The Annals of Statistics
1985, Vol. 13, No. 2, 838–844

BOOK REVIEW

GEOFFREY S. WATSON, *Statistics on Spheres,* Wiley-Interscience, 1983, $x + 238$ pages, $23.50.

REVIEW BY CHRISTOPHER BINGHAM

University of Minnesota

This book is aptly titled. It is concerned with problems arising in the description and analysis of data that consist of points on the $(q-1)$-dimensional surface $\Omega_q = \{\mathbf{u} \| \mathbf{u} \in R_q, \|\mathbf{u}\| = 1\}$ of the unit sphere in q-dimensional space R_q. Such data arise primarily from measurements of direction or orientation in the plane ($q = 2$) or space ($q = 3$). A "direction" may be defined to be a semi-infinite ray in R_q from a fixed origin. Such a ray intersects Ω_q at a unique point $\mathbf{u}$ and conversely such a point uniquely defines a direction. Actually what I have just described is a *vectorial* direction for which $\mathbf{u}$ defines a direction distinct from that defined by $-\mathbf{u}$, for instance the "vanishing direction" of a homing pigeon released from a location distant from its home loft. Also common are *axial* directional data for which $\mathbf{u}$ and $-\mathbf{u}$ cannot be distinguished. An example is the direction of the principal axis of an ellipsoidal pebble found in a stream bed. In the axial case, the natural mathematical ground for directions is the projective plane. However, even in this case the surface of the sphere is a comfortable mathematical environment as long as one bans any operations that treat $\mathbf{u}$ and $-\mathbf{u}$ differently.

Why should there be a specialization of statistics on spheres? Cannot we use the same techniques as we learn in Statistics 101, at least for simple problems? As Fisher was perhaps the first to point out, the problem is that we need to take into account the topology of the sphere. In particular, we need techniques that are immune to the fact that any nonredundant parameterization of Ω_q in terms of reals must have at least one point at which continuity breaks down. In less technical terms, there is a "wrap-around" problem arising from the fact that most common ways of specifying directions are in terms of angles. For the circle Ω_2, indeed, it is hard to imagine another useful one-dimensional parameterization. That this can cause problems for the unwary is obvious by comparing the following two representations for the same batch of data: $\{-35°, 1°, 37°\}$ and $\{325°, 1°, 37°\}$. For these data, any sensible measure of "central tendency" should be close to 1°. However, any reasonable method appropriate for reals will yield very different results for the two representations.

In fact, one can view directional statistics as a particular case of multivariate analysis. Indeed, the approach of studying directions as unit-vectors in R_q is a natural way to avoid wrap-around problems. Thus one can compute a central tendency of a batch of directions as the direction of the vector sum of vectorial unit vectors. For axial data, the eigen structure of the cross-product matrix of

Received October 1984; revised January 1985.

unit vectors may often be used to describe both clustering of directions and the shape of a cluster.

Directional data arise in a great variety of fields—Geology (the orientation of the paleomagnetism in rocks, the direction of the axes of sand dunes, the orientation of tilted earth structures), Biology (the orientation of microorganisms in polarized light, the homing behavior of various animals including Man, the foraging behavior of insects, the directions of capillaries in tissue, the analysis of chromosome patterns), Physics (the direction of particle diffusion), Astronomy (the orientation of planetary and cometary orbits), etc. Directional data methods have even been applied to data that is literally defined by latitude and longitude on the sphere we call the Earth, trying to make sense of patterns of chains of islands.

The statistics of directions have a long history dating back at least to Daniel Bernoulli, who, in carrying out what has been called the first statistical significance test, examined the null hypothesis that the directions normal to the orbital planes of the known planets might have arisen by sampling an isotropic distribution. Of course, as R. A. Fisher pointed out, much of the early development of least squares by Gauss and Laplace was for the analysis of astronomical determinations of the location of planets and stars on the celestial sphere. However, for measurements of such precision there is no need to take the topology of the sphere into account.

Directional statistics entered the main stream of statistics in 1953 with the publication of papers by Fisher (1953) and Gumbel, Greenwood and Durand (1953). Much of the content of these had been anticipated by Arnold (1941) in an unpublished MIT dissertation, but his work seems to have been largely ignored. Both papers dealt with particular cases of what Watson calls the Langevin distribution. This has the density with respect to the uniform probability measure $d\omega_q/\omega_q$ on Ω_q ($d\omega_q$ is invariant measure on Ω_q with total measure $\omega_q = 2\pi^{q/2}/\Gamma(q/2)$)

$$f(\mathbf{u};\, \kappa,\, \boldsymbol{\mu}) = F(\kappa)^{-1}\exp(\kappa \mathbf{u}^T \boldsymbol{\mu}),$$

in which $\boldsymbol{\mu} \in \Omega_q$ is a pole of concentration, $\kappa \geq 0$ is a precision parameter, and $F(\kappa)$ is a normalizing constant expressible in terms of Bessel functions (if one is unfamiliar with special functions when involved in directional statistics, one's ignorance does not last long). In the literature, this family is usually called the Fisher-von Mises or von Mises-Fisher distribution. For $q = 2$, von Mises (1918) applied it to the somewhat far-fetched situation where the data were the fractional parts of atomic weights of the chemical elements, rescaled as angles. Gumbel, Greenwood and Durand (1953) call this case the "circular normal distribution," as it is one candidate (of many) for an analogue on the circle of the usual normal distribution. Fisher (1953) proposed the form with $q = 3$ for the analysis of measurements of the direction of paleomagnetism. His principal interest was to clarify the application of fiducial inference in the presence of a nuisance parameter (see Barnard, 1963; Williams, 1963; Bingham, 1980). Watson gives priority to Langevin who, in 1905, derived the family as an equilibrium distribution arising in statistical mechanics.

Once the Langevin distribution was in the standard literature, its exploration was rapid, especially for the important cases of $q = 2$ and $q = 3$. Many problems parallel to those in normal theory arise naturally. The pole μ is a vector location parameter analogous to the mean and $1/\kappa$ is an analogue of the variance. The primary interest is usually in μ with κ a nuisance parameter. For a single sample there may be interest in testing H_0: $\mu = \mu_0$ or H_0: $\mu \in V$, V a subspace. For more than one sample, there is the obvious problem of testing whether all the μ's are the same, or more generally, lie in the same plane or hyperspace. Just as homogeneity of variance simplifies normal theory in the multi-sample case, so does homogeneity of κ simplify Langevin theory. Exact and approximate inferential methods have been developed for these problems and others (much of the work by Watson and his students). A relatively unexplored area concerns multivariate extensions of the Langevin distribution. Some progress has been made in a regression context (Gould, 1969), and in the bivariate case (Rivest, 1982; Saw, 1984). Mardia's review (1975) suggests other directions. Downs (1972) applied an analogue of the Langevin distribution to orientation data with more complex structure than either axial or vectorial.

Other distributions that have been proposed for vectorial data include the Brownian motion distribution, i.e., the distribution of the location of a particle subject to homogeneous diffusion from a point source on Ω_q (Roberts and Ursell, 1960; Stephens, 1963), and "small circle" distributions (Mardia and Gadsden, 1977; Bingham and Mardia, 1978). These, like the Langevin distribution, are all rotationally symmetric around a single pole μ.

The development of distributions for axial data has lagged somewhat. In the case $q = 2$, there is nothing new to be said, since the projective plane is isomorphic to the circle by simply doubling angles. But for $q \geq 3$ this trick does not work since the projective plane is not isomorphic to the sphere. Selby (1964) suggested densities proportional to $\exp(-\kappa \,|\cos\theta\,|)$ or to $\exp(\kappa \sin\theta)$, where $\cos\theta = \mathbf{u}^T\mu$, while Watson suggested a density proportional to $\exp(-\kappa \cos^2\theta)$ (called now by Watson the Scheidegger-Watson distribution). Depending on the sign of κ, these are either concentrated near the "equator" normal to μ (so-called "girdle" distributions) or near $\pm\mu$ ("bipolar" distributions). All display rotational symmetry about μ. Bingham (1974) independently introduced (for $q = 3$) a generalization of the Scheidegger-Watson distribution that may not be rotationally symmetric around any axis. Its density is proportional to $\exp(\sum_{j=1}^{q} \zeta_j(\mu_j^T\mathbf{u})^2) = \exp(\mathrm{tr}\, A\mathbf{u}\mathbf{u}^T)$. Here A is a $q \times q$ symmetric matrix with spectral decomposition $A = MZM^T$ with orthogonal $M = [\mu_1, \mu_2, \cdots, \mu_q]$, and $Z = \mathrm{diag}[\zeta_1, \cdots, \zeta_q]$. M serves as a generalized location parameter and Z as a set of shape and concentration parameters. For $q = 3$, if $\zeta_1 \ll \zeta_2 \leq \zeta_3$, the density is girdle near the plane normal to μ_1, while if $\zeta_1 \leq \zeta_2 \ll \zeta_3$ it is bipolar near $\pm\mu_3$. For this distribution, too, there are analogues of many of the standard normal theory one- and multi-sample problems. These have been less thoroughly explored than for the Langevin distribution.

Both the Langevin and Bingham distributions can be viewed as conditional distributions of particular multivariate normal distributions for $\mathbf{u}$, given $\mathbf{u} \in \Omega_q$.

Thus if $\mathbf{u}$ is $MVN_q(\mu, \kappa^{-1}Iq)$, $\mu \in \Omega_q$, then conditional on $\| \mathbf{u} \| = 1$, $\mathbf{u}$ is Langevin. Similarly, if $\mathbf{u}$ is $MVN_q(0, -2A^{-1})$ where $A = A^T$ is negative definite, the conditional distribution of $\mathbf{u}$ is Bingham, with the μ's and ζ's the eigenvectors and values of A. This suggests the more general family derived by conditioning an arbitrary $MVN_q(\mu, \Sigma)$. For $q = 3$, Kent (1982) partially classified the distributions so obtained, according to various restrictions placed on μ and Σ. These distributions include the Bingham-Mardia small circle distribution and a useful noncircularly symmetric distribution for vectorial data (Kent's FB_5).

Further, the Langevin and Bingham distributions and their generalizations are, of course, in the exponential family of distributions. Beran (1979) has proposed including higher degree polynomials in $\mathbf{u}$ in the exponent, and has developed a powerful approach to the difficult problems of estimation, largely bypassing the need to work with multi-argument transcendental functions.

There is another important problem in applied directional statistics—testing the hypothesis of isotropy (uniformity on Ω_q). Since Ω_q is compact, the uniform distribution is proper and represents the absence of any structure in the population. In some contexts, rejection of this hypothesis is a principal goal of research. For instance, before it makes sense to study whether the position of the sun affects the typical foraging direction of bees, one needs to establish that the directions are not uniformly random. The most widely applied test is the Rayleigh test based on the squared resultant length $R^2 = \| \sum_{i=1}^{n} \mathbf{u}_i \|^2$. R^2 is asymptotically distributed as $n\chi_q^2/q$ when the underlying distribution is uniform. For vectorial data, Bingham (1974) proposed (for $q = 3$) the statistic $(q(q+2)/2n)\mathrm{tr}(UU^T - nq^{-1}I_q)^2$, where U is the $q \times n$ matrix $[\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n]$. This is asymptotically $\chi^2_{(q+2)(q-1)/2}$ under isotropy. A rich theory of tests for uniformity on spheres and more general compact structures has been developed and exploited (Ajne, 1968; Beran, 1968, 1969b; Giné, 1975; Kuiper, 1960; Prentice, 1978; Stephens, 1964; Watson, 1961, 1976).

In addition to the various parametric approaches, there has been a fair amount of work on extension to the circle of distribution-free methods, primarily rank tests (Beran, 1969a; Mardia, 1967; Rothman, 1971; Schach, 1969; Welner, 1979). In particular, there are tests for a preferred direction, tests of independence, two-sample tests for identity of distribution, etc.

The book under review is indeed an important addition to the literature. It is the closest thing we have to a monograph on directional statistics which can lead research statisticians to the frontiers. It consists of five chapters, each corresponding to a lecture presented at a conference on Statistics on Spheres at the University of Arkansas in March, 1982. No claim is made that it provides enough background and detail to be of much use to the practicing scientist who is a novice at analyzing directional data. For that, Mardia's book (1972) is still indispensible. However, Watson's book contains a wealth of material of interest both to the theoretician and the applied statistician.

In spite of the limitations of the format, much of the work mentioned earlier is covered to some degree—tests of uniformity and parametric methods based on the Langevin, Bingham, and other distributions. However, its greatest importance lies not in describing small sample methods based on parametric families, but in providing the beginnings of a large sample methodology for statistical problems on spheres, making only minimal assumptions such as symmetry.

Chapter 1 is a general introduction for those new to directional data. Types of data and applications are described, and summary statistics and graphical methods of display are presented. Particularly interesting is a detailed look at the directions normal to comet orbits. This is used to illustrate uniformity tests and a graphical assessment of the goodness-of-fit of the Langevin distribution. This last is an important problem that has had too little attention. There is a brief section on density estimation on the sphere using the Langevin density as kernel.

Chapter 2 deals with the uniform distribution. Various results are derived concerning the exact distribution of $\mathbf{X} = \sum_{j=1}^{n} \mathbf{u}_j$, where $\mathbf{u}_j$ are i.i.d. uniform on Ω_q. The Central Limit Theorem gives the large samplè distributions of $\mathbf{X}$ and $M_n = n^{-1} \sum_{j=1}^{n} \mathbf{u}_j \mathbf{u}_j^T$. These are used to derive the asymptotic χ^2 distributions of the Rayleigh and Bingham statistics for testing uniformity. Of some mathematical interest is a look at what happens when n is fixed but q goes to infinity.

Chapter 3 covers an immense amount of territory and serves as an introduction to many of the results in Chapters 4 and 5. It consists of a mixture of facts concerning various parametric families. The discussion of exponential models includes a succinct review of spherical harmonics, another useful theoretical tool. Various routes to constructing distributions are sketched—conditioning (leading to the Langevin, Bingham, and more general families), marginalizing (angular Gaussian), diffusion on Ω_q subject to possibly random stopping times (Brownian motion, Langevin, among others), diffusion to surface from an interior point of a solid sphere (Langevin), and maximization of entropy (Langevin, Bingham). There is a nice proof that the Langevin is the only distribution with density of the form $f(\boldsymbol{\mu}^T\mathbf{u})$ such that $\| \mathbf{X} \|^{-1}\mathbf{X}$ is the MLE of $\boldsymbol{\mu}$.

Chapters 4 and 5 are the most important in the book. They provide the beginnings of general large sample inferential methods based on the sample mean vector $n^{-1}\Sigma\mathbf{u}_i$ (Chapter 4) and the eigen structure of the sample second moment matrix M_n (Chapter 5). Emphasis is on distributions having particular symmetries. Chapter 4 is primarily concerned with methods applicable when the distribution is rotationally symmetric about $\boldsymbol{\mu} \in \Omega_q$, i.e., having densities depending only on $\cos\theta = \boldsymbol{\mu}^T\mathbf{u}$. The population mean direction $\boldsymbol{\mu} \in \Omega_q$ is the object of interest. The distributions emphasized in Chapter 5 have densities of the form $f(\| P_V\mathbf{u} \|)$, where P_V projects into a subspace V. This is a natural generalization of the previous case, with the object of interest now being the subspace V. Single and multi-sample test statistics of various hypotheses (e.g., H_0: $V_1 = V_2 = \cdots = V_k$) are proposed and their null and nonnull asymptotic distributions derived. The results based on the mean vector $\bar{\mathbf{u}}$ are derived fairly readily by standard methods starting with the multivariate central limit theorem. On the other hand, most of the results on the eigen structure of M_n are derived using elegant methods from a book by Kato (1966) on perturbation theory for linear operators. Chapter 4 also discusses large sample methods specifically for the Langevin distribution, and similarly Chapter 5 has a section devoted to the generalized Scheidegger-Watson distribution with density proportional to $\exp(\kappa \| P_V\mathbf{u} \|^2)$.

There are three Appendices. Appendix A presents a lot of facts concerning functions derived from the normalizing functions for the Langevin and Scheidegger-Watson distributions. It is unfortunately marred by a number of algebraic errors. Appendix B summarizes the application of results of Kato (1966) to the spectral analysis of cross-product matrices. And Appendix C, by this reviewer, gives a series expansion in spherical harmonics for the marginal distribution of $\mathbf{u} = \| \mathbf{x} \|^{-1}\mathbf{x}$ when $\mathbf{x}$ is $MVN_q(\mu, \sigma^2 I_q)$.

In summary, Watson's book is a milestone in the literature on spherical distributions. For the specialist it brings together many results and points to paths for new research directions. For the statistician who is new to the subject, it is an excellent introduction to much of what is important in the field. This will be easier with an errata sheet available from Watson.

One of the exciting things about the area of orientation statistics is that there are still many areas where we scarcely have an inkling of what to do. For instance, beyond pairwise correlation measures, I am unaware of anything on genuinely multivariate problems involving several related determinations of direction. Except for a test for serial correlation (Watson and Beran, 1967), there is a dearth of methods that might be used to analyze time series of orientation variates. Appropriate models would find immediate application in geophysics. In fact, given practically any problem area in "flat" statistics—robustness, clustering, modelling, influential observations, to name a few—there is a corresponding problem for spheres. Progress is being made, but there is much to be done. And, of course, when statistics on the sphere are as familiar as $N(0, 1)$, there are worlds of more complicated curved manifolds to conquer.

REFERENCES

AJNE, B. (1968). A simple test for uniformity of a circular distribution. *Biometrika* **55** 343–354.

ARNOLD, K. J. (1941). On spherical probability distributions. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Mass.

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

BERAN, R. J. (1968). Testing for uniformity on a compact homogeneous space. *Appl. Prob.* **5** 177–195.

BERAN, R. J. (1969a). The derivation of nonparametric two-sample tests from tests for uniformity of a circular distribution. *Biometrika* **56** 561–570.

BERAN, R. J. (1969b). Asymptotic theory of a class of tests for uniformity of a circular distribution. *Ann. Math. Statist.* **40** 1196–1206.

BERAN, R. J. (1979). Exponential models for directional data. *Ann. Statist.* **7** 1162–1179.

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

BINGHAM, C. (1980). Distribution on the sphere. In *R. A. Fisher: An appreciation.* (S. E. Fienberg and D. V. Hinkley, eds.) 171–181. Springer-Verlag, New York.

BINGHAM, C. and MARDIA, K. V. (1978). A small circle distribution on the sphere. *Biometrika* **65** 379–389.

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

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

GINÉ M., E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.* **3** 1243–1266.

GOULD, A. L. (1969). A regression technique for angular variates. *Biometrics* **25** 683–696.

GUMBEL, E. J., GREENWOOD, J. A. and DURAND, D. (1953). The circular normal distribution: Theory and tables. *J. Amer. Statist. Assoc.* **48** 131–152.

KATO, T. (1966). *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin.

KENT, J. T. (1982). The Fisher-Bingham distribution on the sphere. *J. Roy. Statist. Soc. Ser. B* **44** 71–80.

KUIPER, N. H. (1960). Tests concerning random points on a circle. *Nederl. Akad. Wetensch. Indag. Math.* **22** 38–47.

MARDIA, K. V. (1967). A non-parametric test for the bivariate two-sample location problem. *J. Roy. Statist. Soc. Ser. B* **29** 320–342.

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

MARDIA, K. V. (1975). Statistics of directional data (with discussion). *J. Roy. Statist. Soc. Ser. B* **37** 349–393.

MARDIA, K. V. and GADSDEN, R. J. (1977). A small circle of best fit for spherical data and areas of vulcanism. *Appl. Statist.* **27** 238–245.

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

RIVEST, L.-P. (1982). Some statistical methods for bivariate circular data. *J. Roy. Statist. Soc. Ser. B* **44** 81–90.

ROBERTS, P. H. and URSELL, R. D. (1960). Random walks on a sphere and on a Riemannian manifold. *Trans. Roy. Soc. Lond.* **252A** 317–356.

ROTHMAN, E. D. (1971). Tests of coordinate independence of a bivariate sample on a torus. *Ann. Math. Statist.* **41** 1962–1969.

SAW, J. G. (1984). Dependent unit vectors. *Biometrika* **70** 665–682.

SCHACH, S. (1969). Nonparametric symmetry tests for circular distributions. *Biometrika* **56** 571–577.

SELBY, B. (1964). Girdle distributions on a sphere. *Biometrika* **51** 381–392.

STEPHENS, M. A. (1963). Random walk on a circle. *Biometrika* **50** 385–390.

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

VON MISES, R. (1918). Uber die 'Ganzzahligheit' der Atomgewichte und Verwandte Frangen. *Physik. Zeitschr.* **7** 153–159.

WATSON, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika* **48** 109–114.

WATSON, G. S. (1976). Optimal invariant tests for uniformity. In *Studies in Probability and Statistics*, (E. J. Willams, ed.) 121–127. North-Holland, Amsterdam.

WATSON, G. S. and BERAN, R. J. (1967). Testing a sequence of unit vectors for serial correlation. *J. Geophysical Res.* **72** 5655–5659.

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

WELLNER, J. A. (1979). Permutation tests for directional data. *Ann. Statist.* **7** 929–943.

WILLIAMS, E. J. (1963). A comparison of the direct and fiducial arguments in the estimation of a parameter. *J. Roy. Statist. Soc. Ser. B* **25** 95–99.

DEPARTMENT OF APPLIED STATISTICS
SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
1994 BUFORD AVENUE
ST. PAUL, MINNESOTA 55108