The Annals of Statistics
1974, Vol. 2, No. 6, 1201–1225

AN ANTIPODALLY SYMMETRIC DISTRIBUTION ON THE SPHERE[1]

BY CHRISTOPHER BINGHAM

The University of Chicago

The distribution $\Psi(\mathbf{x}; Z, M) = \text{const.} \exp(\text{tr}(ZM^T\mathbf{x}\mathbf{x}^TM))$ on the unit sphere in three-space is discussed. It is parametrized by the diagonal shape and concentration matrix Z and the orthogonal orientation matrix M. Ψ is applicable in the statistical analysis of measurements of random undirected axes. Exact and asymptotic sampling distributions are derived. Maximum likelihood estimators for Z and M are found and their asymptotic properties elucidated. Inference procedures, including tests for isotropy and circular symmetry, are proposed. The application of Ψ is illustrated by a numerical example.

1. Introduction. Through an obvious correspondence, a completely specified direction in 3-space is equivalent to a unit vector $\mathbf{x} \in R_3$. $\mathbf{x}$ in turn defines a point on the unit sphere S. Thus a probability distribution that describes random variation in observations of directions can be reduced to a probability distribution $dF(\mathbf{x})$, with $\mathbf{x} \in S$.

DEFINITIONS. A distribution $dF(\mathbf{x})$ on S is said to be *circularly symmetric about axis* $\boldsymbol{\mu} \in S$ if $\mathbf{x}_1^T\boldsymbol{\mu} = \mathbf{x}_2^T\boldsymbol{\mu}$ implies that $dF(\mathbf{x}_1) = dF(\mathbf{x}_2)$. A distribution dF is said to be *antipodally symmetric* if $dF(-\mathbf{x}) = dF(\mathbf{x})$, all $\mathbf{x} \in S$ (i.e., opposite points on S have equal probability).

Many measurements of directions are incomplete in that only an axis is specified and not a direction along that axis. There is an obvious correspondence between distributions that may describe random variation of such measurements and antipodally symmetric distributions on S.

The number of probability distributions on the sphere that have been proposed is small—the Fisher–von Mises distribution (Arnold (1941), Fisher (1953)), the Brownian motion distribution (Arnold (1941), Roberts and Ursell (1960)), and the girdle distribution described independently by Dimroth (1962, 1963) and Watson (1965). These are all circularly symmetric about some axis. Only the Dimroth–Watson distribution is antipodally symmetric, although antipodally

Received September 1971; revised January 1974.

[1] This research was supported in part by Research Grant No. NSF GP 32037x from the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation and by Contract No. Navy N00014-67-A-0009 from the Statistics Branch, Office of Naval Research. Also this research was supported in part in connection with research at Princeton University sponsored by the U.S. Army Research Office (Durham) and by a fellowship from the Connecticut Agricultural Experiment Station held at Yale University.

AMS 1970 *subject classifications*. Primary 62E99; Secondary 33A30, 62F05, 62F10.

Key words and phrases. Distribution on sphere, distribution of directions, distribution of axes, test of isotropy, test of circularity, hypergeometric functions.

symmetric relatives of the others (formed by $\frac{1}{2}(dF(\mathbf{x}) + dF(-\mathbf{x}))$ have been suggested.

We are concerned here with a generalization of the Dimroth–Watson distribution. The family to be examined is, in one parametrization,

$$\Psi(\mathbf{x})\, dS/4\pi = (K(A))^{-1} \exp(\mathbf{x}^T A \mathbf{x})\, dS/4\pi\,, \qquad \mathbf{x} \in S\,, \tag{1.1}$$

where A is a symmetric 3 by 3 matrix, dS represents Lebesgue (invariant) measure on S, and $K(A)$ is a normalizing constant. Ψ is antipodally symmetric, but is not, in general, circularly symmetric about any axis. Ψ can be shown to be identical to a distribution proposed by Breitenberger (1963) as one analogue on the sphere of the ordinary normal distribution. We shall see that $-A$ can be assumed to be positive definite. Thus one way of describing Ψ is as the conditional distribution given $||\mathbf{x}||^2 = 1$, when $\mathbf{x}$ has a trivariate normal distribution with zero mean and covariance matrix $-\frac{1}{2}A^{-1}$.

We can rewrite (1.1) as follows. Since A is symmetric, $A = MZM^T$, where $M = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3]$ is orthogonal and $Z = \operatorname{diag}[\zeta_1, \zeta_2, \zeta_3]$. Then $\mathbf{x}^T A\mathbf{x} = \operatorname{tr} A\mathbf{x}\mathbf{x}^T = \operatorname{tr} MZM^T\mathbf{x}\mathbf{x}^T = \operatorname{tr} ZM^T\mathbf{x}\mathbf{x}^T M = \sum_{i=1}^3 \zeta_i(\boldsymbol{\mu}_i{}^T\mathbf{x})^2$. We will take the standard form of Ψ to be

$$\Psi(\mathbf{x}; Z, M)\, dS/4\pi = (F_{000}(Z))^{-1} \operatorname{etr}(ZM^T\mathbf{x}\mathbf{x}^T M)\, dS/4\pi\,, \tag{1.2}$$

where $\operatorname{etr}(\cdot) \equiv \exp(\operatorname{tr}(\cdot))$. The normalizing constant is independent of M and has been written as a function of Z, in conformity with the notation in Section 2. For definiteness, subscripts will be assigned to the ζ's in such a way that $\zeta_1 \leqq \zeta_2 \leqq \zeta_3$.

The function $F_{000}(Z)$ and related functions are discussed in Section 2. Various particular and limiting cases are discussed in Section 3. Sections 4 and 5 examine the exact and asymptotic sampling distributions of various statistics of interest. In Section 6, maximum likelihood estimators of the parameters are shown to depend on the eigenvalues and eigenvectors of the cross product matrix XX^T of the observations. These estimators are shown to be asymptotically normal and their asymptotic covariance structure is derived. Inference procedures, including tests of isotropy and circular symmetry as well as confidence regions for parameters, are developed in Section 7. The application of Ψ is illustrated by a numerical example in Section 8.

Results concerning asymptotic normality are stated throughout more or less in the form "As $\nu \to \infty$, T is asymptotically $N(\theta, \nu^{-1}\Sigma)$," since this is the form in which they are likely to be applied. Such statements are, of course, to be interpreted as "$\nu^{-\frac{1}{2}}(T - \theta)$ converges in distribution to $N(0, \Sigma)$ as $\nu \to \infty$." Similarly, statements concerning asymptotic independence of random quantities are to be interpreted as meaning that the joint distribution of the variates, suitably normalized, converges to the appropriate product of limiting distributions.

2. The normalizing constant and related functions. The normalizing constant is expressible as

$$F_{000}(Z) = (4\pi)^{-1} \int_S \operatorname{etr}(Z\mathbf{x}\mathbf{x}^T)\, dS = {}_1F_1(\tfrac{1}{2}; \tfrac{3}{2}; Z)\,, \tag{2.1}$$

where ${}_1F_1$ is a confluent hypergeometric function of matrix argument as defined by Herz (1955) and expanded in zonal polynomials by Constantine (1963). $F_{000}(Z)$ is symmetric in the ζ's. A particular case is

$$F_{000}(\operatorname{diag}[\zeta, 0, 0]) = {}_1F_1(\tfrac{1}{2}; \tfrac{3}{2}; \zeta)\,, \tag{2.2}$$

where ${}_1F_1$ is now an ordinary confluent hypergeometric function (Erdélyi (1953), page 248). Readily computable power series and asymptotic series for F_{000} and its derivatives can be found (Bingham (1964)).

LEMMA 2.1. *Let ζ_0 be arbitrary and let $\tilde{Z} = Z - \zeta_0 I_3$, where I_3 is the identity matrix. Then*

$$F_{000}(\tilde{Z}) = \exp(-\zeta_0) F_{000}(Z) \tag{2.3}$$

and

$$\Psi(\mathbf{x}; \tilde{Z}, M) = \Psi(\mathbf{x}; Z, M)\,. \tag{2.4}$$

PROOF. Consider the exponent in (1.2): $\operatorname{tr} ZM^{\prime}\ \mathbf{x}^T M = \operatorname{tr} \tilde{Z} M^T \mathbf{x}\mathbf{x}^T M + \zeta_0 \operatorname{tr} M^T \mathbf{x}\mathbf{x}^T M = \operatorname{tr} \tilde{Z} M^T \mathbf{x}\mathbf{x}^T M + \zeta_0 ||M^T\mathbf{x}||^2 = \operatorname{tr} \tilde{Z} M^T \mathbf{x}\mathbf{x}^T M + \zeta_0$, since an orthogonal transformation is length preserving. Thus

$$\Psi(\mathbf{x}; Z, M)/4\pi = [F_{000}(Z)]^{-1} \exp(\zeta_0) F_{000}(\tilde{Z}) \Psi(\mathbf{x}; \tilde{Z}, M)/4\pi\,.$$

Since $\Psi/4\pi$ integrates to unity on both sides, (2.3) and (2.4) follow. □
Lemma 2.1 points up a degeneracy in the specification of Ψ. The shape parameters ζ_j are determined only up to an additive constant. When uniqueness is convenient we can impose a restraint. For the numerical example in Section 8, $\zeta_3 = 0$ is imposed.

We use the following notations for the derivatives of $F_{000}(Z)$:

$$F_{i_1 i_2 i_3}(Z) = [\textstyle\prod_{j=1}^{3} (\partial/\partial\zeta_j)^{i_j}] F_{000}(Z)\,, \tag{2.5}$$

and, interchangeably,

$$F^{(k)}_{j_1 j_2 \cdots j_k}(Z) = [\textstyle\prod_{i=1}^{k} (\partial/\partial\zeta_{j_i})] F^{(0)}(Z)\,, \qquad F^{(0)}(Z) = F_{000}(Z)\,. \tag{2.6}$$

The logarithmic derivatives of $F_{000}(Z)$ are denoted by

$$Y_{i_1 i_2 i_3}(Z) = [\textstyle\prod_{j=1}^{3} (\partial/\partial\zeta_j)^{i_j}] Y_{000}(Z)\,, \qquad Y_{000}(Z) = \log F_{000}(Z)\,, \tag{2.7}$$

and, interchangeably,

$$Y^{(k)}_{j_1 j_2 \cdots j_k}(Z) = [\textstyle\prod_{i=1}^{k} (\partial/\partial\zeta_{j_i})] Y^{(0)}(Z)\,, \qquad Y^{(0)}(Z) = Y_{000}(Z)\,. \tag{2.8}$$

LEMMA 2.2. *Let $\mathbf{x} = [x_1, x_2, x_3]^T$ be distributed according to $\Psi(\mathbf{x}; Z, I_3)$. Then*

$$F_{i_1 i_2 i_3}(Z)/F_{000}(Z) = E[(x_1^2)^{i_1} (x_2^2)^{i_2} (x_3^2)^{i_3}] \tag{2.9}$$

and

$$Y_{i_1i_2i_3}(Z) = \kappa_{i_1i_2i_3}(x_1^2, x_2^2, x_3^2)\ , \tag{2.10}$$

the joint cumulant of order (i_1, i_2, i_3) *of random vector* $[x_1^2, x_2^2, x_3^2]^T$.

PROOF. From (2.1) and (2.5),

$$F_{i_1i_2i_3}(Z) = (4\pi)^{-1} \int_S x_1^{2i_1}x_2^{2i_2}x_3^{2i_3} \operatorname{etr}(Z\mathbf{x}\mathbf{x}^T)\, dS\ . \tag{2.11}$$

(2.9) now follows from (2.11) and (1.2).

Using (2.2) and Taylor's theorem, (2.10) follows from

$$\log E[\exp(\textstyle\sum_{j=1}^3 t_j x_j^2)] = \log F_{000}(Z+T)/F_{000}(Z) = Y_{000}(Z+T) - Y_{000}(Z)\ ,$$

where $T = \operatorname{diag}[t_1, t_2, t_3]$. □

LEMMA 2.3. (a) *The following equations are valid*:

$$F_{i_1+1,i_2,i_3}(Z) + F_{i_1,i_2+1,i_3}(Z) + F_{i_1,i_2,i_3+1}(Z) = F_{i_1i_2i_3}(Z)\ ; \tag{2.12}$$

$$Y_j^{(1)}(Z) = F_j^{(1)}(Z)/F^{(0)}(Z) > 0\ , \qquad j = 1, 2, 3\ ; \tag{2.13}$$

$$\textstyle\sum_{j=1}^3 Y_j^{(1)}(Z) = 1\ ; \tag{2.14}$$

$$\textstyle\sum_{j=1}^3 Y_{jk}^{(2)}(Z) = 0\ , \qquad k = 1, 2, 3\ ; \tag{2.15}$$

(b) *The matrix* $[Y_{ij}^{(2)}(Z)]_{i,j\leqq 3}$ *is of rank* 2 *and is nonnegative definite. Moreover*

$$Y_{ij}^{(2)}(Z) = F_{ij}^{(2)}(Z)/F^{(0)}(Z) - Y_i^{(1)}(Z)Y_j^{(1)}(Z)\ , \qquad i, j = 1, 2, 3\ ; \tag{2.16}$$

and

$$Y_{jj}^{(2)}(Z) > 0\ , \qquad j = 1, 2, 3\ . \tag{2.17}$$

(c) *Let* $\tilde{Z}$ *be as in Lemma* 2.1. *Then*

$$Y^{(0)}(\tilde{Z}) = -\zeta_0 + Y^{(0)}(Z) \quad \text{and} \quad Y_{j_1j_2\cdots j_k}^{(k)}(\tilde{Z}) = Y_{j_1j_2\cdots j_k}^{(k)}(Z)\ , \quad k \geqq 1\ . \tag{2.18}$$

PROOF. Since $x_1^2 + x_2^2 + x_3^2 = 1$, (2.12) follows from (2.11). (2.13) is clear. Dividing (2.12) with $i_1 = i_2 = i_3 = 0$ by $F_{000}(Z)$ yields (2.14) and then (2.15) follows by differentiation with respect to ζ_k. Lemma 2.2 and the properties of covariance matrices imply (b) since the vector $[x_1^2, x_2^2, x_3^2]^T$ lies in a linear manifold of rank 2. (2.18) follows from (2.3) and the definitions of the Y functions. □

The proof of the following lemma depends on the power series expansion for $F_{000}(Z)$ and will be omitted.

LEMMA 2.4. *For arbitrary* $Z = \operatorname{diag}[\zeta_1, \zeta_2, \zeta_3]$ *and* $i \neq j$,

$$F_i^{(1)}(Z) - F_j^{(1)}(Z) = 2(\zeta_i - \zeta_j)F_{ij}^{(2)}(Z)\ . \tag{2.19}$$

COROLLARY. *When* $\zeta_i \neq \zeta_j$,

$$A_{ij}(Z) = \tfrac{1}{2}[Y_i^{(1)}(Z) - Y_j^{(1)}(Z)]/(\zeta_i - \zeta_j)\ , \tag{2.20}$$

where

$$A_{ij}(Z) = F^{(2)}_{ij}(Z)/F^{(0)}(Z) . \tag{2.21}$$

PROOF. This follows immediately from (2.19) and (2.13). □

3. Particular and limiting cases. When $Z = \zeta_0 I_3$, $\Psi(\mathbf{x}; Z, M) = 1$ and hence $\Psi\, dS/4\pi$ is the uniform or isotropic distribution on S. In particular this is true if $Z = \text{diag}\,[0, 0, 0]$.

When $\zeta_1 \leqq \zeta_2 = \zeta_3$, (2.3) (with $\zeta_0 = \zeta_3$) and (2.2) imply that Ψ reduces to

$$\psi(\mathbf{x}; -\kappa, \boldsymbol{\mu})\, dS/4\pi = {}_1F_1(\tfrac{1}{2}; \tfrac{3}{2}; -\kappa)^{-1} \exp(-\kappa(\boldsymbol{\mu}^T\mathbf{x})^2)\, dS/4\pi , \tag{3.1}$$

where $\kappa = \zeta_2 - \zeta_1 \geqq 0$ and $\boldsymbol{\mu} = \boldsymbol{\mu}_1$. This is the Dimroth–Watson girdle distribution in the form discussed by Watson (1965). For large κ (see Theorem 3.1 below), the latitude of $\mathbf{x}$ (taking $\boldsymbol{\mu}$ as pole) $\lambda \simeq \sin \lambda = \boldsymbol{\mu}^T\mathbf{x}$ is approximately $N(0, (2\kappa)^{-1})$, independently of the longitude which is uniformly distributed on $[0, 2\pi)$. This is, in fact, essentially the form in which Dimroth (1962) introduced the distribution.

When $\zeta_1 = \zeta_2 \leqq \zeta_3$, Ψ reduces to

$$\psi(\mathbf{x}; \tau, \boldsymbol{\mu})\, dS/4\pi = ({}_1F_1(\tfrac{1}{2}; \tfrac{3}{2}; \tau))^{-1} \exp(\tau(\boldsymbol{\mu}^T\mathbf{x})^2)\, dS/4\pi , \tag{3.2}$$

where $\tau = \zeta_3 - \zeta_2 \geqq 0$ and $\boldsymbol{\mu} = \boldsymbol{\mu}_3$. This can be considered to be a polar version of the Dimroth–Watson distribution, being concentrated with circular symmetry around $\boldsymbol{\mu}$ and $-\boldsymbol{\mu}$. By allowing τ in (3.2) (or κ in (3.1)) to vary between $-\infty$ and $+\infty$, both versions of the Dimroth–Watson distribution are obtained.

When $\kappa_1 = \zeta_2 - \zeta_1$ and $\kappa_2 = \zeta_3 - \zeta_1$ are large but not equal, Ψ tends to be concentrated, without circular symmetry, in a "girdle" along the "equator" orthogonal to $\boldsymbol{\mu}_1$ and is a generalization of the Dimroth–Watson girdle distribution. When $\tau_1 = \zeta_3 - \zeta_1$ and $\tau_2 = \zeta_3 - \zeta_2$ are large but not equal, Ψ is concentrated near $\boldsymbol{\mu}_3$ and $-\boldsymbol{\mu}_3$ and generalizes the polar form of the Dimroth–Watson distribution.

THEOREM 3.1. *Define spherical coordinates t_1 and t_2 by $\mathbf{x}^T M = [\cos t_1, \sin t_1 \cos t_2, \sin t_1 \sin t_2]$. Let $\mathbf{x}$ be distributed according to Ψ. Then, as $\kappa = \frac{1}{2}(\kappa_1 + \kappa_2) \to \infty$ while $\delta = \frac{1}{2}(\kappa_1 - \kappa_2)$ remains bounded, $z = \cos t_1$ is asymptotically $N(0, (2\kappa)^{-1})$ independently of t_2 which has the limiting distribution* const $\exp(\delta \cos 2t_2)\, dt_2$.

PROOF.

$$\begin{aligned}\Psi(\mathbf{x}; Z, M)\, dS/4\pi &= \text{const} \exp[(\kappa_1 \cos^2 t_2 + \kappa_2 \sin^2 t_2) \sin^2 t_1] \sin t_1\, dt_1\, dt_2 \\ &= \text{const} \exp(\kappa \sin^2 t_1) \exp(\delta \sin^2 t_1 \cos 2t_2) \sin t_1\, dt_1\, dt_2 \\ &= \text{const} \exp(-\kappa z^2)\, dz \exp[(1 - z^2)\delta \cos 2t_2]\, dt_2 .\end{aligned}$$

As $\kappa \to \infty$ while δ remains bounded $z^2 = O_p(\kappa^{-1})$ and hence the factor $(1 - z^2)$ in the exponent may be ignored. The conclusion follows immediately. □

THEOREM 3.2. *Let $y_1 = \boldsymbol{\mu}_1^T\mathbf{x}$ and $y_2 = \boldsymbol{\mu}_2^T\mathbf{x}$, and let $\mathbf{x}$ be distributed according* *to* Ψ. *Then, as* $\tau_i \to \infty$, $i = 1, 2$, y_1 *and* y_2 *are asymptotically independently* $N(0, (2\tau_i)^{-1})$.

PROOF. $\Psi(\mathbf{x}; Z, M)\, dS/4\pi = \text{const} \exp(-\tau_1 y_1^2 - \tau_2 y_2^2)(1 - y_1^2 - y_2^2)^{-\frac{1}{2}}\, dy_1\, dy_2$. For large τ_1 and τ_2, $y_i = O_p(\tau_i^{-1})$, $i = 1, 2$. Thus the factor $(1 - y_1^2 - y_2^2)^{-\frac{1}{2}}$ can be ignored without asymptotic error. □

Theorems 3.1 and 3.2 to some extent overlap in their application to finite values of the ζ's, because large τ_1 and τ_2 can be viewed as large but bounded τ_2 and $\tau_1 \to \infty$. This last is equivalent to the conditions of Theorem 3.1. This relationship reflects the fact that a distribution that is strongly concentrated in an elliptical pattern near $\boldsymbol{\mu}_3$ and $-\boldsymbol{\mu}_3$ can also be considered to be a girdle distribution with a high degree of circular asymmetry near the "equator" defined by $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$.

4. Exact sampling distributions. Throughout this and subsequent sections, we assume that the 3 by n matrix

$$X = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \tag{4.1}$$

represents a random sample of $n \geqq 3$ unit vectors $\mathbf{x}_i$ distributed according to $\Psi(\mathbf{x}; Z, M)$.

LEMMA 4.1. *The sampling distribution of the cross product matrix* $XX^T = \sum_{j=1}^n \mathbf{x}_j \mathbf{x}_j^T$ *is*

$$dG_n(XX^T; Z, M) = [F_{000}(Z)]^{-n} \operatorname{etr}(ZM^T XX^T M)\, dH_n(XX^T)\,, \tag{4.2}$$

where $dH_n(XX^T) = dG_n(XX^T; 0, I_3)$ *is the sampling distribution of* XX^T *in the case of isotropy.*

PROOF. The sampling distribution of XX^T is

$$dG_n(XX^T; Z, M) = \textstyle\int_{XX^T} \prod_{j=1}^n [\Psi(\mathbf{x}_j; Z, M)\, dS_j/4\pi]\,,$$

where the integral is over the manifold defined by holding XX^T fixed. From (1.2)

$$\begin{aligned}\textstyle\prod_{j=1}^n \Psi(\mathbf{x}_j; Z, M) &= [F_{000}(Z)]^{-n} \exp(\textstyle\sum_{j=1}^n \operatorname{tr} ZM^T \mathbf{x}_j \mathbf{x}_j^T M) \\ &= [F_{000}(Z)]^{-n} \operatorname{etr}(ZM^T XX^T M).\end{aligned}$$

Since $dH_n(XX^T) = \int_{XX^T} \prod_{j=1}^n [dS_j/4\pi]$, the result follows. □
The exact form of $dH_n(XX^T)$ is unknown, although Stephens (1965) has investigated the distribution of the diagonal elements of XX^T.

With probability one, the symmetric matrix XX^T is positive definite and can be factored as

$$XX^T = \hat{M}\Omega\hat{M}^T\,, \tag{4.3}$$

where $\hat{M} = [\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\mu}}_3]$ is orthogonal, $\Omega = \operatorname{diag}[\omega_1, \omega_2, \omega_3]$, and $0 < \omega_1 < \omega_2 < \omega_3 < n$. Equivalently

$$\Omega = \hat{M}^T XX^T \hat{M}\,. \tag{4.4}$$

The columns $\hat{\mu}_i$ of $\hat{M}$ are the eigenvectors of XX^T with eigenvalues ω_i, $i = 1, 2, 3$, and are uniquely determined (with probability one) up to multiplication by ± 1. When convenient we may assume that $\hat{M}$ represents a proper rotation, i.e., $\det \hat{M} = +1$.

The following identity will be useful:

(4.5) $$\operatorname{tr} \Omega = \sum_{j=1}^{3} \omega_j = n .$$

LEMMA 4.2. *In the isotropic case* ($Z = 0$), $\hat{M}$ *and* Ω *are independent. In fact*

$$dH_n(XX^T) = dF_n(\Omega; 0)(8\pi^2)^{-1}(d\hat{M})^+ ,$$

where $(8\pi^2)^{-1}(d\hat{M})^+$ *is the invariant* (Haar) *probability measure on the group* $O^+(3)$ *of proper orthogonal matrices* (James (1954)), *and* $dF_n(\Omega; 0)$ *defines a probability measure on the space*

(4.6) $$\{\Omega = \operatorname{diag}[\omega_1, \omega_2, \omega_3], 0 < \omega_1 < \omega_2 < \omega_3, \omega_1 + \omega_2 + \omega_3 = n\} .$$

PROOF. Isotropy implies that both the marginal distribution of $\hat{M}$ and its distribution conditional on Ω are invariant under $\hat{M} \to H^T\hat{M}$, $H \in O^+(3)$. Thus both distributions must be $(8\pi^2)^{-1}(d\hat{M})^+$ (James (1954)). This implies independence. The indicated decomposition follows immediately. □

THEOREM 4.1. *The joint distribution of* Ω *and* $\hat{M}$ *is*

(4.7) $$[F_{000}(Z)]^{-n} \operatorname{etr}(ZM^T\hat{M}\Omega\hat{M}^TM)\, dF_n(\Omega; 0)(8\pi^2)^{-1}(d\hat{M})^+ ,$$

where dF_n *and* $(d\hat{M})^+$ *are as in Lemma* 4.2.

PROOF. The theorem follows directly from Lemmas 4.1 and 4.2. □

COROLLARY. *The marginal distribution of* Ω *is*

(4.8) $$dF_n(\Omega; Z) = [F_{000}(Z)]^{-n}\, {}_0F_0^{(3)}(Z, \Omega)\, dF_n(\Omega; 0) ,$$

and the conditional distribution of $\hat{M}$ *given* Ω *is*

(4.9) $$\Phi(\hat{M}; MZM^T \mid \Omega)(8\pi^2)^{-1}(d\hat{M})^+ = [{}_0F_0^{(3)}(Z, \Omega)]^{-1} \operatorname{etr}(ZM^T\hat{M}\Omega\hat{M}^TM)(8\pi^2)^{-1}(d\hat{M})^+ ,$$

where

(4.10) $${}_0F_0^{(3)}(Z, \Omega) = (8\pi^2)^{-1} \int_{O^+(3)} \operatorname{etr}(ZH\Omega H^T)(dH)^+$$

is a generalized hypergeometric function of two matrix arguments (James (1964)).

PROOF. (4.8) follows from (4.7) and (4.10) by making a change of variables from $\hat{M}$ to $H = M^T\hat{M}$. (4.9) is then obvious. □

A zonal polynomial expansion for ${}_0F_0^{(3)}$ was given by James (1960 and 1964). It enters as a factor in the distribution of the eigenvalues of a Wishart distributed matrix. Φ can be considered to be an analogue of Ψ on $O^+(3)$, and is, after suitable identification of parameters, the distribution of the eigenvectors of a Wishart distributed matrix conditional on its eigenvalues. Anderson (1965) gives an asymptotic series for ${}_0F_0^{(3)}$ that is valid for large values of $n\Delta_{ij}$, where

(4.11) $$\Delta_{ij} = (\zeta_i - \zeta_j)(\omega_i - \omega_j)/n , \qquad 1 \leqq i < j \leqq 3 .$$

Bingham (1972) has conjectured a series of products of confluent hypergeometric functions. Although the factor $dF_n(\Omega; 0)$ is unknown, likelihood inference procedures derived from (4.8) are possible since the factor depending on Z is known.

5. Asymptotic sampling theory. The results in this section are largely based on the following theorem.

THEOREM 5.1. *Let X be as in* (4.1). *Define $x_{ij}(M)$ by*

$$M^T XX^T M = [\boldsymbol{\mu}_i^T XX^T \boldsymbol{\mu}_j]_{i,j\leq 3} = [x_{ij}(M)]_{i,j\leq 3}\,. \tag{5.1}$$

Then as $n \to \infty$, the vector

$$\mathbf{T}(M) = n^{-1}[x_{11}(M), x_{22}(M), x_{33}(M), x_{12}(M), x_{13}(M), x_{23}(M)]^T \tag{5.2}$$

is asymptotically normal with expectation and covariance matrix

$$E[\mathbf{T}(M)] = [Y_1^{(1)}(Z), Y_2^{(1)}(Z), Y_3^{(1)}(Z), 0, 0, 0]^T\,, \tag{5.3}$$

$$\operatorname{Cov}[\mathbf{T}(M)] = n^{-1}\Gamma(Z) = n^{-1}\,\text{block diag}\,[\Gamma_1(Z), \Gamma_2(Z)]\,, \tag{5.4}$$

where

$$\Gamma_1(Z) = [Y_{ij}^{(2)}(Z)]_{i,j\leq 3} \tag{5.5}$$

$$\Gamma_2(Z) = \operatorname{diag}[A_{12}(Z), A_{13}(Z), A_{23}(Z)]\,, \qquad A_{ij}(Z) \text{ as in } (2.21). \tag{5.6}$$

PROOF. Since $M^T XX^T M = \sum_{j=1}^n (M^T\mathbf{x}_j\mathbf{x}_j^T M)$, $\mathbf{T}(M) = \bar{\mathbf{t}}(M)$ is the average of n bounded i.i.d. random vectors $\mathbf{t}_j(M)$ where $\mathbf{t}_j(M)$ is related to $M^T\mathbf{x}_j\mathbf{x}_j^T M$ in a similar way as $\mathbf{T}(M)$ is to $M^T XX^T M$. By the Central Limit Theorem, $\mathbf{T}(M)$ is asymptotically $N(E[\mathbf{t}_j(M)], n^{-1}\operatorname{Cov}[\mathbf{t}_j(M)])$. When $\mathbf{x} = [x_1, x_2, x_3]^T$ is distributed according to $\Psi(\mathbf{x}; Z, M)$, $M^T\mathbf{x}$ is distributed according to $\Psi(\mathbf{x}; Z, I_3)$. By symmetry, when $M = I_3$, $E[x_i x_j] = 0$, $i \neq j$, and $E[x_i x_j x_k x_l] = 0$, when only one pair of subscripts are the same. The remainder of the result then follows from Lemma 2.2. □

COROLLARY. *Let $Z = 0$ (isotropy), $M \in O^+(3)$ arbitrary, and $\mathbf{T}(M)$ defined by* (5.2). *Then as $n \to \infty$, $\mathbf{T}(M)$ is asymptotically normal with*

$$E[\mathbf{T}(M)] = [\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}, 0, 0, 0]^T\,, \tag{5.7}$$

$$\begin{aligned}\operatorname{Cov}[\mathbf{T}(M)] &= n^{-1}\Gamma(0) \\ &= (45n)^{-1}\,\text{block diag}\,[6I_3 - 2\mathbf{1}\mathbf{1}^T, 3I_3]\,, \qquad \mathbf{1} = [1, 1, 1]^T\,.\end{aligned} \tag{5.8}$$

PROOF. When $Z = 0$, (5.3) and (5.4) reduce to (5.7) and (5.8). □

THEOREM 5.2. *Define*

$$X_U^2 = (15/2n)\operatorname{tr}(XX^T - (n/3)I_3)^2 = (15/2n)\textstyle\sum_{j=1}^3 (\omega_j - n/3)^2\,. \tag{5.9}$$

Then when $Z = 0$, as $n \to \infty$, X_U^2 is asymptotically distributed as $\chi^2(5)$ (chi-squared on 5 degrees of freedom).

PROOF. A generalized inverse (see Rao (1965), page 20) to $\Gamma(0)$ in (5.8) is

$\Gamma^- = (15/2)$ block diag $[I_3, 2I_3]$. Using the preceding corollary and Rao (1965, page 443), letting $\mathbf{T} = \mathbf{T}(I_3)$, $n(\mathbf{T} - E[\mathbf{T}])^T\Gamma^-(\mathbf{T} - E[\mathbf{T}])$ is asymptotically $\chi^2(5)$ since $\Gamma(0)$ has rank 5. Putting $x_{ij} \equiv x_{ij}(I_3)$, this is $(15/2n)\sum_{i=1}^3 (x_{ii} - n/3)^2 + (15/n)\sum_{i<j}^3 x_{ij}^2 = (15/2n)\,\mathrm{tr}\,(XX^T - (n/3)I_3)^2$. □

Theorem 5.2 can also be proved directly from the asymptotic distribution for Ω when $Z = 0$, as given by Anderson and Stephens (1973).

LEMMA 5.1. *Let $\zeta_1 = \zeta_2 < \zeta_3$ (circular symmetry about $\boldsymbol{\mu}_3$). Then as $n \to \infty$, $\frac{1}{2}n^{-1}(x_{11}(M) - x_{22}(M))$ and $n^{-1}x_{12}(M)$ are asymptotically independent $N(0, n^{-1}A_{12}(Z))$ and independent of $n^{-1}x_{33}(M)$ which is $N(Y_3^{(1)}(Z), n^{-1}Y_{33}^{(2)}(Z))$. A similar result holds when $\zeta_1 < \zeta_2 = \zeta_3$.*

PROOF. By symmetry, when $\zeta_1 = \zeta_2$, the joint distributions of $(x_{11}(M), x_{33}(M))$ and $(x_{22}(M), x_{33}(M))$ are identical. This implies that $E[x_{11}(M) - x_{22}(M)] = 0$ and $\mathrm{Cov}\,[x_{11}(M) - x_{22}(M), x_{33}(M)] = 0$. By Theorem 5.1 $\mathrm{Var}\,[\frac{1}{2}(x_{11}(M) - x_{22}(M))] = \frac{1}{4}n(Y_{11}^{(2)}(Z) + Y_{22}^{(2)}(Z) - 2Y_{12}^{(2)}(Z))$. When $\zeta_1 = \zeta_2$ this is $\frac{1}{2}n(Y_{11}^{(2)}(Z) - Y_{12}^{(2)}(Z)) = \frac{1}{2}n(F_{11}^{(2)}(Z) - F_{12}^{(2)}(Z))/F^{(0)}(Z) = \frac{1}{2}nE[(\boldsymbol{\mu}_1^T\mathbf{x})^4 - (\boldsymbol{\mu}_1^T\mathbf{x})^2(\boldsymbol{\mu}_2^T\mathbf{x})^2]$. It is easily checked that when $\zeta_1 = \zeta_2$, $E[(\boldsymbol{\mu}_1^T\mathbf{x})^4] = 3E[(\boldsymbol{\mu}_1^T\mathbf{x})^2(\boldsymbol{\mu}_2^T\mathbf{x})^2]$. Thus $\mathrm{Var}\,[\frac{1}{2}(x_{11}(M) - x_{22}(M))] = nE[(\boldsymbol{\mu}_1^T\mathbf{x})^2(\boldsymbol{\mu}_2^T\mathbf{x})^2] = nF_{12}^{(2)}(Z)/F^{(0)}(Z) = nA_{12}(Z)$. By Theorem 5.1, $n^{-1}x_{12}(M)$ is $N(0, n^{-1}A_{12}(Z))$ and is uncorrelated with $x_{11}(M)$, $x_{22}(M)$, and $x_{33}(M)$. □

Note that when $\zeta_i = \zeta_j$, $i \neq j$, $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ are not well defined. However, they may be assigned in any way that does not depend on the data to make M orthogonal.

THEOREM 5.3. *Let $\omega_1(\boldsymbol{\mu}_3)$ and $\omega_2(\boldsymbol{\mu}_3)$ be the eigenvalues of the two by two submatrix $[x_{ij}(M)]_{i,j\le 2}$. Then when $\zeta_1 = \zeta_2 < \zeta_3$, as $n \to \infty$,*

$$X_C^2(\boldsymbol{\mu}_3) = (4nA_{12}(Z))^{-1}(\omega_1(\boldsymbol{\mu}_3) - \omega_2(\boldsymbol{\mu}_3))^2 \tag{5.10}$$

is asymptotically distributed as $\chi^2(2)$ and is asymptotically independent of $x_{33}(M) = \boldsymbol{\mu}_3^T XX^T\boldsymbol{\mu}_3$. A similar result holds when $\zeta_1 < \zeta_2 = \zeta_3$. Note that $\omega_i(\boldsymbol{\mu}_3)$ $i = 1, 2$, depend only on $\boldsymbol{\mu}_3$ and not on $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$.

PROOF. By the preceding lemma, as $n \to \infty$, $(4nA_{12}(Z))^{-1}[(x_{11}(M) - x_{22}(M))^2 + 4x_{12}^2(M)]$ is asymptotically $\chi^2(2)$ independently of $x_{33}(M)$. But $(x_{11}(M) - x_{22}(M))^2 + 4x_{12}^2(M) = (\mathrm{tr}\,[x_{ij}(M)]_{i,j\le 2})^2 - 4\det[x_{ij}(M)]_{i,j\le 2} = (\omega_1(\boldsymbol{\mu}_3) + \omega_2(\boldsymbol{\mu}_3))^2 - 4\omega_1(\boldsymbol{\mu}_3)\omega_2(\boldsymbol{\mu}_3) = (\omega_1(\boldsymbol{\mu}_3) - \omega_2(\boldsymbol{\mu}_3))^2$. □

The asymptotic distribution of Ω is also of some interest. The distribution in the isotropic case is given by Anderson and Stephens (1973). When $\zeta_1 < \zeta_2 < \zeta_3$, $[\omega_1, \omega_2, \omega_3]^T$ has essentially the same asymptotic distribution as $[x_{11}(M), x_{22}(M), x_{33}(M)]^T$, although correction terms for the mean and covariance matrix can be found. This is made more precise in the following theorem.

THEOREM 5.4. *Let $\zeta_1 < \zeta_2 < \zeta_3$. Then as $n \to \infty$, $\mathbf{W} = n^{-1}[\omega_1, \omega_2, \omega_3]^T$ is* *asymptotically normal with*

$$E[n^{-1}\omega_i] = Y_i^{(1)}(Z) + \tfrac{1}{2}n^{-1}\textstyle\sum_{j\neq i}(\zeta_i - \zeta_j)^{-1} + \tfrac{1}{4}n^{-2}\sum_{j\neq i}[(\zeta_i - \zeta_j)\Delta_{ij}^0]^{-1} + O(n^{-3}), \tag{5.11}$$

$$\operatorname{Var}[n^{-1}\omega_i] = n^{-1}Y_{ii}^{(2)}(Z) - \tfrac{1}{2}n^{-2}\textstyle\sum_{j\neq i}(\zeta_i - \zeta_j)^{-2} + O(n^{-3}), \tag{5.12}$$

$$\operatorname{Cov}[n^{-1}\omega_i, n^{-1}\omega_j] = n^{-1}Y_{ij}^{(2)}(Z) + \tfrac{1}{2}n^{-2}(\zeta_i - \zeta_j)^{-2} + O(n^{-3}), \quad i \neq j, \tag{5.13}$$

where

$$\Delta_{ij}^0 = (\zeta_i - \zeta_j)(Y_i^{(1)}(Z) - Y_j^{(1)}(Z)). \tag{5.14}$$

PROOF. Using the corollary to Theorem 4.1, it may be shown that the cumulant generating function of $[x_{11}(M), x_{22}(M), x_{33}(M)]^T$ conditional on Ω is

$$\log E[\operatorname{etr}(TM^TXX^TM) \mid \Omega] = \log({}_0F_0^{(3)}(Z + T, \Omega)/{}_0F_0^{(3)}(Z, \Omega)),$$

where $T = \operatorname{diag}[t_1, t_2, t_3]$. From Anderson's (1965) asymptotic expansion one finds

$$E[\omega_i - x_{ii}(M) \mid \Omega] = \tfrac{1}{2}\textstyle\sum_{j\neq i}(\zeta_i - \zeta_j)^{-1} + \tfrac{1}{4}n^{-1}\sum_{j\neq i}((\zeta_i - \zeta_j)\Delta_{ij})^{-1} + O(\omega^{-2}),$$

$$\operatorname{Var}[\omega_i - x_{ii}(M) \mid \Omega] = \operatorname{Var}[x_{ii}(M) \mid \Omega] = \tfrac{1}{2}\textstyle\sum_{j\neq i}(\zeta_i - \zeta_j)^{-2} + \tfrac{1}{2}n^{-1}\sum_{j\neq i}(\zeta_i - \zeta_j)^{-2}\Delta_{ij}^{-1} + O(\omega^{-2}),$$

$$\begin{aligned}\operatorname{Cov}[\omega_i - x_{ii}(M), \omega_j - x_{jj}(M) \mid \Omega] &= \operatorname{Cov}[x_{ii}(M), x_{jj}(M) \mid \Omega] \\ &= -\tfrac{1}{2}(\zeta_i - \zeta_j)^{-2} - \tfrac{1}{2}n^{-1}(\zeta_i - \zeta_j)^{-2}\Delta_{ij}^{-1} \\ &\quad + O(\omega^{-2}), \qquad i \neq j,\end{aligned}$$

where Δ_{ij} is as in (4.11). By the usual relationship between conditional and unconditional means, and $E[\Delta_{ij}^{-1}] = (\Delta_{ij}^0)^{-1} + O(n^{-1})$, we have

$$E[\omega_i - x_{ii}(M)] = \tfrac{1}{2}\textstyle\sum_{j\neq i}(\zeta_i - \zeta_j)^{-1} + \tfrac{1}{4}n^{-1}\sum_{j\neq i}((\zeta_i - \zeta_j)\Delta_{ij}^0)^{-1} + O(n^{-2}) = O(1), \tag{5.15}$$

proving (5.11). Similarly, the unconditional variance of $\omega_i - x_{ii}(M)$ is $O(1)$. Thus, by Chebyshev's inequality, $\operatorname{plim}_{n\to\infty} n^{-\frac{1}{2}}(\omega_i - x_{ii}(M)) = 0$, $i = 1, 2, 3$. Theorem 5.1 then implies asymptotic normality. Now $nY_{ii}^{(2)}(Z) = \operatorname{Var}[x_{ii}(M)] = E[\operatorname{Var}[x_{ii}(M) \mid \Omega]] + \operatorname{Var}[E[x_{ii}(M) \mid \Omega]]$. It is readily shown that $\operatorname{Var}[E[x_{ii}(M) \mid \Omega]] = \operatorname{Var}[\omega_i] + O(n^{-1})$. Thus by (5.15), $nY_{ii}^{(2)}(Z) = \frac{1}{2}\sum_{j\neq i}(\zeta_i - \zeta_j)^2 + O(n^{-1}) + \operatorname{Var}[\omega_i] + O(n^{-1})$. (5.12) then follows immediately. (5.13) is established analogously. □

Although $n^{-\frac{1}{2}}\omega_i$ and $n^{-\frac{1}{2}}x_{ii}(M)$ are asymptotically equivalent, we can say something about their differences.

THEOREM 5.5. *Let* $\zeta_1 < \zeta_2 < \zeta_3$. *Define*

$$R^2 = 2(\operatorname{tr} Z\Omega - \operatorname{tr} ZM^TXX^TM) = 2[\textstyle\sum_{j=1}^3 \zeta_j(\omega_j - x_{jj}(M))]. \tag{5.16}$$

Then, as $n \to \infty$,

$$R_0^2 = [1 - (6n)^{-1}\textstyle\sum_{i<j}\Delta_{ij}^{-1}]R^2 = R^2 + O_p(n^{-1}) \tag{5.17}$$

is distributed asymptotically as $\chi^2(3)$.

PROOF. As in the proof of Theorem 5.4, the cumulant generating function for R^2, conditional on Ω, can be shown to be

$$\log E[\exp(tR^2) \mid \Omega] = 2t \operatorname{tr} Z\Omega + \log {}_0F_0^{(3)}((1-2t)Z, \Omega) - \log {}_0F_0^{(3)}(Z, \Omega) .$$

A direct calculation based on Anderson's (1965) asymptotic series for ${}_0F_0^{(3)}$ shows that, conditional on Ω, R_0^2 has the same cumulants as $\chi^2(3)$, except for terms that are $O(\omega^{-2})$. Since $\omega_i = O_p(n)$, the unconditional distribution of R_0^2 is also asymptotically that of $\chi^2(3)$. □

The three degrees of freedom in the preceding result can be isolated asymptotically in several ways. The following seems to be helpful from a standpoint of interpretation. We require a suitable parametrization of $\tilde{M} = M^T\hat{M}$ in the three dimensional manifold $O^+(3)$. Every matrix $M \in O^+(3)$ can be expressed as

$$M = [\cos t_{12}\boldsymbol{\nu}_1 + \sin t_{12}\boldsymbol{\nu}_2, -\sin t_{12}\boldsymbol{\nu}_1 + \cos t_{12}\boldsymbol{\nu}_2, \boldsymbol{\nu}_3] , \tag{5.18}$$

where

$$\begin{aligned} \boldsymbol{\nu}_1 &= [\cos t_{13}, -\sin t_{13} \sin t_{23}, -\sin t_{13} \cos t_{23}]^T \\ \boldsymbol{\nu}_2 &= [0, \cos t_{23}, -\sin t_{23}]^T \\ \boldsymbol{\nu}_3 &= [\sin t_{13}, \cos t_{13} \sin t_{23}, \cos t_{13} \cos t_{23}]^T , \end{aligned} \tag{5.19}$$

with $t_{12} \in (-\pi, \pi]$, $t_{13} \in [-\frac{1}{2}\pi, \frac{1}{2}\pi]$, $t_{23} \in (-\pi, \pi]$. The identity matrix I_3 corresponds to $t_{12} = t_{13} = t_{23} = 0$. Let $\tilde{t}_{12}$, $\tilde{t}_{13}$, and $\tilde{t}_{23}$ be these coordinates for $\tilde{M}$. When $\tilde{t}_{ij}$ is small, it is approximately the angle of deviation of $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{\mu}}_j$ from $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ rotating around the remaining column of M. It can be shown, using the methods of James (1954), that with this parametrization of $O^+(3)$, the invariant probability measure on $O^+(3)$ is

$$(8\pi^2)^{-1}(d\tilde{M})^+ = (8\pi^2)^{-1}(\cos \tilde{t}_{13})\, d\tilde{t}_{12}\, d\tilde{t}_{13}\, d\tilde{t}_{23} . \tag{5.20}$$

Algebraic expansion of the conditional distribution Φ of $\hat{M}$ given Ω (4.9), yields

$$\begin{aligned} &\Phi(\hat{M}; MZM^T \mid \Omega)(d\hat{M})^+/(8\pi^2) \\ &\quad = [8\pi^2 \operatorname{etr}(-Z\Omega)\, {}_0F_0^{(3)}(Z, \Omega)]^{-1} \exp(-n \textstyle\sum_{i<j} \Delta_{ij} \tilde{t}_{ij}^2) \\ &\qquad \times [1 + O(\tilde{t}^2) + O(\tilde{t}^3)O(\omega)] \textstyle\prod_{i<j} d\tilde{t}_{ij} \end{aligned} \tag{5.21}$$

and $R_0^2 = R^2 + O_p(n^{-1}) = R_1 + O_p(n^{-\frac{1}{2}})$ where

$$R_1^2 = 2n \textstyle\sum_{i<j} \Delta_{ij} t_{ij}^2 . \tag{5.22}$$

Anderson's (1965) series for ${}_0F_0^{(3)}(Z, \Omega)$ shows that $8\pi^2 \operatorname{etr}(-Z\Omega)\, {}_0F_0^{(3)}(Z, \Omega) = \pi^{-\frac{3}{2}} \prod_{i<j} (n\Delta_{ij})^{-\frac{1}{2}}[1 + O(\omega^{-1})]$. Thus, (5.21) implies that, conditional on Ω, $(2n\Delta_{ij})^{\frac{1}{2}}\tilde{t}_{ij}$, $1 \leqq i < j \leqq 3$ are asymptotically independent $N(0, 1)$ and thus also have this distribution unconditionally. (5.22) provides the desired partition of $\chi^2(3)$.

6. Estimators of M and Z and their properties. The log likelihood function based on a random sample of size n from Ψ is

$$\begin{aligned} L(M, Z) &= -n \log 4\pi - nY^{(0)}(Z) + \textstyle\sum_{j=1}^n \operatorname{tr} ZM^T\mathbf{x}_j\mathbf{x}_j^T M \\ &= -n \log 4\pi - nY^{(0)}(Z) + \operatorname{tr} ZM^TXX^TM . \end{aligned} \tag{6.1}$$

Lemma 4.2 and (2.18) imply that $L(M, Z - \zeta_0 I_3) = L(M, Z)$ for all ζ_0, and thus all inference procedures based on L are invariant under $Z \to Z - \zeta_0 I_3$. In particular, if $\hat{Z}$ maximizes $L(M, Z)$, then so does $\hat{Z} - \zeta_0 I_3$.

LEMMA 6.1. *For arbitrary* $Z = \operatorname{diag}[\zeta_1, \zeta_2, \zeta_3]$, $\zeta_1 \leqq \zeta_2 \leqq \zeta_3$, $\operatorname{tr} ZM^T XX^T M$ *is maximized by* $\hat{M} = [\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3]$, *where* $\hat{\mu}_i$ *is the eigenvector of* XX^T *associated with eigenvalue* ω_i, *with* $\omega_1 \leqq \omega_2 \leqq \omega_3$.

PROOF. This follows from a result of von Neumann (1937). □

LEMMA 6.2. *For any* M, $L(M, Z)$ *is maximized by any* $\hat{Z}$ *that satisfies the equations*

$$Y_j^{(1)}(\hat{Z}) = n^{-1} x_{jj}(M)\,, \qquad j = 1, 2, 3\,. \tag{6.2}$$

PROOF. By the remark above, we may impose the restriction $\zeta_3 = \hat{\zeta}_3 = 0$. From (6.1) we have $(\partial/\partial\zeta_i)L(M, Z) = -nY_i^{(1)}(Z) + x_{ii}(M)$, $i = 1, 2$, and $(\partial^2/\partial\zeta_i\,\partial\zeta_j)L(M, Z) = -nY_{ij}^{(2)}(Z)$, $i, j = 1, 2$. The positive definiteness of $[Y_{ij}^{(2)}(Z)]_{i,j\leqq 2}$ (Lemma 2.2) implies that $L(M, Z)$ is maximized by $\hat{Z} = \operatorname{diag}[\hat{\zeta}_1, \hat{\zeta}_2, 0]$, where $\hat{Z}$ satisfies (6.2), and that this solution is unique. Also Lemma 2.4 implies that $x_{ii}(M) - x_{jj}(M) = nY_i^{(1)}(\hat{Z}) - nY_j^{(1)}(\hat{Z})$ has the same sign as $\hat{\zeta}_i - \hat{\zeta}_j$. If $\tilde{Z} = \operatorname{diag}[\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\zeta}_3]$ maximizes $L(M, Z)$, then so does $\tilde{Z} - \tilde{\zeta}_3 I_3$ which must thus be $\hat{Z}$ and satisfies (6.2). The invariance of $Y_i^{(1)}(Z)$ under $Z \to Z - \zeta_0 I_3$ then implies that $\tilde{Z}$ also satisfies (6.2). □

The preceding lemmas immediately yield the following theorem.

THEOREM 6.1. *Assume throughout that* $Z = \operatorname{diag}[\zeta_1, \zeta_2, \zeta_3]$, $\zeta_1 \leqq \zeta_2 \leqq \zeta_3$.

(a) *When Z is known, a maximum likelihood estimator* (MLE) *of M is the matrix of eigenvectors* $\hat{M} = [\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3]$ *of* XX^T *with associated eigenvalues* $\omega_1 < \omega_2 < \omega_3$. *When* $\zeta_1 = \zeta_2$ *(resp.* $\zeta_2 = \zeta_3$*) a* MLE *of the single well-defined column of M is* $\hat{\mu}_3$ *(resp.* $\hat{\mu}_1$*).*

(b) *When M is known, then*

1. *If* $x_{11}(M) \leqq x_{22}(M) \leqq x_{33}(M)$, *a* MLE *of Z is any diagonal matrix $\hat{Z}$ satisfying* (6.2);

2. *If* $x_{22}(M) \leqq x_{11}(M)$, *a* MLE *of Z is any matrix of the form*

$$\begin{aligned} \hat{Z} &= \operatorname{diag}[\hat{\zeta}_1, \hat{\zeta}_1, \hat{\zeta}_3]\,, \quad \textit{satisfying } (6.2) \textit{ for } j = 3 \textit{ only}\,, \quad \textit{if} \quad x_{33}(M) \geqq n/3\,, \\ &= \operatorname{diag}[\hat{\zeta}, \hat{\zeta}, \hat{\zeta}]\,, \quad \textit{if} \quad x_{33}(M) < n/3\,; \end{aligned}$$

3. *If* $x_{22}(M) \geqq \max\{x_{11}(M), x_{33}(M)\}$, *then a* MLE *of Z is any matrix of the form*

$$\begin{aligned} \hat{Z} &= \operatorname{diag}[\hat{\zeta}_1, \hat{\zeta}_3, \hat{\zeta}_3]\,, \quad \textit{satisfying } (6.2) \textit{ for } j = 1 \textit{ only}\,, \quad \textit{if} \quad x_{11}(M) \leqq n/3 \\ &= \operatorname{diag}[\hat{\zeta}, \hat{\zeta}, \hat{\zeta}]\,, \quad \textit{if} \quad x_{11}(M) > n/3\,. \end{aligned}$$

(c) *When neither M norZ is known, joint* MLE's *of M and Z are $\hat{M}$ and any $\hat{Z}$ that satisfies*

$$Y_j^{(1)}(\hat{Z}) = n^{-1}\omega_j\,, \qquad j = 1, 2, 3\,. \tag{6.3}$$

(d) *Under the assumption of circular symmetry around* μ_3 (resp. μ_1), *a* MLE *of*

$\boldsymbol{\mu}_3$ (resp. $\boldsymbol{\mu}_1$) *is* (resp. $\hat{\boldsymbol{\mu}}_1$) *and a* MLE *of* Z *is any* $\hat{Z}_C$ *that satisfies*

$$Y_3^{(1)}(\hat{Z}_C) = n^{-1}\omega_3 , \qquad \hat{Z}_C = \operatorname{diag}[\hat{\zeta}_1, \hat{\zeta}_1, \hat{\zeta}_3] \tag{6.4}$$
$$(\text{resp. } Y_1^{(1)}(\hat{Z}_C) = n^{-1}\omega_1, \hat{Z}_C = \operatorname{diag}[\hat{\zeta}_1, \hat{\zeta}_3, \hat{\zeta}_3]) .$$

PROOF. (a) and (b1) follow directly from Lemmas 6.1 and 6.2, as does (c) after observing that $x_{jj}(\hat{M}) = \omega_j$ by (4.4). (b2) and (b3) are straightforward extensions considering maxima on the boundaries of the permissible region. (d) can be proved using Lemma 6.1 and an obvious modification of Lemma 6.2. □

Another approach to estimation of Z when there is no prior knowledge of M, is maximum likelihood estimation based on the marginal distribution of Ω. In analogous situations (e.g., Anderson (1965), or the estimation of σ^2 from a normal sample), the resulting estimators are more nearly unbiased than maximum likelihood estimators based on the joint distribution of the sample. Since in a certain sense (Barnard (1963)), Ω contains all the *available* information concerning Z, it seems plausible that inference concerning Z should be based only on the marginal distribution of Ω. From the corollary to Theorem 4.1, the marginal likelihood function for Z is

$$L_\Omega(Z) = -nY^{(0)}(Z) + \log {}_0F_0^{(3)}(Z, \Omega) + g(\Omega) , \tag{6.5}$$

where $g(\Omega)$ does not depend on Z. Let $\hat{\hat{Z}} = \operatorname{diag}[\hat{\hat{\zeta}}_1, \hat{\hat{\zeta}}_2, \hat{\hat{\zeta}}_3]$ maximize $L_\Omega(Z)$ subject to the restriction that $\zeta_1 \leqq \zeta_2 \leqq \zeta_3$. Then, since $L_\Omega(Z)$, like $L(M, Z)$, is invariant under $Z \to Z - \zeta_0 I_3$, $\hat{\hat{Z}} - \zeta_0 I_3$ also maximizes $L_\Omega(Z)$. It can be shown using Anderson's (1965) asymptotic series for ${}_0F_0^{(3)}(Z, \Omega)$ that, if $\hat{Z}$ maximizes $L(M, Z)$, then approximately

$$\hat{\hat{Z}} \doteq \hat{Z} + \operatorname{diag}[\delta_1, \delta_2, \delta_3] \tag{6.6}$$

maximizes $L_\Omega(Z)$, where

$$\delta_i = (2nB(\hat{Z}))^{-1} Y_{jk}^{(2)}(\hat{Z}) \textstyle\sum_{l \neq i} (\hat{\zeta}_i - \hat{\zeta}_l)^{-1} , \qquad i = 1, 2, 3 \neq j \neq k \neq i \tag{6.7}$$

and

$$\begin{aligned} B(Z) &= Y_{11}^{(2)}(Z) Y_{22}^{(2)}(Z) - (Y_{12}^{(2)}(Z))^2 \\ &= Y_{12}^{(2)}(Z) Y_{13}^{(2)}(Z) + Y_{12}^{(2)}(Z) Y_{23}^{(2)}(Z) + Y_{13}^{(2)}(Z) Y_{23}^{(2)}(Z) . \end{aligned} \tag{6.8}$$

It can further be shown that the use of $\hat{\hat{Z}}$ does, in fact, correspond to removing part of the $O(n^{-1})$ bias of $\hat{Z}$.

The structure of the parameter space makes discussion of the properties of estimators somewhat awkward. The invariance of Ψ under $Z \to Z - \zeta_0 I_3$ and/or $\boldsymbol{\mu}_j \to \pm\boldsymbol{\mu}_j$, $j = 1, 2, 3$, means that such properties as bias and variance are undefined without some constraints to enforce uniqueness. The circularly symmetric case has been considered by Watson. Hence here we will assume the following constraints on Z and $M = [\mu_{ij}]_{i,j \leq 3}$:

$$\zeta_1 < \zeta_2 < \zeta_3 = 0 \tag{6.9}$$

and

$$\mu_{11} > 0 , \qquad \mu_{33} > 0 \qquad \text{and} \qquad M \in O^+(3) . \tag{6.10}$$

When $\mu_{11} = 0$ or $\mu_{33} = 0$, (6.10) is not sufficient to determine M and in that case we could choose a different restriction. $\hat{Z}$ and $\hat{M}$ will also be assumed to satisfy (6.9) and (6.10). With these restrictions, the standard results concerning maximum likelihood estimators hold for Z and M, when suitably parametrized.

LEMMA 6.3. *Let* $\mathbf{t} = [t_1, t_2, t_3]$ *be a parametrization of* $O^+(3)$ *that is regular in a neighborhood of M. Then*

$$(6.11) \qquad (\partial/\partial t_k) \operatorname{tr} ZM^TXX^TM = 2 \operatorname{tr} ZM^TXX^TMD_k = 2 \textstyle\sum_{i>j} d_{ji}^{(k)}(\zeta_i - \zeta_j)x_{ij}(M)\,,$$

where $D_k = [d_{ij}^{(k)}]_{i,j\leq 3} = M^T(\partial/\partial t_k)M$ *is skew symmetric.*

PROOF. Since $M^TM = I_3$, we have $0 = (\partial/\partial t_k)M^TM = D_k^T + D_k$, showing that D_k is skew symmetric. Also $(\partial/\partial t_k) \operatorname{tr} ZM^TXX^TM = \operatorname{tr} ZD_k^TM^TXX^TM + \operatorname{tr} ZM^TXX^TMD_k = 2 \operatorname{tr} ZM^TXX^TMD_k$. □

THEOREM 6.2. *Let Z and M satisfy* (6.9) *and* (6.10). *Let* $\hat{Z}$ *and* $\hat{M}$ *be* MLE's *also satisfying* (6.9) *and* (6.10). *Let* $\mathbf{t}$ *and* $\hat{\mathbf{t}}$ *represent a regular parametrization of* M *and* $\hat{M}$ *as in Lemma* 6.3, *and let* $\mathbf{z} = [\zeta_1, \zeta_2]^T$, $\hat{\mathbf{z}} = [\hat{\zeta}_1, \hat{\zeta}_2]^T$. *Then as* $n \to \infty$, $\hat{\boldsymbol{\theta}} = [\hat{\mathbf{z}}^T, \hat{\mathbf{t}}^T]^T$ *is asymptotically normal with expectation* $\boldsymbol{\theta} = [\mathbf{z}^T, \mathbf{t}^T]^T$ *and covariance matrix*

$$(6.12) \qquad \operatorname{Cov}[\hat{\boldsymbol{\theta}}] = n^{-1} \text{ block diag } [C_1(Z), C_2(Z, M)]\,,$$

where

$$(6.13) \qquad C_1(Z) = B^{-1}(Z)\begin{bmatrix} Y_{22}^{(2)}(Z) & -Y_{12}^{(2)}(Z) \\ -Y_{12}^{(2)}(Z) & Y_{11}^{(2)}(Z) \end{bmatrix}, \qquad B(Z) \text{ as in } (6.8);$$

$$(6.14) \qquad C_2(Z, M) = \tfrac{1}{2}(D(M)^T)^{-1} \operatorname{diag} [1/\Delta_{12}^0, 1/\Delta_{13}^0, 1/\Delta_{23}^0]D(M)^{-1}\,,$$

$$D(M) = \begin{bmatrix} d_{12}^{(1)} & d_{13}^{(1)} & d_{23}^{(1)} \\ d_{12}^{(2)} & d_{13}^{(2)} & d_{23}^{(2)} \\ d_{12}^{(3)} & d_{13}^{(3)} & d_{23}^{(3)} \end{bmatrix}, \qquad d_{ij}^{(k)} \text{ as in Lemma } 6.3,$$

and Δ_{ij}^0 *as in* (5.14).

PROOF. Since Ψ is an exponential family on a compact set, regularity conditions (Wilks (1963), Section 12.7) for the consistency and asymptotic normality of maximum likelihood estimators are easily verified. Thus $\hat{\boldsymbol{\theta}}$ is asymptotically normal with expectation $\boldsymbol{\theta}$ and covariance matrix $\mathscr{I}$ where $\mathscr{I}^{-1} = \operatorname{Cov}[(\partial/\partial\boldsymbol{\theta})L(M, Z)]$. But $(\partial/\partial\zeta_j)L(M, Z) = -nY_j^{(1)}(Z) + x_{jj}(M)$, $j = 1, 2$, and, by Lemma 6.3,

$$(\partial/\partial\mathbf{t})L(M, Z) = 2D(M)\begin{bmatrix} (\zeta_1 - \zeta_2)x_{12}(M) \\ (\zeta_1 - \zeta_3)x_{13}(M) \\ (\zeta_2 - \zeta_3)x_{23}(M) \end{bmatrix}.$$

Theorem 5.1 then implies that $\operatorname{Cov}[(\partial/\partial\mathbf{z})L(M, Z)] = n[Y_{ij}^{(2)}(Z)]_{i,j\leq 2}$,

$$\operatorname{Cov}[(\partial/\partial\mathbf{t})L(M, Z)] = 4D(M) \operatorname{diag} [(\zeta_1 - \zeta_2)^2A_{12}(Z), (\zeta_1 - \zeta_3)^2A_{13}(Z), (\zeta_2 - \zeta_3)^2A_{23}(Z)]D(M)^T\,,$$

and that $(\partial/\partial \mathbf{t})L(M, Z)$ is uncorrelated with $(\partial/\partial \mathbf{z})L(M, Z)$. But by the corollary to Lemma 2.4 and (5.14), $2(\zeta_i - \zeta_j)^2 A_{ij}(Z) = (\zeta_i - \zeta_j)(Y_i^{(1)}(Z) - Y_j^{(1)}(Z)) = \Delta_{ij}^0$. Thus $\mathscr{I}^{-1} = n^{-1}$ block diag $[C_1(Z), C_2(Z, M)]$. □
Theorem 6.2 can also be shown to hold when $\hat{Z}$ satisfies (6.2).

When we use the parametrization $\mathbf{t} = [t_{12}, t_{13}, t_{23}]^T$ introduced in (5.18) and (5.19), we find that

$$D(M) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & \cos t_{12} & -\sin t_{12} \\ \sin t_{13} & \sin t_{12}\cos t_{13} & \cos t_{12}\cos t_{13} \end{bmatrix}.$$

If the coordinates of M and $\hat{M}$ are taken relative to the columns of M, $t_{12} = t_{13} = t_{23} = 0$, $\hat{t}_{ij}$ is identical to $\hat{t}_{ij}$ in Section 5, and $D(M) = \text{diag}[-1, 1, 1]$. This leads to the following:

COROLLARY. *Assume that $\zeta_1 < \zeta_2 < \zeta_3$. Let $\hat{\mathbf{t}} = [\hat{t}_{12}, \hat{t}_{13}, \hat{t}_{23}]^T$ be the parametrization of $\tilde{M} = M^T\hat{M}$ by* (5.18) *and* (5.19). *Then, as $n \to \infty$, $\hat{\mathbf{t}}$ is asymptotically normal with $E[\hat{\mathbf{t}}] = 0$, and*

$$\text{Cov}[\hat{\mathbf{t}}] = (2n)^{-1}\text{diag}[1/\Delta_{12}^0, 1/\Delta_{13}^0, 1/\Delta_{23}^0]. \tag{6.15}$$

This corollary implies that the $(2n\Delta_{ij}^0)^{\frac{1}{2}}\hat{t}_{ij}$ are asymptotically independent $N(0, 1)$. This result differs from that at the end of Section 5 only in the substitution of $Y_i^{(1)}(Z)$ for $n^{-1}\omega_i = Y_i^{(1)}(\hat{Z})$.

7. Statistical inference. Examination of the likelihood function (6.1) makes clear that the elements of XX^T are jointly sufficient for M and Z. Hence inferences concerning the parameters should depend only on these statistics. We consider here only those problems that arise in the analysis of a single random sample. The problems become clearer if we think of Z as a shape parameter and M as a generalized location parameter.

Two important practical questions in analyzing orientation data are the following: (a) Is there evidence that the data do not come from a uniform (isotropic) distribution ($Z = 0$)? and (b) Is there evidence of lack of circular symmetry of the distribution about some axis? If the answer to (a) is "No", then further analysis is not meaningful. If (b) can be answered negatively, then the simpler Dimroth–Watson distribution (in either polar or girdle form) is adequate. One approach would be direct use of likelihood ratio tests based on Ψ. We prefer here tests whose application does not require computation of an unrestricted $\hat{Z}$.

THEOREM 7.1 (Test of isotropy). *Let H_U: $Z = 0$ be the hypothesis of isotropy. Then the procedure: Reject H_U if $X_U^2 > \chi_{1-\alpha}^2(5)$, where X_U^2 is defined by* (5.9), *is a test of H_U of asymptotic size α as $n \to \infty$.*

PROOF. This follows directly from Theorem 5.2. □
When H_U is true, it can be shown that, for large n, X_U^2 is asymptotically equivalent to $-2\log\lambda_U$, where λ_U is the likelihood ratio statistic for testing H_U against general Ψ. The conditions for a $\chi^2(5)$ approximation to $-2\log\lambda_U$ hold here,

thus providing another method of proving Theorem 5.2. Anderson and Stephens (1973) have considered a similar test of isotropy with the Dimroth–Watson distribution as alternative. By Monte Carlo simulation they show that its asymptotic distribution (not χ^2) is approached rapidly as $n \to \infty$. A similar situation is conjectured for X_U^2.

There are two essentially different forms of circular symmetry—polar: $H_{CP}(\zeta_1 = \zeta_2 \leqq \zeta_3)$ and girdle: $H_{CG}(\zeta_1 \leqq \zeta_2 = \zeta_3)$. Often the choice between these is clear and we assume that such a choice has been made. For definiteness we consider the test of H_{CP}.

Lack of circular symmetry around $\boldsymbol{\mu}_3$ in the sample will be reflected by the departure of $\omega_1(\boldsymbol{\mu}_3)$ and $\omega_2(\boldsymbol{\mu}_3)$ from equality where $\omega_i(\boldsymbol{\mu}_3)$, $i = 1, 2$ are as defined in Theorem 5.3. Thus a measure of sample non-circularity about $\boldsymbol{\mu}_3$ is $(\omega_1(\boldsymbol{\mu}_3) - \omega_2(\boldsymbol{\mu}_3))^2$, and when Z and $\boldsymbol{\mu}_3$ are known, $X_C^2(\boldsymbol{\mu}_3)$ (5.10) is a plausible test criterion for H_{CP}. In practice these parameters are not known. However, it can be shown that for large n and $\zeta_1 = \zeta_2 < \zeta_3$

$$n^{-1}(\omega_1(\boldsymbol{\mu}_3) - \omega_2(\boldsymbol{\mu}_3))^2 = n^{-1}(\omega_1 - \omega_2)^2 + O_p(n^{-\frac{1}{2}}) .$$

To the same order we can substitute for Z in $A_{12}(Z)$ either a maximum likelihood estimate $\hat{Z}_C$ satisfying (6.4) or an unrestricted maximum likelihood estimate $\hat{Z}$ satisfying (6.3). Thus asymptotically, under H_{CP}, $X_C^2(\boldsymbol{\mu}_3)$ is equivalent both to

$$X_C^2 = (4nA_{12}(\hat{Z}_C))^{-1}(\omega_1 - \omega_2)^2 \tag{7.1}$$

and to

$$\tilde{X}_C^2 = (4nA_{12}(\hat{Z}))^{-1}(\omega_1 - \omega_2)^2 . \tag{7.2}$$

THEOREM 7.2. *Assume that* $\zeta_1 \leqq \zeta_2 < \zeta_3$. *Then the procedure: Reject* H_{CP} *if* $X_C^2 > \chi^2_{1-\alpha}(2)$ (*or reject* H_{CP} *if* $\tilde{X}_C^2 > \chi^2_{1-\alpha}(2)$) *is a test of* H_{CP} *of asymptotic size* α, *as* $n \to \infty$.

PROOF. This follows from the preceding discussion and Theorem 5.3. □
Using special properties of univariate confluent hypergeometric functions, it can be shown that if $\hat{Z}_C = \operatorname{diag}[0, 0, \hat{\zeta}_C]$, $\hat{\zeta}_C \neq 0$, satisfies (6.4), then

$$4A_{12}(\hat{Z}_C) = 1/(2\hat{\zeta}_C) + (2n)^{-1}(\omega_1 + \omega_2)(1 - 3/(2\hat{\zeta}_C)) . \tag{7.3}$$

Also, by Lemma 2.4, when $\hat{Z}$ satisfies (6.3) and $\omega_1 \neq \omega_2$,

$$4A_{12}(\hat{Z}) = 2n^{-1}(\omega_1 - \omega_2)/(\hat{\zeta}_1 - \hat{\zeta}_2) . \tag{7.4}$$

From (7.4) we find a particularly simple form for $\tilde{X}_C^2$:

$$\tilde{X}_C^2 = \tfrac{1}{2}(\omega_1 - \omega_2)(\hat{\zeta}_1 - \hat{\zeta}_2) . \tag{7.5}$$

Analogous results to the foregoing can be stated for H_{CG} by substituting subscripts 2 and 3 for 1 and 2 throughout.

X_C^2 has the advantage of not requiring the computation of any quantities not needed for applying the Dimroth–Watson distribution. Under the null hypothesis, both X_C^2 and $\tilde{X}_C^2$ can be shown to be asymptotically equivalent to $-2 \log \lambda_C$, where λ_C is the likelihood ratio statistic for testing H_{CP} (or H_{CG}) against unrestricted Ψ. More precisely it can be shown that when $\zeta_1 = \zeta_2 < \zeta_3$, $X_C^2 - \tilde{X}_C^2 = n\,O((\hat{\zeta}_1 - \hat{\zeta}_2)^4)$ and both X_C^2 and $\tilde{X}_C^2$ are $= -2\log\lambda_C + n\,O((\hat{\zeta}_1 - \hat{\zeta}_2)^4)$.

When neither isotropy nor circular symmetry are plausible possibilities, the point estimates of Section 6 should be supplemented by some form of regional estimation of Z. Following Bartlett (1953a, 1953b), a confidence region for an unknown vector parameter $\boldsymbol{\theta}$ can be based on a quadratic form in the score vector $\mathbf{s} = \partial L/\partial\boldsymbol{\theta}$ where L is the log likelihood. The same considerations that make plausible the use of the "marginal" maximum likelihood estimator $\hat{\hat{Z}}$ suggest that the appropriate log likelihood function is $L_\Omega(Z)$ given by (6.4). Define

$$l_i(Z) = (\partial/\partial\zeta_i)L_\Omega(Z)\,, \qquad i = 1, 2, 3\,, \tag{7.6}$$

$$l_{ij}(Z) = (\partial^2/(\partial\zeta_i\,\partial\zeta_j))L_\Omega(Z)\,, \qquad i, j = 1, 2, 3\,. \tag{7.7}$$

Then $E_Z[l_i(Z)] = 0$ and

$$L_{ij}(Z) \equiv \mathrm{Cov}_Z\,[l_i(Z), l_j(Z)] = -E_Z[l_{ij}(Z)]\,, \tag{7.8}$$

where E_Z and Cov_Z represent expectation and covariance when Z is the true parameter. The asymptotic normality of $l_i(Z)$ implies that, when Z is the population parameter,

$$X^2(Z) = \sum\sum_{i,j\leqq 3} L^{ij}(Z)l_i(Z)l_j(Z) \tag{7.9}$$

is approximately $\chi^2(2)$, where the matrix $[L^{ij}(Z)]_{i,j\leqq 3}$ is any generalized inverse (Rao (1965)) to $[L_{ij}(Z)]_{i,j\leqq 3}$. This provides a basis for a confidence region for Z.

THEOREM 7.3. *When $\zeta_1 < \zeta_2 < \zeta_3$, as $n \to \infty$, the region defined by*

$$\{Z \mid \tilde{X}^2(Z) \leqq \chi^2_{1-\alpha}(2)\} \tag{7.10}$$

is asymptotically a confidence region for Z with confidence coefficient $1 - \alpha$, *where*

$$\tilde{X}^2(Z) = \frac{-\tilde{L}_{23}(Z)\tilde{l}_1^2(Z) - \tilde{L}_{13}(Z)\tilde{l}_2^2(Z) - \tilde{L}_{12}(Z)\tilde{l}_3^2(Z)}{\tilde{L}_{12}(Z)\tilde{L}_{13}(Z) + \tilde{L}_{12}(Z)\tilde{L}_{23}(Z) + \tilde{L}_{13}(Z)\tilde{L}_{23}(Z)}\,, \tag{7.11}$$

$$\tilde{l}_j(Z) = \omega_j - nY_j^{(1)}(Z) - \tfrac{1}{2}\sum_{k\neq j}(\zeta_j - \zeta_k)^{-1}\,, \tag{7.12}$$

$$\tilde{L}_{ij}(Z) = nY_{ij}^{(2)}(Z) + \tfrac{1}{2}(\zeta_i - \zeta_j)^{-2}\,, \qquad i \neq j\,. \tag{7.13}$$

PROOF. Employing Anderson's (1965) asymptotic series, it can be shown that

$$l_i(Z) = \omega_i - E_Z[\omega_i] + O_p(n^{-\frac{1}{2}}) = \tilde{l}_i(Z) + O_p(n^{-1})\,, \tag{7.14}$$

and

$$L_{ij}(Z) = \mathrm{Cov}_Z\,[\omega_i, \omega_j] + O(n^{-1}) = \tilde{L}_{ij}(Z) + O(n^{-1})\,. \tag{7.15}$$

The last equalities in (7.14) and (7.15) follow from Theorem 5.4. Since $\sum_{i=1}^3 L_{ij}(Z) = 0$, a generalized inverse to $[L_{ij}(Z)]_{i,j\leqq 3}$ is

$$[L^{ij}(Z)]_{i,j\leqq 3} = (L_{12}(Z)L_{13}(Z) + L_{12}(Z)L_{23}(Z) + L_{13}(Z)L_{23}(Z))^{-1} \times \mathrm{diag}\,[-L_{23}(Z), -L_{13}(Z), -L_{12}(Z)]\,.$$

By (7.14) and (7.15), this implies that $\tilde{X}^2(Z) = X^2(Z) + O_p(n^{-1})$, and hence $\tilde{X}^2(Z)$ is also asymptotically $\chi^2(2)$ when Z is the true parameter. The conclusion is now immediate. □

By Bartlett (1953 b), (7.10) is asymptotically equivalent to the confidence region based on the "marginal" likelihood ratio:

$$\{Z \mid 2(L_\Omega(\hat{\hat{Z}}) - L_\Omega(Z)) \leqq \chi^2_{1-\alpha}(2)\}, \tag{7.16}$$

where $L_\Omega(Z)$ is given by (6.5). This can be usefully determined from a contour plot of $L_\Omega(Z)$. Another confidence region that is asymptotically equivalent to (7.10) is one based on the asymptotic normality of $\hat{Z}$ (Theorem 6.2). Substituting $\hat{Z}$ for Z in the covariance matrix of $\hat{Z}$ does not affect the asymptotic validity. These considerations yield the following region with asymptotic confidence coefficient $1 - \alpha$:

$$\{Z \mid n \sum \sum_{i,j\leqq 3} (\zeta_i - \hat{\zeta}_i)(\zeta_j - \hat{\zeta}_j) Y^{(2)}_{ij}(\hat{Z}) \leqq \chi^2_{1-\alpha}(2)\}. \tag{7.17}$$

This (along with the preceding regions) is valid for any choice of restrictions applied to both Z and $\hat{Z}$ to obtain uniqueness. Under the constraint $\zeta_3 = \hat{\zeta}_3 = 0$, (7.17) defines an ellipse in the (ζ_1, ζ_2)-plane.

When M is the true orientation matrix, (6.1) shows that the diagonal elements $x_{ii}(M)$ of $M^T XX^T M$ are jointly sufficient for Z. Hence the conditional distribution of the $x_{ij}(M)$, $i < j$, given $x_{ii}(M)$, $i = 1, 2, 3$, does not depend on Z and could provide a basis for inference concerning M. Theorem 5.2 implies that the $x_{ij}(M)$ are asymptotically independent of the $x_{ii}(M)$'s and that the conditional and unconditional distributions of the $x_{ij}(M)$'s are asymptotically independent $N(0, nA_{ij}(Z))$. Also, $E[x_{ij}(M)] = 0$, all $i < j$, if and only if M is the true orientation matrix. This motivates the use of

$$X^2(M) = n^{-1} \sum_{i<j} A^{-1}_{ij}(Z) x^2_{ij}(M) \tag{7.18}$$

as a test criterion for the hypothesis H_M: M is the true orientation matrix.

THEOREM 7.4. *Assume* $\zeta_1 < \zeta_2 < \zeta_3$. *Let*

$$\tilde{X}^2(M) = 2 \sum_{i<j} (\omega_i - \omega_j)^{-1} (\hat{\zeta}_i - \hat{\zeta}_j) x^2_{ij}(M). \tag{7.19}$$

Then the procedure: Reject H_M *if and only if* $\tilde{X}^2(M) > \chi^2_{1-\alpha}(3)$ *is a test of* H_M *with asymptotic size* α, *as* $n \to \infty$.

PROOF. The preceding discussion shows that the conclusion is valid for $X^2(M)$ in place of $\tilde{X}^2(M)$. But Lemma 2.4 shows that when $\hat{Z}$ satisfies (6.2), $\tilde{X}^2(M)$ is obtained from $X^2(M)$ by substituting $\hat{Z}$ for Z in $A_{ij}(Z)$. For large n, $A_{ij}(\hat{Z}) = A_{ij}(Z) + O_p(n^{-\frac{1}{2}})$, and, under H_M, $n^{-1}x^2_{ij}(M) = O_p(1)$, $i < j$. Thus $\tilde{X}^2(M) = X^2(M) + O_p(n^{-\frac{1}{2}})$ implying that $\tilde{X}^2(M)$ and $X^2(M)$ have the same asympotic distribution. □

COROLLARY. *As* $n \to \infty$, *the region*

$$\{M \mid \tilde{X}^2(M) \leqq \chi^2_{1-\alpha}(3)\} \tag{7.20}$$

represents a region with asymptotic confidence coefficient $1 - \alpha$.

A more readily understood approximate confidence region for M can be expressed in terms of the parametrization $\hat{\mathbf{t}}$ of $\hat{M}$ near M described in the corollary to Theorem 6.2. The vector $\hat{\mathbf{t}}$ can be considered as a random parametrization of M in the vicinity of $\hat{M}$.

THEOREM 7.5. *Let* $\hat{\mathbf{t}} = [\hat{t}_{12}, \hat{t}_{13}, \hat{t}_{23}]^T$ *be the parametrization of* $\hat{M}$ *in terms of* M *described in the corollary to Theorem* 6.2, *considered as functions of* M. *Then, when* $\zeta_1 < \zeta_2 < \zeta_3$, *as* $n \to \infty$,

$$\{M \mid 2n \textstyle\sum_{i<j} \hat{\Delta}_{ij} \hat{t}_{ij}^2 \leqq \chi^2_{1-\alpha}(3)\} \tag{7.21}$$

is a confidence region for M *with asymptotic confidence coefficient* $1 - \alpha$, *where*

$$\hat{\Delta}_{ij} = n^{-1}(\omega_i - \omega_j)(\hat{\zeta}_i - \hat{\zeta}_j) \,. \tag{7.22}$$

PROOF. By the corollary to Theorem 6.2, when M is the true orientation matrix, $2n \sum_{i<j} \Delta_{ij}^0 \hat{t}_{ij}^2$ is asymptotically $\chi^2(3)$, where Δ_{ij}^0 is as in (5.14). By (6.3), $\hat{\Delta}_{ij}$ is obtained from Δ_{ij}^0 by substituting $\hat{Z}$ for Z. Thus $\hat{\Delta}_{ij} = \Delta_{ij}^0 + O_p(n^{-\frac{1}{2}})$, $i < j$ and the asymptotic distribution is still $\chi^2(3)$. The conclusion is now immediate. □

One difficulty in the preceding results is the problem of picturing confidence regions on $O^+(3)$. One solution is to seek separate confidence regions R_i on the unit sphere for each column $\boldsymbol{\mu}_i$ of M, $i = 1, 2, 3$. Then one can have confidence at least $1 - 3\alpha$ that $\boldsymbol{\mu}_i \in R_i$, $i = 1, 2, 3$. Moreover, for some purposes, confidence regions for the $\boldsymbol{\mu}_i$ separately are desirable. Such R_i could be defined analogously to Theorem 7.4. However, the following, more closely related to Theorem 7.5, appears easier to use.

THEOREM 7.6. *Let* $\zeta_1 < \zeta_2 < \zeta_3$. *For each fixed vector* $\boldsymbol{\mu} \in S$, *define* (*random*) *matrices* $M_i = M_i(\boldsymbol{\mu}) \in O^+(3)$ *with random coordinates* $t_{12}^{(i)}$, $t_{13}^{(i)}$, *and* $t_{23}^{(i)}$ *relative to* $\hat{M}$, $i = 1, 2, 3$, *such that the* ith *column of* M_i *is* $\boldsymbol{\mu}$, *and* $t_{jk}^{(i)} = 0$, $i \neq j < k \neq i$. *Define*

$$X_i^2(\boldsymbol{\mu}) = 2n \textstyle\sum_{j \neq i} \hat{\Delta}_{ij}(t_{ij}^{(i)})^2 \,, \qquad t_{ij}^{(i)} \equiv t_{ji}^{(i)} \qquad \textit{if } i < j \,. \tag{7.23}$$

Then as $n \to \infty$, *the regions*

$$R_i = \{\boldsymbol{\mu} \mid X_i^2(\boldsymbol{\mu}) \leqq \chi^2_{1-\alpha}(2)\} \tag{7.24}$$

are confidence regions for $\boldsymbol{\mu}_i$ *with asymptotic confidence coefficients* $1 - \alpha$.

PROOF. We consider the case $i = 1$. Let $\hat{t}_{jk}$, $j < k$, be the (random) coordinates of true orientation matrix M relative to $\hat{M}$. Then, it can be shown that when $\boldsymbol{\mu}$ is the first column of M, the coordinates $t_{12}^{(1)}$ and $t_{13}^{(1)}$ for $\boldsymbol{\mu}$ satisfy $t_{1j}^{(1)} = \hat{t}_{1j} + O_p(n^{-1})$, $j = 2, 3$. Thus $X_1^2(\boldsymbol{\mu}) = 2n(\hat{\Delta}_{12}\hat{t}_{12}^2 + \hat{\Delta}_{13}\hat{t}_{13}^2) + O_p(n^{-1}) = 2n(\Delta_{12}^0 t_{12}^2 + \Delta_{13}^0 t_{13}^2) + O_p(n^{-\frac{1}{2}})$. The corollary to Theorem 6.2 then implies that $X_1^2(\boldsymbol{\mu})$ is asymptotically $\chi^2(2)$ whenever $\boldsymbol{\mu}$ is the first column of M. Similarly $X_i^2(\boldsymbol{\mu})$ is asymptotically $\chi^2(2)$ when $\boldsymbol{\mu}$ is the ith column of M, $i = 2, 3$. The conclusion is now immediate. □

When the $(n\hat{\Delta}_{ij})^{-1}$ are small, the regions R_i are approximately ellipses centered at $\hat{\mu}_i$ with axes of lengths $(\chi^2_{1-\alpha}(2)/(2n\hat{\Delta}_{ij}))^{\frac{1}{2}}$ in the directions (on the sphere) of $\hat{\mu}_j, j \neq i$.

8. A numerical example. A sample of 150 measurements on the $\underline{c}$ axis of calcite grains from the Taconic Mountains of New York state (Hansen (1963)) was analyzed using Ψ. The data are determinations of undirected axes implying that an analysis based on antipodal symmetry is in order. Figure 1 is an equiareal projection onto the horizontal plane of the data, represented as points in the upper hemisphere. Observe that a "neighborhood" of a point near the circumference includes points near the antipodal point of the circumference.

The cross product matrix XX^T, the orthogonal matrix $\hat{M}$ of its eigenvectors, and its eigenvalues ω_i are displayed in Table 1. Table 2 contains the unrestricted maximum likelihood estimate $\hat{Z}$ (6.3), the maximum "marginal" likelihood estimate $\hat{\hat{Z}}$ (6.6), the estimated covariance matrix of $\hat{\zeta}_1$ and $\hat{\zeta}_2$ (6.12) (this also applies to $\hat{\hat{\zeta}}_1$ and $\hat{\hat{\zeta}}_2$), and the estimated standard deviations $\hat{\sigma}_{ij} = (2n\hat{\Delta}_{ij})^{-\frac{1}{2}}$ of rotation of each pair of eigenvectors around the third (6.15), expressed in degrees. These last have been used to plot Figure 1 the approximate 95% confidence regions for each μ_j that are defined in Theorem 7.6. Table 3 displays values of various test criteria for uniformity and circular symmetry. $X_U{}^2$, $X_C{}^2$, and $\tilde{X}_C{}^2$ are defined by (5.9), (7.1), and (7.2), respectively. λ_U and λ_C are the usual likelihood ratio test criteria for uniformity and circular symmetry.

The observations have been replotted in Figure 2, which represents a hemisphere centered at $\hat{\mu}_3$ projected on the $(\hat{\mu}_1, \hat{\mu}_2)$-plane. The original horizon is drawn as a dashed line with the cardinal compass points identified. The rough elliptical scatter about $\hat{\mu}_3$ suggests that Ψ may be appropriate. Examination of Table 2 confirms that there are significant departures from both uniformity and circularity as would have appeared likely from inspection of Figure 2. The three test criteria for circularity have $\chi^2(2)$ as a null distribution, and the isotropy criteria are both approximately $\chi^2(5)$. Non-circularity and non-uniformity are also evidenced by the relatively small size of the confidence regions for the μ_j in Figure 1.

A contour diagram of the "marginal" likelihood function for Z (6.5) is in Figure 3 (with the constraint that $\zeta_3 = 0$). Superimposed is the boundary of the approximate 95% confidence region for ζ_1 and ζ_2 defined by (7.17). This can be compared with the contour $3.841/2 = 1.921$ down from the maximum of $L_\Omega(Z)$, which is the boundary of (7.16).

Acknowledgments. Many people have contributed in one way or another to this work and I am grateful to them all. I particularly wish to thank Alan T. James who supervised the dissertation on which much of the work is based, and John W. Tukey, Paul Meier, and David L. Wallace who at various times materially aided its progress.

TABLE 1

Cross product matrix XX^T, with eigenvectors $[\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3] = \hat{M}$ and eigenvalues $\omega_1 \leqq \omega_2 \leqq \omega_3$ derived from $N = 150$ measurements of the c axis of calcite grains from limestone in the Taconic Mountains of New York (Hansen (1963))

	XX^T			$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	ω_i
	E-W	N-S	Vert				
E-W	76.5575	18.2147	12.2406	−0.1723	−0.4439	0.8794	23.43215
N-S	18.2147	46.7740	6.8589	−0.1516	0.8940	0.4216	38.19628
Vert	12.2406	6.8589	26.6670	0.9733	0.0606	0.2213	88.37007

TABLE 2

Maximum likelihood estimates $\hat{Z}$, maximum marginal likelihood estimates $\hat{\hat{Z}}$ with their estimated covariance matrix, and standard deviations in degrees of the components of rotation of the orientation matrix $\hat{M}$ for the data summarized in Table 1

	$\hat{Z}$	$\hat{\hat{Z}}$	$\text{Cov}(\hat{Z}) = \text{Cov}(\hat{\hat{Z}})$		i	j	$\hat{\sigma}_{ij}$
ζ_1	−3.518	−3.434	0.17624	0.02003	1	2	8.44°
ζ_2	−1.956	−1.954	0.02003	0.09389	1	3	2.68°
ζ_3	0	0			2	3	4.09°

TABLE 3

Test statistics for circular symmetry and isotropy for the data summarized in Table 1

	Polar	Girdle		
X_C^2	11.058	44.743	X_U^2	115.872
$\tilde{X}_C^2$	11.531	49.060	$-2 \log \lambda_U$	111.697
$-2 \log \lambda_C$	11.294	46.794		

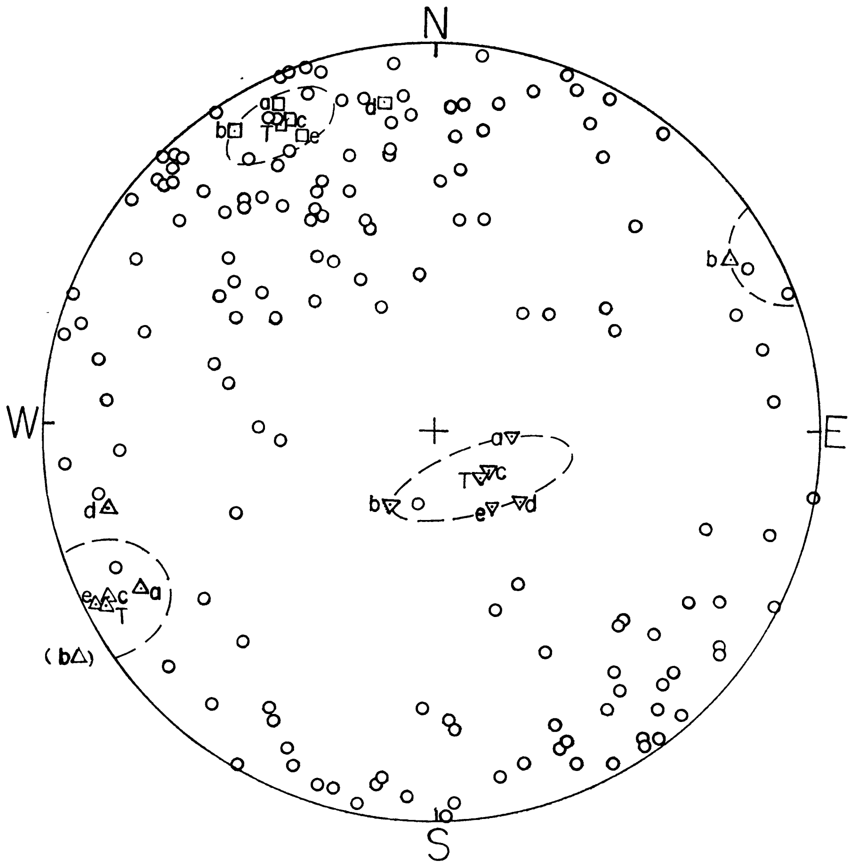

FIG. 1. Orientations of the $\underline{c}$ axis of 150 calcite grains from limestone from the Taconic Mountains of New York (Hansen (1963)) in an equiareal projection on the horizontal plane. Also plotted are maximum likelihood estimates $\hat{\mu}_1$ (▽), $\hat{\mu}_2$ (△), and $\hat{\mu}_3$ (□) for five subsamples (a, b, c, d, e) and for the entire sample (T), together with approximate 95% confidence regions for μ_1, μ_2, and μ_3 (dashed lines).

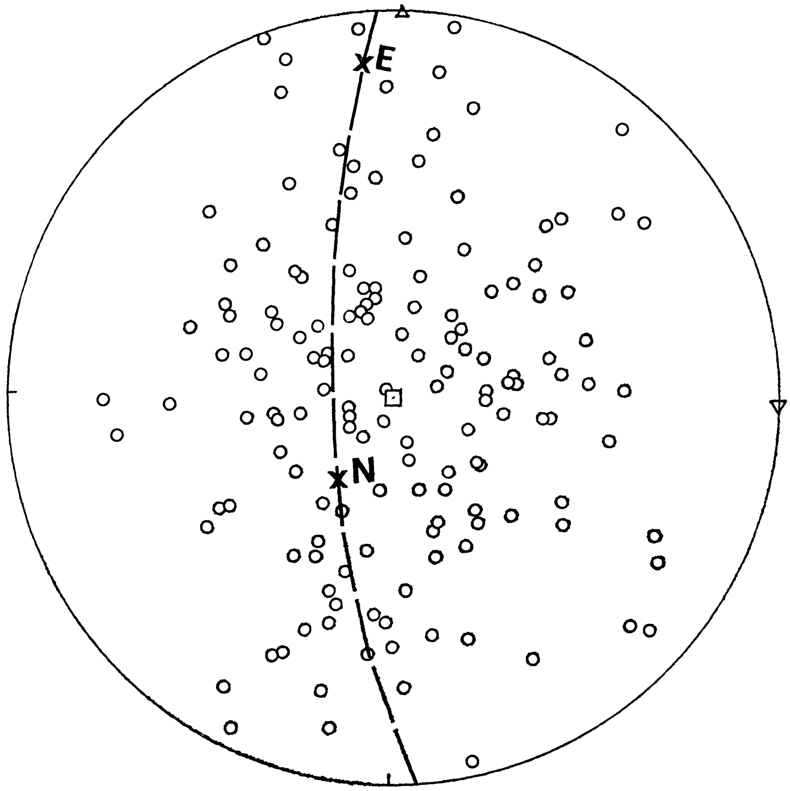

FIG. 2. Data of Figure 1, plotted in an equiareal projection on the plane determined by $\hat{\mu}_1$ (right) and $\hat{\mu}_2$ (top). The dashed line represents the horizontal plane with the $N(\equiv S)$ and $E(\equiv W)$ points labelled.

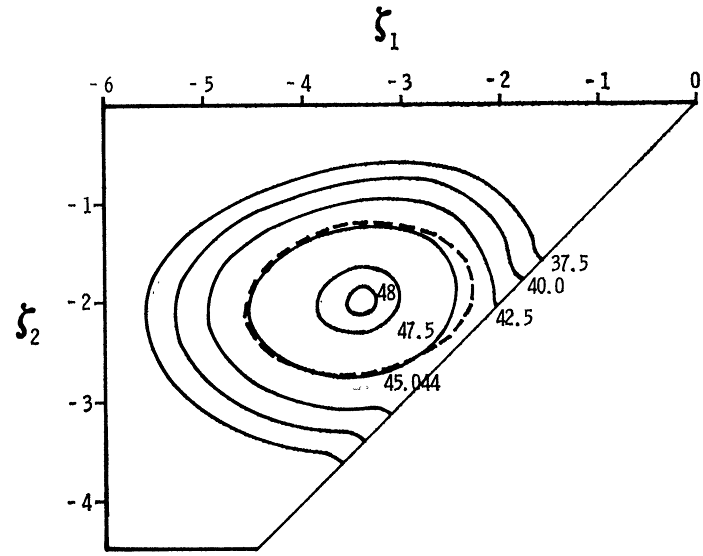

FIG. 3. Contour plot of the log likelihood function $L(Z, \Omega)$ based on the marginal distribution of Ω for the data in Figure 1, using the convention that $\zeta_1 \leqq \zeta_2 \leqq \zeta_3 = 0$. The contour $L = 45.044$ represents the boundary of the 95% confidence region based on the (marginal) likelihood ratio. The dashed line is the boundary of a 95% confidence region based on the asymptotic normality of $\hat{\hat{\zeta}}_1$ and $\hat{\hat{\zeta}}_2$.

REFERENCES

ANDERSON, GEORGE A. (1965). An asymptotic expansion for the distribution of the latent roots of the estimated covariance matrix. *Ann. Math. Statist.* **36** 1153–1173.

ANDERSON, T. W. and STEPHENS, M. A. (1973). Tests for randomness of directions against equatorial and bimodal alternatives. *Biometrika* **59** 613–621.

ARNOLD, KENNETH JAMES (1941). On spherical probability distributions. Ph. D. dissertation, Massachusetts Institute of Technology.

BARNARD, G. A. (1963). Some logical aspects of the fiducial argument. *J. Roy. Statist. Soc. Ser. B* **25** 111–114.

BARTLETT, M. S. (1953a). Approximate confidence intervals. *Biometrika* **40** 12–19.

BARTLETT, M. S. (1953b). Approximate confidence intervals II. More than one unknown parameter. *Biometrika* **40** 306–317.

BARTLETT, M. S. (1955). Approximate confidence intervals III. A bias correction. *Biometrika* **42** 201–204.

BINGHAM, CHRISTOPHER (1964). Distributions on the sphere and on the projective plane. Ph. D. dissertation, Yale Univ.

BINGHAM, CHRISTOPHER (1972). An asymptotic expansion for the distribution of the eigenvalues of a 3 by 3 Wishart matrix. *Ann. Math. Statist.* **43** 1498–1506.

BREITENBERGER, ERNST (1963). Analogues of the normal distribution on the circle and the sphere. *Biometrika* **50** 81–88.

DIMROTH, ERICH (1962). Untersuchungen zum Mechanismus von Blastesis und Syntexis in Phylliten und Hornfelsen des südwestlicken Fichtelgebirges I. Die statistische Auswertung einfacher Gurteldiagramme. *Tschermaks mineralogische und petrographische Mitteilungen* **8** 248–274.

DIMROTH, ERICH (1963). Fortschritte der Gefügestatistik. *N. Jahrbuch für Mineralogie.* Monatshefte 1963 186–192.

ERDÉLYI, A. (ed.) (1953). *Higher Transcendental Functions*, **1**. McGraw-Hill, New York.
FISHER, R. A. (1953). Dispersion on a sphere. *Proc. Roy. Soc. Ser. A* **217** 295–305.
HANSEN, EDWARD C. (1963). Personal communication.
HERZ, CARL S. (1955). Bessel functions of matrix argument. *Ann. of Math.* **61** 474–523.
JAMES, A. T. (1954). Normal multivariate analysis and the orthogonal group. *Ann. Math. Statist.* **25** 40–75.
JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35** 475–501.
RAO, C. RADHAKRISHNA (1965). *Linear Statistical Inference and its Applications*. Wiley, New York.
ROBERTS, P. H. and URSELL, H. D. (1960). Random walk on a sphere and on a Riemannian manifold. *Philos. Trans. Roy. Soc. London Ser. A* **252** 317–356.
STEPHENS, M. A. (1965). The distribution of $S = \sum_{i=1}^{N} \cos^2 \theta_i$: 3 dimensions. *Biometrika* **52** 193–201. (Appendix to Watson (1965)).
VON NEUMANN, J. (1937). Some matrix-inequalities and metrization of matric space. *Forsch. Inst. Math. u. Mech. Univ. Tomsk* **1** 286–299.
WATSON, G. S. (1965). Equatorial distributions on a sphere. *Biometrika* **52** 193–201.
WILKS, SAMUEL S. (1962). *Mathematical Statistics*. Wiley, New York.

DEPARTMENT OF APPLIED STATISTICS
UNIVERSITY OF MINNESOTA
ST. PAUL, MINNESOTA 55101