MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Strong consistency of MLE for finite mixtures of location-scale distributions when the scale parameters are exponentially small

Kentaro Tanaka and Akimichi Takemura

METR 2003–35 October 2003

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

Strong consistency of MLE for finite mixtures of location-scale distributions when the scale parameters are exponentially small

Kentaro Tanaka and Akimichi Takemura
Department of Mathematical Informatics
Graduate School of Information Science and Technology
University of Tokyo

October, 2003

Abstract

In finite mixture of location-scale distributions maximum likelihood estimator does not exist because of the unboundedness of the likelihood function when the scale parameter of some mixture component approaches zero. In order to study the strong consistency of maximum likelihood estimator, we consider the case that the scale parameters of the component distributions are restricted from below by c_n, where $\{c_n\}$ is a sequence of positive real numbers which tends to zero as the sample size n increases. We prove that under mild regularity conditions maximum likelihood estimator is strongly consistent if the scale parameters are restricted from below by $c_n = \exp(-n^d)$, $0 < d < 1$.

1 Introduction

In some finite mixture distributions maximum likelihood estimator does not exist. Let us consider the following example. Denote a normal mixture distribution with M components and parameter $\theta = (\alpha_1, \mu_1, \sigma_1^2, \ldots, \alpha_M, \mu_M, \sigma_M^2)$ by

$$f(x; \theta) = \sum_{m=1}^{M} \alpha_m \phi_m(x; \mu_m, \sigma_m^2),$$

where α_m $(m = 1, \ldots, M)$ are nonnegative real numbers that sum to one and $\phi_m(x; \mu_m, \sigma_m^2)$ are normal densities. Let $x_1, \ldots, x_n$ denote a random sample of size $n \geq 2$ from the density $f(x; \theta_0)$, where θ_0 is the true parameter. The log likelihood function is

$$\sum_{i=1}^{n} \log f(x_i; \theta) = \sum_{i=1}^{n} \log \left\{ \sum_{m=1}^{M} \alpha_m \phi_m(x_i; \mu_m, \sigma_m^2) \right\}.$$

If we set $\mu_1 = x_1$, then the likelihood tends to infinity as $\sigma_1^2 \to 0$. Thus maximum likelihood estimator does not exist.

But when we restrict $\sigma_m \geq c$ $(m = 1, \dots, M)$ by some positive real constant c, we can avoid the divergence of the likelihood. Furthermore, it can be shown that the maximum likelihood estimator is strongly consistent under the restriction.

On the other hand, the smaller σ_1^2 is, the less contribution $\phi_1(x; \mu_1, \sigma_1^2)$ makes to the likelihood at $x_2, \dots, x_n$. Therefore an interesting question here is whether we can decrease the bound $c = c_n$ to zero with the sample size n and yet guarantee the strong consistency of the maximum likelihood estimator. If this is possible, the further question is how fast c_n can decrease to zero. This question is similar to the (so far open) problem stated in Hathaway(1985) [1], which treats mixtures of normal distributions with constraints imposed on the ratios of variances while our restriction is imposed on variances themselves. See also a discussion in section 3.8.1 of McLachlan and Peel(2000) [3].

In the above example, the normality of the component distributions is not essential and the same difficulty exists for finite mixture of general location-scale distributions such as mixtures of uniform distributions. Let b_m $(m = 1, \dots, M)$ denote the scale parameters of the component distributions and consider the restriction $b_m \geq c_n$ $(m = 1, \dots, M)$. Then a question of interest here is whether we can decrease the bound c_n to zero.

For the case of mixture of uniform distributions, in Tanaka and Takemura [2] we proved the the maximum likelihood estimator is strongly consistent if $c_n = \exp(-n^d)$, $0 < d < 1$. Here d can be arbitrarily close to 1 but fixed. In this paper, we prove that the same result holds for general finite mixture of location-scale distributions under very mild regularity conditions (assumptions 1–4 below) on the component densities. As discussed in section 5 the normal density satisfies the regularity conditions and our result implies that MLE is strongly consistent for the normal mixture if $\sigma_m \geq c_n = \exp(-n^d)$, $0 < d < 1$, $m = 1, \dots, M$.

After establishing some relevant lemmas, the proof of this paper follows the same line of arguments as in Tanaka and Takemura [2]. However compared to the case of uniform mixtures, the proof for the general case in this paper is much longer and more subtle with many additional constants we have to keep track of. Therefore in this paper we repeat the arguments in Tanaka and Takemura [2] and give a self-contained proof of the strong consistency of MLE for the general case.

The organization of the paper is as follows. In section 2 we summarize some preliminary results. In section 3 we state our main results in theorems 4 and 5. Section 4 is devoted to the proof of theorems and lemmas. Finally in section 5 we give some discussions.

2 Preliminaries on identifiability of mixture distributions, strong consistency and regularity conditions

A mixture of M densities with parameter $\theta = (\alpha_1, a_1, b_1, \dots, \alpha_M, a_M, b_M)$ is defined by

$$f(x;\theta) \equiv \sum_{m=1}^{M} \alpha_m f_m(x; a_m, b_m),$$

where α_m, $m = 1, \dots, M$, called the mixing weights, are nonnegative real numbers that sum to one and $f_m(x; a_m, b_m)$, called the components of the mixture, are location-scale densities with location parameter a_m and scale parameter b_m. Remark that $f_m(x; a_m, b_m)$ may even belong to different families each other. For example, $f_1(x; a_1, b_1)$ is a normal density, $f_2(x; a_2, b_2)$ is a t-density, $f_3(x; a_3, b_3)$ is a uniform density, etc. Let Θ denote the parameter space. For convenience, we write (a_m, b_m) as η_m.

In general, a parametric family of distributions is identifiable if different values of parameter designate different distributions. In mixtures of distributions, different parameters may designate the same distribution. For example, if $\alpha_1 = 0$, then for all parameters which differ only in η_1, we have the same distribution. Thus mixtures of distributions are not identifiable. Therefore we have to carefully define strong consistency of estimator. The following definition is essentially the same as Redner's(1981) [5]. We assume that the parameter space Θ is a subset of Euclidean space and $\mathrm{dist}(\theta, \theta')$ denotes the Euclidean distance between $\theta, \theta' \in \Theta$. Furthermore we define

$$\mathrm{dist}(U, V) \equiv \inf_{\theta \in U} \inf_{\theta' \in V} \mathrm{dist}(\theta, \theta')$$

for $U, V \subset \Theta$.

Definition 1. (*strongly consistent estimator*)
Let Θ_0 denote the set of true parameters

$$\Theta_0 \equiv \{\theta \in \Theta \mid f(x;\theta) = f(x;\theta_0) \quad a.e.\ x\},$$

where θ_0 is one of parameters designating the true distribution, and let

$$\Theta(\hat\theta) \equiv \{\theta \in \Theta \mid f(x;\theta) = f(x;\hat\theta) \quad a.e.\ x\}.$$

An estimator $\hat\theta_n$ is strongly consistent if

$$\mathrm{Prob}\left(\lim_{n\to\infty} \mathrm{dist}(\Theta(\hat\theta_n), \Theta_0) = 0\right) = 1.$$

Next, we give regularity conditions for strong consistency of maximum likelihood estimator.

Assumption 1. *There exist real constants $v_0, v_1 > 0$ and $\beta > 1$ such that*

$$f_m(x; a_m = 0, b_m = 1) \le \min\{v_0\ ,\ v_1 \cdot |x|^{-\beta}\} \tag{1}$$

for all m.

This assumption means that f_m $(m = 1, \ldots, M)$ are bounded and their tails decrease to zero faster than or equal to $|x|^{-\beta}$, which is a very mild condition.

The following three regularity conditions are standard conditions assumed in discussing strong consistency of MLE. Let Γ denote any compact subset of Θ.

Assumption 2. *For $\theta \in \Theta$ and any positive real number r, let*

$$f(x;\theta,r) \quad \equiv \sup_{\mathrm{dist}(\theta',\theta)\le r} f(x;\theta').$$

For each $\theta \in \Gamma$ and sufficiently small r, $f(x;\theta,r)$ is measurable.

Assumption 3. *For each $\theta \in \Gamma$, if $\lim_{n\to\infty}\theta_n = \theta$, then $\lim_{n\to\infty} f(x;\theta_n) = f(x;\theta)$ except on a set which is a null set and does not depend on the sequence $\{\theta_n\}_{n=1}^{\infty}$.*

Assumption 4.

$$\int |\log f(x;\theta_0)|\, f(x;\theta_0)\mathrm{d}x < \infty. \tag{2}$$

The next theorem can be proved in the same way as in Wald(1949)[6], Redner(1981) [5].

Theorem 1. *Suppose that assumptions 2, 3 and 4 are satisfied. Let S be any closed subset of Γ not intersecting Θ_0. Then*

$$\mathrm{Prob}\left(\lim_{n\to\infty} \frac{\sup_{\theta\in S} f(x_1;\theta)\times\cdots\times f(x_n;\theta)}{f(x_1;\theta_0)\times\cdots\times f(x_n;\theta_0)} = 0\right) = 1\ .$$

The following theorem has been proved by Wald(1949) [6].

Theorem 2. (Wald(1949) [6]) *Let $\tilde{\theta}_n$ be any function of the observations $x_1,\ldots,x_n$ such that*

$$\forall n, \quad \prod_{i=1}^{n} \frac{f(x_i;\tilde{\theta}_n)}{f(x_i;\theta_0)} \ge \delta > 0,$$

then $\tilde{\theta}_n$ is strongly consistent in the sense of definition 1.

Assume that Γ contains the true parameter. If assumptions 2, 3 and 4 are satisfied, then it is readily verified by theorems 1 and 2 that the maximum likelihood estimator restricted to Γ is strongly consistent.

We also state Okamoto's inequality, which will be used in our proof in section 4.3.

Theorem 3. (Okamoto(1958) [4]) *Let Z be a random variable following a binomial distribution* $\mathrm{Bin}(n,p)$*. Then for $\delta > 0$*

$$\mathrm{Prob}\left(\frac{Z}{n} - p \ge \delta\right) < \exp\left(-2n\delta^2\right). \tag{3}$$

3 Main results

Let $\mathscr{I}(K) = \{i_1, i_2, \ldots, i_K\}$ be a subset of $\{1, 2, \ldots, M\}$ with K elements. Let

$$\mathscr{G}_{\mathscr{I}(K)} \equiv \{ \sum_{m \in \mathscr{I}(K)} \alpha_m f_m(x; \eta_m) \mid \sum_{m \in \mathscr{I}(K)} \alpha_m \leq 1 \ , \ \alpha_m \geq 0 \}$$

which is a set of subprobability measures consisting of less than or equal to K components. Because $f_m(x; a_m, b_m)$, $(m = 1, \ldots, K)$, may belong to different families each other, $\mathscr{G}_{\mathscr{I}(K)}$ is not necessarily equal to $\mathscr{G}_{\{1,2,\ldots,K\}}$. However for notational simplicity and without essential loss of generality, in the following we consider only $\{1, 2, \ldots, K\}$ as a subset of $\{1, 2, \ldots, M\}$ with K elements by replacing $i_1 = 1, i_2 = 2, \ldots$, and write $\mathscr{G}_{\{1,2,\ldots,K\}}$ simply as $\mathscr{G}_K$.

Let

$$\mathfrak{A}^{(K)} = \{(\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^K \mid \sum_{m=1}^{K} \alpha_m \leq 1 \ , \ \alpha_m \geq 0\} \quad , \quad (0 \leq K \leq M)$$

and let Ω_m be the parameter space of the m-th component. Then the parameter space of subprobability measures consisting of less than or equal to K components can be written as $\Theta^{(K)} = \mathfrak{A}^{(K)} \times \prod_{m=1}^{K} \Omega_m$.

Let $E_0[\cdot]$ denote the expectation under θ_0. The following theorem is essential to our argument and it is of some independent interest.

Theorem 4. *Suppose that assumptions 1–4 are satisfied and the true model can be represented only by the model which consists of M components. Then there exist real constants $\lambda, \kappa > 0$ such that*

$$E_0 \left[\log \{g + \kappa\}\right] + \lambda < E_0[\log f(x; \theta_0)] \tag{4}$$

for all $g \in \mathscr{G}_L$, $(L \leq M - 1)$.

We now state the main theorem of this paper.

Theorem 5. *Suppose that assumptions 1–4 are satisfied and the true model can be represented only by the model which consists of M components. Let $c_0 > 0$ and $0 < d < 1$. If $c_n = c_0 \cdot \exp(-n^d)$ and*

$$\Theta_n \equiv \{\theta \in \Theta \mid b_m \geq c_n \ , \ (m = 1, \ldots, M)\},$$

then the maximum likelihood estimator restricted to Θ_n is strongly consistent.

4 Proofs

In this section, we prove theorems stated in section 3. The organization of this section is as follows. First in subsection 4.1 we prove theorem 4 which is used to prove theorem 5. Next we prove two lemmas in subsection 4.2. Those two lemmas are also used to prove theorem 5 and can be described separately from the proof of theorem 5. Finally we prove theorem 5 in subsection 4.3.

4.1 Proof of the theorem 4

We prove theorem 4 by contradiction. Suppose that (4) does not hold. Then for any $\lambda, \kappa > 0$, there exists $g \in \mathscr{G}_L$ such that

$$E_0[\log{(g+\kappa)}] + \lambda \geq E_0[\log f(x;\theta_0)].$$

Here, let $\{\lambda_n\}, \{\kappa_n\}$ be positive sequences which decrease to zero. Then for each $\lambda_n, \kappa_n > 0$, there exists $g_n \in \mathscr{G}_L$ such that

$$E_0[\log{(g_n+\kappa_n)}] + \lambda_n \geq E_0[\log f(x;\theta_0)].$$

It follows that

$$\liminf_{n\to\infty} E_0[\log{(g_n+\kappa_n)}] + \lambda_n \geq E_0[\log f(x;\theta_0)]. \tag{5}$$

Now g_n can be written as

$$g_n = \sum_{m=1}^{L} \alpha_m^{(n)} f_m(x; a_m^{(n)}, b_m^{(n)}).$$

Let

$$a'^{(n)}_m \equiv \arctan{(a_m^{(n)})}, \qquad b'^{(n)}_m \equiv \arctan{(b_m^{(n)})}.$$

Then $\{\alpha_1^{(n)}, a'^{(n)}_1, b'^{(n)}_1, \dots, \alpha_L^{(n)}, a'^{(n)}_L, b'^{(n)}_L\}_{n=1}^{\infty}$ are regarded as a sequence in the following compact set.

$$\begin{aligned} &0 \leq \alpha_m \leq 1 \quad , \quad \sum_{m=1}^{L} \alpha_m \leq 1, \\ &-\frac{\pi}{2} \leq a'^{(n)}_m \leq \frac{\pi}{2}, \qquad 0 \leq b'^{(n)}_m \leq \frac{\pi}{2}. \end{aligned} \tag{6}$$

Therefore there exists a subsequence of $\{\alpha_1^{(n)}, a'^{(n)}_1, b'^{(n)}_1, \dots, \alpha_L^{(n)}, a'^{(n)}_L, b'^{(n)}_L\}_{n=1}^{\infty}$ that converges to a point in the set (6). For notational simplicity and without loss of generality, we replace the original sequence with this subsequence, because (5) holds for this subsequence as well. Furthermore we reorder components by asymptotic behavior of their parameters. First, we choose components such that $b'^{(n)}_m \to 0$, and order them as $\{1, \dots, K_b^0\}$ for simplicity. Second, from the remainder, we choose components such that $b'^{(n)}_m \to 1/2$, and order them as $\{K_b^0 + 1, \dots, K_b^\infty\}$. Third, from the remainder, we choose components such that $|a'^{(n)}_m| \to 1/2$, and order them as $\{K_b^\infty + 1, \dots, K_a^\infty\}$. Finally, we order the components that remain after this procedure as $\{K_a^\infty + 1, \dots, L\}$. Let

$$g_\infty(x) = \sum_{m=K_a^\infty+1}^{L} \alpha_m^{(\infty)} f_m(x; a_m^{(\infty)}, b_m^{(\infty)}) \in \mathscr{G}_{L'} \quad , \quad (L' = L - K_a^\infty),$$

where $(\alpha_m^{(\infty)} a_m^{(\infty)}, b_m^{(\infty)}) = \lim_{n\to\infty}(\alpha_m^{(n)}, a_m^{(n)}, b_m^{(n)})$ is finite for $m > K_a^\infty$. By considering the sequence $\{\alpha_m^{(n)}, a_m^{(n)}, b_m^{(n)}\}_{n=n_0}^\infty$ where n_0 is sufficiently large, and replacing n by $n - n_0$ if necessary, we can assume without loss of generality that there exist sufficiently small real constants $\kappa_0 > 0$ and $c_0 > 0$ such that

$$\begin{aligned}
&E_0[\log\{f(x;\theta_0)\}] - E_0\left[\log\left\{g_\infty + \kappa_0\right\}\right] > 0 \ , \\
&\frac{\kappa_0}{4} < \frac{v_0}{c_0(M+1)} \ , \\
&b_m^{(n)} < c_0 \ , \qquad (1 \le m \le K_b^0) \ , \forall n, \\
&b_m^{(n)} > \frac{4Mv_0}{\kappa_0} \ , \qquad (K_b^0 + 1 \le m \le K_b^\infty) \ , \forall n, \\
&c_0 \le b_m^{(n)} \le \frac{4Mv_0}{\kappa_0} \ , \qquad (K_b^\infty + 1 \le m \le L), \ \forall n.
\end{aligned} \tag{7}$$

From assumption 1, $f_m(x; a_m, b_m)$ is bounded by a step function

$$f_m(x; a_m, b_m) \le 1_{[a_m - \nu(b_m), a_m + \nu(b_m))}(x) \cdot v(b_m) + \frac{\kappa_0}{4M},$$

where $\nu(b_m)$ and $v(b_m)$ are defined as

$$\begin{aligned}
\nu(b_m) &\equiv \left(\frac{4Mv_1}{\kappa_0}\right)^{\frac{1}{\beta}} (b_m)^{\left(\frac{\beta-1}{\beta}\right)} \ , \\
v(b_m) &\equiv \frac{v_0}{b_m} \ .
\end{aligned}$$

See figure 1. Hence, from (7), following inequality holds.

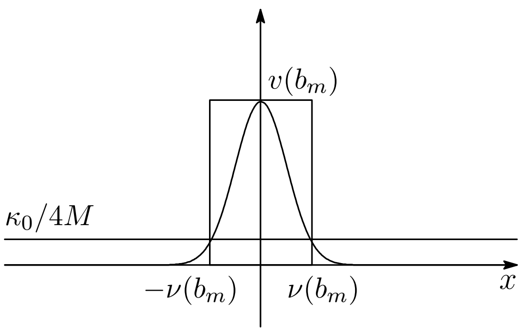

Figure 1: Each component is bounded by step function.

$$\sum_{m=1}^{L} \alpha_m^{(n)} f_m(x; a_m^{(n)}, b_m^{(n)}) \ \le \ \sum_{m=1}^{L} f_m(x; a_m^{(n)}, b_m^{(n)})$$

$$\leq \sum_{m=1}^{L} 1_{[a_m^{(n)}-\nu(b_m^{(n)}),a_m^{(n)}+\nu(b_m^{(n)}))}(x)\cdot v(b_m^{(n)})+\frac{\kappa_0}{4}. \qquad (8)$$

Let $J(\theta^{(n)}) \equiv \bigcup_{m=1}^{K_b^0}[a_m^{(n)}-\nu(b_m^{(n)}), a_m^{(n)}+\nu(b_m^{(n)}))$. Then for $x \in J(\theta^{(n)})$ the terms on the right hand side of (8) can be written as

$$1_{J(\theta^{(n)})}(x)\cdot\left\{\sum_{m=1}^{L} 1_{[a_m^{(n)}-\nu(b_m^{(n)}),a_m^{(n)}+\nu(b_m^{(n)}))}(x)\cdot v(b_m^{(n)})+\frac{\kappa_0}{4}\right\} = \sum_{t=1}^{T(\theta^{(n)})} H(J_t(\theta^{(n)}))\cdot 1_{J_t(\theta^{(n)})}(x),$$

where $J_t(\theta^{(n)})$, $(t=1,\ldots,T(\theta^{(n)}))$, are disjoint intervals, $[a_m^{(n)}-\nu(b_m^{(n)}), a_m^{(n)}+\nu(b_m^{(n)}))$, $(m=1,\ldots,K_b^0)$, are unions of some of $J_t(\theta^{(n)})$'s and $H(J_t(\theta^{(n)}))$ is defined by

$$H(J_t(\theta^{(n)})) = \sum_{m=1}^{L} 1_{[a_m^{(n)}-\nu(b_m^{(n)}),a_m^{(n)}+\nu(b_m^{(n)}))}(x)\cdot v(b_m^{(n)})+\frac{\kappa_0}{4} \quad , \quad (x\in J_t(\theta^{(n)})) \qquad (9)$$

for each $J_t(\theta^{(n)})$. Note that the total number $T(\theta^{(n)})$ of $J_t(\theta^{(n)})$'s satisfies $T(\theta^{(n)}) \leq 2M$, because the change of the height can only occur at $a_m^{(n)}-\nu(b_m^{(n)})$ or $a_m^{(n)}+\nu(b_m^{(n)})$. Let $W(J_t(\theta^{(n)}))$ denote the length of $J_t(\theta^{(n)})$. For convenience we determine the order of $J_1(\theta^{(n)}),\ldots,J_{T(\theta^{(n)})}(\theta^{(n)})$ such that

$$H(J_1(\theta^{(n)})) \leq H(J_2(\theta^{(n)})) \leq \ldots \leq H(J_{T(\theta^{(n)})}(\theta^{(n)})).$$

From the considerations above we have

$$\begin{aligned}
&E_0\left[\log\left\{\sum_{m=1}^{L}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})+\kappa_n\right\}+\lambda_n\right]\\
&= E_0\left[1_{J(\theta^{(n)})}(x)\cdot\log\left\{\sum_{m=1}^{L}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})+\kappa_n\right\}\right]\\
&\quad +E_0\left[1_{\mathbb{R}\setminus J(\theta^{(n)})}(x)\cdot\log\left\{\sum_{m=1}^{L}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})+\kappa_n\right\}\right]+\lambda_n\\
&\leq \int 1_{J(\theta^{(n)})}(x)\cdot\log\left\{\sum_{t=1}^{T(\theta^{(n)})} H(J_t(\theta^{(n)}))\cdot 1_{J_t(\theta^{(n)})}(x)+\kappa_n\right\} f(x;\theta_0)\mathrm{d}x\\
&\quad +\int 1_{\mathbb{R}\setminus J(\theta^{(n)})}(x)\cdot\log\left\{\sum_{m=K_b^0+1}^{L}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})+\frac{\kappa_0}{4}+\kappa_n\right\} f(x;\theta_0)\mathrm{d}x+\lambda_n.
\end{aligned} \qquad (10)$$

Now we evaluate the first term on the right hand side of (10). From (7) we have

$$H(J_t(\theta^{(n)}))-\frac{\kappa_0}{4} \;>\; H(J_t(\theta^{(n)}))-\frac{v_0}{c_0(M+1)}$$

$$> \quad H(J_t(\theta^{(n)})) - \frac{H(J_t(\theta^{(n)}))}{M+1} = \frac{M \cdot H(J_t(\theta^{(n)}))}{M+1}.$$

In (9) at any $x \in J_t(\theta^{(n)})$, $H(J_t(\theta^{(n)})) - \kappa_0/4$ consists of at most M components. Thus there exists at least one component such that

$$v(b_m^{(n)}) \geq \frac{1}{M} \cdot \left(H(J_t(\theta^{(n)})) - \frac{\kappa_0}{4} \right) \geq \frac{H(J_t(\theta^{(n)}))}{M+1}.$$

Therefore we obtain

$$\begin{aligned} \nu(b_m^{(n)}) &= \left(\frac{4Mv_1}{\kappa_0}\right)^{\frac{1}{\beta}} (b_m^{(n)})^{\left(\frac{\beta-1}{\beta}\right)} \\ &= \left(\frac{4Mv_1}{\kappa_0}\right)^{\frac{1}{\beta}} \left(\frac{v_0}{v(b_m^{(n)})}\right)^{\left(\frac{\beta-1}{\beta}\right)} \\ &\leq \left(\frac{4Mv_1}{\kappa_0}\right)^{\frac{1}{\beta}} \left(\frac{v_0(M+1)}{H(J_t(\theta^{(n)}))}\right)^{\left(\frac{\beta-1}{\beta}\right)} . \end{aligned}$$

From this and $W(J_t(\theta^{(n)})) \leq 2\nu(b_m^{(n)})$, it follows that

$$\lim_{n\to\infty} W(J_t(\theta^{(n)})) \cdot \log \{H(J_t(\theta^{(n)}))\} = 0.$$

Let

$$u_0 \equiv \sup_x f(x;\theta_0).$$

Then we have

$$\begin{aligned} &\int 1_{J(\theta^{(n)})}(x) \cdot \log \left\{ \sum_{t=1}^{T(\theta^{(n)})} H(J_t(\theta^{(n)})) \cdot 1_{J_t(\theta^{(n)})}(x) + \kappa_n \right\} f(x;\theta_0)\mathrm{d}x \\ \leq &\sum_{t=1}^{T(\theta^{(n)})} W(J(\theta^{(n)})) \cdot \log \left\{ H(J_t(\theta^{(n)})) + \kappa_n \right\} \cdot u_0 \\ \to &\ 0 , \qquad (n \to \infty) . \end{aligned} \tag{11}$$

Next we evaluate the second term on the right hand side of (10). Let

$$A^{(n)} \equiv \min_{K_b^\infty+1 \leq m \leq K_a^\infty} \left\{ \min \{ |a_m^{(n)} + \nu(b_m^{(n)})|, |a_m^{(n)} - \nu(b_m^{(n)})| \} \right\}.$$

Then the following inequality holds.

$$\int 1_{\mathbb{R}\setminus J(\theta^{(n)})}(x) \cdot \log \left\{ \sum_{m=K_b^0+1}^{L} \alpha_m^{(n)} f_m(x; a_m^{(n)}, b_m^{(n)}) + \frac{\kappa_0}{4} + \kappa_n \right\} f(x;\theta_0)\mathrm{d}x$$

$$
\begin{aligned}
= & \int 1_{\mathbb{R}\setminus J(\theta^{(n)})}(x)\cdot\log\left\{\sum_{m=K_b^0+1}^{K_b^\infty}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})+\sum_{m=K_b^\infty+1}^{K_a^\infty}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})\right.\\
& \qquad\left.+\sum_{m=K_a^\infty+1}^{L}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})+\frac{\kappa_0}{4}+\kappa_n\right\}f(x;\theta_0)\mathrm{d}x\\
\leq & \int 1_{\mathbb{R}\setminus J(\theta^{(n)})}(x)\cdot\log\left\{\sum_{m=K_b^\infty+1}^{K_a^\infty}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})\right.\\
& \qquad\left.+\sum_{m=K_a^\infty+1}^{L}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})+\frac{\kappa_0}{2}+\kappa_n\right\}f(x;\theta_0)\mathrm{d}x\\
\leq & \int 1_{[-A^{(n)},A^{(n)}]\setminus J(\theta^{(n)})}(x)\cdot\log\left\{\sum_{m=K_a^\infty+1}^{L}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})+\frac{3\kappa_0}{4}+\kappa_n\right\}f(x;\theta_0)\mathrm{d}x\\
& +\int 1_{\{(-\infty,-A^{(n)})\bigcup(A^{(n)},\infty)\}\setminus J(\theta^{(n)})}(x)\cdot\log\left\{\sum_{m=K_b^\infty+1}^{K_a^\infty}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})\right.\\
& \qquad\left.+\sum_{m=K_a^\infty+1}^{L}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})+\frac{\kappa_0}{2}+\kappa_n\right\}f(x;\theta_0)\mathrm{d}x. \qquad (12)
\end{aligned}
$$

By bounded convergence theorem, we obtain

$$
\begin{aligned}
& \int 1_{[-A^{(n)},A^{(n)}]\setminus J(\theta^{(n)})}(x)\cdot\log\left\{\sum_{m=K_a^\infty+1}^{L}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})+\frac{3\kappa_0}{4}+\kappa_n\right\}f(x;\theta_0)\mathrm{d}x\\
& \to\int\log\left\{g_\infty+\frac{\kappa_0}{2}\right\}f(x;\theta_0)\mathrm{d}x, \qquad (13)
\end{aligned}
$$

and

$$
\begin{aligned}
& \int 1_{\{(-\infty,-A^{(n)})\bigcup(A^{(n)},\infty)\}\setminus J(\theta^{(n)})}(x)\cdot\log\left\{\sum_{m=K_b^\infty+1}^{K_a^\infty}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})\right.\\
& \qquad\left.+\sum_{m=K_a^\infty+1}^{L}\alpha_m^{(n)} f_m(x;a_m^{(n)},b_m^{(n)})+\frac{\kappa_0}{2}+\kappa_n\right\}f(x;\theta_0)\mathrm{d}x\\
\leq & \left\{M\cdot v(c_0)+\frac{\kappa_0}{2}+\kappa_n\right\}\cdot\int 1_{\{(-\infty,-A^{(n)})\bigcup(A^{(n)},\infty)\}}f(x;\theta_0)\mathrm{d}x\\
\to & \ 0. \qquad (14)
\end{aligned}
$$

From (10), (11), (12), (13), (14), we have

$$E_0[\log\{f(x;\theta_0)\}] \le \limsup_{n\to\infty} E_0[\log(g_n+\kappa_n)] + \lambda_n \le E_0\left[\log\left\{g_\infty + \frac{3\kappa_0}{4}\right\}\right].$$

This is a contradiction to (7). This completes the proof of theorem 4.

4.2 Some lemmas

Here we present two lemmas which are used to prove the main theorem. Before stating the lemmas we bound the true density $f(x;\theta_0)$ from above. Write

$$\theta_0 = (\alpha_{01}, a_{01}, b_{01}, \ldots, \alpha_{0M}, a_{0M}, b_{0M}).$$

Let $u_0 \equiv \sup_x f(x;\theta_0)$. From assumption 1

$$u_0 \le \sum_{m=1}^{M} \alpha_{0m} \frac{v_0}{b_{0m}}.$$

Now let $|\bar{\alpha}| \equiv \max(|a_{01}|, \ldots, |a_{0M}|)$. Then for $|x| \ge 2|\bar{\alpha}|$

$$|x - a_{0m}|^{-\beta} \le (|x| - |\bar{\alpha}|)^{-\beta} \le 2^{\beta}|x|^{-\beta}\ ,\ (m = 1, \ldots, M).$$

Therefore for $|x| \ge 2|\bar{\alpha}|$

$$f(x;\theta_0) \le |x|^{-\beta} 2^{\beta} v_1 \sum_{m=1}^{M} \alpha_{0m} b_{0m}^{\beta-1}.$$

Let

$$u_1 = \max(u_0 \cdot (2|\bar{\alpha}|)^{\beta},\ 2^{\beta} v_1 \sum_{m=1}^{M} \alpha_{0m} b_{0m}^{\beta-1})$$

then

$$f(x;\theta_0) \le \min\{u_0, u_1 \cdot |x|^{-\beta}\}, \quad \forall x \in \mathbb{R}. \tag{15}$$

Now we state and prove two lemmas. For $\bar{\theta} \in \Theta^{(K)}$, $g(x;\bar{\theta}) \in \mathscr{G}_K \backslash \mathscr{G}_{K-1}$ and any positive real number ρ, let

$$g(x;\bar{\theta},\rho) \ \equiv \ \sup_{\mathrm{dist}(\bar{\theta}',\bar{\theta})\le\rho} g(x;\bar{\theta}') = \sup_{\mathrm{dist}(\bar{\theta}',\bar{\theta})\le\rho} \sum_{m=1}^{K} \alpha_m f(x;\eta_m).$$

Lemma 1. *Let $\Gamma^{(K)}$, $(0 \le K \le M)$, denote any compact subset of $\Theta^{(K)}$. For any real constant $\kappa \ge 0$ and any point $\bar{\theta} \in \Gamma^{(K)}$, the following equality holds.*

$$\lim_{\rho\to 0} E_0[\log\{g(x;\bar{\theta},\rho) + \kappa\}] = E_0[\log\{g(x;\bar{\theta}) + \kappa\}] \tag{16}$$

Proof: We treat the case of $\kappa > 0$. The case of $\kappa = 0$ is almost the same as the proof of Lemma 2 in Wald(1949) [6].

From assumption 3 we have

$$\lim_{\rho \to 0} \log\{g(x; \bar{\theta}, \rho) + \kappa\} = \log\{g(x; \bar{\theta}) + \kappa\} \qquad a.e.$$

Now $\Gamma^{(K)}$ is compact and $\kappa > 0$. Hence by assumption 1, $\log\{g(x; \bar{\theta}, \rho) + \kappa\}$ is bounded. Therefore

$$\lim_{\rho \to 0} E_0[\log\{g(x; \bar{\theta}, \rho) + \kappa\}] = E_0[\log\{g(x; \bar{\theta}) + \kappa\}]$$

by bounded convergence theorem. □

Lemma 2. *For any real constant $A_0 > 0$ and $\zeta > 0$, define*

$$A_n \equiv A_0 \cdot n^{\frac{2+\zeta}{\beta-1}}.$$

Let $x_1, \ldots, x_n$ denote a random sample of size n from $f(x; \theta_0)$ and let

$$\begin{aligned} x_{n,1} &\equiv \min\{x_1, \ldots, x_n\} \\ x_{n,n} &\equiv \max\{x_1, \ldots, x_n\}. \end{aligned}$$

Then

$$\mathrm{Prob}\,(x_{n,1} < -A_n \quad \text{or} \quad x_{n,n} > A_n \quad i.o.) = 0\ .$$

Proof: If we show

$$\sum_{n=1}^{\infty} \mathrm{Prob}\,(x_{n,1} < -A_n \quad \text{or} \quad x_{n,n} > A_n) < \infty,$$

then by Borel-Cantelli lemma, lemma 2 follows. Since

$$\mathrm{Prob}\,(x_{n,1} < -A_n \quad \text{or} \quad x_{n,n} > A_n) \le \mathrm{Prob}\,(x_{n,1} < -A_n) + \mathrm{Prob}\,(x_{n,n} > A_n)\,,$$

it suffices to show that

$$\begin{aligned} \sum_{n=1}^{\infty} \mathrm{Prob}\,(x_{n,1} < -A_n) &< \infty, \\ \sum_{n=1}^{\infty} \mathrm{Prob}\,(x_{n,n} > A_n) &< \infty. \end{aligned}$$

Let $F_0(x)$ denotes the distribution function of $f(x;\theta_0)$. Then we have

$$\begin{aligned}\text{Prob}\,(x_{n,1} < -A_n) &= 1-(1-F_0(-A_n))^n,\\ \text{Prob}\,(x_{n,n} > A_n) &= 1-(F_0(A_n))^n,\end{aligned}$$

and

$$\begin{aligned}F_0(-A_n) &\le \int_{-\infty}^{-A_n} u_1\cdot|x|^{-\beta}\mathrm{d}x = \frac{u_1}{\beta-1}\cdot(A_n)^{-\beta+1},\\ 1-F_0(A_n) &\le \int_{A_n}^{\infty} u_1\cdot|x|^{-\beta}\mathrm{d}x = \frac{u_1}{\beta-1}\cdot(A_n)^{-\beta+1}.\end{aligned}$$

By replacing n by $n-n_0$ with a sufficiently large n_0 if necessary, we can assume without loss of generality that

$$\frac{u_1A_0^{-\beta+1}}{\beta-1}\left(n^{\frac{2+\zeta}{\beta-1}}\right)^{-\beta+1} < 1.$$

Then

$$\begin{aligned}\log\left\{(1-F_0(-A_n))^n\right\} &\ge \log\left\{\left(1-\frac{u_1}{\beta-1}\cdot(A_n)^{-\beta+1}\right)^n\right\}\\ &= \log\left\{\left(1-\frac{u_1A_0^{-\beta+1}}{\beta-1}\left(n^{\frac{2+\zeta}{\beta-1}}\right)^{-\beta+1}\right)^n\right\}.\end{aligned}$$

Let $u_2 \equiv u_1A_0^{-\beta+1}/(\beta-1)$ and we have

$$\begin{aligned}\left|\log(1-F_0(-A_n))^n\right| &\le \left|\log\left(1-\frac{u_2}{n^{2+\zeta}}\right)^n\right|\\ &= \frac{u_2}{n^{1+\zeta}}\left|\log\left(1-\frac{u_2}{n^{2+\zeta}}\right)^{\frac{n^{2+\zeta}}{u_2}}\right|\\ &= O(n^{-(1+\zeta)}).\end{aligned}$$

Hence there exists a sufficiently large N and $u_3 > 0$ such that

$$\left|\log(1-F_0(-A_n))^n\right| \le \frac{u_3}{n^{1+\frac{\zeta}{2}}}$$

for all $n > N$. This and $(1-F_0(-A_n))^n \le 1$ imply that for $n > N$

$$\log\left\{(1-F_0(-A_n))^n\right\} \ge -\frac{u_3}{n^{1+\frac{\zeta}{2}}}.$$

Hence by $1-e^{-y} \le y$, we have for $n > N$

$$\text{Prob}(x_{n,1} < -A_n) = 1-(1-F_0(-A_n))^n \le 1-\exp\left(-\frac{u_3}{n^{1+\frac{\zeta}{2}}}\right) \le \frac{u_3}{n^{1+\frac{\zeta}{2}}}.$$

Therefore we obtain

$$\sum_{n>N} \mathrm{Prob}(x_{n,1} < -A_n) = \sum_{n>N} 1 - (1 - F_0(-A_n))^n \le \sum_n \frac{u_3}{n^{1+\frac{\zeta}{2}}} < \infty.$$

The case of $\mathrm{Prob}(x_{n,n} > A_n)$ is also proved by the same argument. □

From the proof of the lemma 2 we obtain the following property

$$\begin{aligned} \mathrm{Prob}\left(x_{n,1} < -A_n \quad \text{or} \quad x_{n,n} > A_n\right) &\le \mathrm{Prob}\left(x_{n,1} < -A_n\right) + \mathrm{Prob}\left(x_{n,n} > A_n\right) \\ &\le \frac{2u_3}{n^{1+\frac{\zeta}{2}}} \end{aligned} \qquad (17)$$

for all sufficiently large n.

4.3 Proof of the main theorem

From now on we follow the line of the proof in Tanaka and Takemura [2], although the details of the proof here is much more complicated. For κ, λ satisfying (4), let κ_0, λ_0 be real constants such that $0 < \kappa_0 \le \kappa$, $0 < \lambda_0 \le \lambda$ and

$$B \equiv \frac{4Mv_0}{\kappa_0} > \max\{b_{01}, \ldots, b_{0M}\}.$$

Note that κ_0, λ_0 also satisfy (4). Because $\{c_n\}$ is decreasing to zero, by replacing c_0 by some c_n if necessary, we can assume without loss of generality that c_0 is sufficiently small to satisfy the following conditions.

$$v(c_0)^{\frac{\beta-1}{\beta}} > e,$$

$$c_0 < \min\{b_{01}, \ldots, b_{0M}\},$$

$$3M \cdot u_0 \cdot 2\nu(c_0) \cdot |\log \kappa_0| < \frac{\lambda_0}{8}, \qquad (18)$$

$$12M \cdot u_0 \cdot \left(\frac{4Mv_1}{\kappa_0}\right)^{\frac{1}{\beta}} \cdot (v_0 \cdot (M+1))^{\frac{\beta-1}{\beta}} \cdot \left(\frac{1}{v(c_0)}\right)^{\frac{\beta-1}{\beta}} \cdot \log\{v(c_0)\} < \frac{\lambda_0}{4}\ , \qquad (19)$$

$$\frac{\kappa_0}{4} < \frac{v(c_0)}{M+1} \quad . \qquad (20)$$

where

$$\begin{aligned} \nu(c_0) &\equiv \left(\frac{4Mv_1}{\kappa_0}\right)^{\frac{1}{\beta}} (c_0)^{\left(\frac{\beta-1}{\beta}\right)}, \\ v(c_0) &\equiv \frac{v_0}{c_0}. \end{aligned}$$

For any set $V \subset \mathbb{R}$, let $P_0(V)$ denote the probability of V under the true density

$$P_0(V) \equiv \int_V f(x;\theta_0)\mathrm{d}x\ .$$

Let $A_0 > 0$ be a positive constant which satisfy

$$P_0(\mathscr{A}_0) \cdot \log\left(\frac{Mv(c_0)+\kappa_0}{\kappa_0}\right) < \frac{\lambda_0}{8}, \tag{21}$$

where

$$\mathscr{A}_0 \equiv (-\infty, -A_0] \cup [A_0, \infty).$$

Let

$$A_n \equiv A_0 \cdot n^{\frac{2+\zeta}{\beta-1}}$$

as in lemma 2. Define

$$\Theta'_n \equiv \{\theta \in \Theta \mid \exists m \text{ s.t. } c_n \leq b_m \leq c_0 \text{ or } |a_m| > A_0 + \nu(c_0)\}$$

and

$$\Gamma_0 \equiv \{\theta \in \Theta \mid c_0 \leq b_m \leq B\ ,\ |a_m| \leq A_0 + \nu(c_0),\ (m = 1, \ldots, M)\ \}\ .$$

Note that $\Theta_0 \subset \Gamma_0$.

In view of theorems 1, 2, for the strong consistency of MLE on Θ_n, it suffices to prove that

$$\lim_{n\to\infty} \frac{\sup_{\theta\in S\cup\Theta'_n} \prod_{i=1}^n f(x_i;\theta)}{\prod_{i=1}^n f(x_i;\theta_0)} = 0, \quad a.e.$$

for all closed $S \subset \Gamma_0$ not intersecting Θ_0. Note that for all S and $\{x_i\}_{i=1}^n$,

$$\sup_{\theta\in S\cup\Theta'_n} \prod_{i=1}^n f(x_i;\theta) = \max\left\{\sup_{\theta\in S}\prod_{i=1}^n f(x_i;\theta)\ ,\ \sup_{\theta\in\Theta'_n}\prod_{i=1}^n f(x_i;\theta)\right\}\ .$$

Furthermore

$$\lim_{n\to\infty} \frac{\sup_{\theta\in S} \prod_{i=1}^n f(x_i;\theta)}{\prod_{i=1}^n f(x_i;\theta_0)} = 0, \quad a.e.$$

holds by theorem 1. Therefore it suffices to prove

$$\lim_{n\to\infty} \frac{\sup_{\theta\in\Theta'_n} \prod_{i=1}^n f(x_i;\theta)}{\prod_{i=1}^n f(x_i;\theta_0)} = 0, \quad a.e. \tag{22}$$

Let $\theta \in \Theta'_n$. Let $K \equiv K(\theta) \geq 1$ be the number of components which satisfy $b_m \leq c_0$ and $L \equiv L(\theta) \geq K(\theta)+1$ be the number of components which satisfy $b_m \leq B$. Without loss of generality, we can set $b_1 \leq b_2 \leq \ldots \leq b_K \leq c_0 < b_{K+1} \leq \ldots \leq b_L \leq B < b_{L+1} \leq \ldots \leq b_M$. Let

$$\mathscr{K} = \mathscr{K}(K, L) \equiv \{m \in \{K+1, \ldots, L\} \mid |a_m| \leq A_0 + \nu(c_0)\} \tag{23}$$

and

$$\mathscr{K}' = \mathscr{K}'(K, L) = \{K+1, \ldots, L\} \backslash \mathscr{K}.$$

Let $|\mathscr{K}|, |\mathscr{K}'|$ denote the number of element in $\mathscr{K}, \mathscr{K}'$. Let

$$\begin{aligned}\Theta'_{n,\mathscr{K}} \;\equiv\; & \{\theta \in \Theta'_n \mid |a_m| > A_0 + \nu(c_0)\ ,\ (m \in \mathscr{K}')\ , \\ & c_n \le b_1 \le \ldots \le b_K \le c_0\ ,\ B \le b_{L+1} \le \ldots \le b_M\}. \end{aligned} \tag{24}$$

As above, it suffices to prove that for each $\mathscr{K}$

$$\lim_{n\to\infty} \frac{\sup_{\theta\in\Theta'_{n,\mathscr{K}}} \prod_{i=1}^n f(x_i;\theta)}{\prod_{i=1}^n f(x_i;\theta_0)} = 0, \quad a.e. \tag{25}$$

We fix K, L and $\mathscr{K} = \mathscr{K}(K, L)$ from now on. Define $\bar{\Theta}_{\mathscr{K}}$ by

$$\begin{aligned}\bar{\Theta}_{\mathscr{K}} \;\equiv\; & \{(\alpha_{m_1}, a_{m_1}, b_{m_1}, \ldots, \alpha_{m_{|\mathscr{K}|}}, a_{m_{|\mathscr{K}|}}, b_{m_{|\mathscr{K}|}}) \in \mathbb{R}^{3|\mathscr{K}|} \mid m_i \in \mathscr{K}\ ,\ (i = 1, \ldots, |\mathscr{K}|), \\ & \sum_{m_i\in\mathscr{K}} \alpha_{m_i} \le 1\ ,\ \alpha_{m_i} \ge 0\}\end{aligned}$$

and for $\bar{\theta} \in \bar{\Theta}_{\mathscr{K}}$, define

$$\begin{aligned}\bar{f}(x;\bar{\theta}) \;&\equiv\; \sum_{m\in\mathscr{K}} \alpha_m f_m(x;\eta_m)\ , \\ \bar{f}(x;\bar{\theta},\rho) \;&\equiv\; \sup_{\mathrm{dist}(\bar{\theta},\bar{\theta}')\le\rho} \bar{f}(x;\bar{\theta}')\ .\end{aligned}$$

Note that $\bar{f}(x;\bar{\theta})$ is a subprobability measure.

Lemma 3. *Let $\mathscr{B}(\bar{\theta}, \rho(\bar{\theta}))$ denote the open ball with center $\bar{\theta}$ and radius $\rho(\bar{\theta})$. Then $\bar{\Theta}_{\mathscr{K}}$ can be covered by a finite number of balls $\mathscr{B}(\bar{\theta}^{(1)}, \rho(\bar{\theta}^{(1)})), \ldots, \mathscr{B}(\bar{\theta}^{(S)}, \rho(\bar{\theta}^{(S)}))$ such that*

$$E_0[\log\{\bar{f}(x;\bar{\theta}^{(s)}, \rho(\bar{\theta}^{(s)})) + \kappa_0\}] + \lambda_0 < E_0[\log\{f(x;\theta_0)\}]\ , \quad s = 1, \ldots, S, \tag{26}$$

.

Proof: From lemma 1 we have

$$\lim_{\rho\to 0} E_0\left[\log\{\bar{f}(x;\bar{\theta},\rho) + \kappa_0\}\right] = E_0\left[\log\{\bar{f}(x;\bar{\theta}) + \kappa_0\}\right].$$

For each $\bar{\theta} \in \bar{\Theta}_{\mathscr{K}}$

$$E_0\left[\log\left\{\bar{f}(x;\bar{\theta}) + \kappa_0\right\}\right] + \lambda_0 < E_0[\log f(x;\theta_0)]$$

holds. Therefore for each $\bar{\theta} \in \bar{\Theta}_{\mathscr{K}}$, there exists a radius $\rho(\bar{\theta}) > 0$ such that

$$E_0[\log\{\bar{f}(x;\bar{\theta},\rho(\bar{\theta})) + \kappa_0\}] + \lambda_0 < E_0[\log\{f(x;\theta_0)\}]. \tag{27}$$

Since

$$\bar{\Theta}_{\mathscr{K}} \subset \bigcup_{\theta\in\bar{\Theta}_{\mathscr{K}}} \mathscr{B}(\bar{\theta}, \rho(\bar{\theta})) \tag{28}$$

and the compactness of $\bar{\Theta}_{\mathscr{K}}$, there exists a finite number of balls $\mathscr{B}(\bar{\theta}^{(1)}, \rho(\bar{\theta}^{(1)})), \dots,$ $\mathscr{B}(\bar{\theta}^{(S)}, \rho(\bar{\theta}^{(S)}))$ which cover $\bar{\Theta}_{\mathscr{K}}$. □

Define

$$\Theta'_{n,\mathscr{K},s} \equiv \{\theta \in \Theta \mid (\alpha_{m_1}, a_{m_1}, b_{m_1}, \dots, \alpha_{m_{|\mathscr{K}|}}, a_{m_{|\mathscr{K}|}}, b_{m_{|\mathscr{K}|}}) \in \mathscr{B}(\bar{\theta}^{(s)}, \rho(\bar{\theta}^{(s)})) \ , \quad (m_1, \dots, m_{|\mathscr{K}|} \in \mathscr{K})\}.$$

We now cover $\Theta'_{n,\mathscr{K}}$ by $\Theta'_{n,\mathscr{K},1}, \dots, \Theta'_{n,\mathscr{K},S}$:

$$\Theta'_{n,\mathscr{K}} = \bigcup_{s=1}^{S} \Theta'_{n,\mathscr{K},s} \ .$$

Again it suffices to prove that for each s , $(s = 1, \dots, S)$,

$$\lim_{n\to\infty} \frac{\sup_{\theta\in\Theta'_{n,\mathscr{K},s}} \prod_{i=1}^{n} f(x_i;\theta)}{\prod_{i=1}^{n} f(x_i;\theta_0)} = 0, \quad a.e. \tag{29}$$

We fix s in addition to $\mathscr{K}$ from now on. Because

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log f(x_i;\theta_0) = E_0[\log f(x;\theta_0)], \quad a.e.$$

(29) is implied by

$$\limsup_{n\to\infty} \frac{1}{n} \sup_{\theta\in\Theta'_{n,\mathscr{K},s}} \sum_{i=1}^{n} \log f(x_i;\theta) < E_0[\log f(x;\theta_0)], \quad a.e. \tag{30}$$

Therefore it suffices to prove (30), which is a new intermediate goal of our proof hereafter.

As discussed in subsection 4.1, from assumption 1, each component is bounded above by a step function

$$f_m(x; a_m, b_m) \le 1_{[a_m-\nu(b_m), a_m+\nu(b_m))} v(b_m) + \frac{\kappa_0}{4M}.$$

Then we have

$$\sum_{m=1}^{K} \alpha_m f_m(x; a_m, b_m) \quad \le \quad \sum_{m=1}^{K} 1_{[a_m-\nu(b_m), a_m+\nu(b_m))} v(b_m) + \frac{\kappa_0}{4}. \tag{31}$$

Let $J(\theta) \equiv \bigcup_{m=1}^{K} [a_m - \nu(b_m), a_m + \nu(b_m))$. We now prove the following lemma.

Lemma 4. *Let $R_n(V)$ denote the number of observations which belong to a set $V \subset \mathbb{R}$. Then for $\theta \in \Theta'_{n,\mathscr{K},s}$*

$$\frac{1}{n} \sum_{i=1}^{n} \log f(x_i;\theta) \quad \le \quad \frac{1}{n} \sum_{i=1}^{n} \log \left\{ \bar{f}(x_i; \bar{\theta}^{(s)}, \rho(\bar{\theta}^{(s)})) + \kappa_0 \right\}$$

$$
\begin{aligned}
&+\frac{1}{n}R_n(\mathscr{A}_0)\cdot\log\left(\frac{Mv(c_0)+\kappa_0}{\kappa_0}\right)\\
&+\frac{1}{n}R_n(J(\theta))\cdot\left\{-\log\left(\frac{\kappa_0}{4}\right)\right\}\\
&+\frac{1}{n}\sum_{x_i\in J(\theta)}\log f(x_i;\theta)\ . \qquad (32)
\end{aligned}
$$

Proof: For $x\notin J(\theta)$, $f(x;\theta)\le\sum_{m=K+1}^{M}\alpha_m f_m(x;\eta_m)+\kappa_0/4$. holds. Therefore

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\log f(x_i;\theta) &\le \frac{1}{n}\sum_{x_i\in J(\theta)}\log f(x_i;\theta)+\frac{1}{n}\sum_{x_i\notin J(\theta)}\log\left\{\sum_{m=K+1}^{M}\alpha_m f_m(x_i;\eta_m)+\frac{\kappa_0}{4}\right\}\\
&= \frac{1}{n}\sum_{i=1}^{n}\log\left\{\sum_{m=K+1}^{M}\alpha_m f_m(x_i;\eta_m)+\frac{\kappa_0}{4}\right\}\\
&\quad +\frac{1}{n}\sum_{x_i\in J(\theta)}\left[\log f(x_i;\theta)-\log\left\{\sum_{m=K+1}^{M}\alpha_m f_m(x_i;\eta_m)+\frac{\kappa_0}{4}\right\}\right]\\
&\le \frac{1}{n}\sum_{i=1}^{n}\log\left\{\sum_{m=K+1}^{L}\alpha_m f_m(x_i;\eta_m)+\frac{\kappa_0}{2}\right\}\\
&\quad +\frac{1}{n}\sum_{x_i\in J(\theta)}\log f(x_i;\theta)-\frac{1}{n}R_n(J(\theta))\cdot\log\left(\frac{\kappa_0}{4}\right)\ .
\end{aligned}
$$

For $x\notin\mathscr{A}_0$

$$
\sum_{m\in\mathscr{K}'}\alpha_m f_m(x;\eta_m)\le\frac{\kappa_0}{4}.
$$

Therefore we obtain

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\log\left\{\sum_{m=K+1}^{L}\alpha_m f_m(x_i;\eta_m)+\frac{\kappa_0}{2}\right\}\\
&= \frac{1}{n}\sum_{x_i\notin\mathscr{A}_0}\log\left\{\sum_{m=K+1}^{L}\alpha_m f_m(x_i;\eta_m)+\frac{\kappa_0}{2}\right\}+\frac{1}{n}\sum_{x_i\in\mathscr{A}_0}\log\left\{\sum_{m=K+1}^{L}\alpha_m f_m(x_i;\eta_m)+\frac{\kappa_0}{2}\right\}\\
&\le \frac{1}{n}\sum_{x_i\notin\mathscr{A}_0}\log\left\{\sum_{m\in\mathscr{K}}\alpha_m f_m(x_i;\eta_m)+\kappa_0\right\}+\frac{1}{n}\sum_{x_i\in\mathscr{A}_0}\log\left\{\sum_{m=K+1}^{L}\alpha_m f_m(x_i;\eta_m)+\kappa_0\right\}\\
&= \frac{1}{n}\sum_{i=1}^{n}\log\left\{\sum_{m\in\mathscr{K}}\alpha_m f_m(x_i;\eta_m)+\kappa_0\right\}\\
&\quad +\frac{1}{n}\sum_{x_i\in\mathscr{A}_0}\left[\log\left\{\sum_{m=K+1}^{L}\alpha_m f_m(x_i;\eta_m)+\kappa_0\right\}-\log\left\{\sum_{m\in\mathscr{K}}\alpha_m f_m(x_i;\eta_m)+\kappa_0\right\}\right]
\end{aligned}
$$

$$\leq \frac{1}{n}\sum_{i=1}^{n}\log\left\{\sum_{m\in\mathscr{K}}\alpha_m f_m(x_i;\eta_m)+\kappa_0\right\}+\frac{1}{n}\sum_{x_i\in\mathscr{A}_0}\left[\log\left\{Mv(c_0)+\kappa_0\right\}-\log\kappa_0\right]$$
$$\leq \frac{1}{n}\sum_{i=1}^{n}\log\left\{\bar{f}(x_i;\bar{\theta}^{(s)},\rho(\bar{\theta}^{(s)}))+\kappa_0\right\}+\frac{1}{n}R_n(\mathscr{A}_0)\cdot\log\left(\frac{Mv(c_0)+\kappa_0}{\kappa_0}\right).$$

□

We want to bound the terms on the right hand side of (32) from above. The first term and the second term are easy. In fact by lemma 3 and the strong law of large numbers we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log\left\{\bar{f}(x_i;\bar{\theta}^{(s)},\rho(\bar{\theta}^{(s)}))+\kappa_0\right\}<E_0[\log f(x;\theta_0)]-\lambda_0,\quad a.e. \tag{33}$$

for the first term. Furthermore by (21) and the strong law of large numbers we have

$$\lim_{n\to\infty}R_n(\mathscr{A}_0)\cdot\log\left(\frac{Mv(c_0)+\kappa_0}{\kappa_0}\right)<\frac{\lambda_0}{8},\quad a.e. \tag{34}$$

for the second term. Next we consider the third term. We prove the following lemma.

Lemma 5.

$$\limsup_{n\to\infty}\sup_{\theta\in\Theta'_{n,\mathscr{K},s}}\frac{1}{n}R_n(J(\theta))\leq 3M\cdot u\cdot 2\nu(c_0),\quad a.e.$$

Proof: Let $\epsilon>0$ be arbitrarily fixed and let $J_0^{(n)}\equiv[-A_n,A_n]$. We divide $J_0^{(n)}$ from $-A_n$

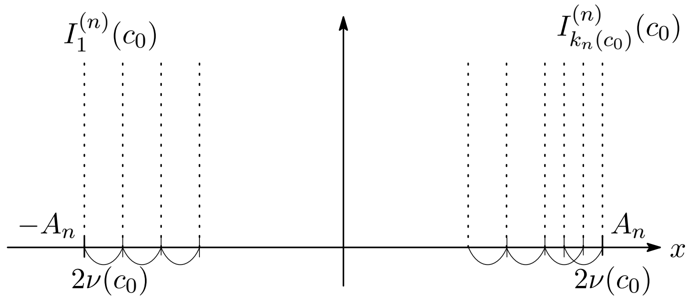

Figure 2: Division of $J_0^{(n)}$ by short intervals of length $2\nu(c_0)$.

to A_n by short intervals of length $2\nu(c_0)$. In right end of the intervals of $J_0^{(n)}$, overlap of two short intervals of length $2\nu(c_0)$ is allowed and the right end of a short interval coincides with the right end of $J_0^{(n)}$. See Figure 2. Let $k_n(c_0)$ be the number of short intervals and let $I_1^{(n)}(c_0),\ldots,I_{k_n(c_0)}^{(n)}(c_0)$ be the divided short intervals. Then we have

$$k_n(c_0)\leq\frac{2A_n}{2\nu(c_0)}+1=\frac{A_n}{\nu(c_0)}+1=\frac{A_0\cdot n^{\frac{2+\zeta}{\beta-1}}}{\nu(c_0)}+1\,. \tag{35}$$

Note that any interval in $J_0^{(n)}$ of length $2\nu(c_0)$ is covered by at most 3 small intervals from $\{I_1^{(n)}(c_0), \ldots, I_{k_n(c_0)}^{(n)}(c_0)\}$. Now consider $J(\theta) = \bigcup_{m=1}^{K}[a_m - \nu(b_m), a_m + \nu(b_m))$. Since $[a_m - \nu(b_m), a_m + \nu(b_m))$, $(m = 1, \ldots, K)$, are intervals of length less than or equal to $2\nu(c_0)$, $J(\theta)$ is covered by at most $3M$ short intervals. Then

$$\begin{aligned}
&\sup_{\theta \in \Theta'_{n,\mathscr{K},s}} \frac{1}{n} R_n(J(\theta)) - 3M \cdot u \cdot 2\nu(c_0) > \epsilon \\
&\Rightarrow \ \{x_{n,1} < -A_n \quad \text{or} \quad x_{n,n} > A_n\} \\
&\quad\ \ \text{or} \\
&\quad\ \ \{1 \le \exists k \le k_n(c_0)\ ,\ \frac{1}{n} R_n(I_k(c_0)) - u_0 \cdot 2\nu(c_0) > \frac{\epsilon}{3M}\}\ . \qquad (36)
\end{aligned}$$

By lemma 2, $\sum_n \mathrm{Prob}(x_{n,1} < -A_n \text{ or } x_{n,n} > A_n) < \infty$ and the first event on the right hand side of (36) can be ignored. We only need to consider the second event. We will use the same logic in the proofs of lemmas 6 and 7 below. Then

$$\begin{aligned}
&\mathrm{Prob}\left(\sup_{\theta \in \Theta'_{n,\mathscr{K},s}} \frac{1}{n} R_n(J(\theta)) - 3M \cdot u_0 \cdot 2\nu(c_0) > \epsilon \right) \\
&\qquad\qquad \le \sum_{k=1}^{k_n(c_0)} \mathrm{Prob}\left(\frac{1}{n} R_n(I_k(c_0)) - u_0 \cdot 2\nu(c_0) > \frac{\epsilon}{3M} \right)\ .
\end{aligned}$$

Since

$$P_0(I_k(c_0)) \le u_0 \cdot 2\nu(c_0), \quad (k = 1, \ldots, k_n(\theta)),$$

$R_n(V) \sim \mathrm{Bin}(n, P_0(V))$ and from (3), we obtain

$$\begin{aligned}
&\mathrm{Prob}\left(\frac{1}{n} R_n(I_k(c_0)) - u \cdot 2\nu(c_0) > \frac{\epsilon}{3M} \right) \\
&\le \ \mathrm{Prob}\left(\frac{1}{n} R_n(I_k(c_0)) - P_0(I_k(c_0)) > \frac{\epsilon}{3M} \right) \\
&\le \ \exp\left(-\frac{2n\epsilon^2}{9M^2} \right).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathrm{Prob}\left(\sup_{\theta \in \Theta'_{n,\mathscr{K},s}} \frac{1}{n} R_n(J(\theta)) - 3M \cdot u_0 \cdot 2\nu(c_0) > \epsilon \right) \\
&\qquad\qquad \le \left(\frac{A_0 \cdot n^{\frac{2+\zeta}{\beta-1}}}{\nu(c_0)} + 1 \right) \cdot \exp\left(-\frac{2n\epsilon^2}{9M^2} \right).
\end{aligned}$$

When we sum this over n, the resulting series on the right converges. Hence by Borel-Cantelli, we have

$$\mathrm{Prob}\left(\sup_{\theta \in \Theta'_{n,\mathscr{K},s}} \frac{1}{n} R_n(J(\theta)) - 3M \cdot u_0 \cdot 2\nu(c_0) > \epsilon \quad i.o. \right) = 0.$$

Because $\epsilon > 0$ was arbitrary, we obtain

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta'_{n,\mathscr{K},s}} \frac{1}{n} R_n(J(\theta)) \le 3M \cdot u_0 \cdot 2\nu(c_0), \quad a.e.$$

□

By this lemma and (18) we have

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta'_{n,\mathscr{K},s}} \frac{1}{n} R_n(J(\theta)) \cdot \left(-\log\frac{\kappa_0}{4}\right) \le 3M \cdot u_0 \cdot 2\nu(c_0) \cdot \left(-\log\frac{\kappa_0}{4}\right) < \frac{\lambda_0}{8} \quad a.e. \tag{37}$$

This bounds the third term on the right hand side of (32) from above.

Finally we bound the fourth term on the right hand side of (32) from above. This is the most difficult part of our proof.

We reuse the same argument as in section 4.1. From (31) we have

$$1_{J(\theta)}(x) \cdot \sum_{m=1}^{M} \alpha_m f_m(x; a_m, b_m) \le 1_{J(\theta)}(x) \cdot \left\{ \sum_{m=1}^{M} 1_{[a_m-\nu(b_m), a_m+\nu(b_m))} \cdot v(b_m) + \frac{\kappa_0}{4} \right\}. \tag{38}$$

For $x \in J(\theta)$, the right hand side of (38) can be written as

$$1_{J(\theta)}(x) \cdot \left\{ \sum_{m=1}^{M} 1_{[a_m-\nu(b_m), a_m+\nu(b_m))}(x) \cdot v(b_m) + \frac{\kappa_0}{4} \right\} = \sum_{t=1}^{T(\theta)} H(J_t(\theta)) \cdot 1_{J_t(\theta)}(x),$$

where $J_t(\theta)$, $(t = 1, \ldots, T(\theta))$, are disjoint intervals, $[a_m - \nu(b_m), a_m + \nu(b_m))$, $(m = 1, \ldots, K)$, are represented by unions of some of $J_t(\theta)$'s and $H(J_t(\theta))$ satisfies

$$H(J_t(\theta)) = \sum_{m=1}^{M} 1_{[a_m-\nu(b_m), a_m+\nu(b_m))}(x) \cdot v(b_m) + \frac{\kappa_0}{4} \quad , \quad (x \in J_t(\theta)).$$

Note that $T(\theta) \le 2M$. Let $W(J_t(\theta))$ denote the length of $J_t(\theta)$. For convenience we determine the order of t such that

$$H(J_1(\theta)) \le H(J_2(\theta)) \le \ldots \le H(J_{T(\theta)}(\theta)) .$$

We now classify the intervals $J_t(\theta),\ t = 1, \ldots, T(\theta)$, by the height $H(J_t(\theta))$. Define c'_n by

$$c'_n = c_0 \cdot \exp\left(-n^{1/4}\right)$$

and define $\tau_n(\theta)$

$$\tau_n(\theta) \equiv \max\{t \in \{1, \ldots, T\} \mid H(J_t(\theta)) \le Mv(c'_n)\} \tag{39}$$

where $v(c_n') \equiv v_0/c_n'$. Then the fourth term on the right hand side of (32) is written as

$$\begin{aligned}
\frac{1}{n}\sum_{x_i \in J(\theta)} \log f(x_i;\theta) &\leq \sum_{t=1}^{T(\theta)} \frac{1}{n} \sum_{x_i \in J_t(\theta)} \log H(J_t(\theta)) \\
&= \frac{1}{n}\sum_{t=1}^{T(\theta)} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) \\
&= \frac{1}{n}\sum_{t=1}^{\tau_n(\theta)} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) \\
&\quad + \frac{1}{n}\sum_{t=\tau_n(\theta)+1}^{T(\theta)} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)). \qquad (40)
\end{aligned}$$

From (20) we have

$$H(J_t(\theta)) - \kappa_0/4 > H(J_t(\theta)) - \frac{v_0}{c_0(M+1)} > H(J_t(\theta)) - \frac{H(J_t(\theta))}{M+1} = \frac{M \cdot H(J_t(\theta))}{M+1}.$$

At any point on $J_t(\theta^{(n)})$, $H(J_t(\theta^{(n)})) - \kappa_0/4$ consists of only at most M components. Therefore there exists at least one component such that

$$v(b_m) \geq \frac{1}{M} \cdot \left(H(J_t(\theta)) - \frac{\kappa_0}{4} \right) \geq \frac{H(J_t(\theta))}{(M+1)}$$

and we obtain

$$\begin{aligned}
\nu(b_m) &= \left(\frac{4Mv_1}{\kappa_0}\right)^{\frac{1}{\beta}} (b_m)^{(\frac{\beta-1}{\beta})} \\
&= \left(\frac{4Mv_1}{\kappa_0}\right)^{\frac{1}{\beta}} \left(\frac{v_0}{v(b_m)}\right)^{(\frac{\beta-1}{\beta})} \\
&\leq \left(\frac{4Mv_1}{\kappa_0}\right)^{\frac{1}{\beta}} \left(\frac{v_0(M+1)}{H(J_t(\theta))}\right)^{(\frac{\beta-1}{\beta})}.
\end{aligned}$$

Let

$$\begin{aligned}
v_2 &\equiv 2 \cdot \left(\frac{4Mv_1}{\kappa_0}\right)^{\frac{1}{\beta}} \cdot (v_0(M+1))^{(\frac{\beta-1}{\beta})}, \\
\tilde{\beta} &\equiv \frac{\beta - 1}{\beta} > 0.
\end{aligned}$$

From $W(J_t(\theta)) \leq 2\nu(b_m)$, it follows that

$$W(J_t(\theta)) \leq 2\nu(b_m) \leq v_2 \cdot \left(\frac{1}{H(J_t(\theta))}\right)^{\tilde{\beta}}. \qquad (41)$$

Furthermore define

$$\xi(y) \equiv v_2 \cdot \left(\frac{1}{y}\right)^{\tilde{\beta}} , \qquad (y > 0) .$$

From (19), and noting that $\log y / y^{\tilde{\beta}}$ is decreasing in $y^{\tilde{\beta}} \geq e$, we have

$$\begin{aligned} 3 \cdot \sum_{t=1}^{\tau_n(\theta)} u_0 \cdot \xi(H(J_t(\theta))) \cdot \log H(J_t(\theta)) &\leq 3 \cdot 2M \cdot u_0 \cdot v_2 \cdot \left(\frac{1}{v(c_0)}\right)^{\tilde{\beta}} \cdot \log \{v(c_0)\} < \frac{\lambda_0}{4}, \\ 3 \cdot \sum_{t=\tau_n(\theta)+1}^{T(\theta)} \frac{2}{n} \log H(J_t(\theta)) &\leq 3 \cdot 2M \cdot \frac{2}{n} \cdot \log \{M \cdot v(c_n)\} \to 0. \end{aligned} \tag{42}$$

Now suppose that the following inequality holds.

$$\begin{aligned} &\limsup_{n\to\infty} \sup_{\theta \in \Theta'_{n,\mathscr{K},s}} \left[\sum_{t=1}^{T(\theta)} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) \right. \\ &\left. - 3 \left\{ \sum_{t=1}^{\tau_n(\theta)} u_0 \cdot \xi(H(J_t(\theta))) \cdot \log H(J_t(\theta)) + \sum_{t=\tau_n(\theta)+1}^{T(\theta)} \frac{2}{n} \log H(J_t(\theta)) \right\} \right] \leq 0, \quad a.e. \end{aligned} \tag{43}$$

Then from (40) and (42), the fourth term on the right hand side of (32) is bounded from above as

$$\limsup_{n\to\infty} \frac{1}{n} \sup_{\theta \in \Theta'_{n,\mathscr{K},s}} \sum_{x_i \in J(\theta)} \log f(x_i;\theta) \leq \frac{\lambda_0}{4} \quad a.e. \tag{44}$$

Combining (33), (34), (37) and (44) we obtain

$$\begin{aligned} \limsup_{n\to\infty} \sup_{\theta \in \Theta'_{n,\mathscr{K},s}} \frac{1}{n} \sum_{i=1}^{n} \log f(x_i;\theta) &\leq (E_0[\log f(x;\theta_0)] - \lambda_0) + \frac{\lambda_0}{8} + \frac{\lambda_0}{8} + \frac{\lambda_0}{4} \\ &\leq E_0[\log f(x;\theta_0)] - \frac{\lambda_0}{2}, \quad a.e. \end{aligned}$$

and (30) is satisfied. Therefore it suffices to prove (43), which is a new goal of our proof.

We now consider further finite covering of $\Theta'_{n,\mathscr{K},s}$. Define

$$\Theta'_{n,\mathscr{K},s,T,\tau} \equiv \{\theta \in \Theta'_{n,\mathscr{K},s} \mid T(\theta) = T \ , \ \tau_n(\Theta) = \tau\} \ .$$

Then

$$\sup_{\theta \in \Theta'_{n,\mathscr{K},s}} \left[\sum_{t=1}^{T(\theta)} \frac{1}{n} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) \right.$$

$$
\left.\left. - 3\left\{\sum_{t=1}^{\tau_n(\theta)} u_0 \cdot \xi(H(J_t(\theta))) \cdot \log H(J_t(\theta)) + \sum_{t=\tau_n(\theta)+1}^{T(\theta)} \frac{2}{n} \log H(J_t(\theta))\right\}\right]\right.
$$

$$
\begin{aligned}
\leq \;& \max_{T=1,\ldots,2M} \max_{\tau=1,\ldots,T} \Bigg[\\
& \sup_{\theta\in\Theta'_{n,\mathscr{K},s,T,\tau}} \left\{\sum_{t=1}^{\tau} \frac{1}{n} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) - 3\sum_{t=1}^{\tau} u_0 \cdot \xi(H(J_t(\theta))) \cdot \log H(J_t(\theta))\right\} \\
& + \sup_{\theta\in\Theta'_{n,\mathscr{K},s,T,\tau}} \left\{\sum_{t=\tau+1}^{T} \frac{1}{n} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) - 3\sum_{t=\tau+1}^{T} \frac{2}{n} \log H(J_t(\theta))\right\}\Bigg].
\end{aligned}
\tag{45}
$$

Suppose that the following inequalities hold for all T and τ.

$$
\begin{aligned}
\limsup_{n\to\infty} \sup_{\theta\in\Theta'_{n,\mathscr{K},s,T,\tau}} & \left[\sum_{t=1}^{\tau} \frac{1}{n} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) \right. \\
& \left. -3\sum_{t=1}^{\tau} u_0 \cdot \xi(H(J_t(\theta))) \cdot \log H(J_t(\theta))\right] \leq 0, \quad a.e.
\end{aligned}
\tag{46}
$$

$$
\limsup_{n\to\infty} \sup_{\theta\in\Theta'_{n,\mathscr{K},s,T,\tau}} \left[\sum_{t=\tau+1}^{T} \frac{1}{n} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) - 3\sum_{t=\tau+1}^{T} \frac{2}{n} \log H(J_t(\theta))\right] \leq 0, \quad a.e.
\tag{47}
$$

Then (43) is derived from (45), (46), (47). Therefore it suffices to prove (46) and (47), which are the final goals of our proof. We state (46) and (47) as two lemmas and give their proofs.

Lemma 6.

$$
\limsup_{n\to\infty} \sup_{\theta\in\Theta'_{n,\mathscr{K},s,T,\tau}} \left[\sum_{t=\tau+1}^{T} \frac{1}{n} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) - 3\sum_{t=\tau+1}^{T} \frac{2}{n} \log H(J_t(\theta))\right] \leq 0 \quad a.e.
$$

Proof: Let $\delta > 0$ be any fixed positive real constant and let $a'_t(\theta)$ denote the middle point of $J_t(\theta)$. Here, we consider the probability of the event that

$$
\sup_{\theta\in\Theta'_{n,\mathscr{K},s,T,\tau}} \left[\sum_{t=\tau+1}^{T} \frac{1}{n} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) - 3\sum_{t=\tau+1}^{T} \frac{2}{n} \log H(J_t(\theta))\right] > 2M\delta.
\tag{48}
$$

Since $H(J_t(\theta)) > Mv(c'_n)$ holds for $t > \tau$, we obtain by (41)

$$
W(J_t(\theta)) \;\leq\; v_2 \cdot \left(\frac{1}{Mv(c'_n)}\right)^{\tilde{\beta}} = v_2 \cdot \left(\frac{c_0}{Mv_0}\right)^{\tilde{\beta}} \cdot \exp\left(-\tilde{\beta} \cdot n^{1/4}\right).
$$

Let

$$
\begin{aligned}
v_3 &\equiv v_2 \cdot \left(\frac{c_0}{M v_0}\right)^{\tilde{\beta}} \\
w_n &\equiv \frac{v_3}{2} \cdot \exp\left(-\tilde{\beta} \cdot n^{1/4}\right).
\end{aligned}
$$

Noting that for $t > \tau$, the length of $J_t(\theta)$ is less than or equal to $2w_n$, the following relation holds.

The event (48) occurs.

$$
\begin{aligned}
\Rightarrow\ & \sup_{\theta \in \Theta'_{n,\mathscr{K},s,T,\tau}} \left[\sum_{t=\tau+1}^{T} \max\left\{0, \left(\frac{1}{n} R_n([a'_t(\theta) - w_n, a'_t(\theta) + w_n]) \right.\right.\right. \\
& \qquad\qquad \left.\left.\left. -3 \cdot \frac{2}{n}\right)\right\} \cdot \log\left(M v(c_n)\right)\right] > 2M\delta \\
\Rightarrow\ & \exists \theta \in \Theta'_{n,\mathscr{K},s,T,\tau}, \exists t > \tau \\
& \max\left\{0, \left(\frac{1}{n} R_n([a'_t(\theta) - w_n, a'_t(\theta) + w_n]) - 3 \cdot \frac{2}{n}\right)\right\} \cdot \log\left(M v(c_n)\right) > \delta \\
\Rightarrow\ & \exists \theta \in \Theta'_{n,\mathscr{K},s,T,\tau}, \exists t > \tau \\
& R_n([a'_t(\theta) - w_n, a'_t(\theta) + w_n]) \geq 6 \\
\Rightarrow\ & \sup_{-\infty < a' < \infty} R_n([a' - w_n, a' + w_n]) \geq 6\ . \qquad (49)
\end{aligned}
$$

Below, we consider the probability of the event that (49) occurs. We divide $J_0^{(n)} = [-A_n, A_n]$ from $-A_n$ to A_n by short intervals of length $2w_n$ as in the proof of lemma 5. Let $k(w_n)$ be the number of short intervals and let $I_1(w_n), \ldots, I_{k(w_n)}(w_n)$ be the divided short intervals. Then we have

$$
k(c'_n) \leq \frac{2A_n}{2w_n} + 1 = \frac{A_0 \cdot n^{\frac{2+\zeta}{\beta-1}}}{\nu(c_0)} + 1\ . \qquad (50)
$$

Since any interval in J_0 of length $2w_n$ is covered by at most 3 small intervals from $I_1(w_n), \ldots, I_{k(w_n)}(w_n)$ and from lemma 2,

$$
\sup_{-\infty < a' < \infty} R_n([a' - w_n, a' + w_n]) \geq 6 \ \Rightarrow\ 1 \leq \exists k \leq k(w_n)\ ,\ R_n(I_k(w_n)) \geq 2\ .
$$

Note that $R_n(I_k(w_n)) \sim \mathrm{Bin}(n, P_0(I_k(w_n)))$ and $P_0(I_k(w_n)) \leq 2w_n u_0$. Therefore from (50) we have

$$
\begin{aligned}
& \sum_{k=1}^{k(w_n)} \mathrm{Prob}\left(R_n(I_k(w_n)) \geq 2\right) \\
& \quad \leq \left(\frac{A_n}{w_n} + 1\right) \cdot \left\{\max_{1 \leq k \leq k(w_n)} \mathrm{Prob}(R_n(I_k(w_n)) \geq 2)\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \left(\frac{A_n}{w_n}+1\right)\sum_{k=2}^{n}\binom{n}{k}(2w_nu_0)^k(1-2w_nu_0)^{n-k}\\
&\leq \left(\frac{A_n}{w_n}+1\right)\sum_{k=2}^{n}\frac{n^k}{k!}(2w_nu_0)^k \leq \left(\frac{A_n}{w_n}+1\right)(2nw_nu_0)^2\sum_{k=0}^{n}\frac{1}{k!}(2nw_nu_0)^k\\
&\leq \left(\frac{A_n}{w_n}+1\right)(2nw_nu_0)^2\exp\left(2nw_nu_0\right) .
\end{aligned}
$$

When we sum this over n, resulting series on the right converges. Hence by Borel-Cantelli and the fact that $\delta > 0$ was arbitrary, we obtain

$$
\limsup_{n\to\infty}\sup_{\theta\in\Theta'_{n,\mathscr{K},s,T,\tau}}\left[\sum_{t=\tau+1}^{T}\frac{1}{n}R_n(J_t(\theta))\cdot\log H(J_t(\theta)) - 3\sum_{t=\tau+1}^{T}\frac{2}{n}\log H(J_t(\theta))\right] \leq 0 \quad a.e.
$$

□

Finally we prove (46).

Lemma 7.

$$
\begin{aligned}
\limsup_{n\to\infty}\sup_{\theta\in\Theta'_{n,\mathscr{K},s,T,\tau}} &\left[\sum_{t=1}^{\tau}\frac{1}{n}R_n(J_t(\theta))\cdot\log H(J_t(\theta))\right.\\
&\left. -3\sum_{t=1}^{\tau}u_0\cdot\xi(H(J_t(\theta)))\cdot\log H(J_t(\theta))\right] \leq 0 \quad a.e.
\end{aligned}
$$

Proof: Let $\delta > 0$ be any fixed positive real constant and let

$$
h_n \equiv \frac{\delta}{12}\left\{u_0\cdot\log\left(Mv(c'_n)\right)\right\}^{-1} . \tag{51}
$$

Since $v(c_0) \leq H(J_t(\theta)) \leq Mv(c'_n)$, we have $\xi(Mv(c'_n)) \leq \xi(H(J_t(\theta))) \leq \xi(v(c_0))$. We divide the interval $[\xi(Mv(c'_n)), \xi(v(c_0))]$ from $\xi(v(c_0))$ to $\xi(Mv(c'_n))$, by short intervals of length h_n. In the left end $\xi(Mv(c'_n))$ of the interval $[\xi(Mv(c'_n)), \xi(v(c_0))]$, overlap of two short intervals of length h_n is allowed and the left end of a short interval is equal to $\xi(Mv(c'_n))$. Let l_n be the number of short intervals of length h_n and define $w_l^{(n)}$ by

$$
2w_l^{(n)} \equiv \begin{cases}\xi(v(c_0)) - (l-1)h_n, & 1 \leq l \leq l_n,\\ \xi(Mv(c'_n)), & l = l_n + 1.\end{cases}
$$

Then we have

$$
l_n \leq \frac{\xi(v(c_0))}{h_n} + 1 . \tag{52}
$$

Let

$$\psi(y) = \xi^{-1}(y) = \left(\frac{v_2}{y}\right)^{1/\tilde{\beta}} \quad , \qquad (y > 0) \ ,$$

where $\xi^{-1}(\cdot)$ is the inverse function of $\xi(\cdot)$. Next we consider the probability of the event that

$$\sup_{\theta \in \Theta'_{n,\mathscr{K},s,T,\tau}} \left[\sum_{t=1}^{\tau} \frac{1}{n} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) \right.$$
$$\left. -3 \sum_{t=1}^{\tau} u_0 \cdot \xi(H(J_t(\theta))) \cdot \log H(J_t(\theta)) \right] > 2M\delta. \qquad (53)$$

For this event the following relation holds.

The event (53) occurs.

$$\Rightarrow \ \exists \theta \in \Theta'_{n,\mathscr{K},s,T,\tau}$$
$$1 \le \exists l(1), \ldots, \exists l(\tau) \le l_n \quad \text{s.t.}$$
$$\psi(2w^{(n)}_{l(1)}) \le H(J_1(\theta)) \le \psi(2w^{(n)}_{l(1)+1}), \ldots, \psi(2w^{(n)}_{l(\tau)}) \le H(J_\tau(\theta)) \le \psi(2w^{(n)}_{l(\tau)+1}),$$
$$\sum_{t=1}^{\tau} \max\left\{0, \left(\frac{1}{n} R_n([a'_t(\theta) - w^{(n)}_{l(t)}, a'_t(\theta) + w^{(n)}_{l(t)}]) \right.\right.$$
$$\left.\left. -3u_0 \cdot 2w^{(n)}_{l(t)+1}\right)\right\} \cdot \log\{\psi(2w^{(n)}_{l(t)+1})\} > 2M\delta$$
$$\Rightarrow \ \exists \theta \in \Theta'_{n,\mathscr{K},s,T,\tau} \ , \ 1 \le \exists t \le \tau$$
$$1 \le \exists l(t) \le l_n \quad \text{s.t.}$$
$$\psi(2w^{(n)}_{l(t)}) \le H(J_t(\theta)) \le \psi(2w^{(n)}_{l(t)+1}),$$
$$\max\left\{0, \left(\frac{1}{n} R_n([a'_t(\theta) - w^{(n)}_{l(t)}, a'_t(\theta) + w^{(n)}_{l(t)}]) \right.\right.$$
$$\left.\left. -3u_0 \cdot 2w^{(n)}_{l(t)+1}\right)\right\} \cdot \log\{\psi(2w^{(n)}_{l(t)+1})\} > \delta$$
$$\Rightarrow \ 1 \le \exists l \le l_n \quad s.t.$$
$$\max\left\{0, \sup_{-\infty < a' < \infty} \left(\frac{1}{n} R_n([a' - w^{(n)}_l, a' + w^{(n)}_l]) \right.\right.$$
$$\left.\left. -3u_0 \cdot 2w^{(n)}_{l+1}\right)\right\} \cdot \log\{\psi(2w^{(n)}_{l+1})\} > \delta$$
$$\Rightarrow \ 1 \le \exists l \le l_n \quad s.t.$$
$$\sup_{-\infty < a' < \infty} \left\{ \left(\frac{1}{n} R_n([a' - w^{(n)}_l, a' + w^{(n)}_l]) - 3u_0 \cdot 2w^{(n)}_l\right) \cdot \log\{\psi(2w^{(n)}_{l+1})\} \right.$$
$$\left. + 3u_0 \cdot (2w^{(n)}_l - 2w^{(n)}_{l+1}) \cdot \log\{\psi(2w^{(n)}_{l+1})\} \right\} > \delta \qquad (54)$$

Then from (51) and lemma 2 the following relation holds.

The event (54) occurs.

$$\Rightarrow \quad 1 \le \exists l \le l_n \ , \quad \sup_{-\infty < a' < \infty} \frac{1}{n} \left(R_n([a' - w_l^{(n)}, a' + w_l^{(n)}]) - 3u_0 \cdot 2w_l^{(n)} \right) \cdot \log\{\psi(2w_{l+1}^{(n)})\} > \frac{\delta}{2}$$

$$\Rightarrow \quad 1 \le \exists l \le l_n \ , \quad \sup_{-A_n < a' < A_n} \frac{1}{n} \left(R_n([a' - w_l^{(n)}, a' + w_l^{(n)}]) - 3u_0 \cdot 2w_l^{(n)} \right) \cdot \log\{\psi(2w_{l+1}^{(n)})\} > \frac{\delta}{2} \tag{55}$$

Below, we consider the probability of the event that (55) occurs. We divide $J_0^{(n)}$ from $-A_n$ to A_n by short intervals of length $2w_l^{(n)}$ as in the proof of lemma 5. Let $k(w_l^{(n)})$ be the number of short intervals and let $I_1(w_l^{(n)}), \ldots, I_{k(w_l^{(n)})}(w_l^{(n)})$ be the divided short intervals. Then we have

$$k(w_l^{(n)}) \le \frac{2A_n}{2w_l^{(n)}} + 1 \ . \tag{56}$$

Since any interval in J_0 of length $2b_l^{(n)}$ is covered by at most 3 small intervals from $\{I_1(w_l^{(n)}), \ldots, I_{k(w_l^{(n)})}(w_l^{(n)})\}$, we have

$$\begin{aligned} &\sup_{-A_n \le a' \le A_n} \left(\frac{1}{n} R_n([a' - w_l^{(n)}, a' + w_l^{(n)}]) - 3u_0 \cdot 2w_l^{(n)} \right) > \frac{\delta}{2} \left(\log\{\psi(2w_{l(t)+1}^{(n)})\} \right)^{-1} \\ &\Rightarrow \max_{k=1,\ldots,k(w_l^{(n)})} \left(\frac{1}{n} R_n(I_k(w_l^{(n)})) - u_0 \cdot 2w_l^{(n)} \right) > \frac{1}{3} \cdot \frac{\delta}{2} \left(\log\{\psi(2w_{l(t)+1}^{(n)})\} \right)^{-1} . \end{aligned} \tag{57}$$

Note that $R_n(I_k(w_l^{(n)})) \sim \mathrm{Bin}(n, P_0(I_k(w_l^{(n)})))$ and $P_0(I_k(w_l^{(n)})) \le u_0 \cdot 2w_l^{(n)}$. Therefore from (3) and (56) we have

$$\begin{aligned} &\mathrm{Prob}\left(\max_{k=1,\ldots,k(w_l^{(n)})} \frac{1}{n} \left(R_n(I_k(w_l^{(n)})) - u_0 \cdot 2w_l^{(n)} \right) > \frac{1}{3} \cdot \frac{\delta}{2} \left\{ \log\left(\psi(2w_{l(t)+1}^{(n)})\right) \right\}^{-1} \right) \\ &\qquad \le \left(\frac{2A_n}{2w_l^{(n)}} + M \right) \cdot \exp\left[-2n \cdot \frac{\delta^2}{36} \left\{ \log\left(\psi(2w_{l(t)+1}^{(n)})\right) \right\}^{-2} \right] \\ &\qquad \le \left(\frac{A_n}{\xi(Mv(c_n'))} + M \right) \cdot \exp\left[-2n \cdot \frac{\delta^2}{36} \left\{ \log\left(Mv(c_n')\right) \right\}^{-2} \right]. \end{aligned} \tag{58}$$

From (52), (54), (55), (57), and (58) we obtain

$$\begin{aligned} &\sum_{l=1}^{l_n} \mathrm{Prob}\left(\sup_{-A_n < a' < A_n} \frac{1}{n} \left(R_n([a' - w_l^{(n)}, a' + w_l^{(n)}]) - 3u_0 \cdot 2w_l^{(n)} \right) \cdot \log\{\psi(2w_{l+1}^{(n)})\} > \frac{\delta}{2} \right) \\ &\le \left(\frac{\xi(v(c_0))}{h_n} + 1 \right) \cdot \left(\frac{A_n}{\xi(Mv(c_n'))} + M \right) \cdot \exp\left[-2n \cdot \frac{\delta^2}{36} \left\{ \log\left(Mv(c_n')\right) \right\}^{-2} \right] \end{aligned} \tag{59}$$

When we sum this over n, the resulting series on the right converges. Hence by Borel-Cantelli and the fact that $\delta > 0$ is arbitrary, we have

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta'_{n,\mathscr{K},s,T,\tau}} \Bigg[\sum_{t=1}^{\tau} \frac{1}{n} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) \\ -3 \sum_{t=1}^{\tau} u_0 \cdot \xi(H(J_t(\theta))) \cdot \log H(J_t(\theta)) \Bigg] \le 0 \quad a.e.$$

□

This completes the proof of theorem 5.

5 Discussions

In this paper we consider the strong consistency of the maximum likelihood estimator for mixtures of location-scale distributions. We treat the case that the scale parameters of the component distributions are restricted from below by $c_n = \exp(-n^d)$, $0 < d < 1$, and give the regularity conditions for the strong consistency of the maximum likelihood estimator.

As in the case of the uniform mixture, it is readily verified that if c_n decreases to zero faster than $\exp(-n)$, then the consistency of the maximum likelihood estimator fails. Therefore the rate of $c_n = \exp(-n^d)$, $0 < d < 1$, obtained in this paper is almost the lower bound of the order of c_n which maintains the consistency.

Although we treat the univariate case in this paper, it is clear that the result obtained in this paper can be extended to the multivariate case under the condition that components are bounded and their tails decrease to zero fast enough if the minimum singular values of the scale matrices of the components are restricted from below by c_n.

Finally let us consider some sufficient conditions for the regularity conditions. For $\eta \in \Omega_m$ and any positive real number r, let

$$f_m(x;\eta,r) \equiv \sup_{\mathrm{dist}(\eta',\eta)\le r} f_m(x;\eta').$$

Let $\tilde{\Omega}_m$ be any compact subset of Ω_m. Consider the following two conditions.

Assumption 5. *For each $\eta_m \in \tilde{\Omega}_m$ and sufficiently small r, $f_m(x;\eta_m,r)$ is measurable.*

Assumption 6. *For each $\eta_m \in \tilde{\Omega}_m$, if $\lim_{n\to\infty} \eta_m^{(n)} = \eta_m$, then $\lim_{n\to\infty} f_m(x;\eta_m^{(n)}) = f_m(x;\eta_m)$ for all x.*

If assumptions 5, 6 hold, then it is easily verified that assumptions 2, 3 hold. Thus assumptions 1, 4, 5, 6 are sufficient conditions for regularity conditions and assumptions 5, 6 are checked more easily. For example, finite mixture density which consists of normal density, t-density and uniform density on an open interval satisfies assumptions 1, 4, 5, 6.

References

[1] Hathaway, R. J. (1985). A constrained formulation of maximum likelihood estimation for normal mixture distributions, *Ann. Statist.*, **13**, pp. 795–800.

[2] Tanaka, K. and Takemura, A. (2003). Strong consistency of MLE for finite uniform mixtures when the scale parameters are exponentially small. *METR Technical Report*, 2003-20, University of Tokyo. Submitted for publication.

[3] McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*, New York: Wiley.

[4] Okamoto, M. (1958). Some inequality relating to the partial sum of binomial probabilities, *Ann. Inst. Statist. Math.,* **10**, pp. 29–35.

[5] Redner, A. R. (1981). Note on the consistency of the maximum likelihood estimation for nonidentifiable distribution, *Ann. Statist.*, **9**, pp. 225–228.

[6] Wald, A. (1949). Note on the consistency of the maximum likelihood estimate, *Ann. Math. Statist.*, **29**, pp. 595–601.