

Likelihood ratio tests in contamination models

MOHAMED LEMDANI¹ and ODILE PONS²

¹*Faculté de Pharmacie de Lille, 3 rue du Professeur Laguesse, 59006 Lille, France.*

E-mail: mlemdani@phare.univ-lille2.fr

²*Biométrie, Institut National de la Recherche Agronomique, 78352 Jouy-en-Josas, France.*

E-mail: op@jouy.inra.fr

We study the asymptotic distribution of the likelihood ratio statistic to test whether the contamination of a known density f_0 by another density of the same parametric family reduces to f_0 . The classical asymptotic theory for the likelihood ratio statistic fails, and we propose a general reparametrization which ensures regularity properties. Under the null hypothesis, the likelihood ratio statistic converges to the supremum of a squared truncated Gaussian process. The result is extended to the case of the contamination of a mixture of p known densities by q other densities of the same family.

Keywords: asymptotic distribution; contamination; homogeneity; likelihood ratio; mixture distribution

1. Introduction

The determination of the number of components in a mixture of distributions of the same parametric form has recently been investigated in a number of papers. However, many of them concern computational aspects of the problem and the question of the asymptotic behaviour of the likelihood ratio (LR) statistic has not been solved. In the natural parametrization of the problem, several difficulties appear under the null hypothesis. One of them is the fact that the null hypothesis lies on the border of the parameter space, whereas it is classically assumed in its interior. This question has been studied by Chernoff (1954) and, more generally, by Self and Liang (1987) and Geyer (1994). Another is the lack of identifiability of some nuisance parameters which are present only under the alternative. Redner (1981) extended the classical result due to Wald (1949) on the consistency of the maximum likelihood estimator (MLE) to a quotient parameter space, and this method applies to mixture models. In regular models with nuisance parameters, Davies (1977) studied the score statistic at a fixed value of the nuisance parameters and then considered the maximum of the Gaussian approximation over these parameters. The lack of identifiability is connected to further difficulties in general mixture models. First, the null hypothesis, which concerns parametric distributions, can be expressed in several different forms in the parameter space. Secondly, for each of them, a component of the score equals zero function and the Fisher information is singular if the nuisance parameters are equal to the true parameter values. To rule out that singularity, Ghosh and Sen (1985) required a separation condition on the parameters of a mixture $(1 - \lambda)f_{\theta_1} + \lambda f_{\theta_2}$. Along the same lines as Davies (1977), they

derived the asymptotic behaviour of the LR test for $\lambda = 0$ against an alternative of a mixture of two separate densities.

Here we study that question in the simplest case of the contamination of known distributions, then we discuss the general case of a mixture. We adapt an approach introduced by Chernoff and Lander (1995) for binomial distributions. It is based on a reparametrization which solves the initial identifiability problems.

Let Θ be an open subset of \mathbb{R}^k with compact closure and let $\mathcal{F} = \{f_\theta, \theta \in \Theta\}$ be a parametric family of densities with respect to some measure ν on a measurable space $(\mathbb{X}, \mathcal{B})$. We assume throughout that the parameter θ is identifiable for \mathcal{F} and that the mixture densities are identifiable in the following sense. Let $p \geq 1$ and $q \geq 1$, $\lambda_1, \dots, \lambda_{p+q}$ be in $[0, 1]$ with $\lambda_1 + \dots + \lambda_{p+q} = 1$, ρ_1, \dots, ρ_p be in $[0, 1]$ with $\rho_1 + \dots + \rho_p = 1$, and let $\theta_1, \dots, \theta_{p+q}$, ξ_1, \dots, ξ_p be in Θ with pairwise distinct ξ_1, \dots, ξ_p . If

$$\sum_{j=1}^{p+q} \lambda_j f_{\theta_j} = \sum_{j=1}^p \rho_j f_{\xi_j}$$

then there exists a permutation π of $\{1, \dots, p+q\}$ such that

$$\begin{aligned} \theta_{\pi(j)} &= \xi_j, & 1 \leq j \leq p, \\ \theta_{\pi(p+l)} &\in \{\xi_1, \dots, \xi_p\} \text{ or } \lambda_{\pi(p+l)} = 0, & 1 \leq l \leq q. \end{aligned}$$

We first study the contamination of a known density f_{θ_0} from \mathcal{F} by a density from the same family, and we test the hypothesis that an observed sample has the density f_{θ_0} (Section 2). In Section 3, we extend this study to the contamination of a mixture of p known densities from \mathcal{F} by a mixture of q other unknown densities from the same family. In order to obtain a unique null hypothesis in Θ , the idea is to mix the natural parameters of the models under the alternative. In order to preserve the continuity of the score function, we next map the parameter space Θ into a compact set $\tilde{\Theta}$ of \mathbb{R}^{k+1} , which is in fact a hypersurface of the topological dimension k . We may then assume that the models are regular in the new parametrizations. We establish locally asymptotic normality in $n^{-1/2}$ -neighbourhoods of the true parameter values and we prove that two times the log-likelihood ratio converges in distribution to the supremum of a squared truncated Gaussian process on $\tilde{\Theta}$. Examples are given for Gaussian densities and binomial densities.

2. The contamination model

We consider the contamination of a known density $f_0 = f_{\theta_0} \in \mathcal{F}$ by some unknown density of \mathcal{F} . The contaminated density is

$$g_{\lambda, \theta} = (1 - \lambda)f_0 + \lambda f_\theta, \quad \lambda \in [0, 1], \theta \in \Theta.$$

Let (X_1, \dots, X_n) be a sample having a density g . We want to test the hypothesis $H_0 : g = f_0$ against the alternative $H_1 : g = g_{\lambda, \theta}$ with $\lambda \neq 0$ and $\theta \neq \theta_0$.

Under the identifiability assumption, the hypothesis H_0 can be expressed as $\lambda = 0$, θ

being a nuisance parameter which disappears under H_0 ; or $H_0 : \theta = \theta_0$ and λ becomes a nuisance parameter; or, finally, $\lambda = 1$ and $\theta = \theta_0$. Since the latter is a special case of the previous one, there is a duality between $\lambda = 0$ and $\theta = \theta_0$ for H_0 which could lead to two different approaches. Moreover, the score function tends to zero and the Fisher information becomes singular if θ tends to θ_0 in the first case and if λ tends to zero in the second one. Therefore, we introduce the mixed parameter

$$\mu = \lambda \|\theta - \theta_0\|$$

which lies in a positive compact subset in \mathbb{R} , and H_0 is equivalent to $\mu = 0$.

Let P_0 be a probability with density f_0 with respect to ν . We will make use of the following notations and conditions:

Condition 1. The map $\theta \mapsto f_\theta$ can ν -almost surely be extended by continuity on $\overline{\Theta}$, the compact closure of Θ , and it is ν -almost surely continuously differentiable on $\overline{\Theta}$, with derivative f'_θ .

Condition 2. The function $\sup_{\lambda \in [0,1]} \sup_{\theta \in \overline{\Theta}} |\log g_{\lambda,\theta}|$ is P_0 -integrable.

Let $\theta^{(0)} = \theta_0$, let $\theta^{(j)}$ be the k -dimensional vector having its first j components equal to those of θ and its $k - j$ last ones equal to those of θ_0 , $\theta^{(j)} = (\theta_1, \dots, \theta_j, \theta_{0j+1}, \dots, \theta_{0k})$, and let $f'_{j,\theta}$ be the value at θ of the j th partial derivative of f with respect to θ . We define the k -dimensional function $\phi = (\phi_1, \dots, \phi_k)^T$ on $\overline{\Theta} \times \mathbb{X}$ by

$$\phi_{j,\theta} = \begin{cases} (\theta_j - \theta_{0j})^{-1} \{f_{\theta^{(j)}} - f_{\theta^{(j-1)}}\} f_0^{-1} & \text{if } \theta_j \neq \theta_{0j} \\ f'_{j,\theta^{(j-1)}} f_0^{-1} & \text{if } \theta_j = \theta_{0j}. \end{cases}$$

The function ϕ is therefore continuous on $\overline{\Theta}$. Let $\mathcal{U} = \{u \in \mathbb{R}^k; \|u\| = 1\}$, and for $u \in \mathcal{U}$ let $\overline{\Theta}_u = \{\theta \in \overline{\Theta}; \theta = \theta_0 + u\|\theta - \theta_0\|\}$ and $K_u = \{\|\theta - \theta_0\|; \theta \in \overline{\Theta}_u\}$ which is a compact in \mathbb{R}_+ , and let

$$\tilde{\Theta} = \{(u, r); u \in \mathcal{U}, \theta_0 + ru \in \overline{\Theta}\}.$$

Then for any $\theta \in \Theta$, $f_\theta = f_0\{1 + (\theta - \theta_0)^T \phi_\theta\}$ and for $\theta = \theta_0 + ru \in \overline{\Theta}_u$, $g_{\lambda,\theta}$ is reparametrized as

$$h_{u,r,\mu} = f_0\{1 + \mu u^T \phi_{\theta_0+ru}\}.$$

Let Y_n be the k -dimensional process defined on $\overline{\Theta}$ by $Y_n(\theta) = n^{-1/2} \sum_i \phi_\theta(X_i)$, $\Sigma_{\theta,\theta'}$ be the covariance matrix $E_{P_0}(\phi_\theta \phi_{\theta'}^T)(X_1)$ and $\Sigma_\theta = \Sigma_{\theta,\theta}$. We will assume the following conditions, which cannot be simply expressed in terms of the densities f_θ because they also include conditions for their derivative with respect to θ :

Condition 3. For any $\theta \in \overline{\Theta}$, the matrix Σ_θ is positive definite and there exists a constant M_0 such that for some $\delta > 0$

$$E_{P_0} \sup_{\theta \in \overline{\Theta}} \|\phi_\theta(X_1)\|^{4+\delta} < M_0.$$

Condition 4. For any positive ε and η there exist $\delta' > 0$ and an integer n_0 such that for any $n \geq n_0$, $P_0(\sup_{\|\theta - \theta'\| < \delta'} \|Y_n(\theta) - Y_n(\theta')\| \geq \varepsilon) \leq \eta$.

We consider the uniform topology on $C(\bar{\Theta})$, the space of continuous functions on $\bar{\Theta}$ with values in \mathbb{R}^k . A uniform metric is also defined on $\tilde{\Theta}$ by $\|(u, r) - (u', r')\| = \|u - u'\| + |r - r'|$ and $C(\tilde{\Theta})$ is the space of real continuous functions on $\tilde{\Theta}$. Using Billingsley's Theorem 12.3 (Billingsley 1968, p. 95) if $K = 1$, and Theorem A.1.19 in Ibragimov and Has'minskiĭ (1981, p. 372) otherwise, a moment condition can often be used to establish Condition 4.

Lemma 1. Under Condition 3, if there exist constants $m \geq s > k$ and a function ℓ on \mathbb{X} with $E_{P_0} \ell(X_1) < \infty$, such that for any $\theta, \theta' \in \bar{\Theta}$ and $x \in \mathbb{X}$, $\|\phi_\theta(x) - \phi_{\theta'}(x)\|^m \leq \|\theta - \theta'\|^s \ell(x)$, then Condition 4 is satisfied. Moreover, under Conditions 3 and 4 the process Y_n is tight on $C(\bar{\Theta})$.

Under H_1 the log-likelihood of the sample (X_1, \dots, X_n) is $L_n(\lambda, \theta) = \sum_{i \leq n} \log g_{\lambda, \theta}(X_i)$ and is denoted $L_n(\theta_0)$ under H_0 . Let $l_n(u, r, \mu) = \sum_{i \leq n} \log(f_0^{-1} h_{u, r, \mu})(X_i)$. The LR test statistic is

$$S_n = 2 \sup_{\theta \in \Theta \setminus \{\theta_0\}} \sup_{\lambda \in]0, 1[} \{L_n(\lambda, \theta) - L_n(\theta_0)\} = 2 \sup_{u \in \mathcal{U}} \sup_{r \in K_u} \sup_{\mu \in [0, r]} l_n(u, r, \mu).$$

Lemma 2. Let A be a compact subset of \mathbb{R}^d and let $\mathcal{V} = \{v_\alpha; \alpha \in A\}$ be a family of real functions defined on \mathbb{X} such that $\sup_{\alpha \in A} |v_\alpha|$ is P_0 -integrable and $\alpha \mapsto v_\alpha(x)$ is continuous for each x . Then $\sup_{\alpha \in A} |n^{-1} \sum_i v_\alpha(X_i) - E_{P_0} v_\alpha(X_1)| \rightarrow 0$ P_0 -almost surely.

Proof. The supremum on $A \subset \mathbb{R}^d$ can be restricted to the rational points $A_\mathbb{Q} = \{\alpha \in A \cap \mathbb{Q}\}$ of A . The map $\alpha \mapsto v_\alpha(x)$ being uniformly continuous on A for each x , Lebesgue's theorem implies that $\int |v_\alpha - v_{\alpha'}| dP_0$ tends to zero as $\|\alpha - \alpha'\| \rightarrow 0$ in $A_\mathbb{Q}$ and the same property holds in A . By the compactness of A , it follows that, for any $\varepsilon > 0$, the bracketing metric entropy $HB_1(\varepsilon, \mathcal{V}, P_0)$ is finite. It is defined by $HB_1(\varepsilon, \mathcal{V}, P_0) = \log NB_1(\varepsilon, \mathcal{V}, P_0)$ with $NB_1(\varepsilon, \mathcal{V}, P_0) = \min\{k \in \mathbb{N}; \exists v_1, \dots, v_k \in \mathcal{L}_1(P_0) \text{ s.t. } \mathcal{V} \subset \cup_{a,b} [v_a, v_b], \int |v_a - v_b| dP_0 < \varepsilon\}$. Then the Glivenko–Cantelli theorem applies: if P_n is the empirical distribution function of the observed sample, $\sup_{\alpha \in A} |\int v_\alpha d(P_n - P_0)| \rightarrow 0$ P_0 -almost surely (Shorack and Wellner 1986, p. 837). □

Lemma 3. Let $\hat{\mu}_n(u, r)$ be the value of μ in $[0, r]$ that maximizes $l_n(u, r, \mu)$ as (u, r) is fixed in $\tilde{\Theta}$. Then $\sup_{(u, r) \in \tilde{\Theta}} \hat{\mu}_n(u, r)$ converges P_0 -almost surely to zero under Conditions 1 and 2.

Proof. We consider the model with the new parametrization $h_{u, r, \mu}$ in $\{(u, r, \mu); (u, r) \in \tilde{\Theta}, \mu \in [0, r]\}$. Conditions 1 and 2 imply that h is continuous with respect to the parameter (u, r, μ) and $|\log h|$ is P_0 -uniformly integrable. Let $l(u, r, \mu) = E_{P_0} \log(f_0^{-1} h_{u, r, \mu})(X_1)$. Lemma 2 implies that $\sup_{(u, r) \in \tilde{\Theta}} \sup_{\mu \in [0, r]} |n^{-1} l_n(u, r, \mu) - l(u, r, \mu)| \rightarrow 0$ P_0 -almost surely.

Let $M_n = \sup_{u, r} \hat{\mu}_n(u, r)$ and $\Omega_\alpha = \{\limsup_n M_n > \alpha\}$. If M_n did not converge P_0 -almost surely to zero, we could find $\alpha > 0$ such that $P_0(\Omega_\alpha) \neq 0$; hence, for any $\omega \in \Omega_\alpha$, $\hat{\mu}_n(\omega, u_n(\omega), r_n(\omega)) > \alpha$ for some sequence $\{u_n(\omega), r_n(\omega)\}_n$. By compactness, for any

$\omega \in \Omega_\alpha$ and from any subsequence of $\{\hat{\mu}_n(\omega, u_n(\omega), r_n(\omega))\}_n$, we could extract a further subsequence $\{\hat{\mu}_{n'}(\omega, u_{n'}(\omega), r_{n'}(\omega))\}_{n'}$ converging to some $M_0(\omega) > \alpha$ and a subsubsequence $\{u_{n''}(\omega), r_{n''}(\omega)\}_{n''}$ from $\{u_{n'}(\omega), r_{n'}(\omega)\}_{n'}$ converging to some $\{u_0(\omega), r_0(\omega)\}$.

For any $(u, r) \in \tilde{\Theta}$, $l(u, r, \cdot)$ is a concave function on $[0, r]$ having a unique maximum at $\mu(u, r) \equiv 0$, therefore $l(u, r, \hat{\mu}_n(u, r)) \leq l(u, r, 0) \equiv 0$. Moreover, for each $(u, r) \in \tilde{\Theta}$ and $n \geq 0$, $l_n(u, r, \hat{\mu}_n(u, r)) \geq l_n(u, r, 0) \equiv 0$. In particular, on Ω_α ,

$$\begin{aligned} 0 &\leq n^{n-1} l_n(u_{n^n}, r_{n^n}, \hat{\mu}_{n^n}(u_{n^n}, r_{n^n})) \\ &\leq \sup_{u,r,\mu} |n^{n-1} l_n(u, r, \mu) - l(u, r, \mu)| + |l(u_{n^n}, r_{n^n}, \hat{\mu}_{n^n}(u_{n^n}, r_{n^n})) - l(u_0, r_0, M_0)| \\ &\quad + l(u_0, r_0, M_0). \end{aligned}$$

By continuity of l and Lemma 2, there would be a subset $\Omega_{0,\alpha}$ of Ω_α having the probability zero such that $l(u_0(\omega), r_0(\omega), M_0(\omega)) = 0$ for any $\omega \in \Omega_\alpha \setminus \Omega_{0,\alpha}$. Using the concavity of $l(u, r, \cdot)$, it follows that $M_0 = 0$ on $\Omega_\alpha \setminus \Omega_{0,\alpha}$, which contradicts the property $P_0(\Omega_\alpha) \neq 0$. □

Theorem 1. *Assume that Conditions 1–4 hold. Then, under H_0 , the LR statistic S_n converges weakly to*

$$\sup_{(u,r) \in \tilde{\Theta}} Z^2(u, r) I_{\{Z(u,r) > 0\}},$$

where Z is a continuous Gaussian process on $\tilde{\Theta}$ having mean zero and covariance function K given by

$$K(u_1, r_1; u_2, r_2) = u_1^T \Sigma_{\theta_1, \theta_2} u_2 (u_1^T \Sigma_{\theta_1} u_1)^{-1/2} (u_2^T \Sigma_{\theta_2} u_2)^{-1/2},$$

where $\theta_1 = \theta_0 + r_1 u_1$ and $\theta_2 = \theta_0 + r_2 u_2$.

Proof. Let $\dot{l}_n(u, r, \cdot)$ be the first derivative of $l_n(u, r, \cdot)$ with respect to μ . Since $\hat{\mu}_n(u, r)$ maximizes $l_n(u, r, \mu)$ under the constraint $\hat{\mu}_n(u, r) \geq 0$, $n^{-1/2} \dot{l}_n(u, r, \hat{\mu}_n(u, r)) = \zeta(u, r)$, where the Lagrange multiplier $\zeta(u, r)$ satisfies

$$\zeta(u, r) \begin{cases} = 0, & \text{if } \hat{\mu}_n(u, r) > 0, \\ \leq 0, & \text{if } \hat{\mu}_n(u, r) = 0. \end{cases} \tag{1}$$

Let

$$a_i(u, r) = u^T \phi_{\theta_0 + ru}(X_i). \tag{2}$$

Using the identity

$$(1 + \mu a_i)^{-1} = 1 - \mu a_i + \mu^2 a_i^2 (1 + \mu a_i)^{-1}, \tag{3}$$

and denoting

$$R_{1n} = n^{-1} \sum_i \frac{a_i^3}{1 + \hat{\mu}_n a_i},$$

we have

$$\zeta(u, r) = n^{-1/2} \dot{l}_n(u, r, \hat{\mu}_n(u, r)) = n^{-1/2} \sum_i \frac{a_i(u, r)}{1 + \hat{\mu}_n(u, r)a_i(u, r)} \tag{4}$$

$$= u^T Y_n(\theta_0 + ru) - n^{1/2} \hat{\mu}_n(u, r) \left\{ n^{-1} \sum_i [u^T \phi_{\theta_0+ru}(X_i)]^2 - \hat{\mu}_n(u, r) R_{1n}(u, r) \right\}. \tag{5}$$

Conditions 1–3 and Lemma 2 imply that

- (a) $\sup_{u \in \mathcal{U}} \sup_{\theta \in \bar{\Theta}} |n^{-1} \sum_i u^T [\phi_\theta(X_i) \phi_\theta^T(X_i) - \sum_0] u|$ converges P_0 -almost surely to zero;
- (b) $\sup_{(u,r) \in \bar{\Theta}} \hat{\mu}_n(u, r) |R_{1n}(u, r)| = o_p(1)$.

To prove (b) we consider, for δ given by Condition 3, constants α and β such that $1/(4 + \delta) < \alpha < 1/4$ and $0 < \beta < \alpha - 1/(4 + \delta)$. We have

$$P_0 \left(n^{-\alpha} \sup_{i \leq n} \sup_{u,r} |a_i(u, r)| > n^{-\beta} \right) \leq M_0 n^{1-(4+\delta)(\alpha-\beta)},$$

hence $n^{-\alpha} \sup_{i \leq n} \sup_{u,r} |a_i(u, r)|$ converges to zero in probability.

Let $R_{1n} = R_{1n}^+ + R_{1n}^-$, with

$$R_{1n}^+ = n^{-1} \sum_{i: a_i > 0} \frac{a_i^3}{1 + \hat{\mu}_n a_i}, \quad R_{1n}^- = n^{-1} \sum_{i: a_i < 0} \frac{a_i^3}{1 + \hat{\mu}_n a_i}.$$

By Lemma 3 and since $0 \leq R_{1n}^+ \leq n^{-1} \sum_i |a_i|^3$ which is a uniform $O_p(1)$ under Condition 3, $\hat{\mu}_n R_{1n}^+ = o_p(1)$ uniformly. To prove a similar convergence for R_{1n}^- , we restrict our attention to the values of (u, r) such that $\hat{\mu}_n(u, r) > 0$, and therefore $\zeta(u, r) = 0$ in (4) and (5). For these values, we write (5) as

$$\begin{aligned} & n^{1/2} \hat{\mu}_n(u, r) \left\{ n^{-1} \sum_i [u^T \phi_{\theta_0+ru}(X_i)]^2 - \hat{\mu}_n(u, r) R_{1n}^+(u, r) \right\} \\ &= u^T Y_n(\theta_0 + ru) + n^{1/2} \hat{\mu}_n^2(u, r) R_{1n}^-(u, r). \end{aligned} \tag{6}$$

By (a) and Condition 3 and because $\hat{\mu}_n R_{1n}^+$ is a uniform $o_p(1)$, the left-hand side of this equality is strictly positive on $\bar{\Theta}$ if n is large enough. Moreover, by Condition 4 and weak convergence of its finite-dimensional distributions, the process Y_n converges weakly to a Gaussian process Y in $C(\bar{\Theta})$, and it follows that $\sup_{u,r} \|Y_n(\theta_0 + ru)\| = O_p(1)$. As $R_{1n}^- \leq 0$, $\sup_{u,r} n^{1/2} \hat{\mu}_n^2(u, r) R_{1n}^-(u, r)$ is necessarily $O_p(1)$; then, from (6), $\sup_{u,r} n^{1/2} \hat{\mu}_n(u, r) = O_p(1)$. As $\hat{\mu}_n(u, r) |R_{1n}^-(u, r)| = n^{1/2} \hat{\mu}_n(u, r) n^{2\alpha-1/2} n^{-2\alpha} |R_{1n}^-(u, r)|$ and

$$\begin{aligned}
 n^{-2\alpha}|R_{1n}^-(u, r)| &\leq n^{-2\alpha} \sup_i \sup_{u,r} |a_i(u, r)|^2 \left\{ -n^{-1} \sum_{i;a_i(u,r) < 0} \frac{a_i}{1 + \hat{\mu}_n a_i}(u, r) \right\} \\
 &= n^{-2\alpha} \sup_i \sup_{u,r} |a_i(u, r)|^2 \left\{ n^{-1} \sum_{i;a_i(u,r) > 0} \frac{a_i}{1 + \hat{\mu}_n a_i}(u, r) \right\}, \quad \text{by (4),} \\
 &\leq n^{-2\alpha} \sup_i \sup_{u,r} |a_i(u, r)|^2 \sup_{u,r} n^{-1} \sum_i |a_i(u, r)|,
 \end{aligned}$$

we deduce that $n^{-2\alpha} \sup_{u,r} |R_{1n}^-(u, r)| = o_p(1)$ and therefore $\sup_{u,r} \hat{\mu}_n(u, r) |R_{1n}^-(u, r)| = o_p(1)$, which ends the proof of convergence property (b).

From (5), (a), (b) and Lemma 3,

$$\zeta(u, r) = u^T Y_n(\theta_0 + ru) - n^{1/2} \hat{\mu}_n(u, r) \{u^T \Sigma_{\theta_0 + ru} u\} + o_p(1),$$

where the $o_p(1)$ holds in P_0 -probability uniformly over $\tilde{\Theta}$. Let Z_n be the process defined on $\tilde{\Theta}$ by

$$Z_n(u, r) = (u^T \Sigma_{\theta_0 + ru} u)^{-1/2} u^T Y_n(\theta_0 + ru).$$

By weak convergence of Y_n to a Gaussian process Y in $C(\bar{\Theta})$ having mean zero and covariance $\Sigma_{\theta_1, \theta_2}$ at θ_1 and θ_2 , the process Z_n converges weakly to a centred Gaussian process Z with covariance $K(\theta_0 + r_1 u_1, \theta_0 + r_2 u_2)$ at $(u_1, r_1; u_2, r_2)$.

If $\hat{\mu}_n(u, r) > 0$, (1) entails that $n^{1/2} \hat{\mu}_n(u, r) = (u^T \Sigma_{\theta_0 + ru} u)^{-1/2} \{Z_n(u, r) + \varepsilon_n(u, r)\}$, where $\sup_{u,r} |\varepsilon_n(u, r)|$ tends to zero in probability, and otherwise $\hat{\mu}_n(u, r) = 0$. Then

$$\begin{aligned}
 n^{1/2} \hat{\mu}_n(u, r) &= (u^T \Sigma_{\theta_0 + ru} u)^{-1/2} \{Z_n(u, r) + \varepsilon_n(u, r)\} 1_{\{Z_n(u,r) + \varepsilon_n(u,r) > 0\}} \\
 &= (u^T \Sigma_{\theta_0 + ru} u)^{-1/2} Z_n(u, r) 1_{\{Z_n(u,r) > 0\}} + o_p(1),
 \end{aligned}$$

since $|z_n(1_{\{z_n + \epsilon_n > 0\}}^{-1} 1_{\{z_n > 0\}})| \leq |z_n 1_{\{|z_n| \leq \epsilon_n\}}| \leq |\epsilon_n|$ which tends to zero.

By integrating (3), we have

$$\begin{aligned}
 l_n(u, r, \hat{\mu}_n(u, r)) &= n^{1/2} \hat{\mu}_n(u, r) u^T Y_n(\theta_0 + ru) - \frac{n}{2} \hat{\mu}_n^2(u, r) n^{-1} \sum_i u^T \phi_\theta(X_i) \phi_\theta^T(X_i) u + R_{2n}(u, r),
 \end{aligned}$$

where

$$R_{2n} = \int_0^{\hat{\mu}_n} \sum_i \frac{a_i^3 \mu^2}{1 + \mu a_i} d\mu = \hat{\mu}_n \sum_i \frac{\mu_n^{*2} a_i^3}{1 + \mu_n^* a_i},$$

using (2) and with $\mu_n^*(u, r)$ in $]0, \hat{\mu}_n(u, r)[$. The functions $\mu \mapsto a_i^3(1 + \mu a_i)^{-1}$ are decreasing; hence, for each $i \leq n$,

$$\frac{a_i^3}{1 + \hat{\mu}_n a_i} \leq \frac{a_i^3}{1 + \mu_n^* a_i} \leq a_i^3,$$

which implies

$$|R_{2n}| \leq \hat{\mu}_n^3 \max \left(\left| \sum_i a_i^3 \right|, \left| \sum_i \frac{a_i^3}{1 + \hat{\mu}_n a_i} \right| \right) = \hat{\mu}_n^3 \max \left(\left| \sum_i a_i^3 \right|, n |R_{1n}| \right).$$

Under Condition 3 and by Lemma 2, $n^{-1}|\sum_i a_i^3|$ is P_0 -almost surely uniformly bounded on $\tilde{\Theta}$. Moreover, $\hat{\mu}_n |R_{1n}|$ converges uniformly to zero in probability and $\sup_{u,r} \hat{\mu}_n(u, r) = O_p(n^{-1/2})$. Therefore,

$$l_n(u, r, \hat{\mu}_n(u, r)) = n^{1/2} \hat{\mu}_n(u, r) u^T Y_n(\theta_0 + ru) - \frac{n}{2} \hat{\mu}_n^2(u, r) u^T \Sigma_{\theta_0+ru} u + o_P(1)$$

uniformly on $\tilde{\Theta}$ and

$$S_n = \sup_{(u,r) \in \tilde{\Theta}} l_n(u, r, \hat{\mu}_n(u, r)) = \sup_{(u,r) \in \tilde{\Theta}} [Z_n^2(u, r) I_{\{Z_n(u,r) > 0\}}] + o_P(1).$$

□

Remark 1. We could have expanded the log-likelihood for μ in a neighbourhood of 0 and with fixed θ in $\bar{\Theta}$. The process $\theta \mapsto n^{-1/2} \sum_i \phi_\theta(X_i)$ is continuous on $\bar{\Theta}$ as well as the associated process \tilde{Y}_n defined on $\tilde{\Theta}$ by $\tilde{Y}_n(u, r) = u^T Y_n(\theta_0 + ru)$. But with the parameter μ , the expansion of the log-likelihood depends on the process $\theta \mapsto n^{-1/2} \sum_i \|\theta - \theta_0\|^{-1} (\theta - \theta_0)^T \phi_\theta(X_i)$, which is not defined at θ_0 , and it could be extended by infinitely many possible values. Thus we cannot study directly weak convergence in $C(\bar{\Theta})$.

Also, the result of Theorem 1 cannot be stated in the original space Θ though for any $r \neq 0$, $Z(u, r)$ is transformed back in the space Θ as $W(\theta) = \{(\theta - \theta_0)^T \Sigma_\theta (\theta - \theta_0)\}^{-1/2} (\theta - \theta_0)^T Y(\theta)$ with $\theta = \theta_0 + ru \neq \theta_0$. The covariance matrix K^W of W is defined by

$$K^W(\theta_1, \theta_2) = (\theta_1 - \theta_0)^T \Sigma_{\theta_1, \theta_2} (\theta_2 - \theta_0) \{(\theta_1 - \theta_0)^T \Sigma_{\theta_1} (\theta_1 - \theta_0)\}^{-1/2} \\ \times \{(\theta_2 - \theta_0)^T \Sigma_{\theta_2} (\theta_2 - \theta_0)\}^{-1/2},$$

for θ_1 and $\theta_2 \neq \theta_0$, and it is undefined if θ_1 or θ_2 equals θ_0 . The distribution of the variable $\sup_{\theta \in \bar{\Theta}, \theta \neq \theta_0} W^2(\theta) I_{\{W(\theta) > 0\}}$ is therefore undefined.

Remark 2. By the arguments used in the proof of Theorem 1, we derive that for any (u, r) in $\tilde{\Theta}$, the log-likelihood ratio at $\mu_n(u, r) = m \{n \Sigma_{\theta_0+ru}\}^{-1/2}$ has an expansion $m \Sigma_{\theta_0+ru}^{-1/2} Y_n(u, r) - \frac{1}{2} m^2 + o(m^2)$, where the o holds uniformly on $\tilde{\Theta}$. From Ibragimov and Has'minskii (1981, p. 120), the optimal convergence rate for μ , for fixed (u, r) , is therefore $n^{-1/2}$. However, when $r \neq 0$ is fixed, estimation of μ at rate $O_p(n^{-1/2})$ is equivalent to estimation of λ at rate $O_p(n^{-1/2})$ and this approach does not allow us to study the asymptotic behaviour of the LR test statistic under the alternative without a separation condition, as in Ghosh and Sen (1985). Since H_0 may be expressed as $\lambda = 0$ or $\theta = \theta_0$, two kinds of alternative should then be considered: either $\theta_n - \theta_0 = O(n^{-1/2})$ with $\lambda \in [0, 1]$; or $\lambda_n - \lambda_0 = O(n^{-1/2})$ with $\theta \in \Theta$ (cf. Lemdani and Pons 1995, for the binomial case).

Proof of Lemma 1. Assume $K = 1$. For θ and θ' in $\bar{\Theta}$, we have $E\{Y_{j,n}(\theta) - Y_{j,n}(\theta')\}^2 \leq E|\phi_{j,\theta}^2(X_1) - \phi_{j,\theta'}(X_1)\phi_{j,\theta'}(X_1)| + E|\phi_{j,\theta'}^2(X_1) - \phi_{j,\theta}(X_1)\phi_{j,\theta'}(X_1)|$ and each term of the right-hand side is bounded by $\sup_{\theta \in \bar{\Theta}} E\{\phi_{j,\theta}(X_1)\}^2 \{E\mathcal{L}(X_1)\|\theta - \theta'\|^s\}^{2/m}$. The process Y_n therefore satisfies Billingsley's moment condition (12.51) (Billingsley 1968, p. 95) under the conditions of Lemma 1. Moreover, for any $\theta \in \bar{\Theta}$, $Y_n(\theta)$ converges in distribution to a Gaussian variable and it is therefore tight on \mathbb{R}^k . Then, from Theorem 12.3 of Billingsley (1968, pp. 95–96), Condition 4 holds true and the process Y_n is tight in $C(\bar{\Theta})$. The proof is similar for $K > 1$. \square

If θ is a one-dimensional parameter, the unit vector set \mathcal{U} reduces to $\{+1, -1\}$ and the test statistic becomes

$$S_n = 2 \max \left[\sup_{\mu > 0, \theta > \theta_0} \sum_i \log\{1 + \mu\phi_\theta(X_i)\}, \sup_{\mu > 0, \theta < \theta_0} \sum_i \log\{1 - \mu\phi_\theta(X_i)\} \right].$$

Then we obtain the following corollary:

Corollary 1. Under Conditions 1–4 and if $\Theta \subset \mathbb{R}$, the statistic S_n converges weakly under H_0 to

$$S = \max \left[\sup_{\theta \geq \theta_0} Z^2(\theta)1_{\{Z(\theta) \geq 0\}}, \sup_{\theta \leq \theta_0} Z^2(\theta)1_{\{Z(\theta) \leq 0\}} \right],$$

where Z is a continuous Gaussian process on $\bar{\Theta}$ having mean zero and covariance function K given by $K(\theta_1, \theta_2) = \Sigma_{\theta_1, \theta_2}(\Sigma_{\theta_1} \Sigma_{\theta_2})^{-1/2}$.

Example 1. Let $\Theta =]a, b[$, with $a > 0$ and $b < 1$, and let $\theta_0 \in \Theta$ and $k \in \mathbb{N}$. Consider the binomial distribution $B(k, \theta_0)$ and the contamination model $(1 - \lambda)B(k, \theta_0) + \lambda B(k, \theta)$, $\lambda \in [0, 1]$ and $\theta \in \Theta$, which extends the model studied by Chernoff and Lander (1995). From Teicher (1963), we assume that $k \geq 2$ so that Condition 1 holds. Under $B(k, \theta)$, the probability of an observation $x \in \{0, \dots, k\}$ is $C_k^x \theta^x (1 - \theta)^{k-x}$ and its density f_θ with respect to the uniform discrete measure ν on $\{0, \dots, k\}$ satisfies Condition 1. For $\theta \neq \theta_0$ the function ϕ is defined as above by $\phi_\theta = (\theta - \theta_0)^{-1}(f_\theta f_{\theta_0}^{-1} - 1)$ if $\theta \neq \theta_0$ and it is extended by continuity at θ_0 as $\phi_{\theta_0}(x) = (x - k\theta_0)\{\theta_0(1 - \theta_0)\}^{-1}$. The covariance function Σ is defined by

$$\Sigma_{\theta, \theta'} = \begin{cases} \{(\theta - \theta_0)(\theta' - \theta_0)\}^{-1} [\{(\theta - \theta_0)(\theta' - \theta_0)\theta_0^{-1}(1 - \theta_0)^{-1} + 1 \}^k - 1], & \text{if } \theta \text{ and } \theta' \neq \theta_0, \\ k\{\theta_0(1 - \theta_0)\}^{-1}, & \text{otherwise;} \end{cases}$$

therefore, $\Sigma_\theta \geq \Sigma_{\theta_0} = k\{\theta_0(1 - \theta_0)\}^{-1} > 0$ for any $\theta \in [a, b]$. Moreover, the integrability properties of Conditions 2 and 3 hold and Condition 4 is also satisfied by Lemma 1, since the function $\theta \mapsto \phi_\theta$ is continuously differentiable on $\bar{\Theta}$ with a derivative ϕ'_θ such that $E \sup_{\theta \in \bar{\Theta}} \|\phi'_\theta(X_1)\|^4 < \infty$, where

$$\phi'_{\theta_0}(x) = f''_0(2f_0)^{-1} = \{(x - k\theta_0)^2 - x(1 - 2\theta_0)\}\{2\theta_0^2(1 - \theta_0)^2\}^{-1}$$

Example 2. Consider the family \mathcal{F} of Gaussian densities $f_{m,\sigma}$ with mean m and variance σ^2 with respect to Lebesgue measure, and suppose that $\theta = (m, \sigma)$ belongs to an open bounded set Θ where the compact closure of $\{\sigma; \exists m \text{ such that } (m, \sigma) \in \Theta\}$ does not contain zero. For any $\theta_0 = (m_0, \sigma_0) \in \Theta$, the two-dimensional function ϕ is defined by

$$\begin{aligned} \phi_{1,\theta}(x) &= \begin{cases} (m - m_0)^{-1}\{f_{m,\sigma_0}(x)f_{m_0,\sigma_0}^{-1}(x) - 1\}, & \text{if } m \neq m_0, \\ \sigma_0^{-2}(x - m_0), & \text{if } m = m_0, \end{cases} \\ \phi_{2,\theta}(x) &= \begin{cases} (\sigma - \sigma_0)^{-1}\{f_{m,\sigma}(x) - f_{m,\sigma_0}(x)\}f_{m_0,\sigma_0}^{-1}(x), & \text{if } \sigma \neq \sigma_0, \\ f_{m,\sigma_0}(x)f_0^{-1}(x)\{(x - m)^2\sigma_0^{-2} - 1\}, & \text{if } \sigma = \sigma_0. \end{cases} \end{aligned}$$

Let $\Sigma_{11,\theta}$, $\Sigma_{12,\theta}$ and $\Sigma_{22,\theta}$ be the elements of the matrix Σ_θ . Its determinant is $\Sigma_{11,\theta}\Sigma_{22,\theta} - \Sigma_{12,\theta}^2 > 0$ for any $\theta \in \bar{\Theta}$ since the functions $\phi_{1,\theta}$ and $\phi_{2,\theta}$ are never almost surely proportional and therefore Σ_θ is positive definite. The integrability assumptions of Conditions 2 and 3 are satisfied, and Lemma 1 applies for the same reason as in Example 1.

3. Testing for the contamination of a mixture of p known distributions by a mixture of q distributions

For integers $p \geq 1$ and $q \geq 1$, let $\mathcal{E}_{p,q}$ be the set of functions from $\{1, \dots, q\}$ into a subset of $\{1, \dots, p\}$ and let $\mathcal{S}_p = \{(\rho_1, \dots, \rho_p)^T \in [0, 1]^p; \rho_1 + \dots + \rho_p = 1\}$ and $\mathcal{S}_p^* = \mathcal{S}_p \cap]0, 1]^p$. We denote a mixture of $p + q$ densities from the family \mathcal{F} by

$$g_{p+q;\lambda,\theta} = \sum_{j=1}^{p+q} \lambda_j f_{\theta_j}, \quad \lambda = (\lambda_1, \dots, \lambda_{p+q}) \in \mathcal{S}_{p+q}, \theta_1, \dots, \theta_{p+q} \in \Theta.$$

Let $\theta^0 = (\theta_1^0, \dots, \theta_p^0) \in \Theta^p$ be known parameter values of densities from \mathcal{F} , θ^0 having pairwise distinct components. For a sample (X_1, \dots, X_n) with density g , we wish to test whether g is a mixture of the known densities $f_{\theta_j^0}$, $1 \leq j \leq p$, against the alternative of a mixture of $p + q$ densities from \mathcal{F} , including the p densities $f_{\theta_j^0}$, that is,

$$H_0 : g = g_{p;\lambda,\theta^0}, \quad \text{where } (\lambda_1, \dots, \lambda_p) \in \mathcal{S}_p^*,$$

$$H_1 : g = g_{p+q;\lambda,\theta}, \quad \text{where } \lambda \in \mathcal{S}_{p+q}^*; \theta_1, \dots, \theta_{p+q} \text{ are pairwise distinct in } \Theta;$$

$$\text{and } \theta = (\theta_1^T, \dots, \theta_{p+q}^T)^T \text{ is such that } \theta = (\theta^{0T}, \xi^T)^T \text{ with } \xi \in \Theta^q,$$

up to a permutation of mixture components. Let $\lambda^0 \in \mathcal{S}_p^*$ be the actual value of the mixture proportions and let $g_0 = \sum_{j=1}^p \lambda_j^0 f_{\theta_j^0}$ be the true density under H_0 . By the identifiability assumption, there exists a map $c \in \mathcal{E}_{p,q}$ such that under H_0 the parameters of a density $g_{p+q;\lambda,\theta}$ of the alternative satisfy

$$\theta_{p+l} = \theta_{c(l)}^0 \text{ or } \lambda_{p+l} = 0, \quad 1 \leq l \leq q.$$

Moreover, $\lambda_j + \sum_{l=1}^q I_{\{c(l)=j\}} \lambda_{p+l} = \lambda_j^0$, for $1 \leq j \leq p$. We adopt the same notation as in the previous section and now

$$S_n = 2 \sup_{\lambda \in \mathcal{S}_{p+q}} \sup_{\theta = (\theta^{0T}, \xi^T)^T \in \Theta^{p+q}} \sum_i \{ \log g_{(p+q;\lambda,\theta)}(X_i) - \log g_0(X_i) \}$$

is denoted

$$2 \sup_{\lambda \in \mathcal{S}_{p+q}} \sup_{\theta = (\theta^{0T}, \xi^T)^T \in \Theta^{p+q}} \{ L_n(p+q; \lambda, \theta) - L_n(\lambda^0, \theta^0) \}.$$

For any map $c \in \mathcal{C}_{p,q}$, let

$$C_c = \{ (\lambda, \theta); \lambda \in \mathcal{S}_{p+q}, \theta \in \Theta^{p+q}, \theta_j = \theta_j^0, \lambda_j + \sum_{l=1}^q I_{\{c(l)=j\}} \lambda_{p+l} = \lambda_j^0, 1 \leq j \leq p, \theta_{p+l} = \theta_{c(l)}^0 \text{ or } \lambda_{p+l} = 0, 1 \leq l \leq q \}$$

and $C = \cup_{c \in \mathcal{C}_{p,q}} C_c$ be the subset of $\mathcal{S}_{p+q} \times \Theta^{p+q}$ yielding the density g_0 . From Redner (1981), for any closed subset S of the parameter space not intersecting C , $\lim_{n \rightarrow \infty} \sup_{(\lambda,\theta) \in S} \prod_{i \leq n} (g_0^{-1} g_{p+q;\lambda,\theta})(X_i) = 0$ almost surely and therefore, if n is large enough,

$$S_n = 2 \sup_{c \in \mathcal{C}_{p,q}} \sup_{g(\lambda,\theta) \in \mathcal{N}(C_c)} \{ L_n(p+q; \lambda, \theta) - L_n(\lambda^0, \theta^0) \},$$

where $\mathcal{N}(C_c)$ is a neighbourhood of C_c .

We first consider c as fixed in $\mathcal{C}_{p,q}$ and we define the $(p-1)$ -dimensional parameter $\Lambda(c)$ and the q -dimensional parameter $\alpha(c)$ in the following way:

$$\Lambda_j(c) = \lambda_j + \sum_{l=1}^q I_{\{c(l)=j\}} \lambda_{p+l} - \lambda_j^0, \quad 1 \leq j \leq p,$$

$$\alpha_l(c) = \lambda_{p+l} \| \theta_{p+l} - \theta_{c(l)}^0 \|, \quad 1 \leq l \leq q.$$

Let $\Lambda(c) = (\Lambda_1, \dots, \Lambda_{p-1})^T(c)$ and $\alpha(c) = (\alpha_1, \dots, \alpha_q)^T(c)$, and let $\mu(c)$ be the $\{p+q-1\}$ -dimensional vector $\mu(c) = (\Lambda^T, \alpha^T)^T(c)$. Since $\sum_{j=1}^p \lambda_j^0 = \sum_{j=1}^{p+q} \lambda_j = 1$, we have $\sum_{j=1}^p \Lambda_j(c) = 0$ and $(\Lambda_1, \dots, \Lambda_p)^T = M\Lambda$, where M is the $p \times (p-1)$ matrix having an identity upper $(p-1) \times (p-1)$ block and -1 for each element of the last row. As the hypothesis H_0 is equivalent to $\mu(c) = 0$, we now use the reparametrization $(\lambda^T, \theta^T)^T \mapsto (\mu^T, \theta^T)^T$, where $\mu = \mu(c)$, and we plug Θ^q into

$$\tilde{\Theta}_{c,q} = \{ (u, r) = (u_1, r_1, \dots, u_q, r_q); u_l \in \mathcal{U}, \theta_{c(l)}^0 + r_l u_l \in \bar{\Theta}, 1 \leq l \leq q \}.$$

With the functions $f_0 = (f_{\theta^0}, \dots, f_{\theta^p})^T$, and $f'_0 = ((f'_{\theta^0})^T, \dots, (f'_{\theta^p})^T)^T$, $g_{p+q;\lambda,\theta}$ is written as

$$\begin{aligned}
 g_{p+q;\lambda,\theta} &= g_0 + \sum_{j=1}^p \left[\lambda_j - \lambda_j^0 + \sum_{l=1}^q \lambda_{p+l} I_{\{c(l)=j\}} \right] f_{\theta_j^0} + \sum_{l=1}^q \lambda_{p+l} (f_{\theta_l} - f_{\theta_{c(l)}^0}) \\
 &= g_0 + \Lambda^T M^T f_0 + \sum_{l=1}^q \lambda_{p+l} (f_{\theta_l} - f_{\theta_{c(l)}^0}).
 \end{aligned}$$

As in Section 2, for any $1 \leq l \leq q$ and $c \in \mathcal{C}_{p,q}$, there exists a k -dimensional function $\phi_{c(l)}$ on $\bar{\Theta} \times \mathcal{R}$ such that for any $\theta \in \bar{\Theta}$, $f_\theta - f_{\theta_{c(l)}^0} = (\theta - \theta_{c(l)}^0)^T \phi_{c(l),\theta} g_0$. Then we consider the family of functions $h_{\mu,u,r,c}$ indexed by $c \in \mathcal{C}_{p,q}$, by $(u, r) = (u_1, r_1, \dots, u_q, r_q) \in \tilde{\Theta}_{c,q}$ and by the parameter μ ,

$$h_{\mu,u,r,c} = g_0 \left(1 + \Lambda^T M^T g_0^{-1} f_0 + \sum_{l=1}^q \alpha_l u_l^T \phi_{c(l),\theta_{c(l)}^0 + r_l u_l} \right).$$

Let $\tilde{Y}_n = (\tilde{Y}_{1n}^T, \tilde{Y}_{2n}^T)^T$ be the $(p + q - 1)$ -dimensional process defined by

$$\tilde{Y}_{1n} = n^{-1/2} \sum_{i=1}^n (M^T g_0^{-1} f_0)(X_i) = n^{-1/2} \sum_{i=1}^n \{g_0^{-1} (f_{\theta_j^0} - f_{\theta_p^0})(X_i)\}_{1 \leq j < p}$$

and, for $c \in \mathcal{C}_{p,q}$ and $(u, r) \in \tilde{\Theta}_{c,q}$, $\tilde{Y}_{2n}(u, r, c)$ is now the q -dimensional vector with components

$$\tilde{Y}_{2n,l}(u, r, c) = n^{-1/2} \sum_i u_l^T \{ \phi_{c(l),\theta_{c(l)}^0 + r_l u_l} \}(X_i), \quad 1 \leq l \leq q.$$

We also denote by $\tilde{\Sigma}(u, r, c)$ the covariance matrix of $\tilde{Y}_1(u, r, c)$, with a block decomposition according to the components of \tilde{Y}_1 ,

$$\tilde{\Sigma}(u, r, c) = \begin{pmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{12}(u, r, c) \\ \tilde{\Sigma}_{21}(u, r, c) & \tilde{\Sigma}_{22}(u, r, c) \end{pmatrix}.$$

Conditions 1–4 are not modified, except that Condition 2 now says that the function $\sup_{\lambda \in \mathcal{S}_{p+q}} \sup_{\theta = (\theta^{0T}, \xi^T)^T \in \bar{\Theta}_{p+q}} |\log g_{p+q;\lambda,\theta}|$ is P_0 -integrable.

Theorem 2. *Assume Conditions 1–4 hold. Then, under H_0 , the LR statistic for testing a mixture of the densities $f_{\theta_j^0}$, $1 \leq j \leq p$, against the alternative of $p + q$ densities from \mathcal{F} including $(f_{\theta_j^0})_{1 \leq j \leq p}$, converges weakly to*

$$\sup_{c \in \mathcal{C}_{p,q}} \sup_{(u,r) \in \tilde{\Theta}_{c,q}} Z^T(u, r, c) Z(u, r, c) I_{\{Z(u,r,c) > 0\}},$$

where Z is defined by $Z = \tilde{\Sigma}_{22}^{-1/2} (\tilde{\Sigma}_{21} Y_1 + \tilde{\Sigma}_{22} Y_2)$, and where $Y = (Y_1^T, Y_2^T)^T$ is a continuous centred $(p + q - 1)$ -dimensional Gaussian process with covariance $\tilde{\Sigma}$.

Proof. By the same arguments as for Theorem 1, for fixed $c \in \mathcal{C}_{p,q}$ and $(u, r) \in \tilde{\Theta}_{c,q}$, the MLE $\hat{\mu}_n(u, r, c)$ of μ under the constraints $\alpha_l \geq 0$, for $1 \leq l \leq q$, is P_0 -almost surely uniformly consistent (as the function h is still linear with respect to the parameter μ , Lemma 3 extends to multidimensional parameters). An expansion of the log-likelihood derivative leads to

$$n^{1/2}\hat{\mu}_n(u, r, c) = \tilde{\Sigma}(u, r, c)^{-1}\tilde{Y}_n(u, r, c)I_n(u, r, c) + \tilde{\Sigma}_{11}^{-1}\tilde{Y}_{1n}(1 - I_n(u, r, c)) + o_p(1),$$

where I_n is the indicator of the set $\{(\tilde{\Sigma}^{-1}\tilde{Y}_n)_p > 0, \dots, (\tilde{\Sigma}^{-1}\tilde{Y}_n)_{p+q-1} > 0\}$ and where the o_p is uniform with respect to $(u, r) \in \tilde{\Theta}_{c,q}$.

Under H_0 , $g_{p,\lambda,\theta^0} = g_0 + \sum_{j=1}^{p-1}(\lambda_j - \lambda_j^0)(f_{\theta_j^0} - f_{\theta_j^0})$ and the MLE of $\lambda = (\lambda_1, \dots, \lambda_p)$ satisfies

$$n^{1/2}\hat{\lambda}_n = \tilde{\Sigma}_{11}^{-1}\tilde{Y}_{1n} + o_p(1).$$

Hence, by a Taylor expansion, the LR statistic satisfies

$$S_n = \sup_{c \in \mathcal{C}} \sup_{p,q} \sup_{(u,r) \in \tilde{\Theta}_{c,q}} \{ \tilde{Y}_n(u, r, c)^T \tilde{\Sigma}^{-1}(u, r, c) \tilde{Y}_n(u, r, c) - \tilde{Y}_{1n} \tilde{\Sigma}_{11}^{-1} \tilde{Y}_{1n} \} I_n + o_p(1).$$

Writing

$$\tilde{\Sigma}^{-1} = \begin{pmatrix} \tilde{\Sigma}^{11} & \tilde{\Sigma}^{12} \\ \tilde{\Sigma}^{21} & \tilde{\Sigma}^{22} \end{pmatrix}$$

and using the relation $\tilde{\Sigma}^{11} = \tilde{\Sigma}_{11}^{-1} + \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21}$, we obtain

$$\tilde{Y}_n(u, r, c)^T \tilde{\Sigma}^{-1}(u, r, c) \tilde{Y}_n(u, r, c) - \tilde{Y}_{1n} \tilde{\Sigma}_{11}^{-1} \tilde{Y}_{1n} = \tilde{Z}_n^T \tilde{Z}_n,$$

where $\tilde{Z}_n = \tilde{\Sigma}_{22}^{-1/2}(\tilde{\Sigma}_{21}\tilde{Y}_{1n} + \tilde{\Sigma}_{22}\tilde{Y}_{2n})$. Hence the LR statistic is

$$S_n = \sup_{c \in \mathcal{C}} \sup_{p,q} \sup_{(u,r) \in \tilde{\Theta}_{c,q}} \tilde{Z}_n(u, r, c)^T \tilde{Z}_n(u, r, c) I_{\{\tilde{Z}_n(u,r,c) > 0\}} + o_p(1).$$

□

4. The mixture of two distributions

It is not obvious that a similar simple reparametrization allows to study the asymptotic behaviour of the LR statistic under general assumptions. We consider the mixture of two densities from the family \mathcal{F} ,

$$g_{\lambda,\theta_1,\theta_2} = (1 - \lambda)f_{\theta_1} + \lambda f_{\theta_2},$$

with $\lambda \in [0, 1]$ and $\theta_1, \theta_2 \in \Theta$. Let (X_1, \dots, X_n) be a sample with density g . The problem is to test the hypothesis of a single unknown density $H_0 : g \in \mathcal{F}$ against the alternative of a true mixture $H_1 : g = g_{\lambda,\theta_1,\theta_2}$ for some $\lambda \neq 0, 1$ and $\theta_1 \neq \theta_2$. By symmetry of $(1 - \lambda, \theta_1)$ and (λ, θ_2) in the mixture, we may assume without loss of generality that $\lambda \in [0, \frac{1}{2}]$. Under this restriction and by the identifiability assumption, H_0 can be viewed as $\lambda = 0$ or $\theta_1 = \theta_2$, when θ_2 or λ are respectively nuisance parameters, and in both cases $\theta_1 = \theta_0$. We assume that there is no restriction on the parameters of f_1 and f_2 : this is, for example, the test of a normal density against a mixture of two normals with a priori different means and variances. Let θ_0 be the true unknown value of the parameter under H_0 , $f_0 = f_{\theta_0}$, and P_0 a probability with density f_0 with respect to ν .

Let $L_n(\lambda, \theta_1, \theta_2)$ be the log-likelihood of (X_1, \dots, X_n) under the mixture density $g_{\lambda,\theta_1,\theta_2}$.

With the restriction $\lambda \in [0, \frac{1}{2}]$ and by Redner’s consistency result, the MLEs have the following properties: $\hat{\theta}_{1n}$ converges to the true parameter value θ_0 , and $\hat{\lambda}_n \rightarrow 0$ or $\hat{\theta}_{2n} \rightarrow \theta_0$, in P_0 -probability. Using the reparametrization $\mu = \lambda\|\theta_1 - \theta_2\|$, the score function for θ_1 determines the score for μ if θ_2 is close to θ_0 , and we cannot ensure that the information matrix remains uniformly positive definite as θ_1 and θ_2 are both in a neighbourhood of θ_0 .

The above approach can be modified with the parameters $\alpha = (1 - \lambda)\theta_1 + \lambda\theta_2$ and $\beta = \lambda\|\theta_2 - \theta_0\|^2 \geq 0$ (θ_0 being unknown, we could prefer to define $\beta = \lambda\|\theta_2 - \theta_1\|^2$ but the problem is similar). With $\lambda \in [0, \frac{1}{2}]$, H_0 is then equivalent to $\beta = 0$; furthermore, $\theta_1 = \theta_0 = \alpha$ under H_0 . The function ϕ is replaced by a $k \times k$ continuous matrix function Ψ on $\bar{\Theta} \times \times$ satisfying $f_\theta - f_0 = (\theta - \theta_0)^T f'_0 + f_0(\theta - \theta_0)^T \Psi_\theta(\theta - \theta_0)$, for any $\theta \in \bar{\Theta}$. If $\theta_2 = \theta_0 + ru$, the mixture density $g_{\lambda, \theta_1, \theta_2}$ is written as $h_{\alpha, \beta, u, r} + \frac{1}{2}(1 - \lambda)(\theta_1 - \theta_0)^T f''_{\theta^*}(\theta_1 - \theta_0)$, where

$$h_{\alpha, \beta, u, r} = f_0\{1 + (\alpha - \theta_0)^T f'_0 f_0^{-1} + \beta u^T \Psi_{\theta_0 + ru} u\}$$

and where θ^* is between θ_1 and θ_0 . Under the assumption that the map $\theta \mapsto f_\theta$ is twice continuously differentiable on $\bar{\Theta}$ and because $E_\nu f''_{\theta_0}(X_1) = 0$, we have $\sup_{\theta^*} \sum_i f''_{\theta^*}(X_i) = o(n)$ P_0 -almost surely, using the arguments of Lemma 2. With further integrability assumptions, $L_n(\lambda, \theta_1, \theta_2) = \sum_{i=1}^n \log(f_0^{-1} h_{\alpha, \beta, u, r})(X_i) + o_P(n\|\theta_1 - \theta_0\|^2)$, uniformly on $\tilde{\Theta}$. However, the components $(1 - \lambda)(\theta_1 - \theta_0)$ and $\lambda(\theta_2 - \theta_0)$ may balance in α and each may have a larger norm than α so that, generally, $\|\theta_1 - \theta_0\|^2$ cannot be compared to $\|\alpha\|^2 + \beta^2$.

In the special case where Θ is convex and where θ_0 lies on the border of the parameter space $\partial\Theta = \bar{\Theta} \setminus \Theta$, there exists a constraint for the sign of the components of $\theta - \theta_0$, for any $\theta \in \Theta$. The same constraint holds for the components of $\alpha - \theta_0$, therefore $|\alpha_j - \theta_{0j}| = (1 - \lambda)|\theta_{1j} - \theta_{0j}| + \lambda|\theta_{2j} - \theta_{0j}|$ and $\|\theta_1 - \theta_0\| \leq 2\|\alpha\|$, with $\lambda \in [0, \frac{1}{2}]$. Then $L_n(\lambda, \theta_1, \theta_2) = \sum_{i=1}^n \log(f_0^{-1} h_{\alpha, \beta, u, r})(X_i) + o_P(n\|\alpha\|^2)$ uniformly on $\tilde{\Theta}$ and an expansion of the first term of the right-hand side of this expression leads to a result similar to Theorem 1, but taking into account the constraint that θ_0 belongs to $\partial\Theta$ (cf. Geyer 1994). As this constraint is very restrictive, we do not detail that case. The limiting behaviour of the LR test for homogeneity in a mixture of two densities from the same family, without any restriction, is still an open question.

Acknowledgement

The authors are most grateful to a referee for helpful and constructive comments that led to a considerable improvement of the paper.

References

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.
 Chernoff, H. (1954) On the distribution of the likelihood ratio. *Ann. Math. Statist.*, **25**, 573–578.
 Chernoff, H. and Lander, E. (1995) Asymptotic distribution of the likelihood ratio test that a mixture of

- two binomials is a single binomial. *J. Statist. Plann. Inference*, **43**, 19–40.
- Davies, R. (1997) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64**, 247–254.
- Geyer, C. (1994) On the asymptotics of constrained M -estimation. *Ann. Statist.*, **22**, 1993–2010.
- Ghosh, J. and Sen, P. (1985) On the asymptotic performance of the log-likelihood ratio statistic for the mixture model and related results. In L. Le Cam and R.A. Olshen (eds), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, pp. 789–806.
- Ibragimov, I.A. and Has'minskii, R.Z. (1981) *Statistical Estimation: Asymptotic Theory*. New York: Springer-Verlag.
- Lemdani, M. and Pons, O. (1995) Tests for genetic linkage and homogeneity, *Biometrics*, **51**, 1033–1041.
- Redner, R. (1981) Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *Ann. Statist.*, **9**, 225–228.
- Self, S. and Liang, K.-Y. (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Soc.*, **82**, 605–610.
- Shorack, G. and Wellner, J. (1986) *Empirical Processes*. New York: Wiley.
- Teicher, H. (1963) Identifiability of finite mixtures. *Ann. Math. Statist.*, **39**, 1265–1269.
- Wald, A. (1949) Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.*, **20**, 595–601.

Received May 1995 and revised December 1997.