/pgf/stepx/.initial=1cm,
/pgf/stepy/.initial=1cm,
/pgf/step/.code=1/pgf/stepx/.expanded=-
10.95415pt,/pgf/stepy/.expanded=-
10.95415pt,
/pgf/step/.value
re-
quired

/pgf/images/width/.estore
in=
/pgf/images/height/.estore
in=
/pgf/images/page/.estore
in=
/pgf/images/interpolate/.cd,.code=,.default=true
/pgf/images/mask/.code=
/pgf/images/mask/matte/.cd,.estore
in=,.value

/pgf/imagesheight=,width=,page=,interpolate=false,mask=,width=14pt,hei
/pgf/imagesheight=,width=,page=,interpolate=false,mask=,width=14pt,hei
/pgf/imagesheight=,width=,page=,interpolate=false,mask=,width=11pt,hei
/pgf/imagesheight=,width=,page=,interpolate=false,mask=,width=11pt,hei
/pgf/imagesheight=,width=,page=,interpolate=false,mask=,width=14pt,hei
/pgf/imagesheight=,width=,page=,interpolate=false,mask=,width=14pt,hei

Joint work with Nottingham colleagues Simon Preston and Michail Tsagris.

Regression Models for Directional Data

Andrew Wood
School of Mathematical Sciences
University of Nottingham

ADISTA14 Workshop, Wednesday 21 May 2014

Joint with Nottingham colleagues Simon Preston and Michail Tsagris.

Supported by EPSRC.

Outline of Talk

1. Introduction and background discussion.
2. Regression models with Fisher error distribution.
3. Regression models with Kent error distribution.
4. Some numerical results.
5. Discussion.

Data Structure

Data structure : $\{y_1, x_1\}, \ldots, \{y_n, x_n\}$,

where

- for each i, y_i is a response vector and x_i is a covariate vector;
- $y_i \in \mathcal{S}^2 = \{y_i \in \mathbf{R}^3 : y_i^\top y_i = 1\}$, i.e. y_i is a unit vector in 3D space;
- the covariate vectors are p-dimensional, i.e. $x_i \in \mathbf{R}^p$;
- it is assumed throughout the talk that $y_1, \ldots, y_n$ are independent;

The main purpose here is to develop parametric regression models on the sphere in which rotational symmetry of the error distribution is not assumed.

Multivariate linear model

Consider the standard multivariate linear model:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E},$$

where

- $\mathbf{Y}$ $(n \times p)$ is the response matrix;
- $\mathbf{X}$ $(n \times q)$ is the covariate matrix;
- $\mathbf{B}$ $(q \times p)$ is the parameter matrix;
- $\mathbf{E}$ $(n \times p)$ is the unobserved error matrix whose rows are assumed to be IID with common distribution $N_p(\mathbf{0}, \mathbf{\Sigma})$.

In this situation, the least squares estimator (and MLE) of $\mathbf{B}$,

$$\hat{\mathbf{B}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y},$$

does not depend on $\mathbf{\Sigma}$.

However, we would still not want to assume that $\mathbf{\Sigma}$ is a scalar multiple of the identity, which is the analogue of the assumption of rotational symmetry of the error distribution on the sphere.

A general covariate vector x_i

A second purpose is to develop models which can handle a general covariate x_i,

i.e. special structure for x_i such as $x_i \in \mathcal{S}^2$ or $x_i \in R$ is not assumed.

The spirit of the modelling here is similar to that often used with generalised linear models (GLMs).

The use of GLMs is sometimes a bit cavalier in the following respects:

- for simplicity and convenience we assume covariate information combines in linear fashion (i.e. through the 'linear predictor') on a suitable scale (i.e. after applying the 'link function');
- we do not expect our models to extrapolate well to regions in which the data are not well represented.

We shall make similar assumptions here.

Fisher error distribution

We will be focusing on the following two error distributions on the sphere:

- Rotationally **symmetric** case: the Fisher (or von Mises-Fisher) distribution.
- Rotationally **asymmetric** case: the Kent distribution.

The Fisher distribution has pdf given by

$$f(y|\kappa, \mu) = c_F(\kappa)^{-1} \exp(\kappa y^\top \mu),$$

where $\mu \in \mathcal{S}^2$ is a unit vector, the mean direction, and $\kappa \geq 0$ is a concentration parameter.

In the case of the Fisher distribution, we can construct a regression model with constant concentration parameter κ, by specifying μ to be of the form

$$\mu = \mu(x, \gamma), \qquad \mu_i = \mu(x_i, \gamma), \quad i = 1, \ldots, n,$$

with γ a parameter vector, and x_i the covariate vector for observation i.

Particular cases of the Fisher regression model

In cases (i)–(iii) below it is assumed that $y_i, x_i \in \mathcal{S}^2$.

Case (i) [Chang, AoS, 1986; Rivest, AoS, 1989]
Here, $\mu_i = Ax_i$, where A is an unknown rotation matrix to be estimated.

Case (ii) [Downs, Biometrika, 2003]
A regression model on the sphere based on Möbius transformations.

Case (iii) [Rosenthal et al., JASA, 2014]
In this case,

$$\mu_i = \frac{Ax_i}{||Ax_i||},$$

where A is a non-singular 3×3 matrix to be estimated.

Particular cases of the Fisher regression model (continued)

Case (iv) In this case,

$$\mu_i = QR_i\delta,$$

where Q is a 3×3 orthogonal matrix, and $\delta \in \mathcal{S}^2$ is a unit vector. One way to define the 3×3 rotation matrix R_i is by

$$R_i = \exp(C_i), \quad \text{where} \quad C_i = \begin{pmatrix} 0 & c_{1i} & c_{2i} \\ -c_{1i} & 0 & c_{3i} \\ -c_{2i} & -c_{3i} & 0 \end{pmatrix}$$

is skew-symmetric, and $c_{ji} = \gamma_j^\top x_i$, $j = 1, 2, 3$.

A second possibility is to define

$$R_i = (I_3 + C_i)(I - C_i)^{-1},$$

which has the advantage that it is not a many-to-one mapping.

The Kent distribution on $\mathcal{S}^2$

Kent (1982) proposed a 5 parameter family which contains the Fisher family as well as rotationally asymmetric distributions.

This has pdf

$$f(y; \mu, \xi_1, \xi_2, \kappa, \beta) = c_K(\kappa, \beta)^{-1} \exp\left[\kappa\mu^\top y + \beta\{(\xi_1^\top y)^2 - (\xi_2^\top y)^2\}\right]$$

where

- μ, ξ_1 and ξ_2 are mutually orthogonal unit vectors;
- $\kappa \geq 0$ and $\beta \geq 0$ are concentration and shape parameters.

The distribution is unimodal if $\kappa > 2\beta$.

Motivation for Kent distributions

- Datasets in directional statistics and shape analysis are often highly concentrated.
- When highly concentrated datasets are projected onto a suitable tangent space, a multivariate normal approximation in the tangent space is often reasonable.
- When the Kent distribution is highly concentrated and unimodal, the induced distribution in the tangent space at the mode is approximately bivariate normal.

Orthonormal basis determined by a unit vector

Consider a unit vector $\begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}$.

Provided $\sin\theta \neq 0$,

$$\begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}, \quad \begin{pmatrix} -\sin\phi \\ \cos\phi \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ -\sin\theta \end{pmatrix}.$$

In cartesian coordinates, assuming $x_1^2 + x_2^2 \neq 0$, we have

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \quad \begin{pmatrix} -y_2/\sqrt{y_1^2 + y_2^2} \\ y_1/\sqrt{y_1^2 + y_2^2} \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} y_1 y_3/\sqrt{y_1^2 + y_2^2} \\ y_2 y_3/\sqrt{y_1^2 + y_2^2} \\ -\sqrt{1 - y_3^2} \end{pmatrix}.$$

Model for axes

Given that $\mu_i = \mu(x_i, \gamma)$, use the previous slide to define an orthonormal triple

$$\mu_i, \quad \xi_{1i}, \quad \xi_{2i}, \qquad \xi_{1i} = \xi_{1i}(x_i, \gamma), \quad \xi_{2i} = \xi_{2i}(x_i, \gamma).$$

Now suppose

$$y_i \sim \mathrm{Kent}(\mu_i, \tilde{\xi}_{1i}, \tilde{\xi}_{2i}, \kappa, \beta), \qquad i = 1, \ldots, n,$$

where

$$\tilde{\xi}_{1i} = (\cos \psi_i)\xi_{1i} + (\sin \psi_i)\xi_{2i}, \qquad \tilde{\xi}_{2i} = -(\sin \psi_i)\xi_{1i} + (\cos \psi_i)\xi_{2i},$$

and $\psi_i = \psi(x_i, \alpha)$, i.e. ψ_i depends on the covariate vector x_i and a further parameter vector α.

Parameters in the Kent regression model

There are three parameter vectors: α; γ; and (κ, β).

In particular:

- The parameter vector γ, along with the covariate vector x_i, determines the mean direction μ_i, and also the orthonormal triple μ_i, ξ_{1i} and ξ_{2i}.
- The parameter vector α, along with x_i, determines the angle ψ_i. Note that the angle ψ_i is measured relative to the coordinate system determined by the orthonormal triple μ_i, ξ_{1i} and ξ_{2i}, which depends on i.
- The parameters $\kappa \geq 0$ and $\beta \geq 0$ are dispersion parameters. If so desired, one could allow κ and β to depend on x_i and further parameters, but for simplicity we assume here than κ and β do not depend on i.

The log-likelihood under the Kent model

The log-likelihood under the Kent model is given by

$$\ell(\alpha, \beta, \gamma, \kappa) = -n \log c_K(\kappa, \beta) + \kappa \sum_{i=1}^{n} y_i^\top \mu_i + \beta \sum_{i=1}^{n} \left\{ \left(y_i^\top \tilde{\xi}_{1i} \right)^2 - \left(y_i^\top \tilde{\xi}_{2i} \right)^2 \right\},$$

where, as above, $\mu_i = \mu(x_i, \gamma)$,

$$\tilde{\xi}_{1i} = (\cos \psi_i) \xi_{1i} + (\sin \psi_i) \xi_{2i}, \qquad \tilde{\xi}_{2i} = -(\sin \psi_i) \xi_{1i} + (\cos \psi_i) \xi_{2i},$$

$\xi_{1i} = \xi_{1i}(x_i, \gamma)$, $\xi_{2i} = \xi_{2i}(x_i, \gamma)$, $\psi_i = \psi_i(x_i, \alpha)$, and $c_K(\kappa, \beta)$ is the Kent normalising constant.

A key point: this construction is generic, in the sense that we can use any suitable rotationally symmetric Fisher regression model for $\mu_i = \mu(x_i, \gamma)$, and any suitable (double) von Mises model for $\psi_i = \psi_i)x_i, \alpha)$.

A structured (or switching) optimisation algorithm

When maximising the log-likelihood, we have found it desirable to cycle between 3 steps:

Step 1: maximise $\ell(\alpha^{(m)}, \beta^{(m)}, \gamma^{(m)}, \kappa^{(m)})$ over γ with $\alpha^{(m)}$, $\beta^{(m)}$, $\kappa^{(m)}$ held fixed, to obtain $\gamma^{(m+1)}$.

Step 2: maximise $\ell(\alpha^{(m)}, \beta^{(m)}, \gamma^{(m+1)}, \kappa^{(m)})$ over α with $\beta^{(m)}$, $\gamma^{(m+1)}$ and $\kappa^{(m)}$ held fixed, to obtain $\alpha^{(m+1)}$.

Step 3: maximise $\ell(\alpha^{(m+1)}, \beta^{(m)}, \gamma^{(m+1)}, \kappa^{(m)})$ over β and κ with $\alpha^{(m+1)}$ and $\gamma^{(m+1)}$ held fixed, to obtain $\beta^{(m+1)}$ and $\kappa^{(m+1)}$.

Comments on the algorithm

- The normalising constant $c_K(\kappa, \beta)$ can be calculated exactly (using the Holonomic gradient method), or approximately, to a high level of accuracy (using e.g. the Kume and Wood (2005) saddlepoint approximation). So Step 3 in the algorithm is relatively straightforward.
- Step 1 can be performed by modifying (because of the addition of the 'quadratic' term) whatever algorithm is used to fit the rotationally symmetric Fisher model.
- Step 2 is equivalent to fitting a weighted (double) von Mises regression model for the ψ_i. The weight for observation i is proportional to

$$\left(y_i^\top \xi_{1i}\right)^2 + \left(y_i^\top \xi_{2i}\right)^2.$$

Step 2 is simplified due to the following result.

A useful lemma

Consider a two-parameter natural exponential family likelihood based on an IID sample of size n:

$$\ell(\theta_1, \theta_2) = -n \log c(\theta_1, \theta_2) + \theta_1 t_1 + \theta_2 t_2,$$

where θ_1 and θ_2 are natural parameters and (t_1, t_2) is the sufficient statistic.

Let

$$h(\bar{t}_1, \bar{t}_2) = \ell\{\hat{\theta}_1(\bar{t}_1, \bar{t}_2), \hat{\theta}_2(\bar{t}_1, \bar{t}_2)\}$$

denote the maximised likelihood, viewed as a function of $\bar{t}_1 = t_1/n$ and $\bar{t}_2 = t_2/n$.

Lemma. $h(\bar{t}_1, \bar{t}_2)$ is an increasing function of $\bar{t}_1$ at $(\bar{t}_1, \bar{t}_2)$ if and only if $\hat{\theta}_1 = \hat{\theta}_1(\bar{t}_1, \bar{t}_2)$ is strictly positive at $(\bar{t}_1, \bar{t}_2)$.

Application of the lemma

An application of the lemma to Step 2 of the computation algorithm shows that, because $\beta \geq 0$, Step 2 is equivalent to maximising over α the 'quadratic' term in the log likelihood, namely

$$\sum_{i=1}^{n} \left\{ \left(y_i^\top \tilde{\xi}_{1i} \right)^2 - \left(y_i^\top \tilde{\xi}_{2i} \right)^2 \right\}.$$

This is equivalent to maximising the log-likelihood of a weighted (double) von Mises regression model.

An alternative approach

Recall the Fisher regression model in Case (iv):

$$y_i \sim \mathrm{Fisher}(\mu_i, \kappa),$$

where $\mu_i = QR_i\delta$ with Q and R_i (3×3) rotational matrices and $delta \in \mathcal{S}^2$.

We could modify to a Kent distribution by writing

$$y_i \sim \mathrm{Kent}(\mu_i, \xi_{1i}, \xi_{2i}, \kappa, \beta),$$

where μ_i, ξ_{1i} and ξ_{2i} are an orthonormal triple determined by

$$QR_i = (\mu_i, \xi_{1i}, \xi_{2i}),$$

with $R_i = R(x_i, \gamma)$.

Animal (seal) tracking data

Some numerical results from the animal movement tracking data found in Jonsen et al. (2005).

Estimated parameters and log-likelihood for the Fisher and Kent regression models assuming linear predictor is linear in time.

Fisher

log-likelihood $= 377.6, \quad \hat{\kappa} = 602.2$

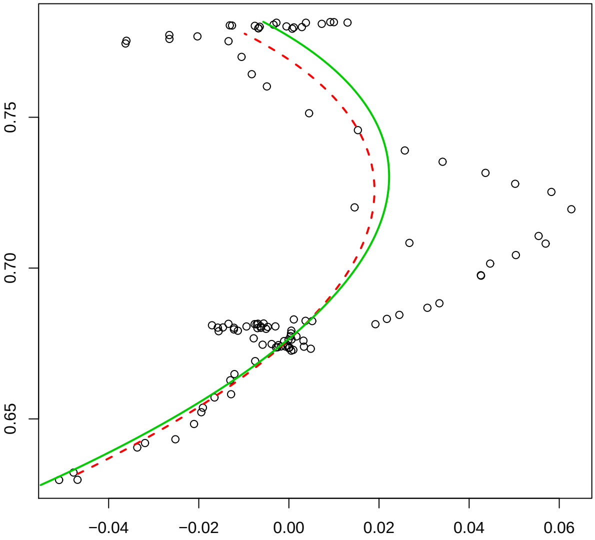
$\hat{\gamma} = (-0.00035, -0.00419, -0.01159)^\top$

Kent

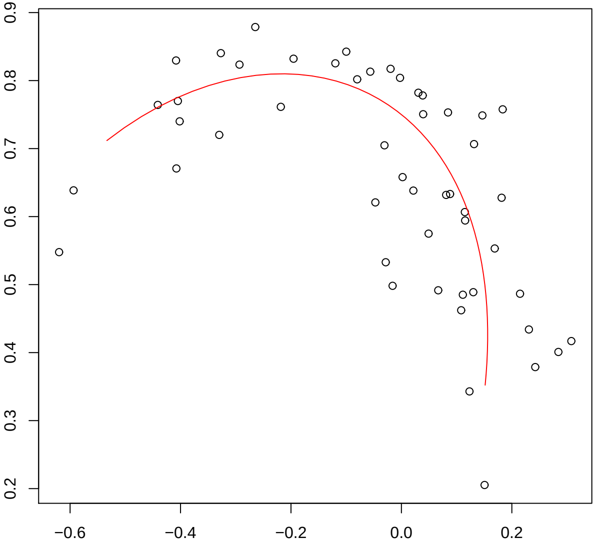
log-likelihood $= 432.8, \quad \hat{\kappa} = 1744.1, \quad \hat{\beta} = 713.367$

$\hat{\gamma} = (-0.00052, -0.00443, -0.01203)^\top$

Animal (seal) tracking data plot



A single simulated example



$\gamma = (-0.02, 0.03, 0.004)^\top$

$\hat{\gamma} = (-0.0192, 0.0304, 0.0164)^\top$

True $\kappa = 50$ and estimated κ, $\hat{\kappa}$, is 43.41 with $n = 50$.

The x_i are scalars from a normal with mean 0 and standard deviation 10.

The latitude goes from 4 to 66.8 degrees and the longitude from 9 to 329 degrees.

Some simulation results

Some simulation results from the Fisher model under Case (iv).

	True betas	-0.02	0.03	0.004	
		Estimated Betas			Estimated kappa
N=50	Kappa=20	-0.0200	0.03010	0.00360	21.432
	Kappa=50	-0.0199	0.03010	0.00370	53.595
N=100	Kappa=20	-0.0201	0.02997	0.00398	20.669
	Kappa=50	-0.01995	0.03002	0.00401	52.074

Table 1: Estimated parameters (averages over 500 simulations).

	Root mean squared errors				
		Estimated Betas			Estimated kappa
N=50	Kappa=20	0.0023	0.0024	0.0090	3.5874
	Kappa=50	0.0016	0.0014	0.0055	8.3020
N=100	Kappa=20	0.00101	0.0009	0.0015	2.1865
	Kappa=50	0.00063	0.00055	0.00084	5.6865

Table 2: Square root of the MSEs of the parameters (500 simulations)

Selected References

Chang, T. (1986). Spherical regression. *Annals of Statistics*, **14**, 907–924.

Downs, T.D. (2003). Spherical regression. *Biometrika*, **90**, 655–668.

Jonsen, I.D., Flemming, J.M., Myers, R.A. (2005). Robust state-space modelling of animal movement data. *Ecology*, **86**, 2874–2880.

Kent, J.T. (1982). The Fisher–Bingham distribution on the sphere. *J.R. Statist. Soc.* B **44**, 285–99.

Rivest, L-P. (1989). Spherical regression for concentrated Fisher-von Mises distributions. *Annals of Statistics*, **17**, 307–317.

Rosenthal, M., Wei, W., Klassen, E. & Srivastava, A. (2014). Spherical regression models using projective linear transformations. *Journal of the American Statistical Association* (to appear).