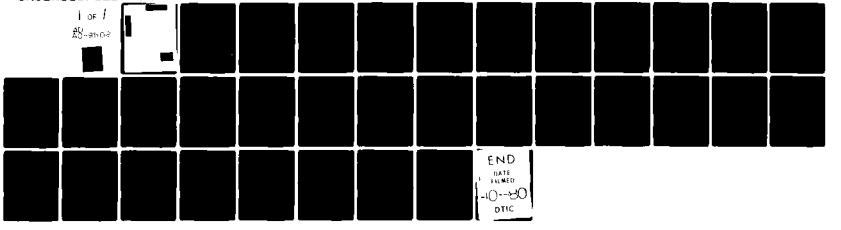
AD-A089 502 STANFORD UNIV CA DEPT OF STATISTICS F/G 12/1
THE VON MISES DISTRIBUTION IN P-DIMENSIONS WITH APPLICATIONS.(U)
AUG 80 M A STEPHENS N00014-76-C-0475
UNCLASSIFIED TR-290 NL



THE VON MISES DISTRIBUTION IN p-DIMENSIONS, WITH APPLICATIONS

By

Michael A. Stephens

TECHNICAL REPORT NO. 290

AUGUST 4, 1980

Prepared under Contract
N00014-76-C-0475 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. INTRODUCTION

The von Mises distribution, in 2-dimensions, and the Fisher distribution, in 3-dimensions, have been extensively used in recent years to describe directional data. In this paper we give the theory of the von Mises distribution in p-dimensions, and suggest some possible applications. Some results along these lines were given by Watson (1956) and by Watson and Williams (1956), but the first complete extension to p-dimensions was given in the author's Ph.D. thesis (Stephens, 1962a, Chapter 9). It was later issued as a technical report (Stephens, 1962b) and some results have since been reproduced by Mardia (1975) and by Degerine (1977); since these sources are relatively inaccessible, we begin with a summary of results, taken essentially from Stephens (1962a,b). In later sections, we develop some new techniques for the analysis of data, and illustrate with an application to data recorded as a set of continuous proportions.

2. THE VON MISES DISTRIBUTION IN p-DIMENSIONS

2.1. The von Mises Distribution.

A typical sample item is recorded as a unit vector from the center O to a point P on the surface of a hypersphere, of unit radius, in p-dimensions. A typical sample then consists of the points P_i, $i = 1,\ldots,N$, or equivalently the vectors OP_i. When $p = 2$, the points are on a circle, and when $p = 3$, they are on a sphere. The vectors can then denote directions, e.g. of

prevailing winds, flights of birds, or magnetization of rocks. Until now, it is in these context that the distribution has been extensively used. Let the unit vector OP , called v , have coordinates $x_1, x_2, \ldots, x_p$ in a suitable rectangular system. It will also be useful to use polar coordinates, consisting, in general, of the radius r (here $r = 1$), and angular coordinates $\theta_1, \theta_2, \ldots, \theta_{p-1}$. The relations between the two sets of coordinates are

$$\left.\begin{aligned} x_1 &= \cos\theta_1 , \\ x_j &= \cos\theta_j \sum_{i=1}^{j=1} \sin\theta_i , \quad (j = 2,\ldots,p-1); \\ x_p &= \prod_{i=1}^{p-1} \sin\theta_i . \end{aligned}\right\} \quad \begin{aligned} &0 \le \theta_j \le \pi \\ &(j = 1,\ldots,p-2); \\ &0 \le \theta_{p-1} \le \pi . \end{aligned}$$

The von Mises density is symmetrical around the modal vector OA ; for convenience in analyzing the distribution, we place this vector along $\theta_1 = 0$. The density per unit area on the hypersphere is then proportional to $\exp(k \cos\theta_1)$ where k is a concentration parameter. The joint density function of the θ_j is

$$f(\theta_1,\theta_2,\ldots,\theta_{p-1}) = C_p(k)\exp(k\cos\theta_1)\sin^{p-2}\theta_1\sin^{p-3}\theta_2\ldots\sin\theta_{p-2} \qquad (1)$$

over the range of θ_j . The constant term is

$$C_p(k) = \frac{k^q}{\{I_{(q)}(k)\}(2\pi)^{p/2}}$$

where q is written for p/2-1 and where $I_m(k)$ is the imaginary Bessel function of order m and argument k. When k is 0, $C_p(0)$ becomes $\Gamma(p/2)/(2\Pi)^{p/2}$ and the density is uniform over the unit hypersphere; the concentration around OA increases with k. With the density as described, x_1 is the component of OP on the modal vector and θ_1 is the angle between OP and the modal vector. An orthogonal transformation allows the density to be transformed to place OA along any chosen vector, but the general form is then very complicated.

2.2. Notation for vectors and related statistics.

In a suitable rectangular system, let vector w_i have components $x_{i1},x_{i2},\ldots,x_{ip}$, written $w_i = (x_{i1},x_{i2},\ldots,x_{ip})$. Two calculations involving vectors are often needed; the scalar product of two vectors and the length of a vector. The scalar product $s(w_i,w_j)$ of vectors w_i and w_j is defined as

$$s(w_i,w_j) = x_{i1}x_{j1} + x_{i1}x_{j2} + \ldots + x_{ip}x_{jp} \qquad (2)$$

and the length ℓ_i of w_i is given by $\ell_i^2 = s(w_i,w_i) = x_{i1}^2 + x_{i2}^2 + \ldots + x_{ip}^2$. A vector w_i is reduced to unit length by dividing by its length; its components are then $(x_{i1}/\ell_i, x_{i2}/\ell_i,\ldots,x_{ip}/\ell_i)$. The scalar product is also $s(w_i,w_j) = \ell_i\ell_j \cos\alpha_{ij}$ where α_{ij} is the smaller angle between vectors w_i and w_j. The values of $s(w_i,w_j)$ and of ℓ_i are not dependent on the coordinate system used. Vectors

and their lengths are often denoted by the same letter but with the vector printed boldface; for ease of printing we shall mostly not use this convention, i.e., vectors will not be written boldface, except to distinguish between a resultant vector $\underset{\sim}{R}$ and its length R.

Suppose a sample of N unit vectors is given, consisting of vectors $OP_i = v_i$, $i = 1,\ldots,N$; a typical vector v_i has components $(x_{i1},x_{i2},\ldots,x_{ip})$, and the polar coordinates are $(\theta_{i1},\theta_{i2},\ldots,\theta_{i(p-1)})$. The resultant, or vector sum, of the set of N vectors has components $X_1,X_2,\ldots,X_p$, where

$$X_j = \sum_{i=1}^{N} x_{ij} .$$

The resultant is denoted by $\underset{\sim}{R} = (X_1,X_2,\ldots,X_p)$ and its length by R; thus $R^2 = X_1^2 + X_2^2 + \ldots X_p^2$. The length of the component of $\underset{\sim}{R}$ on the modal vector OA, when this is known or hypothesized, is often used and will be denoted by X with no subscript. Suppose OA has unit length, and components $a_1,a_2,\ldots,a_p$; then

$$X = s(OA,\underset{\sim}{R}) = a_1X_1 + a_2X_2 + \ldots a_pX_p \qquad (3)$$

The value of X will also be independent of the coordinate system.

The statistics $\underset{\sim}{R}$, R and X are all important statistics for the analysis of a sample. For example, the maximum likelihood estimator (MLE) of the direction of the modal vector is the direction of $\underset{\sim}{R}$ and the MLE $\hat{k}$ of the concentration parameter k is given by the equation

$$\frac{I_{q+1}(\hat{k})}{I_q(\hat{k})} = \frac{R}{N}, \tag{4}$$

with $q = p/2-1$ as before. For k large this equation becomes

$$1 - \frac{p-1}{2\hat{k}} = \frac{R}{N} \quad \text{i.e.} \quad \hat{k} = \frac{N(p-1)}{2(N-R)}. \tag{5}$$

If OA is known, the component X replaces R in (4) and (5).

When several samples of unit vectors are given, questions arise whether they have the same modal vectors, the same concentration parameters, etc. Let the i-th group have modal vector OA_i, and concentration parameter k_i. Let v_{ij}, $j = 1,\ldots,N_i$, be the set of unit vectors in the i-th group, so that N_i is the number of vectors in the group, and let R_i be the length of the resultant vector $\underset{\sim}{R}_i$ of the group. Let $N = \Sigma_i N_i$, and let R be the length of the resultant $\underset{\sim}{R}$ of all the vectors treated as one large group.

2.3. Distributions of statistics X and R.

In the applications to be described, we shall mostly need approximations to the distributions of X and R which hold when k is large; however for completeness the exact distributions of these two statistics will be given. Suppose $J_q(t)$ and $I_q(t)$ are respectively the usual Bessel function and the imaginary Bessel function of order q, and let

$$M_1(R) = \int_0^\infty t^{q(1-N)} 2^{-q} J_q(Rt) \{J_q(G)\}^N t \, dt$$

and

$$M_2(X) = \int_0^\infty \{J_q(t)\}^N t^{-qN} \cos Xt \, dt .$$

Also, put

$$C_1 = \{k^q / I_q(k)\}^N \quad \text{and} \quad C_2 = \{\sqrt{\pi}\Gamma(q + 1/2)\}^{-1} ,$$

The densities of R and of X are (Stephens, 1962a)

$$f_1(R) = C_1 C_2 I_q(k) R^{p/2} M_1(R) \qquad 0 \le R \le N$$

$$f_2(X) = (C_1/\pi) e^{kx} M_2(X) . \qquad -N \le X \le N .$$

These densities involving Bessel functions simplify for odd values of p. Stephens (1962a,b,1967,1969a) discusses the densities for $p = 2$ and 3 in much greater detail and uses them to find tests for k. Stephens (1962a,b) also gives the conditional density of R given X in p-dimensions; this is independent of k as was earlier shown by Watson for $p = 2$ and 3, and so can be used for a test that the modal vector is along a given vector OA_0, when k is unknown. The tests for $p = 2$ and 3 are described in detail in Mardia (1972) and in

Section 9 of Volume 2 of Biometrika Tables for Statisticians; see also Stephens (1975) for a test for the modal vector, when k is known. Other exact results for p-dimensions are given in Stephens (1962a,b).

When $k = 0$, the densities above simplify considerably. The vectors are now uniformly distributed on the hypersphere, and the length R is the length of the final displacement of N unit steps in a random walk from O in p-dimensions. For this reason it has attracted attention long before the use of the von Mises and Fisher distributions. For the exact densities, and for further references, see Stephens (1964, 1969b). When $k = 0$, a simple approximation exists for the distribution of R. Suppose $Z = pR^2/N$ for large N; Z is approximately χ^2 distributed with p degrees of freedom. This result can be used to provide a test for uniformity based on R; the test has come to be known as Rayleigh's test. A recent paper (Prentice, 1978) surveys tests for uniformity in p-dimensions.

2.4. Properties of the von Mises density with large concentration parameter k.

When k is large, there are useful approximations concerning the densities of R and of X. There is then a high probability that θ_1 will be small and so $\cos \theta_1 \approx 1 - \theta_1^2/2$ and $\sin \theta_1 \approx \theta_1$. The density of θ_1 becomes

$$f(\theta_1) \simeq C_p(k)\exp(k)\exp(-k\theta_1^2/2)\theta_1^{p-1}, \quad 0 \le \theta_1 \le \pi , \qquad (6)$$

so that the quantity $k\theta_1^2$ has approximately a χ^2 distribution with $p-1$ degrees of freedom. Since $x_1 = \cos\theta_1$, we write

$$k\theta_1^2 = 2k(1 - \cos\theta_1) = 2k(1 - x_1) \approx \chi^2_{p-1} . \qquad (7)$$

Because of the symmetry around the modal vector, the other coordinates x_j, $j \ge 2$, have identical distributions; for large k these are approximately normal with mean 0 and variance $1/k$. For a tightly clustered sample of vectors v_i we expect $\underset{\sim}{R}$ to point fairly accurately along the modal vector OA, and the length of R to be relatively large. If OA is known, the projection X of $\underset{\sim}{R}$ on OA will also be large. Then clearly $N - X$ and $N - R$ are both measures of the dispersion of the set of vectors. For large k, we have from (7), with $r = p-1$, $\Sigma_i\ 2k(1 - x_{i1}) \approx \chi^2_{Nr}$; this gives $2k(N-X) \approx \chi^2_{Nr}$. Further, the distributional results for x_j, $j = 2,\ldots,p$, lead to the approximate distribution $k(R^2 - X^2)/N \approx \chi_r^2$; since $R \approx X \approx N$, this becomes $2k(R-X) \approx \chi_r^2$.

2.5. Tests for the modal vector and for k.

Watson (1956) and Watson and Williams (1956) have used these identities to devise a technique of analysis for large k, which is analogous to the usual one way analysis of variance for continuous variables. Watson writes the identity

$$2k(N-X) = 2k(N-R) + 2k(R-X) , \qquad (8)$$

which, by analogy with the analysis of variance, becomes, in p-dimensions

$$\chi^2_{Nr} = \chi^2_{(N-1)r} + \chi^2_r . \qquad (9)$$

This leads to the approximation for the statistic Z_1

$$Z_1 = \frac{(N-1)(R-X)}{N-R} \approx F_{r,(N-1)r} .$$

where $F_{s,t}$ is the F distribution with s and t degrees of freedom.

Watson suggested the use of statistic Z_1 in two or three dimensions, to examine whether a given vector OA_o is the modal vector. On the null hypothesis, both R and X will be known, and the hypothesis will be rejected for large values of Z_1, indicating that X is much smaller than R; in that case R does not point in the direction of the vector OA_o. Tests of the null hypothesis that k is a given value k_0 will be based on the χ^2 approximations given above for $2k(N-X)$, when OA is known, or for $2k(N-R)$, when OA is not known. Stephens (1967, 1969a) examined these tests for $p = 2$ and 3 and found them to be very good even for quite low values of k. They will certainly be valid for the large values of k which arise in the applications below.

2.6. Comparison of several modal vectors.

Suppose s different samples of unit vectors are given and we wish to test whether all the samples come from populations with the same modal vector, assuming they have the same value of k. On the null hypothesis we again use the χ^2 approximation for $2k(N-R)$, and apply this result to the individual samples as well as the entire group taken as a whole. We write the following identity

$$2k(N-R) = 2k(N_1-R_1) + 2k(N_2-R_2) + \ldots + 2k(N_s-R_s) + 2k(R_1+R_2+\ldots+R_s-R) \quad (10)$$

and, again by analogy with the analysis of variance we obtain

$$2k\{\Sigma_i(N_i-R_i)\} \approx \chi^2_{(N-s)r} \quad \text{and} \quad 2k(R_1+R_2+\ldots+R_s-R) \approx \chi^2_{(s-1)r}\,, \qquad (11)$$

where again $r = p-1$; hence the statistic

$$Z_2 = \frac{(N-s)(\Sigma_i R_i - R)}{(s-1)(N-\Sigma_i R_i)}$$

will have approximately the F distribution with $(s-1)r$ and $(N-s)r$ degrees of freedom. Therefore to test whether the different groups have the same modal vector, the statistic Z_2 is calculated and compared with this F distribution. Large values of Z_2 will be significant, indicating that the $\underset{\sim}{R}_i$ vectors point in different directions.

The above analysis is essentially a one-way analysis of variance which can be set up in the usual tabular form;

Variance Components		d.f.
Between groups:	$\Sigma_i R_i - R$	$(s-1)r$
Within groups :	$N - \Sigma_i R_i$	$(N-s)r$
Total :	$N-R$	$(N-1)r$.

Note that throughout the table $2k$ has been omitted before the variance components; since only ratios will be used for tests this does not effect the calculations. This is analogous to omitting σ^2 in the terms of squares of an ANOVA table. In a later section we shall give an extension to the above analysis which can be used when, for example, the groups of vectors can be classified according to two criteria.

2.7. Tests for constant k.

In the variance component analysis, k is assumed to be constant for each group, analogous to the assumption of constant variance in an analysis of variance table. For large k this assumption can be tested as follows. Suppose in general, there are s groups of vectors, and it is required to test for constant k for all groups. For group i, calculate $Q_i = N_i - R_i$ and $q_i = (N_i - 1)r = (N_i - 1)(p-1)$; let $T = \Sigma_i Q_i$ and $t = \Sigma_i q_i$. Calculate

$$Z_3 = t \ln T - \Sigma_i q_i \ln Q_i - t \ln t + \Sigma_i q_i \ln q_i$$

and $$C = 1 + \frac{1}{3(s-1)}(\Sigma_i q_i^{-1} - 1/t);$$

finally let $Z_3' = Z_3/C$. The hypothesis of constant k is rejected if Z_3' is significantly large compared with the χ^2 distribution with $s-1$ degrees of freedom. Note that if Z_3 is not significant, Z_3' will not be significant, since C is greater than 1. This test is based on Bartlett's test for homogeneity of variances. Other tests for the same purpose (e.g. those in section 16 of Biometrika Tables for Statisticians, Vol. 1) can be adapted in a similar way.

3. THE ANALYSIS OF CONTINUOUS PROPORTIONS.

3.1. Applications of the von Mises distribution in p dimensions.

The von Mises distribution in p-dimensions can be a useful tool of analysis of multivariate data, where the variables are subject to a constraint which corresponds to the fact that all vectors in the

von Mises sample have unit lengths. An example is the analysis of continuous proportions, where the typical vector v has components $x_1, x_2, \ldots, x_p$ which are proportions of a continuum, e.g. time or volume or mass. Suppose, for example, a subject records the proportion of his day spent in p different activities; let these proportions be $\pi_1, \pi_2, \ldots, \pi_p$, and let $x_i = \sqrt{\pi_i}$. Since $\Sigma_i x_i^2 = \Sigma_i \pi_i = 1$, a typical activity pattern for a subject can be recorded as a unit vector v with components $x_1, x_2, \ldots, x_p$. A group of subjects will be denoted by a set of points on the unit hypersphere, and the population might well be represented by a von Mises density around a central vector. This application occurred to the author when at a Conference some years ago such a data set was being discussed, gathered from the activity patterns of over 200 students at Reading University, and covering over 60 activities. The students had been asked to compile a diary for one week, giving their activities for every fifteen minute period. Thus the vector of activity pattern was available for each student, and in addition a transition matrix was available giving the probability of moving from activity i to activity j. The example in Section 3.3 examines a similar data set, but much simplified.

3.2. The von Mises distribution for continuous proportions.

The von Mises distribution was suggested above as a useful model for a cluster of vectors around a modal direction. For continuous proportions further analytic justification can be given. Suppose for each i the proportion π_i recorded by a subject is a random variable $\pi_i = a_i + \epsilon_i$; a_i is the modal value for the population, and ϵ_i a random fluctuation with a normal distribution with mean 0 and a small variance σ_i^2. Since $\Sigma_i \pi_i = 1$, and $\Sigma_i a_i = 1$, the constraint $\Sigma_i \epsilon_i = 0$ is imposed on the ϵ_i, which are otherwise assumed independent. The

typical component of the unit vector v for the subject is $x_i = \sqrt{\pi_i}$, and the modal vector OA has components $\sqrt{a_i}$, $i = 1,\ldots,p$. The scalar product $s(OA, v)$ gives $\cos\theta_1$; thus

$$\cos\theta_1 = \Sigma_1 x_i\sqrt{a_i} \approx \Sigma_i[a_i + \epsilon_i/(2\sqrt{a_i}) - \epsilon_i^2/\{8(a_i^{3/2})\}]$$

using the binomial expansion of $x_i = \sqrt{(a_i + \epsilon_i)}$. In the simplest case, suppose that all a_i are equal to $1/p$, and that all $\sigma_i^2 = \sigma^2$, a constant. Then $\cos\theta_1$ becomes

$$\cos\theta_1 \approx 1 - p^{3/2}\Sigma_i\epsilon_i^2/8 ;$$

and since $\cos\theta_1 \approx 1 - \theta_1^2/2$ we have $k_1\theta_1^2 \approx \Sigma_i\epsilon_i^2$, where $k_1 = 4/p^{3/2}$. However, $\Sigma_i\epsilon_i^2/\sigma^2$ has the χ^2_{p-1} distribution, one degree of freedom being lost since $\Sigma_i\epsilon_i = 0$. Hence $k\theta_1^2 \approx \chi^2_{p-1}$, where $k = k_1/\sigma^2 = 4\sigma^2/p^{3/2}$. By symmetry, the vector component of v which is not along OA will be uniformly distributed; this result, together with $k\theta_1^2 \approx \chi^2_{p-1}$, indicates that the vector v has the von Mises distribution, at least to a good approximation. When a_i and σ_i^2 are not all equal, the approximation still holds well provided we can take $\Sigma_i\epsilon_i/\sqrt{a_i} \approx 0$ and $\Sigma_i\epsilon_i^2/a_i^{3/2} \approx k_2\chi^2_{p-1}$ where k_2 is a constant. These approximations will be good if $\sigma_i^2 \sim a_i^{3/2}$, a reasonable model, and the von Mises distribution will be probably quite robust for most situations provided the a_i are not too different in value.

3.3. Example.

The data to be analyzed concerns the proportions of time spent in various activities by 130 students at Simon Fraser University. The activities were classified in 8 ways: sleeping, attending lectures, studying, socializing, travelling, family activities, meals, and personal activities and the students were asked to record their activities for one day only. Thus the sample does not represent the overall activity pattern,

but it is used here as an illustration of the general methodology. The complete data set is available from the author.

The first step in analyzing the data is to convert the proportions to x-coordinates. For student i let π_{ij} be the proportion of time spent in activity j, $i = 1,\ldots,N$, $j = 1,\ldots,p$, and let $x_{ij} = \sqrt{\pi_{ij}}$. From the x_{ij}, the component X_j of the resultant $\underset{\sim}{R}$ and the length R of $\underset{\sim}{R}$, are calculated as described in Section 2.2.

We first use the analysis of variance technique to examine whether there appears to be a difference in activity patterns between men and women. The data set is divided into two groups; group 1 for women and group 2 for men. The results for the two groups are given in Table 1. From the ANOVA table of part (a), the value of statistic Z_2 is 0.61, which is not significant at $\alpha = .10$ when compared with the $F_{s,t}$ distribution with $s = 7$ and $t = 896$. With such a large value for t, the percentage point of F, at upper level α, is excellently approximated by $\chi^2_s(\alpha)/s$, where $\chi^2_s(\alpha)$ is the upper tail percentage point of χ^2_s at level α. Here the value required would be $\chi^2_7(\alpha)/7$; for $\alpha = .10$, this is 1.909. Therefore there appears to be no reason to suppose the activity pattern is different between men and women students.

The data were next examined to see if there was a difference is patterns explained by the style of living arrangements of the students. The living styles were classified as: 1, college residence; 2, marriage or marriage style; 3, other, e.g. at home, sharing an apartment, renting a room, etc. The results are given in part (b) of Table 1. The value of Z_2 is now 3.3, to be compared with the $F_{s,t}$ table with $s = 14$, $t = 889$. At the 1%

level, the value of $\chi^2_{14}(.01)/14$ is 2.08, so that Z_2 is significant at this level, and we conclude that there is a difference between groups classified by living styles. The data will be examined in greater detail after some new techniques have been introduced.

4. NEW TECHNIQUES

4.1. Two-way analysis of variance.

In this section we give some new techniques for the examination of data. First the variance component analysis of the preceding section is extended to a two-way layout, and the student data is again used for illustration. In section 4.3, we discuss goodness-of-fit to the von Mises distribution, and in section 4.4 some techniques of clustering and correlation are briefly mentioned.

Suppose the sample items (for example, students) are classified in two ways: by a main classification 1 with I groups, indexed by $i = 1,\ldots,I$, and by classification 2 with J_i groups within group i of classification 1. When a student falls into group i of classification 1 and group j of classification 2, the associated vector of activity proportions will be placed in cell (i,j) in row i, column j, of a two way table. Extending our previous notation, we write v_{ijk} for the k-th vector in cell (i,j). Let N_{ij} be the number of vectors in cell (i,j), and let R_{ij} be the length of the resultant in this cell. Let N be the total number of vectors and let N^* be the number of non-empty cells. Let $R_{i.}$ be the length of the resultant of all vectors in row i, i.e., of the vectors for all students in group i of the first classification, and suppose $R_{..}$ is the length of the resultant

of all the vectors. As before, write r = p-1. A table may be constructed as in Table 2(a). By extension of the previous analysis, we write the following identity

$$2k(N-R_{..}) = 2k\sum_{j=1}^{J_1}(N_{1j}-R_{1j}) + 2k\sum_{j=1}^{J_2}(N_{2j}-R_{2j}) + \cdots + 2k\sum_{j=1}^{J_I}(N_{Ij}-R_{Ij})$$
$$+ 2k\{\sum_{j=1}^{J_1} R_{1j}) - R_{1.}\} + \cdots + 2k\{\sum_{j=1}^{J_I} R_{Ij}) - R_{I.}\} + 2k(\sum_{i=1}^{I} R_{i.}-R_{..});$$

collecting terms, we have

$$2k(N-R_{..}) = 2k\sum_{i=1}^{I}\sum_{j=1}^{J_i}(N_{ij}-R_{ij}) + 2k\{\sum_{j=1}^{J_1}(R_{1j}-R_{1.})\} + \cdots + 2k\{\sum_{j=1}^{J_I}(R_{Ij}-R_{I.})\}$$
$$+ 2k(\sum_{i=1}^{I} R_{i.}-R_{..})$$

with corresponding distributions, for large k :

$$\chi^2_{(N-1)r} = \chi^2_{(N-N^*)r} + \chi^2_{(J_1-1)r} + \cdots + \chi^2_{(J_I-1)r} + \chi^2_{(I-1)r}.$$

The terms may be arranged in a variance component table as in Table 2(b). A final column ("Mean Component") may be added, giving the value of the variance component divided by its degree of freedom. The table allows us to examine differences between rows, or differences between columns within any one row; thus the analysis will be similar to what is usually called a nested analysis of variance. To test the null hypothesis H_0: that there is no difference between rows, we calculate the quotient

$$Z_4 = \frac{(N-N^*)(\sum_{i=1}^{I} R_{i.}-R_{..})}{(I-1)(N-\sum_{i-1}^{I}\sum_{j=1}^{J_i} R_{ij})} \qquad (12)$$

which, on H_0, has an F-distribution with $(I-1)r$ and $(N-N^*)r$ degrees of freedom. The null hypothesis is rejected for a significantly

large value of Z_4. Similarly, to test the null hypothesis H_o: there is no difference between columns within row i, the quotient

$$Z_5 = \frac{(N-N^*)\left(\sum_{j=1}^{J_i} R_{ij} - R_{i.}\right)}{(J-1)\left(N - \sum_{i=1}^{I}\sum_{j=1}^{J_i} R_{ij}\right)} \qquad (13)$$

is calculated. On H_o, Z_5 has an F-distribution with $(J-1)r$ and $(N-1)r$ degrees of freedom, and H_o should be rejected for a significantly large value of Z_5.

4.2. Example.

We continue with the example already begun in the previous section. The original sample is now subdivided by both sex and living arrangements, making a total of six cells. Table 3 shows the sample size of each cell and the resultant length in each cell. Also shown in the table are the resultant length for each row, i.e., for the males and for the females, and for each column, i.e., for the three styles of living arrangements. Other relevant statistics are also given for use with Table 2(b). Tables 3(a) and 3(b) give the tables for two analyses. The column MC gives the mean component, i.e., variance component divided by its degrees of freedom.

In analysis 1, we first test if there appears to be a difference between activity patterns for men and women, using the variance components for "between sexes" and "within groups". The test statistic $Z_4 = 0.00857/0.01376 = 0.62$ and this is clearly not significant when compared to the $F_{7,868}$ distribution. The next

test is for difference between living styles for women only. The statistic $Z_5 = 0.01486/0.01376 = 1.08$, and this is also not significant. The test statistic for difference between living styles for men is $Z_5 = 0.0421/0.01376 = 3.06$; this is significant compared with $F_{14,868} (\approx \chi^2_{14}/14)$ at the $\alpha = 0.005$ level. Thus the previously noted difference between activity patterns for different living styles has been narrowed down to a difference for men.

In analysis 2, the difference between activity patterns for different living styles again shows up in the corresponding Z_4 statistic, $Z_4 = 0.043/0.01376 = 3.13$; but for a difference between sexes within each living style there is no significant Z_5 statistic, confirming the results already found. The results of the more detailed two-way analysis are consistent with each other and sharpen the conclusions gained from the one-way analysis.

Example of the test for constant k. For the cells in Table 3 the values of $\hat{k}$ across the top row are 32.44, 36.63, 38.70, and those across the bottom row are 51.39, 36.53, 37.99. The test statistic Z_3 of Section 2.7 has value $Z_3 = 3.44$, not significant when compared with χ^2_5 , so that the hypothesis of constant k can be maintained.

Example 2. In the above example, the classification within each row was the same, with $J_i = 3$ for both men and women. However, with a nested model, J_i can of course be different for each row. We illustrate with a second example with data kindly provided by Dr. Charles Jones of the Dept. of Sociology, McMaster University. N = 232 respondents

were each asked to rate 35 ethnic or religious groups in Canada, by the criterion of social standing, using a 9-point scale. By a scaling device, the 35-vector of replies was reduced to a vector in 3 dimensions. These vectors could have negative components, in contrast to the continuous proportions data. Thus the basic data set consists of 232 vectors in 3 dimensions; these were divided into 8 groups according to ethnic origin of the respondents. The eight groups have been put into four rows indicated by the nature of the Canadian population, and two of the rows have been subclassified. The data is given in Table 4, with the Analysis of Variance. For a test of significance between rows, the test statistic is $Z_2 = 1.28/.063 = 20.31$ and is highly significant when compared with $\chi^2_6/6$ at the 0.005 level. The test statistics for differences in groups within rows 2 and 4 are respectively $Z_5 = .0306/.063$ and $Z_5 = .0419/.063$ and are clearly far from significant.

4.3. Goodness-of-Fit.

The analysis described so far assumes that the observations come from the p-dimensional von Mises distribution. In order to test this assumption, we use two of the distributional results described in Section 2. The results are the distribution of the angle θ_1 between a typical vector v and the modal vector OA, and the distribution of the component of v, say y, at right angles to the

modal vector. Since the modal vector is not precisely known, θ_1 must be estimated by ϕ_1, the angle between v and the resultant $\underset{\sim}{R}$ of the sample. The set of angles ϕ_1 is tested to come from the density for θ_1 given in equation (6), using the usual Pearson χ^2 test. Since ϕ_1 replaces θ_1, it is difficult to determine exactly the degrees of freedom (though k-2 might be indicated for a test involving k cells) and more examination needs to be made of the test in this case. However the statistic will be a helpful guide to the fit of θ_1. The components y_i, i = 1,2,...,N, at right angles to $\underset{\sim}{R}$ should be uniform on the hypersphere of dimension p-1, and the hypothesis that this is so can be tested in many ways (see e.g. Prentice, 1978). For robustness of the analysis, it will be important that they are not clustered around a single mode and for this purpose the Rayleigh test is indicated. The vector component y_i is found as follows. Let $u = \underset{\sim}{R}/R$ be the unit vector along $\underset{\sim}{R}$ and let c_i be the scalar product of v_i and $\underset{\sim}{R}$ (see section 2.2 for these calculations). The component of v_i along $\underset{\sim}{R}$ is then uc_i and y_i, the component at right angles, is $y_i = v_i - uc$. To apply the Rayleigh test, we then reduce each vector y_i to unit length as shown in Section 2.2; let the unit vector be $\underset{\sim}{z}_i$, with components $z_{i1}, z_{i2}, \ldots, z_{ip}$; the resultant Z of this set of vectors has components $Z_1, Z_2, \ldots, Z_p$, where $Z_j = \Sigma_i z_{ij}$, and the length Z of $\underset{\sim}{Z}$ is given by $Z^2 = Z_1^2 + Z_2^2 + \ldots + Z_p^2$. On the null hypothesis that the vectors $\underset{\sim}{z}_i$ are uniform in the p-1 dimensional subspace, the test statistic $T = (p-1)Z^2/N$ is asymptotically distributed as χ^2 with p-1 degrees of freedom; the hypothesis of uniformity is rejected if T is larger than $\chi^2_{p-1}(\alpha)$. The two tests above, taken together, provide a good omnibus test that the original sample of vectors v_i comes from the von Mises distribution. The

distributional tests are applied to each cell of the two way analysis of variance table described above, analogous to applying tests for normality in the usual analysis of variance table. If, of course, there were to be a significant difference in the modal vector, say between men and women, a test for the von Mises distribution applied to the complete sample, including both men and women, might well be rejected. On the other hand, if each group is found to have a von Mises distribution, the test for common modal vector can be applied, and if accepted, the overall sample should have a von Mises distribution. For the student data in Table 3, all the groups gave far from significant values for the test statistics for goodness-of-fit, so that the von Mises distribution appears to fit the data well.

4.4. Clustering and correlation.

The scalar product $s_{ij} = s(v_i, v_j)$ is a convenient measure of the closeness of the vectors v_i and v_j; s_{ij} takes values between -1 and 1. We can call s_{ij} a proximity measure, and the matrix S, with entries s_{ij}, a proximity matrix. A cluster may then be defined as containing all points for which s_{ij} is greater than r_0, for a suitable r_0, or by using some similar algorithm. This proximity measure has been used on some economic data in an M.Sc. thesis (Holguin, 1980); data on the proportions of different wood products produced by Canadian and U.S. lumber companies were provided by Dr. R. Schwindt of this University and the companies were examined both for differences between groups and also to find clusters. In a second data set, taken from a U.N. publication, countries were clustered according to the proportions of certain staple foods in the national diet.

When two sets of vectors can be logically paired, the techniques of correlation developed by Stephens (1979) may be useful. These have not been illustrated here because they did not appear to be applicable to the student data.

5. CONCLUDING REMARKS

(a) It is hoped that the methodology developed above will be useful for the analysis of data which it is convenient to record as a set of unit vectors; the example illustrated is that of continuous proportions. The special feature of the technique is that it incorporates the constraint expressed by $\Sigma_i \pi_i = 1$ in a natural way. Note that the order of proportions, e.g. the order of labelling student activities, does not affect the analysis. For reasons of space, only one such example has been discussed in detail but analyses similar to the above could be applied to the proportions of different minerals in an ore deposit, calculated by volume or by mass, or the proportions of different products in the total output of a company, the proportions of the area of a city used for different purposes, etc. Several examples on these lines have been suggested to the author, and they will be followed up in later case studies.

(b) For proportions analysis, the components of sample vectors are naturally all positive, so that the vectors in p-dimensions are tightly clustered; hence we obtain the high k values seen with the student data, and the analysis of variance technique works very well. Nevertheless, the robustness of these methods needs further exploration. For example, the effect must be determined of specifying too many components for v_i, e.g. too many student activities, or too many

subdivisions of output of a lumber company, so that some components are zero for many of the sample vectors. The Rayleigh test should detect if the effective reduction in dimensions which this produces has a strong influence.

(c) The methodology described is not primarily intended for proportions which come from counted data, such as proportions of a sample of voters expressing different political preferences. Proportions of this type are usually examined for homogeneity in contingency tables (though not of course using the proportions themselves), and each group of voters represents an independent sample. Even if they were expressed as unit vectors, the model behind the counted data is such that it would not necessarily be appropriate to regard the vectors as from a von Mises distribution. However, if the samples of voters could be naturally grouped, say by regions, and especially if each voter sample were of the same size, some of the above techniques might be useful in exploring the data. For example the scalar product proximity measure s_{ij} discussed in Section 4.4 could be a practical measure of the similarity of two patterns of voter preferences.

The pitfalls of proportions, which have often been emphasized in discussions of contingency tables, should again be stressed. The author has seen, for example, a proposal to analyze proportions of land use in major cities, one of the examples briefly mentioned above, using contingency tables, on the grounds that the numbers were"counted data"; they had been obtained by superimposing a fine grid on the map and counting squares. It is not always easy, especially for the applied worker with only a limited knowledge of a contingency table model, to distinguish between the two types of data when presented as proportions.

(d) It may be seen that there is no shortage of interesting examples where the above methodology may be applied. The best test of its effectiveness will be in these practical applications and especially in comparisons with other techniques of analysis. In this way any difficulties of application, especially concerning robustness of the methods, will hopefully come to light. A number of such comparisons have been started, and it is hoped in a later paper to report on several case studies in different fields.

This work was supported by the National Research Council of Canada (now the National Science and Engineering Research Council), and also by the U.S. Office of Naval Research, and both agencies are thanked for their support. The author also is grateful to Professor G.S. Watson for his guidance in the early stages of this work, and to Mr. J. Holguin for recent helpful conversations and for assistance with the computations.

TABLE 1

Analysis of activity pattern of 130 students.

Overall resultant length (130 vectors) R = 117.199.

Part (a) Analysis of differences between men and women.

Group	N_i	R_i	$\hat{k}_i$
1 (Women)	56	50.504	35.67
2 (Men)	74	66.754	37.74
Total	130	117.258	

ANOVA Table

Variance Component		Value	d.f.	Mean Component
Between groups	$\Sigma_i R_i - R$	0.059	7	.0086
Within groups	$N - \Sigma_i R_i$	12.742	896	.0142
Total	N - R	12.801	903	

Part (b) Analysis of differences by living style.

Group	N_i	R_i	$\hat{k}_i$
1	18	16.317	37.43
2	28	25.190	34.87
3	84	76.293	38.13
Total	130	117.801	

ANOVA Table.

Variance Component		Value	d.f.	Mean Component
Between groups	$\Sigma_i R_i - R$	0.602	14	.043
Within groups	$N - \Sigma_i R_i$	12.199	889	.0137
	N - R	12.801	903	

TABLE 2

Two way analysis of variance for resultant vectors.

(a) Table of resultants.

Classification 2

(Columns)

Classification 1 (rows)	1	2	3	4	. . .	Vector Total
1	R_{11}	R_{12}	R_{13}	R_{14}		$R_{1.}$
2	R_{21}	R_{22}	R_{23}			$R_{2.}$
I	R_{I1}					$R_{I.}$
						$R_{..}$

Note that the vector total at the end of a row or column is not the arithmetic total of the entries of that row or column.

(b) ANOVA Table.

Variance Component	Value	d.f.
Between rows	$\sum_{i=1}^{I} R_{i.} - R_{..}$	$(I-1)r$
Between cols. within row 1	$\sum_{j=1}^{J_1} R_{1j} - R_{1.}$	$(J_1-1)r$
⋮		
Between cols. within row I	$\sum_{j=1}^{J_I} R_{Ij} - R_{I.}$	$(J_I-1)r$
Within groups	$N - \sum_{i=1}^{I} \sum_{j=1}^{J_i} R_{ij}$	$(N-N^*)r$
Total	$N-R_{..}$	$(N-1)r$

TABLE 3

(a) Results for 130 students classified by sex and by three living styles.

Each cell shows the number of students and the resultant lengths for the group. $R_{i.}$ = length of resultant of all students in row i , and $R_{.j}$ is the resultant length for all students in column j . $R_{..}$ is the length of the resultant of all 130 vectors.

Living style	1	2	3	$R_{i.}$
Sex F	9 8.029	13 11.758	34 30.925	50.504
M	9 8.387	15 13.563	50 45.394	66.754
$R_{.j}$	16.317	25.190	76.293	

$\Sigma_j R_{1j} = 50.712$ $\Sigma_i R_{i1} = 16.416$ $\Sigma_j R_{.j} = 117.800$

$\Sigma_j R_{2j} = 67.344$ $\Sigma_i R_{i2} = 25.221$ $\Sigma_i R_{i.} = 117.258$

$\Sigma_{ij} R_{ij} = 118.058$ $\Sigma_i R_{i3} = 76.319$ $R_{..} = 117.198$

(b) ANOVA Table for analysis 1.

Variance Component		Value	d.f.	M.C.	Test Statistic
Between sexes	$\Sigma R_{i.} - R_{..}$	0.060	7	.0086	.62
Between styles, F	$\Sigma R_{1j} - R_{1.}$	0.208	14	.0149	1.08
Between styles, M	$\Sigma R_{2j} - R_{2.}$	0.590	14	.0421	3.06
Within groups	$N - \Sigma_{ij} R_{ij}$	11.942	868	.0138	

(c) ANOVA Table for analysis 2.

Variance Component		Value	d.f.	M.C.	Test Statistic
Between styles	$\Sigma R_{\cdot j} - R_{\cdot\cdot}$	0.602	14	.043	3.13
Between sexes, Style 1	$\Sigma R_{i1} - R_{\cdot 1}$	0.099	7	.014	1.02
Between sexes, Style 2	$\Sigma R_{i2} - R_{\cdot 2}$	0.131	7	.019	1.36
Between Sexes, Style 3	$\Sigma R_{i3} - R_{\cdot 3}$	0.026	7	.0037	.270
Within groups	$N - \Sigma_{ij} R_{ij}$	11.942	868	.0138	

TABLE 4

Vectors determined from sociological ratings.

Each cell shows the number of respondents and the resultant length for the group.

Respondents				
Canadian	33 28.59			
	English	Irish	Scots	$R_{2.} = 91.42$
British	62 55.19	15 12.86	26 23.49	$N_{2.} = 103$
French	41 32.75			
	German	Russian	Others	$R_{4.} = 50.72$
Others	9 8.25	5 4.58	41 38.06	$N_{4.} = 55$

$R_{..} = 195.79$

ANOVA table			df	M.C.	Test Statistic
Between rows	203.48 - 195.79 = 7.69		6	1.28	20.31
Within row 1			—		
row 2	91.54 - 91.42 = .12	$r(J_2-1)$	= 4	.03	.49
row 3			—		
row 4	50.89 - 50.72 = .17	$r(J_4-1)$	= 4	.04	.67
Within groups	232 - 203.77 = 28.23	2(232-8)	= 448	.063	

REFERENCES

DEGERINE, S. (1977). Abstract, Proceedings of the European Meeting of Statisticians, Grenoble.

HOLGUIN, J. (1980). The application of directional methods in p-dimensions. M.Sc. thesis, Department of Mathematics, Simon Fraser University.

MARDIA, K.V. (1972). Statistics of Directional Data. London: Academic Press.

MARDIA, K.V. (1975). Distribution Theory for the von Mises-Fisher Distribution and its Application. Statistical Distributions in Scientific Work. (G.P. Patil et al. (eds.) 1, 113-130. D. Reidel: Dordrecht, Holland.

PRENTICE, M.J. (1978). On invariant tests of uniformity for directions and orientation. Ann. Statist. 6, 169-176.

STEPHENS, M.A. (1962a). The Statistics of Directions: The von Mises and Fisher Distributions. Ph.D. Thesis, Mathematics Department, University of Toronto.

STEPHENS, M.A. (1962b). The von Mises and Fisher distributions in higher dimensions. Technical Report, Dept. of Statistics, The John Hopkins University.

STEPHENS, M.A. (1964). The testing of unit vectors for randomness. J. Am. Statist. Assoc., 59, 160-167.

STEPHENS, M.A. (1967). Tests for the dispersion and for the modal vector of a distribution on a sphere. Biometrika, 54, 211-223.

STEPHENS, M.A. (1969a). Tests for the von Mises distribution. Biometrika, 56, 149-160.

STEPHENS, M.A. (1969b). Tests for randomness of directions against two circular alternatives. J. Am. Statist. Assoc., 64, 280-289.

STEPHENS, M.A. (1975). A new test for the modal vector for the Fisher distribution. Biometrika, 62, 171-174.

WATSON, G.S. (1956). Analysis of dispersion on a sphere. Mon. Not. R. Astr. Soc.: Geophys. Suppl. 7, 153-9.

WATSON, G.S. & WILLIAMS, E.J. (1956). On the construction of significance tests on the circle and the sphere. Biometrika. 43, 344-352.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(19) 34

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 290	2. GOVT ACCESSION NO.: AD-A089502
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): THE VON MISES DISTRIBUTION IN p-DIMENSIONS WITH APPLICATIONS.	5. TYPE OF REPORT & PERIOD COVERED: (9) TECHNICAL REPORT.
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): (10) MICHAEL A. STEPHENS (15)	8. CONTRACT OR GRANT NUMBER(s): N00014-76-C-0475
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, Stanford University, Stanford, CA 94305	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-042 267
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Statistics and Probability Program Code 436, Arlington, Virginia 22217 (11)	12. REPORT DATE: AUG 1980
	13. NUMBER OF PAGES: 30
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office): (14) TR 290	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

von Mises distribution; directional data; analysis of p-dimensional data; analysis of continuous proportions; multivariate data analysis.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE REVERSE SIDE.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

332580

290

THE VON MISES DISTRIBUTION IN p-DIMENSIONS, WITH APPLICATIONS

The theory of the von Mises distribution in p-dimensions is summarized, and it is shown how one or more sets of unit vectors, with p components, can be analyzed using this distribution. In particular, some extensions are made on the analysis of variance techniques first developed by Watson for two and three dimensions. An application is given to the analysis of continuous proportions for which the techniques supply a good methodology. Other examples are also briefly discussed.