Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Q1 On nomenclature for, and the relative merits of, two formulations of skew distributions

Q2 Adelchi Azzalini [a,*], Ryan P. Browne [b], Marc G. Genton [c], Paul D. McNicholas [d]

[a] Department of Statistical Sciences, University of Padua, 35121 Padova, Italy
[b] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1
[c] CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia
[d] Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada, L8S 4L8

ARTICLE INFO

Article history:
Received 30 November 2015
Received in revised form 3 December 2015
Accepted 3 December 2015
Available online xxxx

Keywords:
Flexibility
Model-based clustering
Multivariate distribution
Skew-normal distribution
Skew-*t* distribution

ABSTRACT

We examine some skew distributions used extensively within the model-based clustering literature in recent years, paying special attention to claims that have been made about their relative efficacy. Theoretical arguments are provided as well as real data examples.

1. Introduction

In recent years, much work in model-based clustering has replaced the traditional Gaussian assumption by some more flexible parametric family of distributions. In this context, (Lee and McLachlan, 2014), and other work following therefrom, utilize two formulations of the multivariate skew-normal (MSN) distribution as well as analogous formulations of the multivariate skew-*t* (MST) distribution for clustering, referring to these formulations as “restricted” and “unrestricted”, respectively. This nomenclature carries obvious implications and, rather than delving into semantics, it will suffice here to quote from Lee and McLachlan (2014, Section 2.2), who contend that “the unrestricted multivariate skew-normal (uMSN) distribution can be viewed as a simple extension of the rMSN distribution…” Here, rMSN denotes the “restricted” MSN distribution, and rMST and uMST are used similarly. The purpose of this note is to refute the claim that uMSN distribution is merely a simple extension of the rMSN distribution or, equivalently, the claim that uMST distribution is a simple extension of the rMST distribution. Furthermore, we investigate whether or not one formulation can reasonably be considered superior to the other.

2. Background

When one departs from the symmetry of the multivariate normal or other elliptical distributions, the feature that arises most readily is skewness. This explains the widespread use of the prefix ‘skew’ which recurs almost constantly in this

[*] Corresponding author.
E-mail address: adelchi.azzalini@unipd.it (A. Azzalini).

http://dx.doi.org/10.1016/j.spl.2015.12.008

context. A recent extensive account is provided by Azzalini (2014). This activity has generated an enormous number of formulations, sometimes arising with the same motivation and target, or nearly so. A natural question in these cases is which of the competing alternatives is preferable, either universally or for some given purpose. To be more specific, start by considering the multivariate skew-normal (SN) distribution proposed by Azzalini and Dalla Valle (1996), examined further by Azzalini and Capitanio (1999) and by much subsequent work. Note that, although the latter paper adopts a different parameterization of the earlier one, the set of distributions that they encompass is the same; we shall denote this construction as the classical skew-normal. Another form of skew-normal distribution has been studied by Sahu et al. (2003), which we shall refer to as the SDB skew-normal, by the initials of the author names. The classical and the SDB set of distributions coincide only for dimension $d = 1$; otherwise, the two sets differ and not simply because of different parameterizations. For $d > 1$, the question then arises about whether there is some relevant difference between the two formulations from the viewpoint of suitability for statistical work, both on the side of formal properties and on the side of practical analysis. This question is central to the present note because what we call the classical formulation is what Lee and McLachlan call rMSN, and the SDB formulation is their uMSN.

Analogous formulations arise when the normal family is replaced by the wider elliptical class in the underlying parent distribution, leading to the so-called skew-elliptical distributions. A special case that has received much attention is the skew-*t* family (Branco and Dey, 2001; Azzalini and Capitanio, 2003). Again, the classical skew-*t* has a counterpart given by another skew-*t* considered by Sahu et al. (2003), and the same questions as above hold. As before, what we call the classical formulation of the skew-*t* distribution is what Lee and McLachlan call rMST, and the SDB is their uMST.

Because of their role as the basic constituent for more elaborate formulations, we start by discussing the two forms of skew-normal distributions. The density and the distribution function of a $N_d(0, \Sigma)$ variable are denoted $\varphi_d(\cdot; \Sigma)$ and $\Phi_d(\cdot; \Sigma)$, respectively; the $N(0, 1)$ distribution function is denoted $\Phi(\cdot)$. The classical skew-normal density function is

$$f_c(x) = 2\,\varphi_d(x - \xi; \Omega)\Phi\{\alpha^\top \omega^{-1}(x - \xi)\}, \tag{1}$$

for $x \in \mathbb{R}^d$, with parameter set (ξ, Ω, α). Here ξ is a d-dimensional location parameter, Ω is a symmetric positive definite $d \times d$ scale matrix, α is a d-dimensional slant parameter, and ω is a diagonal matrix formed by the square roots of the diagonal elements of Ω. Various stochastic representations exist for (1). One is as follows: if

$$\begin{pmatrix} X_0 \\ X_1 \end{pmatrix} \sim N_{d+1}(0, \Omega^*), \qquad \Omega^* = \begin{pmatrix} \bar\Omega & \delta \\ \delta^\top & 1, \end{pmatrix}$$

where Ω^* is a correlation matrix, then

$$Y_c = \xi + \omega(X_0|X_1 > 0) \tag{2}$$

has distribution (1) with $\Omega = \omega\bar\Omega\omega$ and $\alpha = (1 - \delta^\top\bar\Omega^{-1}\delta)^{-1/2}\bar\Omega^{-1}\delta$. Here and in the following, given a random variable X and an event E, the notation $(X|E)$ denotes a random variable which has the distribution of X conditional on the event E; the Kolmogorov representation theorem ensures that such a random variable exists.

Another stochastic representation is the following: if δ is a d-vector with elements in $(-1, 1)$, then (1) is the density function of

$$Y_c = \xi + \omega\left\{[I_d - \mathrm{diag}(\delta)^2]^{1/2}\, V_0 + \delta|V_1|\right\}, \tag{3}$$

where V_0 and V_1 are independent normal variates of dimension d and 1, respectively, with 0 mean value, unit variances, and Q3 $\mathrm{cor}(V_0)$ is suitably related to α and Ω; full details are given on p. 128–9 of Azzalini (2014) among other sources. For the SDB skew-normal, we adopt a very minor change from the symbols of Sahu et al. (2003), but retain the same parameterization. Given real values $\lambda_1, \ldots, \lambda_d$, let $\lambda = (\lambda_1, \ldots, \lambda_d)^\top$ and $\Lambda = \mathrm{diag}(\lambda)$, and write the SDB density as

$$f_s(x) = 2^d\,\varphi_d(x - \xi; \Delta + \Lambda^2) \times \Phi_d\{\Lambda(\Delta + \Lambda^2)^{-1}(x - \xi); I_d - \Lambda(\Delta + \Lambda^2)^{-1}\Lambda\}, \tag{4}$$

where Δ is a symmetric positive-definite matrix. This density is associated with the following stochastic representation. For independent variables $\varepsilon \sim N_d(\xi, \Delta)$ and $Z \sim N_d(0, I_d)$, consider the transformation

$$Y_s = \Lambda(Z|Z > 0) + \varepsilon, \tag{5}$$

where $Z > 0$ means that the inequality is satisfied component-wise; then Y_s has density (4).

3. Comparing the formulations

A qualitative comparison of the formal properties of the two distributions lends several annotations. Some of these have already been presented by (Sahu et al., 2003), but they are included here for completeness.

1. The number of individual parameter values is $2d + d(d + 1)/2$ in both cases.
2. The two families of distributions coincide only for $d = 1$, as noted by (Sahu et al., 2003), and neither one is a subset of the other for $d > 1$.

Please cite this article in press as: Azzalini, A., et al., On nomenclature for, and the relative merits of, two formulations of skew distributions. Statistics and Probability Letters (2015), http://dx.doi.org/10.1016/j.spl.2015.12.008