Modelling Asymmetrically Distributed Circular Data

Arthur Pewsey
Departamento de Matemáticas
Universidad de Extremadura

1. Statistical Context

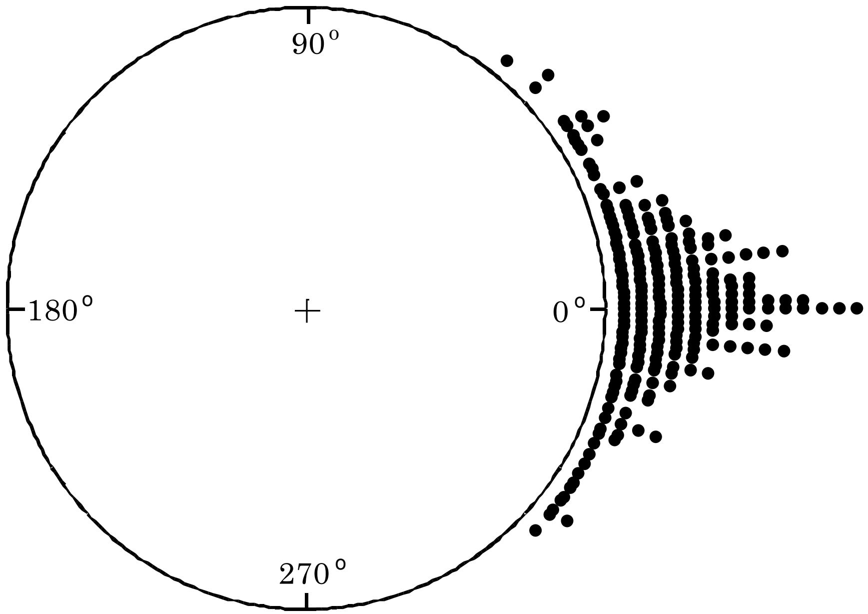
$$\left.\begin{array}{r}\text{Circular Data}\\ \text{Spherical Data}\end{array}\right\}\ \text{Directional Data}$$

$\downarrow$

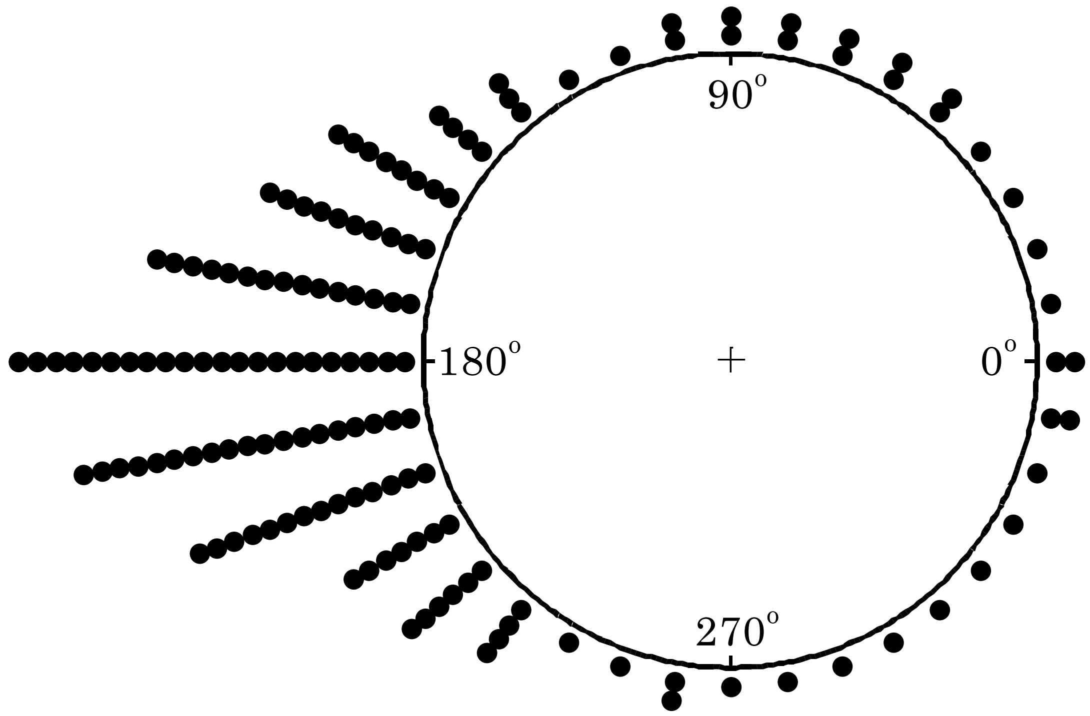
Modelling data on different manifolds

(Kanti Mardia, John Kent)

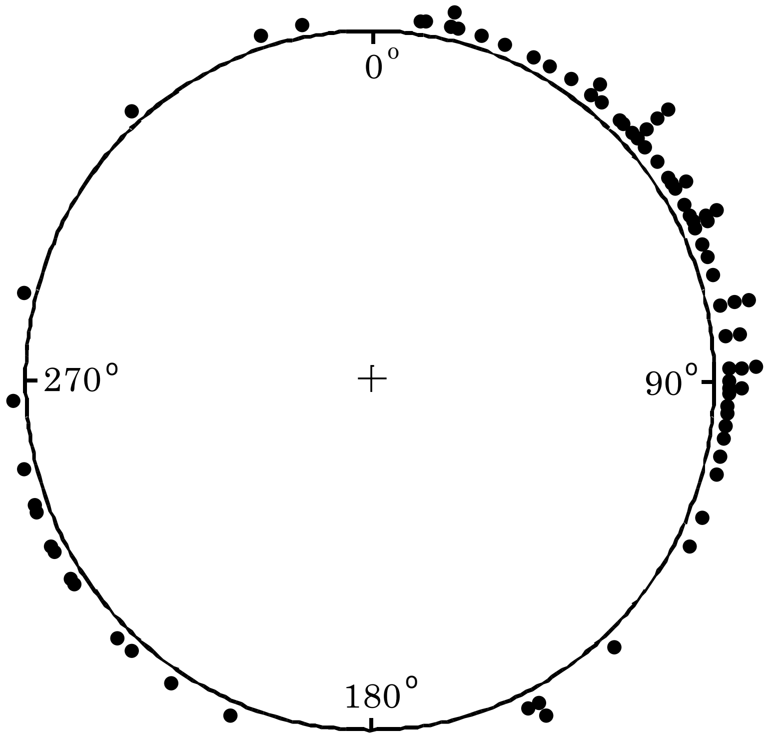
2. Some Data



Data Set 1. **Initial headings of 230 Chinese painted quail on exit from a straight, 1m long, corridor. The zero direction corresponds to the orientation of the corridor.**



Data Set 2. **Orientations of 730 red wood ants in relation to a black target placed at an angle of 180° from the zero direction. Each dot represents the direction followed by five ants.**



Data Set 3. **Orientations, measured in a clockwise direction from North (in degrees), of 76 turtles after egg laying.**

3. Data Sources

Animal orientation experiments

Earth Sciences: Paleocurrents, Paleomagnetic directions

Medicine: Posture, Flexion of limbs

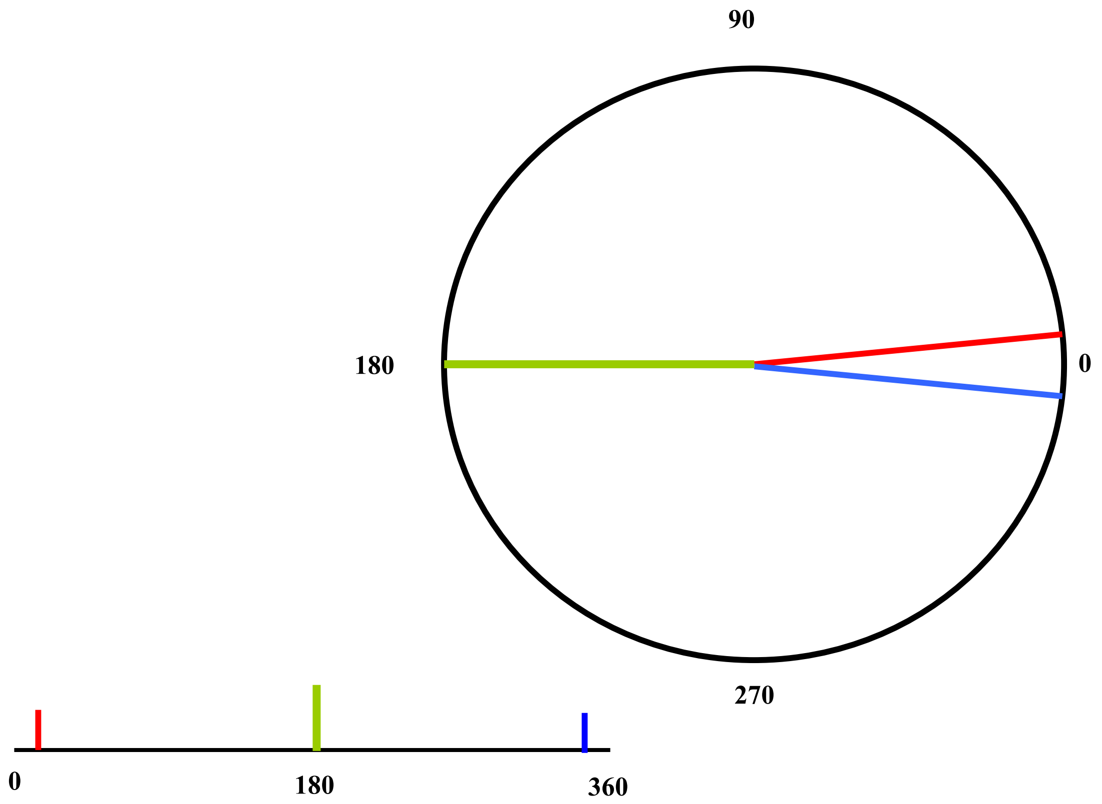
Meteorology: Wind directions, Times of day of thunderstorms

Ecology: Direction of wind (transportation of pollutants)

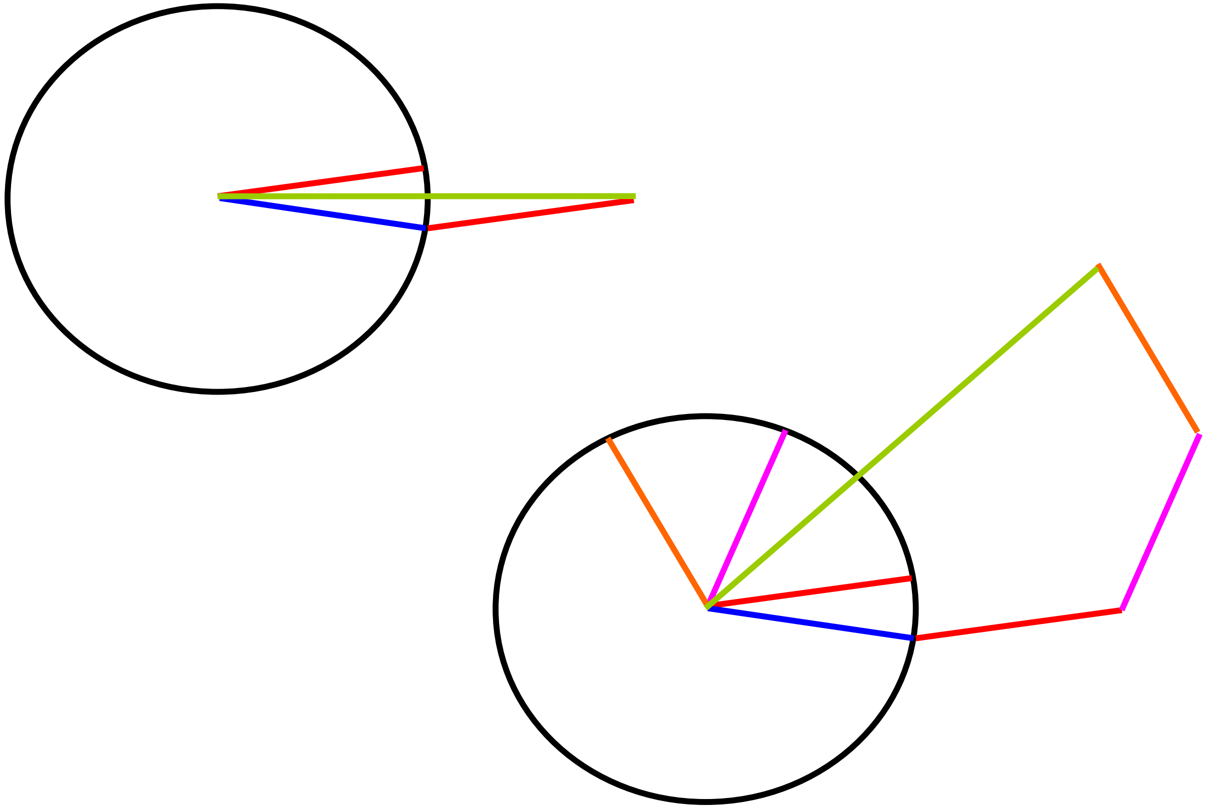
Psychology: Mental maps (representing surroundings)

Image Analysis: Machine vision, Orientation of textures, Orientation of ridges on fingerprints

4. Linear/Circular Geometry



5. Vector Resolution



6. Circular Statistics

θ - random angle; $\theta_1, \ldots, \theta_n$ - random sample of n observations

Sample mean resultant length $\bar{R} = \sqrt{a_1^2 + b_1^2}$

$$a_p = \frac{1}{n}\sum_{i=1}^{n} \cos p\,\theta_i\,, \qquad b_p = \frac{1}{n}\sum_{i=1}^{n} \sin p\,\theta_i$$

*p*th order trigonometric moments about the zero direction

If $\bar{R} = 0$, sample mean direction is undefined

If $\bar{R} > 0$, $\bar{\theta} = \begin{cases} \tan^{-1}\left(b_1/a_1\right) & \text{if } a_1 \geq 0 \\ \pi + \tan^{-1}\left(b_1/a_1\right) & \text{if } a_1 < 0 \end{cases}$ where $\tan^{-1}(x) \in [-\pi/2, \pi/2]$

Sample trigonometric moments about $\bar{\theta}$

$$\bar{a}_p = \frac{1}{n}\sum_{i=1}^{n} \cos p\left(\theta_i - \bar{\theta}\right) \quad \text{and} \quad \bar{b}_p = \frac{1}{n}\sum_{i=1}^{n} \sin p\left(\theta_i - \bar{\theta}\right)$$

$$\bar{b}_1 = \frac{1}{n}\sum_{i=1}^{n} \sin\left(\theta_i - \bar{\theta}\right) = 0$$

$$\bar{b}_2 = \frac{1}{n}\sum_{i=1}^{n} \sin 2\left(\theta_i - \bar{\theta}\right)$$

Measure of circular skewness

Batschelet (1965)

7. Population Characteristics

Characteristic function

$$\{\psi_p : p = 0, \pm 1, \ldots\} \text{ for } \psi_p = \alpha_p + i\beta_p.$$

$$\alpha_p = E(\cos p\theta), \; \beta_p = E(\sin p\theta)$$

Trigonometric moments of θ about the zero direction

$$p = 1: \quad \psi_1 = \alpha_1 + i\beta_1 = \rho\, e^{i\mu}$$

μ, mean direction; ρ, mean resultant length

Trigonometric moments about μ

$$\bar{\alpha}_p = E\{\cos p(\theta - \mu)\} \quad \text{and} \quad \bar{\beta}_p = E\{\sin p(\theta - \mu)\}$$

8. Symmetric Models

8.1 Non-wrapped

Uniform

No preferred direction; Limiting distribution of circular analogue of CLT

$$f(\theta) = \frac{1}{2\pi}, \qquad 0 \le \theta < 2\pi,$$

von Mises $VM(\mu, \kappa)$

Mardia & Jupp (1999), five constructions; Assumed advanced parametric methods

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi\, I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\},$$

$0 \le \mu < 2\pi$, mean direction; $\kappa > 0$, concentration parameter; $I_p(.)$, modified Bessel function - first kind, order p; $\rho = A(\kappa) = I_1(\kappa)/I_0(\kappa) \in [0,1]$, mean resultant length.

8.2 Wrapped Schmidt(1917)

General Result

If the linear random variable X has c.f. $\psi(t)$ then the c.f. of $\Theta = X(\bmod 2\pi)$ is $\{\psi(p) : p = 0, \pm 1, \ldots\}$.

Wrapped Normal $WN(\mu, \sigma)$

Position of particle under Brownian motion on the circle. $VM(\mu, \kappa) \approx WN(\mu, \sigma)$, $\sigma = \sqrt{-2\log[A(\kappa)]}$. Discrimination Pewsey & Jones (2003).

$$f(\theta; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \sum_{k=-\infty}^{\infty} \exp\left\{ \frac{-(\theta - \mu - 2\pi k)^2}{2\sigma^2} \right\}.$$

Wrapped Cauchy

Heavier "tails" than wrapped normal; density can be expressed in closed form.

$$f(\theta;\mu,\rho)=\frac{1}{2\pi}\left(\frac{1-\rho^2}{1+\rho^2-\rho\cos(\theta-\mu)}\right),\qquad \rho\in[0,1].$$

Symmetric wrapped α-stable Mardia (1972)

Includes wrapped normal, wrapped Cauchy, … (but not von Mises)

$$f(\theta;\mu,\alpha,\varsigma)=\frac{1}{2\pi}+\frac{1}{\pi}\sum_{k=1}^{\infty}\exp\left(-\varsigma^{\alpha}k^{\alpha}\right)\cos\{k(\theta-\mu)\},$$

where $\alpha\in(0,1)\cup(1,2]$ and $\varsigma\geq 0$.

Jones-Pewsey Jones & Pewsey (2003)

Closed form; Includes von Mises, wrapped Cauchy, … (but not wrapped normal)

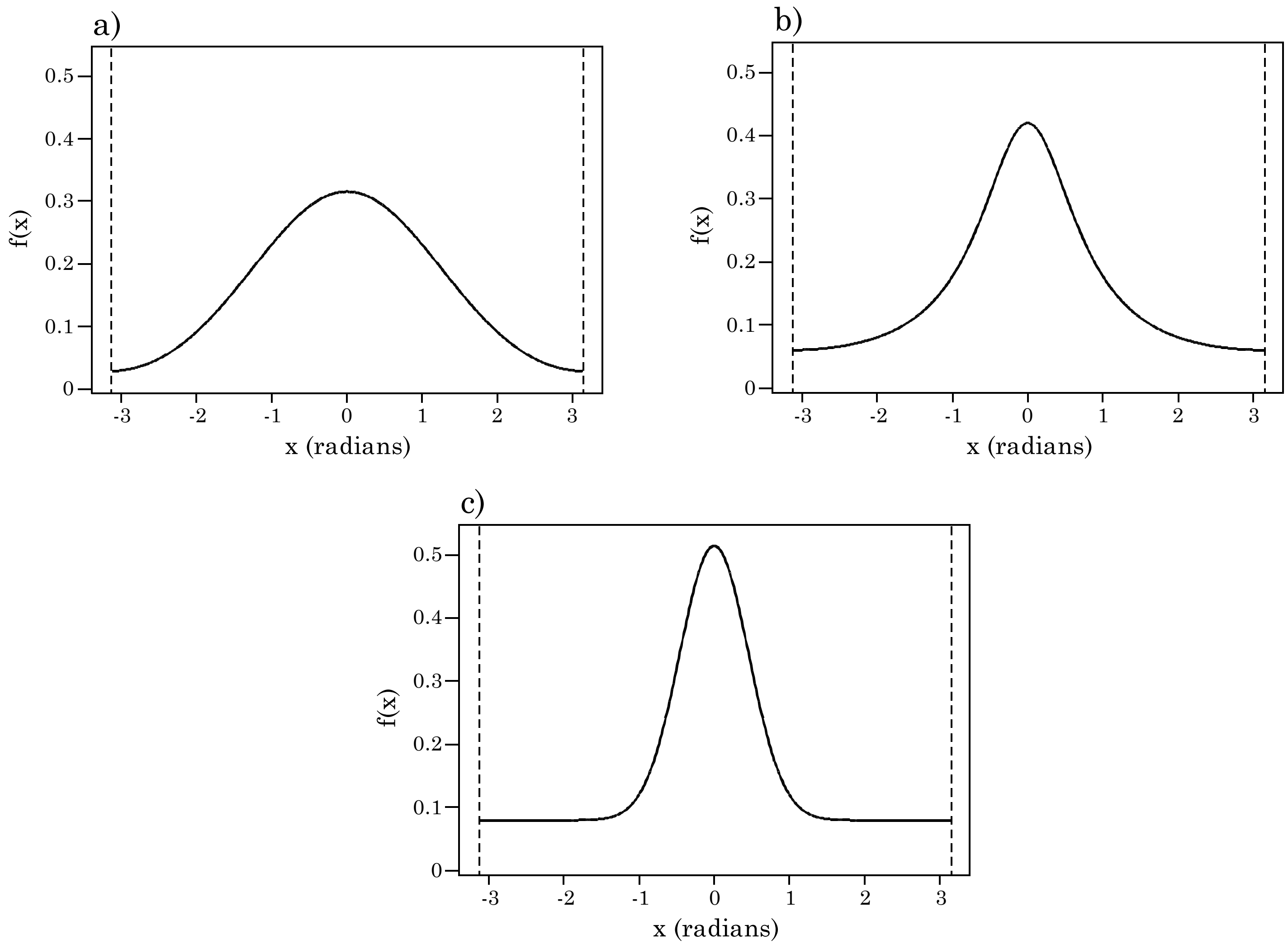
$$f(\theta;\mu,\kappa,\psi)=\frac{\{1+\tanh(\kappa\psi)\cos(\theta-\mu)\}^{1/\psi}}{2\pi\, P_{1/\psi}(\cosh(\kappa\psi))},$$

$\kappa \geq 0$, $-\infty < \psi < \infty$, $P_\nu(\cdot)$ **associated Legendre function - first kind, degree** ν**, order 0.**

Useful Mixture

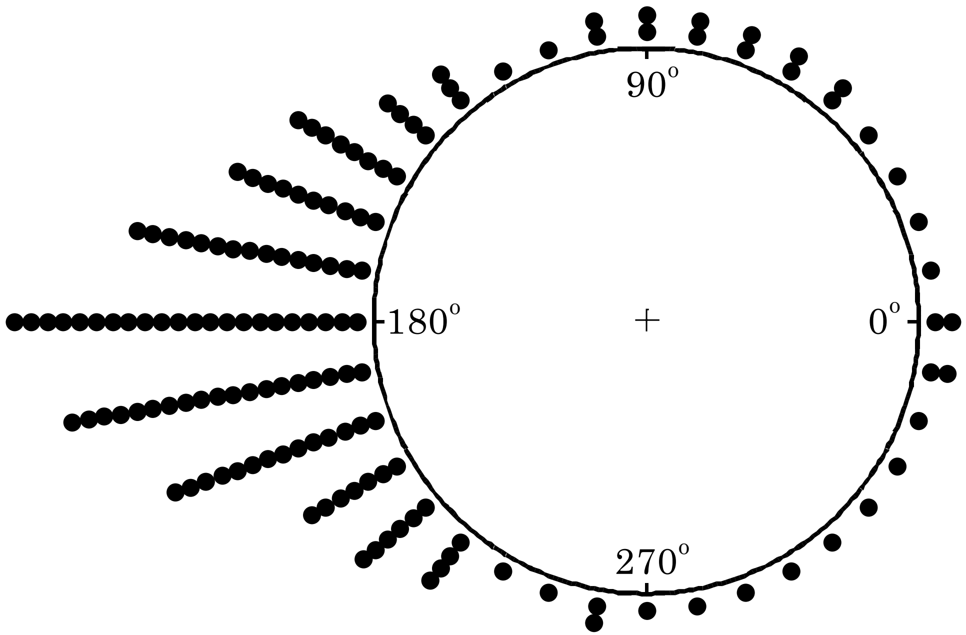
Modelling heavy tails; Uniform "background" & base "foreground".

$$g(\theta)=pf(\theta)+(1-p)\frac{1}{2\pi}$$



Symmetric densities with mean direction 0 and mean resultant length 0.45: a) wrapped normal; b) wrapped Cauchy; c) uniform and wrapped normal mixture.

9. An Example

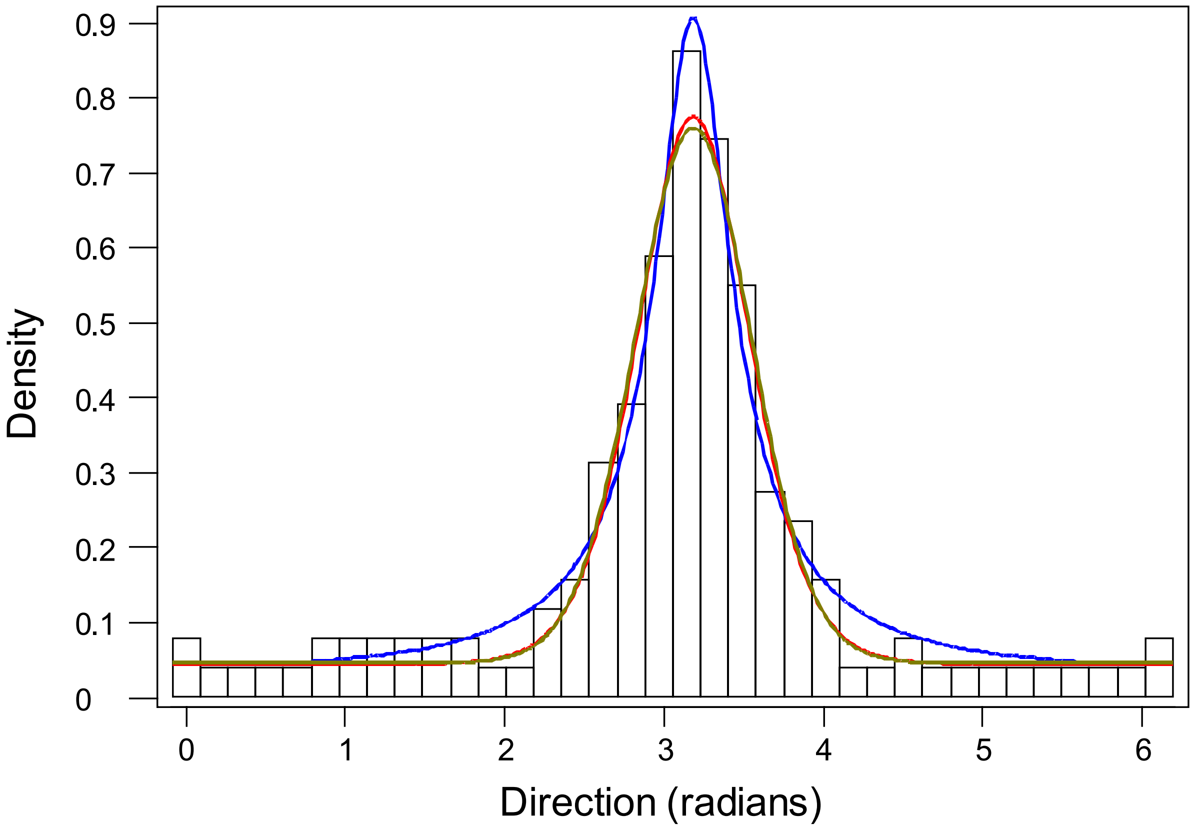


Data Set 2. **Orientations of 730 red wood ants in relation to a black target placed at an angle of 180° from the zero direction. Each dot represents the direction followed by five ants.**

9.1 Summary of Fits

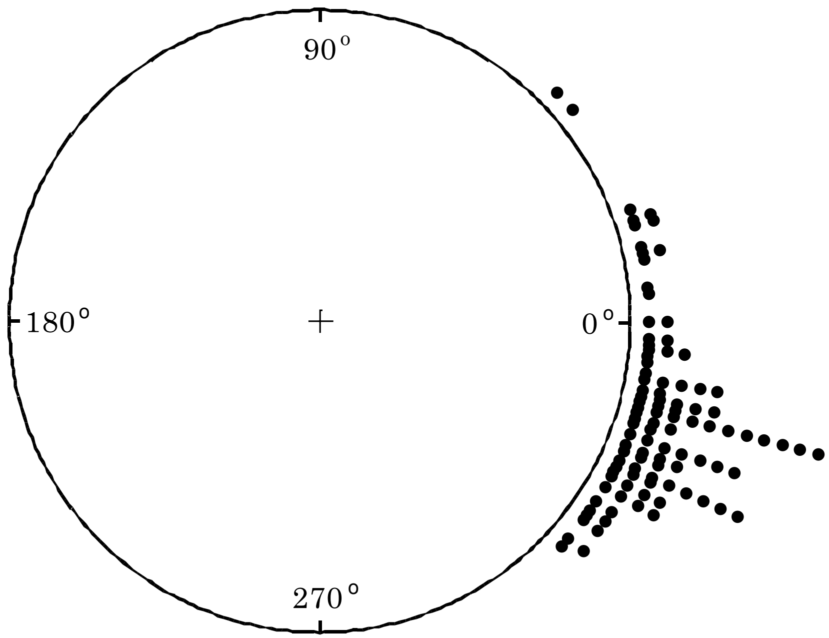
Distribution	ψ or α	μ	κ or ς	p	Log-likelihood	χ^2 g-o-f
JP	-1.30	-3.1	1.6	1	-926.55	0.07
JP+U	-0.10	-3.1	5.01	0.67	-918.72	0.32
SWS+U	1.95	-3.1	0.93	0.67	-918.89	0.30
VM+U	0	-3.1	7.54	0.66	-918.80	0.35
WN+U	2	-3.1	0.93	0.66	-918.90	0.34

9.2 Graphical Representation of Fits

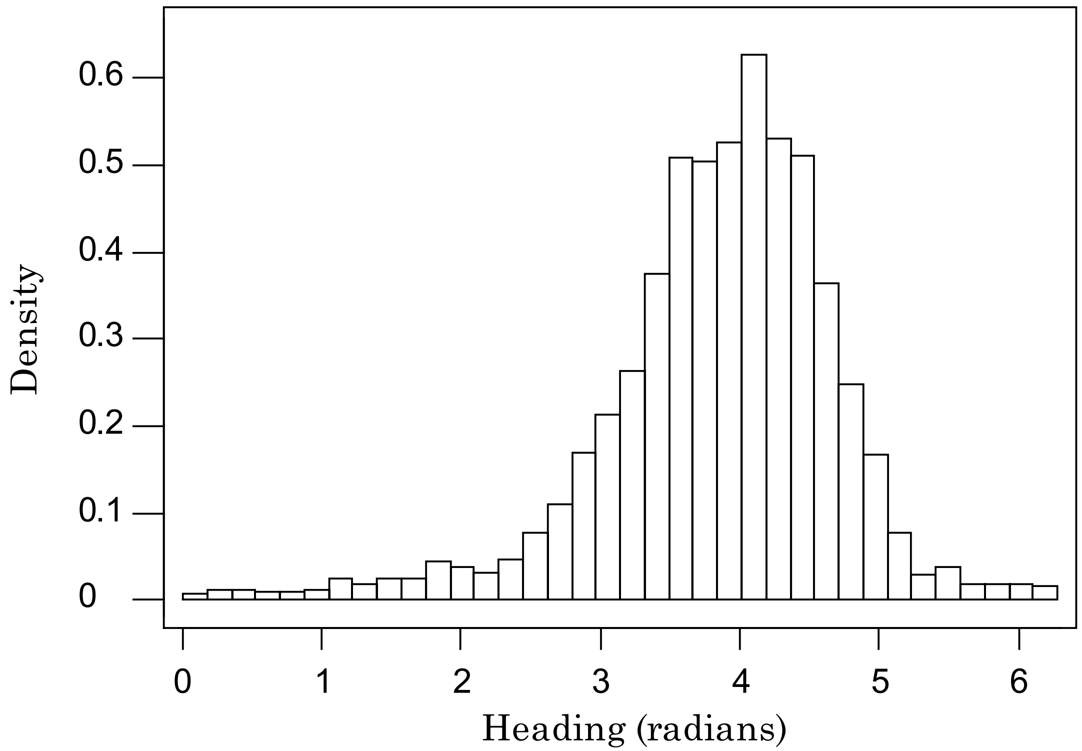


Linear histogram of red ant orientation data together with densities: Jones-Pewsey (Blue); Jones-Pewsey + Uniform (Red); von Mises + Uniform (Green)

10. Some Asymmetric Data



Data Set 4. **Initial headings of 100 Chinese painted quail on exit from a dog-leg corridor. The zero direction corresponds to the orientation of the last 0.5m section of the corridor.**



Data Set 5. **Linear histogram of 1827 bird-flight headings taken from** Bruderer & Jenni (1990).

11. “Problem” of Asymmetry

11.1 Asymmetry as a General Issue

Virtually all of established methodology for the analysis of circular data assumes symmetry.

Symmetry is a tacit assumption in many areas of Statistics (indeed in the Arts and Sciences in general).

Asymmetry not a “problem” unique to the modelling of circular data.

11.2 Asymmetric Data on the Line

Box-Cox transformation: bias on transforming back

Increased interest in modelling data on the scale they were originally observed.

Azzalini & co-authors (1985, 1996, 1999,…)

Jones (2001, 2002,…)

Arnold & co-authors (2000, 2002,…)

11.3 Asymmetric Compositional Data

MOVE/STAY methodology of Aitchinson, Pawlowsky-Glahn and co-workers (Girona/Barcelona).

11.4 Asymmetric Circular Data

Modelling on the original "scale" is an absolute necessity as circle is compact. There does not even exist an equivalent to "standardisation" on the circle. Any form of transformation other than rotation or reflection changes the relative positions of the observations.

12. Models Capable of Modelling Asymmetry

12.1 Papakonstantinou (1979)

$$f(\theta;\kappa,\nu)=\frac{1}{2\pi}+\frac{\kappa}{2\pi}\sin\left(\theta+\nu\sin\theta\right),$$

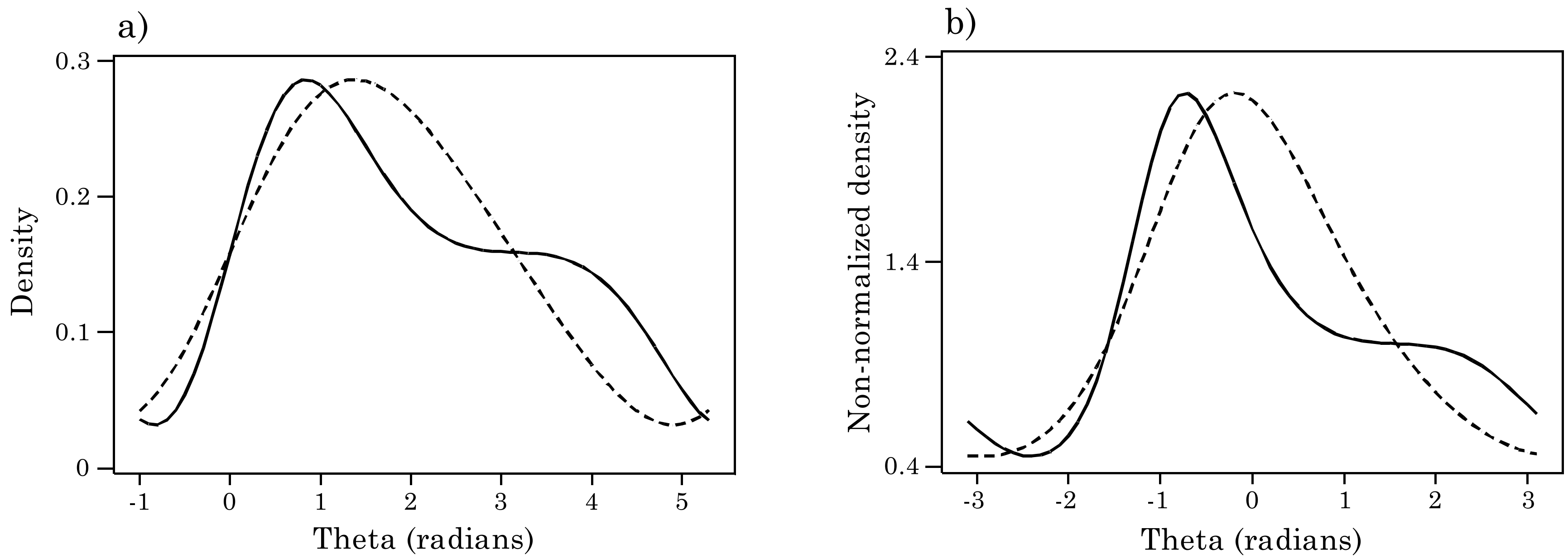
where $|\kappa|\le 1$ and $|\nu|<1$. Shape depends on both κ and ν, the second determining skewness.

12.2 Batschelet (1981)

Extension of von Mises distribution.

$$f(\theta;\kappa,\nu)=c\exp\{\kappa\cos(\theta+\nu\cos\theta)\},$$

$|\nu|<1$, c a normalising constant. Again, ν is skewness parameter.



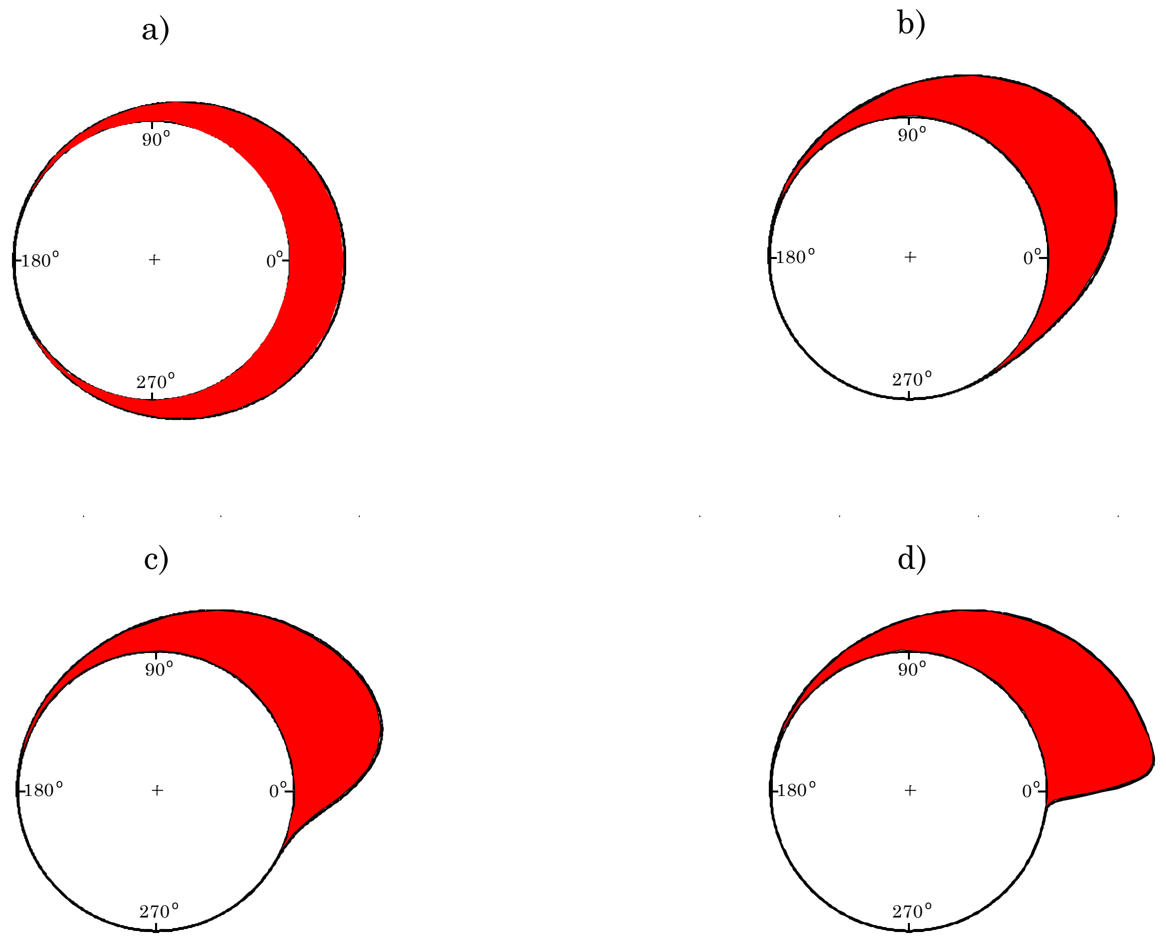
Linear plots of a) Papakonstantinou and b) Batschelet densities. Both pairs of curves correspond to the choices κ = 0.8 and ν = 0.2 (dashed curve) and ν = 0.98 (solid curve). In b) the normalising constant *c* has been set equal to 1.

12.3 Wrapped Skew-normal Pewsey(2000, 2004a)

Wrapped normal and wrapped half-normal as special cases.

$$f(\theta;\xi,\eta,\lambda)=\frac{2}{\eta}\sum_{k=-\infty}^{\infty}\phi\left(\frac{\theta+2\pi k-\xi}{\eta}\right)\Phi\left\{\lambda\left(\frac{\theta+2\pi k-\xi}{\eta}\right)\right\},$$

$\phi(\cdot)$ and $\Phi(\cdot)$, density and distribution function of standard normal distribution: $-\infty<\xi<\infty$, $\eta>0$, and $-\infty<\lambda<\infty$, location, scale and skewness parameters, respectively.



WSNC$(0,1,\lambda)$ densities with: a) $\lambda = 0$ (wrapped standard normal); b) $\lambda = 2$; c) $\lambda = 5$; d) $\lambda = 20$.

12.4 Wrapped α-stable

Capable of modelling varying degrees of asymmetry.

$$f(\theta;\alpha,\beta,\varsigma)=\frac{1}{2\pi}+\frac{1}{\pi}\sum_{k=1}^{\infty}\exp\left(-\varsigma^{\alpha}k^{\alpha}\right)\cos\left\{k(\theta-\mu)-\varsigma^{\alpha}k^{\alpha}\beta\tan\frac{\alpha\pi}{2}\right\},$$

$\alpha\in(0,1)\cup(1,2]$, $|\beta|\le 1$, $\varsigma\ge 0$. **Work in progress.**

12.5 Asymmetric Jones-Pewsey

Work in progress.

13. Modelling Strategies

13.1 Strategy A

Test for Symmetry Pewsey (2002)	
Symmetric	Asymmetric
Fit individual (symmetric) models or flexible symmetric family (e.g. JP or SWS)	Fit individual (asymmetric) models or flexible (asymmetric) family (e.g. WS)

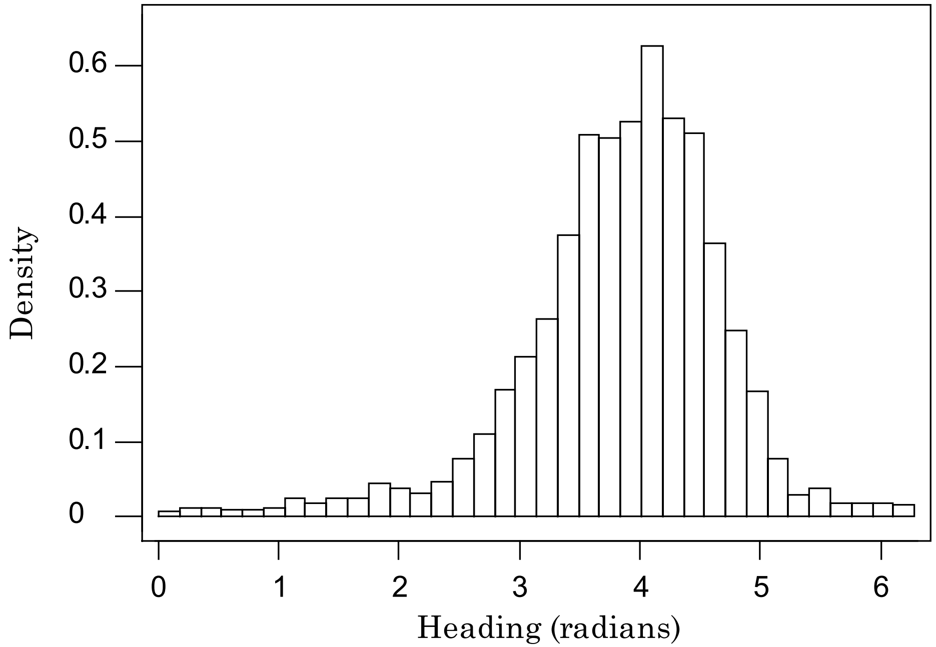
13.2 Strategy B

Fit a flexible family capable of modelling both symmetry and asymmetry

(e.g. WS)

Model refinement using usual likelihood based machinery

14. An Example



Data Set 5. **Linear histogram of 1827 bird-flight headings taken from** Bruderer & Jenni (1990).

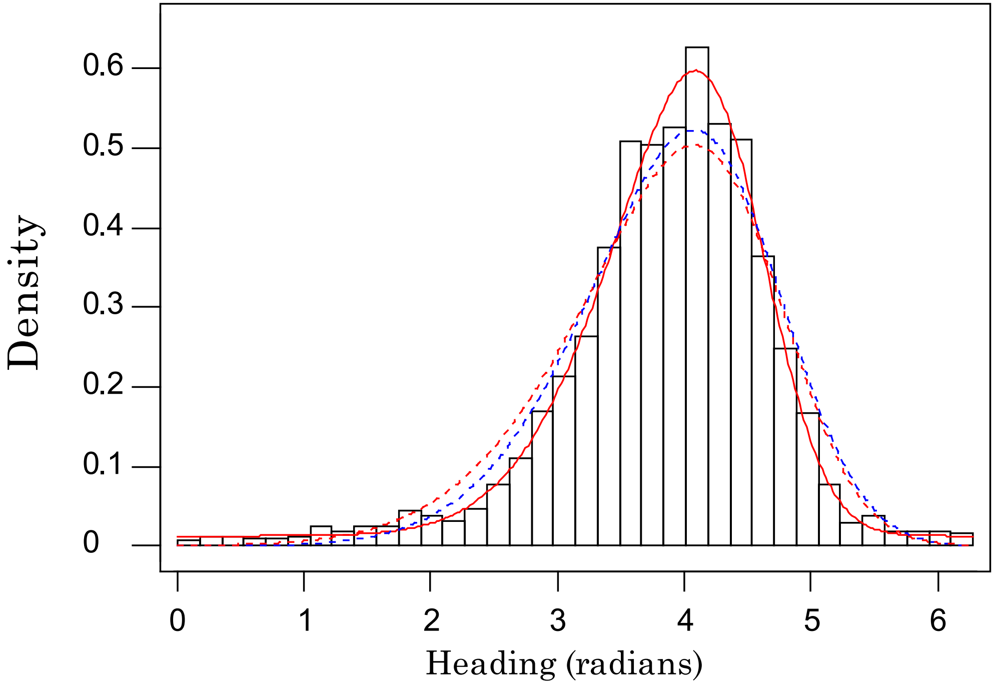
14.1 Summaries of Fits

$$f(\theta; p, \xi, \eta, \lambda) = \frac{(1-p)}{2\pi} + p\frac{2}{\eta}\sum_{r=-\infty}^{\infty} \phi\left(\frac{\theta + 2\pi r - \xi}{\eta}\right)\Phi\left\{\lambda\left(\frac{\theta + 2\pi r - \xi}{\eta}\right)\right\}$$

Test for symmetry *p*-value ≈ 0						
Method	p	ξ	η	λ	Log-likelihood	p-value χ^2 g-o-f
MM	1	4.7	1.1	-1.8	-	0
ML	1	4.7	1.2	-2.2	-2202.1	0
ML	0.9	4.6	0.9	-2.1	-2128.0	0.2

Estimated mean direction: 3.8 radians (218 degrees)
95% profile likelihood based CIs for p and λ: (0.876, 0.925) and (-2.63, -1.56)

14.2 Graphical Representation of Fits



Histogram of bird-flight headings together with fitted densities: ML WSNC+U solution (solid red); ML WSNC solution (broken red); MM WSNC solution (broken blue).

References

Arnold, B. C. & Beaver, R. J. (2000) Hidden truncation models. *Sankhya*, Series A, 62, 23-35.

Arnold, B. C. & Beaver, R. J. (2002) Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, 11, 7-54.

Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.

Azzalini, A. & Capitanio, A. (1999) Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society*, Series B, 61, 579-602.

Azzalini, A. & Dalla Valle, A. (1996) The multivariate skew-normal distribution. *Biometrika*, 83, 715-726.

Batschelet, E. (1965) *Statistical Methods for the Analysis of Problems in Animal Orientation and Certain Biological Rhythms*. Washington DC: American Institute of Biological Sciences.

Batschelet, E. (1981) *Circular Statistics in Biology*. London: Academic Press.

Bruderer, B. & Jenni, L. (1990) Migration across the Alps. In *Bird Migration: Physiology and Ecophysiology* (ed. E. Gwinner), pp. 60-77. Berlin: Springer-Verlag.

Jones, M. C. (2001) A skew *t* distribution. In *Probability and Statistical Models With Applications* (eds. C. A. Charalambides, M. V. Koutras & N. Balakrishnan), pp. 269-278. London: Chapman & Hall/CRC.

Jones, M. C. (2002) Marginal replacement in multivariate densities, with application to skewing spherically symmetric distributions. *Journal of Multivariate Analysis*, 81, 85-99.

Jones, M. C. & Pewsey, A. (2003) A family of symmetric distributions on the circle. To appear

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

Mardia, K. V. & Jupp, P. E. (1999) *Directional Statistics*. Chichester: Wiley.

Papakonstantinou, V. (1979) *Beiträge zur zirkulären Statistik*. PhD dissertation. University of Zurich, Switzerland.

Pewsey, A. (2000) The wrapped skew-normal distribution on the circle. *Communications in Statistics – Theory and Methods*, 29, 2459-2472.

Pewsey, A. (2002) Testing circular symmetry. *The Canadian Journal of Statistics*, 30, 591-600.

Pewsey, A. (2004a) Modelling asymmetrically distributed circular data using the wrapped skew-normal distribution. *Environmental & Ecological Statistics*. To appear.

Pewsey, A. & Jones, M.C. (2003) Discrimination between the von Mises and wrapped normal distributions: just how big does the sample size have to be? To appear.

Schmidt, W. (1917) Statistische Methoden beim Gefügestudium Kristalliner Schiefer, *Sitz. Kaiserl. Akad. Wiss. Wien, Math. – nat. Kl. Abt.* 1, 126, 515-538.