



**HAL**  
open science

# Shape constraint free two-sample contamination model testing

Xavier Milhaud, Denys Pommeret, Yahia Salhi, Pierre Vandekerkhove

► **To cite this version:**

Xavier Milhaud, Denys Pommeret, Yahia Salhi, Pierre Vandekerkhove. Shape constraint free two-sample contamination model testing. 2021. hal-03201760

**HAL Id: hal-03201760**

**<https://hal.science/hal-03201760v1>**

Preprint submitted on 19 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 SHAPE CONSTRAINT FREE TWO-SAMPLE  
2 CONTAMINATION MODEL TESTING

3 Xavier Milhaud,<sup>(1)</sup> Denys Pommeret,<sup>(1,2)</sup> Yahia Salhi<sup>(1)</sup> and  
4 Pierre Vandekerkhove<sup>(3)</sup>

5 <sup>(1)</sup>Univ Lyon, UCBL, ISFA LSAF EA2429, F-69007, Lyon, France

6 <sup>(2)</sup>Aix-Marseille University, Campus de Luminy, 13288 Marseille cedex 9, France

7 <sup>(3)</sup>Université Gustave Eiffel, LAMA (UMR 8050), 77420 Champs-sur-Marne, France

8 April 19, 2021

9 **Abstract**

10 In this paper, we consider two-component mixture distributions having one known com-  
11 ponent. This type of model is of particular interest when a known random phenomenon  
12 is contaminated by an unknown random effect. We propose in this setup to compare the  
13 unknown random sources involved in two separate samples. For this purpose, we introduce  
14 the so-called IBM (Inversion-Best Matching) approach resulting in a relaxed semiparametric  
15 Cramér-von Mises type two-sample test requiring very minimal assumptions (shape constraint  
16 free) about the unknown distributions. The accomplishment of our work lies in the fact that  
17 we establish a functional central limit theorem on the proportion parameters along with the  
18 unknown cumulative distribution functions of the model when Patra and Sen [22] prove that  
19 the  $\sqrt{n}$ -rate cannot be achieved on these quantities in the basic one-sample case. An in-  
20 tensive numerical study is carried out from a large range of simulation setups to illustrate  
21 the asymptotic properties of our test. Finally, our testing procedure is applied to a real-life  
22 application through pairwise post-covid mortality effect testing across a panel of European  
23 countries.

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

25 **Key words.** finite mixture model, semiparametric estimation, Cramér-von Mises, mortality.

# 1 Introduction

Let us consider the semiparametric two-component mixture model with cumulative distribution function (cdf)

$$L(x) = (1 - p)G(x) + pF(x), \quad x \in \mathbb{R}, \quad (1)$$

where  $G$  is a known cdf and where the unknown parameters are the mixture proportion  $p \in ]0, 1[$  and the cdf  $F$  which is not supposed to belong to any parametric family. This model, sometimes so-called *contamination* or *admixture* model, has been widely investigated in the last decades, see for instance Bordes and Vandekherkove [4], Matias and Nguyen [21], Cai and Jin [6] or Celisse and Robin [7] 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 done in Broët *et al.* [5]; ii) the false discovery rate problem (used to assess and control multiple error rates such as in Efron and Tibshirani [9]), see McLachlan *et al.* [19]; iii) astronomy, in which this model arises when observing variables such as metallicity and radial velocity of stars as in Walker *et al.* [26]; iv) biology to model trees diameters, see Podlaski and Roesch [23]; v) kinetics to model plasma data, see Klingenberg *et al.* [14], vi) genomics to represent populations formed by admixture between ancestral founding populations as in Chakraborty and Weiss [8], among many other fields of applications. We recommend also the excellent survey on semiparametric mixture models by Xiang *et al.* [28] to have a panoramic view on this last generation of mixture models.

In this paper, the data of interest is made of two i.i.d. samples  $X_1 = (X_{1,1}, \dots, X_{1,n_1})$  and  $X_2 = (X_{2,1}, \dots, X_{2,n_2})$  with respective cdfs:

$$\begin{cases} L_1(x) = (1 - p_1)G_1(x) + p_1F_1(x), & x \in \mathbb{R} \\ L_2(x) = (1 - p_2)G_2(x) + p_2F_2(x), & x \in \mathbb{R}, \end{cases} \quad (2)$$

where  $p_1, p_2$  are the unknown mixture proportions and  $F_1, F_2$  are unknown cdfs component we propose to name *nodular* distributions. For simplicity matters, we denote  $n = n_1$  and consider that  $n_2 = \kappa n$  where  $\kappa \geq 1$ . In this work, similarly to Patra and Sen [22], we will consider situations where the  $G_i$ 's and  $F_i$ 's distributions are: i) absolutely continuous with respect to the Lebesgue measure, supported over  $\mathbb{R}, \mathbb{R}^+$  or intervals of  $\mathbb{R}$ ; ii) finite discrete or  $\mathbb{N}$ -discrete distributions such as Binomial or Poisson; iii) a mixture of a discrete and an absolutely continuous distribution. All our results will be still valid in such frameworks. Given the above model, our goal is now to answer the following statistical problem:

$$H_0 : F_1 \text{ is equal to } F_2 \quad \text{against} \quad H_1 : F_1 \text{ is different from } F_2 \quad (3)$$

without assigning any specific parametric family to the  $F_i$ 's. Such a problem arises in various applications. For instance, studying the genetic make up of populations has been the subject of much attention. The *admixture* behaviour is of paramount importance in comparing the populations with known ancestors and exhibits linkage relationships between them, see Loh *et al.* [17]. In a more general sense, such a testing problem arises when a known random phenomenon is contaminated by an unknown “nodular” random effect, which may correspond to non-observed heterogeneity. This can be also the case, for example, during crisis

56 where some populations are clearly impacted when others are not concerned yet. The objective  
 57 of the test is then to compare two independent populations that are contaminated. The case of  
 58 the coronavirus disease's example is appealing in the sense that the excess of mortality is clearly  
 59 affecting populations over the world in a different manner. Its nodular impact, among others, can  
 60 be treated under the statistical problem exposed in (3), where the known component refers the  
 61 baseline mortality observed during the recent years.

62 The testing strategy we propose here is very different from the one proposed in Milhaud et  
 63 al. [20] where a semiparametric penalized  $\chi^2$ -type test is used. The latter test is based on a  
 64  $\sqrt{n}$ -consistent estimation  $\widehat{\mathbf{p}}_n = (\widehat{p}_{1,n}, \widehat{p}_{2,n})$  of  $\mathbf{p} = (p_1, p_2)$  along with a pairwise comparison of  
 65 the  $F_i$ 's  $\widehat{\mathbf{p}}_n$ -plugged in orthogonal polynomial basis expansion coefficients estimation. The  $\sqrt{n}$ -  
 66 consistency of  $\widehat{\mathbf{p}}_n$  is satisfied under both  $H_0$  or  $H_1$  when the zero-symmetry of the  $f_i$ 's, pdf of  
 67 the  $F_i$ 's is assumed, according to Bordes and Vandekerkhove [4]. However when considering the  
 68 general Patra and Sen setup (see [22], Theorem 3), the  $\sqrt{n}$ -consistency is not theoretically achieved.  
 69 Despite that, Milhaud et al. [20] proposed interestingly to plug-in the  $\widehat{\mathbf{p}}_n$  estimate of Patra and  
 70 Sen [22] in their testing approach to study its numerical performance in practice. To overcome the  
 71 lack of  $\sqrt{n}$ -consistency under  $H_0$  or  $H_1$  in the Patra and Sen [22] setup, and after all get a complete  
 72 valid asymptotic theory, we decided to rethink from scratch the two-sample testing problem (3).

73 Our idea relies basically on the definition of two parametric function families:

$$\mathcal{F}_i = \left\{ F_i(x, L_i, p_i) := \frac{L_i(x) - (1 - p_i)G_i(x)}{p_i}, \quad p_i \in \Theta_i, \quad x \in \mathbb{R} \right\}, \quad i = 1, 2, \quad (4)$$

74 where  $\Theta_i$  is a compact set of  $\mathbb{R}^+ \setminus \{0\}$ . For simplicity matters and without loss of generality we  
 75 will consider  $\Theta_1 = \Theta_2 = [\delta_1, \delta_2]$ , where  $0 < \delta_1 < 1 < \delta_2 < +\infty$ . It is important to notice here that  
 76 the  $\mathcal{F}_i$ 's are not constrained to contain exclusively cumulative distribution functions and that the  
 77 parametric space  $\Theta_i$  associated to the  $p_i$ 's is not necessarily a  $[\delta, 1 - \delta]$ -type subset,  $0 < \delta < 1$ ,  
 78 of the natural  $]0, 1[$  mixture proportion support anymore. On the other hand, the key point is  
 79 that the true  $F_i$ 's belong respectively to these classes by picking the true value of the parameters  
 80  $p_i^* \in \Theta_i$ , when  $\delta \geq \delta_1 > 0$  are taken small enough, *i.e.*

$$F_i(x) = F_i(x, L_i, p_i^*), \quad x \in \mathbb{R}. \quad (5)$$

81 For convenience we introduce  $\mathcal{F}$  the set of all the probability cumulative distribution functions.  
 82 Consider now the following discrepancy measure

$$d(\theta) = \int_{\mathbb{R}} (F_1(x, L_1, p_1) - F_2(x, L_2, p_2))^2 dU(x), \quad (6)$$

83 with  $\theta = (p_1, p_2) \in \Theta = [\delta_1, \delta_2]^2$  (because these quantities are not specifically viewed as mix-  
 84 ture proportions anymore), measuring possible departures between the functions  $F_1(\cdot, L_1, p_1)$  and  
 85  $F_2(\cdot, L_2, p_2)$ . We denote  $\theta = (p_1, p_2)$  contrarily to  $\mathbf{p}$  previously to stress out the fact that the fitting  
 86 in  $(p_1, p_2)$  is done jointly now when it was done  $\mathcal{F}_i$ -wisely,  $i = 1, 2$ , in Patra and Sen setup [22] or  
 87 Bordes and Vandekerkhove [4]. The integrating cdf  $U$  should obviously be ideally chosen in order  
 88 to focus (highly weight) on domains where the  $F_i(\cdot, L_i, p_i)$ 's clearly depart from each other to help  
 89 on the final test decision. Nevertheless since the structure of the  $F_i$ 's is constantly changing as the

90  $p_i$ 's vary in the parametric space, we propose in practice to consider for  $U$  rather flat distributions  
 91 encompassing the support of the observations.

92 It is worth to notice that under  $H_0$ , there exists  $p_1 = p_1^*$  and  $p_2 = p_2^*$  such that  $d(\theta^*) = d(p_1^*, p_2^*) = 0$ .  
 93 Suppose now that under some regularity and identifiability-type conditions we could prove that:

$$\left\{ \begin{array}{l} \arg \min_{\theta \in \Theta} d(\theta) = \theta^* \\ d(\theta^*) = 0, \end{array} \right. \quad \text{under } H_0, \quad \text{and} \quad \left\{ \begin{array}{l} \arg \min_{\theta \in \Theta} d(\theta) = \theta^c \\ d(\theta^c) > 0, \end{array} \right. \quad \text{under } H_1, \quad (7)$$

94 we would directly have under  $H_1$ :

$$\inf_{\theta \in ]\delta, 1-\delta]^2: F_i(\cdot, L_i, p_i) \in \mathcal{F}, i=1,2} d(\theta) \geq \inf_{\theta \in \Theta} d(\theta) = d(\theta^c) > 0. \quad (8)$$

95 Note that the search of the infimum in the left hand side of (8) matches what we would normally  
 96 expect in a classical semiparametric estimation problem (mixing proportions in  $]\delta, 1-\delta[$ , for  $\delta > 0$   
 97 small enough, and  $F_i$ 's in the cdfs range), when the second infimum have much more relaxed  
 98 constraints (mixing proportions in a compact set  $\Theta$  of  $(\mathbb{R}^+)^2$  embedding  $]\delta, 1-\delta]^2$  and no specific  
 99 constraints on the  $F_i$ 's) that we claim to be sufficient to solve our two sample testing problem.  
 100 This relaxation in the optimisation problem, see (7-8), that was a blocking point to achieve the  
 101  $\sqrt{n}$ -consistency of the estimator  $\hat{p}_n$  in Patra and Sen [22] –method returning a parameter  $p \in ]0, 1[$   
 102 and a true isotonic regression-based cdf estimate for  $F$ , is the first key idea of our paper. It is  
 103 important to notice at this stage that if we let the parameters  $p_i$ ,  $i = 1, 2$ , go together to infinity  
 104 in the parametric expressions (4), the functions  $F_i(\cdot, L_i, p_i)$  mechanically flatten to 0 which makes  
 105 the discrepancy measure  $d(\theta) \rightarrow 0$ , see expression (6). As illustrated in Appendix B, we then  
 106 have to face two types of situations: i) there exists a local minima of  $d(\theta)$ ,  $\theta^*$  under  $H_0$  or  $\theta^c$   
 107 under  $H_1$ , in the interior of the parametric space, and then the testing problem is non-trivial and  
 108 should be addressed, or ii) the optimization of  $d(\theta)$  shows that we bump into the boundaries of the  
 109 parametric space, *i.e.* one of the component of  $\theta^c$  is equal to  $\delta_2$  because the main way to reduce  
 110 the contrast  $d(\theta)$  is to make  $\theta$  large, and then the testing problem is not even worth to be adressed  
 111 because there is no “reasonable”  $\theta^c = (p_1^c, p_2^c)$  close to the probability weights domain  $[0, 1]^2$  that  
 112 make  $F_1(x, L_1, p_1^c)$  close to  $F_2(x, L_2, p_2^c)$ .

113 Now the empirical estimate  $d_n(\cdot)$  of  $d(\cdot)$  obtained by replacing the  $L_i$ 's by the accessible em-  
 114 pirical cdfs  $\hat{L}_i$  in (6), would naturally lead us to find, respectively under  $H_0$  or  $H_1$ , the true value  
 115 of the parameter  $\theta^*$ , respectively the  $(\mathcal{F}_1, \mathcal{F}_2)$ -models distance minimizer  $\theta^c$ , by looking at:

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} d_n(\theta). \quad (9)$$

116 We propose to call IBM-method the semiparametric estimation strategy based on the “Inversion”  
 117 step (4) and the “Best Matching” step (9) between the  $\mathcal{F}_1$  and  $\mathcal{F}_2$  families to look at the closest  
 118 they can possibly be. At this stage of the reasoning it still looks hard to figure out how the finding  
 119 of an “only”  $H_0$ -consistent estimation method could solve our two-sample testing problem (3). The  
 120 next part of the intuition consists in looking at the asymptotic behavior of the stochastic process  
 121  $(a_n d_n(\hat{\theta}_n))_{n \geq 1}$ , for a well chosen increasing sequence of real numbers  $(a_n)_{n \geq 1}$  such that  $a_n \rightarrow +\infty$   
 122 as  $n \rightarrow +\infty$ . This way we could ideally expect a clear hypothesis separation coming from:

$$\begin{aligned} a_n d_n(\hat{\theta}_n) &= a_n [d_n(\hat{\theta}_n) - d(\theta^*)] \xrightarrow{\mathcal{L}} Z_0, \quad \text{under } H_0 \\ a_n d_n(\hat{\theta}_n) &= a_n [d_n(\hat{\theta}_n) - d(\theta^c)] + a_n d(\theta^c) \xrightarrow{a.s.} +\infty \quad \text{under } H_1, \end{aligned}$$

123 with  $Z_0$  an identified limiting random variable (which distribution could be at least tabulated).  
 124 Unfortunately since under  $H_1$  we could not probably get access to the limiting distribution of a  $Z_0$ -  
 125 reference distribution to build-up our test we could try to figure out a way to exhibit a distribution  
 126 under  $H_1$  that could be in the regime of a  $Z_0$  distribution (to decide which magnitude of deviation  
 127 from 0 could be considered as excessive or not in a test perspective). This is the second key idea  
 128 of our paper. By taking  $a_n = n$  and by making a closer analysis of  $nd_n(\widehat{\theta}_n)$  under  $H_1$  we observe  
 129 the following asymptotic behavior (see also the end of Appendix A):

$$\begin{aligned} nd_n(\widehat{\theta}_n) &= U_n^0 \xrightarrow{\mathcal{L}} Z(\theta^*, L_1, L_2), \quad \text{under } H_0 \\ nd_n(\widehat{\theta}_n) &= U_n^1 + V_n^1, \quad \text{with } U_n^1 \xrightarrow{\mathcal{L}} Z(\theta^c, L_1, L_2) \quad \text{and} \quad V_n^1 \xrightarrow{a.s.} +\infty, \quad \text{under } H_1, \end{aligned}$$

130 where the random variables  $Z(\theta^*, L_1, L_2)$  and  $Z(\theta^c, L_1, L_2)$ , corresponding to a parametrized closed  
 131 form stochastic integral, could be consistently sampled (and thus tabulated) under both  $H_0$  or  $H_1$   
 132 by generating  $Z(\widehat{\theta}_n, \widehat{L}_1, \widehat{L}_2)$ -type random variables. It is important to mention that these last  
 133 limiting random variables are strictly connected to the *inner convergence* phenomenon arising  
 134 either under  $H_0$  or  $H_1$ , as expressed in Theorem 2, see convergences (18-19), based on notation  
 135 (14).

136 Finally, by considering an empirical sample-based  $(1 - \alpha)$ -quantile of the stochastic integral  
 137  $Z(\widehat{\theta}_n, \widehat{L}_1, \widehat{L}_2)$ , denoted  $\widehat{q}_{1-\alpha}$ , we decide to consider the following  $H_0$ -rejection rule:

$$nd_n(\widehat{\theta}_n) \geq \widehat{q}_{1-\alpha} \quad \Rightarrow \quad H_0 \text{ is rejected.} \quad (10)$$

138 The above decision rule expresses the following principle: if the test statistic  $nd_n(\widehat{\theta}_n)$  is too far  
 139 from the *inner convergence* regime we could legitimately suspect a difference between  $F_1$  and  $F_2$ ,  
 140 as illustrated in the right side of Fig. 1, and then reject  $H_0$ .

141 Our paper is organized as follows: In Section 2 we analyze the model (2) identifiability, and  
 142 suggest an IBM-parameter picking principle relevant under both  $H_0$  and  $H_1$ . Section 3 is dedicated  
 143 to assumptions and asymptotic results showing the theoretical validity of our testing procedure  
 144 under the condition  $G_1 \neq G_2$ . In Section 3.2 we numerically validate the finite sample size  
 145 properties of the central limit theorem stated in Theorem 2. In section 4 we investigate the  
 146 empirical levels and powers of our test through Monte Carlo simulations. In Section 5 we present  
 147 an original application in which we compare pairwise the excess of mortality due to COVID-19  
 148 across a panel of European countries. Finally, Section 6 contains a discussion where we present two  
 149 further leads of research based on dependent two-sample models: i) we introduce the contaminant  
 150 distribution independence component testing along with the complete concordance/discordance  
 151 testing problem arising in  $z$ -scores modeling, ii) we introduce the homogeneity testing problem  
 152 in the so-called blending process (temporal contamination model). All the proofs and technical  
 153 material are relegated in Appendix sections A–E. Note that Appendix E is devoted to the non  
 154 identifiable situation where  $G_1 = G_2$ , in which the testing problem (3) can still surprisingly be  
 155 addressed by using a parametrization trick.

## 156 2 Identifiability under $G_1 \neq G_2$

157 Consider models (2) with generic proportions parameter  $\theta = (p_1, p_2) \in \Theta$  and denote by  $\theta^* =$   
 158  $(p_1^*, p_2^*) \in ]0, 1[^2$  the true proportions parameter value. By isolating the expressions of  $F_1$  and  $F_2$   
 159 under  $\theta$  we obtain for all  $x \in \mathbb{R}$ :

$$F_1(x, L_1, p_1) = \frac{L_1(x) - (1 - p_1)G_1(x)}{p_1} \quad \text{and} \quad F_2(x, L_2, p_2) = \frac{L_2(x) - (1 - p_2)G_2(x)}{p_2}. \quad (11)$$

160 Let us investigate now the situations where possibly  $F_1(x, p_1) = F_2(x, p_2)$  :

$$\begin{aligned} F_1(x, L_1, p_1) = F_2(x, L_2, p_2) &\Leftrightarrow \frac{L_1(x) - (1 - p_1)G_1(x)}{p_1} = \frac{L_2(x) - (1 - p_2)G_2(x)}{p_2} \\ &\Leftrightarrow \frac{(p_1 - p_1^*)G_1(x) + p_1^*F_1(x)}{p_1} = \frac{(p_2 - p_2^*)G_2(x) + p_2^*F_2(x)}{p_2} \\ &\Leftrightarrow \frac{p_1 - p_1^*}{p_1}G_1(x) = \frac{p_2 - p_2^*}{p_2}G_2(x) + \frac{p_2^*}{p_2}F_2(x) - \frac{p_1^*}{p_1}F_1(x). \end{aligned} \quad (12)$$

Under  $H_0$ ,  $F_1 = F_2 = F$ , we simply obtain

$$\frac{p_1 - p_1^*}{p_1}G_1(x) = \frac{p_2 - p_2^*}{p_2}G_2(x) + \left( \frac{p_2^*}{p_2} - \frac{p_1^*}{p_1} \right) F(x).$$

161 Hence, if  $G_1 \notin \text{span}(G_2, F)$ , which at least requires  $G_1 \neq G_2$  and frames our present study, we  
 162 necessarily have  $p_1 = p_1^*$  and  $p_2 = p_2^*$ . On the other hand, under  $H_1$ ,  $F_1 \neq F_2$ , if the cdfs family  
 163  $\{G_1, G_2, F_1, F_2\}$  is linearly independent, equation (12) is impossible since it would imply  $p_1^* = 0$   
 164 and  $p_2^* = 0$  which is in contradiction with  $\theta^* \in ]0, 1[^2$ , and therefore  $F_1(x, p_1) \neq F_2(x, p_2)$  for all  
 165  $\theta \in \Theta$ . That being said, in order to consistently pick the right  $\theta^*$  under  $H_0$  and select under  $H_1$  a  
 166  $\theta$  such that  $F_1(x, p_1) \neq F_2(x, p_2)$  (the property being actually true for all  $\theta \in \Theta$ ), we propose to  
 167 investigate the location of the minimum contrast parameter  $\theta^c$  :

$$\theta^c = \arg \min_{\theta \in \Theta} d(\theta), \quad \text{where} \quad d(\theta) = \int_{\mathbb{R}} D^2(x, L_1, L_2, \theta) dU(x), \quad (13)$$

168 with

$$D(x, L_1, L_2, \theta) = F_1(x, L_1, p_1) - F_2(x, L_2, p_2), \quad \text{and} \quad F_i(x, L_i, p_i) = \frac{L_i(x) - (1 - p_i)G_i(x)}{p_i}, \quad i = 1, 2, \quad (14)$$

169 where  $U$  is a continuous distribution which support encompasses the support of the  $L_i$ 's. For  
 170 simplicity, we will denote hereafter  $D(x, \theta) = D(x, L_1, L_2, \theta)$  and  $F_i(x, p_i) = F_i(x, L_i, p_i)$ ,  $i = 1, 2$ ,  
 171 except when the role of the  $L_i$ 's is central in our study.

## 172 3 Asymptotic results

173 To look at the proofs related to our theoretical results, the reader is referred to Appendix A. We  
 174 introduce here two assumptions connected to the identifiability and definite positiveness of the

175  $d$ -Hessian matrix.

176  
177 **(I)** The cdfs family  $\mathcal{E}_1 = \{G_1, F_1, G_2, F_2\}$  is linearly independent.

178  
179 **(II)** The family of functions  $\mathcal{E}_2 = \{G_1G_2, F_1F_2, G_i^2, F_i^2, G_iF_i; i = 1, 2\}$  is linearly independent.

180  
181 We consider in the above assumptions that, by convention, under  $H_0$  ( $F_1 = F_2 = F$ ) the cdfs  
182 family  $\mathcal{E}_1$  reduces to  $\{G_1, F, G_2\}$  and  $\mathcal{E}_2$  reduces to  $\{G_1G_2, F^2, G_i^2, G_iF; i = 1, 2\}$ . We assume now  
183 the following technical assumption:

184  
185 **(A)**  $\theta^*$  under  $H_0$  ( $\theta^c = \theta^*$ ), or  $\theta^c$  under  $H_1$  ( $\theta^c \neq \theta^*$ ), belongs to  $\overset{\circ}{\Theta}$  the interior of the compact  
186 parametric space  $\Theta$ .

187  
188 In order to consistently estimate  $\theta^*$  under  $H_0$  and make sure that under  $H_1$  ( $F_1 \neq F_2$ ) the  
189 functions  $F_1(\cdot, p_1)$  and  $F_2(\cdot, p_2)$  will also differ from each other for any value of  $\theta \in \Theta$ , we consider  
190 the following estimator:

$$\widehat{\theta}_n = \arg \min_{\theta \in \Theta} d_n(\theta), \quad (15)$$

191 where

$$d_n(\theta) = \int_{\mathbb{R}} D^2(x, \widehat{L}_1, \widehat{L}_2, \theta) dU(x), \quad \text{with} \quad \widehat{L}_i(x) = \frac{1}{n_i} \sum_{k=1}^{n_i} \mathbb{I}_{X_{i,k} \leq x} \quad i = 1, 2.$$

### 192 3.1 Theoretical results

193 In the sequel, we denote by  $\dot{\ell}(\vartheta)$  and  $\ddot{\ell}(\vartheta)$  the gradient vector and hessian matrix of any real  
194 function  $\ell$  (when it makes sense) with respect to argument  $\vartheta \in \mathbb{R}^2$ . The notation  $A^T$  refers to the  
195 transpose matrix of matrix  $A$ .

196 **Lemma 1.** (i) The mapping  $\theta \mapsto d(\theta)$  is  $\mathcal{C}^2$  over  $\Theta$  both under  $H_0$  or  $H_1$ .

197 (ii) Assume that conditions **(I)** and **(A)** hold. If  $U$  is strictly increasing on an interval  $I_U$  that  
198 encompasses the support of the  $L_i$ 's and  $G_i$ 's,  $i = 1, 2$ , then under  $H_0$ ,  $d$  is a contrast function,  
199 i.e. for all  $\theta \in \Theta$ ,  $d(\theta) \geq 0$  and  $d(\theta) = 0$  if and only if  $\theta = \theta^* \in \overset{\circ}{\Theta}$ .

200 (iii) Assume that conditions **(I)** and **(A)** hold. If  $U$  is strictly increasing on an interval  $I_U$  that  
201 encompasses the support of the  $L_i$ 's and  $G_i$ 's,  $i = 1, 2$ , and if for any given case under  $H_1$   
202 there exists one single point  $\theta^c \in \overset{\circ}{\Theta}$  such that  $\theta^c = \arg \min_{\theta \in \Theta} d(\theta)$ , then  $d(\theta^c) > 0$ .

203 (iv) We have under  $H_0$  or  $H_1$  that:

$$\sup_{\theta \in \Theta} |d_n(\theta) - d(\theta)| = o_{a.s.}(n^{-1/2+\alpha}), \quad \text{for all } \alpha > 0. \quad (16)$$



204 (v) Assume that conditions **(I)**, **(II)**, and **(A)** hold. Then under  $H_0$  we have:

$$\ddot{d}(\theta^*) = 2 \int_{\mathbb{R}} \dot{D}(x, \theta^*) \dot{D}^T(x, \theta^*) dU(x) > 0. \quad (17)$$

205 (vi) Assume that conditions **(I)**, **(II)**, and **(A)** hold. Then under  $H_1$  we have:

$$\ddot{d}(\theta^c) = 2 \int_{\mathbb{R}} \ddot{D}(x, \theta^c) D(x, \theta^c) + \dot{D}(x, \theta^c) \dot{D}(x, \theta^c)^T dU(x) > 0.$$

206 Let us denote  $\|\cdot\|_2$  the Euclidean distance in  $\mathbb{R}^2$ , and  $\theta^c = \theta^*$  if the assumption  $H_0$  is specified.

207 **Theorem 1.** If conditions **(I)**, **(II)** and **(A)** hold, we have under  $H_0$  or  $H_1$  that  $\|\widehat{\theta}_n - \theta^c\|_2 =$   
 208  $o_{a.s.}(n^{-1/4+\alpha})$  for all  $\alpha > 0$ .

209 **Theorem 2.** i) If conditions **(I)**, **(II)** and **(A)** hold, we have under  $H_0$  or  $H_1$ :

$$\sqrt{n} \begin{bmatrix} \widehat{p}_1 - p_1^c \\ \widehat{p}_2 - p_2^c \\ D_n(\cdot) - D(\cdot) \end{bmatrix} \xrightarrow{\mathcal{L}} W(\theta^c, \cdot), \quad \text{as } n \rightarrow +\infty, \quad (18)$$

210 where  $D_n(\cdot) = D(\cdot, \widehat{L}_1, \widehat{L}_2, \widehat{\theta}_n)$ , and  $W(\theta^c, \cdot) = (W_1(\theta^c), W_2(\theta^c), W_3(\theta^c, \cdot))^T$  is a centered 3-dimensional  
 211 Gaussian process with covariance matrix  $\Sigma_W = M(\theta^c, \cdot) \Sigma_L(\cdot, \cdot) M(\theta^c, \cdot)^T$  where  $M(\theta^c, \cdot)$  is defined  
 212 in (38) and  $\Sigma_L(\cdot, \cdot)$  in (41).

213 ii) If conditions **(I)**, **(II)** and **(A)** hold, we have respectively as  $n \rightarrow +\infty$ :  
 214

$$nd_n(\widehat{\theta}_n) = \mathbf{U}_n^0 \xrightarrow{\mathcal{L}} Z(\theta^*) = \int_{\mathbb{R}} (W_3(\theta^*, x))^2 dU(x), \quad \text{under } H_0, \quad (19)$$

$$nd_n(\widehat{\theta}_n) = \mathbf{U}_n^1 + \mathbf{V}_n^1, \quad \text{with } \mathbf{U}_n^1 \xrightarrow{\mathcal{L}} Z(\theta^c) = \int_{\mathbb{R}} (W_3(\theta^c, x))^2 dU(x) \text{ and } \mathbf{V}_n^1 = O_{a.s.}(n), \text{ under } H_1 \quad (20)$$

215 Let us remind that the above stochastic integrals distribution can be simulated by standard  
 216 Monte Carlo methods, see for instance [13], and thus fully tabulated. As detailed in the Introduction  
 217 the use of the above theorem in our testing perspective consists in rejecting  $H_0$  if the statistic  
 218  $nd_n(\widehat{\theta}_n)$  exceeds  $\widehat{q}_{1-\alpha}$ , where  $\widehat{q}_{1-\alpha}$  is the approximated  $(1 - \alpha)$ -quantile of the limiting random  
 219 variable  $\int_{\mathbb{R}} (W_3(\theta^c, x))^2 dU(x)$  (given that by convention  $\theta^c = \theta^*$  under  $H_0$ ). We propose, in order  
 220 to prevent from obvious not competing situations, to check if a  $\theta^c$   $(1 - \alpha)$ -domain of confidence,  
 221 denoted  $\mathcal{I}_{1-\alpha}(\theta^c)$ , intersects somehow the  $]0, 1[$  proportions domain. Let us consider

$$\mathcal{I}_{1-\alpha}(\theta^c) = I_{1-\alpha/2}(p_1^c) \times I_{1-\alpha/2}(p_2^c), \quad \text{where } I_{1-\alpha/2}(p_i^c) = \left[ I_{i,1-\alpha/2}^-, I_{i,1-\alpha/2}^+ \right], \quad (21)$$

$$\text{with } I_{i,1-\alpha/2}^- = \widehat{p}_i - \frac{\sqrt{\Sigma_{W(\widehat{\theta}_n)}[i, i]}}{\sqrt{n}} \phi\left(1 - \frac{\alpha}{4}\right) \quad \text{and} \quad I_{i,1-\alpha/2}^+ = \widehat{p}_i + \frac{\sqrt{\Sigma_{W(\widehat{\theta}_n)}[i, i]}}{\sqrt{n}} \phi\left(1 - \frac{\alpha}{4}\right),$$

222 where  $\phi(\cdot)$  denotes the quantile function of the  $\mathcal{N}(0, 1)$  distribution, and  $\Sigma[i, i]$  the  $i$ -th diagonal  
 223 term of matrix  $\Sigma$ . In the sequel we denote either  $\bar{A}$  the complimentary of a generic probability  
 224 event  $A$  and  $\bar{I}$  the complimentary of a generic domain  $I$  of  $\mathbb{R}^d$ ,  $d \leq 2$ . Noticing that according to  
 225 Theorem 2 we have  $P(I_{1-\alpha/2}(p_i^c)) \approx 1 - \alpha/2$  as  $n \rightarrow +\infty$ ,  $i = 1, 2$ , we can write

$$\begin{aligned} P(\overline{\theta^c \in \mathcal{I}_{1-\alpha}(\theta^c)}) &= P(\theta^c \in \overline{\mathcal{I}_{1-\alpha}(\theta^c)}) \\ &\leq P(\overline{p_1^c \in I_{1-\alpha/2}(p_1^c)}) + P(\overline{p_2^c \in I_{1-\alpha/2}(p_2^c)}) \\ &= P(p_1^c \in \overline{I_{1-\alpha/2}(p_1^c)}) + P(p_2^c \in \overline{I_{1-\alpha/2}(p_2^c)}) \\ &\approx \alpha, \end{aligned}$$

226 which leads to

$$P(\theta^c \in \mathcal{I}_{1-\alpha}(\theta^c)) \geq 1 - \alpha \quad \text{approximately as } n \rightarrow +\infty. \quad (22)$$

227 Finally a simple “green light” criterion to proceed to the test could be the checking of the condition:

$$I_{i,1-\alpha/2}^- < 1, \quad i = 1, 2 \quad (\text{green light testing criterion}). \quad (23)$$

228 **Remark 3.** *Although the same underlying idea can be used, the case where  $G_1 = G_2$  is slightly*  
 229 *different and requires a “picking trick” as the parameters are no longer identifiable even under  $H_0$*   
 230 *(see Appendix E for further details).*

### 231 3.2 Convergence Monte Carlo assessment

232 Introduce the random vector  $(P_1, P_2, D_z)^T = \sqrt{n} (\hat{p}_1 - p_1^c, \hat{p}_2 - p_2^c, D_n(z) - D(z))^T$  at any point  $z \in$   
 233  $\text{support}(X_1, X_2)$ . Recall that  $z$  represents one location point of the empirical process trajectory.  
 234 Theorem 2 states that this vector is asymptotically Gaussian, and that its first two components  
 235 are consistent towards  $\theta^c$  (both under  $H_0$  or  $H_1$ ). Our goal is to check this by comparing numerical  
 236 approximations of our theoretical expressions to Monte Carlo experiments. To this aim, we consider  
 237  $K = 200$  simulations of two samples  $X_1$  and  $X_2$  with cdfs given by (2), both following two-  
 238 component mixtures of Gaussian distributions. More precisely, the  $k$ -th simulation provides  $X_1^k$   
 239 and  $X_2^k$  ( $k = 1, \dots, K$ ), where  $X_1^k$  and  $X_2^k$  are respectively drawn from mixtures with parameters  
 240  $n_1 = n_2 = 5,000$ ,  $p_1^* = 0.4$ ,  $p_2^* = 0.6$ ,  $F_1 = F_2$  are  $\mathcal{N}(1, 1)$  cdfs, when  $G_1, G_2$  are respectively  
 241  $\mathcal{N}(2, 0.7)$  and  $\mathcal{N}(3, 1.2)$  cdfs. Note that we are here under the null, but remember that such  
 242 comparisons were also made on very different setups involving  $H_1$ -type frameworks, with  $n_1 \neq n_2$   
 243 and distributions supported over  $\mathbb{R}^+$ ,  $\mathbb{N}$  or bounded intervals of  $\mathbb{R}$ .

244 Estimating  $\theta = (p_1, p_2)$  by (15) from each of the  $K$  simulated couples  $(X_1, X_2)$ , we obtain  
 245  $(0.402, 0.601)$  as empirical mean of  $\hat{\theta}_n = (\hat{p}_1, \hat{p}_2)$ , illustrating the asymptotic consistency of our  
 246 estimators. Kolmogorov-Smirnov tests on the components of the vector  $(P_1, P_2, D_z)^T$  validate  
 247 that the three estimators are asymptotically Gaussian, with  $p$ -values always greater than 0.7. To  
 248 validate the explicit covariance structure between the estimators, it is necessary to fix  $z$  and to  
 249 compare the empirical covariances (computed from the Monte Carlo simulations) to the theoretical  
 250 ones. Appendix C shows the obtained results in the aforementioned parametric setup for  $z = 2$ .

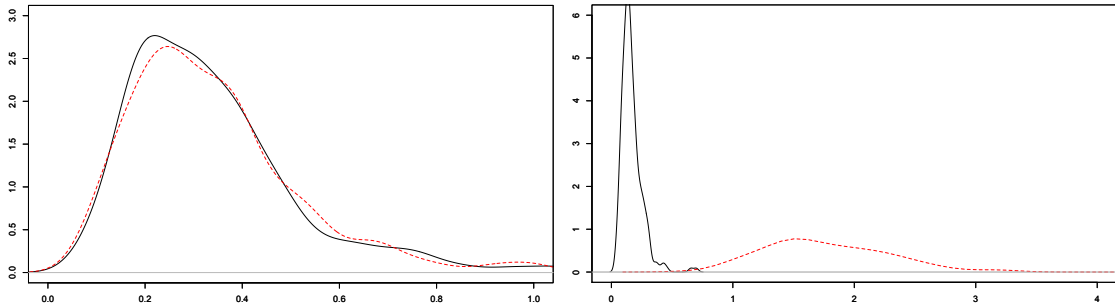


Figure 1: On the left panel, theoretical distribution  $Z(\theta^*)$  (solid) and empirical version  $U_n^0$  (dotted). On the right, distribution  $Z(\theta^c)$  (solid) and empirical contrast distribution  $U_n^1 + V_n^1$  (dotted).

251 Clearly, all the tests made through these comparisons for different values of  $z$  show the validity of  
 252 formulas (38)-(41), which confirms the theoretical consistency given in part i) of Theorem 2.

253 It now remains to have a closer look at the behaviour of the statistic  $nd_n(\hat{\theta}_n)$ , see formulas (19)  
 254 and (20). The theorem states that the empirical distribution of  $U_n^0$  under  $H_0$  (obtained through the  
 255 Monte Carlo procedure providing as many realizations of  $nd_n(\hat{\theta}_n)$  as the  $K$  experiments) should  
 256 converge to some explicit random variable  $Z(\theta^*)$ . Also, the same kind of regime for  $U_n^1$  should be  
 257 observed under  $H_1$ . However, in the latter case, the discrepancy measure dramatically increases  
 258 due to the term  $V_n^1$ , exhibiting the departure from the null hypothesis. This phenomenon is  
 259 well illustrated by Figure 1, where one can see that the empirical distributions suit the expected  
 260 behaviours provided by the random variables  $Z(\theta^*)$  and  $Z(\theta^c)$ . Indeed, under  $H_1$ , the empirical  
 261 distribution of  $nd_n(\hat{\theta}_n)$  is far from  $Z(\theta^c)$ , showing the impact of the drift  $V_n^1$ . This way, the  
 262 tabulated distributions  $Z(\theta^*)$  and  $Z(\theta^c)$  and their appropriate  $(1 - \alpha)$ -quantile can be used to  
 263 fruitfully answer our testing problem.

## 264 4 Test performance

265 In this section, we study the empirical levels and powers of the test in various situations. To  
 266 this aim, we generate  $X_1$  and  $X_2$  from (2) on various supports, and the behaviour of the test  
 267 is investigated in more or less challenging setups. Depending on the case under study, mixture  
 268 components can be easily detected or not, either because of the importance of the mixture weight  
 269  $p_i$ ,  $i = 1, 2$ , or due to the specified mixture components features. The idea is to get some insights  
 270 about the strengths and weaknesses of our testing procedure. Each time we evaluate the empirical  
 271 level (respectively power) of the test, the 95-th percentile of the test was previously assessed  
 272 using 150 trajectories in the computation of the stochastic integral appearing on the right-hand  
 273 side of (19) and (20). Then the testing procedure (10) is performed  $K$  times to get the result,  
 274 through the  $K$  simulations of the  $k$ -th samples  $X_1^k$  and  $X_2^k$  and the associated test statistic  $nd_n(\hat{\theta}_n^k)$   
 275 ( $k = 1, \dots, K$ ). Here, we take  $K = 100$ , fix similar sample sizes  $n_1 = n_2 = n$  for conciseness, and  
 276 make  $n$  vary.

277 **4.1 Empirical levels ( $F_1 = F_2 = F$ )**

278 In terms of distributions depending on the support, we consider Gaussian-Gaussian mixtures on  
 279  $\mathbb{R}$ , Gamma-Exponential ones on  $\mathbb{R}^+$ , Negative-Binomial-Poisson on  $\mathbb{N}$ , and Logit-Uniform on  $[0, 1]$ .  
 280 To check the significance of the test in real-life situations, we have chosen to make the component  
 281 weights  $(p_i)_{i=1,2}$  vary from 10% to 60%. The asymptotic properties of the test can be checked  
 282 by considering different values for the number  $n$  of observations (ranging from 500 to 10,000).  
 283 However, our experiments show that the number of observations does not have a big impact on  
 284 the level of the test, provided that there are at least around 300 observations for the mixture  
 285 component to test. That is why we have chosen to present here only the results corresponding to  
 286  $n = 2,000$  observations, which lightens our results presentation.

287 For each support ( $\mathbb{R}, \mathbb{R}^+, \mathbb{N}, [0, 1]$ ), four very different frameworks are studied (see Figure 9 in  
 288 Appendix F.1, with corresponding mixture parameters stored in Table 2). We will denote from  
 289 (a) to (d) these four different cases, corresponding to: (a)  $G_1$  not so different from  $G_2$ , and  $G_1$   
 290 and  $G_2$  close to  $F$ ; (b)  $G_1$  very different from  $G_2$ , with both distributions close to  $F$ ; (c)  $G_1$  not  
 291 so different from  $G_2$ , with both distributions far from  $F$ ; (d)  $G_1$  very different from  $G_2$ , with  
 292  $G_1$  close to  $F$  and  $G_2$  far from  $F$ . The global simulation scheme thus encompasses 144 different  
 293 setups in all (4 supports, 4 cases, and 9 combinations for  $p_1$  and  $p_2$ ). We remind that for each of  
 294 these 144 possibilities, the testing procedure (10) is performed 100 times, which leads to give an  
 295 approximation of the empirical level of the test in all of the aforementioned situations.

296 The overall results are summarized thanks to the heatmap in Figure 2, with dark zones indi-

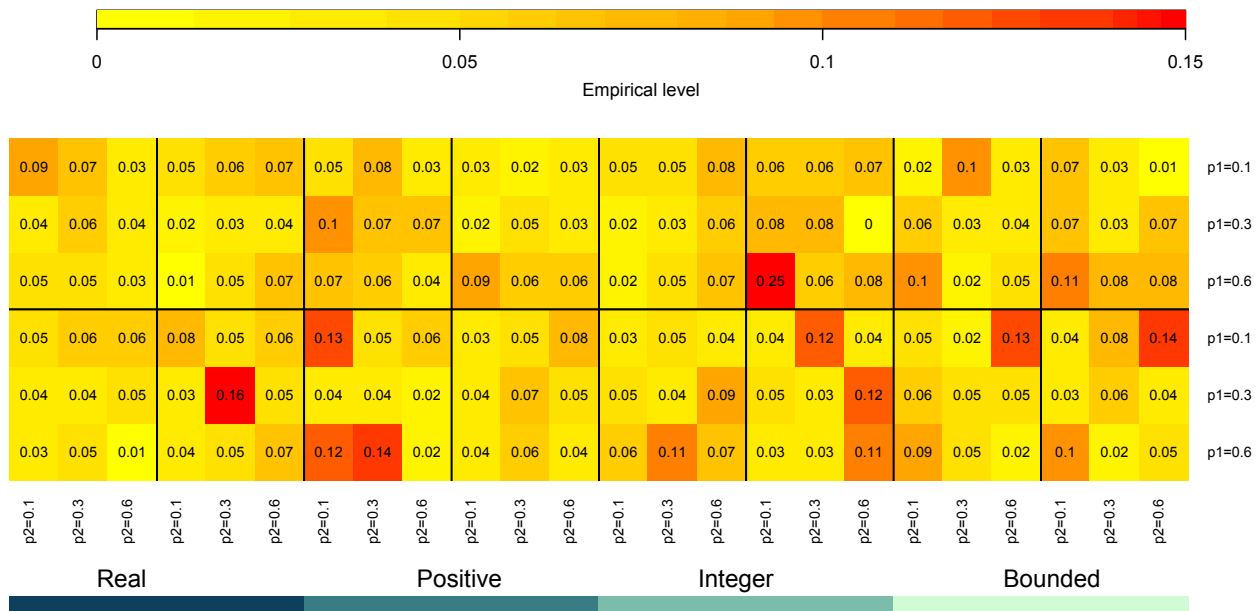


Figure 2: Heatmap of empirical level (under  $H_0$ ) for different supports, different components weights, and different parameters for component distributions. For each support, cases (a) to (d) are given from top left to bottom right.

297 cating unsatisfactory results. For one given support, the four panels from top left to bottom right  
 298 correspond to cases (a) to (d). For instance, case (c) with mixtures of Gaussian distributions is  
 299 the bottom left 3x3 square of the heatmap. One can see that most of the setups under study  
 300 lead to satisfactory empirical levels of the test, close to the theoretical 5%. Indeed, since each  
 301 simulation enables to compare the empirical test statistic to the 95th percentile of the calibrated  
 302 distribution  $U^0$ , it is expected that the level of the test fluctuates around 5%. In practice only  
 303 12 over 144 approximations of the level exceed 10%, which means that less than 9% of the setups  
 304 under study provide mixed conclusions. Looking more carefully at the results, the problematical  
 305 situations mostly arise when at least one of the proportions  $p_i$  equals 10%. It is very likely that  
 306 the main reason explaining this drop of efficiency is the lack of observations to perform the test  
 307 about the unknown components. The low component weight affected to the unknown part of the  
 308 distribution leads to underrepresent the observations useful for the test to be efficient. In some  
 309 very rare frameworks, although  $p_1$  and  $p_2$  equal at least 30%, the empirical level remains “high”  
 310 (e.g. the case of mixing Negative Binomial and Poisson distributions, case (d), with  $p_1 = 0.3$  and  
 311  $p_2 = 0.6$ , where the empirical level equals 12%). In such cases, the choice of the mixture compo-  
 312 nents parameters (Table 2 in Appendix F.1) has a crucial impact and can affect the estimation of  
 313 the component weights, which spreads to the overall quality of the test.

## 314 4.2 Empirical powers

315 In the same spirit, one can analyse the heatmap that illustrates the empirical power of the test in  
 316 Figure 3, still considering the same mixture distributions as previously (Gaussian-Gaussian on  $\mathbb{R}$ ,  
 317 Gamma-Exponential on  $\mathbb{R}^+$ , Negative-Binomial-Poisson on  $\mathbb{N}$ , and Logit-Uniform on  $[0, 1]$ ).

318 However, the difference lies in the different frameworks studied, illustrated by Figure 10 in  
 319 Appendix F.2. Hereafter, we denote them as follows: case (a)  $F_1$  and  $F_2$  have the same distribution,  
 320 with very different means; case (b)  $F_1$  and  $F_2$  have the same distribution, with close means; case  
 321 (c)  $F_1$  and  $F_2$  have the same distribution, with same means but very different variances; case (d)  
 322  $F_1$  and  $F_2$  have the same distribution, with same means and close variances. We obviously expect  
 323 here that the most difficult case to detect is the latter one. Here, the sample size has a major  
 324 impact on the results, which explains why the heatmap is provided for results corresponding to a  
 325 sensitively higher sample size  $n = 3,000$ . To understand how crucial the number of observations  
 326 is, Figure 4 depicts the connection between the empirical power of the test and  $n$ . In fact we  
 327 can observe very heterogeneous behaviours depending on the support and component weights,  
 328 especially in the case where alternatives are very difficult to distinguish, that is, in the case where  
 329  $F_1$  and  $F_2$  have the same two first order moments (see case (e) in Table 3 of Appendix F.2 for  
 330 further details about mixture distributions and parameters). Not surprisingly, departures from  
 331 the null hypothesis can be detected provided that the number of observations is large enough,  
 332 otherwise the power of the test remains low (especially when  $F_1$  and  $F_2$  are very similar, see cases  
 333 (b) and (d)). Indeed, low proportions  $p_i$  ( $i = 1, 2$ ) leads to deteriorate the accuracy of the estimates  
 334  $\hat{p}_i$ , which favour situations where  $\hat{\theta}_n$  can be “far” from  $\theta^c$  (minimization of the contrast is solved  
 335 by escaping from  $]0, 1[^2$ ). The natural consequence of this phenomenon is that extreme quantiles  
 336 (e.g. 95th percentile) of the tabulated random variable  $U^1$  tend to be larger, which mechanically  
 337 lowers the power of the test.

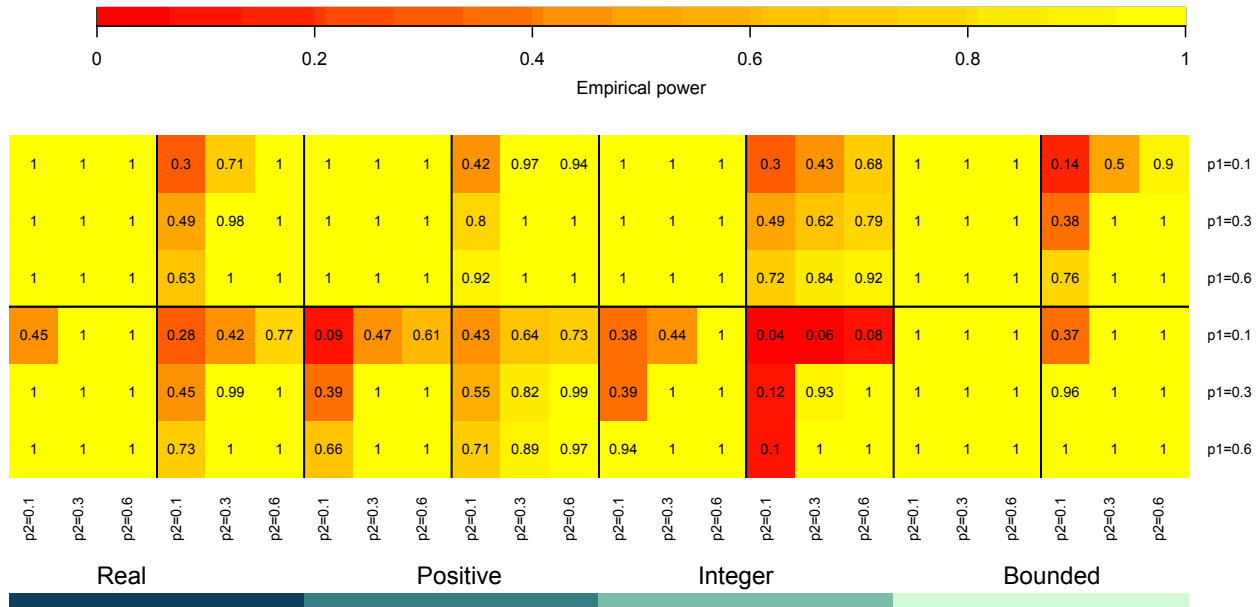


Figure 3: Heatmap of empirical power (under  $H_1$ ) for different supports, different components weights, and different parameters for component distributions. For each support, cases (a) to (d) are given from top left to bottom right.

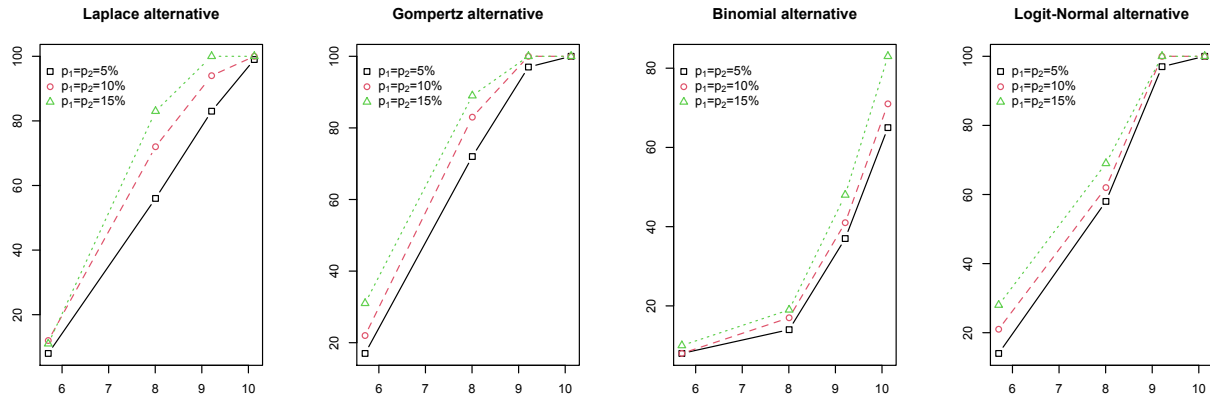


Figure 4: Empirical power depending on the sample size ( $n = 300; 3,000; 10,000; 25,000$  in logarithmic scale), on various supports ( $\mathbb{R}, \mathbb{R}^+, \mathbb{N}, [0, 1]$ ), when  $F_1$  and  $F_2$  have same mean and variance (other parameters are listed in Table 3 of Appendix F.2, case (e), see also Fig. 11).

## 5 Application to COVID-19 excess mortality

339 There is an abundant literature investigating the impact of the 2019 coronavirus disease (COVID-  
 340 19) on the mortality across countries, see for instance Beaney *et al.* [1]. We generally witness a  
 341 wide variation in mortality across countries, leading to questioning the extent to which one can  
 342 proceed to pairwise comparative studies. In our application, we will be looking at the nodular

343 impact of the COVID-19 and compare the latter across a panel of European countries. Formally,  
 344 we investigate the age distribution of deaths (the distribution of the proportion of deaths per age  
 345 group among all deaths) and we study the changes between 2019 and 2020 for France, Belgium,  
 346 Germany, Italy, Netherlands and Spain from the Short-Term Mortality Fluctuations (STMF) data  
 347 series compiled by the Human Mortality Database (HMD). The datasets contain death records  
 348 aggregated over age groups: 0-14, 15-64, 65-74, 75-85 and 85+. We restrain our study to the first  
 349 25 weeks (and ages over 15 years-old) of each considered year as shown in Figure 5. Figure 6  
 350 shows the distribution of the proportion of deaths per age class for years 2019 and 2020 (total of  
 351 proportions equals to 1), indicating the empirical probability for a death to be in each age class.

352 It is assumed that the differences in the observed mortality between 2019 and 2020 is imputed  
 353 (directly or indirectly) to the COVID-19. The 2020 population is then a two-component mixture  
 354 composed by the previous 2019 population plus a latent population subject to the impact of the  
 355 COVID-19 crisis. In other words, model (1) has an appealing application to capture the excess of  
 356 mortality due the COVID-19. It is then legitimate to assume a second unknown nodular component  
 357 driving the mortality due to the COVID-19 during the considered period. More precisely, we will  
 358 assume that the known cdf is the one observed over 2019, *i.e.* the multinomial distribution  $G$ , and  
 359 look to compare the distribution  $F$  of the excess mortality across countries. This excess mortality  
 360 can be regarded as a measure that encompasses all causes of death and provides a metric of the  
 361 overall mortality impact in 2020.

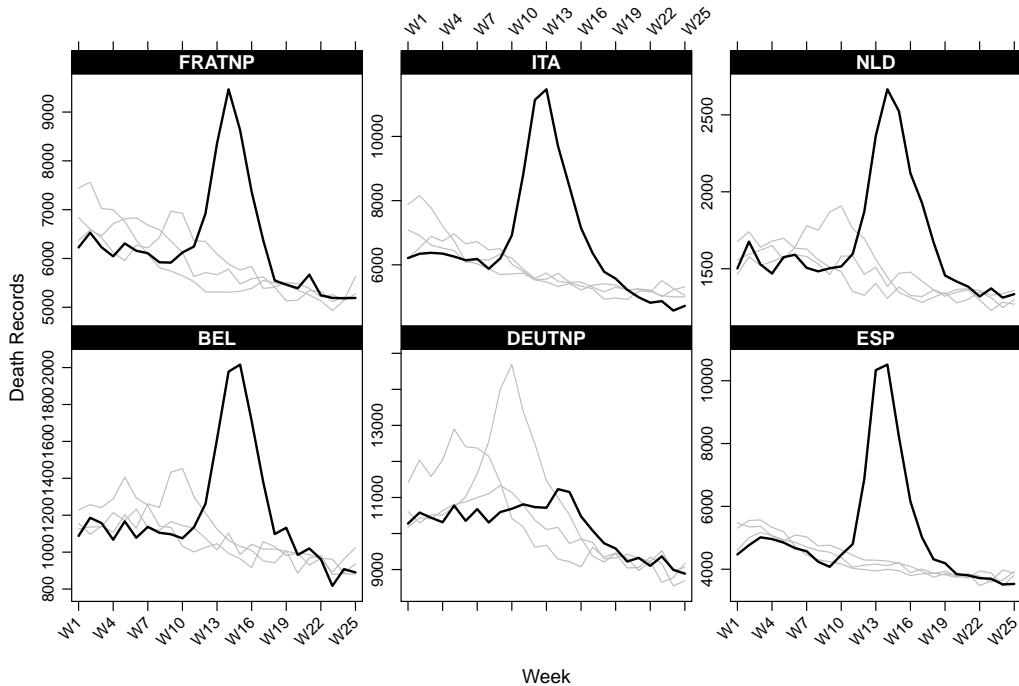


Figure 5: Total death records of individuals aged over 15 years-old across years 2017, 2018, 2019 (gray curves) and 2020 (black curve).

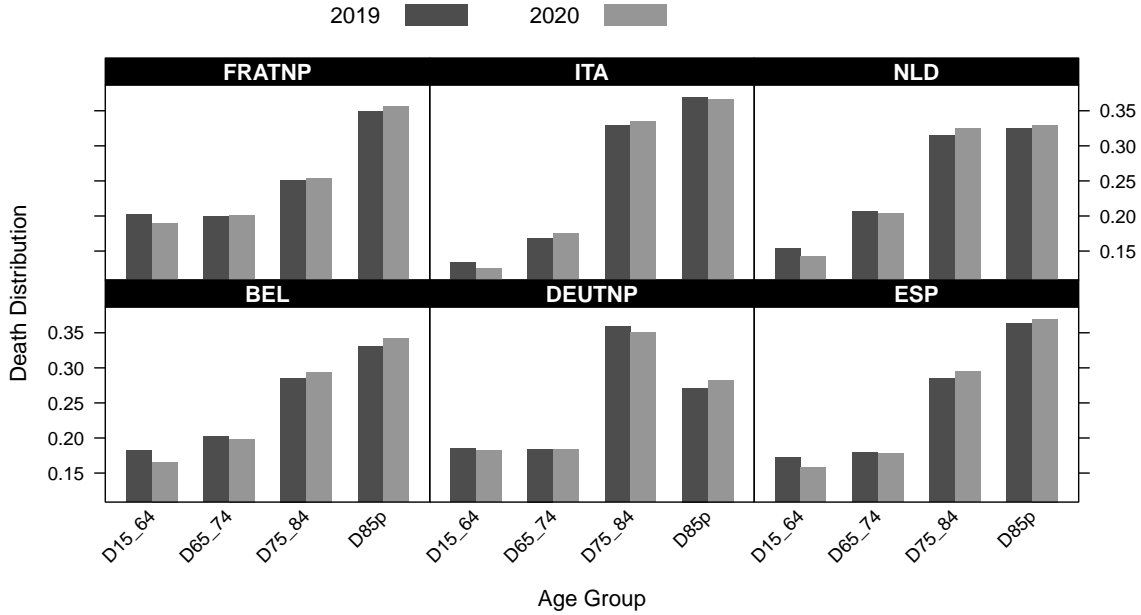


Figure 6: Distribution of the proportion of deaths per age group among all deaths (2019 and 2020).

362 In Table 1 we report the outputs of the testing procedure developed in this paper for the  
 363 aforementioned countries. The known component  $G_i$  is described as the multinomial distribution  
 364 computed in 2019 for each country. We shall stress out that in this application we are clearly in  
 365 presence of two distinct known cdfs, i.e.  $G_1 \neq G_2$ , which is our basic assumption to implement  
 366 our procedure. Also, in this case, we choose the discrete uniform distribution for the integrating  
 367 cdf  $U$ , see (13). In order to avoid any departure of the estimated proportion  $p_i$  to the infinity  
 368 and thus make the discrepancy measure go to 0, we bounded the parametric space over which we  
 369 optimize. Here, we set the upper bound equal to  $\delta_2 = 3$ , which seems to be large enough given the

Population		$p_1$	$p_2$	Green light	Test statistic	95% quantile	$p$ -value	Decision
1	2							
Spain	Italy	0.1404	0.3058		1.3471	14.6930	0.75	$H_0$
Spain	France	3	0.7388		-	-	-	$H_1$
Spain	Germany	3	0.0997		-	-	-	$H_1$
Spain	Netherlands	0.6701	3		-	-	-	$H_1$
Spain	Belgium	3	3		-	-	-	$H_1$
Netherlands	Italy	0.0722	0.1523		20.3887	90.3917	0.65	$H_0$
Netherlands	France	3	0.3031		-	-	-	$H_1$
Netherlands	Germany	0.94	0.1638		14.9862	2.8878	-	$H_1$
Netherlands	Belgium	3	0.6364		-	-	-	$H_1$
Italy	France	0.32	3		-	-	-	$H_1$
Italy	Germany	3	0.1710		-	-	-	$H_1$
Italy	Belgium	3	3		-	-	-	$H_1$
Belgium	France	3	3		-	-	-	$H_1$
Belgium	Germany	0.5742	0.1908354		1.7297	1.344	-	$H_1$

Table 1: Pairwise testing of the excess of mortality behavior across some European countries.



370 simulation results. First, we can see from Table 1 that some estimated proportions in the pairwise  
 371 analysis bump into this boundary, which clearly indicates to reject the null hypothesis  $H_0$ . Among  
 372 these countries, we see that only Spain and Italy (with  $p$ -value equal to 76%), on one hand, and  
 373 Netherlands and Italy (56%), on the other hand, shares the same excess mortality profile. However,  
 374 the equality between Spain and Netherlands is rejected. This result can be interpreted as follows:  
 375 the estimation procedure tends to find a common cdf, that is  $F_1 = F_2 = F$ , under the null. In  
 376 the Spain-Italy comparison case, this common component represents 30.58% of the Italian deaths  
 377 population and 14.04% of the Spanish population's deaths. Their representations are given in  
 378 Figure 7 where we can see the very close patterns of these two multinomial distributions. In the  
 379 Netherlands-Italy comparison case, the common component is slightly different. It represents only  
 380 15.23% of the Italian deaths population and 7.22% of the Netherlands one. Their estimations are  
 381 given in Figure 7 where it can be seen a slight difference between the two last classes which are less  
 382 numerous and therefore more sensible with a larger variability. We also have to take into account  
 383 that the Netherlands second component estimation is based on only 7.22% of the observations.  
 384 In conclusion, the excess mortality Italian profile seems to be very similar to the Spain one with  
 385 a very large  $p$ -value. A part of this excess mortality seems to be similar to the Netherlands one,  
 386 with a lower  $p$ -value. But our test procedure do not retain the equality between Netherlands and  
 387 Spain excess mortalities.

388 Eventually, the proportion for Spain and Italy approximate respectively 30.58% and 14.04%,  
 389 which is consistent with the reported statistics [1, 18]. The discussion of such a behaviour is,  
 390 however, beyond the scope of this paper. Instead, we can refer to the various discussions in  
 391 the literature that intended to understand the differential impact of the COVID-19 crisis over  
 392 countries looking at the socio-economic and demographic variables. In our case, we depict the  
 393 mortality during this first wave of the pandemic for countries validating the null hypothesis  $H_0$   
 394 (Spain, Italy and Netherlands), see the left panel of Figure 7. We see that these countries exhibit  
 395 a comparable behavior but which cannot be visually validated.

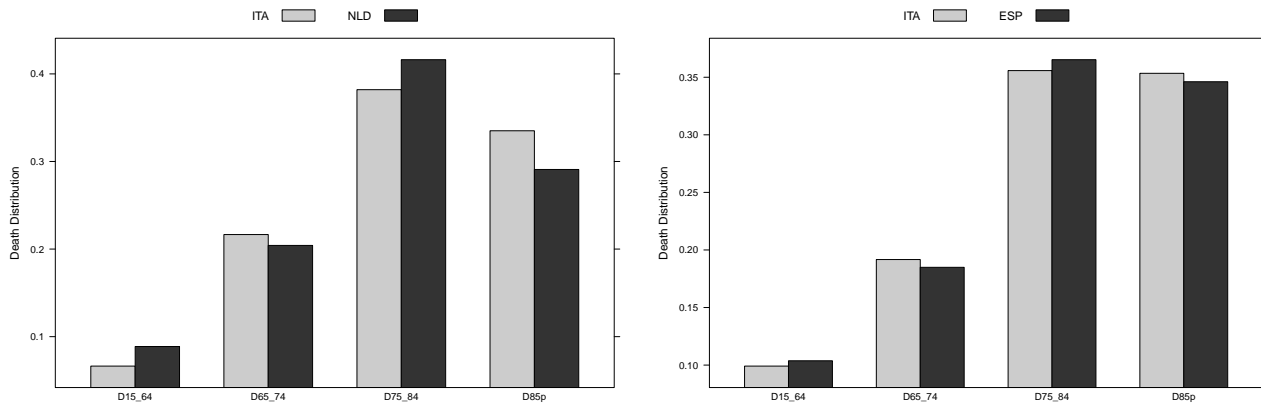


Figure 7: The nodular distribution  $F_i$  due to the COVID-19 crisis for the countries with similar impacts on the mortality : Netherlands-Italy (left) and Spain-Italy (right)

## 396 6 IBM-method and further models

397 In the next two sections we propose to highlight on the range of our method, to describe challeng-  
 398 ing situations (involving dependencies) in which our semiparametric IBM-method could provide  
 399 interesting results. These are ongoing works which are beyond the scope of the current paper.

### 400 6.1 Independence, concordance and discordance

401 Let us consider for simplicity a bivariate contamination model (extension to the  $d$ -variate setup,  
 402  $d \geq 3$ , being straightforward):

$$L(x_1, x_2) = pG(x_1, x_2) + (1 - p)F(x_1, x_2), \quad (x_1, x_2) \in \mathbb{R}^2, \quad (24)$$

403 where  $L$  is the common cdf of an i.i.d. sample  $(X_1, \dots, X_n)$ ,  $G$  is a known cdf when the mixture  
 404 proportion  $p$  and the cdf  $F$  are both unknown. By splitting the observation vector  $X$  into 2  
 405 components  $X = (X_1, X_2)^T$ , we have respective marginal cdfs

$$L_i(x) = pG_i(x) + (1 - p)F_i(x), \quad x \in \mathbb{R}, \quad i = 1, 2. \quad (25)$$

406 An interesting problem is then to test the mutual independence of the nodular components  $X_1$   
 407 and  $X_2$ , *i.e.*

$$H_0 : F = F_1 \otimes F_2 \quad \text{against} \quad H_0 : F \neq F_1 \otimes F_2, \quad (26)$$

408 where  $G \neq G_1 \otimes G_2$  on a  $\mu$ -non null set to avoid trivial testing situations (otherwise independence  
 409 on the  $L$ -components would then reflect the independence on the  $F$ -components). Given the above  
 410 remarks we can define two parametric families (Inversion step):

$$\begin{aligned} \mathcal{F}_1 &= \left\{ F(u_1, u_2; p) = \frac{L(u_1, u_2) - pG(u_1, u_2)}{1 - p}, \quad p \in ]0, 1[ \right\}, \quad \text{and} \\ \mathcal{F}_2 &= \left\{ F_{1 \times 2}(u_1, u_2; p) = F_1(u_1; p)F_2(u_2; p), F_i(\cdot; p) = \frac{L_i(\cdot) - pG_i(\cdot)}{1 - p}, \quad i = 1, 2, \quad p \in ]0, 1[ \right\}, \end{aligned}$$

411 and build a contrast function (Best Matching step) in the spirit of (13–14)

$$d(p) = \int_{\mathbb{R} \times \mathbb{R}} (F(u_1, u_2; p) - F_{1 \times 2}(u_1, u_2; p))^2 dU(u_1, u_2).$$

412 Using copula techniques to handle global and marginal empirical processes, as it is classically done  
 413 in the “direct” (not mixture component testing) Cramér-von Mises independence testing literature,  
 414 see Genest et al. [10] for recent results and bibliography, we reasonably think that asymptotic  
 415 decision results similar to ii) in Theorem 2 could be established on the test statistic  $nd_n(\hat{p}_n)$  where  
 416  $d_n$  is the empirical version of  $d$  and  $\hat{p}_n$  is the minimum argument of  $d_n$  over  $]0, 1[$ . Note that  
 417 such accomplishment would also help in answering/testing the complete concordance/discordance

418 problem arising in  $z$ -score analysis, see Lai *et al.* [15] and Lai *et al.* [16], where basically model  
 419 (24) can take (among others) two basic forms:

$$\begin{aligned} L &= pG_1(x_1) \otimes G_2(x_2) + (1-p)F_1 \otimes F_2, & (\text{complete concordance}), \\ L &= (p_1G_1 + (1-p_1)F_1) \otimes (p_2G_2 + (1-p_2)F_2), & (\text{complete discordance}). \end{aligned}$$

420 In fact in the above models, slightly more complex contrast functions  $d$ , based on  $F$ -inversions and  
 421 comparison inspired from the previous independence testing strategy, can also be proposed and  
 422 proved to provide a fully tractable Cramér-von Mises test in the spirit of Theorem 2.

## 423 6.2 Blending process

424 As mentioned earlier, the testing methodology we introduce in this paper can be extended to  
 425 temporal contamination models we propose to name blending process. This type of model are  
 426 especially interesting to analyze situations in which a phenomenon has been observed with a  
 427 good stability for a long period of time but turns out to be contaminated by a new trend which  
 428 importance becomes more and more prominent. This type of model would be especially relevant  
 429 to analyze temporal mortality datasets during the COVID-19 crisis as described in Section 5  
 430 (collections of mortality datasets over time would be required instead of one single sample collected  
 431 during a given period of time). By denoting  $G$  the cdf of the well known phenomenon and by  $p_t$ ,  
 432 respectively  $F_t$ , the proportion, resp. the cdf, of the new trend at time  $t$ , the distribution of a  
 433 generic i.i.d sample  $X^t = (X_1^t, \dots, X_{n_t}^t)$  at time  $t \in \mathbb{N}$  could be expressed as follows:

$$L_t(x) = p(t)G(x) + (1-p(t))F_t(x), \quad x \in \mathbb{R}. \quad (27)$$

434 In that setup it could be interesting, following the identifiability and parameter picking strategy  
 435 presented in Section E when  $G_1 = G_2$ , to test the consistency in time of the trend distribution,  
 436 *i.e.*

$$H_0(t_i, t_j) : F_{t_i} = F_{t_j} \quad \text{against} \quad H_1(t_i, t_j) : F_{t_i} \neq F_{t_j}, \quad i \neq j \in \{1, \dots, T\}. \quad (28)$$

437 Note that if the testing problem (28) is very mostly answered positively we could possibly assess  
 438 that  $F$  is independent from  $t$  and then estimate nonparametrically the mixing proportion function  
 439  $p(t)$  based on the condition  $F_t = F$ ,  $t \in \mathbb{N}$  and Remark 5 of Section E. The main technical  
 440 difficulty here is to handle correctly the possible dependencies between samples  $X^{t_i}$  and  $X^{t_j}$ ,  
 441 for  $i \neq j$ , especially when  $t_i$  and  $t_j$  are close. Note that the analog of (59) and (66) need to  
 442 established with a limiting bivariate Gaussian process  $\mathcal{B}$  with no longer independent coordinates  
 443 since the source samples  $X^{t_i}$  and  $X^{t_j}$  are dependent. The paper by Gribovka and Lopez [12] on  
 444 non-parametric copula estimation under bivariate censoring looks to provide interesting ideas to  
 445 solve this problem.

## 446 7 Conclusion

447 In this work, we address the comparison testing of the unknown components of a two sample  
448 contamination model. We introduce the so-called IBM (Inversion-Best Matching) approach that  
449 results into a relaxed semi-parametric Cramér-von Mises type two-sample test with very minimal  
450 assumptions about the unknown components. Indeed, we do not require any shape constraints on  
451 the unknown distributions, such as symmetry, tail conditions etc. which are very often key technical  
452 identifiability conditions arising in univariate semiparametric mixture models. We establish in  
453 particular a functional joint central limit theorem on the proportion parameters (with consistency  
454 under  $H_0$ ) along with the best fitted differences between the unknown cdfs, which is unachievable  
455 in the basic univariate case as shown by Patra and Sen [22]. An intensive numerical study has been  
456 carried out from a large range of simulation setups to illustrate the asymptotic properties of our  
457 test. This includes examples using Gaussian distributions but also more challenging distributions  
458 supported on  $\mathbb{R}^+$ ,  $\mathbb{N}$  or  $[0, 1]$  which are considered as very non-standard in the mixture models  
459 literature. Finally, our testing procedure is applied to a real-life case attempting to fill the gap in  
460 understanding the disparities of the excess of mortality during the COVID-19 crisis, which allows  
461 to test pairwise the excess of mortality across a panel of European countries.

462 This work could be extended in many interesting ways, among which the case of paired samples  
463 which would particularly be interesting for time-varying models consideration (as in [11]). We  
464 should, indeed, consider realizations of such samples while tackling the problem of the underlying  
465 dependence structure. Also, coming back to the COVID-19 case, it is fortunate to develop a  
466 more adapted scheme to pairwise testing for the underlying populations. In fact, a clustering  
467 procedure would be beneficial along with a  $K$ -sample testing procedure based on the result of this  
468 paper. This could bring in a new challenging problem. Finally, given the ability of the test to  
469 accommodate very different frameworks, an upcoming R package implementing a wide variety of  
470 two-sample testing methods for contaminated models will be very soon available to researchers as  
471 well as practitioners.

## 472 Funding

473 This work was conducted within the Research Chair DIALog under the aegis of the Risk Founda-  
474 tion, a joint initiative by CNP Assurances and ISFA, Université Claude Bernard Lyon 1 (UCBL).  
475 Y. Salhi was supported by the BNP Paribas Cardif Chair "New Insurees, Next Actuaries" as well  
476 as the CY Initiative of Excellence (grant "Investissements d'Avenir" ANR-16-IDEX-0008), Project  
477 "EcoDep" PSI-AAP2020-0000000013.

## 478 References

- 479 [1] Beaney, T., Clarke, J. M., Jain, V., Golestaneh, A. K., Lyons, G., Salman, D., & Majeed, A.  
480 (2020). Excess mortality: the gold standard in measuring the impact of COVID-19 worldwide?.  
481 J. R. Soc. Med., 113(9), 329-334.

- 482 [2] Bordes, L., Delmas, C. and Vandekerkhove, P. (2006b). Semiparametric estimation of a two-  
483 component mixture model when a component is known. *Scand. J. Stat.*, **33**, 733–752.
- 484 [3] Bordes, L. Mottelet, S. and Vandekerkhove, P. (2006a). Semiparametric estimation of a two  
485 components mixture model. *Ann. Statist.*, 34, p.1204-1232.
- 486 [4] Bordes, L. and Vandekerkhove, P. (2010). Semiparametric two-component mixture model  
487 when a component is known: an asymptotically normal estimator. *Math. Meth. Stat.*, **19**,  
488 22–41.
- 489 [5] Broët, P., Lewin, A., Richardson, S., Dalmasso, C. and Magdelenat, H. (2004). A mixture  
490 model-based strategy for selecting sets of genes in multiclass response microarray experiments.  
491 *Bioinformatics*, **20**, 2562–2571.
- 492 [6] Cai, T.T. and Jin, J. (2010). Optimal rates of convergence for estimating the null density and  
493 proportion of nonnull effects in large-scale multiple testing. *Ann. Stat.*, **38**, 100–145.
- 494 [7] Celisse, A. and Robin, S. (2010). A cross-validation based estimation of the proportion of true  
495 null hypotheses. *J. Stat. Plan. Infer.*, **140**, 3132–3147.
- 496 [8] Chakraborty, R., & Weiss, K. M. (1988). Admixture as a tool for finding linked genes and  
497 detecting that difference from allelic association between loci. *Proc. Natl. Acad. Sci. U.S.A.* ,  
498 85(23), 9119-9123.
- 499 [9] Efron, B. and Tibshirani, R. (2002). Empirical Bayes methods and false discovery rates. *Genet.*  
500 *Epidemiol.*, **23**, 70–86.
- 501 [10] Genest, C., Nešlehová, J.G., Rémillard, B. and Murphy, O.A. (2019) Testing for independence  
502 in arbitray distribution. *Biometrika*, **106**, 47–68.
- 503 [11] Ghattas, B., Pommeret, D., Reboul, L. and Yao, A. F. (2011). Data driven smooth test for  
504 paired populations. *J. Stat. Plan. Infer.*, **141**, 262–275.
- 505 [12] Gribovka, S. and Lopez, O. (2015). Non-parametric copula estimation under bivariate censor-  
506 ing. *Scand. J. Statist.*, **42**, 924–946.
- 507 [13] Higham, D. J. (2001). An Algorithmic Introduction to Numerical Simulation of Stochastic  
508 Differential Equations. *SIAM Review*, **43** (3), 525–546.
- 509 [14] Klingenberg, C., Pirner, M. and Puppò, G. (2017). A consistent kinetic model for a two-  
510 component mixture with an application to plasma. *Kinet. Relat. Models*, **10**, 445–465.
- 511 [15] Lai Y., Adam, B-L, Podolsky, R. and She, J.-X. (2007). A mixture model approach to the  
512 tests of concordance and discordance between two large scale experiments with two-sample  
513 groups. *Bioinformatics*, **23**, 1243–250.

- 514 [16] Lai, Y., Zang, F., Nayak, T.K., Modarres, R., Lee, N.H. and McCaffrey, T.H. (2017). An  
515 efficient concordant integrative analysis of multiple large-scale two-sample expression data  
516 sets. *Bioinformatics*, **33**, 3852–3860.
- 517 [17] Loh, P. R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J. K., Reich, D., & Berger,  
518 B. (2013). Inferring admixture histories of human populations using linkage disequilibrium.  
519 *Genetics*, 193(4), 1233-1254.
- 520 [18] Mannucci, E., Nreu, B., & Monami, M. (2020). Factors associated with increased all-cause  
521 mortality during the COVID-19 pandemic in Italy. *Int. J. Infect. Dis*, 98, 121-124.
- 522 [19] McLachlan, G.J., Bean, R.W., and Ben-Tovim Jones, L. (2006). A simple implementation of  
523 a normal mixture approach to differential gene expression in multiclass microarrays. *Bioin-*  
524 *formatics*, **22**, 1608–1615.
- 525 [20] Milhaud, X., Pommeret, D., Salhi, Y. and Vandekerkhove, P. (2020). Semiparametric two-  
526 sample mixture components comparison test. *Preprint: hal-02491127*.
- 527 [21] Nguyen, V.H. and Matias, C. (2014). On efficient estimators of the proportion of true null  
528 hypotheses in a multiple testing setup. *Scand. J. Stat.*, **41**, 1167-1194.
- 529 [22] Patra, R.K. and Sen, B. (2016). Estimation of a Two-component Mixture Model with Appli-  
530 cations to Multiple Testing. *J. Roy. Statist. Soc., Series B*, **78**, 869–893.
- 531 [23] Podlaski, R. and Roesch, F.A. (2014). Modelling diameter distributions of two-cohort for-  
532 est stands with various proportions of dominant species: A two-component mixture model  
533 approach. *Math. Biosci.*, **249**, 60–74.
- 534 [24] Pommeret, D. and Vandekerkhove, P. (2019). Semiparametric density testing in the contam-  
535 ination model. *Electron. J. Stat.*, **13**, 4743-4793.
- 536 [25] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*.  
537 Wiley, New York.
- 538 [26] Walker, M., Mateo, M., Olszewski, E., Sen, B. and Woodroffe M. (2009). Clean kinematic  
539 samples in dwarf spheroidals: An algorithm for evaluating membership and estimating distri-  
540 bution parameters when contamination is present. *Astron. J.*, **137**, 3109–3138.
- 541 [27] Van der Vaart, A. W. (2000). *Asymptotic statistics (Vol. 3)*. *Cambridge university press*.
- 542 [28] Xiang, S., Yao, W. and Yang, G. (2018) An overview of Semiparametric Extensions of finite  
543 Mixture Models. *Statist. Sci.*, **34**, 391–404.

## 544 A Proofs

545 **Proof of Lemma 1.** Note that, since  $p_i \in \Theta_i = [\delta_1, \delta_2]$ ,  $0 < \delta_1 < 1 < \delta_2 < +\infty$ , we have:

$$|F_i(x, p_i)| \leq \frac{1 + \tilde{\delta}}{\delta_1}, \quad (x, p_i) \in \mathbb{R} \times \Theta_i, \quad i = 1, 2 \quad \text{where} \quad \tilde{\delta} = \max(|1 - \delta_1|, |1 - \delta_2|). \quad (29)$$

546 (i) Noticing that according to (29) the mapping  $\theta \mapsto |D(x, \theta)|$  is bounded over  $\Theta$  by  $2(1 + \tilde{\delta})/\delta_1$   
 547 and continuous at any point  $\theta \in \Theta$  for all fixed  $x \in \mathbb{R}$ , the wanted result is a direct consequence  
 548 of the Lebesgue dominated convergence Theorem. The same type of proof, straightforward but  
 549 painful, hold on the gradient and the Hessian of  $D(x, \theta)$ , see Section D for close form expressions.

550

(ii) Under  $H_0$ , if  $\theta = \theta^*$  then we have  $d(\theta) = 0$ . To prove the reciprocal let us remark that  $d(\theta) = 0$  implies that  $F_1(\cdot, \theta) = F_2(\cdot, \theta)$   $\mu$ -almost everywhere ( $\mu$ -a.e.) because  $U$  is strictly increasing over  $I_U$  that includes the support of the  $L_i$ 's and  $G_i$ 's. Now according to (I) and (12) we necessarily have:

$$\begin{cases} (p_1 - p_1^*)/p_1 = 0 \\ (p_2 - p_2^*)/p_2 = 0 \\ p_2^*/p_2 - p_1^*/p_1 = 0, \end{cases} \quad \Leftrightarrow \quad p_1 = p_1^* \quad \text{and} \quad p_2 = p_2^*,$$

551 which concludes the proof.

552

(iii) Under  $H_1$  ( $F_1 \neq F_2$ ), if we assume that  $F_1(\cdot, p_1) = F_2(\cdot, p_2)$  then we similarly obtain:

$$\begin{cases} (p_1 - p_1^*)/p_1 = 0 \\ (p_2 - p_2^*)/p_2 = 0 \\ p_2^*/p_2 = 0 \\ p_1^*/p_1 = 0, \end{cases}$$

553 which is impossible since the  $p_i$ 's and  $p_i^*$ 's are supposed to be strictly greater than 0. Hence, for  
 554 any  $\theta \in \Theta$ , we have  $d(\theta) > 0$  which implies the wanted result.

555

556 iv) Let consider  $\theta$  and  $\theta'$  two distinct points in  $\Theta$ . We have according to (29):

$$\begin{aligned} |d(\theta) - d(\theta')| &\leq \int_{\mathbb{R}} |D(x, \theta) + D(x, \theta')| \times |D(x, \theta) - D(x, \theta')| dU(x) \\ &\leq \frac{4(1 + \tilde{\delta})}{\delta_1} \int_{\mathbb{R}} |D(x, \theta) - D(x, \theta')| dU(x). \end{aligned}$$

557 Since for  $i = 1, 2$  we have

$$|F_i(x, p_i) - F_i(x, p'_i)| = |L_i - G_i| \left| \frac{p'_i - p_i}{p_i p'_i} \right| \leq \frac{2}{\delta_1^2} |p_i - p'_i|,$$

558 it comes

$$|D(x, \theta) - D(x, \theta')| \leq \sum_{i=1}^2 |F_i(x, p_i) - F_i(x, p'_i)| \leq \frac{2}{\delta_1^2} \sum_{i=1}^2 |p_i - p'_i| \leq \frac{4}{\delta_1^2} \|\theta - \theta'\|_2,$$

559 which leads to

$$|d(\theta) - d(\theta')| \leq \frac{16(1 + \tilde{\delta})}{\delta_1^3} \|\theta - \theta'\|_2,$$

560 and proves the Lipschitz property for  $\theta \mapsto d(\theta)$  over  $\Theta$ . Let us denote  $\widehat{D}(x, \theta) = D(x, \widehat{L}_1, \widehat{L}_2, \theta)$   
561 where  $D(\cdot)$  is defined in (14), and notice that

$$|d_n(\theta) - d(\theta)| \leq \int_{\mathbb{R}} \left| \widehat{D}^2(x, \theta) - D^2(x, \theta) \right| dU(x).$$

562 For all  $x \in \mathbb{R}$ , we have:

$$\begin{aligned} \left| \widehat{D}^2(x, \theta) - D^2(x, \theta) \right| &\leq \left| \widehat{D}(x, \theta) + D(x, \theta) \right| \times \left| \widehat{D}(x, \theta) - D(x, \theta) \right| \\ &\leq \frac{2(1 + \tilde{\delta})}{\delta_1} \left| \widehat{D}(x, \theta) - D(x, \theta) \right| \\ &\leq \frac{2(1 + \tilde{\delta})}{\delta_1} \sum_{i=1}^2 \left| \widehat{L}_i(x, \theta) - L_i(x, \theta) \right|, \end{aligned}$$

563 which leads to

$$|d_n(\theta) - d(\theta)| \leq \frac{2(1 + \tilde{\delta})}{\delta_1} \sum_{i=1}^2 \left\| \widehat{L}_i(\cdot, \theta) - L_i(\cdot, \theta) \right\|_{\infty}.$$

564 Noticing that, for  $i = 1, 2$ ,  $\|\widehat{L}_i(\cdot, \theta) - L_i(\cdot, \theta)\|_{\infty} = O_{a.s.}(\sqrt{n_i^{-1} \log \log(n_i)})$  (see Shorack and Well-  
565 ner [25], p. 766), we obtain the wanted result.

566

567 v) For any  $\theta \in \Theta$ , we have

$$\begin{aligned} \ddot{d}(\theta) &= 2 \int_{\mathbb{R}} \ddot{D}(x, \theta) D(x, \theta) + \dot{D}(x, \theta) \dot{D}(x, \theta)^T dU(x) \\ &= 2 \int_{\mathbb{R}} M(x, \theta) dU(x), \end{aligned}$$

568 where, as detailed in Appendix D,  $M(x, \theta)$  is a  $2 \times 2$  symmetric real-valued matrix for all  $x$  fixed  
569 in  $\mathbb{R}$ . As a consequence, we have for all  $x \in \mathbb{R}$  and any vector  $v \in \mathbb{R}^2$ :

$$v^T M(x, \theta) v \geq 0 \quad \Rightarrow \quad v^T \ddot{d}(\theta) v = 2 \int_{\mathbb{R}} v^T M(x, \theta) v dU(x) \geq 0,$$



570 but also

$$v^T \ddot{d}(\theta)v = 0 \quad \Rightarrow \quad M_{1,1}(x, \theta)v_1^2 + 2M_{1,2}(x, \theta)v_2v_1 + M_{2,2}(x, \theta)v_2^2 = 0 \quad \mu - a.e.$$

571 Now according to assumption **(II)** and decomposition (48), the above nullity condition implies in  
 572 particular that the coefficients associated to  $F_i^2$ 's are null, *i.e.*  $a_3(p_i, v_i)(p_i^*)^2 = 3v_i^2(p_i^*)^2/p_i^4 = 0$ ,  
 573  $i = 1, 2$ , which implies  $v_1 = v_2 = 0$  and concludes the proof of the definite-positiveness of  $\dot{d}(\theta)$  for  
 574 any  $\theta \in \Theta$ .

575 **Proof of Theorem 1** By Lemma 1 v) or vi) there exists  $\gamma > 0$  such that for all  $v \in \mathbb{R}^2$ ,  
 576  $v^T \ddot{d}(\theta^c)v > \gamma \|v\|_2^2$ . By a two order Taylor expansion of  $d$  at point  $\theta^c \in \overset{\circ}{\Theta}$  we can find  $\eta > 0$  such  
 577 that for all  $v$  satisfying  $\|v\| < \eta$  and  $\theta^c + v \in \overset{\circ}{\Theta}$ , we have

$$d(\theta^c + v) \geq \frac{\gamma}{4} \|v\|_2^2. \quad (30)$$

579 Let us consider now  $B(\theta^c, \eta_n)$  the Euclidean ball centered at point  $\theta^c$  with radius  $\eta_n > 0$ . Following  
 580 the proof in Bordes et al. [3] we show the following events inclusion:

$$\limsup_n \left\{ \widehat{\theta}_n \notin B(\theta^c, \eta_n) \right\} \subseteq \limsup_n \left\{ \inf_{\theta \in \Theta \setminus B(\theta^c, \eta_n)} d(\theta) < \xi_n \right\} \cup \limsup_n \left\{ \xi_n \leq 2 \sup_{\theta \in \Theta} |d_n(\theta) - d(\theta)| \right\},$$

581 for any arbitrary sequence  $\xi_n$ . Choosing now  $\xi_n = n^{-1/2+\alpha}$  and  $\eta_n = n^{-1/4+\beta/2}$ , with  $0 < \alpha < \beta$   
 582 taken arbitrarily small, it follows from (30) and the uniform almost sure rate of  $d_n$  given in Lemma  
 583 1 (iv), that

$$P \left( \limsup_n \left\{ \inf_{\theta \in \Theta \setminus B(\theta^c, \eta_n)} d(\theta) < \xi_n \right\} \right) = 0,$$

584 and

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

585 In conclusion,  $\widehat{\theta}_n$  converges almost surely towards  $\theta^c$  at rate  $n^{-1/4+\alpha}$ ,  $\alpha > 0$  chosen arbitrarily small.

586 **Proof of Theorem 2.** i) By a Taylor expansion of  $\dot{d}_n$  about  $\theta^c \in \overset{\circ}{\Theta}$  we have:  
 587

$$\ddot{d}_n(\tilde{\theta}_n) \sqrt{n}(\tilde{\theta}_n - \theta^c) = -\sqrt{n} \dot{d}_n(\theta^c), \quad (31)$$

588 where  $\tilde{\theta}_n$  lies in the line segment with extremities  $\widehat{\theta}_n$  and  $\theta^c$ . Now writing that  $\dot{d}(\theta) = (\dot{d}_1(\theta), \dot{d}_2(\theta))^T$ ,

$$\begin{aligned} \dot{d}_1(\theta) &= 2E_U \left( \frac{(2-p_1)G_1L_1}{p_1^3} - \frac{(1-p_1)G_1^2}{p_1^3} - \frac{L_1^2}{p_1^3} - \frac{G_1L_2}{p_1^2p_2} + \frac{(1-p_2)G_1G_2}{p_1^2p_2} + \frac{L_1L_2}{p_1^2p_2} - \frac{(1-p_2)L_1G_2}{p_1^2p_2} \right) \\ \dot{d}_2(\theta) &= 2E_U \left( \frac{(2-p_2)G_2L_2}{p_2^3} - \frac{(1-p_2)G_2^2}{p_2^3} - \frac{L_2^2}{p_2^3} - \frac{G_2L_1}{p_2^2p_1} + \frac{(1-p_1)G_2G_1}{p_2^2p_1} + \frac{L_2L_1}{p_2^2p_1} - \frac{(1-p_1)L_2G_1}{p_2^2p_1} \right), \end{aligned}$$

589 we look at

$$\begin{aligned} \dot{d}_{1,n}(\theta) - \dot{d}_1(\theta) &= 2 \left( \frac{2-p_1}{p_1^3} T_{1,1} - \frac{1-p_1}{p_1^3} T_{1,2} - \frac{1}{p_1^3} T_{1,3} - \frac{1}{p_1^2 p_2} T_{1,4} + \frac{1-p_2}{p_1^2 p_2} T_{1,5} + \frac{1}{p_1^2 p_2} T_{1,6} - \frac{1-p_2}{p_1^2 p_2} T_{1,7} \right) \\ \dot{d}_{2,n}(\theta) - \dot{d}_2(\theta) &= 2 \left( \frac{2-p_2}{p_2^3} T_{2,1} - \frac{1-p_2}{p_2^3} T_{2,2} - \frac{1}{p_2^3} T_{2,3} - \frac{1}{p_2^2 p_1} T_{2,4} + \frac{1-p_1}{p_2^2 p_1} T_{2,5} + \frac{1}{p_2^2 p_1} T_{2,6} - \frac{1-p_1}{p_2^2 p_1} T_{2,7} \right), \end{aligned}$$

590 where

$$\begin{aligned} T_{1,1}(G_1, L_1) &= E_U \left( G_1 (\widehat{L}_1 - L_1) \right) \\ T_{1,2}(G_1) &= E_U (G_1^2 - G_1^2) = 0 \\ T_{1,3}(L_1) &= E_U (\widehat{L}_1^2 - L_1^2) = E_U \left( (\widehat{L}_1 - L_1)(\widehat{L}_1 + L_1) \right) = E_U \left( (\widehat{L}_1 - L_1)(2L_1 + o_{a.s.}(1)) \right) \\ T_{1,4}(G_1, L_2) &= E_U (G_1(\widehat{L}_2 - L_2)) \\ T_{1,5}(G_1, G_2) &= E_U (G_1 G_2) - E_U (G_1 G_2) = 0 \\ T_{1,6}(L_1, L_2) &= E_U (\widehat{L}_1 \widehat{L}_2 - L_1 L_2) = E_U \left( \widehat{L}_1(\widehat{L}_2 - L_2) + L_2(\widehat{L}_1 - L_1) \right) \\ &= E_U \left( (L_1 + o_{a.s.}(1))(\widehat{L}_2 - L_2) \right) + E_U \left( L_2(\widehat{L}_1 - L_1) \right) \\ T_{1,7}(G_2, L_1) &= E_U \left( G_2(\widehat{L}_1 - L_1) \right) = T_{1,4}(G_2, L_1) \\ T_{2,1}(G_2, L_2) &= E_U \left( G_2(\widehat{L}_2 - L_2) \right) = T_{1,1}(G_2, L_2) \\ T_{2,2}(G_2) &= E_U (G_2^2) - E_U (G_2^2) = 0 \\ T_{2,3}(L_2) &= E_U \left( (\widehat{L}_2 - L_2)(\widehat{L}_2 + L_2) \right) = E_U \left( (\widehat{L}_2 - L_2)(2L_2 + o_{a.s.}(1)) \right) = T_{1,3}(L_2) \\ T_{2,4}(L_1, G_2) &= E_U \left( G_2(\widehat{L}_1 - L_1) \right) = T_{1,4}(L_1, G_2) \\ T_{2,5}(G_1, G_2) &= E_U (G_2 G_1) - E_U (G_2 G_1) = 0 \\ T_{2,6}(L_1, L_2) &= E_U \left( \widehat{L}_2 \widehat{L}_1 - L_2 L_1 \right) = T_{1,6}(L_1, L_2) \\ T_{2,7}(G_1, L_2) &= E_U \left( G_1(\widehat{L}_2 - L_2) \right). \end{aligned}$$

591

592 For a generic cdf  $Y$  and a generic  $N$ -sample based empirical process  $\mathbb{V} = \sqrt{N}(\widehat{V} - V)$ , define

593  $\varphi(Y, \mathbb{V}) = \int_{\mathbb{R}} Y(x) \mathbb{V}(x) dU(x)$ . Introducing  $\mathcal{S} = (G_1, G_2, L_1, L_2)$ , let us consider

$$\begin{aligned} \Psi_{1,1}(\mathcal{S}, \theta) &= 2 \left( \frac{2-p_1}{p_1^3} G_1 - \frac{2}{p_1^3} L_1 + \frac{1}{p_1^2 p_2} L_2 - \frac{1-p_2}{p_1^2 p_2} G_2 \right) \\ \Psi_{1,2}(\mathcal{S}, \theta) &= 2 \left( \frac{1}{p_1^2 p_2} (L_1 - G_1) \right) \\ \Psi_{2,1}(\mathcal{S}, \theta) &= 2 \left( \frac{2-p_2}{p_2^3} G_2 - \frac{2}{p_2^3} L_2 + \frac{1}{p_2^2 p_1} L_1 - \frac{1-p_1}{p_2^2 p_1} G_1 \right) \\ \Psi_{2,2}(\mathcal{S}, \theta) &= 2 \left( \frac{1}{p_2^2 p_1} (L_2 - G_2) \right), \end{aligned}$$

594 and

$$\begin{aligned}\Phi_{1,1}(\mathbb{L}_1, \theta) &= \varphi(\Psi_{1,1}(S, \theta), \mathbb{L}_1), & \Phi_{1,2}(\mathbb{L}_2, \theta) &= \varphi(\Psi_{1,2}(S, \theta), \mathbb{L}_2), \\ \Phi_{2,1}(\mathbb{L}_2, \theta) &= \varphi(\Psi_{2,1}(S, \theta), \mathbb{L}_2), & \Phi_{2,2}(\mathbb{L}_1, \theta) &= \varphi(\Psi_{2,2}(S, \theta), \mathbb{L}_1).\end{aligned}$$

595 Note that the first and fourth, respectively the second and third, expression depends only on the  
596 randomness of  $\mathbb{L}_1$ , resp.  $\mathbb{L}_2$ . We resume the above remarks into the following basic expression:

$$\sqrt{n}(\dot{d}_n(\theta) - \dot{d}(\theta)) = \mathbf{\Phi}(\mathbb{L}_1, \mathbb{L}_2, \theta) + o_{a.s.}(1), \quad (32)$$

597 where, according to  $\sqrt{n} = \sqrt{\kappa n} / \sqrt{\kappa} = \zeta \sqrt{n_2}$  with  $\zeta = 1/\sqrt{\kappa}$ :

$$\mathbf{\Phi}(\mathbb{L}_1, \mathbb{L}_2) = \begin{bmatrix} \Phi_{1,1}(\mathbb{L}_1, \theta) + \zeta \Phi_{1,2}(\mathbb{L}_2, \theta) \\ \zeta \Phi_{2,1}(\mathbb{L}_2, \theta) + \Phi_{2,2}(\mathbb{L}_1, \theta) \end{bmatrix} \quad (33)$$

598 Since the empirical processes  $\mathbb{L}_1$  and  $\mathbb{L}_2$  are independent, by the Donsker Theorem [27, The-  
599 orem 19.3, p. 266], the vector  $[\mathbb{L}_1, \mathbb{L}_2]$  converges in distribution in the space  $D[-\infty, \infty]$  to a  
600 bi-dimensional zero-mean Gaussian process  $\mathcal{B}$ , i.e.

$$\begin{bmatrix} \mathbb{L}_1 \\ \mathbb{L}_2 \end{bmatrix} \rightsquigarrow \mathcal{B} = \begin{bmatrix} \mathcal{B}_1 \\ \mathcal{B}_2 \end{bmatrix}, \quad (34)$$

601 where  $\mathcal{B}$  is a bi-dimensional gaussian process with diagonal correlation matrix  $\rho = \text{diag}(\rho_1, \rho_2)$ ,  
602 where  $\rho_1(x, y) = L_1(x \wedge y)(1 - L_1(x \vee y))$  and  $\rho_2(x, y) = L_2(x \wedge y)(1 - L_2(x \vee y))$ .

603

604 Moreover,

$$\begin{aligned}\sqrt{n}[D(x, \hat{L}_1, \hat{L}_2, \hat{\theta}_n) - D(x, L_1, L_2, \theta^c)] &= \sqrt{n}[F_1(x, \hat{L}_1, \hat{p}_1) - F_1(x, L_1, p_1^c)] \\ &\quad - \sqrt{n}[(F_2(x, \hat{L}_2, \hat{p}_2) - F_2(x, L_2, p_2^c))].\end{aligned} \quad (35)$$

605 Let decompose closely, for  $i = 1, 2$ , the terms  $F_i(x, \hat{L}_i, \hat{p}_i) - F_i(x, L_i, p_i^c)$ :

$$\begin{aligned}\sqrt{n}[F_i(\cdot, \hat{L}_i, \hat{p}_i) - F_i(\cdot, L_i, p_i^c)] &= \sqrt{n} \left[ \left( \frac{\hat{L}_i}{\hat{p}_i} - \frac{L_i}{p_i^c} \right) - \left( \frac{1 - \hat{p}_i}{\hat{p}_i} - \frac{1 - p_i^c}{p_i^c} \right) G_i \right] \\ &= \sqrt{n} \left[ \frac{\hat{L}_i - L_i}{\hat{p}_i} \right] + \sqrt{n} \left( \frac{p_i^c - \hat{p}_i}{p_i^c \hat{p}_i} \right) (L_i - G_i) \\ &= \zeta_i \frac{1}{p_i^c} \mathbb{L}_i - \left( \frac{L_i - G_i}{(p_i^c)^2} \sqrt{n} [\hat{p}_i - p_i^c] \right) + o_P(1),\end{aligned} \quad (36)$$

606 where by convention  $\zeta_1 = 1$  and  $\zeta_2 = \zeta = \frac{1}{\sqrt{\kappa}}$ . It is also easy to prove that  $\ddot{d}_n(\tilde{\theta}_n) \xrightarrow{a.s.} \ddot{d}(\theta^c) > 0$ , as  
607  $n \rightarrow +\infty$ . Indeed, let us now consider the decompositions (45), (46) and (47) in Appendix D, we

608 have

$$\begin{aligned}
\left| [\ddot{d}_n(\tilde{\theta}_n) - \ddot{d}(\theta^c)]_{i,j} \right| &\leq \int_{\mathbb{R}} \left| M_{i,j}(x, \hat{L}_1, \hat{L}_2, G_1, G_2, \tilde{\theta}_n) - M_{i,j}(x, L_1, L_2, G_1, G_2, \theta) \right| dU(x) \\
&\leq \int_{\mathbb{R}} \left| M_{i,j}(x, \hat{L}_1, \hat{L}_2, G_1, G_2, \tilde{\theta}_n) - M_{i,j}(x, L_1, L_2, G_1, G_2, \tilde{\theta}_n) \right| dU(x) \\
&\quad + \int_{\mathbb{R}} \left| M_{i,j}(x, L_1, L_2, G_1, G_2, \tilde{\theta}_n) - M_{i,j}(x, L_1, L_2, G_1, G_2, \theta) \right| dU(x) \\
&\leq C \left( \mathcal{P}(\tilde{\theta}_n) \left[ \sum_{i=1}^2 \|\hat{L}_i - L_i\|_{\infty} \right] + |\mathcal{P}(\tilde{\theta}_n) - \mathcal{P}(\theta^c)| \right), \tag{37}
\end{aligned}$$

609 where  $\mathcal{P}(\theta) = \sum_{k=0}^4 p_1^{-k} p_2^{-4+k}$  is a  $\mathbb{R}^2 \rightarrow \mathbb{R}$  continuous mapping. Now by using on (37) the  
610 Glivenko-Cantelli theorem and the almost sure convergence of  $\hat{\theta}_n$  towards  $\theta^c$ , we obtain the wanted  
611 result.

612 In order to synthetically summarize results (31), (32), (33) and (35–36) for the Central Limit  
613 Theorem relative to our quantities of interest, we define the following matrix-type relation:

$$\sqrt{n} \begin{bmatrix} \hat{p}_1 - p_1^c \\ \hat{p}_2 - p_2^c \\ D_n(\cdot) - D(\cdot) \end{bmatrix} = M(\theta^c, \cdot) \begin{bmatrix} \Phi_{1,1}(\mathbb{L}_1, \theta^c) \\ \Phi_{2,2}(\mathbb{L}_1, \theta^c) \\ \mathbb{L}_1 \\ \Phi_{2,1}(\mathbb{L}_2, \theta^c) \\ \Phi_{1,2}(\mathbb{L}_2, \theta^c) \\ \mathbb{L}_2 \end{bmatrix} + o_{a.s.}(1), \quad \text{with } M(\theta^c, \cdot) = L(\cdot, \theta^c) J^{-1}(\theta^c) \mathcal{C} \tag{38}$$

614 where

$$\mathcal{C} = \begin{bmatrix} -1 & 0 & 0 & 0 & -\zeta & 0 \\ 0 & -1 & 0 & -\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad J(\theta) = \begin{bmatrix} \ddot{d}(\theta) & 0_{2 \times 2} \\ 0_{2 \times 2} & \text{Id}_{2 \times 2} \end{bmatrix}, \quad L(\cdot, \theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\frac{L_1(\cdot) - G_1(\cdot)}{p_1^2} & \frac{L_2(\cdot) - G_2(\cdot)}{p_2^2} & \frac{1}{p_1} & -\frac{\zeta}{p_2} \end{bmatrix}. \tag{39}$$

615 We finally have

$$\begin{bmatrix} \Phi_{1,1}(\mathbb{L}_1, \theta^c) \\ \Phi_{2,2}(\mathbb{L}_1, \theta^c) \\ \mathbb{L}_1 \\ \Phi_{2,1}(\mathbb{L}_2, \theta^c) \\ \Phi_{1,2}(\mathbb{L}_2, \theta^c) \\ \mathbb{L}_2 \end{bmatrix} \xrightarrow{\mathcal{L}} Z = \begin{bmatrix} \Phi_{1,1}(\mathcal{B}_1, \theta^c) \\ \Phi_{2,2}(\mathcal{B}_1, \theta^c) \\ \mathcal{B}_1 \\ \Phi_{2,1}(\mathcal{B}_2, \theta^c) \\ \Phi_{1,2}(\mathcal{B}_2, \theta^c) \\ \mathcal{B}_2 \end{bmatrix}, \tag{40}$$

616 where  $Z$  is Gaussian random vector of  $\mathbb{R}^6$  with covariance matrix  $\Sigma_L = E(ZZ^T)$ . Since  $\mathcal{B}_1$  and  $\mathcal{B}_2$   
617 are two independent (limiting) Gaussian processes, we have:

$$\Sigma_L(x, y) = \begin{bmatrix} \Sigma_1(x, y) & 0_{3 \times 3} \\ 0_{3 \times 3} & \Sigma_2(x, y) \end{bmatrix}. \tag{41}$$

618 Because

$$\begin{aligned}
E\left(\int_{\mathbb{R}} f_1(x)\mathcal{B}(x)dU(x) \times \int_{\mathbb{R}} f_2(y)\mathcal{B}(y)dU(y)\right) &= E\left(\int_{\mathbb{R}^2} f_1(x)f_2(y)\mathcal{B}(x)\mathcal{B}(y)dU(x)dU(y)\right) \\
&= \int_{\mathbb{R}^2} f_1(x)f_2(y)E(\mathcal{B}(x)\mathcal{B}(y))dU(x)dU(y) \\
&= \int_{\mathbb{R}^2} f_1(x)f_2(y)\text{cov}(\mathcal{B}(x), \mathcal{B}(y))dU(x)dU(y) \\
&= \int_{\mathbb{R}^2} f_1(x)f_2(y)\rho(x, y)dU(x)dU(y),
\end{aligned}$$

619 it comes that  $\Sigma_1(x, y) = (\sigma_1(i, j; x, y))_{1 \leq i, j \leq 3}$  where

$$\begin{aligned}
\sigma_1(1, 1; x, y) &= \sigma_1(1, 1) = \int_{\mathbb{R}^2} \Psi_{1,1}(S, \theta, u)\Psi_{1,1}(S, \theta, v)\rho_1(u, v)dU(u)dU(v) \\
\sigma_1(1, 2; x, y) &= \sigma_1(1, 2) = \sigma_1(2, 1) = \int_{\mathbb{R}^2} \Psi_{1,1}(S, \theta, u)\Psi_{2,2}(S, \theta, v)\rho_1(u, v)dU(u)dU(v) \\
\sigma_1(1, 3; x, y) &= \sigma_1(1, 3; y) = \int_{\mathbb{R}} \Psi_{1,1}(S, \theta, u)\rho_1(u, y)dU(u) \\
\sigma_1(2, 2; x, y) &= \sigma_1(2, 2) = \int_{\mathbb{R}^2} \Psi_{2,2}(S, \theta, u)\Psi_{2,2}(S, \theta, v)\rho_1(u, v)dU(u)dU(v) \\
\sigma_1(2, 3; x, y) &= \sigma_1(2, 3; y) = \int_{\mathbb{R}} \Psi_{2,2}(S, \theta, u)\rho_1(u, y)dU(u) \\
\sigma_1(3, 3; x, y) &= \rho_1(x, y) \\
\sigma_1(3, 1; x, y) &= \sigma_1(3, 1; x) = \sigma_1(1, 3; x) \\
\sigma_1(3, 2; x, y) &= \sigma_1(3, 2; x) = \sigma_1(2, 3; x),
\end{aligned}$$

620 and  $\Sigma_2 = (\sigma_2(i, j; x, y))_{1 \leq i, j \leq 3}$  where

$$\begin{aligned}
\sigma_2(1, 1; x, y) &= \sigma_2(1, 1) = \int_{\mathbb{R}^2} \Psi_{2,1}(S, \theta, u)\Psi_{2,1}(S, \theta, v)\rho_2(u, v)dU(u)dU(v) \\
\sigma_2(1, 2; x, y) &= \sigma_2(1, 2) = \sigma_2(2, 1) = \int_{\mathbb{R}^2} \Psi_{2,1}(S, \theta, u)\Psi_{1,2}(S, \theta, v)\rho_2(u, v)dU(u)dU(v) \\
\sigma_2(1, 3; x, y) &= \sigma_2(1, 3; y) = \int_{\mathbb{R}} \Psi_{2,1}(S, \theta, u)\rho_2(u, y)dU(u) \\
\sigma_2(2, 2; x, y) &= \sigma_2(2, 2) = \int_{\mathbb{R}^2} \Psi_{1,2}(S, \theta, u)\Psi_{1,2}(S, \theta, v)\rho_2(u, v)dU(u)dU(v) \\
\sigma_2(2, 3; x, y) &= \sigma_2(2, 3; y) = \int_{\mathbb{R}} \Psi_{1,2}(S, \theta, u)\rho_2(u, y)dU(u) \\
\sigma_2(3, 3; x, y) &= \rho_2(x, y) \\
\sigma_2(3, 1; x, y) &= \sigma_2(3, 1; x) = \sigma_2(1, 3; x) \\
\sigma_2(3, 2; x, y) &= \sigma_2(3, 2; x) = \sigma_2(2, 3; x).
\end{aligned}$$

621 Note that all the above matrices can be estimated consistently, see Appendix D.2.

622 ii) Let us now decompose  $nd_n(\widehat{\theta}_n)$  :

$$\begin{aligned}
nd_n(\widehat{\theta}_n) &= \int_{\mathbb{R}} n(D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\theta}_n) - D(x, L_1, L_2, \theta^c) + D(x, L_1, L_2, \theta^c))^2 dU(x) \\
&= \int_{\mathbb{R}} (\sqrt{n}[D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\theta}_n) - D(x, L_1, L_2, \theta^c)])^2 dU(x) \\
&\quad + 2\sqrt{n} \int_{\mathbb{R}} \sqrt{n}[D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\theta}_n) - D(x, L_1, L_2, \theta^c)]D(x, L_1, L_2, \theta^c) dU(x) \\
&\quad + n \int_{\mathbb{R}} D^2(x, L_1, L_2, \theta^c) dU(x). \tag{42}
\end{aligned}$$

623 Note that under  $H_0$ ,  $\theta^c = \theta^*$ , and we simply have:

$$\begin{aligned}
nd_n(\widehat{\theta}_n) &= \int_{\mathbb{R}} (\sqrt{n}D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\theta}_n))^2 dU(x) \\
&= \int_{\mathbb{R}} (\sqrt{n}[D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\theta}_n) - D(x, L_1, L_2, \theta^*)])^2 dU(x) \\
&= \mathbf{U}_n^0, \tag{43}
\end{aligned}$$

624 since  $D(\cdot, L_1, L_2, \theta^*) = 0$  almost everywhere. Next, notice that it is easy to show that the map  
625  $\mathbf{U}_n(\mathbb{D})$  is Hadamard differentiable from the domain of càd-làg functions of bounded variation into  
626  $\mathbb{R}$  [27, Lemma 20.10]. This combined with the result for weak convergence of the empirical process  
627  $\mathbb{D}$ , in the third row of (18), yields the desired result.

628 Under  $H_1$ , we have  $\int_{\mathbb{R}} D^2(x, L_1, L_2, \theta^c) dU(x) > 0$ , which leads to

$$\begin{aligned}
nd_n(\widehat{\theta}_n) &= \int_{\mathbb{R}} (\sqrt{n}[D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\theta}_n) - D(x, L_1, L_2, \theta^c)])^2 dU(x) \\
&\quad + n \int_{\mathbb{R}} D^2(x, L_1, L_2, \theta^c) dU(x) + o_P(n) \\
&= \mathbf{U}_n^1 + \mathbf{V}_n^1. \tag{44}
\end{aligned}$$

629 Given the asymptotic convergence analysis under both  $H_0$  or  $H_1$ , the random variable within  
630 brackets involved in (42) and (44) can be analyzed closely. Being able to tabulate its asymptotic  
631 distribution whatever  $H_0$  or  $H_1$  is true, we will be able to accept  $H_0$  if the test statistic  $nd_n(\widehat{\theta}_n)$  is  
632 lower than the  $(1 - \alpha)$ -quantile of that asymptotic distribution, otherwise reject  $H_0$ .

## 633 B Optimization of the discrepancy measure

634 As discussed in Section 1, we can face two types of situations when looking for the solution of our  
635 optimization problem (15) : i) there exists a local minima of  $d(\theta)$ ,  $\theta^*$  under  $H_0$  or  $\theta^c$  under  $H_1$ ,  
636 in the interior of the parametric space, and then the testing problem is non-trivial and should be

637 addressed, or ii) the optimization of  $d(\theta)$  shows that we bump into the boundaries of the parametric  
638 space, i.e. one of the component of  $\theta^c$  is equal to the upper-bound of the parametric space because  
639 the main way to reduce the contrast  $d(\theta)$  is to make  $\theta$  large. In the latter case, the testing problem  
640 is not even worth addressing (we clearly are under  $H_1$ ). This phenomenon is illustrated on Figure 8,  
641 where the two first graphs show the situation i) and the last one depicts the case ii). It is clear  
642 that a unique global minimizer belonging to the natural parametric space  $]0, 1[$  exists under  $H_0$ .  
643 On the contrary, under  $H_1$ , local minima out of this natural parametric space can exist and lead  
to select a minimizer tending to infinity.

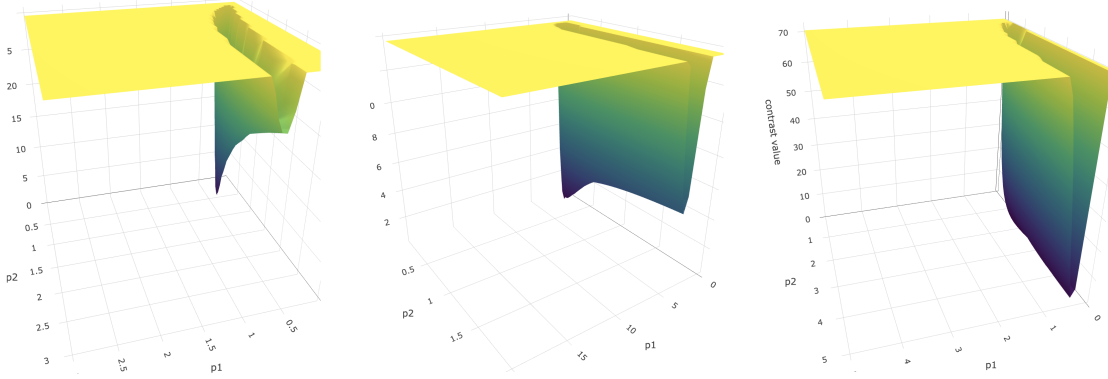


Figure 8: Example of the surface of the discrepancy  $d(\theta)$  where  $p_i (i = 1, 2)$  belongs to  $\Theta_i$ . From left to right: under  $H_0$ , under  $H_1$  with a solution belonging to the natural parametric space  $]0, 1[$ , and under  $H_1$  where  $\hat{p}_2$  (minimizer of  $d(\theta)$ ) is very likely to be very far from the interval  $]0, 1[$ .

644

## 645 C Validation of the estimators variance-covariance

646 To validate the explicit covariance structure between the estimators, it is necessary to fix  $z$  and  
647 to compare the empirical covariances obtained by Monte Carlo simulations to the theoretical  
648 ones. Here, we illustrate the case of two-component Gaussian mixtures with parameters given in  
649 Section 3.2. Moreover, we take  $z = 2$  in formulas (38)-(41). Hereafter, we provide the numerical  
650 approximations for the theoretical version of the covariance structure:

$$\begin{bmatrix} \text{Var}(P_1) & \text{Cov}(P_1, P_2) & \text{Cov}(P_1, D_z) \\ \text{Cov}(P_2, P_1) & \text{Var}(P_2) & \text{Cov}(P_2, D_z) \\ \text{Cov}(D_z, P_1) & \text{Cov}(D_z, P_2) & \text{Var}(D_z) \end{bmatrix} = \begin{bmatrix} 2.0086503 & 1.1209473 & -0.1861036 \\ 1.1209473 & 1.6979287 & -0.2714744 \\ -0.1861036 & -0.2714744 & 0.5082710 \end{bmatrix},$$

651 to be compared to its Monte Carlo version:

$$\begin{bmatrix} \widehat{\text{Var}}(P_1) & \widehat{\text{Cov}}(P_1, P_2) & \widehat{\text{Cov}}(P_1, D_z) \\ \widehat{\text{Cov}}(P_2, P_1) & \widehat{\text{Var}}(P_2) & \widehat{\text{Cov}}(P_2, D_z) \\ \widehat{\text{Cov}}(D_z, P_1) & \widehat{\text{Cov}}(D_z, P_2) & \widehat{\text{Var}}(D_z) \end{bmatrix} = \begin{bmatrix} 2.031565 & 1.091873 & -0.170923 \\ 1.091873 & 1.639897 & -0.29527 \\ -0.170923 & -0.29527 & 0.5158633 \end{bmatrix}.$$

652 We can see that the approximations of both quantities are very close, which ensures the validity  
653 of  $i$ ) in Theorem 2 in this case. Of course, we checked that for many different values of  $z$ , as well  
654 as many different frameworks in terms of mixture distributions.

## 655 D Further details for computations

### 656 D.1 Derivatives

657 Let us compute the gradient vector  $\dot{D}(\theta)$  and the Hessian matrix  $\ddot{D}(\theta)$  for any  $\theta \in \Theta$ ,

$$\begin{aligned}\dot{D}(\theta) &= \left( \frac{G_1(x) - L_1(x)}{p_1^2}, -\frac{G_2(x) - L_2(x)}{p_2^2} \right)^T \\ \ddot{D}(\theta) &= \text{diag} \left( -2\frac{G_1(x) - L_1(x)}{p_1^3}, 2\frac{G_2(x) - L_2(x)}{p_2^3} \right).\end{aligned}$$

Dropping for simplicity the dependence on  $(x, \theta)$  it comes now that

$$\dot{D}(x, \theta)\dot{D}(x, \theta)^T = \begin{pmatrix} \frac{(G_1(x)-L_1(x))^2}{p_1^4} & -\frac{(G_1(x)-L_1(x))(G_2(x)-L_2(x))}{p_1^2 p_2^2} \\ -\frac{(G_1(x)-L_1(x))(G_2(x)-L_2(x))}{p_1^2 p_2^2} & \frac{(G_2(x)-L_2(x))^2}{p_2^4} \end{pmatrix}$$

658 and

$$\begin{aligned}\ddot{D}(\theta)D(\theta) &= \text{diag} \left( -2\frac{G_1(x) - L_1(x)}{p_1^3}, 2\frac{G_2(x) - L_2(x)}{p_2^3} \right) \\ &\quad \times \left[ \frac{L_1(x) - (1 - p_1)G_1(x)}{p_1} - \frac{L_2(x) - (1 - p_2)G_2(x)}{p_2} \right].\end{aligned}$$

659 Let us compute now the terms of the matrix  $M(x, \theta) = \dot{D}(x, \theta)\dot{D}(x, \theta)^T + \ddot{D}(\theta)D(\theta)$ . Dropping  
660 for simplicity the dependence on  $(x, \theta)$  It comes now that

$$\begin{aligned}M_{1,1} &= \frac{G_1^2 - 2G_1L_1 + L_1^2}{p_1^4} - \frac{2}{p_1^4}G_1L_1 + \frac{2(1-p_1)}{p_1^4}G_1^2 + \frac{2}{p_1^3p_2}G_1L_2 - \frac{2(1-p_2)}{p_1^3p_2}G_1G_2 \\ &\quad + \frac{2}{p_1^4}L_1^2 - \frac{2(1-p_1)}{p_1^4}L_1G_1 - \frac{2}{p_1^3p_2}L_1L_2 + \frac{2(1-p_2)}{p_1^3p_2}L_1G_2 \\ &= \frac{-6 + 2p_1}{p_1^4}G_1L_1 + \frac{3 - 2p_1}{p_1^4}G_1^2 + \frac{2}{p_1^3p_2}G_1L_2 - \frac{2(1-p_2)}{p_1^3p_2}G_1G_2 \\ &\quad + \frac{3}{p_1^4}L_1^2 - \frac{2}{p_1^3p_2}L_1L_2 + \frac{2(1-p_2)}{p_1^3p_2}L_1G_2,\end{aligned}\tag{45}$$



$$\begin{aligned}
M_{2,2} &= \frac{G_2^2 - 2G_2L_2 + L_2^2}{p_2^4} + \frac{2}{p_2^3p_1}G_2L_1 - \frac{2(1-p_1)}{p_2^3p_1}G_2G_1 - \frac{2}{p_2^4}G_2L_2 + \frac{2(1-p_2)}{p_2^4}G_2^2 \\
&\quad - \frac{2}{p_2^3p_1}L_2L_1 + \frac{2(1-p_1)}{p_2^3p_1}L_2G_1 + \frac{2}{p_2^4}L_2^2 - \frac{2(1-p_2)}{p_2^4}L_2G_2 \\
&= \frac{2}{p_2^3p_1}G_2L_1 - \frac{2(1-p_1)}{p_2^3p_1}G_2G_1 + \frac{-6+2p_2}{p_2^4}G_2L_2 + \frac{3-2p_2}{p_2^4}G_2^2 \\
&\quad - \frac{2}{p_2^3p_1}L_2L_1 + \frac{2(1-p_1)}{p_2^3p_1}L_2G_1 + \frac{3}{p_2^4}L_2^2, \tag{46}
\end{aligned}$$

$$M_{1,2} = M_{2,1} = -\frac{1}{p_1^2p_2^2}G_1G_2 + \frac{1}{p_1^2p_2^2}G_1L_2 + \frac{1}{p_1^2p_2^2}L_1G_2 - \frac{1}{p_1^2p_2^2}L_1L_2. \tag{47}$$

661 Since now the term  $v^T M v = M_{1,1}v_1^2 + M_{2,2}v_2^2 + 2M_{1,2}v_1v_2$  and denoting

$$\begin{aligned}
a_1(p_1, v_1) &= \frac{(-6+2p_1)v_1^2}{p_1^4}, & a_2(p_1, v_1) &= \frac{(3-2p_1)v_1^2}{p_1^4}, & a_3(p_1, v_1) &= \frac{3v_1^2}{p_1^4}, \\
b_1(p_1, p_2, v_1, v_2) &= \left( \frac{2v_1^2}{p_1^3p_2} + \frac{2(1-p_1)v_2^2}{p_2^3p_1} + \frac{2v_1v_2}{p_1^2p_2^2} \right), \\
b_2(p_1, p_2, v_1, v_2) &= \left( -\frac{2(1-p_2)v_1^2}{p_1^3p_2} - \frac{2(1-p_1)v_2^2}{p_2^3p_1} - \frac{2v_1v_2}{p_1^2p_2^2} \right), \\
b_3(p_1, p_2, v_1, v_2) &= \left( -\frac{2v_1^2}{p_1^3p_2} - \frac{2v_2^2}{p_2^3p_1} + \frac{2v_1v_2}{p_1^2p_2^2} \right),
\end{aligned}$$

662 we obtain the following developed expression:

$$\begin{aligned}
v^T M v &= [a_1(p_1, v_1)G_1L_1 + a_1(p_2, v_2)G_2L_2] + [a_2(p_1, v_1)G_1^2 + a_2(p_2, v_2)G_2^2] \\
&\quad + [a_3(p_1, v_1)L_1^2 + a_3(p_2, v_2)L_2^2] + [b_1(p_1, p_2, v_1, v_2)G_1L_2 + b_1(p_2, p_1, v_2, v_1)L_1G_2] \\
&\quad + b_2(p_1, p_2, v_1, v_2)G_1G_2 + b_3(p_1, p_2, v_1, v_2)L_1L_2 \\
&= [a_1(p_1, v_1)(1-p_1^*) + a_2(p_1, v_1) + a_3(p_1, v_1)(1-p_1^*)^2]G_1^2 \\
&\quad + [a_2(p_2, v_2)(1-p_2^*) + a_2(p_2, v_2) + a_3(p_2, v_2)(1-p_2^*)^2]G_2^2 \\
&\quad + [a_1(p_1, v_1)p_1^* + 2a_3(p_1, v_1)(1-p_1^*)]G_1F_1 + [a_1(p_2, v_2)p_2^* + 2a_3(p_2, v_2)(1-p_2^*)]G_2F_2 \\
&\quad + a_3(p_1, v_1)(p_1^*)^2F_1^2 + a_3(p_2, v_2)(p_2^*)^2F_2^2 \\
&\quad + [b_1(p_1, v_1, p_2, v_2)(1-p_2^*) + b_1(p_2, v_2, p_1, v_1)(1-p_1^*) + b_3(p_1, v_1, p_2, v_2)(1-p_1^*)(1-p_2^*)]G_1G_2 \\
&\quad + b_3(p_1, v_1, p_2, v_2)p_2^*(1-p_1^*)G_1F_2 + b_3(p_1, v_1, p_2, v_2)p_1^*(1-p_2^*)F_1G_2 \\
&\quad + b_3(p_1, v_1, p_2, v_2)p_1^*p_2^*F_1F_2. \tag{48}
\end{aligned}$$

663 This decomposition relates to condition **(II)** insuring the positive definiteness of  $\ddot{d}$  under both  $H_0$   
664 or  $H_1$ .

## 665 D.2 Matrices consistent estimation

666 The notations used here are taken from the notations introduced throughout the paper. All the  
667 matrices involved in our theoretical results can be estimated consistently as follows.

$$\begin{aligned} \widehat{J}_{1,1}(\theta^c) &= M_{1,1}(\widehat{S}, \widehat{\theta}_n), \widehat{J}_{1,2}(\theta^c) = M_{1,2}(\widehat{S}, \widehat{\theta}_n) = \widehat{J}_{2,1}(\theta^c), \widehat{J}_{2,2}(\theta^c) = M_{2,2}(\widehat{S}, \widehat{\theta}_n) \\ \widehat{L}_{3,1}(\theta^c) &= \mathcal{I}_1(\widehat{\theta}), \widehat{L}_{3,2}(\theta^c) = -\mathcal{I}_2(\widehat{\theta}), \widehat{L}_{3,3}(\theta^c) = \frac{1}{\widehat{p}_1}, \widehat{L}_{3,4}(\theta^c) = \mathcal{I}_1(\widehat{\theta}), \widehat{L}_{3,5}(\theta^c) = -\mathcal{I}_2(\widehat{\theta}), \widehat{L}_{3,6}(\theta^c) = \frac{1}{\widehat{p}_2}, \end{aligned}$$

668 and

$$\begin{aligned} \widehat{\sigma}_1(1, 1) &= \int_{\mathbb{R}^2} \Psi_{1,1}(S, \widehat{\theta}, x) \Psi_{1,1}(S, \widehat{\theta}, y) \widehat{\rho}_1(x, y) dU(x) dU(y) \\ \widehat{\sigma}_1(1, 2) &= \widehat{\sigma}_1(2, 1) = \int_{\mathbb{R}^2} \Psi_{1,1}(S, \widehat{\theta}, x) \Psi_{2,2}(S, \widehat{\theta}, y) \widehat{\rho}_1(x, y) dU(x) dU(y) \\ \widehat{\sigma}_1(1, 3) &= \widehat{\sigma}_1(3, 1) = \int_{\mathbb{R}^2} \Psi_{1,1}(S, \widehat{\theta}, x) D(S, \widehat{\theta}, y) \widehat{\rho}_1(x, y) dU(x) dU(y) \\ \widehat{\sigma}_1(2, 2) &= \int_{\mathbb{R}^2} \Psi_{2,2}(S, \widehat{\theta}, x) \Psi_{2,2}(S, \widehat{\theta}, y) \widehat{\rho}_1(x, y) dU(x) dU(y) \\ \widehat{\sigma}_1(2, 3) &= \widehat{\sigma}_1(3, 2) = \int_{\mathbb{R}^2} \Psi_{2,2}(S, \widehat{\theta}, x) D(S, \widehat{\theta}, y) \widehat{\rho}_1(x, y) dU(x) dU(y) \\ \widehat{\sigma}_1(3, 3) &= \int_{\mathbb{R}^2} D(S, \widehat{\theta}, x) D(S, \widehat{\theta}, y) \widehat{\rho}_1(x, y) dU(x) dU(y), \end{aligned}$$

669 and  $\Sigma_2 = (\sigma_2(i, j))_{1 \leq i, j \leq 3}$  where

$$\begin{aligned} \widehat{\sigma}_2(1, 1) &= \int_{\mathbb{R}^2} \Psi_{2,1}(S, \widehat{\theta}, x) \Psi_{2,1}(S, \widehat{\theta}, y) \widehat{\rho}_2(x, y) dU(x) dU(y) \\ \widehat{\sigma}_2(1, 2) &= \widehat{\sigma}_1(2, 1) = \int_{\mathbb{R}^2} \Psi_{2,1}(S, \widehat{\theta}, x) \Psi_{1,2}(S, \widehat{\theta}, y) \widehat{\rho}_2(x, y) dU(x) dU(y) \\ \widehat{\sigma}_2(1, 3) &= \widehat{\sigma}_1(3, 1) = \int_{\mathbb{R}^2} \Psi_{2,1}(S, \widehat{\theta}, x) D(S, \widehat{\theta}, y) \widehat{\rho}_2(x, y) dU(x) dU(y) \\ \widehat{\sigma}_2(2, 2) &= \int_{\mathbb{R}^2} \Psi_{1,2}(S, \widehat{\theta}, x) \Psi_{1,2}(S, \widehat{\theta}, y) \widehat{\rho}_2(x, y) dU(x) dU(y) \\ \widehat{\sigma}_2(2, 3) &= \widehat{\sigma}_1(3, 2) = \int_{\mathbb{R}^2} \Psi_{1,2}(S, \widehat{\theta}, x) D(S, \widehat{\theta}, y) \widehat{\rho}_2(x, y) dU(x) dU(y) \\ \widehat{\sigma}_2(3, 3) &= \int_{\mathbb{R}^2} D(S, \widehat{\theta}, x) D(S, \widehat{\theta}, y) \widehat{\rho}_2(x, y) dU(x) dU(y), \end{aligned}$$

670 where  $\widehat{\rho}_1(x, y) = \widehat{L}_1(x \wedge y)(1 - \widehat{L}_1(x \vee y))$  and  $\widehat{\rho}_2(x, y) = \widehat{L}_2(x \wedge y)(1 - \widehat{L}_2(x \vee y))$ . The above  
671 quantities are consistent estimates of their closed form counterparts given in the proof of Theorem  
672 4. The proof of consistency for each of these terms is lengthy but straightforward. They basically  
673 use the dominated convergence theorem noticing that for all  $(i, j) \in \{1, 2\}^2$  and  $(x, y)$  fixed in  $\mathbb{R}^2$ ,  
674 the terms  $\Psi_{i,j}(S, \widehat{\theta}, x)$ ,  $D(S, \widehat{\theta}, y)$ , are bounded and almost surely consistent estimates  $\Psi_{i,j}(S, \theta^c, x)$ ,  
675  $D(S, \theta^c, x)$  according to Theorem 1, and  $\widehat{\rho}_i(x, y)$  is an almost surely uniformly consistent estimate  
676 of  $\widehat{\rho}_i(x, y)$  due to the Glivenko-Cantelli theorem.

## 677 E Special case: $G_1 = G_2$

678 In this section we propose to investigate how the case  $G_1 = G_2$ , previously analyzed in Milhaud  
 679 et al. [20] (under the  $F_i$ 's symmetry assumption), can be handled simply using the IBM-method  
 680 described in Sections 1 and 2. First, the basic identifiability conditions in this particular setup are  
 681 slightly simplified but also weakened. In fact, rewriting conditions (11) and (12) when  $G_1 = G_2 = G$   
 682 leads to

$$F_1(x, L_1, p_1) = \frac{L_1(x) - (1 - p_1)G(x)}{p_1} \quad \text{and} \quad F_2(x, L_2, p_2) = \frac{L_2(x) - (1 - p_2)G(x)}{p_2}. \quad (49)$$

683 Let us investigate now the situations where possibly  $F_1(x, L_1, p_1) = F_2(x, L_2, p_2)$ .

$$\begin{aligned} F_1(x, L_1, p_1) = F_2(x, L_2, p_2) &\Leftrightarrow \frac{L_1(x) - (1 - p_1)G_1(x)}{p_1} = \frac{L_2(x) - (1 - p_2)G_2(x)}{p_2} \\ &\Leftrightarrow \frac{(p_1 - p_1^*)G(x) + p_1^*F(x)}{p_1} = \frac{(p_2 - p_2^*)G(x) + p_2^*F(x)}{p_2} \\ &\Leftrightarrow \left( \frac{p_1 - p_1^*}{p_1} - \frac{p_2 - p_2^*}{p_2} \right) G(x) = \frac{p_2^*}{p_2} F_2(x) - \frac{p_1^*}{p_1} F_1(x). \end{aligned} \quad (50)$$

Under  $H_0$ ,  $F_1 = F_2 = F$ , we simply obtain

$$\left( \frac{p_1 - p_1^*}{p_1} - \frac{p_2 - p_2^*}{p_2} \right) G(x) = \left( \frac{p_2^*}{p_2} - \frac{p_1^*}{p_1} \right) F(x).$$

If  $\{G, F\}$  is free ( $G \neq F$  on non-null measure set) we necessarily have

$$\frac{p_2^*}{p_2} - \frac{p_1^*}{p_1} = 0 \Leftrightarrow \frac{p_1}{p_2} = \frac{p_1^*}{p_2^*},$$

684 which means that we have an infinite number of solutions. Actually this conclusion is not an  
 685 obstacle to our testing approach. In fact it is enough to arbitrarily fix a value for  $p_1$ , let say  $\check{p}_1$   
 686 in  $]\delta, 1[$ ,  $0 < \delta < 1$ , and investigate  $p_2$  such that  $p_2 = (p_2^*\check{p}_1)/p_1^*$  which will automatically become  
 687 the only parameter  $\lambda = p_2$  under the  $\check{p}_1$ -setup, not necessarily belonging to  $[\delta, 1]$  but probably to  
 688 a larger interval  $\Theta' = [\delta, A]$  for  $A > \delta$  large enough, such that  $F_1(\cdot, L_1, \check{p}_1) = F_2(\cdot, L_2, \lambda)$   $\mu - a.e.$

689 Under  $H_1$ ,  $F_1 \neq F_2$ , if we suppose that  $\{G, F_1, F_2\}$  is free, condition (50) leads to  $p_1^* = p_2^* = 0$  which  
 690 is impossible. Given the above identifiability discussion, the natural extension of our method is to  
 691 consider a one single parameter fitting approach.

692 In order to consistently pick the right  $\lambda^* = (p_2^*\check{p}_1)/p_1^*$  under  $H_0$  and select under  $H_1$  a  $\lambda$  such  
 693 that  $F_1(x, \check{p}_1) \neq F_2(x, \lambda)$ , we propose to investigate the location of the minimum  $(\mathcal{F}_1, \mathcal{F}_2)$ -distance  
 694 parameter  $\lambda^c$  such that

$$\lambda^c = \arg \min_{\lambda \in [\delta, A]} d(\lambda), \quad \text{where} \quad d(\lambda) = \int_{\mathbb{R}} D^2(x, L_1, L_2, \lambda) dU(x) \quad (51)$$

695 with

$$D(x, L_1, L_2, \lambda) = F_1(x, L_1, \check{p}_1) - F_2(x, L_2, \lambda).$$

696 Similarly to Section 2, we propose the following natural estimator of  $\lambda^c$  in (51), *i.e.*:

$$\widehat{\lambda}_n = \arg \min_{\lambda \in ]\delta, A]} d_n(\lambda), \quad \text{where} \quad d_n(\lambda) = \int_{\mathbb{R}} D^2(x, \widehat{L}_1, \widehat{L}_2, \lambda) dU(x). \quad (52)$$

697 According to the above identifiability discussion, Lemma 1 and Theorem 1 in Section 3 are still  
 698 valid and we can establish the counterpart of Theorem 2. To answer the testing problem in the  
 699 case  $G_1 = G_2$ , we would thus use the following result.

700

701 **(I')** The cdfs family  $\mathcal{E}_1 = \{G, F_1, F_2\}$  is linearly independent.

702

703 **(II')** The family of functions  $\mathcal{E}_2 = \{G^2, F_1 F_2, F_i^2, G F_i; i = 1, 2\}$  is linearly independent.

704

705 **(A')**  $\lambda^*$  under  $H_0$  ( $\lambda^c = \theta^*$ ), or  $\lambda^c$  under  $H_1$  ( $\lambda^c \neq \lambda^*$ ), belong to  $] \delta, A[$  the interior of the compact  
 706 parametric space  $\Theta'$ .

707

708 **Theorem 4.** *i) If conditions (I'), (II') and (A') hold, we have under both  $H_0$  or  $H_1$ :*

$$\sqrt{n} \begin{bmatrix} \widehat{\lambda} - \lambda^c \\ D_n(\cdot) - D(\cdot) \end{bmatrix} \xrightarrow{\mathcal{L}} W(\lambda^c, \cdot), \quad \text{as } n \rightarrow +\infty, \quad (53)$$

709 where  $W(\lambda^c, \cdot) = (W_1(\lambda^c), W_2(\lambda^c))^T$  is a 2-dimensional centered Gaussian process with covariance  
 710 matrix  $\Sigma_W = M(\lambda^c, \cdot) \Sigma_L(\cdot, \cdot) M(\lambda^c, \cdot)^T$  where  $M(\lambda^c, \cdot)$  is defined in (63) and  $\Sigma_L(\cdot, \cdot)$  in (66).

711 *ii) Under assumptions (I) and (II) we have respectively as  $n \rightarrow +\infty$ :*

$$n d_n(\widehat{\lambda}_n) = \mathbf{U}_n^0 \xrightarrow{\mathcal{L}} \int_{\mathbb{R}} (W_2(\lambda^*, x))^2 dU(x), \quad \text{under } H_0 \quad (54)$$

$$n d_n(\widehat{\lambda}_n) = \mathbf{U}_n^1 + \mathbf{V}_n^1, \quad \text{with } \mathbf{U}_n^1 \xrightarrow{\mathcal{L}} \int_{\mathbb{R}} (W_2(\lambda^c, x))^2 dU(x) \text{ and } \mathbf{V}_n^1 = O_P(n), \quad \text{under } H_1. \quad (55)$$

712 **Remark 5.** *It is important to notice that if we have at our disposal a consistent estimate of the*  
 713 *proportion  $p_1$ , let say by using  $\widehat{p}_1$  the  $\sqrt{n}$ -consistent estimator of Bordes et al. [4] in the symmetric*  
 714 *case or the  $\widehat{p}_1$  estimator of Patra and Sen [22] in the general case, we can use it to calibrate a*  
 715 *consistent value for  $\check{p}_1$  and finally obtain a consistent couple  $(\check{p}_1, \widehat{\lambda}_n) \rightarrow (p_1^*, p_2^*)$ , with according*  
 716 *rates of convergence under  $H_0$ .*

717 *Proof.* i) By a Taylor expansion of  $\dot{d}_n$  around  $\lambda^c$  we have

$$\ddot{d}_n(\widetilde{\lambda}_n) \sqrt{n} (\widehat{\lambda}_n - \lambda^c) = -\sqrt{n} \dot{d}_n(\lambda^c), \quad (56)$$

718 where  $\widetilde{\lambda}_n$  lies in the line segment with extremities  $\widehat{\lambda}_n$  and  $\lambda^c$ . In particular, we have

$$\dot{d}(\lambda) = 2E_U \left( \frac{(2-\lambda)GL_2}{\lambda^3} - \frac{(1-\lambda)G^2}{\lambda^3} - \frac{L_2^2}{\lambda^3} - \frac{GL_1}{\lambda^2 \check{p}_1} + \frac{(1-\check{p}_1)G^2}{\lambda^2 \check{p}_1} + \frac{L_2 L_1}{\lambda^2 \check{p}_1} - \frac{(1-\check{p}_1)L_2 G}{\lambda^2 \check{p}_1} \right),$$

719 and

$$\dot{d}_n(\lambda) - \dot{d}(\lambda) = 2 \left( \frac{2-\lambda}{\lambda^3} T_1 - \frac{1-\lambda}{\lambda^3} T_2 - \frac{1}{\lambda^3} T_3 - \frac{1}{\lambda^2 \check{p}_1} T_4 + \frac{1-\check{p}_1}{\lambda^2 \check{p}_1} T_5 + \frac{1}{\lambda^2 \check{p}_1} T_6 - \frac{1-\check{p}_1}{\lambda^2 \check{p}_1} T_7 \right).$$

720 where

$$\begin{aligned} T_1(G, L_2) &= E_U \left( G(\widehat{L}_2 - L_2) \right) \\ T_2(G) &= E_U(G^2) - E_U(G)^2 = 0 \\ T_3(L_2) &= E_U \left( (\widehat{L}_2 - L_2)(\widehat{L}_2 + L_2) \right) = E_U \left( (\widehat{L}_2 - L_2)(2L_2 + o_{a.s.}(1)) \right) \\ T_4(L_1) &= E_U \left( G(\widehat{L}_1 - L_1) \right) \\ T_5(G) &= E_U(G^2) - E_U(G)^2 = 0 \\ T_6(L_1, L_2) &= E_U \left( \widehat{L}_2 \widehat{L}_1 - L_2 L_1 \right) \\ T_7(G, L_2) &= E_U \left( G(\widehat{L}_2 - L_2) \right). \end{aligned}$$

721 We can also denote by  $\mathcal{S} = (G, L_1, L_2)$  and consider

$$\Psi_1(\mathcal{S}, \lambda) = 2 \left( \frac{2-\lambda}{\lambda^3} G - \frac{2}{\lambda^3} L_2 + \frac{1}{\lambda^2 \check{p}_1} L_1 - \frac{1-\check{p}_1}{\lambda^2 \check{p}_1} G \right), \quad \Psi_2(\mathcal{S}, \lambda) = 2 \left( \frac{1}{\lambda^2 \check{p}_1} (L_2 - G) \right),$$

722 along with

$$\Phi_1(\mathbb{L}_2, \lambda) = \varphi(\Psi_1(\mathcal{S}, \lambda), \mathbb{L}_2), \quad \Phi_2(\mathbb{L}_1, \lambda) = \varphi(\Psi_2(\mathcal{S}, \lambda), \mathbb{L}_1).$$

723 Note that the first, respectively the second, above expression depends only on  $\mathbb{L}_2$ , resp.  $\mathbb{L}_1$ .

$$\sqrt{n}(\dot{d}_n(\lambda) - \dot{d}(\lambda)) = \Phi(\mathbb{L}_1, \mathbb{L}_2, \lambda) + o_{a.s.}(1), \quad (57)$$

724 where according to  $\sqrt{n} = \sqrt{\kappa n} / \sqrt{\kappa} = \zeta \sqrt{n_2}$  with  $\zeta = 1/\sqrt{\kappa}$ :

$$\Phi(\mathbb{L}_2, \mathbb{L}_1, \lambda) = \Phi_2(\mathbb{L}_1, \lambda) + \zeta \Phi_1(\mathbb{L}_2, \lambda). \quad (58)$$

725 Since the empirical processes  $\mathbb{L}_1$  and  $\mathbb{L}_2$  are independent, by the Donsker theorem:

$$\begin{bmatrix} \mathbb{L}_1 \\ \mathbb{L}_2 \end{bmatrix} \rightsquigarrow \mathcal{B} = \begin{bmatrix} \mathcal{B}_1 \\ \mathcal{B}_2 \end{bmatrix}, \quad (59)$$

726 where  $\mathcal{B}$  is a bi-dimensional gaussian process with diagonal correlation matrix  $\rho = \text{diag}(\rho_1, \rho_2)$   
727 with  $\rho_1(x, y) = L_1(x \wedge y)(1 - L_1(x \vee y))$  and  $\rho_2(x, y) = L_2(x \wedge y)(1 - L_2(x \vee y))$ . Moreover,

$$\begin{aligned} \sqrt{n}[D_n(x, \widehat{L}_1, \widehat{L}_2, \widehat{\lambda}_n) - D(x, L_1, L_2, \lambda^c)] &= \sqrt{n}[F_1(x, \widehat{L}_1, \check{p}_1) - F_1(x, L_1, \check{p}_1)] \\ &\quad - \sqrt{n}[(F_2(x, \widehat{L}_2, \widehat{\lambda}) - F_2(x, L_2, \lambda))]. \end{aligned} \quad (60)$$

728 Let analyze separately the two terms in the right hand side of the above equality.

$$\sqrt{n}[F_1(\cdot, \widehat{L}_1, \check{p}_1) - F_1(\cdot, L_1, \check{p}_1)] = \frac{1}{\check{p}_1} \mathbb{L}_1, \quad (61)$$

729 and

$$\begin{aligned} \sqrt{n}[F_2(\cdot, \widehat{L}_2, \widehat{\lambda}) - F_2(\cdot, L_2, \lambda^c)] &= \sqrt{n} \left[ \left( \frac{\widehat{L}_2}{\widehat{\lambda}} - \frac{L_2}{\lambda^c} \right) - \left( \frac{1 - \widehat{\lambda}}{\widehat{\lambda}} - \frac{1 - \lambda^c}{\lambda^c} \right) G \right] \\ &= \sqrt{n} \left[ \frac{\widehat{L}_2 - L_2}{\widehat{\lambda}} \right] + \sqrt{n} \left( \frac{\lambda^c - \widehat{\lambda}}{\lambda^c \widehat{\lambda}} \right) (L_2 - G) \\ &= \frac{\zeta}{\lambda^c} \mathbb{L}_2 - \left( \frac{L_2 - G}{(\lambda^c)^2} \sqrt{n} [\widehat{\lambda} - \lambda^c] \right) + o_P(1). \end{aligned} \quad (62)$$

730 It is also easy to prove, similarly to the proof of Theorem 2, that  $\ddot{d}_n(\widetilde{\lambda}_n) \xrightarrow{a.s.} \ddot{d}(\lambda^c) > 0$ , as  $n \rightarrow +\infty$ .

731 In order to synthetically summarize results (56), (57), (58) and (60–62) and build up the Central  
732 Limit Theorem relative to our quantities of interest, we define the following matrix-type relation:

$$\sqrt{n} \begin{bmatrix} \widehat{\lambda} - \lambda^c \\ D_n(\cdot) - D(\cdot) \end{bmatrix} = M(\lambda^c, \cdot) \begin{bmatrix} \Phi_2(\mathbb{L}_1, \lambda^c) \\ \mathbb{L}_1 \\ \Phi_1(\mathbb{L}_2, \lambda^c) \\ \mathbb{L}_2 \end{bmatrix} + o_{a.s.}(1), \quad \text{with } M(\lambda^c, \cdot) = L(\cdot, \lambda^c) J^{-1}(\lambda^c) \mathcal{C}, \quad (63)$$

733 where

$$\mathcal{C} = \begin{bmatrix} -1 & 0 & -\zeta & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad J(\lambda) = \begin{bmatrix} \ddot{d}(\lambda) & 0_{1 \times 2} \\ 0_{2 \times 1} & \text{Id}_{2 \times 2} \end{bmatrix}, \quad L(\cdot, \lambda) = \begin{bmatrix} 1 & 0 & 0 \\ \frac{L_2(\cdot) - G(\cdot)}{\lambda^2} & \frac{1}{\check{p}_1} & -\frac{\zeta}{\lambda} \end{bmatrix}. \quad (64)$$

734 Finally, we have

$$\begin{bmatrix} \Phi_2(\mathbb{L}_1, \lambda^c) \\ \mathbb{L}_1 \\ \Phi_1(\mathbb{L}_2, \lambda^c) \\ \mathbb{L}_2 \end{bmatrix} \xrightarrow{\mathcal{L}} Z = \begin{bmatrix} \Phi_2(\mathcal{B}_1, \lambda^c) \\ \mathcal{B}_1 \\ \Phi_1(\mathcal{B}_2, \lambda^c) \\ \mathcal{B}_2 \end{bmatrix}, \quad (65)$$

735 where  $Z$  is Gaussian random vector of  $\mathbb{R}^4$  with covariance matrix  $\Sigma_L = E(ZZ^T)$ .

736 Since  $\mathcal{B}_1$  and  $\mathcal{B}_2$  are two independent (limit) Gaussian processes we have

$$\Sigma_L(x, y) = \begin{bmatrix} \Sigma_1(x, y) & 0_{2 \times 2} \\ 0_{2 \times 2} & \Sigma_2(x, y) \end{bmatrix}, \quad (66)$$

737 where  $\Sigma_1(x, y) = (\sigma_1(i, j; x, y))_{1 \leq i, j \leq 3}$  with

$$\begin{aligned} \sigma_1(1, 1; x, y) &= \sigma_1(1, 1) = \int_{\mathbb{R}^2} \Psi_2(S, \theta, u) \Psi_2(S, \theta, v) \rho_1(u, v) dU(u) dU(v) \\ \sigma_1(1, 2; x, y) &= \sigma_1(1, 2; y) = \int_{\mathbb{R}} \Psi_2(S, \theta, u) \rho_1(u, y) dU(u) \\ \sigma_1(2, 2; x, y) &= \rho_1(x, y) \\ \sigma_1(2, 1; x, y) &= \sigma_1(2, 1; x) = \sigma_1(1, 2; x) \end{aligned}$$

738 and  $\Sigma_2 = (\sigma_2(i, j; x, y))_{1 \leq i, j \leq 3}$  with

$$\begin{aligned}\sigma_2(1, 1; x, y) &= \sigma_2(1, 1) = \int_{\mathbb{R}^2} \Psi_1(S, \theta, u) \Psi_1(S, \theta, v) \rho_2(u, v) dU(u) dU(v) \\ \sigma_2(1, 2; x, y) &= \sigma_1(1, 3; y) = \int_{\mathbb{R}} \Psi_1(S, \theta, u) \rho_2(u, y) dU(u) \\ \sigma_2(2, 2; x, y) &= \rho_2(x, y) \\ \sigma_2(2, 1; x, y) &= \sigma_2(2, 1; x) = \sigma_2(1, 2; x).\end{aligned}$$

739 ii) Let us now decompose  $nd_n(\widehat{\lambda}_n)$ .

$$\begin{aligned}nd_n(\widehat{\lambda}_n) &= \int_{\mathbb{R}} n(D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\lambda}_n) - D(x, L_1, L_2, \lambda^c) + D(x, L_1, L_2, \lambda^c))^2 dU(x) \\ &= \int_{\mathbb{R}} (\sqrt{n}[D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\lambda}_n) - D(x, L_1, L_2, \lambda^c)])^2 dU(x) \\ &\quad + 2\sqrt{n} \int_{\mathbb{R}} \sqrt{n}[D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\lambda}_n) - D(x, L_1, L_2, \lambda^c)] D(x, L_1, L_2, \lambda^c) dU(x) \\ &\quad + n \int_{\mathbb{R}} D^2(x, L_1, L_2, \lambda^c) dU(x).\end{aligned}\tag{67}$$

740 Note that under  $H_0$ ,  $\lambda^c = \lambda^*$ , and we simply get:

$$\begin{aligned}nd_n(\widehat{\lambda}_n) &= \int_{\mathbb{R}} (\sqrt{n}D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\lambda}_n))^2 dU(x) \\ &= \int_{\mathbb{R}} (\sqrt{n}[D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\lambda}_n) - D(x, L_1, L_2, \lambda^*)])^2 dU(x) \\ &= \mathbf{U}_n^0.\end{aligned}\tag{68}$$

741 since  $D(\cdot, L_1, L_2, \lambda^*) = 0$  almost everywhere. Next, notice that it is easy to show that the map  
742  $\mathbf{U}_n(\mathbb{D})$  is Hadamard differentiable from the domain of càd-làg functions of bounded variation into  
743  $\mathbb{R}$  [27, Lemma 20.10]. This combined with the result for weak convergence of the empirical process  
744  $\mathbb{D}$ , in the second row of (53), yields the desired result.

745 Under  $H_1$ , we have  $\int_{\mathbb{R}} D^2(x, L_1, L_2, \lambda^c) dU(x) > 0$ , which leads to

$$\begin{aligned}nd_n(\widehat{\lambda}_n) &= \int_{\mathbb{R}} (\sqrt{n}[D(x, \widehat{L}_1, \widehat{L}_2, \widehat{\lambda}_n) - D(x, L_1, L_2, \lambda^c)])^2 dU(x) \\ &\quad + n \int_{\mathbb{R}} D^2(x, L_1, L_2, \lambda^c) dU(x) + o_P(n) \\ &= \mathbf{U}_n^1 + \mathbf{V}_n^1.\end{aligned}\tag{69}$$

746 Given the asymptotic convergence analysis under both  $H_0$  or  $H_1$ , the random variable within  
747 brackets involved commonly in (67) and (69) can once again be analyzed closely, leading to answer  
748 our statistical test.  $\square$

## 749 F Monte Carlo parametric setups

750 Notation:  $\mathcal{N}(a, b)$ : Gaussian with mean  $a$ , and variance  $b > 0$ ;  $\mathcal{G}(a, b)$ : gamma with shape  $a > 0$ ,  
 751 and scale  $b > 0$ ;  $\mathcal{E}(a)$ : exponential with rate  $a > 0$ ;  $\mathcal{Beta}(a, b)$ : beta with shape parameters  $a > 0$ ,  
 752  $b > 0$ ;  $\mathcal{NB}(a)$ : negative binomial with mean  $a$ , and convolution parameter  $b > 0$ ;  $\mathcal{P}(a)$ : Poisson  
 753 with mean  $a > 0$ ;  $\mathcal{U}(a, b)$  uniform on  $(a, b)$ .

### 754 F.1 Empirical levels - Study under $H_0$

755 We give here an illustration of the different frameworks studied in our simulations to assess the  
 756 empirical level of our test. The component weights have systematically been fixed to 50% for  
 757 illustration purpose, see Figure 9. Table 2 provides the corresponding parameters for each mixture  
 component, also used within the assessment of empirical levels in our global simulation study.

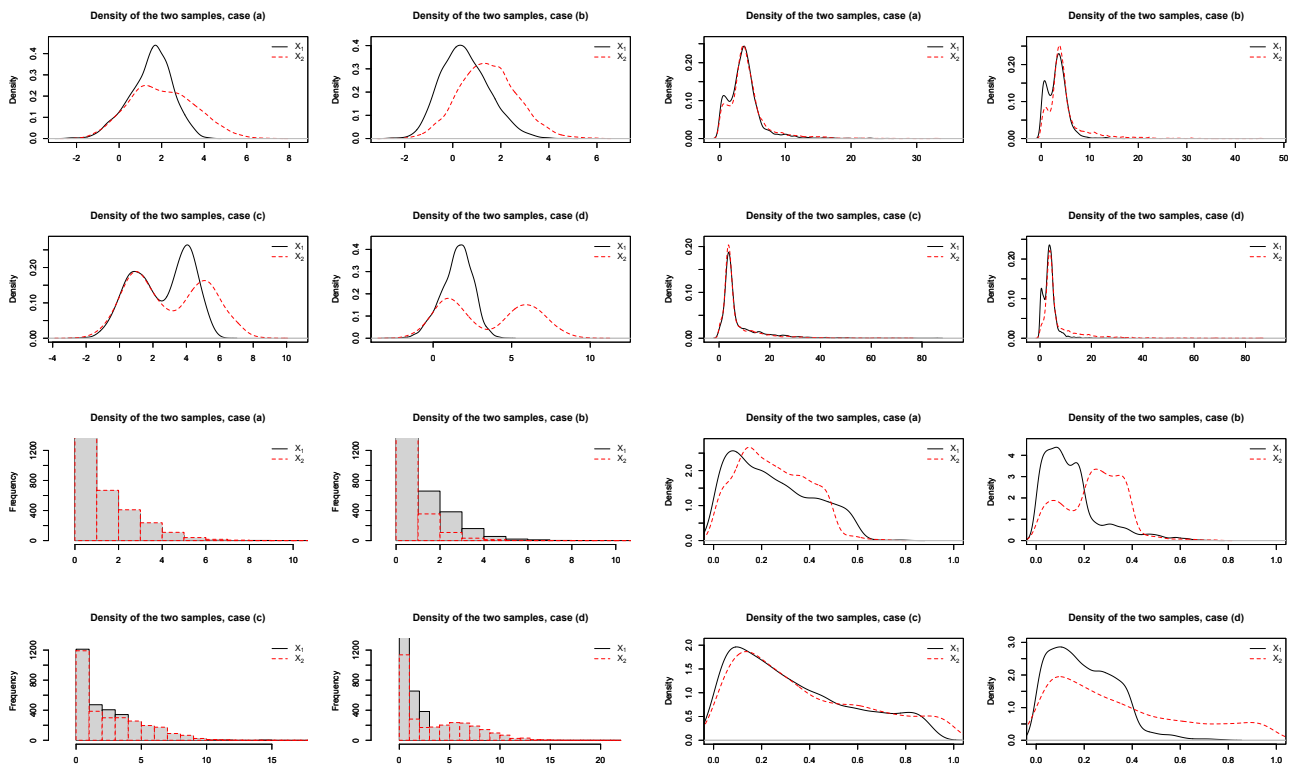


Figure 9: Densities of mixture distributions in the four frameworks under study (on each support:  $\mathbb{R}$ ,  $\mathbb{R}^+$ ,  $\mathbb{N}$ ,  $[0, 1]$  from top left to bottom right by series of 4 plots). Case a):  $G_1$  close to  $G_2$  with  $G_1$  and  $G_2$  close to  $F_1$  (or  $F_2$ ); Case b):  $G_1$  far away from  $G_2$ , with  $G_1$  and  $G_2$  close to  $F_1$ ; Case c):  $G_1$  close to  $G_2$  with  $G_1$  and  $G_2$  far from  $F_1$ ; Case d):  $G_1$  far from  $G_2$  with  $G_1$  close to  $F_1$ .

758



Table 2: Parameters of mixture components corresponding to Fig. 9, used for the assessment of empirical levels of the test. Recall that we are under  $H_0$ , meaning that  $F_1 = F_2 = F$ .

	On $\mathbb{R}$				On $\mathbb{R}^+$			
	Case (a)	Case (b)	Case (c)	Case (d)	Case (a)	Case (b)	Case (c)	Case (d)
$F$	$\mathcal{N}(1, 1)$	$\mathcal{N}(1, 1)$	$\mathcal{N}(1, 1)$	$\mathcal{N}(1, 1)$	$\mathcal{G}(16, 4)$	$\mathcal{G}(16, 4)$	$\mathcal{G}(16, 4)$	$\mathcal{G}(16, 4)$
$G_1$	$\mathcal{N}(2, 0.7)$	$\mathcal{N}(0, 0.7)$	$\mathcal{N}(4, 0.7)$	$\mathcal{N}(2, 0.7)$	$\mathcal{E}(1/3.5)$	$\mathcal{E}(1/2.5)$	$\mathcal{E}(1/12)$	$\mathcal{E}(1/3.5)$
$G_2$	$\mathcal{N}(3, 1.2)$	$\mathcal{N}(2, 1.2)$	$\mathcal{N}(5, 1.2)$	$\mathcal{N}(6, 1.2)$	$\mathcal{E}(1/4.5)$	$\mathcal{E}(1/5.5)$	$\mathcal{E}(1/10)$	$\mathcal{E}(1/10)$
	On $\mathbb{N}$				On $[0, 1]$			
	Case (a)	Case (b)	Case (c)	Case (d)	Case (a)	Case (b)	Case (c)	Case (d)
$F$	$\mathcal{NB}(1, 10)$	$\mathcal{NB}(1, 10)$	$\mathcal{NB}(1, 10)$	$\mathcal{NB}(1, 10)$	$Beta(1.2, 5)$	$Beta(1.2, 5)$	$Beta(1.2, 5)$	$Beta(1.2, 5)$
$G_1$	$\mathcal{P}(2)$	$\mathcal{P}(2)$	$\mathcal{P}(4)$	$\mathcal{P}(2)$	$\mathcal{U}(0, 0.6)$	$\mathcal{U}(0, 0.2)$	$\mathcal{U}(0, 0.9)$	$\mathcal{U}(0, 0.4)$
$G_2$	$\mathcal{P}(2.5)$	$\mathcal{P}(0.5)$	$\mathcal{P}(5)$	$\mathcal{P}(7)$	$\mathcal{U}(0.1, 0.5)$	$\mathcal{U}(0.2, 0.4)$	$\mathcal{U}(0.1, 1)$	$\mathcal{U}(0, 1)$
$p_1 = p_2$	50%	50%	50%	50%	50%	50%	50%	50%

## 759 F.2 Empirical powers - Study under $H_1$

760 We give here an illustration of the different frameworks studied in our simulations to assess the  
761 empirical power of our test. The component weights have systematically been fixed to 50% for  
762 illustration purpose, see Figure 10 and corresponding parameters in Table 3.

Table 3: Parameters corresponding to Fig. 10 and Fig. 11, used for the assessment of empirical powers. Case (a) : different means for  $F_1$  and  $F_2$  ; Case (b) : close means between  $F_1$  and  $F_2$  ; Case (c) : same means, different variances for  $F_1$  and  $F_2$  ; Case (d) : same means, close variances between  $F_1$  and  $F_2$  ; Case (e) : same means and variances for  $F_1$  and  $F_2$ , but different distributions.

	On $\mathbb{R}$					On $\mathbb{R}^+$				
	Case (a)	Case (b)	Case (c)	Case (d)	Case (e)	Case (a)	Case (b)	Case (c)	Case (d)	Case (e)
$F_1$	$\mathcal{N}(1, 1)$	$\mathcal{N}(1, 1)$	$\mathcal{N}(1, 1)$	$\mathcal{N}(1, 1)$	$\mathcal{N}(1, 1)$	$\mathcal{G}(16, 4)$	$\mathcal{G}(16, 4)$	$\mathcal{G}(16, 4)$	$\mathcal{G}(16, 4)$	$\mathcal{G}(1.47, 0.56)$
$F_2$	$\mathcal{N}(2.5, 1)$	$\mathcal{N}(1.5, 1)$	$\mathcal{N}(1, 1.5)$	$\mathcal{N}(1, 1.1)$	$\mathcal{L}(1, \sqrt{0.5})$	$\mathcal{G}(8, 4)$	$\mathcal{G}(14, 4)$	$\mathcal{G}(8, 2)$	$\mathcal{G}(22, 5.5)$	$Gomp(0.1, 0.3)$
$G_1$	$\mathcal{N}(2, 0.7)$	$\mathcal{N}(2, 0.7)$	$\mathcal{N}(2, 0.7)$	$\mathcal{N}(2, 0.7)$	$\mathcal{N}(2, 0.7)$	$\mathcal{E}(1/3)$	$\mathcal{E}(1/3)$	$\mathcal{E}(1/3)$	$\mathcal{E}(1/3)$	$\mathcal{E}(1)$
$G_2$	$\mathcal{N}(3, 1.2)$	$\mathcal{N}(3, 1.2)$	$\mathcal{N}(3, 1.2)$	$\mathcal{N}(3, 1.2)$	$\mathcal{N}(3, 1.2)$	$\mathcal{E}(1/6)$	$\mathcal{E}(1/6)$	$\mathcal{E}(1/6)$	$\mathcal{E}(1/6)$	$\mathcal{E}(1/8)$
	On $\mathbb{N}$					On $[0, 1]$				
	Case (a)	Case (b)	Case (c)	Case (d)	Case (e)	Case (a)	Case (b)	Case (c)	Case (d)	Case (e)
$F_1$	$\mathcal{NB}(1, 10)$	$\mathcal{NB}(1, 10)$	$\mathcal{NB}(1, 10)$	$\mathcal{NB}(1, 2)$	$\mathcal{NB}(3, 100)$	$Beta(1.2, 5)$	$Beta(1.2, 5)$	$Beta(1.2, 5)$	$Beta(1.2, 5)$	$Beta(5, 2)$
$F_2$	$\mathcal{NB}(4, 10)$	$\mathcal{NB}(2, 10)$	$\mathcal{NB}(1, 1)$	$\mathcal{NB}(1, 20)$	$\mathcal{B}(50, 0.6)$	$Beta(0.5, 5)$	$Beta(1, 5)$	$Beta(12, 50)$	$Beta(2.4, 10)$	$LogitN(0.9, 0.8)$
$G_1$	$\mathcal{P}(2)$	$\mathcal{P}(2)$	$\mathcal{P}(2)$	$\mathcal{P}(2)$	$\mathcal{P}(2)$	$\mathcal{U}(0, 0.8)$	$\mathcal{U}(0, 0.8)$	$\mathcal{U}(0, 0.8)$	$\mathcal{U}(0, 0.8)$	$\mathcal{U}(0, 0.8)$
$G_2$	$\mathcal{P}(6)$	$\mathcal{P}(6)$	$\mathcal{P}(6)$	$\mathcal{P}(6)$	$\mathcal{P}(6)$	$\mathcal{U}(0.2, 1)$	$\mathcal{U}(0.2, 1)$	$\mathcal{U}(0.2, 1)$	$\mathcal{U}(0.2, 1)$	$\mathcal{U}(0.2, 1)$
$p_1 = p_2$	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%

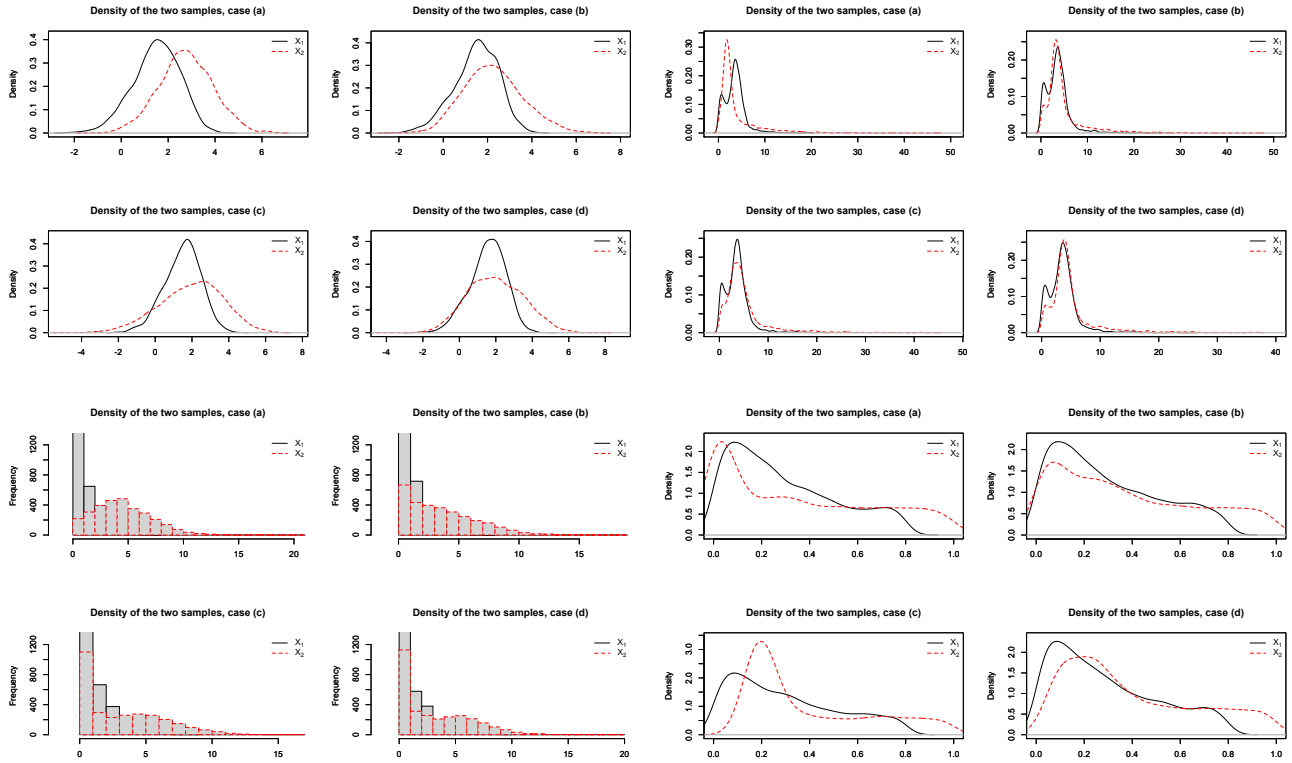


Figure 10: The four frameworks to assess the empirical power of the test depending on the support.

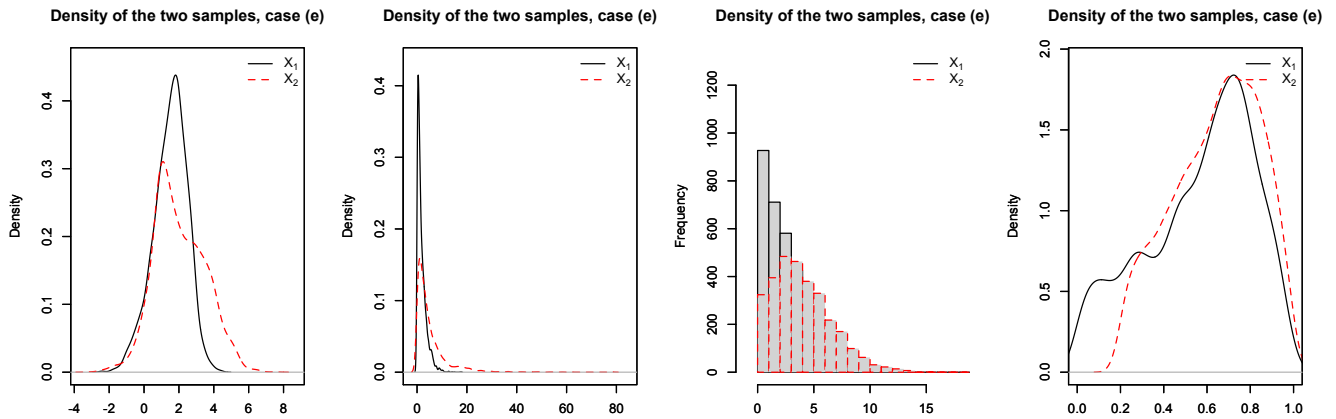


Figure 11: Cases where the distributions  $F_1$  and  $F_2$  have the same first two moments, for different supports  $(\mathbb{R}, \mathbb{R}^+, \mathbb{N}, [0, 1])$ . Parameters of such mixtures are summarized in Case e), see Table 3.