

# On the consistency of MLE in finite mixture models of exponential families

N. Atienza<sup>a,\*</sup>, J. García-Heras<sup>b,1</sup>, J.M. Muñoz-Pichardo<sup>b,2</sup>, R. Villa<sup>c,2</sup>

<sup>a</sup>*Depto. de Matemática Aplicada I, University of Sevilla, Spain*

<sup>b</sup>*Depto. de Estadística e I.O., University of Sevilla, Spain*

<sup>c</sup>*Depto. de Análisis Matemático, University of Sevilla, Spain*

## Abstract

Finite mixtures of densities from an exponential family are frequently used in the statistical analysis of data. Modelling by finite mixtures of densities from different exponential families provide more flexibility in the fittings, and get better results. However, in mixture problems, the log-likelihood function very often does not have an upper bound and therefore a global maximum does not always exist. Redner and Walker (1984. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* 26, 195–239) provide conditions to assure the existence, consistency and asymptotic normality of the maximum likelihood estimator.

These conditions are not generally easy to check, even for mixtures of densities from exponential families and, especially, from different exponential families. In this paper, results are given which make verification of the conditions easier in both cases.

*MSC:* 62F12; 62E20

*Keywords:* Consistency; Exponential families; Mixtures

## 1. Introduction

In numerous studies in different fields of applications, the description of a observed distribution ranges between different possible models (see Marazzi et al., 1998). Finite mixture models are very useful in such cases due to their broad flexibility. For this reason they have been much studied both theoretically and in practice; see Everitt and Hand (1981), Titterton et al. (1985), Lindsay (1995), McLachlan and Peel (2001). The components of the mixture are very often homogeneous. Zasada and Cieszewski (2005) compare three homogeneous mixture models, with Normal, Lognormal and Gamma components, to describe tree diameters. In such cases, when the fitting of experiences ranges between different possible models, better results are expected if those heterogeneous distributions are mixed in the model (e.g., see Al-Hussaini and Abd-El-Hakim, 1990; Rachev and SenGupta, 1993; Chandrasekhar and Natarajan, 1996; Grodzenskii and Domrachev, 2002; Andersen et al., 2003).

---

\* Corresponding author.

*E-mail address:* [natienza@us.es](mailto:natienza@us.es) (N. Atienza).

<sup>1</sup> Supported by Ministerio de Educación y Ciencia of Spain, Plan Nacional I+D, Ref. MTM2004-01433.

<sup>2</sup> Supported by DGES Grant #BFM2003-01297.

Finite mixtures of distributions from exponential families, such as Negative Exponential, Gamma, Lognormal, Loggamma..., are very useful in fitting observed data which exhibit skewness. In this paper we study finite mixtures from different exponential families of distributions when the modelling of experiences ranges among different possible models. This behavior is observed in variables such as “length of hospital stay”, a widely used indicator of the hospital activity. [Marazzi et al. \(1998\)](#) assess the adequacy of three widely used models (Lognormal, Weibull and Gamma) for describing the distributions of length of stay. They adjusted the three models to stays grouped by some Diagnosis Related Groups (DRG). The descriptions provided by the Gamma and the Weibull models were very similar and so the Weibull model could be omitted. Often stays from a given DRG could be described with the same model during several years, and then better fittings are expected if mixtures from the appropriate model are considered. However, the same DRG could be described by a different model, depending on the period or on the country. The theoretical tools exposed in this paper can be used in order to consider heterogeneous mixtures of Gamma and Lognormal distributions (which are exponential families). For instance, better fitting can be obtained if a mixture of a Gamma and a Lognormal distribution is considered, instead of considering a mixture of two distribution from the same family (see [Atienza, 2005](#)).

Identifiability is required for the estimation procedures to be well-defined. This problem has been widely investigated (see [Teicher, 1963](#); [Henna, 1994](#) or [Atienza et al., 2005a](#)).

However, there is one problem which gives rise to serious theoretical difficulties: the consistency of maximum likelihood estimators (MLE). ML estimation is very popular, partly since it fits into the philosophy of likelihood-based inference, and partly because it has the advantage that, under certain conditions, the estimates have desirable properties. See [McLachlan and Peel \(2001\)](#) for further discussions on MLE and other estimates.

The consistency problem was studied by several authors. [Chanda \(1954\)](#) generalizes a result by [Cramér \(1946\)](#) and proves, under some regularity conditions stated below (C1 and C2), that there exists a unique solution of the likelihood equations which is consistent and asymptotically normally distributed. Using the same conditions [Peters and Walker \(1978\)](#) show that there is a unique strongly consistent solution of the likelihood equations, which locally maximizes the log-likelihood functions. [Redner \(1981\)](#) gives an extension of a previous work by [Wald \(1949\)](#) on the strong consistency of MLE in the non-identifiable case, using some integrability conditions (C3 and C4 below). A compendium of all these previous results are given in [Redner and Walker \(1984\)](#). If the parameter is identifiable but on the boundary of the parameter space, [Self and Liang \(1987\)](#) prove the consistency under similar conditions as Chanda’s. [Feng and McCulloch \(1992\)](#) state consistency and asymptotic normality in identifiable situations by an unrestricted MLE, solving the problem of the true parameter to be in the boundary of the parameter space. Non-identifiable and boundary problems are studied by [Feng and McCulloch \(1996\)](#). They state the existence of locally consistent solutions which approach to a non-identifiable subset of the parameter space, as long as the density functions in the family can be extended to all the values of the parameter in  $\mathbb{R}^v$ . They do not deal with strong consistency or asymptotic normality. In the same general situation, [Cheng and Liu \(2001\)](#) extend the initial argument by [Wald \(1949\)](#) to give the existence of locally consistent solutions by a compactification technique in the univariate case. In that paper the authors impose some integrability conditions similar to C1, C3 and C4 given below, whereas condition C2 about the information matrix is replaced by a null behavior of the components in the infinity.

Consistency problems have been studied in many particular cases: see [Jewell \(1982\)](#), [Hathaway \(1985\)](#), [Pfanzagl \(1988\)](#), [Leroux \(1992\)](#), [Van De Geer \(2003\)](#) and [Ciuperca et al. \(2003\)](#).

This paper provides an easy way to verify consistency conditions C1, C3 and C4 for mixtures of different exponential type families. Thus, the study on the consistency and asymptotic normal behavior of the MLE is reduced in these cases to the check of some simple integrability conditions on the components, together with condition C2, which has to be considered separately.

The paper is organized as follows. Section 2 contains the two aforementioned results of Redner and Walker together with the notation and the conditions as they were given in [Redner and Walker \(1984\)](#). We replace two of these conditions by other sufficient ones which are easier to check.

In Section 3, the widely used exponential families are introduced (see [Redner and Walker, 1984](#); [Titterington et al., 1990](#) or [McLachlan and Krishnan, 1997](#)), and it is shown that a finite mixture of components from the same exponential family always verifies the first consistency condition. Sufficient conditions are given for the third and fourth consistency conditions to be verified. The second condition is not dealt with since a particular study for each family may be required. As a generalization, in Section 4 we extend the previous results to mixtures whose components

belong to different exponential families, and apply this theoretic study to mixtures from Gamma and Lognormal distributions.

The Appendix contains technical results used in the proofs of the theorems.

## 2. Consistency conditions and theorems

Let  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$  be a sample of  $N$  observations on independent, identically distributed  $n$ -dimensional random variables with density function given by  $q(\mathbf{x}; \Psi) = \sum_{j=1}^k \pi_j f_j(\mathbf{x}; \theta_j)$ , where

$$\pi_k = 1 - \sum_{j=1}^{k-1} \pi_j, \quad \pi = (\pi_1, \dots, \pi_{k-1}) \in \mathcal{C}_k = \left\{ (\pi_1, \dots, \pi_{k-1}) : \sum_{j=1}^{k-1} \pi_j \leq 1, \pi_j \geq 0 \right\},$$

and each component of the mixture  $f_j$  is differentiable with respect to its parameters  $\theta_j \in \Theta_j \subset \mathbb{R}^{d_j}$ . Let  $\Omega$  denote the parametric space  $\{(\pi, \theta_1, \dots, \theta_k) : \pi \in \mathcal{C}_k, \theta_j \in \Theta_j \text{ for } j = 1, \dots, k\}$ . From now on, the elements of  $\Omega$  will be denoted for convenience by  $\Psi = (\psi_1, \dots, \psi_v)$  where  $v$  is the dimension of  $\Omega$ . For  $\Psi \in \Omega$  and sufficiently small  $r > 0$ , let  $N_r(\Psi)$  denote the closed ball of radius  $r$  about  $\Psi$  in  $\Omega$  and we define

$$q_r(\mathbf{x}; \Psi) = \sup_{\Psi' \in N_r(\Psi)} q(\mathbf{x}; \Psi') \quad \text{and} \quad q_r^*(\mathbf{x}; \Psi) = \max\{1, q_r(\mathbf{x}; \Psi)\}. \quad (1)$$

The consistency conditions collected by Redner and Walker (1984) are now presented and denoted as C1, C2, C3 and C4. Furthermore,  $\Psi^* \in \Omega$  represents the ‘‘true’’ parameter value to be estimated, which is supposed to be identifiable.

C1. For all  $\Psi \in \Omega$ , for almost all  $\mathbf{x} \in \mathbb{R}^n$  and for  $i, j, s = 1, \dots, v$ , the partial derivatives  $\partial q / \partial \psi_j$ ,  $\partial^2 q / \partial \psi_j \partial \psi_i$  and  $\partial^3 q / \partial \psi_j \partial \psi_i \partial \psi_s$  exist and satisfy

$$\left| \frac{\partial q(\mathbf{x}; \Psi)}{\partial \psi_j} \right| \leq h_j(\mathbf{x}), \quad \left| \frac{\partial^2 q(\mathbf{x}; \Psi)}{\partial \psi_j \partial \psi_i} \right| \leq h_{ij}(\mathbf{x}), \quad \left| \frac{\partial^3 \log q(\mathbf{x}; \Psi)}{\partial \psi_j \partial \psi_i \partial \psi_s} \right| \leq h_{ijs}(\mathbf{x}),$$

where  $h_j$  and  $h_{ij}$  are integrable and  $\int h_{ijs}(\mathbf{x}) q(\mathbf{x}; \Psi^*) d\mathbf{x} < +\infty$ .

C2. The Fisher information matrix  $I(\Psi)$  given by

$$I(\Psi) = \int [\nabla_{\Psi} \log q(\mathbf{x}; \Psi)] [\nabla_{\Psi} \log q(\mathbf{x}; \Psi)]^T q(\mathbf{x}; \Psi) d\mathbf{x},$$

is well-defined and positive definite at  $\Psi^*$ , where  $\nabla_{\Psi}$  denotes the gradient of the first partial derivatives with respect to the components of  $\Psi$ .

C3. For each  $\Psi \in \Omega$  and sufficiently small  $r > 0$ ,

$$\int \log q_r^*(\mathbf{x}; \Psi) q(\mathbf{x}; \Psi^*) d\mathbf{x} < +\infty.$$

C4.  $\int |\log q(\mathbf{x}; \Psi^*)| q(\mathbf{x}; \Psi^*) d\mathbf{x} < +\infty$ .

Under the first two conditions, Redner and Walker establish that for each sufficiently small neighborhood of  $\Psi^*$  in  $\Omega$  and large  $N$  there is, with probability 1, a unique solution  $\Psi^N$  of the likelihood equations, which locally maximizes the log-likelihood function. Moreover,  $N^{1/2}(\Psi^N - \Psi^*)$  is asymptotically normally distributed with mean zero and covariance matrix  $I(\Psi^*)^{-1}$ .

On the other hand, if C3 and C4 are satisfied,  $\Omega'$  is any compact subset of  $\Omega$  containing  $\Psi^*$ ,  $E = \{\Psi \in \Omega' : q(\cdot; \Psi) = q(\cdot; \Psi^*) \text{ almost everywhere}\}$  and  $D$  is any closed subset of  $\Omega' \setminus E$ ,

$$\lim_{N \rightarrow +\infty} \sup_{\Psi \in D} \frac{\prod_{i=1}^N q(\mathbf{x}_i; \Psi)}{\prod_{i=1}^N q(\mathbf{x}_i; \Psi^*)} = 0$$

with probability 1. Roughly speaking, if  $\Omega'$  is any compact subset of  $\Omega$  which contains  $\Psi^*$ , then with probability 1,  $\Psi^N$  is a maximum likelihood estimate in  $\Omega'$  for sufficiently large  $N$ . Furthermore, this maximum likelihood estimate is unique in  $\Omega'$ , up to a label switching.

Taking into account the difficulty in checking consistency conditions C3 and C4 since they are related to the mixture density  $q(\mathbf{x}; \Psi)$ , we propose two new analogous conditions C5 and C6, related to the components of the mixture, which initially simplify the verification process.

C5. For each  $(\theta_1, \dots, \theta_k) \in \Theta = \Theta_1 \times \dots \times \Theta_k$  and sufficiently small  $r > 0$ ,

$$\int (\log f_{ir}^*(\mathbf{x}; \theta_i)) f_j(\mathbf{x}; \theta_j^*) \, d\mathbf{x} < +\infty$$

for any  $i, j = 1, \dots, k$ , where  $f_{ir}^*$  is defined in the same way as  $q_r$  in (1).

C6. For  $i, j = 1, \dots, k$ ,  $\int |\log f_i(\mathbf{x}; \theta_i^*)| f_j(\mathbf{x}; \theta_j^*) \, d\mathbf{x} < +\infty$ .

**Theorem 1.** *If C5(C6) is verified, then C3(C4) is also verified.*

**Proof.** By application of Lemma 1 (see Appendix),

$$\begin{aligned} \log q_r^*(\mathbf{x}; \Psi) &\leq \sup_{\Psi' \in N_r(\Psi)} \log \sum_{i=1}^k \pi'_i \max\{1, f_i(\mathbf{x}; \theta'_i)\} \\ &\leq \sup_{(\theta'_1, \dots, \theta'_k) \in N_r(\theta'_1, \dots, \theta'_k)} \sum_{i=1}^k \log \max\{1, f_i(\mathbf{x}; \theta'_i)\} \\ &= \sum_{i=1}^k \log f_{ir}^*(\mathbf{x}; \theta_i). \end{aligned}$$

Thus, C5 follows from

$$\int (\log q_r^*(\mathbf{x}; \Psi)) q(\mathbf{x}; \Psi^*) \, d\mathbf{x} \leq \sum_{i,j=1}^k \int \pi_j^* (\log f_{ir}^*(\mathbf{x}; \theta_i)) f_j(\mathbf{x}; \theta_j^*) \, d\mathbf{x}.$$

On the other hand, applying Lemma 1 (see Appendix),

$$\int |\log q(\mathbf{x}; \Psi^*)| q(\mathbf{x}; \Psi^*) \, d\mathbf{x} \leq \int \sum_{i,j=1}^k |\log f_i(\mathbf{x}; \theta_i^*)| \pi_j^* f_j(\mathbf{x}; \theta_j^*) \, d\mathbf{x},$$

which proves the theorem for C6.  $\square$

### 3. Consistency conditions for mixtures from exponential families

Following the definition given in Barndorff-Nielsen (1978), a parametric family of densities  $\mathcal{E}(a, \mathbf{t}, b)$  is said to be an exponential family associated to the functions  $a$ ,  $\mathbf{t}$  and  $b$  if its members are written as:

$$f(\mathbf{x}; \theta) = a(\mathbf{x})b(\theta)e^{\mathbf{t}(\mathbf{x})\theta^T} \quad \mathbf{x} \in D \subset \mathbb{R}^n, \quad \theta \in \Theta \subset \mathbb{R}^d,$$

where  $\Theta$  is open,  $D$  measurable,  $a : D \subset \mathbb{R}^n \rightarrow [0, +\infty)$ ,  $b : \Theta \rightarrow (0, +\infty)$  and  $\mathbf{t} : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^d$ .

As it will be shown later, mixtures from many exponential parametric families used in practice verify consistency conditions. However, no general result has been found in the literature in this direction. In fact, the example below shows an exponential family for which not all of the conditions hold. Although the example is not excessively important from a practical point of view, we present it here for the sake of completeness.

Let  $\mathcal{F}$  be the exponential family given by

$$\mathcal{F} = \left\{ f(x; \theta) = \frac{\theta - 1}{x(\log \frac{1}{x})^\theta} : x \in \left(0, \frac{1}{e}\right], \theta \in (1, +\infty) \right\}.$$

The mixture  $q(x; \Psi^*) = \frac{1}{2} f(x; \frac{3}{2}) + \frac{1}{2} f(x; 4)$  does not verify condition C3.

It is therefore natural to ask when finite mixtures from an exponential family verify the consistency conditions. We prove that condition C1 is always verified by this kind of mixtures, and present sufficient conditions under which C3 and C4 hold. The second condition C2 is not considered here due to an apparent need for individual study for each specific family. Hence, the verification of C2 is needed in each particular problem for proving consistency.

The following definition will be needed throughout the paper.

**Definition 1.** Given  $t : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^d$  with  $t(\mathbf{x}) = (t_1(\mathbf{x}) \dots t_d(\mathbf{x}))$  and  $h$  a non-negative integer, the family of functions  $\Pi_h(t)$  given by

$$\Pi_h(t) = \left\{ T : D \subset \mathbb{R}^n \rightarrow \mathbb{R} : T(\mathbf{x}) = \prod_{i=1}^d t_i^{v_i}(\mathbf{x}), v_i \in \mathbb{Z}^+, \sum_{i=1}^d v_i \leq h \right\}$$

is called the family of product functions of at most  $h$  components of  $t$ .

Here and subsequently,  $V_d = \{\varepsilon \in \mathbb{R}^d / \varepsilon_i = \pm 1; i = 1, \dots, d\}$ . Furthermore,  $Q_d(\theta^*, r)$  denotes the hypercube  $\prod_{i=1}^d [\theta_i^* - r, \theta_i^* + r]$  and for each  $\varepsilon \in V_d$ ,  $\theta(\varepsilon) \in \mathbb{R}^d$  (whose coordinates are  $\theta_i^* + \varepsilon_i r$ ) denotes its vertices.

**Theorem 2.** Let  $\mathcal{E}(a, t, b)$  be an exponential family and  $\Psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C}_k \times \Theta^k$ . Therefore any finite mixture defined from  $k$  components of  $\mathcal{E}(a, t, b)$  verifies consistency condition C1 on a compact set  $\mathcal{C}'_k \times \Theta'_1 \times \dots \times \Theta'_k \subset \mathcal{C}_k \times \Theta^k$  containing  $\Psi^*$ .

**Proof.** Let  $p(\mathbf{x}; \Psi) = \sum_{i=1}^k \pi_i f_i(\mathbf{x}; \theta_i)$  be a finite mixture defined from  $k$  components of  $\mathcal{E}(a, t, b)$ .

Since  $\Theta$  is open, there exists an  $r > 0$  such that  $Q_d(\theta_i^*, r) \subset \Theta$  for every  $i = 1, \dots, k$ . Hence, choosing any compact set  $\mathcal{C}'_k \subset \mathcal{C}_k$  which contains  $\pi^*$ , we see that the set  $\mathcal{C}'_k \times Q_d(\theta_1^*, r) \times \dots \times Q_d(\theta_k^*, r)$  is compact and it contains  $\Psi^*$ .

For the conditions of the derivatives up to the second order, using linearity of the derivative, triangular inequality and symmetry with respect to the index  $i = 1, \dots, k$ , it suffices to prove that for every  $\theta \in Q_d(\theta^*, r)$  (for simplicity, we omit the indices of the densities),

$$|f(\mathbf{x}; \theta)| \leq g(\mathbf{x}), \quad \left| \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta_j} \right| \leq g_j(\mathbf{x}), \quad \left| \frac{\partial^2 f(\mathbf{x}; \theta)}{\partial \theta_j \partial \theta_s} \right| \leq g_{js}(\mathbf{x}) \quad (2)$$

with  $g, g_j, g_{js}$  integrable functions with respect to  $\mathbf{x}$  and independent from  $\theta$ .

From Theorem 7 (see Appendix) there exists an  $M > 0$  such that

$$|b(\theta)|, \quad \left| \frac{\partial b(\theta)}{\partial \theta_i} \right|, \quad \left| \frac{\partial^2 b(\theta)}{\partial \theta_i \partial \theta_j} \right|, \quad \left| \frac{\partial^3 b(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_u} \right| \leq M$$

for each  $\theta \in \bigcup_{i=1}^k Q_d(\theta_i^*, r)$ .

For the first inequality in (2), from Lemma 2 (see Appendix), for each  $\theta \in \bigcup_{i=1}^k Q_d(\theta_i^*, r)$  we have

$$|f(\mathbf{x}; \theta)| \leq M \sum_{i=1}^k \sum_{\varepsilon \in V_d} a(\mathbf{x}) e^{t(\mathbf{x})(\theta_i(\varepsilon))^T} = g(\mathbf{x}),$$

where  $\theta_i(\varepsilon) = \theta_i^* + r\varepsilon$  and  $g(\mathbf{x})$  is integrable and independent from  $\theta$ .

For the second inequality in (2), differentiating  $f$  gives

$$\begin{aligned} \left| \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta_j} \right| &\leq \left| a(\mathbf{x}) \frac{\partial b(\theta)}{\partial \theta_j} e^{t(\mathbf{x})\theta^T} \right| + |a(\mathbf{x}) b(\theta) t_j(\mathbf{x}) e^{t(\mathbf{x})\theta^T}| \\ &\leq M |a(\mathbf{x}) e^{t(\mathbf{x})\theta^T}| + M |a(\mathbf{x}) t_j(\mathbf{x}) e^{t(\mathbf{x})\theta^T}| \end{aligned}$$

for each  $\theta \in \bigcup_{i=1}^k Q_d(\theta_i^*, r)$  and Lemma 2 (see Appendix) yields

$$\leq M \sum_{i=1}^k \left[ \sum_{\varepsilon \in V_d} |a(\mathbf{x}) e^{\mathbf{t}(\mathbf{x})(\theta_i(\varepsilon))^T}| + \sum_{\varepsilon \in V_d} |a(\mathbf{x}) t_j(\mathbf{x}) e^{\mathbf{t}(\mathbf{x})(\theta_i(\varepsilon))^T}| \right] = g_j(\mathbf{x})$$

where  $g_j(\mathbf{x})$  is integrable from Theorem 6 (see Appendix) and independent from  $\theta$ . The proof for the last inequality in (2) is similar.

It remains to bound the third derivatives of the logarithm. We have

$$\begin{aligned} \frac{\partial^3 \log p}{\partial \psi_i \partial \psi_j \partial \psi_u} &= \frac{1}{p} \frac{\partial^3 p}{\partial \psi_i \partial \psi_j \partial \psi_u} - \frac{1}{p^2} \left[ \frac{\partial^2 p}{\partial \psi_i \partial \psi_j} \frac{\partial p}{\partial \psi_u} + \frac{\partial^2 p}{\partial \psi_j \partial \psi_u} \frac{\partial p}{\partial \psi_i} + \frac{\partial^2 p}{\partial \psi_i \partial \psi_u} \frac{\partial p}{\partial \psi_j} \right] \\ &\quad + \frac{2}{p^3} \frac{\partial p}{\partial \psi_i} \frac{\partial p}{\partial \psi_j} \frac{\partial p}{\partial \psi_u}. \end{aligned} \quad (3)$$

The absolute values of the quotients of the partial derivatives of  $p$  up to the third order and  $p$  are bounded by a sum of quotients of the form

$$\left| \frac{1}{p} \frac{\partial^3 f}{\partial \theta_i \partial \theta_j \partial \theta_u} \right|, \quad \left| \frac{1}{p} \frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \right|, \quad \left| \frac{1}{p} \frac{\partial f}{\partial \theta_i} \right| \quad \text{and} \quad \frac{f}{p},$$

which can be expressed as a sum of functions of the form  $|(f/p)T(\mathbf{x})B(\theta)|$  where  $T \in \Pi_h(\mathbf{t})$  and  $B(\theta)$  is a continuous function of the parameter  $\theta$ , independent from  $\mathbf{x}$  and bounded on  $\bigcup_{i=1}^k Q_d(\theta_i^*, r)$ . Thus, we can bound the sum of these expressions by a sum of functions of the form  $C|T(\mathbf{x})|$  with  $T \in \Pi_h(\mathbf{t})$  and  $C$  a positive constant.

Hence, expression (3) is bounded by a sum of functions of the form  $C|T(\mathbf{x})|$ ,  $C^2|T(\mathbf{x})|$  and  $C^3|T(\mathbf{x})|$ , which are independent from the parameter  $\Psi$ . We consider  $h_{iju}$  to be the previous sum. From Theorem 6 (see Appendix),  $T(\mathbf{x})p(\mathbf{x}; \Psi^*)$  is integrable and therefore  $h_{iju}(\mathbf{x})p(\mathbf{x}; \Psi^*)$  is also integrable.  $\square$

We now propose a sufficient condition under which any mixture of functions from an exponential family verifies consistency conditions C3 and C4.

**Theorem 3.** *Let  $\mathcal{E}(a, \mathbf{t}, b)$  be an exponential family and  $\Psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C}_k \times \Theta^k$ . If  $\int |\log a(\mathbf{x})| f(\mathbf{x}; \theta_j^*) d\mathbf{x} < +\infty$  for  $j = 1, \dots, k$ , then any finite mixture of  $k$  components of  $\mathcal{E}(a, \mathbf{t}, b)$  verify consistency conditions C3 and C4 on any compact set  $\Omega' \subset \Omega$  such that  $\Psi^* \in \Omega'$ .*

**Proof.** We may assume that  $\Omega'$  has the form  $\mathcal{C}'_k \times \Theta'_1 \times \dots \times \Theta'_k$ .

By Theorem 1, C3 will be proved once we prove C5 on  $\Theta' = \Theta'_1 \times \dots \times \Theta'_k$ . Let  $r > 0$  be sufficiently small so that  $\bar{N}_r(\theta, r) \subset \Theta_i$  for every  $\theta \in \Theta'_i$ , for  $i = 1, \dots, k$  and let  $\Theta''_i = \bigcup_{\theta \in \Theta'_i} \bar{N}_r(\theta)$ . Hence, for every  $\theta \in \Theta''_i$ ,  $\mathbf{t}(\mathbf{x})\theta^T \leq C \sum_{s=1}^d |t_s(\mathbf{x})|$ , with  $C$  an upper bound of the absolute values of the components of  $\theta$  on  $\Theta''_i$ .

On the other hand, using Theorem 7 (see Appendix), let  $M$  be an upper bound for  $|b|$  on  $\Theta''_i$ . Hence, for any  $\theta' \in \Theta'_i$ ,  $f_r^*(\mathbf{x}; \theta') \leq \max\{1, M a(\mathbf{x}) e^{C \sum_{s=1}^d |t_s(\mathbf{x})|}\}$  and therefore, for every  $\theta' \in \Theta'_i$ ,  $\log f_r^*(\mathbf{x}; \theta') \leq \log M + |\log a(\mathbf{x})| + C \sum_{s=1}^d |t_s(\mathbf{x})|$ .

From Theorem 6 (see Appendix),  $t_s(\mathbf{x}) f(\mathbf{x}; \theta_j^*)$  is integrable, for  $j = 1, \dots, k$ , and by assumption  $\log a(\mathbf{x}) f(\mathbf{x}; \theta_j^*)$  is also integrable for  $j = 1, \dots, k$ . Hence,  $\log f_r^*(\mathbf{x}; \theta_j^*) f(\mathbf{x}; \theta_j^*)$  is integrable as desired. By Theorem 1, C4 will be proved once we prove C6, which follows from the inequality

$$|\log f(\mathbf{x}; \theta_j^*)| f(\mathbf{x}; \theta_j^*) \leq \left( |\log a(\mathbf{x})| + |\log b(\theta_j^*)| + \sum_{s=1}^d |\theta_j^{*s} t_s(\mathbf{x})| \right) f(\mathbf{x}; \theta_j^*). \quad \square$$

#### 4. Consistency conditions for mixtures from the union of exponential families

As we have seen in the introduction, mixtures from different exponential families can be used to improve the fitting of data from scientific experiences. So far, we have results which assure the verification of consistency conditions for mixtures generated by functions of the same exponential family. In this section, we propose results to assure the verification in the case in which functions from different exponential families are used and we illustrate these results with a finite mixture of functions from the Lognormal and Gamma families. These mixtures are very useful to fit observed data from the variable “length of hospital stay”, as it was mentioned in the introduction.

To start with, the family of product functions of components of more than one vector function is presented below, in the same way as Definition 1.

**Definition 2.** Given  $k$  functions  $t_i : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^{d_i}$ , for  $i = 1, \dots, k$ , with  $t_i(\mathbf{x}) = (t_{i1}(\mathbf{x}), \dots, t_{id_i}(\mathbf{x}))$  and  $h$  a non-negative integer, the family of functions  $\Pi_h(t_1, \dots, t_k)$  given by

$$\left\{ T : D \subset \mathbb{R}^n \rightarrow \mathbb{R} : T(\mathbf{x}) = \prod_{i=1}^k \prod_{j=1}^{d_i} t_{ij}^{v_{ij}}(\mathbf{x}), v_{ij} \in \mathbb{Z}^+ \cup \{0\}, \sum_{i=1}^k \sum_{j=1}^{d_i} v_{ij} \leq h \right\}$$

is called the family of product functions of at most  $h$  components of  $t_1, \dots, t_k$ .

**Theorem 4.** Let  $\mathcal{E}_i = \mathcal{E}(a_i, t_i, b_i)$  be  $k$  exponential families defined on  $\Theta_i \subset \mathbb{R}^{d_i}$ . Let  $\Psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C}_k \times \Theta_1 \times \dots \times \Theta_k$ . If

$$(H1) \int |T(\mathbf{x})| f_j(\mathbf{x}; \theta_j^*) d\mathbf{x} < +\infty, \quad T \in \Pi_3(t_i, t_l, t_s) \text{ for } j, i, l, s \in \{1, \dots, k\},$$

then any finite mixture of the form  $p(\mathbf{x}; \Psi) = \sum_{i=1}^k \pi_i f_i(\mathbf{x}; \theta_i)$  where  $f_i(\mathbf{x}; \theta_i) \in \mathcal{E}(a_i, t_i, b_i)$  verifies the consistency condition C1 on some compact set  $\Omega' \subset \Omega$  which contains  $\Psi^*$ .

**Proof.** Obviously, the inequalities from condition C1 given in (2) follow from the proof of Theorem 2.

The same reasoning given in that proof is also valid in order to bound the third derivatives. Therefore  $|\partial^3 \log p / \partial \psi_u \partial \psi_v \partial \psi_w|$  is bounded by a sum of expressions of the form  $C|T(\mathbf{x})|$  with  $T \in \Pi_3(t_i)$ ,  $C^2|T(\mathbf{x})|$  with  $T \in \Pi_3(t_i, t_l)$  and  $C^3|T(\mathbf{x})|$  with  $T \in \Pi_3(t_i, t_l, t_s)$  where  $C$  is a positive constant. We consider  $h_{uvw}(\mathbf{x})$  as the previous sum. By (H1),  $\int h_{uvw}(\mathbf{x}) f_j(\mathbf{x}; \theta_j^*) d\mathbf{x} < +\infty$  for  $1 \leq j \leq k$ . Multiplication by  $\pi_j^*$  and addition in  $j$  show that  $h_{uvw}(\mathbf{x}) p(\mathbf{x}; \Psi^*)$  is integrable.  $\square$

**Remark 1.** For simplicity of notation, in this theorem each component of the mixture belongs to a concrete family. It is obvious that these families do not necessarily have to be different. So hypothesis (H1) implies condition of consistency C1 in the finite mixtures of the union of exponential families.

**Remark 2.** Regarding to Theorem 2, a new linking hypothesis (H1) between the families is introduced.

**Theorem 5.** Let  $\mathcal{E}_i = \mathcal{E}(a_i, t_i, b_i)$  be  $k$  exponential families defined on  $\Theta_i \subset \mathbb{R}^{d_i}$  and let  $\Psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C}_k \times \Theta_1 \times \dots \times \Theta_k$ . If for each  $1 \leq i, j \leq k$

$$(H2) \int |\log a_i(\mathbf{x})| f_j(\mathbf{x}; \theta_j^*) d\mathbf{x} < +\infty \text{ and}$$

$$(H3) \int |t_{is}(\mathbf{x})| f_j(\mathbf{x}; \theta_j^*) d\mathbf{x} < +\infty, \text{ where } t_{is}(\mathbf{x}) \text{ is any component of } t_i,$$

then any finite mixture  $p(\mathbf{x}; \Psi) = \sum_{i=1}^k \pi_i f_i(\mathbf{x}; \theta_i)$  with  $f_i(\mathbf{x}; \theta_i) \in \mathcal{E}(a_i, t_i, b_i)$  verifies consistency conditions C3 and C4 on any compact set  $\Omega' \subset \Omega$  which contains  $\Psi^*$ .



**Proof.** The basis of the proof is similar to that in Theorem 3. Let  $r > 0$  be sufficiently small such that  $\bar{N}_r(\theta, r) \subset \Theta_i$  for every  $\theta \in \Theta'_i$ , for  $i = 1, \dots, k$  and let  $\Theta''_i = \bigcup_{\theta \in \Theta'_i} \bar{N}_r(\theta)$ . Hence for every  $\theta_i \in \Theta''_i$ ,  $\theta_i^T \mathbf{t}_i(\mathbf{x}) \leq C_i \sum_{s=1}^{d_i} |t_{is}(\mathbf{x})|$  where  $C_i$  bounds the absolute values of the components of  $\theta_i$  on  $\Theta''_i$ .

On the other hand, using Theorem 7 (see Appendix), let  $M > 1$  be such that  $|b_i(\theta_i)| \leq M$  for  $\theta_i \in \Theta''_i$ . Hence,  $f_{ir}^*(\mathbf{x}; \theta'_i) \leq \max\{1, M a_i(\mathbf{x}) e^{C_i \sum_{s=1}^{d_i} |t_{is}(\mathbf{x})|}\}$  for any  $\theta'_i \in \Theta'_i$ , and therefore  $\log f_{ir}^*(\mathbf{x}; \theta'_i) \leq \log M + |\log a_i(\mathbf{x})| + C_i \sum_{s=1}^{d_i} |t_{is}(\mathbf{x})|$ .

By hypothesis the functions  $t_{is}(\mathbf{x}) f_j(\mathbf{x}; \theta_j^*)$  and  $\log a_i(\mathbf{x}) f_j(\mathbf{x}; \theta_j^*)$  are integrable and therefore  $\log f_{ir}^*(\mathbf{x}; \theta_i) f_j(\mathbf{x}; \theta_j^*)$  is integrable, which completes the proof for C3.

From Theorem 1, in order to prove C4, it is enough to examine C6, which follows from

$$\begin{aligned} & |\log f_i(\mathbf{x}; \theta_i^*)| f_j(\mathbf{x}; \theta_j^*) \\ & \leq \left( |\log a_i(\mathbf{x})| + |\log b_i(\theta_i^*)| + \sum_{s=1}^{d_i} |\theta_{is}^* t_{is}(\mathbf{x})| \right) f_j(\mathbf{x}; \theta_j^*). \quad \square \end{aligned}$$

**Remark 3.** Following a similar reasoning as in Remark 1, hypothesis (H2) and (H3) imply consistency conditions C3 and C4 in finite mixtures of the union of exponential families.

**Remark 4.** Regarding to Theorem 3, two new linking hypotheses between the families are included: (H2) and (H3).

The following example illustrate the usefulness of these theorems.

**Example.** Let  $\mathcal{P}$  denote the family of finite mixtures generated from the union of two exponential families, Gamma and Lognormal, given respectively by:

$$\mathcal{G} = \left\{ f(x; u, v) = \frac{v^u}{\Gamma(u)} x^{u-1} e^{-vx}; u > 0, v > 0, x > 0 \right\},$$

where  $a_1(x) = 1/x$ ,  $\mathbf{t}_1(x) = (\log x, -x)$ ,  $b_1(u, v) = v^u/\Gamma(u)$ , and by

$$\mathcal{L} = \left\{ f(x; \mu, \sigma) = \frac{1}{(2\pi)^{1/2} \sigma x} \exp \left\{ -\frac{(\log x - \mu)^2}{2\sigma^2} \right\} : \mu \in \mathbb{R}, \sigma > 0, x > 0 \right\}$$

where  $a_2(x) = 1/x$ ,  $\mathbf{t}_2(x) = (\log x, -\log^2 x)$ ,  $b_2(\alpha_1, \alpha_2) = \pi^{-1/2} \alpha_2^{1/2} e^{-\alpha_1^2/(4\alpha_2)}$  for  $(\alpha_1, \alpha_2) = (\mu/\sigma^2, 1/(2\sigma^2)) \in \mathbb{R} \times (0, +\infty)$ .

Atienza et al. (2005a) proved the identifiability of  $\mathcal{P}$ . Condition C2 can be easily checked in this case. In order to prove C1, C3 and C4, taking into account what the components of  $\mathbf{t}_i(x)$  can be and since  $|\log a_i(x)| = |\log x|$  for  $i = 1, 2$ , we have only to prove that  $x^s (\log x)^h f_i(x; \theta_i)$  is integrable for  $s, h \in \mathbb{Z}^+ \cup \{0\}$ , where  $f_i(x; \theta_i)$  is a density function from any of the previous families.

If  $s = 0$ , these functions can be obtained from the product of components of  $\mathbf{t}_i(x)$  of any of the families. From Theorem 6 (see Appendix), these are integrable functions. If  $s \neq 0$ , the factor  $x^s$  can be written as  $e^{s \log x}$  and hence, for the Lognormal family and the Gamma family, resp.,

$$\begin{aligned} (\log x)^h e^{s \log x} f(x; \alpha_1, \alpha_2) &= (\log x)^h f(x; \alpha_1 + s, \alpha_2) \\ (\log x)^h e^{s \log x} f(x; u, v) &= (\log x)^h f(x; u + s, v) \end{aligned}$$

which from Theorem 6 (see Appendix) are both integrable.

## 5. Conclusions and open questions

This paper provides an easy way to verify consistency conditions C1, C3 and C4 for mixtures of different exponential type families. This reduces the study on the consistency and asymptotic normal behavior of the MLE for these classes of



mixture of distributions to the check of some simple integrability conditions on the functions which define the families involved in the mixture, besides condition C2 related to the information matrix. This previous condition seems hard to be studied for mixtures from general exponential type families. For the sake of solving the general setting, it could be interesting to give conditions under which C2 holds. Nevertheless, it is not difficult to check for mixtures from particular families.

## Acknowledgements

The authors would like to thank the Associate Editor and the referees for helpful suggestions and constructive comments.

## Appendix

The proofs of the following results are technical and can be found in [Atienza et al. \(2005b\)](#).

**Lemma 1.** For any  $a_1, \dots, a_k \geq 0$  and  $\pi_1, \dots, \pi_k \in [0, 1]$  such that  $\sum_{j=1}^k \pi_j = 1$ ,

$$\left| \log \sum_{j=1}^k \pi_j a_j \right| \leq \sum_{j=1}^k |\log a_j|.$$

**Lemma 2.** Let  $t : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^d$ ,  $\theta^* \in \mathbb{R}^d$  and  $r > 0$ . Then

$$e^{t(x)\theta^T} \leq \sum_{\varepsilon \in V_d} e^{t(x)\theta(\varepsilon)^T}$$

for each  $x \in D$  and  $\theta \in Q_d(\theta^*, r)$ , where  $\theta(\varepsilon) = \theta_i^* + \varepsilon_{ir}$  for  $\varepsilon \in V_d$ .

**Theorem 6.** Let  $\Theta \subset \mathbb{R}^d$  be open and  $D \subset \mathbb{R}^n$  measurable. Let  $f : D \times \Theta \rightarrow \mathbb{R}$  be given by  $f(x; \theta) = a(x)e^{t(x)\theta^T}$  where  $a : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  and  $t : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^d$  are two measurable functions such that for each  $\theta \in \Theta$  the function  $x \in D \rightarrow f(x; \theta)$  is integrable on  $D$ . Then the function  $T(x)f(x; \theta)$  is integrable on  $D$ , for every  $\theta \in \Theta$ , with  $h \in \mathbb{Z}^+ \cup \{0\}$  and  $T \in \Pi_h(t)$ .

**Theorem 7.** Let  $\mathcal{E}(a, t, b)$  be an exponential family defined on  $\Theta \subset \mathbb{R}^d$ . The corresponding function  $b$  is continuous, and has continuous partial derivatives of all orders on  $\Theta$ .

## References

- Al-Hussaini, E.K., Abd-El-Hakim, N.S., 1990. Estimation of parameters of the inverse Gaussian–Weibull mixture model. *Comm. Statist. Theory Methods* 19 (5), 1607–1622.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71 (2), 579–625.
- Atienza, N., 2005. Fitting the variable “Length of Hospital Stay” with mixtures from different distribution families (Available at <http://www.personal.us.es/natienza/inves/docu/fitting.pdf>).
- Atienza, N., García-Heras, J., Muñoz-Pichardo, J.M., 2005a. A new condition for identifiability of finite mixture distributions, *Metrika*, to appear, doi:10.1007/s00184-005-0013-z.
- Atienza, N., García-Heras, J., Muñoz-Pichardo, J.M., Villa, R., 2005b. Technical Report: Some Integrability Results on Exponential Families (Available at <http://www.personal.us.es/natienza/inves/docu/technical1.pdf>).
- Barndorff-Nielsen, O., 1978. *Information and Exponential Families in Statistical Theory*. Wiley, New York.
- Chanda, S., 1954. A note on the consistency and maxima of the roots of the likelihood equations. *Biometrika* 41, 56–61.
- Chandrasekhar, S., Natarajan, R., 1996. Confidence limits for steady-state availability of systems with a mixture of Exponential and Gamma operating time and Lognormal repair time. *Microelectron. Reliab.* 36 (9), 1303–1304.
- Cheng, R.C.H., Liu, W.B., 2001. The consistency of estimators in finite mixture models. *Scand. J. Statist.* 28, 603–616.
- Ciuperca, G., Ridolfi, A., Idier, J., 2003. Penalized maximum likelihood estimator for normal mixture. *Scand. J. Statist.* 30, 45–59.

- Cramér, S., 1946. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, NY.
- Everitt, B.S., Hand, D.J., 1981. *Finite Mixture Distributions*. Chapman & Hall, London.
- Feng, Z.D., McCulloch, C.E., 1992. Statistical inference using maximum likelihood estimation and the generalized likelihood ratio when the true parameter is on the boundary of the parameter space. *Statist. Probab. Lett.* 13 (4), 325–332.
- Feng, Z.D., McCulloch, C.E., 1996. Using Bootstrap likelihood ratios in finite mixture models. *J. Roy. Statist. Soc. Ser. B* 58, 609–617.
- Grodzenskii, S.Y., Domrachev, V.G., 2002. Estimation of the parameters of a mixture of exponential and Weibull distributions with progressive censoring. *Measurement Tech.* 45 (f11), 1115–1118.
- Hathaway, R.J., 1985. A constrained formulation of maximum likelihood estimation for normal mixture distributions. *Ann. Statist.* 13, 795–800.
- Henna, J., 1994. Examples of identifiable mixture. *J. Japan Statist. Soc.* 24, 193–200.
- Jewell, N.P., 1982. Mixtures of exponential distributions. *Ann. Statist.* 10 (2), 479–484.
- Leroux, S., 1992. Consistent estimation of a mixing distribution. *Ann. Statist.* 20 (3), 1350–1360.
- Lindsay, B.G., 1995. *Mixture Models: Theory, Geometry and Applications*. IMS,
- Marazzi, A., Paccaud, F., Ruffieux, C., Beguin, C., 1998. Fitting the distributions of length of stay by parametric models. *Medical Care* 36 (6), 915–927.
- McLachlan, G.J., Krisnan, T., 1997. *The EM Algorithm and Extensions*. Wiley, New York.
- McLachlan, G.J., Peel, D., 2001. *Finite Mixture Models*. Wiley, New York.
- Peters, B.C., Walker, H.F., 1978. An iterative procedure for obtaining maximum-likelihood estimates of the parameters for a mixture of normal distributions. *SIAM J. Appl. Math.* 35, 362–378.
- Pfanzagl, S., 1988. Consistency of maximum likelihood estimators for certain nonparametric families, in particular: mixtures. *J. Statist. Plann. Inference* 19, 137–158.
- Rachev, S.T., SenGupta, A., 1993. Laplace-Weibull mixtures for modeling price changes. *Management Sci.* 39 (8), 1029–1038.
- Redner, R.A., 1981. Note on the consistency of the maximum-likelihood estimate for non-identifiable distributions. *Ann. Statist.* 9, 225–228.
- Redner, R.A., Walker, H.F., 1984. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* 26, 195–239.
- Self, S., Liang, K., 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* 82 (398), 605–610.
- Teicher, H., 1963. Identifiability of finite mixtures. *Ann. Math. Statist.* 34, 1265–1269.
- Titterton, D.M., Smith, A.F.M., Makov, U.E., 1985. *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York.
- Van De Geer, S., 2003. Asymptotic theory for maximum likelihood in nonparametric mixture models. *Comput. Statist. Data Anal.* 41, 453–464.
- Wald, A., 1949. Note on the consistency of the maximum-likelihood estimate. *Ann. Math. Statist.* 20, 595–600.
- Zasada, M., Cieszewski, C.J., 2005. A finite mixture distribution approach for characterizing tree diameter distributions by natural social class in pure even-aged Scots pine stands in Poland. *Forest Ecol. Manage.* 204 (2–3), 145–158.