Sankhyā : The Indian Journal of Statistics
2011, Volume 73-A, Part 2, pp. 218-230

Methods of Moments Estimation in Finite Mixtures

Patrick J. Farrell, A. K. Md. Ehsanes Saleh and Zhengmin Zhang
Carleton University, Ottawa, Canada

Abstract

In this paper, we study the method of moments estimation in a finite mixture where the mixing distribution is supported on $[0, 1]$, and the number of support points is fixed. We prove the validity of the modification method proposed by Lindsay (1989). Further, we demonstrate that the modified estimator is weakly consistent and that its convergence rate is $N^{-\frac{1}{2}}(\log\log N)^{\frac{1}{2}}$.

AMS (2000) *subject classification.* Primary 62G05; Secondary 62G20.
Keywords and phrases. Convergence rate, finite mixture, method of moments, modification, weak consistency.

1 Introduction

Following Zhang (2008), in this paper, we wish to estimate the mixing distribution G in a finite mixture

$$f(x; G) = \sum_{i=1}^{n+1} \pi_i f(x; \lambda_i), \tag{1.1}$$

where $n \geq 1$ is a fixed integer, $f(x; \lambda)$, $\lambda \in [0, 1]$ is a parametric family and

$$G = \sum_{i=1}^{n+1} \pi_i \delta(\lambda_i)$$

is an $(n + 1)$-point distribution on $[0, 1]$.

There are many methods available to estimate G, including the method of moments. The method of moments offers relative simplicity in computation when compared to others, such as maximum likelihood estimation. It reduces the problem of estimating G to the estimation of the first $2n + 1$ moments of G. Generally speaking, the estimation of these moments simply requires solving a polynomial equation and a system of linear equations, which can be accomplished with the use of computer software.

Lindsay (1989) studied the method of moments in finite mixtures where the number of support points of the mixing distribution is fixed. He considered several cases in which the estimated distribution violates the constraints on the parameter space, and proposed a method, without proof, to correct the estimated distribution function so that it satisfies all the constraints. His modification method not only adjusts the estimator while matching some initial moments, but also reduces one or two degrees of the polynomial equation in obtaining the support points of the estimator. We provide a proof of Lindsay's (1989) modification result in this paper.

Despite its computational simplicity, one important concern about the method of moments relative to maximum likelihood estimation is with regards to large sample performance. In particular, we focus here on a comparison of the convergence rates of the modified method of moments estimator and the maximum likelihood estimator when the sample size N is large. Many statisticians have studied the convergence rate of the maximum likelihood estimator. According to Chen (1995), the best convergence rate of the maximum likelihood estimator of G in (1) is $N^{-\frac{1}{2}}$. By contrast, there does not appear to be any result relating to the convergence rate of the modified method of moments estimator of G in the literature. In this paper, we derive a result for this convergence rate. In particular, we show that the convergence rate of the modified method of moments estimator of G is $N^{-\frac{1}{2}}(\log\log N)^{\frac{1}{2}}$, which is not too slow compared with $N^{-\frac{1}{2}}$, the best convergence rate of the maximum likelihood estimator. This result would suggest that the method of moments is valuable in estimating G in model (1.1).

The remainder of this paper is organized as follows. In Section 2, we provide a preliminary lemma that allows us in Section 3 to prove Lindsay's (1989) modification result, as well as the weak consistency of the resulting estimator. In Section 4, we establish the convergence rate of $N^{-\frac{1}{2}}(\log\log N)^{\frac{1}{2}}$ of the modified estimator.

2 A Preliminary Lemma

A sequence of real numbers $\mathbf{m} = (m_0, m_1, \cdots)$ is called a moment sequence if there exists a distribution function (a bounded nondecreasing function) $\varphi(x)$ on R such that

$$m_n = \int_{-\infty}^{\infty} x^n d\varphi(x),\ n = 0, 1, \cdots .$$

For any sequence of real numbers $\mathbf{m} = (m_0, m_1, \cdots)$, we define $\Delta_n(\mathbf{m}),\ \Delta_n^{(1)}(\mathbf{m})$ and $\Delta_n^{(2)}(\mathbf{m})$ as

$$\Delta_n(\mathbf{m}) := \begin{vmatrix} m_0 & m_1 & \cdots & m_n \\ m_1 & m_2 & \cdots & m_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ m_n & m_{n+1} & \cdots & m_{2n} \end{vmatrix} = |m_{i+j}|_{i,j=0}^{n};$$

$$\Delta_n^{(1)}(\mathbf{m}) := \begin{vmatrix} m_1 & m_2 & \cdots & m_{n+1} \\ m_2 & m_3 & \cdots & m_{n+2} \\ \cdots & \cdots & \cdots & \cdots \\ m_{n+1} & m_{n+2} & \cdots & m_{2n+1} \end{vmatrix} = |m_{i+j+1}|_{i,j=0}^{n};$$

$$\Delta_n^{(2)}(\mathbf{m}) := \begin{vmatrix} m_0 - m_1 & m_1 - m_2 & \cdots & m_n - m_{n+1} \\ m_1 - m_2 & m_2 - m_3 & \cdots & m_{n+1} - m_{n+2} \\ \cdots & \cdots & \cdots & \cdots \\ m_n - m_{n+1} & m_{n+1} - m_{n+2} & \cdots & m_{2n} - m_{2n+1} \end{vmatrix}$$
$$= |m_{i+j} - m_{i+j+1}|_{i,j=0}^{n};$$

$n = 0, 1, 2, \cdots$.

LEMMA 2.1. *Let $\varphi(x)$ be a distribution function with exactly $n+1$ support points $x_1 < x_2 < \cdots < x_{n+1}$ and let $\mathbf{m} = (m_0, m_1, \cdots)$ be its moment sequence.*

(a) Then

$$\Delta_k^{(1)}(\mathbf{m}) > 0; \Delta_k^{(2)}(\mathbf{m}) > 0, k \le n$$

if and only if

$$x_1 < x_2 < \cdots < x_{n+1} \in (0, 1).$$

(b) If

$$\Delta_{k-1}^{(1)}(\mathbf{m}) > 0, \Delta_n^{(1)}(\mathbf{m}) \le 0; \Delta_k^{(2)}(\mathbf{m}) > 0, k \le n,$$

then

$$x_1 \le 0,\ x_2 < \cdots < x_{n+1} \in (0, 1).$$

(c) If

$$\Delta_k^{(1)}(\mathbf{m}) > 0; \Delta_{k-1}^{(2)}(\mathbf{m}) > 0, \Delta_n^{(2)}(\mathbf{m}) \le 0, k \le n,$$

then

$$x_1 < \cdots < x_n \in (0, 1),\ x_{n+1} \ge 1.$$

(d) If

$$\Delta_k^{(1)}(\mathbf{m}) > 0, \Delta_n^{(1)}(\mathbf{m}) \le 0; \Delta_k^{(2)}(\mathbf{m}) > 0, \Delta_n^{(2)}(\mathbf{m}) \le 0, k \le n-1,$$

then

$$x_1 \le 0,\ x_2 < \cdots < x_n \in (0, 1),\ x_{n+1} \ge 1.$$

PROOF. For (a), consider the integral

$$\int_{-\infty}^{\infty} x P_n^2(x) d\varphi(x),$$

where $P_n(x)$ is a real polynomial of degree less than or equal to n. If

$$\Delta_k^{(1)}(\mathbf{m}) > 0, k \leq n,$$

then

$$\int_{-\infty}^{\infty} x P_n^2(x) d\varphi(x) > 0$$

for any $P_n(x) \not\equiv 0$. For any $1 \leq i \leq n+1$, choose a special $P_n(x)$ such that $x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_{n+1}$ are the roots of $P_n(x)$. If we assume that w_i is the weight of x_i, then

$$\int_{-\infty}^{\infty} x P_n^2(x) d\varphi(x) = x_i w_i P_n^2(x_i) > 0,$$

which implies $x_i > 0$. Conversely, if $x_i > 0, i = 1, \cdots, n+1$, then

$$\int_{-\infty}^{\infty} x P_n^2(x) d\varphi(x) > 0$$

for any $P_n(x) \not\equiv 0$. So

$$\Delta_k^{(1)}(\mathbf{m}) > 0, k \leq n.$$

Now we turn to consider the integral

$$\int_{-\infty}^{\infty} (1-x) P_n^2(x) d\varphi(x),$$

where $P_n(x)$ is a real polynomial of degree less than or equal to n. If

$$\Delta_k^{(2)}(\mathbf{m}) > 0, k \leq n,$$

then

$$\int_{-\infty}^{\infty} (1-x) P_n^2(x) d\varphi(x) > 0$$

for any $P_n(x) \not\equiv 0$. For any $1 \leq i \leq n+1$, if we choose a special $P_n(x)$ such that $x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_{n+1}$ are the roots of $P_n(x)$, then

$$\int_{-\infty}^{\infty} (1-x) P_n^2(x) d\varphi(x) = (1-x_i) w_i P_n^2(x_i) > 0,$$

which implies $1 - x_i > 0$ or $x_i < 1$. Conversely, if $x_i < 1, i = 1, \cdots, n+1$, then

$$\int_{-\infty}^{\infty} (1-x) P_n^2(x) d\varphi(x) > 0$$

for any $P_n(x) \not\equiv 0$. So

$$\Delta_k^{(2)}(\mathbf{m}) > 0, k \leq n.$$

For (b), from the proof of (a), we know $x_i < 1, i = 1, \cdots, n+1$. Since

$$\Delta_n^{(1)}(\mathbf{m}) \le 0,$$

$x_1 \le 0$. If $x_2 \le 0$, by choosing an appropriate real polynomial $P_{n-1}(x)$ of degree $n-1$ such that $P_{n-1}(x_i) = 0, 3 \le i \le n+1$, then we have

$$\int_{-\infty}^{\infty} x P_{n-1}^2(x) d\varphi(x) = x_1 w_1 P_{n-1}^2(x_1) + x_2 w_2 P_{n-1}^2(x_2) \le 0,$$

which contradicts

$$\Delta_k^{(1)}(\mathbf{m}) > 0, k \le n-1.$$

Thus $x_2 > 0$, so we can conclude that $x_1 \le 0, x_2 < \cdots < x_{n+1} \in (0,1)$.

The proofs for (c) and (d) are similar to the proof for (b). □

3 Estimation and Modification

Lindsay (1989) has shown that if $f(x;\lambda)$ is a member of the quadratic variance natural exponential family (QVEF) (the variance is at most a quadratic function of the mean), and parameter λ is its mean value, then the estimator $\hat{a}_{Ni}$ of the i-th moment m_i of the mixing distribution G is easily found such that

$$\hat{a}_{Ni} \to m_i$$

almost surely as the sample size $N \to \infty$.

Morris (1982, 1983) demonstrated that the normal, Poisson, gamma, binomial, negative binomial, and NEF-GHS (natural exponential family generated by the hyperbolic secant) distributions are the six univariate natural exponential families (NEF) with quadratic variance functions (QVF). Throughout this paper, we assume

$$\{f(x;\lambda), \lambda \in [0,1]\}$$

is a QVEF and λ is the mean value. Since

$$\Delta_k(\hat{\mathbf{a}}_N) > 0, k \le n$$

when N is large enough, by Mammana's (1954) result, there exists an $(n+1)$-point distribution $\hat{G}_N$ fitting moments $\hat{a}_{N0}, \hat{a}_{N1}, \cdots, \hat{a}_{N2n+1}$. This distribution $\hat{G}_N$ converges to G weakly.

Suppose that in model (1.1) the true mixing distribution G has support point $\lambda = 0$ or $\lambda = 1$. Then, when N is sufficiently large, with positive probability at least one of support points of the estimated distribution $\hat{G}_N$ lie outside of $[0,1]$. Lindsay (1989) proposed a modification such that the corrected estimator $\tilde{G}_N$ fits the initial $2n+1$ or $2n$ moments. Here we give the proof of this modification as well as the weak consistency of the modified estimator.

Let $t_1 < t_2 < \cdots < t_{n+1}$ be the support points of $\hat{G}_N$. According to Lemma 2.1,

(i) If

$$\Delta_k^{(1)}(\hat{\mathbf{a}}_N) > 0; \Delta_k^{(2)}(\hat{\mathbf{a}}_N) > 0, k \le n,$$

then

$$t_1 < t_2 < \cdots < t_{n+1} \in (0,1).$$

(ii) If

$$\Delta_{k-1}^{(1)}(\hat{\mathbf{a}}_N) > 0, \Delta_n^{(1)}(\hat{\mathbf{a}}_N) \le 0; \Delta_k^{(2)}(\hat{\mathbf{a}}_N) > 0, k \le n,$$

then

$$t_1 \le 0,\ t_2 < \cdots < t_{n+1} \in (0,1).$$

(iii) If

$$\Delta_k^{(1)}(\hat{\mathbf{a}}_N) > 0; \Delta_{k-1}^{(2)}(\hat{\mathbf{a}}_N) > 0, \Delta_n^{(2)}(\hat{\mathbf{a}}_N) \le 0, k \le n,$$

then

$$t_1 < \cdots < t_n \in (0,1),\ t_{n+1} \ge 1.$$

(iv) If

$$\Delta_k^{(1)}(\hat{\mathbf{a}}_N) > 0, \Delta_n^{(1)}(\hat{\mathbf{a}}_N) \le 0; \Delta_k^{(2)}(\hat{\mathbf{a}}_N) > 0, \Delta_n^{(2)}(\hat{\mathbf{a}}_N) \le 0, k \le n-1,$$

then

$$t_1 \le 0,\ t_2 < \cdots < t_n \in (0,1),\ t_{n+1} \ge 1.$$

For case (i), no modification is needed. For case (ii), an $(n+1)$-point modified distribution $\tilde{G}_N$ with $t_1 = 0$ that fits the first $2n+1$ moments $\hat{a}_{N0}, \hat{a}_{N1}, \cdots, \hat{a}_{N2n}$ is derived as follows. If $t_2, \cdots, t_{n+1}$ are the other n support points of $\tilde{G}_N$, and $p_2, \cdots, p_{n+1}$ are the corresponding weights, then

$$\hat{a}_{Ni} = \sum_{j=2}^{n+1} p_j t_j^i,\ i = 1, \cdots, 2n. \tag{3.1}$$

Define

$$\hat{m}_{Ni} := \hat{a}_{Ni},\ i \le 2n,$$
$$\hat{m}_{Ni} := \sum_{j=2}^{n+1} p_j t_j^i,\ i \ge 2n+1.$$

Then $\hat{\mathbf{m}}_N = (\hat{m}_{Ni})_{i=0}^{\infty}$ is the moment sequence of $\tilde{G}_N$. Since $t = 0$ is a solution of

$$\begin{vmatrix} 1 & t & \cdots & t^{n+1} \\ \hat{m}_{N0} & \hat{m}_{N1} & \cdots & \hat{m}_{Nn+1} \\ \hat{m}_{N1} & \hat{m}_{N2} & \cdots & \hat{m}_{Nn+2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{m}_{Nn} & \hat{m}_{Nn+1} & \cdots & \hat{m}_{N2n+1} \end{vmatrix} = 0,$$

we have

$$\Delta_n^{(1)}(\hat{\mathbf{m}}_N) = 0.$$

Following Theorem 1.3 in Shohat and Tamarkin (1963), $t_2, \cdots, t_{n+1} > 0$. If in equation (3.1) we take $p_j t_j > 0$ as the weight of t_j and consider the n-point distribution that has its first $2n$ moments $\hat{a}_{Ni},\ i = 1, \cdots, 2n$ (where $\hat{a}_{Ni}$ is the $(i-1)$-th moment), then all $t_j,\ 2 \le j \le n+1$, are solutions of

$$\begin{vmatrix} 1 & t & \cdots & t^n \\ \hat{a}_{N1} & \hat{a}_{N2} & \cdots & \hat{a}_{Nn+1} \\ \hat{a}_{N2} & \hat{a}_{N3} & \cdots & \hat{a}_{Nn+2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{a}_{Nn} & \hat{a}_{Nn+1} & \cdots & \hat{a}_{N2n} \end{vmatrix} = 0. \tag{3.2}$$

For case (iii), we create a distribution $\tilde{G}_N$ with $t_{n+1} = 1$ that fits the initial $2n+1$ moments $\hat{a}_{N0}, \hat{a}_{N1}, \cdots, \hat{a}_{N2n}$. If we let $t_1, \cdots, t_n$ be the other n support points of $\tilde{G}_N$, then

$$\hat{a}_{Ni} = \sum_{j=1}^{n} p_j t_j^i + p_{n+1},\ i = 0, 1, \cdots, 2n.$$

Suppose $\hat{\mathbf{m}}_N = (\hat{m}_{Ni})_{i=0}^{\infty}$ is the moment sequence of $\tilde{G}_N$. Because

$$\Delta_k^{(2)}(\hat{\mathbf{m}}_N) > 0,\ k \le n-1,$$

by the theory of quadratic form,

$$\int_{-\infty}^{\infty} (1-t) P_{n-1}^2(t) d\tilde{G}_N(t) > 0$$

for any less than or equal to $(n-1)$ degree polynomial $P_{n-1}(t) \not\equiv 0$, which implies that $t_1, \cdots, t_n < 1$. Since

$$\hat{a}_{Ni} - \hat{a}_{Ni+1} = \sum_{j=1}^{n} p_j (1-t_j) t_j^i,\ i = 0, 1, \cdots, 2n-1,$$

if we take $p_j(1-t_j) > 0$ as the weight of t_j and consider an n-point distribution that fits the initial $2n$ moments $\hat{a}_{Ni} - \hat{a}_{Ni+1},\ i = 0, 1, \cdots, 2n-1$, then all $t_j, 1 \le j \le n$ are the solutions of

$$\begin{vmatrix} 1 & t & \cdots & t^n \\ \hat{a}_{N0} - \hat{a}_{N1} & \hat{a}_{N1} - \hat{a}_{N2} & \cdots & \hat{a}_{Nn} - \hat{a}_{Nn+1} \\ \hat{a}_{N1} - \hat{a}_{N2} & \hat{a}_{N2} - \hat{a}_{N3} & \cdots & \hat{a}_{Nn+1} - \hat{a}_{Nn+2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{a}_{Nn-1} - \hat{a}_{Nn} & \hat{a}_{Nn} - \hat{a}_{Nn+1} & \cdots & \hat{a}_{N2n-1} - \hat{a}_{N2n} \end{vmatrix} = 0. \tag{3.3}$$

If there is a support point derived above in case (ii) larger than 1, or there is a support point derived above in case (iii) less than zero, or case (iv) occurs, we only fit the first $2n$ moments $\hat{a}_{N0}, \hat{a}_{N1}, \cdots, \hat{a}_{N2n-1}$. We create a modified distribution $\tilde{G}_N$ with $t_1 = 0,\ t_{n+1} = 1$. Let $t_2, \cdots, t_n$ be the other $n-1$ support points of $\tilde{G}_N$. Then

$$\hat{a}_{Ni} = \sum_{j=2}^{n} p_j t_j^i + p_{n+1},\ i = 1, \cdots, 2n-1.$$

Suppose $\hat{\mathbf{m}}_N = (\hat{m}_{Ni})_{i=0}^{\infty}$ is the moment sequence of $\tilde{G}_N$. Since $t = 0,\ t = 1$ are the solutions of

$$\begin{aligned}
&\begin{vmatrix} 1 & t & \cdots & t^{n+1} \\ \hat{m}_{N0} & \hat{m}_{N1} & \cdots & \hat{m}_{Nn+1} \\ \hat{m}_{N1} & \hat{m}_{N2} & \cdots & \hat{m}_{Nn+2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{m}_{Nn} & \hat{m}_{Nn+1} & \cdots & \hat{m}_{N2n+1} \end{vmatrix} \\
= \quad & (-1)^n \begin{vmatrix} 1 & 1-t & \cdots & (1-t)t^n \\ \hat{m}_{N0} & \hat{m}_{N0} - \hat{m}_{N1} & \cdots & \hat{m}_{Nn} - \hat{m}_{Nn+1} \\ \hat{m}_{N1} & \hat{m}_{N1} - \hat{m}_{N2} & \cdots & \hat{m}_{Nn+1} - \hat{m}_{Nn+2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{m}_{Nn} & \hat{m}_{Nn} - \hat{m}_{Nn+1} & \cdots & \hat{m}_{N2n} - \hat{m}_{N2n+1} \end{vmatrix} \\
= \quad & 0,
\end{aligned}$$

we have

$$\Delta_n^{(1)}(\hat{\mathbf{m}}_N) = 0,\ \Delta_n^{(2)}(\hat{\mathbf{m}}_N) = 0.$$

Following Lemma 2.1(d), $t_2, \cdots, t_n \in (0, 1)$. Since

$$\hat{a}_{Ni} - \hat{a}_{Ni+1} = \sum_{j=2}^{n} p_j (1 - t_j) t_j t_j^{i-1},\ i = 1, 2, \cdots, 2n-2,$$

if we take $p_j(1-t_j)t_j > 0$ as the weight of t_j and consider the $(n-1)$-point distribution that has the first $2n-2$ moments $\hat{a}_{Ni} - \hat{a}_{Ni+1},\ i = 1, \cdots, 2n-2$ (where $\hat{a}_{Ni} - \hat{a}_{Ni+1}$ is the $(i-1)$-th moment), then all $t_j,\ 2 \le j \le n$ are the solutions of

$$\begin{vmatrix} 1 & t & \cdots & t^{n-1} \\ \hat{a}_{N1} - \hat{a}_{N2} & \hat{a}_{N2} - \hat{a}_{N3} & \cdots & \hat{a}_{Nn} - \hat{a}_{Nn+1} \\ \hat{a}_{N2} - \hat{a}_{N3} & \hat{a}_{N3} - \hat{a}_{N4} & \cdots & \hat{a}_{Nn+1} - \hat{a}_{Nn+2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{a}_{Nn-1} - \hat{a}_{Nn} & \hat{a}_{Nn} - \hat{a}_{Nn+1} & \cdots & \hat{a}_{N2n-2} - \hat{a}_{N2n-1} \end{vmatrix} = 0. \qquad (3.4)$$

REMARK 3.1. In fact, when case (ii), (iii), or (iv) occurs, we can simply truncate $\hat{G}_N$ to get

$$G_N^*(t) = \begin{cases} \hat{G}_N(t), & 0 \le t < 1, \\ 0, & t < 0, \\ 1, & t \ge 1. \end{cases}$$

Clearly, G_N^* also converges to G weakly. However, as Lindsay (1989) pointed out, G_N^* no longer fits any of the moments. Moreover, in order to obtain G_N^*, we must solve the equation

$$\begin{vmatrix} 1 & t & \cdots & t^{n+1} \\ \hat{a}_{N0} & \hat{a}_{N1} & \cdots & \hat{a}_{Nn+1} \\ \hat{a}_{N1} & \hat{a}_{N2} & \cdots & \hat{a}_{Nn+2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{a}_{Nn} & \hat{a}_{Nn+1} & \cdots & \hat{a}_{N2n+1} \end{vmatrix} = 0 \tag{3.5}$$

to obtain the $n+1$ support points of $\hat{G}_N$, then process truncation. In Lindsay's (1989) modification, in order to obtain the support points of $\tilde{G}_N$, we need only solve the polynomial equation (3.2), (3.3), or (3.4) which are one or two degrees less than equation (3.5), thus reducing computational complexity. Following Mammana's (1954) result, the corresponding weights can be calculated by solving a system of linear equations.

THEOREM 3.1. *The above modified estimator $\tilde{G}_N$ is weakly consistent.*

PROOF. (a) If $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_{n+1} < 1$, then when N is sufficiently large, only case (i) happens. So $\tilde{G}_N$ is identical with $\hat{G}_N$ and the weak convergence of $\tilde{G}_N$ to G is obvious.

(b) If $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_{n+1} < 1$, then only case (i) or case (ii) can occur when N is large enough. If case (i) occurs, then the situation is the same as when $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_{n+1} < 1$. If case (ii) occurs, then one support point of $\tilde{G}_N$ is 0, and all the other support points of $\tilde{G}_N$ are the roots of (3.2). The mixing distribution G has the same property, with $\hat{\mathbf{a}}_N$ replaced by $\mathbf{m}$ in (3.2). Thus, regardless of whether case (i) or case (ii) occurs, the estimator $\tilde{G}_N$ converges weakly to G.

(c) When $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_{n+1} = 1$ or $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_{n+1} = 1$, the proofs are analogous to (b).

Hence, in any case $\tilde{G}_N$ is consistent. □

4 Convergence Rate

In this section, we will give the convergence rate of the modified method of moments estimate $\tilde{G}_N$. Note that Chen (1995) established the best possible rate $N^{-\frac{1}{4}}$ of convergence associated with the maximum likelihood estimate in finite mixture models when the exact number of support points of the mixing distribution is unknown, while the $N^{-\frac{1}{2}}$ convergence rate is achievable when the exact number of support points is known. He used the quantity

$$d(G_1, G_2) = \int_\Lambda |G_1(\lambda) - G_2(\lambda)| d\lambda$$

to measure the discrepancy of the two mixing distributions G_1 and G_2, where Λ is a compact set. We adopt the same metric. By applying the law of the iterated logarithm of Hartman and Wintner (1941), we can prove

THEOREM 4.1. *The convergence rate of* $\tilde{G}_N$ *is* $N^{-\frac{1}{2}}(\log\log N)^{\frac{1}{2}}$.

PROOF. To begin with, assume that

$$0 < \lambda_1 < \lambda_2 < \cdots < \lambda_{n+1} < 1.$$

When N is sufficiently large, only case (i) occurs. The estimator $\tilde{G}_N$ matches the first $2n+2$ moments

$$\hat{a}_{N0}, \hat{a}_{N1}, \cdots, \hat{a}_{N2n+1}.$$

Let

$$0 < t_1 < t_2 < \cdots < t_{n+1} < 1$$

be the support points of $\tilde{G}_N$ and

$$p_1, p_2, \cdots, p_{n+1}$$

be the respective masses. Then

$$\begin{aligned}
\hat{a}_{N0} - m_0 &= \sum_{j=1}^{n+1} (p_j - \pi_j), \\
\hat{a}_{Ni} - m_i &= \sum_{j=1}^{n+1} (p_j t_j^i - \pi_j \lambda_j^i) \\
&= \sum_{j=1}^{n+1} \left[\left(\sum_{k=0}^{i-1} p_j \lambda_j^k t_j^{i-1-k} \right)(t_j - \lambda_j) + \lambda_j^i (p_j - \pi_j) \right],
\end{aligned}$$

for $i = 1, \cdots, 2n+1$. These moment equations can be written in matrix form

$$\begin{pmatrix} \hat{a}_{N0} - m_0 \\ \hat{a}_{N1} - m_1 \\ \vdots \\ \vdots \\ \vdots \\ \hat{a}_{N2n+1} - m_{2n+1} \end{pmatrix} = A_N \begin{pmatrix} t_1 - \lambda_1 \\ t_2 - \lambda_2 \\ \vdots \\ t_{n+1} - \lambda_{n+1} \\ p_1 - \pi_1 \\ \vdots \\ p_{n+1} - \pi_{n+1} \end{pmatrix}. \tag{4.1}$$

Denote the limiting matrix of A_N by A, then

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ \pi_1 & \pi_2 & \cdots & \pi_{n+1} & \lambda_1 & \lambda_2 & \cdots & \lambda_{n+1} \\ 2\pi_1\lambda_1 & 2\pi_2\lambda_2 & \cdots & 2\pi_{n+1}\lambda_{n+1} & \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_{n+1}^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (2n+1)\pi_1\lambda_1^{2n} & (2n+1)\pi_2\lambda_2^{2n} & \cdots & (2n+1)\pi_{n+1}\lambda_{n+1}^{2n} & \lambda_1^{2n+1} & \lambda_2^{2n+1} & \cdots & \lambda_{n+1}^{2n+1} \end{pmatrix}.$$

We will show that matrix A is invertible. However, to do so, since all $\lambda_1, \cdots, \lambda_{n+1} \neq 0$, we only need to prove that the matrix

$$B(n; \lambda_1, \cdots, \lambda_{n+1}) = \begin{pmatrix} 1 & \cdots & 1 & \lambda_1 & \cdots & \lambda_{n+1} \\ 2\lambda_1 & \cdots & 2\lambda_{n+1} & \lambda_1^2 & \cdots & \lambda_{n+1}^2 \\ 3\lambda_1^2 & \cdots & 3\lambda_{n+1}^2 & \lambda_1^3 & \cdots & \lambda_{n+1}^3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (2n+2)\lambda_1^{2n+1} & \cdots & (2n+2)\lambda_{n+1}^{2n+1} & \lambda_1^{2n+2} & \cdots & \lambda_{n+1}^{2n+2} \end{pmatrix}$$

is invertible. By induction, we can prove that

$$\begin{aligned} |B(n; \lambda_1, \cdots, \lambda_{n+1})| = (-1)^{3n+5}\lambda_1^2(\lambda_2 - \lambda_1)^4 \cdots (\lambda_{n+1} - \lambda_1)^4 \cdots \\ (-1)^{3(n-1)+5}\lambda_2^2(\lambda_3 - \lambda_2)^4 \cdots (\lambda_{n+1} - \lambda_2)^4 \cdots \\ (-1)^{3+5}\lambda_n^2(\lambda_{n+1} - \lambda_n)^4(-1)^1\lambda_{n+1}^2, \end{aligned}$$

which is nonzero since $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_{n+1} < 1$. Thus, when N is sufficiently large, A_N is also invertible. By (4.1)

$$\begin{pmatrix} t_1 - \lambda_1 \\ t_2 - \lambda_2 \\ \vdots \\ t_{n+1} - \lambda_{n+1} \\ p_1 - \pi_1 \\ \vdots \\ p_{n+1} - \pi_{n+1} \end{pmatrix} = A_N^{-1} \begin{pmatrix} \hat{a}_{N0} - m_0 \\ \hat{a}_{N1} - m_1 \\ \vdots \\ \vdots \\ \vdots \\ \hat{a}_{N2n+1} - m_{2n+1} \end{pmatrix}. \tag{4.2}$$

Suppose $f(x; \lambda)$ is the density function of a random variable Y. According to Morris (1982), there is a polynomial $\gamma_i(Y)$ of degree i in Y that is an unbiased estimator of λ^i. Further, let X be the random variable having mixture density $f(x; G)$. Then $\gamma_i(X)$ is an unbiased estimator of the i-th moment m_i of G. Therefore we may take $\hat{a}_{Ni}$ as

$$\frac{\sum_{k=1}^N \gamma_i(X_k)}{N}$$

for a sample $X_1, \cdots, X_N$ from $f(x; G)$. By the strong law of large numbers,

$$\hat{a}_{Ni} \to m_i$$

almost surely.

Suppose that σ_i^2 is the variance of $\gamma_i(X)$. Then, appealing to the law of the iterated logarithm of Hartman and Wintner (1941), which asserts that with probability one the set of limit points of the sequence

$$\left((2N \log\log N)^{-\frac{1}{2}} \sum_{k=1}^{N} \frac{\gamma_i(X_k) - m_i}{\sigma_i} \right)_{N \geq 3}$$

coincides with $[-1,1]$, there are constants c_1, c_2 such that when N is sufficiently large

$$c_1 \le \frac{\hat{a}_{Ni} - m_i}{N^{-\frac{1}{2}}(\log\log N)^{\frac{1}{2}}} \le c_2$$

almost surely. Therefore we come to the conclusion that

$$\hat{a}_{Ni} - m_i \to 0$$

at rate $N^{-\frac{1}{2}}(\log\log N)^{\frac{1}{2}}$. From (4.2),

$$t_i - \lambda_i \to 0,$$

$$p_i - \pi_i \to 0$$

at rate $N^{-\frac{1}{2}}(\log\log N)^{\frac{1}{2}}$. Therefore,

$$d(\tilde{G}_N, G) \to 0$$

at rate $N^{-\frac{1}{2}}(\log\log N)^{\frac{1}{2}}$ almost surely.

Note that under the assumptions

$$0 = \lambda_1 < \lambda_2 < \cdots < \lambda_{n+1} < 1;$$
$$0 < \lambda_1 < \lambda_2 < \cdots < \lambda_{n+1} = 1;$$
$$0 = \lambda_1 < \lambda_2 < \cdots < \lambda_{n+1} = 1,$$

in a similar fashion, we can show that with probability one

$$d(\tilde{G}_N, G) \to 0$$

at rate $N^{-\frac{1}{2}}(\log\log N)^{\frac{1}{2}}$. □

Acknowledgement. The authors are grateful to the referee, Co-editor, and Editor for their valuable comments and suggestions. This research forms a portion of the doctoral thesis of Zhengmin Zhang, and was supported by funds from the Natural Sciences and Engineering Research Council of Canada.

References

CHEN, J. (1995). Optimal rate of convergence for finite mixture models. *Ann. Statist.*, **23**, 221–233.

HARTMAN, P. and WINTNER, A. (1941). On the law of the iterated logarithm. *Amer. J. Math.*, **63**, 169–176.

LINDSAY, B.G. (1989). Moment matrices: Applications in mixtures. *Ann. Statist.*, **17**, 722–740.

Mammana, C. (1954). Sul problema algebraico dei momenti. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, **8**, 133–140.

Morris, C.N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.*, **10**, 65–80.

Morris, C.N. (1983). Natural exponential families with quadratic variance functions: Statistical theory. *Ann. Statist.*, **11**, 515–529.

Shohat, J.A. and Tamarkin, J.D. (1963). *The Problem of Moments.* American Mathematical Society Mathematical Surveys, vol. II. American Mathematical Society, New York.

Zhang, Z. (2008). Estimation in Mixture Models. PhD Thesis, School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada.

Patrick J. Farrell, A. K. Md. Ehsanes Saleh and Zhengmin Zhang
School of Mathematics and Statistics
Carleton University
1125 Colonel By Drive
Ottawa, Ontario, Canada, K1S 5B6
E-mail: pfarrell@math.carleton.ca
esaleh@math.carleton.ca
zzm47@yahoo.ca

Paper received September 2008; revised June 2010.