Modelling and simulation of directional distributions

John Kent
University of Leeds

j.t.kent@leeds.ac.uk
http://maths.leeds.ac.uk/~john

ADISTA 2014

The simulation work is joint with
Asaad Ganeiber and Kanti Mardia

Overview

Directional data analysis

- originally concerned with data on the circle ($S_1 \subset \mathbb{R}^2$) and sphere ($S_2 \subset \mathbb{R}^3$).
- extended to more general spheres ($S_p \subset \mathbb{R}^q,\ p+1=q \geq 2$), real and complex projective spaces, Stiefel manifolds, Grassman manifolds, spaces of rotation matrices, shape spaces, product spaces... .

Models

How to construct statistical models on such spaces?

- at one extreme should include the uniform distribution
- under high concentration should mimic the multivariate normal distribution in a tangent space approximation about a point
- typically based on exponential families
- inference often straightforward except for the normalizing constant

Example — von Mises distribution on the circle S_1.

In angular coordinates the density takes the form

$$f(\theta) \propto \exp\{\kappa \cos(\theta - \alpha)\} \text{ wrt } d\theta$$

where the angle α is a location parameter and $\kappa \geq 0$ is a concentration parameter. For $\kappa = 0$ the distribution is uniform; for large κ,

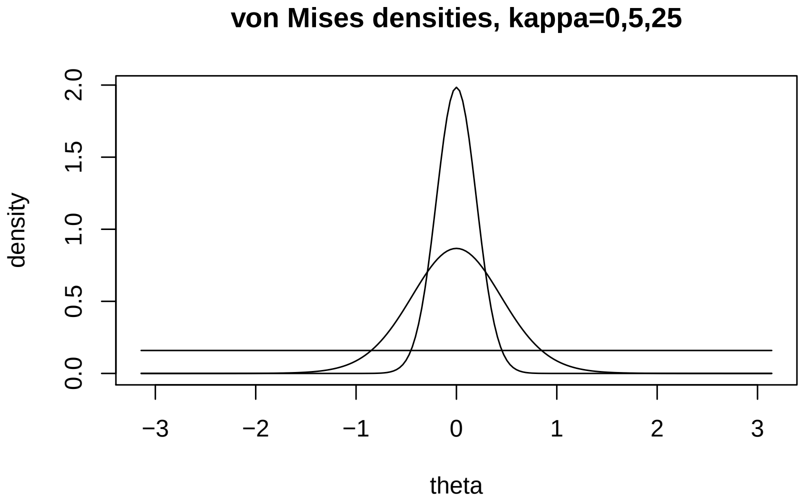
$$\theta - \alpha \approx N(0, 1/\kappa)$$

The density can also be written Euclidean coordinates. Let $\mathbf{x} = (\cos\theta, \sin\theta)^T$ and $\boldsymbol{\mu} = \kappa\boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_0 = (\cos\alpha, \sin\alpha)^T$. Then

$$f(\mathbf{x}) \propto \exp\{\mathbf{x}^T\boldsymbol{\mu}\} \text{ wrt uniform density,}$$

Note the exponent is *linear* in $\mathbf{x}$.

Figure of von Mises density



von Mises-Fisher distribution on S_p

Recall $S_p = \{\mathbf{x} \in \mathbb{R}^q : \mathbf{x}^T\mathbf{x} = 1\}$ is the set of unit vectors in $q = p + 1$ dimensions.

Let $\mathbf{x}$ denote a random direction on S_p and let $\boldsymbol{\mu} = \kappa\boldsymbol{\mu}_0$ where $\boldsymbol{\mu}_0 \in S_p$ denotes a mean direction. In Euclidean coordinates with a general mean direction, the von Mises-Fisher density becomes

$$f(\mathbf{x}) \propto \exp\{\mathbf{x}^T\boldsymbol{\mu}\} \text{ wrt uniform density,}$$

Note again the exponent is *linear* in $\mathbf{x}$.

The use of a uniform base density avoids the need for messy polar coordinates (e.g. $\sin\theta d\theta d\phi$ on S_2; $\theta =$ colatitude, $\phi =$ longitude).

Example — Bingham and FB distributions on S_p

In Euclidean coordinates with a $q \times q$ symmetric parameter matrix A, the Bingham density is defined by

$$f(\mathbf{x}) \propto \exp\{-\mathbf{x}^T A \mathbf{x}\} \text{ wrt uniform density,}$$

Note the density is symmetric ($f(\mathbf{x}) = f(-\mathbf{x})$) and the exponent is *quadratic* in $\mathbf{x}$.

Note A and $A + cI$ define the same distribution since $c\mathbf{x}^T I \mathbf{x} = c$ is constant for $\mathbf{x} \in S_p$. It is convenient to choose c so that the smallest eigenvalue of A equals 0.

We can combine the Fisher and Bingham distributions to get the Fisher-Bingham distribution

$$f(\mathbf{x}) \propto \exp\{\mathbf{x}^T \boldsymbol{\mu} - \mathbf{x}^T A \mathbf{x}\} \text{ wrt uniform density,}$$

Example — matrix Fisher distribution on $SO(s)$

The special orthogonal group of $s \times s$ rotation matrices is defined by

$$SO(s) = \left\{ X \in \mathbb{R}^{s \times s} : \det X = 1, \ X^T X = I_s \right\},$$

A natural distribution here (with parameter matrix $F(s \times s)$) is the matrix Fisher distribution,

$$f(X) \propto \left\{ \mathrm{tr}\left(F^T X \right) \right\}, \quad X \in SO(s),$$

with respect to the underlying invariant "Haar" measure. It is unimodal about a fixed rotation matrix determined by F.

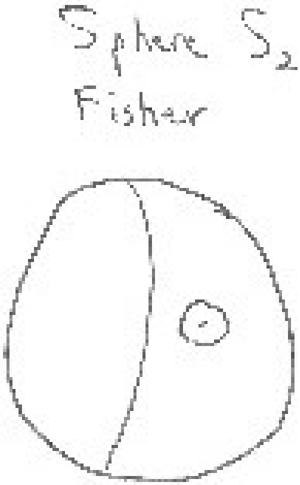
There is a magical identity in differential geometry based on quaternions, which states that the matrix Fisher distribution on $SO(3)$ can be identified with the Bingham distribution on S_3.

Applications

For each distribution, here is a typical application and its limiting distribution under high concentration.

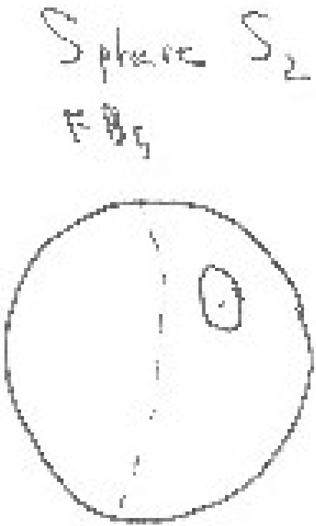
- von Mises on S_1. Vanishing angles of pigeons. Univariate normal.
- Fisher on S_2. Directions of magnetization in rocks. Isotropic bivariate normal.
- Bingham on S_2. Axes of bedding planes in rocks. General bivariate normal (on axes).
- Fisher-Bingham on S_2. Directions of magnetization in rocks. General bivariate normal (in FB5 case).
- matrix Fisher on $SO(3)$ ($=$ Bingham on S_3). Aligning objects in 3 dimensions. General trivariate normal.

Concentrated Fisher distribution



Concentrated Bingham distribution

Concentrated Fisher-Bingham distribution



Concentrated matrix Fisher distribution

Inference issues

- To some extent inference is easy because these distributions are exponential families (sometimes curved).
- However, the normalizing constant can be intractable for the more complicated distributions in higher dimensions (though saddlepoint approximations of Kume and Wood are a great help in some situations, including the Fisher-(and/or)-Bingham models considered here).
- In modern MCMC calculations in Bayesian models for which directional distributions are a building block, simulation is important.

Simulation

- Simulation methods have typically been developed on an ad hoc basis.
- Efficient acceptance-rejection methods available for the simpler distributions (especially von Mises-Fisher in any dimension and Bingham on S_2).
- Messier MCMC methods for the more complicated distributions (e.g. due to Hoff, Habeck, Kume&Walker)
- Directional models have become a key ingredient in modern Bayesian models, e.g. for protein structure simulation, alignment problems.
- Hence a need for efficient AR simulation methods. We shall focus on the Bingham distribution on S_p (including matrix Fisher on $SO(3)$).

Desirable properties of a simulation method

As the parameter matrix A varies, the Bingham distribution ranges from uniform to highly concentrated, including partially concentrated versions inbetween.

We want a method of simulation that is reasonably efficient across the whole parameter space.

Angular central Gaussian distribution

The first step is to find a good envelope for the Bingham density

$$f(\mathbf{x}) \propto \exp\{-\mathbf{x}^T A\mathbf{x}\}, \quad \mathbf{x} \in S_p.$$

We shall use the angular central Gaussian (ACG) density

$$g(\mathbf{x}) \propto (\mathbf{x}^T \Sigma \mathbf{x})^{-q/2}, \quad \mathbf{x} \in S_p$$

where $\Sigma(q \times q)$ is symmetric positive definite. Note Σ and $c\Sigma$ define the same distribution.

The ACG distribution can be easily simulated as follows:

if $\mathbf{x} \sim N_q(0, \Sigma)$, then $\mathbf{y} = \mathbf{x}/||\mathbf{x}|| \sim \text{ACG}(\Sigma)$.

Behavior of the Bingham distribution

Recall the Bingham distribution in $\mathbb{R}^3$ can be used to model data concentrated about the north (and south) pole, say, with ellipse-shaped variability about the pole. E.g., there may be more variation along the Greenwich meridian (0^o, 180^o longitude) and less in the perpendicular direction (90^o and 270^o longitude). It is analogous to the bivariate normal distribution.

The ACG distribution has the same qualitative behavior, but with longer tails. It is analogous to the bivariate Cauchy distribution.

The acceptance/rejection method of simulation — 1

- We want to simulate from a density $f(x)$.
- We can simulate from a density $g(x)$.
- We can find a bound $f(x) \leq Mg(x)$ for all x, where $M \geq 1$.
- Then the A/R method enables us to simulate from f

The acceptance/rejection method of simulation — 2

(a) Simulate $X \sim g$ independently of $W \sim \text{unif}(0,1)$.
(b) If $W < Mf(X)/g(X)$, then accept X.
(c) Otherwise go back to step (a).

The number of trials needed is geometric with mean $M \geq 1$. The efficiency is defined by $1/M$. We want M to be as close to 1 as possible.

If the normalizing constants for f and/or g are not known, then the method can still be used, though the efficiency must be computed by simulation.

The acceptance/rejection method of simulation — 3

That is, if

$$f(x) = c_f f^*(x), \quad g(x) = c_g g^*(x)$$

and

$$f^*(x) \leq M^* g^*(x)$$

where f^* and g^* are known functions, and the bound M^* is known, but the normalizing constants c_f and/or c_g are not known, then the A/R algorithm can still be used, with step (b) replaced by

(b*) If $W < M^* f^*(X)/g^*(X)$, then accept X.

However, the efficiency must be computed by simulation.

A key concave inequality

Since

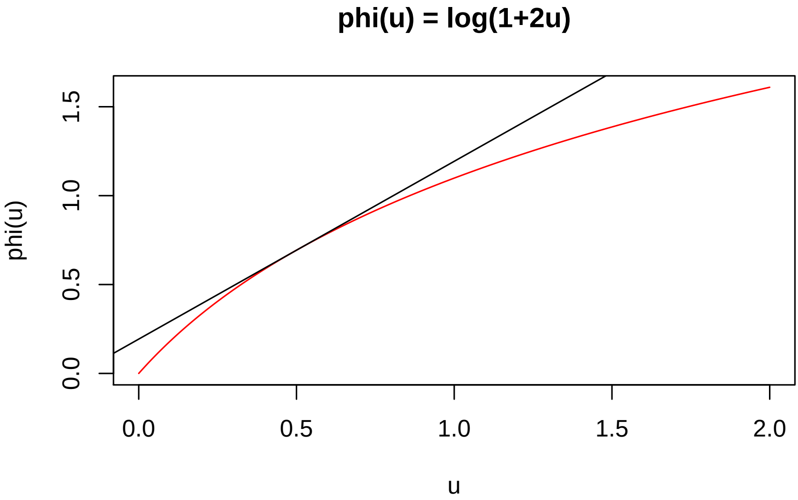
$$\phi(u) = \log(1 + 2u/b), \quad u \geq 0,$$

is a concave function ($0 < b < 2$ is a tuning parameter) and $\phi'(u) = 1/(b/2 + u)$ decreases from $2/b$ (> 1) to 0 as u ranges from 0 to ∞, we can majorize $\phi(u)$ by a linear function with slope 1,

$$\phi(u) \leq \psi(u) = a + u,$$

where a depends on b and is chosen to make the two function touch one another.

Figure showing majorization



A/R for univariate Normal < Cauchy

Changing sign and exponentiating

$$\log(1 + 2u/b) \leq a + u$$

yields

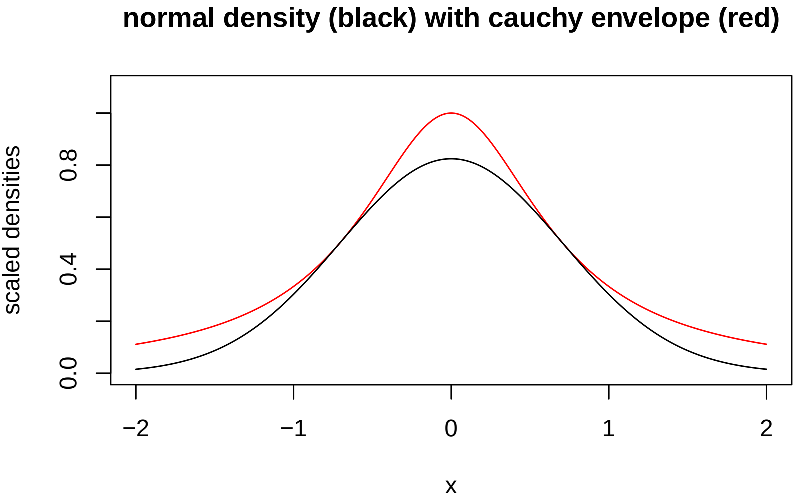
$$e^{-u} < e^{a} \frac{1}{1 + 2u/b}$$

Substituting $u = \frac{1}{2}x^2$ yields

$$\exp(-\frac{1}{2}x^2) < e^{a} \frac{1}{1 + x^2/b},$$

i.e. the normal density is bounded above by a multiple of the Cauchy. Hence we can use A/R to simulate the normal using a Cauchy envelope. (Note this is a toy example!)

Figure showing majorization



Multivariate version

With a mild change, the same argument works for the multivariate normal distribution, $N_p(\mu, \Sigma)$, dominated by the multivariate Cauchy distribution in $\mathbb{R}^p$. The optimal value of b and the efficiency can be computed analytically (and do not depend on the mean μ or the covariance matrix Σ).

Efficiency: $\mathrm{MVN}_p < \mathrm{MVC}_p$

p	1	2	3	4	5	10	50	100
eff.	66%	52%	45%	40%	36%	26%	12%	9%

The efficiency is reasonable for small p but deteriorates if p is very large.

A/R: Bingham < ACG

Represent the Bingham distribution on the unit sphere S_{q-1} in $\mathbb{R}^q$ using the version of the parameter matrix A for which the eigenvalues satisfy $\lambda_1 \geq \cdots \geq \lambda_{q-1} \geq \lambda_q = 0$. Then substituting $u = \mathbf{x}^T A \mathbf{x}$ in the modified concave inequality

$$e^{-u} < \frac{e^a}{\{1 + 2u/b\}^{q/2}}$$

yields

$$\exp\{-\mathbf{x}^T A \mathbf{x}\} \leq \frac{e^a}{\{1 + 2\mathbf{x}^T A \mathbf{x}/b\}^{q/2}}$$
$$= \frac{e^a}{\{\mathbf{x}^T \Sigma \mathbf{x}\}^{q/2}}, \quad \Sigma = I_q + 2A/b.$$

That is the Bingham density is bounded above by a multiple of the ACG density.

Efficiency: Bingham $<$ ACG

In the directional case it is still possible to find the optimal value of b. However, the efficiency of the algorithm must be computed by simulation for each choice of parameters.

It turns out that the efficiencies for the Normal $<$ Cauchy situation are *lower* bounds for the efficiencies for the Bingham $<$ ACG situation.

There is a good reason for this result. Under high concentration,

$$\text{Bingham} \approx \text{MV Normal}$$
$$\text{ACG} \approx \text{MV Cauchy.}$$

Fisher-Bingham simulation

Bingham simulation methods can also be used for the Fisher and Fisher-Bingham distributions. The key is the simple identity

$$
\begin{aligned}
0 &\leq (\alpha - \beta \boldsymbol{\mu}_0^T \mathbf{x})^2 \\
&= \alpha^2 - 2\alpha\beta \boldsymbol{\mu}_0^T \mathbf{x} + \beta^2 \mathbf{x}(\boldsymbol{\mu}_0 \boldsymbol{\mu}_0^T)\mathbf{x}
\end{aligned}
$$

for $\alpha, \beta > 0$, which implies

$$2\alpha\beta \boldsymbol{\mu}_0^T \mathbf{x} \leq \alpha^2 - \mathbf{x}^T A \mathbf{x}$$

where $A = -\beta^2 \boldsymbol{\mu}_0 \boldsymbol{\mu}_0^T$.

Optimize efficiency over the choice of α and β such that $2\alpha\beta = \kappa$.

The price is an (acceptable) loss in efficiency by a factor of about 2 (at least in the unimodal aligned case).

Summary of some manifolds

Sphere: $S_p = \{x \in \mathbb{R}^q : \; x^T x = 1\}, \; p \geq 1$ — the unit vectors in $\mathbb{R}^q, q = p + 1$

Real projective space: $\mathbb{R}P_p = S_p/\{1, -1\}$ — quotient space in which two antipodal points or "directions" $\pm x$ are identified with one another to represent the same "axis"

Special orthogonal group: $SO(s)$ is the space of $s \times s$ rotation matrices, $SO(s) = \{X \in \mathbb{R}^{s \times s} : X^T X = I_s, \; |X| = 1\}$.

Stiefel manifold: $V_{r,s} = \{X_1 \in \mathbb{R}^{s \times r} : X_1^T X_1 = I_r\}$ — space of $s \times r$ column orthonormal matrices X_1 $(1 \leq r \leq s)$

Grassmann manifold: $\mathcal{G}_{r,s}$ — the set of all r-dimensional subspaces of $\mathbb{R}^s$

Summary of recommended methods

Distribution	Space	Simulation method
von Mises-Fisher	S_p	Wood 1987
Bingham	S_p or $\mathbb{R}P_p$	BACG
complex Bingham	S_{2p+1} or $\mathbb{C}P_p$	Kent et al. 2004
complex Bingham quartic	S_{2p+1} or $\mathbb{C}P_p$	??
aligned Fisher-Bingham	S_p	BACG-based
general Fisher-Bingham	S_p	" or MCMC
matrix Fisher	$\mathcal{V}_{r,s}$ or $SO(s),\ s \geq 4$	MCMC
matrix Fisher	$SO(3)$	BACG
balanced matrix Bingham	$\mathcal{V}_{r,s}$ or $\mathcal{G}_{r,s}$	BACG
general matrix Bingham	$\mathcal{V}_{r,s}$	MCMC
matrix Fisher-Bingham	$\mathcal{V}_{r,s}$	MCMC
similar	product spaces	MCMC

Conclusions

- The ACG distribution is a very effective envelope for the Bingham distribution, almost regardless of dimension.
- This automatically gives a simple method to simulate the matrix Fisher distribution on $SO(3)$.
- Once it is possible to simulate from the Bingham distribution, it is straightforward to give reasonably efficient methods to simulate from the Fisher-Bingham distribution.
- However, the ACG approach is not a panacea! It does not (yet?) work for similar types of distribution on more complicated manifolds (e.g. Stiefel, Grassmann, product, ...).
- Since directional distributions are building blocks in more sophisticated statistical models (usually analyzed by MCMC), it is important to have efficient simulation methods.