Article:
Mardia, K.V. orcid.org/0000-0003-0090-6235 (2025) Fisher's legacy of directional statistics, and beyond to statistics on manifolds. Journal of Multivariate Analysis, 207. 105404. ISSN 0047-259X

https://doi.org/10.1016/j.jmva.2024.105404

Fisher's legacy of directional statistics, and beyond to statistics on manifolds

Kanti V. Mardia[a,*]

[a] *University of Leeds and University of Oxford*

Abstract

It is not an exaggeration to say that R.A. Fisher is the Albert Einstein of Statistics. He pioneered almost all the main branches of statistics, but it is not as well known that he opened the area of Directional Statistics with his 1953 paper introducing a distribution on the sphere which is now known as the Fisher distribution. He stressed that for spherical data one should take into account that the data is on a manifold. We will describe this Fisher distribution and reanalyze his geological data. We also comment on the two goals he set himself in that paper, and on how he reinvented the von Mises distribution on the circle. Since then, many extensions of this distribution have appeared bearing Fisher's name such as the von Mises-Fisher distribution and the matrix Fisher distribution. In fact, the subject of Directional Statistics has grown tremendously in the last two decades with new applications emerging in life sciences, image analysis, machine learning and so on. We give a recent new method of constructing the Fisher type distributions on manifolds which has been motivated by some problems in machine learning. The number of directional distributions has increased since then, including the bivariate von Mises distribution and we describe its connection to work resulting in the 2024 Nobel-winning AlphaFold (in Chemistry). Further, the subject has evolved as Statistics on Manifolds which also includes the new field of shape analysis, and finally, we end with a historical note pointing out some correspondence between D'Arcy Thompson and R.A. Fisher related to shape analysis.

Keywords: Distributions on manifolds, Fisher distribution, Machine Learning, Remanent magnetism, von Mises distribution, Wrapped tangent distributions.
2020 MSC: Primary; 62H11, 62R30, 62E10, Secondary: 62H15, 62F03

1. Introduction

Fisher opened the area of Directional Statistics with his pioneering 1953 paper, Fisher [18], introducing what is now known as the Fisher distribution on the sphere. He stressed that for spherical data one should take into account that the data is on a manifold. Since then, many extensions of this distribution have appeared each bearing Fisher's name, as we will describe. In fact, the subject of Directional Statistics has grown tremendously in the last two decades with new applications emerging in life sciences, image analysis, machine learning, and so on.

One of the features of Fisher's work is that the starting point was often a motivating application brought to Fisher by scientists or applied statisticians. For Directional Statistics, this was the problem of pole reversal raised by the geologists Mr J. Hospers and Professor S.K. Runcorn.

In Sections 2 and 3, we give an overview of the subject and discuss in particular the extension of the Fisher distribution on the sphere to the hypersphere, which is now known as the von Mises-Fisher distribution, examined in Section 4. In Section 2.1, we describe some features of the seminal paper of Fisher [18] and his reasoning on why linear statistics is not meaningful in the context of his practical applications. Also we describe how this work was taken up by Geoffrey Watson (starting with Watson [55] which was written in May 1955) who

[1] Based on "The 40th Fisher Memorial Lecture" delivered in November 2022, Oxford. For the link to the talk, visit *http : //www.senns.uk/FisherWeb.html*. The first topic of the lecture is covered in Mardia [33] and this is the second topic.

*Corresponding author. Email address: K.V.Mardia@leeds.ac.uk

made it more accessible and popular. Historically, it is important to learn that though Fisher's paper appeared in 1953, he already formed these ideas in the 1920's. It was the pole-reversal geological application that made him go back and reactivate this work (Section 2.2). We reanalyse his geological data in Section 6.

In general, the maximum likelihood methods for directional distributions are not computationally straightforward. A new approach, the score matching estimate, will be presented for the von Mises-Fisher distribution in Section 5. However, in his paper, Fisher produced estimators using the distribution of some summary statistics, and we examine closely his approach in Section 5.1. He dealt with two cases, known pole and known axis; in the second case, his estimation method is somewhat questionable and has not been used. It is not commonly known that Fisher derived the von Mises distribution independently; his aim was to promote his fiducial argument (see Section 7). He mentioned that his paper has two goals and we assess these in Section 8.

The subject of Statistics on Manifolds is still evolving, including a new method to construct the Fisher type matrix distribution which has been motivated by some problems in machine learning. We examine this new approach in Section 9. Moreover, in Section 10, we give a glimpse of my journey into Directional Statistics, including work with colleagues in Denmark and Leeds on protein structure prediction, whose methods helped inform the 2024 Nobel prize-winning AlphaFold technology. Finally, in Section 11, we give a historical note pointing out some correspondence between D'Arcy Thompson (the pioneer of shape analysis) and R.A. Fisher where we could have gained Fisher's insight into shape analysis; however, this collaborative work did not materialise.

2. Statistics on manifolds/directional statistics

Big data, high dimensional data, and sparse data are all new frontiers of statistics. Changing technologies have created this flood of data and it's associated challenges, and have led to a substantial need for new modelling strategies and data analysis. There are data which are essentially not Euclidean and the data sit on a manifold. Even for the simplest non-Euclidean manifold, the circle, with angular data, using the arithmetic average cannot make sense, as is well known. Consider that the arithmetic average of the two angles 1° and 359° is

$$\frac{1^\circ + 359^\circ}{2} = 180^\circ \ .$$

Of course, it should be 0°. That is, the non-Euclidean setting throws up many major challenges, both mathematical and statistical, and so more care is needed. In simple terms, Statistics on Manifolds deals with non-Euclidean variables driven mainly by the underlying geometrical space. Examples include circle, sphere, torus, and shape spaces.

The subject of Directional Statistics has grown tremendously, especially since the 1980's, with advances in Statistics on Manifolds leading to new distributions on the hypersphere, torus, Stiefel manifold, Grassmann manifold and so on. The progress in this area can be seen through several books published since then in the 1980s and later: Nick Fisher et al. [16], Nick Fisher [15], Mardia and Jupp [36], Jammalamadaka and Sengupta [22] and Ley and Verdebout [24, 25]. There has been a recent special issue of Sankhyā edited by Bharath and Dey [4]. Further, Pewsey and García-Portugués [47] have given a comprehensive survey of Directional Statistics and in the discussion to the paper Mardia [31] has given a brief history of the subject. Particularly, to note that the methods of Principal Component Analysis on the torus are now well established (see, for example, Mardia et al. [42], Wiechers et al. [59]). Another major development is the new area of discrete distributions in Directional Statistics, see [38] and Mardia and Sriram [39]; these two papers, in particular, take forward Karl Pearson's challenges of the 1890's related to his roulette wheel problem (which did not draw the attention of Fisher).

2.1. Fisher 1953's paper: A landmark paper in directional statistics, and the role of Geoffrey Watson

The paper of R.A. Fisher entitled "Dispersion on a sphere" appeared in 1953, motivated by remanent magnetism data. He began with justifying the need for such work, as the following quotes from the paper shows:

- " The theory of errors was developed by Gauss primarily in relation to the needs of astronomers and surveyors, making rather accurate angular measurements. Because of this accuracy it was appropriate to develop the theory in relation to an infinite linear continuum, or, as multivariate errors came into view, to a

Euclidean space of the required dimensionality. The actual topological framework of such measurements, the surface of a sphere, is ignored in the theory as developed, with a certain gain in simplicity.

- It is, therefore, of some little mathematical interest to consider how the theory would have had to be developed if the observations under discussion had in fact involved errors so large that the actual topology had had to be taken into account.
- The question is not, however, entirely academic, for there are in nature vectors with such large natural dispersions. The remanent magnetism found in igneous and sedimentary rocks, . . . , do show such considerable dispersion that an adequate theory for the combination of such observations is now needed.
- My examples are drawn from the very fine body of data on the remanent magnetism of Icelandic lava flows, historic and prehistoric, put at my disposal by Mr J. Hospers . . . "

In fact, this subject took off when the 1953 paper came to the attention of Geoffrey Watson, one of the pioneers of Directional Statistics. It is interesting to record how this happened. In his conversation with Rudy Beran and Nick Fisher (Beran and Fisher [3]), Watson says:

"We sent a food parcel to his daughter, Joan, in England, at his request. As a "thank you" note he sent me a reprint of his 1953 paper which is not easy reading . . . Anyhow I had a look at it and suddenly saw that I could clarify things." (This is referring to Joan Box, a daughter of Fisher who wrote his biography, Box [9]).

We want to recall that Fisher visited Melbourne University while Watson was Senior Lecturer in Statistics from 1951 and an Acting Head of the Department; Watson looked after him during his visit, Watson [56]. In fact, Fig. 1 shows one of the Fisher's photos taken during this period by Watson in a picnic; this was given to the author personally by Geoffrey.

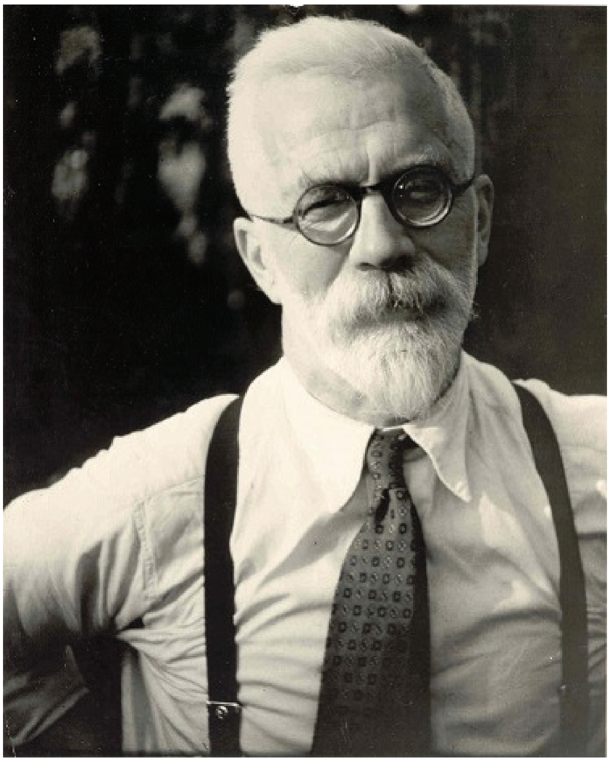


Fig. 1: Fisher 1953 in a Picnic in Australia; Photo by Geoffrey Watson – Gifted this version to the author in 1990.

Following that, Geoffrey Watson and Michael Stephen made several key contributions in 1956 – 1970 (starting from Watson's first paper of 1956, Watson [55], written in May, 1955). In 1970-1980 the subject came into the limelight, partly with my 1972 directional book Mardia [27] but perhaps more so with my discussion paper in the Journal of the Royal Statistical Society (Mardia [28]) which included many eminent discussants, David Cox, Henry Daniel, David Kendall, John Kingman, among others. There was something of a lull in the subject during 1990 to 1999, though a few books came out, see Mardia [31] for more details. From 2000, a resurgence of interest led to many new advances, mainly because of the recognition by Image Analysts and Life Scientists of its importance. This momentum now continues as most of statisticians do regard Directional Statistics as a mainstream statistical topic.

2.2. When was the material of the 1953 paper written?

There has been some discussion on why and when was the material written for the 1953 paper. We first quote Box [9], p. 439 on the roles of Fisher, Hospers and Runcorn (Hospers was Runcorn's student who collected the data):

- Runcorn was at once struck by the fact that these lavas grouped themselves into those along the present field and those opposite, in a manner strongly suggesting that the direction of magnetization of the earth had been reversed at some periods of its history.

- Runcorn explained these results to Fisher and asked him how the particular statistical problems of testing homogeneity should be solved. Fisher, pulling out some old notes he had made before 1930 in connection with his paper of 1929 "Tests of significance in harmonic analysis", Fisher [17], set to work to apply the method he had then devised to the data now in hand.

- Hospers quickly used the method in three very important papers which laid the basis for the paleomagnetic work in the Tertiary Age.

Hospers used Fisher's 1953 paper's preprint for his three papers described above. In Section 6, we give full details of Hospers' data used by Fisher.

Another version on this point is given as follows in Nick Fisher et al. [16], pp. 12–13, quoting George Barnard (letter to Nick Fisher, 30 June 1981.)

- Fisher sent me an offprint of his paper "Dispersion on a sphere" when it came out, ... while I saw you could use the length of the vector sum to test isotropy, I had not seen how to do estimation. He replied that it was easier for him, since he had, in the twenties, asked himself what would be the analogue, on the sphere, of the normal density in the plane, and had made some notes on it.

- When approached by (I presume) Hospers he was able to go to his filing cabinet and pull out the notes, and answer the question on the spot ... the figure 1922 comes into my head, for Fisher's first work. This seems rather early; but he was in touch with Eddington, the astronomer, in 1920, and may well have gone on thinking about the problems he had raised.

Another account of this paper has been given by Persi Diaconis in 1988 (Diaconis [12], p. 171) which we now quote.

- I cannot resist reporting some background on Fisher's motivation for working with the distribution discussed above. This story was told to me in 1984 by the geologist Colin B.B. Bull. Dr. Bull was a student in Cambridge in the early 1950's, ... Fisher asked what area Bull worked in. Bull explained that a group of geologists was trying to test Wegener's theory of continental drift. Wegener had postulated that our current continents used to nest together.

- They searched for data that were closer to geology. They had hit on the distribution of magnetization angle in rocks. This gave points naturally distributed on the sphere. They had two distributions (from matching points on two continents) and wanted to test if the distributions were the same.

- Fisher took a surprisingly keen interest in the problem and set out to learn the relevant geology. In addition to writing his famous paper (which showed the distributions were different) he gave a series of talks at the geology department to make sure he'd got it right.

- Why did Fisher take such a keen interest? A large part of the answer may lie in Fisher's ongoing war with Harold Jeffries. They had been rudely battling for at least 30 years over the foundations of statistics. Jeffries has never really accepted (as of 1987!) continental drift. It is scarcely mentioned in Jeffries' book on geophysics.

These extracts point to the fact that Fisher has developed the distribution around 1922 and his subsequent contacts with geologists in the early 1950's led to his 1953's paper.

3. The Fisher distribution and other directional distributions

We first recall some manifolds before giving some relevant directional distributions.

Sphere. The sphere

$$S_p = \{\boldsymbol{x} \in \mathbb{R}^q : \ \boldsymbol{x}^T\boldsymbol{x} = 1\}, \ q = p + 1, \tag{1}$$

represents the space of unit vectors or "directions" in $\mathbb{R}^q$. We have the circle when $p = 1$.

Real projective space. The real projective space consists of the "axes" or "unsigned directions" $\pm\boldsymbol{x}$. In some sense this space is half of a sphere; it can also be represented as the space of rank 1 projection matrices,

$$\mathbb{R}P_p = \{\boldsymbol{P} \in \mathbb{R}^{q\times q} : \boldsymbol{P} = \boldsymbol{P}^T, \ \boldsymbol{P}^2 = \boldsymbol{P}, \ \text{tr}\ \boldsymbol{P} = 1\}. \tag{2}$$

A rank one projection matrix ($\boldsymbol{P}$) can be written as $\boldsymbol{P} = \boldsymbol{x}\boldsymbol{x}^T$ where $\boldsymbol{x}$ is a unit vector.

Rotation matrices. The special orthogonal group of $r \times r$ rotation matrices ($\boldsymbol{X}$) is defined by

$$SO(r) = \left\{\boldsymbol{X} \in \mathbb{R}^{r\times r} : \ \det \boldsymbol{X} = 1, \ \boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}_r\right\}. \tag{3}$$

On each of these spaces, there is a unique uniform distribution which is invariant under rotations. Further each of these spaces is naturally embedded in a Euclidean space. A natural "linear-exponential" family of distributions can be generated by letting the density (with respect to the uniform measure) be proportional to the exponential of a linear function of the embedded variables. We now list some distributions with their names and space as listed in Table 1.

- The von Mises-Fisher distribution of $\boldsymbol{x}$ (with respect to Lebesgue measure) on S_p, parameter $\boldsymbol{\alpha} \in \mathbb{R}^q, \ q = p + 1$, has its density:

  $$f(\boldsymbol{x}) \propto \exp(\boldsymbol{\alpha}^T\boldsymbol{x}), \quad \boldsymbol{\alpha}, \boldsymbol{x} \in \mathbb{R}^q, \ \boldsymbol{x}^T\boldsymbol{x} = 1,$$

  where usually we write $\boldsymbol{\alpha} = \kappa\boldsymbol{\mu}$ with $\boldsymbol{\mu}^T\boldsymbol{\mu} = 1, \ \kappa \geq 0$. The distribution is analogous to (for the concentrated case) a p-dimensional isotropic normal distribution. More details on this distribution are given in Section 4.

- The Bingham distribution of $\boldsymbol{x}$ (with respect to Lebesgue measure) on S_p, symmetric parameter matrix $\boldsymbol{A}(q \times q)$; $\boldsymbol{A}$ and $\boldsymbol{A} + \lambda\boldsymbol{I}_q$ define the same distribution, and has its density:

  $$f(\boldsymbol{x}) \propto \exp(-\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}) = \exp\{-\text{tr}(\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T)\}, \quad \boldsymbol{x} \in \mathbb{R}^q, \ \boldsymbol{x}^T\boldsymbol{x} = 1,$$

  that is, quadratic-exponential on S_p and linear-exponential on $\mathbb{R}P_p$. The distribution is analogous to (for the concentrated case) a p-dimensional general normal distribution.

- The matrix Fisher distribution of $\boldsymbol{X}$ on $SO(r)$, parameter matrix $\boldsymbol{F}(r \times r)$ has its density,

  $$f(\boldsymbol{X}) \propto \exp\left\{\text{tr}\left(\boldsymbol{F}^T\boldsymbol{X}\right)\right\}, \quad \boldsymbol{X} \in SO(r),$$

  with respect to the underlying invariant "Haar" measure. It is unimodal about a fixed rotation matrix determined by $\boldsymbol{F}$. For the concentrated case, the distribution can be connected to the Wishart distribution.

For further details on these distributions, see, for example, Mardia and Jupp [36].

Table 1: Some common directional distributions in exponential family: spaces and names.

Space	Notation	Distributions
circle	S_1	von Mises ($p = 1$)
sphere	S_p	Fisher ($p = 2$)
		von Mises-Fisher ($p \geq 1$),
		Fisher-Bingham
real projective space	$\mathbb{R}P_p$	Bingham
special orthogonal group	$SO(r)$	matrix Fisher

4. The von Mises-Fisher distribution

The von Mises-Fisher distribution of $\boldsymbol{x}$ on S_p has the density with respect to Lebesgue measure

$$f(\boldsymbol{x}) = c_p(\kappa)\exp(\kappa\boldsymbol{\mu}^T\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^{p+1},\ \kappa \geq 0,\ \boldsymbol{\mu}^T\boldsymbol{\mu} = 1,\ \boldsymbol{x}^T\boldsymbol{x} = 1, \tag{4}$$

where

$$c_p(\kappa) = \frac{\kappa^{(p-1)/2}}{(2\pi)^{(p+1)/2} I_{(p-1)/2}(\kappa)}, \tag{5}$$

and $I_\nu(.)$ is the modified Bessel function of the first kind and order ν. For $p = 1$, it is the von Mises distribution and for $p = 2$, the Fisher distribution. The distribution has a mode at the mean direction $\boldsymbol{\mu}$ and for large concentration parameter κ, it has a p-dimensional isotropic normal distribution. For $\kappa = 0$, we have the uniform distribution on the hypersphere. It is also known as the Langevin distribution since a generalised form was given by Langevin in 1905 in the context of the theory of magnetism (Langevin [23]).

The Fisher distribution. Let θ denote the colatitude $0 \leq \theta \leq \pi$ and ϕ be the longitude $0 \leq \phi \leq 2\pi$ in the spherical polar coordinates. Fisher (1953) with $p = 2$ took the north pole as the mean direction ($\boldsymbol{\mu} = (0, 0, 1)^T$) so θ then represented the angular displacement from the true mean direction, and the Fisher density of θ in (4) simplifies to

$$\frac{\kappa}{2\sinh\kappa}\exp\{\kappa\cos\theta\}\sin\theta, \quad 0 \leq \theta \leq \pi,\ \kappa \geq 0, \tag{6}$$

and θ is independent of ϕ which is distributed uniformly on a circle. In general, the distribution has a mode at the mean direction. Let us write

$$x = \sqrt{\kappa}\,\theta\ \cos\phi,\ y = \sqrt{\kappa}\,\theta\ \sin\phi \tag{7}$$

then for large κ, (x, y) has an isotropic bivariate normal distribution with zero means and unit variances. For $\kappa = 0$, we have the uniform distribution on the sphere.

Summary statistics on a sphere. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$, be n observations on S_p. Then the location of these observations can be summarised by their sample mean vector in R^q, which is

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i. \tag{8}$$

Write the vector $\bar{\mathbf{x}}$

$$\bar{\mathbf{x}} = \bar{R}\bar{\mathbf{x}}_0,\ 0 \leq \bar{R} \leq 1, \tag{9}$$

where $\bar{\mathbf{x}}_0$ is the sample mean direction and $\bar{R}$ ($= ||\bar{\mathbf{x}}||$) is the mean resultant length. Note that $\bar{\mathbf{x}}$ is the centre of gravity with direction $\bar{\mathbf{x}}_0$, and $\bar{R}$ is its distance from the origin. For the circular case, we have $\bar{\mathbf{x}}_0^T = \bar{R}(\cos\bar{\theta}, \sin\bar{\theta})$ where $\bar{\theta}$ is the mean direction.

For further details on the von Mises-Fisher distribution and data analysis, see, for example, Mardia and Jupp [36]. In the next section, we deal with some estimators for this distribution.

5. Estimation for the von Mises-Fisher distribution

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$, be a random sample drawn from the von Mises-Fisher distribution given by (4) then the maximum likelihood estimates (MLE) of $\boldsymbol{\mu}, \kappa$ are given by

$$\hat{\boldsymbol{\mu}}_{MLE} = \bar{\mathbf{x}}_0, \ \hat{\kappa}_{MLE} = A_p^{-1}(\bar{R}), \tag{10}$$

where

$$A_p(\kappa) = I_{(p+1)/2}(\kappa)/I_{(p-1)/2}(\kappa).$$

Assuming that $\boldsymbol{\mu}$ is known and writing in the polar coordinates with $\boldsymbol{\mu}^T \boldsymbol{x} = \cos\theta$, then Mardia et al. [37] have shown that the Score Matching Estimate (SME) of κ is simply

$$\hat{\kappa}_{SME} = p\frac{\sum \cos\theta_i}{\sum \sin^2\theta_i};$$

there are now no Bessel functions in this expression and it has been proved that there is only a moderate loss of efficiency compared to the MLE.

5.1. Fisher's Estimation

We now write the probability density function (pdf) of the Fisher distribution given by (4) in 3D as

$$f(\mathbf{x}) = B(\kappa)\exp(\kappa\boldsymbol{\mu}^T\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^3, \kappa \geq 0, \boldsymbol{\mu}^T\boldsymbol{\mu} = 1, \mathbf{x}^T\mathbf{x} = 1, \tag{11}$$

where, for simplicity, we have written here the constant $c_2(\kappa)$ in (5) as $B(\kappa)$ which has a simple expression given by

$$B(\kappa) = \kappa/(2\sinh\kappa).$$

The pdf is again with respect to the Lebesgue measure. Fisher [18] in his Section 2.1 dealt with Case 1: the known pole case (i.e. the mean direction $\boldsymbol{\mu}$ is given) and in his Section 2.2 dealt with Case 2: the known axis case (i.e. $\boldsymbol{\mu}$ is an axis) to estimate κ. In both these cases, he focused on the distribution of some summary statistics to get an estimate of κ. In Case 1, this approach leads to the maximum likelihood estimate (mle) of κ but not for Case 2. This paper predates what is now known as the Fisher-Neyman factorization theorem and sufficient statistic. In the following, we will use Fisher [18]'s notation wherever possible so the results can be compared with his paper.

5.1.1. Case 1: Known pole

If we know the true pole $\boldsymbol{\mu}$, then a sufficient statistic is

$$x = \sum \cos\theta_i = \sum \boldsymbol{\mu}^T\mathbf{y}_i, \tag{12}$$

where $\mathbf{y}_i$ is i-th observed direction vector, $i \in \{1, \ldots, N\}$. Fisher [18] has shown that the distribution of x is

$$g_N(x) = B(\kappa)^N \exp\{\kappa x\} P(x, N), \tag{13}$$

where

$$P(x, N) = \frac{1}{(N-1)!}\{(N-x)^{N-1} - N(N-2-x)^{N-1} + \cdots - \cdots (-1)^r\binom{N}{r}(N-2r-x)^{N-1}\}, \tag{14}$$

and r is the largest integer less than $\frac{1}{2}(N-x)$. Now, Fisher maximised (13) with respect to κ to get his estimate of κ, that is, this estimate is the solution of the equation

$$\coth(\kappa) - 1/\kappa = x/N, \quad \kappa > 0. \tag{15}$$

This is, of course, the MLE estimate of κ which we would have got by working directly on (11). Further, we note that the Fisher-Neyman factorization theorem does not need the distribution (13) of the sufficient statistic x for the estimation but was used by Fisher rather then going directly to the likelihood. He used his same procedure for Case 2 as we now show.

5.1.2. Case 2: Known axis

Fisher [18] continued to use the summary statistics x for known axis given by (12) in the form defined in his paper (see, the extract in Fig. 2 of this section in his paper), though it is no longer a sufficient statistic. If all we know is that the true axis $\pm\boldsymbol{\mu}$, then $x = \sum \cos\theta_i = \sum \boldsymbol{\mu}^T \mathbf{y}_i$ is not observable, and for this case, we have to include both $+x$ and $-x$ given in (11) finding the probability of x given by (12). One approach is to arbitrarily choose

2·2. *Known axis*

For a given pole, x (and k) may be either positive or negative, but negative values may be interpreted as positive values referred to the antipole. If, therefore, not the pole but only the axis is given, x may be taken to be always positive.

The distribution of x will then be

$$\frac{2}{(N-1)!}\left(\frac{\kappa}{2\sinh\kappa}\right)^N \cosh\kappa x\left\{(N-x)^{N-1} - N(N-2-x)^{N-1} + \ldots + (-)^r \frac{N!}{(N-r)!\,r!}(N-2r-x)^{N-1}\right\}.$$

The logarithmic likelihood of κ is now

$$L = -N(\ln\sinh\kappa - \ln\kappa) + \ln\cosh\kappa x,$$

the score, which equated to zero gives the equation of estimation, is

$$\partial L/\partial\kappa = -N(\coth\kappa - 1/\kappa) + x\tanh\kappa x,$$

and the amount of information is

$$-\partial^2 L/\partial\kappa^2 = N(1/\kappa^2 - \operatorname{cosech}^2\kappa) - x^2\operatorname{sech}^2\kappa x,$$

which does, indeed, depend only on κ and k, but cannot be so expressed explicitly.

Fig. 2: Section 2.2 of Fisher [18] showing his approach to estimation using the distribution of the summary statistic x for the known axis case.

a direction for $\boldsymbol{\mu}$ and then x is defined for any N. When $N > 1$, we must pick a head and a tail for $\boldsymbol{\mu}$ and then $x = \sum \boldsymbol{\mu}^T \mathbf{y}_i$ is well defined, but we do not know whether its sign is correct. Both possibilities $x = \sum \boldsymbol{\mu}^T \mathbf{y}_i$ and $x = -\sum \boldsymbol{\mu}^T \mathbf{y}_i$ need to be included in its distribution, then the pdf of x is

$$g_N(x,\kappa) + g_N(-x,\kappa) = B(\kappa)^N \exp(\kappa x) P(x,N) + B(\kappa)^N \exp(-\kappa x) P(-x,N), \tag{16}$$

where $g_N(\cdot,\kappa)$ is given by (13). It can be seen from (14) that $P(x,N) = P(-x,N)$ in (16) leading to (see, Fig. 2) his pdf as

$$2B(\kappa)^N \cosh(\kappa x) P(x,N). \tag{17}$$

Fisher maximises this sampling distribution (17) with respect to κ to get an estimate of κ (see, Fig. 2), leading to the solution of the equation

$$\coth(\kappa) - 1/\kappa = (x/N)\tanh(\kappa x/N). \tag{18}$$

Note that it can be shown that for large N, the estimate for Case 1 given by (15) is the same as that from (18).

The MLE for the axial case. Let $\boldsymbol{\nu}$ be the upper hemisphere representation of the known axis and $x = \sum(\boldsymbol{\nu}^T \mathbf{y}_i)$. Then x is well defined but we do not know if $x = \sum(\boldsymbol{\mu}^T \mathbf{y}_i)$ or $x = -\sum(\boldsymbol{\mu}^T \mathbf{y}_i)$ where $\boldsymbol{\mu}$ is the true pole.

Fisher does not give the MLE for this case as his aim again was to use an appropriate summary statistics. We can formulate the estimation problem in this case as follows to get the MLE. There are two unknown parameters, $\kappa \geq 0$ and, say, λ with values -1 and $+1$ such that $\boldsymbol{\mu} = \lambda\boldsymbol{\nu}$. The parameter κ is estimated by the solution to (15) with $x = \sum(\boldsymbol{\nu}^T \mathbf{y}_i)$, and λ is estimated by $\mathrm{sign}(\sum(\boldsymbol{\nu}^T \mathbf{y}_i))$.

We note that in practice, the case when only the axis of the mean direction is known (Case 2) is rare and the estimate of Fisher [18] has not been used. Indeed, even in this own paper, he does not give any example for this case.

6. Hospers' remanent magnetism data sets

Fisher [18] used two Hospers' remanent magnetism data to illustrate his analysis which we revisit; we label these as Remanent magnetism Data 1 and Remanent magnetism Data 2. In fact, Hospers, in a paper published

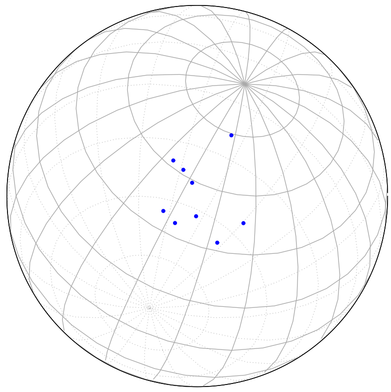


Fig. 3: Hospers‘ Remanent magnetism Data 1: Spherical representation.

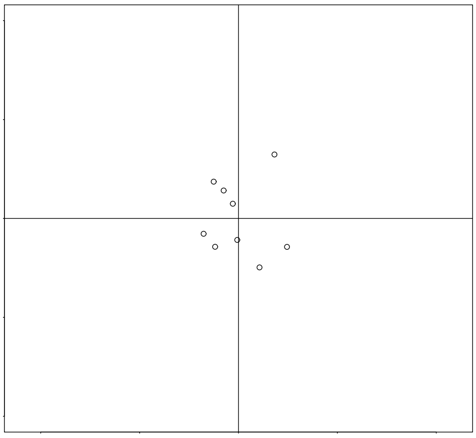


Fig. 4: Hospers‘ Remanent magnetism Data 1: The Lambert equal area projection.

in 1951 in Nature (Hospers [21]) on remanent magnetism of rocks thanks Fisher as follows:

“Prof. R.A. Fisher for the calculation of the estimates of precision. . . ”

That is, Hospers and Fisher had already been applying the methodology to Hospers’ data several years before Fisher published his 1953 paper.

Remanent magnetism Data 1. This data is from the Iceland lava flow of 1947–1948 with sample size $n = 9$ where the full data is given in the paper with the mean direction and is plotted in Fig. 3. It is found that $\hat{\kappa}_{MLE} = 39.53$ so it is highly concentrated and we might have used linear statistics. One way is to use the tangent projection at the mean direction and then use a linear method. Let (θ, ϕ) be the spherical polar coordinates as used in (6). It can be shown that the Lambert equal area projection (see, for example, Mardia and Jupp [36], p. 160) is given by $(2\sin(\theta/2)\cos\phi, 2\sin(\theta/2)\sin\phi)$. Fig. 4 shows the data under this projection and one could use these projected values as drawn from a bivariate normal distribution as κ is large. Equivalently, we can simply take the observed values of (θ, ϕ) from its mean direction and carry out the analysis assuming these are drawn from a bivariate normal distribution. Fisher must have realised this as a possible option for this data after the analysis, and his remanent magnetism Data 2 described below in the paper which is not concentrated so it provides a better illustration.

Remanent magnetism Data 2. For the early Quaternary data of Hospers, $n = 45$, Fisher examined the Hospers's hypothesis that these observations are almost diametrically opposite to the simple dipole field (current) which has the mean direction $(+0.9724, +0.2334, 0)^T$. That is, we need to test the reversal hypothesis, and in the modern terminology, our null hypothesis is

$$H_0 : \boldsymbol{\mu} = (-0.9724, -0.2334, 0)^T.$$

From these 45 observations, the sample mean direction is

$$\hat{\boldsymbol{\mu}}_{MLE} = (-0.9545, -0.2978, +0.0172)^T$$

which is very close to $\boldsymbol{\mu}$ under H_0. Note that the angle between $\boldsymbol{\mu}$ and $\hat{\boldsymbol{\mu}}_{MLE}$ is 3.9° only. In fact, using the Fisher distribution, it is found that H_0 is accepted with a large p-value. Here $\hat{\kappa}_{MLE} = 7.51$ so the data is not concentrated and linear statistics will not be appropriate. Further details are available, for example, in Mardia and Jupp [36].

The topic continues to be of interest and Watson [57] has given more details of Fisher [18] related to paleomagnetism and continental drift. In Rao [48] p. 135, C.R. Rao has commented that

"Fisher (1953) used his model to estimate the true direction (θ) of remnant rock magnetism in lava flow assuming that the observations collected over a geographical area are independent. He did not consider the possibility of spatial correlations which may have some effect on estimation."

Very recently Scealy et al. [49] have reassessed Fisher [18], allowing for the site differences in Hospers type remanent magnetism data.

We end this section with a Fisher's photo (see, Fig. 5) showing one of his field trips for geological data. We quote Box [9], p. 445,

"Fisher continued to the last to be fascinated by the geophysical explorations and to encourage and befriend the explorers."

Note that his geological explorations are not as well known like his genetic experiments (see, for example, one of his genetic experiments on mice, Box [9], pp. 379–381).

7. Fisher on the von Mises distribution

Fisher in his second edition of "Statistical Methods and Scientific Inference", 1959, p. 137, goes back to support his fiducial argument. In so doing, he incidentally derives the von Mises distribution as follows. Readers not familiar with fiducial inference of Fisher, see for example, Efron and Hastie [14], p. 200. From a practical point of view, confidence intervals and fiducial limits have the same objective but their derivations differ.

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$, be a random sample drawn from a bivariate normal distribution with the mean vector $\boldsymbol{\mu}$, unit variances and zero correlation so the density is given by

$$f(\mathbf{x}) = c \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{x} - \boldsymbol{\mu})\}, \quad \mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^2,$$

where c is a normalizing constant. Assume $\boldsymbol{\mu}^T = (\rho \sin\theta,\ \rho\cos\theta)$ with ρ known; the aim is to find the confidence interval (fiducial limits) for $\boldsymbol{\mu}$. The sample mean $\bar{\mathbf{x}}$ is the minimal sufficient statistics for $\boldsymbol{\mu}$ and the ancillary statistic is the resultant $\bar{R} = ||\bar{\mathbf{x}}||$. Then he shows that the distribution of $\hat{\theta}$ given $\bar{R}$ is the von Mises distribution

$$f(\hat{\theta}|\bar{R}) = \frac{\exp\{n\bar{R}\rho\cos(\hat{\theta} - \theta)\}}{2\pi I_0(n\bar{R}\rho)}.$$

He seems not to be aware of the von Mises paper (von Mises [43]).

There are two cases for the conditional distribution depending on $\bar{R}$:

Case A: $\bar{R} < \rho$, and Case B: $\bar{R} > \rho$; and $\bar{R} = \rho$ is a singular case. Suppose under Case A and Case B, we have observed $\bar{R}_1$ and $\bar{R}_2$ respectively. These will give $\hat{\theta}_1$ and $\hat{\theta}_2$ respectively from the von Mises distribution so we

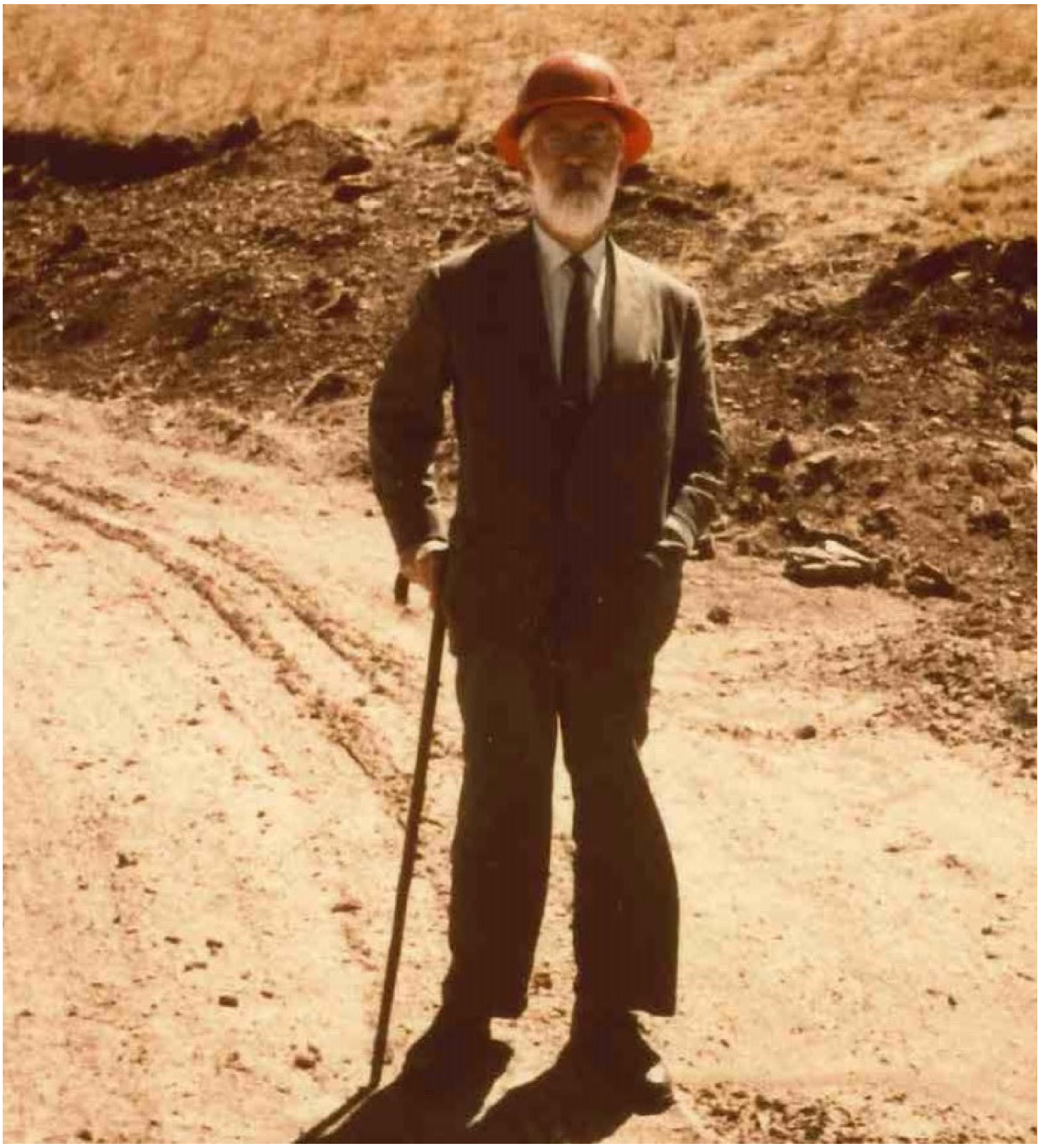


Fig. 5: Fisher 1953 circa in a geological field trip in Australia; author's personal collection.

have for Case A: $\hat{\theta} = \hat{\theta}_1$, and for Case B: $\hat{\theta} = \hat{\theta}_2$. Now our inference can be based on $f(\hat{\theta}_1|\bar{R}_1)$ and $f(\hat{\theta}_2|\bar{R}_2)$. Hinkley [20] has provided more details.

We note an important point which is usually not recognized is that the way Fisher reached the von Mises distribution here has implicitly given a way to construct directional distributions by conditioning an appropriate multivariate normal distribution, for example, the Fisher distribution can be obtained by conditioning a trivariate normal distribution (see, Mardia and Jupp [36], p. 173).

8. The two primary goals of Fisher's 1953 paper

We note from the paper Fisher [18], there are possible two primary goals of the paper.

Goal 1. "To provide methodology for the analysis of more or less widely dispersed measurements of direction such as frequently arise in geology.”

and

Goal 2. “Finally, it is the opinion of the author that certain misapprehensions as to the nature of inductive inference have arisen in examples drawn from the theory of the normal distribution, by reason of the peculiar characteristics of that distribution, and that the examination of these questions, in an analogous though analytically different situation, will exhibit them in a clearer light.”

The Fisher distribution has become a principal tool for analysing spherical data so Goal 1 is achieved but its role as a non-standard example of fiducial inference has not received comparable attention so the assessment of his Goal 2 still continues, see, for example, Bingham [5].

9. Wrapped tangent distributions

For very concentrated data on a sphere, Gauss used a tangent projection, leading to linear statistics. We have already pointed out that Fisher [18] introduced his spherical distribution when the data is not concentrated.

Another approach is to construct directional distributions by wrapping a multivariate distribution in the tangent space of a manifold using an “exponential map” with a base point on the manifold (see, Mardia and Jupp [36], Section 13.4.2). In particular, this gives rise to a distribution on the sphere as an alternative to the Fisher distribution. Recently, this construction has been used for the matrix Fisher type distribution in Benton et al. [2], motivated by problems in machine learning on manifolds. However, the matrix Fisher distribution given in Section 3 is the established distribution in this area.

We will show that their matrix distribution has some serious limitations (though not for a very concentrated data). We first treat spherical case and then comment on their matrix distribution. Let $\boldsymbol{x}$ be a point in the tangent space in $\mathbb{R}^{q-1}$ which can be mapped to a point $\boldsymbol{y}$ on the sphere S_{q-1} as follows; for the circular case ($q = 2$), the tangent space is simply a straight line. Let $f(\boldsymbol{x})$ be a pdf in the tangent space and we will obtain the “wrapped tangent” pdf $g(\boldsymbol{y})$ (with respect to the uniform measure on the sphere) using its exponential map of $\boldsymbol{x}$ on the sphere with origin at the north pole $(0, \ldots, 0, 1)$.

We can write $\boldsymbol{x}$ as $\boldsymbol{x} = r\boldsymbol{v}$ where $r \geq 0$ and $\boldsymbol{v}$ is a unit vector $\in \mathbb{R}^{q-1}$, i.e., a point in S_{q-2}. Further, the point $\boldsymbol{y}$ $\in S_{q-1}$ can be written as

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{v} \sin\theta \\ \cos\theta \end{pmatrix}, \quad 0 \leq \theta \leq \pi. \tag{19}$$

It can be shown that

$$d\boldsymbol{x} = r^{q-2} dr[d\boldsymbol{v}], \quad [d\boldsymbol{y}] = \sin^{q-2}\theta \, d\theta[d\boldsymbol{v}],$$

where $[d\boldsymbol{v}]$ is the uniform measure on S_{q-2}. An isotropic pdf (with respect to $d\boldsymbol{x}$) in $\mathbb{R}^{q-1}$ can be written as $f^*(r) = f(\boldsymbol{x})$ so that

$$\int f^*(r) d\boldsymbol{x} = \int f^*(r) r^{q-2} dr \int [d\boldsymbol{v}] = 1.$$

Similarly, an axially symmetric pdf (wrt $[d\boldsymbol{y}]$) on S_{q-1} can be written as $g^*(\theta) = g(\boldsymbol{y})$ so that

$$\int g^*(\theta)[d\boldsymbol{y}] = \int g^*(\theta) \sin^{q-2}\theta \, d\theta[d\boldsymbol{v}].$$

For $r \geq 0, 0 \leq \theta \leq \pi$, we define the exponential map $\phi(r) = \theta$ as

$$\text{either} \quad r \mod 2\pi = \theta, \text{ or } \quad r \mod 2\pi = -\theta. \tag{20}$$

Note that a point $\boldsymbol{x}$ with its radial part r wraps to a point $\boldsymbol{y}$ with colatitude θ. Take any such point $\boldsymbol{x}$ in the tangent plane and put a thin annulus in $\mathbb{R}^{q-1}$ with radii r and $r + dr$, with volume $r^{q-2}\pi_{q-2} dr$ where π_{q-2} is the surface area of the unit sphere S_{q-2}. This maps to an ”annulus” between two small circles with colatitudes θ and $\theta \pm d\theta$ ($d\theta = dr$). The small circle annulus has volume $\sin^{q-2}\theta \, \pi_{q-2} \, d\theta$. Hence contribution of f at $\boldsymbol{x}$ to g at θ is

$$\{r^{q-2} / \sin^{q-2}\theta\} f^*(r).$$

Now on the mapping (20) we get two different r, leading to the pdf of $\boldsymbol{y}$ given by

$$g(\boldsymbol{y}) = (1/\sin^{q-2}\theta)\sum_{k=0}^{\infty}\{r_{1,k}^{q-2} f(r_{1,k}\boldsymbol{v}) + r_{2,k}^{q-2} f(-r_{2,k}\boldsymbol{v})\}, \quad q \geq 3, \tag{21}$$

where $r_{1,k} = \theta + 2\pi k$, $r_{2,k} = 2\pi(k+1) - \theta$. Except for the term involving $r_{1,0}$ at $\theta = 0$, all the remaining terms have a singularity at $\theta = 0$ and at $\theta = \pi$. In particular, this wrapped distribution on the sphere cannot be unimodal for $q \geq 3$.

For the spherical case ($q = 3$), with $f(\boldsymbol{x})$ the normal distribution, we find that the pdf of the colatitude θ is proportional to

$$\frac{1}{\sin\theta}\sum_{k=0}^{\infty}[(\theta + 2\pi k)\exp\{-(\theta + 2\pi k)^2)/2\sigma^2\} + (2\pi(k+1) - \theta)\exp\{-(2\pi(k+1) - \theta)^2/2\sigma^2\}], \tag{22}$$

where $0 \leq \theta \leq \pi$. Fig. 6 gives plots of the density (22) for three values $\sigma^2 \in \{0.1, 1.0, 2.0\}$. We can see visually that as σ^2 increases the bi-modality increases; though for the concentrated data with $\sigma^2 = 0.1$, it looks unimodal in this figure though there are singularities which are seen clearly for $\sigma^2 = 1.0$ and $\sigma^2 = 2.0$. We can can draw similar figures for $\sigma^2 > 2$ to confirm that as $\sigma^2 > 2$ increases the bi-modality increases.

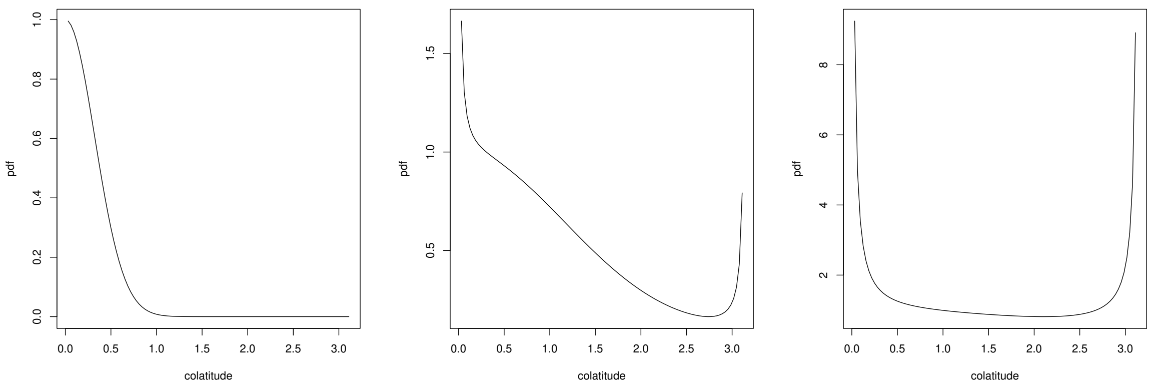


Fig. 6: The pdf of colatitude θ of the wrapped tangent normal distribution with $\sigma^2 \in \{0.1, 1.0, 2.0\}$, respectively.

For the circular case $q = 2$, the wrapped tangent pdf can be written as

$$g^*(\theta) = \sum_{k=0}^{\infty}\{f(\theta + 2\pi k) + f(2\pi(k+1) - \theta)\} \tag{23}$$

so this reduces to the standard circular wrapped distribution except that we have $0 \leq \theta \leq \pi$ as θ deemed to be "colatitude" for the circle. The standard representation by θ^* is the full circle so we have $-\pi \leq \theta^* \leq \pi$ and in the Euclidean representation $(\cos\theta^*, \sin\theta^*)$ is equivalent to $(\cos\theta, \pm\sin\theta)$. In this case, there are no singularities.

Benton et al. [2] have used the wrapped normal distribution on $SO(3)$ using the exponential map to illustrate some of their work in machine learning. However, the most common distribution used on $SO(3)$ is the Fisher matrix distribution which has many desirable properties (see, for example, Mardia and Jupp [36]) whereas there are again inherent singularities in this wrapping for the following reasons (see, also Mardia [32] for some further details). Since $SO(3)$ can be identified with S^3 (after identifying antipodal points, see Mardia and Jupp [36], p. 285), any calculation on $SO(3)$ can be reformulated as one on S^3 and the above discussion on singularities for the spherical pdf applies. This singularity of the pdf is not a practical issue if $f(\cdot)$ is highly concentrated near the origin on the tangent plane, but it is an issue for more diffused distributions and particularly can be an issue when used in mixtures, see, for example, Mardia [32].

Note that the construction above can be generalised as follows. Given a base point m on the Riemannian manifold and a vector $\boldsymbol{x}$ in the tangent space with pdf $f(\cdot)$, the exponential map $\exp_m(\boldsymbol{x})$ yields a point on a manifold and we can go from $f(\cdot)$ to the pdf $g(\cdot)$ on the manifold.

10. My journey into directional statistics

Recently, Mardia [31] has given a glimpse of some of my involvement with Directional Statistics. Few more points will be added here including on how I myself came into this field. In fact, regarding my coming into this field, I have mentioned the following in my first conversation with Mukhopadhyay [45], p. 124:

"In Newcastle, I began developing nonparametric methods by way of Hotelling's T^2 test. But, I was never too keen on working with ranks and asymptotics. In the latter part of 1964, I started thinking about some simple tests. I wanted to have a slick way of doing bivariate nonparametrics and not lose much power. I centered the two distributions, projected them on circles and worked with the uniform scores. Then I examined how these scores in the two populations were distributed. "

Further I added:

" When I did this sort of thing fully in my thesis, I did not know anything about Geoffrey Watson's work on directional data. I did not even know what "directional data" was. Then Robin Plackett pointed out to me that there was a short note (Wheeler and Watson, 1964) proposing a test that came to be known as the "Wheeler–Watson test."... It turned out that I had independently derived the Wheeler–Watson test."

My non-parametric test was published in Mardia [26] but at that point Wheeler–Watson test (Wheeler and Watson [58]) was not known to me so it was not cited in my paper, and the referees of the paper were also not aware of this connection.

There are many other details in Mukhopadhyay [45] including writing of my book Mardia [27], how I was helped by Toby Lewis, and the challenge in getting permission to Stephen's tables for this book. Regarding the tables, I said in the conversation:-

"This book needed many tables and I requested permission from Michael Stephens to reproduce some of the tables from his published works. He was hesitant because he was also writing a book in the same area. As many of those tables were from the journal Biometrika, I then approached its Editor, E. S. Pearson, for permission to reproduce the tables. Pearson said that Michael Stephens could be justified in being hesitant and he hinted that there could be a conflict of interest here because some of these tables were going to be included in the forthcoming E. S. Pearson–H. O. Hartley (1972) volume. He was not too sure that he should give me a "go ahead." I was kept in suspense while I waited with an almost finished book!"

Finally, I went to see E. S. Pearson in 1970/1971 and the outcome I recorded in my conversation:

"When I saw Pearson, I sensed that he was not very comfortable with the whole episode and he was not happy about how the events turned out and became so complicated. He was a very kind person. ... He then suggested that I should recalculate Stephens's tables as much as possible, but he would permit me to reproduce the difficult parts of his tables."

It was some relief and I followed his advice to some extent but Batschelet's book (Batschelet [1]) with the necessary charts also came to the rescue. One example for the permission is related to confidence intervals for the concentration parameter for the von Mises distribution. These tables are given in Stephens [52] (a paper in Biometrika) which would have taken time for me to recompute so used instead the related chart from Batschelet [1] in the first edition (Mardia [27], Appendix 2.11) which were kept even in the second edition (Mardia and Jupp [36]).

This episode might remind some readers of the problem which Fisher had with Karl Pearson in reproducing some tables for his 1925 book as described in Stigler [53]:

"Surely R. A. Fisher played a major role in the canonization of the 5% level as a criterion for statistical significance, although broader social factors were involved. Fisher needed tables for his 1925 book and, evidently, Karl Pearson would not permit the free reproduction of the Biometrika tables, so Fisher computed his own."

The subject continues to grow with several conferences in this area. One of the major sessions in this area took place in 1989 in the European Meeting of Statisticians, Leuven where several leaders in the subject participated including Ted Chang, Nick Fisher, Peter Jupp, John Kent, Toby Lewis, Kanti Mardia, Michael Stephens and Geoffrey Watson (see Fig. 7). The regular conferences in this area include my "Leeds Annual Statistical Research(LASR)" Workshops for many decades, Triennial Conferences "Advances in Directional Statistics (ADISTA)" starting from 2014, "Virtual Symposium on Directional Statistics" from 2000 by Florian Paff.

Most recently, in October 2024, the Nobel Prize in Chemistry was awarded to researchers at DeepMind and the University of Washington for their work on protein structure prediction, an area in Bioinformatics. With

Fig. 7: A rare gathering: August 1989, the European Meeting of Statisticians, Leuven Specialists at its Directional Statistics' session. (From left to right: Michael Stephens, Kanti Mardia, Geoffrey Watson, Nick Fisher, Peter Jupp, Ted Chang, Toby Lewis, John Kent.)

Walter Gilks, I published in 2005 a forward-looking article in Significance,"Meeting the statistical needs of 21st-century science" Mardia and Gilks [35]. The article has the headline

"Kanti Mardia and Walter Gilks consider the future role of statistics in scientific explanation and prediction, through views expressed by eminent scientists, philosophers and statisticians and through their own experience, particularly in the field of bioinformatics."

With various collaborators, I have been pursuing the field of Bioinformatics from 1999 on, with particular attention to protein structure prediction/ Protein Bioinformatics. This research is influenced by my work in Directional Statistics, as I now describe.

10.1. The 2024 Nobel Prizes in Chemistry, and my Work

Some of my work with the highest impact in Directional Statistics turned out to be my bivariate distribution on the torus, now known as the bivariate von Mises distribution, that I introduced in my 1975 discussion paper (Mardia [28]) and more recent developments from it. This impact has been on cutting edge applications in Bioinformatics related to protein structure prediction, which broadly speaking, concerns predicting three dimensional atomic configuration (protein structure/shape) given the amino acids of the protein (simplest stated as a sequence of one-letter codes for the 20 amino acids). Proteins are the workhorses of all living systems; these are large molecules whose function depends on how they *fold* from one dimension (sequences) into structures (3 dimensions). To understand how these molecules work, computer modeling (based on statistics, directly or indirectly) aims to predict the folded protein structure; see, for example, Mardia [30]. My collaboration with researchers in Copenhagen and Leeds led to the method we called PHAISTOS, for protein structure prediction using a probabilistic model of local structure. The CASP competition is the gold standard for evaluating protein structure prediction methods, and our entry into CASP of 2008 (Boomsma et al. [6]), for which we won the best poster award, is described as follows in Borg et al. [8]:

"Here we present our framework, PHAISTOS, and our initial attempts of predicting protein structure from sequence. We tested our approach rigorously by participating in the 8th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP8); a biennial double-blind experiment in protein structure prediction (Moult [44]). We submitted structure predictions for 5 different targets. Two of these targets turned out to be intrinsically unstructured proteins, without a fixed structure. Nonetheless, we managed to predict some important substructures present in these proteins. For the remaining three globular proteins, we successfully predicted the fold for two of them. These results are very encouraging, especially considering the preliminary state of the nonlocal energy function and the fact that we only use the protein sequence as input for the predictions."

PHAISTOS is mainly based on our 2008 paper (Boomsma et al. [7]), published in Proceedings of the National Academy of Sciences (PNAS), which provides the first probabilistic model to predict the structure "locally". The model allows simulation (generation) of realistic protein shapes, and therefore it is a generative, probabilistic model of local protein structure.

We now give its connection with the latest news on the Nobel Prize in Chemistry for 2024

$$https://www.nobelprize.org/prizes/chemistry/2024/summary/$$

The Nobel Prize in Chemistry 2024 was divided, one half awarded to David Baker "for computational protein design", the other half jointly to Demis Hassabis and John Jumper "for protein structure prediction". For a popular and accessible account of this news, see Callaway [10].

In their seminal AlphaFold paper of 2019 (Senior et al. [50]), Demis Hassabis and John Jumper describe the work which went on to win the Nobel Prize. They cite two early important papers, namely our 2008 PNAS paper ([7]) and a second paper of 2010 describing a generative model ([60]), which also cites our PNAS paper. Indeed, our PNAS paper is the first paper making that fundamental step in protein prediction structure methodology.

The relevant paragraph in Demis Hassabis and John Jumper (Senior et al. [50]), page 4 reads as follows.

"Most fragment assembly methods construct fragments by looking up likely fragments based on a database of structures or angles extracted from the Protein Data Bank[15,27], but previous work has also investigated generative [4] and neural-network [36] models of protein structure. " Here, the cited paper "4" is our PNAS paper (Boomsma et al. [7]) and "36" is [60].

Interestingly, the Nobel prize winner David Baker handled our PNAS paper (Boomsma et al. [7]) as the assigned editor.

How and when I became interested in Bioinformatics and how the subject became our priority for more than two decades (starting from 1999) are described in my second conversation with Nitis Mukhopadhyay [46]. We now give some important updates, which are relevant to this section.

In 1975, Mardia [28] introduced the "full" bivariate distribution on the torus (represented by angles (ϕ, ψ)), now known as the bivariate von Mises distribution, with the pdf

$$\begin{aligned} f(\phi,\psi) &= c(\kappa_1,\kappa_2,A)\exp\{\kappa_1\cos(\phi-\mu)+\kappa_2\cos(\psi-\nu)+ \\ &\quad [\cos(\phi-\mu),\sin(\psi-\mu)]A[\cos(\phi-\nu),\sin(\psi-\nu)]^T\}, \end{aligned} \tag{24}$$

where the angles $\phi, \psi \in (-\pi, \pi]$ lie on the torus, and A is a 2×2 matrix; Mardia [29] has obtained an explicit expression for the normalising constant $c(.)$. This model has eight parameters. Various submodels with five parameters have appeared, aiming to mimic the bivariate normal distribution. In our PHAISTOS work, including in the PNAS paper, we used the bivariate cosine model which we introduced in 2007 (Mardia et al. [40]), with density

$$f_c(\phi,\psi) \propto \exp\{\kappa_1\cos(\phi-\mu)+\kappa_2\cos(\psi-\nu)-\kappa_3\cos(\phi-\mu-\psi+\nu)\}; \tag{25}$$

where μ and ν are the mean parameters of ϕ and ψ respectively, and $\kappa_1 \geq 0$ and $\kappa_2 \geq 0$ are the concentration parameters of ϕ and ψ respectively, and κ_3 measures dependence.

Later, we pushed to extend our local methods for protein structure prediction to global prediction using the reference ratio method (RRM), also known as the mean force potential method, for examples, 2011 in Mardia et al. [34] and 2014 in Valentin et al. [54]. This work is ongoing but our progress is slow, partly due to lack of funding. The same RRM methodology to go from local to global prediction was used by Hassabis and Jumper in another seminal paper of 2020 related to deep learning, see Senior et al. [51], Supplement, page 32; this paper also gives some additional details of their entry in CASP13.

It is worth pointing out that the LASR Proceedings highlighted these developments. The figure on the front cover of the 2001 Proceeding is from Demchuk et al. [11], captioned "A picture of the human protein TNF-β consisting of 220 atoms and 707 torsional angles which can be modelled using circular probability distributions;

this is one of the key mediators of AIDS pathogenesis"; the figure on the front cover of the 2009 Proceedings is from Borg et al. [8], captioned "Experimentally determined and predicted structures for two target proteins in the CASP2008 competition"; while the 2011 front cover is from Mardia et al. [34], captioned "The reference ratio method applied to sample compact protein structures". Some of the LASR Proceedings also highlight the pioneering contributions in Molecular Biology by William T Astbury during his long research career at the University of Leeds (1928-1961) who coined the term 'Molecular Biology'. Some details are given in in my second conversation with Nitis Mukhopadhyay [46].

To give a context to this advancement, it is important to understand the terms "Local Protein Structure Prediction" versus "Global Protein Structure Prediction" which we briefly describe.

Local Protein Structure Prediction focuses only on short-range features of the protein backbone. It predicts the three dihedral angles (ϕ, ψ, ω) (also called conformational angles), and possibly the secondary structure (helix, beta sheet, coil), for each amino acid independently or based on its immediate neighbors. We model (ϕ, ψ) by the bivariate cosine model (25) whereas the angle omega, modeling the peptide bond, with binary distribution taking values 0 or π. This local approach does not consider interactions between amino acids that are far apart in the sequence or the overall folding of the protein in 3 dimensions. The result is a detailed description of the backbone's shape in small segments, but without the protein folding into a final 3 dimensional structure.

Global Protein Structure Prediction considers the entire protein chain, including both the "backbone" and the"side chains", to predict the full 3 dimensional atomic arrangement of the protein molecule. It accounts for long-range interactions, such as "hydrophobic packing", "hydrogen bonding" between distant amino acids, and overall compactness. This level of prediction generates the protein's final folded shape, which might be spherical/globular or something else.

To close this section, let me explain our use of the word PHAISTOS. The "phaistos" is a particular disc of fired clay from the island of Crete, dating back to the 2nd millennium BC, Many attempts have been made to decipher the code behind the script engraved on the disk. This name "PHAISTOS" contrasts with the name "ROSETTA" given to the protein design method of David Baker; their program generates many putative solutions and ranks them by energy, see,

https://docs.rosettacommons.org/docs/latest/Home

The Rosetta Stone is a stele of granodiorite inscribed in 196 BC during the Ptolemaic dynasty of Egypt.

Our PHAISTOS is the first probabilistic software framework for local protein structure prediction and is similar in spirit to the later-developed AlphaFold and AlphaFold2 for global protein structure prediction from DeepMind. Table 2 gives a brief time-line of this narrative.

11. R.A. Fisher and D'Arcy Thompson and different manifolds

D'Arcy Thompson is the pioneer of Shape Analysis (see, for example, Dryden and Mardia [13] p. 2.); a subject which has a different manifold than what we have mentioned for Directional Statistics. There seems to be no record of what Thompson wrote to Fisher about a student (Miss Walker) but the letter Fig. 8 from Fisher mentions her, as well as points to investigating the age problem in fish. Fisher wrote on 6 April 1933 to D'Arcy Thompson (see Fig. 9) that the development of adequate techniques as required in the area would be undertaken. But sadly this collaboration via the student did not happen.

We end this section with quotes from Peter Green from Green [19] (his speech at the unveiling of a plaque to R.A. Fisher in London, on May 17th, 2002) which are very pertinent to this paper.

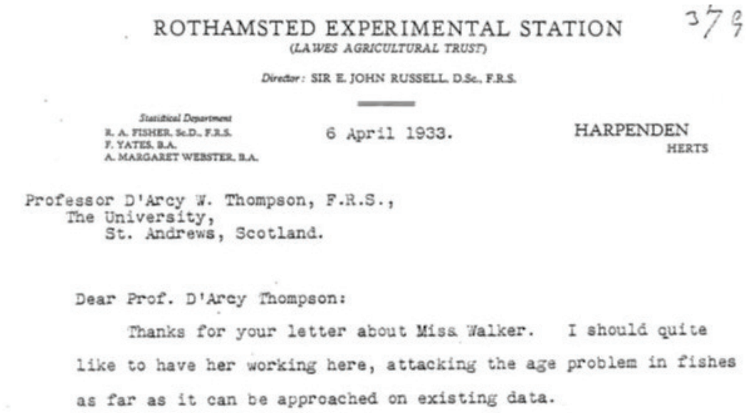
"Even with 40 years of intense research activity in statistics since Fisher's death, it is striking how much of his work is still essentially contemporary. How was Fisher able to achieve so much? The key seems to be his great mathematical ability allied to a passion for contributing to society through science, rather than the pursuit of narrowly technical excellence. That view of statistics in its empirical context is one that continues to be the British tradition, and one that we should be proud to maintain."

Further he said about Fisher

"He has many memorials, including now the plaque that we are unveiling today to commemorate his boyhood home in London. But his greatest memorial is the discipline he did so much to create–recognizably Fisher's statistics, alive and well in the 21st century."

Table 2: A brief time-line of my high impact joint work in protein structure prediction using Directional Statistics in relation to the 2024 Nobel Prizes in Chemistry.

1975	Mardia [28] introduces the bivariate distribution on torus (now known as the bivariate von Mises distribution).
1999	Harshinder Singh from the National Institute for Occupational Safety and Health, West Virginia University, invites me to collaborate on a problem in Protein Bioinformatics.
2001	We publish Demchuk et al. [11], my first paper on Protein Bioinformatics, addressing the protein TNF-beta, the key mediators of AIDS pathogenesis.
2002	Mardia and Westhead [41] is my first Bioinformatics paper with Westhead Group, Biology Faculty, University of Leeds.
2007	Mardia et al. [40] introduce the bivariate cosine model on the torus, a submodel of my bivariate von Mises distribution of 1975.
2008	Collaboration with Thomas Hamelryck and his group in Copenhagen, leading to PHAISTOS Boomsma et al. [6], our probabilistic local structural prediction method, that given a sequence of amino acids, predicts the 3 dimensional protein shapes. This is our introduction of a probabilistic generative model and is **a step change in protein structure prediction methodology.** We enter PHAISTOS into CASP8 and are awarded the prize for the best poster.
2008	We develop Torus-DBN (Dynamic Bayesian Network) based on the bivariate cosine model; Torus-DBN provides local protein structure prediction. Published in Boomsma et al. [7] in Proceedings of National Academy of Sciences (PNAS). **Future Nobel laureate David Baker is the assigned editor for this PNAS paper.**
2009	We review our CASP8 entry for our LASR Workshop paper Borg et al. [8], pointing out our above average performance; the cover page of this LASR Proceedings has the motifs of our predicted protein structures.
2011	Mardia et al. [34] paper introduces use of the reference ratio method (RRM) / potentials of mean force for the first time to do global prediction from local prediction, **a step change in protein structure prediction methodology.**.
2013	Mardia [30] paper ”Statistical approaches to three key challenges in protein structural bioinformatics” appears in the Journal of the Royal Statistical Society, Series C; this paper highlights prediction and is written for a statistical audience.
2014	Our further attempt to do global prediction from local prediction by RRM, Valentin et al. [54], appears in Proteins: structure, function, and bioinformatics.
2019	In their seminal AlphaFold paper, Senior et al. [50], Demis Hassabis and John Jumper cite our 2008 PNAS paper as one of the two earliest important papers utilizing generative modelling for protein structure prediction; AlphaFold is a global generative model.
2020	Hassabis and Jumper discuss their use of our RRM methodology to go from local to global prediction in Senior et al. [51].
2024	The Nobel Prize in Chemistry is awarded to Demis Hassabis and John Jumper, and to David Baker. **Hassabis and Jumper submitted their (global) AlphaFold predictions to CASP13, whereas we submitted our (local) PHAISTOS predictions to CASP8.**



ROTHAMSTED EXPERIMENTAL STATION 379
(LAWES AGRICULTURAL TRUST)

Director: SIR E. JOHN RUSSELL, D.Sc., F.R.S.

Statistical Department
R. A. FISHER, Sc.D., F.R.S.
F. YATES, B.A.
A. MARGARET WEBSTER, B.A.

6 April 1933.

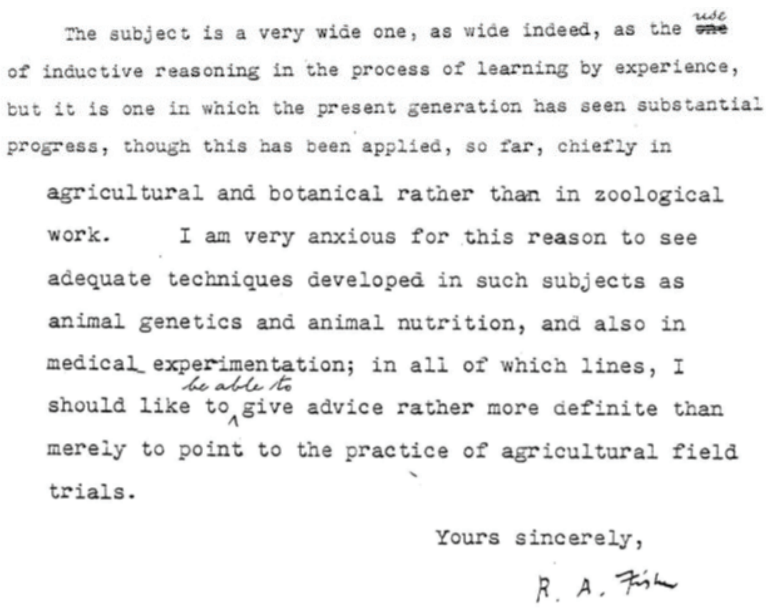
HARPENDEN
HERTS

Professor D'Arcy W. Thompson, F.R.S.,
The University,
St. Andrews, Scotland.

Dear Prof. D'Arcy Thompson:

Thanks for your letter about Miss Walker. I should quite like to have her working here, attacking the age problem in fishes as far as it can be approached on existing data.

Fig. 8: Letter from Fisher to D'Arcy Thompson in 1933.



The subject is a very wide one, as wide indeed, as the ~~one~~ use of inductive reasoning in the process of learning by experience, but it is one in which the present generation has seen substantial progress, though this has been applied, so far, chiefly in agricultural and botanical rather than in zoological work. I am very anxious for this reason to see adequate techniques developed in such subjects as animal genetics and animal nutrition, and also in medical experimentation; in all of which lines, I should like to be able to give advice rather more definite than merely to point to the practice of agricultural field trials.

Yours sincerely,

R. A. Fisher

Fig. 9: Letter (continued) from Fisher to D'Arcy Thompson in 1933.

12. Acknowledgments

I wish to thank Colin Goodall for his kind help in editing the drafts of this paper. I wish also to thank Christopher Bingham, Wouter Boomsma, Thomas Hamelryck, John Kent, Noble Laureate Michael Levitt, Christophe Ley, Charles Taylor and Xiangyu Wu for their helpful comments, and to the Fisher Memorial Trust for inviting me to give the talk. Thanks are also due to the Leverhulme Trust for the Emeritus Fellowship.

References

[1] Batschelet, E. (1965) *Statistical Methods for the Analysis of Problems in Animal Orientation and Certain Biological Rhythms*. Amer. Inst. Biol. Sciences, Washington.

[2] Benton, J., Shi, Y., De Bortoli, V., Deligiannidis, G. and Doucet, A. (2024) From denoising diffusions to denoising markov models (with discussion). *Journal of Royal Statistical Society B*, **85**, 273–339.

[3] Beran, R. J. and Fisher, N. (1998) A conversation with Geoff Watson. *Statistical Science*, **13**, 75–93.

[4] Bharath, K. and Dey, D. (2019) Statistics on non-Euclidean Spaces and Manifolds [Special issue]. *Sankhyā A*, **81**.
[5] Bingham, C. (1980) Distribution on the sphere. *In S. E. Fienberg, D. V. Hinkley (eds) R. A. Fisher: An Appreciation. Lecture Notes in Statistics.*, **1**, 171–181.
[6] Boomsma, W., Borg, M., Frellsen, J., Harder, T., Stovgaard, K., Ferkinghoff-Borg, J., Krogh, A., Mardia, K. V. and Hamelryck, T. (2008) PHAISTOS: Protein structure prediction using a probabilistic model of local structure. *Proceedings of Critical Assessment of Techniques for Protein Structure Prediction*, **CASP8**, 82–83.
[7] Boomsma, W., Mardia, K. V., Taylor, C. C., Ferkinghoff-Borg, J., Krogh, A. and Hamelryck, T. (2008) A generative, probabilistic model of local protein structure. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 8932–8937.
[8] Borg, M., Boomsma, W., Ferkinghoff-Borg, J., Frellsen, J., Harder, T., Mardia, K. V., Rogen, P., Stovgaard, K. and Hamelryck, T. (2009) A Probabilistic Approach to Protein Structure Prediction: PHAISTOS in CASP8. In *LASR 2009 Proceedings, Leeds Univ. Press, Leeds* (eds. A. Gusnanto, K. V. Mardia and C. J. Fallaize), 65–70.
[9] Box, J. F. (1978) *R A Fisher, The Life of a Scientist*. Wiley, New York.
[10] Callaway, E. (2024) Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures. *Nature*, **634**, 525–526.
[11] Demchuk, E., Hnizo, V., Mardia, K. V., Sharp, D. S. and Singh, H. (2001) Statistics and molecular structure of biological macromolecules. In *LASR 2001 Proceedings - Functional and Spatial Data Analysis, Leeds Univ. Press, Leeds* (eds. K. V. Mardia and R. G. Aykroyd), 9–14.
[12] Diaconis, P. (1998) *Group Representations in Probability and Statistics.* Lecture notes-monograph series, volume **11**, Institute of Mathematical Statistics.
[13] Dryden, I. L. and Mardia, K. V. (2016) *Statistical Shape Analysis with Applications in R, Second Edition*. Wiley.
[14] Efron, B. and Hastie, T. (2016) *Computer Age Statistical Inference: Algorithms, Evidence and Data Science*. Cambridge University Press.
[15] Fisher, N. I. (1993) *Statistical Analysis of Circular Data*. Cambridge University Press.
[16] Fisher, N. I., Lewis, T. and Embleton, B. J. J. (1987) *Statistical Analysis of Spherical Data*. Cambridge University Press.
[17] Fisher, R. A. (1929) Tests of significance in harmonic analysis. *Proceedings of the Royal Society, Series A*, **125**, 54–59.
[18] — (1953) Dispersion on a sphere. *Proc. Roy. Soc., A*, **217**, 349–393.
[19] Green, P. J. (2003) Notes on the life and work of R. A. Fisher. *Journal of the Royal Statistical Society D (The Statistician)*, **52**, 299–301.
[20] Hinkley, D. V. (1980) Fisher's development of conditional inference. *In S. E. Fienberg, D. V. Hinkley (eds) R. A. Fisher: An Appreciation. Lecture Notes in Statistics.*, **1**, 101–108.
[21] Hospers, J. (1951) Remanent magnetism of rocks and the history of the geomagnetic field. *Nature*, **168**, 1111–1112.
[22] Jammalamadaka, S. R. and Sengupta, A. (2001) *Topics in Circular Statistics*. Chapman and Hall/CRC.
[23] Langevin, P. (1905) Magnétisme et théorie des électrons. *Ann. de Chim. et de Phys.*, **5**, 70–127.
[24] Ley, C. and Verdebout, T. (2017) *Modern Directional Statistics*. Chapman and Hall/CRC.
[25] — (2018) *Applied Directional Statistics, Modern Methods and Case Studies*. Chapman and Hall/CRC.
[26] Mardia, K. V. (1967) A non-parametric test for the bivariate two-sample location problem. *Journal of Royal Statistical Society B*, **29**, 320–342.
[27] — (1972) *Statistics of Directional Data*. Academic Press, London.
[28] — (1975) Statistics of directional data (with discussion). *Journal of Royal Statistical Society B*, **37**, 349–393.
[29] — (2010) Bayesian analysis for bivariate von Mises distributions. *Journal of Applied Statistics*, **37**, 515–528.
[30] — (2013) Statistical approaches to three key challenges in protein structural bioinformatics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **62**, 487–514.
[31] — (2021) Comments on: Recent advances in directional statistics. *Test*, **30**, 59–63.
[32] — (2024) Discussion to the paper:"From denoising diffusions to denoising markov models" by Benton et al (2023). *Journal of Royal Statistical Society B*, **85**, 311–312.
[33] — (2024) Fisher's pioneering work on discriminant analysis and its impact on AI. *Journal of Multivariate Analysis*, **203**, 105341.
[34] Mardia, K. V., Frellsen, J., Borg, M., Ferkinghoff-Borg, J. and Hamelryck, T. (2011) A statistical view on the reference ratio method. In *LASR 2011 Proceedings, Leeds Univ. Press, Leeds* (eds. A. Gusnanto, K. V. Mardia, C. J. Fallaize and J. Voss), 55–61.
[35] Mardia, K. V. and Gilks, W. (2005) Meeting the statistical needs of 21st-century science. *Significance*, **2**, 162–165.
[36] Mardia, K. V. and Jupp, P. E. (2000) *Directional Statistics (First edtion K. V. Mardia, 1972).* Wiley.
[37] Mardia, K. V., Kent, J. T. and Laha, A. K. (2016) Score matching estimators for directional distributions. *ArXiv (link).*
[38] Mardia, K. V. and Sriram, K. (2020) Families of discrete circular distributions with some novel applications. *Arxiv (version 1: (link)).*
[39] — (2023) Families of discrete circular distributions with some novel applications. *Sankhyā A*, **85**, 1–42.
[40] Mardia, K. V., Taylor, C. C. and Subramaniam, G. K. (2007) Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics*, **63**, 505–12.
[41] Mardia, K. V. and Westhead, D. R. (2002) New major challenges in bioinformatics. In *LASR 2002 Proceedings, Leeds Univ. Press, Leeds* (eds. R. Aykroyd, K. V. Mardia and P. McDonnell), 9–15.
[42] Mardia, K. V., Wiechers, H., Eltzner, B. and Huckemann, S. F. (2022) Principal component analysis and clustering on manifolds. *Journal of Multivariate Analysis*, **188**, 104862.
[43] von Mises, R. (1918) Über die "Ganzzahiligkeit" der Atomgewichte und verwvandte Fragen. *Physikalische Zeitschrift*, **19**, 490–500.
[44] Moult, J. (2006) Rigorous performance evaluation in protein structure modelling and implications for computational biology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **361**, 453–458.
[45] Mukhopadhyay, N. (2002) A conversation with Kanti Mardia. *Statistical Science*, **17**, 113–148.
[46] — (2015) A conversation with Kanti Mardia: Part II. In *Geometry Driven Statistics: Festschrift Volume in honour of K. V. Mardia* (eds. I. L. Dryden and J. T. Kent), 59–85. Wiley.
[47] Pewsey, A. and García-Portugués, E. (2021) Recent advances in directional statistics. *Test*, **30**, 1–58.
[48] Rao, C. R. (2008) Life and work of Ronald Aylmer Fisher. *Journal of Statistical Theory and Practice*, **2**, 131–141.

[49] Scealy, J. L., Heslop, D. Liu, J. and Wood, A. T. A. (2022) Directions old and new: Palaeomagnetism and Fisher (1953) meet modern statistics. *International Statistical Review*, **90**, 237–258.

[50] Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ..., Jones, D. T., Silver, D., Kavukcuoglu, K. and Hassabis, D. (2019) Protein structure prediction using multiple deep neural networks in the 13th critical assessment of protein structure prediction (CASP13). *Proteins: structure, function, and bioinformatics*, **87**, 1141–1148.

[51] — (2020) Improved protein structure prediction using potentials from deep learning. *Nature*, **577**, 708–710.

[52] Stephens, M. A. (1969) Tests for the von mises distribution. *Biometrika*, **56**, 149–160.

[53] Stigler, S. (2008) Fisher and the 5% Level. *Chance*, **21**, 12–12.

[54] Valentin, J., Andreetta, C., Boomsma, W., Bottaro, S., Ferkinghoff-Borg, J., Frellsen, J., Mardia, K. V., Tian, P. and Hamelryck, T. (2014) Formulation of probabilistic models of protein structure in atomic detail using the reference ratio method. *Proteins: structure, function, and bioinformatics*, **82**, 288–299.

[55] Watson, G. S. (1956) Analysis of dispersion on a sphere. *Monthly Notices Royal Astronomical Society, Geophysical Supplement*, **7**, 153–159.

[56] — (1990) A boy from the bush. In *The Art of Statistical Science: A Tribute to G.S. Watson* (ed. K. V. Mardia), xxix–xlvii. Wiley.

[57] — (1998) On the role of statistics in the palaeomagnetic proof of continental drift. *Canadian Journal of Statistics*, **26**, 383–392.

[58] Wheeler, S. and Watson, G. S. (1964) A distribution-free two-sample test on a circle. *Biometrika*, **51**, 256–257.

[59] Wiechers, H., Eltzner, B., Mardia, K. V. and Huckemann, S. F. (2023) Learning torus PCA-based classification for multiscale RNA correction with application to SARS-CoV-2 . *Journal of Royal Statistical Society, Series C*, **72**, 271–293.

[60] Zhao, F., Peng, J. and Xu, J. (2010) Fragment-free approach to protein folding using conditional neural fields. *Bioinformatics (Oxford, England)*, **26**, i310–i317.