BULLETIN OF THE
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
Volume 11, Number 4, December 1983

ON A CLASS OF CONSISTENT ESTIMATORS FOR FINITE MIXTURES OF DISTRIBUTIONS

BY

WEN-TAO HUANG (黃文濤) AND HSIAO-TIEN PAO (包曉天)

Abstract. A class of estimators based on kernel density estimate is proposed for estimating finite maxing probabilities. Strong consistency and asymptotic normality of the proposed estimators have been shown. Rate of convergence of the estimators is also given. Some Monte Carlo studies for small sample size and some comparisons with other known estimator are also given.

0. **Introduction.** The problem of estimating the mixing distribution of a finite mixture has gone over about eighty years. Since K. Pearson [7] started the study of estimating the mixing parameters and some unknown parameters of the mixture of two normal distributions in 1894, there has been a lot of literatures pursuing the same type of estimation problems. Most of the methods have been applied are, respectively, moment method, maximum likelihood method and Bayes approach. However, few desirable statistical properties have been shown (see [5]). Recently, Choi and Bulgren [3], Deely and Kruse [4], Yakowitz [14] and Blum and Sursala [1] respectively constructed different types of estimators for the finite mixing distributions with specified kernels. It has been shown that the estimators are all consistent. However, for the practical point of view, these estimators are not satisfactory for small sample size by the Monte Carlo study.

In this paper, a class of consistent estimators based on the kernel density estimate is proposed and some comparisons have been made with the result of [3] for the Monte Carlo studies. In these studies, it is found that the speed of convergence of the

Received by the editors June 9, 1982.
Key words: finite mixing probabilities estimation, kernel density estimate, strong consistency, asymptotic normality, rate of convergence, Monte Carlo studies.

proposed estimator is more rapid than that of the result of [3] for small sample size. It is to be emphasized that the Monte Carlo results show that the proposed estimator is practically applicable when the sample size is small even for 10 or 25.

1. **A class of discrete mixing estimators.** Let $\{f(x;\theta), \theta\in\Theta\}$ be a family of identifiable densities ([10], [11], [12], [13], [1]). Let $G(\theta)$ be a finite mixing distribution such that the mixture $f_G(x)$ is given by

$$(1.1)\qquad f_G(x)=\int_\Theta f(x;\theta)\,dG(\theta)=\sum_{i=1}^k g_i f(x;\theta_i)$$

where $f(x;\theta_i)$ is completely specified. Based on a sample $\underset{\sim}{X}=(X_1, X_2,\cdots, X_n)$ of size n from a population with mixture density $f_G(x)$, it is required to estimate the mixing distribution $\underset{\sim}{G}=(g_1, g_2, \cdots, g_k)$ which is a unknown vector parameter in $f_G(x)$. For notational convenience, we denote $\underset{\sim}{G}_0$ as the true mixing distribution associated with the sample $\underset{\sim}{X}$. We also denote $f_i(x)$ for $f(x;\theta_i)$ $(i=1, 2,\cdots, k)$ for simplicity.

Let $\hat{f}_n(x)$ be a consistent kernel estimate for the mixture $f_{G_0}(x)$ according to [6]. Let $w(x)$ denote a positive weight function such that $\int w(x)\,dx<\infty$. Define

$$(1.2)\quad S_n(G)=\int_{-\infty}^{\infty}(f_G(x)-\hat{f}_n(x))^2\,w(x)\,dx \qquad \text{for}\quad \underset{\sim}{G}\in\langle 0, 1\rangle^k$$

where

$$(1.3)\qquad \langle 0, 1\rangle^k=\left\{(g_1, g_2,\cdots, g_k):\ 0\le g_i\le 1,\ \sum_1^k g_i=1\right\}.$$

Since the system of linear simultaneous equation which are obtained by setting the respective partial derivatives of $S_n(G)$ with respect to g_i equal to zero has a unique solution, say G_n. And, the matrix whose element is the second partial derivative of $S_n(G)$ given by

$$\frac{\partial^2}{\partial g_i\,\partial g_j}S_n(G)=\int(f_i(x)-f_k(x))(f_j(x)-f_k(x))\,w(x)\,dx \qquad i, j=1, 2,\cdots, k$$

is positive definite by a generalization of the Cauchy inequality.

Accordingly, there exists a unique $\hat{G}_n \in \langle 0, 1\rangle^k$ which is a statistic depending on $\underset{\sim}{X}$ such that

$$S_n(\hat{G}_n) = \inf_{G \in \langle 0,1\rangle^k} S_n(G). \tag{1.4}$$

For given consistent estimate $\hat{f}_n(x)$ for $f_G(x)$, positive weight function $w(x)$ and a sample $\underset{\sim}{X}$, we propose $\hat{G}_n = G_n(\underset{\sim}{X})$ as an estimate for the mixing distribution G. Suppose

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{1.5}$$

is a consistent kernel density estimate for f_G given by [6]. For $k=2$, $K(x)$ is chosen to be the standard normal density and $w(x) = \exp(-x^2)$. Then, $\hat{G}_n = (\hat{g}_{1n}, \hat{g}_{2n})$ is given by

$$\hat{g}_{1n} = a_n/b_n \quad \text{and} \quad \hat{g}_{2n} = 1 - \hat{g}_{1n} \tag{1.6}$$

where

$$a_n = \left(\int f_2^2(x) + \int f_1(x)\hat{f}_n(x) - \int f_2(x)\hat{f}_n(x) - \int f_1(x) f_2(x)\right)^2 e^{-x^2}\, dx$$

$$b_n = \left(\int f_1^2(x) + \int f_2^2(x) - 2\int f_1(x) f_2(x)\right) e^{-x^2}\, dx$$

$$\hat{f}_n(x) = n^{-3/4} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp[-\sqrt{n}\,(x - X_i)^2/2]\,.$$

When $k=3$, and $K(x)$ and $w(x)$ remain unchanged, $\hat{G}_n$ is given by

$$\hat{g}_{1n} = \begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix} \Big/ \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \qquad \hat{g}_{2n} = \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} \Big/ \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \tag{1.7}$$

where

$$a_1 = \left(\int f_1^2(x) + \int f_3^2(x) - 2\int f_1(x) f_2(x)\right) e^{-x^2}\, dx$$

$$b_1 = \left(\int f_3^2(x) + \int f_1(x) f_2(x) - \int f_2(x) f_3(x) - \int f_1(x) f_3(x)\right) e^{-x^2}\, dx$$

$$c_1 = \left(\int f_1(x) \hat{f}_n(x) - \int f_3(x) \hat{f}_n(x) - \int f_1(x) f_3(x) + \int f_3^2(x)\right) e^{-x^2}\, dx$$

$$a_2 = b_1$$

$$b_2 = \left(\int f_2^2(x) - 2\int f_2(x) f_3(x) + \int f_3^2(x)\right) e^{-x^2}\, dx$$

$$c_2 = \left(\int f_2(x) \hat{f}_n(x) - \int f_3(x) \hat{f}_n(x) - \int f_2(x) f_3(x) + \int f_3^2(x)\right) e^{-x^2}\, dx.$$

2. **Some asymptotic properties of $\hat{G}_n$.** Let $\sigma_{ij}(n)$ denote the covariance between the ith and jth components of the column vector of $((\partial/\partial g_i) S_n(G))$. Define

$$J_{ij} = \int_{-\infty}^{\infty} f_i(x) f_j(x)\, w(x)\, dx \qquad i, j = 1, 2, \cdots, k \tag{2.1}$$

$$A = (J_{ij}) \tag{2.2}$$

$$\Sigma_n^* = (\sigma_{ij}(n)) \tag{2.3}$$

$$\Sigma_n = A^{-1} \Sigma_n^* A^{-1}. \tag{2.4}$$

We have the following asymptotic properties of $\hat{G}_n$.

MAIN THEOREM: (i) (*Consistency*) *If the kernel density* $f(x;\theta)$ *is continuous in* x *for each* θ, *and is continuous in* θ *for each* x, *then,* $\hat{G}_n$ *defined by* (1.4) *converges to the true* G_0 *with probability one.*

(ii) (*Asymptotic normality*) *If* A *defined by* (2.2) *is non-singular,* $\sqrt{n}\, \Sigma_n^{-1/2} (\hat{G}_n - G_0)$ *defined by* (2.4) *converges in distribution to the standard multivariate normal* $N(\underset{\sim}{0}, \underset{\sim}{1})$.

(iii) (*Rate of convergence of* $\hat{G}_n$) *If* A *is non-singular, then for the true mixing* G_0,

$$|\hat{G}_n - G_0| < O\left(\frac{\log n}{\sqrt{n}} h^{-1} V(\hat{f}_n) + ch^2\right) \quad \textit{with probability } 1,$$

where $V(\hat{f}_n)$ *denotes the variation of the kernel function* $K(x)$ *associated with* $f_n(x)$, h *is related to* n *given by* (1.5) *and* c *is a constant. If* $h = n^{-1/6}$, *the rate is almost* $O(n^{-1/3})$.

Proof of the Main Theorem. (i) Let $\|\cdot\|$ denote a sup norm. By definition of (1.2) and (1.4), for a true G_0, we have

$$S_n(\hat{G}_n) \leq S_n(G_0) \leq \|f_{G_0} - \hat{f}_n\| \int w(x)\, dx.$$

It follows from definition of $\hat{f}_n$ and [6] that $\|f_{G_0} - \hat{f}_n\| \to 0$ wpl. We have thus

$$S_n(\hat{G}_n) \longrightarrow 0 \quad \text{wpl} \tag{2.5}$$

and

$$S_n(G_0) \longrightarrow 0 \quad \text{wpl.} \tag{2.6}$$

For any y, by the Schwartz inequality, we have

$$\begin{aligned}\left(\int_{-\infty}^{y} |f_G(x) - \hat{f}_n(x)|\, w(x)\, dx\right)^2 &\leq \int_{-\infty}^{y} |f_G(x) - \hat{f}_n(x)|\, w(x)\, dx \int_{-\infty}^{y} w(x)\, dx \\ &\leq S_n(G) \int_{-\infty}^{\infty} w(x)\, dx\end{aligned}$$

Since the right side is independent of y and $\int_{-\infty}^{\infty} w(x)\, dx$ is finite, it follows from (2.6) that

$$\sup_y \int_{-\infty}^{y} |f_G(x) - \hat{f}(x)|\, w(x)\, dx \longrightarrow 0 \quad \text{wpl} \tag{2.7}$$

and

$$\sup_y \int_{-\infty}^{y} |f_{\hat{G}_n}(x) - f(x)|\, w(x)\, dx \longrightarrow 0 \quad \text{wpl.} \tag{2.8}$$

Since

$$\begin{aligned}\sup_y \int_{-\infty}^{y} |f_{\hat{G}_n}(x) - f_{G_0}(x)|\, w(x)\, dx &\leq \sup_y \int_{-\infty}^{y} |f_{\hat{G}_n}(x) - \hat{f}(x)|\, w(x)\, dx \\ &\quad + \sup_y \int_{-\infty}^{y} |f_{G_0}(x) - \hat{f}(x)|\, w(x)\, dx\end{aligned}$$

it follows thus

$$\sup\left(\int_{-\infty}^{y} f_{\hat{G}_n}(x)\,dx - \int_{-\infty}^{y} f_{G_0}(x)\,dx\right) \longrightarrow 0 \qquad \text{wpl.}$$

By [8], we have then $\hat{G}_n \to G_0$ wpl. This completes the proof of (i).

(ii) Since the density estimate is given by (1.5), we have the column vector of the first partial derivative with respect to each component of $\underset{\sim}{G}$ is given by

$$\frac{1}{2}\dot{S}_n(G) = \frac{1}{n}\begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_k \end{pmatrix}$$

where

$$S_i = \sum_{r=1}^{n} \int \left\{ \sum_{j=1}^{k} g_j f_j(x) - \frac{1}{h} K\left(\frac{x - X_r}{h}\right)\right\} f_i(x)\, w(x)\, dx$$

The asymptotic normality of $\dot{S}_n(G)$ can be proved by firstly showing that each of its rows converges in distribution to a normal random variable. It follows from the central limit theroem that the random vector $\sqrt{n}\,\dot{S}_n(G_0)$ has an asymptotic k-variate normal distribution. To show it, we note that the ith row of $\frac{1}{2}\dot{S}_n(G_0)$ is given by $1/n \sum_{1=j}^{n} Y_{ij}$, where $Y_{i1}, Y_{i2}, \cdots, Y_{in}$ are *iid* and Y_{ij} is given by

$$\begin{aligned} Y_{ij} &= Y_{ij}(G_0) \\ &= \int \left\{ \sum_{r=1}^{k} g_r f_r(x) - \frac{1}{h} K\left(\frac{x - X_j}{h}\right)\right\} f_i(x)\, w(x)\, dx\,. \end{aligned}$$

We also note that

$$\begin{aligned} E\,Y_{ij} &= \iint \left\{ f_{G_0}(x) - \frac{1}{h} K\left(\frac{x-y}{h}\right)\right\} f_i(x)\, f_{G_0}(y)\, dx\, dy \\ &= \int f_{G_0}(x)\, f_i(x)\, dx - \frac{1}{h} \iint K\left(\frac{x-y}{h}\right) f_i(x)\, f_{G_0}(y)\, dx\, dy \\ &= \int \{f_{G_0}(x) - E\hat{f}_n(x)\}\, f_i(x)\, dx. \end{aligned}$$

Since $E\hat{f}_n(x) \to f_{G_0}(x)$ for each x, it follows thus

$$E\,Y_{ij}(G_0) \longrightarrow 0 \qquad (j = 1, 2, \cdots, k)\,.$$

Again, by straight computations, we obtain

$$\text{Var } Y_{ij}(G_0)$$
$$= \iint \text{Cov}\left(\frac{1}{h} K\left(\frac{u-X_j}{h}\right), \frac{1}{h} \cdot K\left(\frac{v-X_j}{h}\right)\right) \tag{2.9}$$
$$f_i(u) f_i(v) w(u) w(v)\, du\, dv .$$

Observing that this is finite, it ensures that the ith row of $\sqrt{n}\, \dot{S}_n(G_0)$ converges to a random variable with normal distribution $N(0, 4V_i^2)$ where V_i^2 is given by (2.9). Extending to the k-dimensional case, we have thus

$$\frac{1}{2}\sqrt{n}\, \dot{S}_n(G_0) \xrightarrow{\mathscr{L}} N(\underset{\sim}{0}, \Sigma_n^*)$$

where Σ_n^* denotes the covariance matrix of random variables Y_{i1} and Y_{j1}. The matrix whose elements are the second partial derivatives of $\frac{1}{2} S_n(G_0)$ with respect to each component of $\underset{\sim}{G_0}$ is a constant matrix given by $\frac{1}{2} \ddot{S}_n(G_0) = (a_{ij})$ where

$$a_{ij} = \int f_i(x) f_j(x) w(x)\, dx \qquad i, j = 1, 2, \cdots, k.$$

Using Taylor expansion, we have

$$\dot{S}_n(\hat{G}_n) = \dot{S}_n(G_0) + \ddot{S}_n(\tilde{G})(\hat{G}_n - G_0) \tag{2.10}$$

where $\tilde{G}$ is a convex combination of $\hat{G}_n$ and G_0. Since $\dot{S}_n(\hat{G}_n) = 0$, we have

$$\sqrt{n}\, \Sigma_n^{-1/2}(\hat{G}_n - G_0) \longrightarrow N(\underset{\sim}{0}, \underset{\sim}{1})$$

where $\Sigma_n = A^{-1} \Sigma^* A^{-1}$. This completes the proof of (ii).

(iii) By (2.10) we have

$$0 = \dot{S}_n(\hat{G}_n) = \dot{S}_n(G_0) + \ddot{S}_n(\tilde{G})(\hat{G}_n - G_0) \tag{2.11}$$

Since $S_n(\tilde{G})$ is bounded, hence, the rate of convergence of $(\hat{G}_n - G_0)$ to zero is the same as that of $\dot{S}_n(G_0)$. To find the upper bound of the ith row of the column vector $\dot{S}_n(G_0)$, we note that

$$\left|\int (f_{G_0}(x) - \hat{f}_n(x)) f_i(x) w(x)\, dx\right|$$
$$\leq \left\{\sup_z \left|\int_{-\infty}^{z} (f_{G_0}(y) - \hat{f}_n(y))\, dy\right|\right\} \cdot \{\sup_x f_i(x) w(x)\}$$

Since $K(x)$ is of bounded variation, it follows from [9] that

$$\sup\left|\int_{-\infty}^{x}(f_{G_0}(y)-\hat{f}_n(y)\,dy\right|=O(b_n^{-1})\quad \text{wpl}$$

where $h^2(n)b_n=O(1)$ and $\sum\exp(-\lambda n/b_n^2)<\infty$ for all $\lambda>0$. If $\hat{f}_n(x)=(1/nh_n)\sum_{i=1}^n K((x-X_i)/h_n)$ and $h_n=n^{-\theta}$ for $0<\theta<\frac{1}{2}$, then, by taking $b_n=n^{2\theta-\varepsilon}$, $\varepsilon>0$, $0<\theta<\frac{1}{2}$, the conditions are satisfied. Hence, for $h_n=n^{-\theta}$, $0<\theta<\frac{1}{2}$, $|\hat{G}_n-G_0|\leq O(n^{2\theta-\varepsilon})$ for $\varepsilon>0$. This completes the proof.

3. **Some Monte Carlo results.** We study a special case of $k=2$ and 3 with $f_i(x)=n(x;\ i-1,\ 1)$ $(i=1,2,3)$ where $n(x;\ \theta,\ \sigma^2)$ denotes a normal density with mean θ and variance σ^2. We take $\hat{f}_n(x)=(1/na_n)\sum_{i=1}^n K((x-X_i)/a_n)$ where $a_n=n^{-1/4}$ and $K(x)=n(x;\ 0,\ 1)$ and we also take $w(x)=e^{-x^2}$. We consider the cases of small sample size with $n=10$, 25, 50. For each case, we use computer CDC3170 for a simulation of a sample of size n from the mixture density with associated mixing distribution $\underset{\sim}{G}_0$ which is prefixed beforehand. We use the proposed estimator $\hat{G}_n$ defined by (1.4) to compute the mixing distribution. This process repeats 500 times and then take its arithmetic average as our estimate for $\underset{\sim}{G}_0$. Each time, based on the same simulated data, we also apply the estimator given by [3] (C-B estimator) and this procedure also repeats 500 times and take its average for the estimate of $\underset{\sim}{G}_0$.

In Table 1, associated with mixing probabilities g_1, g_2 and normal means θ_1, θ_2, the tabulations in part I and part II are, respectively, the estimates of g_1 applying the proposed estimator $\hat{G}_n$ and the C-B estimator. The mean square errors with respect to the average and with respect to the true mixing are tabulated, respectively, in the columns of MSE and TMSE. In Table 2, we consider the case of $k=3$. In most cases, it is found that the MSE and TMSE in part I are always less than the associated MSE and TMSE in part II for the case $k=2$ and $n=10$. For $k=3$, it is found that the proposed estimator has less MSE and TMSE than that of the C-B estimator for $n=10$ and 25.

We acknwledge Mr. Pei-Ding Chen for the empirical results of the Table 1 and Table 2 from [2].

Table 1

	n	Part I			Part II		
		$\hat{g}_1$	MSE	TMSE	$\hat{g}_1$	MSE	TMSE
$g_1 = g_2 = 0.5$	10	0.48514	0.34143	0.34165	0.81895	0.44490	0.54662
$\theta_1 = 0,\ \theta_2 = 0.5$	25	0.51586	0.21138	0.21163	0.63146	0.18070	0.19798
	50	0.50402	0.13233	0.13234	0.54607	0.09165	0.09377
$g_1 = g_2 = 0.5$	10	0.48286	0.10978	0.11007	0.67410	0.12387	0.15418
$\theta_1 = 0,\ \theta_2 = 1$	25	0.48517	0.07375	0.07397	0.56651	0.05136	0.05578
	50	0.48553	0.04716	0.04725	0.53762	0.02490	0.02631
$g_1 = g_2 = 0.5$	10	0.46766	0.03314	0.03418	0.59117	0.03288	0.04119
$\theta_1 = 0,\ \theta_2 = 4$	25	0.46776	0.01544	0.01669	0.52706	0.01267	0.01340
	50	0.47644	0.00914	0.00970	0.51450	0.00648	0.00669
$g_1 = 0.4,\ g_2 = 0.6$	10	0.41753	0.34323	0.34354	0.71975	0.44570	0.54794
$\theta_1 = 0,\ \theta_2 = 0.5$	25	0.44150	0.20560	0.20732	0.53282	0.17814	0.19578
	50	0.42151	0.13362	0.13408	0.44632	0.09208	0.09423
$g_1 = 0.4,\ g_2 = 0.6$	10	0.41233	0.10977	0.10992	0.57551	0.14808	0.17888
$\theta_1 = 0,\ \theta_2 = 1$	25	0.41125	0.07226	0.07238	0.46741	0.05003	0.05457
	50	0.40711	0.05019	0.05024	0.42431	0.02615	0.02674
$g_1 = 0.4,\ g_2 = 0.6$	10	0.37961	0.02977	0.03018	0.48806	0.03328	0.04104
$\theta_1 = 0,\ \theta_2 = 4$	25	0.37193	0.01420	0.01499	0.42634	0.01301	0.01370
	50	0.37967	0.00793	0.00834	0.41332	0.00601	0.00619

Table 2

n	Part I						Part II					
	$\hat{g}_1$	MSE	TMSE	$\hat{g}_2$	MSE	TMSE	$\hat{g}_1$	MSE	TMSE	$\hat{g}_2$	MSE	TMSE
$g_1 = g_2 = 0.1,\quad g_3 = 0.8,\quad \theta_1 = 0,\quad \theta_2 = 1,\quad \theta_3 = 2$												
10	0.12725	0.06098	0.06172	0.07277	0.19676	0.19751	0.05061	0.31775	0.32018	0.30459	0.97463	1.01649
25	0.11681	0.03697	0.03725	0.09519	0.10748	0.10751	0.09771	0.06386	0.06386	0.16278	0.25633	0.26027
50	0.10920	0.02650	0.02659	0.11096	0.07268	0.07280	0.09706	0.02975	0.02976	0.14486	0.12449	0.12650
$g_1 = g_2 = 0.4,\quad g_3 = 0.2,\quad \theta_1 = 0,\quad \theta_2 = 1,\quad \theta_3 = 2$												
10	0.43836	0.11435	0.11582	0.22921	0.26270	0.29187	0.47355	0.21802	0.22343	0.41634	0.59342	0.59369
25	0.43495	0.06596	0.06718	0.26507	0.14406	0.16226	0.43301	0.07322	0.07431	0.40102	0.18871	0.18871
50	0.43982	0.04078	0.04236	0.21940	0.08257	0.09711	0.42473	0 03359	0.03420	0.39176	0.08674	0.08681
$g_1 = g_2 = 0.1,\quad g_3 = 0.8,\quad \theta_1 = 0,\quad \theta_2 = 2,\quad \theta_3 = 4$												
10	0.08582	0.01151	0.01171	0.11518	0.04188	0.04211	0.04496	0.48869	0.49172	0.22124	0.66237	0.67707
25	0.09146	0 00621	0.00628	0.10515	0.02381	0.02384	0.10598	0.01758	0.01761	0.12275	0.04683	0.04735
50	0.09510	0.00382	0.00385	0.10800	0.01473	0.01479	0.10610	0.00577	0.00581	0.11261	0.01974	0.01990
$g_1 = g_2 = 0.4,\quad g_3 = 0.2,\quad \theta_1 = 0,\quad \theta_2 = 2,\quad \theta_3 = 4$												
10	0.37306	0.03816	0.03889	0.38364	0.13712	0.13738	0.45784	0.05056	0.05390	0.39864	0.08672	0.08672
25	0.37818	0.02063	0.02111	0.39623	0.08395	0.08396	0.41994	0.01838	0.01878	0.40417	0.02786	0.02788
50	0.38943	0.01265	0.01276	0.39271	0.05825	0.05830	0.41501	0.00830	0.00852	0.39680	0.01243	0.01244

REFERENCES

1. J.R. Blum and V. Susarla, *Estimation of a mixing distribution function*, Ann. Probability, 5 (1977), 200-209.

2. P.D. Chen, *Estimations of the mixing distributions based on the kernel density estimates with empirical studies*, Master thesis, Tamkang University (1981).

3. K. Choi and W.G. Bulgren, *An estimation procedure for mixtures Mixtures of distribution*, J. Roy. Statist. Soc. Ser. B 30 (1968), 444-460.

4. J.J. Deely and R.L. Kruse, *Construction of sequences estimating the mixing distribution*, Ann. Math. Statist., 39 (1968), 286-288.

5. S.S. Gupta and W.T. Huang, *On mixtures of distributions: A survey and some new results on ranking and selection*, Sankhyā Ser. B 43 (1981).

6. E.A. Nadaraya, *On non-parametric estimates of density functions and regression curves*, Theor. Probability, Appl., 10 (1964), 186-190.

7. K. Pearson, *Contributions to the mathematical theory of evolution*, Philos. Trans. Roy. Soc. London, 185 (1894), 71-110.

8. H. Robbins, *The empirical Bayes approach to slatistical decision problems*, Ann. Math. Statist., 35 (1969), 1-20.

9. E.F. Schuster, *On the rate of convergence of an estimate of a functional of a probability density*, Scand. Actuarial J. 2 (1974), 103-107.

10. H. Teicher, *On the mixture of distributions*, Ann. Math. Statist., 31 (1960), 360-369.

11. ______, *Identifiability of mixtures*, Ann. Math. Statist., 32 (1961), 244-248.

12. ______, *Identifiability of finite mixture*, Ann. Math. Statist., 34 (1963), 1265-1269.

13. S. Yakowitz and J. Spragins, *On the identifiability of finite mixture*, Ann. Math. Statist., 39 (1968), 209-214.

14. S.J. Yakowitz, *A consistent estimator for the identification of finite mixtures*, Ann. Math. Statist., 40 (1969), 1728-1735.

INSTITUTE OF MATHEMATICS, ACADEMIA SINICA, NANKANG, TAIPEI, TAIWAN, REPUBLIC OF CHINA.

PEPARTMENT OF MATHEMATICS, NATIONAL CHIAO TUNG UNIVERSITY, TAIPEI, TAIWAN, REPUBLIC OF CHINA.