

Consistency of Maximum Likelihood Estimators in Finite Mixture Models of the Union of the Union of W-Type Families

N. ATIENZA, J. GARCÍA-HERAS, AND
J. MUÑOZ-PICHARDO

Departamento de Matemática Aplicada I, Escuela de Ingeniería
Informática, Sevilla, Spain

In finite mixture models, maximum likelihood estimators have good properties, such as efficiency, consistency, and asymptotic normality under some uniform integrability assumptions on the mixture and its derivatives up to the third order. These conditions are frequently not easy to check because complex computations on bounding a lot of derivatives are involved. We give results implying these conditions for a new class of families of distributions, W-type families, which make it easier to check the conditions in many cases. Many useful and known families of distributions such as Weibull, Generalized Gamma, Log-gamma, inverse Log-gamma, inverse Gaussian, and all of the exponential families are W-type families. Hence, these results have broad applications.

Keywords Consistency; Exponential families; Generalized Gamma; Mixtures; Weibull.

Mathematics Subject Classification AMS 2000 62F12; 62E20.

1. Introduction

In the last few decades, finite mixture models have often been used for describing complex statistical problems (see Everitt and Hand, 1981; McLachlan and Basford, 1988; McLachlan and Peel, 2000; Titterton et al., 1985 for references). Many scientific studies have used finite mixtures of normal distributions as models in real experiences. In other cases, mixtures of alternative families have been considered preferable when, for instance, the intrinsic range of the variable is not the entire real line, or when skewness is exhibited by the data being modeled. Finite mixtures

Address correspondence to Nieves Atienza, Departamento de Matemática Aplicada I, Escuela de Ingeniería Informática, Av. Reina Mercedes s/n. 41012 Sevilla, Spain; E-mail: natiENZA@us.es

of statistical distributions such as negative exponential, Gamma, Weibull, Erlang, etc. have shown their usefulness in these cases. The components of the mixture are often homogeneous. However, mixture distributions need not consist of components from the same family. For instance, a mixture of a Weibull and an Negative Exponential distribution is considered by Johnson et al. (1994) and Balakrishnan and Rao (2001), a mixture of a Weibull and a Wald distribution by Johnson et al. (1994), and an inverse Gaussian–Weibull mixture model by Al-Hussaini and Abd-El-Hakim (1989). In order to provide more flexibility, Atienza (2003) proposes to use finite mixtures from different parametric families of distributions when the modeling of experiences ranges among different possible models. For example, variables which can be modeled by Gamma, Weibull, or Log-normal distributions appear in numerous studies in different fields of applications, such as medicine (McKeegan, 2002), microbiology (Whiting and Golden, 2002), veterinary science (Singer et al., 2001), meteorology (Costa et al., 2000; Stephenson et al., 1999), epidemiology (Cooley et al., 1996), sedimentology (Kondolf and Adhikari, 2000), and mechanics (Ahn et al., 2000).

However, this fact can give rise to additional difficulties in processing, particularly in maximum likelihood (ML) estimation. 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. Redner and Walker (1984) provided conditions to assure the existence, consistency, and asymptotic normality of these estimators in finite mixtures of distributions. These conditions are generally not easy to check. The paper of Atienza et al. (2003) is centered on studying these conditions for the case of finite mixtures generated from the same exponential family or from the union of different exponential families. Therefore, the consistency problem for ML estimators of finite mixtures of distributions such as Log-normal, Gamma, Erlang, etc. is solved.

There exist real situations in which finite mixtures generated from non exponential families are considered. This is the case of the Weibull distribution, and more generally of the Generalized Gamma distribution (Caroni, 2002; Radhakrishna et al., 1992). A general approach to this problem is obtained by considering a new class of families of distributions, \mathcal{W} -type families. This new class contains all the exponential families and the generalized Gamma family. In particular, it contains the negative exponential, Gamma, Log-normal, Weibull, Log-gamma, and Log-gamma inverse distributions, among others. All these distributions are frequently used to model experiences with asymmetric data.

As a first step, necessary conditions for the consistency of finite mixtures of distributions generated from the same \mathcal{W} -type family are studied. For the sake of getting better results, the study of consistency conditions in finite mixtures generated from the union of different \mathcal{W} -type families becomes necessary.

In the second section, \mathcal{W} -type families of distributions are defined and a related result, which will be necessary throughout the rest of the article, is stated. Section 3 provides conditions under which the existence of ML estimators with good properties is verified. Finally, the last two sections provide results which make it easier to check these consistency conditions in the case of finite mixtures generated from a \mathcal{W} -type family and from the finite union of different \mathcal{W} -type families, respectively. In both cases the usefulness of these results is shown with applications to examples where the simplicity of checking our conditions is illustrated, reducing

drastically the amount of calculations with respect to direct application of the known consistency conditions. In the Appendix we include two lemmas which are used throughout the article.

It should be mentioned that an analysis of the consistency problem recently has been presented by Cheng and Liu (2001) for the case where the data come from an embedded distribution with fewer components than that being fitted.

2. \mathcal{W} -Type Families and Some Technical Results

\mathcal{W} -type families arise while looking for a kind of family of distributions which includes exponential parametric families and the Generalized Gamma family. This new type of family allows us to tackle the consistency problem for finite mixtures from a broad set of distributions, either for the same family or from different families. In this section \mathcal{W} -type families are introduced, and some useful results related to them are given.

Definition 2.1. Let $\Theta = \Delta_1 \times \Delta_2 \times \Delta_3 \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_1}$ be an open set, $D \subset \mathbb{R}^n$ a measurable set, and $a : D \subset \mathbb{R}^n \rightarrow [0, +\infty)$, $\mathbf{t} : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^{d_1}$, $b : \Theta \rightarrow (0, +\infty)$, and $c : \Delta_2 \subset \mathbb{R}^{d_2} \rightarrow \mathbb{R}$ be functions such that

$$f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^+ / f(\mathbf{x}) = a(\mathbf{x})b(\boldsymbol{\theta}) \exp\{\boldsymbol{\alpha}^\top \mathbf{t}(\mathbf{x}) + c(\boldsymbol{\beta})e^{\lambda^\top \mathbf{t}(\mathbf{x})}\}$$

are density functions for all $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) \in \Theta$. The set of distributions given by

$$\mathcal{W}(a, \mathbf{t}, b, c) = \{f(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) \in \Theta\}$$

is called a \mathcal{W} -type family of distributions associated with the functions a , \mathbf{t} , b , and c .

Note that every exponential family is a \mathcal{W} -type family, thus, examples of distribution families which are a \mathcal{W} -type family are the Negative Exponential, Gamma (\mathcal{G}), and Log-normal (\mathcal{L}). Moreover, families such as Generalized Gamma (\mathcal{G}_g), Weibull (W), Log-gamma, and Log-gamma inverse are also examples of \mathcal{W} -type families. In Table 1, some of these functions, together with their \mathcal{W} -type representations, are shown.

The following definition is needed for the proof of Theorem 2.1.

Definition 2.2. Given $\mathbf{t} : \mathbb{R}^n \rightarrow \mathbb{R}^d$ with $\mathbf{t}(\mathbf{x}) = (t_1(\mathbf{x}), \dots, t_d(\mathbf{x}))$ and h a non negative integer, the family of functions $\Pi_h(\mathbf{t})$ given by

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

is called the family of product functions of at most h components of \mathbf{t} .

The following theorem will be used to study the uniform integrability of the partial derivatives of any density function belonging to a \mathcal{W} -type family with respect to all of its parameters.

From now on, for $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^t$ and $r > 0$, $\mathcal{Q}_d(\boldsymbol{\theta}, r)$ denotes the hypercube $\prod_{i=1}^d [\theta_i - r, \theta_i + r]$.

Table 1
 \mathcal{W} -type representations of some distributions

Distrib.	Parameters	Functions
$\mathcal{G}[j, k]$	$\alpha = \left(j, -\frac{1}{k} \right)$ $\beta = \beta$ $\lambda = \lambda$	$a(x) = \frac{1}{x}, b(\theta) = \frac{(-\alpha_2)^{\alpha_1}}{\Gamma(\alpha_1)}, \mathbf{t}(x) = \begin{pmatrix} \log x \\ x \end{pmatrix}, c(\beta) \equiv 0$
$W[l, d]$	$\alpha = l$ $\beta = d^l$ $\lambda = l$	$a(x) = \frac{1}{x}, b(\theta) = \frac{\alpha}{\beta}, t(x) = \log x, c(\beta) = \frac{-1}{\beta}$
$\mathcal{L}[\mu, \sigma]$	$\alpha = \left(\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2} \right)$ $\beta = \beta$ $\lambda = \lambda$	$a(x) = \frac{1}{x}, b(\theta) = \frac{-\alpha_1^2}{4\sqrt{\pi\alpha_2}}, \mathbf{t}(x) = \begin{pmatrix} \log x \\ -\log^2 x \end{pmatrix}, c(\beta) \equiv 0$
$\mathcal{G}_g[m, n, r]$	$\alpha = mr$ $\beta = n^r$ $\lambda = r$	$a(x) = \frac{1}{x}, b(\theta) = \frac{\lambda}{\Gamma(\alpha/\lambda)\beta^{\alpha/\lambda}}, t(x) = \log x, c(\beta) = \frac{-1}{\beta}$

Theorem 2.1. *Under the notations stated above, if the function*

$$f(\cdot, \alpha, \beta, \lambda) : \mathbf{x} \in D \mapsto a(\mathbf{x}) \exp\{\alpha^\top \mathbf{t}(\mathbf{x}) + c(\beta) e^{\lambda^\top \mathbf{t}(\mathbf{x})}\}$$

is integrable on D for every $\theta = (\alpha, \beta, \lambda) \in \Theta$, where $c : \Delta_2 \rightarrow \mathbb{R}$ has a continuous partial derivative of first order which does not vanish in Δ_2 , then for every $\theta \in \Theta$, non negative integer h and $T \in \Pi_h(\mathbf{t})$:

- (1) *the functions $\mathbf{x} \in D \mapsto T(\mathbf{x}) e^{m\lambda^\top \mathbf{t}(\mathbf{x})} f(\mathbf{x}, \alpha, \beta, \lambda)$ are integrable for any non negative integer m ;*
- (2) *given $r > 0$ such that $Q_{d_1}(\alpha, 3r) \subset \Delta_1$, for $0 \leq m \leq 3$ and for every $\lambda_1 \in Q_{d_1}(\lambda, r)$, the functions $\mathbf{x} \in D \mapsto T(\mathbf{x}) e^{m\lambda_1^\top \mathbf{t}(\mathbf{x})} f(\mathbf{x}, \alpha, \beta, \lambda)$ are integrable.*

Proof. In order to prove (1) it is assumed that $\partial c / \partial \beta_1$ is continuous and does not vanish in Δ_2 . Since $T \in \Pi_h(\mathbf{t})$,

$$T(\mathbf{x}) f(\mathbf{x}, \alpha, \beta, \lambda) = \prod_{i=1}^d (t_i(\mathbf{x}))^{v_i} f(\mathbf{x}, \alpha, \beta, \lambda) = \frac{\partial^{h'}}{\partial \alpha_1^{v_1}, \dots, \partial \alpha_k^{v_d}} f(\mathbf{x}, \alpha, \beta, \lambda)$$

where $h' = \sum_{i=1}^d v_i \leq h$. Each partial derivative of $f(\mathbf{x}, \alpha, \beta, \lambda)$ with respect to any component of α , namely $t_i(\mathbf{x}) f(\mathbf{x}, \alpha, \beta, \lambda)$, has the same form as $f(\mathbf{x}, \alpha, \beta)$, with $a(\mathbf{x})$ being replaced by $a(\mathbf{x}) t_i(\mathbf{x})$. Hence, it will suffice to show that the partial derivative of $f(\mathbf{x}, \alpha, \beta, \lambda)$ with respect to any component of α is integrable since a repeated application of these would imply the integrability of $T(\mathbf{x}) f(\mathbf{x}, \alpha, \beta, \lambda)$. Without any loss of generality this will be proved for the last component of α , denoted by z . Fix $\theta^* = (\alpha_{(d_1)}^*, z^*, \beta^*, \lambda^*) \in \mathbb{R}^{d_1-1} \times \mathbb{R} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_1}$,

with $\boldsymbol{\alpha}_{(d_1)}^* = (\alpha_1^*, \dots, \alpha_{d_1-1}^*)$, and consider $f_{(\boldsymbol{\alpha}_{(d_1)}^*, \boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)}(\mathbf{x}, z) : D \times [z_1, z_2] \rightarrow \mathbb{R}$ given by $f_{(\boldsymbol{\alpha}_{(d_1)}^*, \boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)}(\mathbf{x}, z) = f(\mathbf{x}, \boldsymbol{\alpha}_{(d_1)}^*, z, \boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)$, where $z_1, z_2 \in \mathbb{R}$ are such that $z^* \in [z_1, z_2]$ and $(\boldsymbol{\alpha}_{(d_1)}^*, z, \boldsymbol{\beta}^*, \boldsymbol{\lambda}^*) \in \Theta$ for every $z \in [z_1, z_2]$. This function obviously verifies condition (1) from Lemma A.2 (see Appendix). Furthermore,

$$\frac{\partial f_{(\boldsymbol{\alpha}_{(d_1)}^*, \boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)}(\mathbf{x}, z)}{\partial z} = a(\mathbf{x})t_{d_1}(\mathbf{x}) \exp\{(\boldsymbol{\alpha}_{(d_1)}^*)^\top \mathbf{t}_{(d_1)}(\mathbf{x}) + zt_d(\mathbf{x}) + c(\boldsymbol{\beta}^*)e^{(\boldsymbol{\lambda}^*)^\top \mathbf{t}(\mathbf{x})}\}$$

with $\mathbf{t}_{(d_1)} = (t_1, \dots, t_{d_1-1})$, which exists and is measurable. Finally, by continuity, it is obvious that $z \in [z_1, z_2] \mapsto a(\mathbf{x})t_{d_1}(\mathbf{x}) \exp\{(\boldsymbol{\alpha}_{(d_1)}^*)^\top \mathbf{t}_{(d_1)}(\mathbf{x}) + zt_d(\mathbf{x}) + c(\boldsymbol{\beta}^*)e^{(\boldsymbol{\lambda}^*)^\top \mathbf{t}(\mathbf{x})}\}$ is integrable on $[z_1, z_2]$ and therefore (2) from Lemma A.2 is satisfied. Applying Lemma A.2, for almost all $z \in [z_1, z_2]$, we get the integrability of the function

$$\mathbf{x} \in D \mapsto \frac{\partial f_{(\boldsymbol{\alpha}_{(d_1)}^*, \boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)}}{\partial z}(\mathbf{x}, z).$$

However, integrability for almost all $z \in [z_1, z_2]$ does not assure integrability for $z = z^*$. To this end, choose z', z'' such that $z^* \in [z', z''] \subset [z_1, z_2]$ and

$$\frac{\partial f_{(\boldsymbol{\alpha}_{(d_1)}^*, \boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)}}{\partial z}(\mathbf{x}, z') \quad \text{and} \quad \frac{\partial f_{(\boldsymbol{\alpha}_{(d_1)}^*, \boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)}}{\partial z}(\mathbf{x}, z'')$$

are integrable with respect to \mathbf{x} . The inequality

$$\left| \frac{\partial f_{(\boldsymbol{\alpha}_{(d_1)}^*, \boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)}}{\partial z}(\mathbf{x}, z^*) \right| \leq |a(\mathbf{x})t_{d_1}(\mathbf{x}) \exp\{(\boldsymbol{\alpha}_{(d_1)}^*)^\top \mathbf{t}_{(d_1)}(\mathbf{x}) + z't_d(\mathbf{x}) + c(\boldsymbol{\beta}^*)e^{(\boldsymbol{\lambda}^*)^\top \mathbf{t}(\mathbf{x})}\}| \\ + |a(\mathbf{x})t_{d_1}(\mathbf{x}) \exp\{(\boldsymbol{\alpha}_{(d_1)}^*)^\top \mathbf{t}_{(d_1)}(\mathbf{x}) + z''t_d(\mathbf{x}) + c(\boldsymbol{\beta}^*)e^{(\boldsymbol{\lambda}^*)^\top \mathbf{t}(\mathbf{x})}\}|,$$

which is pointwise true for only one of the addends (depending on whether $t_d(\mathbf{x}) < 0$ or $t_d(\mathbf{x}) \geq 0$) by monotonicity, proves that $t_{d_1}(\mathbf{x})f(\mathbf{x}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)$ is integrable. By applying this process h' times, we deduce that $T(\mathbf{x})f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})$ is integrable for all $\boldsymbol{\theta} \in \Theta$.

Differentiating with respect to β_1 (for which c has a continuous partial derivative different from 0 what implies the monotonicity of c in the component β_1) and applying Lemma A.2, we conclude that the function

$$\frac{\partial f}{\partial \beta_1}(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = e^{\boldsymbol{\lambda}^\top \mathbf{t}(\mathbf{x})} \frac{\partial c}{\partial \beta_1}(\boldsymbol{\beta})f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

is integrable for almost every β_1 . A similar reasoning as above and the monotonicity of c show that the previous function is integrable for every β_1 . Thus, the function $e^{\boldsymbol{\lambda}^\top \mathbf{t}(\mathbf{x})}f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})$ is also integrable.

To prove that the functions $T(\mathbf{x})e^{m\boldsymbol{\lambda}^\top \mathbf{t}(\mathbf{x})}f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})$, $T \in \Pi_h(\mathbf{t})$ are integrable, it suffices to apply the same argument to the partial derivative of $f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})$ m times with respect to β_1 and according to the factors appearing in T , with respect to the components of $\boldsymbol{\alpha}$ divided previously by $\partial c / \partial \beta_1 (\neq 0)$.

The proof of (2) is based on writing $T(\mathbf{x})e^{m\boldsymbol{\lambda}_1^\top \mathbf{t}(\mathbf{x})}f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})$ in the form:

$$T(\mathbf{x})a(\mathbf{x}) \exp\{(\boldsymbol{\alpha} + m\boldsymbol{\lambda}_1)^\top \mathbf{t}(\mathbf{x}) + c(\boldsymbol{\beta})e^{\boldsymbol{\lambda}^\top \mathbf{t}(\mathbf{x})}\}.$$

Adding and subtracting $m\lambda^\top \mathbf{t}(\mathbf{x})$ in the exponent yields:

$$\begin{aligned} T(\mathbf{x})a(\mathbf{x}) \exp\{(\boldsymbol{\alpha} + m\lambda_1 - m\lambda)^\top \mathbf{t}(\mathbf{x}) + m\lambda^\top \mathbf{t}(\mathbf{x}) + c(\boldsymbol{\beta})e^{\lambda^\top \mathbf{t}(\mathbf{x})}\} \\ = T(\mathbf{x})e^{m\lambda^\top \mathbf{t}(\mathbf{x})}f(\mathbf{x}, \boldsymbol{\alpha} + m(\lambda_1 - \lambda), \boldsymbol{\beta}, \lambda). \end{aligned}$$

Since $(\lambda_1 - \lambda)m < 3r$ and, by hypothesis, $\mathcal{Q}_{d_1}(\boldsymbol{\alpha}, 3r) \subset \Delta_1$, we have, $\boldsymbol{\alpha} + m(\lambda_1 - \lambda) \in \Delta_1$. By applying (1) from this theorem to these new values of the parameters, the proposed result is achieved. \square

This result will be used in the proof of the theorems in Sec. 4 and 5.

3. Mixture Distributions. Known Results on Consistency

Consider a finite mixture of distributions whose density is given by

$$p(\mathbf{x}, \Psi) = \sum_{j=1}^k \pi_j f_j(\mathbf{x}, \theta_j),$$

where $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{k-1}) \in \mathcal{C}_k = \{(\pi_1, \dots, \pi_{k-1}) : \sum_{j=1}^{k-1} \pi_j < 1, \pi_j > 0\}$, $\pi_k = 1 - \sum_{j=1}^{k-1} \pi_j$ 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 Ω denote the parametric space $\mathcal{C}_k \times \Theta_1 \times \dots \times \Theta_k$ and let $\Psi = (\psi_1, \dots, \psi_v)$ denote the elements of Ω , where v is the dimension of Ω .

Redner and Walker (1984) state that, under two consistency conditions (here called **C1** and **C2**), in any sufficiently small neighborhood of the value of the parameter, there is a unique strongly consistent solution Ψ^N of the likelihood equations, and this solution at least locally maximizes the log-likelihood function and is asymptotically normally distributed.

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

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

where h_j and h_{ij} are integrable and h_{ijs} satisfies

$$\int h_{ijs}(\mathbf{x})p(\mathbf{x}; \Psi^*)d\mathbf{x} < +\infty.$$

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

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

is well defined and positive definite at Ψ^* , where ∇_Ψ denotes the gradient of the first partial derivatives with respect to the components of Ψ .

Two questions remain unresolved by the theorem above regarding Ψ^N : (1) “Is Ψ^N really an ML estimator?” (2) “Is there any other ML estimator besides Ψ^N ”

which leads to a limiting density different from $p(x; \Psi^*)$?. Next we present other two conditions, **C3** and **C4**, given in Atienza et al. (2003). These two conditions are simpler in applications than the conditions given in (Redner and Walker, 1984).

For each $\theta_i \in \Theta_i$ and sufficiently small $r > 0$, let $N_r(\theta_i)$ denotes the closed ball of radius r about θ_i in Θ_i and let

$$f_{ir}(x; \theta_i) = \sup_{\theta'_i \in N_r(\theta_i)} f_i(x; \theta'_i) \quad \text{and} \quad f_{ir}^*(x; \theta_i) = \max\{1, f_{ir}(x; \theta_i)\}.$$

C3. For each $\theta = (\theta_1, \dots, \theta_k) \in \Theta = \Theta_1 \times \dots \times \Theta_k$ and sufficiently small $r > 0$,

$$\int (\log f_{ir}^*(x; \theta_i)) f_j(x; \theta_j^*) dx < +\infty \quad \text{for } i, j = 1, \dots, k.$$

C4. For $i, j = 1, \dots, k$, $\int |\log f_i(x, \theta_i^*)| f_j(x, \theta_j^*) dx < +\infty$.

If $E = \{\Psi \in \Omega' : p(x; \Psi) = p(x; \Psi^*) \text{ for almost every } x \in \mathbb{R}^n\}$, under conditions **C3** and **C4**, in any compact $\Omega' \subset \Omega$ which contains Ψ^* , for each closed set $D \subset \Omega' \setminus E$, with probability 1,

$$\lim_{N \rightarrow +\infty} \sup_{\Psi \in D} \frac{\prod_{i=1}^N p(x_i; \Psi)}{\prod_{i=1}^N p(x_i; \Psi^*)} = 0.$$

That is, for any closed set far from Ψ^* in any compact set contained in Ω and for sufficiently large N , the maximum value of the likelihood function in this closed set is considerably smaller than the likelihood function at Ψ^* . Thus, the likelihood function has the largest local maximum at the solution Ψ^N , whose existence is already established in (Redner and Walker, 1984), and this solution has good properties of consistency and asymptotic normality (see Atienza et al., 2003).

From a practical point of view, consistency conditions **C1**, **C3**, and **C4** are not easy to check, especially when numerous parameters are involved in the mixture. Section 4 is devoted to provide conditions which assure those consistency conditions above in mixtures from a wide set of families of distributions (\mathcal{W} -type families). The second condition, concerning the Fisher information matrix, is not considered here since an individual study in each specific family is apparently needed.

4. Consistency for Finite Mixtures of \mathcal{W} -Type Families

From now on, Θ^k will denote the product space $\prod_{i=1}^k \Theta_i$. For $\mathbf{y} \in \mathbb{R}^{d_1}$ and every $\boldsymbol{\varepsilon} \in \mathbb{R}^{d_1}$, a vector of signs with components $\varepsilon_i = \pm 1$, we denote by $\mathbf{y}(\boldsymbol{\varepsilon}) \in \mathbb{R}^{d_1}$ the vertices of the hypercube $Q_{d_1}(\mathbf{y}^*, r)$, with coordinates $y_i(\boldsymbol{\varepsilon}) = y_i^* + \varepsilon_i r$.

Theorem 4.1. *Let $\mathcal{W}(a, \mathbf{t}, b, c)$ be a \mathcal{W} -type family and $\Psi^* = (\boldsymbol{\pi}, \boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_k^*) \in \mathcal{C}_k \times \Theta^k$ with $\boldsymbol{\theta}_i^* = (\boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*, \boldsymbol{\lambda}_i^*)$ for $i = 1, \dots, k$. If*

(H1) *b has continuous partial derivatives up to the third order, and*

(H2) *either $c \equiv 0$ or c has continuous partial derivatives up to the third order with a partial derivative of first order not vanishing in $\boldsymbol{\beta}_i^*$ for every $i = 1, \dots, k$, then any finite mixture of k components from $\mathcal{W}(a, \mathbf{t}, b, c)$ verifies consistency condition **C1** on a compact set $\widehat{\mathcal{C}}_k \times \widehat{\Theta}_1 \times \dots \times \widehat{\Theta}_k \subset \mathcal{C}_k \times \Theta^k$ containing Ψ^* .*

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

Since Θ is open, there exists an $r > 0$ such that $Q_d(\theta_i^*, r) \subset \Theta$ for every $i = 1, \dots, k$. Hence, for any compact set $\widehat{\mathcal{C}}_k \subset \mathcal{C}_k$ containing π^* , the subset $\widehat{\mathcal{C}}_k \times Q_d(\theta_1^*, r) \times \dots \times Q_d(\theta_k^*, r) \subset \mathcal{C}_k \times \Theta^k$ is a compact set containing Ψ^* .

With respect to the conditions concerning the derivatives up to the second order, by linearity of the derivative and triangular inequality, it suffices to prove that for every $\theta \in Q_d(\theta^*, r)$ (for the sake of simplicity, we have omitted the indices of the densities):

$$|f(\mathbf{x}, \theta)| \leq g(\mathbf{x}), \quad (1)$$

$$\left| \frac{\partial f(\mathbf{x}, \theta)}{\partial \theta_j} \right| \leq g_j(\mathbf{x}), \quad (2)$$

$$\left| \frac{\partial^2 f(\mathbf{x}, \theta)}{\partial \theta_j \partial \theta_s} \right| \leq g_{js}(\mathbf{x}), \quad (3)$$

with g, g_j, g_{js} integrable functions with respect to \mathbf{x} and independent of θ .

We can suppose $b(\theta) \equiv 1$, considering $f(\mathbf{x}, \theta)/b(\theta)$ if necessary because b has continuous partial derivatives up to the third order and the theorem is stated in a compact set.

Throughout the proof, let $M > 0$ denote an upper bound on $\widehat{\Theta}$ of all the continuous functions appearing in the sequel, which depend on c and its derivatives up to the third order.

(1) is proved using Lemma A.1 (see the Appendix) and taking β_0 the value of β which maximizes $c(\beta)$. In fact, for every $\theta \in \widehat{\Theta}$

$$|f(\mathbf{x}, \theta)| \leq \sum_{\boldsymbol{\varepsilon} \in V_{d_1}} \left(\sum_{s \in \{-1, 1\}} a(\mathbf{x}) \exp\{\mathbf{t}(\mathbf{x})\boldsymbol{\alpha}(\boldsymbol{\varepsilon}) + c(\beta_0) e^{\lambda(s\boldsymbol{\varepsilon})^\top \mathbf{t}(\mathbf{x})}\} \right) = g(\mathbf{x}),$$

which is integrable and independent of θ .

To prove (2), differentiating with respect to the parameters, we obtain

$$\begin{aligned} \left| \frac{\partial f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \alpha_i} \right| &= |f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) t_i(\mathbf{x})|, \quad i = 1, \dots, d_1, \\ \left| \frac{\partial f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \beta_j} \right| &= \left| f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) \frac{\partial c(\boldsymbol{\beta})}{\partial \beta_j} e^{\lambda^\top \mathbf{t}(\mathbf{x})} \right|, \quad j = 1, \dots, d_2 \quad \text{and} \\ \left| \frac{\partial f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \lambda_u} \right| &= |f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) c(\boldsymbol{\beta}) t_u(\mathbf{x}) e^{\lambda^\top \mathbf{t}(\mathbf{x})}|, \quad u = 1, \dots, d_1, \end{aligned}$$

which are integrable functions, by Theorem 2.1. Furthermore, these functions are bounded respectively by:

$$\begin{aligned} g_{1i}(\mathbf{x}) &= g(\mathbf{x}) t_i(\mathbf{x}), \\ g_{2j}(\mathbf{x}) &= M \sum_{\boldsymbol{\varepsilon}, s} e^{\lambda^\top(\boldsymbol{\varepsilon}) \mathbf{t}(\mathbf{x})} f(\mathbf{x}, (\boldsymbol{\alpha}(\boldsymbol{\varepsilon}), \boldsymbol{\beta}_0, \lambda(s\boldsymbol{\varepsilon}))) \quad \text{and} \\ g_{3u}(\mathbf{x}) &= |t_u(\mathbf{x})| M \sum_{\boldsymbol{\varepsilon}, s} e^{\lambda(\boldsymbol{\varepsilon})^\top \mathbf{t}(\mathbf{x})} f(\mathbf{x}, (\boldsymbol{\alpha}(\boldsymbol{\varepsilon}), \boldsymbol{\beta}_0, \lambda(s\boldsymbol{\varepsilon}))), \end{aligned}$$

which are also integrable functions by (2) in Theorem 2.1 and independent of θ . Hence (2) is proved.

(3) follows by the same method as above, using functions of the form

$$f(\mathbf{x}, \theta)T(\mathbf{x})e^{m\lambda^\top \mathbf{r}(\mathbf{x})}C(\boldsymbol{\beta}), \quad m = 1, 2, 3$$

with $T \in \Pi_3(\mathbf{r})$ and $C(\boldsymbol{\beta})$ a continuous function, bounded by M . From Theorem 2.1, it is known that these sums are integrable and, by Lemma A.1, they can be bounded by integrable functions independent of θ .

It remains to prove that the conditions for the third derivatives of the logarithm are satisfied. We take into account that

$$\begin{aligned} \frac{\partial^3 \log p}{\partial \psi_i \partial \psi_j \partial \psi_u} &= \frac{1}{p} \frac{\partial^3 p}{\partial \psi_i \partial \psi_j \partial \psi_u} - \frac{1}{p^2} \left[\frac{\partial^2 p}{\partial \psi_i \partial \psi_j} \frac{\partial p}{\partial \psi_u} + \frac{\partial^2 p}{\partial \psi_j \partial \psi_u} \frac{\partial p}{\partial \psi_i} + \frac{\partial^2 p}{\partial \psi_i \partial \psi_u} \frac{\partial p}{\partial \psi_j} \right] \\ &\quad + \frac{2}{p^3} \frac{\partial p}{\partial \psi_i} \frac{\partial p}{\partial \psi_j} \frac{\partial p}{\partial \psi_u}, \end{aligned} \quad (4)$$

and that 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 written as a sum of functions of the form

$$\left| \frac{f}{p} T(\mathbf{x}) e^{m\lambda^\top \mathbf{r}(\mathbf{x})} B(\theta) \right|$$

where $T \in \Pi_3(\mathbf{r})$, $m \leq 3$, and $B(\theta)$ is a continuous function, independent of \mathbf{x} and bounded in absolute value by M' on $\bigcup_{i=1}^k Q_d(\theta_i^*, r)$. Bounding p in each of the denominators by

$$p(\mathbf{x}, \boldsymbol{\Psi}) = \sum_{i=1}^k \pi_i f(\mathbf{x}, \theta_i) \geq \alpha_0 f(\mathbf{x}, \theta_h),$$

where $f(\mathbf{x}, \theta_h)$ is, in each case, the term appearing in the numerator of the fraction and $\alpha_0 > 0$ is the minimum of π_1, \dots, π_k when $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{k-1}) \in \widehat{\mathcal{E}}_k$, we can bound the sum of these expressions by a sum of functions of the form

$$\left| \frac{1}{\alpha_0} T(\mathbf{x}) e^{m\lambda^\top \mathbf{r}(\mathbf{x})} M' \right|, \quad \text{with } T \in \Pi_3(\mathbf{r}).$$

Hence, (4) is bounded by a sum of functions of the form

$$\left| \frac{1}{\alpha_0} T(\mathbf{x}) e^{m\lambda^\top \mathbf{r}(\mathbf{x})} M' \right|, \quad \left| \frac{1}{\alpha_0^2} T(\mathbf{x}) e^{m\lambda^\top \mathbf{r}(\mathbf{x})} M'^2 \right| \quad \text{and} \quad \left| \frac{1}{\alpha_0^3} T(\mathbf{x}) e^{m\lambda^\top \mathbf{r}(\mathbf{x})} M'^3 \right|,$$

which are independent of $\boldsymbol{\Psi}$. The sum of these functions is considered as the desired function h_{uvw} . From Theorem 2.1, the product $h_{uvw}(\mathbf{x})p(\mathbf{x}, \boldsymbol{\Psi}^*)$ is integrable. \square

Theorem 4.2. Let $\mathcal{W}(a, \mathbf{t}, b, c)$ be a \mathcal{W} -type family and $\boldsymbol{\psi}^* = (\boldsymbol{\pi}, \boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_k^*) \in \mathcal{C}_k \times \Theta^k$ with $\boldsymbol{\theta}_i^* = (\boldsymbol{\alpha}_i^*, \boldsymbol{\beta}_i^*, \boldsymbol{\lambda}_i^*)$ for $i = 1, \dots, k$. If

(H3) b is continuous,

(H4) c is continuous with a continuous partial derivative of first-order non vanishing in $\boldsymbol{\beta}_i^*$, for any $i = 1, \dots, k$, and

(H5) $\int |\log a(\mathbf{x})| f(\mathbf{x}, \boldsymbol{\theta}_i^*) d\mathbf{x} < +\infty$, $i = 1, \dots, k$,

then any finite mixture of k components from $\mathcal{W}(a, \mathbf{t}, b, c)$ verifies consistency conditions **C3** and **C4** on any compact set $\widehat{\Omega} \subset \Omega$ containing $\boldsymbol{\psi}^*$.

Proof. Taking $r > 0$ such that $Q_d(\boldsymbol{\theta}_i^*, 2r) \subset \Theta$, we will prove that the theorem is verified for all $\boldsymbol{\theta}_i \in Q_d(\boldsymbol{\theta}_i^*, r)$.

Using the proof of Lemma A.1 taking into account that $c(\boldsymbol{\beta})$ is continuous, it is deduced that for any $\mathbf{x} \in D$ there exists $\boldsymbol{\varepsilon}(\mathbf{x}) \in V_d$ such that

$$\exp\{\boldsymbol{\alpha}^\top \mathbf{t}(\mathbf{x}) + c(\boldsymbol{\beta}) e^{\boldsymbol{\lambda}^\top \mathbf{t}(\mathbf{x})}\} \leq \exp\{\boldsymbol{\alpha}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x}) + M e^{\boldsymbol{\lambda}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})}\}$$

for any $\boldsymbol{\theta}_i \in Q_d(\boldsymbol{\theta}_i^*, 2r)$, where $\boldsymbol{\alpha}(\boldsymbol{\varepsilon})$ and $\boldsymbol{\lambda}(\boldsymbol{\varepsilon})$ are, respectively, the vertices of the hypercubes $Q_d(\boldsymbol{\alpha}_i^*, 2r)$ and $Q_d(\boldsymbol{\lambda}_i^*, 2r)$. The value $M > 1$ is an upper bound for b and $|c|$. Then

$$f(\mathbf{x}, \boldsymbol{\theta}) \leq M a(\mathbf{x}) \exp\{\boldsymbol{\alpha}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x}) + M e^{\boldsymbol{\lambda}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})}\} \quad \forall \boldsymbol{\theta} \in Q_d(\boldsymbol{\theta}_i^*, 2r)$$

and since $N_r(\boldsymbol{\theta}_i) \subset Q_d(\boldsymbol{\theta}_i^*, 2r)$ for any $\boldsymbol{\theta}_i \in Q_d(\boldsymbol{\theta}_i^*, r)$,

$$f_r(\mathbf{x}, \boldsymbol{\theta}_i) \leq \max\{1, M a(\mathbf{x}) \exp\{\boldsymbol{\alpha}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x}) + M e^{\boldsymbol{\lambda}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})}\}\}.$$

Taking logarithms, we obtain

$$\begin{aligned} \log f_r^*(\mathbf{x}, \boldsymbol{\theta}_i) &\leq \max\{0, \log(M a(\mathbf{x}) \exp\{\boldsymbol{\alpha}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x}) + M e^{\boldsymbol{\lambda}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})}\})\} \\ &\leq \log M + \max\{0, \log a(\mathbf{x})\} + \max\{0, \boldsymbol{\alpha}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x}) + M e^{\boldsymbol{\lambda}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})}\} \\ &\leq \log M + |\log a(\mathbf{x})| + C \sum_{i=1}^k |t_i(\mathbf{x})| + M \sum_{\boldsymbol{\varepsilon}} e^{\boldsymbol{\lambda}(\boldsymbol{\varepsilon})^\top \mathbf{t}(\mathbf{x})} \end{aligned}$$

where C is a constant which bounds the absolute values of the coordinates on $Q_{d_1}(\boldsymbol{\alpha}_i^*, 2r)$.

From Theorem 2.1, $|t_i(\mathbf{x})| f(\mathbf{x}, \boldsymbol{\theta}_i^*)$ and $e^{\boldsymbol{\lambda}(\boldsymbol{\varepsilon})^\top \mathbf{t}(\mathbf{x})} f(\mathbf{x}, \boldsymbol{\theta}_j^*)$ are integrable and, by hypothesis, $|\log a(\mathbf{x})| f(\mathbf{x}, \boldsymbol{\theta}_j^*)$ is also integrable. Thus we have proved the integrability of $\log f_r^*(\mathbf{x}, \boldsymbol{\theta}_i) f(\mathbf{x}, \boldsymbol{\theta}_j^*)$ as desired.

Condition **C4** follows by bounding $|\log f(\mathbf{x}, \boldsymbol{\theta}_i^*)| f(\mathbf{x}, \boldsymbol{\theta}_j^*)$ by

$$\left[|\log a(\mathbf{x})| + |\log b(\boldsymbol{\theta}_i^*)| + \sum_{s=1}^d |\alpha_{is}^* t_s(\mathbf{x})| + M e^{(\boldsymbol{\lambda}_i^*)^\top \mathbf{t}(\mathbf{x})} \right] f(\mathbf{x}, \boldsymbol{\theta}_j^*),$$

which is a sum of integrable functions. □

Using the previous conditions for \mathcal{W} -type families, we can now tackle the problem of checking conditions **C1**, **C3**, and **C4** for finite mixtures from parametric families such as the Generalized Gamma family given by:

$$\mathcal{G}_g = \left\{ f(x, m, n, r) = \frac{rx^{mr-1}}{n^{mr}\Gamma(m)} e^{-(\frac{x}{n})^r}; m, n, r > 0, x \in (0, +\infty) \right\}$$

which is a \mathcal{W} -type family (see Table 1).

Since the parametric space is $\Theta = (0, +\infty) \times (0, +\infty) \times (0, +\infty)$ and the functions b and c are continuous with continuous derivatives in Θ , the set of finite mixtures $\mathcal{H}_{\mathcal{G}_g}$ from the family \mathcal{G}_g , verifies consistency condition **C1**. In order to verify **C3** and **C4**, taking into account that $|\log a(x)| = |\log x|$, it suffices to prove that $f(x, m, n, r) \log x$ is integrable, which is true by Theorem 2.1, since $\mathbf{t}(x) = \log x$.

Similar analyses can be applied to finite mixtures from other parametric families such as the Log-gamma and Log-gamma inverse families.

5. Consistency for Mixtures from the Union of \mathcal{W} -Type Families

As we have noted in the introduction, many real experiences can be modeled using finite mixtures of elements from different families. Thus, we can obtain better results in the model with the observed data. So far, we have obtained results which assure the verification of consistency conditions for finite mixtures from the same \mathcal{W} -type family. This section is devoted to provide results in order to simplify the verification when functions from different \mathcal{W} -type families are used. We illustrate these results with a finite mixture of functions from the Log-normal and the Generalized Gamma distributions. To this end, the family of product functions of components of several vector functions is defined below in the same way as Definition 2.2. Finally, the generalizations of Theorems 4.1 and 4.2 to the case of finite mixture functions from different \mathcal{W} -type families are stated.

Definition 5.1. Given k functions $\mathbf{t}_i : \mathbb{R}^n \rightarrow \mathbb{R}^{d_i}$, $\mathbf{t}_i(\mathbf{x}) = (t_{i1}(\mathbf{x}), \dots, t_{id_i}(\mathbf{x}))$ and h a non negative integer, the family of functions $\Pi_h(\mathbf{t}_1, \dots, \mathbf{t}_k)$ given by

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

is called the family of product functions of at most h components of $\mathbf{t}_1, \dots, \mathbf{t}_k$.

Theorem 5.1. Let $\boldsymbol{\psi}^* = (\boldsymbol{\pi}^*, \boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_k^*) \in \Omega = \mathcal{C} \times \Theta_1 \times \dots \times \Theta_k$ and let $\mathcal{W}_i(a_i, \mathbf{t}_i, b_i, c_i)$ be k \mathcal{W} -type families defined, respectively, on $\Theta_i \subset \mathbb{R}^{d_i}$ with $i \in \{1, \dots, k\}$ such that

(H1)_i b_i has continuous partial derivatives up to the third order in their domains,

(H2)_i either $c_i \equiv 0$ or c_i has continuous partial derivatives up to the third order in its domain with a first partial derivative not vanishing at $\boldsymbol{\beta}_i^*$ for every $i = 1, \dots, k$, and

(H6) for $i, l, s, j \in \{1, \dots, k\}$, $m_i, m_l, m_s \in \mathbb{Z}$ such that $m_i + m_l + m_s \leq 3$ and

$$c_r = 0 \Rightarrow m_r = 0 \text{ for all } r = i, l, s$$

and for all $T \in \Pi_3(\mathbf{t}_i, \mathbf{t}_l, \mathbf{t}_s)$ it is verified that

$$\int |T(\mathbf{x})| e^{m_i \hat{\lambda}_i^\top \mathbf{t}_i(\mathbf{x})} e^{m_l \hat{\lambda}_l^\top \mathbf{t}_l(\mathbf{x})} e^{m_s \hat{\lambda}_s^\top \mathbf{t}_s(\mathbf{x})} f_j(\mathbf{x}, \boldsymbol{\theta}_j) d\mathbf{x} < +\infty,$$

then any finite mixture $p(\mathbf{x}, \boldsymbol{\psi}) = \sum_{i=1}^k \pi_i f_i(\cdot, \boldsymbol{\theta}_i)$ with $f_i(\mathbf{x}, \boldsymbol{\theta}_i) \in \mathcal{W}_i(a_i, \mathbf{t}_i, b_i, c_i)$ verifies consistency condition **C1** on a compact set $\widehat{\mathcal{C}}_k \times \widehat{\Theta}_1 \times \cdots \times \widehat{\Theta}_k \subset \mathcal{C}_k \times \Theta_1 \times \cdots \times \Theta_k$ containing $\boldsymbol{\psi}^*$.

Remark 5.1. Compared to Theorem 4.1, a linking hypothesis among the families (H6) is needed.

Proof. Conditions (1), (2), and (3) are proved in the same way as in Theorem 4.1. For the third derivatives, note that the absolute value of $|\partial^3 \log p / \partial \psi_u \partial \psi_v \partial \psi_w|$ is bounded by a sum of expressions of the forms

$$\begin{aligned} & \sum_{\boldsymbol{\varepsilon}} \left| \frac{M}{\alpha_0} T(\mathbf{x}) e^{m_i \hat{\lambda}_i(\boldsymbol{\varepsilon})^\top \mathbf{t}_i(\mathbf{x})} \right| \quad \text{with } T \in \Pi_3(\mathbf{t}_i) \text{ or,} \\ & \sum_{\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}'} \left| \frac{M}{\alpha_0^2} T(\mathbf{x}) e^{m_i \hat{\lambda}_i(\boldsymbol{\varepsilon})^\top \mathbf{t}_i(\mathbf{x})} e^{m_l \hat{\lambda}_l(\boldsymbol{\varepsilon}')^\top \mathbf{t}_l(\mathbf{x})} \right| \quad \text{with } T \in \Pi_3(\mathbf{t}_i, \mathbf{t}_l) \text{ or} \\ & \sum_{\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}', \boldsymbol{\varepsilon}''} \left| \frac{M}{\alpha_0^3} T(\mathbf{x}) e^{m_i \hat{\lambda}_i(\boldsymbol{\varepsilon})^\top \mathbf{t}_i(\mathbf{x})} e^{m_l \hat{\lambda}_l(\boldsymbol{\varepsilon}')^\top \mathbf{t}_l(\mathbf{x})} e^{m_s \hat{\lambda}_s(\boldsymbol{\varepsilon}'')^\top \mathbf{t}_s(\mathbf{x})} \right| \quad \text{with } T \in \Pi_3(\mathbf{t}_i, \mathbf{t}_l, \mathbf{t}_s) \end{aligned}$$

where $m_r \in \{0, 1, 2, 3\}$ such that $m_i + m_l + m_s \leq 3$.

Note that if $c_r \equiv 0$ then $m_r = 0$, since the factor $e^{m_r \mathbf{t}_r(x) \lambda_r}$ is obtained by differentiating f_r with respect to components of the parametric vectors $\boldsymbol{\beta}_r$ or λ_r .

From Theorem 2.1, all of these summands are independent of the parameters and integrable with respect to $p(\mathbf{x}, \boldsymbol{\psi}^*)$. The sum of them h_{uvw} verifies that $h_{uvw}(\mathbf{x}) p(\mathbf{x}, \boldsymbol{\psi}^*)$ is integrable. \square

Theorem 5.2. Let $\boldsymbol{\psi}^* = (\boldsymbol{\pi}^*, \boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_k^*) \in \Omega = \mathcal{C} \times \Theta_1 \times \cdots \times \Theta_k$ and let $\mathcal{W}_i(a_i, \mathbf{t}_i, b_i, c_i)$ be k \mathcal{W} -type families defined respectively on $\Theta_i \subset \mathbb{R}^{d_i}$ with $i \in \{1, \dots, k\}$ such that

(H3)_{*i*} b_i is continuous in Θ_i ,

(H4)_{*i*} either $c_i \equiv 0$ or c_i has first continuous partial derivatives, one of them is not vanishing at $\boldsymbol{\beta}_i^*$, and,

(H7) for $i, l \in \{1, \dots, k\}$,

$$\int |\log a_i(\mathbf{x})| f_i(\mathbf{x}, \boldsymbol{\theta}_i^*) d\mathbf{x} < +\infty \quad \text{and}$$

$$\int |t_{is}(\mathbf{x})| e^{m_i t_i(\mathbf{x}) \lambda_i} f_i(\mathbf{x}, \boldsymbol{\theta}_i^*) d\mathbf{x} < +\infty \quad \text{with } m_i = \begin{cases} 0, 1, 2, 3 & \text{if } c_i \neq 0 \\ 0 & \text{if } c_i = 0 \end{cases}$$

Then any finite mixture $p(\mathbf{x}, \boldsymbol{\psi}) = \sum_{i=1}^k \pi_i f_i(\mathbf{x}, \boldsymbol{\theta}_i)$ with $f_i(\cdot, \boldsymbol{\theta}_i) \in \mathcal{W}_i(a_i, \mathbf{t}_i, b_i, c_i)$ verifies consistency conditions **C3** and **C4** on any compact set $\widehat{\Omega} \subset \Omega$ containing $\boldsymbol{\psi}^*$.

Remark 5.2. Compared to Theorem 4.2, this result for the union of \mathcal{W} -type families introduces a linking hypothesis (H7) among the families.

Proof. The proof runs as in Theorem 4.2. \square

To illustrate the usefulness of the previous theorems, we consider a family of finite mixtures studied in Atienza (2003) to model the variable length of hospital stay. This variable was studied by Marazzi et al. (1998), where a procedure to decide between Log-normal, Gamma, and Weibull distributions was provided. This method does not always determine one single model as valid. In Atienza (2003), a model of finite mixtures of these three distributions is proposed in order to describe this variable and a revision for other asymmetric variables, which can be also fitted with one of the Log-normal, Weibull, or Gamma families, is made. This model provides an extra flexibility which in turn allows to get a better description of the behavior of these variables.

According to this model, the family of finite mixtures generated by the union of the Log-normal, Gamma, and Weibull families, given in Table 1, is considered. The Gamma and Weibull families are both particular cases of Generalized Gamma families, while the Log-normal family can be regarded as an exponential family, which is a \mathcal{W} -type family. Hence, these mixtures are contained in the set of all finite mixtures generated from the union of two different \mathcal{W} -type families, given by $\mathcal{G}_{\mathcal{G}} = \mathcal{W}(a_1, t_1, b_1, c_1)$ with a_1, t_1, b_1, c_1 as in Table 1, and $\mathcal{L} = \mathcal{W}(a_2, t_2, b_2, c_2)$ with $a_2(x) = 1/x$, $t_2(x) = (\log x, -\log^2 x)$, $b_2(\alpha_1, \alpha_2) = \pi^{-1} \alpha_1^{1/2} e^{-\alpha_1/(4\alpha_2)}$ and $c_2 \equiv 0$ for $(\alpha_1, \alpha_2) = (\mu/\sigma^2, 1/(2\sigma^2)) \in \mathbb{R} \times (0, +\infty)$.

In order to see that the family of finite mixtures with components from the union $\mathcal{L} \cup \mathcal{G}_{\mathcal{G}}$ verifies consistency conditions **C1**, **C3**, and **C4**, taking into account what the components of $t_i(x)$ can be and $|\log a_i(x)| = |\log x|$ for $i = 1, 2$, it suffices to prove that $x^s \log^h x f_i(x, \theta_i)$ is integrable for non negative integers s, h and for $f_i(x, \theta_i)$ a density function from any of the previous families.

If $s = 0$, these functions can be obtained by multiplying $f_i(x, \theta_i)$ by components of $t_i(x)$ of any of the families. By Theorem 2.1, these are integrable functions.

If $s \neq 0$, the factor x^s can be written as $e^{s \log x}$, and hence

$$(\log x)^h e^{s \log x} f(x, \alpha_1, \alpha_2) = (\log x)^h f(x, \alpha_1 + s, \alpha_2)$$

in the case of the Log-normal family, and

$$(\log x)^h e^{s \log x} f(x, a, b) = (\log x)^h f(x, a + s, b)$$

in the case of the Generalized Gamma family, which are in any case integrable from Theorem 2.1.

Appendix

Lemma A.1. Let $t : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^d$, $(\alpha^*, \lambda^*) \in \mathbb{R}^d \times \mathbb{R}^d$ and $r > 0$. For every $x \in D$, and $(\alpha, \lambda) \in \mathcal{Q}_d(\alpha^*, r) \times \mathcal{Q}_d(\lambda^*, r)$ and $K \in \mathbb{R}$

$$\begin{aligned} & \exp\{\alpha^\top t(x) + K e^{\lambda^\top t(x)}\} \\ & \leq \sum_{\varepsilon \in V_d} \left(\exp\{\alpha(\varepsilon)^\top t(x) + K e^{\lambda(\varepsilon)^\top t(x)}\} + \exp\{\alpha(\varepsilon)^\top t(x) + K e^{\lambda(-\varepsilon)^\top t(x)}\} \right) \end{aligned}$$

where $V_d = \{\varepsilon \in \mathbb{R}^d / \varepsilon_i = \pm 1 \forall i = 1, \dots, d\}$.

Proof. Given $\mathbf{x} \in D$, consider $\boldsymbol{\varepsilon}(\mathbf{x}) = [\varepsilon_1(\mathbf{x}), \dots, \varepsilon_d(\mathbf{x})]^\top$ such that

$$\varepsilon_i(\mathbf{x}) = \begin{cases} 1 & \text{if } t_i(\mathbf{x}) \geq 0, \\ -1 & \text{if } t_i(\mathbf{x}) < 0, \end{cases} \quad (5)$$

which, obviously, belongs to V_d . Then

$$e^{\boldsymbol{\alpha}^\top \mathbf{t}(\mathbf{x})} \leq e^{\boldsymbol{\alpha}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})} \quad \text{and} \quad e^{\lambda(-\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})} \leq e^{\lambda^\top \mathbf{t}(\mathbf{x})} \leq e^{\lambda(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})}.$$

Hence

$$Ke^{\lambda^\top \mathbf{t}(\mathbf{x})} \leq Ke^{\lambda(s(K)\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})},$$

where $s(K)$ is the sign of K . Hence,

$$\begin{aligned} \exp\{\boldsymbol{\alpha}^\top \mathbf{t}(\mathbf{x}) + Ke^{\lambda^\top \mathbf{t}(\mathbf{x})}\} &\leq \exp\{\boldsymbol{\alpha}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x}) + Ke^{\lambda(s(K)\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})}\} \\ &\leq \exp\{\boldsymbol{\alpha}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x}) + Ke^{\lambda(+\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})}\} \\ &\quad + \exp\{\boldsymbol{\alpha}(\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x}) + Ke^{\lambda(-\boldsymbol{\varepsilon}(\mathbf{x}))^\top \mathbf{t}(\mathbf{x})}\}. \end{aligned}$$

Thus, adding in all of the vertices,

$$\begin{aligned} &\exp\{\boldsymbol{\alpha}^\top \mathbf{t}(\mathbf{x}) + Ke^{\lambda^\top \mathbf{t}(\mathbf{x})}\} \\ &\leq \sum_{\boldsymbol{\varepsilon} \in V_{d_1}} \left(\exp\{\boldsymbol{\alpha}(\boldsymbol{\varepsilon})^\top \mathbf{t}(\mathbf{x}) + Ke^{\lambda(+\boldsymbol{\varepsilon})^\top \mathbf{t}(\mathbf{x})}\} + \exp\{\boldsymbol{\alpha}(\boldsymbol{\varepsilon})^\top \mathbf{t}(\mathbf{x}) + Ke^{\lambda(-\boldsymbol{\varepsilon})^\top \mathbf{t}(\mathbf{x})}\} \right) \quad \square \end{aligned}$$

Lemma A.2. Let $I \subset \mathbb{R}$ be a compact interval, $D \subset \mathbb{R}^n$ a measurable set and $f : D \times I \rightarrow \mathbb{R}$ so that for every $\theta \in I$:

- (1) the function $\mathbf{x} \in D \mapsto f(\mathbf{x}, \theta)$ is integrable;
- (2) there exists $\partial f(\mathbf{x}; \theta) / \partial \theta$ for almost every $\mathbf{x} \in D$, such that:

- (2a) $(\mathbf{x}, \theta) \mapsto \partial f(\mathbf{x}, \theta) / \partial \theta$ is measurable, and
- (2b) for almost every $\mathbf{x} \in D$, the function

$$\theta \in I \mapsto \frac{\partial f}{\partial \theta}(\mathbf{x}, \theta)$$

is integrable on I ; and

- (3) the set $Z = \{\mathbf{x} : \exists \theta_1, \theta_2 \in I / \partial f(\mathbf{x}, \theta_1) / \partial \theta > 0 \text{ and } \partial f(\mathbf{x}, \theta_2) / \partial \theta < 0\}$ is a null set.

Then, for almost every $\theta \in I$, the function

$$\mathbf{x} \in D \mapsto \frac{\partial f}{\partial \theta}(\mathbf{x}, \theta),$$

is integrable.

The proof can be found in Atienza et al. (2003).

Acknowledgments

The authors have been supported by DGES grant #BFM2001-3844.

References

- Ahn, Y., Nicholson, D. W., Wang, M. C., Ni, P. (2000). Inverse method for identifying the underlying crack distribution in plates with random strengths. *Acta Mechanica* 144:137–154.
- Al-Hussaini, E. K., Abd-El-Hakim, N. S. (1989). Failure rate of the inverse Gaussian-Weibull mixture model. *Ann. Instit. Statist. Math.* 41:617–622.
- Atienza, N. (2003). *Mixturas de Distribuciones: Modelización de Experiencias con Asimetría en los Datos*. Ph.D. Thesis. University of Seville, Spain.
- Atienza, N., Garcia-Heras, J., Muñoz-Pichardo, J., Villa, R. (2003). Consistency of maximum likelihood estimator in finite mixture models of the union of exponential families. Preprint.
- Balakrishnan, N., Rao, C. R. (2001). *Advances in Reliability*. Handbook of Statistics, 20. Amsterdam: North-Holland Publishing Co.
- Caroni, C. (2002). The correct “ball bearings” data. *Lifetime-Data-Anal.* 8(4):395–399.
- Cheng, R. C. H., Liu, W. B. (2001). The consistency of estimators in finite mixture models. *Scand. J. Statist.* 28(4):603–616.
- Cooley, P. C., Myers, L. E., Hamill, D. N. (1996). A meta-analysis of estimates of the AIDS incubation distribution. *Eur. J. Epidemiology* 12(3):229–235.
- Costa, A. A., de Oliveira, C. J., de Oliveira, J. C. P., Sampaio, A. J. D. (2000). Microphysical observations of warm cumulus clouds in Ceara, Brazil. *Atmospheric Res.* 54:167–199.
- Everitt, B. S., Hand, D. J. (1981). *Finite Mixture Distributions*. London: Chapman and Hall.
- Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions*. Vol. 1. 2nd ed. New York: Wiley.
- Kondolf, G. M., Adhikari, A. (2000). Weibull vs. lognormal distributions for fluvial gravels. *J. Sedimentary Res.* 70:456–460.
- Marazzi, A., Paccaud, F., Ruffieux, C., Beguin, C. (1998). Fitting the distributions of length of stay by parametric models. *Med-Care.* 36(6):915–927.
- McKeegan, D. E. F. (2002). Spontaneous and odour evoked activity in single avian olfactory bulb neurones. *Brain Res.* 929(1):48–58.
- McLachlan, G. J., Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.
- McLachlan, G. J., Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.
- Radhakrishna, C., Dattatreya-Rao, A. V., Anjaneyulu, G. V., Peters, B. C., Walker, H. F. (1992). Estimation of parameters in two-component mixture generalized gamma distribution. *Commun. Statist. Theor. Meth.* 21(6):1799–1805.
- Redner, R. A., Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* 26:195–239.
- Singer, R. S., MacLachlan, N. J., Carpenter, T. E. (2001). Maximal predicted duration of viremia in bluetongue virus-infected cattle. *J. Veterinary Diagnostic Investigation* 13:43–49.
- Stephenson, D. B., Kumar, K. R., Doblaz-Reyes, F. J., Royer, J. F., Chauvin, E., Pezzulli, S. (1999). Extreme daily rainfall events and their impact on ensemble forecasts of the Indian monsoon. *Month. Weather Rev.* 70:1954–1966.
- Titterton, D. M., Smith, A. F. M., Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.
- Whiting, R. C., Golden, M. H. (2002). Variation among *Escherichia coli* O157: H7 strains relative to their growth, survival, thermal inactivation, and toxin production in broth. *Int. J. Food Microbiol.* 75:127–133.