilon_1^2).
\end{aligned}$$

From the above expressions and by definition of $R_{2,n}^{(n)}$ in (10.11) we can mimic the previous arguments to show that a.s. $R_{2,n}^{(n)} \approx \delta_n^2$ and that

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

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

11. Supplementary material

We study in this section the asymptotic behavior of the semiparametric estimator $(\bar{p}_n, \bar{\mu}_n)$ introduced in Bordes and Vandekerkhove (2010) when model (1.4) is no longer fixed but depends on n through the following transformation:

$$g^{(n)}(x) = p f_0(x) + (1-p) f^{(n)}(x - \mu), \quad x \in \mathbb{R}, \tag{11.1}$$

where $(f^{(n)})_{n \geq 1}$ is a sequence of ν-pdfs converging towards the limiting pdf f. For simplicity matters we suppose the following convolution approximation structure on the $f^{(n)}$'s:

$$f^{(n)}(x) = \int_{\mathbb{R}} f(u) \frac{1}{\delta_n} f_1\left(\frac{x-u}{\delta_n}\right) du, \quad x \in \mathbb{R}.$$

Using the classical change of variable $z = (x-u)/\delta_n$ we easily prove that:

$$\begin{aligned} |f^{(n)}(x) - f(x)| &= \left| \int_{\mathbb{R}} [f(x - z\delta_n) - f(x)] f_1(z) dz \right| \\ &\leq \int_{\mathbb{R}} |f(x - z\delta_n) - f(x)| f_1(z) dz \\ &\leq \|f'\|_\infty \delta_n \int_{\mathbb{R}} |z| f_1(z) dz, \end{aligned}$$

which shows that $\|f^{(n)} - f\|_\infty = O(\delta_n)$ when f' is supposed to be bounded and f_1 has a moment of order one. For simplicity we will call model (1.4) the "asymptotic model" of the model sequence (11.1). In this framework, for each $n \geq 1$, we consider a sample $(X_1^n, \ldots, X_n^n)$ iid from the n-local model g_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 (2010) semiparametric estimator based on the "corrupted" sample $(X_1^n, \ldots, X_n^n)$, we consider the *coupling*:

$$\begin{cases} X_i^n & := (1 - U_i)Y_i + U_i Z_i^n, \quad i = 1, \ldots, n \\ X_i & := (1 - U_i)Y_i + U_i Z_i, \quad i \geq 1, \end{cases} \tag{11.2}$$

where $(U_i)_{i\geq 1}$ is an iid sequence of $\mathcal{B}(p)$ random variables, $(Y_i, Z_i, \varepsilon_i)_{i\geq 1}$ is an iid sequence of random variables distributed according $f_0 \otimes f(\cdot - \mu) \otimes f_1$, and $Z_i^n := Z_i + \delta_n \varepsilon_i$ is by construction distributed according to $f^{(n)}$. Note that we have the following stochastic bound:

$$|X_i^n - X_i| \leq \delta_n |\varepsilon_i|, \quad i = 1, \ldots, n. \tag{11.3}$$

12. Estimation method

The cumulative distribution function (cdf) $G^{(n)}$ associated with model (11.1) is defined by

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

where $G^{(n)}$, F_0 and $F^{(n)}$ are cdfs corresponding to the dfs $g^{(n)}$, f_0 and $f^{(n)}$ respectively. Let us denote by ϑ the Euclidean part (p, μ) of the model parameters taking values in Θ. Assume that the asymptotic model (1.4) 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 (2010) procedure. Let us define

$$F^{(n)}(x) = \frac{1}{p}\left(G^{(n)}(x + \mu) - (1 - p)F_0(x + \mu) \right), \quad \forall x \in \mathbb{R}. \tag{12.1}$$

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 (12.1) and the *almost*-symmetry of $F^{(n)}$,

$$H^{(n)}(x;\vartheta_0,G^{(n)}) := H_1^{(n)}(x;\vartheta_0,G^{(n)}) - H_2^{(n)}(x;\vartheta_0,G^{(n)}) \approx 0, \quad (12.2)$$

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.* (2006a) 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 (1.4), see Bordes and Vandekerkhove (2010, p.24).

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

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

and $\widetilde{G}_n^{(n)}(x) = \displaystyle\int_{-\infty}^x \widehat{g}_n^{(n)}(t)dt$ 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 (5.1). 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 (2010, Section 2). Finally we propose to estimate ϑ_0 by

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

13. Identifiability, consistency and asymptotic normality

13.1. 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 df f_0 we consider one assumption, for which the semi-parametric identifiability of the model (1.4) parameters is obtained, see Bordes *et al.*, (2006b, 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).

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

(ii) 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).

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

Comments. The two first conditions in **(B)** (i) are necessary to obtain the pointwize consistency of 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 (1986, p. 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

$$\|\widetilde{G}_n - G\|_\infty = O_{a.s.}\left(\left(\frac{\log\log n}{n}\right)^{-1/2}\right). \tag{13.1}$$

Lemma 3. *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:

$$\begin{aligned}
\left|\widehat{g}_n^{(n)}(x) - \widehat{g}_n(x)\right| &= \left|\frac{1}{nh_n}\sum_{i=1}^n \left(K\left(\frac{x - X_i^n}{h_n}\right) - K\left(\frac{x - X_i}{h_n}\right)\right)\right|, \quad x \in \mathbb{R}\\
&\leq \frac{1}{nh_n}\sum_{i=1}^n \left|K\left(\frac{x - X_i^n}{h_n}\right) - K\left(\frac{x - X_i}{h_n}\right)\right|\\
&\leq \frac{1}{nh_n^2}\sum_{i=1}^n ||K'||_\infty \, |X_i^n - X_i|\\
&\leq \frac{||K'||_\infty \delta_n}{h_n^2} \times \left(\frac{\sum_{i=1}^n |\varepsilon_i|}{n}\right),
\end{aligned} \tag{13.2}$$

where the last inequality comes from (11.3). 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 to the proof (ii). □

13.2. 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 4. *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 $\vartheta\in\Theta$:

$$\begin{aligned}\left|d_n^{(n)}(\vartheta)-d_n(\vartheta)\right| &= \left|\frac{1}{n}\sum_{i=1}^n\left(H^2(X_i^n;\vartheta,\widetilde{G}_n^{(n)})-H^2(X_i;\vartheta,\widetilde{G}_n)\right)\right|\\ &\le \frac{1}{n}\sum_{i=1}^n\left|H^2(X_i^n;\vartheta,\widetilde{G}_n^{(n)})-H^2(X_i^n;\vartheta,\widetilde{G}_n)\right|\\ &\quad+\frac{1}{n}\sum_{i=1}^n\left|H^2(X_i^n;\vartheta,\widetilde{G}_n)-H^2(X_i;\vartheta,\widetilde{G}_n)\right|\\ &\le O_{a.s.}\left(\|\widetilde{G}_n^{(n)}-\widetilde{G}_n\|_\infty\right)\\ &\quad+O_{a.s.}\left(\frac{1}{n}\sum_{i=1}^n\left|\widetilde{G}_n(X_i^n+\mu)-\widetilde{G}_n(X_i+\mu)\right|\right). \qquad (13.3)\end{aligned}$$

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

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

where according to Silverman (1978) $\|\widetilde{g}_n-g\|_\infty = o_{a.s.}(1)$. Similarly to (13.2), 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 (13.3) is a $O_{a.s}(\delta_n/h_n)$ according to Lemma 3 (i), we obtain the wanted result.

For the proof of result (ii), let proceed similarly to Bordes and Vandekerkhove (2010) 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 (2010), we have at point ϑ_0:

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

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

$$\Delta_2(\mathbf{X}^{(n)},\mathbf{X}^n,\widetilde{G}_n):=\frac{2}{n}\sum_{i=1}^n\left|\mathcal{H}(X_i^n;\vartheta_0,\widetilde{G}_n)-\mathcal{H}(X_i;\vartheta_0,\widetilde{G}_n)\right|.$$

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:

$$\begin{aligned}
\Delta_1(\mathbf{X}^{(n)},\widetilde{G}_n^{(n)},\widetilde{G}_n)\leq{}&\frac{2}{n}\sum_{i=1}^n\left|H(X_i^n;\vartheta_0,\widetilde{G}_n^{(n)})-H(X_i^n;\vartheta_0,\widetilde{G}_n)\right|\\
&\times\left|\frac{\partial H}{\partial\mu}(X_i^n;\vartheta_0,\widetilde{G}_n^{(n)})-\frac{\partial H}{\partial\mu}(X_i^n;\vartheta_0,\widetilde{G}_n)\right|\\
&+\frac{2}{n}\sum_{i=1}^n\left|H(X_i^n;\vartheta_0,\widetilde{G}_n)-H(X_i^n;\vartheta_0,G)\right|\\
&\times\left|\frac{\partial H}{\partial\mu}(X_i^n;\vartheta_0,\widetilde{G}_n^{(n)})-\frac{\partial H}{\partial\mu}(X_i^n;\vartheta_0,\widetilde{G}_n)\right|\\
&+\frac{2}{n}\sum_{i=1}^n\left|\frac{\partial H}{\partial\mu}(X_i^n;\vartheta_0,\widetilde{G}_n)-2f(X_i^n)\right|\\
&\times\left|H(X_i^n;\vartheta_0,\widetilde{G}_n^{(n)})-H(X_i^n;\vartheta_0,\widetilde{G}_n)\right|\\
&+\frac{4}{n}\sum_{i=1}^n|f(X_i^n)|\times\left|H(X_i^n;\vartheta_0,\widetilde{G}_n^{(n)})-H(X_i^n;\vartheta_0,\widetilde{G}_n)\right|\\
\leq{}&c_1\left(\|\widetilde{G}_n^{(n)}-\widetilde{G}_n\|_\infty+\|\widetilde{G}_n-G\|_\infty\right)\|\widetilde{g}_n^{(n)}-\widetilde{g}_n\|_\infty\\
&+c_2\left(\|\widetilde{g}_n-g\|_\infty+\|f\|_\infty\right)\|\widetilde{G}_n^{(n)}-\widetilde{G}_n\|_\infty\\
={}&O_{a.s}\left(\frac{\delta_n^2}{h_n^3}\right)+O_{a.s}\left(\frac{\delta_n}{h_n}\right).
\end{aligned}$$

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:

$$\begin{aligned}
&\left|H(x;\vartheta_0,\widetilde{G}_n)-H(x';\vartheta_0,\widetilde{G}_n)\right|\\
&\leq\frac{1}{p_0}\left|[\widetilde{G}(x+\mu)-\widetilde{G}(x'-\mu)]+[\widetilde{G}(-x+\mu)-\widetilde{G}(-x'-\mu)]\right|\\
&+\frac{1-p_0}{p_0}\left|[F_0(x+\mu)-F_0(x'-\mu)]+[F_0(-x+\mu)-F_0(-x'-\mu)]\right|\\
&\leq\frac{2}{p_0}(\|\widetilde{g}_n-g\|_\infty+\|g\|_\infty)|x-x'|+\frac{2(1-p_0)}{p_0}\|f_0\|_\infty|x-x'|, \qquad (13.4)
\end{aligned}$$

and

$$\begin{aligned}
&\left|\frac{\partial H}{\partial \mu}(x;\vartheta_0,\widetilde{G}_n)-\frac{\partial H}{\partial \mu}(x';\vartheta_0,\widetilde{G}_n)\right| \\
&\leq \frac{1}{p_0}\left|[\widetilde{g}_n(x+\mu)-\widetilde{g}_n(x'-\mu)]+[\widetilde{g}_n(-x+\mu)-\widetilde{g}_n(-x'-\mu)]\right| \\
&+\frac{1-p_0}{p_0}\left|[f_0(x+\mu)-f_0(x'-\mu)]+[f_0(-x+\mu)-f_0(-x'-\mu)]\right| \\
&\leq \frac{2}{p_0}(\|\widetilde{g}_n'-g'\|_\infty+\|g'\|_\infty)|x-x'|+\frac{2(1-p_0)}{p_0}\|f_0'\|_\infty|x-x'|. \qquad (13.5)
\end{aligned}$$

Using (13.4) and (13.5) we obtain

$$\begin{aligned}
\Delta_2(\mathbf{X}^{(n)},\mathbf{X}^n,\widetilde{G}_n) \quad \leq \quad & \frac{2}{n}\sum_{i=1}^n\left|H(X_i^n;\vartheta_0,\widetilde{G}_n)-H(X_i;\vartheta_0,\widetilde{G}_n)\right| \\
&\times\left|\frac{\partial H}{\partial \mu}(X_i^n;\vartheta_0,\widetilde{G}_n)-\frac{\partial H}{\partial \mu}(X_i;\vartheta_0,\widetilde{G}_n)\right| \\
&+\frac{2}{n}\sum_{i=1}^n\left|H(X_i;\vartheta_0,\widetilde{G}_n)-H(X_i;\vartheta_0,G)\right| \\
&\times\left|\frac{\partial H}{\partial \mu}(X_i^n;\vartheta_0,\widetilde{G}_n)-\frac{\partial H}{\partial \mu}(X_i;\vartheta_0,\widetilde{G}_n)\right| \\
&+\frac{2}{n}\sum_{i=1}^n\left|\frac{\partial H}{\partial \mu}(X_i;\vartheta_0,\widetilde{G}_n)-2f(X_i)\right| \\
&\times\left|H(X_i^n;\vartheta_0,\widetilde{G}_n)-H(X_i;\vartheta_0,\widetilde{G}_n)\right| \\
&+\frac{4}{n}\sum_{i=1}^n|f(X_i)|\times\left|H(X_i^n;\vartheta_0,\widetilde{G}_n)-H(X_i;\vartheta_0,\widetilde{G}_n)\right| \\
= \quad & O_{a.s}(\delta_n^2)+O_{a.s}(\delta_n^2)+O_{a.s}(\delta_n\|\widetilde{G}_n-G\|_\infty)+O_{a.s}(\delta_n),
\end{aligned}$$

which by (13.1) concludes the proof for (ii). For the proof of result (iii) we use the following decomposition at any point $\vartheta\in\Theta$:

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

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

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

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

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

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

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

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

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

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

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 (1978) 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 (1978) uniform consistency result on the kernel estimator and its derivatives and (11.3), 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 4, 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 5.
(i) Suppose that Conditions **(K)**, **(B)** *and* **(I)** *are satisfied,* Θ *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* $\delta > 0$.

(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*

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

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

$$\sqrt{n}\left(\frac{\delta_n^2}{h_n^3} + \frac{\delta_n}{h_n}\right) \longrightarrow 0, \quad and \quad \frac{\delta_n}{h_n^3} \longrightarrow 0, \quad as \quad n \to +\infty, \tag{13.6}$$

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 (11.2) 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 (2010). 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 (2010) . 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 4 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 4 (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 (2010). □

Remark 6. *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 (13.6) is satisfied if:*

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

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

Remark 7. *The conditions imposed in (13.6) 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 $(X_1^n, \ldots, X_n^n)$, 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 (1978).*

14. Asymptotic behavior of the MLE

In this section we propose to derive the asymptotic covariance matrix involved in the Central Limit Theorem associated with maximum likelihood estimator. 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 $\ell_\phi(x)$:

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

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:

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

and

$$\ddot{f}^\mu_{(\mu,s)}(x) = -\frac{1}{s}f_{(\mu,s)}(x) + \frac{x-\mu}{s}\dot{f}^\mu_{(\mu,s)}(x) = -\frac{1}{s}f_{(\mu,s)}(x) + \left(\frac{x-\mu}{s}\right)^2 f_{(\mu,s)}(x)$$
$$\ddot{f}^s_{(\mu,s)}(x) = \left[\frac{1}{2s^2} - \frac{(x-\mu)^2}{s^3}\right]f_{(\mu,s)}(x) + \left[-\frac{1}{2s} + \frac{(x-\mu)^2}{2s^2}\right]\dot{f}^s_{(\mu,s)}(x)$$

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

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

Given the above expressions we can derive under standard conditions, see van der Vaart (1998, p.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

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

References

[1] 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.

[2] 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.

[3] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B*, **57**, 289–300.

[4] 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.

[5] 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.

[6] Broeït, P., Lewin, A., Richardson, S., Dalmasso, C. and Magdelenat, H. (2004). A mixture model-based strategy for selecting sets of genes in multiclass response microarray experiments. *Bioinformatics*, **20**, 2562–2571.

[7] Do, K-A., Müller, P. and Tang, F. (2005). A Bayesian mixture model for differential gene expression. *Appl. Statist*, **54**, 627–644.

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

[9] Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Amer. Statist. Assoc.*, **99**, 96–104.

[10] Efron, B. and Tibshirani, R. (2002). Empirical Bayes methods and false discovery rates for microarrays. *Genet. Epidemiol.*, **23**, 70–86.

[11] Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.*, **96**, 1151–1160.

[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.*

[13] Ghattas B. Pommeret, D. Reboul, L. and Yao, A.F. (2011) Data driven smooth test for paired populations. *Journal of Statistical Planning and Inference* 141: 262–275.

[14] Gottardo, R., Raftery, A. E. and Yeung, K.. and Bumgarner, R.E. (2006).

Bayesian robust inference for differential gene expression in cDNA microarrays with multiple samples. *Biometrics*, **62**, 10–18.

[15] Guan, Z., Wu, B. and Zhao, H. (2008). Nonparametric estimator of false discovery rate based on Bernstein polynomials. *Statistica Sinica*, **18**, 905–923.

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

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

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

[19] Lönnstedt, I. and Speed, T. (2002). Replicated microarray data. *Statist. Sinica*, **12**, 31–46.

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

[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] Neyman, J. (1937). Smooth Test for Goodness of Fit, *Skandinavisk Aktuarietidskrift*, **20**, 149–199.

[23] Newton, M.A. Noueiry, A., Sarkar, D. and Ahlquist, P. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics*, **5**, 155–176.

[24] Pan, W., Lin, J. and Le, C.T. (2003). A mixture model approach to detecting differentially expressed genes with microarray data. *Functional and Integrative Genomics*, **3**, 117–124.

[25] 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.

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

[27] van der Vaart, A.W. (1998) . *Asymptotic statistics*. Cambridge University Press.

[28] 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.

[29] Wylupek, G. (2010). Data driven k sample tests. *Technometrics* 52: 107–123.

[30] Yang Y., Aghababazadeh F.A. and Bickel D.R. (2013). Parametric estimation of the local false discovery rate for identifying genetic associations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **10**, 98–108.

[31] Zhao, Y. and Pan, W. (2003). Modified nonparametric approaches to detecting differentially expressed genes in replicated microarray experiments. *Bioinformatics*, **19**, 1046–1054.

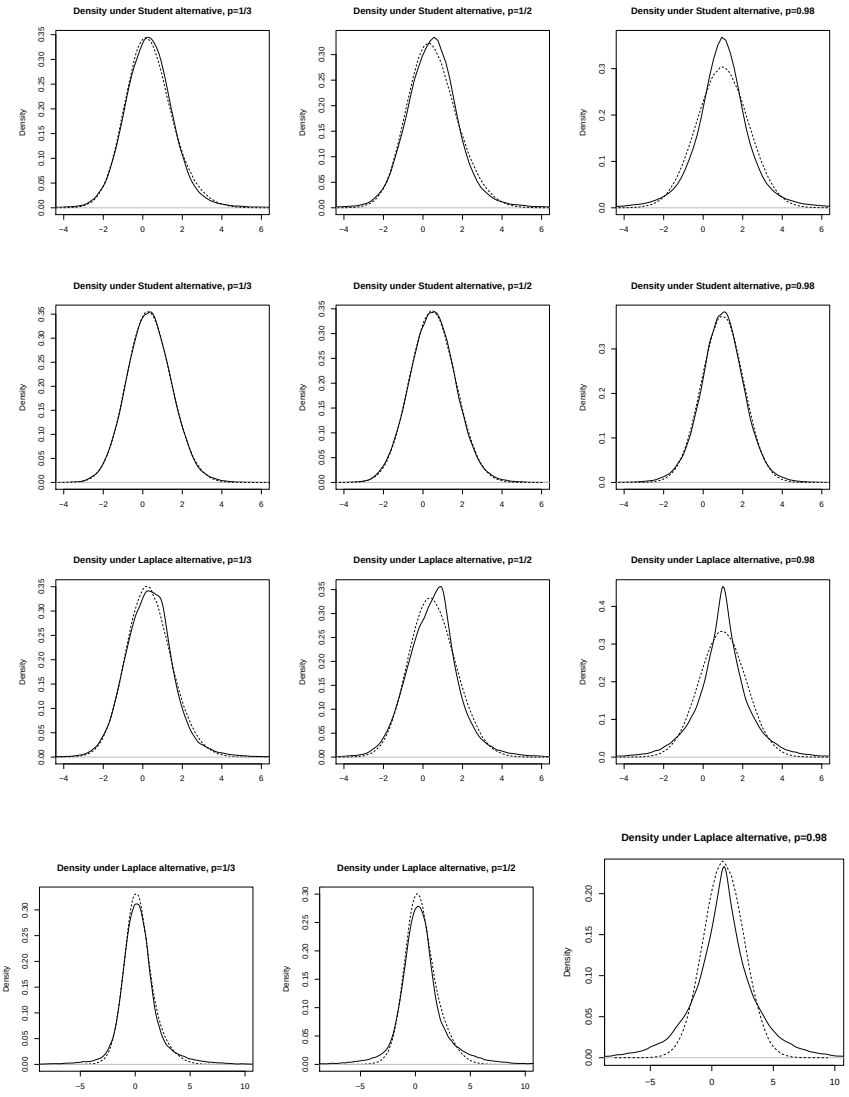

Fig 11: Plot of the graph of g in model (1.4) under several conditions:
Row 1: 1-shifted Student $t(3)$ alternative distribution (plain) and a null-type Gaussian distribution with similar parameters $\mathcal{N}(1,3)$ (dashed).
Row 2: 1-shifted Student $t(10)$ alternative distribution (plain) and a null-type 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 $p = 1/3,\ 1/2,\ 0.98$.