Semiparametric two-sample admixture components comparison test: The symmetric case

Xavier Milhaud, Denys Pommeret, Yahia Salhi, Pierre Vandekerkhove

Semiparametric two-sample admixture components comparison test: the symmetric case

Xavier Milhaud[a], Denys Pommeret[a,b], Yahia Salhi[a], Pierre Vandekerkhove[c]

[a] *Univ Lyon, UCBL, ISFA LSAF EA2429, F-69007, Lyon, France*
[b] *Aix Marseille Univ, CNRS, Centrale Marseille, I2M, 13288 Marseille cedex 9, France*
[c] *Université Gustave Eiffel, LAMA (UMR 8050), 77420 Champs-sur-Marne, France*

Abstract

In this paper, we consider admixture models which are two-component mixture distributions having one known component. This is the case when a gold standard reference component is well known, and when a population contains such a component plus another one with different features. When two populations are drawn from such models, we propose a penalized χ^2-type testing procedure allowing a pairwise comparison of the unknown components, *i.e.* to test the equality of their residual features densities, under a symmetry condition. A numerical study is carried out from a large range of simulation setups to illustrate the asymptotic properties of our test. Moreover the testing procedure is applied on a real-world case: galaxy velocities datasets, where stars heliocentric velocities mixed with the Milky Way are compared.

Keywords:
admixture, Chi-squared test, contamination, finite mixture model, galaxy data, semiparametric estimator, two-sample.

1. Introduction

Let us consider the two-component mixture model with probability density function (pdf) h defined by

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

where g is a known pdf, and the unknown parameters are the mixture proportion $p \in]0,1[$ and the pdf f. This model, sometimes so-called *admixture* or *contamination* model, has been widely investigated in the last decades, see for instance Bordes and Vandekherkove [2], Matias and Nguyen [17], Cai and Jin [4] or Celisse and Robin [5] among others. Numerous applications of model (1) can be found in topics such as: i) genetics regarding the analysis of gene expressions from microarray experiments as in Broët *et al.* [3]; ii) the false discovery rate problem (used to assess and control multiple error rates as in Efron and Tibshirani [7]), see McLachlan *et al.* [14]; iii) astronomy, in which this model arises when observing variables such as metallicity and radial velocity of stars as considered in Walker *et al.* [22]; iv) biology to model trees diameters, see Podlaski and Roesch [19]; v) kinetics to model plasma data, see Klingenberg *et al.* [9].

In this paper, the data of interest is made of two i.i.d. samples $X = (X_1, \ldots, X_{n_1})$ and $Y = (Y_1, \ldots, Y_{n_2})$ of size n_1 and n_2 with respective probability density functions:

$$\begin{cases} h_1(x) = (1-p_1)g_1(x) + p_1 f_1(x), & x \in \mathbb{R}, \\ h_2(x) = (1-p_2)g_2(x) + p_2 f_2(x), & x \in \mathbb{R}, \end{cases} \tag{2}$$

where p_1, p_2 are the unknown mixture proportions and f_1, f_2 are the unknown component densities with respect to a given reference measure ν. Given the above model, our goal is now to answer the following statistical problem:

$$H_0: \quad f_1 \text{ is equal to } f_2 \qquad \text{against} \qquad H_1: \quad f_1 \text{ is different from } f_2 \ , \tag{3}$$

without assigning any specific parametric family to the unknown components f_i's. The main shape constraint used throughout this paper is, similarly to Bordes and Vandekerkhove [2], the fact that the f_i's are symmetric with respect to (w.r.t.) a non-null location parameter, *i.e.* there exists $\mu_i \in \mathbb{R}^*$ such that $f_i(x + \mu_i) = f_i(-x + \mu_i)$, $i = 1, 2$, for all $x \in \mathbb{R}$.

This problem is a natural extension of a recent work by Pommeret and Vandekherkove [20] to the two sample case. Basically our test procedure consists in expanding the two unknown densities in an orthogonal polynomial basis, and then in comparing, with an *ad hoc* method, their coefficients up to a parsimonious rank selected according to a data-driven technique detailed later on in the paper.

Our method can be used in many areas as soon as the unkwown densities are supposed to be symmetric w.r.t. a location parameter, including the Gaussian case, but also Uniform and Laplace, among others. As a practical illustration of our work, we analyze kinematic datasets from two Milky Way dwarf spheroidal (dSph) satellites: Carina and Sextans, see for instance Walker *et al.* [22]. More specifically, we consider the heliocentric velocities (HV) of stars in these satellites, which are the velocities defined with respect to the solar system. These measurements are mixed with the HV of stars in the Milky Way. Since the Milky Way is largely observed, see Robin *et al.* [21], we can assume that, as required in our model (2), its velocity distribution is perfectly known. One interesting problem is then to compare the HV distributions of both satellites Carina and Sextans through such mixture models with a common Milky Way known component. We are therefore left with a two sample admixture components comparison problem with a common known and well documented component g, *i.e.* $g_1 = g_2 = g$ in (2). Moreover, since the f_i's distributions are generally considered as Gaussian in the astronomical literature, we can therefore also reasonably assume that their (technically required) symmetry with respect to a location parameter holds.

The remainder of the paper is organized as follows. In Section 2, we introduce the testing problem and describe our methodology. In Section 3, we state the assumptions and asymptotic results under the null hypothesis, along with the test divergence under the alternative. Section 4 provides details about the adequate polynomial decomposition depending on the nature of the distributions support. In Section 5, we implement a simulation-based study to evaluate the empirical level and power of the test. Finally, Section 6 is devoted to a real-world application based on a kinematic dataset with galactical heliocentric velocities comparisons. A discussion closes the paper, when proofs and additional exploratory simulations involving the Patra and Sen [18] estimator approach are relegated in Appendix.

2. Testing problem

Our test procedure is based on the expansion coefficients comparison of the two probability density functions h_1 and h_2, defined in (2), in an orthonormal polynomial basis. Such an approach was originally introduced by Neyman [16] and extended in a data-driven context by Ledwina [12]. Our test procedure will permit to asymptotically detect any departure between two expansion coefficients, screened pairwise, along the indices. As exhibited in Kallenberg and Ledwina (1995) (see also the more recent study in Ghattas et al. [8]), Neyman tests detect very well any departure from the null, from the first coefficient associated to the first element of the orthogonal basis, namely its first projection, to the k-th one. In addition, the nonparametric estimation of f_1 and f_2 can also be obtained by such a projection procedure (up to an inversion step). It is then natural to compare their coefficients in the spirit of the smooth Neyman type test.

Remark 1. *For technical reasons, we assume in the sequel that*

$$n_1/(n_1+n_2) \to a \in]0,1[\text{ as } n_1, n_2 \to +\infty. \tag{4}$$

This condition is not restrictive and is obviously fulfilled when dealing with real-world datasets, which corresponds to finite sample applications.

Let $\mathcal{Q} = \{Q_k; k \in \mathbb{N}\}$ be an orthonormal basis of the $L^2(\nu)$ space, where ν is the reference measure for the densities expressed in (2). We write $\mathcal{Q} = \{Q_k; k \in \mathbb{N}\}$, where $Q_0 = 1$ and

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

with $\delta_{jk} = 1$ if $j = k$ and 0 otherwise.

We assume that the following integrability conditions are satisfied

$$\int_{\mathbb{R}} h_1^2(x)\nu(dx) < \infty \quad \text{and} \quad \int_{\mathbb{R}} h_2^2(x)\nu(dx) < \infty.$$

Then, for all $x \in \text{supp}(\nu)$, we have for $i = 1, 2$

$$h_i(x) = \sum_{k\geq 0} h_{i,k} Q_k(x) \quad \text{with} \quad h_{i,k} = \int_{\mathbb{R}} Q_k(x) h_i(x) \nu(dx),$$

$$g_i(x) = \sum_{k\geq 0} g_{i,k} Q_k(x) \quad \text{with} \quad g_{i,k} = \int_{\mathbb{R}} Q_k(x) g_i(x) \nu(dx),$$

$$f_i(x) = \sum_{k\geq 0} f_{i,k} Q_k(x) \quad \text{with} \quad f_{i,k} = \int_{\mathbb{R}} Q_k(x) f_i(x) \nu(dx),$$

and, from (2), we deduce that

$$h_{i,k} = (1-p_i) g_{i,k} + p_i f_{i,k}.$$

Note that there is no restriction on the support of f_1 and f_2, excepted that it must be known. The null hypothesis can be rewritten as $f_{1,k} = f_{2,k}$, for all $k \geq 1$. Or equivalently

$$H_0 : \; p_2(h_{1,k} - (1-p_1) g_{1,k}) = p_1(h_{2,k} - (1-p_2) g_{2,k}), \qquad k \geq 1. \tag{5}$$

Since the pdfs g_1 and g_2 are known, the coefficients $g_{i,k}$, $i = 1, 2$, are automatically known. For all $k \geq 1$, the coefficients $h_{i,k}$ can be estimated empirically by:

$$\widehat{h}_{1,k} = \frac{1}{n_1} \sum_{j=1}^{n_1} Q_k(X_j) \quad \text{and} \quad \widehat{h}_{2,k} = \frac{1}{n_2} \sum_{j=1}^{n_2} Q_k(Y_j). \tag{6}$$

The estimation of the proportions p_i, for $i = 1, 2$, involved in model (2) will depend on some technical conditions. In fact, the following assumptions allow to semiparametrically identify model (2) and to estimate the parameters p_1 and p_2 of crucial importance in our testing method, see expression (8) hereafter.

(A1) The regularity and identifiability conditions required in Bordes and Vandekerkhove [2] and Bordes *et al.* [1] are satisfied. Regarding specifically the identifiability conditions we will suppose either:

a) The densities g_i and f_i, for $i = 1, 2$, are respectively supposed to be odd and symmetric w.r.t. a non-null location parameter μ_i, *i.e.* there exists $\mu_i \in \mathbb{R}^*$ such that for all $x \in \mathbb{R}$, $f_i(x + \mu_i) := f_i^S(x) = f_i^S(-x)$, with 2-nd order moments supposed to satisfy

$$m(g_i) \neq m(f_i^S) + \mu_i \frac{2 \pm k}{3k}, \quad \text{for } k \in \{1, 2, \cdots\} \text{ and } i = 1, 2, \tag{7}$$

where $m(f)$ generically denotes the second order moment according to the f density.

b) The densities g_i and f_i, for $i = 1, 2$, are respectively supposed to be strictly positive over $\mathbb{R}$, and symmetric about a non-null location parameter μ_i, *i.e.* there exists $\mu_i \in \mathbb{R}^*$ such that for all $x \in \mathbb{R}$, $f_i(x + \mu_i) := f_i^S(x) = f_i^S(-x)$, both having first order moments and satisfying the following tail conditions:

$$\text{for all } \beta \in \mathbb{R}: \quad \lim_{x \to +\infty} \frac{f_i^S(x - \beta)}{g_i(x)} = 0, \quad \text{or} \quad \lim_{x \to -\infty} \frac{f_i^S(x - \beta)}{g_i(x)} = 0, \quad i = 1, 2.$$

The central role of the above conditions in the semiparametric literature are detailed in the recent and very well documented survey by Xiang et al. [24].

Remark 2. *The parameters p_1 and p_2 can also be consistently estimated by the Patra and Sen [18] estimator without any specific symmetry assumption. However, we do not have the $\sqrt{n}$-consistency of these estimators, see Theorem 4 in Patra and Sen [18]. This last point makes the use of the Patra and Sen estimator theoretically inadequate for our testing procedure, but it could possibly turn out to be interesting in practice to address also asymmetric cases. In Appendix C.1.2 and Appendix C.2.2, we numerically investigate this alternative methodology based on various schemes using Monte-Carlo simulations.*

Henceforth, we will denote by $\widehat{p}$ the Bordes and Vandekerkhove [2] estimator of p under **(A1)**. Hence, to answer the H_0 testing problem (3), we consider the following double-sourced differences

$$\widehat{R}_k := \widehat{p}_2(\widehat{h}_{1,k} - (1 - \widehat{p}_1)g_{1,k}) - \widehat{p}_1(\widehat{h}_{2,k} - (1 - \widehat{p}_2)g_{2,k}), \quad k \geq 1, \tag{8}$$

allowing to detect any possible departure from the null hypothesis.
The basic idea to construct our test statistic is to combine the standard central limit theorem satisfied by the empirical estimators $\widehat{h}_{1,k}$ and $\widehat{h}_{2,k}$, with the asymptotic normality of $\widehat{p}_1$ and $\widehat{p}_2$ proved in [2]. To overcome the complex dependence between the estimators of (p_1, p_2) and the estimators of the coefficients associated to h_1 and h_2, we split each sample into two independent sub-samples of size n_1', n_1'' for X and n_2', n_2'' for Y, with $n_1' + n_1'' = n_1$ and $n_2' + n_2'' = n_2$, respectively. For simplicity we fix $n_1' = n_1'' = n_1/2$ and $n_2' = n_2'' = n_2/2$. Then, we use the first sub-samples to estimate the coefficients of h_1 and h_2, and the second sub-samples to estimate the proportions p_1 and p_2. We detail this construction in Appendix B. For all $k \geq 1$, we obtain convergent estimators $\widehat{w}_k$ of the asymptotic variance of $\sqrt{\widetilde{n}}\widehat{R}_k$, where $\widetilde{n} = (n_1 n_2)/(n_1 + n_2)$. Then for all $k \geq 1$, we define $\widehat{U}_k = (\widehat{R}_1, \ldots, \widehat{R}_k)$, and

$$\widehat{T}_k \;=\; \widetilde{n}\widehat{U}_k \widehat{D}_k^{-1} \widehat{U}_k^{\top}, \tag{9}$$

where $\widehat{D}_k$ is a diagonal matrix of normalization having the form $\mathrm{diag}(\widehat{d}_1, \ldots, \widehat{d}_k)$, with

$$\widehat{d}_j = \max\left(\widehat{w}_j, e(n_1, n_2)\right), \quad 1 \leq j \leq k, \tag{10}$$

where $e(n_1, n_2)$ is a trimming term satisfying $e(n_1, n_2) \to 0$ as n_1, n_2 tend to infinity, and added to avoid instability in the evaluation of $\widehat{D}_k^{-1}$.

Following Ledwina [12] and Kallenberg and Ledwina [10], we suggest a data-driven procedure to select automatically the number of coefficients needed to answer the testing problem. Formally, we introduce the following penalized rule to select the rank k of the statistic $\widehat{T}_k$:

$$S(n_1, n_2) = \min\left\{ \underset{1 \leq k \leq d(n_1, n_2)}{\mathrm{argmax}} \left(s(n_1, n_2)\widehat{T}_k - \beta_k \mathrm{pen}(n_1, n_2) \right) \right\}, \tag{11}$$

where $d(n_1, n_2) \to +\infty$ as $n_1, n_2 \to +\infty$, $\mathrm{pen}(n_1, n_2)$ is a penalty term such that $\mathrm{pen}(n_1, n_2) \to +\infty$ as $n_1, n_2 \to +\infty$, the β_k's are penalization factors, and $s(n_1, n_2)$ is a normalization factor. In practice, we will consider $\beta_k = k$, $k \geq 1$, and $\mathrm{pen}(n_1, n_2) = \log(n_1 n_2/(n_1 + n_2))$, $n_1, n_2 \geq 1$. The rate $s(n_1, n_2)$ depends on the almost sure $o_{a.s.}(n_i^{-1/4+\alpha})$, $\alpha > 0$, $i = 1, 2$, convergence rate of the estimators of p_1 and p_2 (it appears in the proof of Lemma 1). Accordingly, we fix

$$s(n_1, n_2) \;=\; \left(\frac{n_1 n_2}{n_1 + n_2} \right)^{s-1}, \text{ with } 0 < s < 1/2. \tag{12}$$

Finally the data-driven test statistic under consideration is $T(n_1, n_2) = \widehat{T}_{S(n_1, n_2)}$.

3. Additional assumptions and main results

To test consistently (3), based on the statistic $T(n_1, n_2)$, we will suppose the following conditions to hold.

(A2) The coefficient order upper bound $d(n_1, n_2)$ involved in (11) satisfies

$$d(n_1, n_2) = o(\log(n_1 n_2/(n_1 + n_2)) e(n_1, n_2)).$$

(A3) There exist nonnegative constants M_1, M_2 such that for all $k \geq 1$, under H_0

$$\frac{1}{k}\sum_{j=1}^{k}\mathbb{E}(Q_j^2(X)) < M_1, \qquad \text{and} \qquad \frac{1}{k}\sum_{j=1}^{k}\mathbb{E}(Q_j^2(Y)) < M_2.$$

Remark 3. *Note that assumption* **(A3)** *is for example satisfied in the Gaussian case as stated in Pommeret and Vandekerkhove [20], see Lemma 1 p. 4754 along with its proof in Appendix B, when the Q_j's are the $\mathcal{N}(0,1)$-orthogonal Hermite polynomials.*

The next two results state respectively the asymptotic behavior of the selected rank $S(n_1, n_2)$ and the test statistic $T(n_1, n_2)$. For related proofs, see Appendix A.

Lemma 1. *If* **(A1)** *is satisfied, and if* **(A2)** *and* **(A3)** *hold, then, under H_0, $S(n_1, n_2)$ converges in probability towards 1 as $n_1, n_2 \to +\infty$.*

From Lemma 1, $T(n_1, n_2)$ and $\widehat{T}_1$, see definition (9), have the same limiting distribution. Moreover, under assumption **(A1)**, the estimators $\widehat{p}_1$ and $\widehat{p}_2$ are asymptotically Gaussian under the null hypothesis. Thus, we deduce the limit distribution of the test statistic from the previous lemma.

Lemma 2. *If assumptions of Lemma 1 are satisfied, then, under H_0, $\widehat{T}_1$ converges in law towards a χ^2-distribution with one degree of freedom as $n_1, n_2 \to +\infty$.*

Theorem 1. *Assume that* **(A1-3)** *hold, then, under H_0, $T(n_1, n_2)$ converges in law towards a χ^2-distribution with one degree of freedom as $n_1, n_2 \to +\infty$.*

We consider now the collection of H_1-type alternatives defined as follows: there exists $q \in \mathbb{N}^*$ such that

$$H_1(q): \quad f_{1,j} = f_{2,j},\ j = 1, \ldots, q-1, \quad \text{and} \quad f_{1,q} \neq f_{2,q},$$

which describes a departure between f_1 and f_2 at the q-th order coefficient. If we let

$$\delta(k) := p_2(h_{1,k} - (1-p_1)g_{1,k}) - p_1(h_{2,k} - (1-p_2)g_{2,k}), \qquad k \geq 1, \tag{13}$$

then the alternative hypothesis $H_1(q)$ tells that $\delta(q)$ is the first non null coefficient along this series. We can now state the following proposition that describes the asymptotic drift of the test statistic under $H_1(q)$.

Proposition 1. *Assume that* **(A1-3)** *hold. Then, under $H_1(q)$, $S(n_1, n_2) \to s \geq q$ and $T(n_1, n_2) \to +\infty$ as $n_1, n_2 \to +\infty$, that is, for all $\varepsilon > 0$, $\mathbb{P}(T(n_1, n_2) < \varepsilon) \to 0$.*

Remark 4. *Let us point out that all the asymptotic results provided in this section are still valid when the cumulative distribution functions (cdfs) G_1 and G_2 associated respectively to g_1 and g_2 are not exactly known but can be estimated using separate (independent from X and Y) i.i.d. samples $U = (U_1, \ldots, U_{N_1})$ and $V = (V_1, \ldots, V_{N_2})$ drawn from g_1 and g_2. In fact since the estimation method proposed by Bordes and Vandekerkhove [2] is essentially based on almost sure uniform*

control of convergence of the empirical cdf, all their proofs can be revisited by substituting the cdfs G_1 and G_2 by their smoothed empirical versions $\widetilde{G}_1$ and $\widetilde{G}_2$ without impacting their asymptotic results provided that $N_1/(N_1+N_2) \to A \in]0,1[$ as $N_1, N_2 \to +\infty$ and $n_1 = o(N_1)$ along with $n_2 = o(N_2)$. Similarly, the required moments associated to the probability density functions g_1 and g_2 can be empirically estimated by

$$\widehat{g}_{1,k} = \frac{1}{N_1}\sum_{j=1}^{N_1} Q_k(U_j) \quad \textit{and} \quad \widehat{g}_{2,k} = \frac{1}{N_2}\sum_{j=1}^{N_2} Q_k(V_j), \tag{14}$$

with negligible bias, compared to the ones associated to $\widehat{h}_{1,k}$ and $\widehat{h}_{2,k}$. Therefore the use of these empirical estimators in the proofs (instead of g_1 and g_2) still ensures the asymptotic validity of our results.

4. Choice of the reference measure and test construction

4.1. Choice of the adequate reference measure

In this section, we propose to advice on the most relevant reference measure ν to be used for the computation of coefficients $h_{i,k}$, $g_{i,k}$ and $f_{i,k}$, $i = 1, 2$ and $k \geq 0$, given in Section 2 depending on the model setup. Note that we will use some of the reference measures described below in the supplementary simulations provided in Appendix C.1.2 and Appendix C.2.2 in which the Patra and Sen [18] estimator of p is used despite the fact that it is not $\sqrt{n}$-consistent.

i) Real line support: the Gaussian measure. When the support of both unknown mixture components is the real line, we can choose for ν the standard normal distribution. The set $\{Q_k, k \in \mathbb{N}\} = \{H_k,\ k \in \mathbb{N}\}$ is constructed from the orthonormal Hermite polynomials, defined for all $x \in \mathbb{R}$ by:

$$H_0 = 1,\ H_1(x) = x,\ \sqrt{k+1}H_{k+1}(x) = xH_k(x) - \sqrt{k}H_{k-1}(x), \quad k \geq 1. \tag{15}$$

ii) Real line support: the Lebesgue measure. When the support is the real line, another choice for ν is the Lebesgue measure on $\mathbb{R}$. In that case we can choose $\{Q_k, k \in \mathbb{N}\} = \{\mathcal{H}_k,\ k \in \mathbb{N}\}$ the set of orthogonal Hermite functions, defined for all $x \in \mathbb{R}$ by:

$$\mathcal{H}_k(x) = H_k(x)\sqrt{f_{\mathcal{N}(0,1)}(x)}, \quad k \geq 0.$$

where H_k is the k-th Hermite polynomial defined in (15).

iii) Positive real line support: the Gamma measure. When the support of both unknown mixture components is the positive real line, we can chose for ν a gamma distribution $\Gamma(1, \alpha)$, with $\alpha > -1$. The set $\{Q_k, k \in \mathbb{N}\}$ is then constructed from the orthogonal Laguerre polynomials defined for all $x \in \mathbb{R}$ by:

$$\begin{aligned} \mathcal{L}_0^\alpha(x) = 1, \quad \mathcal{L}_1^\alpha(x) &= -x + \alpha + 1, \\ -x\mathcal{L}_k^\alpha(x) &= (k+1)\mathcal{L}_{k+1}^\alpha(x) - (2k+\alpha+1)\mathcal{L}_k^\alpha(x) + (k+\alpha)\mathcal{L}_{k-1}^\alpha(x), \quad k \geq 1. \end{aligned}$$

iv) Discrete support: the Poisson measure. If the common support is the set of integers then the choice of ν can be the Poisson distribution with mean $\alpha > 0$ and with associated orthogonal Charlier polynomials defined by:

$$\mathcal{C}_0^\alpha = 1, \quad \mathcal{C}_1^\alpha(x) = (\alpha - x)/\alpha, \qquad x\mathcal{C}_n^\alpha(x) = -\alpha\mathcal{C}_{n+1}^\alpha(x) + (n+\alpha)\mathcal{C}_n^\alpha(x) - n\mathcal{C}_{n-1}^\alpha(x), \quad k \geq 1.$$

v) Bounded support. If the supports are a bounded interval $]a, b[$, $a < b$, we can use a uniform measure for ν and its associated Legendre polynomials. For instance, when $]a, b[=]-1, 1[$ these polynomials are defined for all $x \in \mathbb{R}$ by:

$$L_0 = 1, \quad L_1(x) = x, \quad (k+1)L_{k+1}(x) = (2k+1)xL_k(x) - kL_{k-1}(x), \quad k \geq 1.$$

vi) Wavelets. Another approach is to consider an orthogonal basis of wavelets, say $\{\phi_i, \psi_{i,j}; i, j \in \mathbb{Z}\}$, see Daubechies [6]. Note here that the measure ν is the Lebesgue one. Hence, the density expansions would take the following generic form:

$$f = \sum_{i\in\mathbb{Z}} \langle f, \phi_i\rangle\overline{\phi_i} + \sum_{i\in\mathbb{N}, j\in\mathbb{Z}} \langle f, \psi_{ij}\rangle\overline{\psi_{ij}},$$

with a double sum, which turns out to be heavier to implement in practice.

4.2. Construction of the test statistic

Under the null hypothesis H_0, Theorem 1 gives the asymptotic distribution of the test statistic. The computation of the test statistic first requires the choice of $d(n_1, n_2)$, $e(n_1, n_2)$ and $s(n_1, n_2)$. A previous study (see Pommeret and Vandekherkove [20]) showed that the empirical levels and powers were overall weakly sensitive to $d(n_1, n_2)$ if taken large enough. From that preliminary study, and based on our simulations, we recommend to make the following distinction:

- In case of large enough sample sizes, that is if $n_1\widehat{p}_1 > 30$ and $n_2\widehat{p}_2 > 30$, we decided to set $d(n_1, n_2)$ equal to 10. The trimming $e(n_1, n_2)$ is set equal to $(\log(\max(n_1, n_2)))^{-1}$. The power of the normalization $s(n_1, n_2) = (n_1n_2/(n_1 + n_2))^{s-1}$ is fixed close enough to $-1/2$, with s equal to $2/5$, which seemed to provide good empirical levels.

- In case of small sample sizes or low proportions, that is, if $n_1\widehat{p}_1 < 30$ and/or $n_2\widehat{p}_2 < 30$, the test will be used only with a small value of d (the number of polynomials), for instance $d = 3$. In that case, the estimators are too unstable to evaluate consistently the expansion factors beyond the third degree. Hence, the test consists in the densities comparison up to their third degree in the polynomial expansions. If the null hypothesis is rejected, the densities are considered as different. However, when the null hypothesis is not rejected, it only means that f_1 and f_2 are to be declared to have the same expectation, variance and skewness (but can be different densities in reality).

5. Monte Carlo simulations

Recall first that the samples X and Y are drawn respectively from mixture densities h_1 and h_2, defined in (2), where p_1, p_2 are the unknown mixing proportions and f_1, f_2 are the unknown component densities with respect to a given reference measure ν. Hereafter, simulations are performed to evaluate the empirical level of the test when the densities f_1 and f_2 are symmetric. This level corresponds to the H_0 rejection probability when H_0 is true ($f_1 = f_2$). In practice, this level is expected to reach 5% asymptotically, since one compares our test statistic to the 95-percentile of the χ^2-distribution (see Theorem 1). We also assess the power of the test, i.e. the probability to reject H_0 given that H_0 is false, which informs on the test ability to detect departures from the null hypothesis.

Usually, statisticians first investigate on low levels of the test under different setups before analyzing its power. To have a deeper understanding on the strengths and weaknesses of our test, various simulation schemes are considered including finite mixture models with different component distributions and weights. Also, to check whether the test quality remains acceptable under various settings, we make the parameters of the component distributions vary. Basically, we introduce two opposite situations. In the former, the two component densities f_i and g_i are in close proximity; whereas they are far apart in the latter. For each case of the simulation study, the test is performed one hundred times to evaluate the empirical level (or power). We also fix $n = n_1 = n_2$ for conciseness when presenting the results, acknowledging that the case where $n_1 \neq n_2$ naturally arises with real datasets (see Section 6).

5.1. Empirical levels

Our objective is first to check whether the results significantly differ when changing the component weights and the component distributions of the mixture models. Regarding the weights, we focus on values ranging from 10% to 70%. Such weights are usual in most applications, where the unknown mixture component can be prevalent or not. The challenge is that the test remains effective although the p_i's (unknown component proportions) are low, meaning that few observations would be assigned to them.

Let us start with two-component mixtures of Gaussian distributions and remind that, under the null hypothesis, $f = f_1 = f_2$. The test is conducted in the following cases: a) $g_1 = g_2$ with g_1 far from (ff) f; b) $g_1 = g_2$ with g_1 close to (ct) f; c) $g_1 \neq g_2$ with g_1 ff g_2 and g_1, g_2 ff f; d) $g_1 \neq g_2$ with g_1 ct f and g_2 ff f (the case where $g_1 \neq g_2$ with g_1 ct g_2 is similar to b)). As an illustration, Fig. 1 depicts densities obtained from each of the aforementioned situations. Note that it is sometimes not obvious that h_i follows a typical "bumpy" mixture distribution (due to component proximity, see cases b) and d)), which is all the more interesting when testing H_0.

Tab. 1 summarizes the empirical level of the test depending on the situation. The results are overall satisfactory since the empirical levels are roughly equal to 5% whatever the context. The worst cases, i.e. empirical levels significantly higher than 5%, correspond to situations where at least one unknown mixture component has a low weight. It is indeed difficult to get accurate estimates of the mixture weights in such cases, which obviously impacts the quality of the test because very few observations relate to the unknown component density to be tested. Despite being sensitive to the component weights, our test procedure does not seem to be highly sensitive to the

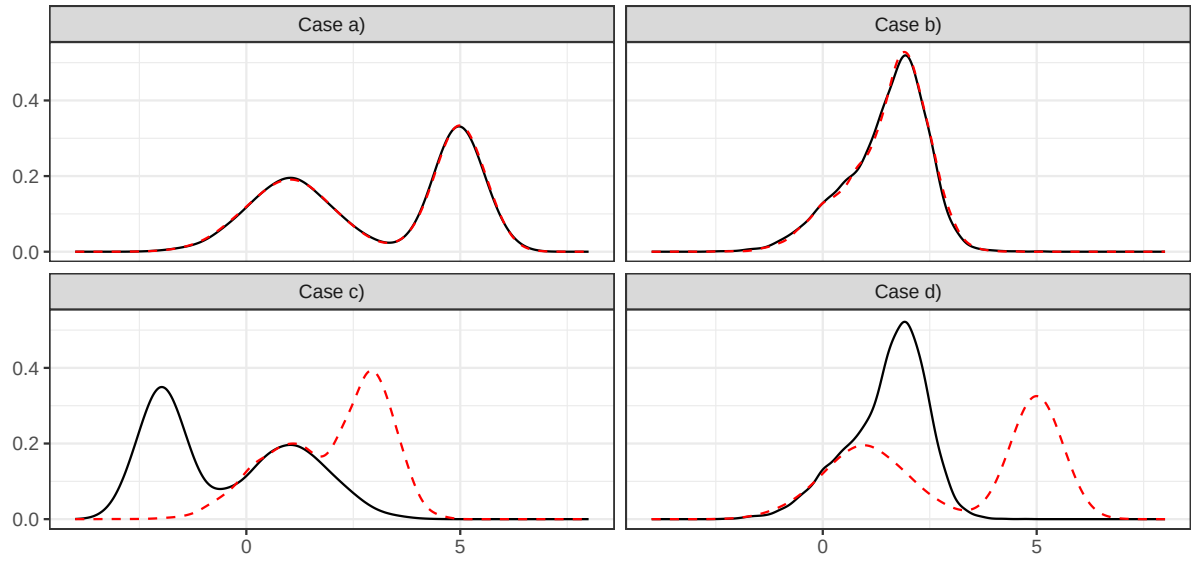

Figure 1: Under H_0. Densities of X (solid) and Y (dashed) in : a) $g_1 = g_2$ with g_1 far from (ff) f; b) $g_1 = g_2$ with g_1 close to (ct) f; c) $g_1 \neq g_2$ with g_1 ff g_2 and g_1, g_2 ff f; d) $g_1 \neq g_2$ with g_1 ct f and g_2 ff f.

number of observations itself. Finally let us notice that the empirical levels tend, as expected, to overall asymptotically decrease.

Tab. 2 shows the obtained results with other types of distributions for f; namely Student, Laplace or Uniform. Our goal here is to check whether the use of different unknown distributions in the mixture models (2) has a noticeable impact on the quality of our procedure. For one given distribution, Tab. 2 stores the worst case (in terms of empirical level) associated to the aforementioned situations a), b), c) and d). Whatever the distribution considered, the worst

Table 1: Empirical level (in %) of the test with two-component Gaussian mixtures in settings a), b), c), and d), see also Fig. 1. Corresponding mixture parameters are stored in Tab. C.5, see Appendix C.1.

			Case a)			Case b)			Case c)			Case d)		
			p_2			p_2			p_2			p_2		
			0.1	0.25	0.7	0.1	0.25	0.7	0.1	0.25	0.7	0.1	0.25	0.7
$n = 1,000$	p_1	0.1	4	8	2	2	6	4	5	7	7	7	5	3
		0.25	3	7	6	5	3	6	2	7	4	6	3	9
		0.7	4	3	5	5	3	9	5	5	6	13	6	6
$n = 4,000$	p_1	0.1	11	6	6	6	6	8	1	4	5	7	5	7
		0.25	7	5	5	4	7	7	1	8	4	1	6	4
		0.7	5	3	4	3	6	3	7	6	5	8	6	5
$n = 10,000$	p_1	0.1	5	6	9	4	7	7	8	3	5	9	3	5
		0.25	3	4	4	6	3	3	4	8	7	4	3	9
		0.7	4	3	4	8	3	5	5	7	1	7	6	4

Table 2: Empirical level (in %) of the test corresponding to case d), identified as the situation providing the worst results whatever the distribution of f. Parameters are given in Tab. C.6 of Appendix C.1.

			Student			Laplace			Uniform		
			p_2			p_2			p_2		
			0.1	0.25	0.7	0.1	0.25	0.7	0.1	0.25	0.7
$n = 1,000$	p_1	0.1	11	14	6	12	10	6	6	10	9
		0.25	9	8	4	8	7	9	9	8	7
		0.7	13	6	8	8	11	7	10	4	5
$n = 4,000$	p_1	0.1	6	3	4	9	5	3	5	2	5
		0.25	8	7	2	8	6	4	3	3	4
		0.7	6	5	1	10	7	6	6	5	3
$n = 10,000$	p_1	0.1	7	4	6	6	5	9	5	6	3
		0.25	4	1	4	5	3	6	4	2	6
		0.7	8	2	5	5	6	3	3	5	4

case stands for case d). In such a situation (see the corresponding density of X in Fig. 1), the mixture weight p_1 is probably hard to estimate correctly. However, contrary to case b), there is no "compensation effect" since p_2 is very likely to be well estimated (the mixture components in Y are well separated). This asymmetric behaviour when estimating the weights p_1 and p_2 deteriorates the quality of the test. This trend tends to vanish when increasing the sample size, which shows the asymptotic detection efficiency of our method on a challenging case. Moreover, let us notice that as long as the sample sizes n_1, n_2 are large enough, the type of distribution considered in the mixture densities (2) has a poor impact on the quality of our testing procedure (other distributions, not presented here, were also tested with similar results). Nevertheless, some bad results happen when the sample size is too small, which can be likely connected to the choice of the distribution parameters themselves (see Tab. C.6 in Appendix C.1). More precisely, high variances of the component distributions seem to cause troubles when estimating the weights p_1 and p_2. Again, this concern tends to naturally disappear when increasing the sample size, thanks to the strong consistency of the semiparametric estimators used in our procedure, see Bordes and Vandekerkhove [2].

5.2. Empirical powers

We now evaluate the ability of our test to detect departures from the null hypothesis. As a starting point, consider the situations where f_1 and f_2 belong to the same distribution family, but have different moments. In the following, the difference can originate from the expectation (case e)) or the variance. When the variances of f_1 and f_2 differ, we are interested in two frameworks: either the difference is big (case f)) or small (case g)). Lastly, we analyse the behaviour of the test when f_1 and f_2 belong to different distribution families with same two first order moments (case h)).

Similarly to Section 5.1, Tab. 3 provides the empirical power of our testing procedure related to Gaussian mixtures in the four aforementioned cases. As it can be noticed, the power of the test is very strongly influenced by the number of observations and is much more sensitive to the sam-

Table 3: Empirical power of the test in two-component Gaussian mixtures in various settings. Mixture parameters are listed in Tab. C.9 of Appendix C.2).

			Case e) $\mathbb{E}[f_1] \neq \mathbb{E}[f_2]$			Case f) $\mathbb{V}(f_1) \neq \mathbb{V}(f_2)$			Case g) $\mathbb{V}(f_1) \simeq \mathbb{V}(f_2)$			Case h) $\mathcal{N}(\mu, \sigma)$ vs $\mathcal{L}(\theta, \nu)$		
			p_2			p_2			p_2			p_2		
			0.1	0.25	0.7	0.1	0.25	0.7	0.1	0.25	0.7	0.1	0.25	0.7
		0.1	51	62	70	41	59	72	13	6	12	4	2	8
$n = 1,000$	p_1	0.25	69	93	100	62	97	100	9	18	34	2	6	6
		0.7	83	100	100	83	100	100	14	41	97	3	5	5
		0.1	100	99	100	96	100	100	17	21	39	1	7	2
$n = 4,000$	p_1	0.25	100	100	100	100	100	100	28	69	93	2	6	3
		0.7	100	100	100	100	100	100	36	96	100	1	3	5
		0.1	100	100	100	100	100	100	31	58	68	3	5	5
$n = 10,000$	p_1	0.25	100	100	100	100	100	100	63	98	100	8	5	4
		0.7	100	100	100	100	100	100	88	100	100	8	10	4

ple sizes than when considering empirical level performances. Indeed, detecting some differences between f_1 and f_2 sometimes requires a lot of data and is more conservative. Concretely, as soon as the difference lies in the skewness, the kurtosis, or higher order moments of the distributions, it becomes very hard to get high powers (except when the size of the data becomes huge). As an example, our trials show that at least $25,000$ observations are needed to reach acceptable powers (70%) in case h) of Tab. 3 (with $p_1 = p_2 = 0.1$ and other parameters listed in Tab. C.9 of Appendix C.2). Of course, for the same reason, the weights also play a key role. Indeed, they somehow represent the exposure of the unknown component densities with a clear impact on our results since basically the bigger the weights are, the higher the power of the test is. Thanks to Tab. 4, we also realize that the type of distribution under study for f_1 and f_2 is not as much impacting as the

Table 4: Empirical power ($n = 1,000$); depending on the distributions for f_1 and f_2, the weights and the order of the moment differentiating f_1 from f_2. The case e) is not considered in the Student case since the mean is fixed (equal to zero). Parameters are stored in Tab. C.10, see Appendix C.2.

			Case e)			Case f)			Case g)			Case h)		
			p_2			p_2			p_2			p_2		
			0.1	0.25	0.7	0.1	0.25	0.7	0.1	0.25	0.7	0.1	0.25	0.7
		0.1	-	-	-	45	47	72	7	17	15	8	5	6
Student	p_1	0.25	-	-	-	64	93	96	10	21	28	7	4	6
		0.7	-	-	-	76	100	100	13	39	94	9	5	9
		0.1	48	57	64	41	48	69	12	8	16	4	5	5
Laplace	p_1	0.25	61	88	100	67	89	97	14	29	33	7	8	6
		0.7	78	97	100	71	93	100	17	49	95	5	8	9
		0.1	59	66	79	52	69	79	12	7	13	4	7	6
Uniform	p_1	0.25	75	94	99	72	100	100	15	34	41	9	8	6
		0.7	88	100	100	86	100	100	22	53	100	7	10	4

sample size or the unknown component weight.

Apart from these statements, our simulations also enable to verify empirically Proposition 1. Under the alternative hypothesis, the selected order of the test statistic should be greater than or equal to the moment order differentiating them. Among the 100 times the test was performed (with $n = 5,000$), more than 80% of the tests have selected the right order following the penalization rule (11); i.e. $k = 1$ in case e), $k = 2$ in case f) or g), and $k \geq 3$ in case h). Let us mention that more than 90% of the tests selected the first rank ($k = 1$) in the decomposition when testing under the null hypothesis, which corroborates our theoretical results.

Finally, additional simulations are presented in the appendices, illustrating that our test strategy can be generalized to other frameworks with practical interest. However, the reader has to keep in mind that the symmetry assumption is not satisfied in such cases, which means that the use of the Patra and Sen estimates for the p_i's make these applications theoretically non valid.

6. Application on Galaxies dataset

We consider two datasets from the SIMBAD Astronomical Database (Observatoire Astronomique de Strasbourg). They gather records of stars heliocentric velocities measurements coming from two dwarf spheroidal (dSph) galaxies: Carina and Sextans. These dSph galaxies are low luminosity galaxies that are companions of the Milky Way and thus are respectively contaminated with Milky Way stars in the field of view.

Hence, the Carina and Sextans measurements at our disposal are mixed with Milky Way stars heliocentric velocity measurements across the stellar landscape. In that sense, the distribution of these measurements can be viewed as an admixture/contamination model (1). Since the Milky Way is very largely observed, see Robin *et al.* [21], it is commonly accepted that its heliocentric velocity (HV) can be expressed as random variable with known probability density function $g = g_1 = g_2$ as in (2). Therefore, we can assume that our two samples are drawn from (2), where f_1 and f_2 stand respectively for the unknown density of the Carina and Sextans galaxy. Fig. 2 shows the probability density estimations of the heliocentric velocity measurements associated to Milky Way and its companions Carina and Sextans. It is based on $N = 170,601$ observations without

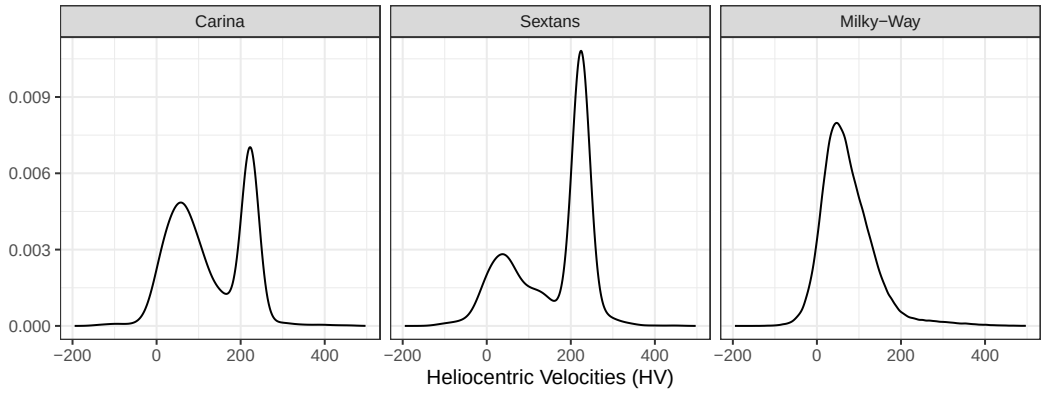

Figure 2: The probability density estimations of the heliocentric velocities of the Carina (contaminated), Sextans (contaminated) and Milky Way galaxies.

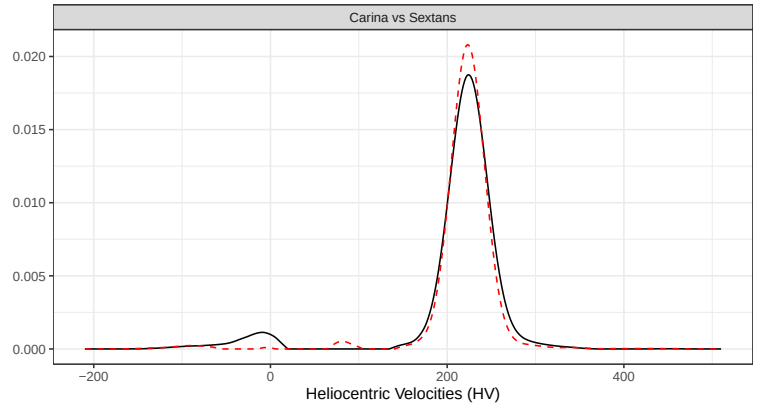

Figure 3: Decontaminated density estimations of the heliocentric velocities of the Carina (dashed) and Sextans (solid) galaxies using Bordes and Vandekerkhove [2] with $(\widehat{p}_1, \widehat{p}_2) = (0.44, 0.56)$.

contamination for Milky Way obtained by Magellan telescope [23], $n_1 = 2,400$ contaminated observations from Carina and $n_2 = 1,488$ contaminated observations from Sextans. It is similar to Fig. 1 shown in [23] where Gaussian assumptions on the densities were used. Such assumptions are not necessary required for our method, and only the knowledge of the moments of the Milky Way HV allows us to implement our test. Note that given the above sample sizes, the technical condition $n_1 = o(N)$ along with $n_2 = o(N)$ required to ensure the validity of our testing method, as discussed in Remark 4, is fully satisfied in this application.

Next, we aim to test if both Carina and Sextans heliocentric velocities have the same distribution, i.e. $f_1 = f_2$. Using the semiparametric estimation procedure in Bordes and Vandekherkove [2] we obtain $\widehat{p}_1 = 0.4446$ and $\widehat{p}_2 = 0.5693$, which means that 44.46% of the Carina HV and 56.93% of the Sextans HV are captured through these datasets. Note that these proportions may vary depending on the position of the data reception. Here they are captured simultaneously and therefore comparable, with the same source of contamination which is from Milky Way. Also, we should note that the estimated proportions differ from the one reported, for example, in Tab. 6 of Patra and Sen [18]. Indeed, the dataset used is larger and was not obtained from the same telescopes. In addition, the corresponding location estimators for Carina and Sextans are respectively: $\widehat{\mu}_1 = 224.5$ and $\widehat{\mu}_2 = 226.4$. Our testing procedure selects the first rank, that is $S(n_1, n_2) = 1$, and provides a test statistic value $\widehat{T}_1 = 0.02567$ with a p-value equal to 0.87. This means that there is no reason to reject the null hypothesis, or more practically, that we can reasonably decide that the Carina and Sextans HV distributions are similar. Note that this conclusion can be visually validated by looking at the decontaminated Carina and Sextans HV densities (see Fig. 3), obtained from the plug-in $\widehat{f}$ inversion formula in Bordes and Vandekerkhove [2] Section 2, where only very slight bumps on the left side tail do not fit exactly. These tiny differences are possibly artefacts due to the aforementioned inversion formula using the approximate $\widehat{p}_i$'s and kernel density estimates $\widehat{h}_i$'s instead of the true p_i's and h_i's.

Discussion

In this work we both theoretically and numerically addressed the two-sample comparison testing problem for two-component mixture models having one known component. More precisely, we proposed a penalized χ^2-type testing procedure allowing a pairwise comparison of the unknown components under a symmetry assumption. We implemented our methodology, with satisfactory results, on a large range of situations, as summarized in Tab. 1-C.11, including Gaussian distributions as well as Laplace, Student or Uniform distributions. We then used our testing procedure on heliocentric velocity comparison for Carina and Sextan galaxies. This real dataset application successfully demonstrates the utility and interpretability of our testing procedure validated by the features comparison of the decontaminated densities, see Fig. 3. The testing procedure requires a splitting technique in order to evaluate the variance of the test statistic, which reduces the sample size by half. However, this splitting technique can turn out to be very challenging when it comes to handle small size samples as it can worsen the underlying weights and moments estimations. To overcome this issue, we think that the use of a bootstrap technique adapted to M-estimators is a promising lead of research (beyond the scope of this paper).

When the symmetry condition **(A1)** cannot be assumed, a possible alternative could be to use the estimators of p_1 and p_2 introduced by Patra and Sen [18] which are proved to not be $\sqrt{n}$-consistent (contrarily to the ones we use for the symmetric case) but seem to provide reasonable results since their use in our testing procedure leads to empirical levels pretty close to 5% across our additional set of simulations, see Appendix C.1.2 and Appendix C.2.2.

We think that this work could be extended in many interesting ways. First we could consider the case where the two samples are paired, with $n_1 = n_2$, as in Ghattas *et al.* [8]. We could probably adapt our testing procedure to obtain results similar to those in Proposition 1 when Theorem 1 could be extended to the paired case by considering the central limit theorem applied on the sequence of random variables $(Q_1(X_j) - Q_1(Y_j))_{j\geq 1}$. This would be particularly interesting for paired time-varying models. Another interesting problem would be the K-sample version of this procedure, which would enable us to deal with time series applications. Moreover, extensions to censored and truncated data would also be of particular interest, especially for insurance applications where it is customary to be confronted with incomplete observations. Finally we started to incorporate our testing procedure in a R package project and plan to enrich it with further developments connected with the admixture testing problem including the above mentionned topics but also the concordance testing problem happening in the z-scores analysis, see Lai et al. [11], which is part of an ongoing work.

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

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

[3] Broët, 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.

[4] Cai, T.T. and Jin, J. (2010). Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. *Ann. Stat.*, **38**, 100–145.

[5] Celisse, A. and Robin, S. (2010). A cross-validation based estimation of the proportion of true null hypotheses. *J. Stat. Plan. Infer.*, **140**, 3132–3147.

[6] Daubechies, I. (1992). *Ten lectures on wavelets*. SIAM: CBMS-NSF Regional Conf. Ser. Appl. Math.

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

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

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

[10] Kallenberg, W.C. and Ledwina, T. (1995). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Ann. Stat.*, **23**(5), 1594-1608.

[11] Lai, Y., Adam, B., Podolsky, R. and She, J. (2007). A mixture model approach to the tests of concordance and discordance between two large-scale experiments with two-sample groups. *Bioinformatics.*, **23**, 1243–1250.

[12] Ledwina, T. (1994). Data-driven version of Neyman's smooth test of Fit. *JASA*, **89**, 1000–1005.

[13] Lokas, E.L. (2009). The mass and velocity anisotropy of the Carina, Fornax, Sculptor and Sextans dwarf spheroidal galaxies. *Mon. Not. R. Astron. Soc.*, 394, 102–106.

[14] 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.

[15] Manton, K.G., Stallard, E. and Vaupel, J. W. (1986). Alternative models for the heterogeneity of mortality risks among the aged. *JASA*, **81**, 635-644.

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

[17] Nguyen, V.H. and Matias, C. (2014). On efficient estimators of the proportion of true null hypotheses in a multiple testing setup. *Scand. J. Stat.*, **41**, 1167-1194.

[18] Patra, R.K. and Sen, B. (2016). Estimation of a Two-component Mixture Model with Applications to Multiple Testing. *J. Roy. Statist. Soc., Series B*, **78**, 869–893.

[19] Podlaski, R. and Roesch, F.A. (2014). Modelling diameter distributions of two-cohort forest stands with various proportions of dominant species: A two-component mixture model approach. *Math. Biosci.*, **249**, 60–74.

[20] Pommeret, D. and Vandekerkhove, P. (2019). Semiparametric density testing in the contamination model. *Electron. J. Stat.*, **13**, 4743-4793.

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

[22] Walker, M., Mateo, M., Olszewski, E., Sen, B. and Woodroofe M. (2009). Clean kinematic samples in dwarf spheroidals: An algorithm for evaluating membership and estimating distribution parameters when contamination is present. *Astron. J.*, **137**, 3109–3138.

[23] Walker, M.G., Mateo, M., Olszewski, E.W., Gnedin, O.Y., Wang, X., Sen, B. and Woodroofe, M. (2007). Velocity dispersion profiles of seven dwarf spheroidal galaxies. *Astrophys. J.*, **667**, L53–L56.

[24] Xiang, S., Yao, W. and Yang, G. (2018) An overview of Semiparametric Extensions of finite Mixture Models. *Statist. Sci.*, **34**, 391–404.

Acknowlegments

The authors would like to thank the two anonymous referees for their constructive comments that contributed to significantly improve the paper. This work was conducted within the Research Chair DIALog under the aegis of the Risk Foundation, a joint initiative by CNP Assurances and ISFA, Université Claude Bernard Lyon 1 (UCBL).

Appendix A. Proofs

Proof of Lemma 1. For simplicity matters, let us denote $\widetilde{n} = n_1 n_2/(n_1+n_2)$. To prove the wanted result it is equivalent to prove that $\mathbb{P}(S(n_1,n_2) \geq 2)$ vanishes as $n_1, n_2 \to +\infty$. By definition of $S(n_1,n_2)$, using the positivity of $\widehat{T}_1$, we have

$$\begin{aligned}
\mathbb{P}(S(n_1,n_2)\geq 2) &= \mathbb{P}\left(\max_{2\leq k\leq d(n_1,n_2)}\{\widetilde{n}^{s-1}\widehat{T}_k - k\log\widetilde{n}\} \geq \widetilde{n}^{s-1}\widehat{T}_1 - \log\widetilde{n}\right)\\
&= \mathbb{P}\left(\exists k,\ 2\leq k\leq d(n_1,n_2):\ \widetilde{n}^{s-1}\widehat{T}_k - k\log\widetilde{n} \geq \widetilde{n}^{s-1}\widehat{T}_1 - \log\widetilde{n}\right)\\
&= \mathbb{P}\left(\exists k,\ 2\leq k\leq d(n_1,n_2):\ \widetilde{n}^{s-1}(\widehat{T}_k - \widehat{T}_1) \geq (k-1)\log\widetilde{n}\right)\\
&= \mathbb{P}\left(\exists k,\ 2\leq k\leq d(n_1,n_2):\ \sum_{j=2}^{k}\widetilde{n}^{s}(\widehat{R}_j)^2/\widehat{d}[j] \geq (k-1)\log\widetilde{n}\right).
\end{aligned}$$

Since $\widehat{d}_j = \max(\widehat{w}_j, e(n_1,n_2))$ we get

$$\mathbb{P}(S(n_1,n_2)\geq 2) \;\leq\; \mathbb{P}\left(\exists k,\ 2\leq k\leq d(n_1,n_2):\ \sum_{j=2}^{k}\widetilde{n}^s(\widehat{R}_j)^2 \geq e(n_1,n_2)(k-1)\log\widetilde{n}\right).$$

We now use the fact that if a sum of positive terms, say $\sum_{j=2}^{k} r_j$ is greater than a constant r, then necessarily there exists a term r_j such that $r_j > r/(k-1)$, to get that

$$\begin{aligned}\mathbb{P}(S(n_1,n_2)\geq 2) &\leq \mathbb{P}\left(\exists k, 2\leq k\leq d(n_1,n_2), \exists j, 2\leq j\leq k, \widetilde{n}^s(\widehat{R}_j)^2 \geq e(n_1,n_2)\log\widetilde{n}\right)\\ &\leq \mathbb{P}\left(\sum_{k=2}^{d(n_1,n_2)}\widetilde{n}^s(\widehat{R}_k)^2 \geq e(n_1,n_2)\log\widetilde{n}\right).\end{aligned}$$

Now decomposing $\widehat{R}_k$ as follows:

$$\begin{aligned}R_k &= \widehat{h}_{1,k}\widehat{p}_2 - \widehat{h}_{2,k}\widehat{p}_1\\ &= (\widehat{h}_{1,k}-p_1\alpha_{1,k})\widehat{p}_2 - (\widehat{h}_{2,k}-p_2\alpha_{2,k})\widehat{p}_1 + \alpha_{1,k}p_1(\widehat{p}_2-p_2) + \alpha_{2,k}p_2(\widehat{p}_1-p_1),\end{aligned}$$

where $\alpha_{i,k} = \int_{\mathbb{R}} Q_k(z)h_i(z)\nu(dz)$, we combine twice the inequality $(a+b)^2\leq 2(a^2+b^2)$, for all $(a,b)\in\mathbb{R}^2$, with $\mathbb{P}(U^2+V^2\geq z)\leq \mathbb{P}(U^2\geq z/2)+\mathbb{P}(V^2\geq z/2)$, for any couple of random variables (U,V), we deduce that

$$\begin{aligned}\mathbb{P}(S(n_1,n_2)\geq 2) &\leq \mathbb{P}\left(\sum_{k=2}^{d(n_1,n_2)}\widetilde{n}^s(\widehat{h}_{1,k}-p_1\alpha_{1,k})^2\widehat{p}_2^2 \geq e(n_1,n_2)\log\widetilde{n}/16\right)\\ &+ \mathbb{P}\left(\sum_{k=2}^{d(n_1,n_2)}\widetilde{n}^s(\widehat{h}_{2,k}-p_2\alpha_{2,k})^2\widehat{p}_1^2 \geq e(n_1,n_2)\log\widetilde{n}/16\right)\\ &+ \mathbb{P}\left(\sum_{k=2}^{d(n_1,n_2)}\widetilde{n}^s\alpha_{1,k}^2p_1^2(\widehat{p}_2-p_2)^2 \geq e(n_1,n_2)\log\widetilde{n}/16\right)\\ &+ \mathbb{P}\left(\sum_{k=2}^{d(n_1,n_2)}\widetilde{n}^s\alpha_{2,k}^2p_2^2(\widehat{p}_1-p_1)^2 \geq e(n_1,n_2)\log\widetilde{n}/16\right).\end{aligned}$$

We study now these four quantities separately. First, using the Markov inequality, we obtain

$$\begin{aligned}\mathbb{P}\left(\sum_{k=2}^{d(n_1,n_2)}\widetilde{n}^s(\widehat{h}_{1,k}-p_1\alpha_{1,k})^2\widehat{p}_2^2 \geq e(n_1,n_2)\log\widetilde{n}/16\right) &\leq \frac{16\widetilde{n}^s}{e(n_1,n_2)\log\widetilde{n}}\sum_{k=2}^{d(n_1,n_2)}\sum_{j=1}^{n_1}\frac{\mathbb{V}(Q_k(X_j))}{n_1}\\ &\leq \frac{16M_1d(n_1,n_2){n_1}^{s-1}}{e(n_1,n_2)\log\widetilde{n}}\left(\frac{n_2}{n_1+n_2}\right)^s,\end{aligned}$$

which tends to zero as n_1, n_2 tend to infinity. Similarly,

$$\mathbb{P}\left(\sum_{k=2}^{d(n_1,n_2)} \widetilde{n}^s(\widehat{h}_{2,k} - p_2\alpha_{2,k})^2\widehat{p}_1^2 \geq e(n_1,n_2)\log\widetilde{n}/16\right) \to 0,$$

as n_1, n_2 tend to infinity. We now consider the two last quantities. Since in addition we have

$$\begin{aligned} p_i^2\alpha_{i,k}^2 &\leq p_i^2\int_{\mathbb{R}} Q_k(z)^2 h_i(z)\nu(dz) \\ &= p_i\left(\int_{\mathbb{R}} Q_k(z)^2 f_i(z)\nu(dz) - (1-p_i)\int_{\mathbb{R}} Q_k(z)^2 g_i(z)\nu(dz)\right) \\ &\leq \int_{\mathbb{R}} Q_k(z)^2 f_i(z)\nu(dz \leq M_1, \end{aligned}$$

it comes that

$$\begin{aligned} &\mathbb{P}\left(\textstyle\sum_{k=2}^{d(n_1,n_2)} \widetilde{n}^s\alpha_{1,k}^2 p_1^2(\widehat{p}_2 - p_2)^2 \geq e(n_1,n_2)\log\widetilde{n}/16\right) \\ &\qquad \leq \mathbb{P}\left((\widehat{p}_2 - p_2)^2 \geq \frac{e(n_1,n_2)\log\widetilde{n}}{16M_1 d(n_1,n_2)\widetilde{n_2}^s}\left(\frac{n_1}{n_1+n_2}\right)^{-s}\right), \end{aligned}$$

which tends to zero since $(\widehat{p}_2 - p_2)^2 = o_{a.s.}(n_2^{1/2+\alpha})$ for all $\alpha > 0$, see Bordes and Vandekerkhove [2]. The same conclusion holds for the last quantity, which is

$$\mathbb{P}\left(\sum_{k=2}^{d(n_1,n_2)} \widetilde{n}^s\alpha_{2,k}^2 p_2^2(\widehat{p}_1 - p_1)^2 \geq e(n_1,n_2)\log\widetilde{n}/16\right) \to 0,$$

and leads us to the wanted result $\mathbb{P}(S(n_1,n_2) \geq 2) \to 0$ as $n_1, n_2 \to +\infty$. □

Proof of Lemma 2. Recall that $n_1/(n_1+n_2) \to a$ as n tends to infinity and that each sample is split into two sub-samples of size $n_1/2$ and $n_2/2$ respectively. From Lemma 1, under H_0 the statistic $T_{S(n_1,n_2)}$ has the same limiting distribution as $\widehat{T}_1$. Combining the independence of $\widehat{p}_1, \widehat{p}_2, \widehat{h}_{1,1}$ and $\widehat{h}_{2,1}$ with their asymptotic normality we have the following convergence in law:

$$\sqrt{\frac{n_1 n_2}{n_1+n_2}}\begin{pmatrix} \widehat{p}_1 - p_1 \\ \widehat{p}_2 - p_2 \\ \widehat{h}_{1,1} - h_{1,1} \\ \widehat{h}_{2,1} - h_{2,1} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma),$$

where $\Sigma = 2\ \mathrm{diag}\left((1-a)\ V_{\widehat{p}_1}, a\ V_{\widehat{p}_2}, (1-a)\ V_1, a\ V_2\right)$, where $V_{\widehat{p}_1}$ and $V_{\widehat{p}_2}$ are the asymptotic variances of $\sqrt{n_1}\widehat{p}_1$ and $\sqrt{n_2}\widehat{p}_2$, respectively, $V_1 = \mathbb{V}(Q_1(X))$ and $V_2 = \mathbb{V}(Q_1(Y))$. The factor 2 which multiplies the diagonal matrix is due to the splitting procedure. Write $L(x,y,z,w) =$

$y(z-(1-x)g_{1,1})-x(w-(1-y)g_{2,1})$. Under the null, (5) implies that $L(p_1,p_2,h_{1,1},h_{2,1})=0$ and then we have

$$\widehat{R}_1 \;=\; L(\widehat{p}_1,\widehat{p}_2,\widehat{h}_{1,1},\widehat{h}_{2,1}) = L(\widehat{p}_1,\widehat{p}_2,\widehat{h}_{1,1},\widehat{h}_{2,1}) - L(p_1,p_2,h_{1,1},h_{2,1}).$$

By applying the Delta method we get the following convergence in law

$$\sqrt{\frac{n_1 n_2}{n_1+n_2}}\Big(L(\widehat{p}_1,\widehat{p}_2,\widehat{h}_{1,1},\widehat{h}_{2,1}) - L(p_1,p_2,h_{1,1},h_{2,1})\Big) \;\xrightarrow{\mathcal{L}}\; \mathcal{N}(0,w_1),$$

where $w_1 = V^\top \Sigma V$, with V the gradient vector of L. We obtain:

$$\begin{aligned} V^\top &= (p_2 g_{1,1} - (h_{2,1}-(1-p_2)g_{2,1}), -p_1 g_{2,1} + (h_{1,1}-(1-p_1)g_{1,1}), p_2, -p_1), \\ w_1 &= 2(1-a)\, V_{\widehat{p}_1}\left(p_2 g_{1,1} - (h_{2,1}-(1-p_2)g_{2,1})\right)^2 + 2(1-a)\, V_1\, p_2^2 \\ &\quad +2a\, V_{\widehat{p}_2}\left(p_1 g_{2,1} - (h_{1,1}-(1-p_1)g_{1,1})\right)^2 + 2a\, V_2\, p_1^2, \end{aligned}$$

that can be consistently estimated by replacing p_1, p_2 by $\widehat{p}_1, \widehat{p}_2$, $V_{\widehat{p}_1}$ and $V_{\widehat{p}_2}$ by their estimators given in [2], and replacing V_1, V_2 and $h_{1,1}, h_{2,1}$ by their empirical estimators. We finally obtain, according to the Slutsky's theorem, the wanted convergence in law

$$\widehat{T}_1 = \frac{n_1 n_2}{n_1+n_2}\widehat{R}_1^2/\widehat{d}_1 \xrightarrow{\mathcal{L}} \chi_1^2,$$

where $\widehat{d}_1$ expression is given in (10). □

Proof of Theorem 1. Since $T(n_1,n_2) = \widehat{T}_{S(n_1,n_2)}$, the proof follows directly from Lemmas 1 and 2 . □

Proof of Proposition 1. We first prove that under $H_1(q)$ we have $\mathbb{P}(S(n_1,n_2)<q)\to 0$ when n_1 and n_2 tend to infinity. We denote again $\widetilde{n} = n_1 n_2/(n_1+n_2)$.

$$\begin{aligned} \mathbb{P}\left(S(n_1,n_2)=k\right) &\le \mathbb{P}\left(\widetilde{n}^{s-1}\widehat{T}_k - k\log(\widetilde{n}) \ge \widetilde{n}^{s-1}\widehat{T}_q - q\log(\widetilde{n})\right) \\ &= \mathbb{P}\left(\widetilde{n}^{s-1}\widehat{T}_q - \widetilde{n}^{s-1}\widehat{T}_k \le (q-k)\log(\widetilde{n})\right) \\ &= \mathbb{P}\left(\widetilde{n}^{s}\sum_{j=k+1}^{q}\widehat{R}_j^2/\widehat{d}_j \le (q-k)\log(\widetilde{n})\right) \\ &\le \mathbb{P}\left(\widetilde{n}^{s}\widehat{R}_q^2/\widehat{d}_q \le (q-k)\log(\widetilde{n})\right) \\ &= \mathbb{P}\left(\frac{\sqrt{\widetilde{n}^s}|\widehat{R}_q|}{\sqrt{\widehat{d}_q\log(\widetilde{n})}} \le \sqrt{(q-k)}\right) \end{aligned}$$

with $0<s<1/2$. We can now apply the following decomposition

$$\begin{aligned} \frac{\sqrt{\widetilde{n}^s}\widehat{R}_q}{\sqrt{\widehat{d}_q\log(\widetilde{n})}} &= \frac{1}{\sqrt{\widetilde{n}^{1-s}\log(\widetilde{n})}} \times \frac{\sqrt{\widetilde{n}}}{\sqrt{\widehat{d}_q}}\Big(\widehat{R}_q - \delta(q)\Big) + \frac{\sqrt{\widetilde{n}^s}}{\widehat{d}_q\log(\widetilde{n})}\delta(q) \\ &= A + B, \end{aligned}$$

where $\delta(q)$ is defined in (13) and $\widehat{d}_q$ is a consistent estimator of the q-th diagonal term of variance matrix D given in (10). Mimicking the proof of Lemma 2 we can show that $\sqrt{\widetilde{n}}\left(\widehat{R}_q - \delta(q)\right)$ is asymptotically Gaussian and then A converges to a Dirac at point zero. Moreover B converges to $+\infty$ since $\delta(q) > 0$. Then for all $k < q$, we have

$$\mathbb{P}\left(S(n_1, n_2) = k\right) \to 0,$$

along with $T_q > \widetilde{n}\widehat{R}_q^2/\widehat{d}_q \to +\infty$ as n_1, n_2 tend to infinity.

Appendix B. Estimation of the asymptotic variance of $\widehat{R}_k$

To overcome the complex dependence between the estimators of p_1, p_2 and the estimators of the coefficients associated with h_1 and h_2, we split each sample into two independent sub-samples of size n_1', n_1'' for X and n_2', n_2'' for Y, with $n_1' + n_1'' = n_1$ and $n_2' + n_2'' = n_2$, respectively. We then use the first sub-samples to estimate the coefficients of h_1 and h_2, and the second sub-samples to estimate the proportions p_1 and p_2. For simplicity we fix with $n_1' = n_1'' = n_1/2$ and $n_2' = n_2'' = n_2/2$, respectively. Write $\widetilde{n} = (n_1 n_2)/(n_1 + n_2)$ and recall that $n_1/(n_1 + n_2) \to a$. We then use the following convergences in law (due accordingly to the central limit theorem or to the asymptotic normality of the Bordes and Vandekerkhove [2] estimators):

$$\begin{aligned}
\sqrt{\widetilde{n}}(\widehat{h}_{1,k} - h_{1,k}) &\xrightarrow{\mathcal{L}} 2bN(0, V_{1,k}), \qquad \sqrt{\widetilde{n}}(\widehat{h}_{2,k} - h_{2,k}) \xrightarrow{\mathcal{L}} 2bN(0, V_{2,k}), \\
\sqrt{\widetilde{n}}(\widehat{p}_1 - p_1) &\xrightarrow{\mathcal{L}} 2aN(0, V_{\widehat{p}_1}), \qquad \sqrt{\widetilde{n}}(\widehat{p}_2 - p_2) \xrightarrow{\mathcal{L}} 2aN(0, V_{\widehat{p}_2}),
\end{aligned}$$

where $b = 1 - a$, $V_{1,k} = \mathbb{V}(Q_k(X))$, $V_{2,k} = \mathbb{V}(Q_k(Y))$, and $V_{\widehat{p}_1}$, $V_{\widehat{p}_2}$ are the asymptotic variances of $\sqrt{n_1}\widehat{p}_1$ and $\sqrt{n_2}\widehat{p}_2$, respectively. To estimate $V_{1,k}$ and $V_{2,k}$, we use the following empirical estimators:

$$\begin{aligned}
\widehat{V}_{1,k} &= \frac{1}{n_1'}\sum_{i=1}^{n_1'} Q_j(X_i)^2 - \left(\frac{1}{n_1'}\sum_{i=1}^{n_1'} Q_j(X_i)\right)^2, \\
\widehat{V}_{2,k} &= \frac{1}{n_2'}\sum_{i=1}^{n_2'} Q_j(Y_i)^2 - \left(\frac{1}{n_2'}\sum_{i=1}^{n_2'} Q_j(Y_i)\right)^2.
\end{aligned}$$

Combining the independence of the basic estimators (obtained by the splitting step) and the Delta method we finally get the convergence in law:

$$\sqrt{\widetilde{n}}\widehat{R}_k \xrightarrow{\mathcal{L}} N(0, W_k),$$

where

$$\begin{aligned}
W_k &= 2bV_{1,k}p_2^2 + 2bV_{\widehat{p}_1}\left((h_{2,k} - (1-p_2)g_{2,k}) - p_2 g_{1,k}\right)^2 + 2aV_{2,k}p_1^2 \\
&\quad + 2aV_{\widehat{p}_2}\left((h_{1,k} - (1-p_1)g_{1,k}) - p_1 g_{2,k}\right)^2.
\end{aligned}$$

The variances $V_{\widehat{p}_1}$ and $V_{\widehat{p}_2}$ and their estimators are entirely described in Appendix A of Bordes and Vandekerkhove [2]. These estimators are denoted by $\widehat{V}_{\widehat{p}_1}$ and $\widehat{V}_{\widehat{p}_2}$. Finally we get the following estimator $\widehat{w}_k$ of $\sqrt{\widetilde{n}}\widehat{R}_k$:

$$\begin{aligned}\widehat{w}_k &= 2b\widehat{V}_{1,k}\widehat{p}_2^2 + 2b\widehat{V}_{\widehat{p}_1}\left((\widehat{h}_{2,k} - (1-\widehat{p}_2)g_{2,k}) - \widehat{p}_2 g_{1,k}\right)^2 + 2a\widehat{V}_{2,k}\widehat{p}_1^2 \\ &\quad +2a\widehat{V}_{\widehat{p}_2}\left((\widehat{h}_{1,k} - (1-\widehat{p}_1)g_{1,k}) - \widehat{p}_1 g_{2,k}\right)^2.\end{aligned}$$

Appendix C. Additional tables on Monte Carlo experiments

Appendix C.1. Empirical levels

Appendix C.1.1. Symmetric case based on Bordes and Vandekherkove [2] estimators

Table C.5: Parameters corresponding to Tab. 1 and Fig. 1.

	Case a)	Case b)	Case c)	Case d)
f	$\mathcal{N}(1,1)$	$\mathcal{N}(1,1)$	$\mathcal{N}(1,1)$	$\mathcal{N}(1,1)$
g_1	$\mathcal{N}(5,0.5)$	$\mathcal{N}(2,0.5)$	$\mathcal{N}(-2,0.5)$	$\mathcal{N}(2,0.5)$
g_2	$\mathcal{N}(5,0.5)$	$\mathcal{N}(2,0.5)$	$\mathcal{N}(3,0.5)$	$\mathcal{N}(5,0.5)$
p_1	50%	50%	50%	50%
p_2	50%	50%	50%	50%

Table C.6: Parameters corresponding to Tab. 2, in case d).

	Student t			Laplace $\mathcal{L}$			Uniform $\mathcal{U}$		
	parameters	$\mathbb{E}$	$\mathbb{V}ar$	parameters	$\mathbb{E}$	$\mathbb{V}ar$	parameters	$\mathbb{E}$	$\mathbb{V}ar$
f	$t(\nu=4)$	0	2	$\mathcal{L}(0,1)$	0	2	$\mathcal{U}(0,1)$	0.5	1/12
g_1	$\mathcal{N}(0,1)$	0	1	$\mathcal{N}(0,1)$	0	1	$\mathcal{N}(0,1)$	0	1
g_2	$\mathcal{N}(5,0.5)$	5	0.25	$\mathcal{N}(5,0.5)$	5	0.25	$\mathcal{N}(5,0.5)$	5	0.25

Appendix C.1.2. Extended study based on Patra and Sen [18] estimators

Let us stress out that our testing methodology is based on $\sqrt{n}$-consistent estimators of p_1 and p_2, but this is not the case anymore when considering the relaxed shape constraints in Patra and Sen [18]. However, in practice and due to the fact that Lemma 1 remains valid under this last framework (the almost sure rates in Patra and Sen [18] insure the correct rank selection under H_0), it is interesting to look at our test performances in setups different than those studied until now, especially for practical applications. Tab. C.7 summarizes some results obtained when considering different types of distribution supports ($\mathbb{R}^+$, $\mathbb{N}$, and $[0,1]$).

Table C.7: Empirical level (in %) of the test corresponding to case d), identified as the situation providing the worst results whatever the support. Parameters are given in Tab. C.8 below.

Support:			$\mathbb{R}^{+\star}$			$\mathbb{N}$			$[0,1]$		
			p_2			p_2			p_2		
			0.1	0.25	0.7	0.1	0.25	0.7	0.1	0.25	0.7
$n = 1,000$	p_1	0.1	12	10	15	6	8	9	6	10	9
		0.25	13	19	10	8	9	9	9	8	4
		0.7	22	26	21	7	7	8	5	4	4
$n = 4,000$	p_1	0.1	8	3	8	9	12	3	4	2	5
		0.25	3	2	1	12	8	3	4	2	7
		0.7	18	12	11	5	11	6	4	2	5
$n = 10,000$	p_1	0.1	5	8	6	6	3	6	5	4	3
		0.25	3	6	5	4	3	6	4	2	7
		0.7	7	3	6	2	1	4	7	5	7

Table C.8: Parameters corresponding to Tab. C.7, in case d). Notations used: $\mathcal{G}$ = Gamma, $\mathcal{E}$ = Exponential, $\mathcal{P}$ = Poisson, $\mathcal{BN}$ = Negative Binomial, and $\mathcal{U}$ = Uniform.

	Support $\mathbb{R}^+$			Support $\mathbb{N}$			Support $[0,1]$		
	parameters	$\mathbb{E}$	$\mathbb{V}ar$	parameters	$\mathbb{E}$	$\mathbb{V}ar$	parameters	$\mathbb{E}$	$\mathbb{V}ar$
f	$\mathcal{G}(16,4)$	4	1	$\mathcal{BN}(1,10)$	1	1.1	$\mathcal{B}eta(1.2,5)$	0.2	0.02
g_1	$\mathcal{E}(1/4)$	4	16	$\mathcal{P}(1)$	1	1	$\mathcal{U}(0,0.4)$	0.2	0.013
g_2	$\mathcal{E}(2)$	0.5	0.25	$\mathcal{P}(4)$	4	4	$\mathcal{U}(0.05,1)$	0.55	0.075

Appendix C.2. Empirical powers

We first give here the parameters involved in the simulations for the symmetric case, see Tab. C.9 and C.10. Then, in the spirit of the previous section, we investigate the power of the test in nonsymmetric cases, see Tab. C.11, and provide the corresponding distributions along with their associated parameters in Tab. C.12.

Appendix C.2.1. Symmetric case based on Bordes and Vandekherkove [2] estimators

Table C.9: Parameters corresponding to Tab. 3, for cases e), f), g) and h) in the Gaussian case.

	e) $\neq$ means	f) very $\neq$ variances	g) slightly $\neq$ variances	h) $\neq$ distributions
$g_1 = g_2$	$\mathcal{N}(5,0.5)$	$\mathcal{N}(5,0.5)$	$\mathcal{N}(5,0.5)$	$\mathcal{N}(5,0.5)$
f_1	$\mathcal{N}(1,1)$	$\mathcal{N}(1,1)$	$\mathcal{N}(1,1)$	$\mathcal{N}(1,\sqrt{2})$
f_2	$\mathcal{N}(2,1)$	$\mathcal{N}(1,3)$	$\mathcal{N}(1,\sqrt{2})$	$\mathcal{L}aplace(1,1)$

Table C.10: Parameters corresponding to Tab. 4 (Student (t), Laplace ($\mathcal{L}$) and Uniform ($\mathcal{U}$) distributions).

	e) $\neq$ means	f) very $\neq$ variances	g) slightly $\neq$ variances	h) $\neq$ distributions
$g_1 = g_2$	$\mathcal{N}(5, 0.5)$	$\mathcal{N}(5, 0.5)$	$\mathcal{N}(5, 0.5)$	$\mathcal{N}(5, 0.5)$
f_1	–	$t(4)$	$t(8)$	$t(3)$
f_2	–	$t(3)$	$t(7)$	$\mathcal{N}(0, \sqrt{3})$
f_1	$\mathcal{L}(1, \sqrt{0.5})$	$\mathcal{L}(1, 0.5)$	$\mathcal{L}(1, \sqrt{0.5})$	$\mathcal{L}(0, 1)$
f_2	$\mathcal{L}(3, \sqrt{0.5})$	$\mathcal{L}(1, 1)$	$\mathcal{L}(1, \sqrt{0.6})$	$t(4)$
f_1	$\mathcal{U}(0, 5)$	$\mathcal{U}(0, 5)$	$\mathcal{U}(0, 5)$	$\mathcal{U}(0, 5)$
f_2	$\mathcal{U}(0, 10)$	$\mathcal{U}(1, 4)$	$\mathcal{U}(0.2, 4.8)$	$\mathcal{L}(2.5, 1)$

Appendix C.2.2. Extended study based on Patra and Sen [18] estimators

Tab. C.11 shows that similar conclusions hold in comparison with the symmetric case when changing the support and the distributions of the mixture components. Some slight differences are very likely to come from the choice of the parameters and the types of the component distributions (see Tab. C.12 below), but results are mainly in line with our expectations. Moreover, Tab. C.11 gives additional information about a sort of lower bound on the powers relatively to the studied cases. Indeed, the sample size n chosen here is equal to $1,000$ which turns out to be the poorest informed case over our simulations setups.

Table C.11: Empirical power of the test ($n = 1,000$); depending on the support, the weights and the order of the moment differentiating f_1 from f_2. Parameters are stored in Tab. C.12.

			Case e)			Case f)			Case g)			Case h)		
			p_2			p_2			p_2			p_2		
			0.1	0.25	0.7	0.1	0.25	0.7	0.1	0.25	0.7	0.1	0.25	0.7
$\mathbb{R}^+$	p_1	0.1	58	75	85	29	28	33	9	7	5	8	7	9
		0.25	82	100	100	60	92	98	3	6	7	11	5	10
		0.7	84	100	100	61	100	100	8	14	17	9	4	6
$\mathbb{N}$	p_1	0.1	22	26	27	14	22	43	5	11	7	14	5	5
		0.25	30	73	94	18	76	96	6	6	19	5	10	6
		0.7	43	97	100	16	91	100	12	11	78	12	4	14
$[0, 1]$	p_1	0.1	16	22	30	19	45	43	6	5	10	10	7	8
		0.25	26	72	95	45	100	100	16	33	62	9	20	20
		0.7	37	90	100	38	100	100	7	60	100	12	34	83

Table C.12: Parameters corresponding to Tab. C.11. Notations: $\mathcal{G}$ = Gamma, $\mathcal{E}$ = Exponential, $\mathcal{P}$ = Poisson, $\mathcal{BN}$ = Negative Binomial, and $\mathcal{U}$ = Uniform, $\mathcal{L}og\mathcal{N}$ = Logit Normal, $\mathcal{G}o$ = Gompertz.

	Support $\mathbb{R}^+$		Support $\mathbb{N}$		Support $[0,1]$	
	e)	f)	e)	f)	e)	f)
f_1	$\mathcal{G}(16,4)$	$\mathcal{G}(8,2)$	$\mathcal{BN}(1,10)$	$\mathcal{BN}(2,10)$	$\mathcal{B}eta(0.8,5)$	$\mathcal{B}eta(12,50)$
g_1	$\mathcal{E}(1/1.1)$	$\mathcal{E}(1/1.1)$	$\mathcal{P}(5)$	$\mathcal{P}(5)$	$\mathcal{U}(0,1)$	$\mathcal{U}(0,1)$
f_2	$\mathcal{G}(16,5)$	$\mathcal{N}(32,8)$	$\mathcal{BN}(2,10)$	$\mathcal{BN}(2,0.5)$	$\mathcal{B}eta(1.2,5)$	$\mathcal{B}eta(1.2,5)$
g_2	$\mathcal{E}(1/1.1)$	$\mathcal{E}(1/1.1)$	$\mathcal{P}(5)$	$\mathcal{P}(5)$	$\mathcal{U}(0,1)$	$\mathcal{U}(0,1)$
	g)	h)	g)	h)	g)	h)
f_1	$\mathcal{G}(8,2)$	$\mathcal{G}(1.47,0.56)$	$\mathcal{BN}(2,10)$	$\mathcal{BN}(3,100)$	$\mathcal{B}eta(1.2,5)$	$\mathcal{B}eta(5,2)$
g_1	$\mathcal{E}(1/1.1)$	$\mathcal{E}(1/1.1)$	$\mathcal{P}(5)$	$\mathcal{P}(3)$	$\mathcal{U}(0,1)$	$\mathcal{U}(0,1)$
f_2	$\mathcal{G}(10,2.5)$	$\mathcal{G}o(0.1,0.3)$	$\mathcal{BN}(2,2)$	$\mathcal{B}(50,0.06)$	$\mathcal{B}eta(2.4,10)$	$\mathcal{L}og\mathcal{N}(0.9,0.8)$
g_2	$\mathcal{E}(1/1.1)$	$\mathcal{E}(1/1.1)$	$\mathcal{P}(5)$	$\mathcal{P}(3)$	$\mathcal{U}(0,1)$	$\mathcal{U}(0,1)$