MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Strong consistency of MLE for finite uniform mixtures when the scale parameters are exponentially small

Kentaro Tanaka and Akimichi Takemura

METR 2003–20 May 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 uniform mixtures 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

May, 2003

Abstract

We consider maximum likelihood estimation of finite mixture of uniform distributions. We prove that the maximum likelihood estimator is strongly consistent, if the scale parameters of the component uniform distributions are restricted from below by $\exp(-n^d)$, $0 < d < 1$, where n is the sample size.

1 Introduction

Consider a mixture of two uniform distributions

$$(1-\alpha)f_1(x; a_1, b_1) + \alpha f_2(x; a_2, b_2),$$

where $f_m(x; a_m, b_m)$, $m = 1, 2$, are uniform densities with parameter (a_m, b_m) on the half-open intervals $[a_m - b_m, a_m + b_m)$ and $0 \le \alpha \le 1$. For definiteness and convenience we use the half-open intervals in this paper, although obviously the intervals can be open or closed. For simplicity suppose that $a_1 = 1/2, b_1 = 1/2, \alpha = \alpha_0$ are known and the parameter space is

$$\{(a_2, b_2) \mid 0 \le a_2 - b_2 \ ,\ a_2 + b_2 \le 1\}$$

so that the support of the density is $[0, 1)$. Let $x_1, \ldots, x_n$ denote a random sample of size $n \ge 2$ from the true density $(1-\alpha_0)f_1(x; 1/2, 1/2) + \alpha_0 f_2(x; a_{2,0}, b_{2,0})$. If we set $a_2 = x_1$, then likelihood tends to infinity as $b_2 \to 0$. (Figure 1)

When we restrict that $b_2 \ge c$, where c is a positive real constant, then we can avoid the divergence of the likelihood and the maximum likelihood estimator is strongly consistent provided that $b_{2,0} \ge c$. But there is a problem of how small we have to choose c to ensure $b_{2,0} \ge c$. 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

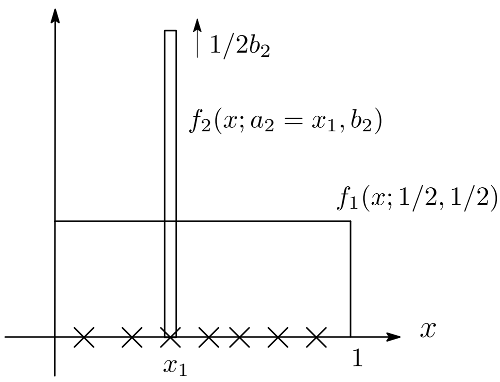

Figure 1: The likelihood tends to infinity as $b_2 \to 0$ at $a_2 = x_1$.

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), which treats mixtures of normal distributions with constraints imposed on the ratios of variances. See also a discussion in section 3.8 of McLachlan and Peel(2000).

Figure 2 depicts an example of likelihood function. Random sample of size $n = 40$ is generated from $0.6 \cdot f(x; 0.5, 0.5) + 0.4 \cdot f(x; 0.6, 0.2)$ and the model is $0.6 \cdot f(x; 0.5, 0.5) + 0.4 \cdot f(x; a, b)$. We see that although the likelihood function diverges to infinity as $b_2 \downarrow 0$, the peaks of the likelihood function are very narrow. This suggests that the bound c_n can decrease to zero fairly quickly. In fact we prove that c_n can decrease exponentially fast to zero for the mixture of M uniform distributions. More precisely we prove that the maximum likelihood estimator is strongly consistent if $c_n = \exp(-n^d)$, $0 < d < 1$.

The organization of the paper is as follows. In section 2 we summarize some preliminary results. In section 3 we state our main result in theorem 3.2. Section 4 is devoted to the proof of theorem 3.2. Finally in section 5 we give some discussions.

2 Preliminaries on identifiability of mixture distributions and strong consistency

In this section, we consider the identifiability and strong consistency of finite mixtures. The properties of finite mixtures treated in this section concerns general finite mixture distributions.

A mixture of M densities with parameter $\theta = (\alpha_1, \eta_1, \dots, \alpha_M, \eta_M)$ is defined by

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

where α_m, $m = 1, \dots, M$, called the mixing weights, are nonnegative real numbers that sum to one and $f_m(x; \eta_m)$ are densities with parameter η_m. $f_m(x; \eta_m)$ are called the components of the mixture. Let Θ denote the parameter space.

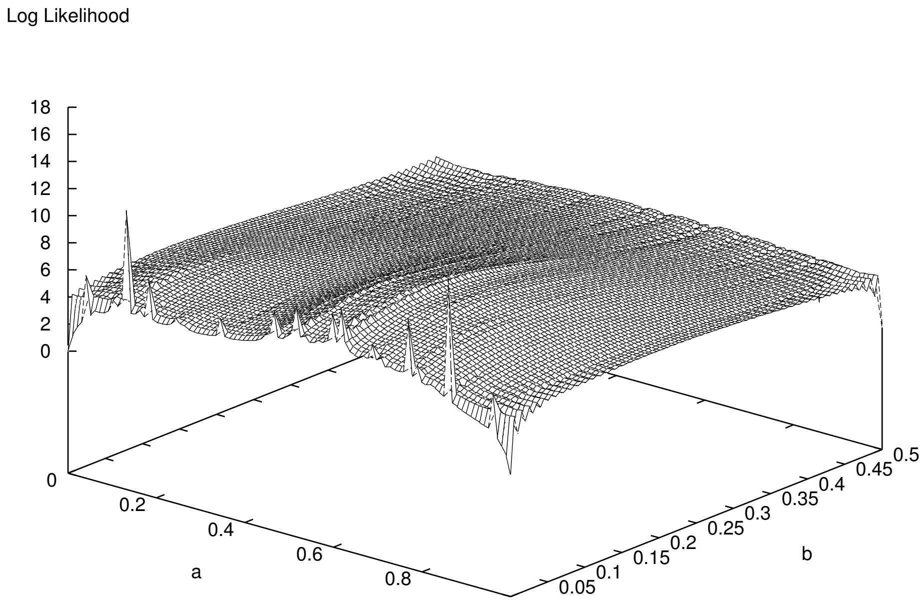

Figure 2: An example of log likelihood function for $n = 40$

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 θ_1, we have the same distribution. In unidentifiable case, true model may consist of two or more points in the parameter space. Therefore we have to carefully define strong consistency of estimator. The following definition is essentially the same as Redner's(1981). We suppose that the parameter space Θ is a subset of Euclidean space and $\text{dist}(\theta, \theta')$ denotes the Euclidean distance between $\theta, \theta' \in \Theta$.

Definition 2.1. (*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. An estimator $\hat{\theta}_n$ is strongly consistent if

$$\text{Prob}\left(\lim_{n\to\infty} \text{dist}(\hat{\theta}_n, \Theta_0) = 0\right) = 1$$

for all Θ_0.

In this paper two notations $\text{Prob}(A) = 1$ and $A, a.e.$, will be used interchangeably.

In finite mixture case, regularity conditions for strong consistency of maximum likelihood estimator are given in Redner(1981). When the components of the mixture are the densities of continuous distributions and the parameter space is Euclidean, the conditions become as follows. Let Γ denote a subset of the parameter space.

Condition 1. Γ is a compact subset of Euclidean space.

For $\theta \in \Gamma$ and any positive real number r, let

$$\begin{aligned} f(x;\theta,r) &= \sup_{\mathrm{dist}(\theta',\theta)\le r} f(x;\theta'), \\ f^*(x;\theta,r) &= \max(1, f(x;\theta,r)) \ . \end{aligned}$$

Condition 2. For each $\theta \in \Gamma$ and sufficiently small r, $f(x;\theta,r)$ is measurable and

$$\int \log(f^*(x;\theta,r)) f(x;\theta_0)dx < \infty \ . \tag{2.1}$$

Condition 3. 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 $\{\theta_n\}_{n=1}^{\infty}$.

Condition 4.

$$\int |\log f(x;\theta_0)|\, f(x;\theta_0)dx < \infty. \tag{2.2}$$

The following two theorems have been proved by Wald(1949), Redner(1981).

Theorem 2.2. (Wald(1949), Redner(1981)) *Suppose that* Conditions 1, 2, 3 and 4 *are satisfied. Let S be any closed subset of Γ not intersecting Θ_0. Then*

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

Theorem 2.3. (Wald(1949)) *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 $\mathrm{Prob}(\lim_{n\to\infty}\tilde\theta_n = \theta_0) = 1$.

If Conditions 1, 2, 3 and 4 are satisfied, then it is readily verified by theorems 2.2 and 2.3 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.

Theorem 2.4. (Okamoto(1958)) *Let Z be a random variable following a binomial distribution* $\mathrm{Bin}(n,p)$. *Then for $\delta > 0$*

$$\mathrm{Prob}\left(\frac{Z}{n} - p \ge \delta \right) < \exp(-2n\delta^2). \tag{2.3}$$

3 Main result

Here, we generalize the problem stated in introduction to the problem of mixture of M uniform distributions and then state our main theorem.

A mixture of M uniform densities with parameter θ is defined by

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

where $f_m(x;\eta_m) \equiv f_m(x;a_m,b_m)$, $m = 1,\ldots,M$, are uniform densities with parameter $\eta_m = (a_m, b_m)$ on half-open intervals $[a_m - b_m, a_m + b_m)$ and α_m are mixing weights. The parameter space $\Theta \subset \mathbb{R}^{3M}$ is defined by

$$\Theta \equiv \{(\alpha_1, a_1, b_1, \ldots, \alpha_M, a_M, b_M) \mid 0 \le \alpha_1, \ldots, \alpha_M \le 1 \ , \ \sum_{m=1}^{M} \alpha_m = 1 \ , \ b_1, \ldots, b_M > 0\} \ .$$

Let $\theta_0 \equiv (\alpha_{0,1}, a_{0,1}, b_{0,1}, \ldots, \alpha_{0,M}, a_{0,M}, b_{0,M})$ be the true parameter and let

$$f(x;\theta_0) = \sum_{m=1}^{M} \alpha_{0,m} \ f_m(x; a_{0,m}, b_{0,m})$$

be the true density. Denote the minimum and the maximum of the support of $f(x;\theta_0)$ by

$$\begin{aligned} L_{\min} &= \min(a_{0,1} - b_{0,1}, \ldots, a_{0,M} - b_{0,M}), \\ L_{\max} &= \max(a_{0,1} + b_{0,1}, \ldots, a_{0,M} + b_{0,M}), \end{aligned}$$

and let

$$L = L_{\max} - L_{\min}.$$

Let Θ_c be a constrained parameter space

$$\Theta_c \equiv \{\theta \in \Theta \mid b_m \ge c > 0 \ , \ m = 1, \ldots, M\},$$

where c is a positive real constant. We can easily see that Conditions 1, 2, 3 and 4 are satisfied with Θ_c. Therefore if $\theta_0 \in \Theta_c$, then the maximum likelihood estimator restricted to Θ_c is strongly consistent (Redner(1981)). But there is a problem of how small c must be to ensure $\theta_0 \in \Theta_c$ as discussed in section 1.

Since the support of uniform density is compact, the following lemma holds.

Lemma 3.1. *For any parameter $\theta = (\alpha_1, a_1, b_1, \ldots, \alpha_M, a_M, b_M) \in \Theta$, there exist a parameter $\theta' = (\alpha_1, a'_1, b'_1, \ldots, \alpha_M, a'_M, b'_M) \in \Theta$ satisfying*

$$L_{\min} \le a'_1, \ldots, a'_M \le L_{\max}, \quad 0 < b'_1, \ldots, b'_M \le L$$

such that

$$\sum_{m=1}^{M} \alpha_m f_m(x; a'_m, b'_m) \ge \sum_{m=1}^{M} \alpha_m f_m(x; a_m, b_m), \quad \forall x \in [L_{\min}, L_{\max}).$$

By lemma 3.1, the maximum likelihood estimator is restricted to a bounded set in $\Theta \subset \mathbb{R}^{3M}$. Let $\{c_n\}_{n=0}^{\infty}$ be a monotone decreasing sequence of positive real numbers converging to zero and define Θ_n by

$$\Theta_n \equiv \{\theta \in \Theta \mid L_{\min} \le a_m \le L_{\max},\ 0 < c_n \le b_m \le L,\ m = 1, \dots, M\} .$$

We are now ready to state our main theorem.

Theorem 3.2. *Suppose that true model $f(x;\theta_0)$ can be represented only by the model which consists of M components. Let $c_0 > 0$ and $0 < d < 1$. If $c_n = c_0 \exp(-n^d)$, then the maximum likelihood estimator restricted to Θ_n is strongly consistent.*

Because in theorem 3.2 we are thinking of d close to 1, we assume $d > 1/4$ hereafter. The next section is devoted to a proof of this theorem.

4 Proof of the strong consistency

Here we present a proof of theorem 3.2. The whole proof is long and we divide it into smaller steps. Intermediate results will be given in a series of lemmas.

Define

$$\Theta'_n \equiv \{\theta \in \Theta \mid \exists m \text{ s.t. } c_n \le b_m \le c_0\}$$

and

$$\Gamma_0 \equiv \{\theta \in \Theta \mid c_0 \le b_m \le L\ ,\ m = 1, \dots, M\} .$$

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 $\Theta_0 \subset \Gamma_0$.

In view of theorems 2.2, 2.3, 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 2.2. Therefore it suffices to prove

$$\lim_{n\to\infty} \frac{\sup_{\theta \in \Theta'_n} \prod_{i=1}^n f(x_i;\theta)}{\prod_{i=1}^n f(x_i;\theta_0)} = 0, \quad a.e. \tag{4.1}$$

Note that in the argument above the supremum of the likelihood function over $S \cup \Theta'_n$ is considered separately for S and Θ'_n. S and Θ'_n form a covering of $S \cup \Theta'_n$. In our proof, we consider finer and finer finite coverings of Θ'_n. As above, it suffices to prove that the ratio of the supremum of the likelihood over each member of the covering to the likelihood at θ_0 converges to zero almost everywhere.

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

$$\Theta'_{n,K} \equiv \{\theta \in \Theta'_n \mid b_1 \leq b_2 \leq \cdots \leq b_K \leq c_0 < b_{K+1} \leq \cdots \leq b_M\} .$$

Our first covering of Θ'_n is given by

$$\Theta'_n = \bigcup_{K=1}^{M} \Theta'_{n,K}.$$

As above, it suffices to prove that for each K, $1 \leq K \leq M$,

$$\lim_{n\to\infty} \frac{\sup_{\theta\in\Theta'_{n,K}} \prod_{i=1}^n f(x_i;\theta)}{\prod_{i=1}^n f(x_i;\theta_0)} = 0, \quad a.e. \tag{4.2}$$

We fix K from now on. Define $\bar{\Theta}_K$ by

$$\begin{aligned} \bar{\Theta}_K \;\; \equiv \;\; & \{(\alpha_{K+1}, a_{K+1}, b_{K+1}, \ldots, \alpha_M, a_M, b_M) \in \mathbb{R}^{3(M-K)} \\ & \mid \sum_{m=K+1}^{M} \alpha_m \leq 1 \ , \ \alpha_m \geq 0 \ , \ c_0 \leq b_m \leq L \ , \ m = K+1, \ldots, M\} \end{aligned}$$

and for $\bar{\theta} \in \bar{\Theta}_K$, define

$$\begin{aligned} \bar{f}(x;\bar{\theta}) \;\; &\equiv \;\; \sum_{m=K+1}^{M} \alpha_m f_m(x;\eta_m) \ , \\ \bar{f}(x;\bar{\theta},\rho) \;\; &\equiv \;\; \sup_{\mathrm{dist}(\bar{\theta},\bar{\theta}')\leq\rho} \bar{f}(x;\bar{\theta}') \ . \end{aligned}$$

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

Lemma 4.1. *Let $B(\bar{\theta}, \rho(\bar{\theta}))$ denote the open ball with center $\bar{\theta}$ and radius $\rho(\bar{\theta})$. Then $\bar{\Theta}_K$ can be covered by a finite number of balls $B(\bar{\theta}^{(1)}, \rho(\bar{\theta}^{(1)})), \ldots, B(\bar{\theta}^{(S)}, \rho(\bar{\theta}^{(S)}))$ such that*

$$E_0[\log \bar{f}(x;\bar{\theta}^{(s)}, \rho(\bar{\theta}^{(s)}))] < E_0[\log f(x;\theta_0)] \ , \quad s = 1, \ldots, S, \tag{4.3}$$

where $E_0[\cdot]$ denotes the expectation under θ_0.

Proof: The proof is the same as in Wald (1949). For all $\bar{\theta} \in \bar{\Theta}_K$, there exists a positive real number $\rho(\bar{\theta})$ which satisfies

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

Since $\bar{\Theta}_K \subset \bigcup_{\bar{\theta}} B(\bar{\theta}, \rho(\bar{\theta}))$ and $\bar{\Theta}_K$ is compact, there exists a finite number of balls $B(\bar{\theta}^{(1)}, \rho(\bar{\theta}^{(1)})), \ldots, B(\bar{\theta}^{(S)}, \rho(\bar{\theta}^{(S)}))$ which cover $\bar{\Theta}_K$. □

Define

$$\Theta'_{n,K,s} \equiv \{\theta \in \Theta'_{n,K} \mid (\alpha_{K+1}, a_{K+1}, b_{K+1}, \ldots, \alpha_M, a_M, b_M) \in B(\bar{\theta}^{(s)}, \rho(\bar{\theta}^{(s)}))\}$$

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

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

Again it suffices to prove that for each s, $s = 1, \ldots, S$,

$$\lim_{n\to\infty} \frac{\sup_{\theta \in \Theta'_{n,K,s}} \prod_{i=1}^n f(x_i; \theta)}{\prod_{i=1}^n f(x_i; \theta_0)} = 0, \quad a.e. \tag{4.4}$$

We fix s in addition to K from now on.

Because

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

(4.4) is implied by

$$\limsup_{n\to\infty} \frac{1}{n} \sup_{\theta \in \Theta'_{n,K,s}} \sum_{i=1}^n \log f(x_i; \theta) < E_0[\log f(x; \theta_0)], \quad a.e. \tag{4.5}$$

Therefore it suffices to prove (4.5), which is a new intermediate goal of our proof hereafter.

Choose G, $0 < G < 1$, such that

$$\lambda \equiv E_0[\log f(x; \theta_0)] - E_0[\log \{\bar{f}(x; \bar{\theta}^{(s)}, \rho(\bar{\theta}^{(s)})) + G\}] > 0 \ . \tag{4.6}$$

Let $u \equiv \max_x f(x; \theta_0)$. Because $\{c_n\}$ is decreasing to zero, by replacing c_0 by some c_n if necessary, we can again assume without loss of generality that c_0 is small enough to satisfy

$$\begin{aligned} 2c_0 &< e^{-1}, \\ 3M \cdot u \cdot 2c_0 \cdot (-\log G) &< \frac{\lambda}{4}, \qquad (4.7) \\ 2M \cdot u \cdot 2c_0 \cdot \log \frac{1}{2c_0} &< \frac{\lambda}{12} \ . \qquad (4.8) \end{aligned}$$

We now prove the following lemma.

Lemma 4.2. *Let $J(\theta)$ denote the support of $\sum_{m=1}^{K} \alpha_m f_m(x;\eta_m)$ and let $R_n(V)$ denote the number of observations which belong to a set $V \subset \mathbb{R}$. Then for $\theta \in \Theta'_{n,K,s}$*

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} \log f(x_i;\theta) \quad &\leq \quad \frac{1}{n}\sum_{i=1}^{n} \log\left\{\bar{f}(x_i;\bar{\theta}^{(s)},\rho(\bar{\theta}^{(s)})) + G\right\} \\ &\quad + \frac{1}{n}\sum_{x_i\in J(\theta)} \log f(x_i;\theta) + \frac{1}{n}R_n(J(\theta))\cdot(-\log G)\ . \qquad (4.9)\end{aligned}$$

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

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} \log f(x_i;\theta) &= \frac{1}{n}\sum_{x_i\in J(\theta)} \log f(x_i;\theta) + \frac{1}{n}\sum_{x_i\notin J(\theta)} \log\left\{\sum_{m=K+1}^{M} \alpha_m f_m(x_i;\eta_m)\right\} \\ &\leq \frac{1}{n}\sum_{i=1}^{n} \log\left\{\sum_{m=K+1}^{M} \alpha_m f_m(x_i;\eta_m) + G\right\} \\ &\quad + \frac{1}{n}\sum_{x_i\in J(\theta)} \left[\log f(x_i;\theta) - \log\left\{\sum_{m=K+1}^{M} \alpha_m f_m(x_i;\eta_m) + G\right\}\right] \\ &\leq \frac{1}{n}\sum_{i=1}^{n} \log\left\{\bar{f}(x_i;\bar{\theta}^{(s)},\rho(\bar{\theta}^{(s)})) + G\right\} \\ &\quad + \frac{1}{n}\sum_{x_i\in J(\theta)} \log f(x_i;\theta) - \frac{1}{n}R_n(J(\theta))\log G\ .\end{aligned}$$

□

We want to bound the terms on the right hand side of (4.9) from above. The first term is easy. In fact by (4.6) and the strong law of large numbers we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \log\left\{\bar{f}(x_i;\bar{\theta}^{(s)},\rho(\bar{\theta}^{(s)})) + G\right\} = E_0[\log f(x;\theta_0)] - \lambda, \quad a.e. \qquad (4.10)$$

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

Lemma 4.3.

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

Proof: Let $\epsilon > 0$ be arbitrarily fixed and let J_0 be the support of the true density. J_0 consists of some intervals. We divide J_0 from $L_{\min}$ to $L_{\max}$ by short intervals of length $2c_0$. In each right end of the intervals of J_0, overlap of two short intervals of length $2c_0$ is allowed and the right end of a short interval coincides with the right end of an interval of J_0. See Figure 3. Let $k(c_0)$ be the number of short intervals and let $I_1(c_0),\dots,I_{k(c_0)}(c_0)$

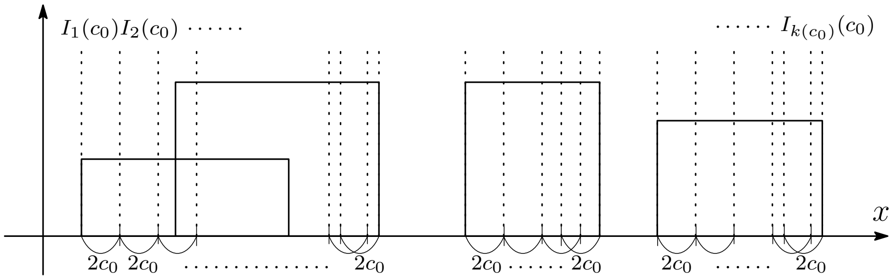

Figure 3: Division of J_0 by short intervals of length $2c_0$.

be the divided short intervals. Because J_0 consists of at most M intervals, we have

$$k(c_0) \leq \frac{L}{2c_0} + M \ .$$

Note that any interval in J_0 of length $2c_0$ is covered by at most 3 small intervals from $\{I_1(c_0), \ldots, I_{k(c_0)}(c_0)\}$. Now consider $J(\theta)$, the support of $\sum_{m=1}^{K} \alpha_m f_m(x; \eta_m)$. The support of each $f_m(x; \eta_m)$, $1 \leq m \leq K$, is an interval of length less than or equal to $2c_0$. Therefore $J(\theta)$ is covered by at most $3M$ short intervals. Then

$$\begin{aligned} &\sup_{\theta \in \Theta'_{n,K,s}} \frac{1}{n} R_n(J(\theta)) - 3M \cdot u \cdot 2c_0 > \epsilon \\ &\Rightarrow \quad 1 \leq \exists k \leq k(c_0) \ , \ \frac{1}{n} R_n(I_k(c_0)) - u \cdot 2c_0 > \frac{\epsilon}{3M} \ . \end{aligned} \tag{4.11}$$

From (4.11), we have

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

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

$$P_0(V) \equiv \int_V f(x; \theta_0) dx \ .$$

Then

$$P_0(I_k(c_0)) \leq u \cdot 2c_0, \quad k = 1, \ldots, k(\theta) \ . \tag{4.12}$$

Since $R_n(V) \sim \text{Bin}(n, P_0(V))$ and from (2.3), we obtain

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

Therefore

$$\text{Prob}\left(\sup_{\theta\in\Theta'_{n,K,s}} \frac{1}{n}R_n(J(\theta)) - 3M\cdot u\cdot 2c_0 > \epsilon\right) \le \left(\frac{L}{2c_0} + M\right)\exp\left(-\frac{2n\epsilon^2}{9M^2}\right).$$

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

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

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

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

□

By this lemma and (4.7) we have

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta'_{n,K,s}} \frac{1}{n}R_n(J(\theta))\cdot(-\log G) \le 3M\cdot u\cdot 2c_0\cdot(-\log G) < \frac{\lambda}{4}. \tag{4.13}$$

This bounds the third term on the right hand side of (4.9) from above.

Finally we bound the second term on the right hand side of (4.9) from above. This is the most difficult part of our proof. For $x \in J(\theta)$ write $f(x;\theta) = \sum_{m=1}^{M} \alpha_m f_m(x;\eta_m)$ as

$$f(x;\theta) = \frac{1}{n}\sum_{t=1}^{T(\theta)} H(J_t(\theta))1_{J_t(\theta)}(x), \tag{4.14}$$

where $J_t \equiv J_t(\theta)$ are disjoint half-open intervals, $1_{J_t(\theta)}(x)$ is the indicator function,

$$H(J_t(\theta)) = f(x;\theta), \quad x \in J_t(\theta),$$

is the height of $f(x;\theta)$ on $J_t(\theta)$ and $T(\theta)$ is the number of the intervals $J_t(\theta)$. Note that $T(\theta) \le 2M$, because $f(x;\theta)$ changes its height only at $a_m - b_m$ or $a_m + b_m$, $m = 1,\ldots,M$. For convenience we determine the order of t such that

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

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

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

and define $\tau_n(\theta)$

$$\tau_n(\theta) \equiv \max\{t \in \{1, \ldots, T\} \mid H(J_t(\theta)) \le \frac{M}{2c'_n}\}. \tag{4.15}$$

Then the second term on the right hand side of (4.9) is written as

$$\begin{aligned} \frac{1}{n} \sum_{x_i \in J(\theta)} \log f(x_i;\theta) &= \sum_{t=1}^{T(\theta)} \frac{1}{n} \sum_{x_i \in J_t(\theta)} \log H(J_t(\theta)) \\ &= \frac{1}{n} \sum_{t=1}^{T(\theta)} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) \\ &= \frac{1}{n} \sum_{t=1}^{\tau_n(\theta)} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) \\ &\qquad + \frac{1}{n} \sum_{t=\tau_n(\theta)+1}^{T(\theta)} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)). \end{aligned} \tag{4.16}$$

From (4.7), (4.8), and noting that $\log x/x$ is decreasing in $x \ge e$, we have

$$\begin{aligned} 3 \sum_{t=1}^{\tau_n(\theta)} \frac{u}{H(J_t(\theta))} \log H(J_t(\theta)) &\le 3 \cdot 2M \cdot u \cdot 2c_0 \cdot \log \frac{1}{2c_0} < \frac{\lambda}{4}, \\ \sum_{t=\tau_n(\theta)+1}^{T(\theta)} 3 \cdot \frac{2}{n} \log H(J_t(\theta)) &\le 3 \cdot 2M \cdot \frac{2}{n} \cdot (n^d - \log \frac{M}{2c_0}) \to 0. \end{aligned} \tag{4.17}$$

Suppose that the following inequality holds.

$$\begin{aligned} \limsup_{n\to\infty} \sup_{\theta \in \Theta'_{n,K,s}} & \Bigg[\sum_{t=1}^{T(\theta)} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) \\ & - 3 \left\{ \sum_{t=1}^{\tau_n(\theta)} \frac{u}{H(J_t(\theta))} \log H(J_t(\theta)) + \sum_{t=\tau_n(\theta)+1}^{T(\theta)} \frac{2}{n} \log H(J_t(\theta)) \right\} \Bigg] \le 0, \quad a.e. \end{aligned} \tag{4.18}$$

Then from (4.16) and (4.17), the second term on the right hand side of (4.9) is bounded from above as

$$\limsup_{n\to\infty} \frac{1}{n} \sup_{\theta \in \Theta'_{n,K,s}} \sum_{x_i \in J(\theta)} \log f(x_i;\theta) \le \frac{4}{\lambda}. \tag{4.19}$$

Combining (4.10), (4.13) and (4.19) we obtain

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

and (4.5) is satisfied. Therefore it suffices to prove (4.18), which is a new goal of our proof.

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

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

Then

$$\begin{aligned}&\sup_{\theta\in\Theta'_{n,K,s}} \Bigg[\sum_{t=1}^{T(\theta)}\frac{1}{n}R_n(J_t(\theta))\log H(J_t(\theta)) \\ &\qquad -3\left\{\sum_{t=1}^{\tau_n(\theta)}\frac{u}{H(J_t(\theta))}\log H(J_t(\theta))+\sum_{t=\tau_n(\theta)+1}^{T(\theta)}\frac{2}{n}\log H(J_t(\theta))\right\}\Bigg] \\ &\le \max_{T=1,\ldots,2M}\max_{\tau=1,\ldots,T}\Bigg[\\ &\qquad \sup_{\theta\in\Theta'_{n,K,s,T,\tau}}\left\{\sum_{t=1}^{\tau}\frac{1}{n}R_n(J_t(\theta))\log H(J_t(\theta))-3\sum_{t=1}^{\tau}\frac{u}{H(J_t(\theta))}\log H(J_t(\theta))\right\} \\ &\qquad +\sup_{\theta\in\Theta'_{n,K,s,T,\tau}}\left\{\sum_{t=\tau+1}^{T}\frac{1}{n}R_n(J_t(\theta))\log H(J_t(\theta))-3\sum_{t=\tau+1}^{T}\frac{2}{n}\log H(J_t(\theta))\right\}\Bigg]\ .\end{aligned} \tag{4.20}$$

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

$$\begin{aligned}\limsup_{n\to\infty}\sup_{\theta\in\Theta'_{n,K,s,T,\tau}}\Bigg[&\sum_{t=1}^{\tau}\frac{1}{n}R_n(J_t(\theta))\log H(J_t(\theta)) \\ &-3\sum_{t=1}^{\tau}\frac{u}{H(J_t(\theta))}\log H(J_t(\theta))\Bigg]\le 0, \quad a.e.\end{aligned} \tag{4.21}$$

$$\limsup_{n\to\infty}\sup_{\theta\in\Theta'_{n,K,s,T,\tau}}\left[\sum_{t=\tau+1}^{T}\frac{1}{n}R_n(J_t(\theta))\log H(J_t(\theta))-3\sum_{t=\tau+1}^{T}\frac{2}{n}\log H(J_t(\theta))\right]\le 0, \quad a.e. \tag{4.22}$$

Then (4.18) is derived from (4.20), (4.21), (4.22). Therefore it suffices to prove (4.21) and (4.22), which are the final goals of our proof. We state (4.21) and (4.22) as two lemmas and give their proofs.

Lemma 4.4.

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

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

$$\sup_{\theta\in\Theta'_{n,K,s,T,\tau}} \left[\sum_{t=\tau+1}^{T} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) - 3 \sum_{t=\tau+1}^{T} \frac{2}{n} \log H(J_t(\theta)) \right] > 2M\delta. \tag{4.23}$$

Noting that for $t > \tau$, the length of $J_t(\theta)$ is less than or equal to $2c'_n$, the following relation holds for this event.

$$\begin{aligned}
&\text{The event (4.23) occurs.}\\
\Rightarrow\ & \sup_{\theta\in\Theta'_{n,K,s,T,\tau}} \left[\sum_{t=\tau+1}^{T} \max\left\{0, \left(\frac{1}{n} R_n([a'_t(\theta) - c'_n, a'_t(\theta) + c'_n]) \right.\right.\right.\\
&\qquad\qquad\qquad \left.\left.\left. -3\cdot\frac{2}{n}\right)\right\} \log \frac{M}{2c_n} \right] > 2M\delta\\
\Rightarrow\ & \exists\theta \in \Theta'_{n,K,s,T,\tau}, \exists t > \tau\\
& \max\left\{0, \left(\frac{1}{n} R_n([a'_t(\theta) - c'_n, a'_t(\theta) + c'_n]) - 3\cdot\frac{2}{n}\right)\right\} \log \frac{M}{2c_n} > \delta\\
\Rightarrow\ & \exists\theta \in \Theta'_{n,K,s,T,\tau}, \exists t > \tau\\
& R_n([a'_t(\theta) - c'_n, a'_t(\theta) + c'_n]) \geq 6\\
\Rightarrow\ & \sup_{L_{\min}\leq a'\leq L_{\max}} R_n([a' - c'_n, a' + c'_n]) \geq 6\ .
\end{aligned} \tag{4.24}$$

Below, we consider the probability of the event that (4.24) occurs. We divide J_0 from $L_{\min}$ to $L_{\max}$ by short intervals of length $2c'_n$ as in the proof of lemma 4.3. Let $k(c'_n)$ be the number of short intervals and let $I_1(c'_n), \ldots, I_{k(c'_n)}(c'_n)$ be the divided short intervals. Then we have

$$k(c'_n) \leq \frac{L}{2c'_n} + M\ . \tag{4.25}$$

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

$$\sup_{L_{\min}\leq a'\leq L_{\max}} R_n([a' - c'_n, a' + c'_n]) \geq 6 \Rightarrow 1 \leq \exists k \leq k(c'_n)\ ,\ R_n(I_k(c'_n)) \geq 2\ .$$

Note that $R_n(I_k(c'_n)) \sim \mathrm{Bin}(n, P_0(I_k(c'_n)))$ and $P_0(I_k(c'_n)) \le 2c'_n u$ Therefore from (4.24) and (4.25) we have

$$\begin{aligned}
\mathrm{Prob}\left(\sup_{\theta\in\Theta'_{n,K,s,T,\tau}} \left\{ \sum_{t=\tau+1}^{T} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) - 3 \sum_{t=\tau+1}^{T} \frac{2}{n} \log H(J_t(\theta)) \right\} > 2M\delta \right) \\
\le \left(\frac{L}{2c'_n} + M \right) \sum_{k=2}^{n} \binom{n}{k} (2c'_n u)^k (1-2c'_n u)^{n-k} \\
\le \left(\frac{L}{2c'_n} + M \right) \sum_{k=2}^{n} \frac{n^k}{k!} (2c'_n u)^k \\
\le \left(\frac{L}{2c'_n} + M \right) (2nc'_n u)^2 \sum_{k=0}^{n} \frac{1}{k!} (2nc'_n u)^k \\
\le \left(\frac{L}{2c'_n} + M \right) (2nc'_n u)^2 \exp(2nc'_n u) \, .
\end{aligned}$$

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

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

□

Finally we prove (4.21).

Lemma 4.5.

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

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

$$h_n \equiv \frac{\delta}{12} \left\{ u \log \left(\frac{M}{c'_n} \right) \right\}^{-1} . \tag{4.26}$$

We divide $[c'_n/M, c_0]$ from c_0 to c'_n/M by short intervals of length h_n. In the left end c'_n/M of the interval $[c'_n/M, c_0]$, overlap of two short intervals of length h_n is allowed and the left end of a short interval is equal to c'_n/M. Let l_n be the number of short intervals of length h_n and define $b_l^{(n)}$ by

$$b_l^{(n)} \equiv \begin{cases} c_0 - (l-1)h_n, & 1 \le l \le l_n, \\ c'_n/M, & l = l_n + 1. \end{cases}$$

Then we have

$$l_n \le \frac{c_0}{h_n} + 1 \ . \tag{4.27}$$

Next, we consider the probability of the event that

$$\sup_{\theta \in \Theta'_{n,K,s,T,\tau}} \left[\sum_{t=1}^{\tau} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) - 3 \sum_{t=1}^{\tau} \frac{u}{H(J_t(\theta))} \log H(J_t(\theta)) \right] > 2M\delta. \tag{4.28}$$

For this event the following relation holds.

$$\begin{aligned}
&\text{The event (4.28) occurs.} \\
&\Rightarrow \ \exists \theta \in \Theta'_{n,K,s,T,\tau} \\
&\quad\ 1 \le \exists l(1), \cdots, \exists l(\tau) \le l_n \quad \text{s.t.} \\
&\quad\ 2b^{(n)}_{l(1)+1} \le \frac{1}{H(J_1(\theta))} \le 2b^{(n)}_{l(1)}, \cdots, 2b^{(n)}_{l(\tau)+1} \le \frac{1}{H(J_\tau(\theta))} \le 2b^{(n)}_{l(\tau)}, \\
&\quad\ \sum_{t=1}^{\tau} \max\left\{0, \left(\frac{1}{n} R_n([a'_t(\theta) - b^{(n)}_{l(t)}, a'_t(\theta) + b^{(n)}_{l(t)}]) - 3u \cdot 2b^{(n)}_{l(t)+1}\right)\right\} \log \frac{1}{2b^{(n)}_{l(t)+1}} > 2M\delta \\
&\Rightarrow \ \exists \theta \in \Theta'_{n,K,s,T,\tau} \ , \ 1 \le \exists t \le \tau \\
&\quad\ 1 \le \exists l(t) \le l_n \quad \text{s.t.} \\
&\quad\ 2b^{(n)}_{l(t)+1} \le \frac{1}{H(J_t(\theta))} \le 2b^{(n)}_{l(t)}, \\
&\quad\ \max\left\{0, \left(\frac{1}{n} R_n([a'_t(\theta) - b^{(n)}_{l(t)}, a'_t(\theta) + b^{(n)}_{l(t)}]) - 3u \cdot 2b^{(n)}_{l(t)+1}\right)\right\} \log \frac{1}{2b^{(n)}_{l(t)+1}} > \delta \\
&\Rightarrow \ 1 \le \exists l \le l_n \quad s.t. \\
&\quad\ \max\left\{0, \sup_{L_{\min} \le a' \le L_{\max}} \left(\frac{1}{n} R_n([a' - b^{(n)}_l, a' + b^{(n)}_l]) - 3u \cdot 2b^{(n)}_{l+1}\right)\right\} \log \frac{1}{2b^{(n)}_{l+1}} > \delta \\
&\Rightarrow \ 1 \le \exists l \le l_n \quad s.t. \\
&\quad\ \sup_{L_{\min} \le a' \le L_{\max}} \left\{ \left(\frac{1}{n} R_n([a' - b^{(n)}_l, a' + b^{(n)}_l]) - 3u \cdot 2b^{(n)}_l\right) \log \frac{1}{2b^{(n)}_{l+1}} \right. \\
&\qquad\quad \left. + 3u(2b^{(n)}_l - 2b^{(n)}_{l+1}) \log \frac{1}{2b^{(n)}_{l+1}} \right\} > \delta
\end{aligned} \tag{4.29}$$

Then from (4.26) the following relation holds.

$$\begin{aligned}
&\text{The event (4.29) occurs.} \\
&\Rightarrow \ 1 \le \exists l \le l_n \ , \ \sup_{L_{\min} \le a' \le L_{\max}} \frac{1}{n} \left(R_n([a' - b^{(n)}_l, a' + b^{(n)}_l]) - 3u \cdot 2b^{(n)}_l\right) \log \frac{1}{2b^{(n)}_{l+1}} > \frac{\delta}{2}
\end{aligned} \tag{4.30}$$

Below, we consider the probability of the event that (4.30) occurs. We divide J_0 from $L_{\min}$ to $L_{\max}$ by short intervals of length $2b_l^{(n)}$ as in the proof of lemma 4.3. Let $k(b_l^{(n)})$ be the number of short intervals and let $I_1(b_l^{(n)}),\ldots,I_{k(b_l^{(n)})}(b_l^{(n)})$ be the divided short intervals. Then we have

$$k(b_l^{(n)}) \le \frac{L}{2b_l^{(n)}} + M\ . \tag{4.31}$$

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

$$\begin{aligned}
&\sup_{L_{\min}\le a'\le L_{\max}} \left(\frac{1}{n}R_n([a'-b_l^{(n)},a'+b_l^{(n)}]) - 3u\cdot 2b_l^{(n)}\right) > \frac{\delta}{2}\left(\log\frac{1}{2b_{l+1}^{(n)}}\right)^{-1} \\
\Rightarrow &\max_{k=1,\ldots,k(b_l^{(n)})} \left(\frac{1}{n}R_n(I_k(b_l^{(n)})) - u\cdot 2b_l^{(n)}\right) > \frac{1}{3}\cdot\frac{\delta}{2}\left(\log\frac{1}{2b_{l+1}^{(n)}}\right)^{-1}\ .
\end{aligned}\tag{4.32}$$

Note that $R_n(I_k(b_l^{(n)})) \sim \mathrm{Bin}(n, P_0(I_k(b_l^{(n)})))$ and $P_0(I_k(b_l^{(n)})) \le u\cdot 2b_l^{(n)}$. Therefore from (2.3) and (4.31) we have

$$\begin{aligned}
\mathrm{Prob}&\left(\max_{k=1,\ldots,k(b_l^{(n)})} \frac{1}{n}\left(R_n(I_k(b_l^{(n)})) - u\cdot 2b_l^{(n)}\right) > \frac{1}{3}\cdot\frac{\delta}{2}\left(\log\frac{1}{2b_{l+1}^{(n)}}\right)^{-1}\right) \\
&\le \left(\frac{L}{2b_l^{(n)}} + M\right)\exp\left\{-2n\cdot\frac{\delta^2}{36}\left(\log\frac{1}{2b_{l+1}^{(n)}}\right)^{-2}\right\} \\
&\le \left(\frac{L}{2c_n'} + M\right)\exp\left\{-2n\cdot\frac{\delta^2}{36}\left(\log\frac{1}{2c_n'}\right)^{-2}\right\}.
\end{aligned}\tag{4.33}$$

From (4.27), (4.29), (4.30), (4.32), (4.33), we obtain

$$\begin{aligned}
\mathrm{Prob}&\left(\sup_{\theta\in\Theta'_{n,K,s,T,\tau}}\left[\sum_{t=1}^{\tau}\frac{1}{n}R_n(J_t(\theta))\log H(J_t(\theta)) - 3\sum_{t=1}^{\tau}\frac{u}{H(J_t(\theta))}\log H(J_t(\theta))\right] > 2M\delta\right) \\
&\le \left(\frac{c_0}{h_n}+1\right)\left(\frac{L}{2c_n'} + M\right)\exp\left\{-2n\cdot\frac{\delta^2}{36}\left(\log\frac{1}{2c_n'}\right)^{-2}\right\}.
\end{aligned}$$

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

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

□

This completes the proof of theorem 3.2.

5 Some discussions

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 $c_n = \exp(-n^d)$, $d < 1$, obtained in this paper is almost the lower bound of the order of c_n which maintains the consistency.

We expect that our result can be extended to other finite mixture cases, especially for densities which are Lipschitz continuous when the scale parameters are fixed. The problem studied in this paper is similar to the question stated in Hathaway(1985) which treats the normal mixtures and the constraint is imposed on the ratios of variances. Methods used in this paper may be useful to solve the question.

References

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

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

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

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

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