Journal of Multivariate Analysis 159 (2017) 67–81

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Skew-rotationally-symmetric distributions and related efficient inferential procedures

Christophe Ley [a], Thomas Verdebout [b,*]

[a] Department of Applied Mathematics, Computer Science and Statistics, Ghent University (UGent), Belgium

[b] Département de Mathématique and ECARES, Université libre de Bruxelles (ULB), Belgium

ARTICLE INFO

Article history:
Received 22 September 2016
Available online 4 May 2017

AMS subject classifications:
62H11
62H15
62F05

Keywords:
Directional statistics
Rotationally symmetric distributions
Skew-symmetric distributions
Tests for rotational symmetry

ABSTRACT

Most commonly used distributions on the unit hypersphere $\mathcal{S}^{k-1} = \{\mathbf{v} \in \mathbb{R}^k : \mathbf{v}^\top \mathbf{v} = 1\}$, $k \geq 2$, assume that the data are rotationally symmetric about some direction $\boldsymbol{\theta} \in \mathcal{S}^{k-1}$. However, there is empirical evidence that this assumption often fails to describe reality. We study in this paper a new class of skew-rotationally-symmetric distributions on $\mathcal{S}^{k-1}$ that enjoy numerous good properties. We discuss the Fisher information structure of the model and derive efficient inferential procedures. In particular, we obtain the first semi-parametric test for rotational symmetry about a known direction. We also propose a second test for rotational symmetry, obtained through the definition of a new measure of skewness on the hypersphere. We investigate the finite-sample behavior of the new tests through a Monte Carlo simulation study. We conclude the paper with a discussion about some intriguing open questions related to our new models.

1. Introduction

Directional data $\mathbf{X}_1, \ldots, \mathbf{X}_n$ on unit spheres, or simply *spherical data*, are observations taking values on the non-linear manifold $\mathcal{S}^{k-1} = \{\mathbf{v} \in \mathbb{R}^k : \mathbf{v}^\top \mathbf{v} = 1\}$ for some integer $k \geq 2$. Over the past decade, there has been a strong surge of interest in directional statistics, thanks in part to the publication of cornerstone reference books [8,20] and the emergence of new applications in structural bioinformatics, genetics, cosmology and machine learning. Meanwhile, the use of spherical data continues to spread in more traditional fields such as paleomagnetism, meteorology or studies of animal behavior.

In the literature on spherical data, the distribution of the $\mathbf{X}_i$'s is commonly assumed to be rotationally symmetric about some location parameter $\boldsymbol{\theta} \in \mathcal{S}^{k-1}$. The probability density function (pdf) with respect to the usual surface area measure on spheres is then taken to be of the form

$$\mathbf{x} \mapsto f_{\boldsymbol{\theta};k}(\mathbf{x}) = c_{f,k} f(\mathbf{x}^\top \boldsymbol{\theta}), \quad \mathbf{x} \in \mathcal{S}^{k-1} \tag{1}$$

where the angular function $f : [-1, 1] \to \mathbb{R}^+$ is absolutely continuous and $c_{f,k}$ is a normalizing constant. The terminology "angular function" is closely related to rotational symmetry as it reflects the fact that the distribution of $\mathbf{X}_i$ only depends on the angle (colatitude angle in case $k = 3$) between $\boldsymbol{\theta}$ and $\mathbf{X}_i$ for each $i \in \{1, \ldots, n\}$. A classical example of such a distribution is the *Fisher–von Mises–Langevin (FvML)* distribution with density

$$\mathbf{x} \mapsto \left(\frac{\kappa}{2}\right)^{k/2-1} \frac{1}{2\pi^{k/2} I_{k/2-1}(\kappa)} \exp(\kappa \mathbf{x}^\top \boldsymbol{\theta}), \quad \mathbf{x} \in \mathcal{S}^{k-1}$$

where $\kappa > 0$ is a concentration parameter and $I_{k/2-1}$ the modified Bessel function of the first kind and of order $k/2 - 1$.

* Corresponding author.
E-mail address: tverdebo@ulb.ac.be (T. Verdebout).

http://dx.doi.org/10.1016/j.jmva.2017.02.010

In practice, however, not all real-life phenomena can be represented by symmetric models. For instance, Leong and Carlile [13] provide evidence that rotational symmetry is a too strong assumption in neurosciences while Mardia [19] explains that in bioinformatics, especially in protein structure prediction, data can be skewed. Motivated by these examples, we study in the present paper a spherical adaptation of the celebrated skew-symmetric distributions on $\mathbb{R}^k$. The vast research stream related to these distributions was initiated in the seminal paper [3] by Azzalini, who investigated the scalar skew-normal density $2\phi(x-\mu)\Phi\{\delta(x-\mu)\}$, $x, \mu, \delta \in \mathbb{R}$, with ϕ and Φ respectively the standard Gaussian density and distribution function. Here μ is a location parameter and δ is a skewness parameter. As an upshot of several generalization efforts, [4,27] proposed the aforementioned multivariate skew-symmetric distributions with pdf

$$2f_k(\mathbf{x}-\boldsymbol{\mu})\Pi_k(\mathbf{x}-\boldsymbol{\mu}, \boldsymbol{\delta}), \quad \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\delta} \in \mathbb{R}^k \tag{2}$$

where f_k is a centrally symmetric pdf (i.e., $f_k(-\mathbf{x}) = f_k(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^k$) and $\Pi_k : \mathbb{R}^k \times \mathbb{R}^k \to [0, 1]$ satisfies $\Pi_k(-\mathbf{x}, \boldsymbol{\delta}) + \Pi_k(\mathbf{x}, \boldsymbol{\delta}) = 1$ and $\Pi_k(\mathbf{x}, \mathbf{0}) = 1/2$ for all $\mathbf{x}, \boldsymbol{\delta} \in \mathbb{R}^k$. This multiplicative perturbation of symmetry enjoys numerous attractive features, including elegant random number generation procedures and a very simple normalizing constant. Our spherical adaptation will enjoy the same advantages. The usefulness of the perturbation approach on the sphere was already put forward by Jupp et al. [10], who recently provided a general analysis of this approach.

We also derive in this paper various results that are important when asymptotic inference within the new model is considered. More precisely, we show that (i) the Fisher information for this model is singular if and only if the kernel rotationally symmetric density is FvML and (ii) how to construct a new semi-parametric test for rotational symmetry about a fixed center $\boldsymbol{\theta}$ that will be optimal within the entire class of rotationally symmetric distributions. More precisely we show that the classical Watson score test (see [28]) for spherical location is locally and asymptotically optimal, in the Le Cam sense, for testing the null hypothesis of rotational symmetry against the proposed skew alternatives. To the best of our knowledge this result is the first to consider the problem of rotational symmetry from a semi-parametric angle. Moreover, the test we obtain is uniformly optimal against the new class of skew distributions on the sphere. We accompany this test by a natural competitor, obtained through the definition of a novel measure of skewness on the sphere. The derivation of its asymptotic distribution is of independent interest.

The paper is organized as follows. In Section 2 we formally define the skew distributions and discuss how our construction is linked to skew densities on the circle; we further establish a stochastic representation allowing us to generate data from our model. In Section 3 we provide the score functions for location and skewness and the corresponding Fisher information matrix; we also investigate the underlying Fisher singularity issue. Turning our attention towards inferential issues, we build in Section 4 the announced uniformly optimal semi-parametric test for rotational symmetry about a fixed location $\boldsymbol{\theta} \in \mathcal{S}^{k-1}$. In Section 5 we propose a measure of skewness on the sphere and derive from this measure another test for rotational symmetry. The finite-sample behavior of the new tests is investigated through a Monte Carlo simulation study in Section 6. We discuss interesting open questions related to our skew-rotationally-symmetric distributions in Section 7. Finally, Appendix A contains a crucial theoretical development required for the tests of Section 4, and Appendix B collects the technical proofs.

2. Skew-rotationally-symmetric distributions

As mentioned in the Introduction, we adapt to the spherical setting the skew-symmetric construction from $\mathbb{R}^k$, yielding the *skew-rotationally-symmetric (SRS)* distributions. Starting with a rotationally symmetric density (kernel) $f_{\boldsymbol{\theta};k}$ with central direction $\boldsymbol{\theta} \in \mathcal{S}^{k-1}$ as defined in (1), the idea of the construction consists in nesting $f_{\boldsymbol{\theta};k}$ into a larger family of distributions whose only rotationally symmetric member is the kernel $f_{\boldsymbol{\theta};k}$. Let $\boldsymbol{\Upsilon}_{\boldsymbol{\theta}}$ stand for a $k \times (k-1)$ semi-orthogonal matrix such that

$$\boldsymbol{\Upsilon}_{\boldsymbol{\theta}}\boldsymbol{\Upsilon}_{\boldsymbol{\theta}}^\top = \mathbf{I}_k - \boldsymbol{\theta}\boldsymbol{\theta}^\top \quad \text{and} \quad \boldsymbol{\Upsilon}_{\boldsymbol{\theta}}^\top\boldsymbol{\Upsilon}_{\boldsymbol{\theta}} = \mathbf{I}_{k-1},$$

where $\mathbf{I}_\ell$ is the $\ell \times \ell$ identity matrix. Consider a skewing function $\Pi : \mathbb{R} \to [0, 1]$, i.e., a monotone increasing continuous function satisfying $\Pi(-y) + \Pi(y) = 1$ for all $y \in \mathbb{R}$. Skewness is introduced by multiplication of the rotationally symmetric kernel with such a skewing function, turning $f_{\boldsymbol{\theta};k}$ into

$$\mathbf{x} \mapsto f_{\boldsymbol{\theta},\boldsymbol{\delta};k}(\mathbf{x}) = 2c_{f,k}f(\mathbf{x}^\top\boldsymbol{\theta})\Pi(\boldsymbol{\delta}^\top\boldsymbol{\Upsilon}_{\boldsymbol{\theta}}^\top\mathbf{x}), \quad \mathbf{x} \in \mathcal{S}^{k-1} \tag{3}$$

with $\boldsymbol{\delta} \in \mathbb{R}^{k-1}$. The following lemma shows that (3) is a density on $\mathcal{S}^{k-1}$.

Lemma 1. *Let $d\sigma_{k-1}(\mathbf{x})$ stand for the usual surface measure on $\mathcal{S}^{k-1}$. Then for all $\boldsymbol{\theta} \in \mathcal{S}^{k-1}$ and all $\boldsymbol{\delta} \in \mathbb{R}^{k-1}$,*

$$\int_{\mathcal{S}^{k-1}} f_{\boldsymbol{\theta},\boldsymbol{\delta};k}(\mathbf{x})\, d\sigma_{k-1}(\mathbf{x}) = 1.$$

A short proof of this result is given in Appendix B. This now allows us to formally define our family of skew distributions on $\mathcal{S}^{k-1}$.