PROBABILITY AND MATHEMATICAL STATISTICS

A Series of Monographs and Textbooks

STATISTICS OF DIRECTIONAL DATA

This is a volume in
PROBABILITY AND MATHEMATICAL STATISTICS

A Series of Monographs and Textbooks

Editors: Z. W. Birnbaum and E. Lukacs

A complete list of titles in this series appears at the end of this volume.

STATISTICS OF DIRECTIONAL DATA

K. V. MARDIA

Department of Mathematical Statistics
The University, Hull, England

ACADEMIC PRESS, INC.

(Harcourt Brace Jovanovich, Publishers)

London Orlando San Diego New York
Toronto Montreal Sydney Tokyo

ACADEMIC PRESS INC. (LONDON) LTD.
24/28 Oval Road,
London NW1

United States Edition published by
ACADEMIC PRESS, INC.
Orlando, Florida 32887

Library of Congress Catalog Card Number: 72–77853
ISBN: 0–12–471150–2

PRINTED IN THE UNITED STATES OF AMERICA

84 85 86 87 9 8 7 6 5 4 3 2

To PAVAN

This page intentionally left blank

PREFACE

The aim of this book is to give a systematic account of statistical theory and methodology for observations which are directions. The directions are usually regarded as points on a circle in 2 dimensions or on a sphere in 3 dimensions. While unifying the work of many researchers, the presentation keeps the requirements of students of mathematical statistics and of scientific workers from various disciplines very much in the foreground. To this end, the book gives the underlying theory of each technique, and applications are illustrated by working through a number of real life examples. Some chapters are devoted exclusively to applications. This presentation is adopted because of the belief that students will appreciate the techniques fully only after grasping the motivation behind them, whereas scientific workers will need some knowledge of the underlying assumptions in order to apply the methods safely.

Chapters 1–7 deal with the statistics of circular data while Chapters 8 and 9 deal with the statistics of spherical data. Chapters 1 and 2 are primarily concerned with the diagrammatical representations of circular data and diagnostic tools, respectively. Chapter 1 contains examples of angular data from various scientific fields. Mathematical justifications of the results used in Chapter 2 are differed until Chapter 3. Chapters 3 and 4 are on probability theory on the circle; several probability models are discussed in Chapter 3 including the von Mises distribution which has the same statistical role on the circle as the normal distribution on the line (see Section 3.4.10). Chapter 4 includes certain sampling distributions for von Mises populations. Chapters 5–7 are on inference on the circle and they deal with estimation, hypothesis testing for samples from von Mises distributions and non-parametric methods respectively. Chapter 8 deals with diagrammatical representations, diagnostic tools, probability models and certain sampling distributions on the sphere. The Fisher distribution is of central importance on the sphere and Chapter 9 deals with inference problems associated with samples from this distribution. Appendix 1 provides a list of formulae for Bessel functions which are collected together to facilitate the manipulation of the von Mises distribution. Appendices 2 and 3 give tables for fitting, estimation and testing problems on the circle and on the sphere respectively. The bibliography is fairly exhaustive as far as theoretical papers are concerned and also serves as an author index. In addition to a subject index, an index of the principal notation is also given with an indication of their meaning.

The book assumes a basic knowledge of mathematical statistics at undergraduate level. Calculus and matrix algebra are used at a level similar to that required for undergraduate statistics courses. To understand only the statistical applications, elementary training in mathematics and statistics is sufficient.

The material of this book can be used as (i) a graduate text (ii) a research monograph (iii) a user's manual or (iv) an elementary methods text. We now suggest how to use this book for each of these four purposes and note some points which may prove helpful.

(i) *As a graduate text*

A course of two hours per week for two terms should cover Chapters 3–9 (excluding Sections 4.6, 8.2, 8.3, 8.6.3b), but if a shorter course is preferred so as to supplement an existing course, there are several possibilities available. For examples, a course on distribution theory may include Chapter 3 (Sections 3.1–3.3, Sections 3.4.2–3.4.5, 3.4.8d–e, 3.4.9, 3.5, 3.6), Chapter 4 and Chapter 8 (Sections 8.5–8.8), while a course on statistical inference may include Chapter 3 (Sections 3.3, 3.4.4, 3.4.9), Chapters 5–7, Chapter 8 (Sections 8.4, 8.5) and Chapter 9. The results from the excluded sections may be quoted where relevant. I have found in giving the latter course to M.Sc. students that a minimum of two hours per week for one term is required.

While treating the circular case, we have alluded to the corresponding methods on the line and it may be useful to elaborate these points. In a similar way, when dealing with the spherical case, the corresponding results for the circular case are referred to for details and illumination. No mathematical examples are spelled out formally but at various places we have left out detailed derivations and the filling in of these details using the hints provided, constitutes good example material. The von Mises and the Fisher distributions are particular cases of a distribution on a p-dimensional hypersphere (see Section 8.8). As a consequence, various results for the two distributions can be unified but this is not done here partly because it has no known application for higher dimensions. Nevertheless, this unification also provides material for good examples.

(ii) *As a research monograph*

Specialists will find a good coverage of existing results as well as various new results. A serious attempt is made to unify the results and many proofs have been simplified. Several new results are included, primarily to fill gaps in the subject. [In particular, the following results are new. Sections 2.5.1, 2.6.1–3, 2.7.2, 3.3.3–4, 3.4.9f (entropy), 3.4.9h, 3.5.2, 3.7; Eqns (4.2.6), (4.2.28); Section 4.3.3 (formulation), Eqns (4.5.7), (4.5.12); Sections 4.5.5, 4.6.2, 4.7.1; Eqns (4.7.11), (4.8.8); Sections 4.9.4, 5.1, 5.6, 6.2.2a, 6.2.2b(1), 6.3.1c, 6.3.2b–2c (cases I, II); optimum properties in Sections 6.2.2b(2), 6.2.3a, b; tests defined by (6.3.37), (6.3.38), (6.4.6), (6.4.11), (6.4.12); Eqns (7.4.15), (7.4.17), (8.5.21), (8.6.8), (8.6.17), (8.6.39), (9.3.7), (9.3.16); Section 9.4.2a (cases I, II), tests defined by (9.4.20), (9.4.21), (9.5.3), (9.5.8)–(9.5.10).]. Almost all theoretical contributions are reviewed.

(iii) *As a user's manual*

It is assumed that the reader has a familiarity with statistical terms on the line such as probability distributions (normal, χ^2 and F distributions in particular), confidence intervals and statistical tests to the extent that he will be able to formulate his problem in a statistical framework. For example, whether he needs a one-sample or a two-sample test, whether he is interested in comparing two means or two variances etc. etc. In addition to the usual mathematical knowledge required for elementary statistics, his familiarity with the trigonometric functions is assumed.

Analogues of almost every univariate method are available for the circular and the spherical cases. There are also techniques which arise only for the circular and the spherical cases such as testing for uniformity. The introductory remarks to Chapters 6–9 contain comments on such situations. All techniques given in the book are illustrated fully with the help of real data, drawn from many fields of application. Appropriate tables are provided in Appendices 2 and 3. Although these examples deal with data obtained from biology, geology, meteorology, medicine, crystallography, astronomy ect., workers in other fields should have no difficulty in translating them into terms which are more familiar to their own disciplines. Section 1.5 may be useful for this purpose since it summarizes applications from various fields.

The von Mises distribution is as important on the circle as the normal distribution is on the line (Section 3.4.9) while the Fisher distribution is important on the sphere (Section 8.5). Examples of fitting these distributions are provided (Sections 5.4, 9.2). There are various one-sample non-parametric tests and the comparisons given in Section 7.2.6 will be useful for deciding which one to use for any given problem. There exist various co-ordinate systems on the sphere and Section 8.2 relates some of them to the standard polar co-ordinates.

As yet there are no standard statistical packages available on computers for directional data techniques but programming them is not too difficult. For large data, as well as for techniques involving eigen values and vectors, such programming is desirable but for small samples, the use of a desk calculator is adequate. The amount of labour involved will be minimized by using the tabular methods illustrated in the examples.

(iv) *As an elementary methods text*

A course which assumes a basic knowledge of statistical methods on the line can be developed. Chapters 1 and 2, Section 3.4.9 (explaining the von Mises distribution with the help of the figures), the solved examples in Chapters 5–7, Sections 8.1, 8.4, the figures of Section 8.5 and the solved examples in Chapter 9 should provide good material. The course must give emphasis to geometrical pictures. Some prior knowledge of the trigonometric functions is needed. Various points mentioned under (iii) are also relevant.

The author will be most grateful to readers who draw his attention to any errors, or obscurities in the book or suggest any other improvements.

February, 1972 K. V. Mardia

ACKNOWLEDGEMENTS

I am deeply indebted to all those statisticians who developed techniques for analysing directional data and to all those scientific workers who drew attention to the need for such techniques; in particular, I should mention E. Batschelet, R. A. Fisher, E. J. Gumbel, E. Irving, M. A. Stephens, G. S. Watson and E. J. Williams.

I am also grateful to those authors and editors who generously granted me permission to reproduce certain charts, figures and tables which add greatly to the value of the book and, likewise, it is a pleasure to thank those authors who sent me copies of their unpublished work. Where such material appears in the book, credit is given to the original source.

My greatest debt is to Professor T. Lewis for his valuable help and encouragement throughout this project. Thanks are also due to my other colleagues Dr. M. S. Bingham, Dr. E. A. Evans and Dr. J. W. Thompson for their support, and to Professor E. Batschelet, Dr. C. Bingham and Dr. J. S. Rao for their help. I am also grateful to Professor E. S. Pearson for his helpful advice. Parts of the draft were read by Dr. M. S. Bingham and Mr. B. D. Spurr and their comments led to substantial improvements. Finally, my thanks go to Miss T. Blackmore and Miss L. Robinson for typing the difficult manuscript with great skill.

K. V. M.

This page intentionally left blank

CONTENTS

This page intentionally left blank

INTRODUCTION

1. The Background

The interest in developing techniques to analyse directional data is as old as the subject of mathematical statistics itself. Indeed, the theory of errors was developed by Gauss primarily to analyse certain directional measurements in astronomy. It is a historical accident that the observational errors involved were sufficiently small to allow Gauss to make a linear approximation and, as a result, he developed a linear rather than a directional theory of errors. In many applications, however, we meet directional data which cannot be treated in this manner, e.g. orientation data in biology, dip and declination data in geology, seasonal fluctuation data in medicine and wind direction data in meteorology. The temptation to employ conventional linear techniques can lead to paradoxes; for example, the arithmetic mean of the angles $1°$ and $359°$ is $180°$ whereas by geometrical intuition the mean ought to be $0°$.

2. Directions

Directions may be visualized in a space of any number of dimensions but practical situations almost invariably give rise to directions in two or three dimensional space where they may be represented by points on the circumference of a circle or on the surface of a sphere respectively. In general, directions can be regarded as points on the surface of a hypersphere. Sometimes, the orientation of an undirected line is of interest, e.g. in determining a crystal axis, or in dealing with hinge lines. Such observations can be described as *axes* rather than directions. In representing axial data on the surface of a hypersphere, the undirected lines are extended to cut the hypersphere and no distinction is made between the two diametrically opposite points which correspond to any given undirected line. Alternatively, axial data can be represented as a set of points on the surface of a semi-hypersphere.

3. Linear versus Directional Statistics

Since the circle is a closed curve but the line is not, we anticipate differences between the theories of statistics on the line and on the circle. For example, it is necessary to define distribution functions, characteristic functions and moments in a way that takes account of the natural periodicity of the circle. The compactness of the circle leads to a simpler treatment of the convergence in law of random variables. The different algebraic structures of the circle and the line, the circle having only one operation (viz., addition modulo 2π) and the line having two operations (viz., addition and multiplication), produce different forms of central limit theorems and stability. On the circle, there are difficulties in ordering observations and this necessitates the use of special tools in non-parametric methods. Similar remarks apply to the hyperspherical case when it is compared with the usual multivariate analysis for Euclidean spaces.

4. Historical Notes

Early developments in the subject were mainly for uniformly distributed random vectors. As early as 1734, Daniel Bernoulli discussed a solution to the problem of whether the close coincidence of the orbital planes of the six planets then known could have arisen by chance. Each orbital plane can be identified by its normal which in turn corresponds to a point on the surface of a sphere. Bernoulli's assertion then amounts to the hypothesis that these points are uniformly distributed on the surface of the unit sphere. A natural test-statistic is the resultant length of the normal vectors to the orbital planes. Rayleigh (1880) was the first to study the distribution of the resultant length of such vectors (in two dimensions) for a problem in sound although statisticians were not generally aware of his solution until 1905 when he responded to a letter of K. Pearson in *Nature* which posed the problem of the isotropic random walk on a circle. Rayleigh's solution was approximate but an exact solution was produced promptly by Kluyver (1905). K. Pearson (1906) provided another approximate solution. Rayleigh (1919) gave an exact solution to the problem of the uniform random walk on the sphere together with an approximation for large samples.

The underlying population distributions in the above work were of course uniform. Non-uniform distributions started appearing only after 1900. Interest in Brownian motion on the circle and the sphere led to wrapped normal distributions (cf. Perrin, 1928). Von Mises (1918) in investigating whether the atomic weights were integers subject to errors introduced a distribution on the circle by using a characterization analogous to the Gauss characterization of the normal distribution on the line. Langevin (1905) in the study of magnetism introduced a distribution on the sphere which was shown by Arnold (1941) to possess a Gauss-type characterization on the sphere. Meanwhile, the need for techniques of analysing directional data was strongly felt especially by the earth scientists (see Steinmetz, 1962 for references). However, real progress was made neither in statistical inference for this distribution on the sphere, nor in the subject of orientation analysis as a whole, until an epoch making paper of R. A. Fisher appeared in 1953. He was attracted to this field by a problem of Hospers (1955) in paleomagnetism. At the same time, E. J. Gumbel, D. Durand and J. A. Greenwood were producing results for the von Mises distribution and Fisher's method helped Greenwood and Durand (1955) to make progress towards a distribution theory for the circular case.

A remarkable paper by Watson and Williams (1956) not only unified the inference problems for the von Mises and the Fisher distributions but also brought a wealth of new results and ideas. Since then, thanks mostly to G. S. Watson and his co-workers, the growth and dissemination of the subject have been rapid; Watson introducing analysis of variance type techniques, various parametric and non-parametric tests, M. A. Stephens making various other major contributions to the small sample theory and applications, and J. Beran unifying the treatment of non-parametric tests. E. Irving, who had attracted the attention of G. S. Watson in 1956 to the subject, illuminated various techniques on the sphere in his book of 1964 which was written especially for geologists. E. Batschelet unified and simplified methodology for the circular case in his 1965 monograph primarily for biologists. There are other notable contributions to this field in the last two decades including those of B. Ajne, T. W. Anderson, C. Bingham, E. Breitenberger, E. J. Burr, E. Dimroth, T. D. Downs, A. L. Gould, J. L. Hodges, Jr., N. L. Johnson, N. H. Kuiper, U. R. Maag, E. S. Pearson, C. R. Rao, J. S. Rao, S. Schach, B. Selby and G. J. G. Upton.

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.

R. A. Fisher

1

ANGULAR DATA AND FREQUENCY DISTRIBUTIONS

1.1 INTRODUCTION

Angular observations arise from random experiments in various different ways. They may be direct measurements such as wind directions or vanishing angles of migrating birds. They may arise indirectly from the measurement of times reduced modulo some period and converted into angles, e.g. the incidence rate of a particular disease in each calendar month over a number of years. Rounding errors in numerical calculations converted into angles also form such observations.

We may regard the angular observations as observations on a circle of unit radius. A single observation $\theta°$ $(0° < \theta° \leqslant 360°)$ measured in degrees is then a unit vector and the data can be described as circular data. Further, $\theta°$ represents the angle made by the vector with the positive x-axis in the anti-clockwise direction. The cartesian co-ordinates of the vector are $(\cos\theta°, \sin\theta°)$ while the polar co-ordinates are $(1, \theta°)$. If the vector is not directed, i.e. if the angles $\theta°$ $(0 < \theta° \leqslant 180°)$ and $180° + \theta°$ are not distinguished, the data can be described as axial data.

1.2. DIAGRAMMATICAL REPRESENTATION

1.2.1. Ungrouped Data

The angular observations can be represented in two ways. They can be represented by points on the circumference of a unit circle, the same mass being assigned to each observation. Figure 1.1 illustrates this method for the following example.

Example 1.1. A roulette wheel was allowed to revolve and its stopping positions were measured in angles with a fixed direction. The measurements in 9 trials were 43°, 45°, 52°, 61°, 75°, 88°, 88°, 279°, 357°. Its representation in Fig. 1.1 shows that the wheel seems to have a *preferred* direction.

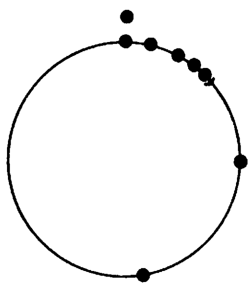

FIG. 1.1. Circular plot of the roulette data of Example 1.1.

Alternatively, we can represent the data by drawing the radii of a unit circle, obtained by joining the origin to the observed points on the circumference. Figure 1.2 shows this representation of Example 1.1. Its relation with the rose diagram (see Section 1.2.2) can be noted; the vectors are of unit length in our case. The first method is analogous to that which is commonly used for data on a line and is preferred.

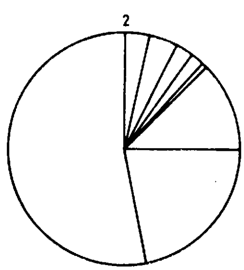

FIG. 1.2. Rose diagram of the roulette data of Example 1.1.

1.2.2. Grouped Data

Circular Histograms. Angular data can be grouped by adopting the same procedure as on the real line. The range (0°, 360°) can be divided into a certain number of class-intervals and the frequency corresponding to each class can be counted. Choice of class-limits and length of class-intervals require the same consideration as for linear data. However, in the circular case there can be an interval such as 330°–30° which contains simultaneously the angles 359° and 0°.

TABLE 1.1 Vanishing angles of 714 British mallards (adapted from Matthews, 1961)

Direction	Number of birds	Direction	Number of birds
0°–	40	180°–	3
20°–	22	200°–	11
40°–	20	220°–	22
60°–	9	240°–	24
80°–	6	260°–	58
100°–	3	280°–	136
120°–	3	300°–	138
140°–	1	320°–	143
160°–	6	340°–	69
		Total	714

Table 1.1 shows the frequencies of the vanishing angles of 714 non-migratory British mallards with 0° as the north. The birds were displaced under sunny conditions from Slimbridge, Gloucestershire, by distances of between 30 km. and 250 km. in different directions over one year. We can represent this data on a histogram similar to that used on a line. We take a unit circle and, corresponding to each interval, construct a block on its circumference whose area is proportional to the frequency in that interval. Figure 1.3 gives a *circular histogram* of the data given in Table 1.1. The corresponding frequency polygon can be constructed by joining the mid-points of the summits of the blocks. The latter presentation on a circle is complicated and does not elicit additional information.

Linear Histograms. Another useful representation is to unroll the circular histogram so that it sits on a segment of width 360°. The point of cut used in unrolling the circle should be selected carefully. If the data has a mode (a preferred direction) then it is wiser to use a cut such that the centre of the *linear* histogram approximately corresponds to this mode. A cut near the mode would give an erroneous impression of the data. Further,

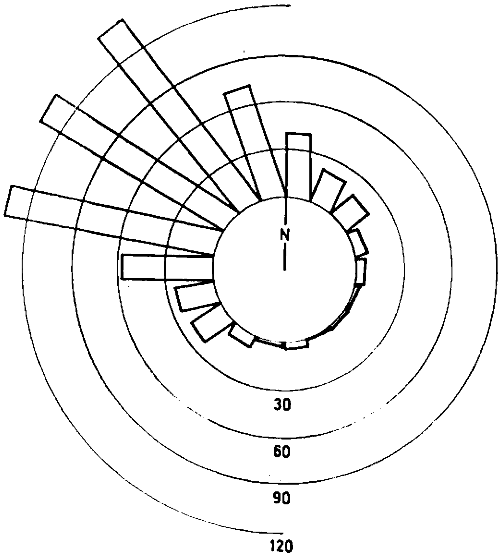

FIG. 1.3. Circular histogram of the mallard data of Table 1.1.

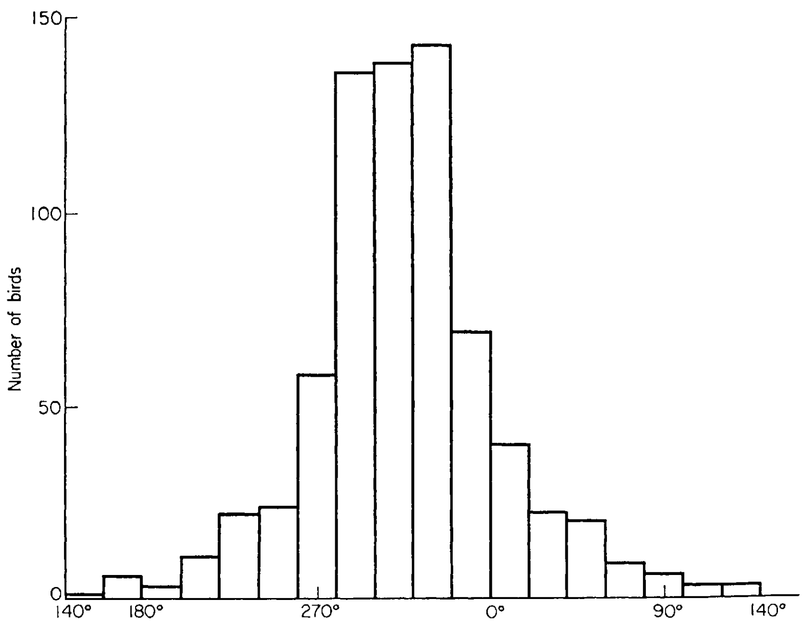

FIG. 1.4. Linear histogram of the mallard data of Table 1.1.

the ends of the axis exist only as a convenience and a linear histogram for circular data should be judged after imagining it to be wrapped around the circumference of a circle. To emphasize this fact, the first block can be repeated at the other end. Further, the linear histogram is preferred to the circular histogram partly because of our competence to interpret such histograms. Figure 1.4 shows a linear histogram of the data in Table 1.1.

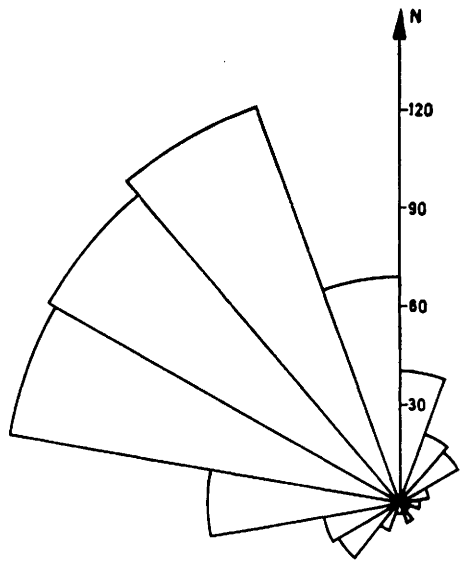

FIG. 1.5. Rose diagram of the mallard data of Table 1.1.

Rose Diagrams. Another natural representation is a rose diagram. Corresponding to each interval, we construct a sector with apex at the origin, radius proportional to the class frequency and arc substending the class interval. Figure 1.5 gives a rose diagram of the data in Table 1.1. The angles are measured in the clockwise direction with the north as 0°. If the observations take values in the interval (0°, 180°) then linear histograms can be drawn as usual. To draw a rose diagram for such data, we can construct that half of the diagram corresponding to the range 0° to 180° and obtain the other half by reflection through the origin. Figure 1.6 gives a rose diagram for the data in Table 1.6.

The area of each sector in the type of rose diagram described above varies as the square of the frequency. In order to make the areas proportional to the frequencies instead of the frequencies squared, the square roots of the frequencies should be taken as the radii. The resulting diagram can be described as an equi-areal rose diagram. The graphic comparisons between observed and expected frequencies in such presentations is hardly satisfactory, since a given arithmetic difference between the frequencies is represented by decreasing intervals on the polar radii as the frequency increases. Of course, linear histograms conserve areas and are comparatively simple to construct.

Rose diagrams and circular histograms are sometimes described as polar-wedge diagrams.

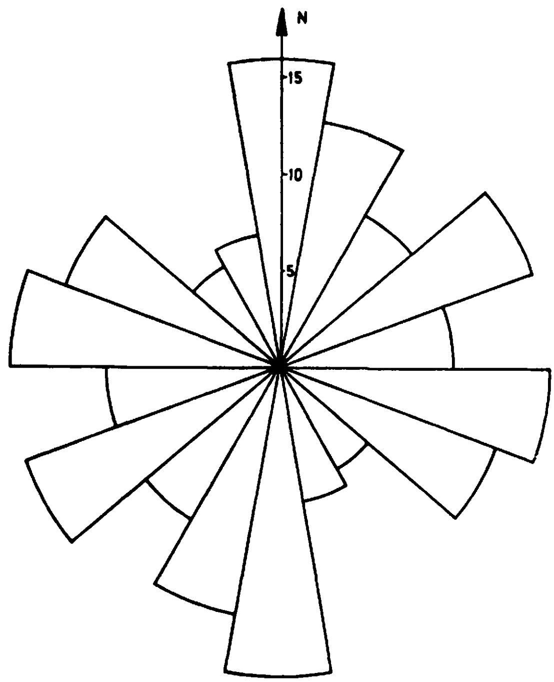

FIG. 1.6. Rose diagram of the pebbles data of Table 1.6.

1.3 INTERRELATIONS BETWEEN DIFFERENT UNITS OF ANGULAR MEASUREMENT

1.3.1. Radians

For theoretical purposes, it is preferred that the angle θ° in degrees is converted into the angle θ in radians. We, of course, have

$$\theta^\circ = 180\,\theta/\pi, \qquad \theta = \pi\theta^\circ/180. \tag{1.3.1}$$

The range of θ is $0 < \theta \leqslant 2\pi$ corresponding to the range $0^\circ < \theta^\circ \leqslant 360^\circ$ of

$\theta°$. If an arc of a circle with unit radius substends an angle of θ radians at the centre of the circle, then the length of the arc is θ.

TABLE 1.2 Orientations of sand-grains of Recent Gulf Coast beach (Curray, 1956)

Class interval	Number of grains	Class interval	Number of grains
0°–	244	90°–	401
10°–	262	100°–	382
20°–	246	110°–	332
30°–	290	120°–	322
40°–	284	130°–	295
50°–	314	140°–	230
60°–	326	150°–	256
70°–	340	160°–	263
80°–	371	170°–	281
		Total	5439

The observations may be concentrated on $(0°, 90°)$ or $(0°, 180°)$. For example, axial data are concentrated on (0°, 180°). Table 1.2 shows orientation of the least projection elongations of sand grains in thin sections, cut parallel to the laminations, of Recent Gulf Coast beach sand. This data with angles in $(0°, 180°)$ can be converted to the range of $(0°, 360°)$ by doubling each angle. Further, the range 0° to $360°/l$ of $\theta°$ can be converted to the range $(0, 2\pi)$ by

$$\theta = l\pi\theta°/180. \qquad (1.3.2)$$

1.3.2. Time Period

Now, we consider conversion of the measurements of times reduced modulo some period into angles. The length of the time period can be identified with 360°. The usual time period is either a day or a year. We first consider data where the period is a day. Table 1.3 shows the number of occasions on which thunder was heard during each two hourly interval of the day at Kew (England) during summer from 1910–1935. In this case, 15° corresponds to 1 hour and 1° corresponds to 4 minutes, so that the conversion is straightforward. The conversion of the data in Table 1.3 is shown in the second column of the table.

TABLE 1.3 The number of occasions on which thunder was heard at Kew in the summers of 1910–1935 (adapted from Bishop, 1947)

G.M.T.	Angle	Frequency	G.M.T.	Angle	Frequency
00–	0°–	26	12–	180°–	133
02–	30°–	24	14–	210°–	149
04–	60°–	14	16–	240°–	122
06–	90°–	15	18–	270°–	80
08–	120°–	14	20–	300°–	61
10–	150°–	65	22–	330°–	22
				Total	725

Next, we consider the case when the period is a year. Table 1.4 shows the number of occurrences of rainfall of 1″ or more per hour in the U.S.A. during 1908–1937, classified according to the twelve months. For a non-leap year, the t^{th} day can be related to the angles by

$$\theta^\circ = 360 \times t/365. \tag{1.3.3}$$

A similar conversion can be done for a leap year. However, when the data is classified into months, the class-intervals and the mid-points obtained by using (1.3.3) are not easy to handle. For example, the class-interval for the

TABLE 1.4 The number of occurrences of rainfall of 1″ or more per hour in the U.S.A., 1908–37 (Dyck and Mattice, 1941)

Month	Angle	Frequency (Unadjusted)	Frequency (Final adjusted)
JAN.	0°–	101	100
FEB.	30°–	94	103
MARCH	60°–	232	229
APRIL	90°–	406	414
MAY	120°–	685	676
JUNE	150°–	1225	1248
JULY	180°–	1478	1458
AUG.	210°–	1384	1365
SEPT.	240°–	907	924
OCT.	270°–	383	378
NOV.	300°–	195	199
DEC.	330°–	145	143
	Total	7235	7235

month of February is 30° 35′–58° 12′. In addition, the lengths of the months vary. To overcome these difficulties, we can adjust the frequencies such that they correspond to 360 days with each month of the same length. Then 1° will correspond to 1 day. The adjustment is as follows. Let the observed frequencies for January, March, May, July, August, October and December be multiplied by

$$c = 30/31 = 0{\cdot}96774$$

and let the frequency for February be multiplied by

$$d = 30/28 = 1{\cdot}07143.$$

Let n_i be the original frequencies and let n_i' be the frequencies so adjusted. We then have reduced the year to 360 days but $N \neq N'$ where $N = \Sigma n_i$ and $N' = \Sigma n_i'$. To preserve the sum N, the final adjusted frequencies are obtained on multiplying n_i' by N/N'.

For the data in Table 1.4, we have

$$N = 7235, \qquad N' = 7099{\cdot}5, \qquad N/N' = 1{\cdot}01908.$$

These corrections are essential when the differences between the frequencies are of the order 10 per cent or less, since the differences between the lengths of the months can themselves produce such irregularities. In particular, the month of February has 10 per cent fewer days than January.

For the data in Table 1.4, the adjusted frequencies are shown in the fourth column and the corresponding class-intervals in terms of the angles are given in the second column. In the month of February, the adjusted frequency is 103 compared with the unadjusted frequency of 94. For January, the adjusted frequency is now 100 and so the month with minimum frequency becomes January instead of February.

The above cycles are the most frequent in practice but the same principle can be used to reduce any other periodic data into angles.

1.4 FORMS OF FREQUENCY DISTRIBUTIONS

The forms of the circular distributions appearing in practice can roughly be identified from the linear histograms just as in the case of linear data. As pointed out before, the linear histogram should be obtained from the circular histogram by cutting the circle at a suitable point such that the maximum concentration on the linear histogram appears around its centre. The data considered in this section will be visualized in this way.

1.4.1. Unimodal Distributions

Each of the frequency distributions given in Table 1.1 to Table 1.4 is unimodal. The frequencies tail off uniformly giving rise to a minimum in the corresponding circular histogram. This point of minimum is called the antimode in contrast to the mode which is, of course, the point of maximum. The frequency distribution in Table 1.1 and Table 1.2 are somewhat symmetrical. The distribution of the directions of birds in Table 1.1 has a high peak and the frequencies tail off uniformly to zero. The birds have a preferred direction between 280°–340°, i.e. there is a tendency to select a N–W course. The distributions in Table 1.3 and Table 1.4 are slightly asymmetrical and are positively skew since the rise to the maximum is more rapid than the fall. In Table 1.3 the main period of thunderstorm activity is 2 p.m. to 4 p.m.

TABLE 1.5 Azimuths of cross-beds in the upper Kamthi river (Sengupta and Rao, 1966)

Azimuth	Frequency	Azimuth	Frequency
0°–	75	180°–	0
20°–	75	200°–	21
40°–	15	220°–	8
60°–	25	240°–	24
80°–	7	260°–	16
100°–	3	280°–	36
120°–	3	300°–	75
140°–	0	320°–	90
160°–	0	340°–	107
		Total	580

Table 1.5 gives an example of an asymmetrical distribution which is negatively skew. The data gives azimuths of cross-beds in the upper Kamthi river, India. As on the line, symmetrical distributions on the circle are comparatively rare.

TABLE 1.6 Horizontal directions of 100 outwash Wisconsin pebbles (adapted from Krumbein, 1939)

Mid-points	0°	20°	40°	60°	80°	100°	120°	140°	160°	Total
Frequency	16	13	9	14	9	14	12	6	7	100

Table 1.6 gives horizontal directions of outwash pebbles from a late Wisconsin outwash terrace along Fox River, near Cary, Illinois. The directions are distributed almost uniformly over the range of (0°, 180°). (In fact, the observed value of χ^2 under this hypothesis is 8·72. The 5% value of χ^2

with 8 degrees of freedom is 15·5.) Table 1.7 shows the month of onset of lymphatic leukemia in the U.K. during 1946–1960. The distribution is unimodal but does not tail off to zero.

TABLE 1.7 Month of onset of lymphatic leukemia in the U.K., 1946–60 (Lee, 1963)

Month	Number of cases	Month	Number of cases
Jan.	40	July	51
Feb.	34	Aug.	55
March	30	Sept.	36
April	44	Oct.	48
May	39	Nov.	33
June	58	Dec.	38
		Total	506

In the strict sense, there are no J or U shaped distributions on a circle if the observations are distributed on the complete range of $(0, 2\pi)$. However, a J-shaped or U-shaped linear histogram will arise from a unimodal circular distribution if the cut point is selected near to the mode.

1.4.2. Multi-modal Distributions

So far we have discussed unimodal distributions. Multi-modal distributions also occur in practice. Table 1.8 and Table 1.9 give typical examples. Table 1.8 shows orientations of 76 turtles after treatment. It can be observed that

TABLE 1.8 Orientations of 76 turtles after treatment (Gould's data cited by Stephens, 1969f)

Direction	Number of turtles	Direction	Number of turtles
0°–	8	180°–	1
30°–	18	210°–	5
60°–	18	240°–	6
90°–	12	270°–	1
120°–	1	300°–	1
150°–	3	330°–	2
		Total	76

the distribution is bimodal and the two modes are roughly 180° apart. The dominant mode is in the interval 60°–90° and the subsidiary mode is in the interval 240°–270°. The data indicates that the turtles have a preferred direction (the homeward direction) but a substantial minority seem to prefer

TABLE 1.9 Geostrophic wind directions at Crawley for the 25 months, November–March, 1957–61 (Findlater *et al.* 1966)

Direction	Frequency	Direction	Frequency
0°–	27	180°–	27
20°–	23	200°–	43
40°–	44	220°–	69
60°–	42	240°–	69
80°–	25	260°–	53
100°–	20	280°–	38
120°–	20	300°–	37
140°–	11	320°–	39
160°–	19	340°–	40
		Total	646

the direction exactly opposite this homeward direction. Table 1.9 shows wind directions at Crawley, England. The data is bimodal and the two modes are again 180° apart, implying two opposite regimes of wind direction.

1.5 FURTHER EXAMPLES OF ANGULAR DATA

Some actual examples giving rise to angular data are described in Sections 1.2–1.4. We now give further examples involving angular data. Although these experiments belong to different scientific disciplines and therefore seem to be quite different, they all give rise to angular data and are abstractly equivalent from the statistical point of view.

1.5.1. Geology

Angular data appear in investigations of various geological processes since these involve transporting matter from one place to another in time. (The term fabrics, or vectorial fabrics, is sometimes used by geologists to describe geological orientation data.) Studies of the directions of remnant magnetism are used to interpret paleomagnetic current and possible magnetic pole wandering during geological times. Studies of the orientations of fractures and fabric elements in deformed rocks are used to interpret tectonic forces. Orientations of cross-bedding and other structures, and particle long axes in undeformed sediments, are used to interpret the directions of depositing currents of wind or water. For further details, see Curray (1956) and Pincus (1953, p. 584). Tables 1.2, 1.5 and 1.6 give some examples in the above category.

Data for the above situations are likely to consist of measurements of planar and linear features at given geographical points. Planar features may include foliation planes, bedding planes, planes of cross-stratification, cleavage planes, fold axial planes, joints and faults while linear features may include fold and other tectonic axes, cleavage-bedding intersections, mineral, fossile or pebble elongations, and other sedimentary and tectonic lineations (Loudon, 1964). Watson (1970) has given an excellent account of geological terms and concepts.

The above measurements can be expressed in terms of azimuth and angle of dip so that circular distributions appear only as marginal distributions. The joint distribution will be discussed in Chapters 8 and 9 which also contain various examples.

1.5.2. Meteorology

A natural source of angular data is, of course, wind directions and we have already considered such data in Table 1.9. A distribution of wind directions may arise either as a marginal distribution of the wind speed and direction as in Table 1.9, or as a conditional distribution for a given speed as in Table 1.10. Wind directions are usually represented clockwise on the map from north (0°) to east (90°), south (180°), west (270°) and back to north (360°). The bearings N32°E and S29°E are therefore 32° and 151° respectively. This representation is used in Table 1.9 and Table 1.10.

TABLE 1.10 Wind directions at Larkhill with speeds between 27 and 41 knots, March–May, 1940–45 (Tucker, 1960)

Direction	N	NE	E	SE	S	SW	W	NW	Total
Mid-points	0°	45°	90°	135°	180°	225°	270°	315°	—
Frequency	20	5	7	4	16	32	43	26	153

The thunderstorms data of Table 1.3 and the rainfall data of Table 1.4 are examples where the circular distribution also appear naturally. Similar data appears for the monthly run-off for a watershed, for the monthly evaporation from a reservoir and in other hydrologic cycles.

1.5.3. Biology

The study of bird orientation in homing or migration leads to angular data. The birds are released from a point and the bearings of their flights, called vanishing angles, are recorded just as they vanish in the distance. Table 1.1 gives an example. There are various important experiments on birds to answer questions such as whether the sun-azimuth-compass hypothesis is obeyed, whether celestial cues are utilized for orientation, whether the flight directions are random under certain conditions etc. For an excellent discussion of various developments and investigations in this field, the reader is referred to Schmidt-Koenig (1965). Chapters 6 and 7 contain various examples from this area.

There are similar investigations on orientation of surface animals. The turtle data in Table 1.8 is such an example.

TABLE 1.11 Angles between the swimming directions of *Daphnia* and the plane of polarization of light (Waterman and Jander's data, cited in Waterman, 1963)

Direction	Frequency	Direction	Frequency
0°–	65	90°–	208
10°–	17	100°–	81
20°–	12	110°–	73
30°–	16	120°–	43
40°–	22	130°–	50
50°–	51	140°–	35
60°–	58	150°–	24
70°–	67	160°–	29
80°–	105	170°–	44
		Total	1000

The study of the effects of polarized light on the orientation of marine animals also gives rise to angular data. Table 1.11 gives data consisting of angles between the swimming directions of *Daphnia* and the plane of polarization, the degree of polarization being 27°. The distribution is bimodal and the modes are roughly 90° apart. This typical characteristic of bimodal data has already been noted in Section 1.4. For another type of investigation under polarized light, giving rise to angular data, see Jaffe (1956).

1.5.4. Geography

Orientation data appear naturally when readings consist of longitudes and latitudes. For example, in the study of the occurrence of earthquakes in a region, the longitude and latitude of each shock (its epicentre) are recorded.

On the other hand, the data may be cyclic as in studying the variation of the number of earthquakes from year to year or from day to day after a large main shock. Angular observations also arise in daily determinations of microseismic directions at a particular location (Jensen, 1959).

1.5.5. Economic Time-series

If an economic time-series is a random perturbation of a periodic phenomenon with known period, the random times associated with it could be described by circular distributions. Table 1.12 gives a series of the production of buses in the U.S.A. which forms a genuine circular distribution consisting of discrete, countable units. Although only few series of prices, production or trade may sensibly be viewed in this way, further examples have been given by Gumbel (1954).

TABLE 1.12 The production of buses in the U.S.A. in 1948–50 (from *Automobile Facts and Figures*, 1951, AMA, New York)

Month	Frequency	Month	Frequency
Jan.	1120	July	1435
Feb.	985	Aug.	1589
March	1223	Sept.	1453
April	1200	Oct.	1523
May	1197	Nov.	1251
June	1527	Dec.	1192
		Total	15695

1.5.6. Physics

Von Mises (1918) proposed to test the hypothesis that atomic weights are integers subject to error. All weights can be reduced to angular deviations such as 8·25 to 90°, 8·75 to 270°, 9·00 to 0°, 9·25 to 90° and so on. The physical problem is not clear but statistically it reduces to testing whether the circular distribution has a mode at 0° (see Example 6.1).

Angular data also appears in determining a preferred direction for optical axes of crystals in rock specimens. The distribution of the resultant of a random sample of unit vectors arises in representing sound waves or molecular links (Rayleigh, 1919). It also arises in various problems related to the interference among oscillations with random phases (Beckmann, 1959). Studies of rotary Brownian motion also involve problems in this area.

Angular data also arises in experiments with a bubble chamber, where points representing events are observed through a circular window. If each point is moved radially to the circumference of the circle, we have angular observations.

1.5.7. Psychology

The perception of direction under varying experimental conditions leads to angular data (Ross *et al.*, 1969). For example, to simulate zero-gravity in space travel, experiments on divers and swimmers are performed under water to assess their ability in perceiving the true horizontal and vertical in the absence of a visual cue. The data consists of their deviations from the true direction. Studies to compare relative performances in perceiving the true vertical under water and on land also give rise to angular data.

1.5.8. Medicine

The number of deaths due to a disease or the number of onsets of a disease in each month over years forms a circular distribution. The data in Table 1.7 is such an example.

Monthly death rates can be regarded as constituting a circular distribution under the assumption that the underlying population is stationary so that the rates are proportional to the numbers of deaths. Under a similar assumption, mean percentages of deaths due to a particular disease constitutes a circular distribution. Gumbel (1954) has given various examples.

Cardiology is a field in which angular variates are prominent especially in the vector cardiogram which is the three-dimensional analogue of the usual one-dimensional cardiogram (see Downs and Liebman, 1969; Gould, 1969). Gould (1969) has given an excellent discussion of the problems involved and Downs and Liebman (1969) contains various references dealing with vector cardiographic data.

1.5.9. Astronomy

The theory of errors as developed by Gauss was primarily for the analysis of astronomical measurements which consisted of points on the celestial sphere. The surface of the celestial sphere can be approximated locally by a tangent plane with the probability concentrated in the neighbourhood of the point of contact. Hence, the actual manifold, the surface of the sphere, was ignored in the theory and this lead to the development of the theory of Statistics on Euclidean spaces.

Bernoulli (1734) enquired whether the close coincidence of the orbital planes of the six planets known at that time could have arisen by chance.

Each orbital plane can be regarded as a point on the sphere (see Example 9.4) and his hypothesis is equivalent to the statement that the points are distributed uniformly on the sphere. Watson (1970) has discussed this problem for the nine planets. Pólya (1919) enquired whether the stars are distributed at random over the celestial sphere.

1.5.10. Sampling

A roulette wheel usually has 37 equally spaced positions on the circumference of a circle. In an unbiased wheel, each position is equally likely. In general, we can consider a wheel with scale graduated from 0 to 2π. The stopping position of the wheel gives rise to a random point on the circle.

Consider a circular disc on which beads are dropped, while the tray is agitated in its plane. The chance of a bead coming off depends on the angle made by the disc with the horizontal plane.

The generation of random numbers with a given base leads to a circular distribution in view of the periodicity of the remainders. We will see in Section 4.3.4 that the behaviour of the distribution of the first digit selected at random from a large compendium such as a census register can be explained by regarding it as a distribution on a circle.