Ann. Inst. Statist. Math.
Vol. 46, No. 3, 573–586 (1994)

DECENTERED DIRECTIONAL DATA*

BERNARD BOULERICE AND GILLES R. DUCHARME

Département de mathématiques et statistique, Université de Montréal, C. P. 6128, Succ. A, Montréal, Québec, Canada H3C 3J7

(Received March 29, 1993; revised September 28, 1993)

Abstract. Directional data analysis usually assumes that the observations are recorded according to a coordinate system whose origin coincides with the center of their support. However, it may happen that the observer does not sit at this center and record the directions from another point. The object of this paper is to investigate the statistical behavior of such decentered directions. First we derive the family of distributions of these directions and produce statistical procedures that recover some information about the underlying process. An important special case is explored in details and compared with the Langevin model. Finally, an example is given where the introduced family of models makes physical sense and well fits the observations.

Key words and phrases: Directional data, goodness-of-fit, group family, Langevin distribution, location shift, maximum likelihood theory, rotational symmetry, uniform distribution, von Mises Fisher distribution.

1. Introduction

Statistical procedures for directional data frequently assume that the data follows a von Mises-Fisher-Langevin distribution. This distribution has a density, with respect to the area element $d\omega_p$ on $S_p = \{\boldsymbol{x} \in \mathcal{R}^p \mid \boldsymbol{x}'\boldsymbol{x} = 1\}$, given by

$$f_{\kappa,\theta}(\boldsymbol{x}) = \frac{1}{a_p(\kappa)} e^{\kappa\theta'\boldsymbol{x}} \tag{1.1}$$

where $a_p(\kappa) = (2\pi)^{p/2}/\kappa^{p/2-1} I_{p/2-1}(\kappa)$ and $I_q(\cdot)$ denotes the modified Bessel function of the first kind of order q. The Langevin distribution, hereafter denoted $L_p(\theta, \kappa)$, is a member of the exponential family and turns out to be mathematically very tractable. It is rotationally symmetric about the axis $\theta \in S_p$ and ranges from the uniform distribution on S_p when $\kappa = 0$ to the point mass distribution at θ as $\kappa \to \infty$. These and other characterizations have made it as popular for directional data analysis as the normal distribution for Euclidean data. An additional

* This research was supported by the Natural Sciences and Engineering Research Council of Canada and by the FCAR and FRSQ funds of Québec.

reason for this popularity is that in many cases, the data can be assumed to arise from a rotationally symmetric distribution. But, as pointed out by Mardia *et al.* (1984), there is usually little ground *a priori* to justify the Langevin assumption. As a consequence, the Langevin model will often provide a crude approximation and thus suffer from lack of fit. In particular, experience shows that the distribution is often too "peaked". This problem is compounded by the relative scarceness of goodness of fit tests for this model against general alternatives (Fisher and Best (1984), Mardia *et al.* (1984), and also Rivest (1986)).

In this paper, we propose an approach generating families of models for directional data that can arise from the physical context of an experiment and, as such, may be more plausible. In Section 2, we develop the motivation and present the general form of the models. In Section 3, some properties of the distribution in an important subset of the family are given and compared to those of the Langevin model. Inference problems on this family are considered in Section 4 and Section 5 presents a goodness-of-fit test. Section 6 proposes a modification to Section's 4 model and we conclude with an example in Section 7.

2. Group families for decentered directional data

2.1 *The experimental context*

Directional data are usually recorded as the coordinates of points on a sphere whose origin coincides with the position of the observer. However, in some cases, experimental conditions can prevent this coincidence from happening. For example, asteroids are observed as they hit the atmosphere whose center is roughly located at the center of the earth. But, an observer will most likely be located on the surface of the earth and may record the impact according to a coordinate system where he sits at the origin. A more extreme case occurs when the observer is located outside the sphere as when, for example, an astronomer records impact point of asteroids on another celestial body.

In such experiments, the observer is only given to observe a directional vector whose origin has been shifted to a point $\boldsymbol{c}$. More precisely, instead of observing the datum $\boldsymbol{Y}$ as measured from the center of the sphere, the observer is given to observe

$$\boldsymbol{X} = \frac{\boldsymbol{Y} + \boldsymbol{c}}{\|\boldsymbol{Y} + \boldsymbol{c}\|}, \tag{2.1}$$

where $\|\cdot\|$ denotes the Euclidean norm and $\boldsymbol{c}$, the position of the center relative to the observer, is unknown. The problem then becomes that of recovering the information on $\boldsymbol{Y}$ from that given by $\boldsymbol{X}$. In this paper, we will restrict ourselves to the case where the center of support of the observations lies inside the unit sphere centered on the observer, that is $\boldsymbol{c} \in B_p = \{\boldsymbol{x} \in \mathcal{R}^p \mid \boldsymbol{x}'\boldsymbol{x} < 1\}$.

Note that other context can lead to model (2.1). One interesting case is the problem of detecting the unknown source $\boldsymbol{c}$ of a signal. At location $\boldsymbol{x}_i$, a detector picks up a direction $\boldsymbol{u}_i$ that points toward $\boldsymbol{c}$, except for random variation. For this problem, Jupp *et al.* (1987) propose a model that assumes $\boldsymbol{u}_i$ has mean direction $(\boldsymbol{x}_i - \boldsymbol{c})/\|\boldsymbol{x}_i - \boldsymbol{c}\|$ with the $\boldsymbol{x}_i$ fixed. When the detectors are (roughly) located on

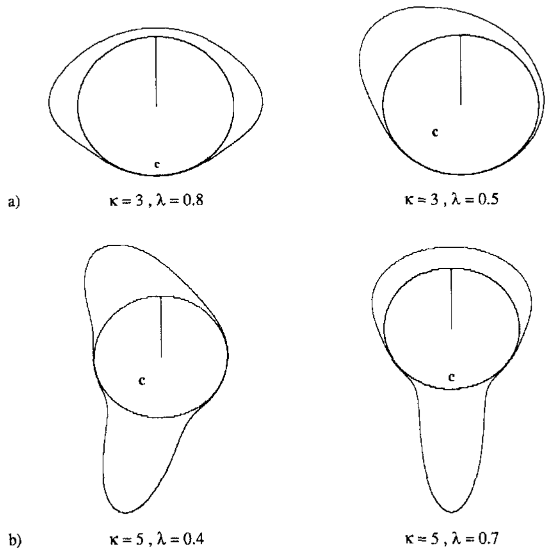

Fig. 1. a) Decentered von Mises $L_2(3,(0,1))$, b) Decentered Dimroth-Watson on the circle with concentration $\kappa = 5$ about the axis $(0,1)$ for various values of $\boldsymbol{c}$.

a circle covering $\boldsymbol{c}$ and the source emits randomly (e.g. radioactive material), the locations $\boldsymbol{x}_i$ corresponding to those detectors that have been triggered, become random directional vectors. Thus, (2.1) can be looked at as a mixture model from Jupp *et al.*'s (1987) proposal.

2.2 *General models*

Suppose that the unobserved directional vector $\boldsymbol{Y}$ possesses a density g_μ, $\mu \in \Omega$, with respect to $d\omega_p$. Then the direction $\boldsymbol{X}$ given by (2.1) has a density given by

$$f_\mu(\boldsymbol{x};\boldsymbol{c}) = \psi(\boldsymbol{x};\boldsymbol{c}) g_\mu(\boldsymbol{z}(\boldsymbol{x};\boldsymbol{c})) d\omega_p \tag{2.2}$$

where

$$\boldsymbol{z}(\boldsymbol{x};\boldsymbol{c}) = \boldsymbol{x}\left[\boldsymbol{c}'\boldsymbol{x} + \sqrt{1 - \|\boldsymbol{c}\|^2 + (\boldsymbol{c}'\boldsymbol{x})^2}\right] - \boldsymbol{c} \tag{2.3}$$

is the inverse of (2.1) and $\psi(\boldsymbol{x};\boldsymbol{c})$ denotes its Jacobian on S_p, which may be obtained through standard methods of differential geometry (see Do Carmo (1976)) and is given by

$$\psi(\boldsymbol{x};\boldsymbol{c}) = \frac{\left[\boldsymbol{c}'\boldsymbol{x} + \sqrt{1 - \|\boldsymbol{c}\|^2 + (\boldsymbol{c}'\boldsymbol{x})^2}\right]^{p-1}}{\sqrt{1 - \|\boldsymbol{c}\|^2 + (\boldsymbol{c}'\boldsymbol{x})^2}}. \tag{2.4}$$

Thus, from the family $\{g_\mu, \mu \in \Omega\}$, transformation (2.1) induces the group family

$$\{f_\mu(\cdot;\boldsymbol{c}), \mu \in \Omega, \boldsymbol{c} \in B_p\} \tag{2.5}$$

that we will refer to as a decentered family.

This family encompasses a much larger variety of distribution types than the original family $\{g_\mu, \mu \in \Omega\}$. For example, if g_μ is rotationally symmetric about the axis μ, the resulting family (2.5) contains unimodal, bimodal and girdle type distributions which need not be rotationally symmetric, depending on the location of $\boldsymbol{c}$ in B_p. Note that rotational symmetry is retained only when $\boldsymbol{c}$ lies along the axis of μ. Figure 1 shows the densities that can be obtained from a von Mises and Dimroth-Watson distribution on the circle, for a few values of $\boldsymbol{c}$.

Thus besides its justification arising from physical considerations, transformation (2.1) and the resulting family of distributions (2.5) offer a new approach to modeling complex patterns of data on S_p.

3. Directional Decentered Uniform (DCU) model

In this and the next two sections, we will suppose that $\boldsymbol{Y}$ is uniformly distributed on S_p such that $g_\mu = 1/\omega_p$, where $\omega_p = 2\pi^{p/2}/\Gamma(p/2)$. Writing $\boldsymbol{c} = \lambda\theta$ with $\theta \in S_p$ and $\lambda \in [0, 1[$, the density of $\boldsymbol{X}$ in (2.2) may be written

$$f(\boldsymbol{x}; \lambda, \theta) = \frac{1}{\omega_p} \frac{\left[\lambda\theta'\boldsymbol{x} + \sqrt{1 - \lambda^2\{1 - (\theta'\boldsymbol{x})^2\}}\right]^{p-1}}{\sqrt{1 - \lambda^2\{1 - (\theta'\boldsymbol{x})^2\}}} \tag{3.1}$$

from which we see that it is unimodal and rotationally symmetric about its mode θ. Now, from Watson (1983), we find that for such rotationally symmetric distributions, $E(\boldsymbol{X}) = E(T)\theta$ and

$$\mathrm{Var}(\boldsymbol{X}) = \mathrm{Var}(T)\theta\theta' + \frac{1 - E(T^2)}{p - 1}(\boldsymbol{I}_p - \theta\theta'), \tag{3.2}$$

where $T = \theta'\boldsymbol{X}$. Moreover, the density of T, $h_p(t; \lambda)$ say, supported on $[-1, 1]$, depends only on λ and has the form

$$h_p(t; \lambda) = \frac{\omega_{p-1}}{\omega_p} \frac{\left[\lambda t + \sqrt{1 - \lambda^2(1 - t^2)}\right]^{p-1}}{\sqrt{1 - \lambda^2(1 - t^2)}} (1 - t^2)^{(p-3)/2}. \tag{3.3}$$

Expanding the numerator with Newton's binomial formula then integrating term by term using the integral formula (15.3.1) of Abramowitz and Stegun (1965), we find after tedious algebra that

$$E[T^k] = \frac{\Gamma\left(\frac{p}{2}\right)}{\sqrt{\pi}} \sum_{j=0}^{p} \binom{p-1}{j} \frac{\Gamma\left(\frac{j+k+1}{2}\right)}{\Gamma\left(\frac{p+j+k}{2}\right)} \lambda^j \times {}_2F_1\left(-\frac{(p-j-2)}{2}, \frac{p-1}{2}, \frac{p+j+k}{2}, \lambda^2\right) I_e(j+k) \tag{3.4}$$

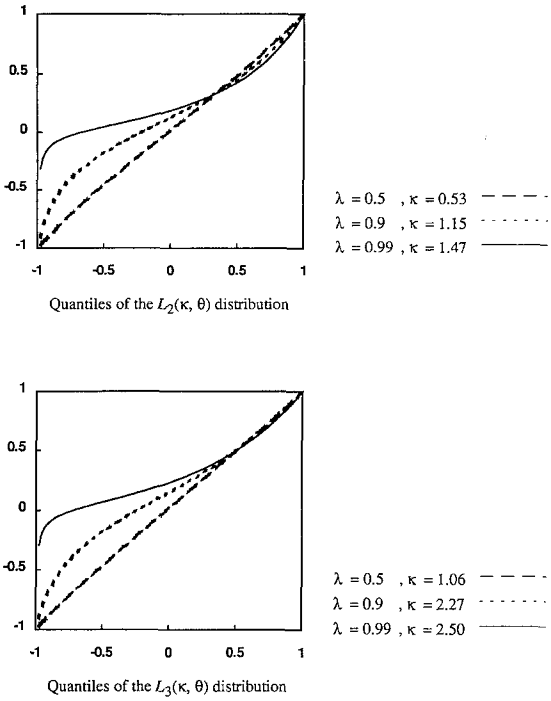

Fig. 2. Q-Q plots of the distribution of T for the DCU distribution versus the nearest Langevin distribution in sup-norm.

where $I_e(i) = 1$ if i is even, 0 otherwise, and ${}_2F_1$ denotes Gauss's hypergeometric function (Abramowitz and Stegun (1965), Chapter 15).

The behavior of $\boldsymbol{X}$ as $\lambda \to 0$ becomes that of the uniform on S_p, a behavior also shared by the Langevin distribution $L_p(\kappa, \theta)$ as $\kappa \to 0$. On the other hand, as $\lambda \to 1$, $h_p(t; \lambda)$ puts all its masses on $[0, 1]$ so that the distribution of $\boldsymbol{X}$ becomes concentrated on the half-sphere centered at θ. This behavior is markedly different from that of the Langevin distribution which converges, as $\kappa \to +\infty$, to the point mass distribution at θ. Thus the DCU distribution (3.1) cannot be used to model very concentrated samples and the behavior of the two distributions will differ most for large values of λ. Figure 2 presents for various values of λ and p, Q-Q plots of (3.3) versus the distribution of T when $\boldsymbol{X}$ follows a Langevin distribution with κ being chosen to minimize the sup-norm distance between the two distribution functions. One can see from these figures that (3.1) is much less concentrated than the Langevin distribution when $\lambda > 0.5$ and may provide a better fit to data with large dispersion. On the other hand, when $\lambda \leq 0.5$ the Langevin model will provide an adequate fit to (3.1) for most practical purposes.

4. Inference in the DCU model

Let $\boldsymbol{A}_i = \boldsymbol{I}_p - \boldsymbol{X}_i\boldsymbol{X}_i'$. Differentiation of the likelihood function with respect to $\boldsymbol{c}$ yields the likelihood system

$$(4.1)\qquad (p-1)\sum_{i=1}^{n}\frac{\left(\boldsymbol{X}_i - \dfrac{\boldsymbol{A}_i\boldsymbol{c}}{\sqrt{1-\boldsymbol{c}'\boldsymbol{A}_i\boldsymbol{c}}}\right)}{\boldsymbol{c}'\boldsymbol{X}_i + \sqrt{1-\boldsymbol{c}'\boldsymbol{A}_i\boldsymbol{c}}} + \sum_{i=1}^{n}\frac{\boldsymbol{A}_i\boldsymbol{c}}{1-\boldsymbol{c}'\boldsymbol{A}_i\boldsymbol{c}} = 0$$

which must be solved in $\boldsymbol{c}$ by iterations. Let $\hat{\boldsymbol{c}}$ denotes the solution to (4.1). From standard asymptotic results on the behavior of MLE's (see Lehmann (1983), Theorem 4.1),

$$(4.2)\qquad n^{1/2}(\hat{\boldsymbol{c}} - \boldsymbol{c}) \xrightarrow[n\to\infty]{\mathcal{L}} N_p\{0, \mathcal{I}^{-1}(\boldsymbol{c})\}$$

where $\mathcal{I}(\boldsymbol{c})$ stands for Fisher's information matrix. It is shown in the Appendix A.1 that

$$(4.3)\qquad \mathcal{I}^{-1}(\boldsymbol{c}) = c_{1,p}^{-1}(\lambda)\theta\theta' + (p-1)c_{2,p}^{-1}(\lambda)(\boldsymbol{I}_p - \theta\theta')$$

with

$$(4.4)\quad c_{1,p}(\lambda) = \frac{(p-1)}{3p(p+2)(1-\lambda^2)} \times \Big\{\lambda^2(p-2)(p+1)(2\lambda^2+p-1)\,{}_2F_1\left(2,\frac{1}{2},\frac{p+4}{2},\lambda^2\right) - (p+2)(2\lambda^4(p-2)+\lambda^2(p-2)(p+1)-3(p-1)) \times {}_2F_1\left(2,\frac{1}{2},\frac{p+2}{2},\lambda^2\right)\Big\}$$

and

$$(4.5)\quad c_{2,p}(\lambda) = \frac{(p-1)}{3p(p+2)} \times \Big\{(p+2)(3(p-1)^2-\lambda^2(p-2))\,{}_2F_1\left(2,\frac{1}{2},\frac{p+2}{2},\lambda^2\right) - \lambda^2(p-2)(\lambda^2+2(p-1))\,{}_2F_1\left(2,\frac{3}{2},\frac{p+4}{2},\lambda^2\right)\Big\}.$$

In particular if $p=2$ we have $c_{1,2}(\lambda) = (1-\lambda^2)^{-3/2}/2$ and $c_{2,2}(\lambda) = (1-\lambda^2)^{-1/2}/2$. Now, the MLE's of $\lambda = \|\boldsymbol{c}\|$ and $\theta = \boldsymbol{c}/\|\boldsymbol{c}\|$ are $\hat{\lambda} = \|\hat{\boldsymbol{c}}\|$ and $\hat{\theta} = \hat{\boldsymbol{c}}/\|\hat{\boldsymbol{c}}\|$ respectively. Thus from (4.2) and the delta theorem we get when $\boldsymbol{c} \neq \boldsymbol{0}$,

$$(4.6)\qquad n^{1/2}(\hat{\theta} - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} N_p\left\{\boldsymbol{0}, \frac{(p-1)}{\lambda^2 c_{2,p}(\lambda)}(\boldsymbol{I}_p - \theta\theta')\right\}.$$

Note that $c_{1,p}^{-1}(\lambda)$ is positive and bounded away from zero for each $\lambda \in [0,1[$ whereas $(p-1)/(\lambda^2 c_{2,p}(\lambda))$ is positive and tends to $+\infty$ as $\lambda \to 0$. Thus when

$\lambda = 0$ (i.e. $\boldsymbol{c} = \mathbf{0}$), $\boldsymbol{X}$ has a uniform distribution over S_p and a test of the null hypothesis $H_0 : \boldsymbol{c} = \mathbf{0}$ against $H_1 : \boldsymbol{c} \in B_p$ can be constructed using (4.2). We get that under the null hypothesis $n\hat{\boldsymbol{c}}'\mathcal{I}(\mathbf{0})\hat{\boldsymbol{c}} = n(p-1)^2\hat{\lambda}^2/p$ is asymptotically distributed as a χ^2_p.

5. Goodness-of-fit test

In this section, we propose a goodness-of-fit test for the DCU model. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a sample of directional vector on S_p. We are concerned with testing whether or not the sample has density (3.1) with $\boldsymbol{c}$ unknown. We will consider here alternatives of the form (2.2) with g_μ being the density of a Langevin distribution with unknown parameters. Other cases may be considered but this choice is motivated by the fact that the ensuing family of models contains the null distribution when $\kappa = 0$, and covers a wide range of density types (see Subsection 2.1).

It is shown in the Appendix A.2 that, in this case, Rao's score test statistic (Rao (1973), p. 418) depends on the mean resultant length of the transformed directions

$$\hat{\boldsymbol{Y}}_i = \boldsymbol{X}_i\left[\hat{\boldsymbol{c}}'\boldsymbol{X}_i + \sqrt{1 - \|\hat{\boldsymbol{c}}\|^2 + (\hat{\boldsymbol{c}}'\boldsymbol{X}_i)^2}\right] - \hat{\boldsymbol{c}}, \qquad i = 1, \ldots, n \tag{5.1}$$

where $\hat{\boldsymbol{c}}$ denotes the MLE of $\boldsymbol{c}$. Let $\hat{\bar{\boldsymbol{Y}}} = \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{Y}}_i$. It is also shown in the Appendix A.2 that under the null hypothesis

$$n^{1/2}\hat{\bar{\boldsymbol{Y}}} \xrightarrow[n\to\infty]{\mathcal{L}} N_p\{\mathbf{0}, \Sigma(\boldsymbol{c})\} \tag{5.2}$$

with

$$\Sigma(\boldsymbol{c}) = \sigma_{1,p}(\lambda)\theta\theta' + \sigma_{2,p}(\lambda)(\boldsymbol{I}_p - \theta\theta') \tag{5.3}$$

and

$$\sigma_{1,p}(\lambda) = \frac{1}{p} - \frac{(\rho_{1,p}(\lambda) - 1)^2}{c_{1,p}(\lambda)} \tag{5.4}$$

with $\rho_{1,p}(\lambda) = ((1-\lambda^2)/p)_2F_1(1, 3/2, (p+2)/2, \lambda^2)$, while

$$\sigma_{2,p}(\lambda) = \frac{1}{p} - \frac{(\rho_{2,p}(\lambda) - (p-1))^2}{(p-1)c_{2,p}(\lambda)} \tag{5.5}$$

with $\rho_{2,p}(\lambda) = ((p-1)/p)_2F_1(1, 1/2, (p+2)/2, \lambda^2)$.

Thus the test reject the null hypothesis for large values of the test statistic

$$\hat{R}^2 = n\,\hat{\bar{\boldsymbol{Y}}}'\Sigma^{-1}(\hat{\boldsymbol{c}})\,\hat{\bar{\boldsymbol{Y}}} \tag{5.6}$$

which is asymptotically distributed as a χ^2_p under the null model.

Note that $\Sigma^{-1}(\hat{\boldsymbol{c}})$ in (5.6) can be replaced with any other consistent estimate of $\Sigma^{-1}(\boldsymbol{c})$ without affecting its asymptotic behavior. Note also that when the distribution of the sample is actually the null model, the transformed data (5.1) should be nearly uniformly distributed on S_p. This can be used to make a graphical assessment of the goodness of fit of the DCU model (see Section 7).

6. Modified DCU model (MDCU)

The DCU model is not appropriate in situations where the underlying phenomenon generates directional data that are restricted to a subset of the sphere. This set-up is common in the analysis of astrophysical data (Mardia and Edwards (1982)) where it may be of interest to verify whether the datum $\boldsymbol{Y}$ in (2.1) comes from a uniform distribution over a given subset of the sphere.

Consider the case where $\boldsymbol{Y}$ is uniformly distributed on the girdle of S_p defined by $G_p = \{\boldsymbol{x} \in S_p \mid a < \nu'\boldsymbol{x} < b\}$ for some axis ν. Then $\boldsymbol{X}$ in (2.1) is distributed according to what we call a modified DCU model (MDCU) that has density

$$f(\boldsymbol{x}; \boldsymbol{c}, \nu, a, b) = \frac{1}{c_p(a,b)} \psi(\boldsymbol{x}; \boldsymbol{c}) I(a < \nu' \boldsymbol{z}(\boldsymbol{x}; \boldsymbol{c}) < b) d\omega_p \tag{6.1}$$

with $c_p(a,b) = \omega_{p-1} \int_{\arcsin(a)}^{\arcsin(b)} \cos(\theta)^{p-2} d\theta$ and $I(\cdot)$ denotes the indicator function and $\boldsymbol{z}(\cdot; \boldsymbol{c})$ is defined in (2.3).

We will assume a, b and ν known, so that the MLE of $\boldsymbol{c}$ is still taken as a solution to (4.1), provided the resulting transformed directions of (5.1) all lie within G_p. In this case, we have that

$$n^{1/2}(\hat{\boldsymbol{c}} - \boldsymbol{c}) \xrightarrow[n\to\infty]{\mathcal{L}} N_p\{\boldsymbol{0}, \boldsymbol{D}^{-1}(\boldsymbol{c})\} \tag{6.2}$$

where $\boldsymbol{D}(\boldsymbol{c})$ is given in the Appendix A.3 and when the MDCU model holds,

$$n^{1/2} \hat{\bar{\boldsymbol{Y}}} \xrightarrow[n\to\infty]{\mathcal{L}} N_p\{\boldsymbol{0}, \Lambda(\boldsymbol{c})\}. \tag{6.3}$$

Thus the goodness-of-fit test developed in Section 5 can still be used, provided we supply a consistent estimator, $\tilde{\Lambda}$ say, of $\Lambda(\boldsymbol{c})$ under the MDCU model. This is done in the Appendix A.3. Finally we obtain the Rao's score statistic, $\tilde{R}^2 = n \hat{\bar{\boldsymbol{Y}}}' \tilde{\Lambda}^{-1} \hat{\bar{\boldsymbol{Y}}}$, which is to be compared to the appropriate quantile of a χ_p^2 distribution.

7. Example

Fisher *et al.* (1987) (data set B3) have digitized the arrival directions of 148 low mu showers of cosmic rays (Bolivian Air Shower Joint Experiment) from Toyoda *et al.* (1965). A simple analysis of this data set leads the first authors to accept the hypothesis of rotational symmetry. Moreover an equal-area projection of the data directions, Fig. 3(a), suggests that the distribution is of girdle type.

On the other hand, Toyoda *et al.* (1965) and also Kiraly *et al.* (1979) noticed that in the range of energy of the particles considered here, these rays may be supposed of a galactic origin. The fact that the galactic plane does not cover the sphere totally, as we see it in the equatorial plane, must be taken into account. Indeed, from astronomical considerations described in Kiraly *et al.* (1979), we may be brought to hypothesize that the original vector $\boldsymbol{Y}$ is uniformly distributed over the subset of the sphere not containing declinations above 65° and below −65°.

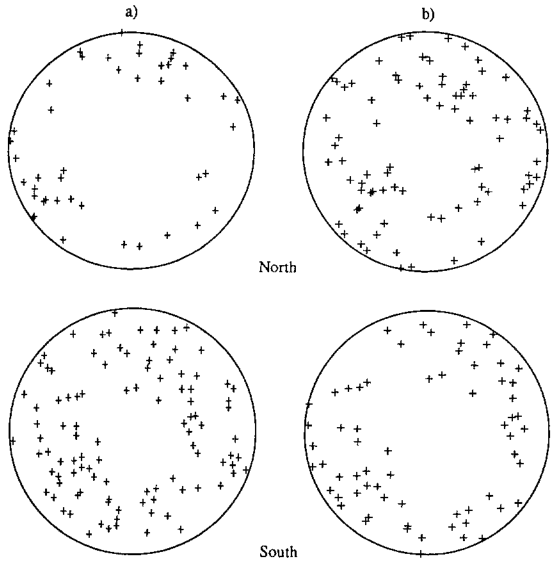

Fig. 3. Equal-area projection of a) the original data and b) the transformed data (5.1) along the polar axis for the two hemispheres.

That is, from the physical context of the experiment, we may want to test whether the data are distributed according to a MDCU model with $b = -a = \sin(65°)$ and $\nu = (0, 0, 1)'$.

For this data set, the solution to (4.1) is $\hat{\boldsymbol{c}} = (-0.116 \ 0.026 \ -0.414)'$ with $\|\hat{\boldsymbol{c}}\| = 0.43$. The resulting transformed directions in (5.1) all lie in G_p. Thus, $\hat{\boldsymbol{c}}$ is the MLE of $\boldsymbol{c}$ which, using (6.2), can be declared significantly different from $\mathbf{0}$ since $n\hat{\boldsymbol{c}}'\boldsymbol{D}(\mathbf{0})\hat{\boldsymbol{c}} = 30.85 > \chi_3^2(1\%) = 11.35$. If our model is correct, the transformed data in (5.1) should be nearly uniformly distributed on G_p, which seems plausible in view of an equal-area projection of the transformed points, Fig. 3(b). However, a Q-Q plot (Fig. 4) suggests that there remains a little excess of observations in a direction that correspond to the Virgo cluster. This fact was also noticed by Toyoda *et al.* (1965). We require a test to conclude if this excess is statistically significant. The goodness-of-fit test of Section 6 can be used. We obtain the estimate

$$\tilde{\Lambda} = \begin{bmatrix} 0.040 & -0.026 & 0.021 \\ -0.026 & 0.045 & -0.037 \\ 0.021 & -0.037 & 0.152 \end{bmatrix}$$

so that the modified Rao's score statistic $\tilde{R}^2 = 5.875 < \chi_3^2(5\%)$. Thus we do not reject the hypothesis of a MDCU model at the 5% level.

An alternative approach for modeling the arrival directions of cosmic rays has been proposed by Mardia and Edwards (1982). They suggest the use of weighted distributions assuming uniform rotation of an off-center cap along a line of constant colatitude. Applying their model (3.2) to this data set leads to a likelihood conditional equation that numerically yields two distinct sets of MLE. One of them

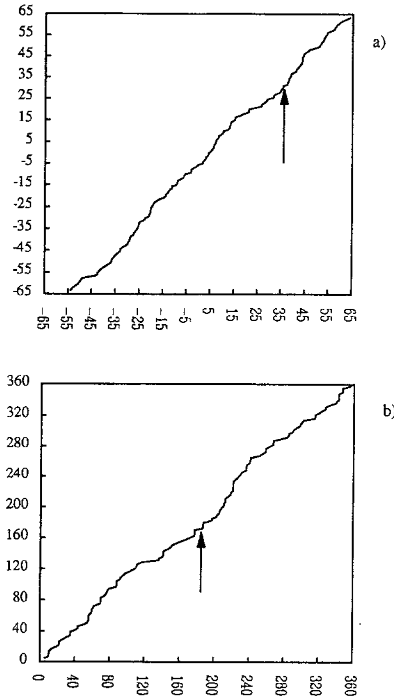

Fig. 4. *Q-Q* plots of the quantiles for the transformed data (5.1) versus the corresponding theoretical quantiles of the MDCU model in a) declination and b) right ascension. (The arrow indicates the position of the Virgo cluster in the new coordinates.)

corresponds to a girdle distribution with axis (82.16°, 44.84°) in colatitude and right ascension, while the other is bipolar about the axis (96.97°, 133.07°) which is almost perpendicular to the first one. The value of the (conditional) likelihood ratio statistic for testing their null hypothesis of a "weighted" uniform distribution is 6.12 compared with $\chi^2_3(5\%) = 7.81$. Here, the null hypothesis essentially states that the longitudes are uniformly distributed. The fact that we do not reject this hypothesis agrees with Toyoda *et al.*'s (1965) empirical finding. It is also coherent with the goodness of fit result of the MDCU model. Indeed $\hat{c}$ being very close to the south pole axis, the longitudes of the data points transformed through (5.1) are very close to the longitudes of the original data. It is therefore not surprising to find that the original longitudes are uniform since the MDCU model calls for uniformity of the transformed longitudes. The MDCU model is appealing here since the behavior of the colatitudes is not arbitrary and naturally arises from the context of the experiment.

Acknowledgements

We wish to thank two anonymous referees for their constructive suggestions, which have greatly improved the paper.

Appendix

A.1 *Computation of the Fisher's information matrix for the DCU model*

Let $\boldsymbol{L}(\boldsymbol{x};\boldsymbol{c}) = [\partial/\partial c_i \log(f(\boldsymbol{x};\boldsymbol{c}))]_{p\times 1}$ with $f(\boldsymbol{x};\boldsymbol{c})$ as in (3.1). It is easy to verify that $E[\boldsymbol{L}(\boldsymbol{X};\boldsymbol{c})] = \boldsymbol{0}$ for each $\boldsymbol{c}$ in B_p, and then $\mathcal{I}(\boldsymbol{c}) = E[\boldsymbol{L}(\boldsymbol{X};\boldsymbol{c})\boldsymbol{L}(\boldsymbol{X};\boldsymbol{c})']$. But $\boldsymbol{X}$ has the same distribution as $(\boldsymbol{Y}+\boldsymbol{c})/\|\boldsymbol{Y}+\boldsymbol{c}\|$, when $\boldsymbol{Y}$ is uniformly distributed on S_p. Making this substitution in $\boldsymbol{L}(\boldsymbol{X};\boldsymbol{c})$ we obtain

$$\begin{aligned}
\text{(A.1.1)}\qquad & \boldsymbol{L}((\boldsymbol{Y}+\boldsymbol{c})/\|\boldsymbol{Y}+\boldsymbol{c}\|;\boldsymbol{c}) \\
&= (p-1)\frac{\boldsymbol{Y}}{(1+\boldsymbol{c}'\boldsymbol{Y})} - \frac{(\boldsymbol{c}'\boldsymbol{Y}+\boldsymbol{c}'\boldsymbol{c})}{(1+\boldsymbol{c}'\boldsymbol{Y})^2}\boldsymbol{Y} + \frac{\boldsymbol{c}}{(1+\boldsymbol{c}'\boldsymbol{Y})} \\
&= \frac{\{(p-2)(1+\boldsymbol{c}'\boldsymbol{Y}) + (1-\boldsymbol{c}'\boldsymbol{c})\}}{(1+\boldsymbol{c}'\boldsymbol{Y})^2}\boldsymbol{Y} + \frac{\boldsymbol{c}}{(1+\boldsymbol{c}'\boldsymbol{Y})}.
\end{aligned}$$

If $\boldsymbol{c} = \lambda\theta$, with $\lambda \in [0,1[$ and $\theta \in S_p$, then $U = \boldsymbol{Y}'\theta$ has density given in Watson ((1983), p. 45). Hence writing $\boldsymbol{Y} = U\theta + (1-U^2)^{1/2}\xi_{p-1}$ with ξ_{p-1} being uniformly distributed over the sphere orthogonal to θ and independent of U, we get

$$\begin{aligned}
\text{(A.1.2)}\quad \mathcal{I}(\boldsymbol{c}) &= E[\{(p-2)(1+\lambda U) + (1-\lambda^2)\}^2 U^2/(1+\lambda U)^4]\theta\theta' \\
&\quad + E[\{(p-2)(1+\lambda U) + (1-\lambda^2)\}^2 (1-U^2)/(1+\lambda U)^4] \\
&\quad \times E[\xi_{p-1}\xi_{p-1}'].
\end{aligned}$$

But

$$\begin{aligned}
\text{(A.1.3)}\qquad & E[\{(p-2)(1+\lambda U) + (1-\lambda^2)\}^2 U^2/(1+\lambda U)^4] \\
&= (p-2)^2 E[U^2/(1+\lambda U)^2] \\
&\quad + 2(p-2)(1-\lambda^2)E[U^2/(1+\lambda U)^3] \\
&\quad + (1-\lambda^2)^2 E[U^2/(1+\lambda U)^4].
\end{aligned}$$

Using the integral formula for the Gauss hypergeometric function (Abramowitz and Stegun (1965)) we obtain

$$\begin{aligned}
\text{(A.1.4)}\qquad & E[U^2/(1+\lambda U)^2] \\
&= \frac{\omega_{p-1}}{\omega_p}\int_{-1}^{1} u^2(1-u^2)^{(p-3)/2}/(1+\lambda u)^2 du \\
&= \frac{\omega_{p-1}}{\omega_p} 2\int_{0}^{1} (1+\lambda^2u^2)u^2(1-u^2)^{(p-3)/2}/(1-\lambda^2u^2)^2 du \\
&= \frac{1}{p}\,{}_2F_1(2, 3/2, (p+2)/2, \lambda^2) \\
&\quad + \frac{3\lambda^2}{p(p+2)}\,{}_2F_1(2, 5/2, (p+4)/2, \lambda^2).
\end{aligned}$$

The same approach yields the other terms. Then, using identities connecting the hypergeometric functions (Abramowitz and Stegun (1965), Chapter 15), the simplified expressions $c_{1,p}(\cdot)$ and $c_{2,p}(\cdot)$ in (4.4) and (4.5) follow.

A.2 *Rao's score test for the DCU model*

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a sample of random directions having density of the form (2.2) with $g_\mu(\boldsymbol{y}) = e^{\mu' \boldsymbol{y}}/a_p(\|\mu\|)$, the Langevin density with parameter $\mu \in \mathcal{R}^p$. Then the null hypothesis of Section 5 is just $\mu = \mathbf{0}$. Let

$$\text{(A.2.1)} \qquad \begin{bmatrix} \boldsymbol{U}_c(\boldsymbol{X}) \\ \boldsymbol{U}_\mu(\boldsymbol{X}) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial}{\partial \boldsymbol{c}} \log(f_\mu(\boldsymbol{X}; \boldsymbol{c})) \\ \dfrac{\partial}{\partial \mu} \log(f_\mu(\boldsymbol{X}; \boldsymbol{c})) \end{bmatrix}.$$

Then Rao's score statistic (Rao (1973)) for testing $\mu = \mathbf{0}$ against the composite hypothesis $\mu \in \mathcal{R}^p$ is given by $\hat{R}^2 = \boldsymbol{V}^{*\prime}\mathcal{I}^{*-1}\boldsymbol{V}^*$, where

$$\text{(A.2.2)} \qquad \boldsymbol{V}^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{bmatrix} \boldsymbol{U}_c(\boldsymbol{X}_i) \\ \boldsymbol{U}_\mu(\boldsymbol{X}_i) \end{bmatrix} \Bigg|_{\mu=\mathbf{0}, \boldsymbol{c}=\hat{\boldsymbol{c}}} = \begin{bmatrix} \mathbf{0} \\ \sqrt{n}\, \hat{\bar{\boldsymbol{Y}}} \end{bmatrix},$$

and

$$\text{(A.2.3)} \qquad \mathcal{I}^* = E \begin{bmatrix} \boldsymbol{U}_c(\boldsymbol{X}_1)\boldsymbol{U}_c(\boldsymbol{X}_1)' & \boldsymbol{U}_c(\boldsymbol{X}_1)\boldsymbol{U}_\mu(\boldsymbol{X}_1)' \\ \boldsymbol{U}_c(\boldsymbol{X}_1)\boldsymbol{U}_\mu(\boldsymbol{X}_1)' & \boldsymbol{U}_\mu(\boldsymbol{X}_1)\boldsymbol{U}_\mu(\boldsymbol{X}_1)' \end{bmatrix} \Bigg|_{\mu=\mathbf{0}, \boldsymbol{c}=\hat{\boldsymbol{c}}}$$

$$= \begin{bmatrix} \mathcal{I}(\hat{\boldsymbol{c}}) & \Sigma_{12}(\hat{\boldsymbol{c}}) \\ \Sigma_{12}(\hat{\boldsymbol{c}}) & \dfrac{1}{p}\boldsymbol{I}_p \end{bmatrix},$$

with $\mathcal{I}(\boldsymbol{c})$ given by (4.3). A little algebra yields

$$\text{(A.2.4)} \qquad \Sigma_{12}(\boldsymbol{c}) = (\rho_{1,p}(\lambda) - 1)\theta\theta' + \frac{(\rho_{2,p}(\lambda) - (p-1))}{(p-1)}(\boldsymbol{I}_p - \theta\theta')$$

where $\rho_{1,p}(\lambda)$ and $\rho_{2,p}(\lambda)$ are given in (5.4) and (5.5). The results of Section 5 then follow easily.

A.3 *Consistent estimation of $\Lambda(\boldsymbol{c})$ for the modified DCU model*

Let $\{\boldsymbol{Y}_i, i = 1, \ldots, n\}$ be independent and uniformly distributed over G_p, so the $\boldsymbol{X}_i = (\boldsymbol{Y}_i + \boldsymbol{c})/\|\boldsymbol{Y}_i + \boldsymbol{c}\|$, $i = 1, \ldots, n$ are distributed according to the MDCU model of Section 6. Then, using a first order Taylor's expansion, we get

$$\text{(A.3.1)} \qquad \hat{\bar{\boldsymbol{Y}}} = \bar{\boldsymbol{Y}} - \tilde{\boldsymbol{B}}(\boldsymbol{c})(\hat{\boldsymbol{c}} - \boldsymbol{c}) + \boldsymbol{o}_P(\|\hat{\boldsymbol{c}} - \boldsymbol{c}\|),$$

where

$$\text{(A.3.2)} \qquad \tilde{\boldsymbol{B}}(\boldsymbol{c}) = \boldsymbol{I}_p - \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Y}_i + \boldsymbol{c})\boldsymbol{Y}_i'/(1 + \boldsymbol{c}'\boldsymbol{Y}_i)$$

which converges in probability to

$$\text{(A.3.3)} \qquad \boldsymbol{B}(\boldsymbol{c}) = \boldsymbol{I}_p - E[(\boldsymbol{Y}_1 + \boldsymbol{c})\boldsymbol{Y}_1'/(1 + \boldsymbol{c}'\boldsymbol{Y}_1)].$$

Thus we may write

$$\text{(A.3.4)} \qquad \hat{\bar{\boldsymbol{Y}}} = \bar{\boldsymbol{Y}} - \boldsymbol{B}(\boldsymbol{c})(\hat{\boldsymbol{c}} - \boldsymbol{c}) + \boldsymbol{o}_P(\|\hat{\boldsymbol{c}} - \boldsymbol{c}\|).$$

Then, we may show, using the implicit function theorem, that $\hat{\boldsymbol{c}}$ as a solution to the system (4.1) can be written as

$$(\hat{\boldsymbol{c}} - \boldsymbol{c}) = \boldsymbol{D}(\boldsymbol{c})^{-1} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{W}_i(\boldsymbol{c}) + \boldsymbol{o}_P(\|\hat{\boldsymbol{c}} - \boldsymbol{c}\|) \tag{A.3.5}$$

with

$$\boldsymbol{W}_i(\boldsymbol{c}) = (p-1) \frac{\boldsymbol{Y}_i}{(1 + \boldsymbol{c}'\boldsymbol{Y}_i)} - \boldsymbol{Y}_i \frac{\|\boldsymbol{Y}_i + \boldsymbol{c}\|}{(1 + \boldsymbol{c}'\boldsymbol{Y}_i)^2} + \frac{(\boldsymbol{Y}_i + \boldsymbol{c})}{(1 + \boldsymbol{c}'\boldsymbol{Y}_i)} \tag{A.3.6}$$

and

$$\boldsymbol{D}(\boldsymbol{c}) = \mathrm{Var}[\boldsymbol{W}_1(\boldsymbol{c})]. \tag{A.3.7}$$

Combining (A.3.4) and (A.3.5) yields

$$\hat{\hat{\boldsymbol{Y}}} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{Y}_i - \boldsymbol{B}(\boldsymbol{c}) \boldsymbol{D}(\boldsymbol{c})^{-1} \boldsymbol{W}_i(\boldsymbol{c})) + \boldsymbol{o}_P(\|\hat{\boldsymbol{c}} - \boldsymbol{c}\|). \tag{A.3.8}$$

It is clear then, that

$$n^{1/2} \hat{\hat{\boldsymbol{Y}}} \xrightarrow[n \to \infty]{\mathcal{L}} N_p\{\boldsymbol{0}, \Lambda(\boldsymbol{c})\} \tag{A.3.9}$$

with

$$\begin{aligned} \Lambda(\boldsymbol{c}) = {} & \mathrm{Var}[\boldsymbol{Y}_1] + \boldsymbol{B}(\boldsymbol{c}) \boldsymbol{D}(\boldsymbol{c})^{-1} \boldsymbol{B}(\boldsymbol{c})' \\ & - \mathrm{Cov}[\boldsymbol{Y}_1, \boldsymbol{W}_1(\boldsymbol{c})] \boldsymbol{D}(\boldsymbol{c})^{-1} \boldsymbol{B}(\boldsymbol{c})' \\ & - \boldsymbol{B}(\boldsymbol{c}) \boldsymbol{D}(\boldsymbol{c})^{-1} \mathrm{Cov}[\boldsymbol{Y}_1, \boldsymbol{W}_1(\boldsymbol{c})]. \end{aligned} \tag{A.3.10}$$

Now when $\boldsymbol{Y}_1$ is uniformly distributed on S_p, (A.3.10) simplifies to (5.3). But in the case where $\boldsymbol{Y}_1$ is uniformly distributed over G_p, specified in Section 6, we have

$$\mathrm{Var}[\boldsymbol{Y}_1] = \mathrm{Var}[T] \nu \nu' + \frac{E[1 - T^2]}{p - 1} (\boldsymbol{I}_p - \nu \nu') \tag{A.3.11}$$

where T has density $f(t) = \omega_{p-1}(1 - t^2)^{(p-3)/2} / c_p(a, b)$ if $t \in [a, b]$.

Let $\hat{\boldsymbol{W}}_i(\hat{\boldsymbol{c}})$ be as in equation (A.3.6) with $\boldsymbol{Y}_i$ and $\boldsymbol{c}$ replaced by $\hat{\boldsymbol{Y}}_i$ and $\hat{\boldsymbol{c}}$. Then, consistent estimators of $\boldsymbol{B}(\boldsymbol{c})$, $\boldsymbol{D}(\boldsymbol{c})$ and $\mathrm{Cov}[\boldsymbol{Y}_1, \boldsymbol{W}_1(\boldsymbol{c})]$ are given by

$$\hat{\boldsymbol{B}}(\hat{\boldsymbol{c}}) = \boldsymbol{I}_p - \frac{1}{n} \sum_{i=1}^{n} (\hat{\boldsymbol{Y}}_i + \hat{\boldsymbol{c}}) \hat{\boldsymbol{Y}}_i' / (1 + \hat{\boldsymbol{c}}' \hat{\boldsymbol{Y}}_i), \tag{A.3.12}$$

$$\hat{\boldsymbol{D}}(\hat{\boldsymbol{c}}) = \frac{1}{n} \sum_{i=1}^{n} \hat{\boldsymbol{W}}_i(\hat{\boldsymbol{c}}) \hat{\boldsymbol{W}}_i(\hat{\boldsymbol{c}})' \tag{A.3.13}$$

and

$$S = \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{Y}}_i \hat{\boldsymbol{W}}_i(\hat{\boldsymbol{c}})' \tag{A.3.14}$$

respectively. Replacing these estimators in (A.3.10), we get a consistent estimator $\tilde{\Lambda}$ of $\Lambda(\boldsymbol{c})$.

References

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions*, N. B. S. applied mathematics series, 55, U. S. Government Printing Office, Washington.

Do Carmo, M. P. (1976). *Differential Geometry of Curves and Surfaces*, Prentice-Hall, New Jersey.

Fisher, N. I. and Best, D. J. (1984). Goodness-of-fit tests for Fisher's distribution on the sphere, *Austral. J. Statist.*, **26**, 142–150.

Fisher, N. I., Lewis, T. and Embleton, B. J. J. (1987). *Statistical Analysis of Spherical Data*, Cambridge University Press, Cambridge.

Jupp, P. E., Spurr, B. D., Nichols, G. J. and Hirst, J. P. P. (1987). Statistical estimation of the apex of a sediment distribution system from paleocurrent data, *Math. Geol.*, **19**, 319–333.

Kiraly, P., Kota, J., Osborne, J. L., Stapley, N. R. and Wolfendale, A. W. (1979). The anisotropy of cosmic rays from 10^{11} to 10^{20} eV, *La Rivista del Nuovo Cimento della Societa Italiana di Fisica*, Vol. 2, Ser. 3, numero 7.

Lehmann, E. L. (1983). *Theory of Point Estimation*, 2nd ed., Wiley, New York.

Mardia, K. V. and Edwards, R. (1982). Weighted distributions and rotating caps, *Biometrika*, **69**, 323–330.

Mardia, K. V., Holmes, D. and Kent, J. T. (1984). A goodness-of-fit test for the von Mises-Fisher distribution, *J. Roy. Statist. Soc. Ser. B*, **46**, 72–78.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley, New York.

Rivest, L.-P. (1986). Modified Kent's statistics for testing goodness-of-fit for the Fisher distribution in small concentrated samples, *Statist. Probab. Lett.*, **4**, 1–4.

Toyoda, Y., Suga, K., Murakami, K., Hasegawa, H., Shibata, S., Domingo, V., Escobar, I., Kamata, K., Bradt, H., Clark, G. and La Pointe, M. (1965). Studies of primary cosmic rays in the energy region 10^{14} eV to 10^{17} eV (Bolivian Air Shower Joint Experiment), *Proceedings of the International Conference on Cosmic Rays*, Vol. 2, London, September 1965, 708–711.

Watson, G. N. (1983). *Statistics on Spheres*, Wiley, New York.