ESAIM: PS July 2009, Vol. 13, p. 276–282
DOI: 10.1051/ps:2008014

ESAIM: Probability and Statistics
www.esaim-ps.org

ESTIMATION AND TESTS IN FINITE MIXTURE MODELS OF NONPARAMETRIC DENSITIES

Odile PONS[1]

Abstract. The aim is to study the asymptotic behavior of estimators and tests for the components of identifiable finite mixture models of nonparametric densities with a known number of components. Conditions for identifiability of the mixture components and convergence of identifiable parameters are given. The consistency and weak convergence of the identifiable parameters and test statistics are presented for several models.

Résumé. Dans les modèles de mélanges de densités non paramétriques, une question est de déterminer le comportement asymptotique d'estimateurs et de statistiques de test sur les composantes identifiables. Des modèles de mélanges non paramétriques d'un nombre connu de densités sont considérés. Des conditions pour l'identifiabilité et pour les convergences des paramètres et fonctions identifiables sont présentées. Le comportement des statistiques de test est décrit et des estimateurs des composantes des densités sont définis dans plusieurs cas.

Mathematics Subject Classification. 62G10, 62H10.

Received April 25, 2007. Revised January 16, 2008.

1. Introduction

Consider a real random variable X on a probability space (Ω, A, P_0) and $\mathcal{F}$ a family of densities with respect to a probability μ on (Ω, A), such that the density $f_0 = dP_0/d\mu$ belongs to $\mathcal{F}$ and the functions of $\mathcal{F}$ are $L_2(\mu)$-integrable, it is a metric space with the metric $d_2(f, f') = \|f - f'\|_{L_2(\mu)}$.

For a fixed number p, the mixture of p components of $\mathcal{F}$ is defined by a vector of p unknown densities $f_{(p)} = (f_1, \ldots, f_p)$ of $\mathcal{F}^p$ with unknown proportions in $\mathcal{S}_p = \{\lambda_{(p)} = (\lambda_1, \ldots, \lambda_p) \in]0,1[^p; \sum_{j=1}^p \lambda_j = 1\}$. The density of the mixture is $g_{f_{(p)},\lambda_{(p)}} = \sum_{j=1}^p \lambda_j f_j$ and we consider the functional set $\mathcal{G}_p = \{g_{f_{(p)},\lambda_{(p)}}; f_{(p)} \in \mathcal{F}^p, \lambda_{(p)} \in \mathcal{S}_p\}$. The densities and the proportions of a mixture $g_{f_{(p)},\lambda_{(p)}}$ of $\mathcal{G}_p$ are identifiable if for any $f'_{(p)}$ in $\mathcal{F}^p$ and $\lambda'_{(p)}$ in $\mathcal{S}_p$, the equality $\sum_{j=1}^p \lambda_j f_j = \sum_{j=1}^p \lambda'_j f'_j$ is satisfied μ-a.s. if there exists a permutation π of $\{1, \ldots, p\}$ such that $\lambda'_j = \lambda_{\pi(j)}$ and $f'_j = f_{\pi(j)}$ μ-a.s., for $j = 1, \ldots, p$. The identifiability condition is written as a condition on the parameters of a nonparametric $\mathcal{F}$, it requires that the densities cannot be confounded and a mixture of two components f_1 and f_2 of $\mathcal{F}$ does not belong to $\mathcal{F}$.

For a parametric mixture of two components $g = \lambda f + (1-\lambda) f_0$ of $\mathcal{G}_2$, the likelihood ratio test (LRT) for the hypothesis $H_0 : g = f_0$ is studied through a reparametrization because the information matrix is not positive

Keywords and phrases. Mixture models, nonparametric densities, test, weak convergence.

[1] pons.odile@free.fr

Article published by EDP Sciences

definite and either λ or f may be considered as nuisance parameters under H_0. Conditions and proof of the convergence of the LRT as f_0 is known (admixture or contamination model) were given in [4,5] for parametric families of two or $p \geq 2$ components, for a test of a known density of $\mathcal{G}_q$ against $\mathcal{G}_p$, $q < p$. The results were generalized to an unknown density f_0 by plugging an estimator of f_0 in the same statistic, for identifiable families, and they are easily adapted for testing sub-models of the semi-parametric mixture $g = \sum_{j=1}^{p} \lambda_j f(\eta(\cdot - \mu))$, f in $\mathcal{F}$ and $\theta = (\eta, \mu)$ in a Borel compact subset of $\mathbb{R}^2$. The methods are extended here to nonparametric density families $\mathcal{G}_2$ with identifiable components, the parameter of interest for an admixture is $\lambda d_2(f, f_0)$ which is zero if and only if H_0 holds, for a mixture it is $\lambda d_2(f_1, f_2)$, with $\lambda \leq 1/2$. Wald's method [8] for the proof of consistency for the parametric case is no more sufficient and the asymptotic distribution is not Gaussian, as it was known for parametric distributions. The conditions are sufficient to remove the assumption of uniform convergence of small order terms present in most papers about expansions of the LRT in parametric models [1–4,6]. Under the alternative, direct estimators of λ and f are built and they satisfy the usual convergence. The tests are extended to finite mixtures of nonparametric densities and the estimators to mixtures of two nonparametric symmetric densities.

2. Admixture models for a density

Let $(X_1, \ldots, X_n)$ be an i.i.d. sample with density having two mixture components $g = \lambda f + (1-\lambda) f_0$ in $\mathcal{G}_{2,0}$, the subset of $\mathcal{G}_2$ of the mixtures with a known f_0. Denote $d_{\chi^2}(f, f') = \int (f - f')^2 f'^{-1} d\mu$ and assume that

A1: λ and $\mathcal{F}$ are identifiables: if $\lambda f + (1-\lambda) f_0 = \lambda' f' + (1-\lambda') f_0$, μ-a.s., with λ and λ' in $[0,1]$, f and f' in $\mathcal{F}$, then $\lambda = \lambda'$ and $f = f'$, μ-a.s.,

A2: There exists a function φ_0 in $L_2(P_0)$ s.t.

$$\lim \sup_{d_2(f,f_0)\to 0} d_2^{-1}(f, f_0) \| f_0^{-1}(f - f_0) - d_2(f, f_0)\varphi_0 \|_{L_2(P_0)} \to 0.$$

For f in $\mathcal{F}$ and $\lambda \in]0,1[$, denote $\mathcal{U} = \{d_2(f, f_0); f \in \mathcal{F}\} \subset \mathbb{R}_+$,

$$\alpha_f = \lambda d_2(f, f_0), \qquad \varphi_f = d_2^{-1}(f, f_0) f_0^{-1}(f - f_0) \text{ if } f \neq f_0, \tag{2.1}$$

and $\varphi_f = \varphi_0$ defined by **A2** if $f = f_0$, $D\mathcal{F} = \overline{\{\varphi_f; f \in \mathcal{F}\}}$, the $L_2(P_0)$-closure defined using **A2** of the set of functions φ_f. Let $\mathbb{E}_0$ the expectation under P_0, further conditions are

A3: $\mathbb{E}_0 \sup_{f\in\mathcal{F}} |\log f| < \infty$ and $\mathbb{E}_0 \sup_{\varphi\in D\mathcal{F}} \|\varphi(X)\|^3 < \infty$,

A4: $\sup_{\varphi\in D\mathcal{F}} |n^{-1/2} \sum_{i=1}^{n} \varphi(X_i) - G_\varphi|$ tends to zero in P_0-probability, with G_φ a centred Gaussian process with variance $\Sigma_\varphi = \mathbb{E}_0\varphi^2(X)$ and covariance $\Sigma_{\varphi_1,\varphi_2} = \mathbb{E}_0\varphi_1(X)\varphi_2(X)$.

Remark 1. The $L_2(P_0)$-norm of the convergence in A2 is a d_{χ^2}-derivability of $\mathcal{F}$ at f_0. It is not symmetric with respect to its arguments f and f_0, unlike the L_2 distance with respect to μ, and it is weaker than the Hellinger-derivability of $\mathcal{F}$. It appears naturally for the convergence under the null hypothesis of the variance of the likelihood ratio after the reparametrization. The function φ_0 is related to the direction of the approximation of f_0 by f.

Remark 2. By A2–A3 and the definition of $D\mathcal{F}$, $E\varphi(X) = 0$ and the weak convergence of $n^{-1/2} \sum_{i=1}^{n} \varphi(X_i)$ holds for every φ of $D\mathcal{F}$, and for $n^{-1/2} \sum_{i=1}^{n} \{\sup_{D\mathcal{F}} \varphi(X_i) - \mathbb{E}_0 \sup_{D\mathcal{F}} \varphi(X)\}$ by integrability of $\sup_{D\mathcal{F}} \varphi^2(X)$. Condition A4 is satisfied under a stronger condition on the dimension of $D\mathcal{F}$.

2.1. Tests for the density f_0

Under the alternative, a density is written

$$g_{\lambda,f} = f_0\{1 + \lambda f_0^{-1}(f - f_0)\} = f_0(1 + \alpha_f \varphi_f),\ \alpha_f \geq 0.$$

The parameters set for a test of f_0 against an admixture alternative is $\{\alpha_f \geq 0; f \in \mathcal{F}\}$ and the hypothesis H_0 is equivalent to the existence of a function f in $\mathcal{F}$ such that $\alpha_f = 0$, or $\inf_{f\in\mathcal{F}} \alpha_f = 0$. For a sample $(X_i)_{1\leq i\leq n}$ of X, the LRT statistic is written

$$\begin{aligned} T_n &= 2 \sup_{\lambda\in]0,1[,\ f\in\mathcal{F}} \sup \sum_{1\leq i\leq n} \{\log g_{\lambda,f}(X_i) - \log f_0(X_i)\} \\ &= 2 \sup_{\alpha\in\mathcal{U},\ \varphi\in D\mathcal{F}} \sup \sum_{1\leq i\leq n} l_n(\alpha,\varphi) \end{aligned}$$

with $l_n(\alpha,\varphi) = \sum_{1\leq i\leq n} \log\{1+\alpha\varphi(X_i)\}$.

Lemma 2.1. *The estimator $\widehat{\alpha}_n(\varphi) = \arg\max_{\alpha\in\mathcal{U}} l_n(\alpha,\varphi)$ converges P_0-a.s. to zero, uniformly on $\mathcal{U}$.*

Proof. For any $\varphi \in D\mathcal{F}$, the MLE $\widehat{g}_{\varphi,n} = f_0\{1+\widehat{\alpha}_n\varphi\}$ of a density $g_{\alpha,\varphi} = f_0\{1+\alpha\varphi\}$, with $\alpha \geq 0$ and $\widehat{\alpha}_n \geq 0$ converges at the parametric rate, therefore $\sup_{\mathbb{R}} |\widehat{g}_{\varphi,n} - f_0| \leq \sup_{\mathbb{R}} |\widehat{g}_n - f_0|$ for any other estimators $\widehat{g}_n$ without this parametric model. For every $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ s.t. for n large enough

$$\begin{aligned} P_0\{\sup_{\varphi\in D\mathcal{F}} \widehat{\alpha}_n(\varphi) > \varepsilon\} &\leq P_0\{\sup_{\varphi\in D\mathcal{F}} n^{-1}\{l_n(\widehat{\alpha}_n(\varphi),\varphi) - l_n(0,\varphi)\} > \delta_\varepsilon\} \\ &\leq P_0\{\int_{f_0>0} \sup_{\varphi\in D\mathcal{F}} \log\{1+\widehat{\alpha}_n(\varphi)\varphi\}\, d\widehat{P}_n > \delta_\varepsilon\}, \end{aligned}$$

for the empirical d.f. $\widehat{P}_n$ of P_0. Let $\bar{g}_n = \sup_{\varphi\in D\mathcal{F}} f_0\{1+\widehat{\alpha}_n(\varphi)\varphi\}$, then

$$\begin{aligned} P_0\{\sup_{\varphi\in D\mathcal{F}} \widehat{\alpha}_n(\varphi) > \varepsilon\} &\leq P_0\{\int_{f_0>0} \log\frac{\bar{g}_n}{f_0}\, d\widehat{P}_n > \delta_\varepsilon\} \\ &\leq P_0\{\int_{f_0>0} \log\frac{\widehat{g}_n}{f_0}\, d\widehat{P}_n > \delta_\varepsilon\} \\ &\leq P_0\{\int_{f_0>0} \log\frac{\widehat{g}_n}{f_0}\, d(\widehat{P}_n - P_0) > \delta_\varepsilon\} \end{aligned}$$

and this sequence tends to zero by the weak convergence of $\widehat{P}_n$. □

Let Y_n the process defined by $Y_n(\varphi) = n^{-1/2}\sum_{i=1}^n \varphi(X_i)$, with variance-covariance Σ, $Z_n = \Sigma^{-1/2}Y_n$ and Z be a centred and continuous Gaussian process on $D\mathcal{F}$, with covariance $\Sigma_{\varphi_1}^{-1/2}\Sigma_{\varphi_1,\varphi_2}\Sigma_{\varphi_2}^{-1/2}$.

Theorem 2.2. *Under H_0, $\sup_{\varphi\in D\mathcal{F}} \|n^{1/2}\widehat{\alpha}_n(\varphi) - (\Sigma_\varphi)^{-1/2}Z(\varphi)1_{\{Z(\varphi)>0\}}\|$ converges in probability to zero and T_n converges weakly to the variable $\sup_{\varphi\in D\mathcal{F}} Z^2(\varphi)1_{\{Z(\varphi)>0\}}$.*

Proof. Let φ in $D\mathcal{F}$, $\dot{l}_n(\cdot,\varphi)$ the derivative of $l_n(\cdot,\varphi)$ with respect to α and $\widehat{\alpha}_n(\varphi)$ under the constraint $\widehat{\alpha}_n \geq 0$, then $\widehat{\alpha}_n(\varphi) = 0$ or it is solution of

$$\begin{aligned} 0 &= n^{-1/2}\dot{l}_n(\varphi,\widehat{\alpha}_n(\varphi))1_{\{\widehat{\alpha}_n(\varphi)>0\}} \\ &= n^{-1/2}\sum_i \frac{a_i(\varphi)}{1+\widehat{\alpha}_n(\varphi)a_i(\varphi)}1_{\{\widehat{\alpha}_n(\varphi)>0\}}. \end{aligned} \tag{2.2}$$

For every $\varphi \in D\mathcal{F}$, let $a_i \equiv a_i(\varphi) = \varphi(X_i)$ and $R_{1n} = n^{-1}\sum_i a_i^3(1+\widehat{\alpha}_n a_i)^{-1}$. If $\widehat{\alpha}_n(\varphi) > 0$, the score equation is simply written

$$0 = Y_n(\varphi) - n^{1/2}\widehat{\alpha}_n(\varphi)\{n^{-1}\sum_i a_i^2(\varphi) - \widehat{\alpha}_n(\varphi)R_{1n}(\varphi)\}. \tag{2.3}$$

Under $\sup_{\varphi\in D\mathcal{F}} |n^{-1}\sum_i\{\varphi^2(X_i)-\Sigma_\varphi\}|$ converges P_0-a.s. to zero and (2.3) becomes

$$0 = Y_n(\varphi) - n^{1/2}\widehat{\alpha}_n(\varphi)\left\{\Sigma_\varphi + o_p(1) - \widehat{\alpha}_n(\varphi)R_{1n}(\varphi)\right\}.$$

As proved in [7]:

$$n^{1/2}\widehat{\alpha}_n(\varphi) \quad = \quad (\Sigma_\varphi)^{-1/2} Z_n(\varphi) 1_{\{Z_n(\varphi)>0\}} + o_p(1). \tag{2.4}$$

By integration of R_{1n}, $l_n(\widehat{\alpha}_n(\varphi),\varphi) = n^{1/2}\widehat{\alpha}_n(\varphi)Y_n(\varphi) - \frac{1}{2}\widehat{\alpha}_n^2(\varphi)\sum_i \varphi^2(X_i) + R_{2n}(\varphi)$ where

$$R_{2n} = \sum_i \int_0^{\widehat{\alpha}_n(\varphi)} \frac{a_i^3\alpha^2}{1+\alpha a_i}\,d\alpha = \sum_i \widehat{\alpha}_n \frac{\alpha_n^{*2}a_i^3}{1+\alpha_n^* a_i}$$

with $\alpha_n^*(\varphi)$ between 0 and $\widehat{\alpha}_n(\varphi)$. As the functions $\alpha \mapsto a_i^3(1+\alpha a_i)^{-1}$ are decreasing and from the bound established for R_{1n}, $n^{-1}|R_{2n}| \le \widehat{\alpha}_n^3 n^{-1}|\sum_i a_i^3|$. Then $\sup_{D\mathcal{F}} n^{-1}|R_{2n}(\varphi)| = o_p(1)$ and

$$l_n(\widehat{\alpha}_n(\varphi),\varphi) = n^{1/2}\widehat{\alpha}_n(\varphi)Y_n(\varphi) - \frac{n}{2}\widehat{\alpha}_n^2(\varphi)\Sigma_\varphi + o_p(1)$$

uniformly on $D\mathcal{F}$. The expression (2.4) of $\widehat{\alpha}_n(\varphi)$ implies a uniform approximation of T_n as $T_n = 2\sup_{\varphi\in D\mathcal{F}} l_n(\widehat{\alpha}_n(\varphi),\varphi) = \sup_{\varphi\in D\mathcal{F}}[Z_n^2(\varphi)1_{\{Z_n(\varphi)>0\}}] + o_P(1)$. □

With p additional components in the mixture under the alternative, the metric is written $d_2(f,f') = \sum_{k=1}^p \|f_k - f_k'\|_{L_2(\mu)}$ for two densities $f=(f_1,\ldots,f_p)$ and $f'=(f_1',\ldots,f_p')$. The density $g_{\lambda,f} = (1-\lambda)f_0 + \sum_{k=1}^p \lambda_k f_k$ becomes

$$g_{\lambda,f} = f_0\{1+\sum_{k=1}^p \lambda_k f_0^{-1}(f_k - f_0)\} = f_0(1+\alpha_f^T\varphi_f),$$

where the coefficient of f_0 satisfies $\lambda = \sum_{k=1}^p \lambda_k$, $\alpha_f = (\alpha_{fk})_{1\le k\le p}$ and $\varphi_f = (\varphi_{fk})_{1\le k\le p}$ are given by

$$\alpha_{fk} = \lambda_k d_2(f_k, f_{0k}), \qquad \varphi_{fk} = d_2^{-1}(f_k,f_{0k})f_{0k}^{-1}(f_k - f_{0k}) \text{ if } f\neq f_0, \tag{2.5}$$

and $\varphi_{0k} = f_{0k}^{-1}f_{0k}'$. The hypothesis H_0 is equivalent to $\|\alpha_f\| = 0$ and the result of Lemma 2.1 still holds. The process $Y_n(\varphi)$ is a p-dimensional process with a diagonal variance $\Sigma = (\Sigma_{\varphi_k})_{1\le k\le p}$. and Theorem 2.2 adapts for each component with the same proof, using tensor product and norms of the vectors already defined when necessary. Denoting $1_{\{Z(\varphi)>0\}}$ for the vector of indicators $1_{\{Z_k(\varphi)>0\}}$, we obtain

Theorem 2.3. *Under H_0, $\sup_{\varphi\in D\mathcal{F}} |n^{1/2}\|\widehat{\alpha}_n(\varphi)\| - \|(\Sigma_\varphi)^{-1/2}Z(\varphi)\|1_{\{Z(\varphi)>0\}}|$ converges in probability to zero and T_n converges weakly to the variable $\sup_{\varphi\in D\mathcal{F}}\|Z(\varphi)\|^2 1_{\{Z(\varphi)>0\}}$.*

When f_0 is already a known mixture of functions in $\mathcal{F}$, the result is extended as in [5] for mixtures of several parametric densities.

2.2. Estimation of the densities and mixture coefficients

Under the alternative, $\lambda_0 \notin \{0,1\}$ and the model for the mixture density is $g_0 = \lambda f + (1-\lambda)f_0$ where λ belongs to a close subset of $]0,1[$. We consider a class of functions $\mathcal{F}$ with compact supports not all confounded with the support $\operatorname{supp} F_0$ of the distribution function F_0, then both λ_0 and f are identifiable and may be estimated. If all the functions of $\mathcal{F}$ have the same support, further conditions are necessary for identifiability of the components of g_0. Estimators of λ and f are obtained after the estimation of the different supports.

Let $\widehat{G}_n$ the empirical estimator of the d.f. of the variable X, with support $\operatorname{supp}\widehat{G}_n = [\min_i X_i, \max_i X_i]$, and

$$\widetilde{G}_n(x) = n^{-1}\sum_{i=1}^n 1_{\{X_i\in \operatorname{supp} F_0\}}1_{\{X_i\le x\}} - F_0(x)$$

a consistent estimator of $(G_0 - F_0)1_{\{\operatorname{supp} F\cap \operatorname{supp} F_0\}} = \lambda(F - F_0)1_{\{\operatorname{supp} F\cap \operatorname{supp} F_0\}}$, with support estimated by $\operatorname{supp}\widetilde{G}_n$. An estimator of λ_0 is deduced from a classification of the observations into three intervals,

$$I_{1n} = \operatorname{supp}\widetilde{G}_n,\ I_{2n} = \operatorname{supp} F_0 \setminus \operatorname{supp}\widetilde{G}_n,\ I_{3n} = \operatorname{supp}\widehat{G}_n \setminus \operatorname{supp} F_0.$$

The following identities

$$\begin{aligned}
\int_{I_{2n}} \mathrm{d}\widehat{G}_n &= (1-\lambda)\int_{I_{2n}} \mathrm{d}F_0,\\
\int_{I_{1n}} 1_{x\le t}\,\mathrm{d}\widehat{G}_n(x) &= \lambda\int_{I_{1n}} 1_{x\le t}\,\mathrm{d}\widehat{F}_n(x) + (1-\lambda)\int_{I_{1n}} 1_{x\le t}\,\mathrm{d}F_0(x),\\
\int_{I_{3n}} 1_{x\le t}\mathrm{d}\widehat{G}_n(x) &= \lambda\int_{I_{3n}} 1_{x\le t}\,\mathrm{d}\widehat{F}_n(x),
\end{aligned}$$

imply simple expressions for the estimators of λ_0 and F on I_{1n} and I_{3n}

$$\begin{aligned}
\widehat{\lambda}_n &= 1 - \frac{\int_{I_{1n}}\mathrm{d}\widehat{G}_n}{\int_{I_{1n}}\mathrm{d}F_0}, \qquad (2.6)\\
\int_{I_{1n}} 1_{x\le t}\,\mathrm{d}\widehat{F}_n(x) &= \widehat{\lambda}_n^{-1}\left\{\int_{I_{1n}} 1_{x\le t}\,\mathrm{d}\widehat{G}_n(x) - \frac{\int_{I_{1n}}\mathrm{d}\widehat{G}_n}{\int_{I_{1n}}\mathrm{d}F_0}\int_{I_{1n}} 1_{x\le t}\,\mathrm{d}F_0(x)\right\},\\
\int_{I_{3n}} 1_{x\le t}\,\mathrm{d}\widehat{F}_n(x) &= \widehat{\lambda}_n^{-1}\int_{I_{3n}} 1_{x\le t}\mathrm{d}\widehat{G}_n(x).
\end{aligned}$$

Let K a symmetric kernel in $L_2(\mu)$ and h a window, $\widehat{g}_n(x) = n^{-1}\sum_{i=1}^n K_h(x - X_i)$ an estimator of g_0, then the unknown density of the mixture is estimated by

$$\begin{aligned}
\widehat{f}_n(x) &= \widehat{\lambda}_n^{-1}\{\widehat{g}_n(x) - \frac{\int_{I_{1n}}\mathrm{d}\widehat{G}_n}{\int_{I_{1n}}\mathrm{d}F_0} f_0(x)\},\ \text{ if } x\in I_{1n},\\
\widehat{f}_n(x) &= \widehat{\lambda}_n^{-1}\widehat{g}_n(x),\ \text{ if } x\in I_{3n}. \qquad (2.7)
\end{aligned}$$

Assume that the support of G is bounded and $\sigma_{1,\lambda}^2 = \int x^2\,\mathrm{d}(1-G)^n(x) - (\int x\,\mathrm{d}(1-G)^n(x))^2 < \infty$ if $\inf\ \operatorname{supp} F_0 < \inf\ \operatorname{supp} F$, and $\sigma_{2,\lambda}^2 = \int x^2\,\mathrm{d}G^n(x) - (\int x\,\mathrm{d}G^n(x))^2 < \infty$ if $\inf\ \operatorname{supp} F_0 > \inf\ \operatorname{supp} F$, then $n^{1/2}(\max_i X_i - \max \operatorname{Supp} G)$ and $n^{1/2}(\min_i X_i - \min \operatorname{Supp} G)$ converge weakly to centered Gaussian variables with variances $\sigma_{1,\lambda}^2$ and $\sigma_{1,\lambda}^2$. It follows that:

Proposition 2.4. *The estimators $\widehat{\lambda}_n$, $\widehat{F}_n$ and $\widehat{f}_n$ are uniformly P_0-a.s. consistent. The variable $n^{1/2}(\widehat{\lambda}_n - \lambda)$ converges weakly to a centered Gaussian variable, $n^{1/2}(\widehat{F}_n - F)$ converges weakly to a transformed Brownian bridge and if $\mathcal{F}$ is a subset of C_2 and $h = o(n^{-1/5})$, $n^{2/5}(\widehat{f}_n - f)$ converges weakly to a centered Gaussian process.*

3. Mixture model of nonparametric densities

Consider a mixture of two unknown distributions with densities in a family $\mathcal{F}$ of concave unimodal and symmetric densities on $\mathbb{R}$,

$$g_{\lambda,f_1,f_2} = \lambda f_1 + (1-\lambda) f_2.$$

An identifiable mixture of two symmetric densities of $\mathcal{F}$ does not belong to $\mathcal{F}$ even with overlapping densities since the mixture density is not concave or not symmetric, except when they have the same mean and in that case they are unidentifiable. Let H_0 the null hypothesis of a single unknown distribution of $\mathcal{F}$. The distribution f_0 under H_0 may be estimated by a symmetric estimator $\widehat{f}_{0n}$: let θ_0 the center of symmetry of f_0, estimated by the median $\widehat{\theta}_n$ of the sample and consider the transformed sample $X'_i = X_i$ if $X_i \geq \widehat{\theta}_n$, $X'_i = 2\widehat{\theta}_n - X_i$ if $X_i < \widehat{\theta}_n$ hence $(X'_1,\dots,X'_n)$ is a sample with asymptotic distribution $a_0^{-1} F_0 1_{]-\infty,\theta_0]}$, with $a_0 = \int_{-\infty}^{\theta_0} \mathrm{d}F_0$. The density f_0 is estimated by a kernel estimator $\widehat{f}_{0n}(x) = n^{-1}\sum_{i=1}^n K_h(x - X'_i)$ on $]-\infty, \widehat{\theta}_n]$ and by symmetry, $\widehat{f}_{0n}(x) = \widehat{f}_{0n}(2\widehat{\theta}_n - x)$ on $]\widehat{\theta}_n, \infty[$, it is P_0-a.s. uniformly consistent on $\mathbb{R}$.

In the following, any other constraint on the form of the set $\mathcal{F}$ may obviously replace the symmetry, for example a disymmetry with some proportions between both sides of the functions. The main condition about $\mathcal{F}$ is the identifiability of the components of mixtures of functions belonging to $\mathcal{F}$ and mixtures cannot themselves belong to $\mathcal{F}$. Under H_0, any estimator $\widehat{f}_{0n}$ of the single unknown density f_0 with the relevant constraint may be used.

A test of a mixture $g_{\lambda,f_1,f_2} = \lambda f_1 + (1-\lambda) f_2$ in $\mathcal{G}_2$, with f_1 and f_2 in $\mathcal{F}$, against the alternative of a single unknown distribution of $\mathcal{F}$ may be performed by replacing f_0 by $\widehat{f}_{0n}$ in the expression of T_n, with the statistic

$$S_n = 2 \sup_{\lambda\in]0,1/2],\, f\in\mathcal{F}} \sum_{1\leq i\leq n} \{\log g_{\lambda,f,\widehat{f}_{0n}}(X_i) - \log \widehat{f}_{0n}(X_i)\}.$$

The conditions for convergence of S_n are

A′1: λ and $\mathcal{F}$ are identifiable: if $\lambda f_1 + (1-\lambda) f_2 = \lambda' f'_1 + (1-\lambda') f'_2$, μ-a.s., with λ and $\lambda' \in]0,1/2]$, f_1, f_2, f'_1 and $f'_2 \in \mathcal{F}$, then $\lambda_1 = \lambda_2$, $f_1 = f'_1$ and $f_2 = f'_2$, μ-a.s.,

A′2: for all f in $\mathcal{F}$, the $L_2(P_0)$-derivative at f, φ_f exists:

$$\lim \sup_{d_2(f,f')\to 0} d_2^{-1}(f,f')\|f^{-1}(f'-f) - d_2(f,f')\varphi_f\|_{L_2(P_0)} \to 0.$$

For all f and f' in $\mathcal{F}$, denote $\alpha_{f',f} = \lambda d_2(f,f')$, $\varphi_{f',f} = d_2^{-1}(f,f') f^{-1}(f'-f)$ if $f' \neq f$ and $\varphi_{f,f} = \varphi_f$, $\mathcal{U}_f = \{d_2(f,f'); f'\in\mathcal{F}\} \subset \mathbb{R}_+$; $D\mathcal{F}_f = \overline{\{\varphi_{f',f}; f'\in\mathcal{F}\}}$ and $D\mathcal{F} = \overline{\{\varphi_{f',f}; f, f'\in\mathcal{F}\}}$.

A′3: $\mathbb{E}_0 \sup_{\lambda\in]0,1]}, \sup_{f_1,f_2\in\mathcal{F}} |\log \lambda f_1 + (1-\lambda) f_2| < \infty$ and there exists a sequence of neighborhoods V_{0n} of f_0 in $\mathcal{F}$ containing $(\widehat{f}_{0n})_n$ and converging to f_0 s.t.

$$\mathbb{E}_0 \sup_{f_{0n}\in V_{0n}} \sup_{\varphi_n\in D\mathcal{F}_{f_{0n}}} \|\varphi_n(X)\|^3 < \infty,$$

A′4: $\sup_{f_{0n}\in V_{0n}}, \sup_{\varphi_n\in D\mathcal{F}_{f_{0n}}} |n^{-1/2}\sum_{i=1}^n \varphi_n(X_i) - G_\varphi|$ converges in P_0-probability to zero, with G_φ defined as in A4.

The mixture density is written $g_{\lambda,f,f'} = \lambda f' + (1-\lambda) f = f(1 + \alpha_{f',f}\varphi_{f',f})$ and the parameter set for a test of H_0 is $\{\alpha_{f',f}; f, f'\in\mathcal{F}\}$. A sequence of parameters of smaller size is sufficient for the test with the statistic S_n, it is defined by $\{(\alpha_{f,\widehat{f}_{0n}})_n; f\in\mathcal{F}\}$. Let $u_{n,f} = u(f,\widehat{f}_{0n})$, $\alpha_{n,f} = \lambda u_{n,f}$, $\varphi_{n,f} = \varphi(f,\widehat{f}_{0n})$, and $g_{n,f} = \widehat{f}_{0n}(1 + \alpha_{n,f}\varphi_{n,f})$, $\mathcal{U}_n = \{u_{n,f}; f\in\mathcal{F}\}$ and $D\mathcal{F}_n = \{\varphi_{n,f}; f\in\mathcal{F}\}$. The statistic S_n is then written

$$S_n = 2 \sup_{\alpha_n\in\mathcal{U}_n/2,\, \varphi_n\in D\mathcal{F}_n} \log(1 + \alpha_n\varphi_n).$$

Under conditions A′, the proofs of Section 2 are easily extended, with the same notations.

Lemma 3.1. *For every φ_n in $D\mathcal{F}_n$, the estimator*

$$\widehat{\alpha}_n(\varphi_n) = \arg \max_{\alpha_n \in \mathcal{U}_n/2} l_n(\alpha, \varphi_n)$$

is such that $\limsup_{\varphi_n \in D\mathcal{F}_n} |\widehat{\alpha}_n(\varphi_n)|$ *converges P_0-a.s. to zero.*

Theorem 3.2. *Under H_0,* $\sup_{\varphi_n \in D\mathcal{F}_n} |n^{1/2}\widehat{\alpha}_n(\varphi_n) - (\Sigma_\varphi)^{-1/2} Z_(\varphi) 1_{\{Z_(\varphi)>0\}}|$ *converges in probability to zero and the statistic S_n converges weakly to* $\sup_{\varphi \in D\mathcal{F}} Z^2(\varphi) 1_{\{Z(\varphi)>0\}}$.

As in Section 2, the convergence rate is due to the sum of n i.i.d. variables and does not depend on the convergence rate of φ_n.

With a vector of p additional components $f' = (f_1, \ldots, f_p)$ to an unknown density f with true value f_0 in the mixture under the alternative, $g_{\lambda,f,f'} = (1-\lambda)f + \sum_{k=1}^p \lambda_k f_k$ in $\mathcal{G}_p$ is written with the notations of A_2' as

$$g_{\lambda,f,} = f_0(1 + \alpha^T_{f',f} \varphi_{f',f}),$$

the hypothesis H_0 is equivalent to $\|\alpha_f\| = 0$ and the results of Lemma 2.1 and Theorem 2.3 hold. They are extended to a model with a mixture of identifiable components under H_0, as in Section 2.

Under the alternative of a mixture of two functions of $\mathcal{F}$ with different supports, the components of the true density $g_0 = \lambda_0 f_{1,0} + (1-\lambda_0) f_{2,0}$ may be estimated. For densities with known or estimated supports (densities with estimable center of symmetry), λ_0 is estimated as in (2.6) and both symmetric densities are estimated by an estimator similar to $\widehat{f}_{0n}$ defined in this section. If only one center of symmetry of $f_{1,0}$ or $f_{2,0}$ may be estimated, one density is estimated by this method and the other density is defined by (2.7) where f_0 is replaced by $\widehat{f}_{0n}$. If both supports are unknown, a learning sample is necessary for the estimation of the support of the densities.

References

[1] J.M. Azaïs, E. Gassiat and C. Mercadier, Asymptotic distribution and local power of the likelihood ratio test for mixtures: bounded and unbounded cases. *Bernoulli* **12** (2006) 775–799.

[2] K. Fukumizu, Likelihood ratio of unidendifiable models and multilayer neural networks, *Ann. Statist.* **31** (2003) 833–851.

[3] J.K. Ghosh and P.K. Sen, On the aymptotic performance of the log-likelihood ratio statistic for the mixture model and related results. In *Proc. of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer*, volume II, edited by L.M. Le Cam and R.A. Olshen, (1985), pp. 789–806.

[4] M. Lemdani and O. Pons, Likelihood ratio tests in mixture models. *C. R. Acad. Sci. Paris, Ser. I* **322** (1995) 399–404.

[5] M. Lemdani and O. Pons, Likelihood ratio tests in contamination models, *Bernoulli* **5** (1999) 705–719.

[6] X. Liu and Y. Shao, Asymptotics for likelihood ratio tests under loss of identifiability, *Ann. Statist.* **31** (2003) 807–832.

[7] O. Pons, *Estimation et tests dans les modèles de mélanges de lois et de ruptures*. Hermès-Science Lavoisier, London and Paris (2009).

[8] A. Wald, Note on the consistency of the maximum likelihood estimate, *Ann. Math. Statist.* **20** (1949) 595–601.