Statistics and Probability Letters 116 (2016) 1–5

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Comment on "On nomenclature, and the relative merits of two formulations of skew distributions" by A. Azzalini, R. Browne, M. Genton, and P. McNicholas

Geoffrey J. McLachlan*, Sharon X. Lee

Department of Mathematics, University of Queensland, St. Lucia, Brisbane, Australia

ARTICLE INFO

Article history:
Received 2 February 2016
Received in revised form 4 April 2016
Accepted 5 April 2016
Available online 20 April 2016

Keywords:
Skew t-distribution
Multivariate skew distribution
Finite mixture models
Model based clustering

ABSTRACT

We clarify an apparent misunderstanding in Azzalini et al. (2016) of the nomenclature to distinguish between two formulations of the skew t-distribution. Also, Lee and McLachlan (2014b) have shown how a broader class that encompasses both models can be fitted.

1. Introduction

In this paper, we provide some comments on Azzalini et al. (2016), which we shall refer to as ABGM in the sequel. In ABGM a comparison is given of two different distributions proposed for the modelling of data that have asymmetric and possibly long-tailed clusters. They refer to the two models as the classical and SDB, the latter so named since it was proposed by Sahu et al. (2003). These two distributions were referred to as the restricted multivariate skew t (rMST) and unrestricted multivariate skew t (uMST) distributions by Lee and McLachlan (2013d). We shall continue to use this latter terminology in our comments below.

In our comments we first wish to respond to statements in ABGM that are apparently based on a misunderstanding of the reporting of our results in Lee and McLachlan (2014a) and, in particular, of the nomenclature used therein. The discussion of our work in ABGM is limited to Lee and McLachlan (2014a), and so it does not consider the results presented in our other papers, in particular, Lee and McLachlan (2013a,b,c,d, 2014b, 2015, 2016) although the first two of the latter seven papers are cited in ABGM. It is particularly unfortunate that these papers are not included in the comparison in ABGM as they contain a comparison of the restricted and unrestricted models applied to nine datasets from various fields.

Also, explicit cautionary notes are made in them to guard against any potential misunderstanding of our terminology. For example, in Lee and McLachlan (2013d, Page 244) it is stated that "Note that the use of 'restricted' here refers to restrictions on the random vector in the (conditioning-type) stochastic definition of the skew distribution. It is not a restriction on the parameter space, and so a 'restricted' form of a skew distribution is not necessarily nested within its corresponding 'unrestricted' form". In Lee and McLachlan (2013b), it is cautioned in the last two lines of Page 431 that "It should be

* Corresponding author.
E-mail address: g.mclachlan@uq.edu.au (G.J. McLachlan).

http://dx.doi.org/10.1016/j.spl.2016.04.004
0167-7152/© 2016 Elsevier B.V. All rights reserved.

stressed that the rMST family and uMST family match only in the univariate case, and one cannot obtain (7) [the restricted distribution] from (9) [the unrestricted distribution] when $p > 1$".

We also note that McLachlan and Lee (2014) have obtained improved results for the unrestricted model over those reported in Azzalini et al. (2016) and in their earlier paper Azzalini et al. (2014) for the two real datasets that were analysed by them to form the basis of their claims on the relative superiority of the restricted and unrestricted models.

The deficiencies in these two models have been demonstrated in Lee and McLachlan (2014b, 2016). Briefly, the restricted distribution is limited essentially to modelling skewness concentrated in a single direction in the feature space. This is because it uses a univariate skewing function; that is, a single latent skewing variable is used in its convolution formulation. As a consequence, the realizations of the latent term used in the formulation of the model to represent skewness are confined to lie on a line in the p-dimensional feature space regardless of the value of p. This effectively means that the restricted distribution is limited to modelling skewness that is concentrated in a single direction in the feature space.

The unrestricted distribution on the other hand uses a multivariate skewing function with the feature-specific skewing variables that allow for skewness in the model taken to be uncorrelated. In its formulation, the p-dimensional vector of these skewing variables is premultiplied by a (diagonal) matrix of skewness parameters. The consequent net effect is that the feature-specific latent terms representing skewness in the model are uncorrelated. Thus it is designed to model data in which the skewness is in directions that are parallel to the axes of the features space.

Lee and McLachlan (2014b, 2016) have shown how a distribution belonging to the broader class, the CFUST class, can be fitted with essentially no additional computational effort than for the unrestricted distribution. The CFUST distribution, which includes the restricted and unrestricted distributions as special cases, can model skewness in different directions simultaneously since it uses an arbitrary matrix of skew parameters in its formulation.

With the availability of software for the fitting of mixtures of CFUST distributions (Lee and McLachlan, 2015), users now have the option for letting the data decide as to which model is appropriate for their particular dataset. Or they can fit all three models rMST, uMST, and CFUST (or mixtures of them), and make their own choice between the three using, say, an information-based criterion such as BIC.

Several of the claims in Azzalini et al. (2016) were made in their earlier paper Azzalini et al. (2014) and were commented on in McLachlan and Lee (2014).

2. Explanation of nomenclature for skew t-distributions

In an attempt to provide an automated approach to the clustering of flow cytometry data, Pyne et al. (2009) considered the fitting of mixtures of skew t-distributions that belonged to the family of skew t-distributions proposed by Sahu et al. (2003). Members of the latter family have the following convolution-type characterization. The $p \times 1$ random vector $\boldsymbol{Y}$ can be expressed as

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\Delta}|\boldsymbol{U}_0| + \boldsymbol{U}_1, \tag{1}$$

where

$$\begin{bmatrix} \boldsymbol{U}_0 \\ \boldsymbol{U}_1 \end{bmatrix} \sim N_{2p}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \frac{1}{w}\begin{bmatrix} \boldsymbol{I}_p & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{bmatrix}\right). \tag{2}$$

In the above, $\boldsymbol{\mu}$ is a p-dimensional vector, $\boldsymbol{\Delta}$ is a $p \times p$ diagonal matrix, $\boldsymbol{I}_p$ denotes the $p \times p$ identity matrix, $\boldsymbol{\Sigma}$ is a positive definite matrix, and $\mathbf{0}$ is a vector/matrix of zeros with appropriate dimensions. Also, w is the realization of the random variable W distributed as gamma$(\frac{\nu}{2}, \frac{\nu}{2})$, and $|\boldsymbol{U}_0|$ denotes the vector whose ith element is the magnitude of the ith element of the vector $\boldsymbol{U}_0$.

In order to simplify the application of the EM algorithm to fit mixtures of these skew t-distributions, Pyne et al. (2009) imposed the restriction

$$U_{01} = U_{02} = \cdots = U_{0p} \tag{3}$$

on the p latent skewing variables, where $U_{0i} = (\boldsymbol{U}_0)_i \; (i = 1, \ldots, p)$. This produces a distribution equivalent to the skew t-distribution formulated by Branco and Dey (2001) and Azzalini and Capitanio (2003) after reparameterization. Lee and McLachlan (2013d) termed this distribution the restricted multivariate skew t (rMST) distribution to distinguish it from the distribution proposed by Sahu et al. (2003). By default, the latter was referred to as the unrestricted multivariate skew t (uMST) distribution since it can be characterized without any restrictions on the p latent skewing variables in the convolution-type stochastic formulation (1).

By letting the degrees of freedom ν go to infinity in (1), we obtain a similar formulation for the restricted multivariate skew normal (rMSN) and unrestricted multivariate skew normal (uMSN) distributions. In the sequel we focus only on skew t-distributions since the situation is similar for skew normal distributions. One slight difference is that although the joint distribution of independent univariate skew normal random variables is the unrestricted skew normal distribution, the joint distribution of independent univariate skew t-random variables is only equal to the unrestricted skew t-distribution in the limit as the degrees of freedom ν in the marginal skew t-distributions becomes infinite.