Statistics of Directional Data

Author(s): K. V. Mardia

Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, 1975, Vol. 37, No. 3 (1975), pp. 349-393

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/2984782

Statistics of Directional Data

By K. V. MARDIA

Leeds University

[Read before the ROYAL STATISTICAL SOCIETY at a meeting organized by the RESEARCH SECTION on Wednesday, March 19th, 1975, Professor R. L. PLACKETT in the Chair]

SUMMARY

Directional data analysis is emerging as an important area of statistics. Within the past two decades, various new techniques have appeared, mostly to meet the needs of scientific workers dealing with directional data. The paper first introduces the two basic models for the multi-dimensional case known as the von Mises–Fisher distribution and the Bingham distribution. Their sampling distribution theory depends heavily on the isotropic case and some developments are discussed. An optimum property of an important test for the von Mises–Fisher case is established. A non-parametric test is proposed for the hypothesis of independence for observations on a torus. In addition to some numerical examples on the preceding topics, five case studies are given which illuminate the power of this new methodology. The case studies are concerned with cancer research, origins of comets, arrival times of patients, navigational problems and biological rhythms. Some unsolved problems are also indicated.

Keywords: ARRIVAL TIMES; BINGHAM DISTRIBUTION; BIOLOGICAL RHYTHMS; BIRD NAVIGATION; CANCER CELLS; CHARACTERIZATION; CORRELATION ON TORUS; DIRECTIONAL DATA; INDEPENDENCE; ORIGIN OF COMETS; RANDOM WALK; UNIFORM SCORES; VON MISES–FISHER DISTRIBUTION

1. INTRODUCTION

THERE are various statistical problems which arise in the analysis of data when the observations are directions. Directional data are often met in astronomy, biology, geology, medicine and meteorology, such as in investigating the origins of comets, solving bird navigational problems, interpreting paleomagnetic currents, assessing variation in the onset of leukaemia, analysing wind directions, etc.

The directions are regarded as points on the circumference of a circle in two dimensions or on the surface of a sphere in three dimensions. In general, directions may be visualized as points on the surface of a hypersphere but observed directions are obviously angular measurements.

The subject has recently been receiving increasing attention, but the field is as old as the subject of mathematical statistics itself. Indeed, the theory of errors was developed by Gauss primarily to analyse certain directional measurements in astronomy. The breakthrough in the subject is marked by a pioneering paper of R. A. Fisher which appeared only two decades ago (Fisher, 1953); his work was motivated by a paleomagnetic problem posed by a geophysicist, J. Hospers. Since then, thanks mostly to G. S. Watson and M. A. Stephens, the development of the subject has been rapid. It is interesting to note that Karl Pearson was involved with a bird-migration problem leading to the isotropic random walk on a circle as early as 1905. Also, von Mises introduced an important circular distribution in 1918 to study the deviation of atomic weights from integral values. For further historical notes, we refer to Mardia (1972a, pp. xvii–xix).

The main aim of this paper is to cover certain new topics which either illuminate the methodology or raise new data-analytic problems. However, the paper does not aim to give a review of the subject as a whole since this has already been covered by Mardia (1972a) for the circular and the spherical cases. Some new mathematical results are given, but considerations of space have made it necessary to present their proofs in other papers.

Some important distributions on a p-dimensional hypersphere are introduced in Section 2. These include the von Mises–Fisher distribution which has certain characterizations analogous to those of the linear normal distribution. Some other desirable characterizations lead to the Brownian motion distribution on a hypersphere. However, the von Mises–Fisher distribution leads to tractable maximum likelihood estimates and sampling distributions in problems of hypothesis testing, whereas the Brownian motion distribution does not. Therefore, the von Mises–Fisher distribution is widely studied and used by statisticians. We discuss some information theoretic characterizations of basic models, and provide an appropriate model for two correlated vectors each having a von Mises–Fisher distribution. Section 3 gives some important results on sampling distribution theory for the von Mises–Fisher population, which leans heavily on the isotropic random walk on a hypersphere. The method of derivation is briefly outlined. An inference problem related to the population mean direction for the von Mises–Fisher distribution is treated in Section 4. Sections 2–4 have resulted from an attempt to unify some of the parametric work; Watson and Williams (1956) were the first to attempt such a unification for the von Mises–Fisher distribution. Section 5 deals with some problems associated with the Bingham distribution which is appropriate when the data are axial. Section 6 looks into the concept of dependence for two circular variables and proposes a non-parametric test to test for independence. Section 7 gives some additional case studies while Section 8 indicates some unsolved problems and outlines some problems not discussed here.

2. Basic Models

Let $\boldsymbol{l}^{\mathrm{T}} = (l_1, \ldots, l_p)$ be a unit random vector taking values on the surface of a p-dimensional hypersphere S_p of unit radius and having its centre at the origin. The vector $\boldsymbol{l}$ can be viewed as a vector of direction cosines. Some important directional distributions are as follows. Their discussion for $p = 2$ and $p = 3$ will be found in Mardia (1972a) and for general p in Mardia (1975a, b).

2.1. *The von Mises–Fisher Distribution*

A random vector $\boldsymbol{l}$ is said to have a p-variate von Mises–Fisher distribution if its probability density function (p.d.f.) is given by

$$c_p(\kappa) \exp(\kappa \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{l}), \quad \kappa > 0, \quad \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{\mu} = 1, \quad \boldsymbol{l} \in S_p, \tag{2.1}$$

where κ is the concentration parameter, $\boldsymbol{\mu}$ is the population mean-direction vector (i.e. $E(\boldsymbol{l}) = \rho\boldsymbol{\mu}$, $\rho > 0$), and

$$c_p(\kappa) = \kappa^{\frac{1}{2}p-1} / \{(2\pi)^{\frac{1}{2}p} I_{\frac{1}{2}p-1}(\kappa)\}, \tag{2.2}$$

with $I_r(\kappa)$ denoting the modified Bessel function of the first kind and order r. If $\boldsymbol{l}$ has its p.d.f. of the form (2.1.), we will say that it is distributed as $M_p(\boldsymbol{\mu}, \kappa)$. For $p = 2$ and $p = 3$, (2.1) reduces to p.d.f.'s of the von Mises and the Fisher distributions respectively. These particular cases explain the nomenclature for $p > 3$ but the distribution (2.1) was first introduced by Watson and Williams (1956).

For $\kappa = 0$, the distribution reduces to the uniform distribution on S_p with p.d.f.

$$f(\boldsymbol{l}) = c_p, \quad \boldsymbol{l} \in S_p, \tag{2.3}$$

where

$$c_p = c_p(0) = \Gamma(\tfrac{1}{2}p)/(2\pi^{\frac{1}{2}p}). \tag{2.4}$$

For $\kappa > 0$, the distribution has a mode at $\boldsymbol{l} = \boldsymbol{\mu}$. The larger the value of κ, the greater is the clustering around the mean-direction vector. For this reason, it is used as a model when the data are suspected to be unimodal.

It is sometimes convenient to consider the density of $\boldsymbol{l}$ in terms of the spherical polar co-ordinates $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1, \ldots, \theta_{p-1})$ with the help of the transformation

$$\boldsymbol{l} = \mathbf{u}(\boldsymbol{\theta}), \quad 0 < \theta_i \leqslant \pi, \quad i = 1, \ldots, p-2, \quad 0 < \theta_{p-1} \leqslant 2\pi, \tag{2.5}$$

where

$$\mathbf{u}(\boldsymbol{\theta}) = (u_1(\boldsymbol{\theta}), \ldots, u_p(\boldsymbol{\theta}))^{\mathrm{T}},$$

$$u_j(\boldsymbol{\theta}) = \cos\theta_j \prod_{i=0}^{j-1} \sin\theta_i, \quad j = 1, \ldots, p, \quad \sin\theta_0 = \cos\theta_p = 1. \tag{2.6}$$

The density (2.1) then reduces to

$$g(\boldsymbol{\theta};\, \boldsymbol{\mu}_0, \kappa) = c_p(\kappa) \exp\{\kappa \mathbf{u}^{\mathrm{T}}(\boldsymbol{\mu}_0)\, \mathbf{u}(\boldsymbol{\theta})\}\, a_p(\boldsymbol{\theta}), \tag{2.7}$$

where $\boldsymbol{\mu}_0 = (\mu_{0,1}, \ldots, \mu_{0,p-1})^{\mathrm{T}}$ denotes the spherical polar co-ordinates of $\boldsymbol{\mu}$, and

$$a_p(\boldsymbol{\theta}) = \prod_{j=2}^{p-1} \sin^{p-j} \theta_{j-1}. \tag{2.8}$$

For $p = 2$, we will say that θ_1 is distributed as $M(\mu_0, \kappa)$.

2.2. *The Bingham Distribution*

Another important model arises when the observations are axial so the p.d.f. of $\boldsymbol{l}$ satisfies the antipodal-symmetry property

$$f(\boldsymbol{l}) = f(-\boldsymbol{l}).$$

How to construct an appropriate model? First, consider a simple construction of the von Mises–Fisher distribution. Let $\mathbf{x}$ be distributed as $N_p(\boldsymbol{\mu}, \kappa^{-1}\mathbf{I})$ with $\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} = 1$. Then the conditional distribution of $\mathbf{x}$ given $\mathbf{x}^{\mathrm{T}}\mathbf{x} = 1$ is $M_p(\boldsymbol{\mu}, \kappa)$. Since, for the conditional distribution to be axial, there should be no terms involving $\boldsymbol{\mu}$, the mean vector of the normal distribution, we now assume that $\mathbf{x}$ is $N_p(\mathbf{0}, \boldsymbol{\Sigma})$. Then the conditional distribution of $\mathbf{x}$ given $\mathbf{x}^{\mathrm{T}}\mathbf{x} = 1$ leads to a p.d.f. of the following form:

$$f(\boldsymbol{l};\, \boldsymbol{\mu}, \boldsymbol{\kappa}) = \text{const} \times \exp\{\operatorname{tr}(\boldsymbol{\kappa}\boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{l}\boldsymbol{l}^{\mathrm{T}} \boldsymbol{\mu})\}, \quad \boldsymbol{l} \in S_p, \tag{2.9}$$

where $\boldsymbol{\mu}$ now denotes an orthogonal matrix, $\boldsymbol{\kappa}$ is a diagonal matrix of constants and the normalizing constant depends only on $\boldsymbol{\kappa}$. Since $\operatorname{tr}(\boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{l}\boldsymbol{l}^{\mathrm{T}} \boldsymbol{\mu}) = 1$, the sum of the parameters κ_i is arbitrary, and it is usual to take $\kappa_p = 0$. Different values of $\boldsymbol{\kappa}$ in (2.9) give the uniform distribution, symmetric and asymmetric girdle distributions and bimodal distributions. The density for the general case is due to Bingham (1964) who investigated the distribution extensively for statistical applications for $p = 3$. For $p = 2$, the distribution is called the bimodal distribution of von Mises type.

2.3. *Information Theoretic Characterizations*

Various characterizations of directional distributions are discussed in Bingham and Mardia (1975) and Mardia (1975a). One of the most important characterizations of these distributions involves maximizing entropy under certain constraints. The proof is obtained from the following result of Mardia (1975a) which turns out to be an extension of Theorem 13.2.1 of Kagan *et al.* (1973, p. 409).

Suppose that distributions defined over a space S are to be represented by densities relative to some familiar measure such as Lebesgue, Haar, etc. Let $t_1, \ldots, t_q$ represent q given real valued functions over S such that no linear combination of $t_1, \ldots, t_q$ is constant. If, for a density f,

(i) S^* is the support of $f(s)$ where $s \in S^*$, $S^* \subset S$,
(ii) $E\{t_i(s)\} = a_i$, fixed, $i = 1, \ldots, q$, and
(iii) the entropy is maximized,

then the p.d.f. is of the form

$$f(s) = \exp\left\{b_0 + \sum_{i=1}^{q} b_i\, t_i(s)\right\}, \quad s \in S^*, \tag{2.10}$$

provided there exist $b_0, b_1, \ldots$ such that (2.10) satisfies (i) and (ii). Further, if there exists such a density then it is unique.

Thus the maximum entropy distributions with the fixed mean vector $E(\boldsymbol{l})$ and fixed "moment of inertia" $E(\boldsymbol{l}\boldsymbol{l}^{\mathrm{T}})$ are the von Mises–Fisher and the Bingham distributions respectively. The characterizations for the von Mises distribution and the Fisher distribution were first given by Mardia (1972, pp. 65–66) and Rao (1969, pp. 141–142) respectively. If both $E(\boldsymbol{l})$ and $E(\boldsymbol{l}\boldsymbol{l}^{\mathrm{T}})$ are fixed, we can write down the p.d.f. but it has not so far received any attention.

2.4. *A Generalized von Mises–Fisher Distribution*

We now use this result in constructing a suitable distribution when two unit random vectors $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ are correlated. Obviously, we should specify $E(\boldsymbol{l}_1)$, $E(\boldsymbol{l}_2)$, $E(\boldsymbol{l}_1 \boldsymbol{l}_2^{\mathrm{T}})$ so that the maximum entropy density from (2.10) is

$$\text{const} \times \exp(\mathbf{a}_1^{\mathrm{T}} \boldsymbol{l}_1 + \mathbf{a}_2^{\mathrm{T}} \boldsymbol{l}_2 + \operatorname{tr} \mathbf{A}\, \boldsymbol{l}_1 \boldsymbol{l}_2^{\mathrm{T}}), \quad \boldsymbol{l}_1, \boldsymbol{l}_2 \in S_p. \tag{2.11}$$

We show that the marginal distributions of $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ are not of the von Mises–Fisher form except for trivial cases. Without any loss of generality, let us take $p = 2$. In this case, the p.d.f. can be written as

$$f(\theta, \phi) = C \exp\{\kappa_1 \cos(\theta - \mu) + \kappa_2 \cos(\phi - \nu) + a \cos\theta \cos\phi + b \sin\theta \cos\phi + c \cos\theta \sin\phi + d \sin\theta \sin\phi\}, \tag{2.12}$$

where $0 < \theta$, $\phi \leqslant 2\pi$. The marginal p.d.f. of ϕ is found to be

$$C \exp\{\kappa_2 \cos(\phi - \nu)\}\, I_0(Q^{\frac{1}{2}}), \tag{2.13}$$

where

$$Q = \kappa_1^2 + 2\kappa_1(a\cos\mu + b\sin\mu)\cos\phi + 2\kappa_1(c\cos\mu + d\sin\mu)\sin\phi + (a^2+b^2)\cos^2\phi + (c^2+d^2)\sin^2\phi + (ac+bd)\cos\phi\sin\phi. \tag{2.14}$$

Hence, the p.d.f. of ϕ is of the von Mises form if and only if Q is a constant. For $\kappa_1 \neq 0$, we have (i)

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = K\begin{pmatrix} \sin\mu & -\cos\mu \\ -\sin\mu & \cos\mu \end{pmatrix},$$

where K is a constant and (ii) the rows of $\mathbf{A}$ are orthogonal. Hence, we have $K = 0$. For $K = 0$, the variables θ and ϕ are independently distributed. If $\kappa_1 = 0$ then either $\mathbf{A} = \mathbf{0}$ or const. $\times\,\mathbf{A}$ is orthogonal, and so θ is not von Mises although it can be uniform. Thus, our assertion is established. However, the model (2.12) possesses some important properties (Mardia, 1975a). Let us take $\mathbf{A} = \rho(\kappa_1\kappa_2)^{\frac{1}{2}}\mathbf{B}$ in (2.12) where $\mathbf{B}$ is an orthogonal matrix. The parameter ρ then behaves as a measure of circular dependence. As $\kappa_1, \kappa_2 \to \infty$, $(\kappa_1^{\frac{1}{2}}\theta, \kappa_2^{\frac{1}{2}}\phi)$ tends to a bivariate normal with limiting correlation depending on ρ. For $\rho = 0$, θ and ϕ are independently distributed and each has a von Mises distribution. When one or both of the variables are uniformly distributed, ρ has no meaning. The question of a measure of circular dependence will be discussed further in Section 6.

Two other methods of constructing directional distributions consist of considering the distribution of $\boldsymbol{l}$ to be either a conditional distribution or the marginal distribution of an appropriate multinormal distribution (real or complex). We have already used the first of these methods in Section 2.2.

3. Isotropic Random Walk and Sampling from von Mises–Fisher Populations

3.1. *Introduction*

Let $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_n$ be a random sample from $M_p(\boldsymbol{\mu}, \kappa)$. Inference problems for $\boldsymbol{\mu}$ and κ depend on the sufficient statistics $\sum \boldsymbol{l}_i^{\mathrm{T}}$ which we shall denote in the following different ways.

$$\sum_{i=1}^{n} \boldsymbol{l}_i^{\mathrm{T}} = (R_{x_1}, \ldots, R_{x_p}) = \mathbf{R}_x^{\mathrm{T}} = R\boldsymbol{l}_0^{\mathrm{T}} = R\mathbf{u}^{\mathrm{T}}(\bar{\mathbf{x}}_0), \tag{3.1}$$

Following the terminology of mechanics, $R_{x_1}, \ldots, R_{x_p}$ are described as $x_1, \ldots, x_p$-components of $\boldsymbol{l}_i$, $i = 1, \ldots, n$. The length and direction of the resultant are given by R and $\bar{\mathbf{x}}_0$ respectively, which are described as the sample resultant length and the sample mean-direction vector respectively.

In most cases, the sampling distribution of R and $\bar{\mathbf{x}}_0$ from the von Mises–Fisher population can be derived with the help of the corresponding distributions for the uniform case. For example, if $f_\kappa(\mathbf{R}_x; \boldsymbol{\mu})$ denotes the p.d.f. of $\mathbf{R}_x$ for the von Mises–Fisher case, then, by a simple argument,

$$f_\kappa(\mathbf{R}_x; \boldsymbol{\mu}) = \text{const} \times \exp(\kappa\boldsymbol{\mu}^{\mathrm{T}}\mathbf{R}_x)\, f_0(\mathbf{R}_x),$$

where $f_0(\mathbf{R}_x)$ denotes the p.d.f. of $\mathbf{R}_x$ for the uniform case. In fact, the distribution theory for the uniform case is comparatively well developed as it is related to the problem of the isotropic random walk on a hypersphere. Various results can be derived for the uniform case by using a characteristic function method which we shall describe in Section 3.2.

3.2. *The Characteristic Function Method and the Isotropic Case*

Let $\boldsymbol{l}$ be a unit random vector having a singular density. Suppose that the characteristic function (c.f.) of $\boldsymbol{l}$ is

$$\phi(\mathbf{t}) = E\{\exp(i\mathbf{t}^{\mathrm{T}}\boldsymbol{l})\}, \quad \mathbf{t}^{\mathrm{T}} = (t_1, \ldots, t_p). \tag{3.2}$$

Then the p.d.f. of $\mathbf{R}_x$ can be obtained from the inversion theorem and so on transforming $\mathbf{t}$ to the spherical polar co-ordinates ρ and $\boldsymbol{\Phi}^{\mathrm{T}} = (\Phi_1, \ldots, \Phi_{p-1})$ with the help of the transformation $\mathbf{t} = \rho\mathbf{u}(\boldsymbol{\Phi})$, we can find the p.d.f. of $(R, \bar{\mathbf{x}}_0)$.

For the isotropic random walk with n equal steps starting from the origin in p-dimensions, the direction-vectors $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_n$ of the n-steps can be regarded as a random sample from the population with p.d.f. (2.3). Hence, it is found that (i) R and $\bar{\mathbf{x}}_0$ are independently distributed, (ii) $\bar{\mathbf{x}}_0$ is uniformly distributed on S_p and (iii) the p.d.f. of R is $h_n(R)$ which is the value of (3.3) at $\beta_1 = \ldots = \beta_n = 1$. If the successive steps are of different length $\beta_1, \ldots, \beta_n$, then the p.d.f. of R is

$$h_n(R;\, \beta_1, \ldots, \beta_n) = (2\pi)^{-p}\, c_p^{n-1}\, R^{p-1} \int_0^\infty \rho^{p-1} \left\{ c_p(i\rho R) \prod_{j=1}^{n} c_p(i\rho\beta_j) \right\}^{-1} d\rho, \tag{3.3}$$

where

$$c_p(i\rho) = \rho^{\frac{1}{2}p-1} / \{(2\pi)^{\frac{1}{2}p}\, J_{\frac{1}{2}p-1}(\rho)\}. \tag{3.4}$$

The distribution of R given by (3.3) was first derived by G. N. Watson (1948, p. 421) who extended a method of Kluyver (1906) for $p = 2$; the method involves highly specialized Bessel function theory. For $p = 3$, the distribution has been simplified by Mardia (1972a, pp. 238–240) on relating the integral (3.3) to a c.f. of some uniform variables on a line. The distribution for $p = 2$ and $p = 3$ has received the attention of various workers (see Mardia, 1972a, pp. 96, 240).

3.3. *Distributional Problems for the von Mises–Fisher Case*

3.3.1. *Single sample problems*

We now quote some results from Mardia (1975b) which are obtained in general from Section 3.2. on using the process described in Section 3.1. Let us assume that $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_n$ is a random sample from $M_p(\boldsymbol{\mu}, \kappa)$.

(i)
$$\bar{\mathbf{x}}_0 \,|\, R \sim M_p(\boldsymbol{\mu}, \kappa R). \tag{3.5}$$

The distribution of $\bar{\mathbf{x}}_0$ has not received sufficient attention although it plays an important role.

(ii) The p.d.f. of R_{x_1} is

$$f(R_{x_1}) = \pi^{-1} c_p^n(\kappa) \exp(\kappa\mu R_{x_1})\, b(R_{x_1};\, \lambda), \quad -\infty < R_{x_1} < \infty, \tag{3.6}$$

where

$$b(R_{x_1};\, \lambda) = \int_0^\infty \cos tR_{x_1}\, c_p^{-n}\{i(t^2 - \lambda^2)^{\frac{1}{2}}\} dt, \quad \mu = \cos\mu_{0,1}, \quad \lambda = \kappa \sin\mu_{0,1}. \tag{3.7}$$

(iii) The conditional p.d.f. of $R \,|\, R_{x_1}$ is

$$\pi c_p^{-n} [c_{p-1}\{\lambda(R^2 - R_{x_1}^2)^{\frac{1}{2}}\}]^{-1}\, h_n(R)\, (R^2 - R_{x_1}^2)^{\frac{1}{2}(p-3)} / b(R_{x_1};\, \lambda), \quad 0 < R_{x_1}^2 < R^2. \tag{3.8}$$

Intuitively, it is not clear why the distribution should depend only on λ.

3.3.2. *Multi-sample problem*

Let $R_1, \ldots, R_q$ denote the resultants for q independent random samples of sizes $n_1, \ldots, n_q$ drawn from $M_p(\boldsymbol{\mu}, \kappa)$. Let $n = \sum n_i$. Suppose that R is the resultant for the combined sample. Extending the argument of Mardia (1972a, pp. 243–244) it is found that (Mardia, 1975b)

$$f\kappa(\mathbf{R}^* \mid R) = h_q(R;\, \mathbf{R}^*) \prod_{i=1}^{q} h_{n_i}(R_i)/h_n(R), \tag{3.9}$$

where $h_q(R; \mathbf{R}^*)$ and $h_n(R)$ are given at (3.3) and $\mathbf{R}^* = (R_1, \ldots, R_q)^{\mathrm{T}}$. Hence the density of $\mathbf{R}^* \mid R$ does not depend on κ. This fact leads to its applications for inference on $\boldsymbol{\mu}$ when κ is unknown.

It should be noted that the result (3.9) was derived for $p = 3$ and $q = 2$ by Fisher (1953). For $p = 2$, the result is due to J. S. Rao (1969). For $p = 3$, Mardia (1972a, pp. 242–244) provides the exact solution to the problem and also gives historical points. For $q = 2$, (3.9) can be further simplified (Mardia, 1975b).

4. AN INFERENCE PROBLEM FOR SAMPLES FROM VON MISES–FISHER POPULATIONS

Suppose that $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_n$ is a random sample from $M_p(\boldsymbol{\mu}, \kappa)$. Consider the problem of testing

$$H_0\colon \boldsymbol{\mu}_0 = \mathbf{0} \quad \text{against} \quad H_1\colon \boldsymbol{\mu}_0 \neq \mathbf{0}, \tag{4.1}$$

where κ is unknown. Since R_{x_1} is a complete sufficient statistic for κ when H_0 is true and $R_{x_1}, \ldots, R_{x_p}$ are jointly complete and sufficient for $\kappa, \boldsymbol{\mu}_0$ when H_1 is true, the use of the conditional distribution of $\mathbf{R}_x = (R_{x_1}, \ldots, R_{x_p})^{\mathrm{T}} \mid R_{x_1}$ can only lead to similar tests (see Lehmann, 1959, p. 130). It can be seen that there is no *UMP* similar test.

A UMP invariant test. Let us modify the hypotheses (4.1) to

$$H_0\colon \boldsymbol{\mu} = \mathbf{e}_1, \quad H_1\colon \boldsymbol{\mu} \neq \mathbf{e}_1 \tag{4.2}$$

where $\mathbf{e}_1 = (\pm 1, 0, 0, \ldots, 0)^{\mathrm{T}}$. The problem remains invariant under the orthogonal transformations $\boldsymbol{l}^* = \mathbf{A}\boldsymbol{l}$, where $\mathbf{A}$ is an orthogonal matrix with the first column as $\mathbf{e}_1$. The condition on $\mathbf{A}$ ensures that R_{x_1} is invariant under the transformation. A function of $\mathbf{R}_x$ invariant under the transformation is the resultant length

$$R = \left\{\left(\sum_{i=1}^{n} \boldsymbol{l}_i\right)^{\mathrm{T}} \left(\sum_{i=1}^{n} \boldsymbol{l}_i\right)\right\}^{\frac{1}{2}}.$$

Indeed, following the argument of Lehmann (1959, p. 297), it is seen that R is a maximal invariant. Hence the invariant tests should be based on $R \mid R_{x_1}$. We now show that the region

$$\{R > K \mid R_{x_1}\} \tag{4.3}$$

provides a *UMP* invariant test. By the Neyman–Pearson lemma, the best invariant critical region is given by

$$f(R \mid R_{x_1};\, \boldsymbol{\mu}, \kappa)/f(R \mid R_{x_1};\, \mathbf{e}_1, \kappa) \geqslant K, \tag{4.4}$$

where the p.d.f. of $R \mid R_{x_1}$ is given by (3.8). It is seen that (4.4) reduces to

$$I_{\frac{1}{2}p-3}\{\lambda(R^2 - R_{x_1}^2)^{\frac{1}{2}}\}/b(R_{x_1};\, \lambda) \geqslant K, \quad \lambda \neq 0.$$

Since $b(R_{x_1};\lambda)$ is a factor in the p.d.f. of R_{x_1} given by (3.6) it is positive. Further, for $Z>0$, $I_{\frac{1}{2}p-3}(Z)$ is a monotonically increasing function of Z. Thus the best invariant critical region for all values of λ is given by (4.3). The result follows.

It should be noted that the power of the test contains the "non-centrality parameter" λ. The test was recommended by Watson and Williams (1956) on intuitive grounds. Under the null hypothesis, the p.d.f. of $R\,|\,R_{x_1}$ is an even function of R_{x_1} so that the above test is strictly a test for a prescribed axis rather than of a fixed direction.

Confidence cone. Let R_0 be the observed value of R, and let $R_{x_1,\alpha}>0$ and δ satisfy the equations

$$P(R>R_0\,|\,R_{x_1}=R_{x_1,\alpha} \text{ and lies in the same direction as } \boldsymbol{\mu}_0)=\alpha,$$

$$R_{x_1,\alpha}=R_0\cos\delta,\quad 0<\delta\leqslant\tfrac{1}{2}\pi. \tag{4.5}$$

Let us assume that $\boldsymbol{\mu}_0=\mathbf{0}$. Then $R_{x_1}=R\cos\bar{x}_{0,1}$ where $\bar{x}_{0,1}$ denotes the angle between the true mean direction and the sample mean direction. Consequently, the probability that the true mean direction lies within a cone with vertex at the origin, axis as the sample mean direction and semi-vertical angle δ is $1-\alpha$ provided that $0<\bar{x}_{0,1}<\frac{1}{2}\pi$. Hence, the confidence coefficient is not $1-\alpha$ but $1-\alpha'$ such that

$$1-\alpha'=(1-\alpha)\,p^*, \tag{4 6}$$

where

$$p^*=P(0<\bar{x}_{0,1}<\tfrac{1}{2}\pi),$$

when the underlying population is $M_p(\mathbf{0},\kappa)$.

Exact values of p^* can be calculated from (3.6) since it is equivalent to $P(R_{x_1}>0)$ when $\mu=1$ and $\lambda=0$ in (3.6). For large κ, $p^*\to 1$ so that $\alpha'\to\alpha$. Table 1 gives a

TABLE 1

Confidence cone for $\boldsymbol{\mu}$: *comparison of actual confidence coefficient* $1-\alpha'$, *and pseudo-confidence coefficient* $1-\alpha$

p	$1-\alpha$	$n\to$	10	10	20	20	40	40
2		$\kappa\to$	0·74	1·08	0·53	0·74	0·37	0·53
	0·95		0·90	0·94	0·90	0·94	0·90	0·94
	0·99		0·94	0·98	0·94	0·98	0·94	0·98
3		$\kappa\to$	0·91	1·30	0·64	0·91	0·45	0·64
	0·95		0·90	0·94	0·90	0·94	0·90	0·94
	0·99		0·94	0·98	0·94	0·98	0·94	0·98

comparison of the actual confidence coefficient $1-\alpha'$ and the "pseudo-confidence coefficient" $1-\alpha$ by using the values of p^* from Pearson and Hartley (1972, pp. 127–131). Thus for moderately large values of n or κ, α' is approximately equal to α. Of course, we can construct an exact two-sided cone since then the factor $p^*=1$.

An approximate confidence cone can be obtained using (3.5). Let $\hat{\kappa}$ be the maximum likelihood estimator (m.l.e.) of κ and let $\kappa' = R\hat{\kappa}$. Suppose that θ_1 is distributed as the first marginal variable of $M_p(\boldsymbol{\mu}, \kappa')$. That is, the p.d.f. of θ_1 is

$$f(\theta_1) = \{c_p(\kappa')/c_p\} \exp(\kappa' \cos\theta_1), \quad 0 < \theta < \pi.$$

We can obtain a δ' such that

$$P(\pi - \delta' < \theta_1 < \pi) = \alpha.$$

An approximate $(1-\alpha)$-confidence cone for $\boldsymbol{\mu}$ is the cone with vertex at the origin, the axis as the sample mean direction and the semi-vertical angle δ'.

A question arises whether the true mean direction should point towards the sample mean direction or could it point in the opposite direction? If so, how to determine which one is the appropriate cone?

Mardia (1975b) deals with optimum properties of various other tests including multi-sample tests depending on (3.9).

5. SAMPLING FROM THE BINGHAM DISTRIBUTION

5.1. *Distribution of* $\mathbf{T}$ *and* $\boldsymbol{\tau}$ *for the Isotropic Case*

Let us assume that $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_n$ is a random sample from the Bingham population with p.d.f. (2.9). Inference problems for this case are found to depend on the matrix of sum of squares and products

$$\mathbf{T} = \sum_{i=1}^{n} \boldsymbol{l}_i \boldsymbol{l}_i^{\mathrm{T}} \tag{5.1}$$

and the eigenvalues $\tau_1, \ldots, \tau_p$ of $\mathbf{T}$. As in the von Mises–Fisher case, the sampling distribution theory for the Bingham population depends heavily on the distribution theory for the uniform case. However, the exact distributions of $\mathbf{T}$ and $\boldsymbol{\tau}^{\mathrm{T}} = (\tau_1, \ldots, \tau_p)$ for the uniform case are unknown and this fact has somewhat hampered further progress (cf. Bingham, 1964; Anderson and Stephens, 1972). We briefly outline the solution to this problem without giving rigorous detail. For simplicity, let us assume $p = 3$ so that the p.d.f. of $\boldsymbol{l}$ is given by

$$f(\boldsymbol{l};\, \boldsymbol{\mu}, \boldsymbol{\kappa}) = \{4\pi d(\boldsymbol{\kappa})\}^{-1} \exp\left\{\sum_{i=1}^{3} \kappa_i (\boldsymbol{l}^{\mathrm{T}} \boldsymbol{\mu}_i)^2\right\}, \tag{5.2}$$

where $d(\boldsymbol{\kappa})$ is a confluent hypergeometric function (Mardia, 1972).

We first obtain the distribution of $\mathbf{T}$ by the c.f. method. Let $\mathbf{U} = (u_{ij})$ be a symmetric matrix of order 3×3 with $\mathrm{tr}(\mathbf{U}) = 1$. The c.f. of $\boldsymbol{l}\boldsymbol{l}^{\mathrm{T}}$ is given by

$$\phi(\mathbf{U}) = (4\pi)^{-1} \int_{S_3} \exp(i\boldsymbol{l}^{\mathrm{T}} \mathbf{U} \boldsymbol{l})\, d\boldsymbol{l}. \tag{5.3}$$

There exists an orthogonal matrix $\mathbf{C}$ such that $\mathbf{U} = \mathbf{C}^{\mathrm{T}} \boldsymbol{\Lambda} \mathbf{C}$ where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$. Substituting the value of $\mathbf{U}$ in (5.3) and transforming $\boldsymbol{l}$ to $\boldsymbol{l}^* = \mathbf{C}\boldsymbol{l}$, it is found from (5.2) that $\phi(\mathbf{U}) = d(i\boldsymbol{\Lambda})$. Hence, the c.f. of $\mathbf{T}$ is known, which, when used in the inversion theorem, gives the p.d.f. of $\mathbf{T}$ to be

$$f(\mathbf{T}) = 4(2\pi)^{-3} \int_{\Sigma\lambda_i = 1} \left[\int_{O(3)} \exp\{-i\,\mathrm{tr}(\mathbf{T}\mathbf{C}^{\mathrm{T}} \boldsymbol{\Lambda} \mathbf{C})\}\, d\mathbf{C}\right] d^n(i\boldsymbol{\Lambda}) \prod_{i<j} (\lambda_i - \lambda_j)\, d\boldsymbol{\Lambda}, \tag{5.4}$$

where the inner integral is over the group of orthogonal matrices of order 3 denoted by O(3) and the outer integral is taken over λ_1, λ_2 and λ_3 with $\lambda_1 + \lambda_2 + \lambda_3 = 1$ $\lambda_1 > \lambda_2 > \lambda_3$.

In deriving (5.4) we have used the decomposition

$$d\mathbf{U} = \prod_{i<j} (\lambda_i - \lambda_j)\, d\mathbf{C}\, d\boldsymbol{\lambda};$$

the differential form $d\mathbf{C}$ representing invariant Haar measure on O(3) (see Kshirsagar, 1972, pp. 517–518). Using equation (60) of James (1964), (5.4) reduces to

$$f(\mathbf{T}) = 4(2\pi)^{-3} \sum_{k=0}^{\infty} \sum_{m} \frac{(-i)^k C_m(\mathbf{T})}{k!\, C_m(\mathbf{I}_3)} \int_{\Sigma\lambda_i=1} C_m(\boldsymbol{\Lambda})\, d^n(i\boldsymbol{\Lambda}) \prod_{i<j} (\lambda_i - \lambda_j)\, d\boldsymbol{\Lambda}, \tag{5.5}$$

where the inner summation is taken over all partitions m of k into p or fewer parts, $\mathbf{I}_3$ is the 3×3 identity matrix and $C_m(\cdot)$ is as defined in James (1964).

Let us transform $\mathbf{T}$ in (5.5) with the help of $\mathbf{T} = \mathbf{D}^{\mathrm{T}} \boldsymbol{\tau} \mathbf{D}$, where $\mathbf{D}$ is an orthogonal matrix. Using the Jacobian given in Kshirsagar (1972, Example 83, p. 517), it is found that $\mathbf{D}$ and $\boldsymbol{\tau}$ are independently distributed, and the p.d.f. of $\boldsymbol{\tau}$ is given by

$$4\pi^{-1} \prod_{i<j} (\tau_i - \tau_j) \sum_{k=0}^{\infty} \sum_{m} \frac{(-i)^k C_m(\boldsymbol{\tau})}{k!\, C_m(\mathbf{I}_3)} \int_{\Sigma\lambda_i=1} C_m(\boldsymbol{\Lambda})\, d^n(i\boldsymbol{\Lambda}) \prod_{i<j} (\lambda_i - \lambda_j)\, d\boldsymbol{\Lambda}, \tag{5.6}$$

where $\tau_1 > \tau_2 > \tau_3 > 0$ and $\tau_1 + \tau_2 + \tau_3 = 1$. Hence, the exact distribution of $\boldsymbol{\tau}$ is known but, as in the normal case, (5.6) is not manageable. An approximation can be obtained following Bingham (1972). The asymptotic distribution of $\boldsymbol{\tau}$ has already been obtained by Anderson and Stephens (1972), using an elegant argument.

5.2. *Critical Values of a Test of Uniformity*

To test the hypothesis of uniformity against the alternative hypothesis of a Bingham distribution, we use the criterion (see Mardia, 1972a, p. 276)

$$U = S_u/n = \frac{15}{2} \sum_{i=1}^{3} (\bar{\tau}_i - \tfrac{1}{3})^2, \quad \bar{\tau}_i = \tau_i/n,$$

where the null hypothesis is rejected for large values of U. It is known that S_u is asymptotically distributed as χ_5^2. The exact distribution of U can be formally written from (5.6) but it leads to a formidable numerical integration problem and we therefore utilized a Monte Carlo method. Using 10,000 trials, we obtained simulated critical values $S_n(\alpha)$ of the test, for each value of α in Table 2, and $n = 6, 7, 8, 10, 12, 15, 30, 40, 60, 100$. Using the points $\{n^{-1}, S_n(\alpha)\}$ together with a few points from the chi-squared approximation for large n, we obtained smooth curves for each α on a computer, with the help of a general procedure which ensures consistency in the shapes of the various curves for differing α. The resulting critical values are shown in Table 2, and these should be accurate to within one unit in the second decimal place. For $n > 100$, the chi-squared approximation is adequate.

Example. From 30 measurements of direction on the c-axis of calcite grains from the Taconic mountains of New York (Bingham, 1964; Mardia, 1972a, p. 226), it is found that

$$\tau_1 = 4{\cdot}1764, \quad \tau_2 = 10{\cdot}9639, \quad \tau_3 = 14{\cdot}8595.$$

We have here $U = 0{\cdot}487$. Using Table 2, we note that the hypothesis is (almost) rejected at the 1 per cent level; the 1 per cent critical value is 0.49. The chi-squared approximation gives the 1 per cent value as $0{\cdot}503$.

TABLE 2

Critical values of the U-test for uniformity

n	$\alpha \rightarrow$ 0·001	0·005	0·01	0·025	0·05	0·075	0·1
6	3·21	2·59	2·35	1·99	1·74	1·59	1·47
7	2·82	2·27	2·05	1·74	1·51	1·37	1·27
8	2·50	2·01	1·81	1·53	1·33	1·21	1·12
9	2·24	1·80	1·62	1·37	1·19	1·08	0·99
10	2·03	1·63	1·47	1·24	1·07	0·97	0·90
11	1·85	1·48	1·34	1·13	0·98	0·88	0·82
12	1·70	1·36	1·23	1·04	0·90	0·81	0·75
14	1·46	1·17	1·05	0·89	0·77	0·70	0·64
15	1·36	1·09	0·99	0·83	0·72	0·65	0·60
16	1·28	1·02	0·92	0·78	0·67	0·61	0·56
20	1·02	0·82	0·74	0·62	0·54	0·49	0·45
30	0·68	0·55	0·49	0·42	0·36	0·33	0·30
50	0·41	0·33	0·29	0·25	0·22	0·20	0·18
60	0·34	0·27	0·25	0·21	0·18	0·16	0·15
80	0·25	0·20	0·18	0·16	0·13	0·12	0·11
100	0·20	0·16	0·15	0·12	0·11	0·10	0·09
χ_5^2	20·52	16·75	15·09	12·83	11·07	10·01	9·24

6. DEPENDENCE OF VARIABLES ON A TORUS

6.1. *Rotational Dependence*

Let θ and ϕ be two circular random variables. The standard linear correlation coefficient is supposed to measure the degree of linearity. In fact, two bivariate variables x and y are perfectly correlated if the whole mass is situated on a line. Keeping the algebraic structure of the circle in mind, we may say that θ and ϕ are perfectly correlated if the whole mass is concentrated on

$$(l\theta \pm m\phi + \psi) \bmod 2\pi = 0, \tag{6.1}$$

where l and m are positive integers and ψ is an angular quantity. Relation (6.1) can be understood easily on considering the case of $l = m = 1$. In this case, (6.1) is true if and only if

$$\theta \equiv (\psi + \phi) \bmod 2\pi \tag{6.2}$$

or

$$\theta \equiv (\psi - \phi) \bmod 2\pi. \tag{6.3}$$

In both cases, we can "rotate" ϕ anti-clockwise or clockwise by an angle ψ to match it with θ. The cases (6.2) and (6.3) have also been recently utilized by Downs (1974) in developing his measure of circular correlation. When (6.2) holds, θ and ϕ are said to be positively correlated whereas if (6.3) holds θ and ϕ are negatively correlated.

6.2. *A Circular Rank Correlation Coefficient*

Definition. Let (θ_i, ϕ_i), $i = 1, \ldots, n$, be a random sample on (θ, ϕ) and let

$$\theta_i^* \equiv l\theta_i \bmod 2\pi, \quad \phi_i^* \equiv m\phi_i \bmod 2\pi,$$

where l and m are assumed *known*. Let us assume that the linear ranks of the θ_i^*'s are $1, \ldots, n$ and those of the ϕ_i^*'s are $r_1, \ldots, r_n$. Replace the angles (θ_i^*, ϕ_i^*) by the uniform scores $(2\pi i/n, 2\pi r_i/n)$. Now for (6.3) to hold, the resultant length of the points $(\theta_i^* - \phi_i^*) \bmod 2\pi$ or $(\theta_i^* + \phi_i^*) \bmod 2\pi$, $i = 1, \ldots, n$, should be one for perfect dependence. Thus, we could define a rank correlation coefficient on the circle by

$$r_0 = \max(\bar{R}_1^2, \bar{R}_2^2), \tag{6.4}$$

where

$$n^2 \bar{R}_1^2 = \left[\sum_{i=1}^{n} \cos\{2\pi(i-r_i)/n\}\right]^2 + \left[\sum_{i=1}^{n} \sin\{2\pi(i-r_i)/n\}\right]^2$$

and

$$n^2 \bar{R}_2^2 = \left[\sum_{i=1}^{n} \cos\{2\pi(i+r_i)/n\}\right]^2 + \left[\sum_{i=1}^{n} \sin\{2\pi(i+r_i)/n\}\right]^2.$$

Obviously, $0 \leqslant r_0 \leqslant 1$. Further, r_0 is invariant under changes of zero direction in either θ or ϕ. We have $\bar{R}_1^2 = 1$ for "positive" perfect dependence while $\bar{R}_2^2 = 1$ for "negative" perfect dependence. When the variables are uncorrelated, $\bar{R}_1^2 = \bar{R}_2^2 \simeq 0$ asymptotically.

Null distribution of r_0. Table 3 gives selected critical values of r_0 for $n = 5\,(1)\,10$. For $n > 10$, the following approximation is found adequate. It can be seen that

$$\bar{R}_1^2 = (T_{cc}+T_{ss})^2 + (T_{cs}-T_{sc})^2, \quad \bar{R}_2^2 = (T_{cc}-T_{ss})^2 + (T_{cs}+T_{sc})^2,$$

where

$$T_{cc} = \frac{1}{n}\sum \cos\frac{2\pi i}{n}\cos\frac{2\pi r_i}{n}, \quad T_{cs} = \frac{1}{n}\sum \cos\frac{2\pi i}{n}\sin\frac{2\pi r_i}{n}$$

TABLE 3

Critical values of the r_0-test (upper entry); lower entry: exact level

n	$\alpha \rightarrow$ 0·001	0·01	0·05	0·10
5				1·0000
				0·0833
6		1·0000		0·6944
		0·0167		0·1167
7	1·0000	0·7964	0·6160	0·5223
	0·0028	0·0222	0·0611	0·1000
8	1·0000	0·7286	0·5335	0·4321
	0·0004	0·0115	0·0528	0·0980
9	0·8987	0·6273	0·4474	0·37†
	0·0005	0·0096	0·0515	
10	0·84†	0·59†	0·41†	0·33†
$-\log_e\{1-\sqrt{(1-\alpha)}\}$	7·6007	5·2958	3·6761	2·9697

† Values calculated from the approximation.

and T_{ss}, T_{sc} are similarly defined. Since the T's are linear rank statistics, we can obtain their moments using a result of Hajek and Sidak (1967, pp. 57–58). It is found that the vector $\frac{1}{2}(n-1)^{-\frac{1}{2}}(T_{cc}, T_{ss}, T_{cs}, T_{sc})^{\mathrm{T}}$ is asymptotically $N_4(\mathbf{0}, \mathbf{I})$. Hence $2(n-1)\bar{R}_1^2$ and $2(n-1)\bar{R}_2^2$ are independently distributed asymptotically as χ_2^2. As a consequence, for large n, $U = 2(n-1)r_0$ has the p.d.f.

$$f(u) = \exp(-\tfrac{1}{2}u)\{1-\exp(-\tfrac{1}{2}u)\}, \quad u>0,$$

so the upper percentage point $r_{0,\alpha}$ of r_0 is given by

$$r_{0,\alpha} = -(n-1)^{-1}\log_e\{1-(1-\alpha)^{\frac{1}{2}}\}.$$

The values of the logarithmic terms for some selected values of α are given in Table 3.

Example 1. In a medical experiment, various measurements were taken on 10 medical students several times daily for a period of several weeks (Downs, 1974). The estimated peak times for two successive measurements of diastolic blood pressure (converted into angles) were

$$\theta: \; 30° \; 15° \; 11° \; 4° \; 348° \; 347° \; 341° \; 333° \; 332° \; 285°$$

$$\phi: \; 25° \; 5° \; 349° \; 358° \; 340° \; 347° \; 345° \; 331° \; 329° \; 287°.$$

It is found that $\bar{R}_1^2 = 0{\cdot}731$, $\bar{R}_2^2 = 0{\cdot}004$, $r_0 = 0{\cdot}731$. The 5 per cent value of r_0 is 0·41 and therefore the null hypothesis of independence is rejected. However, if we *incorrectly* apply linear techniques, we find $r = 0{\cdot}53$ which leads to acceptance of the hypothesis at the 5 per cent level of significance under the assumption of normality. It is plausible to expect a dependence on medical grounds. The dependence is positive as reflected by the high value of $\bar{R}_1^2$. Also if we look at the differences in $i-r_i$, we get the same impression.

Example 2. The wind directions at Gorleston on February 1st to 8th, 1968, at 1300 hours each day were

$$220° \; 250° \; 280° \; 330° \; 180° \; 220° \; 140° \; 60°.$$

Is there any serial correlation between these readings? A study for all the data for 1968 suggests that the parent distribution is bimodal. Let $\theta_1, \theta_2, \ldots$ be the angles doubled. By considering the pairs (θ_1, θ_2), (θ_2, θ_3), etc., it is found that $\bar{R}_1^2 = 0{\cdot}2180$, $\bar{R}_2^2 = 0{\cdot}1897$. The 10 per cent value of r_0 for $n = 7$ from Table 3 is 0·5233. Hence, we accept the null hypothesis. If the angles are not doubled, $\bar{R}_1^2 = 0{\cdot}2108$, $\bar{R}_2^2 = 0{\cdot}1217$.

7. Case Studies

We have given some numerical examples to illustrate specific applications in previous sections. We now investigate some broader problems from various fields. For calculations of basic circular and spherical statistics, algorithms of Mardia and Zemroch (1975a, b) were used.

7.1. *Analysis of a Cancer Cell Data*

In studies of cancer (*Rous sarcoma* virus), it is of interest to investigate the orientation of cancer cells (P. Crouch, R. Weiss and H. Goldstein, 1975, unpublished). For this virus, it is possible to regard every cell as a directed line. The cells are plated onto a glass plate, which has a number of parallel equispaced grooves cut into it. The orientation of the cells can then be measured with respect to this grid. It is known that there is a tendency for the cells to align themselves with these grooves, i.e. the grooves have a physical effect on the orientation of the cells. Therefore, it is of interest to study the effect of the grid spacing x (measured as number of grooves per inch) on these orientations. As a control experiment, the cells are plated onto the reverse smooth side of the plate. Orientations are measured as before as it is possible to see the grid through the glass, but obviously now there should be no physical alignment effect.

To investigate the above problems, Mr H. Goldstein of the National Children's Bureau and Dr R. Weiss of the Imperial Cancer Research Institute conducted an experiment to measure orientations of cancer cells for seven values of x (see Table 4).

TABLE 4

Circular statistics for a set of cancer cell data

Serial no.	*No. of lines per inch/ 1,000* $= 10^{-3}x$	*Grooved side* $\bar{x}_0^* \pm \delta$	$\bar{R}$	$\hat{\kappa}$	*Smooth side* $\bar{x}_0^* \pm \delta$	$\bar{R}$	κ
1	1·25	−0·43°±1·5°‡	0·979	23·57	−72·9°±39°	0·188†	0·38
2	3·0	−0·43°±2·5°	0·965	14·67	80·35°±62°	0·138†	0·28
3	5·0	−1·19°±10°	0·591	1·48	−15·36°±46°	0·169†	0·34
4	7·5	−1·20°±4°	0·919	6·45	44·74°±62°	0·139†	0·28
5	10·0	2·91°±7°	0·739	2·28	−60·76±84°	0·058†	0·12
6	15·0	−49·56°±48°	0·163†	0·33	1·43°±78°	0·102†	0·21
7	30·0	4·37°±12°	0·502	1·16	−79·29°±58°	0·145†	0·29

† Uniformity hypothesis accepted at the 1 per cent level of significance.
‡ The 95 per cent confidence interval for μ_0^*.

Denote these values of x by x_i, $i = 1, \ldots, 7$ in ascending order. For each x, the orientation of 40 cancer cells was measured independently in turn (i) on the grooved side and (ii) on the smooth side of such a plate. The cells were not labelled so that the resulting angles are not paired. After doubling the angles, various circular statistics were obtained and are displayed in Table 4. Each value of $\bar{x}_0^*$ corresponds to the direction after dividing the mean direction of the doubled angles in the range $(-180°, 180°)$ by two. Corresponding 95 per cent confidence intervals for $\bar{x}_0^*$ are also shown. For testing the hypothesis of uniformity against a von Mises distribution for the doubled angles for each value of x, note that the 5 per cent critical value of $\bar{R}$ is 0·273 and the 1 per cent value of $\bar{R}$ is 0·336. It is interesting to see that the hypothesis of uniformity is accepted for the smooth sides for all values of x so that orientations on a smooth plate are not *informative* for cancer cells. In contrast, except for $x = 15{,}000$, the hypothesis of uniformity is rejected for the grooved plates. Let the doubled angles on the grooved plate with spacing x_i be a sample from $M(2\mu_{0\,i}^*, \kappa_i)$. It is seen that for each value of i where the distribution was found to be non-uniform, the hypothesis $\mu_{0,i}^* = 0$ is accepted. That is, the mean orientations $\mu_{0,i}^*$ do not depend on the grid spacing x—as was anticipated prior to the experiment. This is clear from the values of $\bar{x}_{0,i}$. The relationship between x_i and $\hat{\kappa}_i$, $i = 1, \ldots, 7$, is oscillatory, dampening down as $x \to \infty$. In view of the small number of points (x_i) available, instead of using a Fourier series, a linear regression of $\log_e \kappa$ on $\log_e x$ was tried. Fig. 1 shows the plot of κ against x with the corresponding fitted curve being

$$\kappa = 96727x^{-1\cdot1730}. \tag{7.1}$$

As $x \to \infty$, we have $\kappa \to 0$. It was expected from physical considerations that we would approach a uniform distribution as the grooves became closer. A uniform distribution was also expected as the grooves became infinitely wide apart ($x \to 0$) as of course then there would be so few grooves that their effect would become negligible. However, there are no data available for $x < 1{,}250$ to pursue this point further.

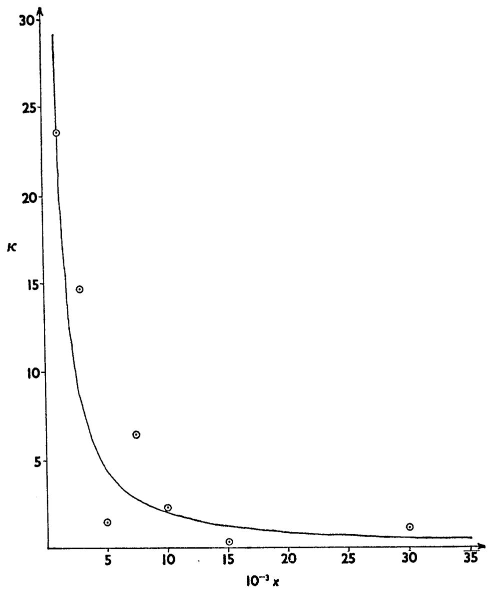

FIG. 1. Relation between the concentration parameter κ and x, the number of grooves per inch/1,000. Fitted curve and the observed points.

An alternative estimation procedure would be to regard the ith sample as drawn from $M(2\mu_0, ax_i^{-b})$ and obtain the maximum likelihood estimators of the parameters by using a numerical procedure.

7.2. *Analysis of Long-period Comet Data*

Tyror (1957) analysed the distribution of the directions of the perihelia of 448 long-period comets. From the accretion theory of Lyttleton (1953, 1961), one expects that

(i) the distribution should be non-uniform, and
(ii) the perihelion points should exhibit a preference for lying near the galactic plane.

Following Tyror (1957), we use a right-handed Cartesian system with the origin at the sun and OX and OY in the plane of the ecliptic (with OX in direction of $\phi = 0$), i.e.

$$l = \sin\theta\cos\phi, \quad m = \sin\theta\sin\phi, \quad n = \cos\theta.$$

In his notation, $\lambda = \phi$ and $\beta = 90° - \theta$. The various statistics in the notation of Mardia (1972) are as follows.

$$R = 63{\cdot}18, \quad (\bar{l}_0, \bar{m}_0, \bar{n}_0) = (-0{\cdot}0541, -0{\cdot}3316, 0{\cdot}9419), \quad \bar{x}_0 = 19{\cdot}63°, \quad \bar{y}_0 = 260{\cdot}73°,$$

$$\mathbf{T}/n = \begin{bmatrix} 0{\cdot}289988 & -0{\cdot}012393 & 0{\cdot}030863 \\ -0{\cdot}012393 & 0{\cdot}389926 & -0{\cdot}011153 \\ 0{\cdot}030863 & -0{\cdot}011153 & 0{\cdot}320086 \end{bmatrix}.$$

The matrix **T** differs only slightly from that given by Tyror (1957). (From our private correspondence, it emerges that the difference could be due to the use of ungrouped data by Tyror.) We find that the eigenvalues of $\mathbf{T}/n$ are

$$\bar{\tau}_1 = 0{\cdot}2705, \quad \bar{\tau}_2 = 0{\cdot}3347, \quad \bar{\tau}_3 = 0{\cdot}3947.$$

Consequently, $S_u = 25{\cdot}93$. Now, S_u is distributed as χ_5^2 for large n and $P(\chi_5^2 > 25{\cdot}93) = 9{\cdot}2 \times 10^{-5}$. Hence the hypothesis of uniformity is rejected.

Tyror used "an equal area" investigation involving counting the number of points per area and the points in the four immediately adjacent areas. His conclusion was the same. However, the above analysis is more appropriate since we have the alternative of a Bingham distribution in mind. The Rayleigh test leads to the same conclusion. Following Mardia (1972a, pp. 225, 277–278) it is found that the distribution is girdle. (Here $\hat{\kappa}_1 = -0{\cdot}9545$, $\hat{\kappa}_2 = -0{\cdot}3978$, $S_g = 5{\cdot}74$, $S_b = 8{\cdot}01$.) Further, the normal to the preferred great circle has the direction cosines,

$$l = 0{\cdot}8522, \quad m = 0{\cdot}0397, \quad n = -0{\cdot}5217 \tag{7.2}$$

which is the direction of the eigenvector corresponding to $\bar{\tau}_1$.

If we now assume the distribution is girdle with p.d.f.

$$c \exp\{-\kappa(\mathbf{l}^{\mathrm{T}} \boldsymbol{\mu})^2\}, \quad \kappa > 0,$$

it remains to test

$$H_0\colon \boldsymbol{\mu}^{\mathrm{T}} = (0{\cdot}8772, -0{\cdot}0536, -0{\cdot}4772) \tag{7.3}$$

which defines the direction of the galactic pole. Let λ be the likelihood ratio criterion for this problem. We have

$$-2\log_e \lambda = -2n \log b(\hat{\kappa}_0) + 2\hat{\kappa}_0 \boldsymbol{\mu}^{\mathrm{T}} \mathbf{T} \boldsymbol{\mu} + 2n \log_e b(\hat{\kappa}_1) - 2\hat{\kappa}_1 \tau_1,$$

where

$$1/b(\kappa) = 2\int_0^1 \exp(-\kappa t^2)\, dt, \quad \hat{\kappa}_1 = D^{-1}(\bar{\tau}_1), \quad \hat{\kappa}_0 = D^{-1}(\boldsymbol{\mu}^{\mathrm{T}} \mathbf{T} \boldsymbol{\mu}/n)$$

with $D(\cdot)$ defined by Mardia (1972, p. 253) and $\boldsymbol{\mu}$ given by (7.3). We find that $-2\log_e \lambda = 0{\cdot}95$. Since n is large, $-2\log_e \lambda$ can be assumed to be distributed as χ_2^2. Hence there is strong evidence that the perihelion points exhibit a preference for lying near the galactic pole. This is not surprising since the angle between the directions (7.2) and (7.3) is only $6{\cdot}1°$. Hence, there is strong evidence that the Lyttleton theory is true. It may be noted that the angle between the mean direction and the apex of the sun's motion is $16{\cdot}3°$. Fig. 2 shows a Schmidt net (Mardia, 1972, p. 215) for the first 432 of these comets with its centre as the galactic pole.

The projection is an equal area projection and therefore we would expect more points towards the edges of the net under a girdle distribution. The figure reflects this feature. (The net is given only for the upper hemisphere. For the lower hemisphere the net is very similar.)

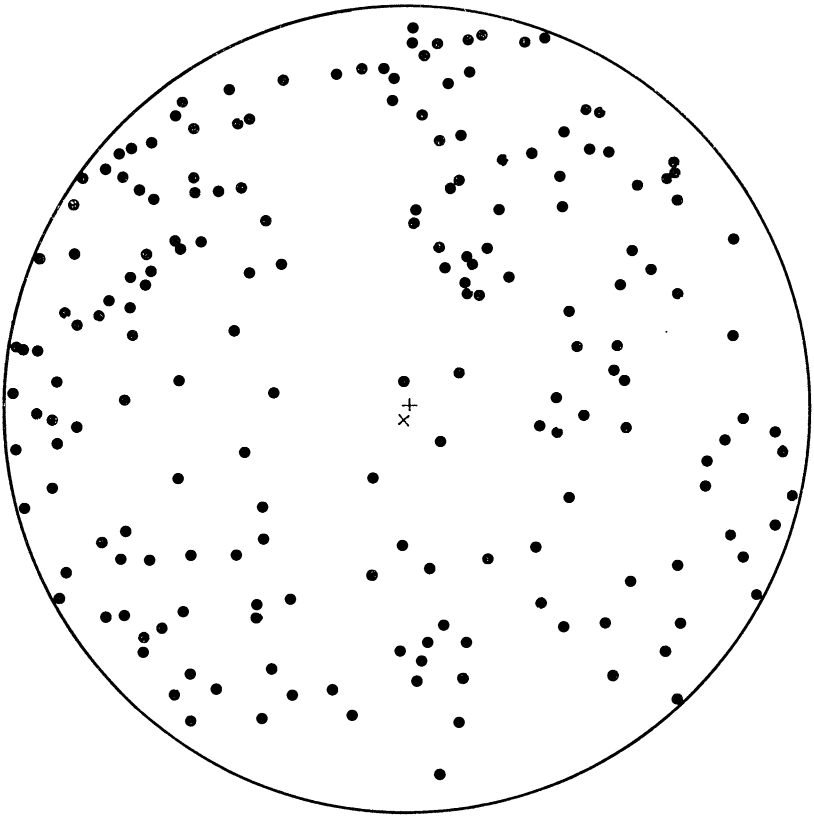

FIG. 2. Schmidt net for Tyror's data. +, Centre of the circle at the galactic pole; ×, estimated direction of the normal to the girdle plane.

7.3. *Cyclic Variation in Poisson Processes and Patient Arrival Data*

In some point processes, there are reasons to believe that there exists a cyclic effect, e.g. the arrival times of patients may have a time-of-day effect. Naturally, this leads to the question of testing a pure Poisson process against a non-homogeneous Poisson process with a cyclic trend. There exist various tests for the Poisson model (Cox and Lewis, 1966, pp. 152–164, 172–173). A simple test based on directional data analysis is as follows. Let 2π be the period suspected and suppose that we observe a pure Poisson process for a predetermined time t_0 which is an integral multiple of 2π. Further, let us suppose that n events are observed. We know (see, for example, Cox and Lewis, 1966, pp. 27–28) that the positions $U_1, \ldots, U_n$ of the events (measured as the distances from the origin) in a Poisson process are independently uniformly distributed over $(0, t_0)$ given that n events have occurred in all.

Since t_0 is an integral multiple of 2π, the random variables

$$\theta_i \equiv U_i \bmod 2\pi, \quad i = 1, \ldots, n \tag{7.4}$$

are independently uniformly distributed over the unit circle. Hence, conditionally on n, under the hypothesis of a pure Poisson process $\theta_1, \ldots, \theta_n$ is a random sample from the distribution with p.d.f.

$$f(\theta) = 1/(2\pi), \quad 0 < \theta \leqslant 2\pi.$$

Under the alternative hypothesis of a non-homogeneous Poisson process, we may assume that θ is distributed as $M(\mu_0, \kappa)$ which implies a cyclic effect. Hence, we are led to the Rayleigh test of uniformity when $\theta_1, \ldots, \theta_n$ are the observations (see Mardia, 1972b).

Example. Cox and Lewis (1966, pp. 254–255) give data regarding arrival times of patients at an intensive care unit. One expects arrivals to be influenced by a time-of-day effect. The number of arrivals is 254, and it is found that $\bar{R} = 0{\cdot}320$. As n is large we may use the chi-squared approximation. Since $2n\bar{R}^2 = 52{\cdot}02$, we reject the hypothesis of uniformity at the 0·1 per cent level. That is, the process is not pure Poisson. We can go one step further to enquire whether the von Mises distribution provides a good fit. It is found that $\hat{\mu}_0 = 96{\cdot}91°$ (= 1728 hours), $\hat{\kappa} = 0{\cdot}676$. The expected and observed frequencies are shown in Table 5. The value of the goodness

TABLE 5

Observed and expected frequencies when a von Mises distribution is fitted to the arrival times

Arrival times	*Observed frequency*	*Expected frequency*	*Arrival times*	*Observed frequency*	*Expected frequency*
$0 \leqslant t < 1$	5	7·9	12–	19	11·4
1–	9	6·7	13–	12	13·4
2–	11	5·9	14–	14	15·3
3–	6	5·3	15–	16	17·0
4–	4	4·9	16–	17	18·2
5–	1	4·8	17–	19	18·6
6–	1	4·9	18–	15	18·1
7–	7	5·3	19–	14	16·9
8–	3	5·9	20–	15	15·2
9–	2	6·8	21–	16	13·2
10–	12	8·0	22–	11	11·2
11–	13	9·5	23–	12	9·4

of fit statistic χ^2 is found to be 29·31. Now, the 10 per cent value of χ^2_{21} is 29·62, and hence we accept the null hypothesis.

7.4. *A Mixture of von Mises Distributions and some Turtle Data*

The bimodal distribution of von Mises type assumes that the two modes are of the same strength and are situated 180° apart. A more general bimodal distribution can be obtained by considering a mixture of two von Mises distributions. That is, the p.d.f. of θ is given by

$$f(\theta;\, \mu_{0,1}, \mu_{0,2}, \kappa_1, \kappa_2, \lambda) = \lambda g(\theta;\, \mu_{0,1}, \kappa_1) + (1-\lambda) g(\theta;\, \mu_{0,2}, \kappa_2), \tag{7.5}$$

where $g(\theta;\, \mu_0, \kappa)$ denotes the p.d.f. of $M(\mu_0, \kappa)$, $0 \leqslant \lambda \leqslant 1$ and $0 < \theta \leqslant 2\pi$. Sufficient conditions for (7.5) to be a bimodal distribution rather than unimodal are given in Mardia and Sutton (1975). The m.l.e. of the five parameters can be obtained numerically with the help of a program of Jones and James (1969). We use this method on some turtle data which have been previously investigated (Stephens, 1969; Boneva *et al.*, 1971, Mardia, 1972a, pp. 11, 129). Boneva *et al.* obtained a circular histospline; the raw data are given in Stephens (1969). Of course histospline analysis does not give a functional form of the density. For the grouped data used in Fig. 7 of Boneva *et al.*, it is found that

$$\lambda = 0{\cdot}85, \quad \mu_{0,1} = 65{\cdot}3°, \quad \mu_{0,2} = 239{\cdot}0°, \quad \kappa_1 = 1{\cdot}94, \quad \kappa_2 = 7{\cdot}76. \tag{7.6}$$

This uses a very wide cell length, but the estimators are comparable to the estimators for the raw data given by

$$\lambda = 0{\cdot}84, \quad \mu_{0,1} = 63{\cdot}5°, \quad \mu_{0,2} = 241{\cdot}2°, \quad \kappa_1 = 2{\cdot}62, \quad \kappa_2 = 8{\cdot}45.$$

Fig. 3 compares the fitted curve from the mixture with estimators given by (7.6) and the histospline. The two curves are similar and the modes are situated at the same

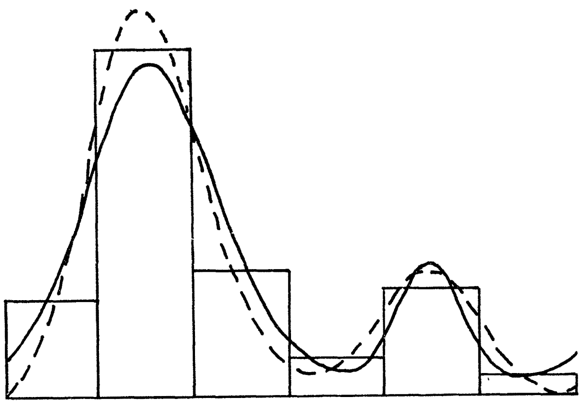

FIG. 3. Histospline (Boneva *et al.*, 1971) and a mixture fitted to turtle data. ----, Histospline; ——, mixture.

points. However, the strengths of the two modes given by the two methods differ slightly.

7.5. *Unknown Period*

In directional data analysis, it is generally assumed that the true period is known. However, there are situations where the true period may be unknown. Sometimes, particularly when mapping the real line onto the circle, it is not always clear which is the best period to use. We consider two problems.

(i) We may be looking, mistakenly, not at the true period, but at an integer multiple of it, for example, we may map the week or the year onto the circle when we really ought to be looking at the day. To ascertain if there is a sub-cycle, let us assume the parental population to have a density of the form

$$f(\theta) = c\exp\{\kappa\cos l(\theta-\mu_0)\}, \quad 0<\theta\leqslant 2\pi, \quad \kappa>0,$$

where l is an integer. The m.l.e. of l is l_0 where $R_{l_0}^2 = \max_l (R_l^2)$. The m.l.e.'s of κ and μ_0 are straightforward. The hypothesis of unimodality ($l = 1$) against the hypothesis of a sub-cycle ($l = l_0$) can be treated by using the likelihood ratio principle.

(ii) Suppose data emerge from a point process where there exists a cycle. Professor J. N. Mills, of Manchester University, provided us with a set of data concerning the times of potassium excretion of a subject living on a 21-hour period rather than the normal period of 24 hours. The main problem is described in Fort and Mills (1970); we perform here a *confirmatory analysis* to the times of potassium excretion only.

Let $t_1, \ldots, t_n$ be the times in hours. For the data, $n = 68$. Let

$$2\pi\theta_i \equiv t_i \bmod 21, \quad 2\pi\theta_i^* \equiv t_i \bmod 24, \quad i = 1, \ldots, n.$$

We wish to test

$$H_0\colon\ \theta_i, \quad i=1,\ldots,n \quad \text{from} \quad M(\mu_0,\kappa), \quad \kappa>0,$$

against

$$H_1\colon\ \theta_i^*, \quad i=1,\ldots,n \quad \text{from} \quad M(\mu_0',\kappa'), \quad \kappa'>0.$$

We can write the likelihood ratio but its distribution is unknown. However, the circular statistics for varying periods should prove helpful. A plot of

$$R_\lambda^* = (n\pi)^{-1}\left[\left\{\sum_{i=1}^n \cos(2\pi t_i/\lambda)\right\}^2 + \left\{\sum_{i=1}^n \sin(2\pi t_i/\lambda)\right\}^2\right] = (n\pi)^{-1}\bar{R}_\lambda^2$$

against the period λ is given in Fig. 4. As expected R_λ^* has a maximum around $\lambda = 21$ hours, but it also has a minimum around $\lambda = 24$ hours. For $\lambda = 24$ hours, $\bar{R} = 0{\cdot}038$ so we accept the hypothesis of uniformity at the 10 per cent level, whereas for $\lambda = 21$ hours $\bar{R} = 0{\cdot}348$, so we reject this hypothesis at the 1 per cent level. The

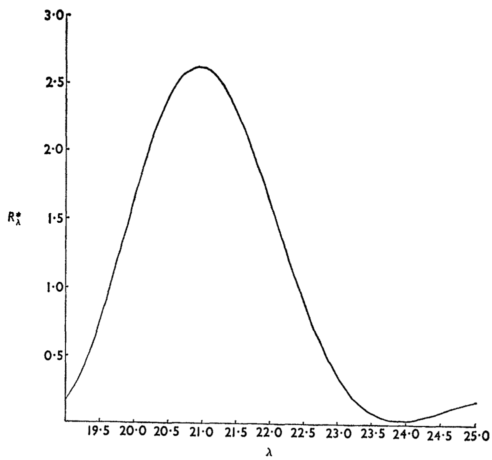

FIG. 4. R_λ^* for the excretion data.

expected and observed frequencies under H_0 were obtained. For a particular grouping, the value of the goodness-of-fit criterion χ^2 was found to be 8·87, distributed as χ_6^2. The 5 per cent value of χ_6^2 is 12·59.

8. CONCLUDING REMARKS

In this final section, we add some miscellaneous comments.

(i) There might exist a limiting process analogous to the central limit theorem leading to the von Mises–Fisher distribution. For the circular case, the Brownian motion distribution appears as a limiting distribution of the mean direction for the circular case (Rukhin, 1971). Hartman and Watson (1974) show that there exists a stopping time distribution for Brownian motion on S_p leading to the von Mises–Fisher distribution. Lewis (1975) has made a significant step forward by showing that von Mises is infinitely divisible for small κ.

(ii) There are various unsolved fundamental estimation problems For example, what does one mean by an "unbiased" estimator of an angular parameter such as the mean direction? When is an estimator best? Some attempts have been made by Mardia (1972, Chapter 5) and Rukhin (1971). The latter reference looks into these problems for a translation family. A Bayesian approach is discussed in Mardia and El-Atoum (1974).

(iii) The definition of circular rank correlation coefficient r_0 given by (6.4) can be modified to the range $(-1, 1)$ so that the positive value of r_0 corresponds to "positive" dependence, etc. However, such a distinction is artificial since we can have $\bar{R}_1^2 = \bar{R}_2^2$. For $n = 6$, and $r_1 = 4$, $r_2 = 1$, $r_3 = 6$, $r_4 = 5$, $r_5 = 2$, $r_6 = 3$, we find $\bar{R}_1^2 = \bar{R}_2^2 = 0{\cdot}25$. Here it is difficult to say whether there is positive or negative dependence.

If the true directions for θ and ϕ are known, say zero, a more appropriate coefficient would be

$$\max\left[\sum \cos\{2\pi(i-r_i)/n\},\ \sum \cos\{2\pi(i+r_i)/n\}\right].$$

Its critical values were obtained simultaneously with the work for r_0 but such examples are not so common. Some other non-parametric circular correlation coefficients are given by Rothman (1971).

On following the argument of Section 6.2, we can define in general a circular correlation coefficient for (θ_i^*, ϕ_i^*), $i = 1, \ldots, n$, as

$$\max(D_+, D_-)/\{(1-\bar{R}_1^2)(1-\bar{R}_2^2)\},$$

where $\bar{R}_1$ and $\bar{R}_2$ are the mean resultant lengths of θ_i^* and ϕ_i^* and

$$n^2 D_\pm = \left\{\sum_{i=1}^{n} \cos(\theta_i^* \pm \phi_i^*) - n\bar{R}_1 \bar{R}_2\right\}^2 + \left\{\sum_{i=1}^{n} \sin(\theta_i^* \pm \phi_i^*)\right\}^2.$$

The mean directions for θ_i^* and ϕ_i^* are selected to be zero. A detailed investigation of this coefficient and related functions will appear elsewhere.

(iv) As in multivariate analysis, it is not surprising that there is a dearth of non-parametric tests of practical value for $p \geqslant 3$. In constructions of circular non-parametric tests, the circular uniform scores of Section 6 have been playing a key role (see Wheeler and Watson, 1964; Mardia, 1967, 1969, 1972b; Mardia and Spurr, 1973).

(v) There have been various other notable developments in the area within the past few years. Downs (1972) deals with orientations characterized by k directions (instead of a single direction) in p-dimensions. Matthews (1974) raises various important statistical problems in bird navigation and Kendall (1974a) provides a satisfactory solution to one of these through pole-seeking Brownian motion. Some practical problems of discriminant analysis are treated in Morris and Laycock (1974). Upton (1974) has given some new approximations.

Since the paper was submitted in 1973, it has been brought to the author's notice that Section 7.3 is related to Lewis P.A.W. (1972) and Section 7.5 to Kendall (1974b).

ACKNOWLEDGEMENTS

I wish to express my thanks to Professor T. Lewis for his valuable help. I have benefited from discussions with various friends including Michael Bingham, Jim Thompson, Peter Zemroch, Harold Peers, Dick Gadsden, Barrie Spurr, Terry Sutton and Shafiq El-Atoum, to whom I am most grateful. My thanks are due to Mr H. Goldstein, Professor R. A. Lyttleton and Professor J. N. Mills for providing interesting case studies, and for their comments. The comet problem was brought to my attention by Professor D. G. Kendall. I am grateful to a referee for his helpful suggestions, and to Professor T. Downs for giving permission to use his data. I am also grateful to Mr P. J. Zemroch for his computing assistance.

REFERENCES

ANDERSON, T. W. and STEPHENS, M. A. (1972). Tests for randomness of directions against equatorial and bimodal alternatives. *Biometrika*, **59**, 613–621.

BINGHAM, C. (1964). *Distributions on the sphere and on the projective plane*. Ph.D. Thesis, Yale University.

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

BINGHAM, M. S. and MARDIA, K. V. (1975). Maximum likelihood characterization of the von Mises distribution. In *Statistical Distributions in Scientific Work* (G. P. Patil, S. Kotz and J. K. Ord, eds), Vol. 3. Boston: D. Reidel.

BONEVA, L. I., KENDALL, D. G. and STEFANOV, I. (1971). Spline transformations: three new diagnostic aids for the statistical data-analyst. *J. R. Statist. Soc.* B, **33**, 1–70.

COX, D. R. and LEWIS, P. A. W. (1966). *Statistical Analysis of Series of Events*. London: Methuen.

CROUCH, P., WEISS, R. and GOLDSTEIN, H. (1975). Orientation of cancer cells on gridlines. (Unpublished.)

DOWNS, T. D. (1972). Orientation statistics. *Biometrika*, **59**, 665–675.

—— (1974). Rotational angular correlations. In *Biorhythms and Human Reproduction* (M. Ferin *et al.*, eds), pp. 97–104. New York: Wiley.

FISHER, R. A. (1953). Dispersion on a sphere. *Proc. Roy. Soc. Lond.* A, **217**, 295–305.

FORT, ANN and MILLS, J. N. (1970). Fitting sine curves to 24 h urinary data. *Nature*, **226**, 657–658.

HAJEK, J. and SIDAK, Z. (1967). *Theory of Rank Tests*. New York: Academic Press.

HARTMANN, P. and WATSON, G. S. (1974). "Normal" distributions functions on spheres and the modified Bessel functions. *Ann. Prob.*, **2**, 593–607.

JAMES, A. T. (1964). Distribution of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.*, **35**, 475–501.

JONES, T. A. and JAMES, W. R. (1969). Analysis of bimodal orientation data. *Math. Geol.*, **1**, 129–135.

KAGAN, A. M., LINNIK, YU. V. and RAO, C. R. (1973). *Characterization Problems in Mathematical Statistics*. New York: Wiley.

KENDALL, D. G. (1974a). Pole seeking Brownian motion and bird navigation. *J. R. Statist. Soc.* B, **36**, 365–417.

—— (1974b). Hunting quanta. *Phil. Trans. R. Soc. Lond.*, A, **276**, 231–266.

KLUYVER, J. C. (1906). A local probability theorem. *Ned. Akad. Wet. Proc.* A, **8**, 341–350.

KSHIRSAGAR, A. M. (1972). *Multivariate Analysis*. New York: Marcel Dekker.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. New York: Wiley.

LEWIS, P. A. W. (1972). Recent results in the statistical analysis of univariate point processes. In *Stochastic Point Processes*, pp. 1–54, (P.A.W. Lewis, ed.), New York: Wiley

LEWIS, T. (1975). Probability functions which are proportional to characteristic functions and the infinite divisibility of the von Mises distribution. In *Perspectives in Probability and Statistics*, the M. S. Bartlett Volume (J. Gani, ed.), Academic Press, London.
LYTTLETON, R. A. (1953). *The Comets and their Origin.* Cambridge University Press.
—— (1961). On the statistical loss of long-period comets from the solar system, I. *Proc. 4th Berkeley Symp. Math. Statist. and Prob.*, **3**, 229–244.
MARDIA, K. V. (1967). A non-parametric test for the bivariate two sample location problem. *J. R. Statist. Soc.* B, **29**, 320–342.
—— (1969). On Wheeler and Watson's test on a circle. *Sankhyā*, A, **31**, 177–190.
—— (1972a). *Statistics of Directional Data.* London: Academic Press.
—— (1972b). A multi-sample uniform scores test on a circle and its parametric competitor. *J. R. Statist. Soc.* B, **34**, 102–113.
—— (1975a). Characterizations of directional distributions. In *Statistical Distributions for Scientific Work* (G. P. Patil, S. Kotz and J. K. Ord, eds), Vol. 3. Boston: D. Reidel.
—— (1975b). Distribution theory for the von Mises–Fisher distribution. In *Statistical Distributions for Scientific Work* (G. P. Patil, S. Kotz and J. K. Ord, eds), Vol. 1. Boston: D. Reidel.
MARDIA, K. V. and EL-ATOUM, S. A. M. (1974). Bayesian inference for the von Mises–Fisher distribution. (Submitted to *Biometrika.*)
MARDIA, K. V. and SPURR, B. D. (1973). Multi-sample tests for multi-modal and axial data. *J. R. Statist. Soc.* B, **35**, 422–436.
MARDIA, K. V. and SUTTON, T. (1975). A mixture of two von Mises distributions. To appear in *Biometrika.*
MARDIA, K. V. and ZEMROCH, P. J. (1975a). Algorithm AS 81. Circular statistics. *Appl. Statist.*, **24**, 147–150.
—— (1975b). Algorithm AS 80. Spherical statistics. *Appl. Statist.* **24**, 144–146.
MATTHEWS, G. V. T. (1974). On bird navigation, with some statistical undertones. *J. R. Statist. Soc.* B, **36**, 349–364.
MORRIS, J. E. and LAYCOCK, P. J. (1974). Discriminant analysis of directional data. *Biometrika*, **61**, 335–342.
PEARSON, E. S. and HARTLEY, H. O. (1972). *Biometrika Tables for Statisticians*, Vol. 2. Cambridge University Press.
RAO, J. S. (1969). Some contributions to the analysis of circular data. Ph.D. Thesis, Indian Statistical Institute, Calcutta.
ROTHMAN, E. D. (1971). Tests of coordinate independence for a bivariate sample on a torus. *Ann. Math. Statist.*, **42**, 1962–1969.
RUKHIN, A. L. (1971). Some statistical decisions about distributions on a circle for large samples. *Sankhyā*, A, **34**, 243–250.
STEPHENS, M. A. (1969). Techniques for directional data. *Tech. Report No.* 150. Dept. of Statist., Stanford University.
TYROR, J. G. (1957). The distribution of the directions of perihelion of long-period comets. *Mon. Not. R. Astr. Soc.*, **117**, 369–379.
UPTON, G. J. G. (1974). New approximations to the distribution of certain angular statistics. *Biometrika*, **61**, 369–374.
WATSON, G. N. (1948). *A Treatise on the Theory of Bessel Functions.* 2nd ed. Cambridge Univ. Press.
WATSON, G. S. and WILLIAMS, E. J. (1956). On the construction of significance tests on the circle and the sphere. *Biometrika*, **43**, 344–352.
WHEELER, S. and WATSON, G. S. (1964). A distribution-free two-sample test on a circle. *Biometrika*, **51**, 256–257.

DISCUSSION OF PROFESSOR MARDIA'S PAPER

Dr P. J. LAYCOCK (University of Manchester Institute of Science and Technology): I am delighted to propose the vote of thanks to Professor Mardia for the scholarly, yet practical, paper he has presented this evening. The paper represents an important step in the establishment of directional data analysis as a field of theoretical and practical importance in statistics.

There is little of a controversial nature in the paper, so I find myself somewhat at a loss for the traditional collection of "pointed" remarks with which to assail the speaker—despite the title of his paper!

I am particularly interested in Professor Mardia's presentation of results in the multi-dimensional and multivariable situation, as (surprisingly ?) there is much of relevance here for some data, on human skulls, which I have recently analysed.

When viewed as a standardized X-ray projection, defined anatomical landmarks within the head can be identified and compared for variations in location. The practice within the clinical subject of orthodontics has been to connect pairs of defined anatomical points to form lines (Broadbent, 1931). The angles between pairs of lines have been measured to identify certain variations in form. The problem has been that as all known points have been variable in location (Dushra, 1948), it has been necessary to define the least variable pair, to act as a baseline from which to measure the more variable landmarks. So we have a situation where the pole is itself just another (variable) direction! See Fig. 1.

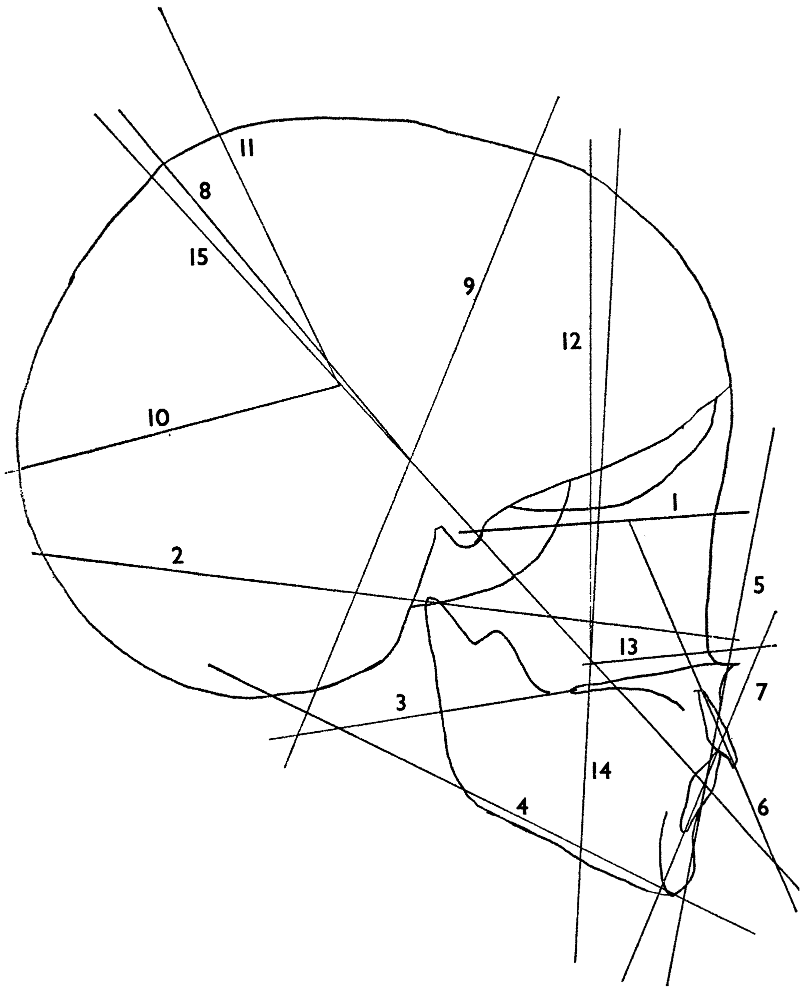

FIG. 1. Fifteen orthodontic planes. 1. Sella-nasion. 2. Frankfurt. 3. Maxillary. 4. Mandibular. 5. Dental base (point A–point B). 6. Upper incisor axis. 7. Lower incisor axis. 8. Skull (greater axis). 9. Skull (lesser axis). 10. Cranium (greater axis). 11. Cranium (lesser axis). 12. Face (greater axis). 13. Face (lesser axis). 14. Profile. 15. Reference.

Let $\theta_{\nu j}$ be the anticlockwise angle from direction ν (as pole) to the jth of q directions. Then (2.11) and (2.12) suggest as the joint distribution of $(\theta_{\nu i}, \ldots, \theta_{\nu q}) = \boldsymbol{\theta}_\nu$,

$$f(\boldsymbol{\theta}_\nu) = C \exp \Big\{ \sum_j \kappa_j \cos(\theta_{\nu j} - \mu_{\nu j})$$

$$+ \sum (a_{jk} \cos \theta_{\nu j} \cos \theta_{\nu k} + b_{jk} \sin \theta_{\nu j} \cos \theta_{\nu k} + c_{jk} \cos \theta_{\nu j} \sin \theta_{\nu k} + d_{jk} \sin \theta_{\nu j} \sin \theta_{\nu k}) \Big\}.$$

Where the second summation is over $j<k$, or more generally $j \leqslant k$ in an extension of (2.11). This distribution will of course be singular, since $\theta_{\nu\nu} \equiv 0$. So we are led to consider the distribution of $\boldsymbol{\theta}_\nu^* = (\theta_{\nu 1}, \theta_{\nu 2}, \ldots, \theta_{\nu,\nu-1}, \ldots, \theta_{\nu q})$. Now $\theta_{\alpha j} = \theta_{\nu j} - \theta_{\nu\alpha} \pmod{2\pi}$, and hence $\boldsymbol{\theta}_\alpha^* = \mathbf{H}_{\alpha\nu} \boldsymbol{\theta}_\nu^*$ where $\mathbf{H}_{\alpha\nu}$ is the $(q-1)\times(q-1)$ matrix whose columns are $(0, 0, \ldots, 0, 1, 0, \ldots, 0)^{\mathrm{T}}$, 1 in jth position, $1 \leqslant j < \alpha$,

$$(-1, -1, \ldots, -1)^{\mathrm{T}}, \quad j = \alpha,$$

$$(0, 0, \ldots, 0, 1, 0, \ldots, 0)^{\mathrm{T}}, 1 \quad \text{in } (j-1)\text{th position}, \quad \alpha < j \leqslant q-1, \quad j \neq \nu,$$

$$(0, 0, \ldots, 0)^{\mathrm{T}}, \quad j = \nu,$$

for $\alpha < \nu$, with obvious alterations to this for $\alpha > \nu$. Note that $\mathbf{H}_{\alpha\nu} = \mathbf{H}_{\nu\alpha}^{-1}$ and $|\mathbf{H}_{\alpha\nu}| = \pm 1$. We clearly require f to be invariant under this transformation, and this implies $a_{jk} = d_{jk}$ and $b_{jk} = -c_{jk}$, giving

$$f(\boldsymbol{\theta}_\nu^*) = C \exp \Big\{ \sum_{j \neq \nu} \kappa_j \cos(\theta_{\nu j} - \mu_{\nu j}) + \sum_{\substack{j<k \\ j,k \neq \nu}} \gamma_{jk} \cos(\theta_{\nu j} - \theta_{\nu k} - \mu_{\nu j} - \mu_{\nu k}) \Big\}$$

which I shall call the "exchangeable Pole" distribution.

If $\gamma_{jk} = 0$ then $\theta_{\nu j}$ and $\theta_{\nu k}$ are (conditionally, at least) independent, whilst if $\gamma_{jk} = \pm\infty$ we have $\theta_{\nu j} = \theta_{\nu k}$, suggesting $\rho_{jk} = 1 - \exp(-\gamma^2)$ as a possible, pole-independent, measure of the "intrinsic" correlation between the directions j and k, with κ_j^{-1} as a measure of the "intrinsic stability" of direction j. I have not yet had time to estimate these parameters and coefficients for any of the skull's data, nor to compare this analysis with some previous work of mine published (as yet) only in a technical report. It seems worth mentioning that a linear model examined in this report was indexed by a direction in $p = 15$ dimensions. Thus calling for a p-dimensional distribution in a Bayesian analysis, for instance.

The circular rank correlation coefficient in (6.4) seems unsatisfactory as it stands, since the arbitrary parameters l and m of (6.1) are assumed known. Professor Mardia appears to have taken $l = m = 1$ in Example 1, though perhaps other values were used. But, as Professor Mardia's paper makes clear, the proper definition of correlation for directional data is still very much an open subject. Confusing results can easily arise. For instance, referring back to the skull's data, let us alternatively define

$$\rho_{jk} = |E\{\exp(\sqrt{(-1)}\, \theta_{jk})\}|.$$

Note first that $1 - \rho_{jk}$ would be what is commonly called the "circular variance" of θ_{jk}. But $0 < \rho_{jk} < 1$, $\rho_{jk} = \rho_{kj}$, $\rho_{jk} = 1$ when directions j and k are fixed in relation to each other, and finally $\rho_{jk} = 0$ when direction j is uniformly distributed with respect to direction k as Pole (or vice versa).

The use of ancillary statistics and conditional arguments is a recurrent theme in the test situation for directional statistics. Although (4.3) provides a "*UMP* invariant test", this is conditional on R_{x_1}, and for circular data this is arguably equivalent to discarding half the available information in the data. The Bayesian approach would seem to be strengthened by default in such a situation, though the fact that Watson and Williams also recommended this particular test, but on *intuitive* grounds, might balance this out.

The case studies at the end of the paper were fascinating, though the exact details of the data collection (and the reasons for doubling the angles) for the cancer cells study still seem obscure. But no doubt that is just me. The opportunities for rigorous statistical analysis of astronomical data are wider than statisticians may realize. The fields of spectral analysis and cluster analysis (see Abell, 1965) spring to mind. Hence it is encouraging to read of Professor Mardia's successful foray into the analysis of comet data.

Finally, I note that the author's "concluding remarks" suggest that we can look forward to many further developments in this field in the next few years.

It gives me pleasure to propose the vote of thanks.

Professor H. E. DANIELS (University of Birmingham): Professor Mardia did us all a service by publishing the right book at the right time, so bringing an important but hitherto specialized field to the attention of all statisticians. It is a young and growing field, not yet too specialized for amateurs to venture into, and it is therefore an admirable topic for discussion at one of our meetings. Today's paper can perhaps be described as an appendix to the book with an emphasis on applications. As such it is to be warmly welcomed.

I have not had a great deal of experience in analysing directional data, so I am glad that Dr Laycock has concentrated on the applications aspect of the paper. I propose to confine my remarks to certain theoretical points which occurred to me as I read the paper.

The first is this matter of characterizations, one of which—by maximum entropy—is discussed in Section 2.3. Characterizations are all very interesting but they provoke in me the question "So what?" Another characterization of the von Mises–Fisher distribution—by maximum likelihood—(mentioned in Professor Mardia's book) is useful because it means that the intuitively reasonable parameter estimator is also the most efficient. But it is not obvious at first sight why having maximum entropy should endear a distribution to us.

One justification could be that it is really a limiting property, though of a different kind from that of the central limit theorem. Kuhn and Grün showed by a familiar argument of statistical mechanics that if a long chain of freely jointed identical links is constrained to have its ends at a fixed distance apart, the most probable distribution of link angles has approximately the von Mises–Fisher form. (It is the appropriate Gibbs distribution.) Maximimizing the log probability of all the link angles of a single long chain, subject to the constraint, approximates to maximizing the entropy in Section 2.3. So there is a physical reason for expecting the von Mises–Fisher form to appear as a limiting distribution in certain situations. But there seems no absolute reason why maximum entropy should be desirable in all circumstances.

My second point does not directly refer to the material of the paper, but is related to it. In a paper read to the Society last year (Kendall, 1974), Professor D. G. Kendall raised the question of "unwrapping" the von Mises distribution. The von Mises distribution on a circle is remarkably similar to the distribution produced by wrapping a normal distribution round the circle. Can we find a distribution which when wrapped round the circle produces the von Mises distribution? Stated in this way the problem is not well defined: because of the ambiguity due to aliasing there are many such distributions. We can make it precise by asking for a distribution generated by a diffusion process, like the one generating the wrapped normal distributions.

An isotropic two-dimensional diffusion, starting from a point on a circle, will generate a von Mises distribution on the circle as time goes on. As Kendall suggested, this can be unwrapped by separating the paths reaching a given point on the circle according to the number of times they have wound round the centre of the circle. A comparable though rather artificial process for the wrapped normal would be a similar diffusion with the "step length" of the paths proportional to the distance from the centre of the circle. This would act like angular diffusion round the circle.

Analytically the unwrapping of the von Mises distribution with density $\exp(\kappa \cos \theta)/I_0(\kappa)$ is achieved by taking its circular characteristic function $I_n(\kappa)/I_0(\kappa)$ and inverting it as a cosine transform with respect to n treated as a continuous variable, thus:

$$g(\theta, \kappa) = \frac{1}{2\pi}\int_0^\infty \frac{I_n(\kappa)}{I_0(\kappa)} \cos n\theta \, dn, \quad -\infty < \theta < \infty.$$

By manipulating contour integrals it can be expressed in the form

$$g(\theta, \kappa) = h(\theta, \kappa)/I_0(\kappa),$$

where

$$h(\theta, \kappa) = \frac{1}{2\pi}\int_a^{\infty} \exp(-\kappa \cosh u) \left\{\frac{\theta-\pi}{u^2+(\theta-\pi)^2}+\frac{\theta+\pi}{u^2+(\theta+\pi)^2}\right\} du$$

$$+\frac{1}{4\pi}\int_{-\pi}^{\pi} \exp(\kappa \cosh a \cos v)$$

$$\times\left\{\frac{a \cos(\sinh a \sin v)+(v-\theta)\sin(\sinh a \sin v)}{a^2+(v-\theta)^2}\right\} dv \quad (a>0).$$

This horrible-looking formula is quite straightforward to compute, though to avoid discontinuities it is best not to take a too near zero. At first glance $g(\theta, \kappa)$ turns out to look rather like a normal density, but closer inspection shows that it has Cauchy-type tails dying away like θ^{-2}.

Can the same sort of unwrapping be done on a sphere? Professor Kendall remarked that as far as he was aware Brownian motion on a sphere was not the result of wrapping anything. In a sense this is true because the centre of a sphere cannot separate the paths in the way that the centre of a circle does. But if the diffusion source is put, not at the obvious place which is the pole $\theta = 0$, but at a point on the equator $\theta = \frac{1}{2}\pi$, the azimuthal angle ϕ becomes the important angle, not θ. We can then regard the paths as winding round an axis through the poles. It will be interesting, though doubtless of no practical importance, to unwrap both the Brownian motion on the sphere and the von Mises–Fisher distribution in this way.

I must conclude by congratulating Professor Mardia on his stimulating paper. I have great pleasure in seconding the vote of thanks.

The vote of thanks was passed by acclamation.

Professor T. Lewis (University of Hull): Just a couple of points. It was kind of Professor Mardia to mention my note showing the infinite divisibility of the von Mises distribution for small values of κ. Since writing it, I have proved that the range of values of κ for which this holds certainly extends up to 0·161. I liked 0·161; but it is not destined to become one of those universal constants such as π, because I heard last week from Mr John Kent, of Cambridge University, that he has proved that the von Mises distribution is infinitely divisible for *all* values of κ. He is here today and, I think, intending to speak in the discussion, and I have probably pinched some of his best lines. But it does give me the opportunity to congratulate him on this important advance.

From the infinite divisibility of the von Mises, all kinds of pleasing inequalities for Bessel functions follow. Here is one:

$$I_m(\kappa)\, I_n(\kappa) < I_0(\kappa) \sqrt{\{I_{m+n}(\kappa)\, I_{m-n}(\kappa)\}}$$

for all integer m, n and, since last week, for all κ. Actually, following Mr Kent's result the inequality may be extended to all real values of the order parameters, in the form

$$I_{|\mu|}(\kappa)\, I_{|\nu|}(\kappa) < I_0(\kappa) \sqrt{\{I_{|\mu+\nu|}(\kappa)\, I_{|\mu-\nu|}(\kappa)\}}.$$

I have not yet thanked Professor Mardia for his most enjoyable paper, full of interesting things, not least—as Dr Laycock and Professor Daniels have already said—his analyses of all these splendid sets of data: comets, medical students, arrangement of the toes, turtles and these people living their lives on a 21-hour cycle, of whom one would like to learn more. Where does Professor Mardia obtain all these exotic data? (though, as he said in his paper, it is Professor Stephens who must be credited with bringing the turtles into the statistical world). Now, I am prepared to bet that nobody who is such an impresario of data as is Professor Mardia can go for long without encountering some suspect or contaminating observations, perhaps misrecordings, perhaps values which have come from another distribution than that of the main sample—the sort of observations which, with ordinary samples on the line, show up as outliers. On the line, the more extreme the value x of the outlier, the greater its separation from the main data mass. But on the circle, supposing a distribution effectively concentrated on part of the circumference, the outlier—as its value θ goes on increasing—sneaks back into the main data mass from the other side and passes itself off as a respectable reading.

I imagine this is particular to the one-dimensional case. For samples from angular distributions in higher dimensions, such as the von Mises–Fisher, presumably an outlier would easily show up because there would be room (e.g. on Dr Laycock's 15-dimensional hypersphere) for it to outlie. On the circle, however, this is not the case, so there is a real problem of data analysis; it calls for some study of spacings and order statistics—e.g. the analogue of Fisher's normal order scores for such distributions as von Mises. Mr David Collett, of the Mathematical Statistics department at Hull, is doing some work on these problems. But underlying all this there are important general issues, and I should dearly like to know what Professor Mardia's views are about them: is there an outlier problem on the circle?; if so, when does it arise, in what circumstances do outliers show up, how can they be identified as discordant and what should be done about them?

Dr G. J. G. UPTON (University of Essex): Professor Mardia continues to supply us with a wealth of interesting data sets, analyses and unsolved problems. My first point concerns the first example of Section 6. This problem is only circular in so far as the data spans a range either side of 0°. It is simple to add 180° to all the observations and then to use the Spearman ranking procedure. The differences in the ranks of θ and ϕ are $0, 0, 1, -1, 2, -1, -1, 0, 0, 0$ and these lead to a value for r_s of $157/165 = 0{\cdot}95$, which far exceeds the upper 1 per cent point. Professor Mardia's test really comes into its own when it is impossible to free the data from its circular nature.

My second comment concerns the unsolved problem (i) of Section 8. I believe that the key to the limiting process lies in the distribution of $\bar{x}_0$, the direction of the resultant of a set of observations. Professor Mardia has pointed out that the conditional distribution of $\bar{x}_0$, given R, is the von Mises distribution, but the unconditional distribution is unfortunately fearsome. If we let the sample size, n, get large and simultaneously let κ get small, then we get back to the usual Normal distribution. However, some form of simultaneous limiting process might lead to a natural circular distribution intermediate between the von Mises and Normal distributions. It is possible to see some sort of physical backcloth for this distribution: we can imagine that any observed vector arises as the resultant of a large number of "error vectors" each having this distribution. In practical terms this limiting distribution seems certain to be virtually indistinguishable from either the von Mises or the wrapped Normal distributions.

Professor J. F. C. KINGMAN (University of Oxford): A parametric analysis of data whose values are necessarily confined to some particular space S, such as the surface of a sphere, requires a plausible family of distributions, at once rich enough to fit the data and simple enough to permit manipulation. Professor Mardia has shown that the von Mises–Fisher distribution and its relatives lay strong claim to this status in the spherical case, but I am uneasy about a curious circularity in the structure of his argument.

In the general setting of Section 2.3, we are given a space S and a particular "uniform" measure μ. We then select functions $t_1, t_2, \ldots, t_q$ defined on S, and ask for distributions with given values of $E\{t_i(s)\}$. The condition of maximum entropy relative to μ leads to an exponential family whose densities with respect to μ are given by (2.10), where $b_1, b_2, \ldots, b_q$

play the role of parameters. For a sample $(s_1, s_2, \ldots, s_n)$, the statistics

$$T_i = \sum_{r=1}^{n} t_i(s_r) \quad (i = 1, 2, \ldots, q)$$

are sufficient for $(b_1, b_2, \ldots, b_q)$, and have probability density

$$f(T_1, T_2, \ldots, T_q) = \exp\left(nb_0 + \sum_{i=1}^{q} b_i T_i\right) f_0(T_1, T_2, \ldots, T_q),$$

where f_0 is the density derived from μ.

Thus the initial choice of the t_i forces us to summarize the data by the sufficient statistics T_i which are the sample sums of the values of t_i. Surely this builds too much on the initial choice of the t_i (and note the summary dismissal of one such choice at the end of Section 2.3). Perhaps therefore Professor Mardia might like to comment on three questions:

(i) Is there any evidence that the von Mises–Fisher or Bingham distributions actually fit typical data in the various applications he has in mind?
(ii) How robust are the procedures to departures from these distributions?
(iii) Is there any statistical meaning to the maximum entropy condition, or is it just a device for lending a spurious respectability to particular exponential families?

Mr JOHN KENT (Cambridge University): I would like to thank Professor Mardia for a very interesting paper. I have two limiting results to present for the von Mises–Fisher distribution which give some probabilistic justification for its use.

First, the von Mises distribution can be derived as the equilibrium distribution of a diffusion process on the circle. This diffusion process, which I shall call a von Mises process, is the circular analogue of the Ornstein–Uhlenbeck (O.U.) process on the line. The O.U. process is defined as follows. If we are at a point x on the line at time t we imagine in the next instant of time dt a diffusing step of size $\pm\sigma(dt)^{\frac{1}{2}}$ and a drift step of size $-\lambda x\, dt$. That is, the further away from the origin the particle is, the more it is pulled back towards the origin. As $t \to \infty$ the diffusing force spreading out the particle balances the drift pulling the particle back to 0. This process asymptotically has a normal distribution with mean 0 and variance $\sigma^2/2\lambda$.

Now we construct a similar process on the circle. We suppose there is a constant diffusing force, as on the line. We would like the drift to pull back towards $\theta = 0$ and to get stronger as we move away from $\theta = 0$. Also we would like continuity in the drift at $\theta = \pi$. With these somewhat contradictory requirements, a plausible choice for the drift is $-\lambda \sin\theta$. The forwards equation for the probability density $p = p(\theta, t)$ of this process is

$$\frac{\sigma^2}{2}\frac{\partial^2 p}{\partial\theta^2} + \lambda\frac{\partial}{\partial\theta}(\sin\theta\, p) = \frac{\partial p}{\partial t}.$$

If the process is in equilibrium then $\partial p/\partial t = 0$. Solving this equation with the appropriate boundary conditions, we find that $p = \exp(\kappa\cos\theta)$ where $\kappa = 2\lambda/\sigma^2$ is the solution. Hence the von Mises distribution is the equilibrium distribution of the von Mises process. Therefore, in this sense, the von Mises distribution is the circular equivalent of the normal distribution on the line.

We can construct an analogous diffusion process on S_n, $n > 1$. Here again, we suppose constant infinitesimal variance and infinitesimal drift proportional to $-\sin\theta_1$ where θ_1 is now the co-latitude in polar co-ordinates (see equation (2.5) in Mardia's paper). If we go through the same reasoning as above we find that the von Mises–Fisher distribution on S_n is the equilibrium distribution.

My second result is an extension of a theorem of T. Lewis. In a paper to appear, I have recently proved that the von Mises distribution is infinitely divisible for all values of $\kappa > 0$. If $X \sim M(0, \kappa)$ then for any m, there exist m independent identically distributed random variables on the circle $\{X_i, i = 1, \ldots, m\}$ such that $X = X_1 + \ldots + X_m$. This result also applies to the unwrapped von Mises distribution mentioned by Professor Daniels. Furthermore, infinite divisibility extends to higher dimensions if we are careful about defining what we mean by the sum of two points on S_n, $n>1$.

Professor M. L. PURI (Universities of Göttingen and Indiana): First, I should like to join my distinguished colleagues in expressing my sincere thanks to Professor Mardia for having provided such a beautiful exposition this evening on some aspects of statistics dealing with directional data problems. Being one of the pioneers in this field, Professor Mardia has conveyed the importance of this area in a most impressive and convincing manner. The applications were extremely instructive to me.

Turning to the specific aspects of the paper, my comments are in the form of inquiries.

My first question relates to the study of robustness of the procedures discussed by Professor Mardia. This question has already been raised by Professor Kingman and needs no further elaboration. A related question concerns the large sample properties of these procedures. Some attempt in this direction has been made by my colleague J. S. Rao (1971) who studied the Bahadur and Pitman efficiencies of some of the procedures for testing uniformity on the circle. Much more needs to be done in this direction.

My second question concerns the non-parametric tests for independence. Professor Mardia has given a procedure based on uniform scores. However, it seems to me that the tests for independence can be described in a relatively general framework by formulating a class of association parameters which are regular functionals of the underlying distribution function. This will provide us with suitable non-parametric competitors of the classical covariance matrix. In the case of linear rank order statistics, the theory has been investigated in full detail in Puri and Sen (1971). With suitable modifications, it can be worked out in the circular case also. This will increase the scope of applications.

Thirdly, Professor Mardia has described the tests of location. What about the related problem of testing the concentration or scale and testing simultaneously the location and scale? It seems that no work has been done in this direction.

My final question concerns the problem of the rates of convergence of the test statistics distributions. This problem is of great practical as well as theoretical interest. In the case of linear rank order statistics, some attempts have been made recently by Jurečková and Puri (1975), and Bergström and Puri (1975), but apparently no progress has been made on this topic in the subject dealt with by Professor Mardia this evening.

Professor EDWARD BATSCHELET (Mathematisches Institut, Universität, Zürich): Professor Mardia's paper contains a remarkable definition of a correlation coefficient, r_0, for a bivariate circular distribution. The procedure resembles that of Spearman's rank correlation coefficient. Let ϕ and θ denote the two correlated random angles. In a sample (ϕ_i, θ_i), $i = 1, 2, ..., n$, ϕ_i and θ_i are ranked relative to arbitrary zero directions. Mardia's correlation coefficient depends on the differences of the ranks of ϕ_i and θ_i much in the same way as Spearman's rank correlation coefficient depends on the differences of ranks of "linear" variates. Therefore, Mardia's r_0 may be interpreted as the circular analogue of Spearman's ρ. Fortunately, both small and large sample distributions of r_0 under the hypothesis of independence of ϕ and ρ are given in the paper.

I was disturbed by some remarks in the discussion that all random angles are treated as circular variates. I do not think that this is correct. A random angle that is clearly restricted to an interval smaller than 2π lacks the circular property since a given angular position cannot be reached by turning clockwise as well as counter-clockwise. Therefore, I claim that a random angle restricted to an interval smaller than 2π should be treated statistically like an ordinary "linear" variate. As an example of such an angular, but not circular, variate let me mention the latitude on earth if the polar regions act as barriers.

Dr J. W. THOMPSON (University of Hull): After reading Professor Mardia's paper it occurred to me that some progress concerning covariance and correlation on the circle might be made if we consider the problem as one of prediction. If we have two angular random variables Θ, Φ with mean directions μ, ν we can try to forecast the deviation of Φ from ν in the light of a known deviation of Θ from μ.

Since mean direction is itself obtained by considering the circle as a subset of the plane, there is no necessary objection to embedding the torus into R^2. We define

$$\Theta^* = \{(\Theta-\mu+\pi) \bmod 2\pi\}-\pi,$$
$$\Phi^* = \{(\Phi-\nu+\pi) \bmod 2\pi\}-\pi$$

and consider the transformation $\Theta \to \sin \frac{1}{2}\Theta^*$, $\Phi \to \sin \frac{1}{2}\Phi^*$. The corresponding relation between the torus and the square $\{(x, y);\ -1 \leqslant x, y \leqslant 1\}$ is 1 : 1 and the signs of deviations from mean directions are preserved. If Θ and Φ have reasonably concentrated unimodal distributions the induced measure on the square will itself be concentrated about the origin.

We can now use standard linear prediction on the square to give an induced non-linear prediction on the torus, namely, use $\sin \frac{1}{2}\Theta^*$ to predict the value of $\sin \frac{1}{2}\Phi^*$. In order to conform with the usual definition of variance on the circle we define

$$\operatorname{cov}(\Theta, \Phi) = E(2 \sin \tfrac{1}{2}\Theta^* \sin \tfrac{1}{2}\Phi^*)$$

so that

$$\operatorname{var} \Theta = \operatorname{cov}(\Theta, \Theta) = E\{1-\cos(\Theta-\mu)\}$$

and

$$\operatorname{corr}(\Theta, \Phi) = \operatorname{cov}(\Theta, \Phi)/(\operatorname{var} \Theta \operatorname{var} \Phi)^{\frac{1}{2}}.$$

Let us consider the consequences for the bivariate Normal distribution $N_2(\mathbf{0}, \mathbf{\Sigma})$, doubly wrapped onto the torus. If $\sigma_1, \sigma_2 \leqslant 1$ most of the R^2 distribution lies inside the square $\{(x, y);\ -\pi \leqslant x, y \leqslant \pi\}$. Adequate approximations are then given by

$$\operatorname{cov}(\Theta, \Phi) \doteqdot 2 \exp\{-\tfrac{1}{8}(\sigma_1^2+\sigma_2^2)\} \sinh(\tfrac{1}{4}\rho\sigma_1\sigma_2),$$
$$\operatorname{var}(\Theta) \doteqdot 1-\exp(-\tfrac{1}{2}\sigma_1^2).$$

The prediction of Φ^* becomes

$$\Phi^* = 2 \sin^{-1}[\sin \tfrac{1}{2}\Theta^* . 2 \exp\{-(\sigma_1^2+\sigma_2^2)/8\} \sinh(\tfrac{1}{4}\rho\sigma_1\sigma_2)/\{1-\exp(-\tfrac{1}{2}\sigma_1^2)\}],$$

which approaches the standard linear prediction if σ_1, σ_2 become sufficiently small.

For relatively concentrated distributions, the prediction seems reasonable but, in principle, more diffuse distributions might lead to poor prediction for large deviations of Θ from μ. Consider

$$f(\theta, \phi) \doteqdot \frac{1}{2\pi} \frac{1}{(2\pi)^{\frac{1}{2}}\tau} \exp[-\tfrac{1}{2}\{(\theta-\phi)^*\}^2/\tau^2\}$$

for $\tau \leqslant 1$. Here Θ, Φ have uniform distributions and the conditional distribution of Φ for $\Theta = \theta$ is wrapped Normal. Mean directions are now arbitrary and taking $\mu = \nu = 0$ will give less accurate results for large $|\Theta|$. The predictor is

$$\hat{\Phi} = 2 \sin^{-1}\{\exp(-\tfrac{1}{8}\tau^2) \sin \tfrac{1}{2}\Theta\}.$$

An obvious alternative predictor would be $\tilde{\Phi} = \Theta$ so it is useful to compare $(\hat{\Phi}-\tilde{\Phi})$ to τ. Numerical investigation shows this ratio to be acceptable. For example, the ratio can be shown to be less than $\frac{1}{2}$ for $|\theta| < 0{\cdot}81\pi$ when $\tau = 1$, whilst the situation improves as τ decreases.

Thus the proposed prediction and its associated covariance seem to have possibilities that may be worthy of further study.

Mr H. GOLDSTEIN (National Children's Bureau): I was partly responsible for the cancer cell data and it is always gratifying to have one's rather crude statistical efforts confirmed in such an elegant way as Professor Mardia has done. Starting from the grid line direction we marked off 24, 15-degree intervals and counted the number of cells with measured angles in each. We tested the resulting frequency distribution for symmetry and uniformity and, I am glad to say, obtained the same results as did Professor Mardia—as shown in Section 7.1.

There is a slight correction to the penultimate paragraph of Section 7.1: the cells were counted—the angles measured—only if they crossed a grid line. Thus, as the lines become infinitely wide apart, we do not go to a uniform angular distribution because the cells which actually cross a line are still the only ones to be counted. What happens is that kappa tends to a constant, and not to zero as in the other situation when the lines become infinitely close together.

Dr Laycock rightly questioned the point of all this. Might I say that the original experiment was to compare cancer cells with non-cancer cells, because it had been observed that the non-cancer cells tended to have a different orientation to the grid lines as compared to the cancer cells. The angles had a more uniform distribution, not clustering around the grid lines quite so much. This was exciting for the experimenters who thought it might lead to at least one Nobel Prize. Unfortunately it has led nowhere, which is the reason why the paper in the references by Crouch, Weiss and Goldstein is listed as "Unpublished"—it is unpublished in 1975, and I fear will remain unpublished!

Professor D. R. Cox (Imperial College, London): Consider tests of the adequacy of a fitted von Mises density. Such tests are of three broad types:

(a) graphical,
(b) broad tests of agreement, based on the distance between observed and fitted distributions,
(c) tests of specific aspects of the fit.

For (a), one possibility is to make an estimated probability integral transformation of the circle and to compare the resulting points visually with a uniform distribution. For (b), one could adopt any of the usual distance measures; possibly the recent work of Durbin (1975) could be adapted to deal with the effect of estimating parameters.

Methods powerful against specific alternatives may be useful especially when there are modest amounts of data. If the density is positive on the whole circle, the log density can be expanded in a Fourier series and in the absence of some more specific alternative, it is natural to consider densities proportional to

$$\exp(\alpha_1 \cos x + \beta_1 \sin x + \alpha_2 \cos 2x + \beta_2 \sin 2x).$$

Agreement with the von Mises density $\alpha_2 = \beta_2 = 0$ is thus tested from the conditional distribution of $(\Sigma \cos 2X_j, \Sigma \sin 2X_j)$ given $(\Sigma \cos X_j, \Sigma \sin X_j)$. To obtain a simple approximation, it is convenient to take the null hypothesis in the form proportional to

$$\exp\{\kappa \cos(x-\mu)\}$$

and to define

$$C_1(\mu) = \Sigma \cos(X_j - \mu), \quad S_1(\mu) = \Sigma \sin(X_j - \mu),$$

$$C_2(\mu) = \Sigma \cos\{2(X_j - \mu)\}, \quad S_2(\mu) = \Sigma \sin\{2(X_j - \mu)\}.$$

The joint distribution of the four random variables is asymptotically multivariate normal and it follows that approximately $C_2(\mu)$ and $S_2(\mu)$ are, conditionally, independently distributed with means

$$n\alpha_2 + \gamma_c(c_1 - n\alpha) \quad \text{and} \quad \gamma_s s_1$$

and variances nv_c and nv_s, say. Thus, omitting some details, we take as test statistics

$$\frac{C_2(\hat{\mu})-n\alpha_2(\hat{\kappa})}{v_c(\hat{\kappa})\sqrt{n}} \quad \text{and} \quad \frac{S_2(\hat{\mu})}{v_s(\hat{\kappa})\sqrt{n}}.$$

The values of γ_c and γ_s are immaterial;

$$v_c = \tfrac{1}{2}+\tfrac{1}{2}\alpha_4-\alpha_2^2-\frac{(\tfrac{1}{2}\alpha_3+\tfrac{1}{2}\alpha_1-\alpha_1\,\alpha_2)^2}{\tfrac{1}{2}+\tfrac{1}{2}\alpha_2-\alpha_1^2},$$

$$v_s = \tfrac{1}{2}(\alpha_1-\alpha_4)-\frac{(\alpha_1-\alpha_3)^2}{1-\alpha_2},$$

where

$$\alpha_r = \alpha_r(\kappa) = I_r(\kappa)/I_0(\kappa).$$

These test respectively symmetrized shape around the pole and symmetry about the pole and as such are analogous to the fourth and third cumulant ratios in Normal theory problems.

With several sets of data the test statistics can be plotted in the usual way. For a combined test we can either take the larger of the two absolute values or the sum of squares, the latter as chi-squared with two degrees of freedom.

Note that the exact conditional distribution of $C_2(\hat{\mu})$ and $S_2(\hat{\mu})$ is independent of the nuisance parameters (μ, κ). The quantities $(\hat{\mu}, \hat{\kappa})$ define the conditional distribution; more refined methods of approximation could be investigated, presumably via the relevant Laplace transforms.

The discussion would have been slightly simpler had μ been known.

Mr R. J. Gadsden (University of Leeds): I should first like to congratulate Professor Mardia on giving such an informative summary of these important new results in the study of directional data analysis.

My own particular interest is in the analysis of long-period comets given in Section 7.2. This study uses data given by Tyror (1957) consisting of observed comets up to and including 1952. Since then Marsden (1972) has produced a more complete catalogue. Using this source of data the sample size increases to 504 comets.

The results obtained in an analysis of this larger sample gives the following values for the test statistics

$$S_u = 21{\cdot}31, \quad S_g = 5{\cdot}84, \quad S_b = 7{\cdot}31.$$

It can be seen that S_g, is now only just below the 5 per cent significance level and the test statistic S_b not so highly significant as before.

Also, the angle between the normal to the preferred great circle and the galactic plane has increased from 6·1 to 10·7°.

Notwithstanding these slight differences, the conclusions are the same as those drawn by Professor Mardia.

Further models are at present under investigation which not only take account of perihelion direction but also of the perihelion distance from the sun. Special reference is made to the "reservoir theory" revived by Oort (1950), as well as to the accretion theory of Lyttleton (1953, 1961).

The following contributions were received in writing, after the meeting:

Professor Christopher Bingham (University of Minnesota, Saint Paul, Minnesota): It is extremely encouraging to see the study of orientation statistics emerge from the rather restricted circles to which it has confined itself. For this we all owe a debt to Professor Mardia whose recent book has brought attention to a growing literature. The present paper carries on this work of proselytizing the study of circular and spherical data. I have a number of comments, some of them more aesthetic than substantive. I will refer to the so-called Bingham distribution as the B distribution.

I find it a very useful convention when dealing with densities on the sphere to express them in the form, say, $f(\mathbf{l})\, dS/A_p$, where dS symbolizes an unspecified differential form giving the usual Lebesgue measure on the sphere S_p in p-space, and

$$A_p = c_p^{-1} = \frac{2\pi^{\frac{1}{2}p}}{\Gamma(\frac{1}{2}p)},$$

the area of the sphere. Thus, for example, I express $M(\boldsymbol{\mu}, \kappa)$ as

$$\{{}_0F_1(\tfrac{1}{2}p;\ \tfrac{1}{4}\kappa^2)\}^{-1} \exp(\kappa\boldsymbol{\mu}^{\mathrm{T}}\, \mathbf{l})\, dS/A_p.$$

Incidentally, I prefer the hypergeometric notation used here for the normalizing constant to the more usual Bessel function notation. The constant reduces to 1 for $\kappa = 0$. How one expresses the differential form dS is a matter of convenience and one can reparametrize the density itself without introducing a Jacobian, except as part of the expression for dS.

It should be remarked that Breitenberger (1963) was probably the first to systematically seek characterization theorems for spherical distributions. He characterized the von Mises and the B distributions in terms of maximum likelihood estimation. No discussion of the von Mises distribution is complete without mention of the exposition of Gumbel *et al.* (1953) and Gumbel (1954). Their rather unfortunate nomenclature—the circular normal distribution—has confused many people. However, modern work on its theory and applications dates from this paper.

One particular case of (2.11) of some interest is when $\mathbf{a}_1 = \kappa_1\, \boldsymbol{\mu}$, $\mathbf{a}_2 = \mathbf{0}$ and $A = \kappa_2\, I_p$. In this case the marginal of $\mathbf{l}_1$ is $M_p(\boldsymbol{\mu}, \kappa_1)$ and, conditionally on $\mathbf{l}_1$, $\mathbf{l}_2$ is $M_p(\mathbf{l}_1, \kappa_2)$. The marginal distribution of $\mathbf{l}_2$ is not, however, von Mises but is

$$\frac{{}_0F_1\{\frac{1}{2}p;\ \frac{1}{4}(\kappa_1^2 + 2\kappa_1\, \kappa_2\, \mathbf{l}_2^{\mathrm{T}}\, \boldsymbol{\mu} + \kappa_2^2)\}}{\{{}_0F_1(\frac{1}{2}p;\ \frac{1}{4}\kappa_1^2)\, {}_0F_1(\frac{1}{2}p;\ \frac{1}{4}\kappa_2^2)\}}\, \frac{dS}{A_p}.$$

An important contributor to the theory of the isotropic case was Lord (1954a, b) in a pair of papers on the use of the Hankel transform in statistics. He showed how to express the distribution of the resultant of a sample from an isotropic distribution in terms of Hankel transforms. Moreover, he shows how the distribution of the projection of the resultant vector on a hyperplane can be deduced from the distribution of the resultant itself, and vice versa. The distribution of R when $p = 3$ is most easily derived by his method.

I think it is well worth pointing out that Fisher considered the inference problem for $M_3(\boldsymbol{\mu}; \kappa)$ to be an important test case for the fiducial method. As such, it has attracted considerable attention from those concerned about the foundations of inference (Barnard, 1963; Williams, 1963; Fraser, 1968, pp. 196–202). The fact that the Watson and Williams test is for an axis rather than for a pole puts somewhat of a damper on my enthusiasm for its optimality properties. If one "corrects" it by taking the "other end" of the cone out of the critical region, one loses similarity. I think there is an error in the stated ascription of size to the confidence cone obtained by inverting the corrected test (i.e. only including a region near the resultant). The problem is that the events $(0 < \bar{x}_{0,1} < \frac{1}{2}\pi)$ and $(R > R_0)$ are not independent. Thus the factor p^* should be $p^* = P(0 < \bar{x}_{0,1} < \frac{1}{2}\pi \mid R > R_0)$, not $P(0 < \bar{x}_{0,1} < \frac{1}{2}\pi)$. For even moderate n and κ, this should be even closer to unity than the stated value.

I am pleased to see an analytic expression for the distribution of the cross-product matrix $\mathbf{T}$ based on a sample from the B distribution. Incidentally, ${}_0F_0^{(p)}$ is to distributions of eigenvalues of sample cross-product matrices of isotropic distributions what Bessel functions (see Lord, 1955a) are to distributions of resultants of sums of isotropic random vectors. My approximation is for ${}_0F_0^{(3)}$ but only when the arguments are real matrices with well-separated eigenvalues. It has applicability to the B distribution (see Bingham, 1974) because the marginal distribution of $\boldsymbol{\tau}$ for arbitrary $\boldsymbol{\kappa}$ is ${}_0F_0^{(p)}(\boldsymbol{\kappa}, \boldsymbol{\tau})\, d^{-n}(\boldsymbol{\kappa}) \times$ (distribution of

$\boldsymbol{\tau}$ when $\boldsymbol{\kappa} = 0$). Thus the likelihood function of $\boldsymbol{\kappa}$ based on the marginal distribution of $\boldsymbol{\tau}$ can be computed even though the complete distribution is intractable.

The analysis of long-period comet data is an interesting application of the B distribution. Incidentally, in a manuscript being prepared for publication, I show that an approximation to $-2\log\lambda$ for this hypothesis, *not* assuming circular symmetry, is

$$2n\{\hat{\kappa}_1 - \tfrac{1}{2}(\hat{\kappa}_2 + \hat{\kappa}_3)\}\,(\tau_1 - \boldsymbol{\mu}_0^{\mathrm{T}}\,\mathbf{T}\boldsymbol{\mu}_0/n)$$

(for bipolar case, $2n\{\hat{\kappa}_3 - \tfrac{1}{2}(\hat{\kappa}_1 + \hat{\kappa}_2)\}\,(\tau_3 - \boldsymbol{\mu}_0^{\mathrm{T}}\,\mathbf{T}\boldsymbol{\mu}_0/n)$). Here this has value 0·950, also distributed approximately as $\chi^2(2)$ under the null hypothesis.

Dr M. S. Bingham (University of Hull): As Professor Mardia points out in his paper, circular analogues of characterizations of the linear Normal distribution sometimes lead to the von Mises distribution but in other cases lead instead to the Brownian motion (wrapped Normal) distribution on the circle. While this dichotomy has led to semantic arguments over which distribution should be called "circular Normal", it seems to have been overlooked that the two types of distribution arise in different contexts. In some problems, as for example in considering a circular Brownian motion, the addition or subtraction of angles plays a fundamental role, whereas in others these operations are not of great importance. In the former case it is necessary to take account of the group structure of the circle under the addition of angles, whereas in the latter case we embed the circle in the plane with its centre at the origin and exploit addition in the plane to define such concepts as the mean direction. Note that addition in the plane is incompatible with adding angles round the circle so that the two approaches treat the circle in fundamentally different ways. The wrapped Normal distribution will tend to arise when addition of angles is considered and the von Mises distribution will arise when it is ignored.

Under the operation of adding angles and with its natural topology, the circle is a compact (topological) group. Other examples of spaces which may be treated as compact groups include hyperspheres and toruses, the latter being relevant for considering the joint distributions of several circular random variables. If we wish to consider the joint distribution of a circular random variable and a linear random variable then we need a probability distribution on the surface of an infinite cylinder of circular cross-section; such a surface can be given the structure of a locally compact abelian group. Pure probabilists have already done a considerable amount of interesting work on the theory of probability distributions on locally compact groups; see, for example, Heyer (1972).

Regarding Mr Kent's circular analogue of the Ornstein–Uhlenbeck process, this seems to provide a heuristic reason why a von Mises distribution with a high concentration will be very close to a wrapped Normal distribution. If the concentration is high then the displacement θ is small with high probability and the restoring force proportional to $\sin\theta$ is approximately proportional to θ. Therefore the Kent diffusion process, with its von Mises displacements, will be approximated by a corresponding wrapped Ornstein–Uhlenbeck process, which has wrapped Normal displacements. Notice that the factor $\sin\theta$ in the restoring force is clearly related to embedding the circle in the plane.

Regarding Professor Lewis' suggestion of looking for outliers in directional data, there is the following interesting difficulty that may arise directly as a result of the compactness of the circle; it seems plausible that one could get say n observations around a circle distributed in such a way that, *whatever* the $(n+1)$st observation, it would not be rejected as an outlier at any reasonable significance level. However, suppose that we have n observations such that we can surmount this difficulty. If the population is unimodal then presumably there will be an arc A such that any future observation lying in that arc would be rejected as an outlier. Would it then be reasonable to unwrap the circle on to a finite interval in the line by cutting it at some point in the arc A and use linear techniques on the resulting observations? If the population is not unimodal, the rejection region A may turn out to be the union of several disjoint arcs and unwrapping by cutting in different arcs may lead to significantly different shapes for the linearized data, so that the validity of the method in this case would be doubtful.

Drs THOMAS D. DOWNS and CLAYTON W. EIFLER (University of Texas): Professor Mardia's paper gives a unified view of the overall present state of affairs in dealing with angular data, as well as proposing some solutions to current problems and enumerating some unsolved problems.

The circular rank correlation coefficient introduced in Section 6.2 is a welcome addition to the statistical methodology of angular data. There is an urgent need for such measures of the extent of angular correlation.

The general circular coefficient introduced in Section 8 measures the extent to which angles are rotationally correlated, and thus is also an indirect measure of the form of the correlation. This coefficient can be shown to be equivalent to that introduced by Downs (1974). Unfortunately, the distribution of this coefficient will in general depend on the degree of clustering of the angular data, and thus its usefulness must be somewhat limited. This problem arises at least in part from the fact that perfect rotational correlation cannot be attained when one angle of the angle pair is more dispersed than the other. The analogy with the linear correlation coefficient, wherein perfect correlation can be attained with unequal variances, thus breaks down.

This difficulty could perhaps be circumvented by angular transformations of the data which equalize the two angular dispersions while preserving their geometric properties. Transformations of this type which are mathematically robust yet still tractable seem hard to find. We would ask Professor Mardia if he knows of any work (fruitful or unfruitful) that has been done in this area. Finally, we thank Professor Mardia for yet another significant contribution to the armamentarium of those of us engaged in the statistical analysis of directional data.

Professor D. G. KENDALL (Statistical Laboratory, University of Cambridge): I was very sorry to have to miss this extremely interesting meeting, especially as Professor Mardia's work touches on my own interests at several points. Spline techniques for studying Schmidt nets seem now to be on the agenda (Section 7.2), and the connection between Professor Mardia's Section 7.5 and what I have called respectively the cosine and modular *quantograms* will also deserve study. As in his problem there is (I assume) no preferred time-origin, it is the modular quantogram that is relevant there. Of course Mardia's problem is different from the one with which I was concerned in his reference (Kendall, 1974b), in that Mardia already has specified candidates for the "period", whereas my question was: is there evidence sufficient to support the existence of *some* period in the interval $[\lambda_0, \lambda_1]$? One or two points might be made here about this paper (Kendall, 1974b). In the first place, it contains a silly mistake; the first displayed formula on p. 261 is *not* the likelihood, because I forgot to include the differential $d\theta_1\, d\theta_2 \ldots d\theta_N$, which is $(2\pi\tau)^N\, dX_1\, dX_2 \ldots dX_N$ in my notation (in which $\tau = 1/\lambda$). The factor τ^N cancels another τ^N, with the result that all logarithmic terms should be deleted from my formulae for the "relative support" (which I denoted by \$). This means that the "isopithons" drawn on my diagrams are incorrect, and are more complicated than they need to be. In fact we find, in the *CQG* case, that the *CQG*-plot with all negative arcs deleted gives the plot of $\surd(2\$)$ against τ, to the order of approximation to which I am working. That is, the highest peak of the *CQG* over the supporting interval $[\tau_0, \tau_1]$ locates the maximum likelihood estimate of τ.

To come now to the test for the significance of the height *of the highest peak*; I dealt with this matter in my (1974b) paper by Monte Carlo simulation, but the (rather surprising) measure of agreement with the Cramér limit-formula (which is distribution-free at the limit, conditionally upon a knowledge of $\surd\sum X_j^2$) suggests that the latter can safely be tried out (if not relied upon) in practice. In this formula the data only occur (apart from the normalized height of the highest peak) in the factor

$$(\tau_1 - \tau_0)\sqrt{\sum X_j^2}, \tag{*}$$

where $(X_1, X_2, ..., X_N)$ are the observations. Now $(\tau_1 - \tau_0)$ is a frequency interval, and dimensional considerations show that we can think of $\sqrt{\sum X_j^2}$ as an inverse frequency, so that it seems appropriate to think of

$$1/\sqrt{\sum X_j^2}$$

as the relevant *bandwidth*. Obviously, the larger $(\tau_1 - \tau_0)$ is, the more likely we are to get a large peak by mere chance, so we should expect to find in the Cramér formula a pure number representing (so to speak) the number of "chances" we have given ourselves to get that high peak. We cannot identify this with $(\tau_1 - \tau_0)$ itself, because that is not a pure number, but we *can* identify it with the pure number (*) above; this is just the frequency range which has been scanned, divided by the "bandwidth". We can think of this ratio (i.e. of (*) above) as the effective "number of observations" in the problem. There are indications that it really *is* $\sqrt{\sum X_j^2}$, and not $\sqrt{E(X^2)}$, which is relevant here, but there are aspects of this which require further study. Some unpublished work by Mr Michael Behrend bears on this question.

Another interesting point is that the *CQG* (which is effectively the first quantity within brackets in Mardia's formula for R_λ^*) is up to a constant multiple nothing other than the *empirical characteristic function* (or rather, its real part; the imaginary part occurs within the second bracket). A nice feature used informally in the (1974b) paper, and subsequently precisely formulated and proved by Mr John Kent, is that for large N we are dealing asymptotically with a standardized gaussian stationary stochastic process, observed over a compact "time"-interval $[\tau_0, \tau_1]$. We have to consider the distribution of the supremum of this stochastic process over this interval; obviously that distribution depends only on the covariance function for the stochastic process. This, however, can be estimated from the appropriate numerical multiple of the *CQG* near to $\tau = 0$. In other words, *the "head" of the CQG supplies an estimate of all that is needed for the study of the stochastic relationships in its "tail"*, and this crude principle is one that seems to have escaped notice until now, even in what ought to be the more familiar context of empirical characteristic functions.

I had hoped to say something also about the stochastic properties of Schmidt nets, but that is perhaps better deferred to another place. I should like to close these remarks by saying how much I admire Professor Mardia's co-ordinated and well thought-out attack on the problems in the much neglected subject of directional statistics. There is some overlap between this statistical subject and the probabilistic subject, stochastic geometry, in that both might be said to be concerned with "stochastics on homogeneous spaces". Dr Ambartsumyan, Professor Kingman, Professor Krickeberg and I are hoping to organize a Symposium in the next year or two, dealing with this general topic, and we should be glad to hear from anyone who has ideas to contribute.

Mr J. Proctor (University of Glasgow): In Section 7.1 of Professor Mardia's paper, his analysis of cancer cell orientation data, he considers several samples of cells, each taken to have a von Mises distribution. The estimate, $\hat{\kappa}$, of the concentration parameter varies with the number of grooves per inch of the plate on which the cells are mounted, and Professor Mardia postulates a regression model to describe this variation.

This situation is similar to an analysis of data that I have recently been working on. The data were measurements made of particle size, shape and orientation, (l, s, θ), in sections of compressed clay. The variables were continuous, l being defined on $(0, \infty)$, s on $(0, 1]$, and θ on $(0°, 180°]$, doubled to $(0°, 360°]$. The problem was to describe the 271 trivariate observations by a density function. Tests indicated that the distributions of l and θ, conditional on values of s, could be taken as independent. That is, the density function decomposed as $g(s)\, h_1(l \mid s)\, h_2(\theta \mid s)$, or $k(l, s)\, h_2(\theta \mid s)$; the form of h_2 is of principal interest here.

The 271 observations were classified into five groups by s values, the group boundaries being 0·3, 0·4, 0·5, 0·6, and the group sizes varying from 34 to 68. It was found that the distribution in each class was von Mises, the preferred directions all being close together. The maximum likelihood estimates for μ and κ, the two von Mises parameters, and the length R of the resultant vector of the observations are shown in the table below. The mid-point, $\bar{s}$, is the average s value of the observations in that class, and n is the class size.

Class	n	R	$\hat{\kappa}$	$\hat{\mu}$	$\bar{s}$
0·0–0·3	61	48·70	2·852	0·47°	0·230
0·3–0·4	68	42·68	1·630	4·11°	0·345
0·4–0·5	64	34·94	1·312	0·75°	0·435
0·5–0·6	44	27·29	1·601	−2·71°	0·535
0·6–1·0	34	10·44	0·645	−5·74°	0·695

Tests for the means indicated that they could all be taken as equal. Models were now postulated for the regression of κ on s; the two particularly considered were that κ be a linear function of s, and that κ be a linear function of $\log s$. Physical considerations suggested that a regression model should have $\kappa(1) = 0$, since s was the ratio of the lengths of the two principal axes of the particle. We wished to discover which, if either, of the two models above gave a satisfactory description of the data, and whether the constraint that the line pass through $(1, 0)$ was acceptable. The analysis was based throughout on the grouped data; otherwise calculations would have been prohibitively complicated.

The method used was a weighted least squares technique. Taking $(\bar{s}, \hat{\kappa})$ as representative of each class, we had five points to which to fit a regression line. Interpolating in Batschelet's tables (Appendix 2.8 of Mardia's book, 1972a), and also using tables in Stephens' paper (1969), equal tails 90 per cent confidence limits for κ in each class were found, denoted by κ_1 and κ_2.

Mardia's book (4.9 and 5.4) gives the results that $\bar{R}(= R/n)$, and hence $\hat{\kappa}$\{which satisfies $A(\hat{\kappa}) = \bar{R}$, where $A(\hat{\kappa}) = I_1(\hat{\kappa})/I_0(\hat{\kappa})$\}, are asymptotically Normally distributed, with variance of order n^{-1}. (Numerical work is currently in progress on the exact distribution of $\hat{\kappa}$ in the ranges that apply to these data.) Since $\hat{\kappa}$ is asymptotically Normal, then, asymptotically, the equal tails 90 per cent confidence limits for $\hat{\kappa}$ will be $3{\cdot}29\sqrt{\{\operatorname{var}(\hat{\kappa})\}}$ apart, and hence an estimate of the variance of $\hat{\kappa}$ can be made as $(\kappa_2-\kappa_1)^2/3{\cdot}29^2$. A weighted least squares analysis, weighting each squared error by the reciprocal of the appropriate variance estimate, was then carried out. Residual sums of squares and their ratios were compared with percentage points of the appropriate χ^2 and F distributions, as a test for the validity of the models. Results were:

	Unconstrained residual R_0 (3 d.f.)	*Constrained residual* R_1 (4 d.f.)	$F_{1,3}$ $(R_1-R_0)/(R_0/3)$
Linear	5·00	5·63	0·378
Logarithmic	3·85	3·99	0·109

Although the roughness of the mathematical treatment would question this as a basis for drawing firm conclusions, it seemed to indicate that a logarithmic model would be better, and also that the constraint was acceptable. The model suggested by this analysis was then used in building the full trivariate model, which was subsequently tested more formally.

In 5.4 of Mardia's book expressions are given, to order n^{-1}, for the bias in $\hat{\kappa}$ as an estimate of κ, and also for the variance of $\hat{\kappa}$. These are $-\frac{1}{2}A''(\kappa)/n\{A'(\kappa)\}^2$ for the bias, and $1/n\{A'(\kappa)\}$ for the variance. The bias and variance can be estimated by substituting $\hat{\kappa}$ in these expressions. Below are shown these estimates, $\hat{b}$ and $\hat{v}$, and also the estimates v^*, based on the confidence limits and used in the analysis.

$\hat{\kappa}$	n	$\hat{b}$	$\hat{v}$	v^*
2·852	61	0·075	0·199	0·187
1·630	68	0·027	0·066	0·071
1·312	64	0·020	0·055	0·067
1·601	44	0·040	0·100	0·102
0·645	34	0·015	0·068	0·084

Using $1/\hat{v}$ as weight in the analysis would give:

	R_0 (3 d.f.)	R_1 (4 d.f.)	$F_{1,3}$
Linear	5·15	5·78	0·371
Logarithmic	3·98	4·13	0·106

This gives slightly larger residuals than the procedure actually used, but conclusions are essentially the same in both cases.

The mathematically rough approach used provides a reasonable method of testing a regression model for κ, particularly as part of a larger piece of model building.

Professor J. S. Rao (Indiana University): It has been a great pleasure for me to read the paper by Professor Mardia, who succeeds admirably in presenting some of the recent research on the topic of directional data. I would like to make two brief comments. In Section 6, r_0 is proposed as a measure of association where it is assumed l and m are known. Are these quantities related to the modality of the respective parent populations or is one to choose l and m that make r_0 a maximum for the given sample? Since these quantities seem analogous to the regression coefficients in the linear case, I wonder if Professor Mardia considered the possibility of estimating them and the changes such estimation would introduce in the distribution of r_0.

My other comment relates to the non-parametric tests of uniformity on the circle, which have been adequately discussed in Mardia (1972a). The circular spacings $D_1, \ldots, D_n$ (i.e. the sample arc lengths made by successive observations) form a maximal invariant under rotations in this situation so that any invariant test is a function of $\{D_i\}$. In this connection Rao and Sethuraman (1970, 1975) have shown that any test function based symmetrically on spacings has a poor asymptotic relative efficiency. However, since all the invariant tests, like for instance those suggested by Kuiper and Watson (see Mardia, 1972a) are functions (though not necessarily symmetric) of $\{D_i\}$, it would be interesting to obtain their null distributions using spacings methods and the empirical spacings processes, along the lines of Rao and Sethuraman (1975). The same type of results could also be used to derive their limiting distributions under a suitable sequence of alternatives that converge to uniformity, thus enabling one to compute the asymptotic relative efficiencies of many such tests of uniformity. Research along these lines is currently in progress. This underlines the fact that spacings have a key role to play in testing uniformity especially on the circle. At the same time, these asymptotic results also hold good for the linear case and throw some light on their relative merit as compared to various other methods available there. I would be interested in hearing any comments Professor Mardia would care to make regarding the relative merits of the non-parametric tests that use uniform scores to which he refers in Section 6 versus those using spacings when both these methods are available, as in testing whether two circular populations are identical. I would like to congratulate Professor Mardia again on an excellent paper that I am sure will stimulate further research in this area of directional data analysis.

Dr A. L. Rukhin (Mathematical Institute, Leningrad): The examples and theory presented by Professor Mardia in his finely crafted paper provide a valuable survey of the statistical analysis of directional data. Another useful approach to the von Mises–Fisher and Bingham distributions seems to me to be the following. Let $p(u)$ be a positive and continuous density given on the p-dimensional sphere S. In directional data analysis the important role is played by the transformation parameter family $\{p_g: p_g(u) = p(g^{-1}u), g \in SO(p)\}$. When deciding what model to choose a statistician would prefer densities p with non-trivial sufficient statistics for the rotation parameters. Exactly this fact is responsible for the author's statement that "the von Mises–Fisher distribution leads to tractable maximum likelihood estimates and sampling distributions in problems of hypothesis testing, whereas the Brownian motion does not". The mentioned densities have the form $p(u) = \exp\{\sum_{l=1}^{m}\langle T_l(u)\,\psi_0, \psi\rangle\}$ where $\{T_l\}$ denotes an irreducible unitary representation of the group $SO(p)$, $T_l(gh) = T_l(g)\,T_l(h)$, and where the vector ψ_0 is invariant under the subgroup $SO(p-1)$, so that the function $\langle T_l(u)\,\psi_0, \psi\rangle$ is defined in fact on the sphere $S = SO(p)/SO(p-1)$. In the case $p = 3$ for the von Mises–Fisher distribution, $p(u) = \exp\{\langle T_1(u)\,\psi_0, \psi_0\rangle\}$ where T_1 has dimension 3 and for the Bingham distribution $p(u) = \exp\{\langle T_2(u)\,\psi_0, \psi\rangle\}$ with 5-dimensional representation T_2 and some ψ defined by parameters κ_1 and κ_2. A corresponding formula holds for the generalized von Mises–Fisher distribution introduced in Section 2.4.

An interesting problem is sequential inference of directional data. The best stopping rule for the estimation of the direction of the von Mises–Fisher sample seems to prescribe that we stop when the norm of the resultant vector $\sum l_i^T$ for the first time exceeds some level, depending on κ and the cost of observation. The problems with unknown κ are of course more difficult.

The author replied as follows:

I am extremely grateful to all discussants for their very interesting and helpful comments. Indeed, I am overwhelmed by the response the paper has received both here and abroad.

Professors Daniels and Kingman question the statistical meaning of the maximum entropy condition in deriving distributions; in the process, Professor Daniels also gives a justification for which I am grateful. Jaynes (1957, 1963) seems to be the first to postulate the "maximum-entropy principle". He regards the principle as providing a constructive criterion for determining probability distributions on the basis of partial knowledge in the form of summary statistics or equivalently constraints related to population parameters such as $E(l)$, etc. He asserts that the maximum-entropy distribution is uniquely determined as the one which is maximally non-committal with regard to missing information (see also Tribus, 1962). Observing that the principle led to expressions formally equivalent to those of statistical mechanics, Jaynes (1957) argues that in the resulting subjective statistical mechanics the usual rules are justified independently of experimental verification because, whether or not the results agree with the experiment, they still represent the best estimates that could have been made on the basis of the information available. Of course, the principle depends on the choice of summary functions. Finch (1973) asserts that theory cannot tell us which summary statistics will lead to results in agreement with experiments and, indeed, one has to experiment to find out which summary statistics provide useful macroscopic descriptions of microscopic phenomena. In the comet data, it is comforting to note that the summary statistics were devised by astronomers! Indeed, this principle is invoked quite frequently by statisticians. The distributions used by Professor D. R. Cox and Dr Laycock in the discussion are immediate examples. For an interesting practical use in urban geography with which statisticians may not be so familiar, I refer to Wilson (1970). It is appropriate to conclude this part of the discussion with the following far-reaching comment by the Chairman (Plackett, 1966):

> "My view is that the differences of opinion should be patiently explored, and that the amount of agreement already in existence should be emphasized. We are dealing with ideas which go rather deeper than logic or mathematics and the important point is, not that we should start from the same assumptions, but that we should reach essentially the same conclusions on given evidence if possible."

We have not investigated the maximum entropy distribution for fixed $E(\mathbf{1})$ and $E(\mathbf{1}\mathbf{1}^{\mathrm{T}})$. The reasons are partially due to the lack of mathematical developments to deal with such situations. In fact, in view of the recent work of Professor Khatri in multivariate analysis, it is possible to investigate the density of $(\mathbf{X}, \mathbf{Y})$ on a Stiefel manifold of the form

$$C \exp\{\operatorname{tr}(\mathbf{FX}) + \operatorname{tr}(\mathbf{GY}) + \operatorname{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{AXB}) + \operatorname{tr}(\mathbf{Y}^{\mathrm{T}}\mathbf{CYD}) + \operatorname{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{SYT}) + \operatorname{tr}(\mathbf{Y}^{\mathrm{T}}\mathbf{UXV})\},$$

where the matrices are appropriately constrained.

Professor Cox's ingenious "tests of von Misesness' deserve every praise. In my book, I also put forward measures of circular skewness (g_1^0) and kurtosis (g_2^0) based on C_2 and S_2 but with scaling factors chosen so as to render them identical to the usual g_1 and g_2 for large κ. The close connection between the wrapped normal and von Mises distributions was also exploited, and with the same argument, it will be convenient to use the measures

$$\{C_2(\bar{x}_0) - (1-\bar{R})^4\}/(n^{\frac{1}{2}}\,\nu_c), \quad S_2(\bar{x}_0)/(n^{\frac{1}{2}}\,\nu_s),$$

where

$$2^{\frac{1}{2}}\,\nu_c = (1-\bar{R})^2(1+\bar{R})^2(1+\bar{R}^2)\,(1+\bar{R}^2+\bar{R}^4)^{\frac{1}{2}},$$

and

$$2^{\frac{1}{2}}\,\nu_s = (1-\bar{R})^{\frac{3}{2}}\{(1+\bar{R}^2)\,(1+\bar{R}^4)\,(1+\bar{R}^2+\bar{R}^4)\}^{\frac{1}{2}}.$$

In g_1^0 and g_2^0, the corresponding scaling factors are proportional to $(1-\bar{R})^2$ and $(1-\bar{R})^{\frac{3}{2}}$, and it may be worth while to compare these tests (after using the asymptotic variances of g_1^0 and g_2^0). Knowing the notorious behaviour of the normal approximation for g_1 and g_2, one would expect n to have to be large (at least for kurtosis) before the normal approximations become operative in the tails.

Various discussants have commented on the concept of a circular correlation coefficient. The pioneering work of Mackenzie (1957) on the regression problem for spherical variables should be emphasized. Consider the regression of $\mathbf{y}$ on $\mathbf{x}$ by $\mathbf{y} = \mathbf{Rx}$ where $\mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}$, $\mathbf{x}^{\mathrm{T}}\mathbf{x} = \mathbf{y}^{\mathrm{T}}\mathbf{y} = 1$. If $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, \ldots, n$, are the observations, he showed, by maximizing $Q = \sum_{i=1}^n \mathbf{y}_i^{\mathrm{T}}\mathbf{R}\mathbf{x}_i$, that $\mathbf{R} = \mathbf{R}_1\mathbf{R}_2^{\mathrm{T}}$ where $\mathbf{R}_2\mathbf{\Lambda}\mathbf{R}_1^{\mathrm{T}}$ is the spectral decomposition of $\mathbf{XY}^{\mathrm{T}}$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$. Further, $Q_0 = \max Q = \sum \lambda_i^{\frac{1}{2}}$ where the λ_i's are the eigenvalues of $\mathbf{XY}^{\mathrm{T}}\mathbf{YX}^{\mathrm{T}}$. Obviously, Q_0/n can provide a measure of dependence. Downs (1974) considers a slightly modified Q, i.e. $Q = \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})^{\mathrm{T}}\mathbf{R}(\mathbf{x}_i - \bar{\mathbf{x}})$, and the resulting Q_0 is scaled to give his measure of dependence. Interestingly for the circular case, the fact that the correlation coefficient, given in Section 8 on intuitive grounds could flow from Mackenzie–Downs' work is indeed incredible although true, since the form given in the paper has not appeared elsewhere! I am grateful to Professor Downs and Mr Eifler for pointing out this connection. The prediction method proposed by Dr Thompson shows deep insight and it will require further examination. His predictor puts the regression of θ on ϕ in a single equation in place of the two in Mackenzie's formulation. Note that

$E(x-a)(y-b)$ has stationary points at $a = E(x)$ and $b = E(y)$ which leads to $\operatorname{cov}(x,y)$. Using this requirement on

$$E\{\sin \tfrac{1}{2}(\theta^* - a) \sin \tfrac{1}{2}(\phi^* - b)\}$$

leads to $\max(D_+, D_-)$ of the paper. It is pleasing to note that various different approaches seem in essence to lead to the correlation coefficient given in Section 8.

For the circular case in Section 6, it is assumed in the paper that we know l and m *a priori*, but when in doubt, Section 7.5 will help in estimating their values and the effect of this estimation on r_0 would need further investigation. This point should have been clearly stated as pointed out by Drs Laycock and Rao. However, for any circular variable, it is usually plausible to make such an assumption (i.e. whether the corresponding population is unimodal or multimodal/axial) and unless we have reasons to believe to the contrary (see Section 6, Example 2), we would take $l = 1$ and $m = 1$ (see Section 6, Example 1).

Professor Batschelet's warning on the temptation of using circular variates where linear techniques are appropriate is an important one. Indeed, by conditioning the choice of zero directions on the observed values, Dr Upton shows how the Spearman rank correlation r_s can be meaningful for circular data although Dr Upton recognizes its pitfall in general. I am grateful to Professor Batschelet for his analogy of r_0 with r_s but a direct extension of r_s for the circular case would proceed as follows. Calculate all possible values of r_s after varying the zero directions (in the negative as well as in the positive direction). Then $\max(r_s)$ or $\min(r_s)$ over all $4n^2$ possible values can be regarded as a direct extension of r_s (cf. Batschelet *et al.*, 1973).

Professor Lewis's penetrating observations on the problem of outliers in directional data analysis are most enlightening. In spite of the finite distance between directions, outliers can arise from recording errors, animals being distracted away from their natural home direction, earthquakes occurring away from established danger zones, rocks in a location becoming disorientated, etc. $\operatorname{Max}_i (n - R_i)/(n+1-R)$ is an obvious criterion to detect an outlier where R_i is the resultant length after omitting the ith observation out of $(n+1)$ observations and R is the resultant length of $n+1$ observations. A plot of the R_i's is informative. One would expect its power to increase with p, such a property being desirable in view of Professor Lewis's comment. I have come across possible recording errors in axial data such as a single observation being given as 321° while all others are reduced modulo 180°. To test a single outlier $\mathbf{1}^*$, the statistic $\mathbf{R}_x^{\mathrm{T}} \mathbf{1}^*/R$ is appropriate since the Bhattacharya distance between $M_p(\boldsymbol{\mu}_1, \kappa)$ and $M_p(\boldsymbol{\mu}_2, \kappa)$ is a function of $\kappa \boldsymbol{\mu}_1^{\mathrm{T}} \boldsymbol{\mu}_2$. Its asymptotic distribution under H_0 can be written down. If there are many outliers in the opposite direction to $\boldsymbol{\mu}$ as in certain navigational problems, it can be shown that the estimate of κ will be seriously influenced in comparison to the estimator of $\boldsymbol{\mu}$. These parametric investigations complement Mr D. Collett's non-parametric work described by Professor Lewis. Dr M. S. Bingham clearly demonstrates the hazard of using linear techniques for outlier problems in directional data.

Concerning Professors Kingman and Puri's query regarding robustness, we note that the general feeling from the study done so far is that the tests for mean directions are robust but the tests for concentration parameters are not. The area of robust estimation for directional data is still an open one. The characteristic function of the unwrapped von Mises distribution was also put forward in Mardia (1972a, p. 63) and it is pleasing to see the corresponding density derived by Professor Daniels. Incidentally, the maximum likelihood characterization for von Mises–Fisher density for all p is given in M. S. Bingham and Mardia (1975).

Dr Laycock's example of directions in $p = 15$ dimensions and its framework is most inspiring and I look forward to his fuller analysis. In the spirit that Bayesian analysis is sometimes more relevant, basic points are examined in Mardia and El-Atoum (1974). I am grateful to Mr Goldstein for describing his original experiment of comparing cancer cells with non-cancer cells, but I think Dr Laycock wanted the exact details of how the data were collected in the revised experiment described in the paper. Obviously, in the revised experiment, the observations are axial which clearly justifies doubling the angles. One of the main aims of the experiment was to obtain a regression curve of the type given in the paper where the *ad hoc* procedure could have hardly been successful; it is indeed a pleasure to see a similar example of Mr Proctor. Should one use the normal theory approximations when $\hat{\kappa}$ is as low as 0·645 and $n = 34$?

Professor Rukhin's and Dr M. S. Bingham's formulations should satisfy some statisticians on the uses of pure probability theory. These connections may help to examine Dr Upton's suggestion of intermediate limiting processes.

I think the type of work Professor Puri describes on circular independence was first formulated by Rothman (1971). As in the linear case, tests for scale would only be asymptotically non-parametric and the first step is to develop such tests when the mean directions are given and Mr El-Atoum has been working on this idea. Mr Gadsden's effort in updating the comet data should be welcomed by astronomers. It is worth while to look into further models which would take into account not only perihelian directions but also the other astronomical co-ordinates. Mr Gadsden has also worked out various sequential procedures and I am pleased to read the comment from Professor Rukhin on the topic.

It is most kind of Professor Kendall to point out the difference between my Section 7.5 and his work. Indeed, Professor Kendall is leading an excellent team in this area and we should express our tribute to him for various significant contributions, including the ones he describes in the discussion.

I am most grateful to Professor C. Bingham for his various valuable comments; my thanks are also due to him for spotting several misprints in the galleys. I agree with him that in higher dimensions it is preferable to work upon hypergeometric rather than Bessel functions, especially as in higher dimensions, hypergeometric functions are related to now well-established zonal polynomials. My joint work with Professor Khatri exploits this idea effectively for distributions on the Stiefel manifold.

Professor C. Bingham enquires whether p^* in (4.6) should be conditioned on $R > R_0$. The formulation as given in the paper is too condensed and thus can be misleading. The confidence cone can be regarded as equivalent to the acceptance region of the following test.

Accept

$$H_{\boldsymbol{\mu}^*}: \boldsymbol{\mu} = \boldsymbol{\mu}^* \quad \text{when} \quad R_{x_1}(\boldsymbol{\mu}^*) = \sum_{i=1}^{n} \mathbf{1}_i^{\mathrm{T}} \boldsymbol{\mu}^* > 0, \quad R \leqslant R_\alpha,$$

where

$$\Pr\{R \leqslant R_\alpha \mid R_{x_1}(\boldsymbol{\mu}^*)\} = 1 - \alpha \quad \text{and} \quad R_\alpha = h(R_{x_1}(\boldsymbol{\mu}^*), \alpha).$$

It can be shown that the acceptance region of this test is of the form

$$\boldsymbol{\mu}^* \in \{\phi < \phi_1(R)\}, \tag{1}$$

where

$$\cos\phi = R_{x_1}(\boldsymbol{\mu}^*)/R, \quad 0 < \phi < \tfrac{1}{2}\pi, \quad \phi_1(R) = \cos^{-1}(Z/R)$$

with Z determined by $R = h(Z, \alpha), Z > 0$. For a given sample, R will be replaced by R_0 in (1) and thus, we are led to (4.6) whether $R > R_0$ or $R < R_0$. For the one-sided fiducial cone of Williams (1963) the argument is much simpler. Fuller details of these points will appear elsewhere.

It was a great relief when Professor Lewis announced that he has proved the infinite divisibility of the von Mises distribution for small κ since there was a rumour emanating from pure probabilists that the contrary was true. Mr Kent should be congratulated for successfully proving the result for all κ. Indeed we are further grateful to Mr Kent for yet another justification of the distribution; Dr M. S. Bingham shows us how this clarifies the close connection between the von Mises and wrapped normal distribution. Again, it is nice to see the further interesting inequality on Bessel functions from Professor Lewis. Here as well as elsewhere it is the first step which counts.

Various important references (especially those given in my book) were not included to keep the paper to a reasonable size and I earnestly hope that this would not be regarded by various authors as belittling their important contributions in any way.

References in the Discussion

Abell, G. O. (1965). Clustering of galaxies. *Ann. Rev. Astron. & Astrophys.*, **3**, 1–12.

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

Batschelet, E., Hillman, D., Smolensky, M. and Halberg, F. (1973). Angular–linear correlation coefficient for rhythmometry and circannually changing human birth rates at different geographic latitudes. *Int. J. Chronobiol.*, **1**, 183–202.

Bergström, H. and Puri, M. L. (1975). Convergence and remainder terms in linear rank statistics. (To appear.)

Bingham, C. (1974). An antipodally symmetric distribution on the sphere. *Ann. Statist.*, **2**, 1201–1225.

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

Broadbent, B. H. (1931). A new X-ray technique and its application to Orthodontia. *Angle Orthodontist*, **1**(2), 45–66.

Bushra, E. (1948). Variations in the human facial pattern in norma lateralis. *Angle Orthodontist*, **8**(3, 4), 100–102.

Durbin, J. (1975). Kolmogorov–Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. *Biometrika*, **62**, 5–22.

Finch, P. D. (1973). Macroscopic descriptions of microscopic phenomena. *Bull. Austr. Math. Soc.*, **9**, 391–410.

Fraser, D. A. S. (1968). *The Structure of Inference*. New York: Wiley.

Gumbel, E. J. (1954). Applications of the circular normal distribution. *J. Amer. Statist. Ass.*, **49**, 267–297.

Gumbel, E. J., Greenwood, J. A. and Durand, D. A. (1953). The circular normal distribution: theory and tables. *J. Amer Statist. Ass.*, **48**, 131–152.

Heyer, H. K. W. (1972). Infinitely divisible probability measures on compact groups. In *Lectures in Operator Algebras* (Vol. 247 of *Lecture Notes in Mathematics*). Berlin: Springer-Verlag.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Phys. Rev.* A, **106**, 620–630.

—— (1963). Information theory and statistical mechanics. In *Statistical Physics*, pp. 81–218. Brandeis University Summer Institute Lectures in Theoretical Physics, 1963, Vol. 3. New York: W. A. Benjamin.

Jurečková, J. and Puri, M. L. (1975). Order of normal approximation for rank test statistics distribution. *Ann. Prob.* **3**, 526–534.

Lord, R. D. (1954a). The use of the Hankel transform in statistics. I. General theory and examples. *Biometrika*, **41**, 44–55.

—— (1954b). The use of the Hankel transform in statistics. II. Methods of computation. *Biometrika*, **41**, 344–350.

MacKenzie, J. K. (1957). The estimation of an orientation relationship. *Acta Crystallog.*, **10**, 61–62.

Marsden, B. G. (1972). *Catalogue of Cometary Orbits*. Smithsonian Astrophysical Laboratory.

Oort, J. H. (1950). The structure of the cloud of comets surrounding the solar system and a hypothesis concerning their origin. *Bull. Astron. Inst. Netherlands*, **2**, 91–110.

Plackett, R. L. (1966). Current trends in statistical inference. *J. R. Statist. Soc.* A, **129**, 249–267.

Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. New York: Wiley.

Rao, J. S. (1971). Bahadur efficiencies of some tests for uniformity on the circle. *Ann. Math. Statist.*, **43**, 468–479.

Rao, J. S. and Sethuraman, J. (1970). Pitman efficiencies of tests based on spacings. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.), pp. 405–415. Cambridge: University Press.

—— (1975). Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. *Ann. Statist.*, **3**, 299–313.

Stephens, M. A. (1969). Tests for the von Mises distribution. *Biometrika*, **56**, 149–160.
Tribus, N. (1962). The use of the maximum entropy estimate in the estimation of reliability. In *Recent Developments in Information and Decision Processes* (E. Machol and P. Gray, eds), pp. 102–140. New York: Macmillan.
Williams, E. J. (1963). A comparison of the direct and fiducial arguments in the estimation of a parameter. *J. R. Statist. Soc.* B, **25**, 95–99.
Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. London: Pion.