Biometrika (1987), **74**, 3, *pp.* 579–89
Printed in Great Britain

Statistical analysis for the angular central Gaussian distribution on the sphere

BY DAVID E. TYLER

Department of Statistics, Rutgers University, New Brunswick, New Jersey 08903, U.S.A.

SUMMARY

The angular central Gaussian distribution is an alternative to the Bingham distribution for modeling antipodal symmetric directional data. In this paper the statistical theory for the angular central Gaussian model is presented. Some topics treated are maximum likelihood estimation of the parameters, testing for uniformity and circularity, and principal components analysis. Comparisons to methods based upon the sample second moments are made via an example.

Some key words: Axial data; Bingham distribution; Bipolar distribution; Directional data; Girdle distribution; Principal component; Principal direction; Test for circularity; Test for uniformity.

1. INTRODUCTION AND SUMMARY

The family of angular central Gaussian distributions on the q-dimensional sphere S_q with radius one is characterized by densities with respect to the uniform measure on S_q of the form

$$f(l; \Lambda) = \alpha_q^{-1} |\Lambda|^{-\frac{1}{2}} (l'\Lambda^{-1} l)^{-\frac{1}{2}q}, \quad l \in S_q, \tag{1}$$

where $\alpha_q = 2\pi^{\frac{1}{2}q}/\Gamma(\frac{1}{2}q)$ is the surface area of S_q, and Λ is a symmetric positive-definite matrix parameter. There is a slight indeterminancy in the parameter Λ; namely $f(l; \Lambda_1) = f(l; \Lambda_2)$ if and only if $\Lambda_1 = \alpha \Lambda_2$ for some positive scalar α. The family of angular central Gaussian distributions is an alternative to the family of Bingham distributions for modelling antipodal symmetric directional data. Although there are many papers on inferential problems concerning the Bingham distributions, corresponding results for the angular central Gaussian distributions appear not to exist. The objective of this paper is to give such results.

The angular central Gaussian model has a number of attractive features, and some of its merits are discussed by Watson (1983, § 3·6). Mardia (1972, § 3·4·7) also discusses the model under the title of offset normal. The family of angular central Gaussians include the uniform distribution on the sphere as a special case, which corresponds to the parameter Λ being proportional to I. An angular central Gaussian with parameter Λ can be transformed to a uniform distribution on the sphere by the transformation $l_0 = \Lambda^{-\frac{1}{2}} l/(l'\Lambda^{-1} l)^{\frac{1}{2}}$. There is no known simple transformation to uniformity for any other antipodal symmetric distribution on the sphere with density proportional to $g(l'Kl)$, where g is some nonnegative function and K is a symmetric matrix parameter. This includes the Bingham model which corresponds to $g(s) = \exp(s)$. The angular central Gaussian derives its name from the property that (1) represents the distribution of the directions of a normal $(0, \Lambda)$ distribution. It also represents the distribution of the directions of any distribution in R^q with concentric elliptical contours defined by $y'\Lambda^{-1}y = c$. In particular, the model (1) has the simple geometric interpretation of representing the distribution of the directions of a uniform distribution on the ellipse $y'\Lambda^{-1}y = 1$, a natural family of alternatives to the uniform distribution on the sphere. The primary motivation for the Bingham model, on the other hand, is that it is a convenient exponential model. Nevertheless, the normalizing constant for the angular central Gaussian distribution, namely $|\Lambda|^{-\frac{1}{2}}$, is simpler than the normalizing constant for the Bingham distribution, which involves the multivariate hypergeometric function ${}_1F_1$, and so maximum likelihood methods are less cumbersome.

In § 2, it is shown that the maximum likelihood estimate of Λ in (1) based on a random sample of size n is the solution to the equation

$$\hat{\Lambda} = qn^{-1} \sum_{i=1}^{n} l_i l_i' / (l_i' \hat{\Lambda}^{-1} l_i). \tag{2}$$

The author has recently studied this estimator in the context of M-estimators of multivariate scatter (Tyler, 1987), in particular a convergent fixed point algorithm for finding the solutions to (2) being given; see § 2 of the present paper.

The asymptotic distribution of the maximum likelihood estimator is given in § 3, and problems of statistical inference are considered in §§ 4–7. In § 4, the fundamental hypothesis that Λ is proportional to I and hence that the distribution is uniform on the sphere is treated. Since the particular axes used in representing directional data are often arbitrary, it is natural to study the principal components decomposition of Λ whenever Λ is not proportional to I. The eigenvectors of Λ are referred to as principal directions or orientation vectors and the roots of Λ, or principal component roots, represent measures of relative concentration in the principal directions. The asymptotic distribution of the roots of $\hat{\Lambda}$ and confidence intervals for the principal component roots are given in § 5. The problem of testing for circularity, which includes as special cases for $q = 3$ the problems of testing for a bipolar and a girdle distribution on the sphere, is discussed in § 6. Section 7 gives confidence intervals and tests for the eigenvectors of Λ.

An example using the data originally analysed by Bingham (1974) is given in § 8; see also Prentice (1984). Both earlier analyses are based on the sample moments of inertia, or sample second moments. Brief comparisons between the methods proposed here and those based on the sample second moments are made.

2. The maximum likelihood estimator

For a random sample $\{l_i, 1 \leq i \leq n\}$ from a distribution having density belonging to the family (1) with Λ unknown, the likelihood function for Λ is proportional to

$$L(\Lambda) = |\Lambda|^{-\frac{1}{2}n} \prod_{i=1}^{n} (l_i' \Lambda^{-1} l_i)^{-\frac{1}{2}q}, \quad \Lambda \in C,$$

where C is the set of symmetric positive-definite matrices of order q. Theorem 1 below states that a maximum likelihood estimate for Λ almost surely exists, corresponds to a solution to (2), and is unique up to multiplication by a positive scalar. The nonuniqueness of the estimate poses no problem since the parameter itself possesses the same indeterminancy. The parameter and estimate can be made unique by placing a common constraint on both. However, it is more convenient to use the unconstrained matrices.

THEOREM 1. *If $n > q(q-1)$ and $\{l_i, 1 \leqslant i \leqslant n\}$ represents a random sample from an angular central Gaussian distribution then the following statements hold with probability one.*

(i) *There exists a $\hat{\Lambda} \in C$ which maximizes $L(\Lambda)$.*

(ii) *The matrix $\hat{\Lambda} \in C$ maximizes $L(\Lambda)$ if and only if it satisfies* (2).

(iii) *Both $\hat{\Lambda} \in C$ and $\tilde{\Lambda} \in C$ maximize $L(\Lambda)$ if and only if $\tilde{\Lambda} = a\hat{\Lambda}$ for some $a > 0$.*

Proof. In this proof, assume all subsets of size q from the sample span R^q. For $n \geqslant q$, this occurs almost surely not only when random sampling from an angular central Gaussian distribution, but from any continuous distribution on S_q.

Let $h(\Gamma) = L(\Gamma^{-1})$ for $\Gamma \in C$. If h obtains its maximum at $\Gamma \in C$, then L obtains its maximum at $\Gamma^{-1} \in C$. Since $h(\Gamma) = h(\alpha\Gamma)$ for any positive scalar α, the domain can be restricted to the bounded set $C_0 = \{\Gamma \in C, \operatorname{tr}(\Gamma) = q\}$. The boundary set of C_0 is the set C_b of singular symmetric positive-semidefinite matrices of order q whose trace equals q. Consider any sequence Γ_k in C_0 such that $\Gamma_k \to \Gamma \in C_b$. Let $\gamma_{1,k} \geqslant \ldots \geqslant \gamma_{q,k} > 0$ represent the eigenvalues of Γ_k and let $\gamma_1 \geqslant \ldots \geqslant \gamma_r > 0$ and $\gamma_{r+1} = \ldots = \gamma_q = 0$ represent the eigenvalues of Γ where $r = \operatorname{rank}(\Gamma)$. By the conditions on the observed directions, there exist at most $m = q - r$ directions such that $\Gamma l_i = 0$, say for convenience $l_1, \ldots, l_m$. Since $\gamma_{q,k}/l_i'\Gamma_k l_i \leqslant 1$ it follows that

$$0 \leqslant h(\Gamma_k) \leqslant \left(\prod_{j=1}^{q-1} \gamma_{j,k}^{\frac{1}{2}n}\right)\left(\prod_{i=m+1}^{n} l_i'\Gamma_k l_i\right)^{-\frac{1}{2}q} \gamma_{q,k}^{\frac{1}{2}(n-qm)}.$$

By the continuity of eigenvalues $\gamma_{j,k} \to \gamma_j$ $(1 \leqslant j \leqslant q)$. Consequently, $h(\Gamma_k) \to 0$ since $\gamma_q = 0$, $l_i'\Gamma l_i \neq 0$ for $m+1 \leqslant i \leqslant n$ and $n > qm$. Since h is a strictly positive continuous and bounded function on the bounded set C_0, it follows that h has a maximum in C_0 and hence part (i) holds.

Let $M(\Lambda) = \log\{L(\Lambda)\}$. The function $M(\Lambda)$ is differentiable, and so part (ii) can be proved by showing that the only critical points of M are those satisfying (2). After some straightforward calculation, we have

$$M(\Lambda + \dot{\Lambda}) = M(\Lambda) + \operatorname{tr}\{H(\Lambda)\dot{\Lambda}\} + O(\|\dot{\Lambda}\|),$$

where $H(\Lambda) = -\frac{1}{2}n\Lambda^{-1} + \frac{1}{2}q\,\Sigma\,\Lambda^{-1}l_il_i'\Lambda^{-1}/l_i'\Lambda^{-1}l_i$. A critical point $\hat{\Lambda}$ of M is obtained if and only if $H(\hat{\Lambda}) = 0$, which holds if and only if (2) holds. Thus, if $\hat{\Lambda} \in C$ maximizes $L(\Lambda)$, then it must satisfy (2). The converse is also true since a solution to (2) is unique up to multiplication by a positive scalar; see Tyler (1987, Corrol. 2·1). Hence the critical point must be a maximum since a maximum exists and $M(\Lambda) = M(\alpha\Lambda)$ for any $\alpha > 0$. The 'uniqueness' of the maximum likelihood estimate also follows from the aforementioned corollary. □

For $n > q(q-1)$ the following fixed point algorithm almost surely generates the solution to (2), and hence the maximum likelihood estimate, which satisfies the constraint $\operatorname{tr}(\hat{\Lambda}) = q$ (Tyler, 1987, Corrol. 2·2). Let $\Lambda_0 = I$ and define the recursive sequence

$$\Lambda_{k+1} = q\left\{\sum_{i=1}^{n} l_il_i'/(l_i'\Lambda_k^{-1}l_i)\right\}\Big/\left\{\sum_{i=1}^{n} 1/(l_i'\Lambda_k^{-1}l_i)\right\}. \qquad (3)$$

The sequence $\Lambda_k \to \hat{\Lambda}$, the solution to (2) whose trace equals q. The convergence is 'monotone' in the sense that the largest and smallest roots of $q\Lambda_k^{-1}\Lambda_{k+1}/\operatorname{tr}(\Lambda_k^{-1}\Lambda_{k+1})$ decrease and increase respectively to 1.

3. Asymptotic distribution

Let $\hat{\Lambda}_0$ represent the solution to (2) which is normalized so that $\mathrm{tr}\,(\Lambda^{-1}\hat{\Lambda}_0)=q$. The asymptotic normality of $\hat{\Lambda}_0$ is established in the next lemma. The normalization employed for $\hat{\Lambda}_0$ results in a fairly tractable form for the asymptotic covariance matrix. The asymptotic distributions associated with different constraints on the estimate can be obtained from the asymptotic distribution of $\hat{\Lambda}_0$. The particular normalization of the parameter Λ itself is unimportant since if Λ is multiplied by a positive scalar then $\hat{\Lambda}_0$ is multiplied by the same scalar.

LEMMA 1. *For a random sample of size n from a distribution with density* (1), $n^{\frac{1}{2}}(\hat{\Lambda}_0-\Lambda)\rightarrow$ N *in distribution as* $n\rightarrow\infty$, *where* $\mathrm{vec}\,(N)$ *is multivariate normal with mean zero and covariance matrix*

$$C(\Lambda)=(1+2q^{-1})\{(I+K_{q,q})(\Lambda\otimes\Lambda)-2q^{-1}\,\mathrm{vec}\,(\Lambda)\,\mathrm{vec}\,(\Lambda)'\}.$$

Proof. The asymptotic distribution given in the Lemma corresponds to the asymptotic distribution of the solution to $\hat{\Lambda}_0=n^{-1}\Sigma\, y_iy_i'/(y_i'\hat{\Lambda}^{-1}y_i)$ whenever $y_1,\ldots,y_n$ represent a random sample from a normal $(0,\Lambda)$ distribution; see Tyler (1987, Th. 3·2 and eqn (3·10)). The lemma then follows directly since if y is normal $(0,\Lambda)$ then $l=y/(y'y)^{\frac{1}{2}}$ has the density (1). □

The notation used in the lemma has recently become somewhat commonplace. For matrix B, $\mathrm{vec}\,(B)$ represents the vector formed by stacking the columns of B, the product $\otimes$ refers to the Kronecker product, and $K_{q,q}$ is the commutation matrix. For further details, see either Magnus & Neudecker (1979) or the recent text by Muirhead (1982).

The asymptotic distribution of $\hat{\Lambda}_0$ can be represented in the following alternative form. As $n\rightarrow\infty$, $n^{\frac{1}{2}}(\Lambda^{-\frac{1}{2}}\hat{\Lambda}_0\Lambda^{-\frac{1}{2}}-I)\rightarrow Z=\Lambda^{-\frac{1}{2}}N\Lambda^{-\frac{1}{2}}$ in distribution, where N is the multivariate normal distribution stated in the lemma and $\Lambda^{-\frac{1}{2}}$ represents the unique symmetric positive-definite square root of Λ^{-1}. The distribution of $\mathrm{vec}\,(Z)$ is multivariate normal with mean zero and covariance matrix $C(I)$. This covariance matrix states that the diagonal elements of Z each have variance $2(q+2)(q-1)/q^2$ and the covariance between any two diagonal elements is $-2(q+2)/q^2$. The distinct off-diagonal elements of Z are uncorrelated with each other and with the diagonal elements, and have variance $(q+2)/q$. Since Z is symmetric and $\mathrm{tr}\,(Z)=0$, its distribution lies in $\frac{1}{2}(q+2)(q-1)$ dimensional real space.

4. A test for uniformity

If Λ is proportional to I in (1), then l is uniformly distributed on the unit q-sphere. A test for uniformity on the sphere against the alternative that (1) holds but with Λ not proportional to I can thus be constructed from the likelihood ratio principle. Hereafter, let $\hat{\Lambda}$ represent the maximum likelihood estimate which is normalized so that $\mathrm{tr}\,(\hat{\Lambda})=q$, a normalization which does not depend on the unknown Λ, and let Λ be normalized so that $\mathrm{tr}\,(\Lambda)=q$. The likelihood ratio statistic for testing H_0: $\Lambda=I$ against the general alternative H_1: $\Lambda\neq I$ is

$$\lambda_n=|\hat{\Lambda}|^{\frac{1}{2}n}\prod_{i=1}^{n}(l_i'\hat{\Lambda}^{-1}l_i)^{\frac{1}{2}q}. \tag{4}$$

The asymptotic distribution of λ_n under H_0 and under the sequence of local alternatives $H_{1,n}$: $\Lambda_n=I+n^{-\frac{1}{2}}D$, where D is symmetric and $\mathrm{tr}\,(D)=0$ follows from the general theory of likelihood ratio tests. Under H_0,

$$-2\log\lambda_n \to \chi^2_{\frac{1}{2}(q+2)(q-1)}, \tag{5}$$

and under the sequence $H_{1,n}$,

$$-2\log\lambda_n \to \chi^2_{\frac{1}{2}(q+2)(q-1)}(\delta), \tag{6}$$

a noncentral chi-squared variable with noncentrality parameter $\delta = \frac{1}{2}q(q+2)^{-1}\operatorname{tr}(D^2)$.

Asymptotically, the likelihood ratio statistic is equivalent to a function of the maximum likelihood estimator $\hat{\Lambda}$ alone. In particular, application of Lemma 1 and the equivalence between Wald's quadratic form criterion and the likelihood ratio statistic gives

$$-2\log\lambda_n \sim n\{\operatorname{vec}(\hat{\Lambda}-I)\}'\{C(I)\}^+\{\operatorname{vec}(\hat{\Lambda}-I)\} \tag{7}$$

under H_0 or under the sequence $H_{1,n}$, where $\{C(I)\}^+$ refers to the Moore–Penrose generalized inverse of $C(I)$. The notation $X_n \sim Y_n$ implies $X_n - Y_n \to 0$ in probability. After some algebraic calculations statement (7) becomes

$$-2\log\lambda_n \sim T_n = \tfrac{1}{2}nq(q+2)^{-1}\operatorname{tr}\{(\hat{\Lambda}-I)^2\} = \tfrac{1}{2}nq(q+2)^{-1}\sum_{i=1}^{q}(\hat{\gamma}_i-1)^2, \tag{8}$$

where $\hat{\gamma}_1 \geqslant \ldots \geqslant \hat{\gamma}_q$ are the eigenvalues of $\hat{\Lambda}$. The results (5), (6) and (8) can also be established directly from Lemma 1 by using Taylor series expansions and noting $\hat{\Lambda} = q\hat{\Lambda}_0/\operatorname{tr}(\hat{\Lambda}_0)$ rather than resorting to the general theory of maximum likelihood.

Other statistics based on $\hat{\Lambda}$ can be constructed for testing the hypothesis of uniformity. Transforming the data from $\{l_i\}$ to $\{Pl_i\}$ for any orthogonal P preserves uniformity and results in transforming the estimate from $\hat{\Lambda}$ to $P\hat{\Lambda}P'$. If one wishes to adhere to the invariance principle, then only those functions of $\hat{\Lambda}$ which are solely functions of its eigenvalues should be considered as possible test statistics. Since $\operatorname{tr}(\hat{\Lambda}) = (\hat{\gamma}_1+\ldots+\hat{\gamma}_q) = q$ is constant, one possible candidate is $|\hat{\Lambda}| = \hat{\gamma}_1\ldots\hat{\gamma}_q$, which is less than one unless $\hat{\gamma}_1 = \ldots = \hat{\gamma}_q = 1$. However, this is asymptotically equivalent to using the likelihood ratio test, since by expanding $\log|\hat{\Lambda}|$ and using (8) it follows that

$$-2\log\lambda_n \sim nq(q+2)^{-1}\log|\hat{\Lambda}| = nq(q+2)^{-1}\sum_{i=1}^{q}\log\hat{\gamma}_i$$

under H_0 or under the sequence $H_{1,n}$. Other possibilities are $\hat{\gamma}_1$ and $\hat{\gamma}_q$ alone. These are not asymptotically equivalent to the likelihood ratio test except whenever $q=2$ in which case $\hat{\gamma}_1 = 2-\hat{\gamma}_2$ or some monotone function of $\hat{\gamma}_1$ are the only possible choices. In general, how different test statistics compare to each other is an open question and is not pursued here. The asymptotic distribution of the roots under H_0 and under the sequence $H_{1,n}$ are given in an earlier expanded version of the current paper.

5. Principal component roots

If uniformity is rejected, then further analysis of the parameter Λ is warranted, and as noted in § 1 it is natural to consider the principal components decomposition of Λ. The asymptotic distribution of the estimated principal component roots, that is the eigenvalues of $\hat{\Lambda}$ is characterized in Theorem 2 below.

Notation used in § 4 still holds. In particular, $\hat{\Lambda}$, $\hat{\Lambda}_0$ and Λ are normalized so that $\mathrm{tr}(\hat{\Lambda})=\mathrm{tr}(\Lambda)=\mathrm{tr}(\Lambda^{-1}\hat{\Lambda}_0)=q$, and $\hat{\gamma}_1\geqslant\ldots\geqslant\hat{\gamma}_q$ represent the eigenvalues of $\hat{\Lambda}$. Let $\gamma_1\geqslant\ldots\geqslant\gamma_q$ represent the eigenvalues of Λ, and define $w_{j,n}=n^{\frac{1}{2}}(\hat{\gamma}_j-\gamma_j)$ for $1\leqslant j\leqslant n$. Due to the nature of the eigenvalue problem, the asymptotic distribution of the roots of $\hat{\Lambda}$ depend on the multiplicity of the corresponding roots of Λ. Suppose

$$\gamma_1=\ldots=\gamma_{q(1)}>\gamma_{q(1)+1}=\ldots=\gamma_{q(2)}>\ldots>\gamma_{q(k-1)+1}=\ldots=\gamma_{q(k)},$$

where $q(k)=q$. This states that Λ consists of k distinct roots with multiplicities $m(1)=q(1)$, $m(2)=q(2)-q(1),\ldots,m(k)=q(k)-q(k-1)$ respectively. Let Z be a $q\times q$ multivariate normal matrix with mean 0 and where the covariance matrix of $\mathrm{vec}(Z)$ is $C(I)$ which is defined in Lemma 1. Let $Z_{(1)},\ldots,Z_{(k)}$ be the nonoverlapping block diagonal matrices of Z of order $m(1),\ldots,m(k)$ respectively.

THEOREM 2. *As* $n\to\infty$, $(w_{1,n},\ldots,w_{q,n})\to(w_1,\ldots,w_q)$ *in distribution, where* $w_i=\gamma_i\{x_i-q^{-1}(\gamma_1x_1+\ldots+\gamma_qx_q)\}$ *and* $x_{q(a-1)+1},\ldots,x_{q(a)}$ *represent the eigenvalues of* $Z_{(a)}$ $(1\leqslant a\leqslant k)$.

Proof. As a special case of Theorem 4.1 of Tyler (1983a), $(x_1,\ldots,x_q)$ represents the limiting distribution of $(\hat{x}_1,\ldots,\hat{x}_q)$, where $\hat{x}_i=n^{\frac{1}{2}}(\hat{\gamma}_{0,i}-\gamma_i)/\gamma_i$ and $\hat{\gamma}_{0,1}\geqslant\ldots\geqslant\hat{\gamma}_{0,q}$ are the eigenvalues of $\hat{\Lambda}_0$. The theorem follows by noting $\hat{\gamma}_i=q\hat{\gamma}_{0,i}/\mathrm{tr}(\hat{\Lambda}_0)=q\hat{\gamma}_{0,i}/(\hat{\gamma}_{0,1}+\ldots+\gamma_{0,q})$ and then expanding $\hat{\gamma}_{0,i}$ about γ_i. □

Let $\Sigma_{j\in(a)}$ represent summation over $j=q(a-1)+1,\ldots,q(a)$, and note that $\Sigma_{j\in(a)}\gamma_jx_j=\gamma_{q(a)}\mathrm{tr}(Z_{(a)})$ and hence $\gamma_1x_1+\ldots+\gamma_qx_q=\gamma_{q(1)}\mathrm{tr}(Z_{(1)})+\ldots+\gamma_{q(k)}\mathrm{tr}(Z_{(k)})$. If γ_i is a distinct root then w_i is a linear combination of variables which are jointly multivariate normal and so has a normal distribution itself. For this case, w_i is normal with mean zero and variance $2q^{-2}(q+2)\{q-2\gamma_i+q^{-1}\mathrm{tr}(\Lambda^2)\}\gamma_i^2$. An asymptotic $1-\alpha$ confidence interval for γ_i is thus given by

$$\hat{\gamma}_i[1\pm z_{\frac{1}{2}\alpha}q^{-1}\{2n^{-1}(q+1)\}^{\frac{1}{2}}\{q-2\hat{\gamma}_i+q^{-1}\mathrm{tr}(\hat{\Lambda}^2)\}^{\frac{1}{2}}]. \tag{9}$$

For the important cases $q=2$ and $q=3$, formulae for confidence intervals for nondistinct roots are not necessary. For $q=2$, if uniformity is rejected then (9) would be appropriate to use for either γ_1 or γ_2. Note that constructing both intervals would be redundant since $\gamma_1+\gamma_2=2$. For $q=3$, if uniformity is rejected, then it is still possible that either $\gamma_1=\gamma_2>\gamma_3$ or $\gamma_1>\gamma_2=\gamma_3$. These possibilities represent a girdle and a bipolar distribution respectively (Mardia, 1972, § 8·4·2a). Whether either possibility holds can be tested by the procedures discussed in § 6. If all three roots are distinct, then (9) applied to any root. For the girdle and bipolar cases, interval (9) can be used for the distinct root. An interval for the multiple root is redundant since $\gamma_1+\gamma_2+\gamma_3=3$.

Theorem 2 gives only a characterization of the asymptotic distribution of the estimated roots. The joint density of the limiting variates $w_1,\ldots,w_q$ are given in an earlier expanded version of this paper, along with formulae for asymptotic confidence intervals for nondistinct roots when $q>3$.

6. TESTS FOR CIRCULARITY

As noted in § 5, besides uniformity on the sphere, two other important cases of (1) for $q=3$ are $\gamma_1=\gamma_2>\gamma_3$ and $\gamma_1>\gamma_2=\gamma_3$ which correspond to girdle and bipolar distributions respectively. In this section, assuming Λ is not proportional to I, an asymptotic test that the girdle distribution model holds and an asymptotic test that the bipolar distribution model holds are given, both against the alternative that (1) holds with $\gamma_1 > \gamma_2 > \gamma_3$.

Using the notation of § 5, either testing problem is a special case of the general problem of testing the hypothesis $H_{0,a}$: $\gamma_{q(a-1)+1} = \ldots = \gamma_{q(a)}$ under the assumption $\gamma_{q(a-1)} \neq \gamma_{q(a-1)+1}$ and $\gamma_{q(a)} \neq \gamma_{q(a)+1}$ in the model (1). The restricted maximum likelihood estimate of Λ under the null hypothesis is difficult to obtain and hence so is the likelihood ratio test. Furthermore, since the null hypothesis is not easily expressible in the form $H(\Lambda) = 0$ for some continuously differentiable function H, the general theory for the likelihood ratio test is not immediately applicable even if the test statistic were known for this problem. For these reasons, the proposed asymptotic tests for $H_{0,a}$ given here are motivated on intuitive grounds.

It is reasonable to consider the differences $\hat{\gamma}_i - \hat{\gamma}_{(a)}$, for $q(a-1)+1 \leq i \leq q(a)$, where $\hat{\gamma}_{(a)}$ is the average of $\hat{\gamma}_i$ over the indicated range. If $H_{0,a}$ is true, these differences should be small. To determine a suitable way of combining these differences in a single index, note that $n^{\frac{1}{2}}(\hat{\gamma}_i - \hat{\gamma}_{(a)})/\hat{\gamma}_{(a)} \to v_i = x_i - m(a)^{-1} \operatorname{tr}(Z_{(a)})$ in distribution under $H_{0,a}$, and that $\Sigma_{i \in (a)} v_i^2 = \operatorname{tr}(Z_{(a)}^2) - m(a)^{-1}\{\operatorname{tr}(Z_{(a)})\}^2$. By the general theory of quadratic forms in normal variables it follows that $\frac{1}{2}q(q+2)^{-1}\Sigma_{i\in(a)} v_i^2$ is distributed as a chi-squared variable on $\nu(a) = \frac{1}{2}\{m(a)+2\}\,\{m(a)-1\}$ degrees of freedom. Thus, an asymptotic chi-squared test statistic for $H_{0,a}$ is given by $T_{a,n}$, where

$$T_{a,n} = \tfrac{1}{2}nq(q+2)^{-1}\{\Sigma_{i\in(a)}(\hat{\gamma}_i - \hat{\gamma}_{(a)})^2/\hat{\gamma}_{(a)}^2\} \to \chi^2_{\nu(a)} \tag{10}$$

in distribution under $H_{0,a}$. This test statistic generalizes the test statistic for testing uniformity given by the right-hand side of (8). For the important case $q = 3$, the test statistic for testing the girdle case simplifies to $0{\cdot}6n(\hat{\gamma}_1 - \hat{\gamma}_2)^2/(\hat{\gamma}_1 + \hat{\gamma}_2)^2$, and for the bipolar case it simplifies to $0{\cdot}6n(\hat{\gamma}_2 - \hat{\gamma}_3)^2/(\hat{\gamma}_2 + \hat{\gamma}_3)^2$. Both are asymptotically χ_2^2 under their respective null hypotheses.

For a fixed alternative, the test statistic $T_{a,n}$ goes to infinity. Under the sequence of local alternatives $H_{1,a,n}$: $\gamma_i = \bar{\gamma}_{(a)}(1 + n^{-\frac{1}{2}}d_i)$, for $q(a-1)+1 \leq i \leq q(a)$, where $\Sigma_{i\in(a)} d_i = 0$, it can be shown that $T_{a,n} \to \chi^2_{\nu(a)}(\delta_a)$ with $\delta_a = \frac{1}{2}q(q+2)^{-1}\{\Sigma_{i\in(a)} d_i^2\}$. For more details, see Tyler (1983a), in particular (4.1), Theorem 4·1 and Theorem 5·2.

7. Principal directions

If γ_1 is a distinct root, then the eigenvector of Λ associated with this largest root can be considered the first principal direction and it is worthy of further study. It indicates the direction towards which most observations tend to fall. In general, a test for a prescribed direction or vector being the eigenvector associated with a presumed distinct root γ_j can be obtained by application of the general results on principal component vectors given by Tyler (1983b). Specifically, if b_j is the eigenvector associated with γ_j normalized so that $b_j'b_j = 1$, then

$$nq(q+2)^{-1}(\hat{\gamma}_j b_j'\hat{\Lambda}^{-1}b_j + \hat{\gamma}_j^{-1}b_j'\hat{\Lambda}b_j - 2) \to \chi^2_{q-1} \tag{11}$$

in distribution. This result follows by applying Theorem 4·1 to statement (3·10) of Tyler (1983b) for $\hat{\Lambda}_0$, and then observing that the constant $\hat{c} = q^{-1}\operatorname{tr}(\Lambda^{-1}\hat{\Lambda}_0)$ in the relationship $\hat{\Lambda} = \hat{c}\hat{\Lambda}_0$ cancels. If the hypothesized principal direction is incorrect then the test statistic (11) goes to infinity as n increases. A limiting noncentral chi-squared distribution for the test statistic under a sequence of local alternatives can also be obtained. For details, see Tyler (1983b).

By (11), a $1-\alpha$ asymptotic confidence region for b_j is given by the set of vectors b such that

$$nb'\hat{B}_j b < \chi^2_{q-1,\alpha}, \quad b'b = 1, \tag{12}$$

where $\hat{B}_j = q(q+2)^{-1}(\hat{\gamma}_j \hat{\Lambda}^{-1} + \hat{\gamma}_j^{-1}\hat{\Lambda} - 2I)$. The matrix $\hat{B}_j$ is singular with $\hat{B}_j b = 0$ if and only if $b = \hat{b}_j$, where $\hat{b}_j$ is an estimated jth principal direction, that is $\hat{\Lambda}\hat{b}_j = \hat{\gamma}_j \hat{\beta}_j$. The confidence region (12) corresponds to points on the unit q-sphere which lie inside the ellipsoid $nb'\hat{B}_j b = \chi^2_{q-1,\alpha}$. This ellipsoid has an infinite major axis in the direction of $\pm\hat{b}_j$, and so is actually an ellipsoidal cylinder. The other axes are in the direction of the other principal directions and have half lengths.

$$d_{ij,\alpha} = \{(nq)^{-1}(q+2)^{-1}\chi^2_{q-1,\alpha}\hat{\gamma}_i\hat{\gamma}_j\}^{\frac{1}{2}}/|\hat{\gamma}_i - \hat{\gamma}_j|, \tag{13}$$

Thus, the angle between the $\pm\hat{b}_j$ axis and the point where (12) intersects the unit q-sphere within the (b_i, b_j) plane is arc sin $(d_{ij,\alpha})$.

Consider the special case $q = 2$ and the first principal direction. The region (12) corresponds to the points on the unit circle between two lines which are parallel and equidistant to the direction $\pm\hat{b}_1$ with the distance being

$$(2n)^{-1} z_{\frac{1}{2}\alpha} \hat{\gamma}_1^{\frac{1}{2}}(2-\hat{\gamma}_1)^{\frac{1}{2}}/(\hat{\gamma}_1 - 1).$$

Let θ_1 and $\hat{\theta}_1$ represent the angle between the horizontal axis and b_1 and $\hat{b}_1$ respectively. The confidence region (12) can be expressed in terms of a confidence region for θ_1. Specifically, an asymptotic $(1-\alpha)$ confidence region for θ_1 is given by the interval

$$\hat{\theta}_1 \pm \text{arc sin}\,\{(2n)^{-1} z_{\frac{1}{2}\alpha} \hat{\gamma}_1^{\frac{1}{2}}(2-\hat{\gamma}_1)^{\frac{1}{2}}/(\hat{\gamma}_1 - 1)\} \tag{14}$$

plus the intervals reflection through the origin.

If γ_j is a multiple root, say of order m, then inferences concerning the m-dimensional eigenspace associated with the root can be made by application of the general results of Tyler (1983b). For the important cases $q = 2$ and $q = 3$, however, (12) is sufficient for constructing confidence regions. For $q = 2$, if uniformity is rejected, then (12) or (14) can be applied to the first principal direction. A confidence region for the second principal direction is redundant since it is orthogonal to the first. For $q = 3$, if uniformity is rejected, then there are three possible cases, $\gamma_1 > \gamma_2 > \gamma_3$, $\gamma_1 = \gamma_2 > \gamma_3$ or $\gamma_1 > \gamma_2 = \gamma_3$. For the first case, (12) applies to any principal direction. For the second or girdle case, (12) applies to the third principal direction. The space of eigenvectors corresponding to $\gamma_1 = \gamma_2$ is orthogonal to the third direction and so does not need to be studied separately. The third or bipolar case is analogous to the girdle case.

8. AN EXAMPLE WITH DISCUSSION

In this section, data analysed by Bingham (1974) using the Bingham model are reanalysed using the angular central Gaussian model (1). The data can be found in Bingham's doctoral dissertation and consist of $n = 150$ measurements of $q = 3$ dimensional c axes of calcite grains from the Taconic Mountains of New York State. Since the data are axial rather than directional, that is undirected axes, its distribution can be represented by points on an arbitrary hemisphere. For axial distributions, the density of an angular central Gaussian distribution on some hemisphere is twice the density given by (1) on the hemisphere. The statistical theory is unaffected.

For this example, the algorithm (3) converges rather quickly to the maximum likelihood estimate. After 20 iterations, the roots of $3\Lambda_{19}^{-1}\Lambda_{20}/\mathrm{tr}\,(\Lambda_{19}^{-1}\Lambda_{20})$ differ from 1 by less than 10^{-6}. The maximum likelihood estimate $\hat{\Lambda}$ normalized so that $\mathrm{tr}\,(\hat{\Lambda})=3$ is given in Table 1, along with its eigenvalues $\hat{\gamma}_1 \geqslant \hat{\gamma}_2 \geqslant \hat{\gamma}_3$ and the corresponding eigenvectors $\hat{b}_1$, $\hat{b}_2$ and $\hat{b}_3$, which are normalized to have length 1. Table 2 gives the values of the likelihood ratio statistic $-2\log\lambda_n$ of (4) for testing uniformity, its asymptotic null equivalent T_n from (8), the test statistic $T_{(1,2),n}$ and $T_{(2,3),n}$ used in testing the circularity hypotheses of a girdle and polar distribution respectively, as in (10), the asymptotic standard errors of the roots $\hat{Y}_i$ assuming the corresponding population roots are distinct, above (9), and finally the statistics arc sin $(d_{ij,\alpha=0\cdot01})$ which characterize 0·99 confidence regions for the principal directions; see (13).

Either test statistic for testing the hypothesis of uniformity, compared to a χ_5^2 distribution, clearly indicates that a uniform model is not appropriate. The asymptotic test statistics for circularity, compared to a χ_2^2 distribution, are both significant at the 0·05 level. However, the test statistic for testing the polar hypothesis is not significant at the 0·01 level. These results suggest that a parsimonious polar model may be a reasonable approximation. A visual inspection of the data tends to support these statements. Figure 1(a) gives an equiareal projection of the data onto the plane determined by $\hat{b}_2$, vertical, and by $\hat{b}_3$, horizontal.

The results of this analysis are similar to the results of Bingham's analysis. The principal directions or orientation vectors and the estimates of their variability are almost identical in both analyses. The tests for uniformity and for circularity obtained by Bingham have slightly higher *p*-values, but are relatively of the same magnitude as the *p*-values obtained here. Prentice (1984) analyses the same data under the more general assumption that the density of l is proportional to $g(l'Kl)$, where g is an unknown nonnegative functional and K is a $q\times q$ symmetric matrix parameter. His conclusions are similar to Bingham's. The major difference between the analysis given here and those given by Bingham (1974) and Prentice (1984) is that the evidence against a polar model is not as strong when assuming the angular Gaussian model. Both Bingham (1974) and Prentice (1984) note that visual inspection suggests the simple polar model is not appropriate. However, visual inspection is subjective and can be misleading. In his doctoral dissertation, Bingham suggests that the data appear to come from 'almost a girdle distribution'.

Table 1. *Maximum likelihood estimate* $\hat{\Lambda}$ *normalized so that* $\mathrm{tr}\,(\hat{\Lambda})=3$, *its eigenvalues* $\hat{\gamma}_1 \geqslant \hat{\gamma}_2 \geqslant \hat{\gamma}_3$ *and corresponding eigenvectors* $\hat{b}_1$, $\hat{b}_2$, $\hat{b}_3$ *for Taconic Mountains calcite data*

	$\hat{\Lambda}$		$\hat{\gamma}_i$	$\hat{b}_i$	$\hat{b}_2$	$\hat{b}_3$
1·73058	0·57920	0·31871	2·09823	0·87357	−0·45619	−0·16962
0·57920	0·86710	0·17051	0·57677	0·43984	0·88917	−0·12619
0·31871	0·17051	0·40232	0·32500	0·20839	0·03564	0·99740

Table 2. *Test statistics for uniformity and circularity, standard errors of* $\hat{\gamma}_i$, *and characterizing half-widths in degrees for* 99% *confidence regions for the principal directions*

	Test statistics	St. error $(\hat{\gamma}_i)$	arc sin $(d_{ij,0\cdot01})$	i	j
$-2\log\lambda_n$	98·1035	0·11664	4·3711	1	2
T_n	82·8386	0·09234	2·8137	1	3
$T_{(1,2),n}$	29·1152	0·05569	10·4438	2	3
$T_{(2,3),n}$	7·0152				

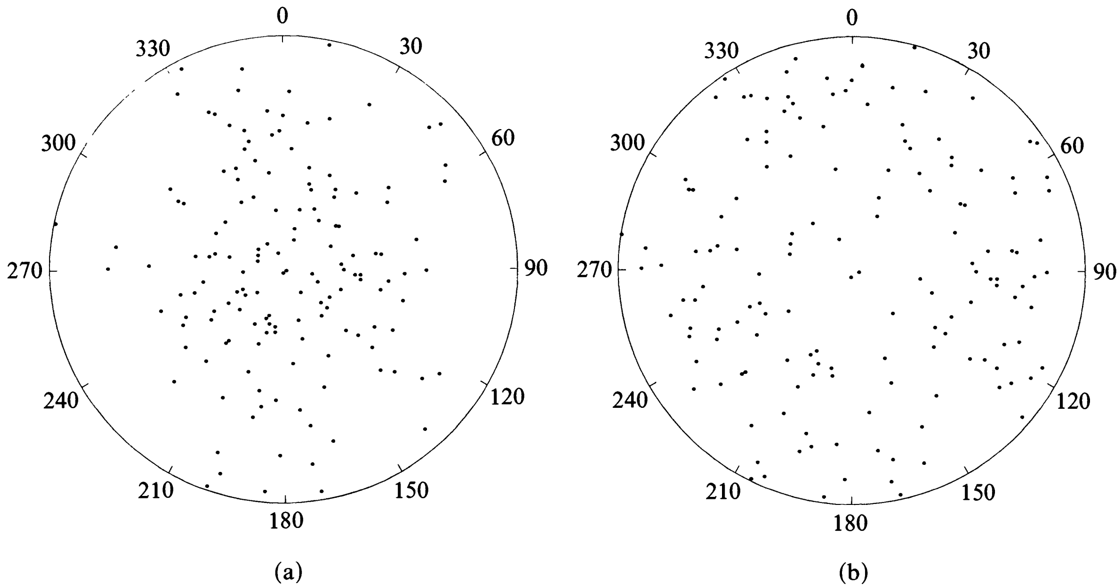

Fig. 1. Equiareal projection onto the plane determined by $\hat{b}_2$, vertical, and by $\hat{b}_3$, horizontal, of (a) calcite data, and (b) transformed 'residual' data.

The estimated principal directions used by both Bingham (1974) and Prentice (1984) correspond to the eigenvectors of $n^{-1}\Sigma\, l_i l_i'$. There is a theoretical reason for this striking similarity between the eigenvectors of this matrix of second moments and the eigenvectors of $\hat{\Lambda}$. Watson (1983, p. 167) has shown that the eigenvectors of $n^{-1}\Sigma\, l_i l_i'$ are consistent estimates of the eigenvectors of the symmetric matrix K in a model with density proportional to $g(l'Kl)$. A somewhat similar argument can be used to show that the eigenvectors of $\hat{\Lambda}$ are also consistent estimates of the eigenvectors of K, and hence the expected similarities.

Summarizing, the analysis given here provides an alternative to the sample second moment matrix for estimating principal directions and for making inferences concerning uniformity and circularity. The tests for uniformity and circularity used by Bingham (1974) and Prentice (1984) are based on the roots of $n^{-1}\Sigma\, l_i l_i'$ or on some asymptotically equivalent statistics. If conclusions based on $\hat{\Lambda}$ differ greatly from those based on the sample second moment matrix, then a more careful look at the data would be warranted.

There are some differences, however, in the interpretation of $\hat{\Lambda}$, the sample second moment matrix, and the maximum likelihood estimate of the Bingham matrix parameter. The roots of the latter two matrices are monotonic functions of each other, and are not directly related to the roots of $\hat{\Lambda}$. Furthermore, unlike for the eigenvectors, it is not apparent whether the roots of $\hat{\Lambda}$ or any function of them are estimating in some sense the same features of the distribution as the roots of the sample second moment matrix. The roots of all three matrices can be viewed as measuring in different ways the concentration of the data in each of the principal directions. The roots of $\hat{\Lambda}$, though, have a simple geometric interpretation. As noted in § 1, under the angular central Gaussian model the data can be viewed as the axes corresponding to points uniformly generated on the ellipse $y'\Lambda^{-1}y = 1$. The estimated ellipse $y'\hat{\Lambda}^{-1}y = 1$ has half-lengths $\hat{\gamma}_i^{\frac{1}{2}}$ along the principal directions $\hat{b}_i$. There is no obvious geometric interpretation for the roots of the other two matrices. Finally, under the angular central Gaussian model, the transformed 'residual' data $l_{0,i} = \hat{\Lambda}^{-\frac{1}{2}} l_i / (l_i' \hat{\Lambda}^{-1} l_i)^{\frac{1}{2}}$ should be approximately uniformly distributed on some hemisphere. Figure 1b gives an equiareal projection of the residual data onto the plane determined by $\hat{b}_2$, vertical, and by $\hat{b}_3$, horizontal. No simple transformations to uniformity based on the other two matrices are known.

ACKNOWLEDGEMENT

The author wishes to thank Professor Javier Cabrera for helpful discussions during the preparation of this paper. Research was supported by a National Science Foundation Grant.

REFERENCES

BINGHAM, C. (1974). An antipodally symmetric distribution on the sphere. *Ann. Statist.* **2**, 1201–25.

MAGNUS, J. R. & NEUDECKER, H. (1979). The commutation matrix: Some properties and applications. *Ann. Statist.* **7**, 381–94.

MARDIA, K. V. (1972). *Statistics of Directional Data.* New York: Academic Press.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory.* New York: Wiley.

PRENTICE, M. J. (1984). A distribution-free method of interval estimation for unsigned directional data. *Biometrika* **71**, 147–55.

TYLER, D. E. (1983a). The asymptotic distribution of principal component roots under local alternatives to multiple roots. *Ann. Statist.* **11**, 1232–42.

TYLER, D. E. (1983b). A class of asymptotic tests for principal component vectors. *Ann. Statist.* **11**, 1243–50.

TYLER, D. E. (1987). A distribution-free *M*-estimator of multivariate scatter. *Ann. Statist.* **15**, 234–51.

WATSON, G. S. (1983). *Statistics on Spheres.* New York: Wiley.

[*Received October* 1984. *Revised November* 1986]