ARTICLE IN PRESS
Journal of Statistical Planning and Inference ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

On consistency of the MLE under finite mixtures of location-scale distributions with a structural parameter

Guanfu Liu [a], Pengfei Li [b], Yukun Liu [c,*], Xiaolong Pu [c]

[a] School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai 201620, China
[b] Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, ON, Canada N2L 3G1,
[c] School of Statistics, East China Normal University, Shanghai 200241, China

ARTICLE INFO

Article history:
Received 30 November 2016
Received in revised form 3 May 2018
Accepted 21 May 2018
Available online xxxx

Keywords:
Consistency
Finite mixture of location-scale distributions
Maximum likelihood estimator
Structural parameter

ABSTRACT

We provide a general and rigorous proof for the strong consistency of maximum likelihood estimators of the cumulative distribution function of the mixing distribution and structural parameter under finite mixtures of location-scale distributions with a structural parameter. The consistency results do not require the parameter space of location and scale to be compact. We illustrate the results by applying them to finite mixtures of location-scale distributions with the component density function being one of the commonly used density functions: normal, logistic, extreme-value, or t. An extension of the strong consistency results to finite mixtures of multivariate elliptical distributions is also discussed.

1. Introduction

Suppose we have an independent and identically distributed (*i.i.d.*) sample $X_1, \dots, X_n$ from the following finite mixture model:

$$g(x; \Psi, \sigma) = \sum_{j=1}^{m} \alpha_j f(x; \mu_j, \sigma) = \int_{\mathbb{R}} f(x; \mu, \sigma) d\Psi(\mu). \tag{1}$$

Here $f(x; \mu, \sigma)$, the component density function, is assumed to come from a location-scale distribution family, namely, $f(x; \mu, \sigma) = \sigma^{-1} f((x-\mu)/\sigma; 0, 1)$ with $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$ being the location and scale parameters, respectively. The positive integer m is called the order of the mixture model, $(\alpha_1, \dots, \alpha_m)$ with $\alpha_j \geq 0$ and $\sum_{j=1}^{m} \alpha_j = 1$ are called the mixing proportions, and $\Psi(\mu) = \sum_{j=1}^{m} \alpha_j I(\mu_j \leq \mu)$ is called the cumulative distribution function of the mixing distribution. The parameter σ, appearing in all m component density functions, is called a structural parameter, and Model (1) is called a finite mixture of location-scale distributions with a structural parameter. Note that $\Psi(\cdot)$ includes unknown μ_j and α_j parameters. Hence, (Ψ, σ) covers all the unknown parameters in (1). In this paper, we investigate the strong consistency of the maximum likelihood estimator (MLE) of (Ψ, σ) under Model (1).

Finite mixtures of location-scale distributions with a structural parameter have many applications. They play an important role in medical studies and genetics. For example, Roeder (1994) applied the finite normal mixture model with a structural parameter to analyze sodium–lithium countertransport activity in red blood cells. Finite mixtures of logistic distributions and of extreme value distributions with a structural parameter are widely used to analyze failure-time data.

* Corresponding author.
E-mail address: ykliu@sfs.ecnu.edu.cn (Y. Liu).

https://doi.org/10.1016/j.jspi.2018.05.003
0378-3758/

For instance, a finite mixture of logistic distributions with a structural parameter was used by Naya et al. (2006) to study the thermogravimetric analysis trace. A finite mixture of extreme value distributions with a structural parameter was found to provide an adequate fit to the logarithm of the number of cycles to failure for a group of 60 electrical appliances (Lawless, 2003; Example 4.4.2). More applications can be found in McLachlan and Peel (2000) and (Lawless, 2003).

The maximum likelihood method has been widely used to estimate the unknown parameters in finite mixture models (McLachlan and Peel, 2000; Chen, 2017). The consistency of the MLE under finite mixture models has been studied by Kiefer and Wolfowitz (1956), Redner (1981), and Chen (2017). As pointed out by Chen (2017), the results in Kiefer and Wolfowitz (1956) require that $g(x; \Psi, \sigma)$ can be continuously extended to a compact space of (Ψ, σ). This turns out to be impossible because $f(x; \mu, \sigma)$ is not well defined at $\sigma = 0$. To make the results in Kiefer and Wolfowitz (1956) applicable to our current setup, we must constrain the parameter σ to be in a compact subset of $\mathbb{R}^+$. The consistency results in Redner (1981) require even more restrictive conditions: the parameter space for (μ, σ) must be a compact subset of $\mathbb{R} \times \mathbb{R}^+$; see Chen (2017) for more discussion. It is worth mentioning that Bryant (1991) established the strong consistency of the estimators obtained by the linear-optimization-based method. His result can be viewed as a generalization of the classical consistency result for MLE. However, it requires that the parameter space be closed and that m be equal to the true order of the mixture model. By utilizing the properties of the normal distribution, Chen (2017) proved the strong consistency of the MLE under finite normal mixture models with a structural parameter without imposing the compactness assumption on the parameter space. To the best of our knowledge, general consistency results for the MLE of (Ψ, σ) under Model (1) are not available in the literature except for the normal mixture model.

Because of the importance of finite mixtures of location-scale distributions with a structural parameter, it is necessary to study the consistency of the MLE of the underlying parameters, (Ψ, σ), under Model (1). The goal of this paper is to provide a general and rigorous proof of this consistency. In Section 2, we present the main consistency results. We emphasize that we do not require the parameter space of (μ, σ) to be compact. The detailed proofs are given in Section 3. Section 4 illustrates the consistency results by applying them to Model (1) with $f(x; \mu, \sigma)$ being one of the commonly used component density functions: normal, logistic, extreme-value, or t. An extension of the consistency results to finite mixtures of multivariate elliptical distributions is discussed in Section 5.

2. Main results

With the *i.i.d.* sample $X_1, \ldots, X_n$ from (1), the log-likelihood function of (Ψ, σ) is given by

$$\ell_n(\Psi, \sigma) = \sum_{i=1}^{n} \log\{g(X_i; \Psi, \sigma)\}.$$

The MLE of (Ψ, σ) is defined as

$$(\hat{\Psi}, \hat{\sigma}) = \arg \max_{\Psi \in \boldsymbol{\Psi}_m,\ \sigma > 0} \ell_n(\Psi, \sigma),$$

where

$$\boldsymbol{\Psi}_m = \left\{ \sum_{j=1}^{m} \alpha_j I(\mu_j \le \mu) : \alpha_j \ge 0,\ \sum_{j=1}^{m} \alpha_j = 1,\ \mu_j \in \mathbb{R} \right\}.$$

In this section, we establish the consistency property of $(\hat{\Psi}, \hat{\sigma})$ without imposing compactness on the parameter space of (μ, σ). To discuss the consistency of $\hat{\Psi}$, we define

$$D(\Psi_1, \Psi_2) = \int_{\mathbb{R}} |\Psi_1(\mu) - \Psi_2(\mu)| \exp(-|\mu|) d\mu. \tag{2}$$

We show that $D(\Psi_1, \Psi_2)$ is a distance on $\boldsymbol{\Psi}_m$ in the Appendix. Suppose $\Psi_0 \in \boldsymbol{\Psi}_m$ is the true cumulative distribution function of the mixing distribution. We say that $\hat{\Psi}$ is strongly consistent if $D(\hat{\Psi}, \Psi_0) \to 0$ almost surely as $n \to \infty$.

The strong consistency of $(\hat{\Psi}, \hat{\sigma})$ depends on the following regularity conditions.

C1. The finite mixture model in (1) is identifiable. That is, if (Ψ_1, σ_1) and (Ψ_2, σ_2) with $\Psi_1 \in \boldsymbol{\Psi}_m$, $\Psi_2 \in \boldsymbol{\Psi}_m$, $\sigma_1 > 0$, and $\sigma_2 > 0$ satisfy

$$\int_{\mathbb{R}} f(x; \mu, \sigma_1) d\Psi_1(\mu) = \int_{\mathbb{R}} f(x; \mu, \sigma_2) d\Psi_2(\mu)$$

for all x, then $\Psi_1 = \Psi_2$ and $\sigma_1 = \sigma_2$.

C2. $\int_{\mathbb{R}} |\log\{g(x; \Psi_0, \sigma_0)\}| \, g(x; \Psi_0, \sigma_0) dx < \infty$, where (Ψ_0, σ_0) is the true value of (Ψ, σ).

C3. There exist positive constants υ_0, υ_1, and β with $\beta > 1$ such that for all x

$$f(x; 0, 1) \le \min\left\{\upsilon_0, \upsilon_1 |x|^{-\beta}\right\}.$$

C4. The function $f(x; 0, 1)$ is continuous with respect to x.