

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/243082597>

Nonparametric Measures of Angular–Angular Association

Article in *Biometrika* · August 1982

DOI: 10.2307/2335405

CITATIONS

24

READS

41

2 authors, including:



Nicholas Fisher

The University of Sydney

129 PUBLICATIONS 6,913 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Statistical graphics [View project](#)



Data Science [View project](#)

Nonparametric measures of angular-angular association

By N. I. FISHER

C.S.I.R.O. Division of Mathematics and Statistics, Sydney, Australia

AND A. J. LEE

Department of Mathematics, University of Auckland, Auckland, New Zealand

SUMMARY

A general model is proposed for association between two angular variables, corresponding to monotone association between two linear variables. A U -statistic analogous to Kendall's tau is developed to estimate the degree of this association and its distributional properties studied. A simple modification of Mardia's rank angular-angular correlation coefficient is proposed as the appropriate analogue of Spearman's rho for assessing this association.

Some key words: Directional data; Rank correlation; Toroidal concordance; U -statistic.

1. INTRODUCTION

Let Θ and Φ be two angular random variables, with joint distribution concentrated on the surface of a torus, and let $(\theta_1, \phi_1), \dots, (\theta_n, \phi_n)$ be n independent realizations of (Θ, Φ) . The problem of testing for general association between Θ and Φ has received little attention in the literature and, to date, no general model for such association akin to a monotone relationship between two real random variables has been proposed.

By adapting the test proposed by Hoeffding (1948b) of the independence of two real continuous random variables, Rothman (1971) obtained a test consistent against all alternative continuous bivariate angular models. He obtained the asymptotic distribution of the test statistic, but not its small-sample distribution. Hillman (1974) suggested a method requiring computation of the maximum and minimum values of the n^2 Spearman's rank correlations obtained by choosing all possible positions for the reference directions of the θ 's and ϕ 's. Hillman tabulated the statistic for $n = 5, \dots, 11$, but the large-sample distribution is unknown. Mardia (1975) suggested that the appropriate analogue, for dependence of Θ and Φ , of a linear relationship between two real random variables, is

$$l\Theta \pm m\Phi + \psi_0 \equiv 0 \pmod{2\pi},$$

where l and m are positive integers and ψ_0 is an unknown constant angle. For the particular case in which l and m can be assumed to be one, Mardia developed an analogue of Spearman's rank correlation coefficient as follows. Denote by r_1, \dots, r_n and s_1, \dots, s_n the separate sample linear ranks of $\theta_1, \dots, \theta_n$ and ϕ_1, \dots, ϕ_n respectively, and set

$$n^2 \bar{R}_1^2 = [\Sigma \cos \{2\pi(r_i - s_i)/n\}]^2 + [\Sigma \sin \{2\pi(r_i - s_i)/n\}]^2,$$

$$n^2 \bar{R}_2^2 = [\Sigma \cos \{2\pi(r_i + s_i)/n\}]^2 + [\Sigma \sin \{2\pi(r_i + s_i)/n\}]^2.$$

Then $r_0 = \max(\bar{R}_1^2, \bar{R}_2^2)$ is a measure of the degree of dependence of the form

$$\theta + \phi + \psi_0 \equiv 0 \pmod{2\pi} \quad (1.1)$$

or

$$\theta - \phi + \psi_0 \equiv 0 \pmod{2\pi} \quad (1.2)$$

and is invariant under choice of reference direction for Θ or Φ , or equivalently, the value of ψ_0 . Clearly, $0 \leq r_0 \leq 1$, with $r_0 = 1$ corresponding either to $\bar{R}_1^2 = 1$ and complete 'positive' dependence as in (1.1), or to $\bar{R}_2^2 = 1$ and complete 'negative' dependence as in (1.2). Mardia provides selected percentiles for the small-sample distributions, and shows that $2(n-1)\bar{R}_1^2$ and $2(n-1)\bar{R}_2^2$ are asymptotically distributed as independent χ_2^2 variates, so that the asymptotic distribution of $2(n-1)r_0$ is known.

In this paper, simple generalizations of (1.1) and (1.2) are proposed as natural models of general association between two angular random variables. The extent of this association can be estimated using a U -statistic analogous to Kendall's tau. In §3, the small-sample and large-sample properties of this statistic are obtained, and in §4 some examples given of its application. In §5.1, some comments are made on these models *vis-à-vis* more general models. In §5.2, a slight modification of Mardia's r_0 statistic is suggested as the Spearman's rho-type statistic corresponding to the Kendall's tau-type statistic considered herein, and its distribution tabulated.

The discussion parallels that given in an earlier paper (Fisher & Lee, 1981) on general association of an angular and a linear random variable; comments on applications of the methods to problems of partial association can be found there.

2. A GENERAL MODEL FOR DEPENDENCE OF TWO ANGULAR RANDOM VARIABLES

Corresponding to the approach in Fisher & Lee (1981) we postulate that a reasonable model for general association between Θ and Φ is any relationship $\phi = g(\theta)$ such that, as θ moves continuously through a complete revolution in a particular sense, either clockwise or anticlockwise, ϕ also moves continuously in a particular sense through a complete revolution. If the senses are the same, the relationship $\phi = g(\theta)$ is said to be toroidally-concordant, or T -concordant, otherwise if one sense is clockwise and the other counter-clockwise, $\phi = g(\theta)$ is T -discordant. Thus $\phi = g(\theta)$ is essentially a warped version of (1.1) or (1.2) according as the relationship is T -concordant or T -discordant.

Now let $p_i = (\theta_i, \phi_i)$ ($i = 1, \dots, n$) be n points on the torus. Define the points to be T -concordant if there exists a T -concordant relationship $\phi = g(\theta)$ such that $\phi_i = g(\theta_i)$ ($i = 1, \dots, n$). The points are defined to be discordant in a corresponding way. Finally, let $P = (\Theta, \Phi)$ be randomly distributed on the torus. We propose to measure association between Θ and Φ by

$$\Delta = \text{pr}(P_1, P_2, P_3 \text{ are } T\text{-concordant}) - \text{pr}(P_1, P_2, P_3 \text{ are } T\text{-discordant}),$$

where P_1, P_2 and P_3 are independent random points distributed as P .

Note that the notion of T -concordance is independent of the choice of reference direction for Θ or Φ , and Δ inherits this property. Any three points on a torus are either T -concordant or T -discordant.

To obtain a U -statistic estimate of Δ , define the kernel

$$\delta(p_1, p_2, p_3) = \begin{cases} 1 & \text{if } p_1, p_2, p_3 \text{ are } T\text{-concordant,} \\ -1 & \text{if } p_1, p_2, p_3 \text{ are } T\text{-discordant.} \end{cases}$$

As with Kendall's tau, δ depends only on the separate-sample ranks of $\theta_1, \theta_2, \theta_3$ and ϕ_1, ϕ_2, ϕ_3 , and is most efficiently calculated from the representation

$$\delta(p_1, p_2, p_3) = \text{sgn}(\theta_1 - \theta_2) \text{sgn}(\theta_2 - \theta_3) \text{sgn}(\theta_3 - \theta_1) \times \text{sgn}(\phi_1 - \phi_2) \text{sgn}(\phi_2 - \phi_3) \text{sgn}(\phi_3 - \phi_1). \tag{2.1}$$

Then the U -statistic for estimating Δ is

$$\hat{\Delta}_n = \binom{n}{3}^{-1} \sum_{1 \leq i < j < k \leq n} \delta(p_i, p_j, p_k),$$

and $\hat{\Delta}_n$ has the following properties:

- (i) $-1 \leq \Delta_n \leq 1$;
- (ii) $\hat{\Delta}_n = +1$ if p_1, \dots, p_n are T -concordant and -1 if they are T -discordant;
- (iii) if Θ and Φ are independent, and P_1, \dots, P_n are distributed independently as P , then $\hat{\Delta}_n(P_1, \dots, P_n)$ is distributed symmetrically about 0.

If there are tied observations amongst $\theta_1, \dots, \theta_n$ or ϕ_1, \dots, ϕ_n , any kernel value $\delta(p_1, p_2, p_3)$ containing a tied pair can be set to zero, and the combinatorial divisor in the definition of $\hat{\Delta}_n$ should be reduced by one for each such zero value.

3. DISTRIBUTION THEORY FOR $\hat{\Delta}_n$

The method of obtaining the sampling properties of $\hat{\Delta}_n$ closely parallels that given in §3 of Fisher & Lee (1981). To derive these properties under the null hypothesis that Θ and Φ are independent, let $P_i = (\Theta_i, \Phi_i)$ ($i = 1, 2, 3$) be independently distributed as (Θ, Φ) , and suppose that Θ and Φ are independent variates on the circle.

Let $P = (\theta, \phi)$ be a point on the torus, with some fixed origin for each coordinate. Let $P^* = (\theta^*, \phi^*)$ be the point $(\frac{1}{2}\theta/\pi, \frac{1}{2}\phi/\pi)$ corresponding to P on the unit square. The transformation $P \rightarrow P^*$ depends of course on the choice of origins on the torus.

Define $\delta^*(P_1^*, P_2^*, P_3^*) = \delta(P_1, P_2, P_3)$; then δ^* can be calculated by (2.1) and depends only on the ranks of θ_i^* and ϕ_i^* , and does not depend on the choice of origin.

Thus the distribution of $\delta(P_1, P_2, P_3) = \delta^*(P_1^*, P_2^*, P_3^*)$ is invariant under monotone transformations of P_i^* , and so we may assume that the P_i^* are uniformly distributed on the unit square. Dropping the $*$ notation and working on the unit square, we define

$$\delta_2(p_1, p_2) = E\{\delta(p_1, p_2, P_3)\}, \quad \sigma_2^2 = \text{var}\{\delta_2(P_1, P_2)\},$$

$$\delta_1(p_1) = E\{\delta_2(p_1, P_2)\}, \quad \sigma_1^2 = \text{var}\{\delta_1(P_1)\},$$

so that $E\{\delta_1(P_1)\} = E\{\delta_2(P_1, P_2)\} = 0$. A routine calculation yields

$$\delta_2(p_1, p_2) = f(\theta_1 - \theta_2) f(\phi_1 - \phi_2),$$

where $f(x) = \text{sgn}(x)(1 - 2|x|)$. Note that

$$\int_0^1 \int_0^1 f(x-y) dx dy = 0$$

and that

$$\delta_1(p_1) = E\{\delta_2(p_1, P_2)\} = \int_0^1 f(\theta_1 - \theta_2) d\theta_2 \int_0^1 f(\phi_1 - \phi_2) d\phi_2 = 0.$$

further,

$$\sigma_1^2 = \text{var} \{ \delta_1(P_1) \} = 0, \quad \sigma_2^2 = \text{var} \{ \delta_2(P_1, P_2) \} = \frac{1}{9}, \quad \sigma_3^2 = \text{var} \{ \delta(P_1, P_2, P_3) \} = 1.$$

Thus (Hoeffding, 1948a) $\hat{\Delta}_n$ is degenerate of order one, and

$$\text{var}(\hat{\Delta}_n) = \binom{n}{3}^{-1} \sum_{c=1}^3 \binom{3}{c} \binom{n-3}{3-c} \sigma_c^2 = \frac{2(n-3)+6}{n(n-1)(n-2)}.$$

Using standard methods (Gregory, 1977; Sproule, 1974) the asymptotic distribution of $n\hat{\Delta}_n$ is found to be identical to that of $3 \sum_v \lambda_v (Z_v^2 - 1)$, where the sum is over $v = 1, \dots, \infty$, where Z_v are independent normal $N(0,1)$ random variables, and the λ_v are the eigenvalues of the integral equation with kernel $f(\theta_1 - \theta_2) f(\phi_1 - \phi_2)$. These eigenvalues are of the form $\lambda_v = (\pm \mu_l) (\pm \mu_m)$ ($l, m = 1, 2, \dots$), where $\pm \mu$ are the eigenvalues of the integral equation with kernel $f(\theta_1 - \theta_2)$, so that $\mu_l = i/(\pi l)$.

Thus the asymptotic distribution of $n\Delta_n$ is that of

$$\frac{3}{\pi^2} \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{lm} W_{lm}, \tag{3.1}$$

where the W_{lm} are independently distributed as the difference of two independent χ_2^2 variates.

The asymptotic distribution of $n\hat{\Delta}_n$ was computed by approximating the characteristic function of (3.1) by that of

$$\frac{3}{\pi^2} \sum_{lm \leq N} \frac{1}{lm} W_{lm} + cZ, \tag{3.2}$$

where c is chosen to make the variance of (3.1) and (3.2) coincide and Z is a $N(0, 1)$ random variable. The distribution function was then computed by numerical inversion.

It is feasible to enumerate the complete distribution for $n = 3, \dots, 7$ by generating all $n!$ permutations (π_1, \dots, π_n) of $(1, \dots, n)$ and computing $n\hat{\Delta}_n$ for each of the $n!$ sets $\{(i, \pi_i); i = 1, \dots, n\}$. These distributions are tabulated in Table 1. For $n > 7$, 5000 random permutations were used to estimate the distribution for each sample size considered. Selected percentiles of these distributions are given in Table 2; beyond $n = 30$, the asymptotic distribution furnishes an adequate approximation. Tests of the hypothesis $\Delta = 0$ against the alternative $|\Delta| = 1$ are then carried out by referring $|n\hat{\Delta}_n|$ to Table 1 and rejecting the hypothesis if $|n\hat{\Delta}_n|$ is too large; similarly one-sided tests can be performed.

TABLE 1. Probability function of $n\hat{\Delta}_n$ ($n = 3, 4, 5, 6, 7$); distributions are symmetric about 0

$n = 3$	x	3	-3										
	$2 \text{ pr}(3\hat{\Delta}_3 = x)$	1	1										
$n = 4$	x	4	0	-4									
	$6 \text{ pr}(4\hat{\Delta}_4 = x)$	1	4	1									
$n = 5$	x	5	2	1	0	...							
	$24 \text{ pr}(5\hat{\Delta}_5 = x)$	1	5	5	2								
$n = 6$	x	6.0	3.6	2.4	1.2	0.0	...						
	$120 \text{ pr}(6\hat{\Delta}_6 = x)$	1	6	12	23	36							
$n = 7$	x	7.0	5.0	3.8	3.4	2.6	2.2	1.8	1.4	1.0	0.6	0.2	...
	$720 \text{ pr}(7\hat{\Delta}_7 = x)$	1	7	14	21	14	21	63	44	28	70	77	

TABLE 2. Selected percentiles $x_{n,\alpha}$ of the distribution of $n\hat{\Delta}_n$; $\text{pr}(n\hat{\Delta}_n \leq x_{n,\alpha}) = 1 - \alpha$, $\text{pr}(|n\hat{\Delta}_n| \leq x_{n,\alpha}) = 1 - 2\alpha$; α , tail probability; n , sample size

n	$\alpha = 0.001$	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.1$
8	6.23	4.74	4.25	3.40	2.78	2.10
9	6.08	4.64	4.16	3.33	2.72	2.07
10	5.96	4.56	4.09	3.28	2.68	2.04
11	5.85	4.50	4.02	3.24	2.65	2.01
12	5.77	4.44	3.97	3.20	2.62	1.99
13	5.70	4.40	3.93	3.17	2.59	1.97
14	5.64	4.36	3.89	3.14	2.57	1.96
15	5.59	4.33	3.86	3.12	2.56	1.95
20	5.40	4.21	3.75	3.04	2.49	1.90
25	5.29	4.14	3.68	2.99	2.46	1.87
30	5.22	4.09	3.64	2.96	2.43	1.86
∞	4.85	3.85	3.42	2.81	2.31	1.77

If Θ and Φ are not independent, $\sigma_1^2 \neq 0$ in general, so the asymptotic distribution of $n^{\frac{1}{2}}(\Delta_n - \Delta)$ will be normal with mean zero and variance $9\sigma_1^2$. It follows that an asymptotic confidence interval for Δ is $\hat{\Delta}_n \pm 3\hat{\sigma}_{1,n}z(\frac{1}{2}\alpha)/\sqrt{n}$, where $z(\frac{1}{2}\alpha)$ is the upper $100(1 - \frac{1}{2}\alpha)$ -percentage point of the $N(0, 1)$ distribution and $\hat{\sigma}_{1,n}^2$ is a consistent estimate of σ_1^2 (Puri & Sen, 1971, p. 58) given by

$$\hat{\sigma}_{1,n}^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{\Delta}_n^{(i)} - \hat{\Delta}_n)^2,$$

where

$$\hat{\Delta}_n^{(i)} = \binom{n-1}{2}^{-1} \sum_{1 \leq i < j \leq n} \delta(P_i, P_i, P_j) \quad (i, j \neq l).$$

4. EXAMPLES

We now apply the techniques described in the previous sections to estimate angular-angular association in 2 sets of data.

Example 1 (Downs, 1974). The peak times for two successive measurements of blood pressure, converted into angles, of 10 medical students were recorded. The estimated association between the two sets of readings is $\hat{\Delta}_{10} = 0.6333$, which is significantly different from zero at the 1% level. An approximate 95% confidence interval for Δ is (0.530, 0.737).

Example 2 (Johnson & Wehrly, 1977). Wind directions were measured at 6 a.m. and 12 noon on each of 21 consecutive days, at a weather station in Milwaukee. The estimated association between the 6 a.m. and 12 noon measurement is $\hat{\Delta}_{21} = 0.2140$ which is significant at the 5% level. An approximate 95% confidence interval for Δ is (0.170, 0.258).

5. DISCUSSION

5.1. More general models

We have only generalized the simplest form of the model $l\Theta \pm m\Phi + \psi_0 \equiv 0 \pmod{2\pi}$ to obtain the notion of toroidal concordance. The statistic $\hat{\Delta}_n$ will presumably have some

worth even if l or m is greater than 1, and unknown. In this latter case, however, one should not expect to obtain a particularly good test of independence of Θ and Φ , since the variables measured will be $\Theta_l = l\Theta \pmod{2\pi}$ and $\Phi_m = m\Phi \pmod{2\pi}$ from which the original Θ and Φ cannot be recovered. There is some comment on this problem in the discussion of Mardia (1975).

5.2. *An analogue of Spearman's rho for assessing T-concordance*

In view of the existence of the notion of positive and negative association, it seems sensible to adapt Mardia's r_0 statistic to the interval $[-1, 1]$ by defining $\hat{\Pi}_n = \bar{R}_1^2 - \bar{R}_2^2$, and so obtaining an appropriate analogue of Spearman's rho. Tables 3 and 4 contain the distribution of $(n-1)\hat{\Pi}_n$ for $n = 3, \dots, 7$, and selected percentiles for larger n , respectively. Because $2(n-1)\bar{R}_1^2$ and $2(n-1)\bar{R}_2^2$ are, asymptotically, distributed independently as χ_2^2 -variates, the asymptotic distribution of $(n-1)\hat{\Pi}_n$ is double exponential with density $\frac{1}{2}e^{-|x|}$ ($-\infty < x < \infty$); some percentiles are also given in Table 4.

TABLE 3. *Probability function of $(n-1)\hat{\Pi}_n$ ($n = 3, 4, 5, 6, 7$); distributions are symmetric about 0*

$n = 3$	x	2	-2									
	$2 \text{ pr}(2\hat{\Pi}_3 = x)$	1	1									
$n = 4$	x	3	0	-3								
	$6 \text{ pr}(3\hat{\Pi}_4 = x)$	1	4	1								
$n = 5$	x	4	1.79	0	...							
	$24 \text{ pr}(4\hat{\Pi}_5 = x)$	1	5	12								
$n = 6$	$12x$	60	40	25	20	15	5	0	...			
	$120 \text{ pr}(5\hat{\Pi}_6 = x)$	1	6	6	3	14	12	36				
$n = 7$	x	6	4.71	3.68	3.47	2.78	2.71	1.93	1.83	1.81	1.54	1.33
	$720 \text{ pr}(6\hat{\Pi}_7 = x)$	1	7	7	7	14	7	28	7	7	28	14
$n = 7$	x	1.07	0.86	0.69	0.59	0.52	0.48	0.38	0.31	0.71	0	
	$720 \text{ pr}(6\hat{\Pi}_7 = x)$	28	42	28	7	7	28	28	7	14	88	

Table 4. *Selected percentiles $y_{n,\alpha}$ of the distribution of $(n-1)\hat{\Pi}_n$; $\text{pr}(n-1)\hat{\Pi}_n \leq y_{n,\alpha} = 1-\alpha$, $\text{pr} |(n-1)\hat{\Pi}_n| \leq y_{n,\alpha} = 1-2\alpha$; α , tail probability; n , sample size*

n	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.10$
8	4.80	4.11	3.26	2.55	1.80
9	4.78	4.09	3.23	2.52	1.78
10	4.77	4.07	3.21	2.50	1.76
11	4.75	4.06	3.19	2.48	1.74
12	4.74	4.04	3.17	2.46	1.73
13	4.73	4.03	3.15	2.45	1.72
14	4.72	4.02	3.14	2.44	1.71
15	4.71	4.02	3.13	2.43	1.71
20	4.67	3.99	3.10	2.40	1.68
25	4.64	3.97	3.07	2.38	1.67
30	4.63	3.96	3.06	2.36	1.66
∞	4.60	3.91	2.99	2.30	1.61

For random samples of bivariate data in the plane, Spearman's rho is just the projection of Kendall's tau into the space of linear rank statistics (Hájek & Sidák, 1967, pp. 60–1). It is interesting to speculate that some corresponding result may be true for $\hat{\Pi}_n$, a quadratic rank statistic, and $\hat{\Delta}_n$.

REFERENCES

- DOWNS, T. D. (1974) Rotational angular-correlations. In *Biorhythms and Human Reproduction*. Eds M. Ferin *et al.*, Chapter 7. New York: Wiley.
- FISHER, N. I. & LEE, A. J. (1981). Nonparametric measures of angular-linear association. *Biometrika* **68**, 629–36.
- GREGORY, G. G. (1977). Large sample theory for U -statistics and tests of fit. *Ann. Statist.* **5**, 110–23.
- HÁJEK, J. & SÍDÁK, Z. (1967). *Theory of Rank Tests*. New York: Academic Press.
- HILLMAN, D. C. (1974). Correlation coefficients for ranked angular variates. In *Chronobiology*. Ed. L. E. Scheving, F. Halberg & J. E. Pauly, pp. 723–30. Tokyo: Igaka Shoin.
- HOEFFDING, W. (1948a). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293–325.
- HOEFFDING, W. (1948b). A nonparametric test of independence. *Ann. Math. Statist.* **19**, 546–57.
- JOHNSON, R. A. & WEHRLY, T. (1977). Measures and models for angular correlation and angular-linear correlation. *J. R. Statist. Soc. B* **39**, 222–9.
- MARDIA, K. V. (1975). Statistics of directional data (with discussion). *J. R. Statist. Soc. B* **37**, 349–93.
- PURI, M. L. & SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. New York, Wiley.
- ROTHMAN, E. D. (1971). Tests of coordinate independence for a bivariate sample on a torus. *Ann. Math. Statist.* **42**, 1962–9.
- SPROULE, R. X. (1974). Asymptotic properties of U -statistics. *Trans. Am. Math. Soc.* **199**, 55–64.

[Received June 1981. Revised January 1982]

