DELTA METHOD APPROACH IN A CERTAIN IRREGULAR CONDITION

Dedicated to the retirement of Professor Nariaki Sugiura from the University of Tsukuba

Yutaka Kano
Faculty of Human Sciences, Osaka University
Suita, Osaka 565-0871, Japan; kano@hus.osaka-u.ac.jp

Key Words: Asymptotic distribution; Differentiability; Factor analysis; The number of components;

ABSTRACT

Let $\boldsymbol{X}_n$ be a sequence of random p-vectors such that $a_n(\boldsymbol{X}_n - \boldsymbol{b}) \xrightarrow{L} \boldsymbol{Z}$, where $a_n \nearrow \infty$, $\boldsymbol{b} \in R^p$ and $\boldsymbol{Z}$ is a continuously distributed random p-vector. Let $\boldsymbol{f}(\cdot)$ be a measurable mapping from a domain of R^p to R^q, where the domain may not include $\boldsymbol{b}$, i.e., $\boldsymbol{f}(\boldsymbol{b})$ may not be defined. Under this setup, we study the asymptotic distribution of $\boldsymbol{f}(\boldsymbol{X}_n)$. Two theorems are developed to obtain the asymptotic distribution. Comprehensive examples are provided to show when and where such an irregular situation takes place and to illustrate the usefulness of these theorems. The examples include the problem of choosing the number of components and noniterative estimation in factor analysis.

1. INTRODUCTION AND MOTIVE EXAMPLES

Let $\{\boldsymbol{X}_n\}$ be a sequence of random p-vectors such that $\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{b}) \xrightarrow{L} \boldsymbol{Z}$, where $\boldsymbol{b} \in R^p$ and $\boldsymbol{Z}$ is typically normally distributed. The delta method approach is particularly useful in obtaining asymptotic distributions of statistics

$\boldsymbol{f}(\boldsymbol{X}_n)$, where $\boldsymbol{f}(\cdot)$ is a measurable mapping from a domain of R^p to R^q. When $\boldsymbol{f}(\boldsymbol{x})$ is (totally) differentiable at $\boldsymbol{x} = \boldsymbol{b}$, the theorem shows that

$$\sqrt{n}\Big(\boldsymbol{f}(\boldsymbol{X}_n) - \boldsymbol{f}(\boldsymbol{b})\Big) \xrightarrow{L} \frac{\partial \boldsymbol{f}(\boldsymbol{b})}{\partial \boldsymbol{x}'}\boldsymbol{Z} \qquad \text{as } n \to \infty.$$

See e.g. Rao (1973, page 388) for a proof. We denote convergence in probability and in distribution by $\xrightarrow{P}$ and $\xrightarrow{L}$, respectively. What happens if the important regularity condition — differentiability — fails? More *irregularly* what if $f(\boldsymbol{x})$ can not be defined at $\boldsymbol{x} = \boldsymbol{b}$? This paper discusses asymptotic theory in such situations.

In multivariate analysis, a typical example of cases where the differentiability condition is violated is asymptotic theory for latent roots or singular values of a random matrix with multiplicity. In the case, sample roots or singular values are continuous but not differentiable. A large number of significant works on the topic have been published (e.g., Anderson 1963; Chikuse 1976; Eaton and Tyler 1991, 1994; Fujikoshi 1977, 1980; Sugiura 1973, 1976, 1990; Tyler 1983; Waternaux 1976). We shall take the simplest case where $nS = (ns_{ij})$ follows $W_2(n, I_2)$, the Wishart distribution with I_2 the parameter matrix and n degrees of freedom. The eigenvalues, $(\hat{\lambda}_1, \hat{\lambda}_2)$, of S are expressed as

$$\frac{1}{2}\left(s_{11} + s_{22} \pm \sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2}\right). \tag{1}$$

It is easily seen that $\hat{\lambda}_i$'s are not differentiable at $S = I_2$. The asymptotic distribution of $(\hat{\lambda}_1, \hat{\lambda}_2)$ and its relation with our theorem will be noted in Section 2. Notice that the eigenvalues are defined at $S = I_2$. In this paper, we discuss a more difficult situation where the function $\boldsymbol{f}(\boldsymbol{x})$ may not be defined at $\boldsymbol{x} = \boldsymbol{b}$.

Asymptotics in many types of nonregular cases have been studied: density functions are not differentiable; support of density functions depends on parameters; rather strong dependency is introduced between observations and so forth (e.g., Akahira and Takeuchi 1995; Roussas 1972; Basawa and Scott 1983; Tanaka 1996). The situation we discuss here takes place when the Fisher information matrix is singular or the probability model is not identified.

Now we shall briefly discuss the problem of choosing the number of components in finite normal mixture distribution as a simple but motive example.

Let $Y_1, \ldots, Y_n$ be a random sample from a population distributed as

$$\pi N(\mu_1, \sigma^2) + (1-\pi)N(\mu_2, \sigma^2), \tag{2}$$

where $\mu_1 \neq \mu_2$. One approach to choosing the number of components is to test the following hypothesis:

$$H_0 : \pi = 0 \text{ or } 1 \quad \text{versus} \quad H_1 : 0 < \pi < 1. \tag{3}$$

Many problems of the method of likelihood ratio (LR) have been reported, among which Hartigan (1985) showed that the LR test statistic diverges to infinity in probability as $n \to \infty$ and Romano and Siegel (1986) showed the MLE does not exist. These anomalies occur partially because the density is independent of the parameter μ_2 for $\pi = 1$ or μ_1 for $\pi = 0$ in H_0, and then the Fisher information matrix is singular.

We shall consider the method of moments for the testing problem. First, we assume that μ_1 and σ^2 are known and the other two parameters π and μ_2 are to be estimated. A more general case will be discussed in Section 2. Let

$$\left\{ \begin{array}{lcl} X_1^{(n)} & = & \frac{1}{n}\sum_{i=1}^n Y_i \quad (= \bar{Y}, \quad \text{say}) \\ X_2^{(n)} & = & \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2, \end{array} \right. \tag{4}$$

and calculate $\mathrm{E}[Y_i] = (1-\pi)(\mu_2-\mu_1)+\mu_1$ and $\mathrm{Var}[Y_i] = \pi(1-\pi)(\mu_2-\mu_1)^2+\sigma^2$. The estimator based on the moment method is a solution to the simultaneous equations:

$$\left\{ \begin{array}{lcl} X_1^{(n)} & = & (1-\pi)(\mu_2-\mu_1) + \mu_1 \\ X_2^{(n)} & = & \pi(1-\pi)(\mu_2-\mu_1)^2 + \sigma^2. \end{array} \right. \tag{5}$$

The estimator for π can be then expressed as

$$\hat{\pi} = \hat{\pi}(\boldsymbol{X}_n) = \frac{X_2^{(n)} - \sigma^2}{(X_1^{(n)} - \mu_1)^2 + (X_2^{(n)} - \sigma^2)}, \tag{6}$$

where $\boldsymbol{X}_n = [X_1^{(n)}, X_2^{(n)}]'$.

Consider separately the cases where $\pi = 1$ and $\pi = 0$ in H_0. First, assume that $\pi = 1$ holds. We then have

$$\begin{aligned} \boldsymbol{X}_n = \left[\begin{array}{c} X_1^{(n)} \\ X_2^{(n)} \end{array} \right] &\xrightarrow{P} \left[\begin{array}{c} \mathrm{E}[X_1^{(n)}] \\ \mathrm{E}[X_2^{(n)}] \end{array} \right] = \left[\begin{array}{c} \mu_1 \\ \sigma^2 \end{array} \right] = \boldsymbol{b}, \quad \text{say}, \\ \sqrt{n}(\boldsymbol{X}_n - \boldsymbol{b}) &\xrightarrow{L} \boldsymbol{Z} = \left[\begin{array}{c} Z_1 \\ Z_2 \end{array} \right] \sim N_2\left(\boldsymbol{0}, \left[\begin{array}{cc} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{array} \right]\right). \end{aligned} \tag{7}$$

It is seen from (6) that $\hat{\pi}(\boldsymbol{X}_n)$ can not be defined at $\boldsymbol{X}_n = \boldsymbol{b}$, the limiting point of $\boldsymbol{X}_n$.

Even though the difficulty about the definition arises, consistency of $\hat{\pi}$ holds and its asymptotic distribution can be easily derived. These asymptotic results are obtained by direct application of Theorem 1, as described in Section 2. A rough proof is stated in this section to help the reader understand the statement and the proof of Theorem 1 more easily.

It follows from (6) and (7) that

$$\begin{aligned}
\hat{\pi} &= \frac{\sqrt{n}(X_2^{(n)} - \sigma^2)}{\sqrt{n}(X_1^{(n)} - \mu_1)^2 + \sqrt{n}(X_2^{(n)} - \sigma^2)} \\
&= \frac{\sqrt{n}(X_2^{(n)} - \sigma^2)}{O_p(n^{-1/2}) + \sqrt{n}(X_2^{(n)} - \sigma^2)} \\
&\xrightarrow{P} 1 \\
\sqrt{n}(\hat{\pi} - 1) &= -\frac{(\sqrt{n}(X_1^{(n)} - \sigma^2))^2}{\sqrt{n}(X_1^{(n)} - \mu_1)^2 + \sqrt{n}(X_2^{(n)} - \sigma^2)} \\
&= -\frac{(\sqrt{n}(X_1^{(n)} - \sigma^2))^2}{O_p(n^{-1/2}) + \sqrt{n}(X_2^{(n)} - \sigma^2)} \\
&\xrightarrow{L} -\frac{Z_1^2}{Z_2}.
\end{aligned}$$

Since the asymptotic distribution of $\hat{\pi}$ does not depend on the unknown parameters, we can use $\hat{\pi}$ to test the hypothesis in (3).

Next, consider the case $\pi = 0$. It can be easily verified that $\hat{\pi}$ defined in (6) converges in probability to zero as n tends to infinity, and that

$$\begin{aligned}
\sqrt{n}\hat{\pi} &= \frac{\sqrt{n}(X_2^{(n)} - \sigma^2)}{(\mu_2 - \mu_1)^2} + o_p(1) \\
&\xrightarrow{L} \frac{Z_2}{(\mu_2 - \mu_1)^2} \sim N\left(0, \frac{2\sigma^4}{(\mu_2 - \mu_1)^4}\right).
\end{aligned}$$

This is a typical consequence of the usual delta theorem. No anomaly occurs because the Fisher information matrix (of (π, μ_2)) is not singular.

In Sections 2 and 3 two different theorems will be given on asymptotic distributions of such statistics. In the both theorems, we assume that each component of $\boldsymbol{f}(\boldsymbol{x})$ is written as $F_i(\boldsymbol{x})/h_i(\boldsymbol{x})$. Theorem 1 requires $F_i(\boldsymbol{x})$ and

$h_i(\boldsymbol{x})$ to have a particular structure, whereas Theorem 2 requires differentiability of $F_i(\boldsymbol{x})$ and $h_i(\boldsymbol{x})$.

Examples on eigenvalues, the problem of choosing the number of components and noniterative statistics in factor analysis are provided to illustrate usefulness of those theorems.

2. THEOREM 1 AND EXAMPLES

The following theorem is obtained.

Theorem 1. Let D be an open domain in R^p, let $\boldsymbol{b} \in D$, and let a_n be a positive sequence which tends to infinity as $n \to \infty$. Let $\boldsymbol{X}_n$ be a sequence of random p-vectors with $P(\boldsymbol{X}_n \in D \setminus \{\boldsymbol{b}\}) = 1$ for each n and let $\boldsymbol{Z}$ be a nondegenerate continuous random p-vector. Suppose that $a_n(\boldsymbol{X}_n - \boldsymbol{b}) \xrightarrow{L} \boldsymbol{Z}$ as $n \to \infty$. Let $\boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), \ldots, f_q(\boldsymbol{x})]'$ be a continuous function from $D \setminus \{\boldsymbol{b}\}$ to R^q. Assume that there exists a continuous function $\boldsymbol{h}(\boldsymbol{x}) = [h_1(\boldsymbol{x}), \ldots, h_q(\boldsymbol{x})]'$ from D to R^q such that for each i $(i = 1, \ldots, p)$, $\lim_{\boldsymbol{x} \to \boldsymbol{b}} f_i(\boldsymbol{x}) h_i(\boldsymbol{x}) = c_i < \infty$ and $h_i(\boldsymbol{X}_n) \neq 0$ with probability one. Define

$$F_i(\boldsymbol{x}) = \begin{cases} f_i(\boldsymbol{x}) h_i(\boldsymbol{x}) & \text{if } \boldsymbol{x} \neq \boldsymbol{b} \\ c_i & \text{if } \boldsymbol{x} = \boldsymbol{b} \end{cases} .$$

Assume further that there exist nonnegative numbers s_i, t_i and continuous functions $\bar{F}_i(\cdot, \cdot)$ and $\bar{h}_i(\cdot, \cdot)$ such that

$$\begin{cases} a^{s_i} F_i(\boldsymbol{x}) &= \bar{F}_i(a(\boldsymbol{x} - \boldsymbol{b}), \boldsymbol{x} - \boldsymbol{b}) \\ a^{t_i} h_i(\boldsymbol{x}) &= \bar{h}_i(a(\boldsymbol{x} - \boldsymbol{b}), \boldsymbol{x} - \boldsymbol{b}) \end{cases} \tag{8}$$

for any $a > 0$ and that $\bar{h}_i(\boldsymbol{Z}, \boldsymbol{0}) \neq 0$ with probability one. Define $G_i(\boldsymbol{Z}) = \dfrac{\bar{F}_i(\boldsymbol{Z}, \boldsymbol{0})}{\bar{h}_i(\boldsymbol{Z}, \boldsymbol{0})}$. Then, we have

$$a_n^{s_i - t_i} f_i(\boldsymbol{X}_n) \xrightarrow{L} G_i(\boldsymbol{Z}) \qquad \text{as} \quad n \to \infty. \tag{9}$$

If $s_i = t_i$ and $G_i(\boldsymbol{Z}) = \beta_i$ a nonstochastic constant with probability one, we then have

$$f_i(\boldsymbol{X}_n) \xrightarrow{P} G_i(\boldsymbol{Z}) = \beta_i \qquad \text{as} \quad n \to \infty. \tag{10}$$

Assume that there exist a nonnegative number u_i and a continuous function $\tilde{F}_i(\cdot,\cdot)$ such that

$$a^{u_i}(F_i(\boldsymbol{x}) - \beta_i h_i(\boldsymbol{x})) = \tilde{F}_i(a(\boldsymbol{x}-\boldsymbol{b}), \boldsymbol{x}-\boldsymbol{b}) \tag{11}$$

for any $a > 0$. Then we have

$$a_n^{u_i - t_i}(f_i(\boldsymbol{X}_n) - \beta_i) \xrightarrow{L} \frac{\tilde{F}_i(\boldsymbol{Z}, \boldsymbol{0})}{\bar{h}_i(\boldsymbol{Z}, \boldsymbol{0})} \qquad \text{as} \quad n \to \infty. \tag{12}$$

Proof. Note that $f_i(\boldsymbol{X}_n) = F_i(\boldsymbol{X}_n)/h_i(\boldsymbol{X}_n)$ with probability one, and evaluate directly the distribution of the statistics as

$$\begin{aligned}
&\Pr\left(\bigcap_{i=1}^{q} a_n^{s_i - t_i} f_i(\boldsymbol{X}_n) < x_i\right) = \Pr\left(\bigcap_{i=1}^{q} \frac{a_n^{s_i} F_i(\boldsymbol{X}_n)}{a_n^{t_i} h_i(\boldsymbol{X}_n)} < x_i\right) \\
&\quad = \Pr\Big(\bigcap_{i=1}^{q} \{a_n^{s_i} F_i(\boldsymbol{X}_n) - x_i a_n^{t_i} h_i(\boldsymbol{X}_n) < 0,\ a_n^{t_i} h_i(\boldsymbol{X}_n) > 0\} \\
&\qquad\qquad \cup \{a_n^{s_i} F_i(\boldsymbol{X}_n) - x_i a_n^{t_i} h_i(\boldsymbol{X}_n) > 0,\ a_n^{t_i} h_i(\boldsymbol{X}_n) < 0\}\Big) \qquad (13) \\
&\quad = \Pr\Big(\bigcap_{i=1}^{q} \{\bar{F}_i(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) - x_i \bar{h}_i(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) < 0,\ \bar{h}_i(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) > 0\} \\
&\qquad\qquad \cup \{\bar{F}_i(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) - x_i \bar{h}_i(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) > 0,\ \bar{h}_i(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) < 0\}\Big) \qquad (14)
\end{aligned}$$

with $\boldsymbol{\xi}_n = a_n(\boldsymbol{X}_n - \boldsymbol{b})$ and $\boldsymbol{\eta}_n = \boldsymbol{X}_n - \boldsymbol{b}$. Assumption in (8) has been used here. Because of the continuity of $\bar{F}_i(\cdot,\cdot)$ and $\bar{h}_i(\cdot,\cdot)$ and the convergence $(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) \xrightarrow{L} (\boldsymbol{Z}, \boldsymbol{0})$, we have

$$\begin{bmatrix} \bar{F}_i(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) \\ \bar{h}_i(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \bar{F}_i(\boldsymbol{Z}, \boldsymbol{0}) \\ \bar{h}_i(\boldsymbol{Z}, \boldsymbol{0}) \end{bmatrix}. \tag{15}$$

Use of (14) and (15) results in

$$\begin{aligned}
&\Pr\left(\bigcap_{i=1}^{q} a_n^{s_i - t_i} f_i(\boldsymbol{X}_n) < x_i\right) \\
&\quad \to \Pr\Big(\bigcap_{i=1}^{q} \{\bar{F}_i(\boldsymbol{Z}, \boldsymbol{0}) - x_i \bar{h}_i(\boldsymbol{Z}, \boldsymbol{0}) < 0,\ \bar{h}_i(\boldsymbol{Z}, \boldsymbol{0}) > 0\} \\
&\qquad\qquad \cup \{\bar{F}_i(\boldsymbol{Z}, \boldsymbol{0}) - x_i \bar{h}_i(\boldsymbol{Z}, \boldsymbol{0}) > 0,\ \bar{h}_i(\boldsymbol{Z}, \boldsymbol{0}) < 0\}\Big) \\
&\quad = \Pr\left(\bigcap_{i=1}^{q} \frac{\bar{F}_i(\boldsymbol{Z}, \boldsymbol{0})}{\bar{h}_i(\boldsymbol{Z}, \boldsymbol{0})} < x_i\right).
\end{aligned}$$

Here we have used $\Pr(\bar{h}_i(\boldsymbol{Z}, \boldsymbol{0}) = 0) = 0$. The result in (12) follows from a similar argument.

Q.E.D.

We shall return to the example of the problem of choosing the number of components described in Section 1. Consider the case $\pi = 1$. The estimator for μ_2 based on the moment method is jointly considered here. Solving equations in (5) for μ_2, we have

$$\hat{\mu}_2 = \frac{(X_1^{(n)} - \mu_1)^2 + \mu_1(X_1^{(n)} - \mu_1) + (X_2^{(n)} - \sigma^2)}{X_1^{(n)} - \mu_1}.$$

Let

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \dfrac{x_2 - \sigma^2}{(x_1 - \mu_1)^2 + (x_2 - \sigma^2)} \\ \dfrac{(x_1 - \mu_1)^2 + \mu_1(x_1 - \mu_1) + (x_2 - \sigma^2)}{x_1 - \mu_1} \end{bmatrix}.$$

We define $h_1(\boldsymbol{x}) = (x_1 - \mu_1)^2 + (x_2 - \sigma^2)$ and $h_2(\boldsymbol{x}) = x_1 - \mu_1$, and then $F_1(\boldsymbol{x}) = x_2 - \sigma^2$ and $F_2(\boldsymbol{x}) = (x_1 - \mu_1)^2 + \mu_1(x_1 - \mu_1) + (x_2 - \sigma^2)$. Recall that $\boldsymbol{\xi}_n = (\xi_1^{(n)}, \xi_2^{(n)})' = (\sqrt{n}(X_1^{(n)} - \mu_1), \sqrt{n}(X_2^{(n)} - \sigma^2))'$ and $\boldsymbol{\eta}_n = (\eta_1^{(n)}, \eta_2^{(n)})' = (X_1^{(n)} - \mu_1, X_2^{(n)} - \sigma^2)'$. Thus, we shall choose $\bar{h}_i$ and $\bar{F}_i$ as below and we have

$$\begin{aligned}
\sqrt{n}h_1(\boldsymbol{x}) &= \xi_1^{(n)}\eta_1^{(n)} + \xi_2^{(n)} = \bar{h}_1(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) \xrightarrow{L} Z_2 \\
\sqrt{n}h_2(\boldsymbol{x}) &= \xi_1^{(n)} = \bar{h}_2(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) \xrightarrow{L} Z_1 \\
\sqrt{n}F_1(\boldsymbol{x}) &= \xi_2^{(n)} = \bar{F}_1(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) \xrightarrow{L} Z_2 \\
\sqrt{n}F_2(\boldsymbol{x}) &= \xi_1^{(n)}\eta_1^{(n)} + \mu_1\xi_1^{(n)} + \xi_2^{(n)} = \bar{F}_2(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n) \xrightarrow{L} \mu_1 Z_1 + Z_2.
\end{aligned}$$

By (9) and (10) we have

$$\begin{aligned}
\hat{\pi} &= f_1(\boldsymbol{X}_n) \xrightarrow{L} \frac{Z_2}{Z_2} = 1 \quad (\text{or} \quad \xrightarrow{P} 1) \\
\hat{\mu}_2 &= f_2(\boldsymbol{X}_n) \xrightarrow{L} \frac{\mu_1 Z_1 + Z_2}{Z_1} = \mu_1 + \frac{Z_2}{Z_1}
\end{aligned}$$

as $n \to \infty$. It follows from (11) and (12) that

$$\sqrt{n}(\hat{\pi} - 1) = -\sqrt{n}\frac{(\boldsymbol{X}_1^{(n)} - \mu_1)^2}{(\boldsymbol{X}_1^{(n)} - \mu_1)^2 + (\boldsymbol{X}_2^{(n)} - \sigma^2)}$$

$$= -\frac{(\xi_1^{(n)})^2}{\xi_1^{(n)}\eta_1^{(n)} + \xi_2^{(n)}} = -\frac{\tilde{F}_1(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n)}{\bar{h}_1(\boldsymbol{\xi}_n, \boldsymbol{\eta}_n)}$$
$$\xrightarrow{L} -\frac{Z_1^2}{Z_2}.$$

In sum, the joint asymptotic distribution of $[\hat{\pi}, \hat{\mu}_2]'$ can be characterized as

$$\begin{bmatrix} \sqrt{n} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\pi} - 1 \\ \hat{\mu}_2 - \mu_1 \end{bmatrix} \xrightarrow{L} \begin{bmatrix} -Z_1^2/Z_2 \\ Z_2/Z_1 \end{bmatrix} \quad \text{as } n \to \infty.$$

Next we shall move to a slightly more general case where μ_1 is unknown in addition to π and μ_2. Assume that μ_1 and μ_2 are distinct. Since three unknown parameters are involved, one more statistic and equation are required in addition to $(X_1^{(n)}, X_2^{(n)})$ in (4) and equations in (5). We shall take as a statistic

$$X_3^{(n)} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^3$$

and the equation is derived as

$$X_3^{(n)} = \pi(1-\pi)(1-2\pi)(\mu_1 - \mu_2)^3. \tag{16}$$

The expectation of the left-hand side in (16) is asymptotically the same as in the right-hand side. When $\pi = 1$, we have

$$\boldsymbol{X}_n = \begin{bmatrix} X_1^{(n)} \\ X_2^{(n)} \\ X_3^{(n)} \end{bmatrix} \xrightarrow{P} \begin{bmatrix} \mathrm{E}[X_1^{(n)}] \\ \mathrm{E}[X_2^{(n)}] \\ \mathrm{E}[X_3^{(n)}] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \sigma^2 \\ 0 \end{bmatrix} = \boldsymbol{b}, \quad \text{say},$$

$$\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{b}) \xrightarrow{L} \boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} \sim N_3\left(\boldsymbol{0}, \begin{bmatrix} \sigma^2 & 0 & 3\sigma^4 \\ 0 & 2\sigma^4 & 0 \\ 3\sigma^4 & 0 & 15\sigma^6 \end{bmatrix}\right).$$

If $\hat{\pi}$ is a solution to the equations in (5) and (16), then $1 - \hat{\pi}$ is also a solution. Existence of two solutions comes from luck of identification of the model (2) when μ_1 and μ_2 are both unknown. In fact, we can not distinguish between $\pi N(\mu_1, \sigma) + (1-\pi)N(\mu_2, \sigma)$ and $(1-\pi)N(\mu_2, \sigma) + \pi N(\mu_1, \sigma)$ (cf. Teicher 1963). To achieve identification we restrict the parameter space of π to be the interval $[\frac{1}{2}, 1]$, and choose a larger solution as an estimate for π.

Here we suppress the superscript (n) for simplicity. Solving (5) and (16) for π, we have

$$\hat{\pi} - 1 = -\frac{4(X_2 - \sigma^2)^3 X_3^2 + 16(X_2 - \sigma^2)^6}{2\left(4(X_2 - \sigma^2)^3 + X_3^2\right)\left(4(X_2 - \sigma^2)^3 + X_3^2 + \sqrt{4(X_2 - \sigma^2)^3 X_3^2 + X_3^4}\right)} \tag{17}$$

and hence

$$\sqrt{n}(\hat{\pi}-1) = -\frac{4\xi_2^3\xi_3^2 + 16\xi_2^5\eta_2}{2\left(4\xi_2^2\eta_2+\xi_3^2\right)\left(4\xi_2^2\eta_2+\xi_3^2+\sqrt{4\xi_2^2\eta_2\xi_3^2+\xi_3^4}\right)} = \frac{\tilde{F}_1(\boldsymbol{\xi},\boldsymbol{\eta})}{\bar{h}_1(\boldsymbol{\xi},\boldsymbol{\eta})}, \quad \text{say,}$$

with $\xi_i = \sqrt{n}(X_i - b_i)$ and $\eta_i = X_i - b_i$. Using Theorem 1, this quantity converges in distribution to

$$\frac{\tilde{F}_1(\boldsymbol{Z},\boldsymbol{0})}{\bar{h}_1(\boldsymbol{Z},\boldsymbol{0})} = -\frac{Z_2^3}{Z_3^2}.$$

The asymptotic distribution is independent of unknown parameters, and hence one can use it to test the hypothesis in (3).

A similar derivation leads to asymptotic distributions of $\hat{\mu}_1$ and $\hat{\mu}_2$. We report here the final result:

$$\begin{bmatrix} \sqrt{n} & 0 & 0 \\ 0 & \sqrt{n} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\pi}-1 \\ \hat{\mu}_1-\mu_1 \\ \hat{\mu}_2-\mu_1 \end{bmatrix} \xrightarrow{L} \begin{bmatrix} -Z_2^3/Z_3^2 \\ Z_1+Z_2^2/Z_3 \\ Z_3/Z_2 \end{bmatrix} \quad \text{as } n\to\infty.$$

It would be interesting to note that when n goes to infinity, the estimator for the indeterminate parameter μ_2 does not concentrate at any specific value and converges in law to the *Cauchy distribution* whether for the two-parameter case or for the three-parameter case.

It would be Day (1969) who first described the moment method in the normal mixture. He did not recommend its use for the reason of efficiency in estimation, however. We suppose that he did not realize the difficulties of the MLE and such beautiful results of the moment method as consistency and easy evaluation of the asymptotic distribution.

We have not taken the parameter space into account when we construct $\hat{\pi}$. The estimator can be outside $[0,1]$ or $[\frac{1}{2},1]$ in the two examples above. It is possible to form modified estimators $\hat{\pi}^*$ to be in these parameter spaces. Testing procedures based on $\hat{\pi}^*$ and $\hat{\pi}$ are equivalent to each other whereas these estimators have different asymptotic distributions. See Hatsuyama (1995) for details.

Under alternative distributions, $\hat{\pi}$ is asymptotically normal, and use of the asymptotic distribution enables us to evaluate statistical power of the test based on $\hat{\pi}$. See again Hatsuyama (1995).

To end the example, we note alternative approaches to the problem of choosing the number of components: Henna (1985)'s nonparametric method, penalty method by Leroux (1992) and Chen and Kalbfleisch (1993), Bayes approach with MCMC due to Richardson and Green (in press) and Roeder and Wasserman (1997).

We shall finalize this section with the eigenvalue problem noted in Section 1. The problem can be dealt with via Theorem 1. Assume that

$$\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{b}) = \sqrt{n}\left(\begin{bmatrix} s_{11} \\ s_{21} \\ s_{22} \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right) \xrightarrow{L} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} \sim N_3(\mathbf{0}, \Gamma),$$

where Γ is the asymptotic covariance matrix. When nS follows the Wishart distribution $W_2(n, I_2)$, $\Gamma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. We do not have to restrict the distribution to the Wishart one, however.

We take $h_i(\cdot) = 1$ because the eigenvalues are defined at $S = I_2$ (or $\boldsymbol{X}_n = \boldsymbol{b}$). It should be noted that $\sqrt{n}(\hat{\lambda}_i - 1)$ can be expressed exactly as functions of $\sqrt{n}(S - I_2)$ (or of $\boldsymbol{\xi}_n = \sqrt{n}(\boldsymbol{X}_n - \boldsymbol{b})$). In fact,

$$\begin{aligned} &\frac{1}{2}\left(\sqrt{n}(s_{11}-1) + \sqrt{n}(s_{22}-1) \pm \sqrt{\{\sqrt{n}(s_{11}-1) - \sqrt{n}(s_{22}-1)\}^2 + 4(\sqrt{n}s_{12})^2}\right) \\ &\quad = \bar{F}(\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{b}), \mathbf{0}), \end{aligned}$$

which converges in distribution to

$$\bar{F}(\boldsymbol{Z}, \mathbf{0}) = \frac{1}{2}(Z_1 + Z_3) \pm \sqrt{(Z_1 - Z_3)^2 + 4Z_2^2}.$$

A more general case can be discussed where S is a $p \times p$ symmetric matrix and $a_n(S - I_p) \xrightarrow{L} Z$, a random matrix of order p. Let $\hat{\lambda}_i(S)$ be the i-th eigenvalue of S, and then

$$a_n(\hat{\lambda}_i(S) - 1) = \hat{\lambda}_i\{a_n(S - I_p)\} \xrightarrow{L} \hat{\lambda}_i(Z)$$

because of the continuity of eigenvalues. The essence of the derivation is to express the eigenvalues as functions of $a_n(S - I_p)$ (or of $\boldsymbol{\xi}_n$) as in the assumption of Theorem 1. Eaton and Tyler (1991) used this idea to derive the asymptotic distribution of eigenvalues in a more general setting.

3. THEOREM 2 AND EXAMPLES

There are some cases where the assumption on $F_i(\boldsymbol{x})$ and $h_i(\boldsymbol{x})$ in (8) are not met. Such an example follows Theorem 2 below. In this section, we shall assume differentiability of $F_i(\boldsymbol{x})$ and $h_i(\boldsymbol{x})$ rather than the assumption in (8). Obviously, the estimators in (1) and (17) are not differentiable at $\boldsymbol{X}_n = \boldsymbol{b}$ for example, so one theorem does not include the other.

For the sake of notational simplicity, we shall write

$$A^{<s>} = \overbrace{A \otimes \cdots \otimes A}^{s \text{ times}} \quad \text{and} \quad d^s F(\boldsymbol{b}) = \overbrace{\left(\frac{\partial}{\partial \boldsymbol{x}'}\right) \otimes \cdots \otimes \left(\frac{\partial}{\partial \boldsymbol{x}'}\right)}^{s \text{ times}} F(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{b}}$$

for a differentiable function $F(\cdot)$. Here the symbol $\otimes$ denotes the (right) Kronecker product. The formal Taylor expansion of $F(\boldsymbol{x})$ about $\boldsymbol{x} = \boldsymbol{b}$ is then expressible in the form:

$$F(\boldsymbol{x}) = \sum_{s=0}^{\infty} \frac{d^s F(\boldsymbol{b})}{s!} (\boldsymbol{x} - \boldsymbol{b})^{<s>} \quad \left(= \sum_{s=0}^{\infty} \frac{1}{s!} \left[\sum_{j=1}^{p} (x_j - b_j) \frac{\partial}{\partial x_j} \right]^s F(\boldsymbol{b}) \right).$$

The following theorem is obtained.

Theorem 2. Let D be an open domain in R^p, let $\boldsymbol{b} \in D$, and let a_n be a positive sequence which tends to infinity as $n \to \infty$. Let $\boldsymbol{X}_n$ be a sequence of random p-vectors with $P(\boldsymbol{X}_n \in D \setminus \{\boldsymbol{b}\}) = 1$ for each n and let $\boldsymbol{Z}$ be a nondegenerate continuous random p-vector. Suppose that $a_n(\boldsymbol{X}_n - \boldsymbol{b}) \xrightarrow{L} \boldsymbol{Z}$ as $n \to \infty$. Let $\boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), \ldots, f_q(\boldsymbol{x})]'$ be a measurable function from $D \setminus \{\boldsymbol{b}\}$ to R^q. Assume that there exists a measurable function $\boldsymbol{h}(\boldsymbol{x}) = [h_1(\boldsymbol{x}), \ldots, h_q(\boldsymbol{x})]'$ from D to R^q such that for each i $(i = 1, \ldots, p)$,

$$\lim_{\boldsymbol{x} \to \boldsymbol{b}} f_i(\boldsymbol{x}) h_i(\boldsymbol{x}) = c_i < \infty. \tag{18}$$

Define

$$F_i(\boldsymbol{x}) = \begin{cases} f_i(\boldsymbol{x}) h_i(\boldsymbol{x}) & \text{if } \boldsymbol{x} \neq \boldsymbol{b} \\ c_i & \text{if } \boldsymbol{x} = \boldsymbol{b} \end{cases}.$$

Assume that $F_i(\boldsymbol{x})$ and $h_i(\boldsymbol{x})$ are differentiable around $\boldsymbol{x} = \boldsymbol{b}$.

Let s_i and t_i be minimum nonnegative integers such that $d^{s_i}F_i(\boldsymbol{b}) \neq \mathbf{0}$ and $d^{t_i}h_i(\boldsymbol{b}) \neq \mathbf{0}$, respectively. Define $G_i(\boldsymbol{Z}) = \dfrac{d^{s_i}F_i(\boldsymbol{b})\boldsymbol{Z}^{<s_i>}/s_i!}{d^{t_i}h_i(\boldsymbol{b})\boldsymbol{Z}^{<t_i>}/t_i!}$. Then, we have

$$a_n^{s_i-t_i} f_i(\boldsymbol{X}_n) \xrightarrow{L} G_i(\boldsymbol{Z}) \qquad \text{as} \quad n \to \infty. \tag{19}$$

Assume further that $s_i = t_i$ and that $d^{t_i}F_i(\boldsymbol{b}) = \beta_i d^{t_i}h_i(\boldsymbol{b})$ for some $\beta_i \in R^1$. Then,

$$f_i(\boldsymbol{X}_n) \xrightarrow{P} G_i(\boldsymbol{Z}) = \beta_i \qquad \text{as} \quad n \to \infty. \tag{20}$$

Let u_i be a minimum nonnegative integer such that $d^{u_i}F_i(\boldsymbol{b}) - \beta_i d^{u_i}h_i(\boldsymbol{b}) \neq \mathbf{0}$. Then we have

$$a_n^{u_i-t_i}(f_i(\boldsymbol{X}_n) - \beta_i) \xrightarrow{L} \frac{(d^{u_i}F_i(\boldsymbol{b}) - \beta_i d^{u_i}h_i(\boldsymbol{b}))\boldsymbol{Z}^{<u_i>}/u_i!}{d^{t_i}h_i(\boldsymbol{b})\boldsymbol{Z}^{<t_i>}/t_i!}. \tag{21}$$

Note 1: Orders of differentiability required are not specified in the theorem. Stating strictly, differentiability of orders lower than or equal to the order where a nonzero derivative appears at first is required for this theorem.

Note 2: When the limit c_i in (18) is nonzero, then $s_i = 0$.

Note 3: The scalar β_i is nonzero since $d^{t_i}F_i(\boldsymbol{b}) \neq \mathbf{0}$ and $d^{t_i}h_i(\boldsymbol{b}) \neq \mathbf{0}$ by definition.

Note 4: The scalar u_i is greater than t_i since $d^{t_i}F_i(\boldsymbol{b}) - \beta_i d^{t_i}h_i(\boldsymbol{b}) = \mathbf{0}$ by assumption.

Proof. The proof is similar to that of Theorem 1. It follows from the differentiability that

$$\begin{aligned}
\begin{bmatrix} a_n^{s_i}F_i(\boldsymbol{X}_n) \\ a_n^{t_i}h_i(\boldsymbol{X}_n) \end{bmatrix} &= \begin{bmatrix} d^{s_i}F_i(\boldsymbol{b})\{a_n(\boldsymbol{X}_n - \boldsymbol{b})\}^{<s_i>}/s_i! \\ d^{t_i}h_i(\boldsymbol{b})\{a_n(\boldsymbol{X}_n - \boldsymbol{b})\}^{<t_i>}/t_i! \end{bmatrix} + o_p(1) \\
&\xrightarrow{L} \begin{bmatrix} d^{s_i}F_i(\boldsymbol{b})\boldsymbol{Z}^{<s_i>}/s_i! \\ d^{t_i}h_i(\boldsymbol{b})\boldsymbol{Z}^{<t_i>}/t_i! \end{bmatrix}.
\end{aligned} \tag{22}$$

Use of (13) and (22) results in

$$\begin{aligned}
&\Pr\left(\bigcap_{i=1}^{q} a_n^{s_i-t_i} f_i(\boldsymbol{X}_n) < x_i\right) \\
&\quad \to \Pr\Bigg(\bigcap_{i=1}^{q} \Big\{ d^{s_i} F_i(\boldsymbol{b})\boldsymbol{Z}^{<s_i>}/s_i! - x_i d^{<t_i>} h_i(\boldsymbol{b})\boldsymbol{Z}^{<t_i>}/t_i! < 0, \\
&\qquad\qquad d^{<t_i>} h_i(\boldsymbol{b})\boldsymbol{Z}^{<t_i>}/t_i! > 0 \Big\} \\
&\qquad \cup \Big\{ d^{s_i} F_i(\boldsymbol{b})\boldsymbol{Z}^{<s_i>}/s_i! - x_i d^{t_i} h_i(\boldsymbol{b})\boldsymbol{Z}^{t_i}/t_i! > 0, \\
&\qquad\qquad d^{t_i} h_i(\boldsymbol{b})\boldsymbol{Z}^{<t_i>}/t_i! < 0 \Big\} \Bigg) \\
&\quad = \Pr\left(\bigcap_{i=1}^{q} \frac{d^{s_i} F_i(\boldsymbol{b})\boldsymbol{Z}^{<s_i>}/s_i!}{d^{t_i} h_i(\boldsymbol{b})\boldsymbol{Z}^{<t_i>}/t_i!} < x_i \right).
\end{aligned}$$

Note that $\Pr(d^{t_i} h_i(\boldsymbol{b})\boldsymbol{Z}^{<t_i>} = 0) = 0$. The result in (21) follows from a similar argument.

Q.E.D.

Let us return to the example of choosing of the number of components described in Section 1. Here we consider the case $\pi = 1$ as before. We can easily verify that

$$\begin{array}{llll}
d^1 h_1(\boldsymbol{b}) = [0,1] & \text{and} & d^1 F_1(\boldsymbol{b}) = [0,1], & \text{so that} \quad s_1 = t_1 = 1, \\
d^1 h_2(\boldsymbol{b}) = [1,0] & \text{and} & d^1 F_2(\boldsymbol{b}) = [\mu_1,1], & \text{so that} \quad s_2 = t_2 = 1.
\end{array}$$

Since $d^1 h_1(\boldsymbol{b}) = d^1 F_1(\boldsymbol{b})$ and so $\beta_1 = 1$, application of (20) leads to

$$\hat{\pi} = f_1(\boldsymbol{X}_n) \xrightarrow{P} 1.$$

From (19), we have

$$\hat{\mu}_2 = f_2(\boldsymbol{X}_n) \xrightarrow{L} \frac{d^1 F_2(\boldsymbol{b})\boldsymbol{Z}}{d^1 h_2(\boldsymbol{b})\boldsymbol{Z}} = \mu_1 + \frac{Z_2}{Z_1},$$

which is exacly the cauchy distribution. Next, the asymptotic distribution of $f_1(\boldsymbol{X}_n)$ will be evaluated. Since $F_1(\boldsymbol{x}) - \beta_1 h_1(\boldsymbol{x}) = -(x_1 - \mu_1)^2$, $d^2 F_1(\boldsymbol{b}) - \beta_1 d^2 h_1(\boldsymbol{b}) = -[2,0,0,0] \neq \boldsymbol{0}$, and hence $u_1 = 2$. Application of (21) leads to

$$\sqrt{n}(\hat{\pi} - 1) = (\sqrt{n})^{u_1 - s_1}(f_1(\boldsymbol{X}_n) - \beta_1) \xrightarrow{L} \frac{(d^2 F_1(\boldsymbol{b}) - \beta_1 d^2 h_1(\boldsymbol{b}))\boldsymbol{Z}/2}{d^1 h_1(\boldsymbol{b})\boldsymbol{Z}} = -\frac{Z_1^2}{Z_2}.$$

A more interesting example is given where Theorem 1 can not apply but Theorem 2 can do. The (exploratory) factor analysis model can be characterized as a covariance structure of an observed random vector $\boldsymbol{Y}$. The covariance structure of the factor analysis model is in the form:

$$\mathrm{Var}(\boldsymbol{Y}) = \Sigma = \Lambda\Lambda' + \Psi, \tag{23}$$

where $\Lambda = (\lambda_{ij})$ is a factor loading matrix of $p \times k$ and $\Psi = \mathrm{diag}(\psi_1, \ldots, \psi_p)$ is a diagonal matrix of order p representing unique variances. The k is said to be the number of factors. The parameters (Λ, Ψ) are unknown and estimated from a sample variance matrix S, based on a sample of size n. Assume that the S has a continuous distribution and that

$$a_n(\boldsymbol{X}_n - \boldsymbol{b}) = a_n(v(S) - v(\Sigma)) \xrightarrow{L} \boldsymbol{Z},$$

where $a_n \nearrow \infty$ and $\boldsymbol{Z}$ is a random vector continuously distributed. The so-called vec-operator $v(S)$ for a symmetric matrix S of order p denotes a $p(p+1)/2$-vector of the elements in the lower triangular part of S (see e.g., Magnus and Neudecker 1988). In practice, the number k of factors is unknown. Let m denote the number of assumed (or extracted) factors. In many cases, m may not be equal to k. When the number of factors is underestimated, i.e. $m < k$, the estimators are biased and the fit of the model will be poor. When it is overestimated, i.e. $m > k$, estimation encounters serious difficulties, such as nonconvergence in iteration. The anomaly occurs because the model is not identifiable. Take $m = k+1$ for example. There are infinitely many parameters for such models when the k-factor model holds as in (23). In fact, it holds that

$$\Sigma = \begin{bmatrix} & c \\ \Lambda & 0 \\ & \vdots \\ & 0 \end{bmatrix} \begin{bmatrix} & c \\ \Lambda & 0 \\ & \vdots \\ & 0 \end{bmatrix}' + \begin{bmatrix} \psi_1 - c^2 & & & \\ & \psi_2 & & \\ & & \ddots & \\ & & & \psi_p \end{bmatrix} \tag{24}$$

for any number c. The direct noniterative estimator of Ihara and Kano (1986) and Kano (1990b) overcomes the identification problem. Assume that $p \geq 2m+1$, and partition the p variables into four groups as Y_1, Y_2-Y_{m+1}, Y_{m+2}-Y_{2m+1} and Y_{2m+2}-Y_p, where the last group can be empty. In this paper, we assume that $m = k+1$ (so that $p \geq 2k+3$), and partition further each of the

two groups of m variables into two subgroups of k variables and one variable. The corresponding partition of a factor loading matrix is made as

$$\Lambda = \begin{bmatrix} \boldsymbol{\lambda}_1 \\ \Lambda_2 \\ \boldsymbol{\lambda}_3 \\ \Lambda_4 \\ \boldsymbol{\lambda}_5 \\ \Lambda_6 \end{bmatrix} \begin{matrix} \} \\ \} \\ \} \\ \} \\ \} \\ \} \end{matrix} \begin{matrix} 1 \\ k \\ 1 \\ k \\ 1 \\ p-(2k+3) \end{matrix} .$$

Assume that the $k \times k$ matrices Λ_2 and Λ_4 are nonsingular. It is a famous sufficient condition for identification for the k-factor model (Anderson and Rubin 1956). Partition Σ, Ψ and S correspondingly. The noniterative estimator for ψ_1 is expressed as

$$\hat{\psi}_1^{(k+1)} = s_{11} - [\boldsymbol{s}_{12}, s_{13}] \begin{bmatrix} S_{42} & \boldsymbol{s}_{43} \\ \boldsymbol{s}_{52} & s_{53} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{s}_{41} \\ s_{51} \end{bmatrix}.$$

The inverse matrix involving the estimator above can be taken with probability one because the distribution of S is continuous. Since

$$\begin{bmatrix} S_{42} & \boldsymbol{s}_{43} \\ \boldsymbol{s}_{52} & s_{53} \end{bmatrix} \xrightarrow{P} \begin{bmatrix} \Sigma_{42} & \boldsymbol{\sigma}_{43} \\ \boldsymbol{\sigma}_{52} & \sigma_{53} \end{bmatrix} = \begin{bmatrix} \Lambda_4 \\ \boldsymbol{\lambda}_5 \end{bmatrix} \begin{bmatrix} \Lambda_2 \\ \boldsymbol{\lambda}_3 \end{bmatrix}',$$

the limiting matrix is singular (its rank is k whereas the order is $k+1$), so that the estimator $\hat{\psi}_1^{(k+1)}$ can not be defined at $S = \Sigma$. Does the estimator have a limiting point? If so, is the limiting point the true unique variance ψ_1? What kind of asymptotic distribution does it have? Theorem 2 answers these questions.

By use of a formula of the inversion of a partitioned matrix (see e.g., Anderson 1984, Theorem A.3.3), we have

$$\begin{aligned} &\hat{\psi}_1^{(k+1)} \\ &\quad = \frac{(s_{11} - \boldsymbol{s}_{12} S_{42}^{-1} \boldsymbol{s}_{41})(s_{53} - \boldsymbol{s}_{52} S_{42}^{-1} \boldsymbol{s}_{43}) - (s_{13} - \boldsymbol{s}_{12} S_{42}^{-1} \boldsymbol{s}_{43})(s_{51} - \boldsymbol{s}_{52} S_{42}^{-1} \boldsymbol{s}_{41})}{s_{53} - \boldsymbol{s}_{52} S_{42}^{-1} \boldsymbol{s}_{43}} \\ &\quad = \frac{F(\boldsymbol{X}_n)}{h(\boldsymbol{X}_n)}, \quad \text{say.} \end{aligned}$$

It should be noted that $F(\boldsymbol{x})$ and $h(\boldsymbol{x})$ do not satisfy the assumption (8) in Theorem 1, so we can not appy Theorem 1 to this problem.

When $S = \Sigma$ (or $\boldsymbol{X}_n = \boldsymbol{b}$), we have

$$\begin{aligned} s_{11} - \boldsymbol{s}_{12} S_{42}^{-1} \boldsymbol{s}_{41} &= \boldsymbol{\lambda}_1 \boldsymbol{\lambda}_1' + \psi_1 - \boldsymbol{\lambda}_1 \Lambda_2' (\Lambda_4 \Lambda_2')^{-1} \Lambda_4 \boldsymbol{\lambda}_1' = \psi_1 \\ s_{ij} - \boldsymbol{s}_{i2} S_{42}^{-1} \boldsymbol{s}_{4j} &= \boldsymbol{\lambda}_i \boldsymbol{\lambda}_j' - \boldsymbol{\lambda}_i \Lambda_2' (\Lambda_4 \Lambda_2')^{-1} \Lambda_4 \boldsymbol{\lambda}_j' = 0 \\ & \qquad\qquad (\text{for } i, j = 1, 3, 5 \text{ and } i \neq j). \end{aligned}$$

in view of the fact that Λ_2 and Λ_4 are nonsingular matrices. It follows that $d^1 h(\boldsymbol{b}) \neq \boldsymbol{0}$ and $d^1 F(\boldsymbol{b}) = \psi_1 d^1 h(\boldsymbol{b})$, which implies from (20) in Theorem 2 that

$$\hat{\psi}_1^{(k+1)} \xrightarrow{P} \frac{d^1 F(\boldsymbol{b}) \boldsymbol{Z}}{d^1 h(\boldsymbol{b}) \boldsymbol{Z}} = \psi_1.$$

This shows consistency of the noniterative estimator for the case where $m = k + 1$, the number of factors is overestimated.

By usual delta theorem, we have

$$\begin{aligned} a_n (s_{11} - \boldsymbol{s}_{12} S_{42}^{-1} \boldsymbol{s}_{41} - \psi_1) &\xrightarrow{L} \boldsymbol{d}_{11} \boldsymbol{Z} \\ a_n (s_{ij} - \boldsymbol{s}_{i2} S_{42}^{-1} \boldsymbol{s}_{4j}) &\xrightarrow{L} \boldsymbol{d}_{ij} \boldsymbol{Z} \end{aligned}$$

where $\boldsymbol{d}_{ij} = \left. \dfrac{\partial (s_{ij} - \boldsymbol{s}_{i2} S_{42}^{-1} \boldsymbol{s}_{4j})}{\partial v(S)'} \right|_{S=\Sigma}$. The asymptotic distribution of $\hat{\psi}_1^{(k+1)}$ is obtained from (21) as

$$\begin{aligned} a_n (\hat{\psi}_1^{(k+1)} - \psi_1) &= \frac{a_n^2 (F(\boldsymbol{X}_n) - \psi_1 h(\boldsymbol{X}_n))}{a_n h(\boldsymbol{X}_n)} \\ &\xrightarrow{L} \frac{(d^2 F(\boldsymbol{b}) - \psi_1 d^2 h(\boldsymbol{b})) \boldsymbol{Z}^{<2>} / 2}{d^1 h(\boldsymbol{b}) \boldsymbol{Z}} \\ &= \frac{(\boldsymbol{d}_{11} \boldsymbol{Z})(\boldsymbol{d}_{53} \boldsymbol{Z}) - (\boldsymbol{d}_{13} \boldsymbol{Z})(\boldsymbol{d}_{51} \boldsymbol{Z})}{\boldsymbol{d}_{53} \boldsymbol{Z}} \\ &= \boldsymbol{d}_{11} \boldsymbol{Z} - \frac{(\boldsymbol{d}_{13} \boldsymbol{Z})(\boldsymbol{d}_{51} \boldsymbol{Z})}{\boldsymbol{d}_{53} \boldsymbol{Z}}. \end{aligned}$$

Equivalents to the derivatives $\boldsymbol{d}_{ij}$'s were given in Kano (1990a).

We can take the same derivation to find the asymptotic distribution of $\hat{\psi}_1^{(m)}$ for a general $m (\geq k + 1)$ as well.

Kano (1991) assumed normality of observations to derive the asymptotic distribution of $\hat{\psi}_1^{(m)}$ for an overestimated number of factors. The proof depends heavily on the Bartlett decomposition of the Wishart matrix nS. In this paper, we have taken a completely different derivation to obtain the asymptotic distribution under general distribution assumption.

Finally we would like to make a comment on the examples employed in this paper. The examples of selecting the number of components and factor analysis seem to differ instinctively. Identification is not achieved in both of the examples, that is, μ_2 and ψ_1 are indeterminate. It appears natural to us that $\hat{\mu}_2$ does have a distribution even in the limit because of the luck of identification. How should we interpret the fact that $\hat{\psi}_1$ is *consistent* ? Although there are many parameter values that perfectly reproduce the same population covariance matrix Σ in (23), the most *reasonable* value is unique, and it is the value when $c = 0$ in (24) because the factor corresponding to the last column of the factor loading matrix in (24) is not a common factor but a unique factor. The noniterative estimator approaches the most reasonable value as the sample size goes to infinity. In contrast, usual estimators including the MLE in factor analysis do not have this property.

Inconsistency of estimators for indeterminate parameters could not be regarded as a result of luck of identification only, and alternative factors may have to be introduced to account for it. We can not yet find any solution to the problem as to when estimators are consistent or inconsistent for indeterminate parameters.

ACKNOWLEDGEMENTS

The author expresses his thanks to Professor Masanobu Taniguchi of Osaka University for his valuable comments on this work. He also thanks reviewers for their comments that help improve the original manuscript greatly.

BIBLIOGRAPHY

Akahira, M. and Takeuchi, K. (1995). *Non-regular Statistical Estimation*. Lecture Notes in Statistics, Vol.107. Springer-Verlag: New York.

Anderson, T. W. (1963). Asymptotic theory of principal component analysis. *Ann. Math. Statist.*, **34**, 122-148.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley: New York.

Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. *Proc. 3rd Berkeley Symp. Math. Statist. Probab.*, **5**, 111-150. University of California Press: Berkeley.

Basawa, I. V. and Scott, D. J. (1983). *Asymptotic Optimal Inference for Non-ergodic Models.* Lecture Notes in Statistics, Vol.17. Springer-Verlag: New York.

Chen, J. and Kalbfleisch, J. D. (1993). Penalized minimum distance estimates in finite mixture models. *Technical Report Series* **94-08**, University of Waterloo.

Chikuse, Y. (1976). Asymptotic distributions of the latent roots of the covariance matrix with multiple population roots. *Jour. Multi. Anal.*, **6**, 237-249.

Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika*, **56**, 463-474.

Eaton, M. L. and Tyler, D. E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix *Ann. Statist.*, **19**, 260-271.

Eaton, M. L. and Tyler, D. E. (1994). The asymptotic distribution of singular values with applications to canonical correlations and corresponding analysis. *Jour. Multi. Anal.*, **50**, 238-264.

Fujikoshi, Y. (1977). An asymptotic expansion of the distributions of the latent roots of the Wishart matrix with multiple population roots. *Ann. Inst. Math. Statist.*, **29A**, 379-387.

Fujikoshi, Y. (1980). Asymptotic expansions of the distributions of the sample roots under nonnormality. *Biometrika*, **67**, 45-51.

Hartigan, J. A. (1985). A failure of likelihood asymptotics for normal mixture. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. 2. edited by L. LeCam and R. A. Olshen, pp. 807-810.

Hatsuyama, M. (1995). Selection of the number of components in a finite mixture distribution. Master's Thesis (In Japanese). Osaka Prefecture University: Sakai, Osaka, Japan.

Henna, J.(1985). On estimating of the number of constituents of a finite mixture of continuous distributions. *Ann. Inst. Statist. Math.*, **37**, 235-240.

Ihara, M. and Kano, Y. (1986). A new estimator of the uniqueness in factor analysis. *Psychometrika*, **51**, 563-566.

Kano, Y. (1990a). Comparative studies of non-iterative estimators based on Ihara and Kano's method in exploratory factor analysis. *Commun. Statist. Theo. Meth.*, **19**, 431-444.

Kano, Y. (1990b). Noniterative estimation and the choice of the number of factors in exploratory factor analysis. *Psychometrika*, **55**, 277-291.

Kano, Y. (1991). The asymptotic distribution of a noniterative estimator in exploratory factor analysis. *Ann. Statist.*, **19**, 272-282.

Leroux, B. (1992). Consistent estimation of a mixture distribution. *Ann. Statist.*, **20**, 1350-1360.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley: Chichester.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications,* 2nd ed. Wiley: New York.

Richardson, S. and Green P. J. (in press). On Baysian analysis of mixtures with an unknown number of components. *Jour. Royal Statist. Soc., Series B.*

Roeder K. and Wasserman, L. (1997). Practical Baysian density estimation using mixture of normals. *Jour. Amer. Statist. Assoc.*, **92**, 894-902.

Romano, J. P. and Siegel, A. F. (1986). *Counterexamples in Probability and Statistics.* Wadsdorth & Brooks/Cole Advanced Books & Software: Belmont, CA.

Roussas, G. G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics.* Cambridge At The University Press: Cambridge.

Sugiura, N. (1973). Derivatives of the characteristic root of a symmetric or Hermitian matrix with two applications in multivariate analysis. *Commun. Statist.*, **1**, 393-417.

Sugiura, N. (1976). Asymptotic expansions of the distributions of the latent roots and latent vector of the Wishart and multivariate F matrices. *Jour. Multi. Anal.*, **6**, 500-525.

Sugiura, N. (1990). Graphs of the distribution of bivariate Wishart roots and their cumulants. *Jour. Japan Statist. Soc.*, **20**, 117-136.

Tanaka, K. (1996). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory.* John Wiley & Sons, Inc.: New York.

Teicher, H. (1963). Identifiability of finite mixtures. *Ann. Math. Statist.*, **34**, 1265-1269.

Tyler, D. E. (1983). The asymptotic distribution of principal component roots under local alternatives to multiple roots. *Ann. Statist.*, **11**, 1232-1242.

Waternaux, C. M. (1976). Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika*, **63**, 639-645.