# THE EM ALGORITHM OF LOCAL ESTIMATIONS FOR FINITE MIXTURE OF REGRESSION MODELS WITH LAPLACE DISTRIBUTION

WANG Ji-xia, WANG Chun-feng, MIAO Yu

(School of Mathematics and Information Sciences, Henan Normal University, Xinxiang 453007, China)

**Abstract:** In this paper, we study the local maximum likelihood estimations for a dinite mixture of regression models with Laplace distribution. By using the kernel regression method and the EM algorithm for maximizing the local weighted likelihood functions, we obtain the local maximum likelihood estimations of the parametric functions, and discuss their asymptotic biases, asymptotic variances and asymptotic normality, which extend the results of the local non-parametric estimations for the finite mixture of regression models.

**Keywords:** finite mixture model; Laplace distribution; EM algorithm; local maximum likelihood estimation; kernel regression

 2010 MR Subject Classification:
 62F10; 62J02

 Document code:
 A
 Article ID:
 0255-7797(2016)04-0667-09

#### 1 Introduction

Mixture models were widely used in social science and econometrics. The work for mixture models were well studied, for example, see [1]. The finite mixture model is a useful class of mixture model. Various efforts were made to explicitly express the finite mixture models, see [2–5].

The main aim of this paper is to provide a finite mixture of regression model with Laplace distribution. The parametric functions are allowed to vary smoothly. Based on the constant fitting, the local maximum likelihood estimations of the unknown parametric functions are obtained. Furthermore, an EM algorithm is proposed to carry out the estimation procedure. The EM algorithm has been used to maximize the likelihood functions when the models contain unobserved latent variables. One main important application of the EM algorithm is to find the maximum likelihood estimations for finite mixture models, see [6–7]. In this paper, we want to evaluate the unknown parametric functions at a set to grid points over an

**Received date:** 2013-12-08 **Accepted date:** 2015-02-26

**Foundation item:** Supported by National Natural Science Foundation of China (11471104); Natural Science Foundation of Henan Educational Committee (2011B110018); Program for Innovative Research Team (in Science and Technology) in University of Henan Province (14IRTSTHN023).

**Biography:** Wang Jixia (1978–), female, born at Pingyu, Henan, associate professor, major in statistic inference and financial statistics.

interval of a given point by using the EM algorithm. In addition, the monotone ascending property of the proposed EM algorithm is proved.

This article is organized as follows. In Section 2, we define the model. The local maximum likelihood estimations of the unknown parametric functions are obtained in Section 3. The EM algorithm for a finite mixture of regression model is provided in Section 4. The monotone ascending property of the proposed EM algorithm is proved in the last section.

#### 2 Model

Assume that  $\{(X_i, Y_i), i = 1, 2, \dots, n\}$  is a random sample from the population (X, Y), where the co-variable X is univariate. Let Z be a latent class variable. Suppose that Z has a discrete distribution  $P\{Z = k | X = x\} = p_k(x)$  given X = x, where  $k = 1, 2, \dots, M$ . Conditioning on Z = k and X = x, Y follows an Laplace distribution with mean  $\mu_k(x)$  and variance  $2\lambda_k^2(x)$ . We further assume that  $p_k(x)$ ,  $\mu_k(x)$  and  $\lambda_k^2(x)$  are unknown but smooth functions. Hence, conditioning on X = x, Y follows a finite mixture of regression models with Laplace distribution as follows:

$$Y|_{X=x} \sim \sum_{k=1}^{M} p_k(x) \frac{1}{2\lambda_k(x)} \exp\left\{-\frac{|y-\mu_k(x)|}{\lambda_k(x)}\right\}.$$
 (2.1)

In this paper, we assume that M is fixed. Model (2.1) is called the finite mixture of regression model with Laplace distribution. Model (2.1) can be viewed as a natural extension of finite mixture of linear regression model. For example, when  $p_k(x)$  and  $\lambda_k^2(x)$  are constants and  $\mu_k(x)$  is linear in x, model (2.1) yields to a finite mixture of linear regression models.

It is well known that identifiability is a critical issue for mixture models. Various efforts were made to study the identifiability of the finite mixture distributions, see [8]. We first introduce the concept of transversality.

**Definition 2.1** Let  $a(x) = (a_1(x), a_2(x))^T$  and  $b(x) = (b_1(x), b_2(x))^T$  be two smooth curves in  $\mathbb{R}^2$ , where  $x \in \mathbb{R}$ ,  $a_i(x)$  and  $b_i(x)$  are differentiable, i = 1, 2. If

$$||a(x) - b(x)||^{2} + ||a'(x) - b'(x)||^{2} \neq 0$$

for any  $x \in R$ , then we say that a(x) and b(x) are transversal.

From Definition 2.1, it can be seen that the transversality of two smooth curves a(x) and b(x) implies if a(x) = b(x), then  $a'(x) \neq b'(x)$ . Now we have the following theorem.

Theorem 2.2 Suppose that three conditions as following hold.

(c1)  $p_k(x) > 0$  are continuous functions, and  $\mu_k(x)$  and  $\lambda_k^2(x)$  are differentiable functions,  $k = 1, 2, \dots, M$ .

(c2) Any two curves  $(\mu_i(x), \lambda_i^2(x))^T$  and  $(\mu_j(x), \lambda_j^2(x))^T$ ,  $i \neq j$ , are transversal.

(c3) The range  $\chi$  of x is an interval in R.

Then model (2.1) is identifiable.

Theorem 2.2 shows a sufficient conditions of the identifiability for finite mixture of regression models.

#### **3** Local Maximum Likelihood Estimations

In this section, we study the local maximum likelihood estimations of the parametric functions  $p_k(x)$ ,  $\mu_k(x)$  and  $\lambda_k^2(x)$ ,  $k = 1, 2, \dots, M$ . The log-likelihood function for the collected data  $\{(X_i, Y_i), i = 1, 2, \dots, n\}$  is

$$\sum_{i=1}^{n} \log \left[ \sum_{k=1}^{M} p_k(X_i) \frac{1}{2\lambda_k(X_i)} \exp\left\{ -\frac{|Y_i - \mu_k(X_i)|}{\lambda_k(X_i)} \right\} \right].$$
 (3.1)

Note that  $p_k(x)$ ,  $\mu_k(x)$  and  $\lambda_k^2(x)$  are parametric functions. In this paper, we will employ the local constant fitting for model (2.1), see [9]. That is, for a given point x, we use local constants  $p_k$ ,  $\mu_k$  and  $\lambda_k^2$  to approximate  $p_k(x)$ ,  $\mu_k(x)$  and  $\lambda_k^2(x)$ , respectively. So the local weighted log-likelihood function for data  $\{(X_i, Y_i), i = 1, 2, \dots, n\}$  is

$$l_n(p,\mu,\lambda;x) = \sum_{i=1}^n \log\left[\sum_{k=1}^M p_k \frac{1}{2\lambda_k} \exp\left\{-\frac{|Y_i - \mu_k|}{\lambda_k}\right\}\right] K_h(X_i - x), \tag{3.2}$$

where  $p = (p_1, p_2, \dots, p_M)^T$ ,  $\mu = (\mu_1, \mu_2, \dots, \mu_M)^T$ ,  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_M)^T$ ,  $K_h(\cdot) = K(\cdot/h)/h$ ,  $K(\cdot)$  be a nonnegative weighted function and h is a properly selected bandwidth. Let  $(\tilde{p}, \tilde{\mu}, \tilde{\lambda})$  be the maximizer of the local weighted log-likelihood function (3.2). Then the local maximum likelihood estimations of  $p_k(x)$ ,  $\mu_k(x)$  and  $\lambda_k^2(x)$  are

$$\tilde{p}_k(x) = \tilde{p}_k, \quad \tilde{\mu}_k(x) = \tilde{\mu}_k, \text{ and } \tilde{\lambda}_k^2(x) = \tilde{\lambda}_k^2,$$
(3.3)

respectively.

Now we study the asymptotic bias, asymptotic variance and asymptotic normality as the following. Let  $\theta = (p^T, (\lambda^2)^T, \mu^T)^T$  and denote

$$\eta(y|\theta) = \sum_{k=1}^{M} p_k \frac{1}{2\lambda_k} \exp\left\{-\frac{|y-\mu_k|}{\lambda_k}\right\}, \quad l(\theta, y) = \log \eta(y|\theta).$$

Furthermore, let  $\theta(x) = (p^T(x), (\lambda^2(x))^T, \mu^T(x))^T$ , and denote

$$I(x) = -E\left[\frac{\partial^2 l(\theta(X), Y)}{\partial \theta \partial \theta^T} | X = x\right]$$

and

$$\Lambda(u|x) = \int_Y \frac{\partial l(\theta(x),y)}{\partial \theta} \eta(y|\theta(u)) dy.$$

For  $k = 1, 2, \cdots, M$ , denote  $\tilde{\mu}_k^* = \{\tilde{\mu}_k - \mu_k\}, (\tilde{\lambda}_k^2)^* = \{\tilde{\lambda}_k^2 - \lambda_k^2\}$ . For  $k = 1, 2, \cdots, M - 1$ , denote  $\tilde{p}_k^* = \{\tilde{p}_k - p_k\}$ . Let  $\tilde{\mu}^* = (\tilde{\mu}_1^*, \tilde{\mu}_2^*, \cdots, \tilde{\mu}_M^*)^T, (\tilde{\lambda}^2)^* = ((\tilde{\lambda}_1^2)^*, (\tilde{\lambda}_2^2)^*, \cdots, (\tilde{\lambda}_M^2)^*)^T,$  $\tilde{p}^* = (\tilde{p}_1^*, \tilde{p}_2^*, \cdots, \tilde{p}_{M-1}^*)^T$  and  $\tilde{\theta}^* = ((\tilde{p}^*)^T, (2\tilde{\lambda}^2)^*)^T, (\tilde{\mu}^*)^T)^T$ . Furthermore, Let  $g(\cdot)$  be the marginal density function of X,  $\nu_0(K) = \int K^2(z)dz$  and  $\kappa_2(K) = \int z^2 K(z)dz$ . Then the asymptotic bias and asymptotic variance of  $\tilde{\theta}^*$  are

$$\operatorname{bias}(\tilde{\theta}^*) = I^{-1}(x) \left\{ \frac{g'(x)\Lambda'_u(x|x)}{g(x)} + \frac{1}{2}\Lambda''_u(x|x) \right\} \kappa_2(K)h^2$$

and

$$\operatorname{Var}(\tilde{\theta}^*) = \frac{\nu_0(K)I^{-1}(x)}{g(x)},$$

respectively.

Under some regularity conditions,  $\tilde{\theta}^*$  has the asymptotic normal distribution. That is, it follows that

$$\sqrt{nh}\{\tilde{\theta}^* - \operatorname{bias}(\tilde{\theta}^*) + o(h)\} \to_L N\left(0, \frac{\nu_0(K)}{I(x)g(x)}\right),$$

where  $\rightarrow_L$  means the convergence in distribution.

The proofs of above results are similar to that of Theorem 2 in Huang, Li and Wang [5]. In this paper, our main aim is the EM algorithm of local estimations for the finite mixture of regression model with Laplace distribution.

#### 4 The EM Algorithm of Local Estimations

For a given point x, the EM algorithm is an effective method to maximize the local weighted log-likelihood function (3.1). In practice, we evaluate the unknown functions at a set of grid points over an interval of x, which requires us to maximize the local weighted log-likelihood function (3.1) at different grid points. First, we introduce component labels for each of the observation, and define a set of local weighted complete log-likelihood function with the same labels. Second, we estimate these labels in the E-step of the EM algorithm. In the M-step of the EM algorithm, we simultaneously update the estimated curves at all grid points for the same probabilistic label obtained in the E-step, which ensure that the resulting functional estimations are continuous and smooth at each iteration of the EM algorithm.

The mixture problem is formulated as an incomplete-data problem in the EM framework. The observed data  $(X_i, Y_i)$ s are viewed as being incomplete, and the unobserved Bernoulli random variables are introduced as following:

$$\xi_{ik} = \begin{cases} 1, & \text{if } (X_i, Y_i) \text{ is in the } k\text{th group,} \\ 0, & \text{otherwise.} \end{cases}$$

Let  $\xi_i = (\xi_{i1}, \xi_{i2}, \dots, \xi_{iM})^T$ , the associated component identity or label of  $(X_i, Y_i)$ . Then  $\{(X_i, Y_i, \xi_i), i = 1, 2, \dots, n\}$  are the complete data, and complete log-likelihood function corresponding to (3.1) is

$$\sum_{i=1}^{n} \sum_{k=1}^{M} \xi_{ik} \left[ \log p_k(X_i) - \log[2\lambda_k(X_i)] - \frac{|Y_i - \mu_k(X_i)|}{\lambda_k(X_i)} \right].$$

For  $x \in \{u_1, u_2, \dots, u_N\}$ , the set of grid points at which the unknown functions are to be evaluated. We define a local weighted complete log-likelihood function as

$$\sum_{i=1}^{n} \sum_{k=1}^{M} \xi_{ik} \left[ \log p_k - \log[2\lambda_k] - \frac{|Y_i - \mu_k|}{\lambda_k} \right] K_h(X_i - x).$$

Note that  $\xi_{ik}$ s do not depend on the choice of x. We have  $\mu_k^{(l)}(\cdot)$ ,  $\lambda_k^{(l)}(\cdot)$  and  $p_k^{(l)}(\cdot)$  in the *l*th cycle of the EM algorithm iteration. Then in the E-step of (l+1)th cycle, the expectation of the latent variable  $\xi_{ik}$  is given by

$$r_{ik}^{(l+1)} = \frac{p_k^{(l)}(X_i) \frac{1}{2\lambda_k^{(l)}(X_i)} \exp\left\{-\frac{|Y_i - \mu_k^{(l)}(X_i)|}{\lambda_k^{(l)}(X_i)}\right\}}{\sum\limits_{k=1}^M p_k^{(l)}(X_i) \frac{1}{2\lambda_k^{(l)}(X_i)} \exp\left\{-\frac{|Y_i - \mu_k^{(l)}(X_i)|}{\lambda_k^{(l)}(X_i)}\right\}}.$$
(4.1)

In the M-step of the (l+1)th cycle, we maximize

$$\sum_{i=1}^{n} \sum_{k=1}^{M} r_{ik}^{(l+1)} \left[ \log p_k - \log[2\lambda_k] - \frac{|Y_i - \mu_k|}{\lambda_k} \right] K_h(X_i - x)$$
(4.2)

for  $x = u_j, j = 1, 2, \dots, N$ . The maximization of equation (4.2) is equivalent to maximizing

$$\sum_{i=1}^{n} \sum_{k=1}^{M} r_{ik}^{(l+1)} \left[\log p_k\right] K_h(X_i - x)$$
(4.3)

and for  $k = 1, 2, \dots, M$ ,

$$\sum_{i=1}^{n} r_{ik}^{(l+1)} \left[ -\log[2\lambda_k] - \frac{|Y_i - \mu_k|}{\lambda_k} \right] K_h(X_i - x),$$
(4.4)

separately. For  $x \in \{u_1, u_2, \dots, u_N\}$ , the solution for maximization of equation (4.3) is

$$p_k^{(l+1)}(x) = \frac{\sum_{i=1}^n r_{ik}^{(l+1)} K_h(X_i - x)}{\sum_{i=1}^n K_h(X_i - x)}.$$
(4.5)

To obtain the solution for maximization of equation (4.4), we first fix the parameter  $\lambda_k$ . Denote  $\hat{\mu}_k$  be the solution for maximization of equation (4.4) with respect to  $\mu_k$ . Let

$$\mu_k^{(l+1)}(x) = \hat{\mu}_k. \tag{4.6}$$

Then let  $\mu_k^{(l+1)}(x)$  be fixed, the solution for maximization of equation (4.4) with respect to  $\lambda_k$  is

$$\lambda_k^{(l+1)}(x) = \frac{\sum_{i=1}^n r_{ik}^{(l+1)} |Y_i - \mu_k^{(l+1)}(x)| K_h(X_i - x)}{\sum_{i=1}^n r_{ik}^{(l+1)} K_h(X_i - x)}.$$
(4.7)

Furthermore, we update  $p_k^{(l+1)}(X_i)$ ,  $\mu_k^{(l+1)}(X_i)$  and  $\lambda_k^{(l+1)}(X_i)$ ,  $i = 1, 2, \dots, n$  by linearly interpolating  $p_k^{(l+1)}(u_j)$ ,  $\mu_k^{(l+1)}(u_j)$  and  $\lambda_k^{(l+1)}(u_j)$ ,  $j = 1, 2, \dots, N$ , respectively. We summarize the EM algorithm as the following.

### The EM Algorithm

**Initial value** Conduct a mixture of polynomial regressions with constant proportions and variance, and obtain the estimations of mean function  $\bar{\mu}_k(x)$ , variance  $\bar{\sigma}_k^2$ , and parameter  $\bar{p}_k$ . Set the initial values  $\mu_k^{(1)}(x) = \bar{\mu}_k$ ,  $\lambda_k^{(1)}(x) = \sqrt{\bar{\sigma}_k^2/2}$  and  $p_k^{(1)}(x) = \bar{p}_k$ .

**E-step** Use equation (4.1) to calculate  $r_{ik}^{(l+1)}$  for  $i = 1, 2, \dots, n$  and  $k = 1, 2, \dots, M$ . **M-step** For  $k = 1, 2, \dots, M$  and  $j = 1, 2, \dots, N$ , evaluate  $p_k^{(l+1)}(u_j)$  in (4.5),  $\mu_k^{(l+1)}(u_j)$ in (4.6) and  $\lambda_k^{(l+1)}(u_j)$  in (4.7). Further, we obtain  $p_k^{(l+1)}(X_i)$ ,  $\mu_k^{(l+1)}(X_i)$  and  $\lambda_k^{(l+1)}(X_i)$ using linear interpolation.

Iteratively update the E-step and the M-step with  $l = 2, 3, \dots$ , until the algorithm converges.

It is well known that the bandwidth selection can be tuned to optimize the performance of the estimated parametric functions. At the end of this section, we select the bandwidth of the local estimations for the parametric functions. We select bandwidth h via the Crossvalidation method, which is discussed in detail in [10].

#### 5 Ascent Property of the EM Algorithm

Note that the EM algorithm for constant parameters possesses an ascent property, which is a desired property. The EM algorithm for the parametric functions in this paper can be viewed as a generalization of the EM algorithm for constant parameters. So it is very interesting to discuss whether the EM algorithm we proposed still preserves the ascent property. Now we first give the following assumptions.

(A1) The sample  $\{(X_i, Y_i), i = 1, 2, \dots, n\}$  is independent and identically distribution from population (X, Y), and the support for X, denoted by  $\chi$ , is a compact subset of R.

(A2) The marginal density function g(x) of X is twice continuously differentiable and positive for all  $x \in \chi$ .

(A3) There exists a function M(y) with  $E[M(y)] < \infty$ , such that for all y, and all  $\theta$  in a neighborhood of  $\theta(x)$ , we have  $\left|\frac{\partial^3 l(\theta, y)}{\partial \theta_j \partial \theta_k \partial \theta_l}\right| < M(y)$ .

(A4) The parametric function  $\theta(x)$  has continuous second derivatives. Furthermore, for  $k = 1, 2, \dots, M, \lambda_k(x) > 0$  and  $p_k(x) > 0$  hold for all  $x \in \chi$ .

(A5) The kernel function  $K(\cdot)$  has a bounded support and satisfies that  $\int K(z)dz = 1$ ,  $\int zK(z)dz = 0$ ,  $\int z^2K(z)dz < \infty$ ,  $\int K^2(z)dz < \infty$  and  $\int |K^3(z)|dz < \infty$ .

Let  $\theta^{(l)} = (p^{(l)}(\cdot), 2\lambda^{2(l)}(\cdot), \mu^{(l)}(\cdot))$  be the estimated functions in the *l*th cycle of the EM algorithm proposed. The local weighted log-likelihood function (3.2) is rewritten as

$$l_n(\theta) = \sum_{i=1}^n l(\theta, Y_i) K_h(X_i - x).$$
 (5.1)

Then we have the following theorem.

**Theorem 4.1** Assume that conditions (A1)–(A5) hold. For any given point x, suppose that  $\theta^{(l)}(\cdot)$  has a continuous first derivative, and  $h \to 0$  and  $nh \to \infty$  as  $n \to \infty$ . Then we

have

$$\liminf_{n \to \infty} \frac{1}{n} \left[ l_n(\theta^{(l+1)}(x)) - l_n(\theta^{(l)}(x)) \right] \ge 0$$
(5.2)

in probability.

**Proof** Suppose that the unobserved data  $\{Z_i, i = 1, 2, \dots, n\}$  is a random sample from population Z. Then, the complete data  $\{(X_i, Y_i, Z_i), i = 1, 2, \dots, n\}$  can be viewed as a sample from (X, Y, Z). Let  $h(y, k | \theta(x))$  be the joint distribution of (Y, Z) given X = x, and g(x) be the marginal density of X. Conditioning on X = x, Y follows a distribution  $\eta(y|\theta(x))$ . Then, the local weighted log-likelihood function (3.2) can be rewritten as

$$l_n(\theta) = \sum_{i=1}^n \log[\eta(Y_i|\theta)] K_h(X_i - x).$$
(5.3)

The conditional probability of Z = k given y and  $\theta$  is

$$f(k|y,\theta) = h(y,k|\theta)/\eta(y|\theta) = \frac{p_k \frac{1}{2\lambda_k} \exp\left\{-\frac{|y-\mu_k|}{\lambda_k}\right\}}{\sum\limits_{k=1}^M p_k \frac{1}{2\lambda_k} \exp\left\{-\frac{|y-\mu_k|}{\lambda_k}\right\}}.$$
(5.4)

For given  $\theta^{(l)}(X_i)$ ,  $i = 1, 2, \dots, n$ , it is clear that  $\int f(k|Y_i, \theta^{(l)}(X_i))dk = 1$ . Then we have

$$l_{n}(\theta) = \sum_{i=1}^{n} \log[\eta(Y_{i}|\theta)] \left[ \int f(k|Y_{i}, \theta^{(l)}(X_{i})) dk \right] K_{h}(X_{i} - x)$$
  
$$= \sum_{i=1}^{n} \left\{ \int \log[\eta(Y_{i}|\theta)] [f(k|Y_{i}, \theta^{(l)}(X_{i}))] dk \right\} K_{h}(X_{i} - x).$$
(5.5)

By equation (5.4), we have

$$\log[\eta(Y_i|\theta)] = \log[h(Y_i, k|\theta)] - \log[f(k|Y_i, \theta)].$$
(5.6)

Thus we have

$$l_{n}(\theta) = \sum_{i=1}^{n} \log[\eta(Y_{i}|\theta)] \left[ \int f(k|Y_{i}, \theta^{(l)}(X_{i})) dk \right] K_{h}(X_{i} - x)$$
  
$$= \sum_{i=1}^{n} \left\{ \int \log[h(Y_{i}, k|\theta)] [f(k|Y_{i}, \theta^{(l)}(X_{i}))] dk \right\} K_{h}(X_{i} - x)$$
  
$$- \sum_{i=1}^{n} \left\{ \int \log[f(k|Y_{i}, \theta)] [f(k|Y_{i}, \theta^{(l)}(X_{i}))] dk \right\} K_{h}(X_{i} - x), \quad (5.7)$$

where  $\theta^{(l)}(X_i)$  is the estimation of  $\theta(X_i)$  at the *l*th iteration. Taking expectation leads to calculating equation (4.1). In the M-step of the EM algorithm, we update  $\theta^{(l+1)}(x)$  such

that

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \int \log[h(Y_i, k|\theta^{(l+1)}(x))][f(k|Y_i, \theta^{(l)}(X_i))]dk \right\} K_h(X_i - x)$$

$$\geq \frac{1}{n} \sum_{i=1}^{n} \left\{ \int \log[h(Y_i, k|\theta^{(l)}(x))][f(k|Y_i, \theta^{(l)}(X_i))]dk \right\} K_h(X_i - x).$$

It suffices to show that

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\int \log\left[\frac{f(k|Y_{i},\theta^{(l+1)}(x))}{f(k|Y_{i},\theta^{(l)}(x))}\right]f(k|Y_{i},\theta^{(l)}(X_{i}))dk\right\}K_{h}(X_{i}-x)\leq0$$
(5.8)

in probability. Let

$$L_{n1} = \frac{1}{n} \sum_{i=1}^{n} \phi(Y_i, X_i) K_h(X_i - x),$$

where

$$\phi(Y_i, X_i) = \int \log\left[\frac{f(k|Y_i, \theta^{(l+1)}(x))}{f(k|Y_i, \theta^{(l)}(x))}\right] f(k|Y_i, \theta^{(l)}(X_i)) dk$$

By using assumptions (A1)–(A4), we have  $f(k|Y, \theta^{(l)}(x)) > a > 0$  for some small value a, and  $E\{[\phi(Y, X)]^2\} < \infty$ . Then by Assumption (A5) and Theorem A in [11], we have

$$\sup_{J} |L_{n1} - g(x)E[\phi(Y, X)]| = o_p(1),$$

where J is a compact interval on which the density of X is bounded below from 0. The proof follows that

$$\begin{split} E[\phi(Y,x)] &= E\left\{\int \log\left[\frac{f(Z|Y,\theta^{(l+1)}(x))}{f(Z|Y,\theta^{(l)}(x))}\right]f(k|Y,\theta^{(l)}(x))dk\right\}\\ &\leq E\left\{\log\left[\int\left[\frac{f(Z|Y,\theta^{(l+1)}(x))}{f(Z|Y,\theta^{(l)}(x))}\right]f(k|Y,\theta^{(l)}(x))dk\right]\right\}.\end{split}$$

This completes the proof of Theorem 5.1.

#### References

- Lindsay B. Mixture models: Theory, geometry and applications[M]. Hayward, CA: Insti. Math. Stat., 1995.
- [2] Cai Z, Jiang B. Almost unbiased ridge estimator for a mixed-effect coefficient linear model[J]. J. Math., 2013, 33(2): 354–358.
- [3] Lu J. Minimum norm estimation for mixed fractional Ornstein-Uhlenbeck type process[J]. J. Math., 2014, 34(3): 597–602.
- [4] Hurn M, Justel A, Robert C. Estimating mixtures of regressions[J]. J. Comput. Graph. Stati., 2003, 12: 55–79.

- [5] Huang M, Li R, Wang S. Nonparametric mixture of regression models[J]. J. Amer. Stati. Assoc., 2013, 108: 929–941.
- [6] Wang Y, Wang J. EM algorithm of estimation of parameter in two-parameter exponential model under type-i censoring sample[J]. J. Henan Normal Univer. (Natur. Sci. Dition), 2012, 40(5): 28–30.
- [7] Yao W. A note on EM algorithm for mixture models[J]. Stati. Prob. Lett., 2013, 83: 519-526.
- [8] Henning C. Identifiability of models for Clusterwise linear regression[J]. J. Class., 2000, 17: 273–296.
- [9] Fan J, Gijbels I. Local polynomial modelling and its applications[M]. London: Chapman Hall, 1996.
- [10] Fan J, Gijbels I. Variable bandwidth and local linear regression[J]. Ann. Stati., 1992, 20: 2008–2036.
- [11] Mack Y, Silverman B. Weak and strong uniform consistency of kernel regression estimates [J]. Prob. The. Rel. Fieids, 1982, 61: 405–415.

# 有限混合Laplace分布回归模型局部估计的EM算法

## 王继霞, 汪春峰, 苗 雨

(河南师范大学数学与信息科学学院,河南新乡 453007)

**摘要**: 本文研究了一类有限混合Laplace分布回归模型的局部极大似然估计问题.利用核回归方法和 最大化局部加权似然函数的EM算法,获得了参数函数的局部极大似然估计量,并讨论了它们的渐近偏差,渐 近方差和渐近正态性.推广了有限混合回归模型下局部非参数估计的结果.

关键词: 有限混合模型; Laplace分布; EM算法; 局部极大似然估计; 核回归 MR(2010) 主题分类号: 62F10; 62J02 中图分类号: O212.1