Bernoulli **12**(5), 2006, 775–799

Asymptotic distribution and local power of the log-likelihood ratio test for mixtures: bounded and unbounded cases

JEAN-MARC AZAÏS[1*], ÉLISABETH GASSIAT[2] and CÉCILE MERCADIER[1**]

[1]*Laboratoire de Statistique et Probabilités, UMR-CNRS C55830, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 4. France. E-mail: *azais@cict.fr; **cecile.mercadier@u-paris10.fr*
[2]*Laboratoire de Mathématiques, Equipe de Probabilités, Statistique et Modélisation, Bât. 425, Université de Paris-Sud, 91405 Orsay Cedex, France. E-mail: Élisabeth.gassiat@math.u-psud.fr*

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, under general assumptions, that the asymptotic local 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 local power under contiguous hypotheses is equal to the asymptotic level.

Keywords: contiguity; extreme values; local power; log-likelihood ratio test; mixture models; number of components

1. Introduction

Mixtures of populations are a modelling tool widely used in applications, and the literature on the subject is vast. For finite mixtures, the first task is to choose the number of components in the mixture. This problem is very important for many applications (e.g., choosing a genetic model for a quantitative trait). It also provides one of several methods for choosing the number of clusters in a clustering procedure.

A number of estimation and testing procedures have been proposed for this purpose; see Titterington *et al.* (1985), Lindsay (1995), McLachlan and Peel (2000), James *et al.* (2001), Gassiat (2002) and the references therein. Asymptotic optimality of the log-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. Simulation studies show that the LRT performs well in various situations (see Goffinet *et al.* 1992). However, the asymptotic distribution and local power of the test must be evaluated to compare with other known tests. In this paper, we focus on the asymptotic properties of the LRT for testing whether

1350–7265

independent and identically distributed (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 $p_0 < p$.

Gassiat (2002) gives a rather weak assumption under which the asymptotic distribution of the LRT statistic is derived in the general situation of testing a small model nested in a larger one. This result holds under the null hypothesis as well as under contiguous hypotheses. In Section 2 we explain what remains to be proven in order to apply 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, and we prove the results for the number of components in a mixture of populations in a parametric set with a possibly unknown nuisance parameter. In this way, we obtain more general results than previously derived in the case where the parameter set is bounded. Specifically,

- they apply to general sets of parametric families with unknown nuisance parameter;
- the asymptotic distribution under contiguous alternatives is considered.

We also recover known results for mixtures of one or two populations under weaker assumptions, as well as known results concerning particular parametric families such as Gaussian or binomial distributions; see Ghosh and Sen (1985), Chernoff and Lander (1995), Dacunha-Castelle and Gassiat (1997, 1999), Lemdani and Pons (1997, 1999), Garel (2001), Chen and Chen (2001) and Mosler and Seidel (2001).

In Sections 3 and 4 we study what happens when the set of parameters becomes increasingly large. For simplicity we restrict our attention to the simplest model: the contamination model for families 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. However, 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 (1985) first noted for normal mixtures. Here, we prove that under some extreme circumstances the LRT can have less local power than moment tests or goodness-of-fit tests. 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:

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

We prove the following:

- For general parametric sets $\mathcal{F}$, $\mathbb{T} = [-T, T]$ and T large enough, under contiguous alternatives, the LRT for (1.1) has asymptotic local power close to the asymptotic level under some smoothness assumptions; see Theorem 4. A set of assumptions is given for which Theorem 4 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 Gaussian density with mean t and variance 1, we have 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 5 that the asymptotic local power under contiguous alternatives is equal to the asymptotic level. This result is related to some results of Hall and Stewart (2004) – our hypotheses are weaker, but we do not determine the exact separation speed.

Proofs for most of the results in Sections 3 and 4 are detailed in Section 5.

Our opinion is that the main consequence of our results for large or unbounded parameter sets is that the theoretical asymptotic study of the LRT for mixtures in the compact case seems to be the more relevant case. Concerning practical applications:

- It is well known that the convergence to the asymptotic distribution is very slow for mixtures of populations in general. For example, for the very simple skewness test, Boistard (2003) showed that $n = 10^3$ observations are needed to attain the asymptotic distribution.
- For maximum likelihood estimates and tests, the problem of the speed of convergence to the asymptotic distribution is very difficult to address because in practice maximum likelihood estimates are computed through iterative algorithms and are only approximate – the most famous being the EM algorithm and its variants. All of 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 due to Seidel *et al.* (2002), that the distribution of the LRT depends heavily on these tuning constants.
- Recently, some results and software have become available to compute the distribution of the maximum of Gaussian processes; see Garel (2001), Delmas (2003) and Mecadier (2005a). In particular, these results show that, if the means are contained in some 'non-huge' set, the asymptotic local power of the LRT under contiguous alternatives is generally better than that of moment tests or of goodness-of-fit tests. Nevertheless, the LRT is not uniformly most powerful.

The only practical consequence of our result is the following: if we have a very large data set and we suspect a possible contamination with very small probability but with large parameter, then it is better to use a moment test or a goodness-of-fit test than an LRT.

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) enables us to find the asymptotic distribution of the LRT for testing a particular model nested 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, these sets are parameterized by a finite-dimensional parameter. It remains to:

- prove that the assumption on the bracket entropy holds;
- identify the isonormal process with a Gaussian field with real parameters, and reduce the asymptotic formula by clever computations.

We first recall the general result of Gassiat (2002) and detail its application to the contamination mixture model. We then state the result for two populations with possibly unknown nuisance parameter and for contamination mixtures with unknown nuisance parameters.

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 $X_1, \ldots, X_n$, and $\mathcal{M}_0 \subset \mathcal{M}$ 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. Throughout the paper:

- λ_n will denote the LRT statistics defined by $\lambda_n = \sup_{g\in\mathcal{M}} \ell_n(g) - \sup_{g\in\mathcal{M}_0} \ell_n(g)$;
- $\|\cdot\|_2$ will denote the norm in $L^2(g_0 \cdot \nu)$.
- g_0 will denote a density in $\mathcal{M}_0$ that is the true (unknown) density of the observations.

When studying $\ell_n(g) - \ell_n(g_0)$, functions of the form $(g - g_0)/g_0$ appear naturally. Define two subsets of the unit sphere in $L^2(g_0 \cdot \nu)$ of such functions when normalized:

$$\mathcal{S} = \left\{ \frac{(g-g_0)/g_0}{\|(g-g_0)/g_0\|_2}, \ g \in \mathcal{M}\backslash\{g_0\} \right\} \quad \text{and} \quad \mathcal{S}_0 = \left\{ \frac{(g-g_0)/g_0}{\|(g-g_0)/g_0\|_2}, \ g \in \mathcal{M}_0\backslash\{g_0\} \right\}.$$

A bracket $[L, U]$ of length ϵ is the set of functions b such that $L \leqslant b \leqslant U$, where L and U are functions in $L^2(g_0 \cdot \nu)$ such that $\|U - L\|_2 \leqslant \epsilon$. Define $H_{[\,],2}(\mathcal{S}, \epsilon)$ to be 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 assumption needed is:

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

This assumption implies, in particular, that $\mathcal{S}$ is Donsker and that its closure is compact. As already stated, when $\mathcal{M}$ is parameterized, $\mathcal{S}$ is also parameterized and smoothness properties will allow us to verify (2). But, in general, the parameterization will not be continuous throughout $\mathcal{S}$. The delicate point may be that one needs to find all possible limit points, in $L^2(g_0 \cdot \nu)$, of sequences $((g_n - g_0)/\ g_0)/\|(g_n - g_0)/g_0\|_2$ when $\|(g_n - g_0)/g_0\|_2$ tends to 0. The set $\mathcal{D}$ ($\mathcal{D}_0$) of limit points of sequences $(g_n - g_0)/g_0)/\|(g_n - g_0)/g_0\|_2$ where $\|(g_n - g_0)/g_0\|_2$ tends to 0, $g_n \in \mathcal{M}\backslash\{g_0\}$ ($g_n \in \mathcal{M}_0\backslash\{g_0\}$), will be parameterized in such a way that Lipschitz properties can be used on subsets.

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 is 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 $\mathcal{D} = \{d_{\mathbf{u}}, \mathbf{u} \in \mathbb{U}\}\mathcal{D}_0 = \{d_{(\mathbf{v},\mathbf{0})}, \mathbf{v} \in \mathbb{U}_0\}$. Define the covariance function $r(\cdot, \cdot)$ on $\mathbb{U} \times \mathbb{U}$ by

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

Then, under (2), 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

$$\sup_{\mathbf{u}\in\mathbb{U}} (\max\{Z(\mathbf{u}), 0\})^2 - \sup_{\mathbf{v}\in\mathbb{U}_0} (\max\{Z(\mathbf{v}, \mathbf{0}), 0\})^2, \tag{3}$$

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 $X_1, \ldots, X_n$ under the null hypothesis. In the particular case where $\mathcal{M}_0$ 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 $\sup_{\mathbf{u}\in\mathbb{U}}(\max\{Z(\mathbf{u}), 0\})^2$.

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 (3) are almost surely finite. In general, $r(\cdot, \cdot)$ is continuous almost everywhere.

It is also proven in Gassiat (2002) that if the densities g_n in $\mathcal{M}\backslash\mathcal{M}_0$ are such that $((g_n - g_0)/g_0)/\|(g_n - g_0)/g_0\|_2$ converges to some $d_{\mathbf{u}_0}$ with $\sqrt{n}\|(g_n - g_0)/g_0\|_2$ 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

$$\sup_{\mathbf{u}\in\mathbb{U}} (\max\{Z(\mathbf{u}) + c \cdot r(\mathbf{u}, \mathbf{u}_0), 0\})^2 - \sup_{\mathbf{v}\in\mathbb{U}_0} (\max\{Z(\mathbf{v}, \mathbf{0}) + c \cdot r((\mathbf{v}, \mathbf{0}), \mathbf{u}_0), 0\})^2. \tag{4}$$

In general, (3) and (4) reduce to the square of only one supremum, due to the particular structure of the Gaussian process.

2.1. Contamination mixture

We consider here the contamination mixture model where the 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,

$$\mathcal{M}_0 = \{f_{\mathbf{0}}\}, \qquad \mathcal{M} = \{g_{\pi,\mathbf{t}} = (1-\pi)f_{\mathbf{0}} + \pi f_{\mathbf{t}}, 0 \leqslant \pi \leqslant 1, \mathbf{t} \in \mathbb{T}\},$$
$$\mathcal{S} = \left\{ d_{\mathbf{t}} = \frac{(f_{\mathbf{t}} - f_{\mathbf{0}})/f_{\mathbf{0}}}{\|(f_{\mathbf{t}} - f_{\mathbf{0}})/f_{\mathbf{0}}\|_2}, \mathbf{t} \in \mathbb{T} \right\}.$$

We shall use the following assumptions (CM) ensuring smoothness and some non-degeneracy:

- $f_{\mathbf{t}} = f_{\mathbf{0}}$ 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 $\mathbf{t} \in \mathbb{T}$.

- There exists $\eta > 0$ such that, for all $\mathbf{t} \in \mathbb{T}$ with $\|\mathbf{t}\| \leqslant \eta$, if $\tau \in \mathbb{R}^k$ is such that $\sum_{i=1}^{k} \tau_i \partial f_{\mathbf{t}} / \partial t_i = 0$ ν-a.e. then $\tau = \mathbf{0}$.
- There exists a positive real η and a function $B \in L^2(f_{\mathbf{0}} \cdot \nu)$ that upper-bounds all the following functions:

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

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

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

$$r(\mathbf{s}, \mathbf{t}) = \int d_{\mathbf{s}} d_{\mathbf{t}} f_{\mathbf{0}} \, \mathrm{d}\nu. \tag{5}$$

Notice that, in 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

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

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

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

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

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

where $Z(\boldsymbol{t})$ is the Gaussian field with covariance r defined by (5) *and*

$$\mu(\mathbf{t}) = \begin{cases} c \cdot r(\mathbf{t}, \mathbf{t}_0) & \text{if } \mathbf{t}_n \to \mathbf{t}_0 \neq 0 \\ c \cdot \bar{r}(\mathbf{t}, \tau) & \text{if } \|\mathbf{t}_n\| \to 0 \text{ and } \mathbf{t}_n/\|\mathbf{t}_n\| \to \tau. \end{cases} \tag{6}$$

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 $\mathbf{t}$ to go $\mathbf{0}$ radially in two opposite directions and using covariance properties in the neighbourhood of $\mathbf{0}$, we see that almost surely $\sup_{\mathbf{t}\in\mathbb{T}}(Z(\mathbf{t}) + m(\mathbf{t})) > 0$, which justifies the equalities in the preceding theorem.

A detailed proof of this theorem may be found in Mercadier (2005b). The theorem applies, for instance, to translation mixtures, Gaussian mixtures, binomial mixtures (with a result equivalent to that of Chernoff and Lander 1995) and mixtures in exponential families (see Mercadier 2005b).

2.2. Two populations against a single one

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

$$\mathcal{M}_0 = \{f_{\mathbf{t}}, \mathbf{t} \in \mathbb{T}\}, \qquad \mathcal{M} = \{g_{\pi,\mathbf{t}_1,\mathbf{t}_2} = (1-\pi)f_{\mathbf{t}_1} + \pi f_{\mathbf{t}_2}, 0 \leqslant \pi \leqslant 1, \mathbf{t}_1 \in \mathbb{T}, \mathbf{t}_2 \in \mathbb{T}\}.$$

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 the following assumptions (TP), ensuring smoothness and some non-degeneracy:

- $(1-\pi)f_{\mathbf{t}_1} + \pi f_{\mathbf{t}_2} = f_{\mathbf{0}}$ ν-a.e. if and only if *either* $\pi = 0$ 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 $\mathbf{t} \in \mathbb{T}$.
- For all $\tau \in \mathbb{R}^k$, for all $\mathbf{t} \in \mathbb{T}$, for all $\mathbf{s} \in \mathbb{T}$, for all $\rho \geqslant 0$, $\rho(f_{\mathbf{s}} - f_{\mathbf{0}}) + \sum_{i=1}^k \tau_i \partial f_{\mathbf{t}}/\partial t_i = 0$ ν-a.e. if and only if $\rho\mathbf{s} = \mathbf{0}$ and $\tau = \mathbf{0}$; and there exists $\eta = 0$ such that, for all $\mathbf{t} \in \mathbb{T}$ with $\|\mathbf{t}\| \leqslant \eta$ if $\tau \in \mathbb{R}^k$ is such that $\sum_{i,j=1}^k \tau_i\tau_j \partial^2 f_{\mathbf{t}}/\partial t_i \partial t_j = 0$ ν-a.e. then $\tau = \mathbf{0}$.
- There exists a function $B \in L^2(f_{\mathbf{0}} \cdot \nu)$ that upper-bounds all of the following functions:

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

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

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

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

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 centred Gaussian variable with variance Σ having entries

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

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

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

Then $\mathcal{S} \subset \mathcal{D}$, $\mathcal{S}_0 \subset \mathcal{D}_0$, $\mathcal{D} = \{d_{\mathbf{t},a,\tau};\ \mathbf{t} \in \mathbb{T}\backslash\{\mathbf{0}\},\ \tau \in \mathbb{R}^k,\ a \geqslant 0,\ a^2 + \tau^{\mathrm{T}}\Sigma\tau + 2a\tau^{\mathrm{T}}C(\mathbf{t}) = 1\}$, with

$$d_{\mathbf{t},a,\tau} = \frac{a(f_{\mathbf{t}} - f_{\mathbf{0}})/f_{\mathbf{0}} + \sum_{i=1}^{k}\tau_i(1/f_{\mathbf{0}})\partial f_{\mathbf{0}}/\partial t_i}{\|a(f_{\mathbf{t}} - f_{\mathbf{0}})/f_{\mathbf{0}} + \sum_{i=1}^{k}\tau_i(1/f_{\mathbf{0}})\partial f_{\mathbf{0}}/\partial t_i\|_2}$$

and $\mathcal{D}_0 = \{d_{\mathbf{0},0,\tau},\ \tau^{\mathrm{T}}\Sigma\tau| = 1\}$. Then $2\lambda_n$ converges under $(f_{\mathbf{0}}\ \ldots,\ \nu)^{\otimes n}$ in distribution to

$$\left(\sup_{\substack{a\geqslant 0, t\in\mathbb{T}, \tau\in\mathbb{R}^k \\ a^2+\tau^{\mathrm{T}}\Sigma\tau+2a\tau^{\mathrm{T}}C(\mathbf{t})=1}} (aZ(\mathbf{t}) + \langle\tau,\ W\rangle)\right)^2 - \left(\sup_{\tau^{\mathrm{T}}\Sigma\tau=1}\langle\tau,\ W\rangle\right)^2. \tag{7}$$

Computation shows that this equals

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

where $N = N(\mathbf{t}) = (\Sigma - C(\mathbf{t})C(\mathbf{t})^{\mathrm{T}})^{-1}$.

Theorem 2. *Assume* (TP)*, let π_n, $\mathbf{t}_1^n$ and $\mathbf{t}_2^n$ be sequences such that*

$$\frac{((1-\pi_n)f_{\mathbf{t}_1^n} + \pi_n f_{\mathbf{t}_2^n} - f_{\mathbf{0}})/f_{\mathbf{0}}}{\|((1-\pi_n)f_{\mathbf{t}_1^n} + \pi_n f_{\mathbf{t}_2^n} - f_{\mathbf{0}})/f_{\mathbf{0}}\|_2}$$

tends to some $d_{\mathbf{t}_0,a_0,\tau_0}$ in $\mathcal{D}$, with $\lim_{n\to+\infty}\sqrt{n}\|((1-\pi_n)f_{\mathbf{t}_1^n} + \pi_n f_{\mathbf{t}_2^n} - f_{\mathbf{0}})/f_{\mathbf{0}}\|_2 = c$ for some positive constant c. Then, $(f_{\mathbf{0}}\cdot\nu)^{\otimes n}$ and $[((1-\pi_n)f_{\mathbf{t}_1^n} + \pi_n f_{\mathbf{t}_2^n})\cdot\nu]^{\otimes n}$ are mutually contiguous, and $2\lambda_n$ converges under $(f_{\mathbf{0}}\cdot\nu)^{\otimes n}$ in distribution to

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

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

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

where $Z(\mathbf{t})$, $C(\mathbf{t})$, $N(\mathbf{t})$ W and Σ are defined above, and if $\mathbf{t}_0 = 0$ then $a_0 = 0$.

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

We refer to Mercadier (2005b) for a detailed proof and the description of applications to particular models.

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 us to reduce the possible nuisance parameters in the definition of the set $\mathcal{S}$ to be in a neighbourhood of the true unknown 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

$$\mathcal{M}_0 = \{f_{\mathbf{0},\alpha}, \alpha \in \mathbb{A}\}, \qquad \mathcal{M} = \{g_{\pi,\mathbf{t},\alpha} = (1-\pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha}, 0 \leqslant \pi \leqslant 1, \mathbf{t} \in \mathbb{T}, \alpha \in \mathbb{A}\}.$$

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 the following assumptions (CMN), ensuring smoothness and some non-degeneracy:

- $(1-\pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha} = f_{\mathbf{0},\alpha_0}$ ν-a.e. if and only if $\alpha = \alpha_0$ and either $\pi = 0$ or $\mathbf{t} = \mathbf{0}$.
- $(\mathbf{t}, \alpha) \to f_{\mathbf{t},\alpha}$ is twice continuously differentiable ν-a.e. at any $(\mathbf{t}, \alpha) \in \mathbb{T} \times \mathbb{A}$.
- There exists $\eta > 0$ such that, for all $\delta \in \mathbb{R}^h$, for all $\mathbf{t} \in \mathbb{T}$, for all $\alpha \in \mathbb{A}$ with $\|\alpha - \alpha_0\| \leqslant \eta$, for all $\rho \geqslant 0$, $\rho(f_{\mathbf{t},\alpha_0} - f_{\mathbf{0},\alpha_0}) + \sum_{i=1}^h \delta_i \partial f_{\mathbf{0},\alpha}/\partial \alpha_i = 0$ ν-a.e. if and only if $\rho\mathbf{t} = \mathbf{0}$ and $\delta = \mathbf{0}$, and for all $\tau \in \mathbb{R}^k$, $\|\mathbf{t}\| \leqslant \eta$, $\|\alpha - \alpha_0\| \leqslant \eta$, $\sum_{i=1}^k \tau_i \partial f_{\mathbf{t},\alpha_0}/\partial t_i + \sum_{i=1}^h \delta_i \partial f_{\mathbf{0},\alpha}/\partial \alpha_i = 0$ 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 of the following functions:

$$\frac{f_{\mathbf{t},\alpha}}{f_{\mathbf{0},\alpha_0}}, \quad \frac{1}{f_{\mathbf{0},\alpha_0}}\left|\frac{\partial f_{\mathbf{t},\alpha}}{\partial t_i}\right|, i = 1, \ldots, k, \quad \frac{1}{f_{\mathbf{0},\alpha_0}}\left|\frac{\partial f_{\mathbf{t},\alpha}}{\partial \alpha_i}\right|, i = 1, \ldots, h,$$
$$(\mathbf{t}, \alpha) \in \mathbb{T} \times \mathbb{A}, \|\alpha - \alpha_0\| \leqslant \eta,$$
$$\frac{1}{f_{\mathbf{0},\alpha_0}}\left|\frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial t_i \partial t_j}\right|, i, j = 1, \ldots, k, \quad \frac{1}{f_{\mathbf{0},\alpha_0}}\left|\frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial t_i \partial \alpha_j}\right|, i = 1, \ldots, k, j = 1, \ldots, h,$$
$$\frac{1}{f_{\mathbf{0},\alpha_0}}\left|\frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial \alpha_i \partial \alpha_j}\right|, i, j = 1, \ldots, h, \qquad (\mathbf{t}, \alpha) \in \mathbb{T} \times \mathbb{A}, \|\alpha - \alpha_0\| \leqslant \eta, \|\mathbf{t}\| \leqslant \eta.$$

Then, since the maximum likelihood estimator of parameter α is consistent, one only needs to verify assumption (2) for

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

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

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

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

$$\mathcal{D}_0 = \left\{ d_{\mathbf{0},0,\delta,\mathbf{0}}, \delta \in \mathbb{R}^h, \|\delta\| = 1 \right\},$$
$$\mathcal{D} = \left\{ d_{\mathbf{t},\rho,\delta,\tau}, \mathbf{t} \in \mathbb{T}, \rho \geqslant 0, \delta \in \mathbb{R}^h, \tau \in \mathbb{R}^k, \rho^2 + \|\delta\|^2 + \|\tau\|^2 = 1 \right\}.$$

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

Again, let

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

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 that of Section 2.1 if we set $f_{\mathbf{0}} = f_{\mathbf{0},\alpha_0}$. Furthermore, let W, Σ and $C(\mathbf{t})$ be the same as in Section 2.1, replacing $\partial f_{\mathbf{0}}/\partial t_i$ by $\partial f_{\mathbf{0},\alpha_0}/\partial t_i$.

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

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

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

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

Also let S be the covariance matrix of W and V, with entries:

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

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

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

Theorem 3. *Assume* (CMN)*, and let* π_n*,* $\mathbf{t}_n$ *and* α_n *be sequences such that*

$$\frac{((1-\pi_n)f_{\mathbf{0},\alpha_n}+\pi_n f_{\mathbf{t}_n,\alpha_n}-f_{\mathbf{0},\alpha_0})/f_{\mathbf{0},\alpha_0}}{\|((1-\pi_n)f_{\mathbf{0},\alpha_n}+\pi_n f_{\mathbf{t}_n,\alpha_n}-f_{\mathbf{0},\alpha_0})/f_{\mathbf{0},\alpha_0}\|_2}$$

tends to some $d_{\mathbf{t}_0,\rho_0,\delta_0,\tau_0}$ *in* $\mathcal{D}$, *with* $\lim_{n\to+\infty}\sqrt{n}\|((1-\pi_n)f_{\mathbf{0},\alpha_n}+\pi_n f_{\mathbf{t}_n,\alpha_n}-f_{\mathbf{0},\alpha_0})/f_{\mathbf{0},\alpha_0}\|_2 = c$ *for some positive constant c. Then, in the above notation,* $(f_{\mathbf{0},\alpha_0}\cdot\nu)^{\otimes n}$ *and* $[((1-\pi_n)f_{\mathbf{0},\alpha_n}+\pi_n f_{\mathbf{t}_n,\alpha_n})\cdot\nu]^{\otimes n}$ *are mutually contiguous, and* $2\lambda_n$ *converges under* $(f_{\mathbf{0}}\cdot\nu)^{\otimes n}$ *in distribution to*

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

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

$$\sup_{\mathbf{t}\in\mathbb{T}}\Bigg[\Bigg(Z(\mathbf{t})+c\rho_0 r(\mathbf{t},\mathbf{t}_0)+cC(\mathbf{t})^{\mathrm{T}}\tau_0+cG(\mathbf{t})^{\mathrm{T}}\delta_0 -\frac{U(\mathbf{t})^{\mathrm{T}}N(\mathbf{t})}{1+U(\mathbf{t})^{\mathrm{T}}N(\mathbf{t})U(\mathbf{t})}\begin{pmatrix}W+c\Sigma\tau_0+c\rho_0C(\mathbf{t}_0)\\ V+c\Gamma\delta_0+c\rho_0G(\mathbf{t}_0)\end{pmatrix}\Bigg)^2\left(1+U(\mathbf{t})^{\mathrm{T}}N(\mathbf{t})U(\mathbf{t})\right)\Bigg]$$
$$+\begin{pmatrix}W+c\Sigma\tau_0+c\rho_0C(\mathbf{t}_0)\\ V+c\Gamma\delta_0+c\rho_0G(\mathbf{t}_0\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\Sigma & S^{\mathrm{T}}\\ S & \Gamma\end{pmatrix}^{-1}\begin{pmatrix}W+c\Sigma\tau_0+c\rho_0C(\mathbf{t}_0)\\ V+c\Gamma\delta_0+c\rho_0G(\mathbf{t}_0\end{pmatrix} -(V+c\Gamma\delta_0+c\rho_0G(\mathbf{t}_0))^{\mathrm{T}}\Gamma^{-1}(V+c\Gamma\delta_0+c\rho_0G(\mathbf{t}_0)),$$

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

For instance, it is easy to apply Theorem 3 to translation mixtures with unknown scale parameter or Gaussian mixtures with unknown variance, as illustrated in Mercadier (2005b).

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

As already stated 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. In this case, the LRT happens to be more powerful than moment tests, as shown in Delmas (2003). Nevertheless:

- the distribution is not independent of 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 the case that results would be nicer if one were able to get rid of the compactness assumption. This section and the next answer tell us that we cannot do so, 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.

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

(K1) $\lim_{n\to+\infty} \sqrt{n}\pi_n \|(f_{t_n} - f_0)/f_0\|_2 = c$ for some positive c;
(K2) 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 $\mathbb{P}_{\pi_n,t_n} = (g_{\pi_n,t_n} \cdot \nu)^{\otimes n}$ and $\mathbb{P}_0 = (f_0 \cdot \nu)^{\otimes n}$. To evaluate the asymptotic local 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 centred Gaussian process defined in Section 2.1 with covariance given by (5). 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 no effect on the extreme behaviour of the process Z here. We shall use Azaïs and Mercadier (2003) to derive the asymptotic distribution of suprema of Gaussian processes. Hence, let

$$M(a, b) = \sup_{t\in(a,b)} (Z(t) + m(t)). \tag{8}$$

Because 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 (2.5) under contiguous alternatives), we wish to characterize asymptotic behaviours of $M(-T, T)$ as $T \to +\infty$.

We therefore introduce further notation and assumptions. Write $r_{ij}(s, t)$ instead of $\partial^{i+j} r(s, t)/\partial^i s \partial^j t$ and define

$$R(t) = \int_0^t \sqrt{r_{11}(s, s)}\,\mathrm{d}s, \qquad a_t = \sqrt{2\log \circ R(t)}, \qquad b_t = a_t - \frac{\log(\pi)}{a_t}. \tag{9}$$

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:

(G1) For all $t \in \mathbb{R}$, $r_{11}(t, t) > 0$ and $\lim_{t\to+\infty} R(t) = +\infty$.
(G2) $r(s, t)\log|R(s) - R(t)| \to 0$ as $|R(s) - R(t)| \to +\infty$.
(G3) For all $\varepsilon > 0$, $\sup_{|R(s)-R(t)|>\varepsilon} |r(s, t)| < 1$.
(G4) r is four times continuously differentiable and $s \to r_{11}(s, s)$ three times continuously differentiable. For all $\gamma > 0$, r^Y_{01} and r^Y_{04} are bounded on $\{(s, t) \in \mathbb{R}^2, |s| > \gamma \text{ and } |t| > \gamma\}$.

(G5) $\sqrt{\log \circ R(t)}\mu(t) \longrightarrow_{t\to+\infty} 0.$

Theorem 4. *Assume* (CM)*,* (G)*,* (K1) *and* (K2)*. Define $M(-T, T)$ by* (8) *and a_T, b_T by* (9)*. Then, as T tends to infinity, $a_T(M(-T, T) - b_T)$ tends in distribution to the Gumbel distribution when $m(t) = 0$ as well as when $m(t) = \mu(t)$. In other words, if $c_{T,\alpha,n}$ is a sequence of thresholds of the test satisfying*

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

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

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

This theorem says that for T large enough, asymptotically, the LRT cannot distinguish the null hypothesis from any contiguous alternatives. The theorem is proven in Section 5.

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

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

Let f_0 be a density on $\mathbb{R}$ satisfying the following assumptions (H), where we denote by $f_0^{(i)}$ the derivative of f_0 of order i:

(H1) For all $x \in \mathbb{R}$, $f_0(x) > 0$, f_0 four times continuously differentiable, and for all $i = 1, \ldots, 4$, there exists $K_i > 0$ such that

$$\left| \frac{f_0^{(i)}}{f_0}(x) \right| \leqslant K_i.$$

(H2) For all $x \in \mathbb{R}$,

$$\lim_{t\to+\infty} \frac{f_0(x+t)}{f_0(t)} = \lim_{t\to-\infty} \frac{f_0(x+t)}{f_0(t)} = 1.$$

(H3) There exists $M > 0$, such that, for all x, $t \in \mathbb{R}$ $f_0(x)f_0(t)/f_0(x+t) \leqslant M$.

(H4) There exists $F \in L^2(\lambda)$ such that $\sup_{|t|\geqslant 1} \log|t|\sqrt{f_0(x+t)} \leqslant F(x)$.

(H5) $\lim_{t\to+\infty} \log(t)\sqrt{f_0(t)} = 0$.

Our result, which is proven in Section 5, is now:

Corollary 1. *Assume* (H)*. Then Theorem 4 applies to the translation mixture model.*

Remarks.

- Among the possible choices of the thresholds are: the asymptotic threshold given by Theorem 1, which does not depend on n; and the exact threshold for each n, which can be calculated by a Monte Carlo method.
- Assumptions (H) are essentially conditions on the tail of f_0. (H4) and (H5) are very weak and hold for all usual distributions. But (H1)–(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 when f_0 is the inverse of a polynomial, and in particular for the Cauchy density.
- The proof relies on the verification of assumptions of Theorem 4. In particular, asymptotic behaviours of the covariance r and its derivatives must 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 4 holds for other densities such as the Gaussian and the normalization of $\cosh^{-1}(x)$ in spite of different justifications.

The LRT needs to be compared with other testing procedures such as sample mean or Kolmogorov–Smirnov testing procedures. Write $\mu_i = \int x^i f_0(x) \mathrm{d}\nu(x)$. Without loss of generality, one can assume that $\mu_1 = 0$. If $\mu_2 < +\infty$ applying Le Cam's third lemma, (i.e. 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

$$\gamma = \begin{cases} \dfrac{c}{\|f_0'/f_0\|_2}, & \text{if } t_n \to 0, \\ \dfrac{c t_0}{\|(f_{t_0} - f_0)/f_0\|_2}, & \text{if } t_n \to t_0 \neq 0. \end{cases}$$

Consequently, the asymptotic local 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 with f_0. Let I denote the identity function on [0, 1] and let $\mathbb{U}$ be a Brownian bridge on [0, 1]. Let $\|\cdot\|_\infty$ denote the supremum norm. The natural normalization of $\mathbb{F}_n$ leads to the definition of the Kolmogorov–Smirnov statistic $\mathbb{K}_n = \sqrt{n}\,\|\mathbb{F}_n - F_0\|_\infty$ and the Cramér–von Mises statistic $\mathbb{W}_n^2 = \int_{-\infty}^{+\infty} n[\mathbb{F}_n(x) - F_0(x)]^2 \,\mathrm{d}F_0(x)$. Set, on [0, 1],

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

where t_n is the translation parameter of the alternative. Hence Δ depends on the asymptotic behaviour of t_n. Then $\mathbb{K}_n$ converges in distribution, as n tends to infinity, to $\|\mathbb{U}\|_\infty$ under $\mathbb{P}_0$ and $\|\mathbb{U} + \Delta\|_\infty$ under $\mathbb{P}_{\pi_n, t_n}$, and $\mathbb{W}_n^2$ converges in distribution, as n tends to infinity, to $\int_0^1 \mathbb{U}^2 \,\mathrm{d}I$ under $\mathbb{P}_0$ and $\int_0^1 (\mathbb{U} + \Delta)^2 \,\mathrm{d}I$ 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 local power is greater than the level.

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

Theorem 5. *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 $g_0 = f_0$ and $g_{\pi,t} = (1-\pi)f_0 + \pi f_t$, $0 \leqslant \pi \leqslant 1$, $t \in \mathbb{T}$. 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 0. λ_n is now given by:

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

Then as n tends to infinity, $2\lambda_n - \log\circ\log n + \log(2\pi^2)$ 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

$$\lim_{n\to+\infty} \big(c_{\alpha,n} - \log\circ\log n + \log(2\pi^2)\big) = \frac{1}{2} G_{1-\alpha}, \tag{10}$$

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

$$\lim_{n\to+\infty} \mathbb{P}_0(\lambda_n > c_{\alpha,n}) = \alpha. \tag{11}$$

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

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

The theorem says that, asymptotically, the LRT cannot distinguish the null hypothesis from any contiguous alternative. This is true if we use asymptotic thresholds (equality for all n in (10)) as well as exact thresholds (equality for all n in (11)). 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 local 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 5. 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 the empirical process S_n defined by

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

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

$$\sup_{t\in\mathbb{R}} S_n(t) = \sup_{|t|\in A_{2,n}} S_n(t) + o_{\mathbb{P}_0}(1), \tag{12}$$

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

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

Combined with (12), the last equality becomes

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

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

$$2\lambda_n = \sup_{|t|\in A_{2,n}} S_0(t)^2 + o_{\tilde{\mathbb{P}}_0}(1) \tag{13}$$

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

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

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 another method here. We define the standardized version of S_0,

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

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

$$r(s, t) = \frac{\exp(st) - 1}{\sqrt{\exp(s^2) - 1}\sqrt{\exp(t^2) - 1}}. \tag{14}$$

We have

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

Now the function r satisfies the conditions of Corollary 1 of Azaïs and Mercadier (2003). Consequently we know the exact order of the maximum: $\sup_{|t|\in A_{2,n}}|Y_0(t)| = O_{\tilde{\mathbb{P}}_0}((\log\circ\log n)^{1/2})$. This last equation can also be deduced from standard results on the maximum of stationary Gaussian processes using the process $\tilde{S}_0$ introduced by Bickel and Chernoff (1993).

On the other hand, the maximum of $1-\sqrt{\rho(t,\,t)}$ on $A_{2,n}$ is obtained at α_n. This permits us to write

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

Finally, this approximation allows us to replace S_0 by Y_0 in (4.4) to obtain

$$2\lambda_n = \sup_{|t|\in A_{2,n}} Y_0(t)^2 + o_{\tilde{\mathbb{P}}_0}(1). \tag{15}$$

With the same idea as before, we define

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

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

$$\log\frac{\mathrm{d}\mathbb{P}_{\pi_n,t_n}}{\mathrm{d}\mathbb{P}_0}(X_1,\,\ldots,\,X_n) = C(\gamma,\,t_0)Y_n(t_n) - \frac{C(\gamma,\,t_0)^2}{2} + o_{\mathbb{P}_0}(1) \tag{16}$$

with

$$C(\gamma,\,t_0) = \begin{cases} \gamma, & \text{if } t_0 = 0,\\ \gamma\dfrac{\sqrt{\mathrm{e}^{t_0^2}-1}}{t_0}, & \text{if } t_0 > 0.\end{cases}$$

Since γ can be supposed positive, t_0 is positive. Using formula (39) of Bickel and Chernoff (1993), we can replace Y_n by Y_0 to obtain

$$\log\frac{\mathrm{d}\mathbb{P}_{\pi_n,t_n}}{\mathrm{d}\mathbb{P}_0}(X_1,\,\ldots,\,X_n) = C(\gamma,\,t_0)Y_0(t_n) - \frac{C(\gamma,\,t_0)^2}{2} + o_{\tilde{\mathbb{P}}_0}(1). \tag{17}$$

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

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

Then, having proved Lemma 1, the theorem follows from Le Cam's third lemma. The proof of Lemma 1 relies on a suitably chosen discretization, following ideas in Azaïs and Mercadier (2003), and an application of the normal comparison lemma as refined in Li and Shao (2002). □

5. Proofs

Proof of Theorem 4. Set $u_{T,x} = x/a_T + \tilde{b}_T$ and $M^V(a,\,b) = \sup_{t\in(a,b)} V_t$ for V the unit-speed transformation of $Z+m$. We have:

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

Now, applying Proposition 4 of Azaïs and Mercadier (2003) with $p = 2$, $D_1 = (-R(T), -\sqrt{R(T)})$ and $D_2 = (\sqrt{R(T)}, R(T))$, we obtain

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

Remark that in Azaïs and Mercadier (2003) sizes of intervals are defined as functions of the level; here the opposite is the case. Furthermore, repeated application of Corollary 1 of Azaïs and Mercadier (2003) enables us to state for $\tau = \sqrt{R(T)}$ and $\tau = R(T)$ the convergence of $a_\tau(M^V(0, \tau) - b_\tau)$ and $a_\tau(M^V(-\tau, 0) - b_\tau)$ to the Gumbel. It follows that $M^V(-\sqrt{R(T)}, \sqrt{R(T)})$ is stochastically negligible compared with $M^V(-R(T), R(T))$ and also that $M(0, \sqrt{R(T)})$ $(M(-\sqrt{R(T)}, 0))$ is stochastically negligible compared with $M^V(0, R(T))$ $(M^V(-R(T), 0))$. Taking the foregoing together, we obtain

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

as T tends to infinity, and which becomes

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

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

Let $G(x) = \exp(-\exp(-\mathrm{x}))$ denote the distribution function of the Gumbel. Corollary 1 of Azaïs and Mercadier (2003) yields, as T tends to infinity,

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

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

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

Proof of Corollary 1. The proof relies on the verification of assumptions of Theorem 4, beginning with (CM). Since f_0 is continuous and positive, for any T, $\inf_{t\in[-T,T]} f_0(t) = \delta_T > 0$. Using (H3), for all $t \in [-T, T]$ and $x \in \mathbb{R}$,

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

and using (H1) and (H3),

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

Let us now prove assumptions (G). Set

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

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

$$\frac{f_0^{(i)}(x-t)f_0^{(j)}(x-s)}{f_0(x)} \leqslant K_iK_j\frac{f_0(x-t)f_0(x-s)}{f_0(x)} \leqslant K_iK_jM^2\frac{f_0(x)}{f_0(t)f_0(s)}$$

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 $(s,\ t) \in (\mathbb{R}\backslash\{0\})^2$ with

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

We first prove (G1). We have, for $t \neq 0$,

$$r_{11}(t,\ t) = \frac{\|f_0'(\cdot-t)/f_0(\cdot)\|_2^2\|(f_0(\cdot-t)-f_0(\cdot))/f_0(\cdot)\|_2^2 - (\langle f_0'(\cdot-t)/f_0(\cdot), (f_0(\cdot-t)-f_0(\cdot))/f_0(\cdot)\rangle_2)^2}{\|(f_0(\cdot-t)-f_0(\cdot))/f_0(\cdot)\|_2^4}$$

which is positive by the Cauchy–Schwarz inequality. Now,

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

Indeed, define the functions

$$A(t) = \int \frac{f_0^2(x)}{f_0(x+t)}\,\mathrm{d}\nu(x), \qquad B(t) = \int \frac{f_0'^2(x)}{f_0(x+t)}\,\mathrm{d}\nu(x), \qquad C(t) = \int \frac{f_0(x)f_0'(x)}{f_0(x+t)}\,\mathrm{d}\nu(x).$$

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

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

By virtue of (H1) and (H3), the integrands of Af_0, Bf_0 and Cf_0 are respectively dominated by $Mf_0(x)$, $K_1^2Mf_0(x)$, and $K_1Mf_0(x)$. By application of (H2) and the Lebesgue theorem, we conclude the proof of (G1) using the fact that $A(t)f_0(t)$, $B(t)f_0(t)$, $C(t)f_0(t)$ converge respectively to

$$\int f_0^2(x)\,\mathrm{d}\nu(x), \qquad \int f_0'^2(x)\,\mathrm{d}\nu(x), \qquad \int f_0(x)f_0'(x)\,\mathrm{d}\nu(x).$$

Thus for a positive constant R,

$$R(t) \sim_{t\to+\infty} R\,t. \tag{18}$$

We turn now to (G2). Considering (18), we need to prove that

$$\lim_{|s-t|\to+\infty} r(s,\ t)\log|s-t| = 0. \tag{19}$$

Using (H3),

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

so that using (H2),

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

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

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

Then, using (H3),

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

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 the Lebesgue theorem using (H4) to obtain (19).

To prove (G5), we observe that it is a consequence of (G2) and formula (6) giving $\mu(t)$.

Moving on to (G3), using (18) and $r_{11} > 0$, one only needs to prove that, for any $\varepsilon > 0$,

$$\sup_{|s-t|>\varepsilon} |r(s,\ t)| < 1. \tag{20}$$

First of all, $r(s,\ t)$ is a continuous function of $(s,\ t)$ and $|r(s,\ t)| < 1$ if $s \neq t$ by the 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 the Lebesgue theorem),

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

Since $|c| \geqslant \varepsilon > 0$, this value differs from 1. Hence we obtain a contradiction with assumptions made on sequences s_n and t_n and (20) is true.

Finally we round off the proof of (G4), having already dealt with the first part. We use the same arguments to prove that $s \mapsto r_{11}(s, s)$ is three times continuously differentiable. Now, this last regularity asociated with (18) allows us to reduce our study to that of functions r_{01} and r_{04}.

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

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

Then the Cauchy–Schwarz inequality leads to

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

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

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

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 $\mathbb{R}^2 \backslash D_\delta$.

Using easy but tedious computations and the Cauchy–Schwarz inequality once more, we have

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

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 apply again and permit us to assert that for any $\delta > 0$ the function $(s, t) \mapsto r_{04}(s, t)$ is bounded on $\mathbb{R}^2 \backslash D_\delta$. □

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

From to (15) and (17), 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 discretization step q_n depending on n in a sense which must 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,

$$\left|\mathbb{P}\left(\sup_{t\in A_{2,n}} Y_0(t) - c_n \leqslant x;\ Y_0(t_0) \leqslant y\right) - \mathbb{P}(\sup_{t\in A_{2,n}} Y_0(t) - c_n \leqslant x)\mathbb{P}(Y_0(t_0) \leqslant y)\right|$$

$$\leqslant 2\,\mathbb{P}\left(\sup_{d\in A_{2,n}^{q_n}} Y_0(d) - c_n \leqslant x;\ \sup_{t\in A_{2,n}} Y_0(t) - c_n > x\right) \tag{21}$$

$$+\left|\mathbb{P}\left(\sup_{d\in A_{2,n}^{q_n}} Y_0(d) - c_n \leqslant x;\ Y_0(t_0) \leqslant y\right) - \mathbb{P}\left(\sup_{d\in A_{2,n}^{q_n}} Y_0(d) - c_n \leqslant x\right)\mathbb{P}(Y_0(t_0) \leqslant y)\right|.$$

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

We define the following modification of the function r:

$$\tilde{r}(t_0,\ t) = 0, \qquad \text{when } t \in A_{2,n}^{q_n},\ t \neq t_0,$$

$$\tilde{r}(s,\ t) = r(s,\ t), \qquad \text{for } s,\ t \in A_{2,n}^{q_n}.$$

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 distributions do not change. This proves that $\tilde{r}$ is a covariance function. We define $\xi(t) = \sup_{u,|u-t_0|>t}|r(u,\ t_0)|$. From (14) we have

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

We restrict our attention to ns such that $c_n > 2|x|$ and $\xi(\alpha_n) < \frac{1}{2}$ so that

$$\frac{(x+c_n)^2}{2(1+\xi(\alpha_n))} \geqslant \frac{c_n^2}{12}.$$

The normal comparison lemma (Li and Shao 2002: Theorem 2.1) gives bounds on terms of the type

$$\mathbb{P}(Y_1 \leqslant u_1,\ \ldots,\ Y_n \leqslant u_n) - \mathbb{P}\big(\tilde{Y}_1 \leqslant u_1,\ \ldots,\ \tilde{Y}_n \leqslant u_n\big),$$

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

$$\mathbb{P}(Y_1 \leqslant u_1,\ \ldots,\ Y_n \leqslant u_n) - \mathbb{P}\big(\tilde{Y}_1 \leqslant u_1,\ \ldots,\ \tilde{Y}_n \leqslant u_n\big)$$

$$\leqslant \frac{1}{2\pi}\sum_{1\leqslant i<j\leqslant n}\big(\arcsin(\rho_{ij}) - \arcsin(\tilde{\rho}_{ij})\big)^+ \exp\left(-\frac{u_i^2+u_j^2}{2(1+\bar{\rho}_{ij})}\right), \tag{22}$$

where $z^+ = \max\{z,\ 0\}$, $\bar{\rho}_{ij} = \max\{|\rho_{ij}|,\ |\tilde{\rho}_{ij}|\}$. Let Const. represent a generic positive constant. Since $\arcsin(x) \leqslant x\pi/2$ for $0 \leqslant x \leqslant 1$, applying inequality (22) 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 obtain:

$$\left|\mathbb{P}\left(\sup_{d\in A_{2,n}^{q_n}} Y_0(d)-c_n\leqslant x;\ Y_0(t_0)\leqslant y\right)-\mathbb{P}\left(\sup_{d\in A_{2,n}^{q_n}} Y_0(d)-c_n\leqslant x\right)\mathbb{P}(Y_0(t_0)\leqslant y)\right|$$

$$\leqslant \text{Const.}\sum_{d\in A_{2,n}^{q_n}}|r(d,\ t_0)|\exp\left(-\frac{(x+c_n)^2+y^2}{2(1+|r(d,\ t_0)|)}\right)$$

$$\leqslant \text{Const.}\sum_{d\in A_{2,n}^{q_n}}|r(d,\ t_0)|\exp\left(-\frac{\mathrm{c}_\mathrm{n}^2}{12}\right)$$

$$\leqslant \frac{\text{Const.}}{q_n}\exp\left(-\frac{c_n^2}{12}\right)\int_{\alpha_n-q_n}^{+\infty}\xi(t)\mathrm{d}t=\frac{\text{Const}}{q_n}\exp\left(-\frac{c_n^2}{12}\right),$$

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

To deal with the first term of (21), 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 $\mathbb{R}$,

$$U_z(B)=\#\{t\in B,\ Y_0(t)=z,\ Y_0'(t)>0\}$$

$$U_z^{q_n}(B)=\#\{l\in\mathbb{Z},\ q_n(l-1)\in B,\ q_nl\in B,\ Y_0(q_n(l-1))<z<Y_0(q_nl)\}.$$

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

$$\mathbb{P}\left(\sup_{d\in A_{2,n}^{q_n}} Y_0(d)-c_n\leqslant x;\ \sup_{t\in A_{2,n}} Y_0(t)-c_n>x\right)$$

$$\leqslant \mathbb{P}(Y_0(\alpha_n)>x+c_n)+\mathbb{P}(Y_0(\alpha_n)\leqslant x+c_n,\ U_{x+c_n}(A_{2,n})\geqslant 1,\ U_{x+c_n}^{q_n}(A_{2,n})=0)$$

$$\leqslant \overline{\Phi}(x+c_n)+\mathbb{E}(U_{x+c_n}(A_{2,n})-U_{x+c_n}^{q_n}(A_{2,n})),$$

where the last upper bound is a result of the Markov inequality. The first term above tends trivially to zero, as for the second if we set $q_n=(\log\circ\log n)^{-\theta}$ with $\theta>\frac{1}{2}$, Condition (U7) of Lemma 2 of Azaïs and Mercadier (2003) is met. It is easy to check that because $\mathbb{E}(U_{x+c_n}(A_{2,n}))$ is bounded we are in the condition of application of that lemma and

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

□

Acknowledgements

The authors would like to thank Eric Gilleland for having corrected the writing and English construction of this paper.

References

Azaïs, J.-M. and Mercadier, C. (2003) Asymptotic Poisson character of extremes in non-stationary Gaussian models, *Extremes*, **6**, 301–318.

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

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.*, **52**, 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. Plann. Inference*, **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. *Sankhyā*, **65**(3), 1–19.

Garel, B. (2001) Likelihood ratio test for univariate Gaussian mixture. *J. Statist. Plann. Inference*, **96**(2), 325–350.

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

Ghosh, J. and Sen, P. (1985) On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In L. LeCam and R.A. Olshen (eds), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, pp. 789–806. Monterey, CA: Wadsworth.

Goffinet, B., Loisel, P. and Laurent, B. (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 L. LeCam and R.A. Olshen (eds), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, pp. 807–810. Monterey, CA: Wadsworth.

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. *Statist. Probab. Lett.*, **33**(1), 15–22.

Lemdani, M. and Pons, O. (1999) Likelihood ratio tests in contamination models. *Bernoulli*, **5**, 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*. Hayward, CA: 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*. New York: Wiley.
Mercadier, C. (2005a) Numerical bounds for the distribution of one- and two-parameter processes. Preprint, Université Toulouse III.
Mercadier, C. (2005b) Extrema de processus stochastiques. Propriétés asymptotiques de tests d'hypothèses. Doctoral thesis, Université Paul Sabatier, Toulouse III.
Mosler, K. and Seidel, W. (2001) Testing for homogeneity in an exponential mixture model. *Austral. N.Z. J. Statist.*, **43**(3), 231–247.
Seidel, W., Mosler, K., and Alker, M. (2000) A cautionary note on likelihood ratio tests in mixture models, *Ann. Inst. Statist. Math.*, **52**, 481–487.
Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics*. New York: Wiley.
Titterington, D.M., Smith, A.F.M. and Makov, U.E. (1985) *Statistical Analysis of Finite Mixture Distributions*. Chichester: Wiley.
van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Received September 2004 and revised December 2005