Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes.

D. Dacunha-Castelle* E. Gassiat†

November 1996, revised June 1998.

Abstract:
In this paper, we address the problem of testing hypothesis using the likelihood ratio test statistic in non identifiable models, with application for model selection in situations where the parametrization for the larger model leads to non identifiability in the smaller model. We give two major applications: the case where the number of populations has to be tested in a mixture, and the case of stationary ARMA(p,q) processes where the order (p,q) has to be tested. We give the asymptotic distribution for the likelihood ratio test statistic when testing the order of the model. In the case of ARMA's order selection, the asymptotic distribution is invariant with respect to the parameters generating the process. A locally conic parametrization is a key tool in deriving the limiting distributions; it allows one to discover the deep similarity between the two problems.

A.M.S. Classification: 62 F 05, 62 H 30, 62 A 10.

Key Words: model selection, tests, non identifiable models, mixtures, ARMA processes.

Aknowledgments: The authors want to thank the referees for their helpful comments.

Abbreviated title: Testing the order of a mixture or an ARMA.

*Laboratoire Modélisation Stochastique et Statistique, Université d'Orsay, Bat. 425, 91 405 Orsay
†Laboratoire Analyse et Probabilité, Université d'Evry-Val d'Essonne, Boulevard des Coquibus, 91 025 Evry

1 Introduction.

In this paper, we propose a general theory for the derivation of the limiting distribution of likelihood ratio test (LRT) statistics in testing problems in which some of the parameters of the alternative hypothesis are no longer identifiable in the null hypothesis. Two famous examples of such situation are the test of the number of components in a mixture, and the test of the order of an ARMA process. Segmented regression model (see Feder, 1975) is another example of such situation.
In such problems, if the model is a regular parametric model, then lack of identifiability leads in general to a non full rank Fisher information matrix, and so standard proofs of the asymptotic chi-squared distribution fail. A simple derivation of the chi-squared theory can be based on an expansion of the likelihood up to order 2, followed by a maximization of this expansion. The major questions that arise in the non-identifiable context are:

- Question 1: Since the parameter is not identifiable, around which point can an expansion be made ?
- Question 2: In the optimization procedure, the inverse of the Fisher information appears, what to do since it is not invertible ?

Our solution to this problem will arise by creating a special "conic" parametrization in which asymptotic expansions can be more readily established. It extends the results given in Dacunha-Castelle and Gassiat (1996) for testing one population against two populations in a mixture.
We start with a discussion of the non-identifiability in these two main examples so as to illustrate the nature of the difficulty.
Let $\mathcal{F} = (f_\gamma)_{\gamma\in\Gamma}$ be a family of probability densities with respect to ν. Let Γ be a compact subset of $\mathbb{R}^k$ for some integer k. Let $\mathcal{G}_p$ be the set of all $p-$mixtures of densities of $\mathcal{F}$:

$$\begin{aligned}\mathcal{G}_p \quad &= \quad \{g_{\pi,\alpha} = \sum_{i=1}^p \pi_i \cdot f_{\gamma^i} \;:\; \pi = (\pi_1,\ldots,\pi_p),\; \alpha = (\gamma^1,\ldots,\gamma^p), \qquad (1)\\ &\forall i = 1,\ldots,p,\; \gamma^i \in \Gamma,\; 0 \le \pi_i \le 1,\; \sum_{i=1}^p \pi_i = 1\}\end{aligned}$$

Obviously, the model is not identifiable for the parameters $\pi = (\pi_1,\ldots,\pi_p)$ and $\alpha = (\gamma^1,\ldots,\gamma^p)$. There exist mixtures g in $\mathcal{G}_p$ which have different representations $g_{\pi,\alpha}$ with different parameters π and α. For instance, we have for any permutation σ of the set $\{1,\ldots,p\}$:

$$\sum_{i=1}^p \pi_i \cdot f_{\gamma^i} = \sum_{i=1}^p \pi_{\sigma(i)} \cdot f_{\gamma^{\sigma(i)}}$$

Another example which may not be avoided by taking some quotient with respect to permutations is:

$$f_{\gamma^0} = \sum_{i=1}^{p} \pi_i \cdot f_{\gamma^0}$$

for any (π_i) such that $\pi_i \geq 0$ and $\sum_{i=1}^{p} \pi_i = 1$.

ARMA processes are given by the recurrence equation:

$$X_n + a_1 X_{n-1} + \ldots + a_p X_{n-p} = \epsilon_n + b_1 \epsilon_{n-1} + \ldots + b_q \epsilon_{n-q} \tag{2}$$

$(X_n)_{n\in\mathbb{N}}$ is a stationary process with $(\epsilon_n)_{n\in\mathbb{N}}$ as innovation process as soon as the complex polynomials $P(z) = 1 + \sum_{j=1}^{p} a_j z^j$ and $Q(z) = 1 + \sum_{j=1}^{q} b_j z^j$ do not have roots inside the complex unit disc and $(\epsilon_n)_{n\in\mathbb{N}}$ is a white noise. If $(\epsilon_n)_{n\in\mathbb{N}}$ is gaussian, then $(X_n)_{n\in\mathbb{N}}$ is a gaussian ARMA process. The spectral density f of such a process is given by:

$$f(x) = \frac{\sigma^2}{2\pi} |\frac{Q}{P}(e^{ix})|^2 \tag{3}$$

where σ^2 is the variance of the noise. Assume now that the true spectral density is

$$f_0(x) = \frac{\sigma_0^2}{2\pi} |\frac{Q_0}{P_0}(e^{ix})|^2 \tag{4}$$

with Q_0 of degree q_0, P_0 of degree p_0, and we want to test (p_0, q_0) against (p, q), where $p \geq p_0$ and $q \geq q_0$, $(p, q) \neq (p_0, q_0)$. The general model is that of stationary processes with spectral densities which have the form (3), where the degree of P is not larger than p and the degree of Q is not larger than q. As soon as $p > p_0$ and $q > q_0$, f_0 has in this model infinitely many representations, for instance by multiplying the representation (4) with the constant 1 written as the quotient of two identical polynomials. With this parametrization of the model, the information matrix for any parameter leading to f_0 has a kernel of dimension $\inf\{p - p_0, q - q_0\}$, see Theorem 3.3 in Azencott and Dacunha-Castelle (1986).

The mixture problem has received extensive consideration in the literature. A complete discussion of previous results may be found in the monograph of Lindsay (1995) together with results covering the multinomial models. As he himself says, "the nature of the limiting distribution for this likelihood ratio test is a long-standing mystery". Ghosh and Sen (1985) gave the asymptotic distribution of the LRT statistic under a strong (and unsatisfactory) separation condition of the parameters of the mixture. Hartigan (1985) proved that, when testing a standard centered gaussian distribution against a mixture of a standard centered gaussian and a standard gaussian distribution

with mean m, with no upper or lower bound on m, the LRT statistic tends to infinity in probability under the null hypothesis, and Bickel and Chernoff (1993) gave the precise asymptotics. More recently, the asymptotic distribution of the LRT statistic for testing various mixtures of binomials has been given, see Chernoff and Lander (1995), Lemdani and Pons (1997).
Testing the dimension of the model for ARMA time series has also received considerable interest in the literature. Concerning the LRT (or pseudo LRT) statistic, the computation of the asymptotic distribution was made by Hannan (1982) for the particular case $p_0 = q_0 = 0$ and $p = q = 1$, introducing a reparametrization of the model adapted to the particular situation. Veres (1987) used again this reparametrization to find the asymptotic distribution of the LRT statistic for testing (p_0, q_0) against $(p_0 + 1, q_0 + 1)$. But this particular reparametrization does not seem to be easily generalised to handle with the general case.

In this paper, we give the asymptotic distribution of the maximum likelihood statistic for *any* mixture model selection and for *any* ARMA model selection problem, so that this allows the construction of a test for the order at a known asymptotic level. To find this asymptotic distribution, we introduce a reparametrization of the model which we call locally conic. The general idea is that a first positive and real parameter θ contains some "distance" to the true model, this is the perturbation parameter, and a second parameter β is some direction of approach to the true model, or in other words the direction of the perturbation. A normalization of the directional vector imposes the directional Fisher informations to be uniformly equal to 1. This gives an answer to question 2. β contains all the non identifiable part of the model. θ contains all the model order information, and is identifiable. θ may be consistently estimated: this gives an answer to question 1, the expansion will be done for θ near 0. The limiting distribution will be obtained as a supremum of a function of a continuous gaussian process on a set $\mathcal{D}$ of directional scores. Those directional scores may be obtained as normalized limits of likelihood ratio functions as one approaches the null from the alternative. The simple idea of such a set was already contained in Lindsay (1995) for the mixture testing problem. The main difference between mixtures and ARMA's model selection is that the limiting distribution, when testing the number of components of a mixture, depends on the null distribution, whereas, when testing the order of an ARMA, it is invariant with respect to the null hypothesis.
The paper is organized as follows: In section 2, we give the definition of a locally conic parametrization. Section 3 is devoted to the problem of testing q against p populations, and section 4 is devoted to the problem of testing an ARMA(p_0, q_0) against an

ARMA(p,q). All technical proofs that are not essential for a comprehensive reading are given in section 5.

2 Locally conic models.

Let $X^{(n)} = (X_1, \ldots, X_n)$ be an n-dimensional real observation with distribution $P_0^{(n)}$ in a set $\mathcal{P}_n$, which is assumed to be dominated by some positive measure $\nu^{(n)}$.
We assume there exists a parametrization of all $\mathcal{P}_n$ through two parameters θ and β: $(\theta, \beta) \in [0, M] \times \mathcal{B}$, where M is a positive real number, $\mathcal{B}$ is a compact Polish space, and there exists a subset $\mathcal{T}$ of $[0, M] \times \mathcal{B}$ such that:

$$\forall n \in \mathbb{N}, \;\; \mathcal{P}_n = \{P^{(n)}_{(\theta,\beta)}, (\theta, \beta) \in \mathcal{T}\}.$$

Here $[0, M] \times \mathcal{B}$ is endowed with the product topology of $\mathbb{R}$ and $\mathcal{B}$, and $\mathcal{T}$ has compact closure $\overline{\mathcal{T}}$.

The parametrization is assumed to be non identifiable in the parameter β for $\theta = 0$, but identifiable in the parameter θ at $\theta = 0$, that is:

- **(LC1)**

$$P^{(n)}_{(\theta,\beta)} = P_0^{(n)} \Longleftrightarrow \theta = 0$$

which in particular implies:

$$\forall \beta \in \mathcal{B}, P^{(n)}_{(0,\beta)} = P_0^{(n)}.$$

For any β in $\mathcal{B}$, define:

$$\theta_\beta = \sup\{t \;:\; [0, t] \times \{\beta\} \subset \overline{\mathcal{T}}\}.$$

Assume moreover:

- **(LC2)** $\forall \beta \in \mathcal{B}$, either $\theta_\beta > 0$, or there exists $c > 0$ such that $[0, c] \times \{\beta\} \cap \overline{\mathcal{T}}$ is empty.

Define now:

$$\tilde{\mathcal{B}} = \{\beta \in \mathcal{B} \;:\; \theta_\beta > 0\}.$$

Assumption **(LC2)** means that, for any accumulation sequence of parameters (θ_k, β_k) leading to $\theta = 0$, the submodels $(P^{(n)}_{(\theta,\beta_k)})_\theta$ are defined in a right neighborhood of 0. Moreover, $\tilde{\mathcal{B}}$ is then the set of all directions β for which the submodel approaches 0.

Such parametrization is called a locally conic parametrization. Models for which there exists a locally conic parametrization are called locally conic models.

3 Testing the number of populations in a mixture.

In this section, $X^{(n)}$ is an n-sample of a mixture of q populations, that is

$$P_0^{(n)} = (g_0 \nu)^{\otimes n}$$

where g_0 is a mixture of q populations in the parametric family $(f_\gamma)_{\gamma \in \Gamma}$, $\Gamma \subset \mathbb{R}^k$:

$$g_0 = \sum_{l=1}^{q} \pi_l^0 f_{\gamma^{l,0}}.$$

The general model is that of p mixtures $\mathcal{G}_p$ given by (1). We assume that $\mathcal{G}_p$ is identifiable in the following weak sense:

$$\sum_{l=1}^{p} \pi_l^0 f_{\gamma^{l,0}} = \sum_{l=1}^{p} \pi_l^1 f_{\gamma^{l,1}} \quad \nu \ a.e. \Longleftrightarrow \sum_{i=1}^{p} \pi_i^0 \cdot \delta_{\gamma_i^0} = \sum_{i=1}^{p} \pi_i^1 \cdot \delta_{\gamma_i^1}$$

as probability distributions on Γ. In other words, $\mathcal{G}_p$ is identifiable if the parameter is the discrete mixing probability distribution on Γ. Teicher (1965) or Yakowitz and Spragins (1968) give sufficient conditions for such weak identifiability, which hold for instance for finite mixtures of gaussian or gamma distributions.

The aim of this section is to derive the limiting distribution of the LRT statistic. Define for any g in $\mathcal{G}_p$:

$$l_n(g) = \sum_{i=1}^{n} \log g(X_i)$$

and the statistic:

$$T_n(p) = \sup_{g \in \mathcal{G}_p} l_n(g) - l_n(g_0).$$

The LRT statistic for testing H_0: " q populations" against H_1: "p populations" is

$$V_n = T_n(p) - T_n(q).$$

Since we shall use partial derivatives of f_γ with respect to γ, we introduce some notations: $D^h_{i_1 \ldots i_h}$ will be the h-th partial derivative operator with respect to $\gamma_{i_1} \ldots \gamma_{i_h}$. So that $D^h_{i_1 \ldots i_h} f_\gamma$ will be the value of this partial derivative of f at point γ.

We introduce the following locally conic parametrization, previously proposed by the authors in Dacunha-Castelle and Gassiat (1996). The idea is to define a perturbation of g_0 in the following way: perturb the q mixture g_0 through a perturbation of the parameters $\gamma^{l,0}$ and the weights π_l^0, and add a perturbation as a $p-q$-mixture with weight tending to 0. This leads to:

$$g_{(\tilde{\theta},\beta)} = \sum_{i=1}^{p-q} \lambda_i \tilde{\theta} f_{\gamma^i} + \sum_{l=1}^{q} (\pi_l^0 + \rho_l \tilde{\theta}) f_{\gamma^{l,0} + \tilde{\theta} \delta^l}$$

Here,

$$\beta = (\lambda_1, \ldots, \lambda_{p-q}, \gamma^1, \ldots, \gamma^{p-q}, \delta^1, \ldots, \delta^q, \rho_1, \ldots, \rho_q)$$

Now, the directional Fisher information is $\| \frac{\partial g_{0,\beta}}{\partial \theta} / g_0 \|_H^2$, where H is the Hilbert space $L^2(g_0 \nu)$, and does not equal 1 uniformly. A normalizing factor $N(\beta)$ is introduced so as to set all directional Fisher informations to 1. This leads to the following definition:

$$g_{(\theta,\beta)} = \sum_{i=1}^{p-q} \lambda_i \frac{\theta}{N(\beta)} f_{\gamma^i} + \sum_{l=1}^{q} (\pi_l^0 + \rho_l \frac{\theta}{N(\beta)}) f_{\gamma^{l,0} + \frac{\theta}{N(\beta)} \delta^l} \tag{5}$$

with

$$N(\beta) = \| \sum_{l=1}^{q} \pi_l^0 \sum_{i=1}^{k} \delta_i^l \frac{D_i^1 f_{\gamma^{l,0}}}{g_0} + \sum_{i=1}^{p-q} \lambda_i \frac{f_{\gamma^i}}{g_0} + \sum_{l=1}^{q} \rho_l \frac{f_{\gamma^{l,0}}}{g_0} \|_H$$

Equation (5) does not completely define a locally conic parametrization. Indeed, it does not define unambiguously (θ, β) for a given mixture. For instance, different sets of parameters may give g_0. First , restrictions are imposed on the β's:

$$\lambda_i \geq 0,\ \gamma^i \in \Gamma,\ i = 1, \ldots, p-q,\ \delta^l \in \mathbb{R}^k,\ \rho_l \in \mathbb{R},\ l = 1, \ldots, q \tag{6}$$

$$\sum_{i=1}^{p-q} \lambda_i + \sum_{l=1}^{q} \rho_l = 0 \ \text{ and } \ \sum_{i=1}^{p-q} \lambda_i^2 + \sum_{l=1}^{q} \rho_l^2 + \sum_{l=1}^{q} \|\delta^l\|^2 = 1 \tag{7}$$

But this is not sufficient. Without further restrictions on β, **(LC1)** does not hold. It is essential to define the set $\mathcal{B}$ of possible β such that $g_{(\theta,\beta)} = g_0 \iff \theta = 0$. We shall do it now.

Let g be any p-mixture:

$$g = \sum_{i=1}^{p} \pi_i \cdot f_{\gamma^i}$$

To describe it through equation (5), one has to associate the parameters of g to those of g_0, that is to give a special order to the parameters. In other words: for any permutation σ of $[1, \ldots, p]$, we define the parameters θ_σ and β_σ such that $g_{(\theta_\sigma, \beta_\sigma)} = g$. This leads to:

$$\beta_\sigma = (\lambda_{1,\sigma}, \ldots, \lambda_{p-q,\sigma}, \gamma^{1,\sigma}, \ldots, \gamma^{p-q,\sigma}, \delta^{1,\sigma}, \ldots, \delta^{q,\sigma}, \rho_{1,\sigma}, \ldots, \rho_{q,\sigma})$$

with:

$$\forall i = 1, \ldots, p-q,\ \ \lambda_{i,\sigma} \cdot \theta_\sigma = \pi_{\sigma(i)} \cdot N(\beta_\sigma)$$

$$\forall i = 1, \ldots, p-q,\ \ \gamma^{i,\sigma} = \gamma^{\sigma(i)}$$

$$\forall i = 1, \ldots, q,\ \ \delta^{i,\sigma} \cdot \theta_\sigma = (\gamma^{\sigma(p-q+i)} - \gamma^{i,0}) \cdot N(\beta_\sigma)$$

$$\forall i=1,\ldots,q,\ \ \rho_{i,\sigma}\cdot\theta_\sigma=(\pi_{\sigma(p-q+i)}-\pi_i^0)\cdot N(\beta_\sigma)$$

It is easily seen that

$$\theta_\sigma=\|\sum_{l=1}^{q}\sum_{i=1}^{k}(\gamma_i^{\sigma(p-q+l)}-\gamma_i^{l,0})\frac{D_i^1 f_{\gamma^{l,0}}}{g_0}+\sum_{i=1}^{p-q}\pi_{\sigma(i)}\frac{f_{\gamma^{\sigma(i)}}}{g_0}+\sum_{l=1}^{q}(\pi_{\sigma(p-q+l)}-\pi_l^0)\frac{f_{\gamma^{l,0}}}{g_0}\|_H$$

The system is not ambiguous on the scale of β_σ because of the normalizing condition (7). The problem is then to choose between the permutations. The idea is to associate step by step the nearest points γ^i involved in g to the set of points $\gamma^{l,0}$ involved in g_0. Look for:

$$\min_{l=1,\ldots,q,i=1,\ldots,p}\|\gamma^{l,0}-\gamma^i\|$$

It is attained for l_1 and i_1. Define then $\sigma(p-q+l_1)=i_1$. Look then for

$$\min_{l=1,\ldots,q,l\neq l_1,i=1,\ldots,p,i\neq i_1}\|\gamma^{l,0}-\gamma^i\|$$

It is attained for l_2 and i_2. Set then $\sigma(p-q+l_2)=i_2$. By induction, define in the same way $\sigma(p-q+l_j)=i_j$ for $j=1,\ldots,q$. In this algorithm, consider only points truly involved in g (eventually less than p points). Then complete the permutation σ in some ordered way. You then have defined a permutation $\sigma(g)$. Define now

$$\mathcal{B}=\{\beta_{\sigma(g)},g\in\mathcal{G}_p\}$$

and also

$$\mathcal{T}=\{(\theta,\beta_{\sigma(g)})\ :\ \theta\leq\theta_{\sigma(g)},\ \ g\in\mathcal{G}\}$$

This induces the set $\tilde{\mathcal{B}}$ as the intersection of all directions approaching 0 in $\mathcal{T}$. By construction, **(LC1)** and **(LC2)** hold and the parametrization is locally conic.
An important point to notice, coming from the normalizing condition (7) is that:

$$\forall(\theta,\beta)\in\mathcal{T},\ \frac{\theta}{N(\beta)}\leq p+2q\sup_{\gamma\in\Gamma}\|\gamma\|^2 \tag{8}$$

Th upper bound in (8) is not tight: for fixed β, some θ satisfying the upper bound will not give a mixture density. But for any β, any θ less than an appropriate constant multiple of $N(\beta)$ will yield a mixture density. Unequality (8) is mentioned to point out that θ is bounded, and tends to 0 as soon as $N(\beta)$ tends to 0.

We shall now need some more assumptions. Define $\mathcal{D}$ the subset of the unit sphere of H of functions of form

$$\frac{1}{N(\beta)}\left(\sum_{l=1}^{q}\pi_l^0\sum_{i=1}^{k}\delta_i^l\frac{D_i^1 f_{\gamma^{l,0}}}{g_0}+\sum_{i=1}^{p-q}\lambda_i\frac{f_{\gamma^i}}{g_0}+\sum_{l=1}^{q}\rho_l\frac{f_{\gamma^{l,0}}}{g_0}\right) \tag{9}$$

with β in $\tilde{\mathcal{B}}$. Define also ξ_d the gaussian process indexed by $\mathcal{D}$ with covariance the usual hilbertian product in H.
We shall use the following assumptions.

- **(P0)** There exists a function h in $L_1(g_0\nu)$ such that: $\forall \gamma \in \Gamma$, $|\log f_\gamma| \leq h$ ν-a.e. Moreover, f_γ possesses partial derivatives till order 5. For all $h \leq 5$, and all $i_1 \dots i_h$,
$$\frac{D^h_{i_1 \dots i_h} f_{\gamma^0}}{g_0} \in L^3(g_0\nu)$$
Moreover, there exists a function m_5 and a positive ϵ such that:
$$\sup_{\|\gamma-\gamma^0\| \leq \epsilon} \left|\frac{D^5_{i_1 \dots i_5} f_\gamma}{g_0}\right| \leq m_5 \quad E_{g_0\nu}[m_5^3] < +\infty$$

- **(P1)** For any integer p_1, p_2, such that $p_1 + p_2 \leq p - q$, for any set of distinct points γ^l, $l = 1, \dots, p_1$, distinct from any $\gamma^{l,0}$, any permutation σ of $[1, \dots, q]$, the set of functions
$$\left(\left(\frac{f_{\gamma^l}}{g_0}\right)_{l=1,\dots,p_1}, \left(\frac{f_{\gamma^{l,0}}}{g_0}\right)_{l=1,\dots,q}, \left(\frac{D^1_i f_{\gamma^{l,0}}}{g_0}\right)_{l=1,\dots,q, i=1,\dots,k}, \left(\frac{D^2_{ij} f_{\gamma^{l,0}}}{g_0}\right)_{l=\sigma(1),\dots,\sigma(p_2), i,j=1,\dots,k}\right.$$
is free in H.

Notice that **(P0)** implies the existence of functions m_j, $j = 1, \dots, 4$ such that for all j:
$$\sup_{\|\gamma-\gamma^0\| \leq \epsilon} \left|\frac{D^j_{i_1 \dots i_j} f_\gamma}{g_0}\right| \leq m_j \quad E_{g_0\nu}[m_j^3] < +\infty.$$
Another important consequence of **(P0)** is:

Proposition 3.1 *Under the assumption **(P0)**, $\mathcal{D}$ is a Donsker class (see Van der Vaart and Wellner, 1996), and ξ_d has continuous sample paths.*

The proof of Proposition 3.1 is given in section 5. It involves the identification of $\overline{\mathcal{D}}$, the (compact) closure of $\mathcal{D}$ in H, which is found by letting some of the γ^l tend to some of the $\gamma^{l,0}$ in (9). $\overline{\mathcal{D}}$ is the subset of the unit sphere of H of functions of form
$$\sum_{l=1}^{p_1} \mu_l \frac{f_{\gamma^l}}{g_0} + \sum_{l=1}^{q} \tilde{\rho}_l \frac{f_{\gamma^{l,0}}}{g_0} + \sum_{l=1}^{q} \sum_{i=1}^{k} \lambda_{l,i} \frac{D^1_i f_{\gamma^{l,0}}}{g_0} + \sum_{l \in L(p_1)} \sum_{u \in J(l)} \tau_u \sum_{i,j=1}^{k} a_i^u a_j^u \frac{D^2_{ij} f_{\gamma^{l,0}}}{g_0} \tag{10}$$
with $p_1 \leq p - q$, $\mu_l \geq 0$, $\sum_{l=1}^{p_1} \mu_l + \sum_{l=1}^{q} \tilde{\rho}_l = 0$, $\tau_u \geq 0$, and where $L(p_1)$ is a subset of $\{1, \dots, q\}$ of cardinality less or equal to $p - q - p_1$, and $(J(l))_{l \in L(p_1)}$ is a partition of $\{1, \dots, p - q\}$.
The following Theorem states the asymptotic distribution of $T_n(p)$:

Theorem 3.2 *Under the assumptions **(P0)**and **(P1)**, $T_n(p)$ converges in distribution to the following variable:*

$$\frac{1}{2}\sup_{d\in\mathcal{D}}\xi_d^2 1_{\xi_d\geq 0}$$

We just give a sketch of proof, the complete proof being given in section 5. The difficult point is to define a (random) partition of the parameter space so that on each partitioning set, the converging parameter converges uniformly to 0. Indeed, this is needed to be able to handle with an expansion of the likelihood and to compute the limiting distribution of the statistic.
First of all, define $\widehat{\theta}_\beta$ a maximizer of $l_n(\theta,\beta)$ for the fixed value of β. Here, $l_n(\theta,\beta)=l_n(g_{(\theta,\beta)})$. Recall the following result, which is proved in Dacunha-Castelle and Gassiat (1996), coming from **(P0)** and the fact that the parametrization is locally conic:

Proposition 3.3 *Define $\eta_n=\sup_{\beta\in\tilde{\mathcal{B}}}\widehat{\theta}_\beta$. Then η_n converges to 0 in probability as n tends to infinity.*

Define now the partition

$$A_n=\{\beta\ :\ \frac{\|\sup_l\delta^l\|}{N(\beta)^2}\leq\frac{1}{\eta_n^\alpha}\}$$

for some $\alpha<3/4$, and

$$B_n=\{\beta\ :\ \exists l=1,\ldots,q,\ \frac{\|\delta^l\|}{N(\beta)^2}\geq\frac{1}{\eta_n^\alpha}\}$$

We have:

Lemma 3.4 *Define $\mathcal{L}_n=\sup_{\beta\in A_n}l_n(\widehat{\theta}_\beta,\beta)-l_n(0)$. Under the assumptions of Theorem 3.2, $\mathcal{L}_n$ converges in distribution to*

$$\frac{1}{2}\sup_{d\in\mathcal{D}}(\xi_d)^2\cdot 1_{\xi_d}\geq 0 \tag{11}$$

and:

Lemma 3.5 *Under the assumptions of Theorem 3.2, $\sup_{\beta\in B_n}l_n(\widehat{\theta}_\beta,\beta)-l_n(0)$ is upper bounded by $\mathcal{L}_n+o(1)$.*

Theorem 3.2 is now a consequence of Lemma 3.4 and Lemma 3.5.

The distribution of the LRT statistic may now be given. Indeed, the limiting distribution of $(T_n(p),T_n(q))$ is

$$(\frac{1}{2}\sup_{d\in\mathcal{D}}\frac{1}{2}\xi_d^2 1_{\xi_d\geq 0},\frac{1}{2}\sup_{d_0\in\mathcal{D}_0}\frac{1}{2}\xi_{d_0}^2 1_{\xi_{d_0}\geq 0})$$

where $\mathcal{D}_0$ is the set of directional scores for the identifiable model of q-mixtures $\mathcal{G}_q$. That is, $\mathcal{D}_0$ is the set of functions of form

$$\frac{1}{N(\beta)}\left(\sum_{l=1}^{q}\pi_l^0\sum_{i=1}^{k}\delta_i^l\frac{D_i^1 f_{\gamma^{l,0}}}{g_0}+\sum_{l=1}^{q}\rho_l\frac{f_{\gamma^{l,0}}}{g_0}\right), \tag{12}$$

and $\mathcal{D}_0$ is of course a subset of $\mathcal{D}$. Define $\mathcal{E}_0$ the positive cone spanned by $\mathcal{D}_0$, which is here a linear space. Define $P^{\mathcal{E}_0^\perp}$ the linear orthogonal projector onto $\mathcal{E}_0^\perp$, the orthogonal space of $\mathcal{E}_0$. Let $\mathcal{U}$ be the set of normalized vectors of $P^{\mathcal{E}_0^\perp}(\mathcal{D})$, that is:

$$\mathcal{U}=\{\frac{u}{\|u\|_H}:u=P^{\mathcal{E}_0^\perp}(d),d\in\mathcal{D}\}.$$

We have:

Theorem 3.6 *Under the assumptions **(P0)** and **(P1)**, V_n converges in distribution to the following variable:*

$$\frac{1}{2}\sup_{u\in\mathcal{U}}\xi_u^2 1_{\xi_u\geq 0}$$

Comments.

- The compactness restriction on Γ is essential as shown by Hartigan (1985). It is though not sufficient to have a tight limit for the LRT: Ciuperca (1998) proves that, for translation mixtures of exponential distributions, the LRT converges to $+\infty$ with probability 1/2. This is due to the non compactness of the set of scores $\mathcal{D}$.

- The limiting distribution depends on the null parameter. Lindsay (1995) gives an approximation theory of the limiting distribution based on Hotelling's tubes.

- Analytic derivations of the distributions of the supremum of the Gaussian process as involved in the Theorems are difficult problems. In similar contexts, Beran and Millar (1987) have proposed stochastic procedures using bootstrapping to find the estimated level of confidence sets when the asymptotic distribution is too intractable. Similar ideas could be used here.

- Verification of assumptions **(P0), (P1)**, is done in Kéribin (1997) for Gaussian and Poisson mixtures.

- Assumption **(P0)** is probably not optimal. It should be possible to prove the result using only derivatives up to order 3. However, one has to be careful to uniformity of convergence of the converging parameter and in the expansions.

In the particular case where $q=1$, $p=2$ and $k=1$, the limiting distribution takes a particular simple form. Define

$$e_1 = \frac{D^1 f_{\gamma^0}}{f_{\gamma^0}} / \| \frac{D^1 f_{\gamma^0}}{f_{\gamma^0}} \|_H .$$

The set $\mathcal{D}_0$ reduces to the set $\{e_1, -e_1\}$, and the set $\mathcal{D}$ is the set of functions of form

$$\lambda e_1 + \tau h(\gamma)$$

sucht that $\tau \geq 0$, $\lambda^2 + \tau^2 = 1$,

$$a(\gamma) = \langle \frac{f_\gamma - f_{\gamma^0}}{f_{\gamma^0}}, e_1 \rangle_H$$

and

$$h(\gamma) = \frac{\frac{f_\gamma - f_{\gamma^0}}{f_{\gamma^0}} - a(\gamma) e_1}{\| \frac{f_\gamma - f_{\gamma^0}}{f_{\gamma^0}} - a(\gamma) e_1 \|_H}.$$

Corollary 3.7 *In case $q=1$, $p=2$ and $k=1$, under the assumptions **(P0)** and **(P1)**, V_n converges in distribution to*

$$\frac{1}{2} \sup_{\gamma \in \Gamma} \xi^2_{h(\gamma)} 1_{\xi_{h(\gamma)} \geq 0}$$

Proof of Theorem 3.6.

Let $\mathcal{C}_+(\mathcal{D})$ be the positive cone spanned by $\mathcal{D}$, and $\mathcal{C}_+(\mathcal{U})$ be the positive cone spanned by $\mathcal{U}$. We have:

$$\mathcal{C}_+(\mathcal{D}) = \mathcal{C}_+(\mathcal{U}) \overset{\perp}{\bigoplus} \mathcal{E}_0.$$

Thus, we easily have

$$\sup_{d \in \mathcal{D}} \xi_d^2 1_{\xi_d \geq 0} = \sup_{u \in \mathcal{U}, d_0 \in \mathcal{D}_0} \sup_{\mu \geq 0, \lambda^2 + \mu^2 = 1} \xi^2_{\lambda d_0 + \mu u} 1_{\xi_{\lambda d_0 + \mu u} \geq 0}.$$

But (ξ_d) is a linear process so that

$$\xi_{\lambda d_0 + \mu u} = \lambda \xi_{d_0} + \mu \xi_u.$$

Now direct computation leads to

$$\sup_{\mu \geq 0, \lambda^2 + \mu^2 = 1} (\lambda \xi_{d_0} + \mu \xi_u)^2 1_{\lambda \xi_{d_0} + \mu \xi_u \geq 0} = \xi^2_{d_0} + \xi^2_u 1_{\xi_u \geq 0}.$$

$\mathcal{D}_0$ is a symmetrical set, so that

$$\sup_{d_0 \in \mathcal{D}_0} \xi^2_{d_0} 1_{\xi_{d_0} \geq 0} = \sup_{d_0 \in \mathcal{D}_0} \xi^2_{d_0}$$

and thus

$$\frac{1}{2} \sup_{d \in \mathcal{D}} \xi_d^2 1_{\xi_d \geq 0} - \frac{1}{2} \sup_{d_0 \in \mathcal{D}_0} \xi^2_{d_0} 1_{\xi_{d_0} \geq 0} = \frac{1}{2} \sup_{u \in \mathcal{U}} \xi_u^2 1_{\xi_u \geq 0}.$$

4 Testing the order of an ARMA process.

In this section, $X^{(n)}$ is an n-realization of a strictly stationary process with spectral density f_0 given by equation (4).
Recall that if X is an ARMA(p,q) process with spectral density f, the Fejer-Riesz canonical representation is:

$$f(e^{ix}) = \frac{\sigma^2}{2\pi} \cdot |\frac{Q}{P}(e^{ix})|^2 = \frac{\sigma^2}{2\pi} \cdot g(e^{ix})$$

where P is a polynomial with p roots of modulus strictly bigger then 1 and Q is a polynomial with q roots of modulus ≥ 1, $P(0) = Q(0) = 1$, and P and Q have real coefficients. We then define the parameter space $F(p,q;\rho,u)$ as the space of all spectral densities of the previous form with all poles and zeros $\geq 1+\rho$, and $0 < u \leq \sigma^2 \leq 1/u$.
For any integrable function h on the torus, define $\widehat{h}_k$ the Fourier coefficient:

$$\widehat{h}_k = \int_{-\pi}^{\pi} e^{-ikx} h(e^{ix}) \frac{dx}{2\pi}$$

Define also the Toeplitz operator of order n, T_n, as the operator which associates to each integrable function h on the torus the $n \times n$ Toeplitz matrix $T_n(h)$:

$$(T_n(h))_{i,j} = \widehat{h}_{i-j}, \quad i,j = 1, \ldots, n$$

Define for any continuous function v the periodogram

$$I_n(v) :=^T X^{(n)} \cdot T_n(v) \cdot X^{(n)} = \int_{-\pi}^{\pi} v(e^{ix}) |\sum_1^n X_k e^{ikx}|^2 \frac{dx}{2\pi}$$

In case the process is a gaussian process, the logarithm of the likelihood is $L_n(f)$ with:

$$-2L_n(f) = n \log 2\pi + \log \det T_n(f) +^T X^{(n)} [T_n(f)]^{-1} X^{(n)}$$

It is well known (see for instance Azencott and Dacunha, 1986) that it is well approximated by the Whittle contrast function C_n, or pseudo likelihood, given by:

$$\begin{aligned} C_n(f) &= n \log 2\pi +^T X^{(n)} \cdot T_n(\frac{1}{f}) \cdot X^{(n)} + n \log \sigma^2 &\quad (13) \\ &= n \log 2\pi + n \log \sigma^2 + I_n(\frac{1}{f}) &\quad (14) \end{aligned}$$

which is a contrast function for the estimation of the parameters even if the process is not gaussian. Define the statistic:

$$U_n(p,q) = \inf_{f \in F(p,q;\rho,u)} C_n(f) - C_n(f_0).$$

The LRT (or pseudo LRT) for testing H_0: "$ARMA(p_0, q_0)$" against H_1: "$ARMA(p, q)$" is

$$W_n = U_n(p, q) - U_n(p_0, q_0).$$

We now define a locally conic parametrization. Let f_0 have zeros $1/u_i, i = 1, \ldots, q_0$ and poles $1/t_i, i = 1, \ldots, p_0$. We assume for the moment that all poles are distinct and all zeros are distinct. Define $r = \min(p - p_0, q - q_0)$. Suppose that $r = p - p_0$. The case where $r = q - q_0$ can be handled in the same manner. Let $s = q - q_0 - r$. As for mixtures, the locally conic parametrization is defined as a perturbation of the spectral density f_0, followed by a choice of $\mathcal{B}$ such that **(LC1)** and **(LC2)** hold. For any non negative θ, define

$$\begin{aligned} f_{(\theta,\beta)}(e^{ix}) &= \left(\frac{\sigma_0^2 + \frac{\theta}{N(\beta)}\delta}{2\pi}\right) \left| \frac{\prod_{i=1}^{q_0}(1 - (u_i + \frac{\theta}{N(\beta)}\mu_i)z)}{\prod_{i=1}^{p_0}(1 - (t_i + \frac{\theta}{N(\beta)}\tau_i)z)} \right|^2 \left| \prod_{i=1}^{r} \frac{(1 - (c_i + \epsilon_i \frac{\theta}{N(\beta)}\gamma_i)z)}{(1 - (c_i + (1 - \epsilon_i)\frac{\theta}{N(\beta)}\gamma_i)z)} \right|^2 \\ &\quad \cdot |(1 + \frac{\theta}{N(\beta)} \sum_{i=1}^{s} \nu_i z^i)|^2 \end{aligned} \tag{15}$$

with $z = e^{ix}$, and where, for $i = 1, \ldots, r$, $\epsilon_i = 1$ if $\inf_j |c_i - t_j| < \inf_j |c_i - u_j|$ and $\epsilon_i = 0$ if $\inf_j |c_i - t_j| \geq \inf_j |c_i - u_j|$. Here

$$\beta = (\delta, \mu, \tau, c, \gamma, \nu)$$

with $\delta \in \mathbb{R}$, $\mu = (\mu_j)_{j=1,\ldots,q_0} \in \mathbb{C}^{q_0}$, $\tau = (\tau_j)_{j=1,\ldots,p_0} \in \mathbb{C}^{q_0}$, $c = (c_j)_{j=1,\ldots,r} \in \mathbb{C}^r$, $|c_j| \leq 1/(1 + \rho)$, $\gamma = (\gamma_j)_{j=1,\ldots,r}\mathbb{C}^r$, $\nu = (\nu_j)_{j=1\ldots,s} \in \mathbb{C}^s$, and such that

$$\|\mu\|^2 + \|\tau\|^2 + \|\gamma\|^2 + \|\nu\|^2 = 1$$

The normalizing factor so as to set the directional Fisher informations to one is given by

$$\begin{aligned} N(\beta) &= \|\frac{\delta}{\sigma_0^2} + \sum_{i=1}^{p_0}(\frac{\tau_i z}{1 - t_i z} + \frac{\overline{\tau_i}}{z - \overline{t_i}}) - \sum_{i=1}^{q_0}(\frac{\mu_i z}{1 - u_i z} + \frac{\overline{\mu_i}}{z - \overline{u_i}}) \\ &\quad + \sum_{i=1}^{r}(\frac{(1 - 2\epsilon_i)\gamma_i z}{1 - c_i z} + \frac{(1 - 2\epsilon_i)\overline{\gamma_i}}{z - \overline{c_i}}) + \sum_{i=1}^{s}(\nu_i z^i + \frac{\overline{\nu_i}}{z^i})\|_H^2 \end{aligned}$$

where H is the Hilbert space $L^2([0, 2\pi]; \frac{dx}{2\pi})$.
Let us now define $\mathcal{B}$. Let f be any function in $F(p, q; \rho, u)$, and let it have zeros $1/v_j$, $j = 1, \ldots, q'$, $q_0 \leq q' \leq q$, and poles w_j, $j = 1, \ldots, p'$, $p_0 \leq p' \leq p$. Let ρ_1 be any permutation of $[1, \ldots, q']$ and ρ_2 any permutation of $[1, \ldots, p']$. Define then $P(z) = \prod_{j=1}^{p'}(1 - w_{\rho_2(j)}z)$, $Q(z) = \prod_{j=1}^{q'}(1 - v_{\rho_1(j)})$. The right choice of the permutations ρ_1 and

ρ_2 will lead to the desired local identifiability without loosing infinite differentiability. For this choice, use the following rule: choose first $\rho_1(j_1)$ such that

$$|v_{\rho_1(j_1)} - u_{j_1}| = \inf_{l,j} |v_l - u_j|$$

Then iterate the rule for the $q_0 - 1$ remaining zeros of f_0. Do the same for the p_0 poles of f_0, so that ρ_2 is defined for p_0 points, and the $\theta\mu_i/N(\beta)$ and $\theta\tau_i/N(\beta)$ are defined. It remains $p' - p_0$ poles and $q' - q_0$ zeros. Do the same coupling for r points relating r remaining poles to the r remaining poles. This defines the $\theta\gamma_i/N(\beta)$ and the ϵ_i (eventually some γ_i are null if $\inf(p' - p_0, q' - q_0) < r$). To end, complete ρ_2 in some way, so that it defines the $\theta\nu_i/N(\beta)$. Recall that the AR and MA polynomials have real coefficients if and only if the conjugate of each zero is a zero. The number of real parameters is then $p + q + 1$.
The set $\mathcal{T}$ is then the set of all obtained (θ, β) when f runs over $F(p, q; \rho, u)$. By construction, **(LC1)** and **(LC2)** hold.
Notice that, at least for small enough θ, the perturbation direction β may take any direction. In other words, the set of real parameters involved in $\tilde{\mathcal{B}}$ spans $\mathbb{R}^{p+q+1}$.
As for mixtures, we have for all (θ, β) the upper bound:

$$\frac{\theta}{N(\beta)} \leq 2(p_0 + q_0 + r + s)$$

We now define the derivative space $\mathcal{D}$ to be the subset of the unit sphere of H of functions of form

$$\begin{aligned} \frac{1}{N(\beta)} \quad & \Big(\frac{\delta}{\sigma_0^2} + \sum_{i=1}^{p_0}\Big(\frac{\tau_i z}{1 - t_i z} + \frac{\overline{\tau_i}}{z - \overline{t_i}}\Big) - \sum_{i=1}^{q_0}\Big(\frac{\mu_i z}{1 - u_i z} + \frac{\overline{\mu_i}}{z - \overline{u_i}}\Big) \\ & + \sum_{i=1}^{r}\Big(\frac{(1 - 2\epsilon_i)\gamma_i z}{1 - c_i z} + \frac{(1 - 2\epsilon_i)\overline{\gamma_i}}{z - \overline{c_i}}\Big) + \sum_{i=1}^{s}\Big(\nu_i z^i + \frac{\overline{\nu_i}}{z^i}\Big)\Big) \end{aligned} \tag{16}$$

with $z = e^{ix}$ and with β in $\tilde{\mathcal{B}}$. Define also ξ_d the gaussian process indexed by $\mathcal{D}$ with covariance the usual hilbertian product in H.
The following theorem states the asymptotic distribution of $U_n(p, q)$:

Theorem 4.1 *$U_n(p, q)$ converges in distribution to the following variable:*

$$-\sup_{d \in \mathcal{D}} \xi_d^2$$

The complete proof is given is section 5. As for mixtures, it relies on partitioning the parameter space so that, on each set, the converging parameter converges uniformly to 0 and the limiting distribution may be computed. We have first, if $\widehat{\theta}_\beta$ is a minimizer of $C_n(f_{(\theta,\beta)})$ for the fixed value of β:

Proposition 4.2 *Define* $\eta_n = \sup_{\beta \in \tilde{\mathcal{B}}} \widehat{\theta}_\beta$. *Then* η_n *converges to* 0 *in probability as* n *tends to infinity.*

Let

$$A_n = \{(\theta, \beta) \in \mathcal{T} \ : \ \frac{\theta}{N(\beta)^2} \leq \eta_n^\alpha, \theta \leq 2\eta_n\}$$

$$B_n = \{(\theta, \beta) \in \mathcal{T} \ : \ \frac{\theta}{N(\beta)^2} \geq \eta_n^\alpha, \theta \leq 2\eta_n\}$$

for some $\alpha \in]0, 1[$.
We have:

Lemma 4.3 $\mathcal{M}_n = \inf_{(\theta,\beta) \in A_n} C_n(f_{(\theta,\beta)}) - C_n(f_0)$ *converges in distribution to the following variable:*

$$-\sup_{d \in \mathcal{D}} \xi_d^2 1_{\xi_d \geq 0} \tag{17}$$

and:

Lemma 4.4 $\inf_{(\theta,\beta) \in A_n} C_n(f_{(\theta,\beta)}) - C_n(f_0)$ *is lower bounded by* $\mathcal{M}_n + o(1)$.

Lemma 4.3, Lemma 4.4, and the fact that $\mathcal{D}$ is a symmetrical set lead to Theorem 4.1.

Comments:

- In case $p_0 = q_0 = 0$ and $p = q = 1$, we recover the result of Hannan (1980), and in case $p_0 + 1 = p$ and $q_0 + 1 = q$, we recover the asymptotic result of Veres, Lemma 1 (1987).

- An important point is the identification of $\overline{\mathcal{D}}$, that is the study of the limit points in $\mathcal{D}$ when $N(\beta)$ tends to 0. This is why the ϵ_i appears in (15), so that when $N(\beta)$ tends to 0, $f_{(\theta,\beta)}$ tends to f_0. This happens when some c_i tend to some t_i with corresponding γ_i tending to the corresponding τ_i, or-and some c_i tend to some u_i with corresponding γ_i tending to the corresponding μ_i. $\overline{\mathcal{D}}$ is the subset of the unit sphere of H of functions of form

$$\frac{\delta}{\sigma_0^2} + \sum_{i=1}^{p_0} \left(\sum_{k=1}^{a_i} \tau_{i,k} (\frac{z}{1 - t_i z})^{a_i} + \overline{\tau_{i,k}} (\frac{1}{z - \overline{t_i}})^{a_i} \right) + \sum_{i=1}^{q_0} \left(\sum_{k=1}^{b_i} \mu_{i,k} (\frac{z}{1 - u_i z})^{b_i} + \overline{\mu_{i,k}} (\frac{1}{z - \overline{u_i}})^{b_i} \right)$$
$$- \sum_{i=1}^{r_1} (\frac{(1 - 2\epsilon_i)\gamma_i z}{1 - c_i z} + \frac{(1 - 2\epsilon_i)\overline{\gamma_i}}{z - \overline{c_i}}) + \sum_{i=1}^{s} (\nu_i z^i + \frac{\overline{\nu_i}}{z^i}) \tag{18}$$

  with $1 \leq a_i$, $i = 1, \ldots, p_0$, $1 \leq b_i$, $i = 1, \ldots, q_0$, $\sum_1^{p_0} a_i + \sum_1^{q_0} b_i - (p_0 + q_0) \leq r$, and $r_1 \leq r - [\sum_1^{p_0} a_i + \sum_1^{q_0} b_i - (p_0 + q_0)]$.

Define $\mathcal{D}_0$ the subset of the unit sphere of H of functions of form:

$$\frac{1}{N(\beta)}\left((\frac{\delta}{\sigma_0^2}+\sum_{i=1}^{p_0}(\frac{\tau_i z}{1-t_i z}+\frac{\overline{\tau_i}}{z-\overline{t_i}})-\sum_{i=1}^{q_0}(\frac{\mu_i z}{1-u_i z}+\frac{\overline{\mu_i}}{z-\overline{u_i}})\right) \tag{19}$$

The limiting distribution of the LRT statistic may now be derived in the same way as for mixtures.

Theorem 4.5 *The asymptotic distribution of W_n is that of:*

$$-\sup_{v\in\mathcal{V}}\xi_v^2 \tag{20}$$

where $\mathcal{V}$ is the set of normalized vectors of the orthogonal projection of $\mathcal{D}$ onto the orthogonal of the space spanned by $\mathcal{D}_0$.
This distribution is that of

$$-\sup_{(\tau_1,\ldots,\tau_r)\in[-T_\rho,T_\rho]^r}\left(Z(\tau_1)^2+\ldots+Z(\tau_r)^2\right)-W$$

where $T_\rho=\frac{1}{2}\log\frac{\rho+2}{\rho}$, $(Z(\tau))$ is the centered gaussian process with covariance function

$$C(s,t)=1/ch(t-s),$$

and W has distribution $\chi^2(s)$ and is independent $(Z(\tau))$.

Proof of Theorem 4.5.
In case $(p,q)=(p_0+1,q_0+1)$, the set $\mathcal{D}$ reduces to the subset $\mathcal{D}_{1,1}$ of the unit sphere of H of functions of form

$$\begin{aligned}\frac{1}{N(\beta)}\quad&(\frac{\delta}{\sigma_0^2}+\sum_{i=1}^{p_0}(\frac{\tau_i z}{1-t_i z}+\frac{\overline{\tau_i}}{z-\overline{t_i}})-\sum_{i=1}^{q_0}(\frac{\mu_i z}{1-u_i z}+\frac{\overline{\mu_i}}{z-\overline{u_i}})\\&+(\frac{(1-2\epsilon)\gamma z}{1-cz}+\frac{(1-2)\overline{\gamma}}{z-\overline{c}})\end{aligned}$$

Let $\mathcal{V}_1$ be the set of normalized vectors of the orthogonal projection of $\mathcal{D}_{1,1}$ onto the orthogonal of the space spanned by $\mathcal{D}_0$. By the result of Veres, Theorem 1 (1987), the distribution of $\sup_{v\in\mathcal{V}_1}\xi_v^2$ is that of $\sup_{\tau\in[-T_\rho,T_\rho]}Z(\tau)^2$.
Now, in case $r=0$ and $s>0$, the set $\mathcal{D}$ reduces to the subset $\mathcal{D}_s$ of the unit sphere of H of functions of form

$$\begin{aligned}\frac{1}{N(\beta)}\quad&(\frac{\delta}{\sigma_0^2}+\sum_{i=1}^{p_0}(\frac{\tau_i z}{1-t_i z}+\frac{\overline{\tau_i}}{z-\overline{t_i}})-\sum_{i=1}^{q_0}(\frac{\mu_i z}{1-u_i z}+\frac{\overline{\mu_i}}{z-\overline{u_i}})\\&+\sum_{i=1}^{s}(\nu_i z^i+\frac{\overline{\nu_i}}{z^i}).\end{aligned}$$

Let $\mathcal{V}_s$ be the set of normalized vectors of the orthogonal projection of $\mathcal{D}_s$ onto the orthogonal of the space spanned by $\mathcal{D}_0$. With $r = 0$, the overparametrized model is identifiable, the chi-square theory applies, and the asymptotic distribution of $\sup_{v\in\mathcal{V}_s} \xi_v^2$ is $\chi^2(s)$. Now looking at the linear form of $\mathcal{D}$, $\mathcal{V}$ is the set of normalized sums of r vectors in $\mathcal{V}_1$, to which s vectors in $\mathcal{V}_s$ are added, that can be decomposed also by orthogonal projection. Considering the linearity of the process ξ_v, Theorem 4.5 follows.

5 Proofs.

5.1 Proof of Theorem 3.1.

Using **(P0)**, $g_{(\theta,\beta)}$ is differentiable with respect to θ till order 5, and we have for all $2 \le h \le 5$, if $g^{(h)}_{(\theta,\beta)}$ is the h-th derivative of g with respect to θ at point (θ,β):

$$\begin{aligned} g^{(h)}_{(\theta,\beta)} &= \frac{h}{N(\beta)^h}\sum_{i_1\dots i_{h-1}=1}^{k}\sum_{l=1}^{q}\rho_l\delta^l_{i_1}\dots\delta^l_{i_{h-1}}D^{h-1}_{i_1\dots i_{h-1}}f_{\gamma^{l,0}+\frac{\theta}{N(\beta)}\delta^l} \\ &+\frac{1}{N(\beta)^h}\sum_{i_1\dots i_h=1}^{k}\sum_{l=1}^{q}(\pi^0_l+\rho_l\frac{\theta}{N(\beta)})\delta^l_{i_1}\dots\delta^l_{i_h}D^h_{i_1\dots i_h}f_{\gamma^{l,0}+\frac{\theta}{N(\beta)}\delta^l}\end{aligned}$$

and also

$$g'_{(\theta,\beta)} = \frac{1}{N(\beta)}\left(\sum_{l=1}^{p-q}\lambda_l f_{\gamma^l} + \sum_{l=1}^{q}\rho_l f_{\gamma^{l,0}+\frac{\theta}{N(\beta)}\delta^l} + \sum_{i=1}^{k}\sum_{l=1}^{q}(\pi^0_l+\rho_l\frac{\theta}{N(\beta)})\delta^l_i D^1_i f_{\gamma^{l,0}+\frac{\theta}{N(\beta)}\delta^l}\right)$$

The proof of the Theorem follows the same lines as that of Theorems 4.2 and 4.3 in Dacunha-Castelle and Gassiat (1996). However, the parametrization is not exactly the same, and has more terms, so that we detail the proof. It will aslo rely on some control Lemma that we set now.

Lemma 5.1 *Under **(P1)**, there exists a constant number a such that for β in $\tilde{\mathcal{B}}$:*

$$\frac{\sup_l \|\delta^l\|^2}{N(\beta)} \le a$$

Proof of Lemma 5.1.

First of all, using assumption **(P1)**, the hermitian matrix of all hilbertian scalar products involving functions in the free system is positive, so that it has a positive smallest eigenvalue σ, and the associated hermitian product is larger than σ multiplied by the usual scalar product in $\mathbb{R}^s$, $s = p + kq + p_2(k^2-1)$.

If $\frac{\|\delta^l\|^2}{N(\beta)}$ is unbounded, there exists a sequence β_n such that

$$\lim_{n\to+\infty}\frac{\|\delta^{l,n}\|}{N(\beta_n)} = +\infty$$

Since $\delta^{l,n}$ is bounded (see (7)), this implies that $N(\beta_n)$ tends to 0. Using assumption **(P1)**, this implies that $\delta^{l,n}$ tends to 0. Let now J_l be the set of indices i such that $\gamma^{i,n}$ tends to $\gamma^{l,0}$. We have:

$$\frac{\|\delta^{l,n}\|^2}{N(\beta_n)} \leq \frac{1}{\sigma} \frac{\|\delta^{l,n}\|^2}{(\|\sum_{j\in J_l}\lambda_j(\gamma^{j,n}-\gamma^{l,0})+\delta^l\|^2+\frac{1}{4}(\sum_{j\in J_l}\lambda_j\|\gamma^{j,n}-\gamma^{l,0})\|^2)^2)^{1/2}(1+o(1))}$$

Now, whether $\|\sum_{j\in J_l}\lambda_j(\gamma^{j,n}-\gamma^{l,0})\| = o(\|\delta^{l,n}\|)$ or $\|\delta^{l,n}\| = o(\|\sum_{j\in J_l}\lambda_j(\gamma^{j,n}-\gamma^{l,0})\|)$ or $\sum_{j\in J_l}\lambda_j(\gamma^{j,n}-\gamma^{l,0}) = -\delta^{l,n}(1+o(1))$ one can see that $\|\delta^{l,n}\|/N(\beta)$ is bounded and the Lemma follows.

Proof of Proposition 3.1.
Let $d(\beta)$ be the function in $\mathcal{D}$

$$d(\beta) = \frac{1}{N(\beta)}\left(\sum_{l=1}^{q}\pi_l^0\sum_{i=1}^{k}\delta_i^l\frac{D_i^1 f_{\gamma^{l,0}}}{g_0}+\sum_{i=1}^{p-q}\lambda_i\frac{f_{\gamma^i}}{g_0}+\sum_{l=1}^{q}\rho_l\frac{f_{\gamma^{l,0}}}{g_0}\right)$$

Let $(d(\beta_n))$ be a sequence of $\mathcal{D}$, and let us search all possible accumulation points in H. If $N(\beta_n)$ stays bounded below by a fixed positive number, **(P0)** implies that the accumulation points have form $d(\tilde{\beta})$. Else, there exists an integer $p_1 \leq p-q-1$ such that for the indices (eventually reordered) $l \leq p_1$, the γ^l do not converge to any of the $\gamma^{l,0}$. By **(P1)**, for the other $p-q-p_1$ indices, the γ^l converge to some $\gamma^{l,0}$. Let $L(p_1)$ be the set of indices of the limit points $\gamma^{l,0}$, and let $(J(l))_{l\in L(p_1)}$ be the partition such that for u in $J(l)$, γ^u converges to $\gamma^{l,0}$. Writing a Taylor expansion, and keeping only the leading terms, we have

$$\begin{aligned} d(\beta) \;=\; & (\frac{1}{N(\beta)}(\sum_{l=1}^{p_1}\lambda_l\frac{f_{\gamma^l}}{g_0}+\sum_{l=1}^{q}(\sum_{u\in J(l)}\lambda_u+\rho_l)\frac{f_{\gamma^{l,0}}}{g_0} \\ & +\sum_{l=1}^{q}\sum_{j=1}^{k}(\sum_{u\in J(l)}\lambda_u(\gamma_j^u-\gamma_j^{l,0})+\pi_l^0\delta_j^l)\frac{D_j^1 f_{\gamma^{l,0}}}{g_0} \\ & +\sum_{l\in L(p_1)}\sum_{j,j'=1}^{k}\frac{1}{2}(\sum_{u\in J(l)}\lambda_u(\gamma_j^u-\gamma_j^{l,0})(\gamma_{j'}^u-\gamma_{j'}^{l,0}))\frac{D_{jj'}^2 f_{\gamma^{l,0}}}{g_0}))(1+o(1)). \end{aligned} \tag{21}$$

$N(\beta)$ has the same expansion. The sequences of coefficients

$$\frac{\lambda_l}{N(\beta)},\quad \frac{\sum_{u\in J_l}\lambda_u+\rho_l}{N(\beta)},\quad \frac{\sum_{u\in J(l)}\lambda_u(\gamma_j^u-\gamma_j^{l,0})+\pi_l^0\delta_j^l}{N(\beta)}$$

$$\frac{\sum_{u\in J(l)}\lambda_u(\gamma_j^u-\gamma_j^{l,0})(\gamma_{j'}^u-\gamma_{j'}^{l,0})}{N(\beta)}$$

are bounded. Let the accumulations points be respectively μ_l, $\tilde{\rho}_l$, $\lambda_{l,i}$ and $\tau_u a_i^u a_j^u$. The accumulation points have thus form (10). Now, expansion (21) and assumption

(P0) allow to prove that $\overline{\mathcal{D}}$ has a $g_0\nu$ square integrable cover function. Let $\mathcal{N}(\epsilon)$ bet the bracketing number, that is the number of ϵ-brackets needed to cover $\overline{\mathcal{D}}$. An ϵ-bracket $[f,g]$ is the set of function d such that for all x, $f(x) \leq d(x) \leq g(x)$, and with $\|f-g\|_H \leq \epsilon$ (see Van der Vaart and Wellner, 1996). Again expansion (21) and assumption **(P0)** allow to prove that $\mathcal{N}(\epsilon)$ is at most of order $1/\epsilon^K$ with $K=(k+1)p$. Van der Vaart and Wellner (1996) p.129 leads to the conclusion.

Proof of lemma 3.4.
First, the following expansion holds for θ tending to 0:

$$l_n(\theta,\beta)-l_n(0)=\sum_{i=1}^n \frac{g_{(\theta,\beta)}-g_0}{g_0}(X_i)-\frac{1}{2}\sum_{i=1}^n(\frac{g_{(\theta,\beta)}-g_0}{g_0})^2(X_i)+\frac{1}{3}\sum_{i=1}^n U_i(\frac{g_{(\theta,\beta)}-g_0}{g_0}(X_i))^3 \tag{22}$$

where $|U_i|\leq 1$. Let us now write an expansion of $g_{(\theta,\beta)}$ till order 2:

$$g_{(\theta,\beta)}(x)=g_0(x)+\theta\cdot g'_{(0,\beta)}(x)+\frac{\theta^2}{2}\cdot g"_{(\theta*,\beta)}(x)$$

for a $\theta*\leq\theta$ and depending on x. Now as θ tends to 0:

$$\begin{aligned} g"_{(\theta*,\beta)}(x) &= \frac{2}{N(\beta)^2}\sum_{l=1}^q\sum_{i=1}^k \rho_l\delta_i^l D_i^1 f_{\gamma^{l,0}}(x)+\frac{1}{N(\beta)^2}\sum_{l=1}^q\sum_{i,j=1}^k \pi_l^0\delta_i^l\delta_j^l D_{ij}^2 f_{\gamma^{l,0}}(x) \\ & 0((\sup_l \frac{\|\delta^l\|^2}{N(\beta)^3})\theta(m_2(x)+m_3(x))g_0(x)) \end{aligned}$$

since δ^l is bounded and using **(P0)**. Write, if $g'_{(0,\beta)}=\frac{\partial}{\partial\theta}g_{(0,\beta)}$:

$$D_n(\beta)=\sum_{i=1}^n \frac{g'_{(0,\beta)}}{g_0}(X_i)$$

$$F_n(\beta)=\sum_{i=1}^n\sum_{l=1}^q\sum_{i=1}^k \rho_l\delta_i^l\frac{D_i^1 f_{\gamma^{l,0}}}{g_0}(X_i)$$

$$G_n(\beta)=\sum_{i=1}^n\sum_{l=1}^q\sum_{i,j=1}^k \pi_l^0\delta_i^l\delta_i^l\frac{D_{ij}^2 f_{\gamma^{l,0}}}{g_0}(X_i)$$

Now, for any $l=1,\ldots,q$:

$$\frac{\|\delta^l\|}{N(\beta)^2}\theta\leq\eta_n^{1-\alpha}$$

and since $(D_n(\beta),F_n(\beta),G_n(\beta))/\sqrt{n}$ converges uniformly in distribution using Proposition 3.1, we have easily

$$\frac{\theta^2}{N(\beta)^2}F_n(\beta)=o(\theta D_n(\beta))$$

and

$$\frac{\theta^2}{N(\beta)^2}G_n(\beta) = o(\theta D_n(\beta))$$

where the $o(\cdot)$ are uniform in probability over β in A_n. Also, applying lemma 5.1 we obtain for any $l = 1, \ldots, q$:

$$\frac{\|\delta^l\|^2}{N(\beta)^3}\theta \leq \eta_n^{1-4\alpha/3}$$

which goes to 0 since $\alpha < 3/4$. We finally get, looking at all terms in (22), for β in A_n and for $\theta \leq 2\eta_n$:

$$l_n(\theta, \beta) - l_n(0) = \left(\theta D_n(\beta) - \frac{\theta^2}{2}n\right)(1 + o(1))$$

where again the $o(\cdot)$ is uniform in probability over β in A_n. Since $\widehat{\theta}_\beta \leq \eta_n$ this obviously leads, by maximizing $\theta D_n(\beta) - \frac{\theta^2}{2}n$ to:

$$l_n(\widehat{\theta}_\beta, \beta) - l_n(0) = \frac{1}{2}\frac{D_n(\beta)^2}{n}1_{D_n(\beta)\geq 0}(1 + o(1))$$

for β in A_n and where the $o(\cdot)$ is uniform in probability over β in A_n. Define

$$\mathcal{D}_n = \left\{\frac{g'_{(0,\beta)}}{g_0},\ \beta \in A_n\right\}$$

We have $\cup_n \mathcal{D}_n = \mathcal{D}$, and using Proposition 3.1

$$\mathcal{L}_n = \frac{1}{2}\sup_{d\in\mathcal{D}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n d(X_i)\right)^2 1_{\frac{1}{\sqrt{n}}\sum_{i=1}^n d(X_i)\geq 0}(1 + o(1)) \tag{23}$$

The conclusion of Lemma 3.4 follows using again Proposition 3.1.

Let us now study what happens on B_n. First of all, notice that on B_n, $N(\beta)$ tends to 0, and using **(P1)** all δ^l tend to 0. We have as an immediate consequence of Lemma 5.1:

Lemma 5.2 *There exists a constant number M such that for β in B_n:*

$$N(\beta) \leq M\eta_n^{2\alpha/3}$$

and for any (i, l):

$$|\delta_i^l| \leq M\eta_n^{\alpha/3}$$

Proof of Lemma 5.2.
Define $\phi_{i,l} = \frac{\delta_i^l}{N(\beta)^2}$. We first have, using Lemma 5.1

$$N(\beta) \leq \frac{a^{1/3}}{|\phi_{i,l}|^{2/3}}$$

Now, on B_n, there exists (i,l) such that $|\phi_{i,l}| \geq 1/k\eta_n^\alpha$, and the first inequality of the Lemma follows. Now, using Lemma 5.1 we have for all (i,l)

$$|\delta_i^l| \leq \sqrt{aN(\beta)}$$

and the second inequality follows .
In other words, on B_n, $N(\beta)$ and all $|\delta_i^l|$ tend uniformly to 0.

Proof of lemma 3.5.
We shall use again expansion (22), but the expansion for $g_{(\theta,\beta)}$ has now to be done till order 5:

$$g_{(\theta,\beta)}(x) = g_0(x) + \theta \cdot g'_{(0,\beta)}(x) + \sum_{i=2}^{4} \frac{\theta^i}{i!} \cdot g^{(i)}_{(0,\beta)}(x) + \frac{\theta^5}{5!} \cdot g^{(5)}_{(\theta*,\beta)}(x)$$

for a $\theta* \leq \theta$ and depending on x.
Define

$$d_1(\beta) = \sum_{l=1}^{q}\sum_{i=1}^{k} \rho_l \delta_i^l \frac{D_i^1 f_{\gamma^{l,0}}}{g_0}$$

$$U(\beta) = \sum_{l=1}^{q}\sum_{i=1}^{k} \rho_l \delta_i^l u(\beta,i,l) = \langle d_1(\beta), \frac{g'_{(0,\beta)}}{g_0}\rangle_H$$

$$S(\beta) = \|d_1(\beta)\|_H^2$$

Define also $P_n(\theta,\beta)$ the polynomial of degree 4 in the variable θ:

$$\begin{aligned} P_n(\theta,\beta) &= \theta D_n(\beta) - \frac{n\theta^2}{2} + \frac{\theta^2}{N(\beta)^2}F_n(\beta) - \frac{n\theta^3}{N(\beta)^2}U(\beta) - \frac{n\theta^4}{2N(\beta)^4}S(\beta) \qquad (24) \\ &= \sum_{j=1}^{4} p_j(n,\beta)\theta^j \end{aligned}$$

The aim is now to prove that for $\theta \leq 2\eta_n$ and for $\beta \in B_n$ we have:

$$l_n(\theta,\beta) - l_n(0) = P_n(\theta,\beta)(1 + o(1)) \qquad (25)$$

where all the $o(\cdot)$ are uniform in probability over β in B_n. From now on, any $o(1)$ will be uniform over B_n. Since $|\delta_i^l|$ tends uniformly to 0, we have

$$\frac{\delta_{i_1}^l \ldots \delta_{i_h}^l}{N(\beta)^h} = o\left(\frac{\delta_{i_1}^l \ldots \delta_{i_{h-1}}^l}{N(\beta)^h}\right)$$

So that we can write for $h \leq 4$:

$$g^{(h)}_{(\theta,\beta)} = \left(\frac{h}{N(\beta)^h} \sum_{i_1 \ldots i_{h-1}=1}^{k} \sum_{l=1}^{q} \rho_l \delta^l_{i_1} \ldots \delta^l_{i_{h-1}} D^{h-1}_{i_1 \ldots i_{h-1}} f_{\gamma^{l,0}+\frac{\theta}{N(\beta)}\delta^l} \right) (1 + o(1)).$$

We obtain

$$l_n(\theta,\beta) - l_n(0) = P_n(\theta,\beta) + R_n$$

where R_n is a sum of terms which are $o(Q_n(\theta,\beta))$, with $Q_n(\theta,\beta) = \sup_{1\leq j\leq 4} |p_j(n,\beta)\theta^j|$, plus terms which may be bounded with one of the following forms:

$$\frac{\theta^h \delta^l_{i_1} \ldots \delta^l_{i_{h-1}}}{N(\beta)^h} \sum_{i=1}^{n} \frac{D^{h-1}_{i_1 \ldots i_{h-1}} f_{\gamma^{l,0}}}{g_0}(X_i) \quad h \geq 3$$

$$\frac{\theta^5 \|\delta^l\|^5}{N(\beta)^5} n \; ; \; \frac{\theta^{h+1} \delta^l_{i_1} \ldots \delta^l_{i_{h-1}}}{N(\beta)^h} n \; ; \; \frac{\theta^{h+l} \delta^l_{i_1} \ldots \delta^l_{i_{h-1}} \delta^l_{j_1} \ldots \delta^l_{j_{l-1}}}{N(\beta)^{h+l}} n \text{ with } h,l \geq 3$$

$$\theta^3 n \; ; \; \frac{\theta^{h+2} \delta^l_{i_1} \ldots \delta^l_{i_{h-1}}}{N(\beta)^h} n \; ; \; \frac{\theta^{h+l+1} \delta^l_{i_1} \ldots \delta^l_{i_{h-1}} \delta^l_{j_1} \ldots \delta^l_{j_{l-1}}}{N(\beta)^{h+l}} n$$

$$\frac{\theta^{h+l+m} \delta^l_{i_1} \ldots \delta^l_{i_{h-1}} \delta^l_{j_1} \ldots \delta^l_{j_{l-1}} \delta^l_{k_1} \ldots \delta^l_{k_{m-1}}}{N(\beta)^{h+l+m}} n \text{ with } h,l,m \geq 2$$

Now, since $\theta/N(\beta)$ and the δ^l_i are bounded, and using Lemma 5.1, Lemma 5.2 and the fact that β is in B_n, each term can be bounded by $o(n\theta^2)$ or $o(n\theta^4\|\delta^l\|^2/N^4(\beta))$. So every term of R_n is $o(Q_n(\theta,\beta))$ uniformly in β. It is not possible now to conclude that (24) holds since we need $o(P_n(\theta,\beta))$ instead of $o(Q_n(\theta,\beta))$ and it could be that $Q_n(\theta,\beta)$ be much larger than $P_n(\theta,\beta)$.
First, we prove that

$$\sup_{\theta\leq 2\eta_n, \beta\in B_n} P_n(\theta,\beta) \leq \mathcal{L}_n(1+o(1)). \tag{26}$$

Let $\phi = \frac{1}{N(\beta)^2}$. Define, considering ϕ and β as different variables (in fact ϕ is a function of β):

$$Y_n(\beta,\phi,\theta) = \theta D_n(\beta) - \frac{n}{2}\theta^2 + (\theta^2 F_n(\beta) - \theta^3 n U(\beta)) \cdot \phi - \frac{\theta^4}{2} n S(\beta) \cdot \phi^2$$

We have $P_n(\theta,\beta) = Y_n(\beta,\phi,\theta)$, so that

$$\sup_{\theta\leq 2\eta_n, \beta\in B_n} P_n(\theta,\beta) \leq \sup_{\beta\in B_n} Z_n(\beta)$$

where $Z_n(\beta) = \sup_{\phi,\theta\geq 0} Y_n(\beta,\phi,\theta)$. Optimizing in ϕ then in θ gives:

$$\phi = \frac{1}{n\theta^2}\frac{F_n(\beta)}{S(\beta)} - \frac{1}{\theta}\frac{U(\beta)}{S(\beta)}$$

$$\theta = \frac{1}{n}\left(\frac{D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)}}{1 - \frac{U(\beta)^2}{S(\beta)}}\right) 1_{D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)} \geq 0}$$

and

$$\begin{aligned} Z_n(\beta) &= \frac{1}{2n}\left(\frac{(D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)})^2}{1 - \frac{(U(\beta))^2}{S(\beta)}}\right) 1_{D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)} \geq 0} \\ &+ \frac{1}{2n}\left(\frac{(F_n(\beta))^2}{S(\beta)}\right). \end{aligned}$$

For any real number λ, and any $\mu \geq 0$ such that $\lambda^2 + \mu^2 = 1$, the function

$$d(\lambda, \mu, \beta) = \lambda d_1(\beta) + \mu \left(\frac{\frac{g'_{(0,\beta)}}{g_0} - d_1(\beta)\frac{U(\beta)}{S(\beta)}}{1 - \frac{U^2(\beta)}{S(\beta)}}\right)$$

is in $\mathcal{D}$, and

$$Z_n(\beta) = \frac{1}{2} \sup_{\mu \geq 0, \lambda^2+\mu^2=1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} d(\lambda, \mu, \beta)(X_i)\right)^2 1_{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} d(\lambda,\mu,\beta)(X_i) \geq 0}$$

by a straightforward computation of the last supremum. Thus

$$\sup_{\theta \leq 2\eta_n, \beta \in B_n} P_n(\theta, \beta) \leq \frac{1}{2} \sup_{d \in \mathcal{D}} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} d(X_i)\right)^2 1_{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} d(X_i) \geq 0}$$

so that, applying (23), (26) follows. Let us now prove that when at least one term $p_j(n, \beta)\theta^j$ of P_n tends to $+\infty$, then this implies that $P_n(\theta, \beta)$ tends to $-\infty$, and that at the optimizing value $(\widehat{\theta}, \widehat{\beta})$, all terms in P_n have the same order and $R_n = o(P_n(\widehat{\theta}, \widehat{\beta}))$. We shall conclude that, in the neighborhood of $(\widehat{\theta}, \widehat{\beta})$, (25) will hold and, applying (26), Lemma 3.5 will follow.
We shall need one technical Lemma which is a simple consequence of **(P1)**.

Lemma 5.3 *There exists $\tau > 0$ such that, with $\phi_{i,l} = \frac{\delta_i^l}{N(\beta)^2}$:*

$$\sum_{l,l'=1}^{q} \sum_{i,i'=1}^{k} \rho_l \rho_{l'} \phi_{i,l} \phi_{i',l'} nv(i, l; i', l') \geq \tau \sum_{l=1}^{q} \sum_{i=1}^{k} \rho_l^2 \phi_{i,l}^2 nv(i, l; i, l)$$

- If $n\theta^2 \to +\infty$, notice that

$$P_n(\theta, \beta) = \sqrt{n}\theta \frac{D_n(\beta)}{\sqrt{n}} - \frac{n}{2}\theta^2 \mathrm{Var}[\frac{D_n(\beta)}{\sqrt{n}} + \frac{\theta F_n(\beta)}{N(\beta)^2\sqrt{n}}]$$

  so that $P_n \to -\infty$.

- If $n\theta^2 = O(1)$,
  - either there exists l such that $\theta\|\delta^l\|/N(\beta)^2 \to +\infty$, in which case Lemma 5.3 proves that the larger term in P_n is $-\theta^4 p_4(n,\beta)$ and $P_n \to -\infty$
  - or $\sup_l \theta\|\delta^l\|/N(\beta)^2 = O(1)$, in which case $R_n = o(n\theta^2)$, so that $R_n = o(1)$. In case the maximum value of P_n is $o(1)$, it is $o(\mathcal{L}_n)$, and in case it is not $o(1)$, (25) holds in the neighborhood of the maximizing value.

5.2 Proof of Theorem 4.1.

We shall study derivatives of $C_n(f_{(\theta,\beta)})$ with respect to θ for fixed β. Define

$$\begin{aligned}
e_{(\theta,\beta)} &= \frac{1}{N(\beta)}\Big(\frac{\delta}{\sigma_0^2 + \frac{\theta\delta}{N(\beta)}} + \sum_{i=1}^{p_0}\Big(\frac{\tau_i z}{1-(t_i + \frac{\theta}{N(\beta)}\tau_i)z} + \frac{\overline{\tau_i}}{z - (\overline{t_i} + \frac{\theta}{N(\beta)}\overline{\tau_i})}\Big)\\
&\quad - \sum_{i=1}^{q_0}\Big(\frac{\mu_i z}{1-(u_i + \frac{\theta}{N(\beta)}\mu_i)z} + \frac{\overline{\mu_i}}{z - (\overline{u_i} + \frac{\theta}{N(\beta)}\overline{\mu_i})}\Big)\\
&\quad - \sum_{i=1}^{r}\Big(\frac{\gamma_i z}{1-(c_i + \frac{\theta}{N(\beta)}\gamma_i)z} + \frac{\overline{\gamma_i}}{z - (\overline{c_i} + \frac{\theta}{N(\beta)}\overline{\gamma_i})}\Big) + \sum_{i=1}^{s}\Big(\nu_i z^i + \frac{\overline{\nu_i}}{z^i}\Big)\Big)
\end{aligned}$$

and $e^{(k)}_{(\theta,\beta)}$ to be the partial derivative of $e_{(\theta,\beta)}$ with respect to θ. $e_{(\theta,\beta)}$ is an element of $\mathcal{D}$. Let $C'_n(\theta,\beta)$ be the partial derivative of $C_n(f_{(\theta,\beta)})$ with respect to θ, and $C_n^{(k)}(\theta,\beta)$ the k-th partial derivative of $C_n(f_{(\theta,\beta)})$ with respect to θ. We have for $k \geq 1$:

$$\begin{aligned}
(-1)^k e^{(k)}_{(\theta,\beta)} &= \frac{1}{N(\beta)^{k+1}}\Big[\sum_{i=1}^{p_0}\Big(\Big(\frac{\tau_i z}{1-(t_i + \frac{\theta}{N(\beta)}\tau_i)z}\Big)^{k+1} + \Big(\frac{\overline{\tau_i}}{z - (\overline{t_i} + \frac{\theta}{N(\beta)}\overline{\tau_i})}\Big)^{k+1}\Big)\\
&\quad - \sum_{i=1}^{q_0}\Big(\Big(\frac{\mu_i z}{1-(u_i + \frac{\theta}{N(\beta)}\mu_i)z}\Big)^{k+1} + \Big(\frac{\overline{\mu_i}}{z - (\overline{u_i} + \frac{\theta}{N(\beta)}\overline{\mu_i})}\Big)^{k+1}\Big)\\
&\quad - \sum_{i=1}^{r}\Big(\Big(\frac{\gamma_i z}{1-(c_i + \frac{\theta}{N(\beta)}\gamma_i)z)^{k+1}} + \Big(\frac{\overline{\gamma_i}}{z - (\overline{c_i} + \frac{\theta}{N(\beta)}\overline{\gamma_i})}\Big)^{k+1}\Big) + \frac{k!(\frac{\delta}{N(\beta)})^{k+1}}{(\sigma_0^2 + \theta\frac{\delta}{N(\beta)})^{k+1}}\Big]
\end{aligned}$$

and also

$$\begin{aligned}
C'_n(\theta,\beta) &= n\frac{\frac{\delta}{N(\beta)}}{\sigma_0^2 + \theta\frac{\delta}{N(\beta)}} - I_n\left(\frac{e_{(\theta,\beta)}}{f_{(\theta,\beta)}}\right)\\
C''_n(\theta,\beta) &= -n\frac{(\frac{\delta}{N(\beta)})^2}{(\sigma_0^2 + \theta\frac{\delta}{N(\beta)})^2} + I_n\left(\frac{e^2_{(\theta,\beta)} - e'_{(\theta,\beta)}}{f_{(\theta,\beta)}}\right)\\
C_n^{(3)}(\theta,\beta) &= n\frac{2(\frac{\delta}{N(\beta)})^3}{(\sigma_0^2 + \theta\frac{\delta}{N(\beta)})^3} + \frac{1}{N(\beta)^2}I_n\left(K_{(\theta,\beta)}\right)
\end{aligned}$$

$$
\begin{aligned}
C_n^{(4)}(\theta,\beta) &= -n\frac{6(\frac{\delta}{N(\beta)})^4}{(\sigma_0^2+\theta\frac{\delta}{N(\beta)})^4}+\frac{1}{N(\beta)^4}I_n\left(S_{(\theta,\beta)}\right)\\
C_n^{(5)}(\theta,\beta) &= n\frac{24(\frac{\delta}{N(\beta)})^5}{(\sigma_0^2+\theta\frac{\delta}{N(\beta)})^5}+\frac{1}{N(\beta)^5}I_n\left(T_{(\theta,\beta)}\right)
\end{aligned}
$$

We shall make use of a Lemma which is a consequence of Theorem 2.5 of Dahlhaus (1988). We prove it below.

Let $\mathcal{G}$ be the subset of H:

$$
\mathcal{G}=\left\{\frac{e_{(0,\beta)}}{f_0},\frac{N(\beta)^2e'_{(0,\beta)}}{f_0},\frac{e^2_{(0,\beta)}}{f_0},K_{(0,\beta)}\ :\ \beta\in\tilde{\mathcal{B}}\right\}\cup\left\{T_{(\theta,\beta)},S_{(\theta,\beta)}\ :\ (\theta,\beta)\in\mathcal{T}\right\}
$$

Define also for any function g in $\mathcal{G}$

$$
I(g)=\frac{1}{2\pi}\int_0^{2\pi}g(\lambda)f(\lambda)d\lambda
$$

Let $\mathcal{X}$ be the set of bounded real functions on $\mathcal{G}$, equipped with the metric generated by the uniform norm $\|x\|=\sup|x(g)|$.

Lemma 5.4 *Let $E_n(g)=\frac{1}{\sqrt{n}}I_n(g)-\sqrt{n}I(g)$, where g is in $\mathcal{G}$. Let $(W(g))_{g\in\mathcal{G}}$ be the centered gaussian process with covariance the Hilbert product in H. Then the empirical spectral process $(E_n(g))_{g\in\mathcal{G}}$ converges weakly on $\mathcal{X}$ to $(W(g))_{g\in\mathcal{G}}$.*

Proof of Lemma 4.3.

Taylor expansion till order 4 with integral remaining term, together with Lemma 5.4 lead to:

$$
\begin{aligned}
C_n(\theta,\beta)-C_n(0) &= -\theta\sqrt{n}E_n(\frac{e_{(0,\beta)}}{f_0})(1+o(1))+\frac{1}{2}n\theta^2(1+o(1))\\
&\quad -\frac{\theta^2}{2N(\beta)^2}\sqrt{n}E_n(\frac{N(\beta)^2e'_{(0,\beta)}}{f_0})(1+o(1))\\
&\quad +\frac{\theta^3}{6N(\beta)^2}na(1+o(1))+\frac{\theta^4}{24N(\beta)^4}nb(1+o(1))
\end{aligned}
$$

where

$$
a=I(K_{(0,\beta)})+\frac{2(\frac{\delta}{N(\beta)})^3}{\sigma_0^6}
$$

and

$$
b=I(S_{(0,\beta)})+\frac{6(\frac{\delta}{N(\beta)})^4}{\sigma_0^8}
$$

and the $o(1)$ are uniform in probability over A_n. Now, on A_n we have

$$\frac{\theta^2}{2N(\beta)^2}\sqrt{n}E_n(\frac{N(\beta)^2 e'_{(0,\beta)}}{f_0}) = O(\theta\sqrt{n}E_n(\frac{N(\beta)^2 e'_{(0,\beta)}}{f_0})\eta_n^{\alpha}) = o(1)$$

using Lemma 5.4, and

$$\frac{\theta^3}{6N(\beta)^2}na = O(n\theta^2\eta_n^{2\alpha}) = o(n\theta^2)$$

where the $o(.)$ are uniform in probability, so that we have

$$C_n(\theta,\beta) - C_n(0) = -\theta\sqrt{n}E_n(\frac{e_{(0,\beta)}}{f_0})(1+o(1)) + \frac{1}{2}n\theta^2(1+o(1))$$

When minimizing over θ this leads to

$$-\frac{1}{2}\frac{E_n(\frac{e_{(0,\beta)}}{f_0})^2}{n}1_{E_n(\frac{e_{(0,\beta)}}{f_0})\geq 0}(1+o(1))$$

The set of functions $\frac{e_{(0,\beta)}}{f_0}$ is exactly $\mathcal{D}$ so that Lemma 4.3 holds.

Let us now study what happens on B_n. As for the mixtures, a key point of the proof will be a control Lemma to be able to stop Taylor's expansion and to have uniformly $o(1)$ remaining terms. We set it now.

Lemma 5.5 *On B_n we have*

$$N(\beta) \leq 2\eta_n^{(1-\alpha)/2}$$

and

$$\frac{\theta}{N(\beta)} \leq M\eta_n^{(1-\alpha)/2r}$$

for some constant number M.

First of all, since on B_n, $\theta \leq 2\eta_n$, we have

$$N(\beta)^2 \leq 2\eta_n^{(1-\alpha)}$$

and the first inequality follows. Now, let us study what happens when $N(\beta)$ tends to 0. There must be at least one c_i tending to some u_i with corresponding γ_i tending to corresponding μ_i, or (and) some c_i tending to some t_i with corresponding γ_i tending to corresponding τ_i. In each case, we have, due to the locally conic structure (coming from the choice of permutation):

$$\frac{\theta}{N(\beta)}|\mu_i|(1+o(1)) \leq (|c_i - u_i|)$$

or (and)

$$\frac{\theta}{N(\beta)}|\tau_i|(1+o(1)) \leq (|c_i - t_i|)$$

Now, looking at the expansion of $N(\beta)$ near 0, it appears that the leading term is at least of order $\min_i(|c_i - u_i|^r, |c_i - t_i|^r)$, so that

$$\frac{\theta}{N(\beta)} = 0(N(\beta)^{1/r})$$

which, when combined with the first inequality, leads to the second inequality.

Proof of Lemma 4.4.

On B_n, we shall write Taylor expansion till order 5 again with integral remaining term, so that, when using Lemma 5.4, we obtain:

$$\begin{aligned} C_n(\theta,\beta) - C_n(0) &= -\theta\sqrt{n}E_n(\frac{e_{(0,\beta)}}{f_0})(1+o(1)) + \frac{1}{2}n\theta^2(1+o(1)) \\ &\quad -\frac{\theta^2}{2N(\beta)^2}\sqrt{n}E_n(\frac{N(\beta)^2 e'_{(0,\beta)}}{f_0})(1+o(1)) \\ &\quad +\frac{\theta^3}{6N(\beta)^2}na(1+o(1)) + \frac{\theta^4}{24N(\beta)^4}nb(1+o(1)) + O(\frac{\theta^5}{N(\beta)^5}n)(1+o(1)) \end{aligned}$$

where the $o(1)$ are in probability uniform over B_n. All $o(1)$ will be now uniform over B_n using Lemma 5.5. We have

$$\frac{\theta^5}{N(\beta)^5}n = o(\frac{\theta^4}{N(\beta)^4}n)$$

Notice that the functions $\frac{e'_{(0,\beta)}}{f_0}$, after normalization, are in $\overline{\mathcal{D}}$.
Define for $\phi = 1/N(\beta)^2$ the polynomial

$$P_n(\theta,\phi) = -\theta\sqrt{n}W_1^n + +\frac{1}{2}n\theta^2 - \frac{\theta^2}{2}\phi\sqrt{n}W_2^n + \frac{\theta^3}{2}n\phi C_{12} + \frac{\theta^4}{8}n\phi^2 C_{22}$$

with

$$W_1^n = E_n(\frac{e_{(0,\beta)}}{f_0}) \;\; W_2^n = E_n(\frac{N(\beta)^2 e'_{(0,\beta)}}{f_0})$$

and, up to a factor $1+o(1)$, C_{12} is the covariance of W_1^n and W_2^n, C_{22} the variance of W_2^n, W_1^n being of unit variance.
As for mixtures we obtain on B_n:

$$C_n(\theta,\beta) = P_n(\theta,\phi)(1+o(1))$$

Minimizing P_n over ϕ, then over θ leads to

$$\phi = \frac{1}{C_{22}}\left(\frac{2W_2^n}{\sqrt{n}\theta^2} - \frac{2C_{12}}{\theta}\right)$$

$$\theta = \frac{1}{\sqrt{n}}\frac{W_1^n - W_2^n\frac{C_{12}}{C_{22}}}{1-\frac{C_{12}^2}{C_{22}^2}} 1_{W_1^n - W_2^n\frac{C_{12}}{C_{22}}\geq 0}$$

with minimum value

$$-\frac{1}{2}\left(\frac{(W_2^n)^2}{C_{22}} + \frac{(W_1^n - W_2^n\frac{C_{12}}{C_{22}})^2}{1-\frac{C_{12}^2}{C_{22}^2}} 1_{W_1^n - W_2^n\frac{C_{12}}{C_{22}}\geq 0}\right)(1+o(1))$$

The end of the proof proceeds as for Lemma 3.5 and will be omitted.

Proof of Lemma 5.4.

The proof proceeds by a verification of the assumptions used in Theorem 2.5 of Dahlhaus (1988), that is his assumption 2.1.

Assumption (a) is verified since the process is an ARMA process and the spectral density has infinitely many derivatives, all bounded.

(b) is verified since the tapering is the constant 1.

Let us now verify (c).

$\mathcal{G}$ is a permissible subset of H (in the sense of Pollard, 1984 Appendix C) since it is a parametric class of functions which is pointwise continuous in the interior, and may be approached by sequences of parameters on the boundary.

Let us recall that for all x:

$$m_0 = \frac{1}{2\pi u}\left(\frac{(1-\frac{1}{1+\rho})^{q_0}}{2^{p_0}}\right)^2 \leq |f_0(x)| \leq \frac{u}{2\pi}\left(\frac{2^{q_0}}{(1-\frac{1}{1+\rho})^{p_0}}\right)^2 = M_0$$

It is then enough to prove that the functions in $\mathcal{G}$ are uniformly pointwise bounded, and to verify the entropy condition.

Moreover, since the functions $S_{(\theta,\beta)}$, $T_{(\theta,\beta)}$, $\frac{N(\beta)^2 e'_{(0,\beta)}}{f_0}$ are bounded functions of bounded parameters, which are continuous both pointwise and in H, with the square of the norm which is a quadratic function of some of the parameters, so that the entropy condition is verified. it is enough to verify the conditions for the set of functions

$$\frac{N(\beta)^2 e^3_{(0,\beta)}}{f_0},\ \frac{N(\beta)^2 e'_{(0,\beta)} e_{(0,\beta)}}{f_0},\ \frac{N(\beta)^2 e"_{(0,\beta)}}{f_0}$$

It is again enough to verify that the functions

$$\frac{e_{(0,\beta)}}{f_0},\ \frac{N(\beta)^2 e"_{(0,\beta)}}{f_0}$$

are uniformly pointwise bounded, and that the set of such functions verify the entropy condition. This in turns implies the entropy condition for the whole set of functions. To see that they are bounded, let us look at a precise expansion when $N(\beta)$ tends to 0. Let for any $i = 1, \ldots, q_0$ $U(i)$ be the set of indices j such that c_j tends to u_i, and $c_j = u_i + \alpha_j(u_i)$, and for any $i = 1, \ldots, p_0$ $T(i)$ be the set of indices j such that c_j tends to t_i, and $c_j = t_i + \alpha_j(t_i)$. Let also J be the complementary set of the union of all $U(i)$ and $T(i)$ in $\{1, \ldots, r\}$. We then have:

$$
\begin{aligned}
e_{(0,\beta)} &= \frac{1}{N(\beta)}[\frac{\delta}{\sigma_0^2} + \sum_{i=1}^{p_0}(\frac{(\tau_i - \sum_{j\in T(i)}\gamma_j)z}{1 - t_i z} + \frac{\overline{\tau_i} - \sum_{j\in T(i)}\overline{\gamma_j}}{z - (\overline{t_i}})\\
&+ \sum_{i=1}^{p_0}(\sum_{j\in T(i)}\gamma_j \sum_{h\geq 2}\frac{\alpha_j^{h-1}z^h}{(1 - t_i z)^h} + \overline{\gamma_j}\sum_{h\geq 2}\frac{\overline{\alpha_j}^{h-1}}{(z - \overline{t_i})^h})\\
&- \sum_{i=1}^{q_0}(\frac{(\mu_i - \sum_{j\in U(i)}\gamma_j)z}{1 - u_i z} + \frac{\overline{\mu_i} - \sum_{j\in U(i)}\overline{\gamma_j}}{z - \overline{u_i}})\\
&- \sum_{i=1}^{p_0}(\sum_{j\in U(i)}\gamma_j \sum_{h\geq 2}\frac{\alpha_j^{h-1})z^h}{(1 - u_i z)^h} + \overline{\gamma_j}\sum_{h\geq 2}\frac{\overline{\alpha_j}^{h-1})}{(z - \overline{u_i})^h})\\
&- \sum_{i\in J}(\frac{\gamma_i z}{1 - (c_i + \frac{\theta}{N(\beta)}\gamma_i)z} + \frac{\overline{\gamma_i}}{z - (\overline{c_i} + \frac{\theta}{N(\beta)}\overline{\gamma_i})}) - \sum_{i=1}^{s}(\nu_i z^i + \frac{\overline{\nu_i}}{z^i}))
\end{aligned}
$$

and also

$$
\begin{aligned}
N(\beta)^2 e"_{(0,\beta)} &= \frac{1}{N(\beta)}[\frac{2\delta^3}{\sigma_0^6} + \sum_{i=1}^{p_0}(\frac{\tau_i^3 z^3}{(1 - t_i z)^3} - \sum_{j\in T(i)}\frac{\gamma_j^3 z^3}{(1 - t_i z)^3}(\sum_{h\geq 0}(\frac{\alpha_j z}{(1 - t_i z))^h})^3\\
&+ \frac{\overline{\tau_i}^3 z^3}{(1 - \overline{t_i} z)^3} - \sum_{j\in T(i)}\frac{\overline{\gamma_j}^3 z^3}{(z - \overline{t_i})^3}(\sum_{h\geq 0}(\frac{\overline{\alpha_j}}{(z - \overline{t_i})})^h)^3)\\
&- \sum_{i=1}^{q_0}(\frac{(\mu_i^3 z^3}{(1 - u_i z)^3} - \sum_{j\in U(i)}\frac{\gamma_j^3 z^3}{(1 - u_i z)^3}(\sum_{h\geq 0}(\frac{\alpha_j z}{(1 - u_i z)})^h)^3\\
&- \frac{\overline{\tau_i}^3 z^3}{(1 - \overline{u_i} z)^3} - \sum_{j\in U(i)}\frac{\overline{\gamma_j}^3 z^3}{(z - \overline{u_i})^3}(\sum_{h\geq 0}(\frac{\overline{\alpha_j}}{(z - \overline{u_i})})^h)^3)\\
&- \sum_{i\in J}(\frac{\gamma_i^3 z^3}{(1 - c_i z)^3} + \frac{\overline{\gamma_i}^3}{(z - \overline{c_i})^3})
\end{aligned}
$$

Looking at the leading terms in the expansions, we obtain that for all β:

$$
\frac{e_{(0,\beta)}}{f_0} \leq \frac{1}{m_f}\left(u + (1 + 2(r + p_0 + q_0)\frac{1}{\rho})^r) + 2s\right)
$$

$$
\frac{N(\beta)^2 e"_{(0,\beta)}}{f_0} \leq \frac{1}{m_f}\left(2u^3 + (1 + 2(r + p_0 + q_0)\frac{1}{\rho})^{3r})\right)
$$

The entropy condition is the following. Let $\mathcal{N}(\epsilon)$ be the number of balls of diameter ϵ in H necessary for covering the set of functions. We have to verify that

$$\int_0^1 \left[\log \frac{\mathcal{N}(\epsilon)^2}{\epsilon}\right]^2 d\epsilon < +\infty$$

The previous expansions allow to find that for the set

$$\left\{ \frac{e_{(0,\beta)}}{f_0}, \frac{N(\beta)^2 e"_{(0,\beta)}}{f_0} \; : \; \beta \in \tilde{\mathcal{B}} \right\}$$

the norm square is a quadratic function of at most K bounded parameters, with $K = r(1 + 2s + 2p_0 + 2q_0)$, so that we have

$$\mathcal{N}(\epsilon) = O(\frac{1}{\epsilon^K})$$

and the condition holds.

References

Azencott, R. and Dacunha-Castelle, D.(1986). *Series of irregular observations forescasting and model building.* Springer.

Beran, R. and Millar, P.W. (1987). Stochastic estimation and testing. *Annals of Stat.*, **15** 1131–1154.

Bickel, P. and Chernoff, H. (1993). Asymptotic distribution of the Likelihood ratio Statistic in a Prototypical Non Regular Problem. *Statistics and Probability: A Raghu Raj Bahadur Festschrift.* J.K. Ghosh, S.K. Mitra, K.R. Parthasarathy and B.L.S. Prakasa Rao eds., Wiley, 83-96.

Chernoff, H. and Lander, E. (1995). Asymptotic distribution of the likelihood ratio test taht a mixture of two binomials is a single binomial. *Journal of Stat. Plann. and Inf.* **43**, 19-40.

Ciuperca, G. (1998). Maximum Likelihood Test for Mixtures on Translation Parameter. *Prépublication de l'Université d'Orsay.*

Dacunha-Castelle, D. and Gassiat, E.(1997). Testing in locally conic models and application to mixture models. *ESAIM Prob. and stat.*, 285-317.

Dahlhaus, R.(1988). Empirical processes and their applications to time series analysis. *Stochastic Processes and their Applications* **30** 69–83.

Feder, P. I.(1975). The log likelihood ratio in segmented regression. *Annals of Stat.* **3**, 1, 84-97.

Ghosh, J.K. and Sen, P.K.(1985). On the asymptotic performance of the log-likelihood ratio statistic for the mixture model and related results. *Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer* vol. II, L.M. Le Cam and R.A. Olshen (eds), Wadsworth, Monterey, 789-806.

Hannan, J.(1980) The estimation of the order of an Arma process. *Annals of Stat.* **8** 5:1071–1081.

Hartigan, J.A. (1985). A failure of likelihood asymptotics for normal mixtures. *Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer* vol. II, L.M. Le Cam and R.A. Olshen (eds), Wadsworth, Monterey, 807-810.

Kéribin, C. (1997). Consistent estimation of the order of mixture models. *Prépublication 61* Evry.

Lemdani, M. and Pons, O. (1997). Likelihood ratio tests for genetic linkage. *Stat. and Prob. Letters* **33**, 15-22.

Lindsay, B.G. (1995). Mixture models: theory, Geometry and Applications.

Pollard, D.(1984). *Convergence of Stochastic Processes.* Springer Verlag.

Teicher, H.(1965). Identifiability of finite mixtures. *Annals of Math. Stat.* **36** 423–439.

Van der Vart, A. W., Wellner, J.A.(1996). *Empirical Processes.* Springer Verlag.

Veres, S.(1987). Asymptotic distribution of likelihood ratios for overparametrized Arma processes. *J. Time Series Anal.* **8** 345–357.

Yakowitz, S.J., Spragins, J.D.(1968). On the identifiability of finite mixtures. *Annals of Math. Stat.* **39** 209–214.