J. Japan Statist. Soc.
Vol. 31 No. 2 2001 187–191

# MARGINAL DISTRIBUTIONS OF FINITE MIXTURES OF MULTIVARIATE NORMAL DISTRIBUTIONS

Jogi Henna*

We give a transformation such that, for a $k$-dimensional finite normal mixture, the number of components and the mixing ratios are preserved on each marginal density through the transformation. Furthermore, under the only assumption of a $k$-dimensional finite normal mixture, we construct a one-dimensional random variable with a finite normal mixture of the true number of components and the true mixing ratios.

*Key words and phrases*: Finite mixture, $k$-dimensional distribution, One-dimensional distribution, Orthogonal matrix.

## 1. Introduction

Let $n(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a $k$-dimensional normal probability density function (pdf) with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. The pdf

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \alpha_i n(\boldsymbol{x} \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \tag{1.1}$$

will be called a finite mixture of $n(\boldsymbol{x} \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ $(i = 1, 2, \ldots, m)$, where $0 < \alpha_i < 1$ and $\sum_{i=1}^{m} \alpha_i = 1$. The usefulness of the mixture model is explained in McLachlan and Basford (1988) and Titterington *et al.* (1985). Even if $\boldsymbol{X}$ has a finite mixture (1.1), the marginal pdf of the $j$-th coordinate $X_j$ of $\boldsymbol{X}$ is not necessarily a finite mixture of $m$ pdf's. A simple example is given as follows:

$$f(\boldsymbol{x}) = \alpha_1 n(\boldsymbol{x} \mid \boldsymbol{\mu}_1, I) + \alpha_2 n(\boldsymbol{x} \mid \boldsymbol{\mu}_2, I) + \alpha_3 n(\boldsymbol{x} \mid \boldsymbol{\mu}_3, I), \tag{1.2}$$

which is a finite mixture of three pdf's, where $\boldsymbol{\mu}_1 = (0,0)'$, $\boldsymbol{\mu}_2 = (0,1)'$, $\boldsymbol{\mu}_3 = (1,1)'$ and $I$ is the $2 \times 2$ identity matrix. Then the pdf of $X_1$ is given as follows:

$$g_1(x_1) = (\alpha_1 + \alpha_2) n(x_1 \mid 0, 1) + \alpha_3 n(x_1 \mid 1, 1), \tag{1.3}$$

which is a finite mixture of two one-dimensional normal pdf's. This can be similarly said for $X_2$.

Now let us consider transformation $\boldsymbol{Y} = \boldsymbol{M}\boldsymbol{X} + \boldsymbol{\rho}$, where $\boldsymbol{M}$ is a $k \times k$ orthogonal matrix and $\boldsymbol{\rho}$ a vector. From a property of normal pdf (Anderson (1984)), it can be seen that a necessary and sufficient condition for $\boldsymbol{X}$ to have a finite mixture (1.1) is for $\boldsymbol{Y}$ to have a finite mixture

$$g(\boldsymbol{y}) = \sum_{i=1}^{m} \alpha_i n(\boldsymbol{y} \mid \boldsymbol{M}\boldsymbol{\mu}_i + \boldsymbol{\rho}, \boldsymbol{M}\boldsymbol{\Sigma}_i\boldsymbol{M}'). \tag{1.4}$$

Received August 8, 2000. Revised February 13, 2001. Accepted June 14, 2001.
*Faculty of Science, University of the Ryukyus, 1 Senbaru, Nishihara-cho, Okinawa 903-0213, Japan.

Taking into account of the above example, it is valuable to find an $\boldsymbol{M}$ such that the marginal pdf of the $j$-th coordinate $Y_j$ of $\boldsymbol{Y}$ is a finite mixture of $m$ one-dimensional normal pdf's. The purpose of this paper is to find the orthogonal matrix $\boldsymbol{M}$ and then obtain $\boldsymbol{Y}$. For the purpose, we give a statement which can be obtained easily from the theory of linear algebra (Gantmacher (1977)).

STATEMENT 1.1. *Let $\mu_{j1}$ and $\mu_{j2}$ be the $j$-th coordinates, referred to the orthogonal coordinate system $(x_1, x_2, \ldots, x_k)$, of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively. Then $\mu_{j1} = \mu_{j2}$ is equivalent to the fact that the $j$-th coordinate axis lies on the $(k-1)$-dimensional hyperplane $\pi$ going through the origin and perpendicular to the straight line going through the points $(\mu_{11}, \mu_{21}, \ldots, \mu_{k1})$ and $(\mu_{12}, \mu_{22}, \ldots, \mu_{k2})$.*

In Section 2, we construct $\boldsymbol{Y}$ such that, when $\boldsymbol{X}$ has the pdf (1.1) with known mean vectors, the number of components and the mixing ratios are preserved on each marginal density through the transformation. In Section 3, when $\boldsymbol{X}$ has the pdf (1.1) with unknown mean vectors, we construct a one-dimensional random variable $Y_j^{(\gamma)}$ with a finite normal mixture of the true number of components and the true mixing ratios, and then give an example to construct such random variable.

## 2. Construction of $\boldsymbol{Y}$ when $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_m$ are known

We first need to obtain an $\boldsymbol{M}$ to give a transformation which satisfies our demand under the assumption that the mean vectors of component pdf's of (1.1) are known. For this purpose, we give the following lemma.

LEMMA 2.1. *For any different $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_m$ and any $\boldsymbol{\rho}$, there exists an $\boldsymbol{M}$ such that $\nu_{j1}, \nu_{j2}, \ldots, \nu_{jm}$ are different for $j = 1, 2, \ldots, k$, where $\nu_{ji}$ is the $j$-th coordinate of $\boldsymbol{\nu}_i = \boldsymbol{M}\boldsymbol{\mu}_i + \boldsymbol{\rho}$ $(i = 1, 2, \ldots, m)$.*

PROOF. Without loss of generality, we assume that $\boldsymbol{\rho} = \boldsymbol{0}$, the zero vector. Referring to the orthogonal coordinate system $(x_1, x_2, \ldots, x_k)$, the $(k-1)$-dimensional hyperplane $\pi_{st}$ going through the origin and perpendicular to the straight line going through the points $(\mu_{1s}, \mu_{2s}, \ldots, \mu_{ks})$ and $(\mu_{1t}, \mu_{2t}, \ldots, \mu_{kt})$ is determined uniquely by $\boldsymbol{\mu}_s$ and $\boldsymbol{\mu}_t$ $(s < t$ and $s, t = 1, 2, \ldots, m)$. Consider a new orthogonal coordinate system $(z_1, z_2, \ldots, z_k)$ which has the origin in common with the coordinate system $(x_1, x_2, \ldots, x_k)$ and neither coordinate axis of which lies on any $\pi_{st}$ for $s < t$ and $s, t = 1, 2, \ldots, m$. Then the $j$-th coordinates $\nu_{j1}, \nu_{j2}, \ldots, \nu_{jm}$ $(j = 1, 2, \ldots, k)$, referred to the coordinate system $(z_1, z_2, \ldots, z_k)$, of $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_m$, respectively, are different by Statement 1.1. Let $\boldsymbol{\nu}_i = \boldsymbol{M}\boldsymbol{\mu}_i$ $(i = 1, 2, \ldots, m)$, where $\boldsymbol{M}$ is the orthogonal matrix which transforms the coordinate system $(z_1, z_2, \ldots, z_k)$ to the coordinate system $(x_1, x_2, \ldots, x_k)$. Then we can see that $\nu_{j1}$, $\nu_{j2}, \ldots, \nu_{jm}$ $(j = 1, 2, \ldots, k)$ are equal to the $j$-th coordinates of $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_m$, respectively. This completes the proof.

THEOREM 2.1. *Suppose that $\boldsymbol{X}$ has the finite mixture* (1.1) *with different $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_m$. Then, for any $\boldsymbol{\rho}$, there exists an $\boldsymbol{M}$ such that the $j$-th coordinate*

*$Y_j$ of $\boldsymbol{Y} = \boldsymbol{MX} + \boldsymbol{\rho}$ has also a finite mixture of $m$ one-dimensional normal pdf's such as*

$$g_j(y_j) = \sum_{i=1}^{m} \alpha_i n(y_j \mid \nu_{ji}, \sigma_{ji}^2) \qquad (j = 1, 2, \ldots, k), \tag{2.1}$$

*where $\nu_{ji}$ is the $j$-th coordinate of $\boldsymbol{\nu}_i = \boldsymbol{M\mu}_i + \boldsymbol{\rho}$, $\sigma_{ji}^2$ the $(j, j)$-th element of $\boldsymbol{M\Sigma}_i\boldsymbol{M}'$ $(i = 1, 2, \ldots, m)$, and $\nu_{j1}, \nu_{j2}, \ldots, \nu_{jm}$ are different.*

PROOF. By the last lemma, there exixts an $\boldsymbol{M}$ such that $\nu_{j1}, \nu_{j2}, \ldots, \nu_{jm}$ are different for $j = 1, 2, \ldots, k$, where $\nu_{ji}$ is the $j$-th coordinate of $\boldsymbol{\nu}_i = \boldsymbol{M\mu}_i + \boldsymbol{\rho}$ $(i = 1, 2, \ldots, m)$. The pdf of $\boldsymbol{Y}$ is given by (1.4). Hence, the marginal pdf of $Y_j$ is given by (2.1) and the means $\nu_{j1}, \nu_{j2}, \ldots, \nu_{jm}$ of $m$ normal pdf's are different. Accordingly, each $Y_j$ has a finite mixture of $m$ components. This completes the proof.

## 3. Construction of $Y_j^{(\gamma)}$ when $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_m$ are unknown

The last theorem guarantees the existence of an $\boldsymbol{M}$ we hope. However, as can be seen from the proof, we use a knowledge of $\boldsymbol{\mu}_i$ $(i = 1, 2, \ldots, m)$ to give an $\boldsymbol{M}$. We want to obtain actually a one-dimensional random variable $Y_j$ with a finite mixture of $m$ one-dimensional pdf's under the only assumption that $\boldsymbol{X}$ has a $k$-dimensional finite normal mixture. Hence, we need to construct an $\boldsymbol{M}$ with no knowledge of $\boldsymbol{\mu}_i$ $(i = 1, 2, \ldots, m)$ other than a fact that $m$ is finite. This difficulty can be solved by constructing matrices $\boldsymbol{M}_\gamma$ $(\gamma = 1, 2, \ldots)$ sequentially by the following.

A procedure of construction of $\boldsymbol{M}_\gamma$:

(i) For $\gamma = 1$, $\boldsymbol{M}_1 = (\boldsymbol{e}_1^{(1)}, \boldsymbol{e}_2^{(1)}, \ldots, \boldsymbol{e}_k^{(1)})$ is the $k \times k$ identity matrix.

(ii) For $\gamma \geq 2$, $\boldsymbol{M}_\gamma = (\boldsymbol{e}_1^{(\gamma)}, \boldsymbol{e}_2^{(\gamma)}, \ldots, \boldsymbol{e}_k^{(\gamma)})$ is a $k \times k$ ortogonal matrix such that $\boldsymbol{e}_1^{(\gamma)}$ is linearly independent of any $k-1$ vectors in $\{\boldsymbol{e}_i^{(\delta)} : 1 \leq i \leq k, 1 \leq \delta \leq \gamma - 1\}$, and $\boldsymbol{e}_j^{(\gamma)}$ is linearly independent of any $k-1$ vectors in $\{\boldsymbol{e}_i^{(\delta)} : 1 \leq i \leq k, 1 \leq \delta \leq \gamma - 1\} \cup \{\boldsymbol{e}_1^{(\gamma)}, \boldsymbol{e}_2^{(\gamma)}, \ldots, \boldsymbol{e}_{j-1}^{(\gamma)}\}$ when $2 \leq j \leq k$.

Let $\{\boldsymbol{M}_\gamma : \gamma = 1, 2, \ldots, \ell\}$ be a set of matrices obtained by the procedure above, then we have the following lemma which shows that an $\boldsymbol{M}_\gamma$ which satisfies our demand can be obtained.

LEMMA 3.1. *For any different $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_m$ and any $\boldsymbol{\rho}$, if $m(m-1)(k-1) < 2\ell k$, then there exist a $\gamma$ and a $j$ such that $\nu_{j1}^{(\gamma)}, \nu_{j2}^{(\gamma)}, \ldots, \nu_{jm}^{(\gamma)}$ are different, where $\nu_{ji}^{(\gamma)}$ is the $j$-th coordinate of $\boldsymbol{\nu}_i^{(\gamma)} = \boldsymbol{M}_\gamma\boldsymbol{\mu}_i + \boldsymbol{\rho}$ $(i = 1, 2, \ldots, m)$.*

PROOF. Without loss of generality, we assume again that $\boldsymbol{\rho} = \boldsymbol{0}$. Let us consider $\ell$ new orthogonal coordinate systems $(z_1^{(\gamma)}, z_2^{(\gamma)}, \ldots, z_k^{(\gamma)})$ $(\gamma = 1, 2, \ldots, \ell)$ which have the origin in common with the orthogonal coordinate system $(x_1, x_2, \ldots, x_k)$. Suppose that the coordinate system $(z_1^{(\gamma)}, z_2^{(\gamma)}, \ldots, z_k^{(\gamma)})$ is transformed to the coordinate system $(x_1, x_2, \ldots, x_k)$ by $\boldsymbol{M}_\gamma$, then all of coordinate axes are different and any $k$ coordinate axes can not lie on the same $(k-1)$-dimensionl hyperplane as can be seen from the construction of the new coordinate systems. Therefore, the number of coordinate axes lying on a hyperplane $\pi_{st}$ is $k-1$ at most. The total number of coordinate axes lying on some hyperplane $\pi_{st}$, for $s<t$, $s=1,2,\ldots,m$, then is half of $m(m-1)(k-1)$ at most. On the other hand, the total number of coordinate axes is $\ell k$. So, if $m(m-1)(k-1)<2\ell k$, then there exists at least one coordinate axis $z_j^{(\gamma)}$ which does not lie on any $\pi_{st}$. Accordingly, the $j$-th coordinates $\nu_{j1}^{(\gamma)},\nu_{j2}^{(\gamma)},\ldots,\nu_{jm}^{(\gamma)}$, referred to the coordinate system $(z_1^{(\gamma)},z_2^{(\gamma)},\ldots,z_k^{(\gamma)})$, of $\boldsymbol{M}_\gamma\boldsymbol{\mu}_1,\boldsymbol{M}_\gamma\boldsymbol{\mu}_2,\ldots,\boldsymbol{M}_\gamma\boldsymbol{\mu}_m$, respectively, are different by Statement 1.1. This completes the proof.

Let $\{\boldsymbol{Y}_\gamma:\gamma=1,2,\ldots,\ell\}$ be a set of $k$-dimensional random vectors, where $\boldsymbol{Y}_\gamma=\boldsymbol{M}_\gamma\boldsymbol{X}+\boldsymbol{\rho}$ and $\boldsymbol{M}_\gamma$ is an orthogonal matrix obtained by the procedure above. Then we can prove the following theorem in the same way to the proof of Theorem 2.1 by substituting $\boldsymbol{M}_\gamma$ for $\boldsymbol{M}$.

THEOREM 3.1. *Suppose that* $\boldsymbol{X}$ *has a finite mixture* (1.1) *with different* $\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\ldots,\boldsymbol{\mu}_m$. *If* $m(m-1)(k-1)<2\ell k$, *then, for any* $\boldsymbol{\rho}$, *there exist a* $\gamma$ *and a* $j$ *such that the* $j$*-th coordinate* $Y_j^{(\gamma)}$ *of* $\boldsymbol{Y}_\gamma=\boldsymbol{M}_\gamma\boldsymbol{X}+\boldsymbol{\rho}$ *also has a finite mixture of* $m$ *one-dimensional normal pdf's such as*

$$g_j(y_j)=\sum_{i=1}^{m}\alpha_i n(y_j\mid\nu_{ji}^{(\gamma)},(\sigma_{ji}^{(\gamma)})^2), \tag{3.1}$$

*where* $\nu_{ji}^{(\gamma)}$ *is the* $j$*-th coordinate of* $\boldsymbol{\nu}_i^{(\gamma)}=\boldsymbol{M}_\gamma\boldsymbol{\mu}_i+\boldsymbol{\rho}$, $(\sigma_{ji}^{(\gamma)})^2$ *the* $(j,j)$*-th element of* $\boldsymbol{M}_\gamma\boldsymbol{\Sigma}_i\boldsymbol{M}'_\gamma$ $(i=1,2,\ldots,m)$, *and* $\nu_{j1}^{(\gamma)}$, $\nu_{j2}^{(\gamma)},\ldots,\nu_{jm}^{(\gamma)}$ *are different.*

The last theorem guarantees that if we construct a sequence of $k$-dimensional random vectors $\{\boldsymbol{Y}_\gamma\}$ sequentially by the method above, then we will eventually obtain a one-dimensional radom variable $Y_j^{(\gamma)}$ which has a finite normal mixture with the true number of components and the true mixing ratios.

*Example.* Let the pdf of $\boldsymbol{X}$ be

$$f(\boldsymbol{x})=\alpha_1 n(\boldsymbol{x}\mid\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)+\alpha_2 n(\boldsymbol{x}\mid\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)+\alpha_3 n(\boldsymbol{x}\mid\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_3), \tag{3.2}$$

where $\boldsymbol{\mu}_i$ $(i=1,2,3)$ are those of (1.2). By the last theorem, we can construct a univariate random variable which satisfies our demand for $\ell=2$. So, any orthogonal matrix $\boldsymbol{M}_2$ other than the identity matrix satisfies this reqirement. Let us have a try with

$$\boldsymbol{M}_2=\begin{pmatrix}1/\sqrt{2} & 1/\sqrt{2}\\ -1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix}.$$

Then the pdf of $\boldsymbol{Y}_2=\boldsymbol{M}_2\boldsymbol{X}$ is

$$g(\boldsymbol{y})=\alpha_1 n(\boldsymbol{y}\mid\boldsymbol{\eta}_1,\boldsymbol{\Sigma}_1^*)+\alpha_2 n(\boldsymbol{y}\mid\boldsymbol{\eta}_2,\boldsymbol{\Sigma}_2^*)+\alpha_3 n(\boldsymbol{y}\mid\boldsymbol{\eta}_3,\boldsymbol{\Sigma}_3^*), \tag{3.3}$$

where $\boldsymbol{\eta}_1 = (0,0)'$, $\boldsymbol{\eta}_2 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})'$, $\boldsymbol{\eta}_3 = (\sqrt{2}, 0)'$ and $\boldsymbol{\Sigma}_i^* = \boldsymbol{M}_2 \boldsymbol{\Sigma}_i \boldsymbol{M}_2'$ $(i = 1,2,3)$. Accordingly, the pdf of the first coordinate $Y_1^{(2)}$ of $\boldsymbol{Y}_2$ is given as follows:

$$(3.4) \quad g_1(y_1) = \alpha_1 n(y_1 \mid 0, (\sigma_{11}^{(2)})^2) + \alpha_2 n\left(y_1 \,\middle|\, \frac{1}{\sqrt{2}}, (\sigma_{12}^{(2)})^2\right) + \alpha_3 n(y_1 \mid \sqrt{2}, (\sigma_{13}^{(2)})^2).$$

**Acknowledgements**

The author wishes to express his hearty thanks to the referee for his valuable comments.

## REFERENCES

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, Second Edition, John Wiley.

Gantmacher, F. R. (1977). *The theory of matrices*, Chelsea Publishing Company.

McLachlan, G. J. and Basford, K. E. (1988). *Mixture Models*, Marcel Dekker, Inc.

Titterington, D. M. *et al.* (1985). *Statistical Analysis of Finite Mixture Distributions*, John Wiley.