应用概率统计　第六卷
第四期　1990年11月

Chinese Journal of Applied Probability
and Statistics Vol. 6. No. 4. Nov. 1990

TWO-PARAMETER EXPONENTIAL FINITE MIXTURE

W. LIANG
(*Institute of Statistical Science, Teibei*)

Y. Q. YIN
(*University of Lowell*)

Abstract

The strong consistency of the constrained maximum likelihood estimator of the two-parameter exponential finite mixture is established.

§ 1. Introduction

The investigation of the maximum likelihood estimator of the exponential finite mixture has been pursued by many authors, among whom we mention [1]—[5] For a comprehensive bibliography we refer to the book [6].

Most existing works on exponential finite mixture deal with the one-parameter case. This paper deals with two-parameter case which is much more complicated. To see the difference between the one-parameter exponential finite mixture and two-parameter exponential finite mixture, we notice the following fact. For the one-parameter exponential density, i.e.

$$\psi(x, \lambda) = \lambda e^{-\lambda x} I_{[0,\infty)}(x), \quad \lambda > 0,$$

where $I_{[0,\infty)}(x)$ is the indicator of the set $[0, \infty)$, the density is bounded on the natural parameter space $(0, \infty)$ for any fixed $x > 0$. On the contrary, two-parameter exponential density, i. e.

$$\phi(x, \lambda, a) = \lambda e^{-\lambda(x-a)} I_{[a,\infty)}(x), \quad \lambda > 0,\ a \in (-\infty, \infty) = R$$

is unbounded on the natural parameter space $\{(\lambda, a): \ \lambda > 0, \ a \in R\}$ for any fixed x. Due to this fact, it is easy to see that the maximum likelihood estimator exists in the case of one-parameter exponential finite mixture and does not exist in the case of two-parameter exponential finite mixture. For example, consider a two-parameter exponential mixture of two components. Without loss of generality, we may suppose $x_1 \leqslant \cdots \leqslant x_n$. Let $w_1 = w_2 = 1/2$, $a_1 = a_2 = x_1$, $\lambda_1 \to \infty$, $\lambda_2 = 1$, then we see that

$$\sup \prod_{i=1}^{n} (w_1\phi(x_i, \lambda_1, a_1) + w_2\phi(x_i, \lambda_2, a_2)) \geqslant \left(\frac{1}{2}\lambda_1\right)\left(\frac{1}{2} e^{-(x_2-x_1)}\right)\cdots\left(\frac{1}{2} e^{-(x_n-x_1)}\right) \to \infty.$$

This shows that the maximum likelihood estimator does not exist.

The density of the two-parameter exponential finite mixture is

Received October 1989.

$$f(x, \theta)=\sum_{k=1}^{K} w_k \phi(x, \lambda_k, a_k),$$

where K, the number of the components, is known, the mixing proportions $w_k \geqslant 0$, $w_1+\cdots+w_K=1$, the parameter $\theta=(w, \lambda, a)$, where $w=(w_1, \cdots, w_K)$, $\lambda=(\lambda_1, \cdots, \lambda_K)$, $a=(a_1, \cdots, a_K)$. The natural parameter space is

$$\Xi=W \times(0, \infty)^K \times R^K$$

where $W=\{(w_1, \cdots, w_K): w_1+\cdots+w_K=1 \text{ and } w_k \geqslant 0, \forall k=1, \cdots, K\}$. The likelihood function is

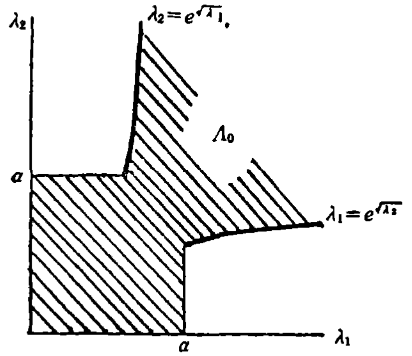
$$L_n(x_1, \cdots, x_n; \theta)=\prod_{i=1}^{n} f(x_i, \theta).$$

Since we cannot consider the maximum likelihood estimator over the natural parameter space Ξ, we shall use, however, the constrained maximum likelihood estimator on some constrained region Ξ_0 which ensures the existence of the maximizer of L_n on it.

In this paper, we find a kind of constrained region Ξ_0, on which the constrained maximum likelihood estimator exists and tends to the true parameter θ_0 as $n \rightarrow \infty$ almost surely, provided the true parameter $\theta_0 \in \Xi_0$. Because we can make the constrained region Ξ_0 arbitrarily approximate the natural space Ξ, the condition $\theta_0 \in \Xi_0$ will not be an impassable chasm in practice.

§ 2. Main result

In the light of the example mentioned above, we know that in order to guarantee the existence of the maximum likelihood estimator, it is necessary to add some constraints on λ space so as to avoid the occurrence of the situation where some $\lambda_k \rightarrow \infty$ and the others keep constant.



Let $g(t)$ be an increasing function defined on $(0, \infty)$, and $g(t)=0(t)$ as $t \rightarrow \infty$. Let α be a positive number. We defined a subset of the λ space $(0, \infty)^K$ as follows:

$$\Lambda_0=\{(\lambda_1, \cdots, \lambda_K): \forall j \neq k, \text{ if } \lambda_j \geqslant \alpha \text{ then. } \lambda_j \leqslant \exp(g(\lambda_k))\}.$$

For example, we take $g(t)=\sqrt{t}$ and $K=2$, then the shape of Λ_0 is drawn as the following figure.

Now, we take the constrained region of the parameter θ to be

$$\Xi_0=W \times \Lambda_0 \times R^K.$$

Theorem. Let the sample space be Ω. For almost all $\omega \in \Omega$, there exists a positive number N_ω such that for any $n>N_\omega$ the likelihood function L_n has a maximizer $\hat{\theta}_n$ on Ξ_0 and $\hat{\theta}_n \rightarrow \theta_0$, provided $\theta_0 \in \Xi_0$.

Proof. For an arbitrary ϵ-neighborhood $U_\epsilon(\theta_0)$ of θ_0, let

$$\Xi'_0=\Xi_0-U_\epsilon(\theta_0).$$

If Ξ'_0 can be expressed as a finite union of subsets, $\Xi'_0=\bigcup_i \Xi_i$, and for almost all $\omega\in\Omega$, there is a positive number N_ω such that for each Ξ_i we have

$$\sup_{\Xi_i} L_n(\theta)<L_n(\theta_0),\quad \forall n>N_\omega,$$

then the theorem is proved. By the strong law of large numbers,

$$L_n^{1/n}(\theta_0)=\exp\Big(\frac{1}{n}\log L_n(\theta_0)\Big)\xrightarrow{\text{a.s.}}\exp(E\log f(x,\theta_0))\overset{\text{say}}{=\!=}\xi>0.$$

So, we need only prove that

$$(*)\qquad \sup_{\Xi_i} L_n^{1/n}(\theta)<\xi,\quad \forall n>N_\omega.$$

On the other hand,

$$\log\sup_{\Xi_i} L_n^{1/n}(\theta)\leqslant\frac{1}{n}\sum_{j=1}^{n}\sup_{\Xi_i}\log f(x_j,\theta)\xrightarrow{\text{a. s.}}E\sup_{\Xi_i}\log f(x,\theta),$$

$$\log L_n^{1/n}(\theta_0)\xrightarrow{\text{a. s.}}E\log f(x,\theta_0).$$

Therefore, it is also enough to prove that

$$(**)\qquad E\sup_{\Xi_i}\log f(x,\theta)<E\log f(x,\theta_0).$$

For $B>0$ and $C>0$, put

$$\Lambda_1=\Lambda_0\cap\{(\lambda_1,\cdots,\lambda_K):\text{ all }\lambda_k\geqslant B\},$$
$$\Lambda_2=\Lambda_0\cap\{(\lambda_1,\cdots,\lambda_K):\text{ all }\lambda_k\leqslant C\}.$$

For every $B>0$, when C is sufficiently large, we have

$$\Xi'_0=\Xi_1\bigcup\Xi_2,$$

where $\Xi_i=\Xi'_0\cap(W\times\Lambda_i\times R^K)$, $i=1, 2$. Hence, in what follows we need only prove that Ξ_1 possesses the property $(*)$ and that Ξ_2 can be expressed as a union of finitely many subsets such that each subset possesses the property $(**)$.

1°. Consider Ξ_1.

Suppose $\theta\in\Xi_1$. Take $K+1$ arbitrary disjoint intervals $I_k=[\beta_k-2,\beta_k+2]$, $k=1,\cdots,K+1$. Let $I'_k=[\beta_k-1,\beta_k+1]$ and $p_k=P\{x\in I'_k\}$, $p=\min\{p_1,p_2,\cdots,p_{K+1}\}$. By the strong law of large numbers, for almost all $\omega\in\Omega$ there is a N_ω such that for $n>N_\omega$ we have

$$\min_{1\leqslant k\leqslant K+1}\#\{x_l:\ x_l\in I'_k\}/n>\frac{1}{2}p.$$

For any $(a_1,\cdots,a_K)$ there must be some I_{k_0} which contains none of $a_1,\cdots,a_K$. For sufficiently large B, we have

$$L_n^{1/n}=[\prod_{x_l\notin I'_{k_0}}f(x_l,\theta)\prod_{x_l\in I'_{k_0}}f(x_l,\theta)]^{1/n}$$

$$\leqslant\Big(\sum_{k=1}^{K}\lambda_k\Big)\Big[\prod_{x_l\in I'_{k_0}}\Big(\sum_{k=1}^{K}w_k\lambda_k e^{-\lambda_k(x_l-a_k)}I_{[a_k,\infty)}(x_l)\Big)^{\frac{1}{n}}\Big]$$

$$\leqslant\Big(\sum_{k=1}^{K}\lambda_k\Big)\Big(\prod_{x_l\in I'_{k_0}}\sum_{k=1}^{K}\lambda_k e^{-\lambda_k}\Big)^{\frac{1}{n}}\leqslant\Big(\sum_{k=1}^{K}\lambda_k\Big)\Big(\sum_{k=1}^{K}\lambda_k e^{-\lambda_k}\Big)^{\frac{p}{2}}\leqslant\Big(\sum_{k=1}^{K}\lambda_k\Big)\Big(\sum_{k=1}^{K}\lambda_k^{p/2}e^{-\lambda_k p/2}\Big)$$

$$=\text{sum}_1+\text{sum}_2$$

where the entries in sum_1 and sum_2 are, respectively, of the form

$$\lambda_k^{1+(p/2)}e^{-\lambda_k p/2} \quad \text{and} \quad \lambda_j(\lambda_k^{p/2}e^{-\lambda_k p/2}), \quad j\neq k.$$

Because $\theta\in E_1$, all $\lambda_k\geqslant B$. Hence the entries of the first form are infinitesimals as $B\to\infty$. As to the entries of the second form, when $B\to\infty$ we have $\lambda_j\geqslant B>\alpha$, hence by definition of Λ_0 we have

$$\lambda_j(\lambda_k^{p/2}e^{-\lambda_k p/2})\leqslant\lambda_k^{p/2}e^{-\lambda_k p/2+g(\lambda_k)}=\lambda_k^{p/2}e^{-\lambda_k p/2+o(\lambda_k)}\to 0.$$

Consequently, $\text{sum}_1+\text{sum}_2<\xi/2$. It implies that

$$\sup_{E_1} L_n^{1/n}<\xi \quad \forall n>N_\omega.$$

This is just what we wanted.

2°. Before considering E_2, we claim that corresponding to each $w^*\in W$ and each $\lambda^*\in\Lambda_2$, the set

$$E^*=E_0'\cap\{\theta:\ w=w^*,\ \lambda=\lambda^*\}$$

can be expressed as a finite union

$$E^*=\bigcup_j E_j^*,$$

where $E_j^*=E_0'\cap\{\theta:\ w=w^*,\ \lambda=\lambda^*,\ a\in A_j^*\}$, $A_j^*=A_{1j}^*\times\cdots\times A_{Kj}^*$, and for each E_j^* we have

$$E\sup_{E_j^*}\log f(x,\ \theta)<E\log f(x,\ \theta_0).$$

We prove the claim as follows. Let $0<m_1<\cdots<m_K$, where m_1 is so large that for every θ with some $|a_k|\geqslant m_1$ we have $\theta\notin U_\epsilon(\theta_0)$. The space R^K of $(a_1,\ \cdots,\ a_K)$ can be expressed as a union of finite numbers of subsets such that each subset is of one of the following two forms

$M_1=\{(a_1,\ \cdots,\ a_K):\ \text{all}\ |a_k|\leqslant m_l\}$

$M_2=\{(a_1,\ \cdots,\ a_K):\ \text{there are some}\ |a_k|\geqslant m_{l+1}\ \text{and all other}\ |a_k|\leqslant m_l\}.$

Hence instead of E^*, we need only consider the sets

$$E_l^*=E_0'\cap\{\theta:\ w=w^*,\ \lambda=\lambda^*,\ a\in M_l\},\ l=$$

Because E_1^* is compact, it can be expressed as a finite union of subsets such that each subset is of the following form

$$E_{1j}^*=E_0'\cap\{\theta:\ w=w^*,\ \lambda=\lambda^*,\ a\in A_j^*\},$$

and

$$E\sup_{E_{1j}^*}\log f(x,\ \theta)<E\log f(x,\ \theta_0).$$

Let the component a of θ_0 be

$$a_0=(a_{01},\ \cdots,\ a_{0K}).$$

We can assume that the A_j^*'s make a lattice covering of M_1, namely, each A_j^* is a rectangle $I_{i1}\times\cdots\times I_{i_K}$, and for each $k\leqslant K$, there is a I_{i_k} which does not intersect the interval $(a_{0k}-\epsilon,\ a_{0k}+\epsilon)$. This will be useful later.

Regarding E_2^*, we consider an arbitrary point $\theta_2^*\in E_2^*$, say $\theta_2^*=(w^*,\ \lambda^*,\ a_1^*,\ \cdots,\ a_K^*)$. Without loss of generality, we can assume that

$$|a_1^*|,\ \cdots,\ |a_k^*|\geqslant m_{l+1} \quad \text{and} \quad |a_{k+1}^*|,\ \cdots,\ \ |a_K^*|\leqslant m_l, \quad k\geqslant 1.$$

Let $\tilde{\theta}_2^*$ be a point with $|a_1|=\cdots=|a_k|=m_1$ and all other components coinciding with those of θ_2^*. Then, $\tilde{\theta}_2^* \in \Xi_1^*$. Hence there is some Ξ_{1j}^* such that $\tilde{\theta}_2^* \in \Xi_{ij}^*$. This implies that

$$\theta_2^* \in \Xi_0^1 \cap \{\theta: w=w^*, \lambda=\lambda^*, |a_1|, \cdots, |a_k| \geqslant m_{i+1}, a_{k+1} \in I_{i_{k+1}}, \cdots, a_K \in I_{i_K}\} \xlongequal{\text{say}} \Xi_{2j}^*$$

Therefore, Ξ_2^* can be expressed as a finite union of subsets with the similar form of Ξ_{2j}^*. By the monotone convergence theorem, we have

$$\lim_{m_{i+1}\to\infty} E \sup_{\Xi_{2j}^*} \log f(x, \theta) = E \log \lim_{m_{i+1}\to\infty} \sup_{\Xi_{2j}^*} f(x, \theta)$$

$$= E \log \lim_{m_{i+1}\to\infty} \sup_{\Xi_{2j}^*} \sum_{l=k+1}^{K} W_l \phi(x, \lambda_l, a_l).$$

Take $I_{i_1}, \cdots, I_{i_k}$ such that I_{i_1} does not intersect $(a_{01}-\epsilon, a_{01}+\epsilon)$. Note that corresponding to each $\theta \in \Xi_{2j}^*$ there exists

$$\theta' \in \Xi_{1j}^* = \Xi_0^1 \cap \{\theta: w=w^*, \lambda=\lambda^*, a \in I_{i_1} \times \cdots \times I_{i_k}\}$$

such that θ' and θ have the same components $a_{k+1}, \cdots, a_K$. Hence,

$$\lim_{m_{i+1}\to\infty} \sup_{\Xi_{2j}^*} \log f(x, \theta) \leqslant E \sup_{\Xi_{1j}^*} \log f(x, \theta) < E \log f(x, \theta_0).$$

This completes the proof of our claim.

3°. Consider $\Xi_2 = \Xi_0' \cap (W \times \Lambda_2 \times R^K)$.

Suppose that $c > r_1 > \cdots > r_K > 0$ and for any θ with some $\lambda_k \leqslant r_1$ we have $\theta \notin U_\epsilon(\theta_0)$. Ξ_2 can be expressed as a finite union of subsets such that each subset is of one of the following two forms:

$$\Xi_{21} = \Xi_0' \cap \{\theta: \text{all } \lambda_k \in [r_i, c]\},$$

$$\Xi_{22} = \Xi_0' \cap \{\theta: \text{there are some } \lambda_k \in [r_i, c] \text{ and all other } \lambda_k \leqslant r_{i+1}\}.$$

First, we consider Ξ_{21}.

From the result of 2°, we know that for $w^* \in W$ and $\lambda^* \in \Lambda_2$ there are neighborhoods $U(W^*)$ and $U(\lambda^*)$ such that

$$E \sup_{\Xi_j^*} \log f(x, \theta) < E \log f(x, \theta_0), \quad \forall j,$$

where

$$\Xi_j^* = \Xi_0' \cap \{\theta: W \in U(w^*), \lambda \in U(\lambda^*), a \in A_j^*\}.$$

Beacuse the projection of Ξ_{21} on the space of (w, λ) is $W \times [r_i, c]^K$ which is compact, we have a finite number of $U(w^*) \times U(\lambda^*)$, say $U(w^{*q}) \times U(\lambda^{*q})$, $q=1, \cdots, Q$, such that

$$\Xi_{21} = \bigcup_q \bigcup_j \Xi_{21qj}$$

where

$$\Xi_{21qj} = \Xi_0' \cap \{w \in U(w^{*q}), \lambda \in U(\lambda^{*q}), a \in A_j^{*q}\},$$

and

$$E \sup_{\Xi_{21qj}} \log f(x, \theta) < E \log f(x, \theta_0).$$

Secondly, we consider Ξ_{22}.

Suppose $\theta^* = (w^*, a^*, \lambda_1^*, \cdots, \lambda_k^*) \in \Xi_{22}$. Without loss of genreality, we assume $\lambda_1^*, \cdots, \lambda_k^* \in [r_i, c]$ and $\lambda_{k+1}^*, \cdots, \lambda_K^* \leqslant r_{i+1}$. Analogous to the reasoning in 2°, let

$$\theta^{**} = (w^*, a^*, \lambda_1^*, \cdots, \lambda_k^*, r_i, \cdots, r_i).$$

Then, $\theta^{**} \in \Xi_{21}$. Hence $\theta^{**} \in$ some Ξ_{21qj}. This implies that

$$\theta^* \in \Xi_0' \cap \{\theta: w^* \in U(w^{*q}),\ a^* \in A_j^{*q},\ \lambda_1^* \in \Lambda_1^{*q},\ \cdots,\ \lambda_k^* \in \Lambda_k^{*q}, \lambda_{k+1}^*,\ \cdots,\ \lambda_K^* \leqslant r_{i+1}\} \stackrel{\text{say}}{=\!=} \Xi_{qjk},$$

where

$$\Lambda_1^{*q} \times \cdots \times \Lambda_K^{*q} = U(\lambda^{*q}).$$

Thus, Ξ_{22} can be expressed as a finite union of subsets Ξ_{qjk} and we have

$$\begin{aligned}\lim_{r_{i+1}\to 0} E \sup_{\Xi_{qjk}} \log f(x, \theta) &= E \log \lim_{r_{i+1}\to 0} \sup_{\Xi_{qjk}} f(x, \theta) \\ &= E \log \lim_{r_{i+1}\to 0} \sup_{\Xi_{qjk}} \sum_{h=1}^{k} w_h \phi(x, \lambda_h, a_h) \\ &= E \log \sup_{\Xi_{i1qj}} f(x, \theta) < E \log f(x, \theta_0).\end{aligned}$$

This terminates the proof of the theorem.

References

[1] Weiner, S., *Samples from mixed-exponential populations*. Mimeo, ARINC Res. Corp., Wishington, D. C. 1962.

[2] Tallis, G. M., Light, R., The use of fractional moments for estimating the parameters of a mixed exponential distribution. *Technometrics*, **10** (1968), 161—175.

[3] Hasselblad, V., Estimation of finite mixtures of distributions from the exponential family. *J. Amer. Statist. Assoc.*, **64** (1969), 1459—1471.

[4] Suchindran, C. M., Lachenbruch, P. A., (1974) Estimates of parameters in a probability model for first livebirth interval. *J. Amer. Statist. Assoc.*, 69, 507—513.

[5] Wilson, D. L., Sargent, R. G., Some results of Monte Carlo experiments in estimating the parameters of the finite mixed exponential distribution. *Proc. Twelfth Ann. Symp. Interface* (Ed. J. F. Gentleman), pp. 461—465. Univ. Waterloo, Canada. 1979.

[6] Titterington, D. M., Smith, A. F. M., Makov. U. E., *Statistical analysis of finite mixture distributions*. John Wiley. 1985.

二参数指数型有限混合分布

梁文骐　　殷涌泉

(统计研究所，台北)　　(Lowell 大学)

本文给出二参数指数型有限混合分布的带约束的极大似然估计的强相合性。