STATISTICS FOR INDUSTRY AND TECHNOLOGY

N. Balakrishnan
N. Kannan
H.N. Nagaraja
Editors

Advances in Ranking and Selection, Multiple Comparisons, and Reliability

Methodology and Applications

Birkhäuser

5

A Simple Classification Rule for Directional Data

Ashis SenGupta[1] **and Supratik Roy**[2]

[1]*Applied Statistics Division, Indian Statistical Institute, Kolkata, India*
[2]*Department of Statistics, University College of Cork, Cork, Republic of Ireland*

Abstract: An intuitive and geometrically motivated chord-length based discriminant statistic is proposed for the classification of a new observation into one of two circular populations when training samples are available from each of them. Assuming that each of the two underlying populations is von Mises, the exact distribution of this statistic is indicated and its relationship to Fisher's discrimination and Cox's Logistic discrimination rules are discussed. The performance of this rule is presented and compared with Fisher's rule in terms of exact error probabilities and apparent error rates. This new rule is illustrated by a real-life data set.

Keywords and phrases: Apparent error rate, classification rule, directional data, logistic discrimination

5.1 Introduction

Consider the problem of classifying a new observation into one of two distinct circular populations. For an introduction to analysis of circular or directional data, see e.g., Mardia (1972), Jammalamadaka and SenGupta (2001). Suppose we have observations as directional data from these two (identifiable) populations given as θ_{ij}, $j = 1, \ldots, n_i$, $i = 1, 2$. We will utilize these observations as training samples to provide estimates of parameters of the above two populations as needed. Let a new observation be denoted by θ. Denote the sample mean directions by $\overline{\theta}_i$, $i = 1, 2$.

Morris and Laycock (1974) have discussed the usual Fisher's discrimination rule for the von Mises or circular normal (CN) populations when the parameters may possibly be unknown. El Khattabi and Streit (1996) have illustrated the use of classical Bayes rule with offset normal distribution on the circle and some other distributions on the sphere. Note that such parametric rules become quite cumbersome even for applications, when invoked for other popular circular distributions, e.g., the family of symmetric wrapped stable distributions [SenGupta and Pal (2001)]. In a somewhat more extended context, Collett and Lewis (1981) have discussed the problem of discriminating between the von Mises and wrapped normal distributions given a set of data assumed to be coming from one (unknown) of these two populations. These works however apply the standard linear techniques and do not address the peculiarity and the distinctive features of directional data.

In Section 5.2 we introduce a very simple and elegant chord-based discrimination rule which is intuitively appealing and geometrically motivated specifically for circular data and which may be used for arbitrary circular distributions with possibly unknown functional forms. The basic idea used here is to find out the average "distance" (in an appropriate sense) from the new observation to the observations in the two known groups. If the distance from one group is less than from the other, then the new observation is classified as belonging to the "closer" population. Though this approach may be used for any circular distribution, here we illustrate it by the von Mises populations. Next we recall that in a linear setup, for the univariate or multivariate normal distributions the Fisher type discrimination rule, which coincides with the Logistic Discrimination (LGD) rule of Cox (1966) with same variances and the Quadratic LGD rule [see e.g., Anderson (1975)] if variances are different, can be viewed as a quadratic distance function i.e., with variance-covariance matrix playing the role of the metric tensor. In Section 5.3 we show that a similar phenomenon holds for the class of directional distributions also. In Section 5.4 we discuss the exact distribution of the discriminant statistic and note how one can compute the threshold value numerically. Section 5.5 presents a study on the efficiency of the chord-based rule and compares it with Fisher's rule in terms of their Apparent Error Rates (AERs). Finally in Section 5.6, the new rule is illustrated by a real-life data set.

5.2 Construction of the Rule

5.2.1 A distance measure

The simplest distance that can be used for circular data is the arc-length, which in the case of the unit circle is equivalent to the radian measure subtended at the center of the circle, i.e., the value of the observation in radians.

But to be a proper distance on the circle, the distance measure must be rotationally invariant, both in terms of magnitude as well as in the sense of rotation. Thus if we have to consider the arc-length in terms of radian measure, we have to transform it in a suitable way, i.e., take absolute value of the difference in angles, modulo 2π. We may also have to consider the minimum of the two arc-lengths into which two points on the circle divides a circle.

These problems do not arise if instead of the arc-length we consider the length of the chord cut off by the two points on the circle. This is always non-negative, invariant under rotation, both in magnitude and displacement. As we shall see, this particular form has also other attractive properties due to its similarity to known descriptive measures, e.g., circular variance [Mardia (1972, p. 21)].

We observe that though the use of chord length as a descriptive measure is quite natural and may have been in use for long, the approach in the following section seems to be the maiden attempt in this direction.

5.2.2 Average distance of a point from a group

Let two points on the unit circle be denoted by θ_1, θ_2. Then the square of the chord-length between the two is given by $2\,(1 - \cos(\theta_1 - \theta_2))$. Based on this we take the distance measure as

$$d_{ij} \;=\; 1 \;-\; \cos(\theta_i - \theta_j). \tag{5.1}$$

Note that d_{ij} has the following properties : It is always non-negative, symmetric in its indices and is invariant under rotation. A measure of deviation between two points on a circle, e.g., two circular observations, the true mean direction and its estimator [SenGupta and Maitra (1998)], etc. may thus be based on it.

The average distance $d_i(\theta)$ of θ from the group i, is given by

$$d_i(\theta) \;=\; 1 \;-\; \frac{1}{n_i}\sum_j \cos(\theta_{ij} - \theta). \tag{5.2}$$

Note that this is similar to the sample circular variance with a shift in the mean direction. Let

$$\overline{C}_i \;=\; \frac{1}{n_i}\sum_j \cos(\theta_{ij}),\; \overline{S}_i \;=\; \frac{1}{n_i}\sum_j \sin(\theta_{ij}),\; \overline{R}_i = \sqrt{\overline{C}_i^2 + \overline{S}_i^2},\; \tan(\overline{\theta}_i) = \frac{\overline{S}_i}{\overline{C}_i}.$$

5.2.3 The chord-based rule

Let the new observation to classify be θ. Let d_{0i} be the distance of θ from $\bar{\theta}_i$, the circular mean for group i, $i = 1, 2$. Define $D(\theta) \;=\; d_{01}(\theta) - d_{02}(\theta)$. Suppose that prior probabilities for the two populations are equal and let c be a real constant. The classification rule is then given as follows:

$$\begin{aligned} &\text{If}\quad D(\theta) < c \quad \text{assign}\;\; \theta \;\;\text{to population 1,} \\ &\qquad\text{and assign}\;\; \theta \;\;\text{to population 2 otherwise.} \end{aligned} \tag{5.3}$$

Now

$$D(\theta) = (\cos(\bar{\theta}_2) - \cos(\bar{\theta}_1))\cos\theta + (\sin(\bar{\theta}_2) - \sin(\bar{\theta}_1))\sin\theta. \tag{5.4}$$

Let

$$\tan(\theta_0) = \frac{\sin(\bar{\theta}_2) - \sin(\bar{\theta}_1)}{\cos(\bar{\theta}_2) - \cos(\bar{\theta}_1)}. \tag{5.5}$$

Note that $\mathrm{P}(\bar{\theta}_1 = \bar{\theta}_2) = 0$, assuming that we are dealing with underlying continuous distributions and hence θ_0 is well defined (with probability 1).

Then (5.4) can be written as

$$D(\theta) = \sqrt{2 - 2\cos(\bar{\theta}_1 - \bar{\theta}_2)}\ \cos(\theta - \theta_0). \tag{5.6}$$

Note that by (5.5), there will be two solutions for θ_0.

The classification rule in (5.3) now reduces to an equivalent but a very simple form as

$$\text{If } \cos(\theta - \theta_0) > K \text{ assign } \theta \text{ to population 1, and assign } \theta \text{ to population 2 otherwise,} \tag{5.7}$$

where K is an appropriate constant.

Remarks.

1. The direction θ_0 is orthogonal to the bisector of $\bar{\theta}_1$ and $\bar{\theta}_2$.

2. As is often done for the sake of simplicity of the classification rule [see, e.g., Rao (1973, p. 575, Eq. (8e.1.8))], we can take $K = 0$. The rule as given by equation (5.7) then simply partitions the circle into sectors of width 180°. In this case, explicitly, the sectors can be specified as one semicircle having θ_0 as its midpoint, and the complementary arc. Note that if the sample mean directions are equal, unequal circular variances have no effect on the rule. In this case θ_0 is simply the mean direction itself. However, when the sample mean directions are not equal, the circular variances do affect θ_0. It is obvious that the rule can be modified trivially to cover the case of unequal prior probabilities also. Finally, in case specified misclassification probabilities are to be maintained, K can be suitably determined by using the distribution of $\cos(\theta - \theta_0)$ as discussed in Section 5.4.

5.2.4 An extension of the chord-based rule

Let $V_1 = d_1(\bar{\theta}_1), V_2 = d_2(\bar{\theta}_2)$, i.e., V_i is the average intragroup "distance" from each other for the observations in group or sample i. Note that V_i is nothing but the sample "circular variance" for sample i, see, e.g., Mardia (1972, p. 21). Define the intra-group average d_{ii} from the sample mean direction as

$$d_{ii} = 1 - \frac{1}{n_i}\sum \cos(\theta_{ij} - \bar{\theta}_i). \tag{5.8}$$

Then $d_{ii} = 1 - \bar{R}_i = V_i$. Take constants $\alpha_i > 0, i = 1, 2$, and β, and define

$$D_1(\theta) = \alpha_1 \left(d_1(\theta) - \frac{d_{11} + d_{22}}{2}\right) - \alpha_2 \left(d_2(\theta) - \frac{d_{11} + d_{22}}{2}\right) + \beta.$$

The classification rule is given by

$$\text{If} \quad D_1(\theta) < 0 \quad \text{assign } \theta \text{ to population 1}$$
$$\text{and assign } \theta \text{ to population 2 otherwise.} \tag{5.9}$$

Now $D_1(\theta)$ reduces to

$$\{\alpha_2 \cos(\bar{\theta}_2) - \alpha_1 \cos(\bar{\theta}_1)\} \cos\theta + \{\alpha_2 \sin(\bar{\theta}_2) - \alpha_1 \sin(\bar{\theta}_1)\} \sin\theta$$
$$+ \frac{1}{2}(\alpha_1 - \alpha_2)(\bar{R}_1 + \bar{R}_2). \tag{5.10}$$

Let

$$\tan(\theta_0) = \frac{\alpha_2 \sin(\bar{\theta}_2), - \alpha_1 \sin(\bar{\theta}_1)}{\alpha_2 \cos(\bar{\theta}_2) - \alpha_1 \cos(\bar{\theta}_1)}. \tag{5.11}$$

Then by (5.11), there will be two solutions for θ_0. However, as is done [see, e.g., Jammalamadaka and SenGupta (2001)] for defining $\bar{\theta}$, θ_0 also may be defined uniquely by taking the quadrant specific arc-tan function by interpreting the numerator and denominator of the ratio in the right-hand side of (5.11) accordingly.

5.3 Relationship of Chord-based Rule with Other Rules

5.3.1 Fisher's rule

Assume that the underlying populations are in the CN family, i.e., $CN(\mu_i, \kappa_i)$, $i = 1, 2$. Recall that the density function corresponding to $CN(\mu, \kappa)$ is given by

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \mu)], \quad 0 \le \mu < 2\pi, \kappa > 0.$$

Note that, given the parameters, the standard Fisher type ('maximum likelihood') function would have the form

$$- \ln I_0(\kappa_1) + \ln I_0(\kappa_2) + \{\kappa_1 \cos(\mu_1) - \kappa_2 \cos(\mu_2)\} \cos(\theta)$$
$$+ \{\kappa_1 \sin(\mu_1) - \kappa_2 \sin(\mu_2)\} \sin(\theta) + \beta \tag{5.12}$$

$$= -\ln I_0(\kappa_1) + \ln I_0(\kappa_2) + \sqrt{\kappa_1^2 + \kappa_2^2 - 2\kappa_1\kappa_2\cos(\mu_1 - \mu_2)}$$
$$\times \cos\left(\theta - \tan^{-1}\frac{\kappa_2 \sin\mu_2 - \kappa_1 \sin\mu_1}{\kappa_2\cos\mu_2 - \kappa_1\cos\mu_1}\right) + \beta. \tag{5.13}$$

Putting $\alpha_i = \kappa_i$ in (5.10) and observing that

$$\frac{d}{d\kappa}\ln I_0(\kappa) = A(\kappa),$$

we have

$$\ln I_0(\kappa) = \int A(\kappa)d\kappa.$$

Recall from the ML estimation of κ for the CN population, that $\bar{R}$ is asymptotically $A(\kappa)$. Note also that $\kappa\bar{R}^2/2$ approximates the integral $\int A(\kappa)\mathrm{d}\kappa$ by a triangle. Note also that

$$\frac{d}{d\kappa}A(\kappa) = 1 - A^2(\kappa) - \frac{A(\kappa)}{\kappa}$$

and hence that for small change in κ, the order of change in $A(\kappa)$ is less than that of κ. Therefore,

$$\ln I_0(\kappa_1) - \ln I_0(\kappa_2) = \int_{\kappa_1}^{\kappa_2} A(\kappa)d\kappa \sim \frac{1}{2}[A(\kappa_1) + A(\kappa_2)](\kappa_2 - \kappa_1),$$

by the trapezoidal rule. Note that, if κ_1 and κ_2 are very close to each other, asymptotically (5.10) approximates (strongly converges to) the corresponding portion of (5.12). The equivalence between Fisher's rule and our rule then becomes clear. Thus although we have kept the rule flexible by introducing the constants α_is, a recommended choice in case von Mises distributions seem to be the underlying populations, is that which is found by substituting the pairs

$$\hat{\kappa}_i = A^{-1}(\bar{R}_i), \; i = 1, 2.$$

5.3.2 Cox's logistic discrimination rule

In the above discussion, the modified rule can easily be identified as a semiparametric rule which approaches the Fisher type rule (ratio of densities) when the underlying populations are circular normals and they are close to each other in terms of population parameters.

Since LGD models the ratio of densities in the case when the log ratio is linear in the underlying random variable, observe that LGD cannot be directly applied to discriminate between two von Mises populations. However, note that a simple generalization of LGD can still be used in such cases, since the log-ratio in this case is linear on the sine and cosine transformations of θ. This also bypasses the rather computationally tricky problem of having to estimate κ,

$A(\kappa)$ and their logarithms, as the constant term in the expression of the LGD subsumes all the Bessel function terms. This can also be approached through the method of Generalized Pseudo Maximum Likelihood estimation [see Roy (1999)]. Also, the rule based on chord lengths as given above, however then need not assume independence of the linear components as done for the LGD rule.

5.4 Exact Distribution of $D(\theta)$

The distribution of $\bar{\theta}$ conditional on $R = r$ is von Mises with mean direction μ and concentration parameter κr. The joint distribution of $C = R\cos\bar{\theta}$, $S = R\sin\bar{\theta}$ is given by

$$f(C,S) = \frac{1}{I_0^n(\kappa)} e^{\kappa(\cos(\mu)C+\sin(\mu)S)}\phi_n(C^2+S^2). \tag{5.14}$$

Here ϕ_n is the density of R^2 when $\theta_1, \dots, \theta_n$ is a random sample from a circular uniform distribution. The joint distribution of $U = \alpha_1\cos(\bar{\theta}_1) - \alpha_2\cos(\bar{\theta}_2)$, $V = \alpha_1\sin(\bar{\theta}_1) - \alpha_2\sin(\bar{\theta}_2)$, given $R_1 = r_1, R_2 = r_2, \alpha_1, \alpha_2$, is given by

$$\begin{aligned} f(U,V) &= \frac{e^{\kappa_1 r_1\cos(\mu_1)\frac{U}{\alpha_1}+\kappa_2 r_2\cos(\mu_2)\frac{V}{\alpha_2}}}{(2\pi)^2 I_0(\kappa_1 r_1) I_0(\kappa_2 r_2)} \\ &\quad \times \int_{\bar{\theta}_2} \exp\left\{(\kappa_1 r_1\cos(\mu_1)\frac{\alpha_2}{\alpha_1} + \kappa_2 r_2\cos(\mu_2))\cos(\bar{\theta}_2)\right. \\ &\qquad\qquad \left. +(\kappa_1 r_1\sin(\mu_1)\frac{\alpha_2}{\alpha_1} + \kappa_2 r_2\sin(\mu_2))\sin(\bar{\theta}_2)\right\} d\bar{\theta}_2. \end{aligned}$$

Combining this with (5.14), we have the joint distribution of (C,S,U,V) (where $\cos(\theta) = C$ and $\sin(\theta) = S$), given $R_1 = r_1, R_2 = r_2, \alpha_1, \alpha_2$, to be

$$\begin{aligned} f(C,S,U,V) &= \frac{e^{\kappa_1 r_1\cos(\mu_1)\frac{U}{\alpha_1}+\kappa_2 r_2\cos(\mu_2)\frac{V}{\alpha_2}+\kappa\cos(\theta-\mu)}}{(2\pi)^3 I_0(\kappa) I_0(\kappa_1 r_1) I_0(\kappa_2 r_2)} \\ &\quad \times \phi_1(1)\int_{\bar{\theta}_2} e^{(\kappa_1 r_1\cos(\mu_1)\frac{\alpha_2}{\alpha_1}+\kappa_2 r_2\cos(\mu_2))\cos(\bar{\theta}_2)} \\ &\quad \times e^{(\kappa_1 r_1\sin(\mu_1)\frac{\alpha_2}{\alpha_1}+\kappa_2 r_2\sin(\mu_2))\sin(\bar{\theta}_2)} d\bar{\theta}_2. \end{aligned} \tag{5.15}$$

To get the distribution of the statistic, the conditional density in (5.15) multiplied by the joint density $h_{n_1}(R_1)h_{n_2}(R_2)$ [for the definition of $h_n(R)$, see e.g., Mardia (1972, p. 94)] has to be integrated over regions of the form $\boldsymbol{d} = aCU + (1-a)SV$. This fact may be used in invoking numerical integration to obtain the constant K of Section 5.2 when specified misclassification probabilities are to be met.

Table 5.1: Comparison of Fisher's and chord-based rules

$n_1 = n_2 = 10, \mu_1 = 0, \kappa_1 = 0.10$

$\kappa_2 = 0.10$

μ_2	$\hat{\mu_1}$	$\hat{\mu_2}$	R_1	R_2	ERR_1	AER_1	ERR_2	AER_2
0.1	0.05	0.12	0.09	0.09	0.15	0.19	0.13	0.25
0.2	0.05	0.17	0.091	0.089	0.14	0.16	0.14	0.25
0.3	0.05	0.26	0.091	0.09	0.137	0.15	0.146	0.24
0.4	0.055	0.36	0.092	0.091	0.135	0.148	0.145	0.24
0.5	0.05	0.48	0.093	0.092	0.13	0.14	0.148	0.21

$\kappa_2 = 0.20$

μ_2	$\hat{\mu_1}$	$\hat{\mu_2}$	R_1	R_2	ERR_1	AER_1	ERR_2	AER_2
0.1	0.05	0.12	0.09	0.18	0.15	0.17	0.25	0.3
0.2	0.05	0.17	0.091	0.19	0.14	0.16	0.22	0.26
0.3	0.05	0.26	0.091	0.19	0.132	0.145	0.2	0.25
0.4	0.055	0.36	0.092	0.192	0.129	0.14	0.19	0.24
0.5	0.05	0.48	0.093	0.192	0.126	0.134	0.18	0.22

$\kappa_2 = 0.30$

μ_2	$\hat{\mu_1}$	$\hat{\mu_2}$	R_1	R_2	ERR_1	AER_1	ERR_2	AER_2
0.1	0.05	0.12	0.09	0.28	0.16	0.17	0.25	0.3
0.2	0.05	0.17	0.091	0.289	0.14	0.16	0.22	0.26
0.3	0.05	0.26	0.091	0.29	0.132	0.145	0.2	0.25
0.4	0.055	0.36	0.092	0.291	0.129	0.14	0.19	0.24
0.5	0.05	0.48	0.093	0.292	0.126	0.134	0.18	0.22

5.5 Efficiency of the Rule

As is apparent, closed form expressions for error probabilities do not exist and the actual values have to be numerically computed for each pair of training samples.

Table 5.1 presents and compares the performances of the Fisher's and the chord-based rules, where we have taken $\mu_1 = 0$ (angles are given in radians) without loss of generality. ERR_1 denotes the calculated error probability from the exact distribution of the modified statistic as given above, ERR_2 the calculated error probability from the Fisher type (ratio of densities) discrimination rule, AER_1 the apparent error rate from the modified statistic as given above, and AER_2 the apparent error rate from the Fisher type discrimination rule. It is clear that our proposed rule outperforms Fisher's rule in terms of both ERR and AER over all the parameter combinations considered.

5.6 A Real-life Example

We now consider the data on pigeon-homing, as referred to in Mardia (1972, pp. 156–157), in which the internal clocks of 10 birds were reset by 6 hours clockwise while the clocks of 9 birds were left unaltered. Assuming that the underlying distributions are von Mises with equal concentration parameters [as in Mardia (1972, p. 157)], we classify each observation in the two samples on the basis of the remaining observations, by comparing the average chord-length distance from each group.

The result shows that the apparent error rate (AER) is 0.0 for the control group, 0.25 for the experimental group and 0.117 for the combined sample.

The AERs or the sample misclassification probabilities show that the rule correctly classifies all the observations in the control group, and 75% in the experimental group.

References

1. Anderson, J. A. (1975). Quadratic logistic discrimination, *Biometrika*, **62**, 149–154.

2. Collett, D. and Lewis, T. (1981). Discriminating between the von Mises and wrapped normal distributions, *Australian Journal of Statistics,* **23**, 71–79.

3. Cox, D. R. (1966). Some procedures associated with the logistic qualitative response curve, In *Research Papers in Statistics* (Ed., F. N. David), pp. 55-71, John Wiley & Sons, New York.

4. El Khattabi, S. and Streit, F. (1996). Identification analysis in directional statistics, *Computational Statistics & Data Analysis,* **23**, 45–63.

5. Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*, World Scientific Publishers, New Jersey.

6. Mardia, K. V. (1972). *Statistics of Directional Data,* Academic Press, London.

7. Morris, J. E. and Laycock, P. J. (1974). Discriminant analysis of directional data, *Biometrika*, **61**, 335–341.

8. Rao, C. R. (1973). *Linear Statistical Inference and Its Applications,* Second edition, John Wiley & Sons, New York.

9. Roy, S. (1999). *On Logistic and Some New Discrimination Rules: Characterizations, Inference and Applications*, Ph.D. Thesis, Indian Statistical Institute, Kolkata, India.

10. SenGupta, A. and Maitra, R. (1998). On best equivariance and admissibility of simultaneous MLE for mean direction vectors of several Langevin distributions, *Annals of the Institute of Statistical Mathematics*, **50**, 715–727.

11. SenGupta, A. and Pal, C. (2001). On optimal tests for isotropy against the symmetric wrapped stable-circular uniform mixture family, *Journal of Applied Statistics*, **28**, 129–143.