Ann. Inst. Statist. Math.
Vol. 57, No. 1, 1–19 (2005)

STRONG CONSISTENCY OF MLE FOR FINITE UNIFORM MIXTURES WHEN THE SCALE PARAMETERS ARE EXPONENTIALLY SMALL

KENTARO TANAKA[1] AND AKIMICHI TAKEMURA[2]

[1] *The Department of Industrial Engineering and Management, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan*
[2] *Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan*

(Received May 23, 2003; revised June 15, 2004)

Abstract. We consider maximum likelihood estimation of finite mixture of uniform distributions. We prove that 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.

Key words and phrases: Mixture distribution, maximum likelihood estimator, consistency.

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. By using half-open intervals, our densities are right continuous and the version of the density is uniquely determined. 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 $x_1, \dots, x_n$ 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 $b_2 \to 0$ (Fig. 1). Hence the maximum likelihood estimator is not consistent. Actually it does not even exist for each finite n.

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 $b_{2,0} \ge c$ since we do not know $b_{2,0}$. 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 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), which treats mixtures of normal distributions with

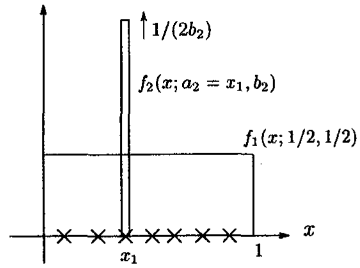

Fig. 1. The likelihood tends to infinity as $b_2 \to 0$ at $a_2 = x_1$.

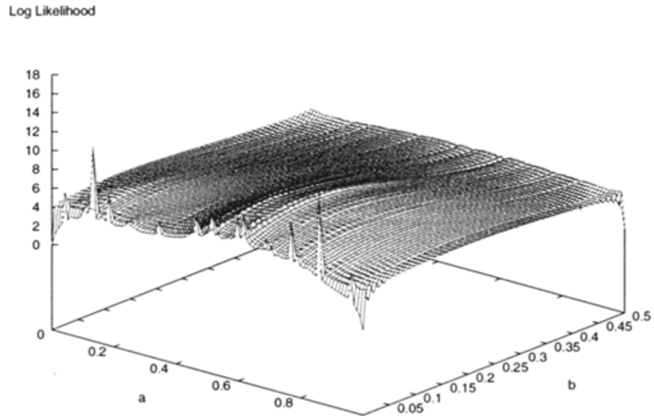

Fig. 2. An example of log likelihood function for $n = 40$.

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 $0.6 \cdot f(x; 0.5, 0.5) + 0.4 \cdot f(x; a, b)$. Despite the limited resolution in Fig. 2, there are actually $n = 40$ peaks of the likelihood function as $b \downarrow 0$. We see that although the likelihood function diverges to infinity at these peaks, the divergence takes place only for very small b and the likelihood function is well-behaved for most of the ranges of b. This suggests that the bound c_n can decrease to zero fairly quickly while maintaining the consistency of maximum likelihood estimator. 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 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 3.1. Proof of Theorem 3.1 is given in Appendix. In Section 4 we give a simulation result and 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 $\theta = (\alpha_1, \eta_1, \ldots, \alpha_M, \eta_M)$ is defined by

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

where α_m, $m = 1, \ldots, M$, 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.

In general, identifiability of a parametric family of densities is defined as follows. Note that in this paper a version of the density is uniquely determined by the right continuity.

Definition 2.1. (Identifiability of a parametric family of densities) A parametric family of densities $\{f(x;\theta) \mid \theta \in \Theta\}$ is identifiable if different values of parameter designate different densities; that is

$$f(x;\theta) = f(x;\theta') \quad \forall x,$$

implies $\theta = \theta'$.

If a parametric family of densities is not identifiable, then it is said to be unidentifiable.

In mixture case, when all components $f_m(x;\eta_m)$, $m = 1, \ldots, M$ belong to the same parametric family, then $f(x;\theta)$ is invariant under the permutations of the component labels. Because of this trivial unidentifiability, the definition of identifiability for the mixture densities can be weakened as described in Teicher (1960), Yakowitz and Spragins (1968), McLachlan and Peel (2000) and so on, so that $\sum_{m=1}^{M} \alpha_m f_m(x;\eta_m) = \sum_{m'=1}^{M'} \alpha'_{m'} f_{m'}(x;\eta'_{m'})$ implies $M = M'$ and for each m there exists some m' such that $\alpha_m = \alpha_{m'}$ and $\eta_m = \eta'_{m'}$. But, even under such a weakened definition, mixtures of density functions still have unidentifiability. For example, if $\alpha_1 = 0$, then for all parameters which differ only in η_1, we have the same density. We also discuss examples of non-trivial unidentifiability of mixtures after Theorem 3.1 below. In any way, mixture model is unidentifiable.

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 $\hat{\theta}_n$, because we should define $\hat{\theta}_n$ to be consistent if $\hat{\theta}_n$ falls in arbitrary small neighborhood of the set of points designating the true model as $n \to \infty$.

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 $\mathrm{dist}(\theta, \theta')$ denotes the Euclidean distance between $\theta, \theta' \in \Theta$.

Definition 2.2. (Strongly consistent estimator) Let T_0 denote the set of true parameters

$$T_0 \equiv \{\theta \in \Theta \mid f(x;\theta) = f(x;\theta_0) \ \forall x\},$$

where θ_0 is one of parameters designating the true distribution. An estimator $\hat{\theta}_n$ is strongly consistent if

$$\mathrm{Prob}\left(\lim_{n\to\infty} \inf_{\theta \in T_0} \mathrm{dist}(\hat{\theta}_n, \theta) = 0\right) = 1.$$

In this paper two notations $\mathrm{Prob}(A) = 1$ and A, a.e. (A holds almost everywhere), will be used interchangeably. The index $_0$ to the parameter always denotes the true parameter.

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

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

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 2.1. (Wald (1949), Redner (1981)) *Suppose that Conditions* 1, 2, 3 *and* 4 *are satisfied. Let S be any closed subset of Γ not intersecting T_0. Then*

$$\mathrm{Prob}\left(\lim_{n\to\infty} \frac{\sup_{\theta\in S} f(x_1;\theta)\times\cdots\times f(x_n;\theta)}{f(x_1;\theta_0)\times\cdots\times f(x_n;\theta_0)} = 0\right) = 1. \tag{2.3}$$

THEOREM 2.2. (Wald (1949), Redner (1981)) *Let $\tilde\theta_n$ be any function of the observations $x_1,\dots,x_n$ 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 $\mathrm{Prob}(\lim_{n\to\infty} \inf_{\theta\in T_0} \mathrm{dist}(\tilde\theta_n,\theta)) = 1$.

If Conditions 1, 2, 3 and 4 are satisfied, then it is readily verified by Theorems 2.1 and 2.2 that maximum likelihood estimator restricted to Γ is strongly consistent.

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

THEOREM 2.3. (Okamoto (1958)) *Let Z be a random variable following a binomial distribution* $\mathrm{Bin}(n,p)$. *Then for $\delta > 0$*

$$\text{Prob}\left(\frac{Z}{n} - p \geq \delta\right) < \exp(-2n\delta^2). \tag{2.4}$$

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

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

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

$$L_{\mathsf{min}} = \min(a_{0,1} - b_{0,1}, \ldots, a_{0,M} - b_{0,M}),$$
$$L_{\mathsf{max}} = \max(a_{0,1} + b_{0,1}, \ldots, a_{0,M} + b_{0,M}),$$

and let

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

Let Θ_c be a constrained parameter space

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

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 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 exists a parameter $\theta' = (\alpha_1, a'_1, b'_1, \ldots, \alpha_M, a'_M, b'_M) \in \Theta$ satisfying*

$$L_{\min} \leq a'_1, \cdots, a'_M \leq L_{\max}, \qquad 0 < b'_1, \ldots, b'_M \leq L$$

such that

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

where equality does not hold if there exists $\alpha_m > 0$ such that $a_m \notin [L_{\min}, L_{\max}]$ or $b_m > L$.

By Lemma 3.1, 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

$$\Theta_n \equiv \{\theta \in \Theta \mid 0 < c_n \leq b_m, m = 1, \ldots, M\}.$$

We are now ready to state our main theorem.

THEOREM 3.1. *Suppose that the true model $f(x; \theta_0)$ can not be represented by any model consisting of less than M components. Let $c_0 > 0$ and $0 < d < 1$. If $c_n = c_0 \exp(-n^d) \leq b_m$ for all b_m, then maximum likelihood estimator (which is restricted to Θ_n) is strongly consistent.*

Proof of this theorem is given in Appendix.

Note that under the assumption of Theorem 3.1 the strong consistency holds even if the true model is unidentifiable in a non-trivial way. We illustrate the assumption of Theorem 3.1 by examples of two-component models. If the true model is $\alpha U(x; 0, \alpha) + (1 - \alpha) U(x; \alpha, 1)$ (see Titterington *et al.* (1985), p. 36) which is unidentifiable and can be represented by one component model, then the assumption of Theorem 3.1 is not satisfied. But if the true model is represented by $\frac{1}{3}U(x; -1, 1) + \frac{2}{3}U(x; -2, 2)$ (see Everitt and Hand (1981), p. 5), which is unidentifiable because $\frac{1}{2}U(x; -2, 1) + \frac{1}{2}U(x; -1, 2)$ represents the same distribution, then the assumption of Theorem 3.1 is satisfied, because it can not be represented by one component model.

Next proposition states that the rate of $c_n = \exp(-n^d)$, $d < 1$, obtained in Theorem 3.1 is almost the lower bound of the order of c_n which maintains the consistency.

PROPOSITION 3.1. *If c_n decreases faster than $\exp(-n)$, i.e., $e^n c_n \to 0$, then the consistency of maximum likelihood estimator restricted to Θ_n fails.*

PROOF. By the strong law of large numbers, mean log likelihood of true model $\frac{1}{n} \log \sum_{i=1}^{n} f(x_i; \theta_0)$ converges to $E[\log f(x; \theta_0)] < \infty$ almost everywhere. Assume that c_n decrease faster than $\exp(-n)$. Take $a_1 = x_1$, $b_1 = c_n$. Fix $\alpha_1 > 0$ and fix other parameters $(\alpha_2, \eta_2, \ldots, \alpha_M, \eta_M)$ such that $\frac{1}{n}\sum_{i=2}^{n} \log\{\sum_{m=2}^{M} \alpha_m f_m(x_i; \eta_m)\}$ converges to a finite limit almost everywhere. Then

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \log \sum_{m=1}^{M} f_m(x_i; \eta_m) \\
&\qquad \geq \frac{1}{n} \log\{\alpha_1 f_1(x_1; a_1 = x_1, b_1 = c_n)\} + \frac{1}{n}\sum_{i=2}^{n} \log\left\{\sum_{m=1}^{M} \alpha_m f_m(x_i; \eta_m)\right\} \\
&\qquad \geq \frac{1}{n} \log\left\{\frac{\alpha_1}{2c_n}\right\} + \frac{1}{n}\sum_{i=2}^{n} \log\left\{\sum_{m=2}^{M} \alpha_m f_m(x_i; \eta_m)\right\} \to \infty.
\end{aligned}
$$

Therefore mean log likelihood of the true model is dominated by that of other models and consistency of maximum likelihood estimator fails. □

4. Some discussions

As stated above in Section 1, the failure of consistency of maximum likelihood estimator is caused by the divergence of the likelihood of the model, where some scale parameters go to zero. Therefore in our setting it is of interest to investigate the behavior of the likelihood of the models on the boundary ($b_m = c_n$) of the restricted parameter space Θ_n. We report a simulation result for the case that the true model is $0.6 \cdot f(x; 0.5, 0.5) + 0.4 \cdot f(x; 0.6, 0.2)$ and a competing model is $0.6 \cdot f(x; 0.5, 0.5) + 0.4 \cdot f(x; a, b = c_n)$ which is on the boundary ($b = c_n$) of the restricted parameter space, where $c_n = \exp(n^{-0.93})$. The second column of Table 1 shows the log likelihood at $\hat{\theta}_n = \theta_0$. The third column shows the log likelihood maximized with respect to $a \in [0, 1]$ (but b is taken to be c_n). In the competing model, with probability tending to 1, the length of the interval $2c_n$ is shorter than the minimum of the distance between realized values. Therefore with probability tending to 1 the support of $f(x; a, b = c_n)$ does not contain two or more realized values for all $a \in [0, 1]$. Therefore the maximum of the likelihood is usually achieved when the support of $f(x; a, b = c_n)$ contains just one realized value. Then $f(x; a, b = c_n) = 0.6 + 0.4/(2c_n)$ on one particular realization and $f(x; a, b = c_n) = 0.6$ on the other $n - 1$ realized values. In this case the maximum of the log likelihood in competing model is given by $\log\{0.6 + 0.4/(2c_n)\} + (n - 1)\log\{0.6\}$. The result in Table 1 is based on one replication for each sample size. If we repeat the simulations, the results are similar. Therefore the result in Table 1 indicates that the log likelihood of the true model gets larger than that of the competing models with $b = c_n$ as the sample size n increases. This simulation result is consistent with Theorem 3.1.

Table 1. Log likelihood of the true model and that of a competing model.

sample size n	log likelihood (true)	log likelihood ($b = c_n$)
10	0.7767	2.305
50	9.769	11.38
100	15.61	20.26
500	56.49	67.11
1000	117.9	104.7
5000	582.6	199.3

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. On the other hand, in Theorem 3.1, it might be difficult to weaken the assumption that there is no representation of the true model with less than M components. 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.

Appendix: Proof of the strong consistency

Here we present a proof of Theorem 3.1. Note that it is sufficient to prove Theorem 3.1 for d arbitrarily close to 1. Therefore we assume $d > 1/4$ hereafter.

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

Define

$$\Theta'_n \equiv \{\theta \in \Theta_n \mid L_{\min} \leq \forall a_m \leq L_{\max}, c_n \leq \forall b_m \leq L, c_n \leq \exists b_m \leq c_0\},$$
$$\Gamma_0 \equiv \{\theta \in \Theta \mid L_{\min} \leq a_m \leq L_{\max}, c_0 \leq b_m \leq L, m = 1, \ldots, M\}.$$

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 $T_0 \subset \Gamma_0$.

In view of Theorems 2.1, 2.2, for the strong consistency of MLE on Θ_n, by Lemma 3.1, *it suffices to prove that*

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

for all closed $S' \subset \Gamma_0$ not intersecting T_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 equation (2.3) with S replaced by S' holds by Theorem 2.1. This implies that it suffices to prove equation (2.3) with S replaced by Θ'_n.

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 equation (2.3) with S replaced by $\Theta'_{n,K}$. We fix K from now on. Define $\bar{\Theta}_K$ by

$$\bar{\Theta}_K \equiv \Bigg\{ (\alpha_{K+1}, a_{K+1}, b_{K+1}, \ldots, \alpha_M, a_M, b_M) \in \mathbb{R}^{3(M-K)} \mid \sum_{m=K+1}^{M} \alpha_m \leq 1, \alpha_m \geq 0,$$
$$L_{\min} \leq a_m \leq L_{\max}, c_0 \leq b_m \leq L, m = K+1, \ldots, M \Bigg\}$$

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

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

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

LEMMA A.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*

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

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, $s = 1, \ldots, S$,

$$\lim_{n\to\infty} \frac{\sup_{\theta\in\Theta'_{n,K,s}} \prod_{i=1}^{n} f(x_i;\theta)}{\prod_{i=1}^{n} f(x_i;\theta_0)} = 0, \qquad \text{a.e.} \tag{A.2}$$

We fix s in addition to K from now on.

Because

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

(A.2) is implied by

$$\text{(A.3)} \qquad \limsup_{n\to\infty} \frac{1}{n} \sup_{\theta\in\Theta'_{n,K,s}} \sum_{i=1}^{n} \log f(x_i;\theta) < E_0[\log f(x;\theta_0)], \qquad \text{a.e.}$$

Therefore it suffices to prove (A.3), which is a new intermediate goal of our proof hereafter.

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

$$\text{(A.4)} \qquad \lambda \equiv E_0[\log f(x;\theta_0)] - E_0[\log\{\bar{f}(x;\bar{\theta}^{(s)},\rho(\bar{\theta}^{(s)})) + G\}] > 0.$$

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

$$2c_0 < e^{-1},$$

$$\text{(A.5)} \qquad 3M \cdot u \cdot 2c_0 \cdot (-\log G) < \frac{\lambda}{4},$$

$$\text{(A.6)} \qquad 2M \cdot u \cdot 2c_0 \cdot \log\frac{1}{2c_0} < \frac{\lambda}{12}.$$

Although G depends on c_0, it can be shown that G and c_0 can be chosen small enough to satisfy these inequalities. We now prove the following lemma.

LEMMA A.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 $\theta \in \Theta'_{n,K,s}$*

$$\text{(A.7)} \qquad \frac{1}{n}\sum_{i=1}^{n} \log f(x_i;\theta) \le \frac{1}{n}\sum_{i=1}^{n} \log\{\bar{f}(x_i;\bar{\theta}^{(s)},\rho(\bar{\theta}^{(s)})) + G\} + \frac{1}{n}\sum_{x_i\in J(\theta)} \log f(x_i;\theta) + \frac{1}{n} R_n(J(\theta)) \cdot (-\log G).$$

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

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

$$\leq \frac{1}{n}\sum_{i=1}^{n}\log\{\bar f(x_i;\bar\theta^{(s)},\rho(\bar\theta^{(s)}))+G\}$$
$$+\frac{1}{n}\sum_{x_i\in J(\theta)}\log f(x_i;\theta)-\frac{1}{n}R_n(J(\theta))\log G. \qquad \square$$

We want to bound the terms on the right hand side of (A.7) from above. The first term is easy. In fact by (A.4) and the strong law of large numbers we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log\{\bar f(x_i;\bar\theta^{(s)},\rho(\bar\theta^{(s)}))+G\}=E_0[\log f(x;\theta_0)]-\lambda, \quad \text{a.e.} \tag{A.8}$$

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

LEMMA A.3.

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

PROOF. Let $\epsilon>0$ be arbitrarily fixed and let J_0 be the support of the true density. J_0 consists of at most M 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 Fig. 3. Let $k(c_0)$ be the number of short intervals and let $I_1(c_0),\dots,I_{k(c_0)}(c_0)$ 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 $\{I_1(c_0),\dots,I_{k(c_0)}(c_0)\}$. 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 the following relation holds.

$$\sup_{\theta\in\Theta'_{n,K,s}}\frac{1}{n}R_n(J(\theta))-3M\cdot u\cdot 2c_0>\epsilon \tag{A.9}$$
$$\Rightarrow 1\leq\exists k\leq k(c_0),\ \frac{1}{n}R_n(I_k(c_0))-u\cdot 2c_0>\frac{\epsilon}{3M}.$$

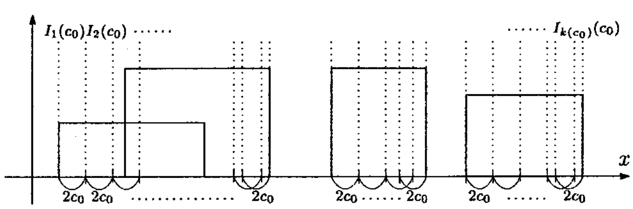

Fig. 3. Division of J_0 by short intervals of length $2c_0$.

From (A.9), we have

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

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

$$P_0(I_k(c_0)) \le u\cdot 2c_0, \qquad k=1,\dots,k(\theta). \tag{A.10}$$

Since $R_n(V) \sim \mathrm{Bin}(n, P_0(V))$ and from (2.4), we obtain

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

Therefore

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

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

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

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

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

By this lemma and (A.5) we have

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta'_{n,K,s}} \frac{1}{n} R_n(J(\theta))\cdot(-\log G) \le 3M\cdot u\cdot 2c_0\cdot(-\log G) < \frac{\lambda}{4}. \tag{A.11}$$

This bounds the third term on the right hand side of (A.7) from above.

Finally we bound the second term on the right hand side of (A.7) 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

$$f(x;\theta) = \frac{1}{n}\sum_{t=1}^{T(\theta)} H(J_t(\theta)) 1_{J_t(\theta)}(x), \tag{A.12}$$

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

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

is the height of $f(x;\theta)$ on $J_t(\theta)$ and $T \equiv T(\theta)$ 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 $a_m + b_m$, $m = 1, \dots, M$. 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 $J_t(\theta)$, $t = 1, \dots, T(\theta)$, by the height $H(J_t(\theta))$. Define c_n' by

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

and define $\tau_n(\theta)$

$$\tau_n(\theta) \equiv \max\left\{ t \in \{1, \dots, T\} \mid H(J_t(\theta)) \le \frac{M}{2c_n'} \right\}. \tag{A.13}$$

Then the second term on the right hand side of (A.7) is written as

$$\begin{aligned}
\frac{1}{n}\sum_{x_i \in J(\theta)} \log f(x_i;\theta) &= \sum_{t=1}^{T(\theta)} \frac{1}{n} \sum_{x_i \in J_t(\theta)} \log H(J_t(\theta)) \\
&= \frac{1}{n}\sum_{t=1}^{T(\theta)} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) \\
&= \frac{1}{n}\sum_{t=1}^{\tau_n(\theta)} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)) \\
&\quad + \frac{1}{n}\sum_{t=\tau_n(\theta)+1}^{T(\theta)} R_n(J_t(\theta)) \cdot \log H(J_t(\theta)).
\end{aligned} \tag{A.14}$$

From (A.5), (A.6), and noting that $\log x/x$ is decreasing in $x \ge e$, we have

$$\begin{aligned}
&3\sum_{t=1}^{\tau_n(\theta)} \frac{u}{H(J_t(\theta))} \log H(J_t(\theta)) \le 3 \cdot 2M \cdot u \cdot 2c_0 \cdot \log \frac{1}{2c_0} < \frac{\lambda}{4}, \\
&\sum_{t=\tau_n(\theta)+1}^{T(\theta)} 3 \cdot \frac{2}{n} \log H(J_t(\theta)) \le 3 \cdot 2M \cdot \frac{2}{n} \cdot \left(n^d - \log \frac{M}{2c_0}\right) \to 0.
\end{aligned} \tag{A.15}$$

Suppose that the following inequality holds.

$$
\begin{aligned}
\text{(A.16)} \qquad \limsup_{n\to\infty} \sup_{\theta\in\Theta'_{n,K,s}} \Bigg[\sum_{t=1}^{T(\theta)} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) \\
- 3 \left\{ \sum_{t=1}^{\tau_n(\theta)} \frac{u}{H(J_t(\theta))} \log H(J_t(\theta)) \right. \\
\left. + \sum_{t=\tau_n(\theta)+1}^{T(\theta)} \frac{2}{n} \log H(J_t(\theta)) \right\} \Bigg] \le 0, \qquad \text{a.e.}
\end{aligned}
$$

Then from (A.14) and (A.15), the second term on the right hand side of (A.7) is bounded from above as

$$
\text{(A.17)} \qquad \limsup_{n\to\infty} \frac{1}{n} \sup_{\theta\in\Theta'_{n,K,s}} \sum_{x_i\in J(\theta)} \log f(x_i;\theta) \le \frac{4}{\lambda}.
$$

Combining (A.8), (A.11) and (A.17) we obtain

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

and (A.3) is satisfied. Therefore it suffices to prove (A.16), 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

$$
\begin{aligned}
\text{(A.18)} \qquad \sup_{\theta\in\Theta'_{n,K,s}} \Bigg[\sum_{t=1}^{T(\theta)} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) \\
- 3 \left\{ \sum_{t=1}^{\tau_n(\theta)} \frac{u}{H(J_t(\theta))} \log H(J_t(\theta)) + \sum_{t=\tau_n(\theta)+1}^{T(\theta)} \frac{2}{n} \log H(J_t(\theta)) \right\} \Bigg] \\
\le \max_{T=1,\dots,2M} \max_{\tau=1,\dots,T} \Bigg[\sup_{\theta\in\Theta'_{n,K,s,T,\tau}} \left\{ \sum_{t=1}^{\tau} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) \right. \\
\left. - 3 \sum_{t=1}^{\tau} \frac{u}{H(J_t(\theta))} \log H(J_t(\theta)) \right\} \\
+ \sup_{\theta\in\Theta'_{n,K,s,T,\tau}} \left\{ \sum_{t=\tau+1}^{T} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) \right. \\
\left. - 3 \sum_{t=\tau+1}^{T} \frac{2}{n} \log H(J_t(\theta)) \right\} \Bigg].
\end{aligned}
$$

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

$$\text{(A.19)} \qquad \limsup_{n\to\infty} \sup_{\theta\in\Theta'_{n,K,s,T,\tau}} \left[\sum_{t=1}^{\tau} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) - 3\sum_{t=1}^{\tau} \frac{u}{H(J_t(\theta))} \log H(J_t(\theta))\right] \le 0, \qquad \text{a.e.}$$

$$\text{(A.20)} \qquad \limsup_{n\to\infty} \sup_{\theta\in\Theta'_{n,K,s,T,\tau}} \left[\sum_{t=\tau+1}^{T} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) - 3\sum_{t=\tau+1}^{T} \frac{2}{n} \log H(J_t(\theta))\right] \le 0, \qquad \text{a.e.}$$

Then (A.16) is derived from (A.18), (A.19), (A.20). Therefore it suffices to prove (A.19) and (A.20), which are the final goals of our proof. We state (A.19) and (A.20) as two lemmas and give their proofs.

LEMMA A.4.

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

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

$$\text{(A.21)} \qquad \sup_{\theta\in\Theta'_{n,K,s,T,\tau}} \left[\sum_{t=\tau+1}^{T} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) - 3\sum_{t=\tau+1}^{T} \frac{2}{n} \log H(J_t(\theta))\right] > 2M\delta.$$

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.

The event (A.21) occurs.

$$\Rightarrow \sup_{\theta\in\Theta'_{n,K,s,T,\tau}} \left[\sum_{t=\tau+1}^{T} \max\left\{0, \left(\frac{1}{n} R_n([a'_t(\theta) - c'_n, a'_t(\theta) + c'_n]) - 3\cdot\frac{2}{n}\right)\right\} \log\frac{M}{2c_n}\right] > 2M\delta$$

$$\Rightarrow \exists\theta \in \Theta'_{n,K,s,T,\tau}, \exists t > \tau$$

$$\max\left\{0, \left(\frac{1}{n} R_n([a'_t(\theta) - c'_n, a'_t(\theta) + c'_n]) - 3\cdot\frac{2}{n}\right)\right\} \log\frac{M}{2c_n} > \delta$$

$$\Rightarrow \exists\theta \in \Theta'_{n,K,s,T,\tau}, \exists t > \tau$$

$$R_n([a'_t(\theta) - c'_n, a'_t(\theta) + c'_n]) \ge 6$$

$$\text{(A.22)} \qquad \Rightarrow \sup_{L_{\min}\le a'\le L_{\max}} R_n([a' - c'_n, a' + c'_n]) \ge 6.$$

Below, we consider the probability of the event that (A.22) occurs. We divide J_0 from $L_{\min}$ to $L_{\max}$ by short intervals of length $2c'_n$ as in the proof of Lemma A.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. Because J_0 consists of at most M intervals, we have

$$k(c'_n) \le \frac{L}{2c'_n} + M. \tag{A.23}$$

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

$$\sup_{L_{\min} \le a' \le L_{\max}} R_n([a' - c'_n, a' + c'_n]) \ge 6 \Rightarrow 1 \le \exists k \le k(c'_n), \qquad R_n(I_k(c'_n)) \ge 2. \tag{A.24}$$

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 (A.22), (A.23) and (A.24) we have

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

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

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

□

Finally we prove (A.19).

LEMMA A.5.

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

PROOF. Let $\delta > 0$ be any fixed positive real constant and let h_n be

$$h_n \equiv \frac{\delta}{12}\left\{u \log\left(\frac{M}{c'_n}\right)\right\}^{-1}. \tag{A.25}$$

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

$$l_n \le \frac{c_0}{h_n} + 1. \tag{A.26}$$

Next, we consider the probability of the event that

$$\sup_{\theta \in \Theta'_{n,K,s,T,\tau}} \left[\sum_{t=1}^{\tau} \frac{1}{n} R_n(J_t(\theta)) \log H(J_t(\theta)) - 3\sum_{t=1}^{\tau} \frac{u}{H(J_t(\theta))} \log H(J_t(\theta))\right] > 2M\delta. \tag{A.27}$$

For this event the following relation holds.

The event (A.27) occurs.

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

$$- 3u \cdot 2b_{l+1}^{(n)}\Big)\Big\} \log \frac{1}{2b_{l+1}^{(n)}} > \delta$$

$$\text{(A.28)} \quad \Rightarrow 1 \leq \exists l \leq l_n \quad \text{s.t.}$$
$$\sup_{L_{\min} \leq a' \leq L_{\max}} \left\{ \left(\frac{1}{n} R_n([a' - b_l^{(n)}, a' + b_l^{(n)}]) - 3u \cdot 2b_l^{(n)} \right) \log \frac{1}{2b_{l+1}^{(n)}} \right.$$
$$\left. + 3u(2b_l^{(n)} - 2b_{l+1}^{(n)}) \log \frac{1}{2b_{l+1}^{(n)}} \right\} > \delta.$$

Then from (A.25) the following relation holds.

$$\text{The event (A.28) occurs.}$$
$$\text{(A.29)} \quad \Rightarrow 1 \leq \exists l \leq l_n,$$
$$\sup_{L_{\min} \leq a' \leq L_{\max}} \frac{1}{n}(R_n([a' - b_l^{(n)}, a' + b_l^{(n)}]) - 3u \cdot 2b_l^{(n)}) \log \frac{1}{2b_{l+1}^{(n)}} > \frac{\delta}{2}.$$

Below, we consider the probability of the event that (A.29) 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 A.3. Let $k(b_l^{(n)})$ be the number of short intervals and let $I_1(b_l^{(n)}), \dots, I_{k(b_l^{(n)})}(b_l^{(n)})$ be the divided short intervals. Then we have

$$\text{(A.30)} \qquad k(b_l^{(n)}) \leq \frac{L}{2b_l^{(n)}} + M.$$

Since any interval in J_0 of length $2b_l^{(n)}$ is covered by at most 3 small intervals from $\{I_1(b_l^{(n)}), \dots, I_{k(b_l^{(n)})}(b_l^{(n)})\}$, the following relation holds.

$$\text{(A.31)} \quad \sup_{L_{\min} \leq a' \leq L_{\max}} \left(\frac{1}{n} R_n([a' - b_l^{(n)}, a' + b_l^{(n)}]) - 3u \cdot 2b_l^{(n)} \right) > \frac{\delta}{2} \left(\log \frac{1}{2b_{l+1}^{(n)}} \right)^{-1}$$
$$\Rightarrow \max_{k=1,\dots,k(b_l^{(n)})} \left(\frac{1}{n} R_n(I_k(b_l^{(n)})) - u \cdot 2b_l^{(n)} \right) > \frac{1}{3} \cdot \frac{\delta}{2} \left(\log \frac{1}{2b_{l+1}^{(n)}} \right)^{-1}.$$

Note that $R_n(I_k(b_l^{(n)})) \sim \mathrm{Bin}(n, P_0(I_k(b_l^{(n)})))$ and $P_0(I_k(b_l^{(n)})) \leq u \cdot 2b_l^{(n)}$. Therefore from (2.4) and (A.30) we have

$$\text{(A.32)} \quad \mathrm{Prob} \left(\max_{k=1,\dots,k(b_l^{(n)})} \frac{1}{n} \left(R_n(I_k(b_l^{(n)})) - u \cdot 2b_l^{(n)} \right) > \frac{1}{3} \cdot \frac{\delta}{2} \left(\log \frac{1}{2b_{l+1}^{(n)}} \right)^{-1} \right)$$
$$\leq \left(\frac{L}{2b_l^{(n)}} + M \right) \exp \left\{ -2n \cdot \frac{\delta^2}{36} \left(\log \frac{1}{2b_{l+1}^{(n)}} \right)^{-2} \right\}$$
$$\leq \left(\frac{L}{2c_n'} + M \right) \exp \left\{ -2n \cdot \frac{\delta^2}{36} \left(\log \frac{1}{2c_n'} \right)^{-2} \right\}.$$

From (A.26), (A.28), (A.29), (A.31), (A.32), we obtain

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

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

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

This completes the proof of Theorem 3.1.

References

Everitt, B. S. and Hand, D. J. (1981). *Finite Mixture Distributions*, Chapman and Hall, London.

Hathaway, R. J. (1985). A constrained formulation of maximum likelihood estimation for normal mixture distributions, *The Annals of Statistics*, **13**, 795–800.

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

Okamoto, M. (1958). Some inequality relating to the partial sum of binomial probabilities, *Annals of the Institute of Statistical Mathematics*, **10**, 29–35.

Redner, A. R. (1981). Note on the consistency of the maximum likelihood estimation for nonidentifiable distribution, *The Annals of Statistics*, **9**, 225–228.

Teicher, H. (1960). On the mixture of distributions, *Annals of Mathematical Statistics*, **31**, 55–73.

Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate, *Annals of Mathematical Statistics*, **29**, 595–601.

Yakowitz, S. J. and Spragins, J. D. (1968). On the identifiability of finite mixtures, *Annals of Mathematical Statistics*, **39**, 209–214.