

Mixtures of ‘Unrestricted’ Skew- t Factor Analyzers

Paula M. Murray, Ryan P. Browne and Paul D. McNicholas*

Department of Mathematics & Statistics, University of Guelph

Abstract

Mixtures of skew- t distributions offer a flexible choice for model-based clustering. A mixture model of this sort can be implemented using a variety of forms of the skew- t distribution. Herein we develop a set of mixtures of skew- t factor analyzers models for clustering of high-dimensional data using a comparatively flexible variant of the skew- t distribution. Methodological details of our approach, which represents an extension of the mixture of factor analyzers model to a flexible skew- t distribution, are outlined and details of parameter estimation are provided. Extension of our mixtures of ‘unrestricted’ skew- t factor analyzers to a family of parsimonious models is also outlined.

1 Introduction

Mixture models have become an increasingly popular tool for clustering since they were used by Wolfe (1963), and the term model-based clustering is commonly used to describe the application of mixture models for clustering. Although much of the work on model-based clustering has been based on Gaussian mixtures, recent years have seen extensive work on non-Gaussian mixture model-based approaches. In fact, there has been a veritable explosion of such work over the past three years or so (e.g., Lin, 2010; Frühwirth-Schnatter and Pyne, 2010; Lee and McLachlan, 2011; Vrbik and McNicholas, 2012; Franczak et al., 2013; Morris et al., 2013; Morris and McNicholas, 2013a,b; Murray et al., 2013a,b; Forbes and Wraith, 2013; Lin et al., 2013; McNicholas et al., 2013). Much of this work has focused on non-elliptical mixtures, including mixtures of skew-normal distributions (e.g., Lin et al., 2007), mixtures of skew- t distributions (e.g., Lin, 2010), mixtures of shifted asymmetric Laplace distributions (Franczak et al., 2013), mixtures of variance-gamma distributions (McNicholas et al., 2013), and mixtures of generalized hyperbolic distributions (Browne and McNicholas, 2013).

*Department of Mathematics & Statistics, University of Guelph, Guelph, Ontario, N1G 2W1, Canada.
E-mail: pmcnicho@uoguelph.ca.

There are several forms of the skew-normal and skew- t distributions, some of which are discussed by Lee and McLachlan (2013). To date, three forms have been studied for model-based clustering. The first, which Lee and McLachlan (2013) refer to as “restricted”, is that developed by Azzalini and Dalla Valle (1996), and the second, which Lee and McLachlan (2013) refer to as “unrestricted”, is that of Sahu et al. (2003). We prefer not to use the terms “restricted” and “unrestricted” because they give the impression that one is a special case of the other. What we see as this unusual use of “unrestricted” is the reason the word appears in single inverted commas in the title of this paper, i.e., to indicate a non-standard usage of the word. Hereafter, we will refer to the form of Azzalini and Dalla Valle (1996) as classical, and to that of Sahu et al. (2003) simply as SDB, from the authors’ initials. The third form that has been studied for model-based clustering is a special and limiting case of the generalized hyperbolic distribution (cf. Murray et al., 2013a,b).

The SDB skew- t distribution was used by Lin (2010) in a mixture modelling framework, exploiting a variant of the expectation-maximization (EM) algorithm (Dempster et al., 1977) for parameter estimation. However, this model requires a Monte Carlo estimation step which, as noted by Lee and McLachlan (2011), is computationally expensive. In an effort to improve computational efficiency of the integral estimation in the E-step for the SDB skew- t mixture model, Lee and McLachlan (2011) write the integrals in the form of a truncated multivariate- t distribution. This is subsequently written as a non-truncated t -distribution and pre-existing statistical packages can be used for computation. Herein we extend the mixture of factor analyzers model (Ghahramani and Hinton, 1997; McLachlan and Peel, 2000) using the SDB skew- t mixture model. The resulting mixture of SDB skew- t factor analyzers (MSDBFA) model is very well suited for application to high-dimensional data.

The remainder of this paper is outlined as follows. In Section 2, background material is provided. Then the MSDBFA model is introduced (Section 3.1) and details of parameter estimation are provided (Section 3.2). A parsimonious family of MSDBFA models is outlined in Section 3.3, and we conclude with discussion (Section 4).

2 Background

2.1 Mixtures of SDB Skew- t Distributions

The density of a finite mixture model is given by

$$f(\mathbf{x} \mid \boldsymbol{\theta}_g) = \sum_{g=1}^G \pi_g f_g(\mathbf{x} \mid \boldsymbol{\theta}_g), \quad (1)$$

where $\pi_g > 0$ is a mixing proportion such that $\sum_{g=1}^G \pi_g = 1$, and $f_g(\mathbf{x} \mid \boldsymbol{\theta}_g)$ is the g th component density with parameters $\boldsymbol{\theta}_g$. The density of the SDB multivariate skew- t distribution

is given by

$$\begin{aligned}
b(\mathbf{x} \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \nu_g, \boldsymbol{\Delta}_g) &= 2^p t_p(\mathbf{x}; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \nu_g) \\
&\times T_p\left(\boldsymbol{\Delta}_g \boldsymbol{\Sigma}_g^{-1}(\mathbf{x} - \boldsymbol{\mu}_g) \sqrt{\left(\frac{\nu_g + p}{\nu_g + d(\mathbf{x} \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)}\right)}; \mathbf{0}, \mathbf{I}_p - \boldsymbol{\Delta}_g \boldsymbol{\Sigma}_g^{-1} \boldsymbol{\Delta}_g, \nu_g + p\right), \quad (2)
\end{aligned}$$

where $\boldsymbol{\Delta}_g = \text{diag}(\boldsymbol{\delta}_g)$ is the skewness, $d(\mathbf{x} \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) = (\mathbf{x} - \boldsymbol{\mu}_g)' \boldsymbol{\Sigma}_g^{-1} (\mathbf{x} - \boldsymbol{\mu}_g)$, $t_p(\cdot)$ is the density of a p -dimensional t -distributed random variable, and $T_p(\cdot)$ is the cumulative distribution function. Using $b(\mathbf{x} \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \nu_g, \boldsymbol{\Delta}_g)$ in (2) as the component density in (1), i.e., setting $f_g(\mathbf{x} \mid \boldsymbol{\theta}_g) = b(\mathbf{x} \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \nu_g, \boldsymbol{\Delta}_g)$ in (1), leads to a mixture of SDB distributions. A mixture model of this form has previously been used for clustering by Lin (2010) and Lee and McLachlan (2011).

2.2 Mixtures of Factor Analyzers

The factor analysis model (Spearman, 1904) assumes that the variation in p observed variables can be explained by q unobserved or latent variables where $q \ll p$. We may write this model as

$$\mathbf{X} = \boldsymbol{\Lambda} \mathbf{U} + \boldsymbol{\epsilon},$$

where \mathbf{X} is a p -dimensional observed data vector, $\boldsymbol{\Lambda}$ is a $p \times q$ matrix of factor loadings, \mathbf{U} is a q -dimensional vector of latent factors such that $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$, where $\boldsymbol{\Psi}$ is a $p \times p$ diagonal matrix with positive entries. It follows that the marginal distribution of \mathbf{X} is multivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Lambda} \boldsymbol{\Lambda}' + \boldsymbol{\Psi}$.

The mixture of factor analyzers model (Ghahramani and Hinton, 1997; McLachlan and Peel, 2000) of the form

$$f(\mathbf{x} \mid \boldsymbol{\mu}_g, \boldsymbol{\Lambda}_g, \boldsymbol{\Psi}_g) = \sum_{g=1}^G \pi_g \phi_g(\mathbf{x} \mid \boldsymbol{\mu}_g, \boldsymbol{\Lambda}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g). \quad (3)$$

The mixture of factor analyzers model is effective for modelling high dimensional data and has been extended in various ways (e.g., McLachlan et al., 2007; McNicholas and Murphy, 2008, 2010; McNicholas, 2010; Baek et al., 2010; Andrews et al., 2011; Steane et al., 2011; Baek and McLachlan, 2011; Andrews and McNicholas, 2011a,b, 2012; Murray et al., 2013a).

2.3 The Expectation-Maximization Algorithm

The expectation-maximization (EM) algorithm (Dempster et al., 1977) is an iterative algorithm for finding maximum likelihood estimates when data are incomplete or are treated as such. The EM algorithm has been widely used for parameter estimation in mixture model based-clustering. The EM algorithm alternates between two steps, an expectation (E-) step

and a maximization (M-) step. In the E-step, the expected value of the complete-data log-likelihood is computed, i.e., updated, given the current parameter estimates. In the M-step, this expected value is maximized with respect to the model parameters, i.e., the parameter estimates are updated. McLachlan and Krishnan (2008) give a detailed review of EM algorithms and their application to mixture models.

To formulate clustering problems in the complete-data framework, we introduce indicator variables Z_{ig} , where $z_{ig} = 1$ if \mathbf{x}_i is in component g and $z_{ig} = 0$ otherwise. The complete-data are then given by the observed $\mathbf{x}_1, \dots, \mathbf{x}_n$ together with the missing $\mathbf{z}_1, \dots, \mathbf{z}_n$, where $\mathbf{z}_i = (z_{i1}, \dots, z_{iG})$ for $i = 1, \dots, n$. For the MUSTFA models, there is another source of missing data, i.e., the latent variables (cf. Section 3.2). These latent variables join the observed data and the missing labels to form the complete-data for the MUSTFA models. We fit our MUSTFA models using an alternating expectation-conditional maximization (AECM) algorithm (Meng and van Dyk, 1997), which allows different complete-data at each conditional maximization (CM) step. Details are given in Section 3.2.

2.4 Parsimonious Mixtures of Skew- t Factor Analyzers

Murray et al. (2013a) introduced a family of mixtures of skew- t factor analyzers for clustering high-dimensional data. This family is an extension of the Gaussian family of McNicholas and Murphy (2008), which comprises the mixture of factor analyzers model and variants thereof. The family of skew- t factor models introduced by Murray et al. (2013a) is based on the skew- t distribution that arises as a special and limiting case of the generalized hyperbolic distribution (cf. Barndorff-Nielsen and Shephard, 2001), and it is referred to as the parsimonious mixtures of skew- t factor analyzers (PMSTFA) family (Table 1). The models in the PMSTFA family are of the general form

$$f(\mathbf{x} | \boldsymbol{\vartheta}) = \sum_{g=1}^G \pi_g \zeta(\mathbf{x} | \boldsymbol{\mu}_g, \boldsymbol{\Lambda}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g, \boldsymbol{\alpha}_g, \nu_g),$$

where $\boldsymbol{\vartheta}$ denotes the model parameters and

$$\begin{aligned} \zeta(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \nu) &= \left[\frac{\nu + \delta(\mathbf{x}, \boldsymbol{\mu} | \boldsymbol{\Sigma})}{\boldsymbol{\alpha}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}} \right]^{(-\nu-p)/4} \\ &\times \frac{\nu^{\nu/2} K_{(-\nu-p)/2} \left(\sqrt{[\boldsymbol{\alpha}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}] [\nu + \delta(\mathbf{x}, \boldsymbol{\mu} | \boldsymbol{\Sigma})]} \right)}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2} \Gamma(\nu/2) 2^{\nu/2-1} \exp\{(\boldsymbol{\mu} - \mathbf{x})' \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}\}}, \end{aligned} \quad (4)$$

is the density of the skew- t distribution of Barndorff-Nielsen and Shephard (2001), where $\boldsymbol{\mu}$ is the location, $\boldsymbol{\Sigma}$ is the scale matrix, $\boldsymbol{\alpha}$ is the skewness, ν is the value for degrees of freedom, and K_λ is the modified Bessel function of the third kind with index parameter λ .

A skew- t random variable \mathbf{X} with the density in (4) can be written

$$\mathbf{X} = \boldsymbol{\mu} + Y \boldsymbol{\alpha} + \sqrt{Y} \mathbf{Q},$$

Table 1: Nomenclature and covariance structure for the members of the PMSTFA family.

ID	Loading Matrix	Error Variance	Isotropic	Free Covariance Parameters
CCC	Constrained	Constrained	Constrained	$[pq - q(q - 1)/2] + 1$
CCU	Constrained	Constrained	Unconstrained	$[pq - q(q - 1)/2] + p$
CUC	Constrained	Unconstrained	Constrained	$[pq - q(q - 1)/2] + G$
CUU	Constrained	Unconstrained	Unconstrained	$[pq - q(q - 1)/2] + Gp$
UCC	Unconstrained	Constrained	Constrained	$G[pq - q(q - 1)/2] + 1$
UCU	Unconstrained	Constrained	Unconstrained	$G[pq - q(q - 1)/2] + p$
UUC	Unconstrained	Unconstrained	Constrained	$G[pq - q(q - 1)/2] + G$
UUU	Unconstrained	Unconstrained	Unconstrained	$G[pq - q(q - 1)/2] + Gp$

where $\mathbf{Q} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ and $Y \sim \text{IG}(\nu/2, \nu/2)$, where $\text{IG}(\cdot)$ denotes the inverse Gamma distribution. It follows that $\mathbf{X} \mid y \sim \mathcal{N}(\boldsymbol{\mu} + y\boldsymbol{\alpha}, y\Sigma)$ and $Y \mid \mathbf{x} \sim \text{GIG}(\boldsymbol{\alpha}'\Sigma^{-1}\boldsymbol{\alpha}, \nu + \delta(\mathbf{x}, \boldsymbol{\mu} \mid \Sigma), -(\nu + p)/2)$, where $\text{GIG}(\cdot)$ denotes the generalized inverse Gaussian (GIG) distribution (Good, 1953; Barndorff-Nielsen and Halgreen, 1977; Blæsild, 1978; Halgreen, 1979; Jørgensen, 1982). The AECM algorithm described by Murray et al. (2013a) for the fitting of the PMSTFA models requires the computation of the following expected values:

$$\begin{aligned}\mathbb{E}[Y_i] &= \sqrt{\frac{\chi}{\psi}} \frac{K_{\lambda+1}(\sqrt{\psi\chi})}{K_{\lambda}(\sqrt{\psi\chi})}, \\ \mathbb{E}[1/Y_i] &= \sqrt{\frac{\psi}{\chi}} \frac{K_{\lambda+1}(\sqrt{\psi\chi})}{K_{\lambda}(\sqrt{\psi\chi})} - \frac{2\lambda}{\chi}, \\ \mathbb{E}[\log(Y_i)] &= \log\sqrt{\frac{\chi}{\psi}} + \frac{1}{K_{\lambda}(\sqrt{\psi\chi})} \frac{\delta}{\delta\lambda} K_{\lambda}(\sqrt{\psi\chi}),\end{aligned}$$

where $\psi, \chi \in \mathbb{R}$. These expectations lead to a very computationally efficient E-Step relative to that of the EM algorithm for other skew- t mixture models. Note that these attractive closed forms follow from the fact that $Y \mid \mathbf{x}$ has the GIG distribution. Extensive details are given by Murray et al. (2013a).

3 Methodology

3.1 The MSDBFA Model

We can write a random variable \mathbf{X} arising from the SDB skew- t distribution, cf. (2), as

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Delta}|\mathbf{V}^*| + \mathbf{R}^*,$$

where $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta})$ is a skewness parameter and

$$\begin{bmatrix} \mathbf{V}^* \\ \mathbf{R}^* \end{bmatrix} \sim t_{2p} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \Xi \end{bmatrix}, \nu \right) \quad (5)$$

with $\Xi = \Sigma - \Delta^2$. It follows that $\mathbf{X} = \boldsymbol{\mu} + (\mathbf{R} \mid \mathbf{V} > \mathbf{0})$, with

$$\begin{bmatrix} \mathbf{V} \\ \mathbf{R} \end{bmatrix} \sim t_{2p} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}_p & \Delta \\ \Delta & \Sigma \end{bmatrix}, \nu \right). \quad (6)$$

By introducing a latent variable $W \sim \text{gamma}(\nu/2, \nu/2)$, it follows that

$$\mathbf{V} \mid w \sim \text{HN}_p \left(\frac{1}{w} \mathbf{I}_p \right),$$

where $\text{HN}_p(\cdot)$ denotes the half-normal distribution, and

$$\mathbf{X} \mid \mathbf{v}, w \sim \mathcal{N} \left(\boldsymbol{\mu} + \Delta \mathbf{v}, \frac{1}{w} \Sigma \right).$$

Thus, we can write

$$\mathbf{X} = \boldsymbol{\mu} + \Delta |\mathbf{V}| + \frac{1}{W} \mathbf{R}, \quad (7)$$

where $\mathbf{R} \sim \mathcal{N}(\mathbf{0}, \Sigma)$. Recall that the factor analysis model is written as

$$\mathbf{R} = \Lambda \mathbf{U} + \boldsymbol{\epsilon}, \quad (8)$$

where $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \Psi)$. Substituting (8) into (7) gives

$$\mathbf{X} = \boldsymbol{\mu} + \Delta |\mathbf{V}| + \frac{1}{W} (\Lambda \mathbf{U} + \boldsymbol{\epsilon}).$$

It follows that $\mathbf{X} \mid \mathbf{v}, w \sim \mathcal{N}(\boldsymbol{\mu} + \Delta \mathbf{v}, (1/w)(\Lambda \Lambda' + \Psi))$.

Therefore, the density of our MSDBFA model can be written

$$\begin{aligned} f(\mathbf{x} \mid \boldsymbol{\vartheta}) &= \sum_{g=1}^G \pi_g b(\mathbf{x} \mid \boldsymbol{\mu}_g, \Lambda_g \Lambda_g' + \Psi_g, \nu_g, \Delta_g) \\ &= \sum_{g=1}^G \pi_g \gamma(w_{ig} \mid \nu_g/2, \nu_g/2) h(\mathbf{v}_{ig} \mid (1/w_{ig}) \mathbf{I}_p) \\ &\quad \times \phi(\mathbf{u}_{ig} \mid \mathbf{0}, \mathbf{I}_q) \phi(\mathbf{x} \mid \boldsymbol{\mu}_g + \Delta_g \mathbf{v}_{ig}, (1/w_{ig})(\Lambda_g \Lambda_g' + \Psi_g)), \end{aligned}$$

where $\gamma(\cdot)$ denotes the density of a gamma distribution, $\phi(\cdot)$ denotes the density of a Gaussian distribution, $h(\cdot)$ denotes the density of a half-normal distribution, and $\boldsymbol{\vartheta}$ denotes all model parameters.

3.2 Parameter Estimation

At each iteration of the AECM algorithm, the expected-value of the complete-data log-likelihood is computed. The complete-data log-likelihood for the MUSTFA model is

$$l_c(\boldsymbol{\vartheta} \mid \mathbf{x}, \mathbf{w}, \mathbf{v}, \mathbf{z}) = \sum_{i=1}^n \sum_{g=1}^G z_{ig} \log \left[\pi_g \gamma(w_{ig} \mid \nu_g/2, \nu_g/2) h(\mathbf{v}_{ig} \mid (1/w_{ig}) \mathbf{I}_p) \right. \\ \left. \times \phi(\mathbf{u}_{ig} \mid \mathbf{0}, \mathbf{I}_q) \phi(\mathbf{x} \mid \boldsymbol{\mu}_g + \boldsymbol{\Delta}_g \mathbf{v}_{ig}, (1/w_{ig})(\boldsymbol{\Lambda}_g \boldsymbol{\Lambda}'_g + \boldsymbol{\Psi}_g)) \right], \quad (9)$$

The E-steps of our AECM algorithm require the following expected values:

$$\begin{aligned} \mathbb{E}[Z_{ig} \mid \mathbf{x}_i] &= \frac{\pi_g b(\mathbf{x}_i \mid \boldsymbol{\mu}_g, \boldsymbol{\Lambda}_g \boldsymbol{\Lambda}'_g + \boldsymbol{\Psi}_g, \nu_g, \boldsymbol{\Delta}_g)}{\sum_{h=1}^G \pi_h b(\mathbf{x}_i \mid \boldsymbol{\mu}_h, \boldsymbol{\Lambda}_h \boldsymbol{\Lambda}'_h + \boldsymbol{\Psi}_h, \nu_h, \boldsymbol{\Delta}_h)} =: \hat{z}_{ig}, \\ \mathbb{E}[W_{ig} \mid \mathbf{x}_i, z_{ig} = 1] &=: e_{1,ig}, \\ \mathbb{E}[W_{ig} \mathbf{V}_{ig} \mid \mathbf{x}_i, z_{ig} = 1] &=: e_{2,ig}, \\ \mathbb{E}[W_{ig} \mathbf{V}_{ig} \mathbf{V}'_{ig} \mid \mathbf{x}_i, z_{ig} = 1] &=: e_{3,ig}, \\ \mathbb{E}[\log W_{ig} \mid \mathbf{x}_i, z_{ig} = 1] &=: e_{4,ig}. \end{aligned}$$

We employ the integral approximation method of Lee and McLachlan (2011), which we believe offers advantages in terms of accuracy and computation time in evaluating these intractable expectations when compared to the Monte Carlo EM algorithm of Lin (2010).

At the first stage of our AECM algorithm, our incomplete-data include the labels z_{ig} , the latent variables w_{ig} , and the latent \mathbf{v}_{ig} . The location $\boldsymbol{\mu}_g$ and the skewness $\boldsymbol{\Delta}_g$ are updated via

$$\begin{aligned} \boldsymbol{\mu}_g &= \frac{\sum_{i=1}^n [e_{1,ig} \mathbf{x}_i - \boldsymbol{\Delta}_g e_{2,ig}]}{\sum_{i=1}^n e_{1,ig}}, \quad \text{and} \\ \boldsymbol{\Delta}_g &= \left((\boldsymbol{\Lambda}_g \boldsymbol{\Lambda}_g + \boldsymbol{\Psi})^{-1} \odot \sum_{i=1}^n e_{3,ig} \right)^{-1} \text{diag} \left((\boldsymbol{\Lambda}_g \boldsymbol{\Lambda}_g + \boldsymbol{\Psi})^{-1} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}_g) e'_{2,ig} \right), \end{aligned}$$

respectively, and the equation

$$\log(\nu_g/2) - \psi(\nu_g/2) - \frac{1}{n} \sum_{i=1}^n (e_{1,ig} - e_{4,ig}) + 1 = 0$$

is solved numerically to obtain the update for ν_g .

The ‘sample covariance’ matrix \mathbf{S}_g is updated by

$$\mathbf{S}_g = \frac{1}{n} \sum_{i=1}^n \left[\boldsymbol{\Delta}_g e'_{3,ig} \boldsymbol{\Delta}'_g - (\mathbf{x}_i - \boldsymbol{\mu}_g) e'_{2,ig} \boldsymbol{\Delta}_g + (\mathbf{x}_i - \boldsymbol{\mu}_g) (\mathbf{x}_i - \boldsymbol{\mu}_g)' e_{1,ig} - \boldsymbol{\Delta}_g e_{2,ig} (\mathbf{x}_i - \boldsymbol{\mu}_g)' \right].$$

At the second stage of our AECM algorithm, the incomplete-data include the labels z_{ig} , the latent variables w_{ig} , the latent \mathbf{v}_{ig} , and the latent factors \mathbf{u}_{ig} . In this step, the factor loading matrix $\mathbf{\Lambda}_g$ and the error variance matrix $\mathbf{\Psi}_g$ are updated via

$$\mathbf{\Lambda}_g = \mathbf{S}_g \boldsymbol{\beta}'_g \boldsymbol{\Theta}_g^{-1} \quad \text{and} \quad \mathbf{\Psi}_g = \text{diag}\{\mathbf{S}_g - \mathbf{\Lambda}_g \boldsymbol{\beta}_g \mathbf{S}_g\},$$

respectively, where $\boldsymbol{\beta}_g = \boldsymbol{\Lambda}'_g (\boldsymbol{\Lambda}_g \boldsymbol{\Lambda}'_g + \mathbf{\Psi}_g)^{-1}$ and $\boldsymbol{\Theta}_g = \mathbf{I}_p - \boldsymbol{\beta}_g \boldsymbol{\Lambda}_g + \boldsymbol{\beta}_g \mathbf{S}_g \boldsymbol{\beta}'_g$.

3.3 Parsimonious MSDBFA Models

Following McNicholas and Murphy (2008), we develop a family of parsimonious models by imposing combinations of the constraints $\mathbf{\Lambda}_g = \mathbf{\Lambda}$, $\mathbf{\Psi}_g = \mathbf{\Psi}$, and $\mathbf{\Psi}_g = \psi_g \mathbf{I}_p$. The resulting family of eight parsimonious MSDBFA (PMSDBFA) models will have scale matrices analogous to those used in the PMSTFA models (Table 1). Note that while the constraints used for the scale matrices in the PMSTFA and PMSDBFA families are the same, the underlying densities are different.

4 Discussion

The mixture of factor analyzers model has been extended to SDB skew- t mixtures. Model development and parameter estimation have been outlined, and parsimonious analogues of the model have been introduced. Work in progress will illustrate the efficacy of our PMSDBFA models for clustering high-dimensional data, and compare the performance of these models to the PMSTFA family. Other work in progress focuses on extending the mixture of common factor analyzers model (Baek et al., 2010) to SDB skew- t mixtures and comparing the resulting model to the common skew- t factor analyzers model of Murray et al. (2013b).

References

- Andrews, J. L. and P. D. McNicholas (2011a). Extending mixtures of multivariate t -factor analyzers. *Statistics and Computing* 21(3), 361–373.
- Andrews, J. L. and P. D. McNicholas (2011b). Mixtures of modified t -factor analyzers for model-based clustering, classification, and discriminant analysis. *Journal of Statistical Planning and Inference* 141(4), 1479–1486.
- Andrews, J. L. and P. D. McNicholas (2012). Model-based clustering, classification, and discriminant analysis via mixtures of multivariate t -distributions. *Statistics and Computing* 22(5), 1021–1029.

- Andrews, J. L., P. D. McNicholas, and S. Subedi (2011). Model-based classification via mixtures of multivariate t -distributions. *Computational Statistics and Data Analysis* 55(1), 520–529.
- Azzalini, A. and A. Dalla Valle (1996). The multivariate skew-normal distribution. *Biometrika* 83(4), 715–726.
- Baek, J., G. McLachlan, and L. Flack (2010). Mixtures of factor analyzers with common factor loadings: Applications to the clustering and visualization of high-dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(7), 1298–1309.
- Baek, J. and G. J. McLachlan (2011). Mixtures of common t -factor analyzers for clustering high-dimensional microarray data. *Bioinformatics* 27(9), 1269–1276.
- Barndorff-Nielsen, O. and C. Halgreen (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* 38, 309–311.
- Barndorff-Nielsen, O. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society B* 63, 167–241.
- Blæsild, P. (1978). The shape of the generalized inverse Gaussian and hyperbolic distributions. Research Report 37, Department of Theoretical Statistics, Aarhus University, Denmark.
- Browne, R. P. and P. D. McNicholas (2013). A mixture of generalized hyperbolic distributions. Arxiv preprint arXiv:1305.1036.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B* 39(1), 1–38.
- Forbes, F. and D. Wraith (2013). A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: application to robust clustering. *Statistics and Computing*. To appear, DOI: 10.1007/s11222-013-9414-4.
- Franczak, B. C., R. P. Browne, and P. D. McNicholas (2013). Mixtures of shifted asymmetric Laplace distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. To appear.
- Frühwirth-Schnatter, S. and S. Pyne (2010). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew- t distributions. *Biostatistics* 11(2), 317–336.
- Ghahramani, Z. and G. Hinton (1997). The EM algorithm for factor analyzers. Technical Report CRG-TR-96-1, University of Toronto, Toronto.

- Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* 40, 237–260.
- Halgreen, C. (1979). Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* 47, 13–18.
- Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. New York: Springer-Verlag.
- Lee, S. and G. McLachlan (2013). On mixtures of skew normal and skew t -distributions. *Advances in Data Analysis and Classification* 7(3), 241–266.
- Lee, S. and G. J. McLachlan (2011). On the fitting of mixtures of multivariate skew- t distributions via the EM algorithm. Arxiv preprint arXiv:1109.4706.
- Lin, T., J. Lee, and W. Hsieh (2007). Robust mixture modeling using the skew- t distribution. *Statistics and Computing* 17(2), 81–92.
- Lin, T.-I, G. J. McLachlan, and S. X. Lee (2013). Extending mixtures of factor models using the restricted multivariate skew-normal distribution. Arxiv preprint arXiv:1307.1748.
- Lin, T.-I. (2010). Robust mixture modeling using multivariate skew- t distributions. *Statistics and Computing* 20(3), 343–356.
- McLachlan, G., R. Bean, and L. B.-T. Jones (2007). Extension of the mixture of factor analyzers model to incorporate the multivariate t -distribution. *Computational Statistics & Data Analysis* 51(11), 5327 – 5338.
- McLachlan, G. J. and T. Krishnan (2008). *The EM Algorithm and Extensions* (2nd ed.). New York: Wiley.
- McLachlan, G. J. and D. Peel (2000). Mixtures of factor analyzers. In *Proceedings of the Seventh International Conference on Machine Learning*, San Francisco, pp. 599–606. Morgan Kaufmann.
- McNicholas, P. D. (2010). Model-based classification using latent Gaussian mixture models. *Journal of Statistical Planning and Inference* 140(5), 1175–1181.
- McNicholas, P. D. and T. B. Murphy (2008). Parsimonious Gaussian mixture models. *Statistics and Computing* 18, 285–296.
- McNicholas, P. D. and T. B. Murphy (2010). Model-based clustering of microarray expression data via latent Gaussian mixture models. *Bioinformatics* 26(21), 2705–2712.
- McNicholas, S. M., P. D. McNicholas, and R. P. Browne (2013). Mixtures of variance-gamma distributions. Arxiv preprint arXiv:1309.2695.

- Meng, X.-L. and D. A. van Dyk (1997). The EM algorithm — an old folk song sung to a fast new tune (with discussion). *Journal of the Royal Statistical Society: Series B* 59, 511–567.
- Morris, K., P. D. McNicholas, and L. Scrucca (2013). Dimension reduction for model-based clustering via mixtures of multivariate t -distributions. *Advances in Data Analysis and Classification* 7(3), 321–338.
- Morris, K. and P. D. McNicholas (2013a). Dimension reduction for model-based clustering via mixtures of shifted asymmetric Laplace distributions. *Statistics and Probability Letters* 83(9), 2088–2093.
- Morris, K. and P. D. McNicholas (2013b). Non-Gaussian mixtures for dimension reduction, clustering, classification, and discriminant analysis. Arxiv preprint arXiv:1308.6315.
- Murray, P. M., R. P. Browne, and P. D. McNicholas (2013a). Mixtures of skew- t factor analyzers. Arxiv preprint arXiv:1305.4301.
- Murray, P. M., P. D. McNicholas, and R. P. Browne (2013b). Mixtures of common skew- t factor analyzers. Arxiv preprint arXiv:1307.5558.
- Sahu, S. K., D. K. Dey, and M. D. Branco (2003). A new class of multivariate skew distributions with applications to bayesian regression models. *Canadian Journal of Statistics* 31(2), 129–150.
- Spearman, C. (1904). The proof and measurement of association between two things. *Journal of Statistical Planning and Inference* 15, 72–101.
- Steane, M. A., P. D. McNicholas, and R. Y. Yada (2011). Model-based classification via mixtures of multivariate t -factor analyzers. *Communications in Statistics – Simulation and Computation* 41(4), 510–523.
- Vrbik, I. and P. D. McNicholas (2012). Analytic calculations for the EM algorithm for multivariate skew-mixture models. *Statistics and Probability Letters* 82(6), 1169–1174.
- Wolfe, J. H. (1963). Object cluster analysis of social areas. Master’s thesis, University of California, Berkeley.