Ann. Inst. Statist. Math.
Vol. 54, No. 3, 585–594 (2002)

LIKELIHOOD RATIO STATISTIC FOR EXPONENTIAL MIXTURES

GABRIELA CIUPERCA

Laboratoire de Probabilités, Combinatoire et Statistique, Domaine de Gerland, Bât. Recherche B, Univ. Lyon 1, 50 Av. Tony-Garnier, 69366 Lyon cedex 07, France,
e-mail: Gabriela.Ciuperca@univ-lyon1.fr

(Received July 17, 2000; revised February 5, 2001)

Abstract. Let $f_0(x)$ be the exponential density and $f_\gamma(x)$ the translation model. Let $(X_i)_{i=1,n}$ be i.i.d. random variables, with density g. We test that g is f_0 against g is a simple mixture, using the LRT statistic. We prove that the LRT diverges to infinity with probability 1/2 and it is equal to 0 with probability 1/2. Therefore, the classical likelihood limiting theory does not hold.

Key words and phrases: Mixture models, likelihood test, exponential distribution.

1. Introduction

Let $X_1, X_2, \dots, X_n$ be independently and identically distributed random variables, having the density g. We suppose that g is a mixture of densities. Let $\mathcal{F} = \{f_\gamma; \gamma \in \Gamma\}$ be a family of densities. The set Γ is a compact subset of $\mathbb{R}^l$ for some integer l. The densities f are known up to the parameter γ. By definition, the set of all p-mixtures of densities of $\mathcal{F}$ is the set $\mathcal{G}_p$ defined by

$$\left\{ g_{\pi,\alpha} = \sum_{i=1}^{p} \pi_i f_{\gamma_i} / \pi = (\pi_1, \dots, \pi_p), \alpha = (\gamma_1, \dots, \gamma_p), \right.$$
$$\left. \forall i = 1, \dots, p, \gamma_i \in \Gamma, 0 \le \pi_i \le 1, \sum_{i=1}^{p} \pi_i = 1 \right\}.$$

The unknown parameters of the mixture are π and α.

The mixture models have enormous importance in applications, see e.g. Everitt and Hand (1981), Titterington *et al.* (1985), McLachlan and Basford (1988), Lindsay (1995) and Roberts *et al.* (1998).

An important problem is assessing the number of components, i.e. to test if g is a mixture of p-densities or a mixture of q-densities, with $q < p$. The Likelihood ratio tests (LRT) are the most commonly used methodology for generating tests in parametric models. The determination of the limiting distribution of the LRT statistic in the mixture model has been for many years an open problem. One of the key difficulties is that the parameters are not identifiable under the null hypothesis, so one cannot apply the standard limiting distribution theory. For example, for 2-mixture of densities, the model is:

$$g_{\pi,\gamma_1,\gamma_2} = (1-\pi) f_{\gamma_1} + \pi f_{\gamma_2} \qquad \pi \in [0,1]. \tag{1.1}$$

The parameters are: π, γ_1 and γ_2. The model is not identifiable for these parameters. There exists mixtures g which have different representations $g_{\pi,\gamma_1,\gamma_2}$ with different parameters π, γ_1 and γ_2. For example, we can write:

$$\forall \pi \in [0,1] \quad \forall \gamma \in \Gamma : \quad f_\gamma = (1-\pi) f_\gamma + \pi f_\gamma.$$

A partial solution for the mixture problem has been given by Ghosh and Sen (1985) for testing one density against 2-mixture of densities. But they impose a rather strong assumption concerning the distance between the parameters. They assumed that the model (1.1) verifies $|\gamma_1 - \gamma_2| \geq \varepsilon$, for a *fixed* positive ε. Redner (1981) proved that the maximum likelihood estimators for finite mixtures with compact parameter space is consistent in the quotient parameter space. Bickel and Chernoff (1993) gave the asymptotic distribution of the LRT statistic in a nonregular problem.

Dacunha-Castelle and Gassiat (1997) have proposed a complete solution without any assumption on the parameters but with conditions of regularity on the densities f. In their paper, a reparametrization: $(\theta, \beta) \in \Theta \times \mathcal{B}$, is introduced. The driving idea is to parametrize in such a way that one of the parameters is identifiable at the previously non identifiable point, so that it is possible to have asymptotic expansions in its neighborhood, and the other parameter contains all the non identifiability. The parameter θ can be thought around the true point as something close to the Kullback distance, the parameter β can be thought as a "direction". It can not be consistently estimated. This parametrization is used to derive the asymptotic distribution of the LRT statistic. The key point for the asymptotic convergence is to assume that the closure $\mathcal{D}$ of the derivatives of the log-likelihood with respect to θ, in any direction β, at $\theta = 0$, is a Donsker class. Roughly speaking, a Donsker class is a set of functions for which the empirical distributions (with i.i.d. variables) verify a uniform central limit theorem, with limit distribution a Gaussian process. Ciuperca (1999) gave accurately conditions for the parametric densities f_γ so that the set $\mathcal{D}$ is Donsker.

A particular case is to test a p-mixture of densities against a density f_{γ_0}, with $\gamma_0 \in \Gamma$. The LRT statistic is:

$$T_n = \sup_{g \in \mathcal{G}_p} (l_n(g) - l_n(f_{\gamma_0})) = \sup_{\pi,\alpha} \sum_{i=1}^{n} \log \left[1 + \frac{g_{\pi,\alpha} - g_0}{g_0}(X_i)\right]$$

where $g_0 = g_{\pi,0} = f_{\gamma_0}$ and $l_n(g)$ is the log-likelihood: $l_n(g) = \sum_{i=1}^{n} \log(g_{\pi,\alpha}(X_i))$.

In this paper, we consider the translation model: $f_\gamma(x) = f_0(x-\gamma)$, $\gamma \in \Gamma$, with f_0 the exponential density. We test $g = f_0$ against $g = (1-\pi)f_0 + \pi f_\gamma$, i.e. g is the exponential density against g is a simple mixture. First, we prove that the set $\mathcal{D}$ is not a Donsker class. It will be shown that the LRT statistic diverges to $+\infty$, with probability $1/2$, and it is equal to 0 with probability $1/2$. Therefore the limiting distribution cannot be used to set critical values. This result is due to the fact that $\mathcal{D}$ is not relatively compact even though the set of parameters is compact. The theoretical result is confirmed be a numerical study.

To the author's knowledge this is the first example of mixture hypothesis such that the LRT statistic diverges to infinity when the parameters belong to a compact set. For testing a mixture against an underlying function f_0, Hartigan (1985) proved that the LRT can converge to $+\infty$ if the parameters space is unbounded.

The choise of the number of components p in a mixture model has been considered using reversible-jump MCMC methods by Richardson and Green (1997). The acceptance probabilities for the split move have the form $\min(1, A)$ where A depends to likelihood ratio. Our result implies that the MCMC algorithm for the exponential simple mixtures chooses every other time to remove a component.

The rest of this paper is organized as follows. Section 2 introduces the model of simple mixtures. In Section 3, we prove that the set $\mathcal{D}$ is not Donsker. An useful asymptotic result concerning the Brownian bridge on $I\!R_+$ is established. Finally, we study the asymptotic behaviour of the likelihood test statistic.

2. Preliminaries on simple mixtures

To fix the problem, we consider $X_1, \ldots, X_n$ i.i.d. random variables with the density g. Let $\mathcal{F} = \{f_\gamma; \gamma \in \Gamma\}$ be a parametric family of densities and $\gamma_0 \in \Gamma$.

DEFINITION 2.1. The model of simple mixtures (or contamination model) of density of $\mathcal{F}$ is:

$$\mathcal{G}_2^0 = \{g_{\pi,\gamma} = (1-\pi) f_{\gamma_0} + \pi f_\gamma / 0 \leq \pi \leq 1, \gamma \in \Gamma\}.$$

The model is a subset of $\mathcal{G}_2$. We test:

$$H_0 : \; g = f_{\gamma_0} \qquad \text{against} \qquad H_1 : \; g \in \mathcal{G}_2^0.$$

The log-likelihood ratio statistic is:

$$\sum_{i=1}^{n} \log \left[1 + \pi \frac{f_\gamma - f_{\gamma_0}}{f_{\gamma_0}} (X_i) \right]. \tag{2.1}$$

Define the Hilbert space $L^2(f_{\gamma_0}\nu)$, with ν a positive measure on $I\!R$. Since the model is not identifiable, in order to test the hypothesis H_0, Dacunha-Castelle and Gassiat (1997) have proposed the parametrization:

$$\theta = \left\| \frac{g_{\pi,\gamma} - g_0}{g_0} \right\|_{L^2(f_{\gamma_0}\nu)} = \pi \left\| \frac{f_\gamma - f_{\gamma_0}}{f_{\gamma_0}} \right\|_{L^2(f_{\gamma_0}\nu)} ; \quad \beta = \gamma. \tag{2.2}$$

Let be the norm in the space $L^2(f_{\gamma_0}\nu)$:

$$N(\beta) = \left\| \frac{f_\beta - f_{\beta_0}}{f_{\beta_0}} \right\|_{L^2(f_{\gamma_0}\nu)}.$$

The LRT takes the form

$$T_n = \max_{\theta,\beta} \sum_{i=1}^{n} \log \left[1 + \frac{\theta}{N(\beta)} \cdot \frac{f_\beta - f_{\beta_0}}{f_{\beta_0}} (X_i) \right].$$

We denote by $g'_{(\theta,\beta)}(x)$ the partial derivatives of $g_{(\theta,\beta)}(x)$ with respect to θ. The set of the derivatives of the log-likelihood with respect to θ at $\theta = 0$ is:

$$\mathcal{D} = \left\{ d(\beta, x) = \frac{g'_{0,\beta}}{g_0}(x) / \beta \in \mathcal{B} \right\} \tag{2.3}$$

with:

$$d(\beta, x) = \frac{f_\beta(x) - f_{\beta_0}(x)}{f_{\beta_0}(x)} \frac{1}{N(\beta)}.$$

Before giving the asymptotic distribution of statistic T_n, we recall the definition of the Donsker class (see e.g. Van der Vaart and Wellner (1996)).

Let $Y_1, \ldots, Y_n$ be i.i.d. random variables with the common distribution P. The empirical measure $\boldsymbol{P}_n$ of $Y_1, \ldots, Y_n$ is the discrete random measure given by

$$\boldsymbol{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$$

with δ the Dirac measure. Given a collection $\mathcal{H} = \{h\}$ of measurable functions, the $\mathcal{H}$-indexed empirical process $\boldsymbol{G}_n$ is given by

$$\boldsymbol{G}_n h = \sqrt{n}\,(\boldsymbol{P}_n - P)\,h = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [h(Y_i) - Ph] \qquad h \in \mathcal{H}$$

with $Ph = \int h dP$.

Definition 2.2. The set of measurable functions $\mathcal{H} \in L^2(P)$ is a P-Donsker class, if almost surely $\sup_{h \in \mathcal{H}} |\boldsymbol{G}_n h - \boldsymbol{G} h| \to 0$ where the limit process $\{\boldsymbol{G}h; h \in \mathcal{H}\}$ is a zero-mean Gaussian process with the covariance function $E[\boldsymbol{G}h_1 \boldsymbol{G}h_2] = Ph_1h_2 - Ph_1Ph_2$.

We have the following asymptotic result.

Theorem 2.1. (Dacunha-Castelle and Gassiat (1997)) *Under the following regularity conditions for $f_\gamma(x)$.*

(M1) *There exists a function u in $L^1(f_{\gamma_0}\nu)$ such that $\forall f \in \mathcal{F}$, $|\log f| \leq u$ ν-a.e.*

(M2) *f_γ is continuously differentiable ν-a.e. with respect to $\gamma = (\gamma_1, \ldots, \gamma_l)$ in the interior on Γ. Moreover, there exists a function v such that*

$$\forall \gamma \in \Gamma, \qquad \left| \frac{1}{f_\gamma} \frac{\partial f_\gamma}{\partial \gamma_i} \right|, \quad i = 1, \ldots, l \quad E_{f_{\gamma_0}\nu}[v^2] < +\infty$$

and under the condition that the set $\mathcal{D}$ defined by (2.3) is a Donsker class, the statistic T_n converges to the supremum of a square Gaussian process

$$\frac{1}{2} \sup_{d \in \mathcal{D}} (\boldsymbol{G}d)^2 \cdot 1_{\boldsymbol{G}d \geq 0}$$

with $\boldsymbol{G}d$ a Gaussian process on $\mathcal{D}$ with covariance the usual scalar product in $L^2(f_{\gamma_0}\nu)$.

3. Exponential case

Let f_0 be the exponential density: $f_0(x) = e^{-x} 1_{x>0}$ and $f_\gamma(x) = f_0(x - \gamma)$. We consider the parameter set $\Gamma = [0, G]$, $G > 0$. We take $\gamma_0 = 0$. For this density, the test becomes:

$$H_0 : \ g = f_0 \quad \text{against} \quad H_1 : \ g \in \mathcal{G}_2^0. \tag{3.1}$$

The log-likelihood ratio statistic is:

$$l_n(\pi,\gamma) - l_n(0) = \sum_{i=1}^{n} \log[1 + \pi(e^{\gamma} 1_{X_i > \gamma} - 1) 1_{X_i > 0}] \tag{3.2}$$

where $l_n(\pi,\gamma) = \sum_{i=1}^{n} \log((1-\pi) f_0(X_i) + \pi f_0(X_i - \gamma))$ and $l_n(0) = l_n(\pi, 0) = \sum_{i=1}^{n} \log f_0(X_i)$.

We take ν the Lebesgue measure on $\mathbb{R}$ and we denote the space $L^2(f_0\nu)$ by $L^2(f_0)$.

PROPOSITION 3.1. *Assume that we test* (3.1) *with f_0 the exponential density and $f_\gamma(x) = f_0(x-\gamma)$. Then the set $\mathcal{D}$ is not relatively compact in $L^2(f_0)$.*

PROOF OF PROPOSITION 3.1. We first observe that for any $\gamma_n \searrow 0$ and $x > \gamma_n$:

$$\frac{f_0(x-\gamma_n) - f_0(x)}{f_0(x)} = O(\gamma_n) \quad \text{and} \quad N(\gamma_n) = O(\sqrt{\gamma_n}).$$

So: $\forall x > \gamma_n$, $\forall \gamma_n \searrow 0$, we have $d(\gamma_n, x) \to 0$ for $n \to \infty$. Now, if $d(\gamma_n, .) \to \bar{d}(\cdot)$ in $L^2(f_0)$ then we can extract a subsequence (γ_{n_k}) of (γ_n) such that: $d(\gamma_{n_k}, x) f_0^{1/2}(x) \to \bar{d}(x) f_0^{1/2}(x)$ pointwise, for almost every x. Then $\bar{d}(x) = 0$ for almost every $x > 0$, incompatible with $\|\bar{d}\|_{L^2(f_0)} = 1$. □

Then, conforming to Van der Vaart and Wellner (1996), the set $\mathcal{D}$ is not a Donsker class. To prove the main result, we need the following lemma:

LEMMA 3.1. *Let S be a Brownian motion and $\mathcal{B}$ a Brownian bridge on $\mathbb{R}_+$. Let $(a_n)_{n\geq 1}$ be the sequence such that: $a_n = n^{1/4}$, $n \geq 1$. Thus:*

$$\overline{\lim_{n\to\infty}} \sup_{a_n \leq t \leq 2a_n} \frac{S(t)}{\sqrt{2t \log(\log t)}} \geq 1 \quad a.s \tag{3.3}$$

$$\overline{\lim_{n\to\infty}} \sup_{\frac{1}{2a_n} \leq t \leq \frac{1}{a_n}} \frac{\mathcal{B}(t)}{\sqrt{2t \log(|\log t|)}} \geq 1 \quad a.s. \tag{3.4}$$

PROOF OF LEMMA 3.1. Using Lemma 2.1 on page 610 from Hanson and Russo (1983) one gets (where $\mathbf{N}$ is the set of positive integers)

$$\overline{\lim_{n\to\infty}} \sup_{a_n \leq t \leq 2a_n} \frac{S(t)}{\sqrt{2t \log(\log t)}} \geq \overline{\lim_{n\to\infty}} \sup_{\substack{a_n \leq t \leq 2a_n \\ t \in \mathbf{N}}} \frac{S(t)}{\sqrt{2t \log(\log t)}}$$

and using the law of the iterated logarithm for Brownian process (see e.g. Theorem 1 on page 72 from Shorack and Wellner (1986)) we have

$$\overline{\lim_{n\to\infty}} \sup_{\substack{a_n \leq t \leq 2a_n \\ t \in \mathbf{N}}} \frac{S(t)}{\sqrt{2t \log(\log t)}} \geq \overline{\lim_{n\to\infty}} \frac{S(n)}{\sqrt{2n \log(\log n)}} = 1 \quad \text{a.s.}$$

The inequality (3.3) follows.

Since $S(t)$ is a Brownian motion on $[0,\infty)$, the process $\{tS(\frac{1}{t}), t \geq 0\}$ is also a Brownian motion (see Shorack and Wellner (1986)). Thus:

$$1 \leq \overline{\lim_{n\to\infty}} \frac{tS(\frac{1}{t})}{\sqrt{2t\log(\log t)t}} \overset{\text{a.s}}{=} \overline{\lim_{n\to\infty}} \sup_{\frac{1}{2a_n}\leq r\leq \frac{1}{a_n}} \frac{S(r)}{r\sqrt{\frac{2}{r}\log(|\log r|)}}.$$

Hence

$$1 \leq \overline{\lim_{n\to\infty}} \sup_{\frac{1}{2a_n}\leq r\leq \frac{1}{a_n}} \frac{S(r)}{\sqrt{2r\log(|\log r|)}} \quad \text{a.s.} \tag{3.5}$$

We can represent $\mathcal{B}(t)$ as $S(t) - t\cdot S(1)$. Then:

$$\overline{\lim_{n\to\infty}} \sup_{\frac{1}{2a_n}\leq t\leq \frac{1}{a_n}} \frac{\mathcal{B}(t)}{\sqrt{2t\log(|\log t|)}} = \overline{\lim_{n\to\infty}} \sup_{\frac{1}{2a_n}\leq t\leq \frac{1}{a_n}} \frac{S(t)-tS(1)}{\sqrt{2t\log(|\log t|)}} \quad \text{a.s.}$$

Inequality (3.5) and

$$\lim_{t\to 0} \frac{t}{\sqrt{2t\log(|\log t|)}} = 0$$

prove inequality (3.4). □

Our main result is present in the following theorem:

THEOREM 3.1. *If f_0 is the exponential density and $f_\gamma(x) = f_0(x-\gamma)$, then the LRT statistic $T_n = \sup_{\pi,\gamma} l_n(\pi,\gamma) - l_n(0)$ converges, as $n\to\infty$, to $+\infty$ with probability 1/2 and it is 0 with probability 1/2.*

PROOF OF THEOREM 3.1. Let F_0 be the distribution function associated to f_0. We consider the random variable: $U_i = F_0(X_i)$. Note that U_i has the uniform distribution. Making the change of variables $F_0(\gamma) = \xi$, we get:

$$\xi = 1 - e^{-\gamma} \quad \text{with} \quad \xi \in [0, 1-e^{-G}]. \tag{3.6}$$

It is easy to see that $1_{X_i<\gamma} = 1_{U_i<\xi}$. Let F_n be the empirical distribution function of $U_1, U_2, \ldots, U_n$. The expression (3.2) of the log-likelihood ratio statistic becomes:

$$nF_n(\xi)\log(1-\pi) + n[1-F_n(\xi)]\log\left(1+\frac{\pi\xi}{1-\xi}\right). \tag{3.7}$$

The new parameters (τ, λ) are defined by:

$$\tau = \pi\xi, \qquad \lambda = \xi \tag{3.8}$$

with τ, $\lambda \in \Lambda$, $\Lambda = \left[0, 1-e^{-G}\right]$. Now, the expression (3.7) takes the form:

$$l_n(\tau,\lambda) - l_n(0) = nF_n(\lambda)\log\left(1-\frac{\tau}{\lambda}\right) + n[1-F_n(\lambda)]\log\left(1+\frac{\tau}{1-\lambda}\right).$$

Let $\hat{\tau}_n^\lambda = \arg\max_\tau l_n(\tau,\lambda)$ be directional estimator of τ when λ is fixed. Its form is: $\hat{\tau}_n^\lambda = \max(\lambda - F_n(\lambda), 0)$. Also, we have:

$$\frac{\partial^2 l_n(\tau,\lambda)}{\partial\tau^2} = -nF_n(\lambda)\frac{1}{(\lambda-\tau)^2} - n[1-F_n(\lambda)]\frac{1}{(1-\lambda-\tau)^2} < 0, \qquad \forall \tau \in \Lambda.$$

Then $\hat{\tau}_n^\lambda$ is a maximizator of $l_n(\tau,\lambda)$. With regard to the sign of $\lambda - F_n(\lambda)$ we remark that $P[\lambda \le F_n(\lambda)] = P[\lambda > F_n(\lambda)] = 1/2$ (see Dacunha-Castelle and Duflo (1990)). We then consider two cases:

Case I. $\lambda \le F_n(\lambda)$. In this case $\hat{\tau}_n^\lambda = 0$ and $T_n = \max_{(\tau,\lambda)\in\Lambda\times\Lambda} l_n(\tau,\lambda) - l_n(0) = 0$.

Case II. $\lambda > F_n(\lambda)$. We have $\hat{\tau}_n^\lambda = (\lambda - F_n(\lambda))1_{\lambda > F_n(\lambda)}$.

The expression of $\sup_\tau [l_n(\tau,\lambda) - l_n(0)]$ becomes:

$$nF_n(\lambda)\log\left[1 - \frac{\lambda - F_n(\lambda)}{\lambda}1_{\lambda>F_n(\lambda)}\right] + n[1-F_n(\lambda)]\log\left[1 + \frac{\lambda - F_n(\lambda)}{1-\lambda}1_{\lambda>F_n(\lambda)}\right]. \tag{3.9}$$

Since $\hat{\tau}_n^\lambda = [\lambda - F_n(\lambda)]1_{\lambda>F_n(\lambda)} \to^P 0$, uniformly on λ, as $n \to \infty$ (see Dacunha-Castelle and Duflo (1990)) we can use Taylor expansion for $\log[1 - [\lambda - F_n(\lambda)]/(1-\lambda)1_{\lambda>F_n(\lambda)}]$. For the expression (3.9), we will check that $[\lambda - F_n(\lambda)]/\lambda \cdot 1_{\lambda>F_n(\lambda)} \mapsto^{a.s} 0$, as $n \to \infty$. But

$$\sup_{\lambda\in\Lambda} \sqrt{n}\cdot[F_n(\lambda) - \lambda] \overset{a.s}{\longmapsto} 0 \quad \text{for} \quad n \to \infty.$$

Then a sufficient condition for $[\lambda - F_n(\lambda)]/\lambda \cdot 1_{\lambda>F_n(\lambda)} \mapsto^{a.s} 0$ is that $\lambda \in [\lambda_n, 1 - e^{-G}]$, where (λ_n) is a positive sequence decreasing to 0 and:

$$\lim_{n\to\infty} \sqrt{n}\cdot\lambda_n = \infty. \tag{3.10}$$

Under these conditions, we make the expansion

$$\begin{aligned} l_n(\hat{\tau}_n^\lambda,\lambda) - l_n(0) &= nF_n(\lambda)\left[-\frac{\lambda - F_n(\lambda)}{\lambda} - \frac{1}{2}\left(\frac{\lambda - F_n(\lambda)}{\lambda}\right)^2\right]1_{\lambda>F_n(\lambda)}[1+o(1)] \\ &\quad + n[1-F_n(\lambda)]\left[\frac{\lambda - F_n(\lambda)}{1-\lambda} - \frac{1}{2}\left(\frac{\lambda - F_n(\lambda)}{1-\lambda}\right)^2\right]1_{\lambda>F_n(\lambda)}[1+o(1)] \\ &= \frac{1}{2}n\frac{[\lambda - F_n(\lambda)]^2}{\lambda(1-\lambda)}1_{\lambda>F_n(\lambda)}\left[1 + O\left(\frac{(\lambda - F_n(\lambda))^3}{\lambda^2}\right)\right] \\ &= \frac{1}{2}n\frac{[\lambda - F_n(\lambda)]^2}{\lambda(1-\lambda)}1_{\lambda>F_n(\lambda)}[1+o(1)]. \end{aligned}$$

Since $(\lambda - F_n(\lambda))\cdot 1_{\lambda>F_n(\lambda)} \mapsto^P 0$, uniformly over λ, as $n\to\infty$, then the $o(1)$ term is uniformly so in λ. We study the behaviour of $\sup_\lambda[[\lambda - F_n(\lambda)]/\sqrt{\lambda}]$ using Komlos-Major-Tusnady theorem (see Shorack and Wellner (1986)): for any $a\in[0,1]$ and $y \ge 0$, we can find a Brownian bridge $\mathcal{B}_n(\lambda)$ such that:

$$P\left\{\sup_{0\le\lambda\le a}[\sqrt{n}|\sqrt{n}(\lambda - F_n(\lambda))1_{\lambda>F_n(\lambda)} - \mathcal{B}_n(\lambda)|] \ge y + C_1\log(na)\right\} \le \Lambda_1\cdot e^{-\lambda_1 y} \tag{3.11}$$

where C_1, Λ_1 and λ_1 are positive constants. Take $a = 1 - e^{-G}$, $y = \log^2 n$, $\mathcal{L}(n) = \log^2 n + C_1 \log[n(1-e^{-G})]$, $\alpha_n = \Lambda_1 e^{-\lambda_1 \log^2 n}$. We then have: $\lim_{n\to\infty} \mathcal{L}(n)/n^{1/4} = 0$ and $\sum_{n=1}^{\infty} \alpha_n < \infty$. Inequality (3.11) leads to:

$$P\left\{\sup_{0\leq\lambda\leq 1-e^{-G}} \left|\sqrt{n}\,(\lambda - F_n(\lambda))\,1_{\lambda>F_n(\lambda)} - \mathcal{B}_n(\lambda)\right| \geq \frac{\mathcal{L}(n)}{\sqrt{n}}\right\} \leq \alpha_n \tag{3.12}$$

which implies that for any sequence λ_n:

$$P\left\{\sup_{\lambda_n\leq\lambda\leq 1-e^{-G}} \left|\sqrt{n}(\lambda - F_n(\lambda))1_{\lambda>F_n(\lambda)} - \mathcal{B}_n(\lambda)\right| \geq \frac{\mathcal{L}(n)}{\sqrt{n}}\right\} \leq \alpha_n.$$

Define the function:

$$f_n(\lambda) = |\sqrt{n}(\lambda - F_n(\lambda))1_{\lambda>F_n(\lambda)} - \mathcal{B}_n(\lambda)|$$

and the event:

$$A_n = \left\{\sup_{\lambda_n\leq\lambda\leq 1-e^{-G}} \frac{f_n(\lambda)}{\sqrt{\lambda}} \geq \frac{\mathcal{L}(n)}{n^{1/4}}\right\}.$$

For $\lambda \geq \lambda_n$, $(\lambda_n)_n$ satisfying the relation (3.10), we have:

$$\left\{f_n(\lambda) \geq \frac{\mathcal{L}(n)}{\sqrt{n}}\right\} \supseteq \left\{\frac{f_n(\lambda)}{\sqrt{\lambda\sqrt{n}}} \geq \frac{\mathcal{L}(n)}{\sqrt{n}}\right\}.$$

Thus:

$$\alpha_n \geq P\left\{\sup_{\lambda_n\leq\lambda\leq 1-e^{-G}} f_n(\lambda) \geq \frac{\mathcal{L}(n)}{\sqrt{n}}\right\} \geq P(A_n).$$

So:

$$P\left\{\overline{\lim_{n\to\infty}}\left[\sup_{\lambda_n\leq\lambda\leq 1-e^{-G}} \left|\sqrt{n}\frac{\lambda - F_n(\lambda)}{\sqrt{\lambda}}1_{\lambda>F_n(\lambda)} - \frac{\mathcal{B}_n(\lambda)}{\sqrt{\lambda}}\right| \geq \frac{\mathcal{L}(n)}{n^{1/4}}\right]\right\} = 0. \tag{3.13}$$

This implies that:

$$P\left\{\overline{\lim_{n\to\infty}}\left[\sup_{\lambda_n\leq\lambda\leq 2\lambda_n} \left|\sqrt{n}\frac{\lambda - F_n(\lambda)}{\sqrt{\lambda}}1_{\lambda>F_n(\lambda)} - \frac{\mathcal{B}_n(\lambda)}{\sqrt{\lambda}}\right| \geq \frac{\mathcal{L}(n)}{n^{1/4}}\right]\right\} = 0. \tag{3.14}$$

In particular, we take $\lambda_n = \frac{1}{2}n^{-1/4}$. The previous lemma and relation (3.4) yield:

$$\overline{\lim_{n\to\infty}} \sup_{\lambda_n\leq\lambda\leq 2\lambda_n} \frac{\mathcal{B}_n(\lambda)}{\sqrt{2\lambda\log(|\log\lambda|)}} \geq 1 \quad \text{a.s.}$$

Hence, we deduce that:

$$\overline{\lim_{n\to\infty}} \sup_{\lambda_n\leq\lambda\leq 2\lambda_n} \frac{\mathcal{B}_n(\lambda)}{\sqrt{\lambda}} = \infty \quad \text{a.s.} \tag{3.15}$$

Table 1. Divergence of the LRT statistic.

n	N	$\sup l_n(\tau) - l_n(0)$
10^3	10^3	6.45
10^3	10^4	29.2
10^4	10^3	6.69
10^4	10^4	36.5
10^5	$3 \cdot 10^4$	79.5

Let (c_n) be a sequence converging to $+\infty$. Using the triangular inequality, we have

$$P\left\{\varliminf_{n\to\infty}\left[\sup_{\lambda\in[\lambda_n,2\lambda_n]}\sqrt{n}\frac{\lambda-F_n(\lambda)}{\sqrt{\lambda}}1_{\lambda>F_n(\lambda)}\geq c_n\right]\right\}$$
$$\geq P\left\{\varliminf_{n\to\infty}\left[\sup_{\lambda\in[\lambda_n,2\lambda_n]}\frac{\mathcal{B}_n(\lambda)}{\sqrt{\lambda}}\geq c_n\right.\right.$$
$$\left.\left.+\sup_{\lambda_n\leq\lambda\leq 2\lambda_n}\left|\sqrt{n}\frac{\lambda-F_n(\lambda)}{\sqrt{\lambda}}1_{\lambda>F_n(\lambda)}-\frac{\mathcal{B}_n(\lambda)}{\sqrt{\lambda}}\right|\right]\right\}.$$

Relations (3.14) and (3.15) imply:

$$P\left\{\varliminf_{n\to\infty}\left[\sup_{\lambda\in[\lambda_n,2\lambda_n]}\sqrt{n}\frac{\lambda-F_n(\lambda)}{\sqrt{\lambda}}1_{\lambda>F_n(\lambda)}\right]=\infty\right\}=1.$$

Then, the conclusion of theorem holds:

$$P(T_n=0,\lambda<F_n(\lambda))=\frac{1}{2}\qquad P\Big(\lim_{n\to\infty}T_n=\infty,\lambda\geq F_n(\lambda)\Big)=\frac{1}{2}.\qquad\square$$

Remark. The methodology here used is adapted for testing an exponential density against a simple mixture. For testing p against q mixtures, $q < p$, it would find another reparametrization. In our opinion, the LRT statistic will converge with non zero some probability to $+\infty$ (since $\mathcal{D}$ is not relatively compact).

A numerical study confirms that the maximum of the relation (3.9) diverges slowly to $+\infty$ for the half of the case and for the other half it is equal to 0. We take the maximum value of (3.9) on a grid of N values for λ. The results are exhibited in Table 1.

Acknowledgements

The author is deeply indebted to Didier Dacunha-Castelle for his substantial assistance. He also wants to thank Elisabeth Gassiat for her valuable suggestions.

References

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 Bahabur Festschrift, 83–96, Wiley, New York.

Ciuperca, G. (1999). Sur le test de maximum de vraisemblance pour le mélange de populations, *Comptes-Rendus de l'Académie des Sciences, Serie I*, **328**(4), 351–356.

Dacunha-Castelle, D. and Duflo, M. (1990). *Probabilités et Statistiques, Tome* 1, *Problèmes à Temps Fixe*, Masson, Paris.

Dacunha-Castelle, D. and Gassiat, E. (1997). Testing in locally conic model and application to mixture models, *ESAIM Probab. Statist.*, **1**, 285–317.

Everitt, B. S. and Hand, D. J. (1981). *Finite Mixture Distributions*, Chapman and Hall, New York.

Ghosh, J. and Sen, P. (1985). On the asymptotic performance of the log-likelihood ratio statistic for the mixture model and related results, *Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, 789–806, Wadsworth, Belmont, California.

Hanson, D. L. and Russo, R. P (1983). Some results on increments of the Wiener process with applications to lag sums of i.i.d. random variables, *Ann. Probab.*, **11**(3), 609–623.

Hartigan, J. A. (1985). A failure of likelihood ratio asymptotics for normal mixtures, *Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, 807–810, Wadsworth, Belmont, California.

Lindsay, B. G. (1995). Mixture models: Theory, geometry and applications, *NSF-CBMS Regional Conference Series in Probability and Statistics*, **5**, IMS, Hayward, California; ASA, Alexandria, Virginia.

McLachlan, G. J. and Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering*, Marcel Dekker, New York.

Redner, R. (1981). Note on the Consistency of the maximum likelihood estimate for nonidentifiable distributions, *Ann. Statist.*, **9**, 225–228.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components, *J. Roy. Statist. Soc. Ser. B*, **59**(4), 731–792.

Roberts, S. J., Husmeier, D., Rezek, I. and Penny, W. (1998). Bayesian approaches to Gaussian mixture modeling, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(11), 887–906.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statitics*, Wiley, New York.

Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Models*, Wiley, New York.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*, Springer, New York.