



HAL
open science

The likelihood ratio test for general mixture models with or without structural parameter

Jean-Marc Azaïs, Élisabeth Gassiat, Cécile Mercadier

► **To cite this version:**

Jean-Marc Azaïs, Élisabeth Gassiat, Cécile Mercadier. The likelihood ratio test for general mixture models with or without structural parameter. *ESAIM: Probability and Statistics*, 2009, 13, pp.301-327. 10.1051/ps:2008010 . hal-00504651v2

HAL Id: hal-00504651

<https://hal.science/hal-00504651v2>

Submitted on 21 Jul 2010

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

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

THE LIKELIHOOD RATIO TEST FOR GENERAL MIXTURE MODELS WITH OR WITHOUT STRUCTURAL PARAMETER

JEAN-MARC AZAÏS¹, ELISABETH GASSIAT² AND CÉCILE MERCADIER³

Abstract. This paper deals with the likelihood ratio test (LRT) for testing hypotheses on the mixing measure in mixture models with or without structural parameter. The main result gives the asymptotic distribution of the LRT statistics under some conditions that are proved to be almost necessary. A detailed solution is given for two testing problems: the test of a single distribution against any mixture, with application to Gaussian, Poisson and binomial distributions; the test of the number of populations in a finite mixture with or without structural parameter.

Résumé. Nous étudions le test du rapport de vraisemblance (TRV) pour des hypothèses sur la mesure mélangeante dans un mélange en présence éventuelle d'un paramètre structurel, et ce dans toutes les situations possibles. Le résultat principal donne la distribution asymptotique du TRV sous des hypothèses qui ne sont pas loin d'être nécessaires. Nous donnons une solution détaillée pour le test d'une simple distribution contre un mélange avec application aux lois Gaussiennes, Poisson et binomiales, ainsi que pour le test du nombre de populations dans un mélange fini avec un paramètre structurel.

1991 Mathematics Subject Classification. 62F05, 62F12, 62H10, 62H30.

The dates will be set by the publisher.

1. INTRODUCTION

1.1. Motivations and aims

Latent variables are used in location-scale problems, in various regression settings with covariate measurement error, in biased sampling models or for modelling some censoring mechanisms. We refer to [Bickel et al., 1993] for the description of several latent variable models. An other example is that of mixtures, see [Lindsay, 1995], [McLachlan and Peel, 2000], [Titterington et al., 1985]. One observes a sample X_1, \dots, X_n , that is independent and identically distributed random variables with a density of the type

$$p_{\gamma, \eta}(x) = \int p_{\gamma}(x|z) d\eta(z). \quad (1)$$

Keywords and phrases: Likelihood ratio test, mixture models, number of components, local power, contiguity

¹ Institut de Mathématiques de Toulouse, UMR 5219, Université Paul Sabatier, 31062 Toulouse Cedex 9, France.

² Équipe Probabilités, Statistique et Modélisation, UMR CNRS 8628, Université Paris-Sud, Bâtiment 425, Université de Paris-Sud, 91405 Orsay Cedex, France.

³ Université de Lyon, Université Lyon 1, CNRS, UMR 5208 Institut Camille Jordan, Bâtiment du Doyen Jean Braconnier, 43, boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France

Here $x \rightarrow p_\gamma(x|z)$ is a family of probability densities with respect to some measure μ on a measurable space $(\mathcal{X}, \mathcal{A})$, γ is called the *structural parameter*. The latent variable Z has distribution η on a measurable space $(\mathcal{Z}, \mathcal{C})$, η is called the *mixing distribution*. In case η has q supporting points z_1, \dots, z_q with weights π_1, \dots, π_q and if in addition z and γ vary in finite dimension spaces, (1) reduces to a parametric family

$$p_\theta(x) = \sum_{i=1}^q \pi_i p_\gamma(x|z_i), \quad (2)$$

in which the parameter is

$$\theta = (\gamma, \pi_1, \dots, \pi_q, z_1, \dots, z_q). \quad (3)$$

When all supporting points z_i are distinct and all weights π_i are non null, q is the number of populations in the mixture.

This paper focuses on testing hypotheses on the mixing distribution using the likelihood ratio test (LRT for short). Let $\mathcal{G}_1 \subset \mathcal{G}_2$ be two sets of probability distributions on \mathcal{Z} , and consider the problem of testing

$$H_0 : \text{“}\eta \in \mathcal{G}_1\text{” against } H_1 : \text{“}\eta \in \mathcal{G}_2 \setminus \mathcal{G}_1\text{”}. \quad (4)$$

In case \mathcal{G}_1 is the set of Dirac masses and \mathcal{G}_2 the set of all probability distributions on \mathcal{Z} , the problem is that of testing whether there is a single population against a mixture of any kind.

In case \mathcal{G}_i is the set of finite measures with q_i supporting points, $q_1 < q_2$, the problem is that of testing whether the number of populations is less or equal to q_1 or at least $q_1 + 1$ but not more than q_2 . When $q_1 = 1$, the question is that of “homogeneity” against “heterogeneity”.

To set the threshold in the LRT at a prescribed level, one has to know the distribution of the likelihood ratio statistic when the true mixing distribution η_0 lies in \mathcal{G}_1 .

In classical parametric statistics, twice the log-likelihood ratio has a chi-square asymptotic distribution or a convex combination of chi-square distributions. Such a result does not apply here, due to lack of identifiability of the parameters in \mathcal{G}_2 and degeneracy of the Fisher information of the model. The challenging question of the asymptotic distribution of the likelihood ratio has received much interest in the past decade, after that [Ghosh and Sen, 1985] raised the question, see [Chernoff and Lander, 1995], [Dacunha-Castelle and Gassiat, 1997], [Dacunha-Castelle and Gassiat, 1999], [Lemdani and Pons, 1997], [Lemdani and Pons, 1999], [Chen and Chen, 2001], [Garel, 2001], [Chen and Chen, 2003], [Chambaz, 2006], [Chen et al, 2004], [Garel, 2005], [Lo, 2005]. Chen et al. (followed by Qin et al.) proposed a simple and clever idea to avoid the degeneracy problems: they add a penalization to the log-likelihood with a factor increasing to infinity as the parameters tend to values where degeneracy occurs. They consequently obtain convex combination of chi-square for the asymptotic distribution of the modified testing statistic, see [Chen et al, 2001], [Chen and Kalbfleisch, 2005], [Qin and Smith, 2004], [Qin and Smith, 2006].

The aim of the current paper is to give a detailed general solution to the asymptotic distribution of the LRT statistic for the testing problem (4). One of the author proved a general result for likelihood ratio statistics under weak assumptions, see Theorem 3.1 in [Gassiat, 2002]. Some applications to mixtures were developed in Section 2 of [Azais et al., 2006]. Here, we solve the precise form of the asymptotic distribution for the previous two problems: testing a single population against any mixture, and testing the number of components in a mixture with or without structural parameter (with the above notations, it means that γ is unknown). This precise form allows to construct numerical tables by simulation or by Gaussian calculation [Mercadier, 2005].

1.2. Intuition

In the parametric case, likelihood procedures for estimating and testing parameters are well understood. Under identifiability and regularity conditions, the maximum likelihood estimator is consistent. Thus it can be

expanded around the true value of the parameter so that it is seen that this difference has asymptotic Gaussian behaviour, and the log-likelihood ratio statistic has asymptotic chi-square behaviour. This comes from two-term Taylor expansion in the classical Wald's argument, and from more intricate arguments in Le Cam's theory, see [van der Vaart, 1998]. In the semi-parametric or non-parametric situation, such a theory does not hold in full generality, [Murphy and van der Vaart, 1997]. One may try to use one dimensional sub-models to obtain the result as follows. Let $(p_t)_{t \geq 0}$ be a sub-model of probability densities with respect to some measure μ such that p_0 is the true density of the observations, and such that the parameter t is identifiable along the sub-model: $p_t = p_0$ if and only if $t = 0$. Under weak integrability conditions, the maximum likelihood estimator (m.l.e.) \hat{t} is consistent: \hat{t} tends to 0 in probability.

Assume moreover that the sub-model is differentiable in quadratic mean, with score function g in the sense of (10) below. Then the score satisfies $\int g p_0 d\mu = 0$, the Fisher information of the sub-model exists and its value is $I(g) = \int g^2 p_0 d\mu$. Under regularity assumptions, it holds that

$$\sqrt{n} \hat{t} = I(g)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(X_i) \right) 1_{\sum_{i=1}^n g(X_i) \geq 0} + o_{\mathbb{P}}(1)$$

where $o_{\mathbb{P}}(1)$ is a random variable tending to 0 in probability as the number n of observations tends to infinity. Moreover, letting $\ell_n(t) = \sum_{i=1}^n \log p_t(X_i)$, it holds that

$$\sup_{t \geq 0} \ell_n(t) - \ell_n(0) = \frac{1}{2I(g)} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(X_i) \right)^2 1_{\sum_{i=1}^n g(X_i) \geq 0} + o_{\mathbb{P}}(1).$$

If now $\{p_\theta, \theta \in \Theta\}$ is a family of probability densities, and \mathcal{S} a set of scores obtained using all possible one-dimensional sub models $(p_{\theta_t})_{t \geq 0}$, then one may think that, if \mathcal{S} is rich enough, and under Donsker-like conditions

$$\sup_{\theta \in \Theta} \ell_n(\theta) - \ell_n(\theta_0) = \frac{1}{2} \sup_{g \in \mathcal{S}} \left[\frac{1}{I(g)} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(X_i) \right)^2 1_{\sum_{i=1}^n g(X_i) \geq 0} \right] + o_{\mathbb{P}}(1), \quad (5)$$

where now $\ell_n(\theta) = \sum_{i=1}^n \log p_\theta(X_i)$ and p_{θ_0} is the density of the observations. Observe that $I(g)$ is the square norm of g in $L^2(p_0\mu)$, so that one may rewrite (5) as

$$\sup_{\theta \in \Theta} \ell_n(\theta) - \ell_n(\theta_0) = \frac{1}{2} \sup_{d \in \mathcal{D}} \left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n d(X_i) \right)^2 1_{\sum_{i=1}^n d(X_i) \geq 0} \right] + o_{\mathbb{P}}(1), \quad (6)$$

where \mathcal{D} is the set of normalized scores: $\mathcal{D} = \{g/\|g\|_2, g \in \mathcal{S}\}$, $\|\cdot\|_2$ being the norm in $L^2(p_0\mu)$.

In the regular parametric identifiable situation, where Θ is a subset of a k -dimensional Euclidean space, the largest set of scores \mathcal{S} is

$$\mathcal{S} = \left\{ \langle U, \dot{\ell}_{\theta_0} \rangle, U \in \mathbb{U} \right\}$$

where $\dot{\ell}_{\theta_0}$ is the score function at θ_0 , $\langle \cdot, \cdot \rangle$ denotes usual scalar product, \mathbb{U} is the full Euclidean space in case θ_0 is in the interior of Θ , and only a sub-cone of it in case θ_0 is on the boundary of Θ . The supremum over \mathcal{D} is easily computed and gives the asymptotic chi-square distribution in case θ_0 is in the interior of Θ , or convex combination of chi-square distribution if θ_0 is on the boundary and Θ is polyhedral.

Consider now a non-identifiable situation with model (1) and the testing problem (4). Let \mathcal{G}_i be the set of finite measures with q_i supporting points, $q_2 = q_1 + q$, $q \geq 1$. Define Θ_1 and Θ_2 the associated sets of

parameters, and LRT statistic

$$\Lambda_n = 2 \left(\sup_{\theta \in \Theta_2} \ell_n(\theta) - \sup_{\theta \in \Theta_1} \ell_n(\theta) \right). \quad (7)$$

Assume that the true density of the observations has finite mixing distribution with q_1 populations, and parameter $\theta_0 = (\gamma^0, \pi_1^0, \dots, \pi_{q_1}^0, z_1^0, \dots, z_{q_1}^0)$. Let $\dot{\ell}_z$ be the vector score for the model $(p_{\gamma^0}(\cdot|z))_z$ at point z , \dot{m}_0 be the vector score for the model $(\sum_{i=1}^{q_1} \pi_i^0 p_{\gamma^0}(\cdot|z_i^0))_\gamma$ at point γ^0 , and let $\dot{\ell}_0$ be the vector obtained by concatenation of $\dot{\ell}_{z_1^0} \frac{p_{\gamma^0}(\cdot|z_1^0)}{p_{\theta_0}(\cdot)}$, \dots , $\dot{\ell}_{z_{q_1}^0} \frac{p_{\gamma^0}(\cdot|z_{q_1}^0)}{p_{\theta_0}(\cdot)}$ and \dot{m}_0 . Then it will be proved later on that scores along one dimensional sub-models for $\eta \in \mathcal{G}_2$ are of form

$$\langle U, \dot{\ell}_0(\cdot) \rangle + \frac{\sum_{i=1}^{q_1} \alpha_i p_{\gamma^0}(\cdot|z_i^0) + \sum_{i=1}^q \rho_i p_{\gamma^0}(\cdot|z_i)}{p_{\theta_0}(\cdot)},$$

where: U is any vector (with the same dimension as $\dot{\ell}_0$), $\alpha_1, \dots, \alpha_{q_1}$ are real numbers, ρ_1, \dots, ρ_q are non negative real numbers such that $\sum_{i=1}^{q_1} \alpha_i + \sum_{i=1}^q \rho_i = 0$, and z_1, \dots, z_q are points in \mathcal{Z} . In the same way, scores along one dimensional sub-models for $\eta \in \mathcal{G}_1$ are of form

$$\langle U, \dot{\ell}_0(\cdot) \rangle + \frac{\sum_{i=1}^{q_1} \alpha_i p_{\gamma^0}(\cdot|z_i^0)}{p_{\theta_0}(\cdot)},$$

where: U is any vector (with the same dimension as $\dot{\ell}_0$), and $\alpha_1, \dots, \alpha_p$ are real numbers such that $\sum_{i=1}^{q_1} \alpha_i = 0$. Define $(W(z))_z$ as the centered Gaussian process with covariance function

$$\Gamma(z, z') = \int \frac{p_{\gamma^0}(x|z)p_{\gamma^0}(x|z')}{p_{\theta_0}(x)} d\mu(x) - 1.$$

Define V as the centered Gaussian vector with variance Σ , the variance of $\dot{\ell}_0(X_1)$, and covariance with $W(z)$ given by $C(z) = \int p_{\gamma^0}(x|z)\dot{\ell}_0(x)d\mu(x)$. Denote $B(U, \alpha, \rho, \mathbf{z})$ the variance of $\langle U, V \rangle + \sum_{i=1}^{q_1} \alpha_i W(z_i^0) + \sum_{i=1}^q \rho_i W(z_i)$ (which is a quadratic form in $U, (\alpha_i), (\rho_i)$). Then if it is possible to apply (6), Λ_n converges in distribution to the random variable Λ :

$$\Lambda = \sup_{z_1, \dots, z_q} \left[\sup_{U, \alpha, \rho \geq 0, \sum_i \rho_i + \sum_i \alpha_i = 0, B(U, \alpha, \rho, \mathbf{z}) = 1} \left(\langle U, V \rangle + \sum_{i=1}^{q_1} \alpha_i W(z_i^0) + \sum_{i=1}^q \rho_i W(z_i) \right) \right]^2 - \left[\sup_{U, \alpha, \sum_i \alpha_i = 0, B(U, \alpha, \mathbf{0}, \cdot) = 1} \left(\langle U, V \rangle + \sum_{i=1}^{q_1} \alpha_i W(z_i^0) \right) \right]^2. \quad (8)$$

Indeed, the supremum of the random variables involved in (6) are in this case easily seen to be non negative. In (8) or equivalently in (40) below, derivation of the suprema inside the brackets involves pure algebraic computations. This will be done in a further section after proving that this intuitive reasoning is indeed true. One may just notice, for the moment, that since the Fisher informations $I(g)$ may tend to 0, for (6) to be true, it is needed that the closure of \mathcal{D} in $L^2(p_{\theta_0}\mu)$ be Donsker, that is the centered process $(\frac{1}{\sqrt{n}} \sum_{i=1}^n d(X_i))_{d \in \mathcal{D}}$ converges uniformly to a Gaussian process, see [van der Vaart and Wellner, 1996] for instance for more about uniform convergence of empirical measures.

1.3. Related questions

Power is an important issue in the validation of a testing procedure. Our methods allow to identify contiguous alternatives and their associated asymptotic power. We shall not insist on this question in this work since, as usual for LRT, there is no general optimality conclusion.

For normal mixtures, [Hartigan, 1985] noted first the unboundness of the LRT when the parameters are unbounded. [Gassiat, Keribin, 2000] proved also this divergence in a mixture with Markov regime, [Ciuperca, 2002] in the contamination model of exponential densities and [Liu et al., 2003] for testing homogeneity against gamma mixture alternatives. [Liu and Shao, 2004] obtained the asymptotic distribution of a normalization of the LRT. [Azais et al., 2006] extended this result to contiguous alternatives and characterized the local power as trivial in several cases. This loss of power is also established in [Hall and Stewart, 2005] for Gaussian models under stronger hypotheses that allow the determination of the separation speed.

The estimation of the number of components in a mixture using likelihood technics is closely related to the LRT. One may use penalized likelihood and estimate the number of components by maximization. The main problem is to calibrate the penalization factor. In case the possible number of populations is a priori bounded, one obtains easily a consistent estimator as soon as it is known that the likelihood statistic remains stochastically bounded, see [Keribin, 2000], [Gassiat, 2002] Section 2, see also [James et al., 2001], [Henna, 2005] and [Chambaz, Garivier, Gassiat 2005] without prior bounds.

1.4. Roadmap

Section 2 gives a rigorous form of the heuristics explained in 1.2 leading to the asymptotic distribution in general testing problems. The main theorem gives sufficient conditions under which the result holds, and it is proved that these assumptions are not far to be necessary. The main part of the section may be viewed as a rewriting under weaker assumptions of Section 3 of [Gassiat, 2002].

Section 3 develops a particular non parametric testing procedure: testing a single population against any mixture. The latent variable is real valued and the structural parameter is known. In this context, the set of scores is exhibited. The asymptotic distribution of the LRT statistic is stated for mixtures of Gaussian, Poisson and binomial distributions. These results are completely new.

Section 4 derives our initial main goal: the application of Theorem 1 for testing the number of components in a mixture with possible unknown structural parameter in all possible situations. Indeed, in the literature one may find many papers that give partial results on that question. Section 4 gives weak simple assumptions to obtain the asymptotic distribution of the LRT statistic in all cases, and a computation of its precise form. The most popular example, that of Gaussian mixtures, is then handled.

A last section is devoted to technical proofs that are not essential at first reading.

2. ASYMPTOTICS FOR THE LRT STATISTIC

Let \mathcal{F} be a set of probability densities with respect to the measure μ on the measurable space $(\mathcal{X}, \mathcal{A})$. Let the observations X_1, \dots, X_n have density p_0 in \mathcal{F} , and denote by \hat{p} the m.l.e., that is an approximate maximizer of the log-likelihood $\ell_n(p) = \sum_{i=1}^n \log p(X_i)$: for all p in \mathcal{F} , $\ell_n(\hat{p}) \geq \ell_n(p) - o_{\mathbb{P}}(1)$, so that it satisfies $\ell_n(\hat{p}) - \ell_n(p_0) = \sup_{p \in \mathcal{F}} \ell_n(p) - \ell_n(p_0) + o_{\mathbb{P}}(1)$.

Note $H^2(p_1, p_2)$ the square Hellinger distance between densities p_1 and p_2 : $H^2(p_1, p_2) = \int (\sqrt{p_1} - \sqrt{p_2})^2 d\mu$, and $K(p_1, p_2)$ the Kullback-Leibler divergence $K(p_1, p_2) = \int p_1 \log(p_1/p_2) d\mu$ in $[0, +\infty]$. Recall the following inequality:

$$H^2(p_1, p_2) \leq K(p_1, p_2). \quad (9)$$

As usual, consistency of the m.l.e. and asymptotic distribution (of the m.l.e. or of the LRT statistic) require assumptions of different kinds. Introduce

Assumption 1. *The set $\{\log p, p \in \mathcal{F}\}$ is Glivenko-Cantelli in $p_0\mu$ probability.*

Then, if Assumption 1 holds, $K(p_0, \hat{p}) = o_{\mathbb{P}}(1)$, and by (9) also $H^2(p_0, \hat{p}) = o_{\mathbb{P}}(1)$. In order to derive the asymptotic distribution of the LRT statistic $\ell_n(\hat{p}) - \ell_n(p_0)$, we introduce one dimensional sub-models in which differentiability in quadratic mean holds with scores in some subset \mathcal{S} of $L^2(p_0\mu)$.

Assumption 2. *The set \mathcal{S} satisfies Assumption 2 if for any $g \in \mathcal{S}$, there is a sub-model $(p_{t,g})_{t \geq 0}$ in \mathcal{F} such that*

$$\int \left(\sqrt{p_{t,g}} - \sqrt{p_0} - \frac{t}{2} g \sqrt{p_0} \right)^2 d\mu = o(t^2), \quad (10)$$

and the Fisher information in the sub-model is non null: $I(g) = \int g^2 p_0 d\mu \neq 0$.

Let for any $g \in \mathcal{S}$ the m.l.e. in the sub-model $(p_{t,g})_{t \geq 0}$ be \hat{t}_g . Since for every $g \in \mathcal{S}$,

$$\ell_n(\hat{p}) - \ell_n(p_0) \geq \ell_n(p_{\hat{t}_g, g}) - \ell_n(p_0) + o_{\mathbb{P}}(1)$$

one may use classical parametric results to obtain:

Proposition 1. *Suppose that Assumptions 1 and 2 hold. Then for any finite subset \mathcal{I} of \mathcal{S} :*

$$\sup_{p \in \mathcal{F}} \ell_n(p) - \ell_n(p_0) \geq \frac{1}{2} \sup_{g \in \mathcal{I}} \left[\frac{1}{I(g)} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(X_i) \right)^2 \mathbf{1}_{\sum_{i=1}^n g(X_i) \geq 0} \right] + o_{\mathbb{P}}(1).$$

Define now the set

$$\tilde{\mathcal{D}} = \left\{ \frac{g}{\sqrt{I(g)}}, g \in \mathcal{S} \right\}.$$

Note that if g is the score in sub-model $(p_{t,g})_{t \geq 0}$ then for positive real a , ag is the score in $(p_{at,g})_{t \geq 0}$ so that we may assume that \mathcal{S} is a cone, and $\tilde{\mathcal{D}}$ is a subset of \mathcal{S} . $\tilde{\mathcal{D}}$ is also a subset of the unit sphere of $L^2(p_0\mu)$.

Define $(W(d))_{d \in \tilde{\mathcal{D}}}$ the centered Gaussian process with covariance function $\Gamma(d_1, d_2) = \int d_1 d_2 p_0 d\mu$. In other words, W is the isonormal process on $\tilde{\mathcal{D}}$. Obviously, for any finite subset \mathcal{I} of $\tilde{\mathcal{D}}$ and any x , under the assumptions of Proposition 1,

$$\liminf_{n \rightarrow +\infty} \mathbb{P} \left(\sup_{p \in \mathcal{F}} \ell_n(p) - \ell_n(p_0) \geq x \right) \geq \mathbb{P} \left(\frac{1}{2} \sup_{d \in \mathcal{I}} W(d)^2 \mathbf{1}_{W(d) \geq 0} \geq x \right)$$

so that as soon as $\sup_{d \in \tilde{\mathcal{D}}} W(d)^2 \mathbf{1}_{W(d) \geq 0}$ is not finite a.s., so is asymptotically $\sup_{p \in \mathcal{F}} \ell_n(p) - \ell_n(p_0)$.

Properties of the isonormal Gaussian process indexed by a subset \mathcal{H} of the Hilbert space $L^2(p_0\mu)$ are understood through entropy numbers. Let \mathcal{H} be a class of real functions and d a metric on it. The ϵ -covering number $N(\epsilon, \mathcal{H}, d)$ is defined to be the minimum number of balls of radius ϵ needed to cover \mathcal{H} . The ϵ -bracketing number $N_{[\cdot]}(\epsilon, \mathcal{H}, d)$ is defined to be the minimum number of brackets of size ϵ needed to cover \mathcal{H} , where a bracket of size ϵ is a set of the form $[l, u] = \{h : l(x) \leq h(x) \leq u(x), \forall x \in \mathcal{X}\}$ and $d(l, u) \leq \epsilon$. The ϵ -covering number is upper bounded by the ϵ -bracketing number.

Suppose that the closure of $\tilde{\mathcal{D}}$ is compact. Remarking that the isonormal process is stationary with respect to the group of isometries on the unit sphere of $L^2(p_0\mu)$, it is known (see Theorem 4.16 of [Adler, 1990]) that a necessary and sufficient condition for finiteness of $\sup_{\tilde{\mathcal{D}}} W(d)$ is the convergence of

$$\int_0^{+\infty} \sqrt{\log(N(\epsilon, \tilde{\mathcal{D}}, d))} d\epsilon \quad (11)$$

where d is the canonical distance in $L^2(p_0\mu)$ since the process is isonormal. Throughout the paper, the canonical distance will be used for bracketing and covering numbers, so that d will be omitted in the notation.

To obtain the convergence result for the LRT statistic, in view of Proposition 1, it is needed that for a rich enough \mathcal{S} , the associated $\tilde{\mathcal{D}}$ be Donsker, in which case the closure of $\tilde{\mathcal{D}}$ is compact and the isonormal process indexed by $\tilde{\mathcal{D}}$ has a.s. uniformly continuous and bounded paths.

When $\tilde{\mathcal{D}}$ is not compact, one could use (if this is the case) that it has parametric description and is locally

compact.

It has been noticed in earlier papers that the LRT statistic may diverge to infinity in p_0 -probability for mixture testing problems, see [Hartigan, 1985], [Ciuperca, 2002], [Gassiat, Keribin, 2000], [Liu and Shao, 2004], [Liu et al., 2003], [Hall and Stewart, 2005], [Azais et al., 2006]. In all these papers, the reason is that the set of normalized scores contains almost an infinite dimensional linear space (one may construct an infinite sequence of normalized scores with Hilbertian product near zero).

We shall now state a sufficient condition under which the asymptotic distribution of the LRT statistic may be derived. For any positive ϵ , define

$$\mathcal{D}_\epsilon = \left\{ \frac{\sqrt{\frac{p}{p_0}} - 1}{H(p, p_0)}, H(p, p_0) \leq \epsilon, p \in \mathcal{F} \setminus \{p_0\} \right\} \quad (12)$$

and \mathcal{D} as the set of all limit points (in $L^2(p_0\mu)$) of sequences of functions in \mathcal{D}_{ϵ_n} , $\epsilon_n \rightarrow 0$. Then the closure of \mathcal{D}_ϵ is $\overline{\mathcal{D}}_\epsilon = \mathcal{D}_\epsilon \cup \mathcal{D}$. Introduce

Assumption 3. *There exists some positive ϵ_0 such that $\overline{\mathcal{D}}_{\epsilon_0}$ is Donsker and has a $p_0\mu$ -square integrable envelope function F .*

A sufficient condition for Assumption 3 to hold is that

$$\int_0^{+\infty} \sqrt{\log N_{[]} (u, \overline{\mathcal{D}}_{\epsilon_0})} du < +\infty,$$

see Theorem 19.5 in [van der Vaart, 1998]. Under Assumption 3, $\overline{\mathcal{D}}_{\epsilon_0}$ is a compact subset of the unit sphere of $L^2(p_0\mu)$. Thus \mathcal{D} is also the compact subset

$$\mathcal{D} = \bigcap_{\epsilon \leq \epsilon_0} \overline{\mathcal{D}}_\epsilon.$$

Under Assumption 3, \mathcal{D} is the "rich enough" set $\tilde{\mathcal{D}}$ that we need to obtain precise asymptotics for the LRT statistic. We shall need differentiability in quadratic mean along sub-models with scores in a dense subset of \mathcal{D} . This will in general be a consequence of smooth parameterization: in case \mathcal{F} may be continuously parameterized, all functions in \mathcal{D} are half score functions along one-dimensional sub-models (since they occur as the Hellinger distance to the true distribution tends to 0) or limit points of such scores when their norm (the Fisher Information along the sub-model) tends to 0.

Theorem 1. *Assume that Assumptions 1 and 3 hold. Assume there exists a dense subset \mathcal{S} of \mathcal{D} for which Assumption 2 holds. Then*

$$\sup_{p \in \mathcal{F}} \ell_n(p) - \ell_n(p_0) = \frac{1}{2} \sup_{d \in \mathcal{D}} \left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n d(X_i) \right)^2 \mathbf{1}_{\sum_{i=1}^n d(X_i) \geq 0} \right] + o_{\mathbb{P}}(1).$$

Proof: The proof follows that of Theorem 3.1 in [Gassiat, 2002]. Since $H(\hat{p}, p_0) = o_{\mathbb{P}}(1)$ by Assumption 1, for any positive ϵ ,

$$\ell_n(\hat{p}) - \ell_n(p_0) = \sup_{p \in \mathcal{F}, H(p, p_0) \leq \epsilon} (\ell_n(p) - \ell_n(p_0)) + o_{\mathbb{P}}(1),$$

so that we can limit our attention to the densities p belonging to

$$\mathcal{F}_{\epsilon_0} = \{p \in \mathcal{F} \setminus \{p_0\} : H(p, p_0) \leq \epsilon_0\}.$$

Define for any $p \in \mathcal{F}_{\epsilon_0}$

$$s_p = \frac{\sqrt{\frac{p}{p_0}} - 1}{H(p, p_0)} \quad \text{and} \quad \bar{s}_p = s_p - \int s_p p_0 d\mu = s_p + \frac{H(p, p_0)}{2}. \quad (13)$$

Step 1 We have for $p \in \mathcal{F}$:

$$\ell_n(p) - \ell_n(p_0) = 2 \sum_{i=1}^n \log \left(\sqrt{\frac{p}{p_0}} \right) (X_i) = 2 \sum_{i=1}^n \log \left(1 + H(p, p_0) s_p(X_i) \right).$$

Throughout the paper, for a real number u , we note $u_- = -u1_{u < 0}$ and $u_+ = u1_{u > 0}$. Since for $u > -1$, $\log(1+u) \leq u - \frac{1}{2}u_-^2$, we have

$$\ell_n(p) - \ell_n(p_0) \leq 2H(p, p_0) \sum_{i=1}^n s_p(X_i) - H^2(p, p_0) \sum_{i=1}^n (s_p(X_i))_-^2.$$

As a consequence, for p such that $\ell_n(p) - \ell_n(p_0) \geq 0$,

$$\sqrt{n}H(p, p_0) \leq 2 \frac{n^{-1/2} \sum_{i=1}^n s_p(X_i)}{n^{-1} \sum_{i=1}^n (s_p(X_i))_-^2} \leq 2 \frac{n^{-1/2} \sum_{i=1}^n \bar{s}_p(X_i)}{n^{-1} \sum_{i=1}^n (s_p(X_i))_-^2}$$

since $s_p \leq \bar{s}_p$. Theorem 2.6.10 of [van der Vaart and Wellner, 1996] and Assumption 3 give that the set $\{(s_p)_-, p \in \mathcal{F}_{\epsilon_0}\}$ is Donsker and has integrable envelope, so that by Lemma 2.10.14 of [van der Vaart and Wellner, 1996], the set of squares is Glivenko-Cantelli, and the right hand side of the previous inequality is uniformly $O_{\mathbb{P}}(1)$ as soon as $\inf_{p \in \mathcal{F}_{\epsilon_0}} \int (s_p)_-^2 p_0 d\mu \neq 0$, which holds and may be proved by contradiction. Thus

$$\sup_{p \in \mathcal{F}_{\epsilon_0}: \ell_n(p) - \ell_n(p_0) \geq 0} H(p, p_0) = O_{\mathbb{P}}(n^{-1/2}). \quad (14)$$

Step 2 Setting $\log(1+u) = u - u^2/2 + u^2 R(u)$ with $R(u) \rightarrow 0$ as $u \rightarrow 0$, it comes:

$$\ell_n(p) - \ell_n(p_0) = 2H(p, p_0) \sum_{i=1}^n s_p(X_i) - H^2(p, p_0) \sum_{i=1}^n s_p^2(X_i) + 2H^2(p, p_0) \sum_{i=1}^n s_p^2(X_i) R(H(p, p_0) s_p(X_i)). \quad (15)$$

Since the envelope function F is square integrable, an application of Markov inequality to the variable $F1_{F \geq \sqrt{n}\eta}$ yields

$$\max_{i=1, \dots, n} F(X_i) = o_{\mathbb{P}}(\sqrt{n}).$$

Also, by Lemma 2.10.14 of [van der Vaart and Wellner, 1996], the set $\{s_p^2, p \in \mathcal{F}_{\epsilon_0}\}$ is Glivenko-Cantelli with $\int s_p^2 p_0 d\mu = 1$. Then it easy to see that the last term in (15) is negligible as soon as $H(p, p_0) = O_{\mathbb{P}}(n^{-1/2})$:

$$\sup_{p \in \mathcal{F}_{\epsilon_0}: \ell_n(p) - \ell_n(p_0) \geq 0} (\ell_n(p) - \ell_n(p_0)) = \sup_{p \in \mathcal{F}_{\epsilon_0}: \ell_n(p) - \ell_n(p_0) \geq 0} \left(2H(p, p_0) \sum_{i=1}^n s_p(X_i) - H^2(p, p_0) \sum_{i=1}^n s_p^2(X_i) \right) + o_{\mathbb{P}}(1).$$

Now, $s_p = \bar{s}_p - H(p, p_0)/2$, so that

$$\sup_{p \in \mathcal{F}_{\epsilon_0}: \ell_n(p) - \ell_n(p_0) \geq 0} (\ell_n(p) - \ell_n(p_0)) = \sup_{p \in \mathcal{F}_{\epsilon_0}: \ell_n(p) - \ell_n(p_0) \geq 0} 2 \left(H(p, p_0) \sum_{i=1}^n \bar{s}_p(X_i) - nH^2(p, p_0) \right) + o_{\mathbb{P}}(1).$$

Using Equation (14), we have that, for any ϵ_n tending to 0 slower than $1/\sqrt{n}$,

$$\sup_{p \in \mathcal{F}_{\epsilon_0} : \ell_n(p) - \ell_n(p_0) \geq 0} (\ell_n(p) - \ell_n(p_0)) = \sup_{p \in \mathcal{F}_{\epsilon_n} : \ell_n(p) - \ell_n(p_0) \geq 0} (\ell_n(p) - \ell_n(p_0)) + o_{\mathbb{P}}(1),$$

and maximizing in $H(p, p_0)$ gives that

$$\begin{aligned} \sup_{p \in \mathcal{F}} (\ell_n(p) - \ell_n(p_0)) &\leq \frac{1}{2} \sup_{p \in \mathcal{F}_{\epsilon_n}} \left(n^{-1/2} \left(\sum_{i=1}^n \bar{s}_p(X_i) \right) 1_{\sum_{i=1}^n \bar{s}_p(X_i) \geq 0} \right)^2 + o_{\mathbb{P}}(1) \\ &= \frac{1}{2} \sup_{d \in \mathcal{D}_{\epsilon_n}} \left(n^{-1/2} \left(\sum_{i=1}^n d(X_i) \right) 1_{\sum_{i=1}^n d(X_i) \geq 0} \right)^2 + o_{\mathbb{P}}(1). \end{aligned}$$

But if we represent weak convergence by an almost sure convergence in a suitable probability space (see for instance Theorem 1.10.3 of [van der Vaart and Wellner, 1996]), we get that for any $\epsilon \leq \epsilon_0$

$$\sup_{d \in \mathcal{D}_{\epsilon}} \left(n^{-1/2} \left(\sum_{i=1}^n d(X_i) \right) 1_{\sum_{i=1}^n d(X_i) \geq 0} \right)^2 = \sup_{d \in \mathcal{D}_{\epsilon}} (W(d) 1_{W(d) \geq 0})^2 + o_{\mathbb{P}}(1),$$

where W is the isonormal Gaussian process. Since \mathcal{D} and $\bar{\mathcal{D}}_{\epsilon}$ are compact, the distance between \mathcal{D} and the complementary of \mathcal{D}_{ϵ} tends to zero as $\epsilon \rightarrow 0$, and the isonormal process W is continuous on $\bar{\mathcal{D}}_{\epsilon_0}$,

$$\sup_{d \in \mathcal{D}_{\epsilon_n}} (W(d) 1_{W(d) \geq 0})^2 = \sup_{d \in \mathcal{D}} (W(d) 1_{W(d) \geq 0})^2 + o_{\mathbb{P}}(1),$$

so that

$$\sup_{p \in \mathcal{F}} \ell_n(p) - \ell_n(p_0) \leq \frac{1}{2} \sup_{d \in \mathcal{D}} \left[\left(n^{-1/2} \sum_{i=1}^n d(X_i) \right)^2 1_{\sum_{i=1}^n d(X_i) \geq 0} \right] + o_{\mathbb{P}}(1). \quad (16)$$

Step 3 We have by Proposition 1

$$\sup_{p \in \mathcal{F}} \ell_n(p) - \ell_n(p_0) \geq \frac{1}{2} \sup_{d \in \mathcal{S}} \left[\left(n^{-1/2} \sum_{i=1}^n d(X_i) \right)^2 1_{\sum_{i=1}^n d(X_i) \geq 0} \right] + o_{\mathbb{P}}(1). \quad (17)$$

But again, the isonormal process W is separable and \mathcal{D} is compact, so that the supremum over \mathcal{S} equals the supremum over \mathcal{D} , and the theorem follows from equations (16) and (17).

Proposition 2. *Let Assumption 3 hold. If $(p_n)_{n \in \mathbb{N}}$ is a deterministic sequence in \mathcal{F} such that $\sqrt{n}H(p_n, p_0)$ tends to a positive constant $c/2$, and s_{p_n} tends to d_0 in \mathcal{D} , then the sequences $(p_n \mu)^{\otimes n}$ and $(p_0 \mu)^{\otimes n}$ are mutually contiguous.*

Proof. Indeed,

$$\begin{aligned}
\ell_n(p_n) - \ell_n(p_0) &= \frac{c}{\sqrt{n}} \sum_{i=1}^n \bar{s}_{p_n}(X_i) - \frac{c^2}{2} + o_{\mathbb{P}}(1) \\
&= cW(s_{p_n}) + o_{\mathbb{P}}(1) - \frac{c^2}{2} + o_{\mathbb{P}}(1) \\
&= cW(d_0) + o_{\mathbb{P}}(1) - \frac{c^2}{2} + o_{\mathbb{P}}(1) \\
&= \frac{c}{\sqrt{n}} \sum_{i=1}^n d_0(X_i) - \frac{c^2}{2} + o_{\mathbb{P}}(1).
\end{aligned}$$

The proposition follows from Example 6.5 of [van der Vaart, 1998].

Then one may apply Le Cam's third Lemma (see for instance [van der Vaart, 1998]) to obtain that under the assumptions of Theorem 1,

$$\sup_{p \in \mathcal{F}} \ell_n(p) - \ell_n(p_0)$$

converges in distribution under $(p_n \mu)^{\otimes n}$ (that is if X_1, \dots, X_n are i.i.d. with density p_n) to

$$\frac{1}{2} \sup_{d \in \mathcal{D}} \left[(W(d) + c\Gamma(d, d_0))^2 1_{W(d) + c\Gamma(d, d_0) \geq 0} \right].$$

Thus Theorem 1 allows to derive asymptotic properties of LRT statistic as the difference of two terms in which the sets \mathcal{D} are defined under the null hypothesis H_0 and the alternative hypothesis H_1 respectively.

3. ONE POPULATION AGAINST ANY MIXTURE

3.1. General result

We now assume that \mathcal{Z} is a closed compact interval of \mathbb{R} . Consider the mixture model with known structural parameter, that is \mathcal{F} is the set of densities:

$$p_\eta(x) = \int p(x|z) d\eta(z). \quad (18)$$

Let the observations X_1, \dots, X_n be i.i.d. with distribution $p_0 = p_{\eta_0}$, $\eta_0 \in \mathcal{G}$, where \mathcal{G} is the set of all probability distributions on $(\mathcal{Z}, \mathcal{C})$.

We want to test that the mixing measure η reduces to a Dirac mass δ_z at some $z \in \mathcal{Z}$ against that it does not:

$$H_0 : \text{"}\exists z_0 \in \mathcal{Z} : p_{\eta_0}(\cdot) = p(\cdot|z_0)\text{"} \text{ against } H_1 : \text{"}\forall z \in \mathcal{Z} : p_{\eta_0}(\cdot) \neq p(\cdot|z)\text{"}. \quad (19)$$

We assume that $p_0 = p(\cdot|z_0)$ for some z_0 in the interior of \mathcal{Z} .

We shall need the following weak local identifiability assumption:

Assumption 4. For any \tilde{z} in \mathcal{Z} , $p_\eta(\cdot) = p(\cdot|\tilde{z})$ if and only if η is the Dirac mass at \tilde{z} .

We shall use:

Assumption 5. For all $x, z \rightarrow p(x|z)$ is continuous, $|\log \sup_z p(\cdot|z)|$ and $|\log \inf_z p(\cdot|z)|$ are $p_0 \mu$ -integrable.

Since \mathcal{Z} is compact, \mathcal{G} is a compact metric space (for the weak convergence topology), $\eta \rightarrow \log \int p(x|z)d\eta(z)$ is continuous for all x and it is easy to see that Assumption 1 holds, so that if $\hat{\eta}$ is the m.l.e.,

$$H(p_{\hat{\eta}}, p_0) = o_{\mathbb{P}}(1).$$

We now assume that $p(\cdot|z)$ is defined for z in some open interval \mathcal{Z}^+ that contains \mathcal{Z} .

Assumption 6. For all x , $z \rightarrow p(x|z)$ is twice continuously differentiable on \mathcal{Z}^+ ,

$$\frac{\sup_{z \in \mathcal{Z}} p(x|z)}{p_0(x)} \in L^2(p_0\mu), \quad \frac{\sup_{z \in \mathcal{Z}} \dot{p}(x|z)}{\inf_{z \in \mathcal{Z}} p(x|z)} \in L^2(p_0\mu),$$

and for some neighborhood \mathcal{V} of z_0 :

$$\frac{\sup_{z \in \mathcal{V}} \ddot{p}(x|z)}{p_0(x)} \in L^2(p_0\mu).$$

Here, $\dot{p}(x|z)$ and $\ddot{p}(x|z)$ denote the first and second derivative of $p(x|z)$ with respect to the variable z .

For any $\nu \in \mathcal{G}$, we shall denote by F_ν its cumulative distribution function. Define for $a \in \mathbb{R}$, $b \geq 0$, $c \geq 0$, $\nu \in \mathcal{G}$ with F_ν continuous at z_0 :

$$d(a, b, c, \nu)(x) = a \frac{\dot{p}(x|z_0)}{p_0(x)} + b \frac{\ddot{p}(x|z_0)}{p_0(x)} + c \left(\frac{p_\nu(x) - p_0(x)}{p_0(x)} \right). \quad (20)$$

We shall need the following assumption:

Assumption 7. $d(a, b, c, \nu) = 0$ μ -a.e. if and only if $a = 0$, $b = 0$, and $c = 0$.

Let now \mathcal{E} be the closure in $L^2(p_0\mu)$ of the set

$$\left\{ \frac{d(a, b, c, \nu)}{\|d(a, b, c, \nu)\|_2} : \|d(a, b, c, \nu)\|_2 \neq 0, a \in \mathbb{R}, b \geq 0, c \geq 0, \nu \in \mathcal{G} \text{ s.t. } F_\nu \text{ continuous at } z_0 \right\}. \quad (21)$$

Throughout the paper, $\|\cdot\|_2$ will denote the norm in $L^2(p_0\mu)$.

Then under Assumption 7, \mathcal{E} is the union of (21) and the set of all limit points in $L^2(p_0\mu)$ of sequences $\frac{d(a_n, b_n, c_n, \nu_n)}{\|d(a_n, b_n, c_n, \nu_n)\|_2}$ for $a_n \rightarrow 0$, $b_n \rightarrow 0$, and $(c_n \rightarrow 0$ or ν_n converging weakly to the Dirac mass at z_0).

Proposition 3. Under Assumptions 4, 5, 6 and 7, the set of all possible accumulation points in $L^2(p_0\mu)$ of sequences of functions in \mathcal{D}_{ϵ_n} , ϵ_n tending to 0, is exactly \mathcal{E} , and there exists a dense subset of \mathcal{E} for which Assumption 2 holds.

Proof. Let \mathcal{A} be the set of all possible accumulation points in $L^2(p_0\mu)$ of sequences of functions in \mathcal{D}_{ϵ_n} , with ϵ_n tending to 0. Define for $a \in \mathbb{R}$, $0 \leq b \leq 1$, $c \geq 0$, $\nu \in \mathcal{G}$ with F_ν continuous at z_0 and $t \geq 0$:

$$p_{\eta_t}(x) = (1-b) \left[(1-ct^2) p(x|z_0 + at^2) + ct^2 p_\nu \right] + \frac{b}{2} [p(x|z_0 - t) + p(x|z_0 + t)].$$

Then using Assumption 6 and Lemma 7.6 in [van der Vaart, 1998], $(p_{\eta_t})_{t \geq 0}$ is differentiable in quadratic mean with score function $d((1-b)a, b, (1-b)c, \nu)$. Since as t tends to 0

$$\frac{\sqrt{\frac{p_{\eta_t}}{p_0}} - 1}{H(p_{\eta_t}, p_0)} \xrightarrow{L^2(p_0\mu)} \frac{d((1-b)a, b, (1-b)c, \nu)}{\|d((1-b)a, b, (1-b)c, \nu)\|_2},$$

we get that

$$\mathcal{E} \subset \mathcal{A}$$

and that there exists a dense subset of \mathcal{E} for which Assumption 2 holds.

Let now

$$\frac{\sqrt{\frac{p_{\eta_t}}{p_0} - 1}}{H(p_{\eta_t}, p_0)}$$

be a sequence converging in $L^2(p_0\mu)$ and such that $H(p_{\eta_t}, p_0)$ tends to 0 as t tends to 0. Then η_t converges weakly to the Dirac mass at z_0 , and for all x , $p_{\eta_t}(x)$ converges to $p_0(x)$.

Notice that for all x , there exists $y_t(x)$ in $(p_{\eta_t}(x), p_0(x))$ such that

$$\sqrt{p_{\eta_t}(x)} - \sqrt{p_0(x)} = \frac{p_{\eta_t}(x) - p_0(x)}{2\sqrt{y_t(x)}}. \quad (22)$$

For any sequence u_t of non negative real numbers, let

$$\rho_t = F_t((z_0 - u_t)^-) + 1 - F_t(z_0 + u_t),$$

with $F_0 = F_{\eta_0}$ and $F_t = F_{\eta_t}$.

Notice the following. For any positive u , $F_t((z_0 - u)^-) + 1 - F_t(z_0 + u)$ tends to 0, so that there exists u_t , a non negative sequence of real numbers decreasing (slowly enough) to 0, such that ρ_t tends to 0. (From now on, unless specifically said, all limits are taken as t tends to 0).

Let

$$m_t = \int_{|z-z_0| \leq u_t} (z - z_0) d\eta_t \quad \text{and} \quad e_t = \frac{1}{2} \int_{|z-z_0| \leq u_t} (z - z_0)^2 d\eta_t.$$

Then m_t and e_t tend to 0. Also, as soon as $\rho_t < 1$, $e_t > 0$, and if this is the case for all small enough t , using Assumption 6,

$$\int_{|z-z_0| \leq u_t} (p(x|z) - p(x|z_0)) d\eta_t = [m_t \dot{p}(x|z_0) + e_t \ddot{p}(x|z_0)] (1 + o(1)). \quad (23)$$

Also,

$$\int_{|z-z_0| > u_t} (p(x|z) - p(x|z_0)) d\eta_t = \int_{|z-z_0| > u_t} \dot{p}(x|z) (F_0(z) - F_t(z)) dz. \quad (24)$$

Define, if $\rho_t \neq 0$, the probability distribution η_t restricted to $|z - z_0| > u_t$, with distribution function G_t :

$$G_t(z) = \frac{F_t(z)}{\rho_t} = \frac{F_t(z) - F_0(z)}{\rho_t}, \quad z < z_0 - u_t,$$

$$G_t(z) = 1 - \frac{1 - F_t(z)}{\rho_t} = 1 - \frac{F_0(z) - F_t(z)}{\rho_t}, \quad z > z_0 + u_t.$$

Let ν_t be the probability distribution with repartition function G_t (which is continuous at z_0). Then using Equation (24),

$$\int_{|z-z_0| > u_t} (p(x|z) - p(x|z_0)) d\eta_t = \rho_t (p_{\nu_t} - p_0) (1 + o(1)), \quad (25)$$

and

$$p_{\eta_t}(x) - p_0(x) = [m_t \dot{p}(x|z_0) + e_t \ddot{p}(x|z_0) + \rho_t (p_{\nu_t} - p_0)] (1 + o(1)),$$

which leads, using Equation (22) to

$$\sqrt{\frac{p_{\eta_t}}{p_0}} - 1 = \frac{1}{2} \left[m_t \frac{\dot{p}(x|z_0)}{p_0} + e_t \frac{\ddot{p}(x|z_0)}{p_0} + \rho_t \frac{(p_{\nu_t} - p_0)}{p_0} \right] (1 + o(1)).$$

Using Assumption 6, the limit also holds in $L^2(p_0\mu)$ so that

$$\frac{\sqrt{\frac{p_{\eta_t}}{p_0}} - 1}{H(p_{\eta_t}, p_0)} = \frac{d(m_t, e_t, \rho_t, \nu_t)}{\|d(m_t, e_t, \rho_t, \nu_t)\|_2} (1 + o(1)).$$

Also, if there exists a sequence u_t such that for small enough t , $\rho_t = 0$, then the limit of $(\sqrt{\frac{p_{\eta_t}}{p_0}} - 1)/H(p_{\eta_t}, p_0)$ is some $d(a, b, 0, \cdot)/\|d(a, b, 0, \cdot)\|_2$ in \mathcal{D} .

We have thus proved that in all cases the limit of $(\sqrt{\frac{p_{\eta_t}}{p_0}} - 1)/H(p_{\eta_t}, p_0)$ is in $\overline{\mathcal{D}}$, and

$$\mathcal{A} \subset \mathcal{E}.$$

In other words, the set of scores \mathcal{D} is the set \mathcal{E} , which is the closure of the set

$$\left\{ \left(\frac{p_\eta(x) - p_0(x)}{p_0(x)}, \left\| \frac{p_\eta(x) - p_0(x)}{p_0(x)} \right\|_2 \neq 0, \eta \in \mathcal{G} \right) \right\}. \quad (26)$$

Then, in such a situation, Assumption 3 is not weaker than the following one, which may be easier to verify:

Assumption 8. \mathcal{E} is Donsker and has a $p_0\mu$ -square integrable envelope function.

Comparing with Theorem 3.1 in [Gassiat, 2002], we thus state:

Theorem 2. Under Assumptions 4, 5, 6, 7, and 8,

$$\sup_{p \in \mathcal{F}} \ell_n(p) - \ell_n(p_0) = \frac{1}{2} \sup_{d \in \mathcal{E}} \left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n d(X_i) \right)_+^2 \right] + o_P(1).$$

3.2. Application to the Gaussian family

In this section, we derive the asymptotic distribution of the LRT statistic for testing that the observations follow a Gaussian distribution (with unknown mean) against that they follow a mixture of Gaussian distributions. That is we consider the situation where μ is the Lebesgue measure on \mathbb{R} and:

$$p(x|z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-z)^2}{2}\right), \forall x \in \mathbb{R}.$$

Here, the variance is considered to be known and is fixed to one without loss of generality, and $\mathcal{Z} = [-M, +M]$. We note $M = \max\{\underline{M}, \overline{M}\}$.

Assumptions 4, 5, 6, and 7 hold.

Let $(H_k)_{k \in \mathbb{N}}$ be the Hermite polynomials, given by:

$$H_k(x) = (-1)^k \exp\left(\frac{x^2}{2}\right) \left[\frac{d^k}{dx^k} \exp\left(-\frac{x^2}{2}\right) \right].$$

By the Taylor formula

$$\frac{p(x|z)}{p_0(x)} = 1 + \sum_{k=1}^{\infty} \frac{(z - z_0)^k}{k!} H_k(x - z_0),$$

and

$$\frac{p_\eta(x) - p_0(x)}{p_0(x)} = \sum_{k=1}^{\infty} \frac{\mathbb{E}[(Z - z_0)^k]}{k!} H_k(x - z_0) \quad (27)$$

where Z is a random variable with distribution η , $\eta \in \mathcal{G}$. Set $m_{k,\eta} = \mathbb{E}[(Z - z_0)^k]$.

The $\{H_k(x - z_0), k \in \mathbb{N}\}$ form an orthogonal system in $L^2(p_0\mu)$, each $H_k(x - z_0)$ having norm $\sqrt{k!}$. Then \mathcal{E} is the closure of the set

$$\left\{ \frac{\sum_{k=1}^{\infty} \frac{m_{k,\eta}}{k!} H_k(x - z_0)}{\left(\sum_{k=1}^{\infty} \frac{(m_{k,\eta})^2}{k!}\right)^{1/2}}, \eta \in \mathcal{G} \right\}.$$

Let $(W_k)_{k \geq 1}$ be a sequence of i.i.d. standard Gaussian random variables. Then

Theorem 3. *The likelihood ratio test statistic $2 \left(\sup_{\eta \in \mathcal{G}} \ell_n(\eta) - \sup_{z \in \mathcal{Z}} \ell_n(\delta_z) \right)$ converges in distribution, as n tends to infinity, to*

$$\sup_{\eta \in \mathcal{G}} \frac{\left(\sum_{k=1}^{\infty} \frac{m_{k,\eta}}{\sqrt{k!}} W_k \right)_+^2}{\sum_{k=1}^{\infty} \frac{(m_{k,\eta})^2}{k!}} - W_1^2.$$

Recall that $u_+ = u \mathbf{1}_{u > 0}$.

Proof: Classical parametric theory gives that $2 \left(\sup_{z \in \mathcal{Z}} \ell_n(\delta_z) - \ell_n(\delta_{z_0}) \right)$ converges in distribution to W_1^2 , as n tends to infinity. Theorem 1 will give that $2 \left(\sup_{\eta \in \mathcal{G}} \ell_n(\eta) - \ell_n(\delta_{z_0}) \right)$ converges in distribution, as n tends to infinity, to

$$\sup_{\eta \in \mathcal{G}} \frac{\left(\sum_{k=1}^{\infty} \frac{m_{k,\eta}}{\sqrt{k!}} W_k \right)_+^2}{\sum_{k=1}^{\infty} \frac{(m_{k,\eta})^2}{k!}}$$

as soon as it is seen that Assumption 8 holds.

Let us first prove that \mathcal{E} has a $p_0\mu$ -square integrable envelope function. We have

$$\left\| \frac{p_\eta}{p_0} - 1 \right\|_2 \geq \mathbb{E} \left| \frac{p_\eta(X)}{p_0(X)} - 1 \right|$$

where X is a random variable with density p_0 . But, for Z a random variable with distribution η ,

$$\begin{aligned} \left| \frac{p_\eta(x)}{p_0(x)} - 1 \right| &= \mathbb{E} \left| \exp \left((Z - z_0)(x - z_0) - \frac{(Z - z_0)^2}{2} \right) - 1 \right| \\ &\geq \mathbb{E} \left[\left((Z - z_0)(x - z_0) - \frac{(Z - z_0)^2}{2} \right) \mathbf{1}_{(Z - z_0)(x - z_0) \geq \frac{(Z - z_0)^2}{2}} \right] \\ &\geq \mathbb{E} \left[\left((Z - z_0)(x - z_0) - \frac{(Z - z_0)^2}{2} \right) \mathbf{1}_{(Z - z_0)(x - z_0) \geq (Z - z_0)^2} \right], \end{aligned}$$

so that

$$\left\| \frac{p_\eta}{p_0} - 1 \right\|_2 \geq m_{2,\eta}/2. \quad (28)$$

On the other hand since

$$\mathbf{1}_{(Z - z_0)(x - z_0) \geq (Z - z_0)^2} \geq \mathbf{1}_{x \geq M} \mathbf{1}_{(Z - z_0) \geq 0} + \mathbf{1}_{x \leq -M} \mathbf{1}_{(Z - z_0) \leq 0}$$

one gets

$$\left\| \frac{p_\eta}{p_0} - 1 \right\|_2 \geq \frac{1}{2} (|m_{1,\eta}|(M - |z_0|) - \frac{m_{2,\eta}}{2}) \mathbb{P}_0(|X| \geq M + |z_0|). \quad (29)$$

Also, remarking that $|\exp(u) - 1| \leq |u| \exp(|u|)$, one obtains easily that

$$\left| \frac{p_\eta(x)}{p_0(x)} - 1 \right| \leq \left(|m_{1,\eta}| |x - z_0| + \frac{m_{2,\eta}}{2} \right) \exp[(M + |z_0|)|x - z_0|]. \quad (30)$$

Using (28) (29) and (30), it follows that there exists a positive constant C such that for any $\eta \in \mathcal{G}$,

$$\frac{\left| \frac{p_\eta(x)}{p_0(x)} - 1 \right|}{\left\| \frac{p_\eta}{p_0} - 1 \right\|_2} \leq C (|x - z_0| + 1) \exp[(M + |z_0|)|x - z_0|]$$

which proves that \mathcal{E} has a $p_0\mu$ -square integrable envelope function.

Let now $f_k = \frac{H_k}{k\sqrt{k!}}$ and set

$$c_{k,\eta} = \frac{km_{k,\eta}/\sqrt{k!}}{(\sum_{k=1}^{\infty} \frac{(m_{k,\eta})^2}{k!})^{1/2}}.$$

Then

$$\sum_{k=1}^{\infty} c_{k,\eta}^2 \leq 1 + 2 \sum_{k=2}^{\infty} \frac{k^2 (m_{k,\eta})^2}{(m_{2,\eta})^2 k!}.$$

Since $|m_{k,\eta}| \leq m_{2,\eta} M^{k-2}$ for $k \geq 2$, we can write

$$\sum_{k=1}^{\infty} c_{k,\eta}^2 \leq 1 + 2 \sum_{k=2}^{\infty} \frac{k^2 (M^{k-2})^2}{k!}$$

and for a constant C that does not depend of η :

$$\sum_{k=1}^{\infty} c_{k,\eta}^2 \leq CM^4 \exp M^2.$$

As a consequence:

$$\mathcal{E} \subset \left\{ \sum_{k=1}^{\infty} c_k f_k, \sum_{k=1}^{\infty} c_k^2 \leq CM^4 \exp M^2 \right\},$$

which is Donsker by Theorem 2.13.2 of [van der Vaart and Wellner, 1996], so that \mathcal{E} is Donsker.

3.3. Application to the Poisson family

The study of mixtures of Poisson distributions is, for example, motivated by ecological sampling problems, see [Misra et al., 2003]. The abundance of a given species can be modelled by a Poisson distribution with parameter depending on the species. If there is a lot of species that are too difficult to identify, the number of individuals on a given location will have a distribution which is a mixture of Poisson distributions.

In this section, we derive the asymptotic distribution of the LRT statistic for testing that the observations follow a Poisson distribution (with unknown mean) against that they follow a mixture of Poisson distributions. That is we consider the situation where:

$$p(x|z) = \frac{z^x}{x!} \exp(-z), \quad \forall x \in \mathbb{N}.$$

Here, $\mathcal{Z} = [\underline{M}, \overline{M}]$ with $\underline{M} > 0$. Again Assumptions 4, 5, 6, and 7 hold. Let $(C_k)_{k \in \mathbb{N}}$ be the Charlier polynomials (see [Chihara, 1978]):

$$C_k(x) = z_0^k \left[\frac{d^k}{dz^k} \left(\frac{z}{z_0} \right)^x \exp(-z + z_0) \right]_{z=z_0}.$$

Then, as in (27)

$$\frac{p_\eta(x) - p_0(x)}{p_0(x)} = \sum_{k=1}^{\infty} \frac{\mathbb{E}[(Z - z_0)^k]}{k! z_0^k} C_k(x) \quad (31)$$

where Z is a random variable with distribution η , $\eta \in \mathcal{G}$. Let us note $m_{k,\eta} = \mathbb{E}[(Z - z_0)^k]$.

The $\{C_k(X), k \in \mathbb{N}\}$ are centered under $p_0\mu$, (μ is here the counting measure on \mathbb{N}), and are the orthogonal polynomials associated to $p_0\mu$, with square norm:

$$\|C_k\|_2^2 = z_0^k k!,$$

so that the square norm of $\frac{p_\eta - p_0}{p_0}$ in $L^2(p_0\mu)$ is:

$$\left\| \frac{p_\eta - p_0}{p_0} \right\|_2^2 = \sum_{k=1}^{\infty} \frac{(m_{k,\eta})^2}{k! z_0^k}.$$

Then \mathcal{E} is the closure of the set

$$\left\{ \frac{\sum_{k=1}^{\infty} \frac{m_{k,\eta}}{k! z_0^k} C_k(x)}{\left(\sum_{k=1}^{\infty} \frac{(m_{k,\eta})^2}{k! z_0^k} \right)^{1/2}}, \eta \in \mathcal{G} \right\}.$$

Let $(W_k)_{k \geq 1}$ be a sequence of i.i.d. standard Gaussian random variables. Then

Theorem 4. *The likelihood ratio test statistic $2 \left(\sup_{\eta \in \mathcal{G}} \ell_n(\eta) - \sup_{z \in \mathcal{Z}} \ell_n(\delta_z) \right)$ converges in distribution, as n tends to infinity, to*

$$\sup_{\eta \in \mathcal{G}} \frac{\left(\sum_{k=1}^{\infty} \frac{m_{k,\eta}}{\sqrt{k! z_0^k}} W_k \right)_+^2}{\sum_{k=1}^{\infty} \frac{(m_{k,\eta})^2}{k! z_0^k}} - W_1^2.$$

Proof: We proceed as in the proof of Theorem 3. Using the inequality

$$\left\| \frac{p_\eta - p_0}{p_0} \right\|_2^2 \geq \frac{(m_{1,\eta})^2}{z_0} + \frac{(m_{2,\eta})^2}{2z_0^2}$$

and the fact that, for some constants A and B

$$\left| \frac{p_\eta(x)}{p_0(x)} - 1 \right| \leq |m_{1,\eta}| \left| \frac{x}{z_0} - 1 \right| + A m_{2,\eta} x^2 B^x$$

one obtains that there exists a positive constant C such that for any $\eta \in \mathcal{G}$,

$$\left| \frac{p_\eta(x)}{p_0(x)} - 1 \right| \leq C x^2 B^x.$$

This proves that \mathcal{E} has a $p_0\mu$ -square integrable envelope function.

Set $f_k = \frac{C_k}{k\sqrt{k!}z_0^{k/2}}$ and

$$c_{k,\eta} = \frac{km_{k,\eta}/\sqrt{k!}z_0^{k/2}}{\left(\sum_{k=1}^{\infty} \frac{(m_{k,\eta})^2}{k!z_0^k}\right)^{1/2}}.$$

It is easy to prove that for a positive constant C that does not depend of η :

$$\sum_{k=1}^{\infty} c_{k,\eta}^2 \leq CM^4 \exp \frac{M^2}{z_0}.$$

As a consequence \mathcal{E} is Donsker since

$$\mathcal{E} \subset \left\{ \sum_{k=1}^{\infty} c_k f_k, \sum_{k=1}^{\infty} c_k^2 \leq CM^4 \exp \frac{M^2}{z_0} \right\}.$$

3.4. Application to the binomial family

Historically mixture of binomial distributions were introduced by [Lord, 1969] in the context of psychological testing: each student has a random score z and sits an n questions test. An other example is the following: consider the admixture of different human populations with different sex ratio, i.e. different probability z of having a male child, and suppose that we consider families with a fixed number of children say N . If we assume the independence between the sex of the different children of the same family, we see that the distribution of the number of males is a mixture of binomial distributions. More precisely if X is the number of males in a family "at random"

$$\mathbb{P}(X = x) = \int_0^1 \binom{N}{x} z^x (1-z)^{N-x} d\eta(z), \quad \forall x \in \{0, 1, \dots, N\} \quad (32)$$

where η is the distribution of the probability of a family to have a male in the admixture of populations. This topic of mixture of binomial distribution has received a lot of attention, see [Wood, 1999] and references therein. The problem we consider is to test whether the probability of having a male is constant in the population or not. Thus \mathcal{Z} is a closed subset of $]0, 1[$, and the conditional density with respect to the counting measure μ on the set $\{0, \dots, N\}$ is

$$p(x|z) = \binom{N}{x} z^x (1-z)^{N-x}.$$

Assumptions 4, 5 and 6 hold.

Since $p(x|z)$ is a polynomial in z with degree N ,

$$p(x|z) - p_0(x) = \sum_{k=1}^N \frac{(z - z_0)^k}{k!} p^{(k)}(x|z_0),$$

where $p^{(k)}(x|z_0)$ is the k -th derivative of $p(x|z)$ with respect to z at point z_0 , so that

$$\frac{p_\eta(x) - p_0(x)}{p_0(x)} = \sum_{k=1}^N \frac{\mathbb{E}[(Z - z_0)^k]}{k!} \frac{p^{(k)}(x|z_0)}{p_0(x)} \quad (33)$$

where Z is a random variable with distribution η .

Let $\mathbf{b} = (b_1, \dots, b_N)^T \in \mathbb{R}^N$, let $f_k(x) = \frac{1}{k!} \frac{p^{(k)}(x|z_0)}{p_0(x)}$ for any $k \in \{1, \dots, N\}$ and set $f = (f_1, \dots, f_N)^T$. It is

easy to see that

$$\langle \mathbf{b}, f \rangle = 0 \mu - a.e. \Leftrightarrow \mathbf{b} = 0.$$

Recall that $\langle \cdot, \cdot \rangle$ is the usual scalar product in Euclidean space.

Indeed, the $\frac{p^{(k)}(x|z_0)}{p_0(x)}$ are polynomial in x with degree k and leading coefficient $\frac{1}{z_0^k(1-z_0)^k}$, and the linear combination has $N+1$ zeros $(0, 1, \dots, N)$. Thus, Assumption 7 holds and if Σ is the variance matrix of $f(X_1)$, Σ is positive definite.

Let now Φ_N be the closed convex hull of the set $\{(z - z_0), (z - z_0)^2, \dots, (z - z_0)^N\}^T, z \in \mathcal{Z}$. Φ_N is in \mathcal{Z}^N the set of possible moments of $(Z - z_0)$ under a distribution with support in \mathcal{Z} . Let \mathcal{U}_N be the set of limit points of $(\frac{b_1}{(\mathbf{b}^T \Sigma \mathbf{b})^{1/2}}, \dots, \frac{b_N}{(\mathbf{b}^T \Sigma \mathbf{b})^{1/2}})^T$, as $\mathbf{b} \in \Phi_N$ tends to the null vector. Then applying Proposition 3:

$$\bar{\mathcal{D}} = \{\langle U, f \rangle, U \in \mathcal{U}_N\}.$$

Set $\ell_n(\eta) = \sum_{i=1}^n \log p_\eta(X_i)$. Let V be a N -dimensional centered Gaussian vector with variance matrix Σ . Then

Theorem 5. *The likelihood ratio test statistic $2 \left(\sup_{\eta \in \mathcal{G}} \ell_n(\eta) - \sup_{z \in \mathcal{Z}} \ell_n(\delta_z) \right)$ converges in distribution, as n tends to infinity, to*

$$\sup_{U \in \mathcal{U}_N} \langle U, V \rangle^2 1_{\langle U, V \rangle \geq 0} - V_1^2.$$

The proof of the Theorem is immediate by applying Theorem 2, but may also be obtained by classical parametric considerations. Indeed, p_η may be finitely parameterized through $m_{k,\eta}$, $k = 1, \dots, N$.

4. FINITE MIXTURE MODELS WITH OR WITHOUT STRUCTURAL PARAMETER

In this section we focus on the testing problem (4) where \mathcal{G}_j , $j = 1, 2$ is the set of finite measures with q_j supporting points, $q_1 < q_2$, so that the problem is that of testing whether the number of populations is less or equal to q_1 or at least $q_1 + 1$ but not more than q_2 . We assume that p_{θ_0} has exactly q_1 populations (the supporting points z_i^0 , $i = 1, \dots, q_1$, are distinct and the weights π_i^0 , $i = 1, \dots, q_1$, are non null).

We shall prove that the LRT statistic converges in distribution to Λ as stated in (8) under natural assumptions, and propose a way to compute Λ .

Weak identifiability of the mixtures is a minimal assumption:

Assumption 9.

$$\sum_{i=1}^r \pi_i p_\gamma(\cdot | z_i) = \sum_{i=1}^{\tilde{r}} \tilde{\pi}_i p_{\tilde{\gamma}}(\cdot | \tilde{z}_i) \quad \mu\text{-a.e.}$$

if and only if $\gamma = \tilde{\gamma}$ and the mixing measures are equal: $\sum_{i=1}^r \pi_i \delta_{z_i} = \sum_{i=1}^{\tilde{r}} \tilde{\pi}_i \delta_{\tilde{z}_i}$.

We consider the class $\mathcal{F}_j = \{p_\theta, \theta \in \Theta_j\}$, $j = 1, 2$, where p_θ is defined by (2) and the parameter sets can be written as

$$\Theta_j = \mathcal{H} \times \Pi_j \times \mathcal{Z}^{q_j}$$

where \mathcal{H} denotes a subset of \mathbb{R}^h , $\Pi_j = \{(\pi_1, \dots, \pi_{q_j}), \pi_i \in (0, 1), \sum_{i=1}^{q_j} \pi_i = 1\}$ and \mathcal{Z} is a compact subset of \mathbb{R}^k for k the common dimension of the z_i^0 .

Assumption 10. *We assume that (i) or (ii) is satisfied:*

- (i) \mathcal{H} is compact.
- (ii) Under p_θ with $\theta \in \Theta_1$ or under p_θ with $\theta \in \Theta_2$ the maximum likelihood estimator of γ is weakly consistent.

If Assumption (10) (ii) holds, one may reduce the parameter space to a compact neighborhood of the true parameter when studying the asymptotics of the LRT statistics, in the same way as for the proof of Theorem 1. If Assumption (10) (i) is assumed, it is useful only to prove that the maximum likelihood estimator of γ is consistent, which is not always necessary for that, see for instance the Gaussian case with the variance as structural parameter.

4.1. Sets of scores

Consider $(J_i)_{i=0,\dots,q_1}$ a partition of $\{1, \dots, q_2\}$. Let $|J_0|$ denotes the cardinal of J_0 . For $\boldsymbol{\pi} = (\pi_{i,j})_{1 \leq i \leq q_1, j \in J_i} \in \mathbb{R}^{q_2 - |J_0|}$, $\gamma \in \mathcal{H}$, $\boldsymbol{\alpha} = (\alpha_{i,j})_{1 \leq i \leq q_1, j \in J_i} \in \mathbb{R}^{q_2 - |J_0|}$, $\mathbf{u} = (u_{i,j})_{i \leq q_1, j \in J_i} \in \mathbb{R}^{k(q_2 - |J_0|)}$, $\boldsymbol{\rho} = (\rho_j)_{j \in J_0} \in \mathbb{R}^{|J_0|}$, and $\mathbf{z} \in \mathcal{Z}^{|J_0|}$, we set $\mathbf{v} = (\gamma, \boldsymbol{\alpha}, \mathbf{u}, \boldsymbol{\rho})$ and

$$p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}}(\cdot) = \sum_{i=1}^{q_1} \left(\sum_{j \in J_i} (\pi_{i,j} + \alpha_{i,j}) p_{\gamma^0 + \gamma}(\cdot | z_i^0 + u_{i,j}) \right) + \sum_{j \in J_0} \rho_j p_{\gamma^0 + \gamma}(\cdot | z_j). \quad (34)$$

Assumption 11. *The map $\theta \mapsto p_\theta$ is continuously differentiable on Θ_2 .*

Moreover, $R(\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}) = \partial_{\mathbf{v}} p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}} / \sqrt{p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}} p_{\theta_0}}$ has $p_{\theta_0} \mu$ -square integrable coordinates, and

$$\limsup_{\mathbf{v} \rightarrow \mathbf{0}} \frac{\|\langle \mathbf{v}, \int_0^1 (R(\boldsymbol{\pi}, t\mathbf{v}, \mathbf{z}) - R(\boldsymbol{\pi}, \mathbf{0}, \mathbf{z})) dt \rangle\|_2}{\|\langle \mathbf{v}, R(\boldsymbol{\pi}, \mathbf{0}, \mathbf{z}) \rangle\|_2} = 0,$$

where we recall that $\|\cdot\|_2$ is the norm in $L^2(p_{\theta_0} \mu)$.

Here, $\partial_{\mathbf{v}} p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}}$ denotes the partial derivative of $p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}}$ with respect to the variable \mathbf{v} .

An application of Lemma 7.6 of [van der Vaart, 1998] gives that, if Assumption 11 holds, the model $(p_{\boldsymbol{\pi}, t\mathbf{v}, \mathbf{z}})_{t \geq 0}$ is differentiable in quadratic mean at $t = 0$ with score function $\langle \mathbf{v}, R(\boldsymbol{\pi}, \mathbf{0}, \mathbf{z}) \rangle$. Thus, we may define the sets \mathcal{S}_j such that Assumption 2 holds for \mathcal{S}_j in model \mathcal{F}_j , $j = 1, 2$ as follows.

Let $\dot{\ell}_{\gamma^0, z}$ be the score function in the model $(p_\gamma(\cdot | z) \mu)_{\gamma \in \Gamma, z \in \mathcal{Z}}$ at point (γ^0, z) for $z \in \mathcal{Z}$. Let $\dot{\ell}_{\gamma^0, [z]} = (\dot{\ell}_{\gamma^0, z})_{1, \dots, h}$ be the score function in the model $(p_\gamma(\cdot | z) \mu)_{\gamma \in \mathcal{H}}$ at point γ^0 . Let $\dot{\ell}_{[\gamma^0], z} = (\dot{\ell}_{\gamma^0, z})_{h+1, \dots, h+k}$ be the score function in the model $(p_{\gamma^0}(\cdot | z) \mu)_{z \in \mathcal{Z}}$ at point z .

Let $\dot{\ell}_0$ be the vector obtained by concatenation of $\left\{ \pi_i^0 \dot{\ell}_{[\gamma^0], z_i^0} \frac{p_{\gamma^0}(\cdot | z_i^0)}{p_{\theta_0}(\cdot)} \right\}_{i=1, \dots, q_1}$ and $\sum_{i=1}^{q_1} \pi_i^0 \dot{\ell}_{\gamma^0, [z_i^0]} \frac{p_{\gamma^0}(\cdot | z_i^0)}{p_{\theta_0}(\cdot)}$.

Denote by q_0 its dimension: $q_0 = kq_1 + h$. Then since for all i , $|\frac{\pi_i^0 p_{\gamma^0}(\cdot | z_i^0)}{p_{\theta_0}(\cdot)}| \leq 1$, $\dot{\ell}_0 \in (L^2(p_{\theta_0} \mu))^{q_0}$.

For any $U \in \mathbb{R}^{q_0}$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{q_1}) \in \mathbb{R}^{q_1}$, $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)$ where ρ_1, \dots, ρ_q are non negative real numbers such that $\sum_{i=1}^{q_1} \alpha_i + \sum_{i=1}^q \rho_i = 0$, and for any $\mathbf{z} = (z_1, \dots, z_q) \in \mathcal{Z}^q$, define

$$s_2(U, \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{z}) = \langle U, \dot{\ell}_0(\cdot) \rangle + \frac{\sum_{i=1}^{q_1} \alpha_i p_{\gamma^0}(\cdot | z_i^0) + \sum_{i=1}^q \rho_i p_{\gamma^0}(\cdot | z_i)}{p_{\theta_0}(\cdot)}, \quad (35)$$

and

$$s_1(U, \boldsymbol{\alpha}) = s_2(U, \boldsymbol{\alpha}, \mathbf{0}, \cdot) = \langle U, \dot{\ell}_0(\cdot) \rangle + \frac{\sum_{i=1}^{q_1} \alpha_i p_{\gamma^0}(\cdot | z_i^0)}{p_{\theta_0}(\cdot)}, \quad (36)$$

with the convention $s_2(U, \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{z}) = s_1(U, \boldsymbol{\alpha}) = 0$ on $\{p_{\theta_0} = 0\}$. Let now

$$\mathcal{S}_2 = \left\{ \frac{s_2(U, \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{z})}{\|s_2(U, \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{z})\|_2}, U, \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{z} : \|s_2(U, \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{z})\|_2 \neq 0 \right\} \quad (37)$$

and

$$\mathcal{S}_1 = \left\{ \frac{s_1(U, \boldsymbol{\alpha})}{\|s_1(U, \boldsymbol{\alpha})\|_2}, U, \boldsymbol{\alpha} : \|s_1(U, \boldsymbol{\alpha})\|_2 \neq 0 \right\}. \quad (38)$$

Indeed, using Assumption 11, for any $z \in \mathcal{Z}$, $\frac{p_{\gamma^0}(\cdot|z)}{p_{\theta_0}(\cdot)} \in L^2(p_{\theta_0}\mu)$. Since $L^2(p_{\theta_0}\mu)$ is closed under linear combinations, we notice that for any $U, \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{z}$ satisfying previous conditions, the function $s_2(U, \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{z}) \in L^2(p_{\theta_0}\mu)$.

Let also for $j = 1, 2$, the set \mathcal{D}_ϵ^j be defined by (12) for the model \mathcal{F}_j and \mathcal{D}_j be the closure of \mathcal{S}_j in $L^2(p_0\mu)$. Then

Proposition 4. *Under Assumptions 9, 10 and 11, we have $\mathcal{D}_j = \bigcap_{\epsilon \leq \epsilon_0} \overline{\mathcal{D}_\epsilon^j}$ for $j = 1, 2$.*

Assumption 12. *There exists $m \in L^2(p_{\theta_0}\mu)$ and $\beta > 0$ such that*

$$\left| \frac{\int_0^1 \langle \mathbf{v}_1, R(\boldsymbol{\pi}_1, t\mathbf{v}_1, \mathbf{z}_1) \rangle dt}{\left\| \int_0^1 \langle \mathbf{v}_1, R(\boldsymbol{\pi}_1, t\mathbf{v}_1, \mathbf{z}_1) \rangle dt \right\|_2} - \frac{\int_0^1 \langle \mathbf{v}_2, R(\boldsymbol{\pi}_2, t\mathbf{v}_2, \mathbf{z}_2) \rangle dt}{\left\| \int_0^1 \langle \mathbf{v}_2, R(\boldsymbol{\pi}_2, t\mathbf{v}_2, \mathbf{z}_2) \rangle dt \right\|_2} \right| \leq m \|(\boldsymbol{\pi}_1, \mathbf{v}_1, \mathbf{z}_1) - (\boldsymbol{\pi}_2, \mathbf{v}_2, \mathbf{z}_2)\|^\beta.$$

Proposition 5. *Under Assumptions 10, 11, 12, Assumption 3 holds for models \mathcal{F}_1 and \mathcal{F}_2 .*

Thus, Assumptions 9, 10, 11 and 12 are sufficient to apply Theorem 1 in order to derive the asymptotic distribution of the LRT statistic. This will be done in the next section.

Remark 1. *Assume*

Assumption 13.

- *The model $(p_\gamma(\cdot|z)\mu)_{\gamma \in \Gamma, z \in \mathcal{Z}}$ is differentiable in quadratic mean at points (γ^0, z) for $z \in \mathcal{Z}$.*
- *The model $((1-t)p_{\theta_0} + tp_{\gamma^0}(\cdot|z)\mu)_{t \geq 0}$ is differentiable in quadratic mean at $t = 0$ for any $z \in \mathcal{Z}$.*

Then using the fact that the set of models that are differentiable in quadratic mean is closed under convex combinations, we can prove that Assumption 2 holds for \mathcal{S}_j in model \mathcal{F}_j for $j = 1, 2$. In some cases this is sufficient to prove that the LRT statistic converges to infinity if the isonormal process on \mathcal{D}_2 is unbounded as explained in Section 2.

4.2. Asymptotic distribution of the LRT statistic

Last results gave conditions under which the asymptotic distribution of the LRT statistic may be derived. In this part, we simplify considerably its form using only algebraic arguments. Indeed, the LRT statistic Λ_n defined by (7) can obviously be written as

$$\Lambda_n = 2 \left(\sup_{\theta \in \Theta_2} \ell_n(\theta) - \ell_n(\theta_0) \right) - 2 \left(\sup_{\theta \in \Theta_1} \ell_n(\theta) - \ell_n(\theta_0) \right)$$

so that using Propositions 4, 5 and applying twice Theorem 1, we obtain under Assumptions 9, 10, 11 and 12:

$$\Lambda_n = \sup_{d \in \mathcal{D}_2} \left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n d(X_i) \right)^2 1_{\sum_{i=1}^n d(X_i) \geq 0} \right] - \sup_{d \in \mathcal{D}_1} \left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n d(X_i) \right)^2 1_{\sum_{i=1}^n d(X_i) \geq 0} \right] + o_{\mathbb{P}}(1).$$

Consequently Λ_n converges in distribution to

$$\Lambda = \sup_{\mathbf{z} \in \mathcal{Z}^q} \sup_{\substack{U, \boldsymbol{\alpha}, \boldsymbol{\rho} \geq 0 \\ \sum_i \alpha_i + \sum_i \rho_i = 0 \\ B(U, \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{z}) = 1}} \left[\left(\langle U, V \rangle + \sum_{i=2}^{q_1} \alpha_i W(z_i^0) + \sum_{i=1}^q \rho_i W(z_i) \right)_+ \right]^2 - \sup_{\substack{U, \boldsymbol{\alpha} \\ \sum_i \alpha_i = 0 \\ B(U, \boldsymbol{\alpha}, \mathbf{0}, \cdot) = 1}} \left[\left(\langle U, V \rangle + \sum_{i=2}^{q_1} \alpha_i W(z_i^0) \right)_+ \right]^2$$

where W is the centered Gaussian process on \mathcal{Z} with covariance function $\Gamma(z, z') = \int \frac{p_{\gamma^0}(x|z)p_{\gamma^0}(x|z')}{p_{\theta_0}(x)} d\mu(x) - 1$, V is the centered Gaussian vector with variance $\Sigma = \int \dot{\ell}_0(x)\dot{\ell}_0(x)^T p_{\theta_0}(x) d\mu(x)$ and covariance with $W(z)$ given by $C(z) = \int p_{\gamma^0}(x|z)\dot{\ell}_0(x) d\mu(x)$, and where $B(U, \boldsymbol{\alpha}, \boldsymbol{\rho}; \mathbf{z}) = \text{Var}(\langle U, V \rangle + \sum_{i=1}^{q_1} \alpha_i W(z_i^0) + \sum_{i=1}^q \rho_i W(z_i))$ which is a quadratic form in $U, (\alpha_i), (\rho_i)$.

In order to present a simplified form of Λ , we need to introduce the notations $\kappa(z) = C(z) - C(z_1^0)$ and $\Delta(z_i, z_j) = \Gamma(z_i, z_j) - \Gamma(z_i, z_1^0) - \Gamma(z_j, z_1^0) + \Gamma(z_1^0, z_1^0)$. For any \mathbf{z} and \mathbf{z}' in \mathcal{Z}^q , we set

$$A_q(\mathbf{z}, \mathbf{z}') = \Gamma_q(\mathbf{z}, \mathbf{z}') - C_q(\mathbf{z})^T \Upsilon_\emptyset^{-1} C_q(\mathbf{z}') \quad (39)$$

where $\Gamma_q(\mathbf{z}, \mathbf{z}') = [\Delta(z_i, z_j')]_{\substack{i=1, \dots, q \\ j=1, \dots, q}}$,

$$\Upsilon_\emptyset = \begin{bmatrix} \Sigma & \kappa(z_2^0) & \dots & \kappa(z_{q_1}^0) \\ \kappa(z_2^0)^T & \Delta(z_2^0, z_2^0) & \dots & \Delta(z_{q_1}^0, z_2^0) \\ \vdots & \vdots & & \vdots \\ \kappa(z_{q_1}^0)^T & \Delta(z_2^0, z_{q_1}^0) & \dots & \Delta(z_{q_1}^0, z_{q_1}^0) \end{bmatrix} \quad \text{and} \quad C_q(\mathbf{z}) = \begin{bmatrix} \kappa(z_1) & \dots & \kappa(z_q) \\ \Delta(z_1, z_2^0) & \dots & \Delta(z_q, z_2^0) \\ \vdots & & \vdots \\ \Delta(z_1, z_{q_1}^0) & \dots & \Delta(z_q, z_{q_1}^0) \end{bmatrix}.$$

Remark that $\Gamma_q(\mathbf{z}, \mathbf{z}') \in \mathcal{M}_q(\mathbb{R})$, $\Upsilon_\emptyset \in \mathcal{M}_{q_0+q_1-1}(\mathbb{R})$ and $C_q(\mathbf{z}) \in \mathcal{M}_{(q_0+q_1-1) \times q}(\mathbb{R})$, so that $A_q(\mathbf{z}, \mathbf{z}') \in \mathcal{M}_q(\mathbb{R})$. Moreover, since $A_q(\mathbf{z}, \mathbf{z})$ is definite positive, we can introduce the notation $A_q(\mathbf{z}, \mathbf{z})^{-1/2}$.

Theorem 6. *Under Assumptions 9, 10, 11 and 12, the LRT statistic converges in distribution to*

$$\Lambda = \sup_{\mathbf{z} \in \mathcal{Z}^q} \left\{ \|Y_q(\mathbf{z})\|^2 \mathbf{1}_{A_q(\mathbf{z}, \mathbf{z})^{-1/2} Y_q(\mathbf{z}) \geq 0} \right\}$$

where $\|\cdot\|$ denotes the norm in \mathbb{R}^q , Y_q is the q -dimensional centered Gaussian field with covariance function $\mathbb{E}[Y_q(\mathbf{z})Y_q(\mathbf{z}')^T] = (A_q(\mathbf{z}, \mathbf{z})^{-1/2})^T A_q(\mathbf{z}, \mathbf{z}') A_q(\mathbf{z}', \mathbf{z}')^{-1/2}$ and where the condition of positivity means that all coordinates of the vector have to be non negative.

Proof: Remark that the supremum inside the brackets of Λ is always positive (use the particular case when U is parallel to V and the other terms vanish). To get rid of the conditions $\sum_i \alpha_i + \sum_i \rho_i = 0$ and $\sum_i \alpha_i = 0$, last formula can be written (with the trivial convention that the sum vanishes when $q_1 = 1$)

$$\Lambda = \sup_{\mathbf{z} \in \mathcal{Z}^q} \left[\sup_{\substack{U, \boldsymbol{\alpha}, \boldsymbol{\rho} \geq 0 \\ B(U, \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{z})=1}} \left(\langle U, V \rangle + \sum_{i=2}^{q_1} \alpha_i (W(z_i^0) - W(z_1^0)) + \sum_{i=1}^q \rho_i (W(z_i) - W(z_1^0)) \right) \right]^2 - \left[\sup_{B(U, \boldsymbol{\alpha}, \mathbf{0}; \dots)=1} \left(\langle U, V \rangle + \sum_{i=2}^{q_1} \alpha_i (W(z_i^0) - W(z_1^0)) \right) \right]^2. \quad (40)$$

For $J \subset \{1, \dots, q\}$ and for $\mathbf{z} \in \mathcal{Z}^q$ define $\mathbf{W}_J(\mathbf{z})$ as the vector that consists of the $|J|$ coordinates $W(z_j) - W(z_1^0)$ for $j \in J$ (ordered as in J). Denoting by $:$ the concatenation, define also the vector

$$\mathcal{W}_J^T(\mathbf{z}) = (V^T : W(z_2^0) - W(z_1^0), \dots, W(z_{q_1}^0) - W(z_1^0) : \mathbf{W}_J(\mathbf{z})^T)$$

and $\Upsilon_J(\mathbf{z}) = \text{Var}(\mathcal{W}_J(\mathbf{z}))$. To simplify the notations, set $\mathbf{W}_q = \mathbf{W}_{\{1, \dots, q\}}$, $\mathcal{W}_q := \mathcal{W}_{\{1, \dots, q\}}$ and $\Upsilon_q = \Upsilon_{\{1, \dots, q\}}$. Then the first inner term in (40) satisfies

$$\sup_{B(U, \alpha, \rho, z_1, \dots, z_q)=1} \left(\langle U, V \rangle + \sum_{i=2}^{q_1} \alpha_i (W(z_i^0) - W(z_1^0)) + \sum_{i=1}^q \rho_i (W(z_i) - W(z_1^0)) \right) = \sup_{\substack{\tau \in \mathbb{R}^{q_0+q_1-1} \times \mathbb{R}_+^q \\ \tau^T \Upsilon_q \tau = 1}} \langle \tau, \mathcal{W}_q(\mathbf{z}) \rangle.$$

Remark that for any J subset of $\{1, \dots, q\}$ and any $\mathbf{z} \in \mathcal{Z}^q$:

$$M_J(\mathbf{z}) := \left(\sup_{\substack{\tau \in \mathbb{R}^{q_0+q_1+|J|-1} \\ \tau^T \Upsilon_J(\mathbf{z}) \tau = 1}} \langle \tau, \mathcal{W}_J(\mathbf{z}) \rangle \right)^2 = \mathcal{W}_J^T(\mathbf{z}) \Upsilon_J^{-1}(\mathbf{z}) \mathcal{W}_J(\mathbf{z}),$$

and that the supremum is attained for $\tau = (\text{const}) \Upsilon_J^{-1}(\mathbf{z}) \mathcal{W}_J(\mathbf{z})$, where (const) is some positive constant. Let $D_J(\mathbf{z})$ be the condition

$$D_J(\mathbf{z}) = \text{"The last } |J| \text{ coordinates of } \Upsilon_J^{-1}(\mathbf{z}) \mathcal{W}_J(\mathbf{z}) \text{ are non-negative"},$$

and set again $D_q = D_{\{1, \dots, q\}}$. Then considering the location where the maximum is attained, we easily get:

$$\left(\sup_{\substack{\tau \in \mathbb{R}^{q_0+q_1-1} \times \mathbb{R}_+^q \\ \tau^T \Upsilon_q \tau = 1}} \langle \tau, \mathcal{W}_q(\mathbf{z}) \rangle \right)^2 = \max_{J \subset \{1, \dots, q\}} M_J(\mathbf{z}) 1_{D_J(\mathbf{z})}.$$

Observe now that $M_J(\mathbf{z}) 1_{D_J(\mathbf{z})}$ is the value at the point $(z_j)_{j \in J}$ of some function depending only on $|J|$. Taking the supremum over $\mathbf{z} = (z_1, \dots, z_q) \in \mathcal{Z}^q$, as it is required by Formula (40), we deduce that the contributions of $M_J(\mathbf{z}) 1_{D_J(\mathbf{z})}$ and $M_{|J|}(\mathbf{z}) 1_{D_{|J|}(\mathbf{z})}$ are the same. So, with the convention that $M_0 1_{D_0} = M_\emptyset$, we get

$$\sup_{\mathbf{z} \in \mathcal{Z}^q} \max_{J \subset \{1, \dots, q\}} M_J(\mathbf{z}) 1_{D_J(\mathbf{z})} = \sup_{\mathbf{z} \in \mathcal{Z}^q} \max_{k=0, \dots, q} M_k(\mathbf{z}) 1_{D_k(\mathbf{z})}.$$

For $k \in \{2, \dots, q\}$, let $(\tilde{z}_1, \dots, \tilde{z}_{k-1}) \in \mathcal{Z}^{k-1}$ and fix $\tilde{z}_k = z_1^0$. Then $M_k(\tilde{\mathbf{z}}) = M_{k-1}(\tilde{\mathbf{z}})$ and the conditions $D_k(\tilde{\mathbf{z}})$ and $D_{k-1}(\tilde{\mathbf{z}})$ are equivalent. Thus $\sup_{\mathbf{z} \in \mathcal{Z}^q} M_k(\mathbf{z}) 1_{D_k(\mathbf{z})} \geq \sup_{\mathbf{z} \in \mathcal{Z}^q} M_{k-1}(\mathbf{z}) 1_{D_{k-1}(\mathbf{z})}$, and consequently

$$\sup_{\mathbf{z} \in \mathcal{Z}^q} \max_{k=0, \dots, q} M_k(\mathbf{z}) 1_{D_k(\mathbf{z})} = \sup_{\mathbf{z} \in \mathcal{Z}^q} \max(M_q(\mathbf{z}) 1_{D_q(\mathbf{z})}, M_\emptyset).$$

On the other hand, we can write

$$\Upsilon_q(\mathbf{z}) = \begin{bmatrix} \Upsilon_\emptyset & C_q(\mathbf{z}) \\ C_q(\mathbf{z})^T & \Gamma_q(\mathbf{z}, \mathbf{z}) \end{bmatrix}$$

where $C_q(\mathbf{z})$ is the covariance of \mathcal{W}_\emptyset with $\mathbf{W}_q(\mathbf{z})$ and $\Gamma_q(\mathbf{z}, \mathbf{z})$ is the variance of $\mathbf{W}_q(\mathbf{z})$. It is well known that

$$\Upsilon_q(\mathbf{z})^{-1} = \begin{bmatrix} \Upsilon_\emptyset^{-1} + \Upsilon_\emptyset^{-1} C_q(\mathbf{z}) A_q(\mathbf{z}, \mathbf{z})^{-1} C_q(\mathbf{z})^T \Upsilon_\emptyset^{-1} & -\Upsilon_\emptyset^{-1} C_q(\mathbf{z}) A_q(\mathbf{z}, \mathbf{z})^{-1} \\ -A_q(\mathbf{z}, \mathbf{z})^{-1} C_q(\mathbf{z})^T \Upsilon_\emptyset^{-1} & A_q(\mathbf{z}, \mathbf{z})^{-1} \end{bmatrix}$$

where $A_q(\mathbf{z}, \mathbf{z})$ can be deduced from (39). Some calculations show that

$$M_q(\mathbf{z}) - M_\emptyset = (\mathbf{W}_q(\mathbf{z}) - C_q(\mathbf{z})^T \Upsilon_\emptyset^{-1} \mathcal{W}_\emptyset)^T [A_q(\mathbf{z}, \mathbf{z})^{-1} (\mathbf{W}_q(\mathbf{z}) - C_q(\mathbf{z})^T \Upsilon_\emptyset^{-1} \mathcal{W}_\emptyset)]$$

and also that

$$D_q(\mathbf{z}) = \text{''All the coordinates of } [A_q(\mathbf{z}, \mathbf{z})^{-1} (\mathbf{W}_q(\mathbf{z}) - C_q(\mathbf{z})^T \Upsilon_\emptyset^{-1} \mathcal{W}_\emptyset)] \text{ are non-negative''}.$$

Then remarking that $\max(M_q(\mathbf{z})1_{D_q(\mathbf{z})}, M_\emptyset)$ can be written as $M_\emptyset + (M_q(\mathbf{z}) - M_\emptyset)1_{D_q(\mathbf{z})}$, we obtain

$$\Lambda = \sup_{\mathbf{z} \in \mathcal{Z}^q} \left\{ (\mathbf{W}_q(\mathbf{z}) - C_q(\mathbf{z})^T \Upsilon_\emptyset^{-1} \mathcal{W}_\emptyset)^T A_q(\mathbf{z}, \mathbf{z})^{-1} (\mathbf{W}_q(\mathbf{z}) - C_q(\mathbf{z})^T \Upsilon_\emptyset^{-1} \mathcal{W}_\emptyset) 1_{A_q(\mathbf{z}, \mathbf{z})^{-1} (\mathbf{W}_q(\mathbf{z}) - C_q(\mathbf{z})^T \Upsilon_\emptyset^{-1} \mathcal{W}_\emptyset) \geq 0} \right\}.$$

To conclude, it suffices to define Y_q as the field $(A_q(\cdot, \cdot)^{-1/2})^T (\mathbf{W}_q(\cdot) - C_q(\cdot)^T \Upsilon_\emptyset^{-1} \mathcal{W}_\emptyset)$.

4.3. Application to the Gaussian family

In this part, we apply our results for testing homogeneity on Gaussian mixture in \mathbb{R} and \mathbb{R}^2 . We present the computation of elements that characterizes the asymptotic distribution of the LRT statistic (according to Theorem 6). Notice that under homogeneity ($q_1 = 1$), we have $\Upsilon_\emptyset = \Sigma$, $\kappa(z) = C(z)$ for all $z \in \mathcal{Z}$, $C_q(\mathbf{z}) = (C(z_1) \dots C(z_q))$ for all $\mathbf{z} = (z_1, \dots, z_q) \in \mathcal{Z}^q$.

4.3.1. Testing homogeneity for univariate Gaussian distributions

We consider the family of Gaussian densities in \mathbb{R} of mean z and variance γ , denoted by $p_\gamma(\cdot|z)$. We assume that the true distribution is $p_{\gamma_0}(\cdot|z_0)$. We set for simplicity $\gamma = \sigma^2$ so that

$$p_\gamma(x|z) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-z)^2}{2\sigma^2}\right), \forall x \in \mathbb{R},$$

for $\sigma \in \mathcal{H}$ a subset of $]0, +\infty[$ and $z \in \mathcal{Z}$ a compact subset of \mathbb{R} . It can be proved that Assumptions 9, 10, 11 and 12 are satisfied. In this framework, the centered Gaussian process W has for covariance function

$$\Gamma(z, z') = \exp\left(\frac{z_0^2 + zz' - zz_0 - z'z_0}{\sigma_0^2}\right) - 1 = \exp(\tilde{z}\tilde{z}') - 1,$$

with $\tilde{z} = (z - z_0)/\sigma_0$ and $\tilde{z}' = (z' - z_0)/\sigma_0$. By regularity of the model, we easily compute that:

$$\dot{\ell}_0 = \left(\dot{\ell}_{[\gamma_0], z_0}^T, \dot{\ell}_{\gamma_0, [z_0]}^T\right)^T = \left(\frac{x - z_0}{\sigma_0^2}, \frac{(x - z_0)^2}{\sigma_0^3} - \frac{1}{\sigma_0}\right)^T.$$

We deduce that V is the centered Gaussian vector with variance Σ and covariance $C(z)$ with $W(z)$ given by

$$\Sigma = \frac{1}{\sigma_0^2} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad C(z)^T = \left(\frac{z - z_0}{\sigma_0^2}, \frac{(z - z_0)^2}{\sigma_0^3}\right).$$

Eventually, we obtain with obvious notations

$$A_q(\mathbf{z}, \mathbf{z}') = \left[\exp(\tilde{z}_i \tilde{z}'_j) - 1 - \tilde{z}_i \tilde{z}'_j - \frac{1}{2} (\tilde{z}_i \tilde{z}'_j)^2 \right]_{\substack{i=1, \dots, q \\ j=1, \dots, q}}.$$

In the particular case of $q = 1$ we find that

$$\Lambda = \sup_{z \in \mathcal{Z}} \{Y^2(z)1_{Y(z) \geq 0}\}$$

with

$$\mathbb{E}(Y(z)Y(z')) = \frac{\exp(\tilde{z}\tilde{z}') - 1 - \tilde{z}\tilde{z}' - \frac{1}{2}(\tilde{z}\tilde{z}')^2}{\sqrt{\exp(\tilde{z}^2) - 1 - \tilde{z}^2 - \frac{1}{2}\tilde{z}^4} \sqrt{\exp((\tilde{z}')^2) - 1 - (\tilde{z}')^2 - \frac{1}{2}(\tilde{z}')^4}}.$$

This expression corresponds to that conjectured by Garel [Garel, 2001] or obtained by Delmas [Delmas, 2003]. Looking to the properties of the process Y around zero, we see as in [Azais et al., 2006] that the indicator function can be removed.

4.3.2. Testing homogeneity for Bivariate Gaussian distributions

We consider \mathcal{Z} a compact subset of \mathbb{R}^2 and \mathcal{H} a subset of the set of the 2×2 positive definite matrices. For $Z \in \mathcal{Z}$, we set $Z = (z_x, z_y)^T$ and we write $\gamma \in \mathcal{H}$ as follow:

$$\gamma = \begin{bmatrix} \sigma_x^2 & r\sigma_x\sigma_y \\ r\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}.$$

Consequently, we consider the parametric family of Gaussian densities given by

$$p_\gamma(X|Z) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp\left(-\frac{\sigma_y^2(x-z_x)^2 - 2r\sigma_x\sigma_y(x-z_x)(y-z_y) + \sigma_x^2(y-z_y)^2}{2\sigma_x^2\sigma_y^2(1-r^2)}\right), \forall X = (x, y) \in \mathbb{R}^2.$$

We assume that the true distribution is $p_0 = p_{\gamma_0}(\cdot|Z^0)$ with $Z^0 = (z_{x,0}, z_{y,0})^T$ and $\gamma^0 = \begin{bmatrix} \sigma_{x,0}^2 & r_0\sigma_{x,0}\sigma_{y,0} \\ r_0\sigma_{x,0}\sigma_{y,0} & \sigma_{y,0}^2 \end{bmatrix}$. In that case, the centered Gaussian process W has for covariance function

$$\Gamma(Z, Z') = \exp\left[-\frac{1}{2}(Z - Z^0)^T \gamma_0^{-1} (Z - Z^0) - \frac{1}{2}(Z' - Z^0)^T \gamma_0^{-1} (Z' - Z^0)\right] - 1.$$

In order to describe $\dot{\ell}_0$ we introduce, according to the parameters $(z_x, z_y, \sigma_x, \sigma_y, r)$, the following functions:

$$\begin{aligned} \ell_{1,Z,\gamma}(X) &= \frac{(x-z_x)}{\sigma_x^2(1-r^2)} - \frac{r(y-z_y)}{\sigma_x\sigma_y(1-r^2)}, & \ell_{2,Z,\gamma}(X) &= \frac{(y-z_y)}{\sigma_y^2(1-r^2)} - \frac{r(x-z_x)}{\sigma_x\sigma_y(1-r^2)}, \\ \ell_{3,Z,\gamma}(X) &= -\frac{1}{\sigma_x} \left(1 - \frac{1}{1-r^2} \left\{ \frac{(x-z_x)^2}{\sigma_x^2} - \frac{r(x-z_x)(y-z_y)}{\sigma_x\sigma_y} \right\}\right), \\ \ell_{4,Z,\gamma}(X) &= -\frac{1}{\sigma_y} \left(1 - \frac{1}{1-r^2} \left\{ \frac{(y-z_y)^2}{\sigma_y^2} - \frac{r(x-z_x)(y-z_y)}{\sigma_x\sigma_y} \right\}\right), \\ \ell_{5,Z,\gamma}(X) &= \frac{r}{1-r^2} - \frac{r}{(1-r^2)^2} \left[\frac{(x-z_x)^2}{\sigma_x^2} + \frac{(y-z_y)^2}{\sigma_y^2} \right] + \frac{1+r^2}{(1-r^2)^2} \left(\frac{x-z_x}{\sigma_x} \right) \left(\frac{y-z_y}{\sigma_y} \right). \end{aligned}$$

Since $\dot{\ell}_{[\gamma^0], Z^0} = (\ell_{1,Z^0,\gamma^0}, \ell_{2,Z^0,\gamma^0})^T$, $\dot{\ell}_{\gamma^0, [Z^0]} = (\ell_{3,Z^0,\gamma^0}, \ell_{4,Z^0,\gamma^0}, \ell_{5,Z^0,\gamma^0})^T$ and $\dot{\ell}_0 = \left(\dot{\ell}_{[\gamma^0], z_0}, \dot{\ell}_{\gamma^0, [z_0]} \right)^T$ we deduce that the Gaussian vector V has for variance matrix:

$$\Sigma = \frac{1}{(1-r_0^2)\sigma_{x,0}^2\sigma_{y,0}^2} \begin{bmatrix} \sigma_{y,0}^2 & -r_0\sigma_{x,0}\sigma_{y,0} & 0 & 0 & 0 \\ -r_0\sigma_{x,0}\sigma_{y,0} & \sigma_{x,0}^2 & 0 & 0 & 0 \\ 0 & 0 & (2-r_0^2)\sigma_{y,0}^2 & -r_0^2\sigma_{x,0}\sigma_{y,0} & -r_0\sigma_{x,0}\sigma_{y,0}^2 \\ 0 & 0 & -r_0^2\sigma_{x,0}\sigma_{y,0} & (2-r_0^2)\sigma_{x,0}^2 & -r_0\sigma_{x,0}^2\sigma_{y,0} \\ 0 & 0 & -r_0\sigma_{x,0}\sigma_{y,0}^2 & -r_0\sigma_{x,0}^2\sigma_{y,0} & \frac{(1+r_0^2)\sigma_{x,0}^2\sigma_{y,0}^2}{(1-r_0^2)} \end{bmatrix}$$

and covariance $C(Z)$ with $W(Z)$:

$$C(Z) = \frac{1}{(1-r_0^2)\sigma_{x,0}^2\sigma_{y,0}^2} \begin{pmatrix} (z_x - z_{x,0})\sigma_{y,0}^2 - r_0(z_y - z_{y,0})\sigma_{x,0}\sigma_{y,0} \\ (z_y - z_{y,0})\sigma_{x,0}^2 - r_0(z_x - z_{x,0})\sigma_{x,0}\sigma_{y,0} \\ (z_x - z_{x,0})^2\sigma_{x,0}^{-1}\sigma_{y,0}^2 - r_0(z_x - z_{x,0})(z_y - z_{y,0})\sigma_{y,0} \\ (z_y - z_{y,0})^2\sigma_{x,0}^2\sigma_{y,0}^{-1} - r_0(z_x - z_{x,0})(z_y - z_{y,0})\sigma_{x,0} \\ \frac{1+r_0^2}{1-r_0^2}\sigma_{x,0}\sigma_{y,0}(z_x - z_{x,0})(z_y - z_{y,0}) - \frac{r_0}{1-r_0^2}\sigma_{x,0}^2(z_y - z_{y,0})^2 - \frac{r_0}{1-r_0^2}\sigma_{y,0}^2(z_x - z_{x,0})^2 \end{pmatrix}.$$

Then Formula (39) allows the computation of A_q .

5. PROOFS

Proof of Proposition 4 Consider some $U, \alpha, \rho, \mathbf{z}$ such that $\|s_2(U, \alpha, \rho, \mathbf{z})\|_2 \neq 0$. Let

$$p_t(\cdot) = \sum_{i=1}^{q_1} (\pi_i^0 + t\alpha_i) p_{\gamma_t}(\cdot | z_{i,t}) + t \sum_{j=1}^q \rho_j p_{\gamma_t}(\cdot | z_j). \quad (41)$$

We know that

$$\frac{\sqrt{\frac{p_t}{p_{\theta_0}}} - 1}{H(p_t, p_{\theta_0})}$$

has a limit in $\bigcap_{\epsilon \leq \epsilon_0} \overline{\mathcal{D}_\epsilon^2}$. Using the property of differentiability in quadratic mean we deduce that $\mathcal{S}_2 \subset \bigcap_{\epsilon \leq \epsilon_0} \overline{\mathcal{D}_\epsilon^2}$. Let us define $\tilde{\Theta}_1$ consisting of the elements $(\gamma^0, \boldsymbol{\pi}, \mathbf{z}) \in \Theta_2$ such that:

- $\exists J_0 \subset \{1, \dots, q_2\} : \forall j \in J_0, \pi_j = 0;$
- $\forall i \in \{1, \dots, q_1\}, \exists J_i \subset \{1, \dots, q_2\} : \forall j \in J_i, z_j = z_i^0$ and $\sum_{j \in J_i} \pi_j = \pi_i^0;$
- $(J_i)_{i=0, \dots, q_1}$ is a partition of $\{1, \dots, q_2\}$.

By Assumption 10 and the remark that follows, it suffices to prove that converging subsequences $\theta_n \in \Theta_2$ such that $H(p_{\theta_n}, p_0) \rightarrow 0$ have their limit, say $\theta = (\bar{\gamma}, \boldsymbol{\pi}, \mathbf{z})$, in $\tilde{\Theta}_1$. By application of Fatou's lemma and first part of Assumption 11, $H(p_{\theta_n}, p_{\theta_0}) \rightarrow 0$ implies that $p_\theta = p_{\theta_0}$ μ -a.e. By Assumption 9, we know that $\bar{\gamma} = \gamma_0$ and $\sum_{i=1}^{q_2} \pi_i \alpha_{z_i} = \sum_{i=1}^{q_1} \pi_i^0 \alpha_{z_i^0}$. Set $J_i = \{j \in \{1, \dots, q_2\}, z_j = z_i^0\}$ and $J_0 = \{1, \dots, q_2\} \setminus (\cup_{i=1}^{q_1} J_i)$. Since we assumed that p_{θ_0} has exactly q_1 populations, we deduce that $(J_i)_{i=0, \dots, q_1}$ is a partition of $\{1, \dots, q_2\}$ and consequently that $\theta \in \tilde{\Theta}_1$.

According to previous remark, it suffices to characterize limits of $s_{p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}}}$ as $\mathbf{v} = (\gamma, \boldsymbol{\alpha}, \mathbf{u}, \boldsymbol{\rho})$ tends to 0, \mathbf{z} tends to \mathbf{z}_{lim} and for vectors $\boldsymbol{\pi}$ such that for all $i = 1, \dots, q_1$ we have $\sum_{j \in J_i} \pi_{i,j} = \pi_i^0$. A Taylor series with Lagrange remainder of the function $\mathbf{v} \mapsto \sqrt{p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}}}$ at the point 0 yields

$$s_{p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}}} = \frac{\sqrt{\frac{p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}}}{p_{\theta_0}}} - 1}{H(p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}}, p_{\theta_0})} = \frac{\int_0^1 \langle \mathbf{v}, R(\boldsymbol{\pi}, t\mathbf{v}, \mathbf{z}) \rangle dt}{\left\| \int_0^1 \langle \mathbf{v}, R(\boldsymbol{\pi}, t\mathbf{v}, \mathbf{z}) \rangle dt \right\|_2}.$$

Now we can prove that $\langle \mathbf{v}, R(\boldsymbol{\pi}, 0, \mathbf{z}_{lim}) \rangle = s_2(U, \bar{\boldsymbol{\alpha}}, \boldsymbol{\rho}, \mathbf{z}_{lim})$ where we set $U = \left(\left(\frac{u_{i,j} \pi_{i,j}}{\pi_i^0} \right)_{i=1, \dots, q_1; j=1, \dots, |J_i|}, \gamma \right)$ and for any $i = 1, \dots, q_1$, $\bar{\alpha}_i = \sum_{j \in J_i} \alpha_{i,j}$. Observing that

$$\lim_{\mathbf{v} \rightarrow 0, \mathbf{z} \rightarrow \mathbf{z}_{lim}} \left\| s_{p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}}} - \frac{s_2(U, \bar{\boldsymbol{\alpha}}, \boldsymbol{\rho}, \mathbf{z}_{lim})}{\|s_2(U, \bar{\boldsymbol{\alpha}}, \boldsymbol{\rho}, \mathbf{z}_{lim})\|_2} \right\| \leq 2 \limsup_{\mathbf{v} \rightarrow 0} \frac{\| \langle \mathbf{v}, \int_0^1 (R(\boldsymbol{\pi}, t\mathbf{v}, \mathbf{z}) - R(\boldsymbol{\pi}, \mathbf{0}, \mathbf{z})) dt \rangle \|_2}{\| \langle \mathbf{v}, R(\boldsymbol{\pi}, \mathbf{0}, \mathbf{z}) \rangle \|_2}$$

and applying second part of Assumption 11, we conclude that $\bigcap_{\epsilon \leq \epsilon_0} \overline{\mathcal{D}_\epsilon^2} \subset \overline{\mathcal{S}_2}$.

Proof of Proposition 5. Applying the characterization obtained along the proof of Proposition 4, it suffices to upper bound the bracketing numbers for the class $\{s_{p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}}}, \mathbf{v} \rightarrow \mathbf{0}\}$. Arguments that follow are inspired by Lemma 6.11 in [van der Vaart, 1999] which gives a sufficient condition for parametric classes to be Donsker. We use brackets of the type $[s_p - em, s_p + em]$ for p of the form of some $p_{\boldsymbol{\pi}, \mathbf{v}, \mathbf{z}}$. Their $L^2(p_{\theta_0} \mu)$ -size is equal to $2e\|m\|_2$. Taking into account Assumption 12, we can prove that the class is covered if $(\boldsymbol{\pi}, \mathbf{v}, \mathbf{z})$ ranges over a grid of mesh width $e^{\frac{1}{\beta}}$ over Θ_2 . So we can deduce that there exists a constant (*const*) depending only on Θ_2 such that the bracketing numbers satisfy

$$N_{[\cdot]}(e\|m\|_2, \{s_p\}, L^2(p_{\theta_0} \mu)) \leq (\text{const}) \left(\frac{\text{diam} \Theta_2}{e^{\frac{1}{\beta}}} \right)^{\dim(\Theta_2)}.$$

REFERENCES

- [Adler, 1990] Adler, R.J. (1990). An introduction to continuity, extrema and related topics for general Gaussian processes. *Institute of Mathematical Statistics Lecture Notes-Monograph Series*, **12**.
- [Azais et al., 2006] Azais, J.-M., Gassiat, E. and Mercadier, C. (2006). Asymptotic distribution and power of the likelihood ratio test for mixtures: bounded and unbounded case. *Bernoulli*, **12**(5), 775-799.
- [Bickel et al., 1993] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. (1993) *Efficient and adaptive estimation for semi-parametric models*. Johns Hopkins Series in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD.
- [Chambaz, 2006] Chambaz, A. (2006). Testing the order of a model. *Ann. Statist.*, **34**(3), 1166-1203 (2006)
- [Chambaz, Garivier, Gassiat 2005] Chambaz, A., Garivier, A. and Gassiat, E. (2005). A MDL approach to HMM with Poisson and Gaussian emissions. Application to order identification. submitted.
- [Chen and Chen, 2001] Chen, H. and Chen, J. (2001), Large sample distribution of the likelihood ratio test for normal mixtures, *Statist. Probab. Lett.*, **2**, 125–133.
- [Chen and Chen, 2003] Chen, H. and Chen, J. (2003), Test for homogeneity in normal mixtures in the presence of a structural parameter. *Statist. Sinica*, **13**, 355–365.
- [Chen et al, 2001] Chen, H., Chen, J. and Kalbfleisch, J.D. (2001), A modified likelihood ratio test for homogeneity in finite mixture models. *J.R. Statist. Soc. B*, **63**, 19–29.
- [Chen et al, 2004] Chen, H., Chen, J. and Kalbfleisch, J.D. (2004), Testing for a finite mixture model with two components. *J.R. Statist. Soc. B*, **66**, 95–115.
- [Chen and Kalbfleisch, 2005] Chen, J. and Kalbfleisch, J.D. (2005), Modified likelihood ratio test in finite mixture models with a structural parameter. *J. Statist. Plann. Inference*, **129**, 93–107.
- [Chernoff and Lander, 1995] 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.
- [Chihara, 1978] Chihara, T. (1978), *An introduction to orthogonal polynomials*, Gordon and Breach, New York.
- [Ciuperca, 2002] Ciuperca, G. (2002), Likelihood ratio statistic for exponential mixtures. *Ann. Inst. Statist. Math.*, **54**, 585–594.
- [Dacunha-Castelle and Gassiat, 1997] 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 and Gassiat, 1999] 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, 2003] Delmas, C. (2003), On likelihood ratio test in Gaussian mixture models, *Sankya* **65**, 3, 513-531.
- [Gassiat, 2002] Gassiat, E. (2002), Likelihood ratio inequalities with applications to various mixtures, *Ann. Inst. H. Poincaré Probab. Statist.*, **6**, 897–906.
- [Gassiat, Keribin, 2000] Gassiat, E. and Keribin, C. (2000). The likelihood ratio test for the number of components in a mixture with Markov regime, 2000. *iESAIM P&S*.
- [Garel, 2001] Garel, B. (2001), Likelihood Ratio Test for Univariate Gaussian Mixture, *J. Statist. Plann. Inference*, **96**(2), 325–350.
- [Garel, 2005] Garel, B. (2005), Asymptotic theory of the likelihood ratio test for the identification of a mixture, *J. Statist. Plann. Inference*, **131**(2), 271–296.
- [Ghosal and van der Vaart, 2001] Ghosal, S. and van der Vaart, A. W. (2001) Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities, *Ann. Statist.*, **29** (5), 1233–1263.
- [Ghosh and Sen, 1985] 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.
- [Hall and Stewart, 2005] Hall, P. and Stewart, M. (2005), Theoretical analysis of power in a two-component normal mixture model. *J. Statist. Plann. Inference* **134** (1), 158–179.

- [Hartigan, 1985] 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.
- [Henna, 2005] Henna, J. (2005), Estimation of the number of components of finite mixtures of multivariate distributions, *Ann. Inst. Statist. Math.*, **57** (4), 655–664.
- [James et al., 2001] James, L.F., Priebe, C.E., and Marchette, D.J. (2001), Consistent Estimation of Mixture Complexity, *Ann. Statist.*, **29**, 1281–1296.
- [Keribin, 2000] Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā Ser. A*, **62**(1) 49–66
- [Lemdani and Pons, 1997] Lemdani, M. and Pons, O. (1997), Likelihood ratio test for genetic linkage, *Statis. Probab. Lett.*, **33**(1), 15–22.
- [Lemdani and Pons, 1999] Lemdani, M. and Pons, O. (1999), Likelihood ratio in contamination models, *Bernoulli*, **5**(4), 705–719.
- [Lindsay, 1995] 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 and Shao, 2004] 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.
- [Liu et al., 2003] Liu, X., Pasarica, C. and Shao, Y. (2003), Testing homogeneity in gamma mixture models, *J. Scand. J. Statist.*, **30**(1), 227–239.
- [Lo, 2005] Lo, Y. (2005), Likelihood ratio tests of the number of components in a normal mixture with unequal variances. *Statis. Probab. Lett.*, **71**(3), 225–235.
- [Lord, 1969] Estimating the true-score distributions in psychological testing (an empirical bayes estimation problem). *Psychometrika*, **34**, 259-299.
- [McLachlan and Peel, 2000] McLachlan, G. and Peel, D. (2000), *Finite mixture models*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York.
- [Mercadier, 2005] Mercadier, C. (2005), toolbox MATLAB. <http://math.univ-lyon1.fr/mercadier/MAGP/>
- [Misra et al., 2003] Misra N. , Singh H. and Harner E.J. Stochastic comparisons of poisson and binomial random variables with their mixtures. *Statis. Probab. Lett.*, **65**, 279-290.
- [Murphy and van der Vaart, 1997] Murphy, S.A. and van der Vaart, A.W. (1997), Semiparametric likelihood ratio inference. *Ann. Statist.*, **25**(4), 1471–1509.
- [Qin and Smith, 2004] Quin, Y. S. and Smith, B. (2004), Likelihood ratio test for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sinica*, **143**, 1165–1177.
- [Qin and Smith, 2006] Quin, Y. S. and Smith, B. (2006), The likelihood ratio test for homogeneity in bivariate normal mixtures. *J. Multivariate Anal.*, **97**, 474–491.
- [Titterton et al., 1985] Titterton, 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, 1998] van der Vaart, A. W. (1998), *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.
- [van der Vaart, 1999] van der Vaart, A. W. (1999), *Semiparametric Statistics*, Lectures on probability theory and statistics, Saint-Flour, 1999. Lecture Notes in Math., Vol. 1781, pp 331–457, Springer, Berlin, 2002.
- [van der Vaart and Wellner, 1996] van der Vaart, A. W. and Wellner, J. A. (1996), *Weak convergence and empirical processes*, Springer Series in Statistics, Springer-Verlag.
- [Wood, 1999] Wood, G. R. (1999), Binomial mixtures: geometric estimation of the mixing distribution. *Ann. Statist.*, **5**, 1706-1721.