Journal of Statistical Planning and Inference 141 (2011) 56–64

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Maximum likelihood estimators in finite mixture models with censored data

Yoichi Miyata

Faculty of Economics, Takasaki City University of Economics, 1300 Kaminamie, Takasaki, Gunma 370-0801, Japan

A R T I C L E   I N F O

Article history:
Received 19 June 2009
Received in revised form
21 April 2010
Accepted 12 May 2010
Available online 21 May 2010

Keywords:
Censored data
Finite mixture
Maximum likelihood estimators
Strong consistency

A B S T R A C T

The consistency of estimators in finite mixture models has been discussed under the topology of the quotient space obtained by collapsing the true parameter set into a single point. In this paper, we extend the results of Cheng and Liu (2001) to give conditions under which the maximum likelihood estimator (MLE) is strongly consistent in such a sense in finite mixture models with censored data. We also show that the fitted model tends to the true model under a weak condition as the sample size tends to infinity.

1. Introduction

Finite mixture models have been much studied both theoretically and practically by several authors (Redner, 1981; Feng and McCulloch, 1996; McLachlan and Peel, 2000; Cheng and Liu, 2001). Feng and McCulloch (1996) have proposed unrestricted MLEs which are consistent under some complicated conditions. Cheng and Liu (2001) have provided easily-verified conditions under which the vector of MLEs will converge to an arbitrary point in the subset representing the true model, allowing the estimators to approach a boundary point of the parameter space. On the other hand, censored and multimodal observations often appear in some fields such as reliability engineering, education, and so on. Chauveau (1995) has discussed a stochastic EM algorithm for the ML fitting of finite mixture models with censored data. However little is known about the strong consistency. In addition, these models cannot be directly applied to Cheng and Liu's results because each of the components does not uniformly converge to zero as the norm of its partial parameters tends to infinity. This paper extends their approach to show the strong consistency of the MLEs for finite mixture models with censored data when the number of components assumed is larger than or equal to the true one.

Section 3 shows the strong consistency of the MLEs and the fitted distributions in finite mixture models with fixed censoring regions. Section 4 provides parameter spaces under which the consistency results hold in a mixture of censored exponential distributions, a mixture of censored normal distributions, and a random censorship mixture model.

2. Definitions and assumptions

Let L^1 and B^+ denote the following spaces of integrable functions on $\mathbb{R}$.

$$L^1 = \left\{ f(x) | f(x) \text{ is measurable, } \|f\| = \int_{\mathbb{R}} |f(x)|\, d\mu < \infty \right\}, \tag{1}$$

E-mail address: ymiyatagbt@tcue.ac.jp

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.05.006

$$B^{+}=\{f(x)|f\in L^1,\|f\|=1,f(x)\geq 0\}. \tag{2}$$

Let Ω_1, Ω_2 be two closed sets in $\mathbb{R}^k$. We denote a metric between the two sets as follows:

$$dis(\Omega_1,\Omega_2)=dis(\Omega_2,\Omega_1)=\inf_{y\in\Omega_2}\inf_{x\in\Omega_1}|x-y|, \tag{3}$$

where $|\cdot|$ is the Euclidean norm. If Ω_1 and Ω_2 are singleton sets, then this metric agrees with the classic Euclidean distance. The following is used later.

Property 1. (i) *$dis(\Omega_1,\Omega_2)=0$ if and only if there are sequences of points, $\{x_n\}\in\Omega_1$ and $\{y_n\}\in\Omega_2$ such that $|x_n-y_n|\to 0$ as $n\to\infty$.*

(ii) *$dis(x_n,\Omega)\to 0$ if and only if there is a sequence $\{y_n\}$ of points in Ω, such that $|x_n-y_n|\to 0$ as $n\to\infty$.*

Let $\mathcal{F}_0=\{f^0(x|\boldsymbol{\theta})|\boldsymbol{\theta}\in\boldsymbol{\Theta}\}$ be a parametric family of one dimensional probability density functions (pdfs) with respect to a $\sigma-$finite measure μ_0 on $\mathcal{R}\subseteq\mathbb{R}$ from which mixtures are to be formed. Let X^0 be a random variable taking values in $\mathcal{R}$ with the pdf $f^0(x|\boldsymbol{\pi},\boldsymbol{\theta})=\sum_{j=1}^{g}\pi_j f^0(x|\boldsymbol{\theta}_j)$, where $f^0(x|\boldsymbol{\theta}_j)\in\mathcal{F}_0$, $\boldsymbol{\Theta}$ is a closed set belonging to $\mathbb{R}^d$,

$$\boldsymbol{\pi}\in\Pi\equiv\left\{(\pi_1,\pi_2,\ldots,\pi_g)\middle|\pi_j\geq 0,\sum_{j=1}^{g}\pi_j=1\right\}\quad\text{and} \tag{4}$$

$$\boldsymbol{\theta}\in\boldsymbol{\Theta}^g\equiv\{(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_g)|\boldsymbol{\theta}_j\in\boldsymbol{\Theta},(j=1,2,\ldots,g)\}, \tag{5}$$

(i.e., $\boldsymbol{\Theta}^g=\boldsymbol{\Theta}\times\cdots\times\boldsymbol{\Theta}$ is the Cartesian product of g copies of $\boldsymbol{\Theta}$). Let $\Gamma=\Pi\times\boldsymbol{\Theta}^g$ be a parameter space. Then Γ is a closed set.

Given a partition $R_0,R_1,\ldots,R_q$ of $\mathcal{R}$, we observe a random variable $X=X^0 1_{[X^0\in R_0]}+\sum_{k=1}^{q}c_k 1_{[X^0\in R_k]}$ taking values in $\mathcal{X}=R_0\cup\{c_1,\ldots,c_q\}$, where $1_{[\cdot]}$ is an indicator function, and c_k is just a code for the event $\{X^0\in R_k\}$. Let μ be the measure on $\mathcal{X}$ which coincides with μ_0 on R_0 and whose restriction of $\{c_1,\ldots,c_q\}$ is the counting measure on this set.

Then the pdf of X with respect to μ is given by

$$f(x|\boldsymbol{\pi},\boldsymbol{\theta})=\sum_{j=1}^{g}\pi_j f(x|\boldsymbol{\theta}_j), \tag{6}$$

where $F^k(\boldsymbol{\theta}_j)=\int_{R_k}f^0(t|\boldsymbol{\theta}_j)\,dt$, and

$$f(x|\boldsymbol{\theta}_j)=f^0(x|\boldsymbol{\theta}_j)1_{[x\in R_0]}+\sum_{k=1}^{q}F^k(\boldsymbol{\theta}_j)1_{[x=c_k]}. \tag{7}$$

We say that Eq. (6) is a finite mixture model with fixed censoring region. For any given $(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)\in\Gamma$ such that $f(x|\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)\in B^{+}$, we define the set

$$\Gamma(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)=\{(\boldsymbol{\pi},\boldsymbol{\theta})|(\boldsymbol{\pi},\boldsymbol{\theta})\in\Gamma,\text{ and } f(x|\boldsymbol{\pi},\boldsymbol{\theta})=f(x|\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)\}. \tag{8}$$

As well known in the ordinary mixture models, $\Gamma(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)$ is not a singleton set, and hence the MLE is not consistent in the sense of converging to a unique point. Therefore we shall use the distance (3) between the MLE and $\Gamma(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)$ to discuss the strong consistency.

In this paper, we allow the true parameter $(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)$ to be a boundary point of Γ. In this case, $\Gamma(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)$ becomes a continuum of parameter values (e.g., see Cheng and Liu, 2001; McLachlan and Peel, 2000, p. 28). Then the true model takes the reduced form $f(x|\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)=\sum_{j=1}^{g_0}\pi_{l_j}^0 f(x|\boldsymbol{\theta}_{l_j}^0)$ with $1\leq g_0\leq g-1$, and $\boldsymbol{\theta}_{l_j}^0\in\boldsymbol{\Theta}$. This means that the true number of components, g_0, is unknown, but the maximum number of g_0 is known.

2.1. Assumptions

We write expectations of $g(x)$ under $f(x|\boldsymbol{\theta}_i^0)$ by $E_{\boldsymbol{\theta}_i^0}[g(X)]=\int_{R_0}g(x)f(x|\boldsymbol{\theta}_i^0)\,dx+\sum_{k=1}^{q}g(c_k)f_i(c_k|\boldsymbol{\theta}_i^0)=\int_{\mathcal{X}}g(x)f(x|\boldsymbol{\theta}_i^0)\,d\mu$. Here we shall give sufficient conditions (a)–(g) under which the main results hold in the model (6).

(a) For any $\boldsymbol{\theta}_j\in\boldsymbol{\Theta}, f(x|\boldsymbol{\theta}_j)\in B^{+}$ $(j=1,\ldots,g)$. Furthermore, $f(x|\boldsymbol{\theta}_j^1)=f(x|\boldsymbol{\theta}_j^2)$ in B^{+} only if $\boldsymbol{\theta}_j^1=\boldsymbol{\theta}_j^2$.
(b) The support of $f(x|\boldsymbol{\theta}_j)$ does not depend on the parameter $\boldsymbol{\theta}_j\in\boldsymbol{\Theta}$.
(c) Let i=1,…,g and j=1,…,g.

- $E_{\boldsymbol{\theta}_i^0}[\log f(X|\boldsymbol{\theta}_j)]>-\infty$ for any $\boldsymbol{\theta}_j\in\boldsymbol{\Theta}$,
- $E_{\boldsymbol{\theta}_i^0}[\log\max\{f(X|\boldsymbol{\theta}_j),1\}]<\infty$ for any $\boldsymbol{\theta}_j\in\boldsymbol{\Theta}$,
- $E_{\boldsymbol{\theta}_i^0}[\log\sup_{|\boldsymbol{\theta}_j'-\boldsymbol{\theta}_j|\leq\rho}\max\{1,f(X|\boldsymbol{\theta}_j')\}]<\infty$ for small $\rho>0$, and any $\boldsymbol{\theta}_j\in\boldsymbol{\Theta}$,
- $E_{\boldsymbol{\theta}_i^0}[\log\sup_{|\boldsymbol{\theta}_j|\geq r}\max\{1,f(X|\boldsymbol{\theta}_j)\}]<\infty$ for large $r>0$.

(d) For any fixed $x\in\mathcal{X}$, $f(x|\boldsymbol{\theta}_j)$ are continuous with respect to $\boldsymbol{\theta}_j\in\boldsymbol{\Theta}$.
(e) $\lim_{|\boldsymbol{\theta}_j|\to\infty}f(x|\boldsymbol{\theta}_j)=0$ for any fixed $x\in R_0$.