Some Statistical Methods for Bivariate Circular Data

Louis-Paul Rivest

Journal of the Royal Statistical Society. Series B (Methodological), Vol. 44, No. 1 (1982), 81-90.

Stable URL:
http://links.jstor.org/sici?sici=0035-9246%281982%2944%3A1%3C81%3ASSMFBC%3E2.0.CO%3B2-1

Journal of the Royal Statistical Society. Series B (Methodological) is currently published by Royal Statistical Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/rss.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Fri Nov 25 15:36:38 2005

J. R. Statist. Soc. B (1982),
44, *No.* 1, *pp.* 81–90

Some Statistical Methods for Bivariate Circular Data

By Louis-Paul Rivest

University of Toronto, Canada†

[Received April 1980. Revised April 1981]

Summary

A function of the smallest singular value of the cross product matrix between two circular unit vectors is suggested as a measure of correlation. It possesses most of the desirable properties for a correlation coefficient given by Jupp and Mardia (1980). The concept of cluster dependence is introduced. A good predictor of one unit vector given the other is shown to depend on the type of dependence observed. The wrapped bivariate normal is characterized.

Keywords: CORRELATION; REGRESSION; WRAPPED NORMAL; PREDICTION; CIRCULAR DATA

1. Introduction

To measure the dependence between two samples of angles is a problem that has attracted the attention of many researchers. (See Jupp and Mardia (1980) for a list of the different proposals.)

Mimicking the inference for multivariate samples, some authors have based their measure of correlation on $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ where Σ_{ij} is the covariance between the ith and the jth random vector. Johnson and Wehrly (1977) suggested use of the largest singular value of that matrix while Jupp and Mardia (1980) defined ρ^2 as its trace. They also showed that ρ^2 is similar to Mardia and Puri's (1978) suggestion:

$$\rho_{ss}^2+\rho_{sc}^2+\rho_{cs}^2+\rho_{cc}^2,$$

where ρ_{ss}^2 denotes the usual correlation coefficient between $\sin(x-\mu_0(x))$ and $\sin(y-\mu_0(y))$..., x and y are the two angles under consideration and $\mu_0(x)$ is the mean direction (Mardia, 1972, p. 45) of x.

In this paper, a function of the smallest singular value of Σ_{12}^*, the cross product matrix, is proposed. For uniform samples, it is analogous to Johnson and Wehrly's proposal, for clustered samples to ρ_{ss} and for highly clustered samples to the classical correlation coefficient.

The concept of cluster dependence is introduced. In Section 3 a new class of bivariate angular distributions is presented for which the construction of a good predictor of x given y is shown to depend on the type of dependence between x and y. The asymptotic distribution of various test statistics is derived in Section 4.

2. Rotational Correlation in the Population

Let u and v be random unit vectors defined on the circle; in a parametric form, $u=(\cos x, \sin x)'$ and $v=(\cos y, \sin y)'$. Let $\Sigma_{12}^*=E(uv')$ and define λ_1 as the largest singular value of Σ_{12}^* and λ_2 as its smallest singular value multiplied by the sign of $\det(\Sigma_{12}^*)$. In this section λ_2 is shown to be a parameter analogous to the covariance for linear data that contains most of the information about the dependence between u and v.

Proposition 2.1. If u and v are independent, $\lambda_1=\sigma(u)\sigma(v)$ and $\lambda_2=0$. Here $\sigma(u)$ and $\sigma(v)$ or $\sigma(x)$ and $\sigma(y)$ denote the resultant length of u and v respectively (see Mardia, 1972, p. 45).

† Present address: Dept. de Mathématique, Université Laval, Ste Foy, Québec, Canada G1K 7P4

© 1982 Royal Statistical Society 0035–9246/82/44081$2.00

Proof. If u and v are independent, $\Sigma_{12}^* = E(u)(E(v))'$ and λ_1^2 and λ_2^2 are the eigenvalues of

$$E(u)(E(v))'E(v)(E(u))' = \sigma^2(v)\,E(u)(E(u))',$$

which are $\sigma^2(v)\sigma^2(u)$ and 0. Q.E.D.

In the second proposition, a convenient parametric expression is obtained for λ_1 and λ_2.

Proposition 2.2.

$$\lambda_1 = (\sigma(x-y)+\sigma(x+y))/2,$$
$$\lambda_2 = (\sigma(x-y)-\sigma(x+y))/2.$$

Proof. It is well known that the largest singular value of a matrix A is equal to the maximum over f and g of $f'Ag$ where f and g are unit vectors, therefore

$$\begin{aligned}\lambda_1 &= \max_{\theta,\psi} E((\cos\theta\cos x+\sin\theta\sin x)(\cos\psi\cos y+\sin\psi\sin y))\\ &= \max_{\theta,\psi} E(\cos(x+y-\theta-\psi)+\cos(x-y-\theta+\psi))/2\\ &= (\sigma(x+y)+\sigma(x-y))/2.\end{aligned}$$

Now if $\lambda_1 \neq 0$, $\lambda_2 = \det(\Sigma_{12})/\lambda_1$. After some algebraic manipulation, it is shown that

$$\begin{aligned}\det(\Sigma_{12}) &= \{E^2(\cos(x-y))+E^2(\sin(x-y))-E^2(\cos(x+y))-E^2(\sin(x+y))\}/4\\ &= \{\sigma^2(x-y)-\sigma^2(x+y)\}/4\\ &= \lambda_1\{\sigma(x-y)-\sigma(x+y)\}/2.\end{aligned}$$

Therefore if $\lambda_1 \neq 0$, $\lambda_2 = \{\sigma(x-y)-\sigma(x+y)\}/2$ which remains true even if $\lambda_1 = 0$. Q.E.D.

If $\sigma(x-y)$ and $\sigma(x+y)$ are not 0, then λ_1 and λ_2 can be written as $E(\cos(x-\theta_0)\cos(y-\psi_0))$ and $E(\sin(x-\theta_0)\sin(y-\psi_0))$ where θ_0 and ψ_0 are defined by

$$\theta_0+\psi_0 = \mu_0(x+y) \quad \text{and} \quad \theta_0-\psi_0 = \mu_0(x-y).$$

Note that if θ_0 and ψ_0 satisfy the last equality so do $\theta_0+\pi$ and $\psi_0+\pi$. According to Proposition 2.1, λ_2 is 0 when x and y are independent. When they are dependent, a normalized version of λ_2 will provide a suitable measure of dependence. The normalization will depend on whether at least one of $\sigma(x-y)$ and $\sigma(x+y)$ is 0 or not.

Definition. If $\sigma(x-y)$ and $\sigma(x+y)$ are non-zero define, $\rho_c^*(x,y) = \rho_c^*$ as

$$\rho_c^* = \frac{\lambda_2}{\max\{E(\sin^2(x-\theta_0)),\ E(\sin^2(y-\psi_0))\}}.$$

If one of $\sigma(x-y)$ and $\sigma(x+y)$ is zero define $\rho_u^*(x,y) = \rho_u^*$ as

$$\rho_u^* = 2\lambda_2.$$

As exemplified in Section 3, $\sigma(x+y)$ and $\sigma(x-y)$ are non-zero when x and y are clustered around their mean direction while one of $\sigma(x+y)$ and $\sigma(x-y)$ is 0 when either x or y does not exhibit such a clustering.

It is easily verified that ρ_u^* and ρ_c^* are rotation invariant and both satisfy $\rho^*(-x,y) = -\rho^*(x,y)$ and $\rho^*(x,y) = \rho^*(y,x)$. It is clear that $|\rho_u^*| < 1$, furthermore $\rho_u^* = 1\ (-1)$ if and only if

$$P(x-y = \mu_0(x-y)(\bmod 2\pi)) = 1(P(x+y = \mu_0(x+y)(\bmod 2\pi)) = 1).$$

For ρ_c^*, one can prove the following.

Proposition 2.3. The coefficient of correlation ρ_c^* has the following properties:
(i) $-1 \leqslant \rho_c^* \leqslant 1$,
(ii) $\rho_c^* = 1$ if and only if $P(x-y = \mu_0(x-y)$ or $x+y = \pi+\mu_0(x+y)(\text{mod } 2\pi)) = 1$,
(iii) $\rho_c^* = -1$ if and only if $P(x+y = \mu_0(x+y)$ or $x-y = \pi+\mu_0(x-y)(\text{mod } 2\pi)) = 1$.

When $\rho_c^* = 1$, $P(x+y = \pi+\mu_0(x+y)(\text{mod}.2\pi)) < \frac{1}{2}$ otherwise $\mu_0(x+y)$ would be equal to $\pi+\mu_0(x+y)(\text{mod}.2\pi)$. That is $\rho_c^* = 1$ implies that $P(x-y = \mu_0(x-y)\ (\text{mod } 2\pi)) > \frac{1}{2}$. A similar comment holds for $\rho_c^* = -1$.

When $\sigma(x+y) = 0$, $\rho_u^* = \sigma(x-y)$ has been introduced by Johnson and Wehrly (1977) as a coefficient of correlation for uniform samples. They showed that ρ_u^* is the maximum of $E(u'Pv)$ over all the rotations P. In a similar way, it is proved that when $\sigma(x-y) = 0$, $-\rho_w^* = \sigma(x+y)$ is the maximum of $E(u'Pv)$ over all the orthogonal transformations P which are not rotations.

When both $\sigma(x+y)$ and $\sigma(x-y)$ are non-null, $\lambda_2 = E(\sin(x-\theta_0)\sin(y-\psi_0))$ so that ρ_c^* is analogous to ρ_{ss} which has been introduced by Mardia and Puri (1978). Since ρ_{ss} is a covariance normalized by a product, not a maximum, $|\rho_{ss}| = 1$ cannot be interpreted as total rotational dependence in the sense of Proposition 2.3. Also the sign of ρ_c^* is related to the regression model introduced by Jupp and Mardia (1980).

Proposition 2.4. If P_0 is an orthogonal matrix such that $E(u'P_0 v)$ is equal to the maximum over all the orthogonal transformations P of $E(u'Pv)$, then $\lambda_2 \neq 0$ if and only if P_0 is uniquely defined. Furthermore $\lambda_2 > 0$ if and only if P_0 is a rotation.

The proof of this result is easily deduced from Lemma 4.2 of Sibson (1978).

It can also be shown that $P_0 v$ is a unit vector with angle $y+\mu_0(x-y)$ when $\lambda_2 > 0$ and angle $-y+\mu_0(x+y)$ when $\lambda_2 < 0$.

Define *the best predictor of u given v* as the unit vector $\lambda(v)$ which maximizes $E(\lambda'(v)u)$. An application of the Cauchy–Schwarz inequality shows that the best predictor of u given v is equal to $E(u\,|\,v)/\|E(u\,|\,v)\|$ when $\|E(u\,|\,v)\| \neq 0$. Define $\mu_0(x\,|\,y)$ as the mean direction of that vector.

In general, $\rho_c^* = 0$ does not imply that u and v are independent. As an example of dependence which is not detected by ρ_c^*, consider r and s, two random unit vectors with mean direction equal to r_0 and s_0 respectively. Suppose that r and s are independent except for a correlation in the clustering around their respective mean direction, that is given r, s is a random unit vector with mean direction $-s_0$ or s_0 whose clustering depends on r. Then the cross product matrix Σ_{12}^{**} between r and s has rank 1, that is $\rho_c^* = 0$; also if r and s are independent $\Sigma_{12}^* = r_0 s_0'$.

Therefore $\rho_c^* = 0$ implies that Σ_{12}^* has rank 1; if in addition $\rho_{sc} = \rho_{cs} = 0$ we can expect the dependence between u and v to be similar to the one between r and s.

The parameter

$$\rho_1^* = \frac{\lambda_1 - E(\cos(x-\theta_0)\,E(\cos(y-\psi_0))}{\{V(\cos(x-\theta_0))\,V(\cos(y-\psi_0))\}^{\frac{1}{2}}},$$

which is analogous to ρ_{cc} is an appropriate measure of the cluster dependence between u and v. Section 3 discusses a class of bivariate distributions where $\rho_c^* = 0$ implies that x and y are at most cluster dependent.

For highly clustered random variables it is now shown that μ_0, σ and ρ_c^* are related to classical parameters for linear variables. For the remainder of this section assume that x and y belong to the interval $[0,2\pi)$.

Proposition 2.5. Let x and y be circular random variables distributed as $F_k(\theta, \phi)$ for which there exist constants θ_x and θ_y (in $(0, 2\pi)$) such that $(X_k, Y_k) = k^{\frac{1}{2}}(x-\theta_x, y-\theta_y)$ converges to a non-degenerate distribution. Then assuming that, X_k^4 and Y_k^4 are uniformly integrable

(Breiman, 1968, p. 91)

(i) $k^{3/2}(E(x)-\mu_0(x))$ is 0(1),

(ii) $k^2(1-\sigma(x)-V(x)/2)$ is 0(1),

(iii) $k(\rho_c^*-\operatorname{cov}(x,y)/\max\{V(x),V(y)\})$ is 0(1).

Proof. Write $\sin(\mu_0(x)-E(x))$ as

$$\begin{aligned}&\sin\mu_0(x)\cos E(x)-\cos\mu_0(x)\sin E(x)\\&\quad=E(\sin x\cos E(x)-\cos x\sin E(x))/\sigma(x).\end{aligned}$$

To prove the first part, it suffices to show that

$$K^{3/2}\,E(\sin x\cos E(x)-\cos x\sin E(x))=K^{3/2}\,E\sin(x-E(x))\text{ is }0(1).$$

Using a Taylor series expansion, the last expression can be written as

$$k^{3/2}(E(x-E(x))+E((x-E(x))^3\,\varepsilon(x)),$$

where $|\varepsilon(x)|<1$. The first term is 0 while the second one is bounded by

$$E(|X_k-E(X_k)|^3),$$

which using Minkowski's inequality and the uniform integrability of X_k^4 is 0(1); hence (i) is proved.

Consider

$$\begin{aligned}&K^2\,E(\cos(x-\mu_0(x))-\cos(x-E(x)))\\&=K^2\,E(\cos(x-\mu_0(x)))(1-\cos(\mu_0(x)-E(x)))\end{aligned}$$

Using (i) this is 0(1), therefore

$$K^2\,E(1-\cos(x-\mu_0(x))-(x-E(x))^2/2)$$

behaves as

$$K^2\,E(1-\cos(x-E(x))-(x-E(x))^2/2).$$

Expending $\cos(x-E(x))$ in a Taylor series one shows that the last expression is 0(1) by an argument similar to the previous one.

Now by (ii), $K^2(\lambda_2-\operatorname{cov}(x,y))$ is 0(1). Also

$$E(\sin^2(x-\theta_0))=(1-E(\cos(2x-\mu_0(x+y)-\mu_0(x-y))))/2$$

and as before

$$K^2\{E(\cos(2x-\mu_0(x+y)-\mu_0(x-y))-\cos(2x-\mu_0(2x)))\}\text{ is }0(1)$$

so that

$$K^2\{E(\sin^2(x-\theta_0)-V(x)\}\text{ is }0(1)$$

which also implies

$$K^2\{\max(E(\sin^2(x-\theta_0)),E(\sin^2(y-\psi_0)))-\max(V(x),V(y))\}\text{ is }0(1).$$

This together with the fact that $\lim K\max(V(x),V(y))\neq 0$ terminates the proof. Q.E.D.

3. A Class of Bivariate Angular Distributions

Define:

$$\begin{aligned}x&=\theta+\psi+\phi\quad(\operatorname{mod}2\pi),\\y&=\theta-\psi\qquad(\operatorname{mod}2\pi),\end{aligned}$$

where θ, ψ and ϕ are independent random variables with densities $g_\alpha(\theta-\mu_0(\theta))$, $g_\beta(\psi-\mu_0(\psi))$ and $g_\gamma(\phi-\mu_0(\phi))$ respectively $\{g_\lambda\}$ is a class of circular densities satisfying

(i) $g_\lambda(z) = g_\lambda(-z)$ for any λ, z,

(ii) $\int_0^{2\pi}\cos(z)\,g_\lambda(z)\,dz$ and $\int_0^{2\pi}\cos(2z)\,g_\lambda(z)\,dz$ are both bigger than 0 or are both null,

(iii) $\int_0^{2\pi}\cos(2z)\,(g_{\lambda_1}(z)-g_{\lambda_2}(z))\,dz = 0$,

if and only if $g_{\lambda_1}(z) = g_{\lambda_2}(z)$ for any z. If $P(\phi = 0) = 1$, if θ has the uniform distribution and if the density of ψ is $h(2(\psi-\mu_0(\psi)))$ where h is a symmetric circular density then it can be shown that the joint density of x and y is

$$h(x, y) = h(x-y-\mu_0(x-y))/2\pi \tag{3.1}$$

which has been suggested by Johnson and Wehrly (1977) to model the dependence between uniform samples.

Also if $P(\psi = 0) = 1$, this is the location mixture model (Gould, 1969). Throughout this section assume $\sigma(\phi) \neq 0$.

Proposition 3.1. Assuming that $\sigma(\theta)$ and $\sigma(\psi)$ are not equal to 0,

(i) $\rho_{sc} = \rho_{cs} = 0$,

(ii) $\rho_c^* = 1$ if and only if $P(x-y = \mu_0(x-y)\,(\mathrm{mod}\,2\pi)) = 1$,

(iii) $\rho_c^* = -1$ if and only if $P(x+y = \mu_0(x+y)\,(\mathrm{mod}\,2\pi)) = 1$,

(iv) $\rho_c^* = 0$ if and only if $E(\sin(x-\theta_0)\,|\,y) = 0$ for any y, or $E(\sin(y-\psi_0)\,|\,x) = 0$ for any x.

Proof. Since $x+y = 2\theta+\phi(\mathrm{mod}\,2\pi)$ and $x-y = 2\psi+\phi(\mathrm{mod}\,2\pi)$,

$$\lambda_2 = (\sigma(2\psi+\phi)-\sigma(2\theta+\phi)/2 = \sigma(\phi)\,(\sigma(2\psi)-\sigma(2\theta))/2$$

Here the result that if Z_1 and Z_2 are independent,

$$\sigma(Z_1+Z_2) = \sigma(Z_1)\,\sigma(Z_2) \quad \text{and} \quad \mu_0(Z_1+Z_2) = \mu_0(Z_1)+\mu_0(Z_2)\,(\mathrm{mod}\,2\pi) \tag{3.2}$$

has been used.

Using properties (i) and (ii) of $\{g_\lambda\}$,

$$\mu_0(x-y) = 2\mu_0(\psi)+\mu_0(\phi)\,(\mathrm{mod}\,2\pi) \quad \text{and} \quad \mu_0(x+y) = 2\mu_0(\theta)+\mu_0(\phi)$$

(which implies that one can take

$$\theta_0 = \mu_0(\theta)+\mu_0(\psi)+\mu_0(\phi)\,(\mathrm{mod}\,2\pi) \quad \text{and} \quad \psi_0 = \mu_0(\theta)-\mu_0(\psi)\,(\mathrm{mod}\,2\pi)).$$

From this, one deduces that

$$E(\sin^2(x-\theta_0)) = \tfrac{1}{2}-\sigma(2\theta)\,\sigma(2\psi)\,\sigma(2\phi)/2$$

and

$$E(\sin^2(y-\psi_0)) = \tfrac{1}{2}-\sigma(2\theta)\,\sigma(2\psi)/2.$$

Now

$$\rho_c^* = \sigma(\phi)\left(\frac{\sigma(2\psi)-\sigma(2\theta)}{1-\sigma(2\psi)\,\sigma(2\theta)}\right)$$

is equal to 1 if and only if $\sigma(\phi) = 1$ and $\sigma(2\psi) = 1$ which implies $P(x-y = \mu_0(x-y)\,(\mathrm{mod}\,2\pi) = 1$. The proof of (iii) is similar.

By (3.2) $\mu_0(x) = \theta_0(\mathrm{mod}\,2\pi)$ and $\mu_0(y) = \psi_0(\mathrm{mod}\,2\pi)$ which implies

$$E(\sin(y-\mu_0(y))\cos(x-\mu_0(x))) = E(\sin(x-\mu_0(x))\cos(y-\mu_0(y))) = 0$$

so that (i) is proved.

According to property (iii) of $\{g_\lambda\}$, $\rho_c^* = 0$ implies $g_\alpha = g_\beta = g$ say. To prove (iv) note that the joint density of $\psi-\mu_0(\psi)$ and $\theta-\mu_0(\theta)$ is $g(\psi)\,g(\theta)$. By a straightforward change of variables, the joint density of $y-\mu_0(y) = \theta-\psi-\mu_0(\theta)+\mu_0(\psi)$ and $\psi-\mu_0(\psi)$ is equal to $g(\psi)\,g(\psi+y)$.

Consider

$$\begin{aligned} E(\sin(x-\theta_0)\mid y) &= E(\sin(2\psi+y+\phi-\mu_0(\theta)-\mu_0(\psi)-\mu_0(\phi)\mid y) \\ &= E\{\sin(2(\psi-\mu_0(\psi))+(y-\mu_0(y)))\cos(\phi-\mu_0(\phi)) \\ &\quad +\sin(\phi-\mu_0(\phi))\cos(2(\psi-\mu_0 t))+(y-\mu_0(y)))\mid y\} \end{aligned} \tag{3.3}$$

To prove that (3.3) is 0 it suffices to prove that $E(\sin(2(\psi-\mu_0(\psi))+y-\mu_0(y))\mid y)=0$. The last expression is proportional to

$$\int_0^{2\pi}\sin(2\psi+z)\,g(\psi)\,g(\psi+z)\,d\psi=\int_0^{2\pi}\sin(2\psi)\,g(\psi-z/2)\,g(\psi+z/2)\,d\psi,$$

where $z=y-\mu_0(y)$.

By the symmetry of g, $g(-\psi-z/2)\,g(-\psi+z/2)=g(\psi-z/2)\,g(\psi+z/2)$ so that the last integral is 0. In a similar way it is shown that

$$E(\sin(y-\psi_0)\mid\theta+\psi)=0$$

Now $E(\sin(y-\psi_0)\mid x)$ is equal to

$$E(E(\sin(y-\psi_0)\mid x)\mid\phi)=E(E(\sin(y-\psi_0)\mid\theta+\psi)\mid\phi)=0. \quad \text{Q.E.D.}$$

Note that for the model under consideration, the coefficient of rotational dependence of Section 2 is ρ_c^* if and only if $\sigma(x)$ and $\sigma(y)$ are not zero; that is if and only if x and y exhibit some clustering.

By (iv) $\rho_c^*=0$ implies that the best predictor of x given y, $\mu_0(x\mid y)$ is $\mu_0(x)$ if $E(\cos(x-\theta_0)\mid y)\geqslant 0$ and $\mu_0(x)+\pi$ if not. To show that when $\rho_c^*=0$ it is possible to have $E(\cos(x-\theta_0)\mid y)<0$ for some y, consider $g_\lambda(z)=c\exp(\lambda\cos(\alpha z))$, $|z|<\pi$ for some fixed α in $(1,2)$. It is easily checked that $\{g_\lambda\}$ satisfies the requirements of this section. One can show that when $\rho_c^*=0$, $E(\cos(x-\theta_0)\mid y=\pi+\mu_0(y))$ is proportional to

$$4\int_0^{\pi/2}\cos(2\psi)\exp(2\lambda\cos(\alpha\psi)\cos(\alpha\pi/2))\,d\psi.$$

For α in $(1,2)$, $\cos(\alpha\pi/2)<0$ so that for ψ in $(0,\pi/4)$,

$$\exp(2\lambda\cos(\alpha\psi)\cos(\alpha\pi/2))<\exp(2\lambda\cos(\alpha(\pi/2-\psi))\cos(\alpha\pi/2))$$

which implies that $E(\cos(x-\theta_0)\mid y=\pi+\mu_0(y))<0$. When $\alpha=1$, i.e. when $\{g_\lambda(z)\}$ is the von Mises family, $\rho_c^*=0$ implies that the conditional distribution of x given that $y=\mu_0(y)+\pi$ is uniform.

When $\rho_c^*>0$, the best predictor of u given v is the unit vector with angle $y+2\mu_0(\psi\mid y)+\mu_0(\phi)$. When ρ_c^* is large 2ψ is highly clustered that is $2\mu_0(\psi\mid y)$ is close to $2\mu_0(\psi)$ so that $y+\mu_0(x-y)$ the predictor of Proposition 2.4 is a good approximation to the best predictor of x given y. Also when ρ_c^* is close to -1, $\mu_0(x+y)-y$ is close to the best predictor of x given y.

Proposition 3.2. Assuming that $E(\cos(p\phi))\neq 0$ for any integer p and that $\sigma(\theta)$ and $\sigma(\psi)$ are not zero, $\rho_c^*=0$ implies that x and y are independent if and only if $\{g_\alpha(\theta)\}$ is the family of wrapped normal densities (Mardia, 1972, p. 55).

Proof. Assume without losing generality that $\mu_0(\theta)=\mu_0(\psi)=0$. Let p and q be integers then

$$E(\exp(i(px+qy)))=E(\exp(i((p+q)\,\theta+(p-q)\,\psi+p\phi))$$

under the assumption $\rho_c^*=0$, this can be written as $f(p+q)\,f(p-q)\,f_0(p)$, where f and f_0 are the characteristic functions of g and g_γ respectively. Now x and y are independent if and only if

$$f(p+q)\,f(p-q)=(f(p)\,f(q))^2.$$

For $q = 1$ and $p = n$, this yields

$$f(n+1)\,f(n-1) = [f(1)]^2\,f^2(n).$$

By assumption $E(\cos(\psi)) \neq 0$, i.e. $f(1) \neq 0$, which together with $f(0) = 1$ implies $f(n) \neq 0$ for all n. If $a_n = \log f(n)$, $\{a_n\}$ satisfies the following difference equation:

$$a_{n+1} - 2a_n + a_{n-1} = 2\log f(1)$$

with the side conditions: $a_0 = 0$ and $a_n = a_{-n}$. This difference equation is similar to the one for the expected duration of the game in the gambler's ruin problem (Bailey, 1964, p. 27) with $p = \frac{1}{2}$, and the general solution is: $a_n = a + bn + \log f(1)\,n^2$. The side conditions yield $a = b = 0$ so that $f(n) = \exp(-\sigma^2 n^2/2)$, where $\log f(1) = -\sigma^2/2$, that is $f(n)$ is the characteristic function of the wrapped normal distribution. Q.E.D.

Now assume that the family $\{g_\alpha\}$ satisfies $\int \cos(\theta)\,g_\alpha(\theta)\,d\theta = 0$ if and only if $g_\alpha(\theta) = 1/2\pi$. If either $\sigma(\theta)$ or $\sigma(\psi)$ is 0, $\rho^*_\omega = 0$ if and only if θ and ψ are uniformly distributed, i.e. for uniform samples $\rho^*_u = 0$ implies independence.

Suppose that θ, ψ and ϕ are distributed as wrapped normal with parameters σ_1^2, σ_2^2 and σ_3^2 respectively. It is easily shown (Johnson and Wehrly, 1977) that the characteristic function of x and y is the characteristic function of a wrapped bivariate normal distribution with mean 0 and covariance matrix:

$$\begin{pmatrix} \sigma_1^2+\sigma_2^2+\sigma_3^2 & \sigma_1^2-\sigma_2^2 \\ \sigma_1^2-\sigma_2^2 & \sigma_1^2+\sigma_2^2 \end{pmatrix}.$$

For this model

$$\rho_c^* = \exp(\sigma_3^2/2)\frac{(\exp(-2\sigma_2^2)-\exp(-2\sigma_1^2))}{1-\exp(-2(\sigma_1^2+\sigma_2^2))}$$

and the sign of ρ_c^* is equal to the sign of the covariance of the bivariate normal distribution.

In the next proposition ρ^2, Jupp and Mardia's (1980) coefficient of correlation, is expressed as a function of ρ_c^*, ρ_u^* and ρ_1^*.

Proposition 3.3. Assuming that $\sigma(x-y)$ and $\sigma(x+y)$ are not 0,

$$\rho^2 = \frac{\max\{V(\sin(x-\theta_0)), V(\sin(y-\psi_0))\}}{\min\{V(\sin(x-\theta_0)), V(\sin(y-\psi_0))\}}(\rho_c^*)^2 + (\rho_1^*)^2$$

If θ or ψ is uniformly distributed: $\rho^2 = 2(\rho_u^*)^2$.

Proof. Since the parameters under consideration are rotation invariant, take $\mu_0(x) = \mu_0(y) = 0$. Then,

$$\Sigma_{12} = \Sigma_{21} = \begin{pmatrix} \lambda_1 - \sigma(x)\,\sigma(y) & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

and

$$\Sigma_{11} = \begin{pmatrix} V(\cos x) & 0 \\ 0 & V(\sin x) \end{pmatrix}$$

so that $\rho^2 = \operatorname{tr}(\Sigma_{11}^{-1}\,\Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21})$ is equal to

$$\operatorname{tr}\begin{pmatrix} \dfrac{(\lambda_1 - \sigma(x)\,\sigma(y))}{\{V(\cos x)\,V(\cos y)\}^{\frac{1}{2}}} & 0 \\ 0 & \dfrac{\lambda_2}{\{V(\sin x)\,V(\sin y)\}^{\frac{1}{2}}} \end{pmatrix}^2$$

which proves the first part of the proposition. Now if either ψ or θ is uniform $\sigma(x) = \sigma(y) = 0$ and $V(\cos x) = V(\sin x) = V(\cos y) = V(\sin y) = \frac{1}{2}$ so that $\rho^2 = (2\lambda_1)^2 + (2\lambda_2)^2 = 2(\rho_u^*)^2$ since for these models $|\lambda_1| = |\lambda_2|$. Q.E.D.

Therefore a large value of ρ^2 is due to either rotational or cluster dependence or both. Since the type of dependence observed is an important factor in the construction of a predictor of u given v a more detailed analysis than estimating ρ^2 is advisable here.

4. Rotational Correlation in the Sample

To carry out inference in a bivariate angular sample $\{x_i, y_i\}$, the following procedure is suggested:

(i) If at least one sample is uniform estimate ρ_u^*; if H_0: $\rho_u^* = 0$ is rejected, $P_0\, v$ the predictor of Proposition 2.4 is the best predictor of u given v under Johnson and Wehrly's models (3.1) and under the model $f(x, y) = h(x + y - \mu_0(x + y))/2\pi$.

(ii) If the two samples are clustered estimate ρ_c^*, if $H_0\, \rho_c^* = 0$ is rejected $P_0\, v$ should be a good approximation of the best predictor of u given v (especially if $|\hat{\rho}_c^*|$ is large). If $H_0\, \rho_c^* = 0$ is not rejected, test for cluster dependence.

Notation: In general, $\hat{\lambda}$ denotes the estimate of λ; r will be used instead of $\hat{\rho}$; $\hat{\mu}_0$ and $\hat{\sigma}$ are replaced by $\bar{x}_0$ and $\bar{R}$.

Proposition 4.1.

(i) If $\sigma(x-y)$ and $\sigma(x+y)$ are not 0,

$$r_c^* = \frac{\bar{R}(x-y) - \bar{R}(x+y)}{2\max\{\Sigma \sin^2(x_i - \hat{\theta}_0), \Sigma \sin^2(y_i - \hat{\psi}_0)\}}$$

is a consistent estimator of ρ_c^*, furthermore if the samples are independent,

$$n^{\frac{1}{2}}\left(\frac{\max\{\Sigma \sin^2(x_i - \hat{\theta}_0), \Sigma \sin^2(y_i - \hat{\psi}_0}{\min\{\Sigma \sin^2(x_i - \hat{\theta}_0), \Sigma \sin^2(y_i - \hat{\psi}_0)\}}\right)^{\frac{1}{2}} r_c^*$$

is asymptotically distributed as $N(0, 1)$.

(ii) If $\sigma(x+y)(\sigma(x-y))$ is 0,

$$r_u^* = \bar{R}(x-y)(-\bar{R}(x+y))$$

is a consistent estimator of ρ_u^*. Assuming that x and y are independent with either x or y uniformly distributed, then $2n(r_u^*)^2$ is asymptotically distributed as χ^2 with two degrees of freedom.

The proof of this statement is routine, the first part relies on the analysis of variance property for circular data (Mardia, 1972, p. 23). For the second part, if x and y are independent $x-y$ and $x+y$ are uniformly distributed provided either x or y is. Then by Mardia (1972, p. 112), $2n\bar{R}^2(x-y)(2n\bar{R}^2(x+y))$ has the prescribed asymptotic distribution.

Remarks (the following comments are made under the assumption that x and y are distributed as in Section 3).

(1) If $\sigma(2\sigma)$ and $\sigma(2\psi)$ are not equal to 0, under $H_0\, \rho_c^* = 0$, the test statistic of (i) is asymptotically distributed as $N(0, \sigma^2)$. In general σ^2 is not equal to one, as expected $\sigma^2 = 1$ for the wrapped normal family.

(2) When θ or ψ is uniformly distributed, to test H_0: $\rho_u^* = 0$ is to test for complete independence. According to Proposition 3.3, $(r_u^*)^2$ and $r^2/2$ are consistent estimates of ρ_u^{*2}. When H_0 is true, their asymptotic expectations are $1/n$ and $2/n$ respectively (Jupp and Mardia have shown that the asymptotic distribution of nr^2 is χ^2 with four degrees of freedom). Therefore when $\rho_u^* = 0$, r_u^* is more accurate than r.

When the samples are independent, the asymptotic distribution of $n^{\frac{1}{2}} r_1^*$ (the sample estimate of ρ_1^*) is a standard normal. This result can be used to test for cluster dependence.

Along the lines of Proposition 2.5 one proves the following.

Proposition 4.3. Let $\{x_{ik}, y_{ik}\}$ be a random sample from a bivariate circular distribution F_k(x_{ik} and y_{ik} are assumed to belong to $(0, 2\pi)$) assuming that there exists θ_x and θ_y in $(0, 2\pi)$ such that $k^{\frac{1}{2}}(x_{ik}-\theta_x,\ y_{ik}-\theta_y)$ converges in distribution then

(i) $k^{3/2}(\bar{x}_{0k}-\bar{x}_k)$ is $0_p(1)$,

(ii) $k^2(1-\bar{R}(x)-\Sigma(x_i-\bar{x}_k)^2/2n)$ is $0_p(1)$,

(iii) $k\left(r_c^*-\dfrac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\max\{\Sigma(x_i-\bar{x})^2, \Sigma(y_i-\bar{y})^2\}}\right)$ is $0_p(1)$.

Here $\bar{x}_{0k}$ and $\bar{x}_k$ denote respectively the mean direction and the mean of the sample.

This proposition shows that the statistical techniques used for linear data might be appropriate for circular data. In particular it shows that the normality assumption for a highly clustered sample of angles is as legitimate as the normality assumption for any linear sample; that is robust methods might be useful for circular data as well.

Also it demonstrates that for highly clustered samples the inference procedure described in Proposition 4.1 reduces to the classical inference on Pearson's correlation coefficient. This suggests that the t distribution should be a good approximation to the distribution of the test statistic in Proposition 4.1.

Example. Consider

$$f(x, y) = (I_0(k_1)\, I_0(k_2))^{-1} \exp\{k_1 \cos(x-y-\mu_1)+k_2 \cos(x+y-\mu_2)\}.$$

This density belongs to the family of Section 3 (let $g_\alpha(\theta) = \exp(\alpha\cos(2\theta))/2\pi I_0(\alpha)$, i.e. 2θ and 2ψ have the von Mises distribution). For $k_2 = 0$, f has been discussed by Johnson and Wehrly (1977). If

$$A = \begin{matrix} k_1\cos\mu_1+k_2\cos\mu_2 & -k_1\sin\mu_1+k_2\sin\mu_2 \\ k_1\sin\mu_1+k_2\sin\mu_2 & k_1\cos\mu_1-k_2\cos\mu_2 \end{matrix}$$

and $u = (\cos x, \sin x)'$, $v = (\cos y, \sin y)'$ then

$$f(u, v) = C(A)^{-1} \exp u'Av,$$

that is, f belongs to Mardia's class of bivariate angular distributions (see Jupp and Mardia, 1980).

Suppose that a sample with density f has been observed, the sufficient statistic is $(\bar{x}_0(x-y), \bar{x}_0(x+y), \bar{R}(x-y)$ and $\bar{R}(x+y))$. To test H_0: $\rho_c^* = 0$ is equivalent to testing the equality of the concentration parameters of two samples coming from von Mises populations (Mardia, 1972, p. 158). For this problem the asymptotic variance of the test statistic of Proposition 4.1 depends on $k = k_1 = k_2$.

To test H_0 $k_1 = k_2 = 0$, i.e. that the samples are independent uniformly distributed, the test "reject if $n(\bar{R}^2(x+y)+\bar{R}^2(x-y))$ is large" is appropriate. It can be shown that this test statistic is asymptotically equivalent to nr^2 which has been suggested by Jupp and Mardia (1980).

This model has attractive properties: it belongs to the exponential family and the inference on ρ_c^*, k_1 and k_2 is simple. Yet, since $f(x+\pi, y+\pi) = f(x, y)$ f is bimodal that is f should be used to model the dependence between bimodal samples only. Also when k_1 and k_2 are not 0, the marginal densities of x and y are not uniform so that to model the dependence between uniform samples one should set either k_1 or k_2 equal to 0 depending on whether positive or negative dependence is expected.

ACKNOWLEDGEMENTS

I thank Professor Stephens for introducing me to this topic and the referee for his constructive comments on an earlier version of that work.

REFERENCES

BAILEY, N. T. J. (1964). *The Elements of Stochastic Processes.* New York and London: Wiley.

BREIMAN, L. (1968). *Probability.* Reading, Mass.: Addison-Wesley.

GOULD, A. L. (1969). A regression technique for angular variates. *Biometrics*, **25**, 683–700.

JOHNSON, R. A. and WEHRLY, T. (1977). Measures and models for angular correlation and angular-linear correlation. *J. R. Statist. Soc.* B, **39**, 222–229.

JUPP, P. E. and MARDIA, K. V. (1980). A general correlation coefficient for directional data and related regression problems. *Biometrika*, **67**, 163–173.

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

MARDIA, K. V. and PURI, M. L. (1978). A spherical correlation coefficient robust against scale. *Biometrika*, **65**, 391–395.

SIBSON, R. (1978). Studies in the robustness of multidimensional scaling: procrustres statistics. *J. R. Statist. Soc.* B, **40**, 234–238.