Asymptotic distribution and power of the likelihood ratio test for mixtures: bounded and unbounded cases.

Jean-Marc Azaïs [1] Élisabeth Gassiat[2*] Cécile Mercadier[1]

September 20, 2004

[1] Laboratoire de Statistique et Probabilités, UMR-CNRS C55830, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 4. France.
[2] Laboratoire de Mathématiques, Equipe de Probabilités, Statistique et Modélisation, Bat425, Université de Paris-Sud, 91405 Orsay Cedex, France. elisabeth.gassiat@math.u-psud.fr.

Abstract

In this paper, we consider the log-likelihood ratio test (LRT) for testing the number of components in a mixture of populations in a parametric family. We provide the asymptotic distribution of the LRT statistic under the null hypothesis as well as under contiguous alternatives when the parameter set is bounded. Moreover, for the simple contamination model we prove that, under general assumptions, the asymptotic power under contiguous hypotheses may be arbitrarily close to the asymptotic level when the set of parameters is large enough. In the particular problem of normal distributions, we prove that, when the unknown mean is not a priori bounded, the asymptotic power under contiguous hypotheses is equal to the asymptotic level.
Keywords: Likelihood ratio test, mixture models, number of components, extreme values, power, contiguity.
Short title: Asymptotic study of the LRT for mixtures

1 Introduction

Mixtures of populations is a modelling tool widely used in applications and the literature on the subject is vast. For finite mixtures, the first task is the choice of the number of components in the mixture. Some estimation or testing procedures have been proposed for this purpose, see for instance the books of Titterington et al. (1985), Lindsay (1995) and McLachlan and Peel (2000) or the papers of James et al. (2001), Gassiat (2002) and references therein. Asymptotic optimality of the likelihood ratio test (LRT) in several parametric contexts is well known. Using the LRT for testing the number of components in a mixture appears quite natural. In one way, simulation studies show that the LRT performs well in various situations (see Goffinet et al., 1992). In another way, the asymptotic distribution and power of the test have to be evaluated to compare with other known tests.

In this paper, we focus on the asymptotic properties of the LRT for testing that i.i.d. observations $X_1, \ldots, X_n$ come from a mixture of p_0 populations in a parametric set of densities $\mathcal{F}$ (null hypothesis H_0) against a mixture of p populations (alternative H_1), where the integers p_0 and p satisfy $p_0 < p$.

In Section 2 we apply results of Gassiat (2002) to obtain the asymptotic distribution of the LRT statistic for testing (H_0) against (H_1) under the null hypothesis as well as under contiguous alternatives. Indeed, Gassiat (2002) gives a quite weak assumption under which the derivation of the asymptotic distribution of the LRT statistic is made in the general situation when one has to test a small model in a larger one, under the null hypothesis as well as under contiguous hypotheses. This applies to the number of components in a mixture of populations in a parametric set with eventually an unknown nuisance parameter. For instance, we apply the method to multidimensional Gaussian distributions with unknown common variance. By this way, we recover known results for mixtures of one or two populations but under weaker assumptions, or known results concerning particular parametric families such as Gaussian or Binomial distributions; see Ghosh and Sen (1985), Dacunha-Castelle and Gassiat (1997, 1999), Garel (2001), Lemdani and Pons (1997, 1999), Chen and Chen (2001), Mosler and Seidel (2001) Chernoff and Lander (1995). We also obtain more general results than previous ones:

- They apply to general sets of parametric families with unknown nuisance parameter.
- Asymptotic distribution under contiguous alternatives is considered.

However, except for smoothness assumptions, the main point is that these asymptotic results require that the parameter set is bounded.

In Sections 3 and 4 we study what happens when the set of parameters becomes larger and larger. For simplicity we restrict our attention to the simplest model: the contamination model for family of distributions indexed by a single real parameter. Indeed, roughly speaking, the LRT statistic converges in distribution to half the square of the supremum of some Gaussian process indexed by a compact set of scores. But when this set of scores is enlarged, the covariance of the Gaussian process is close to 0 for sufficiently distant scores, so that the supremum of the Gaussian process may become arbitrarily large. Thus one also knows that for unbounded sets of parameters, the LRT statistic tends to infinity in probability, as Hartigan first noted for normal mixtures (see Hartigan, 1985). Here, we prove that under some extreme circumstances the LRT can have less power than moment tests or goodness-of-fit tests. At the end of the introduction we draw carefully practical conclusions from this result.

More precisely, let $\mathbb{T}$ be $[-T, T]$ and $\mathcal{F} = \{f_t \ , \ t \in \mathbb{T}\}$ be a parametric set of probability densities on $\mathbb{R}$ with respect to the Lebesgue measure. Using i.i.d. observations $X_1, \ldots, X_n$, we consider the testing problem for the density g of the observations.

$$(H_0) : \text{“}g = f_0\text{”} \quad \text{against} \quad (H_1) : \text{“}g = (1-\pi)f_0 + \pi f_t, \ 0 \le \pi \le 1, t \in \mathbb{T}\text{”}. \tag{1}$$

We prove that:

- For general parametric sets $\mathcal{F}$, $\mathbb{T} = [-T, T]$ and T large enough, under contiguous alternatives, the LRT for (1) has asymptotic power close to the asymptotic level, under some smoothness assumptions, see Theorem 7.
  A set of assumptions is given for which Theorem 7 applies in the case of translation mixtures, that is when $f_t(\cdot) = f_0(\cdot - t)$, see Corollary 1. This is done in Section 3.
- When f_t is the standard Gaussian with mean t we get the normal mixture problem. When the set of means is not a priori bounded, that is $\mathbb{T} = \mathbb{R}$, Liu and Shao (2004) obtained the asymptotic distribution of the LRT under the null hypothesis by using the strong approximation proved in Bickel and Chernoff (1993). We prove in Theorem 8 of Section 4 that the asymptotic power under contiguous alternatives is equal to the asymptotic level.

The way to obtain these results is to gather together: expansion of the LRT obtained in Gassiat (2002) to identify contiguity and apply Le Cam's third Lemma (see van der Vaart, 1998), behaviour of the supremum of a Gaussian process on an interval with bounds tending to infinity as obtained in Azaïs and Mercadier (2004), and the normal comparison inequality as refined in Li and Shao (2002). Proof of most results of Sections 3 and 4 are detailed in Section 5.

Independently of our work, for the Gaussian model with unbounded means, Hall and Stewart (2004) obtained the speed of separation of alternatives that ensures asymptotic power to be bigger than asymptotic level. Their result indicates that it should be $\sqrt{\log \circ \log n}/\sqrt{n}$ contrary to the classical parametric situation, where $1/\sqrt{n}$ is the speed of separation.

Practical application

Such tests that have power less or equal to level are sometimes called "worthless" (see for example van der Vaart, 1998) but this word would be dangerous because practical interpretation of our result must take into account the following points:

- It is well known that for mixtures of population in general the convergence to the asymptotic distribution is very slow. For example for a very simple test, as the skewness test, Boistard (2003) showed that $n = 10^3$ observations are needed to meet the asymptotic distribution.
- For maximum likelihood estimates (MLE) and tests, the problem of the speed of convergence to the asymptotic distribution is very difficult to address since in practice MLE are computed through iterative algorithms and are only approximative. The most famous one is the EM algorithm and its variants. All these algorithms depend on tuning constants, in particular concerning the stopping rule. It is shown for example in table 6.3 of McLachlan and Peel (2000) (based on results by Seidel et al., 2000) that the distribution of the LRT depends heavily on these tuning constants. Simulation results by Liu and Shao (2004) suggest that their asymptotic distribution is not met for $n = 5.10^3$ observations.
- Nowadays some results and softwares are available to compute the distribution of the maximum of Gaussian processes. See for example Garel (2001), Delmas (2003) and Mercadier (2004). In particular these results show that, as soon as the means are contained in some "not huge" set, the asymptotic power under contiguous alternatives of the LRT is generally better than that of moment tests or of goodness-of-fit tests. Nevertheless, the LRT is not uniformly most powerful.

Our result that shows that the LRT is asymptotically less powerful than moment tests is valid in practice only for very large data sets. For all the reasons above it will be very difficult to say precisely when. Simulations have proved that in practice LRT based on Monte-Carlo calculation of threshold or bootstrapping behave well (see Goffinet et al., 1992) for unbounded parameter.
Our opinion is that the main consequence of our result for large or unbounded parameter sets is that the study of the LRT for mixtures in the compact case seems to be the more relevant case.

2 Asymptotic distribution of the LRT for the number of populations in a mixture under null and contiguous hypotheses.

A general theorem in Gassiat (2002) allows to find the asymptotic distribution of the LRT for testing a particular model in a larger one, under the null hypothesis as well as under contiguous alternatives. Roughly speaking, the asymptotic distribution is some function of the supremum of the isonormal process on a set of score functions. The theorem holds under a simple assumption on the bracket entropy of an enlarged set. In many applications, those sets are parameterized by a finite dimensional parameter. In such cases,

- Lipschitz properties allow to compute easily bracket entropies, as in van der Vaart (1998, p. 271). We give some examples in the text.
- The covariance structure of the isonormal process may be computed in an explicit way and identified with the covariance function of a Gaussian field with real parameters.

We shall describe in this section how it applies to mixture models. We first recall the general result of Gassiat (2002) and its application to a very simple contamination mixture model: one has to test between a particular known population with some density $f_{\mathbf{0}}(\cdot)$ and a mixture of this known one and another with density $f_{\mathbf{t}}(\cdot)$, $\mathbf{t}$ a multidimensional parameter. Then we detail the case of two populations with eventually unknown nuisance parameter. A typical example will be that of translation mixtures with eventually unknown scale parameter. We end the section by giving general considerations on how to deal with parametric mixture models, allowing an unknown nuisance parameter and setting a general result in such situations.

Assume one would like to use the LRT for testing (H_0) : "$g \in \mathcal{M}_0$" against (H_1) : "$g \in \mathcal{M}$", where g is the generic density of i.i.d. observations $X_1, \ldots, X_n$, $\mathcal{M}_0 \subset \mathcal{M}$ are sets of densities with respect to some measure ν on $\mathbb{R}^k$ (or more generally on some Polish space). Let $\ell_n(g) = \sum_{i=1}^n \log g(X_i)$ be the log-likelihood, and

$$\lambda_n = \sup_{g \in \mathcal{M}} \ell_n(g) - \sup_{g \in \mathcal{M}_0} \ell_n(g)$$

be the LRT statistic. Let also g_0 be a density in $\mathcal{M}_0$ that will denote the true (unknown) density of the observations. In the first considered examples, and without loss of generality, we will assume that g_0 coincides with f_0.
Throughout the paper we use $\|\cdot\|_2$ to denote the norm in $L^2(g_0 \cdot \nu)$.
When studying $\ell_n(g) - \ell_n(g_0)$, functions $\frac{g-g_0}{g_0}$ appear naturally. Define the set $\mathcal{S}$ as the subset of the unit sphere in $L^2(g_0 \cdot \nu)$ of such functions when normalized:

$$\mathcal{S} = \left\{ \frac{g - g_0}{g_0} / \| \frac{g - g_0}{g_0} \|_2, \ g \in \mathcal{M} \setminus \{g_0\} \right\}, \tag{2}$$

and $\mathcal{S}_0$ its subset when $g \in \mathcal{M}_0$:

$$\mathcal{S}_0 = \left\{ \frac{g - g_0}{g_0} / \| \frac{g - g_0}{g_0} \|_2, \ g \in \mathcal{M}_0 \setminus \{g_0\} \right\}. \tag{3}$$

A bracket $[L, U]$ of length ϵ is the set of functions b such that $L \leq b \leq U$, where L and U are functions in $L^2(g_0 \cdot \nu)$ such that $\|U - L\|_2 \leq \epsilon$. Define $H_{[],2}(\mathcal{S}, \epsilon)$ the entropy with bracketing of $\mathcal{S}$ with respect to the norm $\|\cdot\|_2$, as the logarithm of the number of brackets of length ϵ needed to cover $\mathcal{S}$. To apply the theorem in Gassiat (2002), the only needed assumption is:

$$\int_0^1 \sqrt{H_{[],2}(\mathcal{S}, \epsilon)} d\epsilon < +\infty. \tag{4}$$

This assumption implies in particular that $\mathcal{S}$ is Donsker and that its closure is compact. As said before, when $\mathcal{M}$ is parameterized, $\mathcal{S}$ is also parameterized and smoothness properties will allow to verify (4). But in general the parameterization will not be continuous throughout $\mathcal{S}$. The delicate point may be that one has to find all possible limit points, in $L^2(g_0 \cdot \nu)$, of sequences $\frac{g_n - g_0}{g_0} / \| \frac{g_n - g_0}{g_0} \|_2$ when $\| \frac{g_n - g_0}{g_0} \|_2$ tends to 0. The set $\mathcal{D}$ (resp. $\mathcal{D}_0$) of limit points of sequences $\frac{g_n - g_0}{g_0} / \| \frac{g_n - g_0}{g_0} \|_2$ where $\| \frac{g_n - g_0}{g_0} \|_2$ tends to 0, $g_n \in \mathcal{M} \setminus \{g_0\}$ (resp. $g_n \in \mathcal{M}_0 \setminus \{g_0\}$) will be parameterized in such a way that Lipschitz properties can be used on subsets.
Let us for example see how it applies to the simple contamination mixture model (1). In this case,

$$\mathcal{M}_0 = \{f_0\}, \ \mathcal{M} = \{g_{\pi,t} = (1 - \pi) f_0 + \pi f_t, \ 0 \leq \pi \leq 1, \ t \in [-T, T]\}$$

for a given positive real number T. Since $\mathcal{M}_0$ is a singleton, we do not need to define $\mathcal{S}_0$ and $\mathcal{D}_0$. One has $\frac{g_{\pi,t} - g_0}{g_0} = \pi \frac{f_t - f_0}{f_0}$, so that

$$\mathcal{S} = \left\{ d_t = \frac{\frac{f_t - f_0}{f_0}}{\| \frac{f_t - f_0}{f_0} \|_2}, \ \ t \in [-T, 0) \cup (0, T] \right\}.$$

If $\| \frac{f_t - f_0}{f_0} \|_2 = 0$, which occurs if and only if $t = 0$, then under smoothness assumptions $\| \frac{g_{\pi_n, t_n} - g_0}{g_0} \|_2$ tends to 0 if and only if π_n or t_n tends to 0. Then d_{t_n} has two possible limit points (depending on the sign of t_n), and

$$\mathcal{D} = \left\{ d_t, \ t \in [-T, 0) \cup (0, T], d_{0^-} = \frac{-\frac{f_0'}{f_0}}{\| \frac{f_0'}{f_0} \|_2}, d_{0^+} = \frac{\frac{f_0'}{f_0}}{\| \frac{f_0'}{f_0} \|_2} \right\}.$$

Here derivatives are taken with respect to parameter t. Again under smoothness assumptions, it will be possible to prove, considering $\{d_t, t \in [-T, 0), d_{0^-}\}$ and $\{d_t, t \in (0, T], d_{0^+}\}$ that the number of brackets of length ϵ needed to cover $\mathcal{S}$ is of order at most $O(1/\epsilon)$, so that Assumption (4) holds. (A complete proof is given below for contamination models with multidimensional parameterization).

In general when $\mathcal{M}_0$ contains more than one density, $\mathcal{D}_0 \subset \mathcal{D}$, and if the parameterization is smooth enough, it will be possible to define a set $\mathbb{U}$ in $\mathbb{R}^{k_0} \times \mathbb{R}^{k_1}$ and a set $\mathbb{U}_0$ in $\mathbb{R}^{k_0}$ such that

$$\mathcal{D} = \{d_{\mathbf{u}}, \mathbf{u} \in \mathbb{U}\} \text{ and } \mathcal{D}_0 = \{d_{(\mathbf{v},\mathbf{0})}, \mathbf{v} \in \mathbb{U}_0\}.$$

Define the covariance function $r(\cdot,\cdot)$ on $\mathbb{U} \times \mathbb{U}$ by

$$r(\mathbf{u}_1, \mathbf{u}_2) = \int d_{\mathbf{u}_1} d_{\mathbf{u}_2} g_0 d\nu.$$

Then, under (4), applying Theorem 3.1 in Gassiat (2002),

$$2\,\lambda_n = \sup_{\mathbf{u}\in\mathbb{U}} \left(\max\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} d_{\mathbf{u}}(X_i), 0 \right\} \right)^2 - \sup_{\mathbf{v}\in\mathbb{U}_0} \left(\max\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} d_{(\mathbf{v},\mathbf{0})}(X_i), 0 \right\} \right)^2 + o_{\mathbb{P}_0}(1),$$

so that $2\,\lambda_n$ converges in distribution to

$$\sup_{\mathbf{u}\in\mathbb{U}} \left(\max\left\{Z(\mathbf{u}), 0\right\}\right)^2 - \sup_{\mathbf{v}\in\mathbb{U}_0} \left(\max\left\{Z(\mathbf{v},\mathbf{0}), 0\right\}\right)^2 \tag{5}$$

where $Z(\cdot)$ is the Gaussian process on $\mathbb{U}$ with covariance $r(\cdot,\cdot)$ and $\mathbb{P}_0$ is the joint distribution of the observations $X_1, ..., X_n$ under the null hypothesis. In the particular case when $\mathcal{M}_0$ is reduced to a single element, a direct application of Corollary 3.1 of Gassiat (2002) gives that $2\,\lambda_n$ converges in distribution to

$$\sup_{\mathbf{u}\in\mathbb{U}} \left(\max\left\{Z(\mathbf{u}), 0\right\}\right)^2. \tag{6}$$

It will be seen in the examples below that $r(\cdot,\cdot)$ is in general not continuous everywhere on the closure of $\mathbb{U} \times \mathbb{U}$. $Z(\cdot)$ is not a continuous Gaussian field, though the isonormal process on $\mathcal{D}$ is continuous, so that the suprema involved in (5) are a.s. finite. In general, $r(\cdot,\cdot)$ is continuous almost everywhere. In the simple contamination mixture model (1), for non null s and t,

$$r(s,t) = \int \left(\frac{\frac{f_t - f_0}{f_0}}{\| \frac{f_t - f_0}{f_0} \|_2} \right) \left(\frac{\frac{f_s - f_0}{f_0}}{\| \frac{f_s - f_0}{f_0} \|_2} \right) f_0 d\nu; \tag{7}$$

r is continuous for non zero s and t and admits the following limits

$$r(0^+, 0^+) = r(0^-, 0^-) = 1, r(0^+, 0^-) = -1.$$

It is also proved in Gassiat (2002) that if the densities g_n in $\mathcal{M} \setminus \mathcal{M}_0$ are such that $\frac{g_n - g_0}{g_0} / \| \frac{g_n - g_0}{g_0} \|_2$ converges to some $d_{\mathbf{u}_0}$ with $\sqrt{n} \| \frac{g_n - g_0}{g_0} \|_2$ tending to a positive constant c, then the distributions $(g_0 \cdot \nu)^{\otimes n}$ and $(g_n \cdot \nu)^{\otimes n}$ are mutually contiguous, and $2\,\lambda_n$ converges in distribution under this contiguous alternative to

$$\sup_{\mathbf{u}\in\mathbb{U}} \left(\max\left\{Z(\mathbf{u}) + c \cdot r(\mathbf{u}, \mathbf{u}_0), 0\right\}\right)^2 - \sup_{\mathbf{v}\in\mathbb{U}_0} \left(\max\left\{Z(\mathbf{v},\mathbf{0}) + c \cdot r((\mathbf{v},\mathbf{0}), \mathbf{u}_0), 0\right\}\right)^2. \tag{8}$$

In general (5) and (8) reduce to the square of only one supremum, due to the particular structure of the Gaussian process.
We will see, in the subsequent subsections, examples such as: translation mixtures, exponential families, in particular Bernoulli or Gaussian mixtures.

2.1 Contamination mixture.

We consider here the contamination mixture model where parameter $\mathbf{t}$ may be multidimensional: $\mathbf{t} \in \mathbb{T}$, $\mathbb{T}$ being a compact subset of $\mathbb{R}^k$ such that $\mathbf{0}$ belongs to the interior of $\mathbb{T}$. Let $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ denote the Euclidean norm and scalar product in $\mathbb{R}^k$. Again,

$$\mathcal{M}_0 = \{f_{\mathbf{0}}\},\ \mathcal{M} = \{g_{\pi,\mathbf{t}} = (1-\pi) f_{\mathbf{0}} + \pi f_{\mathbf{t}},\ 0 \le \pi \le 1,\ \mathbf{t} \in \mathbb{T}\},$$

$$\mathcal{S} = \left\{ d_{\mathbf{t}} = \frac{\frac{f_{\mathbf{t}} - f_{\mathbf{0}}}{f_{\mathbf{0}}}}{\|\frac{f_{\mathbf{t}} - f_{\mathbf{0}}}{f_{\mathbf{0}}}\|_2},\quad \mathbf{t} \in \mathbb{T} \right\}.$$

We shall use the following Assumptions (CM), insuring smoothness and some non degeneracy:

(CM)

- $f_{\mathbf{t}} = f_{\mathbf{0}}$ ν a.e. if and only if $\mathbf{t} = \mathbf{0}$.
- $\mathbf{t} \to f_{\mathbf{t}}$ is twice continuously differentiable ν a.e. at any $\mathbf{t} \in \mathbb{T}$.
- $\exists \eta > 0, \forall \tau \in \mathbb{R}^k$, $\forall \mathbf{t} \in \mathbb{T}$ such that $\|\mathbf{t}\| \le \eta$, $\sum_{i=1}^k \tau_i \frac{\partial f_{\mathbf{t}}}{\partial t_i} = 0$ ν a.e. if and only if $\tau = \mathbf{0}$.
- There exists a positive real η and a function $B \in L^2(f_{\mathbf{0}} \cdot \nu)$ that upper bounds all following functions:
$$\frac{f_{\mathbf{t}}}{f_{\mathbf{0}}},\ \frac{1}{f_{\mathbf{0}}}\left|\frac{\partial f_{\mathbf{t}}}{\partial t_i}\right|, i = 1, \ldots, k,\ \mathbf{t} \in \mathbb{T},$$
$$\frac{1}{f_{\mathbf{0}}}\left|\frac{\partial^2 f_{\mathbf{t}}}{\partial t_i \partial t_j}\right|, i, j = 1, \ldots, k,\ \mathbf{t} \in \mathbb{T},\ \|\mathbf{t}\| \le \eta.$$

Notice that in this assumption the real number η is fixed. We shall prove that the condition (4) holds true for $\mathcal{S}$ by splitting it into two sets

$$\mathcal{S}_1 = \{d_{\mathbf{t}},\ \mathbf{t} \in \mathbb{T},\ \|\mathbf{t}\| \ge \eta\} \text{ and } \mathcal{S}_2 = \{d_{\mathbf{t}},\ \mathbf{t} \in \mathbb{T},\ \|\mathbf{t}\| < \eta\}.$$

Since $\|(g_{\pi,\mathbf{t}} - f_{\mathbf{0}})/f_{\mathbf{0}}\|_2 = \pi \|(f_{\mathbf{t}} - f_{\mathbf{0}})/f_{\mathbf{0}}\|_2$ tends to 0 as soon as π or $\mathbf{t}$ tends to 0, it is easy to see that a limit point exists only if either $\mathbf{t}$ converges to a limit different of $\mathbf{0}$ or $\frac{\mathbf{t}}{\|\mathbf{t}\|}$ converges to some τ. One obtains easily

$$\mathcal{D} = \left\{ d_{\mathbf{t}} = \frac{f_{\mathbf{t}} - f_{\mathbf{0}}}{f_{\mathbf{0}}} / \|\frac{f_{\mathbf{t}} - f_{\mathbf{0}}}{f_{\mathbf{0}}}\|_2,\ \mathbf{t} \in \mathbb{T} \right\} \cup \left\{ \bar{d}_\tau = \frac{1}{f_{\mathbf{0}}} \sum_{i=1}^k \tau_i \frac{\partial f_{\mathbf{0}}}{\partial t_i} / \|\frac{1}{f_{\mathbf{0}}} \sum_{i=1}^k \tau_i \frac{\partial f_{\mathbf{0}}}{\partial t_i}\|_2,\ \|\tau\| = 1 \right\}.$$

Set $h_{\mathbf{t}} = \frac{f_{\mathbf{t}} - f_{\mathbf{0}}}{f_{\mathbf{0}}}$. Then, for $i = 1, \ldots, k$, if (CM) holds,

$$\frac{\partial d_{\mathbf{t}}}{\partial t_i} = \frac{\frac{\partial h_{\mathbf{t}}}{\partial t_i}}{\|h_{\mathbf{t}}\|_2} - \int \left(\frac{\frac{\partial h_{\mathbf{t}}}{\partial t_i}}{\|h_{\mathbf{t}}\|_2} \right) \left(\frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_{\mathbf{0}} d\nu \cdot \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2}.$$

This proves that, there exists a constant C such that, for all $\mathbf{t}$ and $\mathbf{s}$ such that $\|\mathbf{t}\| \ge \eta$ and $\|\mathbf{s}\| \ge \eta$, $|d_{\mathbf{t}} - d_{\mathbf{s}}| \le C \cdot B \cdot \|\mathbf{t} - \mathbf{s}\|$, so that the number of brackets of length ϵ needed to cover $\mathcal{S}_1$ is of order at most $O(1/\epsilon^k)$ and that Condition (4) holds true for the set $\mathcal{S}_1$.

Now, for any $\tau \in \mathbb{T}$ such that $\|\tau\| = 1$, one has letting $\mathbf{t} = \lambda \tau$, $\lambda \in \mathbb{R}$,

$$\frac{\partial}{\partial \lambda} (d_{\lambda \tau}) = \frac{\sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{t}}}{\partial t_i}}{\|h_{\mathbf{t}}\|_2} - \int \left(\frac{\sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{t}}}{\partial t_i}}{\|h_{\mathbf{t}}\|_2} \right) \left(\frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_{\mathbf{0}} d\nu \cdot \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2}.$$

But using Taylor expansions, there exists $\hat{\lambda}$, λ^* and $\tilde{\lambda}$ in $[0, \lambda]$, such that

$$h_{\mathbf{t}} = \lambda \sum_{i=1}^k \tau_i \frac{\partial h_{\hat{\lambda}\tau}}{\partial t_i} = \lambda \sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} + \frac{\lambda^2}{2} \sum_{i,j=1}^k \tau_i \tau_j \frac{\partial^2 h_{\lambda^* \tau}}{\partial t_i \partial t_j},$$

$$\sum_{i=1}^{k} \tau_i \frac{\partial h_{\mathbf{t}}}{\partial t_i} = \sum_{i=1}^{k} \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} + \lambda \sum_{i,j=1}^{k} \tau_i \tau_j \frac{\partial^2 h_{\tilde{\lambda}\tau}}{\partial t_i \partial t_j}.$$

All this leads to

$$\frac{\partial}{\partial \lambda}\left(d_{\lambda\tau}\right) = \frac{H_{\mathbf{t}}}{\| \sum_{i=1}^{d} \tau_i \frac{\partial h_{\tilde{\lambda}\tau}}{\partial t_i} \|_2} - \left[\int \frac{H_{\mathbf{t}}}{\| \sum_{i=1}^{d} \tau_i \frac{\partial h_{\tilde{\lambda}\tau}}{\partial t_i} \|_2} \left(\frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_{\mathbf{0}} d\nu \right] \cdot \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2}.$$

with

$$H_{\mathbf{t}} = \frac{1}{\lambda} \sum_{i=1}^{k} \tau_i \frac{\partial h_{\mathbf{t}}}{\partial t_i} - \frac{1}{\lambda^2} h_{\mathbf{t}} = \sum_{i,j=1}^{k} \tau_i \tau_j \frac{\partial^2 h_{\tilde{\lambda}\tau}}{\partial t_i \partial t_j} - \frac{1}{2} \sum_{i,j=1}^{k} \tau_i \tau_j \frac{\partial^2 h_{\lambda^*\tau}}{\partial t_i \partial t_j}.$$

But using (CM), this implies that for some constant C, $\forall \lambda \in (0, \eta]$,

$$\left| \frac{\partial}{\partial \lambda}\left(d_{\lambda\tau}\right) \right| \leq C \cdot B,$$

and that

$$\lim_{\lambda \to 0} \frac{\partial}{\partial \lambda}\left(d_{\lambda\tau}\right) = \partial \overline{d}_{\tau}$$
$$= \frac{1}{2} \frac{\sum_{i,j=1}^{k} \tau_i \tau_j \frac{\partial^2 h_{\mathbf{0}}}{\partial t_i \partial t_j}}{\| \sum_{i=1}^{k} \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} \|_2} - \frac{1}{2} \frac{\sum_{i=1}^{k} \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i}}{\| \sum_{i=1}^{k} \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} \|_2} \int \left(\frac{\sum_{i,j=1}^{k} \tau_i \tau_j \frac{\partial^2 h_{\mathbf{0}}}{\partial t_i \partial t_j}}{\| \sum_{i=1}^{k} \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} \|_2} \right) \left(\frac{\sum_{i=1}^{k} \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i}}{\| \sum_{i=1}^{k} \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} \|_2} \right) f_{\mathbf{0}} d\nu,$$

so that $\lambda \to d_{\lambda\tau}$ is continuously differentiable on $[0, \eta]$. Using the fact that

$$\begin{aligned} \left|d_{\lambda\tau} - d_{\lambda'\tau'}\right| &\leq \left|d_{\lambda\tau} - \overline{d}_{\tau}\right| + \left|\overline{d}_{\tau} - \overline{d}_{\tau'}\right| + \left|d_{\lambda'\tau'} - \overline{d}_{\tau'}\right| \\ &\leq C \cdot B \cdot (\lambda + \lambda') + \left|\overline{d}_{\tau} - \overline{d}_{\tau'}\right| \end{aligned}$$

and that, using (CM), there exists a positive constant $\tilde{C}$ such that

$$\inf_{\|\tau\|=1} \| \sum_{i=1}^{k} \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} \|_2 \geq \tilde{C},$$

we obtain that for some constant C^*, and any τ, τ' such that $\|\tau\| = 1, \|\tau'\| = 1$,

$$\left|\overline{d}_{\tau} - \overline{d}_{\tau'}\right| \leq C^* \cdot B \cdot \|\tau - \tau'\|.$$

It is straightforward to see that the number of brackets of length ϵ needed to cover $\mathcal{S}_2$ is of order at most $O(1/\epsilon^k)$. Thus Assumptions (CM) imply Condition (4).

Define now for all non null $\mathbf{s}$ and $\mathbf{t}$ in $\mathbb{T}$,

$$r(\mathbf{s}, \mathbf{t}) = \int d_{\mathbf{s}} d_{\mathbf{t}} f_{\mathbf{0}} d\nu, \tag{9}$$

and let $Z(\cdot)$ be the Gaussian field on $\mathbb{T} \backslash \{\mathbf{0}\}$ with covariance r. Notice that, on each direction τ such that $\|\mathbf{t}\| \to 0$ with $\mathbf{t}/\|\mathbf{t}\| \to \tau$, one may extend $r(\cdot, \cdot)$ by continuity, setting

$$\bar{r}(\tau, \mathbf{t}) = \bar{r}(\mathbf{t}, \tau) = \int \bar{d}_{\tau} d_{\mathbf{t}} f_{\mathbf{0}} d\nu \quad ; \quad \tilde{r}(\tau, \tau') = \int \bar{d}_{\tau} \bar{d}_{\tau'} f_{\mathbf{0}} d\nu. \tag{10}$$

Let π_n and $\mathbf{t}_n$ be sequences such that

- $\lim_{n \to +\infty} \sqrt{n} \pi_n \|(f_{\mathbf{t}_n} - f_{\mathbf{0}})/f_{\mathbf{0}}\|_2 = c$ for some positive c,
- either $\mathbf{t}_n$ tends to some $\mathbf{t}_0 \neq \mathbf{0}$ and $\sqrt{n}\pi_n$ tends to some positive constant, or $\mathbf{t}_n$ tends to $\mathbf{0}$, and $\mathbf{t}_n/\|\mathbf{t}_n\|$ converges to some limit τ.

Then:

Theorem 1 *Assume (CM). Then* $(f_{\mathbf{0}}\cdot\nu)^{\otimes n}$ *and* $[((1-\pi_n)f_{\mathbf{0}}+\pi_n f_{\mathbf{t}_n})\cdot\nu]^{\otimes n}$ *are mutually contiguous,* $2\,\lambda_n$ *converges under* $(f_{\mathbf{0}}\cdot\nu)^{\otimes n}$ *in distribution to*

$$\sup_{\mathbf{t}\in\mathbb{T}}(\max\{Z(\mathbf{t}),0\})^2=\Big(\sup_{\mathbf{t}\in\mathbb{T}}Z(\mathbf{t})\Big)^2,$$

and under $[((1-\pi_n)f_{\mathbf{0}}+\pi_n f_{\mathbf{t}_n})\cdot\nu]^{\otimes n}$ *to*

$$\sup_{\mathbf{t}\in\mathbb{T}}\big(\max\{Z(\mathbf{t})+\mu(\mathbf{t}),0\}\big)^2=\Big(\sup_{\mathbf{t}\in\mathbb{T}}(Z(\mathbf{t})+\mu(\mathbf{t}))\Big)^2$$

with

$$\mu(\mathbf{t})=c\cdot r(\mathbf{t},\mathbf{t}_0)\ \textit{if}\ \mathbf{t}_n\to\mathbf{t}_0\neq 0,\ \textit{and}\ \mu(\mathbf{t})=c\cdot\bar{r}(\mathbf{t},\tau)\ \textit{if}\ \|\mathbf{t}_n\|\to 0\ \textit{and}\ \mathbf{t}_n/\|\mathbf{t}_n\|\to\tau. \quad (11)$$

Remark: Set $m\equiv 0$ under $(f_{\mathbf{0}}\cdot\nu)^{\otimes n}$ and $m\equiv\mu$ under $[((1-\pi_n)f_{\mathbf{0}}+\pi_n f_{\mathbf{t}_n})\cdot\nu]^{\otimes n}$. Letting $\mathbf{t}$ go to $\mathbf{0}$ radially in two opposite directions and using covariance properties in the neighbourhood of $\mathbf{0}$ we see that almost surely $\sup_{\mathbf{t}\in\mathbb{T}}(Z(\mathbf{t})+m(\mathbf{t}))>0$ what justifies equalities in preceding theorem.

Let us give applications of this theorem to particular models:

2.1.1 Translation mixtures

We consider the translation mixture model, where ν is the Lebesgue measure and

$$f_{\mathbf{t}}(\cdot)=f_{\mathbf{0}}(\cdot-\mathbf{t}).$$

Then, it is easy to see that Theorem 1 applies as soon as the following Assumptions (CTM) hold:

(CTM)

- $f_{\mathbf{0}}$ is positive on $\mathbb{R}^k$,
- $x\to f_{\mathbf{0}}(x)$ is twice continuously differentiable ν a.e.
- There exists a function $B\in L^2(f_0\cdot\nu)$ that upper bounds all following functions:

$$\frac{f_{\mathbf{0}}(x-\mathbf{t})}{f_{\mathbf{0}}(x)},\ \frac{1}{f_0(x)}\left|\frac{\partial f_{\mathbf{0}}}{\partial x_i}(x-\mathbf{t})\right|,i=1,\ldots,k,\ \mathbf{t}\in\mathbb{T},$$

$$\frac{1}{f_0(x)}\left|\frac{\partial^2 f_{\mathbf{0}}}{\partial x_i\partial x_j}(x-\mathbf{t})\right|,i,j=1,\ldots,k,\ \mathbf{t}\in\mathbb{T},\ \|\mathbf{t}\|\le\eta.$$

Indeed, since $\frac{\partial f_{\mathbf{t}}}{\partial t_i}(x)=-\frac{\partial f_0}{\partial x_i}(x-\mathbf{t})$, if τ is such that $\sum_{i=1}^k\tau_i\frac{\partial f_{\mathbf{t}}}{\partial t_i}=0$ ν a.e. for all $\|\mathbf{t}\|\le\eta$, then $\sum_{i=1}^k\tau_i\frac{\partial f_0}{\partial x_i}=0$ ν a.e., so that $f_0(x+\lambda\tau)=f_{\mathbf{0}}(x)$ for all $\lambda\in\mathbb{R}$, which is impossible unless $\tau=\mathbf{0}$.

Here are some examples of situations in which these assumptions are met : $f_{\mathbf{0}}$ being the inverse of a polynomial with degree at least 2, among which the Cauchy density, the Gaussian densities and the normalization of $\mathrm{ch}(x)^{-1}$.

The covariance function r is given for non null $\mathbf{s}$ and $\mathbf{t}$ by

$$r(\mathbf{s},\mathbf{t})=\frac{\displaystyle\int\frac{f_{\mathbf{0}}(x-\mathbf{s})f_{\mathbf{0}}(x-\mathbf{t})}{f_{\mathbf{0}}(x)}d\nu(x)-1}{\sqrt{\displaystyle\int\frac{f_{\mathbf{0}}(x-\mathbf{s})^2}{f_{\mathbf{0}}(x)}d\nu(x)-1}\sqrt{\displaystyle\int\frac{f_{\mathbf{0}}(x-\mathbf{t})^2}{f_{\mathbf{0}}(x)}d\nu(x)-1}},$$

and if the dimension $k=1$, one may define $r(0^+,0^-)=-1$, and for non null t

$$r(0^+,t) = -\,r(0^-,t) = \frac{\displaystyle\int \frac{-f_{\mathbf{0}}'(x) f_{\mathbf{0}}(x-t)}{f_{\mathbf{0}}(x)} d\nu(x)}{\sqrt{\displaystyle\int \frac{f_{\mathbf{0}}'^2(x)}{f_{\mathbf{0}}(x)} d\nu(x)}\sqrt{\displaystyle\int \frac{f_{\mathbf{0}}(x-t)^2}{f_{\mathbf{0}}(x)} d\nu(x) - 1}},$$

where the derivation is with respect to x.

2.1.2 Gaussian mixtures

Without loss of generality we may assume that $f_{\mathbf{0}}$ is standard normal. Let K be a bound for $\|\mathbf{t}\|$, $\mathbf{t} \in \mathbb{T}$. Then the following bounds show that the function B exists for any η

$$\begin{aligned}
\frac{f_{\mathbf{0}}(x-\mathbf{t})}{f_{\mathbf{0}}(x)} = \exp\big(\langle x, \mathbf{t}\rangle - \|\mathbf{t}\|^2/2\big) &\leq \exp\big(K\|x\|\big), \\
\frac{1}{f_{\mathbf{0}}(x)} \left|\frac{\partial f_{\mathbf{0}}}{\partial x_i}(x-\mathbf{t})\right| = |x_i - t_i| \frac{f_{\mathbf{0}}(x-\mathbf{t})}{f_{\mathbf{0}}(x)} &\leq (\|x\|+K)\exp\big(K\|x\|\big), \\
\frac{1}{f_{\mathbf{0}}(x)} \left|\frac{\partial^2 f_{\mathbf{0}}}{\partial x_i \partial x_j}(x-\mathbf{t})\right| = |x_i - t_i||x_j - t_j| \frac{f_{\mathbf{0}}(x-\mathbf{t})}{f_{\mathbf{0}}(x)} &\leq (\|x\|+K)^2\exp\big(K\|x\|\big), \quad i \neq j, \\
\frac{1}{f_{\mathbf{0}}(x)} \left|\frac{\partial^2 f_{\mathbf{0}}}{\partial x_i^2}(x-\mathbf{t})\right| = |(x_i - t_i)^2 - 1| \frac{f_{\mathbf{0}}(x-\mathbf{t})}{f_{\mathbf{0}}(x)} &\leq [1+(\|x\|+K)^2]\exp\big(K\|x\|\big).
\end{aligned}$$

So (CTM) holds, and Theorem 1 applies, as soon as f_0 is some Gaussian density on $\mathbb{R}^k$ and $\mathbb{T}$ is compact. The covariance of the process Z is:

$$r(\mathbf{s},\mathbf{t}) = \frac{\exp(\langle \mathbf{t}, \mathbf{s}\rangle) - 1}{\sqrt{\exp(\|\mathbf{t}\|^2) - 1}\sqrt{\exp(\|\mathbf{s}\|^2) - 1}}.$$

2.1.3 Binomial mixtures

Here ν is the measure with density $\frac{k!}{x!(k-x)!}$ with respect to the counting measure on the set $\{0,1,\ldots,k\}$. We consider the binomial family $Bi(k,\theta)$ with density $\theta^x(1-\theta)^{k-x}$; $x = 0,1,...,k$. Let $\theta_0 \in (0,1)$ and f_t be the density of $Bi(k,\theta_0+t)$. The most relevant case for genetic applications is the case $\theta_0 = 1/2$, see Problem 1 in Chernoff and Lander (1995). We have

$$\begin{aligned}
f_t(x) &= (t+\theta_0)^x(1-t-\theta_0)^{k-x}, \\
\frac{\partial f_t}{\partial t}(x) &= \left(\frac{x}{t+\theta_0} - \frac{k-x}{1-t-\theta_0}\right) f_t(x), \\
\frac{\partial^2 f_t}{\partial t^2}(x) &= \left[\left(\frac{x}{t+\theta_0} - \frac{k-x}{1-t-\theta_0}\right)^2 - \frac{x}{(t+\theta_0)^2} + \frac{k-x}{(1-t-\theta_0)^2}\right] f_t(x).
\end{aligned}$$

It is clear that $f_t(x)$ and $\frac{\partial f_t}{\partial t}(x)$ are uniformly upper bounded and that $\frac{\partial^2 f_t}{\partial t^2}(x)$ is upper bounded for t small enough, proving Assumptions (CM). Direct calculations lead to

$$\frac{f_t - f_0}{f_0}(x) = (1+\frac{t}{\theta_0})^x(1-\frac{t}{1-\theta_0})^{k-x} - 1,$$

$$r(s,t) = \frac{\Gamma(s,t)}{\sqrt{\Gamma(s,s)}\sqrt{\Gamma(t,t)}},$$

with

$$\Gamma(s,t) = \sum_{x=0}^{k} \big[(1+\frac{s}{\theta_0})^x(1-\frac{s}{1-\theta_0})^{k-x} - 1\big]\big[(1+\frac{t}{\theta_0})^x(1-\frac{t}{1-\theta_0})^{k-x} - 1\big]\theta_0^x(1-\theta_0)^{k-x}$$

which is equivalent to the result of Chernoff and Lander (1995).

2.1.4 Mixtures in exponential families

This case generalizes the preceding. Let $f_{\mathbf{t}}$ be a regular exponential family with exhaustive statistic $T(x) = (T_1(x), \ldots, T_k(x))$:

$$f_{\mathbf{t}}(x) = f_{\mathbf{0}}(x) \exp\left(\sum_{i=1}^{k} t_i T_i(x) - \phi(\mathbf{t})\right),$$

and assume $\mathbb{T}$ is a compact subset in the interior of the definition set of the exponential family. Then $\mathbf{t} \to f_{\mathbf{t}}$ is infinitely differentiable on $\mathbb{T}$. Let $F(x) = \sup_{\mathbf{t}\in\mathbb{T}} \exp\left(\sum_{i=1}^{k} t_i T_i(x)\right)$, Assumption (CEM) will be:

(CEM)

- There exists B in $L^2(f_{\mathbf{0}} \cdot \nu)$ that upper bounds all following functions: F, $|T_i|F$, $|T_i T_j|F$, $i, j = 1, \ldots, k$.

One can see easily that (CEM) implies (CM), so that Theorem 1 applies to exponential families as soon as (CEM) holds. Direct calculations again lead to

$$\frac{f_{\mathbf{t}} - f_{\mathbf{0}}}{f_{\mathbf{0}}}(x) = \exp\big(\sum_{i=1}^{k} t_i T_i(x) - \phi(\mathbf{t})\big) - 1,$$

$$r(\mathbf{s}, \mathbf{t}) = \frac{\exp\big(\phi(\mathbf{s}+\mathbf{t}) - \phi(\mathbf{s}) - \phi(\mathbf{t})\big) - 1}{\sqrt{\exp\big(\phi(2\mathbf{s}) - 2\phi(\mathbf{s})\big) - 1}\sqrt{\exp\big(\phi(2\mathbf{t}) - 2\phi(\mathbf{t})\big) - 1}}.$$

2.2 Two populations against a single one

We consider here the case where one wants to test a single population in the family of densities $f_{\mathbf{t}}, \mathbf{t} \in \mathbb{T}$, $\mathbb{T}$ compact subset of $\mathbb{R}^k$ against a mixture of two such populations. That is:

$$\mathcal{M}_0 = \{f_{\mathbf{t}},\ \mathbf{t} \in \mathbb{T}\},$$

and

$$\mathcal{M} = \{g_{\pi,\mathbf{t}_1,\mathbf{t}_2} = (1-\pi) f_{\mathbf{t}_1} + \pi f_{\mathbf{t}_2},\ 0 \le \pi \le 1,\ \mathbf{t}_1 \in \mathbb{T},\ \mathbf{t}_2 \in \mathbb{T}\}.$$

We suppose moreover that $\mathbf{0}$ is an interior point of $\mathbb{T}$ and that $f_{\mathbf{0}}$ is the unknown distribution of the observations (with no loss of generality). We shall use Assumptions (TP), insuring smoothness and some non degeneracy:

(TP)

- $(1-\pi) f_{\mathbf{t}_1} + \pi f_{\mathbf{t}_2} = f_{\mathbf{0}}$ ν a.e. if and only if ($\pi = 0$ and $\mathbf{t}_1 = \mathbf{0}$) or ($\pi = 1$ and $\mathbf{t}_2 = \mathbf{0}$) or ($\mathbf{t}_1 = \mathbf{0}$ and $\mathbf{t}_2 = \mathbf{0}$),
- $\mathbf{t} \to f_{\mathbf{t}}$ is three times continuously differentiable ν a.e. at any $\mathbf{t} \in \mathbb{T}$,
- $\forall \tau \in \mathbb{R}^k$, $\forall \mathbf{t} \in \mathbb{T}$, $\forall \mathbf{s} \in \mathbb{T}$, $\forall \rho \ge 0$, $\rho(f_{\mathbf{s}} - f_{\mathbf{0}}) + \sum_{i=1}^{k} \tau_i \frac{\partial f_{\mathbf{t}}}{\partial t_i} = 0$ ν a.e. if and only if $\rho \mathbf{s} = \mathbf{0}$ and $\tau = \mathbf{0}$,
  and $\exists \eta > 0$, such that $\forall \tau \in \mathbb{R}^k$, $\forall \mathbf{t} \in \mathbb{T}$ with $\|\mathbf{t}\| \le \eta$ $\sum_{i,j=1}^{k} \tau_i \tau_j \frac{\partial^2 f_{\mathbf{t}}}{\partial t_i \partial t_j} = 0$ ν a.e. if and only if $\tau = \mathbf{0}$,
- there exists a function $B \in L^2(f_{\mathbf{0}} \cdot \nu)$ that upper bounds all following functions:

$$\frac{f_{\mathbf{t}}}{f_{\mathbf{0}}},\ \frac{1}{f_{\mathbf{0}}}\left|\frac{\partial f_{\mathbf{t}}}{\partial t_i}\right|,\ \frac{1}{f_{\mathbf{0}}}\left|\frac{\partial^2 f_{\mathbf{t}}}{\partial t_i \partial t_j}\right|, i, j = 1, \ldots, k,\ \mathbf{t} \in \mathbb{T},$$

$$\frac{1}{f_{\mathbf{0}}}\left|\frac{\partial^3 f_{\mathbf{t}}}{\partial t_i \partial t_j \partial t_l}\right|, i, j, l = 1, \ldots, k,\ \mathbf{t} \in \mathbb{T},\ \|\mathbf{t}\| \le \eta.$$

Then $\mathcal{S} \subset \mathcal{D}$, $\mathcal{S}_0 \subset \mathcal{D}_0$, and $\mathcal{D}$ can be parameterized as follows:

$$\mathcal{D} = \left\{ d_{\mathbf{t},a,\tau} = \frac{a\frac{f_{\mathbf{t}}-f_{\mathbf{0}}}{f_{\mathbf{0}}} + \sum_{i=1}^{k} \tau_i \frac{1}{f_{\mathbf{0}}}\frac{\partial f_{\mathbf{0}}}{\partial t_i}}{\|a\frac{f_{\mathbf{t}}-f_{\mathbf{0}}}{f_{\mathbf{0}}} + \sum_{i=1}^{k} \tau_i \frac{1}{f_{\mathbf{0}}}\frac{\partial f_{\mathbf{0}}}{\partial t_i}\|_2},\ \mathbf{t} \in \mathbb{T} \setminus \{\mathbf{0}\},\ \tau \in \mathbb{R}^k,\ a \geq 0,\ a + \|\tau\| = 1 \right\},$$

$$\mathcal{D}_0 = \{d_{\mathbf{0},0,\tau},\ \|\tau\| = 1\}.$$

Let $r(\cdot,\cdot)$ be as in Section 2.1:

$$r(\mathbf{s},\mathbf{t}) = \int \left(\frac{h_{\mathbf{s}}}{\|h_{\mathbf{s}}\|_2}\right)\left(\frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2}\right) f_{\mathbf{0}} d\nu$$

with $h_{\mathbf{t}} = (f_{\mathbf{t}} - f_{\mathbf{0}})/f_{\mathbf{0}}$, and $Z(\cdot)$ the associated Gaussian field.
Let W be the k-dimensional centered Gaussian variable with variance Σ with entries :

$$\Sigma_{i,j} = \int \left(\frac{\frac{1}{f_{\mathbf{0}}}\frac{\partial f_{\mathbf{0}}}{\partial t_i}}{\|\frac{1}{f_{\mathbf{0}}}\frac{\partial f_{\mathbf{0}}}{\partial t_i}\|_2}\right)\left(\frac{\frac{1}{f_{\mathbf{0}}}\frac{\partial f_{\mathbf{0}}}{\partial t_j}}{\|\frac{1}{f_{\mathbf{0}}}\frac{\partial f_{\mathbf{0}}}{\partial t_j}\|_2}\right) f_{\mathbf{0}} d\nu,\ i,j = 1,\dots,k,$$

and for any $\mathbf{t}$, let $C(\mathbf{t})$ be the k-dimensional vector of covariances of $Z(\mathbf{t})$ and W:

$$C(\mathbf{t})_i = \int \left(\frac{\frac{1}{f_{\mathbf{0}}}\frac{\partial f_{\mathbf{0}}}{\partial t_i}}{\|\frac{1}{f_{\mathbf{0}}}\frac{\partial f_{\mathbf{0}}}{\partial t_i}\|_2}\right)\left(\frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2}\right) f_{\mathbf{0}} d\nu,\ i = 1,\dots,k.$$

Then $\mathcal{D}$ can be reparametrized as follows

$$\mathcal{D} = \left\{d_{\mathbf{t},a,\tau};\ \mathbf{t} \in \mathbb{T}\setminus\{\mathbf{0}\},\ \tau \in \mathbb{R}^k,\ a \geq 0,\ a^2 + \tau^T\Sigma\tau + 2a\tau^T C(\mathbf{t}) = 1\right\}.$$

Using the same tricks as for proving Theorem 1, $2\,\lambda_n$ converges under $(f_{\mathbf{0}} \cdot \nu)^{\otimes n}$ in distribution to

$$\left(\sup_{\substack{a \geq 0, t \in \mathbb{T}, \tau \in \mathbb{R}^k \\ a^2 + \tau^T\Sigma\tau + 2a\tau^T C(\mathbf{t}) = 1}} (aZ(\mathbf{t}) + \langle\tau, W\rangle)\right)^2 - \left(\sup_{\tau^T\Sigma\tau=1}\langle\tau, W\rangle\right)^2. \tag{12}$$

Remark that:

$$\left(\sup_{\tau^T\Sigma\tau=1}\langle\tau, W\rangle\right)^2 = W^T\Sigma^{-1}W, \tag{13}$$

and that the supremum is attained for τ colinear to $\Sigma^{-1}W$. Then consider the matrix:

$$\tilde{\Sigma} = \begin{bmatrix} 1 & C(\mathbf{t})^T \\ C(\mathbf{t}) & \Sigma \end{bmatrix}$$

with inverse

$$(\tilde{\Sigma})^{-1} = \begin{bmatrix} \alpha & u^T \\ u & M \end{bmatrix}$$

where $M = M(\mathbf{t}) = \left(\Sigma - C(\mathbf{t})C(\mathbf{t})^T\right)^{-1}$, $u = u(\mathbf{t}) = -M(\mathbf{t})C(\mathbf{t})$, $\alpha = \alpha(\mathbf{t}) = 1 + C(\mathbf{t})^T M(\mathbf{t})C(\mathbf{t})$. Now consider the maximization problem in a and τ:

$$\left(\sup_{a\geq 0, a^2+\tau^T\Sigma\tau+2a\tau^T C(\mathbf{t})=1} (aZ(\mathbf{t}) + \langle\tau, W\rangle)\right)^2. \tag{14}$$

If the maximum is attained for $a > 0$, then by (13) its value is

$$\begin{pmatrix} Z(\mathbf{t}) \\ W \end{pmatrix}^T (\tilde{\Sigma})^{-1} \begin{pmatrix} Z(\mathbf{t}) \\ W \end{pmatrix},$$

which is equal to

$$\alpha\big(Z(\mathbf{t})+\frac{u^TW}{\alpha}\big)^2+W^T\big(M-\frac{uu^T}{\alpha}\big)W=\alpha\big(Z(\mathbf{t})+\frac{u^TW}{\alpha}\big)^2+W^T\Sigma^{-1}W.$$

This is the case when the first coordinate of

$$(\tilde{\Sigma})^{-1}\begin{pmatrix} Z(\mathbf{t}) \\ W \end{pmatrix}$$

is non-negative that is $\alpha Z(\mathbf{t})+\langle u,W\rangle\geq 0$. In the other case ($a=0$) the maximum is $W^T\Sigma^{-1}W$ by (13). Finally we have proved that the supremum in (14) is equal to

$$\left\{\max\left\{Z(\mathbf{t})-\frac{C(\mathbf{t})^TM(\mathbf{t})W}{1+C(\mathbf{t})^TM(\mathbf{t})C(\mathbf{t})},0\right\}\right\}^2\big(1+C(\mathbf{t})^TM(\mathbf{t})C(\mathbf{t})\big)+W^T\Sigma^{-1}W. \tag{15}$$

This implies that the limit of $2\,\lambda_n$ under $(f_{\mathbf{0}}\cdot\nu)^{\otimes n}$ is equal in distribution to

$$\left(\sup_{\mathbf{t}\in\mathbb{T}}\left(Z(\mathbf{t})-\frac{C(\mathbf{t})^TM(\mathbf{t})W}{1+C(\mathbf{t})^TM(\mathbf{t})C(\mathbf{t})}\right)\right)^2\big(1+C(\mathbf{t})^TM(\mathbf{t})C(\mathbf{t})\big)\,.$$

Indeed one may see, letting $\mathbf{t}$ go to $\mathbf{0}$ radially in two opposite directions, that the supremum of the Gaussian process involved in formula (15) is non negative. Let now π_n, $\mathbf{t}_1^n$ and $\mathbf{t}_2^n$ be sequences such that $\frac{(1-\pi_n)f_{\mathbf{t}_1^n}+\pi_nf_{\mathbf{t}_2^n}-f_{\mathbf{0}}}{f_{\mathbf{0}}}/\|\frac{(1-\pi_n)f_{\mathbf{t}_1^n}+\pi_nf_{\mathbf{t}_2^n}-f_{\mathbf{0}}}{f_{\mathbf{0}}}\|_2$ tends to some $d_{\mathbf{t}_0,a_0,\tau_0}$ in $\mathcal{D}$, with $\lim_{n\to+\infty}\sqrt{n}\|\frac{(1-\pi_n)f_{\mathbf{t}_1^n}+\pi_nf_{\mathbf{t}_2^n}-f_{\mathbf{0}}}{f_{\mathbf{0}}}\|_2=c$ for some positive constant c. Then, using the same tricks again:

Theorem 2 *Assume (TP). Then $(f_{\mathbf{0}}\cdot\nu)^{\otimes n}$ and $[((1-\pi_n)f_{\mathbf{t}_1^n}+\pi_nf_{\mathbf{t}_2^n})\cdot\nu]^{\otimes n}$ are mutually contiguous, $2\,\lambda_n$ converges under $(f_{\mathbf{0}}\cdot\nu)^{\otimes n}$ in distribution to*

$$\left(\sup_{\mathbf{t}\in\mathbb{T}}\left(Z(\mathbf{t})-\frac{C(\mathbf{t})^TM(\mathbf{t})W}{1+C(\mathbf{t})^TM(\mathbf{t})C(\mathbf{t})}\right)\right)^2\big(1+C(\mathbf{t})^TM(\mathbf{t})C(\mathbf{t})\big)\,,$$

and under $[((1-\pi_n)f_{\mathbf{t}_1^n}+\pi_nf_{\mathbf{t}_2^n})\cdot\nu]^{\otimes n}$ to

$$\left[\sup_{\mathbf{t}\in\mathbb{T}}\left(aZ(\mathbf{t})+a_0cr(\mathbf{t},\mathbf{t}_0)+cC(\mathbf{t})^T\tau_0-\frac{C(\mathbf{t})^TM(\mathbf{t})(W+c\Sigma\tau_0+ca_0C(t_o))}{1+C(\mathbf{t})^TM(\mathbf{t})C(\mathbf{t})}\right)\right]^2\big(1+C(\mathbf{t})^TM(\mathbf{t})C(\mathbf{t})\big)\,,$$

where if $\mathbf{t}_0=0$ then $a_0=0$.

Notice that, when $\mathbf{t}_0=0$, $d_{0,a_0,\tau_0}=d_{0,0,\tau_0}$, and $\langle d_{0,0,\tau_0},d_{\mathbf{t},a,\tau,},\rangle=cC(\mathbf{t})^T\tau_0+c\Sigma\tau_0$. This is why one has to take $a_0=0$ when $\mathbf{t}_0=0$ in the last formula of Theorem 2.

2.2.1 Examples.

Results of Section 2.2 apply to the same previous examples.

- **Translation mixtures**, under (CTM) with moreover $x\to f_{\mathbf{0}}(x)$ is three times continuously differentiable ν a.e., and the function $B\in L^2(f_{\mathbf{0}}\cdot\nu)$ is also an upper bound for
$$\frac{1}{f_{\mathbf{0}}(x)}\left|\frac{\partial^3f_{\mathbf{0}}}{\partial x_i\partial x_j\partial x_l}(x-\mathbf{t})\right|,i,j,l=1,\ldots,k,\ \mathbf{t}\in\mathbb{T},\ \|\mathbf{t}\|\leq\eta.$$
- **Gaussian mixtures**, in this case W is a standard normal vector and for all $\mathbf{t}\in\mathbb{T}$ $C(\mathbf{t})=\frac{\mathbf{t}}{\sqrt{e^{\|\mathbf{t}\|^2}-1}}$.
- **Bernoulli mixtures**,
- **Mixtures in exponential families**, under (CEM) with moreover: the function $B\in L^2(f_{\mathbf{0}}\cdot\nu)$ is also an upper bound for $|T_iT_jT_l|F$, $i,j,l=1,\ldots,k$. In this

case, W is the Gaussian vector with covariance Σ the correlation matrix of the vector $(T_1(X),\dots,T_k(X))$, when X has density $f_{\mathbf{0}}$. Recall that the variance matrix of the vector $(T_1(X),\dots,T_k(X))$ when X has density $f_{\mathbf{0}}$ is the matrix $D^2\phi$ of second derivatives of the function ϕ at point $\mathbf{0}$, and the vector $C(\mathbf{t})$ is given by

$$C(\mathbf{t})_i = \frac{\frac{\partial\phi}{\partial t_i}(\mathbf{t}) - \frac{\partial\phi}{\partial t_i}(\mathbf{0})}{\sqrt{\exp(\phi(2\mathbf{t}) - 2\phi(\mathbf{t})) - 1}\sqrt{(D^2\phi(\mathbf{0}))_{i,i}}},\ i = 1,\dots,k.$$

2.3 Contamination with unknown nuisance parameter

We consider here the contamination mixture model with some unknown parameter, which is the same for all populations. A typical example may be that of mixtures of Gaussian distributions with the same unknown variance, or translation mixtures with the same unknown scale parameter. We shall assume that the nuisance parameter is identifiable, so that its maximum likelihood estimator is consistent. This will allow to reduce the possible nuisance parameters in the definition of the set $\mathcal{S}$ to be in a neighbourhood of the true unknown one (recall that $\mathcal{S}$ is only a theoretical tool to verify that some theorem apply, and compute the set of normalized scores, so that this does not restrict the model $\mathcal{M}$, for which the nuisance parameter is not restricted to be in a neighbourhood of the true one).
Let $\mathcal{F} = \{f_{\mathbf{t},\alpha}, \mathbf{t}\in\mathbb{T}, \alpha\in\mathbb{A}\}$ be a set of densities with respect to some dominating measure ν, where $\mathbb{T}$ is a compact subset of $\mathbb{R}^k$ and $\mathbb{A}$ is a compact subset of $\mathbb{R}^h$. We consider here the case where

$$\mathcal{M}_0 = \{f_{\mathbf{0},\alpha},\ \alpha\in\mathbb{A}\},$$

and

$$\mathcal{M} = \{g_{\pi,\mathbf{t},\alpha} = (1-\pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha},\ 0\le\pi\le 1,\ \mathbf{t}\in\mathbb{T},\ \alpha\in\mathbb{A}\}.$$

The unknown true distribution of the observations will be $f_{\mathbf{0},\alpha_0}$. We suppose that $(\mathbf{0},\alpha_0)$ is an interior point of $\mathbb{T}\times\mathbb{A}$. We shall use Assumptions (CMN), insuring smoothness and some non degeneracy:

(CMN)

- $(1-\pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha} = f_{\mathbf{0},\alpha_0}$ ν a.e. if and only if $\alpha=\alpha_0$ and $[\pi = 0$ or $\mathbf{t} = \mathbf{0}]$,
- $(\mathbf{t},\alpha)\to f_{\mathbf{t},\alpha}$ is twice continuously differentiable ν a.e. at any $(\mathbf{t},\alpha)\in\mathbb{T}\times\mathbb{A}$,
- $\exists\eta>0$, such that $\forall\delta\in\mathbb{R}^h$, $\forall\mathbf{t}\in\mathbb{T}$, $\forall\alpha\in\mathbb{A}$ with $\|\alpha-\alpha_0\|\le\eta$, $\forall\rho\ge 0$: $\rho(f_{\mathbf{t},\alpha_0} - f_{\mathbf{0},\alpha_0}) + \sum_{i=1}^h \delta_i\frac{\partial f_{\mathbf{0},\alpha}}{\partial\alpha_i} = 0$ ν a.e. if and only if $\rho\mathbf{t} = \mathbf{0}$ and $\delta = \mathbf{0}$, and $\forall\tau\in\mathbb{R}^k$, $\|\mathbf{t}\|\le\eta$, $\|\alpha-\alpha_0\|\le\eta$: $\sum_{i=1}^k \tau_i\frac{\partial f_{\mathbf{t},\alpha_0}}{\partial t_i} + \sum_{i=1}^h \delta_i\frac{\partial f_{\mathbf{0},\alpha}}{\partial\alpha_i} = 0$ ν a.e. if and only if $\tau = \mathbf{0}$ and $\delta = \mathbf{0}$.
- There exists a function $B\in L^2(f_{\mathbf{0},\alpha_0}\cdot\nu)$ that upper bounds all following functions:

$$\frac{f_{\mathbf{t},\alpha}}{f_{\mathbf{0},\alpha_0}},\ \frac{1}{f_{\mathbf{0},\alpha_0}}\left|\frac{\partial f_{\mathbf{t},\alpha}}{\partial t_i}\right|, i=1,\dots,k,\ \frac{1}{f_{\mathbf{0},\alpha_0}}\left|\frac{\partial f_{\mathbf{t},\alpha}}{\partial\alpha_i}\right|, i=1,\dots,h, (\mathbf{t},\alpha)\in\mathbb{T}\times\mathbb{A},\ \|\alpha-\alpha_0\|\le\eta,$$

$$\frac{1}{f_{\mathbf{0},\alpha_0}}\left|\frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial t_i\partial t_j}\right|, i,j=1,\dots,k,\ \frac{1}{f_{\mathbf{0},\alpha_0}}\left|\frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial t_i\partial\alpha_j}\right|, i=1,\dots,k,\ j=1,\dots,h,$$

$$\frac{1}{f_{\mathbf{0},\alpha_0}}\left|\frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial\alpha_i\partial\alpha_j}\right|,\ i,j=1,\dots,h,\ (\mathbf{t},\alpha)\in\mathbb{T}\times\mathbb{A},\ \|\alpha-\alpha_0\|\le\eta,\ \|\mathbf{t}\|\le\eta.$$

Then, since the maximum likelihood estimator of parameter α is consistent, one only needs to verify Assumption (4) for

$$\mathcal{S} = \left\{\frac{(1-\pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}}/\|\frac{(1-\pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}}\|_2,\ 0\le\pi\le 1, \mathbf{t}\in\mathbb{T}, \alpha\in\mathbb{A}, \|\alpha-\alpha_0\|\le\eta\right\},$$

where we restrict our definition to $\pi,\mathbf{t}$ and α such that $(1-\pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha}$ differs from $f_{\mathbf{0},\alpha_0}$.
One has also

$$\mathcal{S}_0 = \left\{\frac{f_{\mathbf{0},\alpha} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}}/\|\frac{f_{\mathbf{0},\alpha} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}}\|_2,\ 0\le\pi\le 1, \alpha\in\mathbb{A}, \|\alpha-\alpha_0\|\le\eta\right\}.$$

Define, for $(\mathbf{t},\rho,\delta,\tau) \in \mathbb{T} \times \mathbb{R}_+ \times \mathbb{R}^h \times \mathbb{R}^k$,

$$H_{\mathbf{t},\rho,\delta,\tau} = \rho(f_{\mathbf{t},\alpha_0} - f_{\mathbf{0},\alpha_0}) + \sum_{i=1}^{h} \delta_i \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i} + \sum_{i=1}^{k} \tau_i \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial t_i},$$

and

$$d_{\mathbf{t},\rho,\delta,\tau} = \frac{H_{\mathbf{t},\rho,\delta,\tau}/f_{\mathbf{0},\alpha_0}}{\|H_{\mathbf{t},\rho,\delta,\tau}/f_{\mathbf{0},\alpha_0}\|_2}.$$

The sets $\mathcal{D}$ and $\mathcal{D}_0$ can be parameterized as follows:

$$\mathcal{D} = \left\{ d_{\mathbf{t},\rho,\delta,\tau}, \mathbf{t} \in \mathbb{T},\ \rho \geq 0, \delta \in \mathbb{R}^h,\ \tau \in \mathbb{R}^k, \rho^2 + \|\delta\|^2 + \|\tau\|^2 = 1 \right\},$$

$$\mathcal{D}_0 = \left\{ d_{\mathbf{0},0,\delta,\mathbf{0}}, \delta \in \mathbb{R}^h,\ \|\delta\| = 1 \right\}.$$

Note that due to the existence of the nuisance parameter which is fixed to α_0, now $\mathcal{D}$ does not contain $\mathcal{S}$.

It will be possible to obtain the asymptotic distributions in the same way as in Section 2.2. Let again

$$r(\mathbf{s},\mathbf{t}) = \int \left(\frac{h_{\mathbf{s}}}{\|h_{\mathbf{s}}\|_2} \right) \left(\frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_{\mathbf{0},\alpha_0} d\nu$$

with $h_{\mathbf{t}} = (f_{\mathbf{t},\alpha_0} - f_{\mathbf{0},\alpha_0})/f_{\mathbf{0},\alpha_0}$, and $Z(\cdot)$ the associated Gaussian field.
Note that this process is the same as the one of Section 2.1 if we set $f_{\mathbf{0}} = f_{\mathbf{0},\alpha_0}$. Let also W, Σ and $C(\mathbf{t})$ be the same as in Section 2.1 replacing $\frac{\partial f_{\mathbf{0}}}{\partial t_i}$ by $\frac{\partial f_{\mathbf{0},\alpha_0}}{\partial t_i}$.

Let V be the h-dimensional centered Gaussian variable with variance Γ:

$$\Gamma_{i,j} = \int \left(\frac{\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i}}{\|\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i}\|_2} \right) \left(\frac{\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_j}}{\|\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_j}\|_2} \right) f_{\mathbf{0},\alpha_0} d\nu,\ i,j = 1,\ldots,h,$$

and for any $\mathbf{t}$, let $G(\mathbf{t})$ be the h-dimensional vector of covariances of $Z(\mathbf{t})$ and V:

$$G(\mathbf{t})_i = \int \left(\frac{\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i}}{\|\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i}\|_2} \right) \left(\frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_{\mathbf{0},\alpha_0} d\nu,\ i = 1,\ldots,h.$$

Let also S be the convariance matrix of W and V, with entries:

$$S_{i,j} = \int \left(\frac{\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i}}{\|\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i}\|_2} \right) \left(\frac{\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial t_j}}{\|\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial t_j}\|_2} \right) f_{\mathbf{0},\alpha_0} d\nu,\ i = 1,\ldots,h,\ j = 1,\ldots,k.$$

Define the matrices $U(\mathbf{t})$ and $N(\mathbf{t})$ by

$$U(\mathbf{t}) = \begin{pmatrix} C(\mathbf{t})^T \\ G(\mathbf{t}) \end{pmatrix},$$

$$N(\mathbf{t}) = \left(\begin{pmatrix} \Sigma & S^T \\ S & \Gamma \end{pmatrix} - U(\mathbf{t})U(\mathbf{t})^T \right)^{-1}.$$

Let π_n, $\mathbf{t}_n$ and α_n be sequences such that $\frac{(1-\pi_n)f_{\mathbf{0},\alpha_n} + \pi_n f_{\mathbf{t}_n,\alpha_n} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}} / \| \frac{(1-\pi_n)f_{\mathbf{0},\alpha_n} + \pi_n f_{\mathbf{t}_n,\alpha_n} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}} \|_2$ tends to some $d_{\mathbf{t}_0,\rho_0,\delta_0,\tau_0}$ in $\mathcal{D}$, with $\lim_{n\to+\infty} \sqrt{n} \| \frac{(1-\pi_n)f_{\mathbf{0},\alpha_n} + \pi_n f_{\mathbf{t}_n,\alpha_n} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}} \|_2 = c$ for some positive constant c. Then, using the same tricks as for proving Theorem 2 :

Theorem 3 *Assume (CMN). Then* $(f_{\mathbf{0},\alpha_0} \cdot \nu)^{\otimes n}$ *and* $[((1-\pi_n)f_{\mathbf{0},\alpha_n} + \pi_n f_{\mathbf{t}_n,\alpha_n}) \cdot \nu]^{\otimes n}$ *are* *mutually contiguous,* $2\,\lambda_n$ *converges under* $(f_{\mathbf{0}}\cdot\nu)^{\otimes n}$ *in distribution to*

$$\left[\sup_{\mathbf{t}\in\mathbb{T}}\left(Z(\mathbf{t})-\frac{U(\mathbf{t})^T N(\mathbf{t})}{1+U(\mathbf{t})^T N(\mathbf{t})U(\mathbf{t})}\begin{pmatrix}W\\V\end{pmatrix}\right)\right]^2\left(1+U(\mathbf{t})^T N(\mathbf{t})U(\mathbf{t})\right)$$
$$+\begin{pmatrix}W\\V\end{pmatrix}^T\begin{pmatrix}\Sigma & S^T\\ S & \Gamma\end{pmatrix}^{-1}\begin{pmatrix}W\\V\end{pmatrix}-V^T\Gamma^{-1}V,$$

and under $[((1-\pi_n)f_{\mathbf{0},\alpha_n}+\pi_n f_{\mathbf{t}_n,\alpha_n})\cdot\nu]^{\otimes n}$ *to*

$$\left[\sup_{\mathbf{t}\in\mathbb{T}}\left(Z(\mathbf{t})+c\rho_0 r(\mathbf{t},\mathbf{t}_0)+cC(\mathbf{t})^T\tau_0+cG(\mathbf{t})^T\delta_0-\frac{U(\mathbf{t})^T N(\mathbf{t})}{1+U(\mathbf{t})^T N(\mathbf{t})U(\mathbf{t})}\begin{pmatrix}W+c\Sigma\tau_0+c\rho_0 C(\mathbf{t}_0)\\V+c\Gamma\delta_0+c\rho_0 G(\mathbf{t}_0)\end{pmatrix}\right)\right]^2$$
$$.\left(1+U(\mathbf{t})^T N(\mathbf{t})U(\mathbf{t})\right)+\begin{pmatrix}W+c\Sigma\tau_0+c\rho_0 C(\mathbf{t}_0)\\V+c\Gamma\delta_0+c\rho_0 G(\mathbf{t}_0)\end{pmatrix}^T\begin{pmatrix}\Sigma & S^T\\ S & \Gamma\end{pmatrix}^{-1}\begin{pmatrix}W+c\Sigma\tau_0+c\rho_0 C(\mathbf{t}_0)\\V+c\Gamma\delta_0+c\rho_0 G(\mathbf{t}_0)\end{pmatrix}$$
$$-(V+c\Gamma\delta_0+c\rho_0 G(\mathbf{t}_0))^T\Gamma^{-1}(V+c\Gamma\delta_0+c\rho_0 G(\mathbf{t}_0)),$$

where $\rho_0=0$ *when* $\mathbf{t}_0=0$.

2.3.1 Translation mixtures with unknown scale parameter

Here $h=1$, ν is the Lebesgue measure and

$$f_{\mathbf{t},\alpha}(\cdot)=\alpha f_{\mathbf{0},1}(\alpha(\cdot-\mathbf{t})),$$

with $\mathbb{A}=[a,A]$ for some $a>0$. Then, it is easy to see that Theorem 3 applies as soon as the following Assumptions (CTMN) hold:

(CTMN)

- $f_{\mathbf{0},1}$ is positive on $\mathbb{R}^k$,
- $x\to f_{\mathbf{0},1}(x)$ is twice continuously differentiable ν a.e.,
- There exists a function $B\in L^2(f_{\mathbf{0},1}\cdot\nu)$ that upper bounds all following functions:

$$\frac{f_{\mathbf{0},1}(x-\mathbf{t})}{f_{\mathbf{0},1}(x)},\ \frac{1}{f_{\mathbf{0},1}(x)}\left|(1+|x_i|)\frac{\partial f_{\mathbf{0},1}}{\partial x_i}(x-\mathbf{t})\right|, i=1,\dots,k,\ \mathbf{t}\in\mathbb{T},$$

$$\frac{1}{f_{\mathbf{0},1}(x)}\left|(1+|x_i||x_j|)\frac{\partial^2 f_{\mathbf{0},1}}{\partial x_i\partial x_j}(x-\mathbf{t})\right|, i,j=1,\dots,k,\ \mathbf{t}\in\mathbb{T},\ \|\mathbf{t}\|\le\eta.$$

These assumptions are met when $f_{\mathbf{0},1}$ is the inverse of a polynomial with degree at least 2, among which the Cauchy density, or the Gaussian densities and the normalization of $\mathrm{ch}(x)^{-1}$.

2.3.2 Gaussian mixtures with unknown variance

Here $h=k(k+1)/2$ since α is the unknown variance. It is easy to see that Assumptions (CTMN) hold, and Theorem 3 applies, as soon as the $f_{\mathbf{t},\alpha}$ are the Gaussian distributions $\mathcal{N}(\mathbf{t},\alpha)$ on $\mathbb{R}^k$, $\mathbb{T}$ is compact, $\mathbb{A}$ is a compact subset of symmetric matrices that are positive definite.

2.4 General mixtures with unknown nuisance parameter

Let $\mathcal{F}=\{f_{\mathbf{t},\alpha},\mathbf{t}\in\mathbb{T},\alpha\in\mathbb{A}\}$ be a set of densities with respect to some dominating measure ν, where $\mathbb{T}$ is a compact subset of $\mathbb{R}^k$ and $\mathbb{A}$ is a compact subset of $\mathbb{R}^h$. The parameter $\mathbf{t}$ will characterize the population in the mixture, the parameter α will be the same for all populations. As a simple example one may think to Gaussian distributions (eventually multidimensional) with $\mathbf{t}$ the mean vector and α the variance matrix. One may define a mixture with p populations as

$$g_{p,\pi,\mathbf{T},\alpha} = \sum_{i=1}^{p} \pi_i f_{\mathbf{t}^i,\alpha}. \tag{16}$$

Here $\pi = (\pi_1, \dots, \pi_p)$ is a vector of non negative real numbers that sum to one, $\mathbf{T} = (\mathbf{t}^1, \dots, \mathbf{t}^p) \in \mathbb{T}^p$ and $\alpha \in \mathbb{A}$. One would like to use the LRT for testing (H_0) : "g is a mixture of p_0 populations" against (H_1) : "g is a mixture of p populations", where g is the density of i.i.d. observations $X_1, \dots, X_n$, and $p_0 < p$. This is the case when

$$\mathcal{M}_0 = \left\{ g_{p_0,\pi_0,\mathbf{T_0},\alpha},\ \pi_0 \in [0,1]^{p_0},\ \mathbf{T}_0 \in \mathbb{T}^{p_0},\ \alpha \in \mathbb{A},\ \sum_{i=1}^{p_0} \pi_i = 1,\ \ i = 1, \dots, p_0 \right\},$$

$$\mathcal{M} = \left\{ g_{p,\pi,\mathbf{T},\alpha},\ \pi \in [0,1]^{p},\ \mathbf{T} \in \mathbb{T}^{p},\ \alpha \in \mathbb{A},\ \sum_{i=1}^{p} \pi_i = 1,\ i = 1, \dots, p \right\}.$$

To understand what happens and how to do computations, the main point is to understand how two mixtures with eventually different number of populations may become close.

The main weak identifiability Assumption (WID) will be that $g_{p,\pi,\mathbf{T},\alpha} = g_{q,\pi',\mathbf{T}',\alpha'}$ if and only if $\alpha = \alpha'$ and $\sum_{i=1}^{p} \pi_i \delta_{\mathbf{t}^i} = \sum_{i=1}^{q} \pi'_i \delta_{\mathbf{t}'^i}$ where δ_z is the Dirac measure at z.

Then, if the parameterization $(\mathbf{t}, \alpha) \to f_{\mathbf{t},\alpha}$ is smooth enough, two mixtures become close if their parameter α becomes close, and their mixing measure becomes close in the weak topology.
Let now $g_{\mathbf{0}} = g_{p_0,\pi_0,\mathbf{T}_0,\alpha_0}$ be a particular mixture in $\mathcal{M}_0$ which has exactly p_0 populations and not fewer, that will denote the true unknown density of the observations. We denote by $\mathbf{t^{0,i}}$ the elements of $\mathbf{T}_0$. Since parameter α is identifiable, its maximum likelihood estimator is consistent under weak smoothness assumptions, so that to define the sets $\mathcal{S}$ and $\mathcal{S}_0$ by (2) and (3), one may restrict α by $\|\alpha - \alpha_0\| \leq \eta$ for some small η. Then, as seen in the previous subsections, the main point is to find $\mathcal{D}$ and $\mathcal{D}_0$, so as to be able to:

- understand how parameterization and smoothness may be used to compute the order of the bracketing entropy,
- define the Gaussian process that is used in the limiting distribution.

For these points, smoothness assumptions and bounding with a square integrable function have to be used together with some non degeneracy of functions that come in the norm appearing in denominator, when this one goes to zero. In fact, if it may be degenerate, it means that one has to go further in the order of the Taylor development until non degeneracy. This, of course, depends on particular examples.
A rather general situation is the following. Let $q = p - p_0$. Denote by $D_{\mathbf{t}} f_{\mathbf{t},\alpha}$ the k-dimensional vector of derivatives of $f_{\mathbf{t},\alpha}$ with respect to $\mathbf{t}$, $D_\alpha f_{\mathbf{t},\alpha}$ the h-dimensional vector of derivatives of $f_{\mathbf{t},\alpha}$ with respect to α, $D^2_{\mathbf{t}} f_{\mathbf{t},\alpha}$ the $k \times k$-dimensional matrix of second derivatives of $f_{\mathbf{t},\alpha}$ with respect to $\mathbf{t}$. Introduce Assumptions (GM):

(GM)

- $(\mathbf{t}, \alpha) \to f_{\mathbf{t},\alpha}$ is three times continuously differentiable ν a.e. at any $(\mathbf{t}, \alpha) \in \mathbb{T} \times \mathbb{A}$,
- $\exists \eta > 0$ such that, for all $\alpha^i \in \mathbb{R}^h,\ \tilde{\mathbf{t}}^i \in \mathbb{R}^k,\ \tau^i \in \mathbb{R}^k,\ \delta^i \in \mathbb{R}^h,\ \pi_i \in \mathbb{R},\ i = 1, .., p_0$, for all $\rho_1, \dots, \rho_q \geq 0$ such that $\|\alpha^i - \alpha_0\| \leq \eta$, $\|\tilde{\mathbf{t}}^i - \mathbf{t}^{0,i}\| \leq \eta$, $\sum_i \rho_i + \sum_j \pi_j = 0$ then : $\sum_{i=1}^{q} \rho_i f_{\mathbf{t}^i,\alpha_0} + \sum_{i=1}^{p_0} \pi_i f_{\mathbf{t}^{i,0},\alpha_0} + \sum_{i=1}^{p_0} \langle \delta^i, D_\alpha f_{\mathbf{t}^{i,0},\alpha^i} \rangle + \sum_{i=1}^{p_0} \langle \tau^i, D_{\mathbf{t}} f_{\tilde{\mathbf{t}}^i,\alpha_0} \rangle = 0$ ν a.e. if and only if
$\sum_{i=1}^{q} \rho_i f_{\mathbf{t}^i,\alpha_0} + \sum_{i=1}^{p_0} \pi_i f_{\mathbf{t}^{0,i},\alpha_0} = 0$, $\delta^1 = \mathbf{0}, \dots, \delta^{p_0} = \mathbf{0}$ and $\tau^1 = \mathbf{0}, \dots, \tau^{p_0} = \mathbf{0}$,
- For any subset J of at most $\inf\{p_0, q\}$ points in $\mathbb{T}$ such that for each one there is one of the $\mathbf{t}^{0,i}$s at distance at most η, for any $(\tau^j)_{j \in J}$ of vectors of $\mathbb{R}^k$, for any $\delta^1, \dots, \delta^{p_0}$ in $\mathbb{R}^h$:
$\sum_{i=1}^{p_0} \langle \delta^i, D_\alpha f_{\mathbf{t}^{0,i},\alpha^i} \rangle + \sum_{j \in J} (\tau^j)^T D^2_{\mathbf{t}} f_{j,\alpha_0} (\tau^j) = 0$ ν a.e. if and only if $\delta^1 = \mathbf{0}, \dots, \delta^{p_0} = \mathbf{0}$ and $\tau^j = \mathbf{0}$, $j \in J$;

- There exists a function $B \in L^2(g_0 \cdot \nu)$ that upper bounds all following functions:

$$\frac{f_{\mathbf{t},\alpha}}{g_0},\ \frac{1}{g_0}\left|\frac{\partial f_{\mathbf{t},\alpha}}{\partial t_i}\right|, i=1,\ldots,k,\ \frac{1}{g_0}\left|\frac{\partial f_{\mathbf{t},\alpha}}{\partial \alpha_i}\right|, i=1,\ldots,h, (\mathbf{t},\alpha)\in\mathbb{T}\times\mathbb{A},\ \|\alpha-\alpha_0\|\leq\eta$$

$$\frac{1}{g_0}\left|\frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial t_i\partial t_j}\right|, i,j=1,\ldots,k,\ (\mathbf{t},\alpha)\in\mathbb{T}\times\mathbb{A},\ \|\alpha-\alpha_0\|\leq\eta$$

$$\frac{1}{g_0}\left|\frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial t_i\partial \alpha_j}\right|, i=1,\ldots,k,\ j=1,\ldots,h,\ \frac{1}{g_0}\left|\frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial \alpha_i\partial \alpha_j}\right|,\ i,j=1,\ldots,h,$$

$$\frac{1}{g_0}\left|\frac{\partial^3 f_{\mathbf{t},\alpha}}{\partial t_i\partial t_j\partial t_l}\right|,\ i,j,l=1,\ldots,k,\frac{1}{g_0}\left|\frac{\partial^3 f_{\mathbf{t},\alpha}}{\partial t_i\partial t_j\partial \alpha_l}\right|,\ i,j=1,\ldots,k,l=1,\ldots,h$$

$$(\mathbf{t},\alpha)\in\mathbb{T}\times\mathbb{A},\ \|\alpha-\alpha_0\|\leq\eta,\ \|\mathbf{t}-\mathbf{t}^i\|\leq\eta \text{ for some } i.$$

Set $\Delta = ((\delta^1)^T,\ldots,(\delta^{p_0})^T)$, with $\delta_i\in\mathbb{R}^h$; $\Theta = ((\tau^1)^T,\ldots,(\tau^{p_0})^T)$, with $\tau_i\in\mathbb{R}^d$; $\mathbf{T} = (\mathbf{t}^1,\ldots,\mathbf{t}^q)\in\mathbb{T}^q$; $\Xi=(\rho_1,\ldots,\rho_q)\in\mathbb{R}^q$; $\Pi=(\pi_1,\ldots,\pi_{p_0})\in[0,1]^{p_0}$,

$$H_{\mathbf{T},\Xi,\Pi,\Delta,\Theta} = \sum_{i=1}^{q}\rho_i f_{\mathbf{t}^i,\alpha_0} + \sum_{i=1}^{p_0}\pi_i f_{\mathbf{t}^{0,i},\alpha_0} + \sum_{i=1}^{p_0}\langle\delta^i, D_\alpha f_{\mathbf{t}^{0,i},\alpha_0}\rangle + \sum_{i=1}^{p_0}\langle\tau^i, D_{\mathbf{t}} f_{\mathbf{t}^{0,i},\alpha_0}\rangle,$$

and

$$d_{\mathbf{T},\Xi,\Pi,\Delta,\Theta} = \frac{H_{\mathbf{T},\Xi,\Pi,\Delta,\Theta}/g_{\mathbf{0}}}{\|H_{\mathbf{T},\Xi,\Pi,\Delta,\Theta}/g_{\mathbf{0}}\|_2}.$$

Define now:

$$\mathcal{K} = \Big\{(\mathbf{T},\Xi,\Pi,\Delta,\Theta):\ \rho_1,\ldots,\rho_q\geq 0;\ \sum_i\rho_i+\sum_i\pi_i=0;$$
$$\|\delta^1\|^2+\ldots+\|\delta^{p_0}\|^2+\|\tau^1\|^2+\ldots+\|\tau^{p_0}\|^2+\sum_i\rho_i^2+\sum_i\pi_i^2=1;\ H_{(\mathbf{T},\Xi,\Pi,\Delta,\Theta)}\neq 0\Big\}.$$

Then:

$$\mathcal{D} = \{d_{\mathbf{T},\Xi,\Pi,\Delta,\Theta}, (\mathbf{T},\Xi,\Pi,\Delta,\Theta)\in\mathcal{K}\},$$

and

$$\mathcal{D}_0 = \{d_{\mathbf{0},\mathbf{0},\mathbf{0},\Delta,\Theta}\}.$$

It will be possible to obtain the asymptotic distributions in the same way as in Section 2.2 under Assumptions (WID) and (GM). Define the Gaussian field $\mathcal{Z}(\mathbf{T},\Xi,\Pi,\Delta,\Theta)$ on $\mathcal{K}$ with covariance

$$r\left((\mathbf{T},\Xi,\Pi,\Delta,\Theta),(\mathbf{T}',\Xi',\Pi',\Delta',\Theta')\right) = \int d_{\mathbf{T},\Xi,\Pi,\Delta,\Theta}\, d_{\mathbf{T}',\Xi',\Pi',\Delta',\Theta'}\, g_{\mathbf{0}} d\nu.$$

Notice that, as in previous sections, $\mathcal{K}$ is not closed, and $r(\cdot,\cdot)$ is not continuous on some limiting points, but may be extended in some sense, as has been done for instance in Section 2.1.

Let also $p_n,\pi_n,\mathbf{T}_n,\alpha_n$ be such that $\sqrt{n}\|\frac{g_{p_n,\pi_n,\mathbf{T}_n,\alpha_n}-g_{\mathbf{0}}}{g_{\mathbf{0}}}\|_2$ tends to some positive constant c, with $\frac{g_{p_n,\pi_n,\mathbf{T}_n,\alpha_n}-g_{\mathbf{0}}}{g_{\mathbf{0}}}/\|\frac{g_{p_n,\pi_n,\mathbf{T}_n,\alpha_n}-g_{\mathbf{0}}}{g_{\mathbf{0}}}\|_2$ tending to $\bar{d}$ in the closure of $\mathcal{D}$.

Theorem 4 *If (WID) and (GM) hold, then $(g_{\mathbf{0}}\cdot\nu)^{\otimes n}$ and $(g_{p_n,\pi_n,\mathbf{T}_n,\alpha_n}\cdot\nu)^{\otimes n}$ are mutually contiguous, $2\,\lambda_n$ converges under $(g_{\mathbf{0}}\cdot\nu)^{\otimes n}$ in distribution to*

$$\left(\sup_{(\mathbf{T},\Xi,\Pi,\Delta,\Theta)\in\mathcal{K}}\mathcal{Z}(\mathbf{T},\Xi,\Pi,\Delta,\Theta)\right)^2 - \left(\sup_{(\mathbf{0},\mathbf{0},\mathbf{0},\Delta,\Theta)\in\mathcal{K}}\mathcal{Z}(\mathbf{0},\mathbf{0},\mathbf{0},\Delta,\Theta)\right)^2,$$

and under $(g_{p_n,\pi_n,\mathbf{T}_n,\alpha_n}\cdot\nu)^{\otimes n}$ *to*

$$\left(\sup_{(\mathbf{T},\Xi,\Pi,\Delta,\Theta)\in\mathcal{K}}\mathcal{Z}(\mathbf{T},\Xi,\Pi,\Delta,\Theta)+c\int d_{\mathbf{T},\Xi,\Pi,\Delta,\Theta}\bar{d}g_{\mathbf{0}}d\nu\right)^2 - \left(\sup_{(\mathbf{0},\mathbf{0},\mathbf{0},\Delta,\Theta)\in\mathcal{K}}\mathcal{Z}(0,0,0,\Delta,\Theta)+c\int d_{\mathbf{0},\mathbf{0},\mathbf{0},\Delta,\Theta}\bar{d}g_{\mathbf{0}}d\nu\right)^2. \quad (17)$$

It is possible to reduce the formula of the asymptotic distributions in Theorem 4 into only one supremum, using linear algebra computations as in previous sections. We shall not give the result for all situations since it involves too long and complicated formula. However, in case $q=1$, the result takes a simpler form that we will give below. For this one needs to define notations. When $q=1$, Ξ is reduced to ρ and $\mathbf{T}$ reduces to $\mathbf{t}$ so that elements of $\mathcal{D}$ may be written as $d_{\mathbf{t},\Pi,\Delta,\Theta}$ with

$$H_{\mathbf{t},\Pi,\Delta,\Theta}=\sum_{i=1}^{p_0}\pi_i(f_{\mathbf{t},\alpha_0}-f_{\mathbf{t}^{0,i},\alpha_0})+\sum_{i=1}^{p_0}\langle\delta^i,D_\alpha f_{\mathbf{t}^{0,i},\alpha_0}\rangle+\sum_{i=1}^{p_0}\langle\tau^i,D_{\mathbf{t}}f_{\mathbf{t}^{0,i},\alpha_0}\rangle.$$

where $\sum_{i=1}^{p_0}\pi_i\geq 0$.
Let W be the $p_0(h+d)$-dimensional centered Gaussian random variable with variance Σ such that for all Δ and Θ,

$$\begin{pmatrix}\Delta\\ \Theta\end{pmatrix}^T\Sigma\begin{pmatrix}\Delta\\ \Theta\end{pmatrix}=\|\frac{H_{0,0,\Delta,\Theta}}{g_0}\|_2^2.$$

Let $Z(\mathbf{t})$ be the (p_0)-dimensional centered Gaussian field with covariance the $p_0\times p_0$ matrix $\Gamma(\cdot,\cdot)$ such that for all $\mathbf{t}_1$, $\mathbf{t}_2$,

$$\Gamma(\mathbf{t}_1,\mathbf{t}_2)_{i,j}=\int\left(\frac{\frac{f_{\mathbf{t},\alpha_0}-f_{\mathbf{t}^{0,i},\alpha_0}}{g_0}}{\|\frac{f_{\mathbf{t},\alpha_0}-f_{\mathbf{t}^{0,i},\alpha_0}}{g_0}\|_2}\right)\left(\frac{\frac{f_{\mathbf{t},\alpha_0}-f_{\mathbf{t}^{0,j},\alpha_0}}{g_0}}{\|\frac{f_{\mathbf{t},\alpha_0}-f_{\mathbf{t}^{0,j},\alpha_0}}{g_0}\|_2}\right)g_0d\nu,$$

and let

$$\Gamma=\Gamma(\mathbf{t})=\Gamma(\mathbf{t},\mathbf{t}).$$

Define $C=C(\mathbf{t})$ the $p_0(h+d)\times p_0$ matrix such that for all $(\mathbf{t},\Pi,\Delta,\Theta)$,

$$\begin{pmatrix}\Delta\\ \Theta\end{pmatrix}^T C(\mathbf{t})\Pi=\sum_{i=1}^{p_0}\pi_i\langle\frac{\frac{f_{\mathbf{t},\alpha_0}-f_{\mathbf{t}^{0,i},\alpha_0}}{g_0}}{\|\frac{f_{\mathbf{t},\alpha_0}-f_{\mathbf{t}^{0,i},\alpha_0}}{g_0}\|_2},\frac{H_{0,0,\Delta,\Theta}}{g_0}\rangle_2,$$

and let $A=A(\mathbf{t})$, $U=U(\mathbf{t})$, $M=M(\mathbf{t})$ be matrices such that

$$M = \left(\Sigma-C\Gamma^{-1}C^T\right)^{-1} \quad (18)$$
$$U = -MC\Gamma^{-1} \quad (19)$$
$$A = \Gamma^{-1}+\Gamma^{-1}C^TMC\Gamma^{-1}. \quad (20)$$

Let $\mathbf{1}$ denote the p_0-dimensional vector with all coordinates equal to 1. Then:

Theorem 5 *Assume (WID) and (GM), and* $p=p_0+1$. *Then* $2\,\lambda_n$ *converges under* $(g_{\mathbf{0}}\cdot\nu)^{\otimes n}$ *in distribution to*

$$\sup_{\mathbf{t}}\left(AZ+U^TW\right)^T\left(A^{-1}-\frac{\mathbf{1}\mathbf{1}^T}{\mathbf{1}^TA\mathbf{1}}1_{(AZ+U^TW)^T\mathbf{1}<0}\right)\left(AZ+U^TW\right)$$

The distribution under contiguous alternatives is rather difficult to express in its full generality so it is omitted for simplicity. The proof of Theorem 5 is given in Section 5.

In the case of Gaussian mixtures with unknown variance, the assumption "$\sum_{i=1}^{p_0}\langle\delta^i, D_\alpha f_{\mathbf{t}^{0,i},\alpha^i}\rangle + \sum_{j\in J}(\tau^j)^T D^2_{\mathbf{t}} f_{j,\alpha_0}(\tau^j) = 0$ ν a.e. if and only if $\delta^1 = \mathbf{0},\dots,\delta^{p_0} = \mathbf{0}$ and $\tau^j = \mathbf{0}$, $j \in J$" does not hold. Indeed, second derivatives with respect to $\mathbf{t}$ are proportional with derivatives with respect to α. In this case, it is necessary to go further in the Taylor development: when taking third derivative with respect to $\mathbf{t}$, the condition of non degeneracy holds. Also, all derivatives till fourth order may be uniformly upper bounded with some function B as needed. Since the limiting points of process Z need not to be known at boundary values of $\mathcal{K}$ to define the asymptotic distribution of λ_n, the following result holds:

Theorem 6 *The asymptotic distributions under the null hypothesis and under contiguous hypotheses given in Theorem 4 and Theorem 5 hold for Gaussian mixtures with the same unknown variance matrix.*

3 The LRT for contamination mixtures when the set of parameters is large.

As already said in the introduction, the asymptotic distribution of the LRT for compact $\mathbb{T}$ and $\mathbb{A}$ can be used in practice for large data sets. The LRT happens in this case to be more powerful than moment tests as shown in Delmas (2003). Nevertheless it suffers from two drawbacks:

- the distribution is not free from the location of the null hypothesis inside $\mathbb{T}$,
- for testing one population against two (or p_0 against p) the LRT with bounded parameter is not invariant by translation or change of scale.

Several solutions to the first point exist. Threshold calculation can be conducted under the "worst" form of the null hypothesis (see Delmas, 2003) or one can use a "Plug-in", that is an estimate of f_0. It remains that results would be nicer if one would be able to get rid of the compactness assumption. This section and the next one answer by the negative, showing that in the simplest case: contamination for translation mixtures on $\mathbb{R}$, the LRT is theoretically less powerful than moment tests under contiguous alternatives. As already said in the introduction, the convergence to this result is very slow, so it is not so relevant in practice. It mainly shows that it is difficult to construct an unbounded asymptotic theory of the LRT.

We consider in this section the contamination mixture model (1) with $\mathbb{T} = [-T, T]$ for a given positive real number T and ν the Lebesgue measure. We use notations and results of Section 2.1. Let π_n and $\mathbf{t}_n$ be sequences such that:

- $\lim_{n\to+\infty}\sqrt{n}\pi_n\|(f_{t_n} - f_0)/f_0\|_2 = c$ for some positive c,
- either t_n tends to some $t_0 \neq 0$ and $\sqrt{n}\pi_n$ tends to some positive constant, or t_n tends to 0, and $t_n/\|t_n\|$ converges to some limit τ.

Let $\mathbb{P}_{\pi_n,t_n} = \Big(g_{\pi_n,t_n}\cdot\nu\Big)^{\otimes n}$ and $\mathbb{P}_0 = \Big(f_0\cdot\nu\Big)^{\otimes n}$. To evaluate the asymptotic power and the asymptotic level for large values of T, one has to investigate the behaviour of suprema of the Gaussian processes $Z(t)$ and $Z(t) + m(t)$ as defined in Theorem 1. Z is the centered Gaussian process defined in Section 2.1 with covariance given by (9). For simplicity we consider this process as defined on the whole real line $\mathbb{R}$. We will use assumptions under which the supremum of $Z(\cdot)$ over $[-T, T]$ tends to infinity as T tends to infinity, and is achieved for some t tending to infinity. So the study of this supremum on $[0, T]$ for large T can be replaced by the study of the supremum on $[1, T]$. The discontinuity of the covariance function r at 0 will have for us no consequence on the extreme behaviour of the process Z. We shall use Azaïs and Mercadier (2004) to derive the asymptotic distribution of suprema of Gaussian processes. Hence let $M(a, b) = \sup_{t\in(a,b)}(Z(t) + m(t))$. Since the asymptotic distribution of $2\,\lambda_n$, under the null hypothesis or under contiguous alternatives, in Theorem 1 can be written as $M(-T, T)^2$ (taking $m(t) = 0$ under the null hypothesis and $m(t) = \mu(t)$ as defined by (11) under contiguous alternatives), we want to characterize asymptotic behaviours of $M(-T, T)$ as $T \to +\infty$. We thus introduce further notations and assumptions.

Write $r_{ij}(s,t)$ instead of $\frac{\partial^{i+j}}{\partial^i s \partial^j t} r(s,t)$ and define $R(t) = \int_0^t \sqrt{r_{11}(s,s)}\, ds$.

Let $a_t = \sqrt{2 \log \circ R(t)}, b_t = a_t - \frac{\log(2\pi)}{a_t}$ and $\tilde{b}_t = a_t - \frac{\log(\pi)}{a_t}$.

Let $V = \{V(t) = Z(R^{-1}(t)) + m(R^{-1}(t)), t \in \mathbb{R}\}$ be the "unit-speed" transformation of $Z + m$ in the sense that the variance of $V^{'}(t)$ equals 1 for all t in $\mathbb{R}$. We denote by r^V its covariance function.

We shall use the following Assumptions (G) on r and μ:

(G)

$$
\begin{array}{ll}
(G1) & \forall\, t \in \mathbb{R}, \quad r_{11}(t,t) > 0 \quad \text{and} \quad \lim_{t \to +\infty} R(t) = +\infty, \\
(G2) & r(s,t) \log |R(s) - R(t)| \to 0 \text{ as } |R(s) - R(t)| \to +\infty, \\
(G3) & \forall\, \varepsilon > 0 \quad \sup_{|R(s)-R(t)|\, >\, \varepsilon} |r(s,t)| < 1, \\
(G4) & \star\ r \text{ is four times continuously differentiable,} \\
 & \star\ s \to r_{11}(s,s) \text{ three times continuously differentiable,} \\
 & \star\ \forall\, \gamma > 0, r^Y_{01} \text{ and } r^Y_{04} \text{ are bounded on } \{(s,t) \in \mathbb{R}^2, |s| > \gamma \text{ and } |t| > \gamma\}, \\
(G5) & \sqrt{\log \circ R(t)}\, \mu(t) \xrightarrow[t \to +\infty]{} 0.
\end{array}
$$

We have:

Theorem 7 *Assume (CM) and (G). Then, as T tends to infinity, $a_T\big(M(-T,T) - \tilde{b}_T\big)$ tends in distribution to the Gumbel distribution when $m(t) = 0$ as well as when $m(t) = \mu(t)$. In other words, if $c_{T,\alpha}$ is the threshold of the test defined by*

$$\lim_{n \to +\infty} \mathbb{P}_0 \left(\lambda_n > c_{T,\alpha}\right) = \alpha,$$

then for any contiguous alternative, the limiting power of the LRT equals its level:

$$\lim_{T \to +\infty} \lim_{n \to +\infty} \mathbb{P}_{\pi_n, t_n} \left(\lambda_n > c_{T,\alpha}\right) = \alpha.$$

Theorem 7 is proved in Section 5.

The theorem says that for T large enough, asymptotically, the LRT cannot distinguish the null hypothesis from any contiguous alternatives.

We shall consider the translation mixture model defined in Section 2.1.1. Let f_0 be a density on $\mathbb{R}$ satisfying Assumptions (H) where we denote by $f_0^{(i)}$ the derivative of f_0 of order i.

(H)

$$
\begin{array}{ll}
(H1) & \forall x \in \mathbb{R},\ f_0(x) > 0,\ f_0 \text{ four times continuously differentiable,} \\
 & \text{and } \forall\, i = 1, \ldots, 4, \quad \exists\, K_i > 0, \forall x \in \mathbb{R}, \left|\frac{f_0^{(i)}}{f_0}(x)\right| \leq K_i, \\
(H2) & \forall x \in \mathbb{R},\ \lim_{t \to +\infty} \frac{f_0(x+t)}{f_0(t)} = \lim_{t \to -\infty} \frac{f_0(x+t)}{f_0(t)} = 1, \\
(H3) & \exists\, M > 0,\ \forall\, x, t \in \mathbb{R},\ \frac{f_0(x) f_0(t)}{f_0(x+t)} \leq M, \\
(H4) & \exists\, F \in L^2(\lambda)\ :\ \sup_{|t| \geq 1} \log |t| \sqrt{f_0(x+t)} \leq F(x), \\
(H5) & \lim_{t \to +\infty} \log(t) \sqrt{f_0(t)} = 0.
\end{array}
$$

Our result is now:

Corollary 1 *Assume (H). Then Theorem 7 applies to the translation mixture model.*

Proof of Corollary 1 is given in Section 5.

Remarks:

- Assumptions $(H1)$ to $(H5)$ are essentially conditions on the tail of f_0. $(H4)$ and $(H5)$ are very weak and hold for all usual distributions. But $(H1)$ to $(H3)$ though rather weak, are more restrictive. They hold for example if $f_0(t) = O(t^{-\alpha})$ for $\alpha > 0$ as $t \to +\infty$ and $f_0(t) = O(t^{-\beta})$ for $\beta > 0$ as $t \to -\infty$. For instance, they hold for f_0 being the inverse of a polynomial and in particular for the Cauchy density.
- The proof relies on the verification of assumptions of Theorem 7. In particular, asymptotic behaviours of the covariance r and its derivatives have to be checked. Assumptions (H) only express sufficient conditions under which the asymptotic analysis is done with some generality. However, though $(H2)$ does not hold for the Gaussian density, we also verified that Theorem 7 holds for other densities such as the Gaussian and the normalization of $\mathrm{ch}(x)^{-1}$ in spite of different justifications.

LRT has to be compared with other testing procedures such as sample mean or Kolmogorov-Smirnov testing procedures.

- Denote by $\mu_i = \int x^i f_0(x) d\nu(x)$. Without loss of generality one can assume that $\mu_1 = 0$. If $\mu_2 < +\infty$ applying Le Cam's third Lemma, that is Theorem 6.6 of van der Vaart (1998),

  $\sqrt{n}\,\overline{X}_n$ converges in distribution, as n tends to infinity, to the Gaussian $N(0, \mu_2)$ under $\mathbb{P}_0$ and to the Gaussian $N(\gamma, \mu_2)$ under $\mathbb{P}_{\pi_n, t_n}$, where $\gamma = c/\|\frac{f_0'}{f_0}\|_2$ if $t_n \to 0$ and $\gamma = ct_0/\|\frac{f_{t_0}-f_0}{f_0}\|_2$ if $t_n \to t_0 \neq 0$.

  Consequently the asymptotic power is greater than the level.

- Remark that, when no condition of moment is available, other tests can be also proposed. Define $\mathbb{F}_n$ the random distribution function and F_0 the distribution function associated to f_0. Let I denote the identity function on $[0,1]$ and let $\mathbb{U}$ be a Brownian bridge on $[0,1]$. Let $\|\cdot\|_\infty$ denotes the supremum norm. The natural normalization of $\mathbb{F}_n$ leads to the definition of the Kolmogorov-Smirnov statistic $\mathbb{K}_n$ and the Cramér-von Mises statistic $\mathbb{W}_n^2$:

  $$\mathbb{K}_n = \sqrt{n}\,\|\mathbb{F}_n - F_0\|_\infty \text{ and } \mathbb{W}_n^2 = \int_{-\infty}^{+\infty} n\,[\mathbb{F}_n(x) - F_0(x)]^2\, dF_0(x).$$

  Set on $[0,1]$

  $$\Delta(x) = \gamma \lim_{n\to+\infty} \frac{F_0\left(F_0^{-1}(x) - t_n\right) - x}{t_n},$$

  where t_n is the translation parameter of the alternative. Hence Δ depends on the asymptotic behaviour of t_n.

  - ⋆ $\mathbb{K}_n$ converges in distribution, as n tends to infinity, to $\|\mathbb{U}\|_\infty$ under $\mathbb{P}_0$ and $\|\mathbb{U} + \Delta\|_\infty$ under $\mathbb{P}_{\pi_n,t_n}$.
  - ⋆ $\mathbb{W}_n^2$ converges in distribution, as n tends to infinity, to $\int_0^1 \mathbb{U}^2\, dI$ under $\mathbb{P}_0$ and $\int_0^1 (\mathbb{U} + \Delta)^2\, dI$ under $\mathbb{P}_{\pi_n,t_n}$.

See Shorack and Wellner (1986) for a version of theses convergences. Simulations show that in both cases, the distribution under $\mathbb{P}_{\pi_n,t_n}$ is stochastically greater than that under $\mathbb{P}_0$. Consequently the asymptotic power is greater than the level.

4 Asymptotic distribution of the LRT for Gaussian contamination mixtures with unbounded mean under contiguous alternatives.

Consider $\mathbb{T} = \mathbb{R}$ (no prior upper bound) and the testing problem (1) with

$$f_t(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-t)^2}{2}\right).$$

Set

$$g_0 = f_0 \text{ and } g_{\pi,t} = (1-\pi)f_0 + \pi f_t,\ 0 \le \pi \le 1, t \in \mathbb{T}.$$

Let π_n and t_n be sequences such that $\lim_{n\to+\infty} \sqrt{n}\pi_n t_n = \gamma \in \mathbb{R}$ and $\lim_{n\to+\infty} t_n = t_0 \in \mathbb{R}$. Note that t_0 can be equal to zero.

λ_n is now given by:

$$\lambda_n = \sup_{\pi\in[0,1],\ t\in\mathbb{R}} \sum_{i=1}^n \log\Big(1 + \pi\left(\exp[tX_i - \frac{t^2}{2}] - 1\right)\Big).$$

Then:

Theorem 8 *As n tends to infinity, $\big(2\,\lambda_n - \log\circ\log n + \log(2\pi^2)\big)$ tends in distribution to the Gumbel distribution under $\mathbb{P}_0$ as well as under $\mathbb{P}_{\pi_n,t_n}$ for any γ and t_0. In other words, let us define as rejection values the region: $(\lambda_n > c_{\alpha,n})$ with*

$$c_{\alpha,n} = \frac{1}{2}\big(G_{1-\alpha} + \log\circ\log n - \log(2\pi^2)\big),$$

where $G_{1-\alpha}$ is the $1-\alpha$ fractile of the Gumbel distribution. We have by definition

$$\lim_{n\to+\infty} \mathbb{P}_0\left(\lambda_n > c_{\alpha,n}\right) = \alpha.$$

Then for any γ and t_0, the limit power of the LRT is

$$\lim_{n\to+\infty} \mathbb{P}_{\pi_n,t_n}\left(\lambda_n > c_{\alpha,n}\right) = \alpha.$$

The theorem says that asymptotically, the LRT cannot distinguish the null hypothesis from any contiguous alternative. Indeed, this has to be compared with other testing procedures such as moment testing procedures. For example, if $\overline{X}_n$ is the sample mean, applying Le Cam's third Lemma, $\sqrt{n}\,\overline{X}_n$ converges in distribution, under $\mathbb{P}_{\pi_n,t_n}$ as n tends to infinity, to the Gaussian $N(\gamma,1)$. Thus the test based on the statistic $\sqrt{n}\,\overline{X}_n$ has an asymptotic power that is strictly greater than the level. As mentioned in the introduction this makes sense in practice only for very large data sets.

Proof of Theorem 8

The separation of the hypotheses is greater when $\gamma \neq 0$. Using Lemma 14.31 of van der Vaart (1998) it is easy to see that this is the only case to consider. Moreover by symmetry we can suppose also that $\gamma > 0$. Let us introduce S_n the empirical process defined by

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big\{ \exp[tX_i - t^2] - \exp(-\frac{t^2}{2}) \Big\}.$$

Liu and Shao (2004, Theorem 1) recall results obtained by Bickel and Chernoff (1993) on the process S_n:

$$\sup_{t\in\mathbb{R}} S_n(t) = \sup_{|t|\in A_{2,n}} S_n(t) + o_{\mathbb{P}_0}(1) \tag{21}$$

where $A_{2,n} = [\alpha_n, \beta_n]$, $\alpha_n = 2\sqrt{\log\circ\log\circ\log n}$ and $\beta_n = \sqrt{\log n/2} - 2\sqrt{\log\circ\log n}$.

Through the proof of their Theorem 2 Liu and Shao (2004) state that

$$2\,\lambda_n \quad = \quad \sup_{t\in\mathbb{R}} S_n(t)^2 + o_{\mathbb{P}_0}(1).$$

Combining with (21), the last equality becomes

$$2\,\lambda_n \quad = \sup_{|t|\in A_{2,n}} S_n(t)^2 + o_{\mathbb{P}_0}(1).$$

Let us denote $\tilde{\mathbb{P}}_0$ the extension of $\mathbb{P}_0$ constructed by Bickel and Chernoff (1993) by Hungarian construction. According to their formula (39), we get

$$2\,\lambda_n = \sup_{|t|\in A_{2,n}} S_0(t)^2 + o_{\tilde{\mathbb{P}}_0}(1) \tag{22}$$

where S_0 is the zero mean non-stationary Gaussian process with covariance function

$$\rho(s,t) = \exp[-\frac{(s-t)^2}{2}] - \exp[-\frac{s^2}{2} - \frac{t^2}{2}].$$

In their paper, Bickel and Chernoff remark that this process is very close to a stationary process namely $\tilde{S}_0$. Because we need it later, we will use here an other way. We define the standardized version of S_0

$$Y_0(t) = \frac{S_0(t)}{\sqrt{\rho(t,t)}} = \frac{S_0(t)}{\sqrt{1-e^{-t^2}}},$$

in order to be able to apply the Normal Comparison Lemma (Li and Shao, 2002, Theorem 2.1). Y_0 is a zero mean non-stationary Gaussian process, with unit variance and covariance function

$$r(s,t) = \frac{\exp\big(st\big)-1}{\sqrt{\exp\big(s^2\big)-1}\sqrt{\exp\big(t^2\big)-1}}. \tag{23}$$

We have

$$0 \le \sup_{|t|\in A_{2,n}} |Y_0(t) - S_0(t)| \le \sup_{|t|\in A_{2,n}} (1-\sqrt{\rho(t,t)}) \sup_{|t|\in A_{2,n}} |Y_0(t)|.$$

Now the function r satisfies conditions of Corollary 1 of Azaïs and Mercadier (2004). Consequently we know the exact order of the maximum

$$\sup_{|t|\in A_{2,n}} |Y_0(t)| = O_{\tilde{\mathbb{P}}_0}\big((\log\circ\log n)^{\frac{1}{2}}\big).$$

This last equation can also be deduced from standard result on the maximum of stationary Gaussian processes using the process $\tilde{S}_0$ introduced by Bickel and Chernoff (1993).
On the other side, the maximum of $1-\sqrt{\rho(t,t)}$ on $A_{2,n}$ is obtained at α_n. This permits us to write

$$0 \le \sup_{|t|\in A_{2,n}} |Y_0(t) - S_0(t)| \le O_{\tilde{\mathbb{P}}_0}\big((\log\circ\log n)^{\frac{1}{2}-4}\big).$$

Finally this approximation allows us to replace S_0 by Y_0 in (22) to get

$$2\,\lambda_n = \sup_{|t|\in A_{2,n}} Y_0(t)^2 + o_{\tilde{\mathbb{P}}_0}(1). \tag{24}$$

With the same idea as before, we define

$$Y_n(t) = \frac{S_n(t)}{\sqrt{1-e^{-t^2}}}.$$

For all t_0 and all γ, using argument close to those that lead to formula (7) in Gassiat (2002) we have

$$\log \frac{d\mathbb{P}_{\pi_n,t_n}}{d\mathbb{P}_0}(X_1,\dots,X_n) = C(\gamma,t_0)Y_n(t_n) - \frac{C(\gamma,t_0)^2}{2} + o_{\mathbb{P}_0}(1) \tag{25}$$

with $C(\gamma, t_0) = \gamma$ if $t_0 = 0$ and $C(\gamma, t_0) = \gamma \frac{\sqrt{e^{t_0^2}-1}}{t_0}$ if $t_0 > 0$. Since γ can be supposed positive, t_0 is positive. A detailed proof of formula (25) is given in Section 5. Using the formula (39) of Bickel and Chernoff (1993), we can replace Y_n by Y_0 to get

$$\log \frac{d\mathbb{P}_{\pi_n,t_n}}{d\mathbb{P}_0}(X_1, \ldots, X_n) = C(\gamma, t_0) Y_0(t_n) - \frac{C(\gamma, t_0)^2}{2} + o_{\tilde{\mathbb{P}}_0}(1). \tag{26}$$

We next use the following lemma, its proof is given in Section 5.

Lemma 1 *For all t_0, $2\lambda_n - \log \circ \log n + \log(2\pi^2)$ and $\log \frac{d\mathbb{P}_{\pi_n,t_n}}{d\mathbb{P}_0}(X_1, \ldots, X_n)$ are asymptotically independent under $\mathbb{P}_0$.*

Then, as soon as one proves Lemma 1 the theorem follows from a generalization of Le Cam's third Lemma. The proof of Lemma 1 relies on a suitably chosen discretization, following ideas in Azaïs and Mercadier (2004), and an application of the normal comparison lemma as refined in Li and Shao (2002).

5 Proofs

5.1 Proof of Theorem 5

To obtain the supremum in the first limit of Theorem 4, one has to compute the supremum of:

$$\left(\begin{pmatrix} \Pi \\ \Delta \\ \Theta \end{pmatrix}^T \begin{pmatrix} Z \\ W \end{pmatrix} \right)^2 \tag{27}$$

under the constraints

$$\begin{pmatrix} \Pi \\ \Delta \\ \Theta \end{pmatrix}^T \tilde{\Sigma} \begin{pmatrix} \Pi \\ \Delta \\ \Theta \end{pmatrix} = 1, \ \Pi^T \mathbf{1} \geq 0, \tag{28}$$

where

$$\tilde{\Sigma} = \begin{pmatrix} \Gamma & C^T \\ C & \Sigma \end{pmatrix}.$$

Consider the sumpremum under the first constraint. Then, similarly to the proof of Theorem 2, the value of the supremum is

$$\left(AZ + U^T W\right)^T A^{-1} \left(AZ + U^T W\right) + W^T \Sigma^{-1} W$$

and it is attained on some Π such that $\Pi^T \mathbf{1}$ has the same sign as $(AZ + U^T W)^T \mathbf{1}$.
If $(AZ + U^T W)^T \mathbf{1} < 0$, then the supremum of (27) under (28) equals the supremum of (27) under the constraints

$$\begin{pmatrix} \Pi \\ \Delta \\ \Theta \end{pmatrix}^T \tilde{\Sigma} \begin{pmatrix} \Pi \\ \Delta \\ \Theta \end{pmatrix} = 1, \ \Pi^T \mathbf{1} = 0. \tag{29}$$

Computation of this supremum using Lagrange multipliers leads to the fact that it is equal to

$$\left(AZ + U^T W\right)^T \left(A^{-1} - \frac{\mathbf{1}\mathbf{1}^T}{\mathbf{1}^T A \mathbf{1}} \right) \left(AZ + U^T W\right) + W^T \Sigma^{-1} W$$

and the Theorem is proved.

5.2 Proof of Theorem 7

Set $u_{T,x} = \frac{x}{a_T} + \tilde{b}_T$ and $M^V(a, b) = \sup_{t \in (a,b)} V_t$ for V the unit-speed transformation of $Z + m$.

⋆ We have

$$\mathbb{P}\Big(M(-T,T) \le u_{T,x}\Big) = \mathbb{P}\Big(M^V\big(-R(T),R(T)\big) \le u_{T,x}\Big).$$

Now, applying with $p=2$, $D_1 = \big(-R(T), -\sqrt{R(T)}\big)$ and $D_2 = \big(\sqrt{R(T)}, R(T)\big)$ Proposition 4 of Azaïs and Mercadier (2004), we obtain

$$\mathbb{P}\big(M^V(D_1 \cup D_2) \le u_{T,x}\big) = \mathbb{P}\big(M^V(D_1) \le u_{T,x}\big)\mathbb{P}\big(M^V(D_2) \le u_{T,x}\big) + o(1).$$

Remark that in Azaïs and Mercadier (2004) sizes of intervals are defined as functions of the level, here it is the opposite which is made. Furthermore, repeated application of Corollary 1 of Azaïs and Mercadier (2004) enables us to state for $\tau = \sqrt{R(T)}$ and $\tau = R(T)$ the convergence of $a_\tau\big(M^V(0,\tau) - b_\tau\big)$ and $a_\tau\big(M^V(-\tau,0) - b_\tau\big)$ to the Gumbel. It follows that $M^V\big(-\sqrt{R(T)}, \sqrt{R(T)}\big)$ is stochastically negligible compared with $M^V\big(-R(T),R(T)\big)$ and also that $M\big(0,\sqrt{R(T)}\big)$ $\big($resp. $M\big(-\sqrt{R(T)},0\big)\big)$ is stochastically negligible compared with $M^V\big(0,R(T)\big)$ $\big($resp. $M^V\big(-R(T),0\big)\big)$. By combining what precedes, we get

$$\begin{aligned}&\mathbb{P}\Big(M^V\big(-R(T),R(T)\big) \le u_{T,x}\Big)\\ &\qquad = \mathbb{P}\Big(M^V\big(0,R(T)\big) \le u_{T,x}\Big)\mathbb{P}\Big(M^V\big(-R(T),0\big) \le u_{T,x}\Big) + o(1),\end{aligned}$$

as T tends to infinity, and which becomes

$$\mathbb{P}\Big(M(-T,T) \le u_{T,x}\Big) = \mathbb{P}\Big(M(0,T) \le u_{T,x}\Big)\mathbb{P}\Big(M(-T,0) \le u_{T,x}\Big) + o(1)$$

when we return to the initial process $Z+m$.

Let $G(x) = \exp\big(-\exp(-x)\big)$ denote the distribution function of the Gumbel. Corollary 1 of Azaïs and Mercadier (2004) yields, as T tends to infinity,

$$\begin{aligned}\mathbb{P}\Big(M(0,T) \le u_{T,x}\Big) &= \mathbb{P}\Big(a_T\big(M(0,T) - \tilde{b}_T\big) \le x\Big) + o(1)\\ &= \mathbb{P}\Big(a_T\big(M(0,T) - b_T\big) \le x + \log(2)\Big) + o(1)\\ &= G(x+\log(2)) + o(1).\end{aligned}$$

Since the same equality holds on $(-T,0)$, one can conclude that

$$\mathbb{P}\Big(M(-T,T) \le u_{T,x}\Big) = G(x+\log(2))^2 + o(1) = G(x) + o(1).$$

5.3 Proof of Corollary 1

The proof relies on the verification of assumptions of Theorem 7.

Proof of (CM): since f_0 is continuous and positive, for any real T,

$$\inf_{t\in[-T,T]} f_0(t) = \delta_T > 0.$$

Using $(H3)$, for all $t \in [-T,T]$ and $x \in \mathbb{R}$,

$$\left|\frac{f_t - f_0}{f_0}(x)\right| \le \sup_{x\in\mathbb{R}}\left|\frac{f_0(x-t)\, f_0(t)}{f_0(x)}\right| \frac{1}{f_0(t)} + 1 \le \frac{M}{\delta_T} + 1,$$

and using $(H1)$ and $(H3)$

$$\left|\frac{f_t'}{f_0}(x)\right| \le K_1\frac{M}{\delta_T}, \ \left|\frac{f_t''}{f_0}(x)\right| \le K_2\frac{M}{\delta_T}.$$

Let us now prove Assumptions (G). Set

$$N(s,t) = \int \frac{f_0(x-t)f_0(x-s)}{f_0(x)} d\nu(x).$$

Differentiation of r, for s and t in $\mathbb{R}\setminus\{0\}$, is a consequence of that of $N(s,t)$. Now, for any integers $i \le 4$ and $j \le 4$, using $(H1)$ and $(H3)$

$$\frac{f_0^{(i)}(x-t)f_0^{(j)}(x-s)}{f_0(x)} \le K_i K_j \frac{f_0(x-t)f_0(x-s)}{f_0(x)} \le K_i K_j M^2 \frac{f_0(x)}{f_0(t)f_0(s)}$$

and $f_0(t)f_0(s)$ is positively lower bounded on the neighbourhood of any (s_0,t_0), which proves that N is differentiable at any $(s,t) \in (\mathbb{R}\setminus\{0\})^2$ with

$$\frac{\partial^{i+j} N}{\partial^i t\, \partial^j s}(s,t) = (-1)^{i+j} \int \frac{f_0^{(i)}(x-t)f_0^{(j)}(x-s)}{f_0(x)} d\nu(x).$$

Proof of (G1): we thus have for $t \neq 0$

$$\begin{aligned} r_{11}(t,t) &= \frac{\int \frac{f_0'^2(x-t)}{f_0(x)} d\nu(x)\Big(\int \frac{f_0^2(x-t)}{f_0(x)} d\nu(x) - 1\Big) - \Big(\int \frac{f_0(x-t)f_0'(x-t)}{f_0(x)} d\nu(x)\Big)^2}{\Big(\int \frac{f_0(x-t)^2}{f_0(x)} d\nu(x) - 1\Big)^2} \\ &= \frac{\|\frac{f_0'(\cdot-t)}{f_0(\cdot)}\|_2^2 \|\frac{f_0(\cdot-t)-f_0(\cdot)}{f_0(\cdot)}\|_2^2 - \Big(\langle \frac{f_0'(\cdot-t)}{f_0(\cdot)}, \frac{f_0(\cdot-t)-f_0(\cdot)}{f_0(\cdot)}\rangle_2\Big)^2}{\|\frac{f_0(\cdot-t)-f_0(\cdot)}{f_0(\cdot)}\|_2^4} \end{aligned}$$

which is positive by Cauchy-Schwarz inequality. Now,

$$\lim_{t\to+\infty} r_{11}(t,t) = \frac{\int f_0'^2\, d\nu \int f_0^2\, d\nu - \Big(\int f_0 f_0'\, d\nu\Big)^2}{\Big(\int f_0'^2\, d\nu\Big)^2}.$$

Indeed, define the functions

$$\begin{aligned} A(t) &= \int \frac{f_0^2(x)}{f_0(x+t)}\, d\nu(x) \\ B(t) &= \int \frac{f_0'^2(x)}{f_0(x+t)}\, d\nu(x) \\ C(t) &= \int \frac{f_0(x)f_0'(x)}{f_0(x+t)}\, d\nu(x). \end{aligned}$$

Then write the function r_{11} under the following form:

$$r_{11}(t,t) = \frac{B(t)f_0(t)\big(A(t)f_0(t) - f_0(t)\big) - \big(C(t)f_0(t)\big)^2}{\big(A(t)f_0(t) - f_0(t)\big)^2}.$$

Thanks to $(H1)$ and $(H3)$, integrands of Af_0, Bf_0 and Cf_0 are respectively dominated by

$$Mf_0(x),\ K_1^2 M f_0(x),\ K_1 M f_0(x).$$

By application of $(H2)$ and Lebesgue Theorem, we conclude using the following convergences:

$$\begin{aligned}
\lim_{t\to+\infty} A(t)f_0(t) &= \int f_0^2(x)\, d\nu(x) \\
\lim_{t\to+\infty} B(t)f_0(t) &= \int f_0^{'2}(x)\, d\nu(x) \\
\lim_{t\to+\infty} C(t)f_0(t) &= \int f_0(x)f_0^{'}(x)\, d\nu(x).
\end{aligned}$$

Thus for a positive constant R

$$R(t) \sim_{t\to+\infty} R\,t. \tag{30}$$

Proof of (G2): considering (30), we have to prove that

$$\lim_{|s-t|\to+\infty} r(s,t)\log|s-t| = 0. \tag{31}$$

Using $(H3)$,

$$\frac{f_0(t)f_0^2(x)}{f_0(x+t)} \le M f_0(x),$$

so that using $(H2)$,

$$\lim_{t\to+\infty} \int \frac{f_0(t)f_0^2(x)}{f_0(x+t)} d\nu(x) = \int f_0^2(x)d\nu(x),$$

and there exists a constant C such that for $|s-t|$ large enough,

$$r(s,t) \le C \int \sqrt{f_0(t)}\sqrt{f_0(s)}\frac{f_0(x-t)f_0(x-s)}{f_0(x)} d\nu(x).$$

Then, using $(H3)$,

$$r(s,t) \le \int C\, M\, \sqrt{f_0(x)}\sqrt{f_0(x+s-t)} d\nu(x).$$

But according to $(H5)$ for any $x\in\mathbb{R}$,

$$\lim_{|s-t|\to+\infty} \log|s-t|\sqrt{f_0(x+s-t)} = 0,$$

and so, one may apply Lebesgue Theorem using $(H4)$ to obtain (31).

Proof of (G5): $(G5)$ is a consequence of $(G2)$ and formula (11) giving $\mu(t)$.

Proof of (G3): Using (30) and the fact that $r_{11} > 0$, one just has to prove that for any $\varepsilon > 0$,

$$\sup_{|s-t|>\varepsilon} |r(s,t)| < 1. \tag{32}$$

First of all, $r(s,t)$ is a continuous function of (s,t) and $|r(s,t)| < 1$ as soon as $s \ne t$ by Cauchy-Schwarz inequality. Thus for any $\varepsilon > 0$, for any compact set K,

$$\sup_{|s-t|>\varepsilon,\ t\in K,\ s\in K} |r(s,t)| < 1.$$

On the other hand because of $(G2)$ for $|s-t|$ sufficiently large $r(s,t)$ is bounded away from 1, so we may suppose that $|s-t|$ is bounded. Suppose that there exists s_n and t_n such that $|s_n - t_n|$ is bounded, $|s_n - t_n| > \varepsilon$ and $r(s_n,t_n) \to 1$. By compactness it would be possible to choose subsequences $s_{\varphi(n)}$ and $t_{\varphi(n)}$ such that $s_{\varphi(n)} - t_{\varphi(n)} \to c$. But using the same tricks as before (using $(H2)$, $(H3)$ and Lebesgue Theorem),

$$\lim_{n\to+\infty} r(s_{\varphi(n)}, t_{\varphi(n)}) = \frac{\int f_0(x)f_0(x+c)d\nu(x)}{\int f_0^2(x)d\nu(x)}.$$

Since $|c| \ge \varepsilon > 0$ this value differs from 1. Hence we get a contradiction with assumptions made on sequences s_n and t_n and (32) is true.

Proof of (G4): The first part of $(G4)$ has been already proved. We use same arguments to prove that $s \mapsto r_{11}(s,s)$ is three times continuously differentiable. Now, this last regularity associated to (30) allow us to reduce our study to that of functions r_{01} and r_{04}.

$\star$ The first derivative $r_{01}(s,t)$ can be written as:

$$\frac{-\langle \frac{f_0'(\cdot - t)}{f_0}, \frac{f_0(\cdot - s)-f_0}{f_0}\rangle_2}{\|\frac{f_0(\cdot - s)-f_0}{f_0}\|_2 \|\frac{f_0(\cdot - t)-f_0}{f_0}\|_2} + \frac{\langle \frac{f_0(\cdot - t)-f_0}{f_0}, \frac{f_0(\cdot - s)-f_0}{f_0}\rangle_{f_0} \langle \frac{f_0'(\cdot - t)}{f_0}, \frac{f_0(\cdot - t)-f_0}{f_0}\rangle_{f_0}}{\|\frac{f_0(\cdot - s)-f_0}{f_0}\|_2 \|\frac{f_0(\cdot - t)-f_0}{f_0}\|_2^3}$$

then Cauchy-Schwarz inequality leads to

$$|r_{01}(s,t)| \leq 2 \frac{\|\frac{f_0'(\cdot - t)}{f_0}\|_2}{\|\frac{f_0(\cdot - t)-f_0}{f_0}\|_2}.$$

This upper bound is a continuous function on t. By making appear $f_0(t)$, it is easily seen that it converges, as t tends to infinity, to

$$2 \frac{\int f_0'^2 \, d\nu}{\int f_0^2 \, d\nu}.$$

Moreover for any $\delta > 0$, the denominator is lower bounded on $D_\delta = \{(s,t), s \in \mathbb{R}, |t| > \delta\}$. Consequently for any $\delta > 0$, $(s,t) \mapsto r_{01}(s,t)$ is bounded on $\mathbb{R}^2 \setminus D_\delta$.

$\star$ Using easy but tedious computations and Cauchy-Schwarz inequality once more, we have:

$$|r_{04}(s,t)| \leq \sum_{i \geq 1} \sum_{j \geq 1} \frac{\prod_{k=1}^4 \|\frac{f_0^{(k)}(\cdot - t)}{f_0}\|_2^{\alpha_{ijk}}}{\|\frac{f_0(\cdot - t)-f_0}{f_0}\|_2^i}$$

where the sums on i and j are finite and where for any i and j: $\sum_{k=1}^4 \alpha_{ijk} = i$. Previous arguments run again and permit us to assert that for any $\delta > 0$ the function $(s,t) \mapsto r_{04}(s,t)$ is bounded on $\mathbb{R}^2 \setminus D_\delta$.

5.4 Proof of formula (25)

Define the functions L and $\tilde{L}$ by $\log(1+u) = u - u^2/2 + u^2 L(u)$ and $\tilde{L}(u) = \sup_{|v|<u} |L(v)|$. It is clear that $\tilde{L}(u) \to 0$ as $u \to 0$ and we have

$$\begin{aligned} \log \frac{d\mathbb{P}_{\pi_n,t_n}}{d\mathbb{P}_0}(X_1,\ldots,X_n) &= \sum_{i=1}^n \log\left(1 + \pi_n(e^{t_n X_i - \frac{t_n^2}{2}} - 1)\right) \\ &= \pi_n \sum_{i=1}^n (e^{t_n X_i - \frac{t_n^2}{2}} - 1) + \frac{\pi_n^2}{2} \sum_{i=1}^n (e^{t_n X_i - \frac{t_n^2}{2}} - 1)^2 + S, \quad (33) \end{aligned}$$

with

$$|S| \leq n \pi_n^2 t_n^2 \frac{1}{n} \sum_{i=1}^n \left(\frac{e^{t_n X_i - \frac{t_n^2}{2}} - 1}{t_n} \right)^2 \tilde{L}\left(\pi_n t_n \max_{i=1,\ldots,n} \left(\frac{e^{t_n X_i - \frac{t_n^2}{2}} - 1}{t_n} \right) \right).$$

Now it suffices to remark that the random variables $\left(\frac{e^{t_n X - \frac{t_n^2}{2}} - 1}{t_n} \right)$, $n = 1, 2, \ldots$ for X of distribution $N(0,1)$ have bounded third moment. Applying the Markov inequality

$$\max_{i=1,\ldots,n} \left(\frac{e^{t_n X_i - \frac{t_n^2}{2}} - 1}{t_n} \right) = o_{\mathbb{P}_0}(\sqrt{n}).$$

Moreover, the class of functions

$$x \to \left(\frac{e^{t_n x - \frac{t_n^2}{2}} - 1}{t_n} \right)^2$$

is Glivenko-Cantelli in probability (indeed, it is the square of a Donsker class, as a consequence of Section 2.1), we get $S = o_{\mathbb{P}_0}(1)$ and

$$\frac{1}{n}\sum_{i=1}^{n}[(e^{t_n X_i - \frac{t_n^2}{2}} - 1)^2 - (e^{t_n^2} - 1)] = o_{\mathbb{P}_0}(1),$$

so that

$$\begin{aligned} \log \frac{d\mathbb{P}_{\pi_n, t_n}}{d\mathbb{P}_0}(X_1, \ldots, X_n) &= \sqrt{n}\pi_n \sqrt{e^{t_n^2} - 1} Y_n(t_n) + n\frac{\pi_n^2}{2}\frac{1}{n}\sum_{i=1}^{n}(e^{t_n X_i - \frac{t_n^2}{2}} - 1)^2 + o_{\mathbb{P}_0}(1) \\ &= \sqrt{n}\pi_n \sqrt{e^{t_n^2} - 1} Y_n(t_n) + n\frac{\pi_n^2}{2}(e^{t_n^2} - 1) + o_{\mathbb{P}_0}(1). \quad (34) \end{aligned}$$

Now setting $C(\gamma, t_0) = \gamma$ if $t_0 = 0$ and $C(\gamma, t_0) = \gamma \frac{\sqrt{e^{t_0^2} - 1}}{t_0}$ we have

$$\sqrt{n}\pi_n \sqrt{e^{t_n^2} - 1} = C(\gamma, t_0) + o(1)$$

and

$$\log \frac{d\mathbb{P}_{\pi_n, t_n}}{d\mathbb{P}_0}(X_1, \ldots, X_n) = C(\gamma, t_0) Y_n(t_n) - \frac{C(\gamma, t_0)^2}{2} + o_{\mathbb{P}_0}(1).$$

5.5 Proof of Lemma 1

Beforehand we set $c_n = (\log \circ \log n)^{\frac{1}{2}}$ and we recall that $A_{2,n} = [\alpha_n, \beta_n]$ with $\alpha_n = 2\sqrt{\log \circ \log \circ \log n}$ and $\beta_n = \sqrt{\log n / 2} - 2\sqrt{\log \circ \log n}$.

According to (24) and (26) we need to prove that $\sup_{t \in A_{2,n}}(Y_0(t) - c_n)$ and $Y_0(t_0)$ are asymptotically independent. To this end, we consider the discretized process $\{Y_0(q_n k),\ k \in \mathbb{Z}\}$ with a step of discretization q_n depending on n in a sense which has to be defined. Let us gather the discretized points of $A_{2,n}$ in $A_{2,n}^{q_n} = \{d_1, \ldots, d_{N(n)}\}$.
By triangular inequalities and simplifications we have for any x and y

$$\begin{aligned} &\Big|\mathbb{P}\Big(\sup_{t \in A_{2,n}} Y_0(t) - c_n \le x; Y_0(t_0) \le y\Big) - \mathbb{P}(\sup_{t \in A_{2,n}} Y_0(t) - c_n \le x)\mathbb{P}(Y_0(t_0) \le y)\Big| \\ &\le 2\,\mathbb{P}\Big(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \le x; \sup_{t \in A_{2,n}} Y_0(t) - c_n > x\Big) + \quad (35) \\ &\Big|\mathbb{P}\Big(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \le x; Y_0(t_0) \le y\Big) - \mathbb{P}\Big(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \le x\Big)\mathbb{P}\Big(Y_0(t_0) \le y\Big)\Big| \end{aligned}$$

The task is now to prove that for fixed x and y each component of the upper bound converges to 0.

$\star$ We define the following modification of the function r

$$\begin{aligned} \tilde{r}(t_0, t) &= 0 \qquad t \in A_{2,n}^{q_n},\ t \neq t_0, \\ \tilde{r}(s, t) &= r(s, t) \qquad s, t \in A_{2,n}^{q_n}. \end{aligned}$$

Note that under the Gaussian distribution defined by $\tilde{r}$ the value of the process at t_0 is independent of the values of the process at other locations whose distribution does not changes. This proves that $\tilde{r}$ is a covariance function. We define $\xi(t) = \sup_{u,\ |u - t_0| > t} |r(u, t_0)|$.

From (23) we have

$$\xi(t) = O\Big(\exp\big(-\frac{t^2}{2}\big)\Big).$$

We restrict our attention to n's such that

$$c_n > 2|x| \ \ ; \ \ \xi(\alpha_n) < 1/2 \text{ so that } \frac{(x+c_n)^2}{2(1+\xi(\alpha_n))} \geq \frac{c_n^2}{12}$$

The Normal Comparison Lemma (Li and Shao, 2002, Theorem 2.1) gives bounds to terms of the type

$$\mathbb{P}\Big(Y_1 \leq u_1, ..., Y_n \leq u_n\Big) - \mathbb{P}\Big(\tilde{Y}_1 \leq u_1, ..., \tilde{Y}_n \leq u_n\Big)$$

where Y and $\tilde{Y}$ are two centered Gaussian vectors with the same variance and possibly different covariances ρ_{ij} and $\tilde{\rho}_{ij}$, $i,j = 1, ..., n$. It says that

$$\begin{aligned} &\mathbb{P}\Big(Y_1 \leq u_1, ..., Y_n \leq u_n\Big) - \mathbb{P}\Big(\tilde{Y}_1 \leq u_1, ..., \tilde{Y}_n \leq u_n\Big) \\ &\qquad \leq \frac{1}{2\pi} \sum_{1 \leq i < j \leq n} \big(\arcsin(\rho_{ij}) - \arcsin(\tilde{\rho}_{ij})\big)^+ \exp\left(-\frac{u_i^2 + u_j^2}{2(1+\bar{\rho}_{ij})}\right) \end{aligned} \qquad (36)$$

where $z^+ = \max\{z, 0\}$, $\bar{\rho}_{ij} = \max\{|\rho_{ij}|, |\tilde{\rho}_{ij}|\}$. Let $(Const)$ represents a generic positive constant. Since $\arcsin(x) \leq x\pi/2$ for $0 \leq x \leq 1$, applying inequality (36) in both directions to the vector Y_0 with covariance r and to the vector $\tilde{Y}_0$ with covariance $\tilde{r}$ corresponding to the points belonging to $\{t_0\} \cup A_{2,n}^{q_n}$ we get:

$$\begin{aligned} &\Big|\mathbb{P}\Big(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \leq x; Y_0(t_0) \leq y\Big) - \mathbb{P}\Big(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \leq x\Big)\mathbb{P}\Big(Y_0(t_0) \leq y\Big)\Big| \\ &\quad \leq \ (Const) \sum_{d \in A_{2,n}^{q_n}} |r(d,t_0)| \exp\Big(-\frac{(x+c_n)^2 + y^2}{2\big(1+|r(d,t_0)|\big)}\Big) \\ &\quad \leq \ (Const) \sum_{d \in A_{2,n}^{q_n}} |r(d,t_0)| \exp\Big(-\frac{c_n^2}{12}\Big) \\ &\quad \leq \ \frac{(Const)}{q_n} \exp\Big(-\frac{c_n^2}{12}\Big) \int_{\alpha_n - q_n}^{+\infty} \xi(t)dt = \frac{(Const)}{q_n} \exp\left(-\frac{c_n^2}{12}\right). \end{aligned}$$

which tends to zero if , for example, $q_n = \big(\log \circ \log n\big)^{-\theta}$ as soon as $\theta > 0$.

$\star$ To deal with the first term of (35), we denote by U_z and $U_z^{q_n}$ the point processes of up-crossings of level z for Y_0 and its q_n-polygonal approximation (linear interpolation) respectively. For any subset B of $\mathbb{R}$,

$$\begin{aligned} U_z(B) &= \#\{t \in B, Y_0(t) = z, Y_0^{'}(t) > 0\}, \\ U_z^{q_n}(B) &= \#\Big\{l \in \mathbb{Z},\, q_n(l-1) \in B,\, q_n l \in B,\, Y_0\big(q_n(l-1)\big) < z < Y_0\big(q_n l\big)\Big\}. \end{aligned}$$

Set Φ the distribution function of the standard Gaussian and $\overline{\Phi} = 1 - \Phi$.

$$\begin{aligned} &\mathbb{P}\Big(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \leq x; \sup_{t \in A_{2,n}} Y_0(t) - c_n > x\Big) \\ &\quad \leq \mathbb{P}\Big(Y_0(\alpha_n) > x + c_n\Big) + \mathbb{P}\Big(Y_0(\alpha_n) \leq x + c_n,\, U_{x+c_n}(A_{2,n}) \geq 1,\, U_{x+c_n}^{q_n}(A_{2,n}) = 0\Big) \\ &\qquad \leq \overline{\Phi}(x + c_n) + \mathbb{E}\Big(U_{x+c_n}(A_{2,n}) - U_{x+c_n}^{q_n}(A_{2,n})\Big) \end{aligned}$$

where the last upper bound is due to Markov inequality. The first term above tends trivially to zero, as for the second if we set $q_n = \big(\log \circ \log n\big)^{-\theta}$ with $\theta > \frac{1}{2}$, Condition (U7) of Lemma

2 of Azaïs and Mercadier (2004) is met. It is easy to check that since $\mathbb{E}\Big(U_{x+c_n}(A_{2,n})\Big)$ is bounded we are in the condition of application of that lemma and

$$\mathbb{E}\Big(U_{x+c_n}(A_{2,n}) - U^{q_n}_{x+c_n}(A_{2,n})\Big) = o(1).$$

References

Azaïs, J.-M. and Mercadier, C. (2004), Asymptotic Poisson character of extremes in non-stationary Gaussian models, *To appear in Extremes.*

Bickel, P. and Chernoff, H. (1993), Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem, *Statistics and Probability: A Raghu Raj Bahadur Festschrift (J.K. Ghosh, S.K. Mitra, K.R. Parthasarathy and B.L.S. Prakasa Rao eds)*, 83–96, Wiley Eastern Ltd.

Boistard, H. (2003), Test of goodness-of-fits for mixture of population, *Technical report*, Université Toulouse III, France.

Chen, H. and Chen, J. (2001), Large sample distribution of the likelihood ratio test for normal mixtures, *Statist. Probab. Lett.*, **2**, 125–133.

Chernoff, H. and Lander, E. (1995), Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial, *J. Statist. Plan. Inf.*, **43**, 19–40.

Dacunha-Castelle, D. and Gassiat, É. (1997), Testing in locally conic models, and application to mixture models, *ESAIM Probab. Statist.*, **1**, 285–317.

Dacunha-Castelle, D. and Gassiat, É. (1999), Testing the order of a model using locally conic parameterization: population mixtures and stationary ARMA processes, *Ann. Statist.*, **27**(4), 1178–1209.

Delmas, C. (2003), On likelihood ratio tests in Gaussian mixture models, *Sankhya*, **65**(3), 1–19.

Gassiat, É. (2002), Likelihood ratio inequalities with applications to various mixtures, *Ann. Inst. H. Poincaré Probab. Statist.*, **6**, 897–906.

Garel, B. (2001), Likelihood Ratio Test for Univariate Gaussian Mixture, *J. Statist. Plann. Inference*, **96**(2), 325–350.

Ghosh, J. and Sen, P. (1985), On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results, *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, vol. II, 789–806, Wadsworth, Belmont, CA.

Goffinet, B., Loisel, P., and B. Laurent. (1992), Testing in normal mixture models when the proportions are known, *Biometrika* **79**, 842–846.

Hall, P. and Stewart, M. (2004), Theoretical analysis of power in a two-component normal mixture model. *Private communication.*

Hartigan, J.A. (1985), A failure of likelihood asymptotics for normal mixtures, In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer (Berkeley, Calif., 1983)*, Vol. II, 807–810, Wadsworth, Belmont, CA.

James, L.F., Priebe, C.E., and Marchette, D.J. (2001), Consistent Estimation of Mixture Complexity, *Ann. Statist.*, **29**, 1281–1296.

Lemdani, M. and Pons, O. (1997), Likelihood ratio test for genetic linkage, *Statis. Probab. Lett.*, **33**(1), 15–22.

Lemdani, M. and Pons, O. (1999), Likelihood ratio tin contamination models, *Bernoulli*, **5**(4), 705–719.

Li, W.V. and Shao, Q. (2002), A normal comparison inequality and its applications, *Probab. Theory Related Fields*, **122**(4), 494–508.

Lindsay, B. G. (1995), *Mixture models: Theory, geometry, and applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 5, Hayward, Calif.: Institute for Mathematical Statistics.

Liu, X. and Shao, Y. (2004), Asymptotics for the likelihood ratio test in two-component normal mixture models, *J. Statist. Plann. Inference*, **123**(1), 61–81.

McLachlan, G. and Peel, D. (2000), *Finite mixture models*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York.

Mercadier, C. (2004), Computing the distribution of the maximum of random processes and fields: how far are the Rice and Euler Characteristics valid. *Preprint*, Université Toulouse III, France, http://www.lsp.ups-tlse.fr/Fp/Mercadier.

Mosler, K. and Seidel, W. (2001), Testing for homogeneity in an exponential mixture model, *Aust. N. Z. J. Stat.*, **43**(3), 231–247.

Seidel, W., Mosler, K., and Alker, M. (2000), A cautionary note on likelihood ratio tests in mixture models, *Ann. Instit. Statist. Math.*, **52**, 481–487.

Shorack, G.R. and Wellner, J.A. (1986), *Empirical processes with applications to statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc. New York.

Titterington, D.M., Smith, A.F.M., and Makov, U.E. (1985), *Statistical analysis of finite mixture distributions*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Ltd.

van der Vaart, A. W. (1998), *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.