



**HAL**  
open science

# Weighted likelihood test for a change in one component of a parametric mixture

Dominique Abgrall

► **To cite this version:**

Dominique Abgrall. Weighted likelihood test for a change in one component of a parametric mixture. 2021. hal-03211913v2

**HAL Id: hal-03211913**

**<https://hal.science/hal-03211913v2>**

Preprint submitted on 30 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.

# Weighted likelihood test for a change in one component of a parametric mixture

Dominique Abgrall\*

April 30, 2021

## Abstract

In this article, we consider a closed sample of independent random variables that follow a finite mixture distribution with parametric components. The sample might contain at most one change in the parameters of the first component. If there is a change, the parameters which describe the distribution of the first component are different before and after the change-point while the other parameters of the mixture remain the same. To test whether there is a change or not, we introduce two alternative hypothesis tests. They are based on weighted likelihood ratios that can be computed with standard inference algorithms. With a technique from Davis et al. (1995), we derive the limit distribution of their statistics under the null hypothesis in the form of quadratic forms of a multidimensional Brownian motion, with the help of a dedicated functional limit theorem. We show that validity conditions of the main result hold for univariate finite Gaussian mixtures within the framework of Hathaway (1985). Numerical applications on simulated data illustrate the advantage of the alternative tests compared to a standard benchmark test. An application to Property and Casualty insurance real data is provided for the alternative tests.

**Keywords:** Change-point detection, parametric mixture, weak limit theorems for dependent sequences, weighted likelihood quotient test, applications to insurance data.

## 1 Introduction

On the one hand, finite parametric mixtures of distributions play a central role in applied statistics, as they allow to describe experiments with different sub-populations (McLachlan and Peel (2000), Lachos Dávila et al. (2018), Frühwirth-Schnatter et al. (2019) and the references therein). These distributions are often associated to questions about the identifiability of the mixture model; the estimation of the mixture parameters with e.g. with the EM algorithm by Dempster et al. (1977); the number of components in the mixture; etc.

On the other hand, the detection of at most one change-point in a closed sample is a standard problem for which techniques already exist, such as the ones in Csörgő and Horváth (1997)<sup>1</sup>. Applications are numerous in finance, climatology, oceanography and anywhere some monitoring is needed (Killick and Eckley (2014)). Dedicated detection techniques for mixtures are not so common because such models raise already so many difficulties when it comes to the inference of the parameters (including the number of components). To our knowledge, very few references seem to specifically detect change-points in mixtures with likelihood ratio-based techniques<sup>2</sup> (Pons (2009), Zou et al. (2015)). It turns out that general and standard change-point detection techniques can be adapted for finite parametric mixtures (Csörgő and Horváth (1997), van der Vaart (1998), Pons (2018)). However, when it comes to numerical applications, we observed that the standard approach raises many computational difficulties.

---

\*Laboratoire de Mathématiques de Bretagne Atlantique, Université de Bretagne Occidentale, dominique.abgrall@gmail.com.

<sup>1</sup>See also Barber et al. (2011), Chen and Gupta (2012) and the references therein.

<sup>2</sup>Some of the existing work is dedicated to a Bayesian framework and therefore not in the scope of this article. See e.g. Giordani and Kohn (2008), Pandya and Jadav (2009), Pandya and Jadav (2010), Wilson et al. (2013), Li et al. (2018) or Ganji and Mostafayi (2019).

In this article, we consider a sample of  $n$  independent random variables that follow a finite mixture distribution with parametric components. The sample might contain *at most one change* (AMOC) in the parameters of the first component. If there is a change, the r.v. are identically distributed before and after the change-point: the parameters which describe the distribution of the first component are different before and after the break-point while the other parameters of the mixture remain the same. For example, a shift occurs in the mean or in the standard deviation of the first component in the case of a Gaussian mixture. We want to test whether there is a change or not.

In order to circumvent the problems raised by the standard technique, we suggest a different approach that takes the form of a weighted likelihood test (WLT)<sup>3</sup>. In particular, the WL test can be computed using standard estimation algorithms. With a technique from Davis et al. (1995), we derive the limit distribution of its statistic under the null hypothesis in the form of a quadratic form of a multidimensional Brownian motion.

We start in Section 2 by the introduction of the model and the validity conditions required for our main result. In particular, we impose that the change-point cannot occur too close to the first nor the last observation of the closed sample. In addition, the Maximum Likelihood Estimator (MLE) for the parameters of the mixture has to be strongly consistent. As in Davis et al. (1995) and Csörgő and Horváth (1997), the test is based on a likelihood ratio. The main difference from the standard approach lies in the presence of weight functions that allow to focus on the first component of the mixture. In Section 3, under the null hypothesis, we first provide asymptotic properties of the MLE (Lehmann and Casella (1998)) before deriving a functional limit result for one term of the log-likelihood ratio in Theorem 3.7. This result is based on multiple applications of the Continuous Mapping Theorem and a Functional Delta Method in the Skorokhod metric space of càd-làg functions (Billingsley (1999), van der Vaart (1998)). In Theorem 3.8, the limit distribution of the test statistic is obtained as a consequence. In Section 4, we suggest an extension of the test (EWLT) where we scale the contribution of the sample to the weighted likelihood ratio. This improves significantly the detection frequency of the test in the case of a change (lower type II error). In Section 5, we show that validity conditions hold for univariate finite Gaussian mixtures within the framework of Hathaway (1985).

Applications in Section 6 consist in two parts. First, with numerical simulations, we illustrate the properties of the WL and EWL tests and compare them to a *benchmark* test (BM) obtained from an application of the standard test (e.g. exposed in Csörgő and Horváth (1997)). Both WL and EWL tests have notably lower type II errors, especially for large samples of over 10 000 observations. Overall, the EWLT performs significantly better than the other candidates. The second application is an illustration of the WL and EWL tests on Property and Casualty insurance real data. The tests are applied for the detection of a change in the variation over six months of the claim amount. For insurance problems, this application indicates that the proposed tests can be used for the monitoring of changes, when they are unexpected, and also to assess their significativity when they are known or suspected.

In Section 7, we give an overview of the conclusions and perspectives of this work.

Additional results, mathematical proofs and further illustrations of the numerical applications are given in the supplementary material.

## 2 Description of the model, assumptions and notations

### 2.1 Model and assumptions

We consider an experiment where we observe a sample of  $n$  independent continuous random variables  $X = (X_i)_{1 \leq i \leq n}$ , defined on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , with values in some set  $\mathcal{X}$ , subset of an Euclidean space, endowed with Lebesgue's measure. Each  $X_i$ ,  $1 \leq i \leq n$ , follows a finite mixture distribution with  $2 < m < \infty$  parametric components  $\mathbb{P}_{\theta^i}$ , where  $\theta^i$  belongs to a convex set of eligible parameters  $\Theta$ .

More precisely, for  $m$  a fixed, deterministic and known integer, the elements of  $\Theta$  are of following type:

---

<sup>3</sup>Weighted likelihood approaches are used in many contexts: see e.g. Dickey (1971), Hu and Zidek (2002), Amisano and Giacomini (2007), Basu et al. (2011), Song et al. (2018) and the references therein.

$\Theta \ni \theta = (p_1, \dots, p_{m-1}, \lambda_1, \dots, \lambda_m)$ , with  $(p_1, \dots, p_{m-1})$  in the open set

$$\Theta_0 := \left\{ (p_1, \dots, p_{m-1}) \in (0, 1)^{m-1}, \sum_{k=1}^{m-1} p_k < 1 \right\}$$

and, for each  $k \in \{1, \dots, m\}$ ,  $\lambda_k \in \Theta_k$ , with  $\Theta_k$  an open convex subset of some  $\mathbb{R}^{d_k}$ , with  $d_k \geq 1$ . Set  $d := m - 1 + \sum_{k=1}^m d_k$ : then  $\Theta = \Theta_0 \times \prod_{k=1}^m \Theta_k$  is an open convex subset of  $\mathbb{R}^d$ . Finally, given  $f_1, \dots, f_m$  some fixed density functions on  $\mathcal{X}$ , the distribution  $\mathbb{P}_\theta$ ,  $\theta \in \Theta$ , admits the density

$$f(x, \theta) := \sum_{k=1}^m p_k f_k(x, \lambda_k), \quad x \in \mathcal{X},$$

with  $p_m := 1 - \sum_{k=1}^{m-1} p_k$ . We first assume that the distributions  $\{\mathbb{P}_\theta, \theta \in \Theta\}$  are all distinct. This means in particular that the mixture should be identifiable<sup>4</sup>. We also add some usual assumptions on the regularity of the components of the mixture:

- ◇ The distributions defined by  $\{f(\cdot, \theta), \theta \in \Theta\}$  have common support, i.e. the set  $\{x \in \mathcal{X}, f(x, \theta) > 0\}$  does not depend on  $\theta$ ;
- ◇ For almost all  $x \in \mathcal{X}$ , the function  $\theta \mapsto f(x, \theta)$  is twice continuously differentiable in  $\theta \in \Theta$  with partial derivatives bounded by a non-negative integrable function of  $x$  that does not depend on  $\theta$ .

In this experiment, the sample is identically distributed before and after the change-point. There is *at most one change* (AMOC), deterministic but unknown, or none. Since we are interested in the limit behavior of the sample of variables when their number tends to infinity, we suppose that the experience takes place in the time interval  $[0, 1]$  and each variable  $X_i$  occurs at time  $i/n$ ,  $1 \leq i \leq n$ .

With this approach, if there is a change-point, it occurs at a time denoted by  $\mathbf{s} \in (0, 1)$  such that the parameters which describe the distribution of the first component are different before and after the break-point while the other parameters of the mixture remain the same. We write  $\boldsymbol{\theta} := \theta^1 = \dots = \theta^{\lfloor \mathbf{s}n \rfloor} \neq \theta^{\lfloor \mathbf{s}n \rfloor + 1} = \dots = \theta^n$ , where, for any  $x \in [0, 1]$ ,  $\lfloor x \rfloor$  denotes the integer part of  $x$ . Since it is not needed to define our test, refer to Section 6.1 for more details on the alternative hypothesis and how it can be described with the parameters  $p_1, \dots, p_{m-1}, \lambda_1, \dots, \lambda_m$ .

If there is no change-point, then  $X$  is an i.i.d. sample with  $\boldsymbol{\theta} = \theta^1 = \dots = \theta^n$ .

For the sake of simplicity, we impose that there is at most one change. A setting with more than one change-point can be extended with the same logic.

In the sequel we suppose that the following holds for the change-point  $\mathbf{s}$ :

**Assumption 2.1.** *If there is a change, the change-point  $\mathbf{s}$  is contained in  $[\bar{s}, 1 - \bar{s}]$  where  $0 < \bar{s} < 1/2$  is deterministic and known, i.e. the change does not occur too close to 0 nor 1.*

Let us construct the following hypothesis test:

1. The null hypothesis  $H_0$  defines the case when there is no change-point;
2. The composite alternative hypothesis  $H_1$ : a change-point occurs at time  $s$ ,  $s \in [\bar{s}, 1 - \bar{s}]$ , i.e. the parameters which describe the distribution of the first component are different before and after  $s$  while the other parameters of the mixture remain the same.

This setting implies that the number of components does not change. Our work provides results when  $H_0$  holds; so they do not depend on what happens after the change. However, the test statistic we shall define is designed to amplify the change in one of the components of the mixture. Hence it makes sense to assume that the distribution after the change is a well defined mixture (identifiable) with the same number of components.

---

<sup>4</sup>See e.g. Redner (1981), Feng and McCulloch (1996), and Section 1.14 in McLachlan and Peel (2000) for discussion on the identifiability of mixtures.

After defining a statistic, we will establish a central limit theorem under the assumption that  $H_0$  holds. This is necessary to be able to determine the rejection domain while controlling the type I error (proportion of false positives). Therefore, in the following, we assume that  $H_0$  holds, i.e.  $X$  is independent identically distributed with distribution  $\mathbb{P}_\theta$ . We also consider that the following assumptions hold:

**Assumption 2.2.** For almost all  $x \in \mathcal{X}$ ,

- ◇ **Regularity:** the density  $f(x, \theta)$  is three times differentiable in  $\theta = (\theta_1, \dots, \theta_d) \in \Theta$ .
- ◇ **Integrability:** for  $1 \leq j \leq d$ ,  $\mathbb{E}_{H_0} \left[ \frac{\partial}{\partial \theta_j} \log f(X_1, \theta) \right] = 0$ .
- ◇ **Continuity:** for any  $1 \leq j, k, l \leq d$ , the applications  $\theta \mapsto \frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} \log f(x, \theta)$  are continuous on  $\Theta$ .
- ◇ **Dominance:** we can find some function  $\kappa_1(x)$ ,  $x \in \mathcal{X}$ , that does not depend on  $\theta$  and is such that, for all  $1 \leq j, k, l \leq d$  and all  $\theta$  in  $\Theta$ ,  $\left| \frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} \log f(x, \theta) \right| \leq \kappa_1(x)$  with  $\mathbb{E}_{H_0} [\kappa_1(X_1)] < \infty$ .

Let  $\mathbf{I}$  be the Fisher information matrix defined by

$$\mathbf{I}_{j,k} := \text{Cov} \left( \frac{\partial}{\partial \theta_j} \log f(X_1, \theta), \frac{\partial}{\partial \theta_k} \log f(X_1, \theta) \right), \quad 1 \leq j, k \leq d. \quad (1)$$

The assumptions we state on  $\mathbf{I}$  are the following:

**Assumption 2.3.** The matrix  $\mathbf{I}$  is positive definite with finite elements and

$$\mathbf{I}_{j,k} = \mathbb{E}_{H_0} \left[ \frac{\partial}{\partial \theta_j} \log f(X_1, \theta) \frac{\partial}{\partial \theta_k} \log f(X_1, \theta) \right] = -\mathbb{E}_{H_0} \left[ \frac{\partial^2}{\partial \theta_j \partial \theta_k} \log f(X_1, \theta) \right].$$

We define the log-likelihood of the sample  $X$  as a function of the  $d$ -dimensional vector  $\theta \in \Theta$  by

$$\theta \mapsto L(X, \theta) := \sum_{i=1}^n \log f(X_i, \theta).$$

It is well defined since, by assumption, the  $X_i(\omega)$ ,  $\omega \in \Omega$  and  $1 \leq i \leq n$ , can only take values in the set  $\{x \in \mathcal{X}, f(x, \theta) > 0\}$ .

With Assumptions 2.2 and 2.3, from the usual Limit Theorems<sup>5</sup> for Maximum Likelihood Estimators (MLE), there exist sequences of solutions of the likelihood equations  $\frac{\partial}{\partial \theta_j} L(X, \theta) = 0$ ,  $1 \leq j \leq d$ , that exist with probability tending to one as  $n \rightarrow \infty$  and are consistent. Let us select one of these consistent sequences of solutions as an estimator for the unknown  $\theta$  and denote it  $\hat{\theta} = (\hat{p}_1, \dots, \hat{p}_{m-1}, \hat{\lambda}_1, \dots, \hat{\lambda}_m)$ .

Note that this precaution is necessary since MLE might not exist for a finite sample<sup>6</sup>. In addition, for numerical applications, it is clear that the search of all roots of the likelihood equations would require unreasonable time. Thus the question of uniqueness is not really the problem. For the computation, one might use the EM algorithm<sup>7</sup> or some other approach<sup>8</sup>. Algorithms, when they converge, ensure to provide some local maximum but there is no guarantee to find the global maximum. We will see in Section 5 that, for univariate Gaussian mixtures, one can find some  $\Theta$  that ensures the existence of the estimator  $\hat{\theta}$  for finite samples.

In addition to the usual regularity conditions given in Assumptions 2.2 and 2.3, we assume that:

**Assumption 2.4.** The estimator  $\hat{\theta}$  is strongly consistent when  $H_0$  holds.

<sup>5</sup>See e.g. Theorem 5.1 in Lehmann and Casella (1998), Section 6.5.

<sup>6</sup>For the existence of MLE of Gaussian mixtures, see e.g. Example 6.10 in Lehmann and Casella (1998), Section 6.6. For a discussion on uniqueness of MLE in general, see Mäkeläinen et al. (1981). Chapter 2 in McLachlan and Peel (2000) gives an overview of these questions for mixtures.

<sup>7</sup>See e.g. Dempster et al. (1977), Wu (1983), Hathaway (1983), Redner and Walker (1984), Benaglia et al. (2009) and the references therein.

<sup>8</sup>For an overview, see e.g. Section 1.13 in McLachlan and Peel (2000), Tanaka (2009), Chen (2017) and the references therein.

This is an important restriction compared to the general case because Assumptions 2.2 and 2.3 only ensure the convergence in probability (Lehmann and Casella (1998)). For the applications in this article, we will consider the Gaussian case and discuss it in Section 5, using the result from Hathaway (1985). For other cases, one might use the classical results in the literature that cover a wide range of reasonable sufficient conditions. One of the first results has been given by Theorem 2 in Wald (1949) when  $\Theta$  is compact. Theorem 4 in Redner (1981) weakens Wald's conditions for  $\Theta$  a compact subset of the quotient topological space of all possible parameters, with a dedicated result for finite mixtures in Theorem 5. Feng and McCulloch (1996) proves it for  $\Theta$  a compact subset of the Euclidean space of possible parameters, i.e. showing that the question of identifiability does not impact the convergence properties of maximum likelihood estimators. Other approaches have been proposed by Kiefer and Wolfowitz (1956) or Redner and Walker (1984). For a wider discussion on mixtures, refer to Section 2.5 in McLachlan and Peel (2000).

**Remark 2.1.** *Since we study asymptotic properties, the main point in Assumption 2.2 is the integrability of the function  $\kappa_1$ . With the continuity condition and the convergence of  $\hat{\theta}$  from Assumption 2.4, almost surely, the boundedness itself follows for  $n$  large enough.*

For  $s \in [\bar{s}, 1 - \bar{s}]$ , by the same logic as for the estimator  $\hat{\theta}$ , we consider the estimators of  $\theta$  over the subsamples  $(X_i)_{1 \leq i \leq \lfloor sn \rfloor}$  and  $(X_i)_{\lfloor sn \rfloor + 1 \leq i \leq n}$ , respectively denoted by  $\hat{\theta}_{0,s}$  and  $\hat{\theta}_{s,1}$ . For a fixed  $s \in (0, 1)$ , they have the same properties as  $\hat{\theta}$  when  $n \rightarrow \infty$ .

To focus our study on only one component of the mixture (the first component), we design a specific weight function that, in the detection statistic, allows to overweight the density of an observation  $x$  when the density of the first component is dominant compared to others. For that purpose we define the weight function at point  $x \in \mathcal{X}$  for  $\theta = (p_1, \dots, p_{m-1}, \lambda_1, \dots, \lambda_m) \in \Theta$  by

$$w(x, \theta) := \frac{p_1 f_1(x, \lambda_1)}{f(x, \theta)}. \quad (2)$$

The function  $w$  is well defined since, by assumption, the random variables  $X_i$ ,  $1 \leq i \leq n$ , can only take values in the set  $\{x \in \mathcal{X}, f(x, \theta) > 0\}$ . In addition, by definition, for any  $x \in \mathcal{X}$  and any  $\theta \in \Theta$ ,  $0 \leq w(x, \theta) \leq 1$ . As a consequence of Assumption 2.2, the application

$$w \log f_1 : (x, \theta) \mapsto w(x, \theta) \log f_1(x, \lambda_1) \quad (3)$$

is twice differentiable in  $\theta \in \Theta$ , and, for all  $1 \leq j, k \leq d$  and almost all  $x \in \mathcal{X}$ , the application  $\theta \in \Theta \mapsto \frac{\partial^2}{\partial \theta_j \partial \theta_k} (w \log f_1)(x, \theta)$  is continuous in  $\theta$ .

Here is an assumption concerning the application  $w$ :

**Assumption 2.5. (Dominance)** *There exists some convex subset  $\Theta' \subset \Theta$  such that  $\theta$  is in the interior of  $\Theta'$ , and an application  $\kappa_2$  from  $\mathcal{X}$  to  $\mathbb{R}$  that does not depend on  $\theta$ , such that, for all  $1 \leq j, k \leq d$ ,  $\theta$  in  $\Theta'$  and for almost all  $x \in \mathcal{X}$ ,  $\left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} (w \log f_1)(x, \theta) \right| \leq \kappa_2(x)$  with  $\mathbb{E}_{H_0}[\kappa_2(X_1)] < \infty$ .*

For the same reasons as for Assumption 2.2, the essential point here is the integrability of the dominating function  $\kappa_2$ .

## 2.2 Definition of the Weighted Likelihood Test (WLT)

We now define the test statistic. First we introduce  $\Lambda_n := (\Lambda_{s,n})_{s \in [\bar{s}, 1 - \bar{s}]}$ , the underlying càd-làg stochastic process of the detection statistic:

$$\begin{aligned} \Lambda_{s,n} &:= \log \left( \frac{\prod_{i=1}^{\lfloor sn \rfloor} f_1(X_i, \hat{\lambda}_{0,s,1})^{w(X_i, \hat{\theta}_{0,s})} \prod_{j=\lfloor sn \rfloor + 1}^n f_1(X_j, \hat{\lambda}_{s,1,1})^{w(X_j, \hat{\theta}_{s,1})}}{\prod_{i=1}^n f_1(X_i, \hat{\lambda}_1)^{w(X_i, \hat{\theta})}} \right) \\ &= \sum_{i=1}^{\lfloor sn \rfloor} w(X_i, \hat{\theta}_{0,s}) \log f_1(X_i, \hat{\lambda}_{0,s,1}) + \sum_{i=\lfloor sn \rfloor + 1}^n w(X_i, \hat{\theta}_{s,1}) \log f_1(X_i, \hat{\lambda}_{s,1,1}) - \sum_{i=1}^n w(X_i, \hat{\theta}) \log f_1(X_i, \hat{\lambda}_1). \end{aligned} \quad (4)$$

Note that, for an observation  $X_i$  with distribution parameter  $\theta$ , the weight  $w(X_i, \theta)$  is the probability that  $X_i$  comes from the first component. Conditionally to this fact, the log-likelihood of  $X_i$  is given by  $\log f_1(X_i, \lambda_1)$ . Thus the expression  $w(X_i, \theta) \log f_1(X_i, \lambda_1)$  in  $\Lambda_{s,n}$  somehow reflects the contribution of the first component in the likelihood of  $X_i$ . As a consequence, the response of the statistic is magnified when a change occurs in the first component.

In addition, the process  $\Lambda_n$  is defined on  $[\bar{s}, 1 - \bar{s}]$  in order to ensure that, for  $n$  large enough, an asymptotic behavior can be obtained for each sum. The **test statistic** is then defined by

$$S_n := \sup_{s \in [\bar{s}, 1 - \bar{s}]} \Lambda_{s,n}. \quad (5)$$

We refer to this test as the WLT (Weighted Likelihood Test). The test procedure states that there is no change-point for the first component (i.e. we accept  $H_0$ ) when  $S_n$  is smaller than some threshold  $L_\alpha$  chosen with respect to a false alarm constraint. This false alarm can be obtained from the probability of false alarm  $\alpha \in (0, 1)$  such that  $L_\alpha$  is the  $\alpha$ -percentile of the limit distribution of  $S_n$  when  $H_0$  holds.

The main purpose of this article is to derive the limit distribution of  $S_n$  when  $H_0$  holds. We follow the work of Davis et al. (1995) and look at  $\Lambda_{s,n}$  as a stochastic process. In the end, we derive the limit distribution of the detection statistic  $S_n$  in Theorem 3.8.

### 2.3 Notations

We denote by  $D_\theta(\cdot)$ ,  $D_\theta^2(\cdot)$  and  $D_\theta^3(\cdot)$  respectively the vector, matrix and hypermatrix differential operators in  $\theta \in \mathbb{R}^d$ . For  $\theta, \tilde{\theta} \in \mathbb{R}^d$ , we denote by  $[\theta, \tilde{\theta}]$  the segment  $[\theta, \tilde{\theta}] := \{\lambda\theta + (1 - \lambda)\tilde{\theta}, \lambda \in [0, 1]\}$ .

$gl_d(\mathbb{R})$  denotes the set of matrices of size  $d \times d$  with real coefficients and  $GL_d(\mathbb{R})$  the set of invertible  $d \times d$ -matrices with real coefficients.

For a given matrix  $M$ , its  $i$ -th line is denoted by  $M_{i,\cdot}$  and its  $j$ -th column is denoted by  $M_{\cdot,j}$ . The same logic is used for hypermatrices: for a given  $J \in \mathbb{R}^{d \times d \times d}$  and  $1 \leq i \leq d$ , we denote by  $J_{i,\dots} := (J_{i,j,k})_{1 \leq j,k \leq d}$  the  $d \times d$ -matrix obtained from  $J$ .

For a given matrix  $M$ , we denote  $(M^{-1})^T$  by  $M^{-1T}$ .

If  $\Sigma^2$  is a covariance matrix, then it is positive semi-definite, and  $\Sigma$  will denote the unique positive semi-definite square root of  $\Sigma^2$ .

In this paper, the weak convergence of càd-làg paths is to be understood as a weak convergence in the Skorohod metric space of càd-làg paths. Refer to the supplementary material for additional technical details.

A glossary of notations is given in Section S6 of the supplementary material.

## 3 Limit distribution of the test statistic

In this section we shall determine the limit distribution of the process  $\Lambda_n$  as  $n$  tends to infinity.

Let us consider the constant

$$\mathbf{u} := \mathbb{E}_{H_0} [D_\theta(w \log f_1)(X_1, \boldsymbol{\theta})] \in \mathbb{R}^d, \quad (6)$$

where the application  $w \log f_1$  is defined in (3). By Assumption 2.5,  $\mathbf{u}$  is finite. We indicate that  $\mathbf{u}$  is not null in general: we numerically established that  $\mathbf{u}$  is strictly positive for some examples of Gaussian mixtures<sup>9</sup>. The càd-làg real-valued process  $\Lambda_n$  defined on  $[\bar{s}, 1 - \bar{s}]$  can be decomposed as follows:

$$\Lambda_{s,n} = Q_{s,n}^1 + Q_{s,n}^2 - Q_{1,n}^1, \quad s \in [\bar{s}, 1 - \bar{s}], \quad (7)$$

---

<sup>9</sup>Using the Strong Law of Large Numbers, a simple numerical simulation for a Gaussian mixture with 3 components shows that, in general, the constant  $\mathbf{u}$  is not null. See Section S4 of the supplementary material for an illustration.

where  $Q_n^1 = (Q_{s,n}^1)_{s \in [\bar{s}, 1]}$  and  $Q_n^2 = (Q_{s,n}^2)_{s \in [\bar{s}, 1 - \bar{s}]}$  are càd-làg real-valued processes defined by

$$Q_{s,n}^1 := \sum_{i=1}^{\lfloor sn \rfloor} \left( w(X_i, \hat{\theta}_{0,s}) \log f_1(X_i, \hat{\lambda}_{0,s,1}) - w(X_i, \boldsymbol{\theta}) \log f_1(X_i, \boldsymbol{\lambda}_1) \right) - \mathbf{u}^T \mathbf{I}^{-1} \sum_{i=1}^{\lfloor sn \rfloor} D_{\boldsymbol{\theta}}(\log f)(X_i, \boldsymbol{\theta}),$$

$$Q_{s,n}^2 := \sum_{i=\lfloor sn \rfloor + 1}^n \left( w(X_i, \hat{\theta}_{s,1}) \log f_1(X_i, \hat{\lambda}_{s,1,1}) - w(X_i, \boldsymbol{\theta}) \log f_1(X_i, \boldsymbol{\lambda}_1) \right) - \mathbf{u}^T \mathbf{I}^{-1} \sum_{i=\lfloor sn \rfloor + 1}^n D_{\boldsymbol{\theta}}(\log f)(X_i, \boldsymbol{\theta}).$$

The process  $Q_n^1$  is defined on  $[\bar{s}, 1]$  in order to include the last term of  $\Lambda_{s,n}$  in (7) while  $Q_n^2$  needs only to be defined on  $[\bar{s}, 1 - \bar{s}]$ . In addition, we remark that the random processes  $(Q_{s,n}^1)_{s \in [\bar{s}, 1 - \bar{s}]}$  and  $Q_n^2$  have a similar structure that differs only by the sub-sample considered. Therefore, in the following, we study the limit of  $Q_n^1$  and simply extends the arguments to obtain the limit of  $\Lambda_n$ .

### 3.1 Limit distribution of $Q_n^1$

The process  $Q_n^1$  is the first of the three terms defining the process  $\Lambda_n$  in (7). We want to derive its limit distribution with a Donsker-type result. In the following, our purpose is to reorganize the terms so that the process  $Q_{s,n}^1$  is somehow the product of a matrix that converges almost surely uniformly in  $s \in [\bar{s}, 1]$  to some constant, and of a vector that converges weakly to some random process. With the help of Slutsky's Theorem for random elements<sup>10</sup>, one can target to derive the limit process of  $Q_{s,n}^1$ .

The difficulty here is that the random variable  $Q_{s,n}^1$  is not a sum of independent terms, because of the presence of the estimator  $\hat{\theta}_{0,s}$ , which depends itself on the whole sub-sample  $(X_1, X_2, \dots, X_{\lfloor sn \rfloor})$ . For that purpose, we expect to develop  $\boldsymbol{\theta} = (p_1, \dots, p_{m-1}, \lambda_1, \dots, \lambda_m) \mapsto w(x, \boldsymbol{\theta}) \log f_1(x, \boldsymbol{\lambda}_1)$  around  $\boldsymbol{\theta} = (\mathbf{p}_1, \dots, \mathbf{p}_{m-1}, \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_m)$ . This is possible with the help of the two following properties of the estimator  $\hat{\theta}_{0,s}$ .

First, from Assumption 2.4, we already know that the estimator  $\hat{\boldsymbol{\theta}}$  converges almost surely to  $\boldsymbol{\theta}$  when  $n \rightarrow \infty$ . With the following result inspired from Proposition 3.3 in Dehling et al. (2014), we extend this convergence property to  $\hat{\theta}_{0,s}$ ,  $s \in [\bar{s}, 1]$  (for a detailed proof, refer to Section S2 in the supplementary material).

**Proposition 3.1.** *Under  $H_0$  and Assumptions 2.1-2.4, the estimator  $\hat{\theta}_{0,s}$  converges almost surely to  $\boldsymbol{\theta}$ , uniformly in  $s \in [\bar{s}, 1]$ .*

This will represent a key property for the main result.

Second, almost surely, it is possible to obtain an explicit form for  $\hat{\theta}_{0,s}$  for  $n$  large enough ( $n$  depending on  $\omega$ ). Indeed, the estimator  $\hat{\boldsymbol{\theta}}$  is a sequence of solutions of the likelihood equations  $D_{\boldsymbol{\theta}}L(X, \boldsymbol{\theta}) = 0$ . Therefore, we can follow the ideas from the proof of the usual limit theorems for maximum likelihood estimators (see e.g. Theorem 5.1 in Lehmann and Casella (1998), Section 6.5). As detailed in Section S2 of the supplementary material, we obtain that, for any  $s \in [\bar{s}, 1]$ , there exists some point  $\hat{\theta}'_{0,s}$  on the segment  $[\hat{\theta}_{0,s}, \boldsymbol{\theta}]$ , such that the matrix

$$\hat{A}_{0,s} := -\frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} \left( D_{\boldsymbol{\theta}}^2(\log f)(X_i, \boldsymbol{\theta}) + \frac{1}{2} \sum_{l=1}^d (\hat{\theta}_{0,s;l} - \boldsymbol{\theta}_l) D_{\boldsymbol{\theta}}^3(\log f)(X_i, \hat{\theta}'_{0,s})_{l,\dots} \right)$$

satisfies  $\hat{A}_{0,s}(\hat{\theta}_{0,s} - \boldsymbol{\theta}) = \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D_{\boldsymbol{\theta}}(\log f)(X_i, \boldsymbol{\theta})$ . The following proposition provides an explicit expression for  $\hat{\theta}_{0,s}$  and the convergence of  $\hat{A}_{0,s}$ .

**Proposition 3.2.** *Under  $H_0$  and Assumptions 2.1-2.4, almost surely, for  $n$  large enough, for all  $s \in [\bar{s}, 1]$ , the matrix  $\hat{A}_{0,s}$  is invertible and*

$$\hat{\theta}_{0,s} - \boldsymbol{\theta} = \hat{A}_{0,s}^{-1} \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D_{\boldsymbol{\theta}}(\log f)(X_i, \boldsymbol{\theta}),$$

with  $\mathbb{E}_{H_0} [D_{\boldsymbol{\theta}}(\log f)(X_1, \boldsymbol{\theta})] = 0$ , and where  $\hat{A}_{0,s}^{-1}$  converges almost surely to  $\mathbf{I}^{-1}$ , uniformly in  $s \in [\bar{s}, 1]$ .

<sup>10</sup>See e.g. Theorem 18.10 in van der Vaart (1998).

By Assumption 2.5 and Proposition 3.1, it follows that, with a Taylor-Lagrange decomposition<sup>11</sup>, almost surely, we can find some  $N \geq 1$  (depending on  $\omega$ ) and some  $\theta'_{0,s} \in [\hat{\theta}_{0,s}, \boldsymbol{\theta}]$  (depending on  $\omega$ ,  $n$  and  $s$ ), such that, for  $n \geq N$  and for  $s \in [\bar{s}, 1]$ ,

$$\begin{aligned} & \sum_{i=1}^{\lfloor sn \rfloor} \left( w(X_i, \hat{\theta}_{0,s}) \log f_1(X_i, \hat{\lambda}_{0,s,1}) - w(X_i, \boldsymbol{\theta}) \log f_1(X_i, \boldsymbol{\lambda}_1) \right) \\ &= \sum_{i=1}^{\lfloor sn \rfloor} \left( D_{\theta}(w \log f_1)(X_i, \boldsymbol{\theta})^T (\hat{\theta}_{0,s} - \boldsymbol{\theta}) + (\hat{\theta}_{0,s} - \boldsymbol{\theta})^T D_{\theta}^2(w \log f_1)(X_i, \theta'_{0,s}) (\hat{\theta}_{0,s} - \boldsymbol{\theta}) \right). \end{aligned}$$

So far, we can rewrite  $Q_{s,n}^1$  as follows

$$\begin{aligned} Q_{s,n}^1 &= \left( \sum_{i=1}^{\lfloor sn \rfloor} D_{\theta}(w \log f_1)(X_i, \boldsymbol{\theta})^T \right) (\hat{\theta}_{0,s} - \boldsymbol{\theta}) + (\hat{\theta}_{0,s} - \boldsymbol{\theta})^T \left( \sum_{i=1}^{\lfloor sn \rfloor} D_{\theta}^2(w \log f_1)(X_i, \theta'_{0,s}) \right) (\hat{\theta}_{0,s} - \boldsymbol{\theta}) \\ &\quad - \mathbf{u}^T \mathbf{I}^{-1} \sum_{i=1}^{\lfloor sn \rfloor} D_{\theta}(\log f)(X_i, \boldsymbol{\theta}). \end{aligned}$$

From Assumption 2.5, we know that  $\mathbb{E}_{H_0} [|D_{\theta}(w \log f_1)(X_1, \boldsymbol{\theta})|] < \infty$ . Thus we can center the right side sum of the first term by  $\mathbf{u}$ :

$$\begin{aligned} Q_{s,n}^1 &= \left( \sum_{i=1}^{\lfloor sn \rfloor} (D_{\theta}(w \log f_1)(X_i, \boldsymbol{\theta})^T - \mathbf{u}^T) \right) (\hat{\theta}_{0,s} - \boldsymbol{\theta}) + (\hat{\theta}_{0,s} - \boldsymbol{\theta})^T \left( \sum_{i=1}^{\lfloor sn \rfloor} D_{\theta}^2(w \log f_1)(X_i, \theta'_{0,s}) \right) (\hat{\theta}_{0,s} - \boldsymbol{\theta}) \\ &\quad + \lfloor sn \rfloor \mathbf{u}^T (\hat{\theta}_{0,s} - \boldsymbol{\theta}) - \mathbf{u}^T \mathbf{I}^{-1} \sum_{i=1}^{\lfloor sn \rfloor} D_{\theta}(\log f)(X_i, \boldsymbol{\theta}). \end{aligned}$$

Choosing  $n$  large enough, Proposition 3.2 provides us an explicit expression of  $\hat{\theta}_{0,s} - \boldsymbol{\theta}$ , which permits us to rewrite  $Q_{s,n}^1$  as follows:

$$\begin{aligned} Q_{s,n}^1 &= \left( \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} (D_{\theta}(w \log f_1)(X_i, \boldsymbol{\theta})^T - \mathbf{u}^T) \right) \lfloor sn \rfloor \hat{A}_{0,s}^{-1} \left( \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D_{\theta}(\log f)(X_i, \boldsymbol{\theta}) \right) \\ &\quad + \left( \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D_{\theta}(\log f)(X_i, \boldsymbol{\theta})^T \right) \lfloor sn \rfloor \hat{A}_{0,s}^{-1 T} \left( \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D_{\theta}^2(w \log f_1)(X_i, \theta'_{0,s}) \right) \hat{A}_{0,s}^{-1} \left( \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D_{\theta}(\log f)(X_i, \boldsymbol{\theta}) \right) \\ &\quad + \mathbf{u}^T (\hat{A}_{0,s}^{-1} - \mathbf{I}^{-1}) \lfloor sn \rfloor \left( \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D_{\theta}(\log f)(X_i, \boldsymbol{\theta}) \right). \end{aligned} \tag{8}$$

Remark that, in this reformulation of  $Q_{s,n}^1$ , we can recognize several sums of centered i.i.d random variables which, multiplied by  $\sqrt{n}$ , can be treated by Donsker's Theorem and produce at the limit a multi-dimensional Brownian motion. Further, there are the variables  $\hat{A}_{0,s}^{-1}$ , which, by Proposition 3.2, converge a.s. to the inverse Fischer information  $\mathbf{I}^{-1}$ . The last line of (8), decomposing  $\lfloor sn \rfloor$  into  $\sqrt{n} \frac{\lfloor sn \rfloor}{n} \sqrt{n}$ , makes appear the term  $\sqrt{n} (\hat{A}_{0,s}^{-1} - \mathbf{I}^{-1})$ . Its limit, when  $n$  tends to  $\infty$  and the link with the other components of  $Q_{s,n}^1$  have to be analyzed separately, before we combine all these terms to compute the limit of the process  $(Q_{s,n}^1)_{s \in [\bar{s}, 1]}$  in terms of a transformation of a Brownian motion.

---

<sup>11</sup>See e.g. Theorem 5.3 in Coleman (2012).

Firstly, for all  $s \in [\bar{s}, 1]$ , we define the triple  $\hat{\xi}_{0,s} := (\hat{\iota}_{0,s}, \hat{u}_{0,s} - \mathbf{u}, \hat{I}_{0,s}) \in (\mathbb{R}^d)^2 \times gl_d(\mathbb{R})$  by

$$\begin{aligned}\hat{\iota}_{0,s} &:= \frac{1}{[sn]} \sum_{i=1}^{[sn]} D_\theta(\log f)(X_i, \boldsymbol{\theta}) \\ \hat{u}_{0,s} &:= \frac{1}{[sn]} \sum_{i=1}^{[sn]} D_\theta(w \log f_1)(X_i, \boldsymbol{\theta}) \\ \hat{I}_{0,s} &:= -\frac{1}{[sn]} \sum_{i=1}^{[sn]} D_\theta^2(\log f)(X_i, \boldsymbol{\theta})\end{aligned}\tag{9}$$

which allows us to rewrite  $Q_{s,n}^1$  as

$$\begin{aligned}Q_{s,n}^1 &= \frac{[sn]}{n} \sqrt{n}(\hat{u}_{0,s} - \mathbf{u})^T \hat{A}_{0,s}^{-1} \sqrt{n} \hat{\iota}_{0,s} + \frac{[sn]}{n} \sqrt{n} \hat{\iota}_{0,s}^T \hat{A}_{0,s}^{-1} \left( \frac{1}{[sn]} \sum_{i=1}^{[sn]} D_\theta^2(w \log f_1)(X_i, \boldsymbol{\theta}'_{0,s}) \right) \hat{A}_{0,s}^{-1} \sqrt{n} \hat{\iota}_{0,s} \\ &\quad + \frac{[sn]}{n} \mathbf{u}^T \sqrt{n} \left( \hat{A}_{0,s}^{-1} - \mathbf{I}^{-1} \right) \sqrt{n} \hat{\iota}_{0,s}.\end{aligned}\tag{10}$$

Another crucial ingredient of our following discussion is the covariance matrix  $\Sigma^2 \in gl_{2d+d^2}(\mathbb{R})$  under  $H_0$  of the triple

$$(D_\theta(\log f)(X_1, \boldsymbol{\theta}), D_\theta(w \log f_1)(X_1, \boldsymbol{\theta}) - \mathbf{u}, -D_\theta^2(\log f)(X_1, \boldsymbol{\theta}) - \mathbf{I})$$

reorganized in a  $2d + d^2$ -dimensional real vector. As a covariance matrix,  $\Sigma^2$  is positive semi-definite. Then  $\Sigma$  will denote the unique positive semi-definite square root of  $\Sigma^2$ .

**Lemma 3.3.** *Set  $\boldsymbol{\xi} := (0, 0, \mathbf{I}) \in (\mathbb{R}^d)^2 \times GL_d(\mathbb{R})$ . Under  $H_0$ , for all  $s \in [\bar{s}, 1]$ ,  $\mathbb{E}_{H_0}[\hat{\xi}_{0,s}] = \boldsymbol{\xi}$ , the sequence of random variables  $\hat{\xi}_{0,s}$  converges a.s. to  $\boldsymbol{\xi}$ , uniformly in  $s \in [\bar{s}, 1]$ , and the process  $\sqrt{n} \left( \hat{\xi}_{0,s} - \boldsymbol{\xi} \right)_{s \in [\bar{s}, 1]}$  converges weakly as follows*

$$\sqrt{n} \left( \hat{\xi}_{0,s} - \boldsymbol{\xi} \right)_{s \in [\bar{s}, 1]} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \left( \frac{1}{s} \Sigma W_s \right)_{s \in [\bar{s}, 1]},$$

where  $W := (W_s)_{s \in [0,1]}$  is a standard  $2d + d^2$ -dimensional Brownian motion and  $\Sigma W_s$  is reorganized as a triple in  $(\mathbb{R}^d)^2 \times gl_d(\mathbb{R})$ .

The next Lemma will help us to handle the term  $(\hat{A}_{0,s}^{-1} - \mathbf{I}^{-1})$  which appears in the last line of the expression (10) of  $Q_{0,s}^1$ . To this aim we introduce the hypermatrix

$$\mathbf{J} := \mathbb{E}_{H_0} [D_\theta^3(\log f)(X_1, \boldsymbol{\theta})] \in \mathbb{R}^{d \times d \times d}.\tag{11}$$

**Lemma 3.4.** *Almost surely, for large  $n$  (depending on  $\omega$ ), the variable  $\hat{A}_{0,s}$  can be written as*

$$\hat{A}_{0,s} = \hat{I}_{0,s} - \frac{1}{2} \sum_{l=1}^d \left( \hat{\iota}_{0,s}^T \left( \hat{A}_{0,s}^{-1} \right)_{\cdot, l} \right) (\hat{J}_{0,s})_{\dots, l},$$

for all  $s \in [\bar{s}, 1]$ , where  $\hat{J}_{0,s}$  is the hypermatrix defined by  $\hat{J}_{0,s} := \frac{1}{[sn]} \sum_{i=1}^{[sn]} D_\theta^3(\log f)(X_i, \boldsymbol{\theta}'_{0,s})$ . In addition, under  $H_0$ , almost surely,  $\hat{J}_{0,s}$  converges to the hypermatrix  $\mathbf{J}$ , uniformly in  $s \in [\bar{s}, 1]$ .

Set now, for all  $s \in [\bar{s}, 1]$  and all  $n \geq 1$ ,

$$\hat{\xi}'_{0,s} := (\hat{\iota}_{0,s}, \hat{u}_{0,s} - \mathbf{u}, \hat{A}_{0,s}^{-1}).$$

It is easy to see with Proposition 3.2 that, when  $n$  tends to infinity,  $\hat{\xi}'_{0,s}$  converges a.s. to

$$\boldsymbol{\xi}' := (0, 0, \mathbf{I}^{-1}).$$

The aim of the next theorem is to establish a Donsker-type result for the process  $(\hat{\xi}'_{0,s})_{s \in [\bar{s}, 1]}$ . The result is obtain with the help of Lemma 3.4.

**Theorem 3.5.** Under  $H_0$ , the process  $\sqrt{n}(\hat{\xi}'_{0,s} - \xi')$  converges weakly to  $(\frac{1}{s}\mathbf{g}(\Sigma W_s))_{s \in [\bar{s}, 1]}$ , where  $\mathbf{g}$  is the linear map defined for  $(\iota, u, I) \in (\mathbb{R}^d)^2 \times gl_d(\mathbb{R})$  by

$$\mathbf{g}(\iota, u, I) := \left( \iota, u, -\mathbf{I}^{-1} \left( I - \frac{1}{2} \sum_{l=1}^d \left( \iota^T \left( \mathbf{I}^{-1 T} \right)_{\cdot, l} \right) \mathbf{J}_{\dots, l} \right) \mathbf{I}^{-1} \right), \quad (12)$$

and  $\Sigma W_s$  is reorganized as a triple in  $(\mathbb{R}^d)^2 \times gl_d(\mathbb{R})$ .

The main theorem states that the process  $Q_n^1$  converges to a quadratic form of the Brownian motion  $W$  introduced in Theorem 3.5. For  $(\iota, u, I; A, J) \in (\mathbb{R}^d)^2 \times (gl_d(\mathbb{R}))^3$ , we set first:

$$q(\iota, u, I; A, J) := u^T A \iota + \iota^T A^T J A \iota + \mathbf{u}^T I \iota \in \mathbb{R}, \quad (13)$$

where we recall the vector  $\mathbf{u} = \mathbb{E}_{H_0} [D_\theta(w \log f_1)(X_1, \theta)]$  defined in (6). We also consider the application  $\mathbf{q}$  defined as follows:

$$\mathbf{q}(z) := q(\mathbf{g}(\Sigma z); \mathbf{I}^{-1}, \mathbf{U}), \quad z \in \mathbb{R}^{2d+d^2}, \quad (14)$$

where  $\mathbf{g}$  is defined by (12),  $\Sigma z$  is reorganized as a triple in  $(\mathbb{R}^d)^2 \times gl_d(\mathbb{R})$ . Further  $\mathbf{U}$  is the matrix defined by

$$\mathbf{U} := \mathbb{E}_{H_0} [D_\theta^2(w \log f_1)(X_1, \theta)]. \quad (15)$$

**Remark 3.6.** The map  $z \in \mathbb{R}^{2d+d^2} \mapsto \mathbf{q}(z)$  is a quadratic form.

We can now state our main result for  $Q_n^1$ .

**Theorem 3.7.** Under  $H_0$ , the process  $Q_n^1$  converges weakly as  $n \rightarrow \infty$  to the process  $(\frac{1}{s}\mathbf{q}(W_s))_{s \in [\bar{s}, 1]}$ .

### 3.2 Limit distribution of the test statistic

The limit distribution of  $\Lambda_n$  is obtain as an extension of the result in Theorem 3.7 for  $Q_n^1$ . In the following Theorem we derive the limit distribution of the test statistic  $S_n$ .

**Theorem 3.8.** Under  $H_0$  and Assumptions 2.1-2.5, the test statistic

$$S_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \sup_{s \in [\bar{s}, 1 - \bar{s}]} \frac{\mathbf{q}(W_s - sW_1)}{s(1-s)}$$

where  $(W_s)_{s \in [0, 1]}$  is a standard  $2d + d^2$ -dimensional Brownian motion and the application  $\mathbf{q}$  is defined in (14).

The limit distribution obtained is somehow similar to the one given by Csörgő and Horváth (1997), Corollary 1.1.1, for the i.i.d. case of an unconstrained log-likelihood ratio test. Since  $\mathbf{q}$  is a quadratic form,  $\mathbf{q}(W_s - sW_1)_{s \in [\bar{s}, 1 - \bar{s}]}$  is also a quadratic form of a Brownian bridge. The introduction of the weights impacts here the dimension of the Brownian bridge that is here  $2d + d^2$ , while, in the standard case, the Brownian bridge is of dimension  $d$ .

### 3.3 Test procedure

In practice, we observe a realization  $X(\omega)$  for some  $\omega \in \Omega$  of the random sample  $X$  with  $n$  observations. We propose the following test procedure:

1. Compute the estimators  $\hat{\theta}$ ,  $(\hat{\theta}_{0,s})_{s \in [\bar{s}, 1 - \bar{s}]}$  and  $(\hat{\theta}_{s,1})_{s \in [\bar{s}, 1 - \bar{s}]}$  where  $\bar{s}$  is known from Assumption 2.1. The three estimators are defined in Section 2.
2. Compute the process  $(\Lambda_{s,n})_{s \in [\bar{s}, 1 - \bar{s}]}$  and the test statistic  $S_n$  using their definitions in Equations (4) and (5).
3. Compute the constants  $\mathbf{I}^{-1}$  defined in (1),  $\mathbf{U}$  defined in (15),  $\mathbf{J}$  defined in (11) and the covariance matrix  $\Sigma$  defined for Lemma 3.3. This requires additional developments: theoretical computation or numerical approximation.

4. Compute the distribution of  $\sup_{s \in [\bar{s}, 1 - \bar{s}]} \frac{\mathbf{q}(W_s - sW_1)}{s(1-s)}$  where  $W$  is a standard Brownian motion and  $\mathbf{q}$  is defined in (14). This requires additional development: theoretical computation or numerical approximation.
5. Compute the threshold  $L_\alpha$  chosen with respect to a false alarm constraint. It can be obtained from the probability of false alarm  $\alpha \in (0, 1)$  such that  $L_\alpha$  is the  $\alpha$ -percentile of the distribution of  $\sup_{s \in [\bar{s}, 1 - \bar{s}]} \frac{\mathbf{q}(W_s - sW_1)}{s(1-s)}$ . We conclude that no change-point occurs if the statistic  $S_n$  is smaller than the threshold: we should reject  $H_0$ .

Since, by definition, the estimators  $\hat{\theta}_{0,s}$ ,  $\hat{\theta}_{s,1}$  and the process  $\Lambda_{s,n}$  are constant piecewise, it is sufficient to estimate them on a finite set of  $s \in [\bar{s}, 1 - \bar{s}]$ .

This test focuses on the first component of the mixture. Obviously, by definition of the mixture and of the test itself, this can be applied to any other component. Therefore, in practice, it might be relevant to run a test on each component. Alongside a standard Likelihood Ratio Test that does not focus on a specific component, the set of tests constitutes a useful detection tool for the industry.

In the next section, we suggest an extended version of this test. Numerical applications indicate that such an extension increases the detection frequency when a change occurs.

## 4 Extension: scaling the contributions in the likelihood ratio (EWLT)

In this section, we introduce an extended version of the test defined in Section 2.2. For a fixed  $s \in [\bar{s}, 1]$ , we define the contribution  $c_{s,n}$  by

$$c_{s,n} := \sum_{i=1}^{\lfloor sn \rfloor} w(X_i, \hat{\theta}_{0,s}) + \sum_{i=\lfloor sn \rfloor + 1}^n w(X_i, \hat{\theta}_{s,1}).$$

Remark that, from its definition in (4), the log-ratio  $\Lambda_{s,n}$  is the difference of  $(Q_{s,n}^1 + Q_{s,n}^2)$  and  $Q_{1,n}^1$ . Then,  $c_{s,n}$  is the contribution of the sample to the term  $(Q_{s,n}^1 + Q_{s,n}^2)$ , and  $c_{1,n} = \sum_{i=1}^n w(X_i, \hat{\theta})$  is the contribution of the sample to the term  $Q_{1,n}^1$ . Under the null hypothesis, by the definition of  $w(\cdot, \cdot)$  in (2), we can show<sup>12</sup> that

$$\frac{c_{s,n}}{n} \xrightarrow[n \rightarrow \infty]{a.s.} \mathbb{E}_{H_0} [w(X_1, \boldsymbol{\theta})] = \int_{\mathcal{X}} \frac{\mathbf{p}_1 f_1(x, \boldsymbol{\lambda}_1)}{f(x, \boldsymbol{\theta})} f(x, \boldsymbol{\theta}) dx = \mathbf{p}_1,$$

uniformly in  $s \in [\bar{s}, 1]$ . It follows that, for a given  $s \in [\bar{s}, 1 - \bar{s}]$ , the average contributions  $c_{s,n}/n$  and  $c_{1,n}/n$  have the same limit under the null hypothesis. Under the alternative hypothesis, we observed in numerical applications that these average contributions can play a significant role in the detection performance. Therefore, we suggest to scale our statistic with the total contributions. We define a new log-ratio process  $\Lambda_n^* := (\Lambda_{s,n}^*)_{s \in [\bar{s}, 1 - \bar{s}]}$  by

$$\Lambda_{s,n}^* := \frac{c_{1,n}}{c_{s,n}} \left( \sum_{i=1}^{\lfloor sn \rfloor} w(X_i, \hat{\theta}_{0,s}) \log f_1(X_i, \hat{\lambda}_{0,s,1}) + \sum_{i=\lfloor sn \rfloor + 1}^n w(X_i, \hat{\theta}_{s,1}) \log f_1(X_i, \hat{\lambda}_{s,1,1}) \right) - \sum_{i=1}^n w(X_i, \hat{\theta}) \log f_1(X_i, \hat{\lambda}_1). \quad (16)$$

The test statistic is then defined by  $S_n^* := \sup_{s \in [\bar{s}, 1 - \bar{s}]} \Lambda_{s,n}^*$ . We refer to this test as the EWLT (Extended Weighted Likelihood Test).

<sup>12</sup>With the help of Lemma S2.9 in the supplementary material.

In order to obtain a limit distribution for  $S_n^*$ , we start by noticing that

$$\begin{aligned}\Lambda_{s,n}^* &= \frac{c_{1,n}}{c_{s,n}} \Lambda_{s,n} + \left( \frac{c_{1,n}}{c_{s,n}} - 1 \right) \sum_{i=1}^n w(X_i, \hat{\theta}) \log f_1(X_i, \hat{\lambda}_1) \\ &= \frac{c_{1,n}}{c_{s,n}} \Lambda_{s,n} - \frac{\frac{1}{n} \sum_{i=1}^n w(X_i, \hat{\theta}) \log f_1(X_i, \hat{\lambda}_1)}{\frac{1}{n} c_{s,n}} \left( \sum_{i=1}^{\lfloor sn \rfloor} w(X_i, \hat{\theta}_{0,s}) + \sum_{i=\lfloor sn \rfloor+1}^n w(X_i, \hat{\theta}_{s,1}) - \sum_{i=1}^n w(X_i, \hat{\theta}) \right).\end{aligned}$$

We already know that the ratio  $c_{1,n}/c_{s,n}$  converges a.s. to 1 uniformly in  $s \in [\bar{s}, 1]$ . With the help of Lemma S2.9 in the supplementary material, we can show that

$$\frac{\frac{1}{n} \sum_{i=1}^n w(X_i, \hat{\theta}) \log f_1(X_i, \hat{\lambda}_1)}{\frac{1}{n} c_{s,n}} \xrightarrow[n \rightarrow \infty]{a.s.} \mathbb{E}_{H_0} [\log f_1(Y, \boldsymbol{\lambda}_1)] =: \boldsymbol{\beta},$$

uniformly in  $s \in [\bar{s}, 1]$ , with  $Y$  a random variable with density  $f_1(\cdot, \boldsymbol{\lambda}_1)$ . In addition, we can assess numerically that, in general,  $\boldsymbol{\beta}$  is not null.

Remark that the sum

$$\sum_{i=1}^{\lfloor sn \rfloor} w(X_i, \hat{\theta}_{0,s}) + \sum_{i=\lfloor sn \rfloor+1}^n w(X_i, \hat{\theta}_{s,1}) - \sum_{i=1}^n w(X_i, \hat{\theta})$$

has the same form as  $\Lambda_{s,n}$  in (4), but without the factor  $\log f_1(X_i, \hat{\lambda}_{\dots,1})$ . Then, we already see that the limit distribution of  $S_n^*$  is obtained with similar arguments that gave us the limit distribution of  $S_n$  in Theorem 3.8. As in Section 3, we start by rewriting  $\Lambda_{s,n}^*$  as follows

$$\Lambda_{s,n}^* = \frac{c_{1,n}}{c_{s,n}} \Lambda_{s,n} - \frac{\frac{1}{n} \sum_{i=1}^n w(X_i, \hat{\theta}) \log f_1(X_i, \hat{\lambda}_1)}{\frac{1}{n} c_{s,n}} (Q_{s,n}^{1*} + Q_{s,n}^{2*} - Q_{1,n}^{1*})$$

where,  $Q_n^{1*} = (Q_{s,n}^{1*})_{s \in [\bar{s}, 1]}$  and  $Q_n^{2*} = (Q_{s,n}^{2*})_{s \in [\bar{s}, 1 - \bar{s}]}$  are càd-làg real-valued processes defined by

$$\begin{aligned}Q_{s,n}^{1*} &:= \sum_{i=1}^{\lfloor sn \rfloor} \left( w(X_i, \hat{\theta}_{0,s}) - w(X_i, \boldsymbol{\theta}) \right) - \mathbf{v}^T \mathbf{I}^{-1} \sum_{i=1}^{\lfloor sn \rfloor} D_\theta(\log f)(X_i, \boldsymbol{\theta}), \quad s \in [\bar{s}, 1], \\ Q_{s,n}^{2*} &:= \sum_{i=\lfloor sn \rfloor+1}^n \left( w(X_i, \hat{\theta}_{s,1}) - w(X_i, \boldsymbol{\theta}) \right) - \mathbf{v}^T \mathbf{I}^{-1} \sum_{i=\lfloor sn \rfloor+1}^n D_\theta(\log f)(X_i, \boldsymbol{\theta}), \quad s \in [\bar{s}, 1 - \bar{s}],\end{aligned}$$

and  $\mathbf{v} := \mathbb{E}_{H_0} [D_\theta w(X_1, \boldsymbol{\theta})] \in \mathbb{R}^d$ . With similar arguments as in Section 3.1,  $Q_{s,n}^{1*}$  and  $Q_{s,n}^{2*}$  can be expressed as functions of the triple  $(\hat{\iota}_{0,s}, \hat{\nu}_{0,s} - \mathbf{v}, \hat{I}_{0,s})$  with  $\hat{\iota}_{0,s}$  and  $\hat{I}_{0,s}$  defined in (9), and

$$\hat{\nu}_{0,s} := \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D_\theta w(X_i, \boldsymbol{\theta}).$$

It follows that  $\Lambda_{s,n}^*$  can be expressed as a function of the quadruple

$$\hat{\xi}_{0,s}^* := (\hat{\iota}_{0,s}, \hat{\nu}_{0,s} - \mathbf{u}, \hat{\nu}_{0,s} - \mathbf{v}, \hat{I}_{0,s})$$

and a random variable that depends on  $s$  and  $n$  and converges a.s. uniformly in  $s$  to some finite constant. We denote by  $\Sigma^*$  the unique positive semi-definite square root of the covariance matrix under the null hypothesis of the quadruple

$$\left( D_\theta(\log f)(X_1, \boldsymbol{\theta}), D_\theta(w \log f_1)(X_1, \boldsymbol{\theta}) - \mathbf{u}, D_\theta w(X_1, \boldsymbol{\theta}) - \mathbf{v}, -D_\theta^2(\log f)(X_1, \boldsymbol{\theta}) - \mathbf{I} \right)$$

reorganized in a  $3d + d^2$ -dimensional real vector. Still under the null hypothesis, the result of Lemma 3.3 can be extended to the process  $(\hat{\xi}_{0,s}^*)_{s \in [\bar{s}, 1]}$  with an application of Donsker's Theorem. With  $\boldsymbol{\xi}^* := (0, 0, 0, \mathbf{I})$ , the process  $\sqrt{n}(\hat{\xi}_{0,s}^* - \boldsymbol{\xi}^*)_{s \in [\bar{s}, 1]}$  converges weakly as follows

$$\sqrt{n} \left( \hat{\xi}_{0,s}^* - \boldsymbol{\xi}^* \right)_{s \in [\bar{s}, 1]} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \left( \frac{1}{s} \Sigma^* W_s \right)_{s \in [\bar{s}, 1]},$$

where  $W := (W_s)_{s \in [0,1]}$  is a standard  $3d + d^2$ -dimensional Brownian motion and  $\Sigma^* W_s$  is reorganized as a quadruple in  $(\mathbb{R}^d)^3 \times gl_d(\mathbb{R})$ .

As for the limit distribution of  $S_n$  in Theorem 3.8, the convergence of  $S_n^*$  is obtained by a functional delta method (Corollary S3.5 given in the supplementary material) and multiple applications of Slutsky's Theorem and the Continuous Mapping Theorem.

To this end, we start by adapting the function  $\mathbf{g}$  defined in (12), for Theorem 3.5. We define the map  $\mathbf{g}^*$  for a quadruple  $(\iota, u, v, I)$  in  $(\mathbb{R}^d)^3 \times gl_d(\mathbb{R})$  by

$$\mathbf{g}^*(\iota, u, v, I) := \left( \iota, u, v, -I^{-1} \left( I - \frac{1}{2} \sum_{l=1}^d \left( \iota^T (I^{-1T})_{\cdot, l} \right) \mathbf{J}_{\dots, l} \right) I^{-1} \right).$$

Remark that, for a fixed  $s$ ,  $\mathbf{g}^*(\Sigma^* W_s)$  is a quadruple in  $(\mathbb{R}^d)^3 \times gl_d(\mathbb{R})$ . Remark also that, when we applied the delta method in the proof of Theorem 3.5, the vector  $u$  did not play a significant role. Here, this is also the case for the vector  $v$ . The result is then extended to the quadruple process  $\sqrt{n}(\hat{\xi}_{0,s}^* - \xi^*)_{s \in [\bar{s}, 1]}$  without additional arguments.

Then, similarly to the function  $q$  defined in (13), we introduce the map  $q^*$ , defined for  $(\iota, u, v, I; A, J, J^*) \in (\mathbb{R}^d)^3 \times (gl_d(\mathbb{R}))^4$  by

$$q^*(\iota, u, v, I; A, J, J^*) = u^T A \iota - \beta v^T A \iota + \iota^T A^T (J - \beta J^*) A \iota + (\mathbf{u}^T - \beta \mathbf{v}^T) I \iota \in \mathbb{R}.$$

Last, we set

$$\mathbf{q}^*(z) := q^*(\mathbf{g}^*(\Sigma^* z); I^{-1}, \mathbf{U}, \mathbf{V}), \quad z \in \mathbb{R}^{3d+d^2}, \quad (17)$$

where  $\Sigma^* z$  is reorganized as a quadruple in  $(\mathbb{R}^d)^3 \times gl_d(\mathbb{R})$  and with  $\mathbf{V}$  the matrix defined by

$$\mathbf{V} := \mathbb{E}_{H_0} [D_{\theta}^2 w(X_1, \theta)]$$

and  $\mathbf{U}$  defined in (15). With the same arguments as in Remark 3.6,  $\mathbf{q}^*$  is also a quadratic form. The following theorem states the limit distribution obtained for  $S_n^*$ .

**Theorem 4.1.** *Under  $H_0$  and Assumptions 2.1-2.5, the test statistic*

$$S_n^* \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \sup_{s \in [\bar{s}, 1-\bar{s}]} \frac{\mathbf{q}^*(W_s - sW_1)}{s(1-s)}$$

where  $(W_s)_{s \in [0,1]}$  is a standard  $3d + d^2$ -dimensional Brownian motion and the application  $\mathbf{q}^*$  is defined in (17).

The test procedure follows the same steps as described in Section 3.3. We will see in Section 6.2 that this extension improves significantly the detection quality. In particular the type II error is smaller compared to the one of the WL test.

**Remark 4.2.** *In Theorems 3.8 and 4.1, and in the results on which they rely, the dimension of the limit Brownian motion  $W$  can be reduced by recognizing that the matrix  $\hat{I}_{0,s}$  defined in (9) is symmetric: it contains  $d(d+1)/2$  distinct elements. This does not seem to have an impact on the numerical properties of the tests.*

In the following section we explicit the test in the case of an univariate Gaussian finite mixture.

## 5 Example: the univariate finite Gaussian mixture

In this section, we assume that the sample  $X = (X_i)_{1 \leq i \leq n}$  follows an univariate Gaussian mixture with  $m$  components. In addition to the weight parameters  $p_1, \dots, p_{m-1}$ , the mixture is defined by the means  $\mu_1, \dots, \mu_m \in \mathbb{R}$  and the standard deviations  $\sigma_1, \dots, \sigma_m \in \mathbb{R}_*^+$  of the  $m$  components. We assume that the set of eligible parameters  $\Theta$  that contains  $\theta$  is the subset of  $\Theta_0 \times (\mathbb{R} \times \mathbb{R}_*^+)^m$  defined by the parameters

$$\theta = (p_1, \dots, p_{m-1}, \mu_1, \sigma_1, \dots, \mu_m, \sigma_m)$$

such that:

1. The means are strictly increasing:  $\mu_1 < \mu_2 < \dots < \mu_m$ .
2. There exists some dispersion boundary  $0 < \mathbf{b} \leq 1$ , deterministic and known, such that the variances are positive and bounded: for all  $i \in \{1, \dots, m\}$ ,  $\sigma_i > 0$  and

$$\min \left\{ \frac{\sigma_j}{\sigma_k}, 1 \leq j, k \leq m \right\} > \mathbf{b}.$$

Note that  $\Theta$  is an open convex<sup>13</sup> subset of  $\mathbb{R}^{3m-1}$ .

**Remark 5.1.** *We impose that the means of the components cannot be equal in order to ensure that the mixture is identifiable and that  $\Theta$  is an open subset of  $\mathbb{R}^{3m-1}$ . If, under  $H_0$ , two means are equal, it is sufficient to define a model that uses one less parameter, i.e. using the same parameter for both means but with different variances.*

The second assumption is a constraint from Hathaway (1985) that ensures the strong consistency of the estimator  $\hat{\theta}$  (Theorem 3.3 in Hathaway (1985)). We obtain the following result.

**Proposition 5.2.** *Under Assumption 2.1 and with the parameter set  $\Theta$  defined above, the validity conditions of Theorems 3.8 and 4.1 hold for a finite Gaussian mixture.*

Remark that Assumption 2.1 does not concern the choice of distribution and therefore remains a preliminary condition to be discussed when applying the WL and EWL tests. In the following section, we give a few applications for the Gaussian case: we start with numerical illustrations of the test, compared to some standard test available in the literature. Then we show how this test can help detect if a change occurs or not in a dataset from the non-life insurance industry.

## 6 Applications

We provide two distinct applications for the case of an univariate finite Gaussian mixture. First, with numerical simulations, we illustrate the properties of the WL and EWL tests compared to a standard *benchmark* test that we shall define in the following subsection. Our main interest lies in the detection of changes in the first component that are not visible to the naked eye (small) but also not too close to 0 (no impact in practice), for large samples (over 10k observations). The second application is an illustration of the WL and EWL tests on a Property and Casualty insurance large dataset (15k observations).

### 6.1 The *benchmark* test

The book of Csörgő and Horváth (1997) gathers standard likelihood-based approaches for the detection of change-points in many different frameworks. For the simple *at most one change* (AMOC) case, one can consider that each parameter  $\theta = (a, b) \in \Theta$  is defined by two sub-parameters  $a$  and  $b$ . We give below a standard likelihood-based hypothesis test that aims to detect if a change occurs in the first sub-parameter  $a$  (see e.g. Section 1.1 in Csörgő and Horváth (1997)). Here  $b$  is called a *nuisance* parameter. We test:

- ◇ the null hypothesis where no change happens, i.e.  $\theta^1 = \dots = \theta^n$ ,
- ◇ against the alternative hypothesis where at most one change occurs, i.e. there exists some  $1 \leq k \leq n$  such that  $a^1 = \dots = a^k \neq a^{k+1} = \dots = a^n$  and  $b^1 = \dots = b^n$ .

The test is defined with the help of the log-likelihood ratio

$$\Lambda_{k,n}^{cs} := \log \left( \frac{\sup_{(a,b),(a',b) \in \Theta} \prod_{i=1}^k f(X_i, (a, b)) \prod_{i=k+1}^n f(X_i, (a', b))}{\sup_{\theta \in \Theta} \prod_{i=1}^n f(X_i, \theta)} \right)$$

---

<sup>13</sup>In particular, for any  $t \in [0, 1]$ , for any  $\theta$  and  $\theta'$  in  $\Theta$ , and for any  $1 \leq j, k \leq m$ , it holds that  $\frac{t\sigma_j + (1-t)\sigma'_j}{t\sigma_k + (1-t)\sigma'_k} > \frac{t\mathbf{b}\sigma_k + (1-t)\mathbf{b}\sigma'_k}{t\sigma_k + (1-t)\sigma'_k} = \mathbf{b}$ .

and the test statistic  $\max_{1 \leq k \leq n} 2\Lambda_{k,n}^{cs}$ . Csörgő and Horváth (1997) provide its limit distribution<sup>14</sup>.

For the detection of a change in the first component for a finite parametric mixture with  $m$  components, we set  $a := (p_1, \lambda_1)$  and  $b := (p_2, \dots, p_{m-1}, \lambda_2, \dots, \lambda_m)$ . In other words, the separation between the parameters of interest  $a$  and the nuisance parameters  $b$  allows the test to focus on a change in the first component. In the WL and EWL tests, this role is played by the weight functions.

This setting for  $a$  and  $b$  also means that we allow the weight parameter of the first component to change ( $p_1 \neq p'_1$ ). Looking at the numerator of  $\Lambda_{k,n}^{BM}$ , for both couples  $(a, b)$  and  $(a', b)$ , the sum of the  $m - 1$  weight parameters has to be strictly below one. We impose first that  $\sum_{k=1}^{m-1} p_k < 1$ . Then, we assume that the relative weight of two components for  $2 \leq i, j \leq m$  is the same before and after  $k$ . The log-likelihood ratio becomes:

$$\Lambda_{k,n}^{BM} := \log \left( \frac{\sup_{a, a', b} \prod_{i=1}^k f(X_i, (a, b)) \prod_{i=k+1}^n f(X_i, (a', b))}{\sup_{\theta \in \Theta} \prod_{i=1}^n f(X_i, \theta)} \right) \quad (18)$$

with  $b' := (\frac{1-p'_1}{1-p_1} p_2, \dots, \frac{1-p'_1}{1-p_1} p_{m-1}, \lambda_2, \dots, \lambda_m)$ . Thus the sum of the  $m - 1$  weight parameters of the couple  $(a', b')$  is also below one. We refer to the test defined by the test statistic  $\max_{1 \leq k \leq n} 2\Lambda_{k,n}^{BM}$  as the *benchmark* test.

## 6.2 Illustration of numerical properties

We illustrate our analysis with simulations of a Gaussian mixture with 3 components such that:

- ◇ The sample size is large:  $n = 1k$  or  $10k$  observations (obs.) as we are interested in applications to large datasets (see the application on real data in Section 6.3).
- ◇ Under the null hypothesis, the mixture used for the illustrations has equal weight parameters (1/3), equal standard deviations (0.25) and respective means -1, 0 and 1. The empirical density of the mixture shows clearly the three components.
- ◇ The change occurs in middle of the sample ( $s = 0.5$ )<sup>15</sup>.
- ◇ The detection threshold is the 90% percentile of test statistic under the null hypothesis (type I error).

The results are obtained by multiple simulations of the random sample with standard algorithms from the *R* software. Note that the detection thresholds are computed with the help of the limit distributions from Theorems 3.8 and 4.1, which is up to 10 000 times faster than re-simulations of the detection statistic under the null hypothesis.

### Results: detection quality under the alternate hypothesis, i.e. a change occurs in the parameters of the first component

The alternate hypothesis is defined in Section 2.1 as the case when the sample contains one change in the first component: the parameters which describe the distribution of the first component are different before and after the change-point while the other parameters of the mixture remain the same. In this application, we illustrate the potential properties of the WL and EWL tests through three kind of changes in the parameters of the first component: a shift between -1 and +1 of the mean, a shift between -0.2 and +0.5 of the standard deviation and a shift between -0.25 and +0.25 of the weight parameter.

The type II error (proportion of false negative) of each test is obtained from multiple re-simulations. Table 1 gives the key impacts that we observed (detailed results are given in Figure S2 in the supplementary material). Note that a high-performance test is characterized by a low type II error.

<sup>14</sup>See Corollary 1.1.1, Theorems 1.3.1. and 1.3.2 in Csörgő and Horváth (1997).

<sup>15</sup>The simulations for the numerical comparison given in this section require an important run time. In order to reduce significantly this run time, we computed the illustrations for  $s = 0.5$  (statistic and threshold) and not over the whole process. We made sure that this does not effect the conclusions. The application on the real data in Section 6.3 is based on a computation of the whole process.

in %	n=1k obs.			n=10k obs.		
	WL	EWL	BM	WL	EWL	BM
Mean +0.1	44.4	21.8	42.4	<b>0.4</b>	<b>0.8</b>	30.8
Mean -0.1	42.4	19.0	29.8	<b>0.8</b>	<b>1.2</b>	50.8
Std dev +0.1	33.0	9.2	73.0	<b>0.4</b>	<b>0.6</b>	43.4
Std dev -0.1	9.8	1.0	22.6	<b>1.0</b>	<b>0.2</b>	45.4
Weight +0.1	67.2	42.0	40.8	<b>0.2</b>	<b>0.6</b>	55.4
Weight -0.1	51.8	27.4	34.2	<b>0.6</b>	<b>0.2</b>	27.8

Table 1: Type II error (in %) for a change in the first component (500 re-simulations). Results are given respectively for the WLT, the EWLT and the benchmark test (BM).

We remind that our main interest lies in the detection of changes in the first component that are not visible to the naked eye (small) but also not too close to 0 (no impact in practice), for large samples (over 10k observations). In that sense, Table 1 indicates that the WLT performs significantly better than the benchmark test for large samples. The EWLT improves the performance of the WLT and makes it the better choice for almost all cases for 1k obs. samples and for all the cases for 10k obs.

The benchmark test fails to detect properly the change, especially with large samples. To our understanding, this is mainly due to the optimization problem in the numerator of  $\Lambda_{k,n}^{BM}$  that the algorithm often fails to solve. Since their estimation algorithms are more robust, the WL and EWL tests have both very low type II errors for small changes in the parameters on a sample of 10k observation.

We also see in Figure S2 in the supplementary material that there is still room for improvement for large changes in the weight and the standard deviation, even if large changes do not have a strong importance since they are visible to the naked eye.

In addition to these performance results under the alternative hypothesis, Figure S5 in the supplementary material shows that, as the sample size increases, the run time needed to compute the benchmark test increases considerably faster compared to the WLT. It follows that one major advantage of the WLT is that it can be computed quickly with standard algorithms, making it especially convenient for large datasets.

We conclude that the WLT is a valid candidate for the detection of a change in the first component of a Gaussian mixture. The EWLT is an improved version that significantly reduces the type II error.

### Results: when a change occurs in the second or the third component

As a nice-to-have, we also studied the detection frequency of each test when a change occurs in the second or third component. We used the same shift ranges as the ones used for a change in the first component. In this context, a high-performance test is characterized by a low detection frequency. The results are given in Table 2 with detailed graphs in Figures S3 and S4 in the supplementary material.

	n=1k obs.			n=10k obs.		
	WL	EWL	BM	WL	EWL	BM
<i>Change in the second component</i>						
Mean +0.1	13	11	25	<b>16</b>	<b>8</b>	59
Mean -0.1	15	10	29	<b>23</b>	<b>6</b>	41
Std dev +0.1	44	20	28	<b>99</b>	<b>61</b>	43
Std dev -0.1	62	44	40	<b>100</b>	<b>100</b>	80
Weight +0.1	7	9	27	<b>14</b>	<b>13</b>	61
Weight -0.1	15	16	13	<b>16</b>	<b>14</b>	24
<i>Change in the third component</i>						
Mean +0.1	13	14	10	<b>15</b>	<b>13</b>	28
Mean -0.1	15	16	13	<b>15</b>	<b>14</b>	65
Std dev +0.1	11	11	10	<b>90</b>	<b>68</b>	18
Std dev -0.1	6	6	12	<b>7</b>	<b>6</b>	33
Weight +0.1	13	12	15	<b>11</b>	<b>10</b>	29
Weight -0.1	8	9	28	<b>12</b>	<b>12</b>	66

Table 2: Detection frequency (in %) for a change in the second and third components (500 re-simulations). Results are given respectively for the WLT, the EWL and the benchmark test (BM).

The main observation is that the EWL shows the best results and, for large samples of 10k observations, the WL is better than the benchmark test. The WL and EWL tests tend to detect a change when there is a change in the standard deviation of the second or third component. This is explained by the role of the weight functions in their detection statistic that zoom on some range of values around the mean of the first component: there are significantly more (or less) points that enter this range, increasing the detection frequency. We still observe that the benchmark test behaves poorly for large samples.

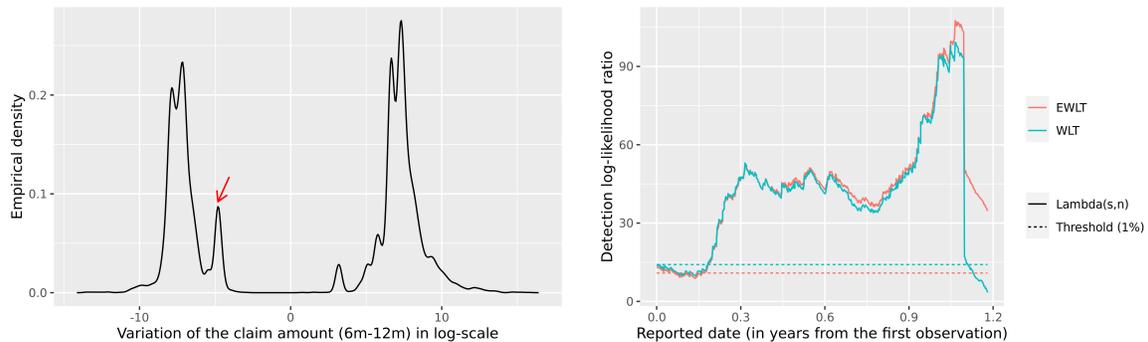
One could expect that a high-performance test would have a detection frequency that stays close to the type I error (10%). This is the case for the WL and EWL tests when a change occurs in the weight parameter (Figures S3f and S4f in the supplementary material). However, from Figure S3 in the supplementary material, we remark that the three tests show systematic patterns for a wide range of possible deviations for a change in the mean and the standard deviation. It follows that there is still some room for improvement regarding this criterion.

In the next section, we study briefly an illustration to a Property and Casualty insurance large dataset (15k observations).

### 6.3 Illustration of the WL and EWL tests on P&C insurance data

In this section, we give an example of application of these two tests to a problem from the insurance industry, in particular the bodily injuries from the motor claims. Each claim is known by the insurer from its declaration date that is the starting point of observation: we set  $t = 0$ ,  $t$  being expressed in years. We denote by  $C_t$  the real-valued random variable that represents the amount that, at time  $t$ , the insurer expects to pay eventually. This amount varies over time when the claim is reviewed, until it is settled. The evolution of the amount  $C_t$  depends on structural factors (e.g. the type of injury), internal factors (e.g. a change in the revision policy) and external factors (e.g. new elements are known by the insurer, a court decision sets the final amount, etc.). For bodily injuries, the claims takes time to resolve (over 2 years in average). After the settlement, we assume that  $C_t$  is fixed and possibly null or negative.

In this application, we know that a change of the revision process happened at some point in the past. The question is then to determine whether or not this change impacted significantly the observed variations of claim amount over time. For that purpose, we consider the random variable  $Z = \text{sgn}(C_1 - C_{0.5}) \log(1 + |C_1 - C_{0.5}|)$  that gives the variation of the claim amount between the 6th and the 12th month in log-scale, where  $\text{sgn}(\cdot)$  is the function that gives the sign of a real number taking respectively the values -1, 0 and 1 when this number is negative, null or positive. From a first analysis of the data, a Kolmogorov-Smirnov hypothesis test does not reject the assumption that observed realizations of  $Z$  before the change follow a finite parametric mixture with 12 components (Figure 1a).



(a) Empirical density of the variation in claims amount between the 6th and the 12th development months, before the change of process (7k obs). (b) Log-ratio processes  $\Lambda_n$  and  $\Lambda_n^*$  of the WL and EWL tests for the detection of a change in the 5th component of the mixture (15k obs).

Figure 1: Application of the WL and EWL tests to a change in the variation of incurred claims amount (Motor bodily injuries claims).

For internal reasons, the insurance company is particularly interested in the 5th component of the mixture, highlighted in Figure 1a by a red arrow. This component represents slight decreases of claim amounts. We applied the WL and EWL tests to a sample of 15k claims where the first third of the sample is known to contain claims that are not impacted by the change of process. Under the null hypothesis that no change occurs in the 5th component of the mixture, both tests reject this hypothesis with a p-value below  $10^{-4}$ . Figure 1b illustrates the underlying processes  $\Lambda_n$  and  $\Lambda_n^*$  and their respective thresholds: the change is significant and, according to both tests, it seems to occur from the time 1.07.

This conclusion allowed the insurance company to investigate further the quantification of the change.

This application shows that the WL and EWL tests can be used in the industry for the monitoring of changes, when they are unexpected but also to assess their significativity when they are known or suspected. For other topics that tackle change-point problems in non-life insurance, we refer for example to Dhaene et al. (2002), Andersen et al. (2009), Kwon and Vu (2017), Peřtová and Peřta (2017) or Maciak et al. (2020).

## 7 Conclusion

In this article, we consider a closed sample of independent random variables that follow a finite mixture distribution with parametric components. The sample might contain at most one change in the parameters of the first component. If there is a change, the r.v. are identically distributed before and after the change-point: the parameters which describe the distribution of the first component are different before and after the break-point while the other parameters of the mixture remain the same.

To test whether there is a change or not, we proposed two alternative tests (WLT and EWLT). Each test statistic is built upon a càd-làg process obtained from a likelihood ratio (see (4) and (16)). The specificity of these tests is that they can be computed using known inference algorithms. The WLT uses weight functions to help the likelihood ratio to *zoom* on the first component. In the EWLT, we added an adjustment that helps improve the type II error, as explained in Section 6.2.

With a technique from Davis et al. (1995), we derived in Theorems 3.8 and 4.1 the limit distribution of the test statistics under the null hypothesis in the form of a quadratic form of a multidimensional Brownian motion, with the help of a dedicated functional limit theorem. In particular, the proof is based on a functional delta method<sup>16</sup> and multiple applications of Slutsky's Theorem and the Continuous Mapping Theorem. We showed in Section 5 that validity conditions of the main result hold for univariate finite Gaussian mixtures

<sup>16</sup>See Corollary S3.5 in the supplementary material.

within the framework of Hathaway (1985).

Numerical applications on simulated data for the Gaussian case showed that second version of the test outperforms significantly a benchmark test exposed in Csörgő and Horváth (1997) and defined by (18): the type II error is considerably reduced (when a change occurs in the first component) and the detection frequency remains low in most cases when a change occurs in another component.

Two issues of the benchmark test are that usual optimization algorithms have an unrealistic convergence run time for large samples, and that they fail to compute properly its statistic. However, in the case of simulated data, we assess that, without this computational issue, the benchmark test would have a lower type II error than our tests. Therefore the extended version of our tests remains so far the best candidate even if a dedicated algorithm for computing more robustly the statistic of the benchmark test would be an improvement.

In addition, in some cases, the three tests still detect a change when a change occurs in another component (i.e. not the first one). Extensions of our work could consider adding a penalization term to the likelihood ratio in order to improve this aspect.

We end the applications by a brief illustration of the proposed tests on variations of claim amounts for bodily injuries motor claims (real data), in the context of a change of process in the claims handling department of an insurance company. A change is detected in the fifth component of the Gaussian mixture with 12 components: the insurance company could therefore assess the change and investigate further its causes.

From the numerical applications, the WL and EWL tests are valid candidates when looking for a change in one component of a finite parametric mixture. In addition, the results obtained under the null hypothesis in Theorems 3.8 and 4.1 allow us to reduce significantly the calibration run time of the detection thresholds: the marginal run time of one simulation is divided by 10 000. Beyond these promising results, the possibilities for other techniques still exist and are worth to be explored.

## References

- Amisano, G. and Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *J. Bus. Econom. Statist.*, 25(2):177–190.
- Andersen, T. G., Davis, R. A., Kreiß, J.-P., and Mikosch, T. V. (2009). *Handbook of financial time series*. Springer-Verlag Berlin Heidelberg.
- Barber, D., Cemgil, A. T., and Chiappa, S., editors (2011). *Bayesian time series models*. Cambridge University Press, Cambridge.
- Basu, A., Shioya, H., and Park, C. (2011). *Statistical inference*, volume 120 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL. The minimum distance approach.
- Benaglia, T., Chauveau, D., Hunter, D., and Young, D. (2009). mixtools: An r package for analyzing finite mixture models. *Journal of Statistical Software*, 32.
- Billingsley, P. (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition. A Wiley-Interscience Publication.
- Chen, J. (2017). Consistency of the MLE under mixture models. *Statist. Sci.*, 32(1):47–63.
- Chen, J. and Gupta, A. K. (2012). *Parametric statistical change point analysis*. Birkhäuser/Springer, New York, second edition. With applications to genetics, medicine, and finance.
- Coleman, R. (2012). *Calculus on normed vector spaces*. Universitext. Springer, New York.
- Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. With a foreword by David Kendall.
- Davis, R. A., Huang, D. W., and Yao, Y.-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *Ann. Statist.*, 23(1):282–304.

- Dehling, H., Franke, B., Kott, T., and Kulperger, R. (2014). Change point testing for the drift parameters of a periodic mean reversion process. *Stat. Inference Stoch. Process.*, 17(1):1–18.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–38. With discussion.
- Dhaene, J., Dhaene, G., Beirlant, J., Claeskens, G., Croux, C., Degryse, H., Dewachter, H., Gijbels, I., Goovaerts, M., Hubert, M., Roodhooft, F., Schouten, W., and Willekens, M. (2002). Managing uncertainty: Financial, actuarial and statistical modeling. *Review of Business and Economics*, L:23–48.
- Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *Ann. Math. Statist.*, 42:204–223.
- Feng, Z. D. and McCulloch, C. E. (1996). Using bootstrap likelihood ratios in finite mixture models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(3):609–617.
- Frühwirth-Schnatter, S., Celeux, G., and Robert, C. P., editors (2019). *Handbook of mixture analysis*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL.
- Ganji, M. and Mostafayi, R. (2019). Bayes and non-Bayes estimation of change point in nonstandard mixture inverse Weibull distribution. *J. Stat. Theory Appl.*, 18(1):79–86.
- Giordani, P. and Kohn, R. (2008). Efficient Bayesian inference for multiple change-point and mixture innovation models. *J. Bus. Econom. Statist.*, 26(1):66–77.
- Hathaway, R. J. (1983). *Constrained maximum likelihood estimation for a mixture of  $m$  univariate normal distributions*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Rice University.
- Hathaway, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *Ann. Statist.*, 13(2):795–800.
- Hu, F. and Zidek, J. V. (2002). The weighted likelihood. *Canad. J. Statist.*, 30(3):347–371.
- Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.*, 27:887–906.
- Killick, R. and Eckley, I. A. (2014). changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, 58(3):1–19.
- Kwon, H. S. and Vu, U. Q. (2017). Consideration of a structural-change point in the chain-ladder method. *Communications for Statistical Applications and Methods*, 24(3):211–226.
- Lachos Dávila, V. H., Cabral, C. R. B., and Zeller, C. B. (2018). *Finite mixture of skewed distributions*. SpringerBriefs in Statistics. Springer, Cham. SpringerBriefs in Statistics—ABE.
- Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition.
- Li, Q., Guo, F., Kim, I., Klauer, S. G., and Simons-Morton, B. G. (2018). A Bayesian finite mixture change-point model for assessing the risk of novice teenage drivers. *J. Appl. Stat.*, 45(4):604–625.
- Maciak, M., Pešta, M., and Peštová, B. (2020). Changepoint in dependent and non-stationary panels. *Statist. Papers*, 61(4):1385–1407.
- Mäkeläinen, T., Schmidt, K., and Styan, G. P. H. (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. *Ann. Statist.*, 9(4):758–767.
- McLachlan, G. and Peel, D. (2000). *Finite mixture models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York.
- Pandya, M. and Jadav, P. V. (2009). Bayes estimation of change point in non-standard mixture distribution of inverse Weibull with unknown proportions. *J. Indian Statist. Assoc.*, 47(1):41–62.

- Pandya, M. and Jadav, P. V. (2010). Bayes estimation of change point in non standard mixture of left-truncated exponential with unknown proportions. *Comm. Statist. Theory Methods*, 39(15):2725–2742.
- Peřtová, B. and Peřta, M. (2017). Change point estimation in panel data without boundary issue. *Risks*, 5(1):7.
- Pons, O. (2009). *Estimation et tests dans les modèles de mélanges de lois et de ruptures*. Hermes Science Publications/Lavoisier, Paris.
- Pons, O. (2018). *Estimations and tests in change-point models*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ.
- Redner, R. (1981). Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *Ann. Statist.*, 9(1):225–228.
- Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.*, 26(2):195–239.
- Song, Z., Liu, Y., Li, Z., and Zhang, J. (2018). A weighted likelihood ratio test-based chart for monitoring process mean and variability. *J. Stat. Comput. Simul.*, 88(7):1415–1436.
- Tanaka, K. (2009). Strong consistency of the maximum likelihood estimator for finite mixtures of location-scale distributions when penalty is imposed on the ratios of the scale parameters. *Scand. J. Stat.*, 36(1):171–184.
- van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.
- Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statistics*, 20:595–601.
- Wilson, R. C., Nassar, M. R., and Gold, J. I. (2013). A mixture of delta-rules approximation to Bayesian inference in change-point problems. *PLoS Comput. Biol.*, 9(7):e1003150, 18.
- Wu, C.-F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.*, 11(1):95–103.
- Zou, C., Fang, X., and Zhai, G. (2015). Ovarian cancer screening based on mixture change-point model. *J. Syst. Sci. Complex.*, 28(2):471–488.