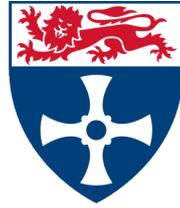


NEWCASTLE UNIVERSITY



MAS8391 MMATHSTAT PROJECT

Bayesian Inference for Directional Statistics and Applications

Author:
Thomas Brereton

Supervisor:
Prof. Richard Boys

April 30, 2015

Contents

1	An Introduction to Directional Statistics	1
1.1	Overview	1
1.2	Coordinate systems	1
1.3	Visualization and programming	2
1.4	Common distributions	4
1.5	The von Mises-Fisher distributions	4
2	Circular Data and the von Mises Distribution	6
2.1	The von Mises distribution	6
2.2	Simulation	6
2.3	Von Mises distributions with different parameter values	8
2.4	Assessing von Misesness	9
2.5	The likelihood function and prior distributions	10
2.6	Inference for κ when $\mu = \mu_0$	11
2.7	Inference for μ when $\kappa = \kappa_0$	14
2.8	Bayesian inference on two parameters	17
2.9	Vanishing directions of homing pigeons	21
3	Spherical Data and the von Mises-Fisher Distribution	25
3.1	Simulation	25
3.2	Visual representation of the von Mises-Fisher distribution	26
3.3	Discussion of priors	27
3.4	The posterior distribution	28
3.5	Metropolis-Hastings algorithm	29
3.6	Cosmic rays	34
4	Conclusions and Further Work	40

Abstract

This report has three main aims. These are:

- to provide an introduction to directional data and its uses;
- to formulate and perform Bayesian analysis on the von Mises-Fisher distributions in both two dimensions and its generalised form in p dimensions;
- to apply the subsequent results to data.

Common problems with directional data, such as identifiability issues and effective visualization, are discussed. The von Mises-Fisher distributions on the p -dimensional sphere are presented and inference for the distribution's parameters is presented via a Bayesian approach.

Markov Chain Monte Carlo methods are applied, and a Metropolis-Hastings within Gibbs algorithm in p dimensions is proposed. The general method is illustrated using two real datasets: one on the homing directions of pigeons in two dimensions, and the other on the arrival locations of low-level muons in three dimensions.

Chapter 1

An Introduction to Directional Statistics

1.1 Overview

Directional data arise when measurements are made regarding a direction, rather than a magnitude. For data relating to the circle, this predominantly arises through measurements on the compass or the clock; natural occurrences include wind directions, animal migrations and waiting times. Meanwhile, data on the sphere can be used to measure event occurrences on the surfaces of celestial objects, such as the formation of a hurricane on Earth or solar flare locations on the sun, as well as smaller objects such as measurements of imperfections on the surface of an eyeball.

Although the study of directional data encompasses many geometries, we will focus solely on analysis on circles, spheres and hyperspheres (collectively named p -dimensional spheres). Additionally, we will only consider the case where the object has a radius of 1. Only data recorded as measurements on the surface of the shape will be considered, i.e all data points have unit length from the origin; thus, the magnitude of results will not be discussed.

1.2 Coordinate systems

Measurements of direction can be analysed either using a Cartesian or an angular system. To define a Cartesian system in p dimensions, we require a vector of length p , with elements denoted $\mathbf{x} = (x_1, x_2, x_3, \dots, x_p)^T$, where

$$-1 \leq x_i \leq 1, \quad \forall i \in 1, 2, \dots, p \quad \text{and} \quad \|\mathbf{x}\| = 1,$$

and $\|\mathbf{x}\|$ is the Euclidean length of the vector \mathbf{x} . These conditions ensure that \mathbf{x} lies on the p -dimensional unit sphere, denoted \mathcal{S}_p .

On the other hand, angular notation uses a parameterization that requires $p - 1$ values, say an angle x for circular data and angles (α, β) for spherical data. Note that, as working with radii of unit length, we can describe the position of the data points as

$$(\cos x, \sin x), \quad \text{or} \quad (\cos \alpha, \sin \alpha \cos \beta, \sin \alpha \sin \beta)$$

for two and three dimensions respectively. These vectors satisfy the same properties

as described for the Cartesian system, as can be seen through Euclidean normalization. However, there still remains an identifiability problem.

Consider circular data, which requires a central direction from which to measure x , and any range of length 2π to specify the direction. We can remove the identifiability problem by using $x \in [-\pi, \pi)$, although another sensible range is $[0, 2\pi)$.

This highlights a fundamental difference between directional statistics and statistics on the real line. Due to the periodic nature of \sin and \cos , a measurement recorded as $\pi + 1$ is equivalent to a measurement of $-\pi + 1$, i.e. a shift of 2π results in the same observation. Consider the simple case of a dataset consisting of these two observations, before range adjustment is applied. The classical interpretation of the mean of the observations is given as 1; however, the mean *direction* is actually $-\pi + 1$, easily seen once the first data point is transformed to be in the correct range. This problem arises in several instances through the Bayesian analysis in Chapters 2 and 3, and will be discussed further then.

Spherical coordinate systems have similar issues, however complexity is increased due to the presence of two angles, both which must be carefully defined. We will take $\alpha \in [0, \pi)$ as our azimuthal (or latitude) angle, and $\beta \in [0, 2\pi)$ as our polar (or longitude) angle. A visualization of this coordinate system is presented in Figure 1.3.2.

In order to explore the merits of both systems, we will perform Bayesian analysis in angular coordinates on the circle in Chapter 2 and Cartesian coordinates on the sphere in Chapter 3.

1.3 Visualization and programming

One of the major challenges involved in directional statistics is the graphical presentation of results. Fortunately, the R package `circular` gives a number of ways to present circular data, as well as a variety of other functions relating to relevant distributions. For spherical data, the R package `movMF` simulates von Mises-Fisher distributions, and so provides a method for checking simulations in three dimensions. Plotting data on the sphere is performed through the package `rgl`, the 3D visualization device system within R.

Figure 1.3.1 shows an example data set of the orientation of pebbles, provided by the package `circular`. Although originally specified in geographics (i.e. as an angle measured in degrees from North), the data have been transformed in order to be suitable for our range of $[-\pi, \pi)$. A stacked histogram, as well as a density estimate, is provided. Note that measurements of North are given at the point described by $-\pi/\pi$, and measurements are reflected to transform from clockwise to anticlockwise.

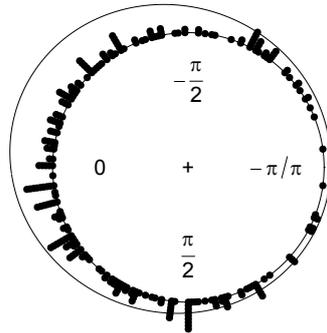


Figure 1.3.1: Orientation of pebbles found in Fox river, Illinois

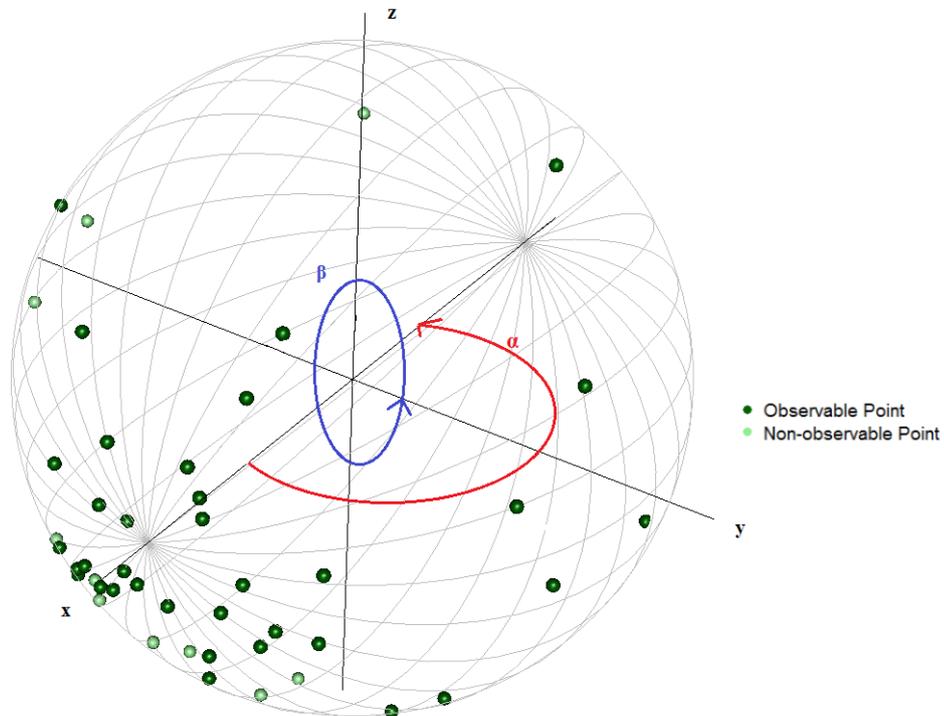


Figure 1.3.2: Visualization of spherical conventions

Visualization of spherical data is understandingly more challenging. The main problem is negating the confusion of which hemisphere a particular point lies in. The most convenient solution is to plot ambiguous points from the same sample in different colours. This is best shown through an example, given in Figure 1.3.2. Points given in light green are on the other side of the sphere, i.e. if the sphere was opaque, these points would not

be visible. Figure 1.3.2 also demonstrates the effect of altering α and β when data is measured in angular system.

In order to provide a more natural, three dimensional interpretation of data on the sphere, shadow lines have been added that explore α values for set β values. These are included solely to make graphical interpretation easier, and play no part in the description of the data or distributions.

1.4 Common distributions

Circular distributions take one of two forms: wrapped distributions and circular-only distributions. Wrapped distributions are formed by taking a distribution on the line and wrapping it around the circumference of the circle of unit radius, i.e. $x = y \bmod 2\pi$, where y is the realization on the real line and x is the corresponding angular value on the circular distribution. Commonly found wrapped distributions, described at length by Mardia and Jupp (2000) and Best and Fisher (1979), include the wrapped Normal, the wrapped Poisson and the wrapped Cauchy distributions. Simulation for such distributions is straightforward and Bayesian analysis in such situations is best performed before wrapping is undertaken.

Analysis for circular-only distributions is often more challenging. Distribution functions are often given in terms of infinite series, due to their periodic nature, and so numeric evaluation and simulation is more difficult. However, most circular-only distributions have a comparable distribution on the real line; for example, the circular uniform distribution provides a good method for analysis of points with no concentration or skew on the circle.

Wrapped distributions do exist on the sphere, although they are not commonly studied. This is due to the requirement of simulation on the two dimensional plane, which is performed in specialist situations when further wrapping to the sphere would likely be counterproductive and so are not discussed here. Spherical-only distributions are our main focus, and are used extensively in the areas described in Section 1.1.

One family of distributions used to describe data on spheres, on all p -dimensions, dominates the mathematical literature; these are the von Mises-Fisher distributions, and Bayesian inference for data assuming these distributions provide the basis for this report.

1.5 The von Mises-Fisher distributions

Mardia and Jupp (2000) describe the von Mises distribution as “perhaps the most useful distributions on the circle”. First published by Richard von Mises in 1918 with respect to the circle, the continuous distribution was introduced in order to study the deviations of measured atomic weights from integral values, before the discovery of isotopes in the 1920s. Since then it has been used extensively in circular statistical analysis, and it can be considered as the circular analogous of the Normal distribution on the real line.

Contributions by R.A. Fisher in the 1950s and other developments lead to the von Mises-Fisher family of distributions, used to describe points in all p dimensions.

Consider a vector \mathbf{x} of unit length, taking values on a p -dimensional hypersphere \mathcal{S}_p with unit radius and center at the origin; then \mathbf{x} has a von Mises-Fisher distribution if it follows the probability density function

$$f(\mathbf{x}|\boldsymbol{\mu}, \kappa) = \left(\frac{\kappa}{2}\right)^{p/2-1} \frac{1}{\Gamma(p/2)I_{p/2-1}(\kappa)} \exp\{\kappa\boldsymbol{\mu}^T\mathbf{x}\}, \quad \mathbf{x} \in \mathcal{S}_p, \quad (1.5.1)$$

where $\kappa \geq 0$, $\boldsymbol{\mu} \in \mathcal{S}_p$ and I_ν denotes the modified Bessel function of the first kind and order ν . Here

- $\boldsymbol{\mu}$ is a measure of location;
- κ is a measure of concentration (points become more clustered around $\boldsymbol{\mu}$ as $\kappa \rightarrow \infty$).

It can be shown that the mean and the median is $\boldsymbol{\mu}$, comparable with the Normal distribution. Mardia and Jupp (2000) present a method for calculating the variance, given as $I_{p/2}(\kappa)/I_{p/2-1}(\kappa)$. Additionally, as $\kappa \rightarrow 0$, the distribution tends to the uniform distribution on \mathcal{S}_p . Provided $\kappa > 0$, it can also be shown that $\boldsymbol{\mu}^T\mathbf{x}$ has a maximum (mode) at $\boldsymbol{\mu}$ and a minimum (antimode) at $-\boldsymbol{\mu}$.

Although the analysis provided in Chapter 3 is applicable in p -dimensions, we will be concerned with $p = 2$ and $p = 3$, i.e. data on the unit circle and sphere. This is because, for $p > 3$,

- it is challenging to present findings in a meaningful manner
- there exist very few data sets

The von Mises-Fisher distributions are denoted $vMF_p(\boldsymbol{\mu}, \kappa)$ for p dimensions. Note that $vMF_2(\boldsymbol{\mu}, \kappa)$ reduces to the von Mises distribution on the circle, discussed in Chapter 2, whilst $vMF_3(\boldsymbol{\mu}, \kappa)$ reduces to the Fisher distribution on the sphere, discussed in Chapter 3.

Chapter 2

Circular Data and the von Mises Distribution

2.1 The von Mises distribution

Consider the von Mises-Fisher distributions given in Equation (1.5.1). In $p = 2$ dimensions the density is

$$f(x|\mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(x - \mu)\}, \quad x \in [-\pi, \pi) \quad (2.1.1)$$

where $\kappa \geq 0$ and $\mu \in [-\pi, \pi)$. The modified Bessel function of order 0, denoted I_0 , is defined as

$$I_0(\kappa) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\{\kappa \cos u\} du. \quad (2.1.2)$$

It is clear that this density integrates to 1. Also, by parameterising $\boldsymbol{\mu}$ and \boldsymbol{x} in terms of angles (discussed in Section 1.2), the exponent of Equation (1.5.1) can be written as

$$\begin{aligned} \kappa \boldsymbol{\mu}^T \boldsymbol{x} &= \kappa(\cos \mu, \sin \mu)^T (\cos x, \sin x) \\ &= \kappa(\cos \mu \cos x + \sin \mu \sin x) \\ &= \kappa \cos(x - \mu) \end{aligned} \quad (2.1.3)$$

using a cosine trigonometric identity.

This is the von Mises distribution, the name given to the von Mises-Fisher distribution on $p = 2$ dimensions, i.e. the circle. The notation for this distribution, vMF_2 , is commonly abbreviated to vM .

2.2 Simulation

Simulation of the von Mises distribution is discussed by Best and Fisher (1979). They begin by discussing a number of difficulties with the efficient generation of pseudo-random observations from the von Mises distribution, including that the cumulative distribution function does not have a closed form (and so inverse transform methods are impractical). Also there is no way of writing a general von Mises density in terms of a standardised von

Mises with say unit concentration. Therefore simulation cannot be focused on the case when $\kappa = 1$ and then transformed; it must be performed for every κ individually.

After considering previous (and generally quite unsuccessful) attempts of simulation, Best and Fisher examined algorithms based on envelope-rejection methods, and decided on a method using the Wrapped Cauchy distribution as an envelope.

The full algorithm for simulation of one x -value from the $vM(\mu, \kappa)$ distribution is given as follows:

1. Define

$$\tau = 1 + \sqrt{1 + 4\kappa^2}, \quad \rho = \frac{\tau - \sqrt{2\tau}}{2\kappa}, \quad r = \frac{1 + \rho^2}{2\rho}.$$

2. Generate u_1, u_2 and u_3 as independent random observations from a uniform $U(0, 1)$ distribution.

3. Set

$$z = \cos(\pi u_1), \quad f = \frac{1 + rz}{r + z}, \quad c = \kappa(r - f).$$

4. If $c(2 - c) > u_2$, go to step 6.
5. If $\log(c/u_2) + 1 - c < 0$, return to step 3.

6. Set

$$x = \text{sign}(u_3 - 0.5) \cos^{-1}(f)$$

where the *sign* function returns the values $-1, 0$ or 1 depending on if the argument is negative, zero or positive respectively.

We illustrate the output of this algorithm by simulating 100 data points from a $vM(0, 1)$ distribution. The output is summarised as points in Figure 2.2.1, and is compared to the $vM(0, 1)$ density. The density describes the distribution of the simulated points well, with a peak at $x = 0$ and decreasing concentration as the angle moves away from this point.

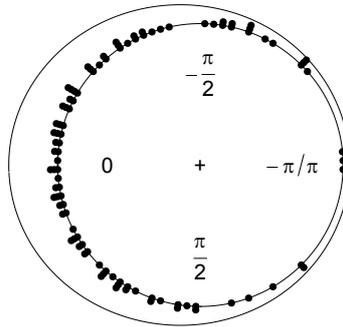


Figure 2.2.1: 100 simulations from a $vM(0, 1)$ density

2.3 Von Mises distributions with different parameter values

Figure 2.3.1 shows a comparison of three von Mises densities with different parameter values.

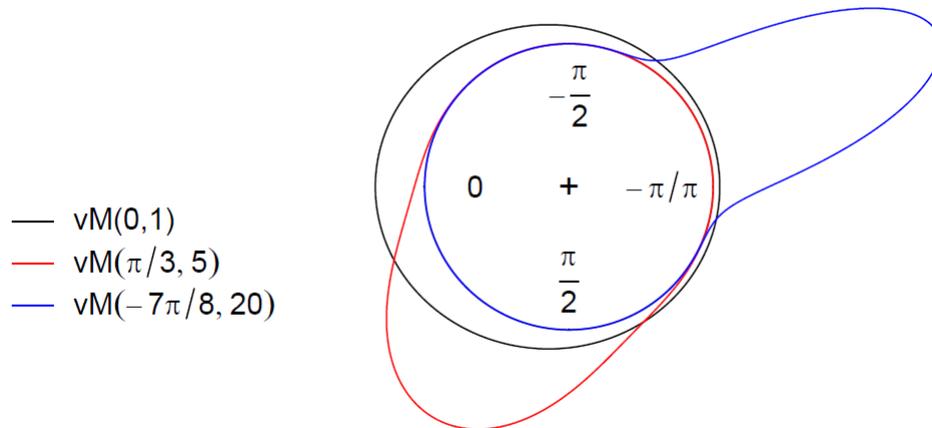


Figure 2.3.1: Comparison of three von Mises densities

We can see that each distribution is centered on their respective μ value, and that the distributions clearly become more concentrated around μ as the von Mises concentration

parameter κ increases.

The symmetric nature of the von Mises distribution is made immediately apparent in Figure 2.3.1. Additionally, the distribution with the highest concentration parameter (blue) graphically resembles the more familiar Normal distribution, supporting the remarks made in Section 1.5.

2.4 Assessing von Misesness

Much of the work within this chapter assumes that the data are consistent with an underlying von Mises distribution, and so it is beneficial to explore ways in which von Misesness can be assessed.

One manner to assess von Misesness is graphically. This can be performed via a von Mises probability-probability plot, referred to within the R package `circular` as a `pp.plot`. This calculates the frequentist MLEs for μ and κ from the data, and plots the corresponding empirical distribution function, as well as the data points, for comparison. Note that this is similar to the method used to create a `qqplot` when testing for Normality with real-line data.

A `pp.plot`, for the data simulated in Section 2.3, is given in Figure 2.4.1.

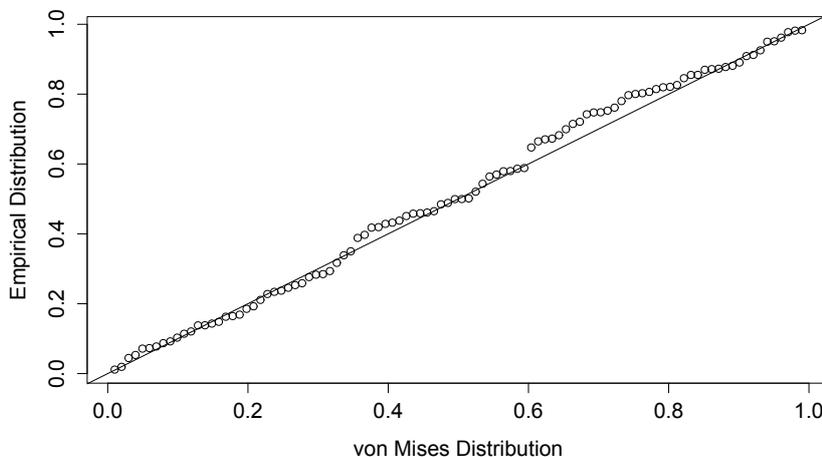


Figure 2.4.1: Von Mises probability-probability plot for 100 simulated $vM(0, 1)$ points

As we expect, the points fit well to the line, suggesting that the simulation has successfully generated values from a $vM(0, 1)$ distribution.

A second way to assess von Misesness is to perform a Watson's goodness-of-fit test. This is also given as a function within R that returns a test statistic and a corresponding p -value.

Performing the test with the simulated data above gives $p > 0.1$, again suggesting that there is little evidence against the von Misesness of the simulated data.

2.5 The likelihood function and prior distributions

The likelihood function for a random sample of n points x_1, x_2, \dots, x_n from a $vM(\mu, \kappa)$ distribution is

$$\begin{aligned} f(\mathbf{x}|\mu, \kappa) &= \prod_{i=1}^n f_X(x_i|\mu, \kappa) = \prod_{i=1}^n \{[2\pi I_0(\kappa)]^{-1} \exp\{\kappa \cos(x_i - \mu)\}\} \\ &= [2\pi I_0(\kappa)]^{-n} \exp\left\{\kappa \sum_{i=1}^n \cos(x_i - \mu)\right\} \\ &\propto [I_0(\kappa)]^{-n} \exp\left\{\kappa \sum_{i=1}^n \cos(x_i - \mu)\right\} \end{aligned} \quad (2.5.1)$$

where $\mu \in [-\pi, \pi)$ and $\kappa \geq 0$.

Before conducting Bayesian analysis, we must first give consideration to a suitable prior distribution for the parameters of interest μ and κ . In this chapter we assume that μ and κ are independent *a priori*.

The concentration parameter κ is defined over the range $0 \leq \kappa < \infty$. Therefore, a suitable prior might be a Gamma distribution, with

$$\kappa \sim Ga(a, b)$$

and prior density

$$\pi(\kappa) = \frac{\kappa^{a-1} \exp\{-b\kappa\}}{b^a \Gamma(a)}, \quad \kappa \geq 0 \quad (2.5.2)$$

where $a, b > 0$.

We now turn our attention to μ . As we require μ to be a value on the unit circle, the most plausible prior distribution would be to use a von-Mises distribution itself. Thus, we might use

$$\mu \sim vM(m, c)$$

with prior density

$$\pi(\mu) = \frac{1}{2\pi I_0(c)} \exp\{c \cos(\mu - m)\}, \quad \mu \in [-\pi, \pi) \quad (2.5.3)$$

where $c \geq 0$ and $m \in [-\pi, \pi)$.

Mardia and El-Atoum (1976) provide an alternative prior for μ , given as a uniform distribution on the circle. However, using a von Mises distribution instead allows for varying certainty regarding the prior mean direction, and using a low concentration parameter allows for the prior to be similar to a uniform distribution if desired.

2.6 Inference for κ when $\mu = \mu_0$

We first consider a single parameter problem in which inference is performed on κ when $\mu = \mu_0$ is a known quantity.

For simplicity we take $\mu_0 = 0$, as this reduces the complexity of the algebra and rotation to other values of μ_0 after analysis is trivial. The likelihood function here is

$$f(\mathbf{x}|\kappa, \mu = 0) \propto [I_0(\kappa)]^{-n} \exp \left\{ \kappa \sum_{i=1}^n \cos x_i \right\}, \quad (2.6.1)$$

where $\kappa \geq 0$.

Applying Bayes' Theorem gives the posterior density of κ as, for $\kappa \geq 0$,

$$\begin{aligned} \pi(\kappa|\mathbf{x}, \mu = 0) &\propto f(\mathbf{x}|\kappa, \mu = 0)\pi(\kappa) \\ &\propto [I_0(\kappa)]^{-n} \exp \left\{ \kappa \sum_{i=1}^n \cos x_i \right\} \kappa^{a-1} \exp \{-b\kappa\} \\ &\propto [I_0(\kappa)]^{-n} \kappa^{a-1} \exp \left\{ \kappa \left[\sum_{i=1}^n (\cos x_i) - b \right] \right\}. \end{aligned} \quad (2.6.2)$$

Unfortunately, this is not the density of a standard distribution. Thus, we will develop Markov Chain Monte Carlo (MCMC) methods in order to draw realizations from the posterior density for κ .

The algorithm used is an adaptation of the commonplace Metropolis-Hastings scheme, as discussed by Gamerman and Lopes (2006). A new value of κ (denoted κ^*) is simulated from a proposal density $q(\kappa^*|\kappa)$ that has the same domain as the posterior. The parameters of the proposal distribution are functions of the current value of κ and a tuning parameter ν , which can be altered to improve the mixing of the chain if necessary. As $\kappa \geq 0$, it seems sensible to use a Log-Normal distribution as a proposal distribution, that is

$$\kappa^* \sim LN(\log \kappa, \nu^2).$$

The full algorithm is:

1. Initialise the chain at some starting value $\kappa = \kappa^{(0)}$ and set $j = 1$. Note that for the majority of cases, the initial values are taken to be the prior mean, i.e. $\kappa^{(0)} = a/b$.

2. Simulate a proposed value of κ , denoted κ^* , where

$$\kappa^* \sim LN(\log \kappa^{(j-1)}, \nu^2).$$

3. Evaluate the acceptance probability $\alpha(\kappa^*|\kappa^{(j-1)}) = \min(1, A)$, where

$$A = \frac{\pi(\kappa^*|\mathbf{x}, \mu = 0)}{\pi(\kappa^{(j-1)}|\mathbf{x}, \mu = 0)} \times \frac{\kappa^*}{\kappa^{(j-1)}}.$$

(the full conditional for κ is given in Equation (2.8.2)). Note that the Jacobian of the ratio of the log-Normal proposal distribution is given as a factor that is the ratio of the proposed value to the current value.

4. Set $\kappa^{(j)} = \kappa^*$ with probability α ; otherwise set $\kappa^{(j)} = \kappa^{(j-1)}$.

5. Set $j := j + 1$ and return to step 2.

For the purposes of testing the algorithm, we begin by specifying a prior distribution as $\kappa \sim Ga(5, 1)$, and simulate 100 data points from a $vM(0, 5)$ distribution. Thus, both our prior and our data suggest that the posterior for κ should be centered around 5. Initially, ν will be taken as 1.

We run the algorithm for 1000 iterations with initial $\kappa_0 = 5$. The corresponding diagnostic plots are given in Figure 2.6.1, and are used to assess the mixing of the posterior density.

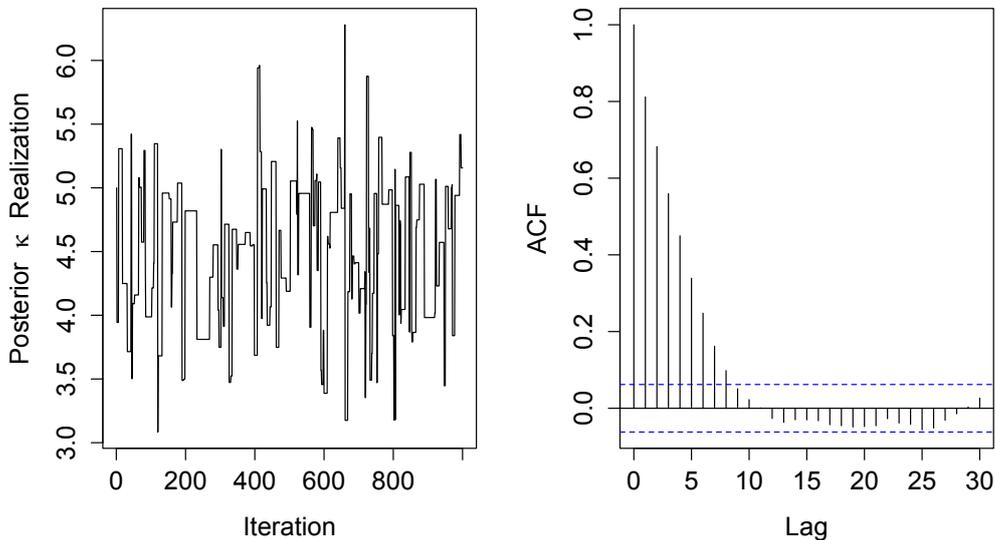


Figure 2.6.1: Trace (left) and autocorrelation (right) plots of posterior κ

The traceplot suggests that the Markov chain has reached stationarity and appears to be hovering around a value of $\kappa = 4.5$. However, it appears that the chain is behaving in a “sticky” manner, i.e. the proposed value κ^* is unlikely to differ from the previous iteration. This suggests that the proposal distribution is simulating values too far from the current κ , and thus the tuning parameter ν should be lowered.

The autocorrelation plot in Figure 2.6.1 suggests that there is significant autocorrelation up to lag 10. This suggests that taking every 10th realization will probably lead to an (almost) un-autocorrelated sample, i.e thin by 10.

After further testing with algorithmic parameters, we re-ran for 10,000 iterations with an improved proposal distribution tuning parameter ν of 0.1. Unfortunately altering ν leads to higher levels of autocorrelation, and so thinning by 50 is now required in order to have a negligible value for the autocorrelation at lag 1 and produce (almost) un-autocorrelated realizations from the posterior.

When using a Metropolis-Hastings scheme, if the initial value of the chain is unlikely in the posterior, there will be a number of realizations recorded before the chain reaches stationarity. These realizations are often discarded as they can distort belief regarding the posterior distribution, especially with few iterations. The discarding of such realizations is known as “burn-in”, and can consist of 100 realizations or so. Nothing of interest arises with respect to such values for the remainder of the report, and so this method will be performed automatically without further mention.

Updated diagnostic plots are provided in Figure 2.6.2.

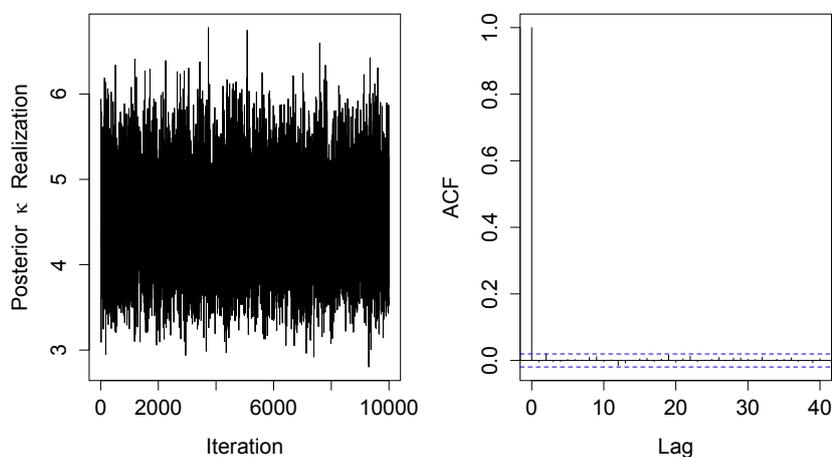


Figure 2.6.2: Diagnostic plots of posterior κ , thinned

The Figure shows very good mixing in κ with very low autocorrelation.

Satisfied with our posterior realizations, kernel density estimation in R is used to plot the posterior probability density for κ , presented in Figure 2.6.3.

The prior is included for comparison, as well as equitailed 95% confidence intervals (CI) for both distributions. These summary statistics, *a priori* and *a posteriori* respectively, are

$$E(\kappa) = 5, \quad 95\%CI = (1.623, 10.242)$$

and

$$E(\kappa) = 4.49, \quad 95\%CI = (3.467, 5.653).$$

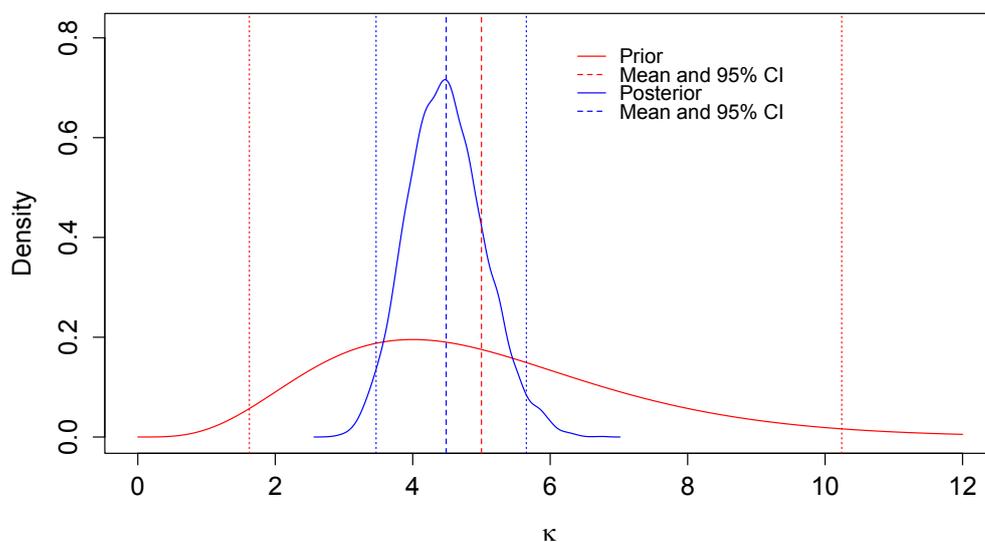


Figure 2.6.3: Prior and posterior κ densities, with means and 95% CIs

The density for the posterior is more concentrated than the prior, suggesting the data has been informative. The summary statistics support that certainty about κ has increased in the posterior, noted by a 74% reduction in the width of the 95% CI.

2.7 Inference for μ when $\kappa = \kappa_0$

As with the previous section, we again monitor one parameter in the posterior; however, the parameter of interest is now μ , and so we set $\kappa = \kappa_0$, a known value.

The full conditional distribution for μ with fixed κ_0 is

$$\begin{aligned} \pi(\mu|\mathbf{x}, \kappa = \kappa_0) &\propto f(\mathbf{x}|\mu, \kappa = \kappa_0)\pi(\mu) \\ &\propto \exp\left\{\kappa_0 \sum_{i=1}^n \cos(x_i - \mu)\right\} \exp\{c \cos(\mu - m)\} \\ &\propto \exp\left\{\kappa_0 \sum_{i=1}^n \cos(x_i - \mu) + c \cos(\mu - m)\right\}, \quad \mu \in [-\pi, \pi]. \end{aligned} \quad (2.7.1)$$

Mardia and El-Atoum (1976) claim that this posterior density takes the form of a von Mises distribution; however, this may not be immediately apparent, and so will be shown here. Assume that the posterior density is a $vM(M, K)$ distribution, and this takes the form

$$\pi(\mu|\mathbf{x}, \kappa = \kappa_0) \propto \exp\{K \cos(x - M)\}$$

up to proportionality. Thus, the problem reduces to finding K, M such that

$$\kappa_0 \sum_{i=1}^n \cos(x_i - \mu) + c \cos(\mu - m) = K \cos(\mu - M) \quad (2.7.2)$$

where K, M satisfy the conditions of a von Mises distribution, i.e. $K \geq 0$ and $-\pi \leq M < \pi$.

We begin by considering the trigonometric identity

$$\cos(a - b) = \cos a \cos b + \sin a \sin b$$

and introduce notation such that

$$d_c = \sum_{i=1}^n \cos x_i, \quad d_s = \sum_{i=1}^n \sin x_i.$$

Thus, the left hand side of Equation (2.7.2) becomes

$$\begin{aligned} &\kappa_0 \sum_{i=1}^n \cos(x_i - \mu) + c \cos(\mu - m) \\ &= \kappa_0 [d_c \cos \mu + d_s \sin \mu] + c [\cos \mu \cos m + \sin \mu \sin m] \\ &= [\kappa_0 d_c + c \cos m] \cos \mu + [\kappa_0 d_s + c \sin m] \sin \mu. \end{aligned} \quad (2.7.3)$$

Expanding the right hand side of Equation (2.7.3) and comparing the coefficients of $\cos \mu$ and $\sin \mu$, we recover a pair simultaneous equations in K and M , namely

$$K \cos M = \kappa_0 d_c + c \cos m, \quad K \sin M = \kappa_0 d_s + c \sin m.$$

By squaring both equations and noting that $\cos^2 M + \sin^2 M = 1$ for all allowable M gives

$$K = \sqrt{(\kappa_0 d_c + c \cos m)^2 + (\kappa_0 d_s + c \sin m)^2}$$

and so we obtain

$$M = \arctan \left(\frac{\kappa_0 d_s + c \sin m}{\kappa_0 d_c + c \cos m} \right).$$

Thus, our posterior distribution for μ is given as

$$\pi(\mu | \mathbf{x}, \kappa = \kappa_0) = \frac{1}{2\pi K} \exp \{K \cos(\mu - M)\}, \quad \mu \in [-\pi, \pi) \quad (2.7.4)$$

where $M \in [-\pi, \pi)$ and $K \geq 0$.

Consider the case with 30 data points simulated from a $vM(\pi/3, \kappa = 3)$ distribution and a prior distribution $\mu \sim vM(\pi/3, 1)$. After simulation and calculation, the posterior parameters are

$$M = 1.075 \quad \text{and} \quad K = 79.44,$$

and so the posterior distribution is

$$\mu | \kappa = 3, \mathbf{x} \sim vM(1.075, 79.44). \quad (2.7.5)$$

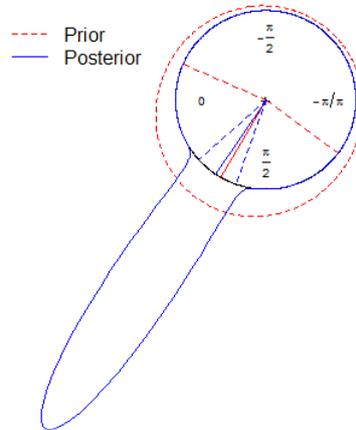


Figure 2.7.1: Prior and posterior μ with 95% CIs

Summary statistics for the prior and posterior density of μ are

$$E(\mu) = 1.047, \quad 95\% \text{CI} = (-0.411, 2.506)$$

and

$$E(\mu|\kappa = 3, \mathbf{x}) = 1.075, \quad 95\%CI = (0.836, 1.314)$$

respectively.

An 84% reduction in the width of the confidence interval suggests that the data has been highly informative in the posterior. The mean of the data is 1.076, and so our posterior mean is significantly closer to the mean of the data than the prior mean.

Figure 2.7.1 shows that efficiently plotting von Mises distributions with such a large concentration parameters can be problematic, simply due to the magnitudinal relationship between the distribution and the unit circle.

2.8 Bayesian inference on two parameters

The joint posterior density for (μ, κ) is

$$\begin{aligned} \pi(\mu, \kappa|\mathbf{x}) &\propto \pi(\mu, \kappa)f(\mathbf{x}|\mu, \kappa) \\ &\propto \pi(\mu)\pi(\kappa)f(\mathbf{x}|\mu, \kappa) \end{aligned}$$

as $\mu \perp \kappa$ a priori. Therefore, for $\kappa \geq 0$ and $\mu \in [-\pi, \pi)$,

$$\begin{aligned} \pi(\mu, \kappa|\mathbf{x}) &\propto e^{c \cos(\mu-m)} \kappa^{a-1} e^{-b\kappa} I_0(\kappa)^{-n} \exp \left\{ \kappa \sum_{i=1}^n \cos(x_i - \mu) \right\} \\ &\propto I_0(\kappa)^{-n} \kappa^{a-1} \exp \left\{ \kappa \left[\sum_{i=1}^n \cos(x_i - \mu) - b \right] + c \cos(\mu - m) \right\}. \end{aligned} \tag{2.8.1}$$

This distribution is non-standard and analysis is not conjugate. Therefore, in order to generate realisations from the distribution, we will employ Markov chain Monte Carlo (MCMC) methods.

The full conditional distributions for μ and κ are

$$\pi(\mu|\mathbf{x}, \kappa) \propto \exp \left\{ \kappa \left[\sum_{i=1}^n \cos(x_i - \mu) \right] + c \cos(\mu - m) \right\} \tag{2.8.2}$$

and

$$\pi(\kappa|\mathbf{x}, \mu) \propto I_0(\kappa)^{-n} \kappa^{a-1} \exp \left\{ \kappa \left[\sum_{i=1}^n \cos(x_i - \mu) - b \right] \right\} \tag{2.8.3}$$

respectively. As expected, Equation (2.8.2) takes the form of Equation (2.6.2), with unknown μ , and equation (2.8.3) takes the form of equation (2.7.1) with unknown κ .

Thus, we can perform a Metropolis within Gibbs scheme, taking a Gibbs step for μ and using a Metropolis-Hastings step for κ .

The full algorithm is:

1. Initialise the chain at some starting values ($\mu = \mu^{(0)}, \kappa = \kappa^{(0)}$) and set $j = 1$. Note that for the majority of cases, the initial values are taken to be the prior means, i.e. $\mu^{(0)} = m, \kappa^{(0)} = a/b$.

2. Simulate a value of $\mu^{(j)}$

$$\mu^{(j)} \sim vM(M, K)$$

where

$$M = \arctan\left(\frac{\kappa^{(j-1)}d_s + c \sin m}{\kappa^{(j-1)}d_c + c \cos m}\right), \quad K = \sqrt{(\kappa^{(j-1)}d_c + c \cos m)^2 + (\kappa^{(j-1)}d_s + c \sin m)^2}$$

3. Simulate a proposed value of κ , denoted κ^* , where

$$\kappa^* \sim LN(\log \kappa^{(j-1)}, \nu)$$

and ν denotes the tuning parameter for κ , fixed for each iteration of the algorithm but can be altered by the user when assessing the mixing of the chain.

4. Evaluate the acceptance probability $\alpha(\kappa^*|\kappa^{(j-1)}) = \min(1, A)$, where

$$A = \frac{\pi(\kappa^*|\mathbf{x}, \mu^{(j)})}{\pi(\kappa^{(j-1)}|\mathbf{x}, \mu^{(j)})} \times \frac{\kappa^*}{\kappa^{(j-1)}}$$

(the full conditional for κ is given in Equation (2.8.2)). Note that the Jacobian of the ratio of the log-Normal proposal distribution is given as a factor that is the ratio of the proposed value to the current value.

5. Set $\kappa^{(j)} = \kappa^*$ with probability α ; otherwise set $\kappa^{(j)} = \kappa^{(j-1)}$.

6. Set $j := j + 1$ and return to step 2.

This algorithm produces realizations from the joint posterior for μ and κ .

Consider the ratio of full conditionals given in step 4. For large data sets, computation of $I_0(\kappa)^{-n}$ creates values close to 0; on the other hand, the exponentials can give extremely large values. Thus, in practice, it is computationally beneficial to work on the log-scale, and so the log acceptance probability becomes $\alpha' = \min(0, A')$, where

$$A' = -n \log \left[\frac{I_0(\kappa^*)}{I_0(\kappa^{(j-1)})} \right] + a \log \left(\frac{\kappa^*}{\kappa^{(j-1)}} \right) + (\kappa^* - \kappa^{(j-1)}) \left[\sum_{i=1}^n \cos(x_i - \mu) - b \right]$$

Let us illustrate the algorithm. The data are a random sample of 20 values simulated from a $vM(-\pi/2, 2)$ distribution. The prior distributions are taken as

$$\mu \sim vM\left(-\frac{\pi}{3}, 1\right), \quad \kappa \sim Ga(3, 2) \quad (2.8.4)$$

independently. Let our tuning parameter ν be 0.1, based on our findings in Section 2.6. We provisionally run the algorithm for 1000 iterations. The subsequent diagnostic plots in Figure 2.9.1 suggest that thinning by around 40 is required in order to produce un-autocorrelated results.

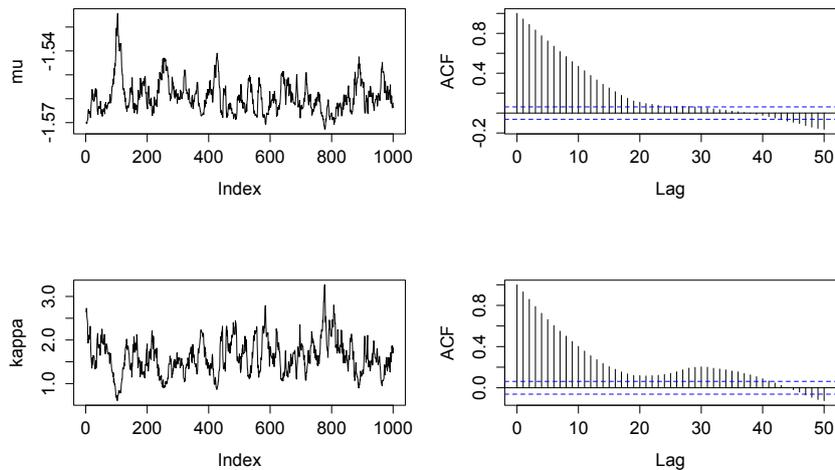


Figure 2.8.1: Diagnostic plots for posterior parameters, unthinned

As in Section 2.6, the diagnostic plots suggest that thinning is required. Thus after further experimentation we re-run our algorithm for 10,000 iterations, but now thinning our realizations by 40. Updated diagnostic plots are given in Figure 2.8.2, which suggest that realizations from the joint posterior are now (almost) un-autocorrelated.

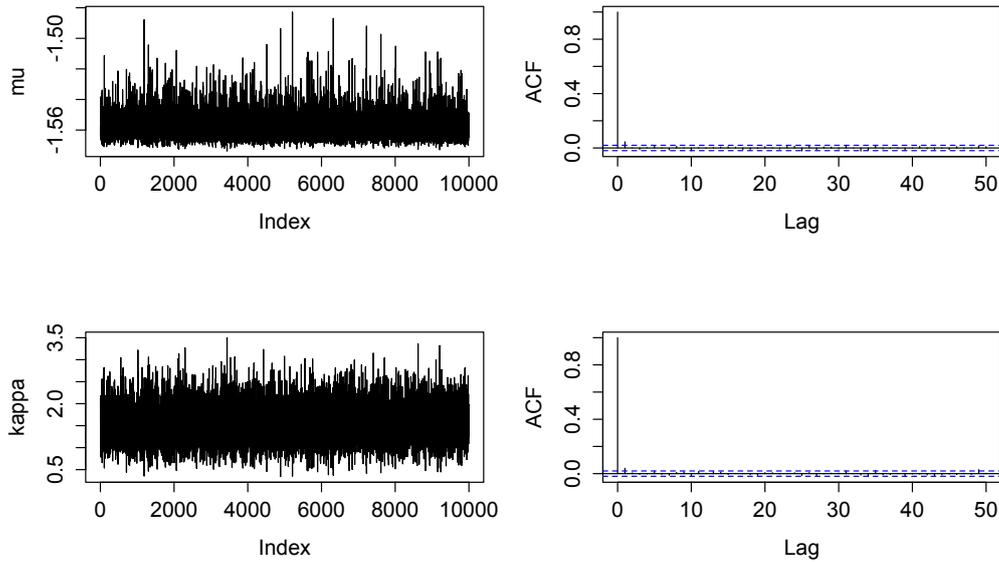


Figure 2.8.2: Diagnostic plots for posterior parameters, thinned

Posterior realizations now appear to be un-autocorrelated. Summary statistics for the posterior distribution are

$$E(\mu|\mathbf{x}) = -1.558, \quad 95\%CI = (-1.574, -1.541)$$

and

$$E(\kappa|\mathbf{x}) = 1.541, \quad 95\%CI = (0.716, 2.366)$$

The mean angle of the simulated data is -1.58 , whilst the prior mean is $\pi/3 \approx -1.047$, again suggesting that the data has been highly informative. The posterior mean of κ is closer to the prior mean of 1.5 than the data distribution concentration 2, although this value is plausible as it lies within the 95% confidence interval. This conflicts with our findings in Section 2.6, and further investigation shows that the difference is due to a lower prior variance in this instance.

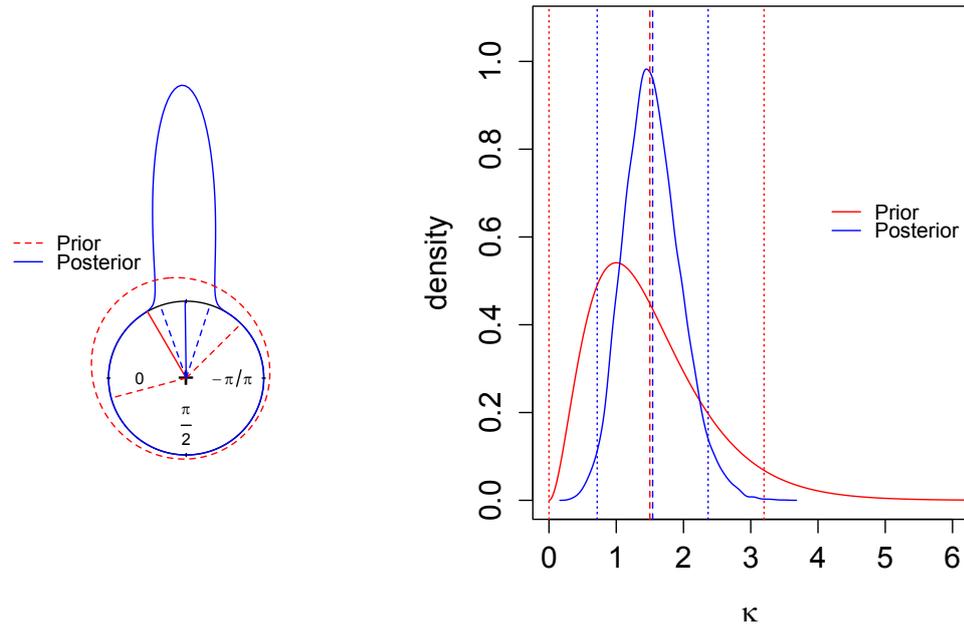


Figure 2.8.3: Prior and posterior densities for μ and κ

Figure 2.8.3 demonstrates characteristics that are similar to those in Figures 2.6.3 and 2.7.1.

2.9 Vanishing directions of homing pigeons

As discussed in the introduction, one of the predominant uses for circular statistics is the study of ecological systems, e.g. animal migration. A dataset provided by Schmidt-Koenig (1975) gives the vanishing directions of 15 homing pigeons, released approximately 16km Northwest of their loft. The term vanishing refers to the point at which the pigeon is no longer visible from the origin (the location of release). The vanishing directions of the pigeons are presented in Figure 2.9.1.

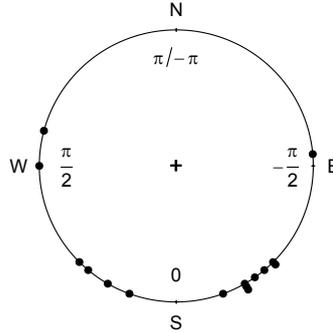


Figure 2.9.1: Geographical directions and corresponding angular values of vanishing directions of pigeons

As the directions are originally given as angles in degrees, a basic transformation has been applied; additionally, South has been set to be the middle of our range, i.e. South corresponds to a direction of 0.

We require appropriate prior distributions for μ and κ . Unfortunately, an expert opinion was not readily available for elicitation, and so prior parameters have been chosen based on limited personal knowledge of the subject. As the pigeons were released Northwest of their loft, it seems sensible to have a prior mean in the direction of their loft, i.e. Southeast; this direction is represented by an angle of $-\pi/4$ within our framework. However, my knowledge of animal migration is limited, and so we use a low concentration parameter, taken to be $1/10$. Thus our prior distribution for μ is

$$\mu \sim vM\left(-\frac{\pi}{4}, \frac{1}{10}\right). \quad (2.9.1)$$

Regarding concentration, it seems intuitive that each homing pigeon is likely to use the same method to “home”. Thus I believe the concentration should be high, and by graphical assessment of various von Mises densities I suggest that a prior mean $E(\kappa) = 15$ is appropriate. However, my knowledge regarding the vanishing directions of pigeons is limited, and so we will use a prior density with a large variance in order to account for the lack of certainty. The decided prior distribution is

$$\kappa \sim Ga(1.5, 0.1). \quad (2.9.2)$$

We can now perform our Bayesian analysis for the two unknown parameters, as discussed in Section 2.8. Several trial runs were performed and assessment of subsequent diagnostic

plots (omitted) suggest that observing every 80th realization only produce realizations with desirable characteristics.

The conditional posterior kernel density estimates for κ and μ are presented in Figures 2.10.2 and 2.10.3, alongside the priors and 95% equi-tailed confidence intervals.

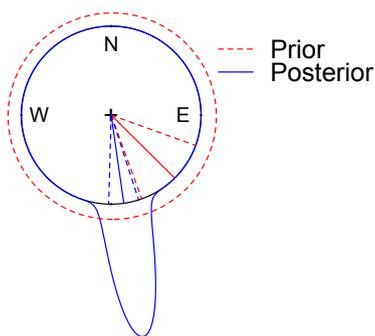


Figure 2.9.2: Prior and posterior μ density estimates with 95% CIs

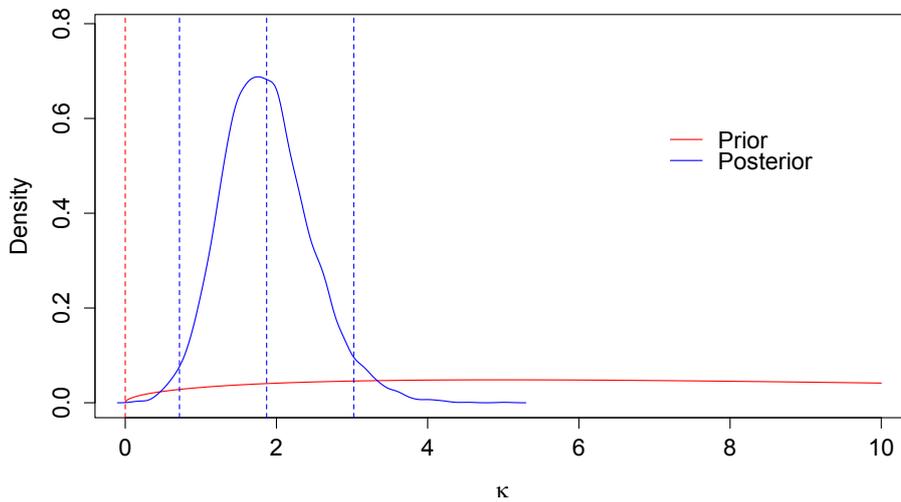


Figure 2.9.3: Prior and posterior κ density estimates with 95% CIs

The summary statistics for μ *a posteriori* are

$$E(\mu|\mathbf{x}) = -0.141, \quad 95\%CI = (-0.310, 0.028).$$

These values suggest that the pigeons begin their journey home by heading South to begin with, before (presumably) heading East to arrive home. This disagrees with my prior distribution, although posterior analysis has been dominated by the data available. Note that the prior and posterior 95% CIs do not even overlap, and so the assumption that pigeons fly directly towards their target is likely to be wrong.

The summary statistics for κ *a posteriori* are

$$E(\kappa|\mathbf{x}) = 1.87, \quad 95\%CI = (0.717, 3.02),$$

suggesting that the direction of flight of one pigeon is similar to the next. This value is somewhat surprisingly low however and does not support the assumption that each pigeon uses an identical method to home.

Chapter 3

Spherical Data and the von Mises-Fisher Distribution

3.1 Simulation

A method for generating $X \sim vMF_p(\boldsymbol{\mu}, \kappa)$ is discussed by Ulrich (1984) where a rejection algorithm is used, which is of a similar form to the case for two dimensions. However, one caveat is that the algorithm provided requires first the simulation of the spherical uniform distribution, i.e. when $\kappa = 0$ in the von Mises-Fisher distribution. Fortunately, as we are dealing with spheres of radius 1 and centered at the origin, this is rather simple, and consists of simulating a p -length vector of $N(0, 1)$ random variables and returning the normalised vector.

A decade after the original publication of the algorithm, Wood (1994) stated that the algorithm specified by Ulrich does not work, without giving specific reasons. He provided an altered specification of the algorithm as:

1. Calculate

$$b = \frac{\sqrt{4\kappa^2 + (p-1)^2} - 2\kappa}{p-1}, \quad x_0 = \frac{1-b}{1+b}, \quad c = \kappa x_0 + (p-1) \log(1-x_0^2)$$

2. Generate

$$Z \sim \text{Beta}\left(\frac{m-1}{2}, \frac{m-1}{2}\right), \quad U \sim U(0, 1)$$

3. Calculate

$$W = \frac{1 - (1+b)Z}{1 - (1-b)Z}$$

4. If $\kappa W + (p-1) \log(1-x_0 W) - c < \log U$ then reject simulations and return to step 1.

5. Generate

$$V \sim SU(p-1)$$

and return

$$X^T = \left(W, \left[\sqrt{1 - W^2} V^T \right] \right).$$

It follows that X is a p -length vector and has a von Mises-Fisher distribution with mean direction $(1, 0, \dots, 0)^T$ and concentration parameter $\kappa \geq 0$.

One feature currently missing from the simulation algorithm is the concept of rotating the final modal direction. This is required to simulate von Mises-Fisher distributions with $\boldsymbol{\mu} \neq (1, 0, 0, \dots, 0)$, and is crucial in order to implement the Bayesian algorithm we are leading towards. The easiest method to alter the mean direction of simulation is to apply rotation by angles. Two possible methods are:

- to use the corresponding rotation matrix on the simulated value of X ; or
- to identify the corresponding α, β angular values of the X values simulated, and to add the desired angles to each draw individually, before transforming back.

We will approach the problem using the latter method, largely due to having previously created functions that easily transforms the points on the unit sphere between Cartesian and Spherical coordinate systems, and so incorporation into the algorithm through R was convenient.

3.2 Visual representation of the von Mises-Fisher distribution

A graphical demonstration of simulations of the von Mises-Fisher distribution is presented in Figure 3.2.1. Considered first is the case where $\alpha = \beta = 0$ (no rotation applied), and so the mean direction is $\boldsymbol{\mu} = (1, 0, 0)$. Three simulations are given, to demonstrate the effect of increasing κ .

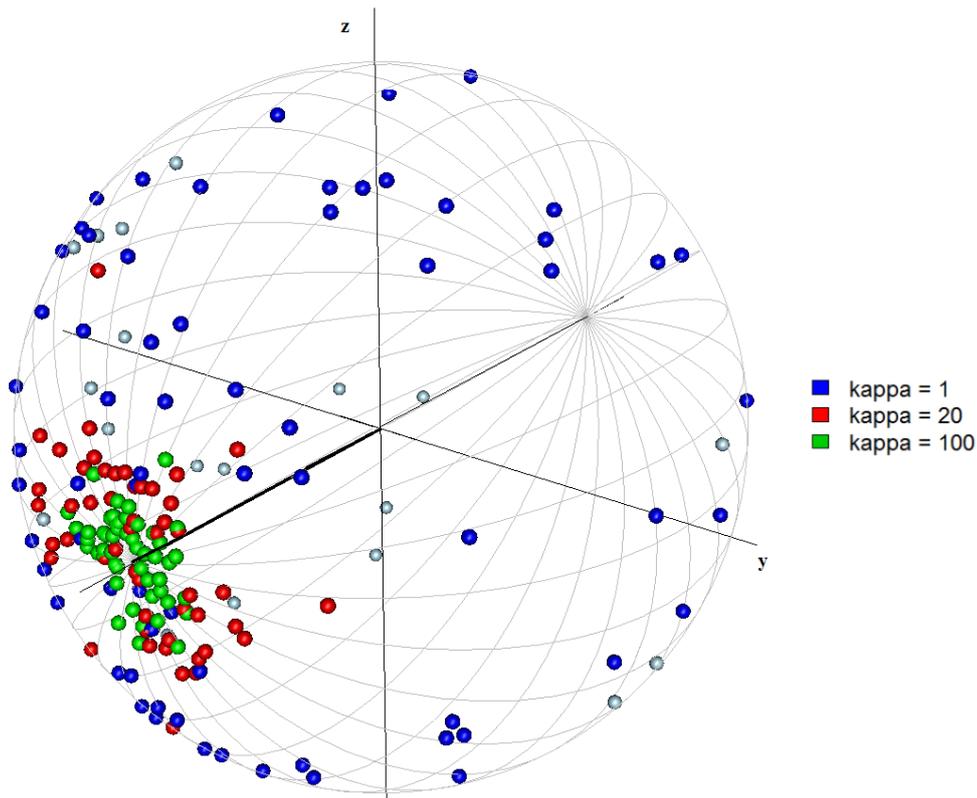


Figure 3.2.1: Simulations of von Mises-Fisher distributions with mean direction $\boldsymbol{\mu} = (1, 0, 0)^T$

As desired, the points become increasingly clustered around the point $(1, 0, 0)$ as our concentration parameter κ increases. Note that for $\kappa = 1$, the points do tend to towards the expected mean direction (noting the significant absence of points located near $(-1, 0, 0)$, the antimode), despite initially appearing evenly distributed.

3.3 Discussion of priors

A discussion of prior distributions is given by Mardia and El-Atoum (1976). As $\boldsymbol{\mu}$ is required to take a value on the surface of the unit sphere it is logical that the von Mises-Fisher distribution itself is suggested as a prior. Also, conjugacy between the prior and posterior for $\boldsymbol{\mu}$ was seen in two dimensions (see Equation (2.8.6)). Therefore, we adopt the prior

$$\boldsymbol{\mu} \sim vMF_p(\mathbf{m}, c), \quad (3.3.1)$$

where $\|\mathbf{m}\| \in \mathcal{S}_p$ and $c \geq 0$ are hyperparameters to be elicited.

As with the two dimensional von Mises distribution, the concentration parameter κ remains a scalar with condition $\kappa \geq 0$. Thus, we continue to assume a Gamma prior for κ (given in Equation (2.6.1)).

3.4 The posterior distribution

The likelihood function for a random sample of n points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ from a $vMF_p(\boldsymbol{\mu}, \kappa)$ distribution is

$$\begin{aligned} f(X|\boldsymbol{\mu}, \kappa) &= \prod_{i=1}^n \left\{ \left(\frac{\kappa}{2}\right)^{p/2-1} \frac{1}{\Gamma(p/2)I_{p/2-1}(\kappa)} \exp\{\kappa \boldsymbol{\mu}^T \mathbf{x}_i\} \right\} \\ &= \left[\left(\frac{\kappa}{2}\right)^{p/2-1} \Gamma(p/2)^{-1} I_{p/2-1}(\kappa)^{-1} \right]^n \exp\{n\kappa \boldsymbol{\mu}^T \bar{\mathbf{x}}\}, \end{aligned} \quad (3.4.1)$$

where $\|\mathbf{x}\| \in \mathcal{S}_p$. Applying Bayes' Theorem gives the joint posterior for density $\boldsymbol{\mu}$ and κ as

$$\begin{aligned} \pi(\boldsymbol{\mu}, \kappa|X) &\propto f(X|\boldsymbol{\mu}, \kappa)\pi(\boldsymbol{\mu}, \kappa) \\ &\propto f(X|\boldsymbol{\mu}, \kappa)\pi(\boldsymbol{\mu})\pi(\kappa) \\ &\propto [\kappa^{p/2-1} I_{p/2-1}(\kappa)^{-1}]^n \exp\{n\kappa \boldsymbol{\mu}^T \bar{\mathbf{x}}\} \exp\{c\mathbf{m}^T \boldsymbol{\mu}\} \kappa^{a-1} e^{-b\kappa} \\ &\propto I_{p/2-1}(\kappa)^{-n} \kappa^{n(p/2-1)+a-1} \exp\{[n\kappa \bar{\mathbf{x}}^T + c\mathbf{m}^T] \boldsymbol{\mu} - b\kappa\} \end{aligned} \quad (3.4.2)$$

where $\|\boldsymbol{\mu}\| \in \mathcal{S}_p$ and $\kappa \geq 0$.

The full conditional distribution of $\boldsymbol{\mu}$ is

$$\begin{aligned} \pi(\boldsymbol{\mu}|X, \kappa) &\propto \exp\{[n\kappa \bar{\mathbf{x}}^T + c\mathbf{m}^T] \boldsymbol{\mu}\} \\ &\propto \exp\{K\mathbf{M}^T \boldsymbol{\mu}\} \end{aligned} \quad (3.4.3)$$

where

$$K = \|n\kappa \bar{\mathbf{x}} + c\mathbf{m}\|, \quad M = \frac{n\kappa \bar{\mathbf{x}} + c\mathbf{m}}{K} \quad (3.4.4)$$

for $K \geq 0$ and $M \in \mathcal{S}_p$. This is the density of a von-Mises Fisher distribution up to proportionality, and so

$$\boldsymbol{\mu}|\kappa, X \sim vMF_p(M, K)$$

Likewise the full conditional distribution for κ is

$$\pi(\kappa|\boldsymbol{\mu}, X) \propto I_{p/2-1}(\kappa)^{-n} \kappa^{n(p/2-1)+a-1} \exp\{(n\boldsymbol{\mu}^T \bar{\mathbf{x}} - b)\kappa\}, \quad \kappa \geq 0. \quad (3.4.5)$$

As for two dimensions (see Equation 2.8.3), this distribution is non-standard and so simulation of realizations requires MCMC methods.

3.5 Metropolis-Hastings algorithm

We now consider the implementation of a Metropolis-Hastings algorithm in order to generate realizations of our full posterior distribution. As discussed in the previous section, the full conditional distribution of $\boldsymbol{\mu}$ has a known distribution, and therefore we can use a Gibbs update this parameter. However, for κ , we must generate a proposed value and use a Metropolis-Hastings algorithm to accept or reject the proposed value. The detailed algorithm is as follows:

- (1) Initialize the algorithm at $j = 1$. Set $(\boldsymbol{\mu}^{(0)}, \kappa^{(0)})$ to some initial starting value, potentially the prior means.
- (2) Generate a proposed value $\kappa^* \sim LN(\log(\kappa^{(j-1)}), \nu^2)$, where ν is a tuning parameter chosen by the user.
- (3) Calculate the acceptance probability $\alpha(\kappa^*, \kappa^{(j-1)}) = \min(1, A)$, where

$$\begin{aligned} A &= \frac{\pi(\kappa^* | \boldsymbol{x}, \boldsymbol{\mu}^{(j-1)})}{\pi(\kappa^{(j-1)} | \boldsymbol{x}, \boldsymbol{\mu}^{(j-1)})} \times \frac{q(\kappa^{(j-1)} | \kappa^*)}{q(\kappa^* | \kappa^{(j-1)})} \\ &= \frac{I_{p/2-1}(\kappa^*)^{-n} (\kappa^*)^{n(p/2-1)+a-1} \exp\{n\kappa^* (\boldsymbol{\mu}^{(j-1)})^T \bar{\boldsymbol{x}} - b\kappa^*\}}{I_{p/2-1}(\kappa^{(j-1)})^{-n} (\kappa^{(j-1)})^{n(p/2-1)+a-1} \exp\{n(\kappa^{(j-1)}) (\boldsymbol{\mu}^{(j-1)})^T \bar{\boldsymbol{x}} - b(\kappa^{(j-1)})\}} \times \left(\frac{\kappa^*}{\kappa^{(j-1)}}\right) \\ &= \left[\frac{I_{p/2-1}(\kappa^*)}{I_{p/2-1}(\kappa^{(j-1)})}\right]^{-n} \left(\frac{\kappa^*}{\kappa^{(j-1)}}\right)^{n(p/2-1)+a} \exp\{n(\kappa^* - \kappa^{(j-1)}) (\boldsymbol{\mu}^{(j-1)})^T \bar{\boldsymbol{x}} - b(\kappa^* - \kappa^{(j-1)})\} \end{aligned}$$

- (4) Set $\kappa^{(j)} = \kappa^*$ with probability $\alpha(\kappa^*, \kappa^{(j-1)})$, otherwise set $\kappa^{(j)} = \kappa^{(j-1)}$.
- (5) Generate $\boldsymbol{\mu}^{(j)} \sim vMF(M, K)$, where

$$M = \frac{n\kappa^{(j)} \bar{\boldsymbol{x}}^T + c\boldsymbol{m}}{K}, \quad K = \|n\kappa^{(j)} \bar{\boldsymbol{x}}^T + c\boldsymbol{m}\|.$$

- (6) Set $j := j + 1$ and return to step (2).

We now illustrate the algorithm above using test data consisting of a random sample of 40 data points simulated from a $vMF_3(\boldsymbol{\mu} = [0.7071, 0.5, 0.5], \kappa = 20)$ distribution (where the mean direction is a result of rotation by $\alpha = \beta = \pi/4$). The data are illustrated in Figure 3.5.1.

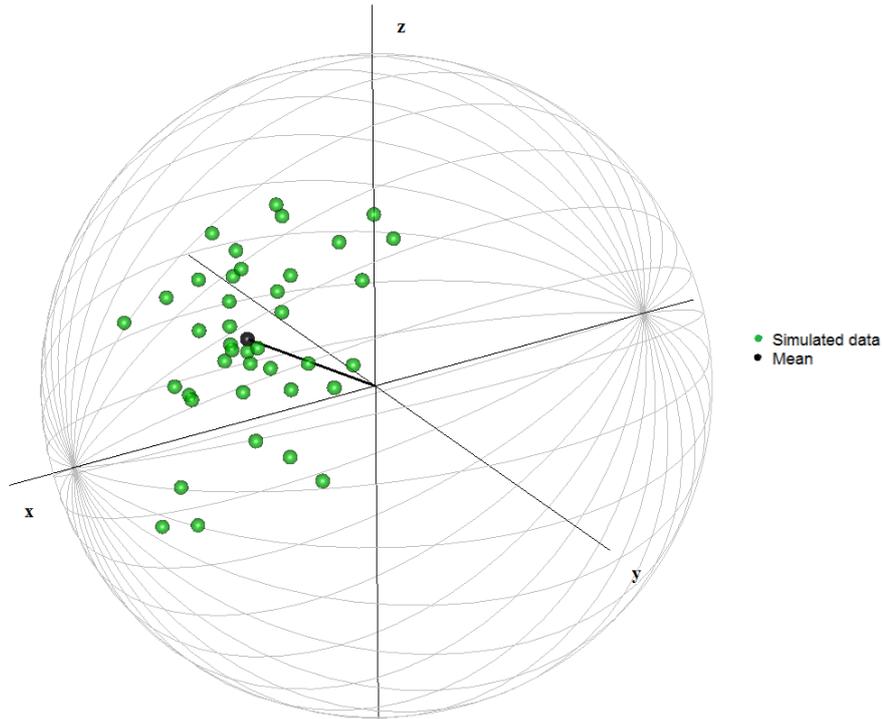


Figure 3.5.1: 40 simulated points from a $vMF_3(\boldsymbol{\mu} = [0.7071, 0.5, 0.5], \kappa = 20)$ distribution

We take our prior distributions to be

$$\boldsymbol{\mu} \sim vMF_3(\boldsymbol{m} = (0, 0, 1)^T, c = 50), \quad \kappa \sim Ga(5, 1/4).$$

We run our algorithm initially for 1,000 iterations. Investigation using diagnostic plots (omitted) into the effect of the magnitude of the tuning parameter ν suggests that, for higher dimensions, a reduction to $\nu = 0.05$ is more appropriate than the previous value of 0.1.

Traceplots and autocorrelation plots for the posterior parameters are provided in Figure 3.5.2.

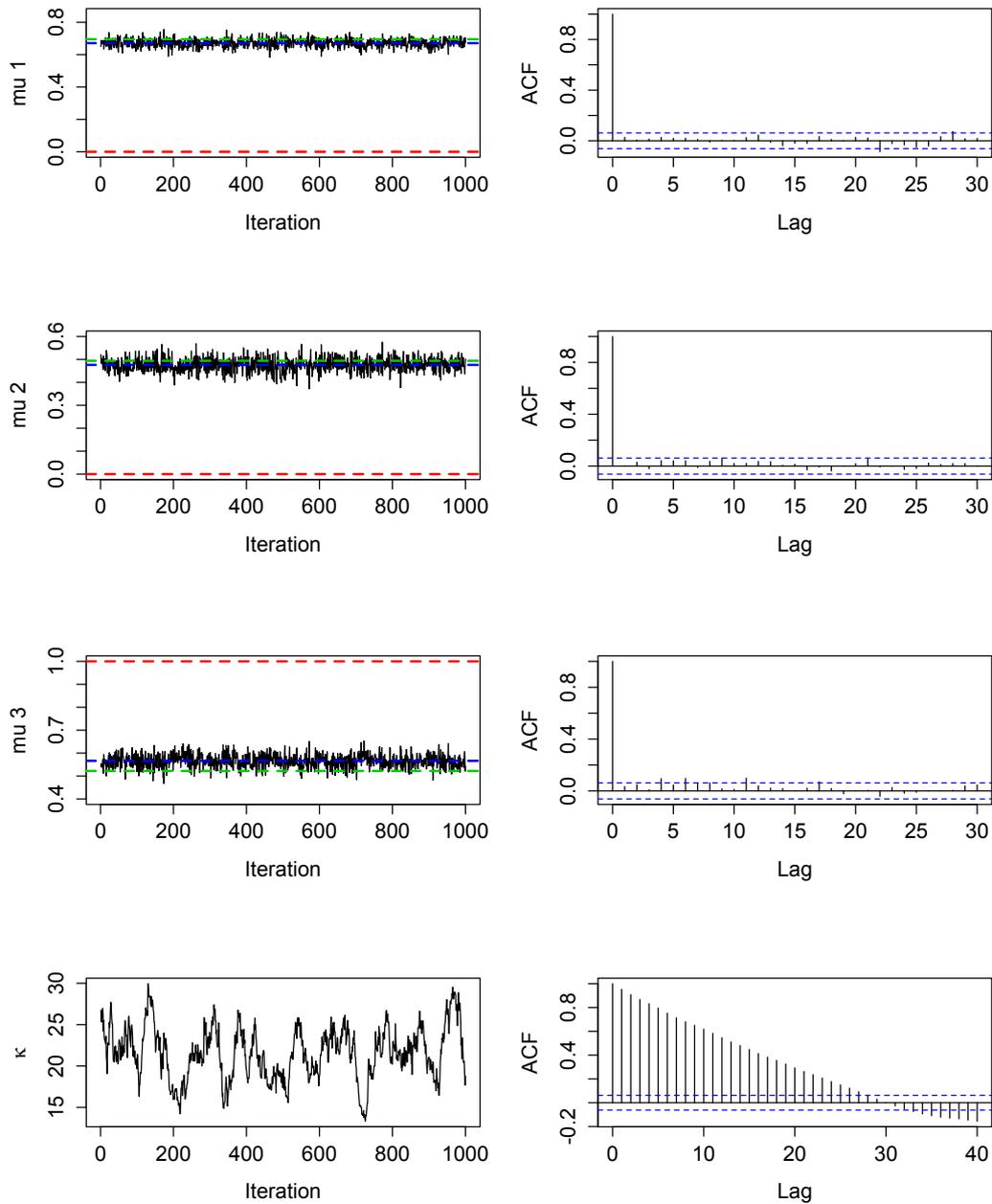


Figure 3.5.2: Trace (left) and autocorrelation (right) plots for posterior realizations, unthinned

Each element of $\boldsymbol{\mu}$ appears to be mixing well, and require no thinning. In addition, the mean of the simulated data (green), the posterior mean (blue) and the prior direction (red) has been highlighted; however, the posterior mean and the mean of the simulated data are so similar that it is difficult to tell the difference graphically. This suggests that the data has been highly informative, which is to be expected due to the high concentration

of the data and the relatively low certainty placed on prior μ .

The autocorrelation plot for κ reveals that realizations are autocorrelated up to approximately lag 30, and thus we need to thin our realizations in order to produce suitable posterior draws. Additionally the traceplot appears to be “snaking”, which again suggests successive draws are correlated.

The previous algorithm specifications (unthinned) result in an effective sample size of 4.3 for κ and so the level of thinning required can be estimated as $1000/4.3 \approx 233$. Experimentation with algorithmic parameters suggest that thinning by 250 is required to produce un-autocorrelated result, a significantly higher level of thinning than suggested by the autocorrelation plot in Figure 3.5.2.

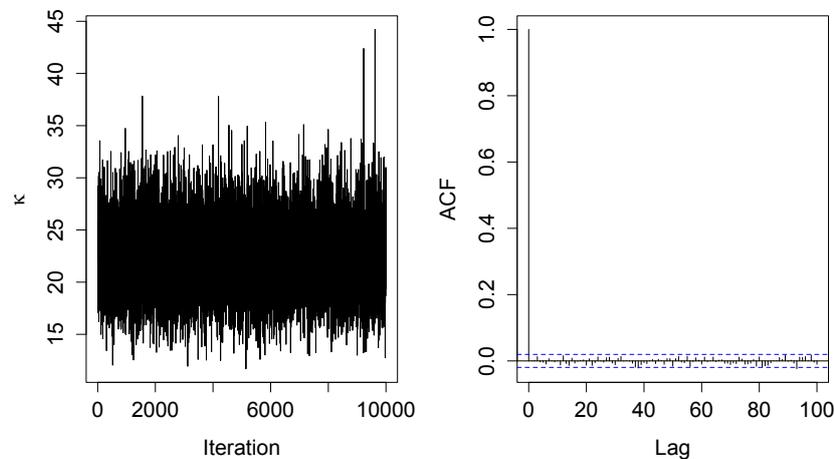


Figure 3.5.3: Diagnostic Plots for posterior κ , thinned

The updated diagnostic plots for κ are given in Figure 3.5.3, once thinning by 250 has been applied. The autocorrelation now has desirable characteristics, with negligible autocorrelation at lag 1, and the traceplot suggests that the chain is centered on a value close to the prior mean. The effective sample size of posterior realizations of κ is now 10,000 and so realizations are (almost) un-autocorrelated.

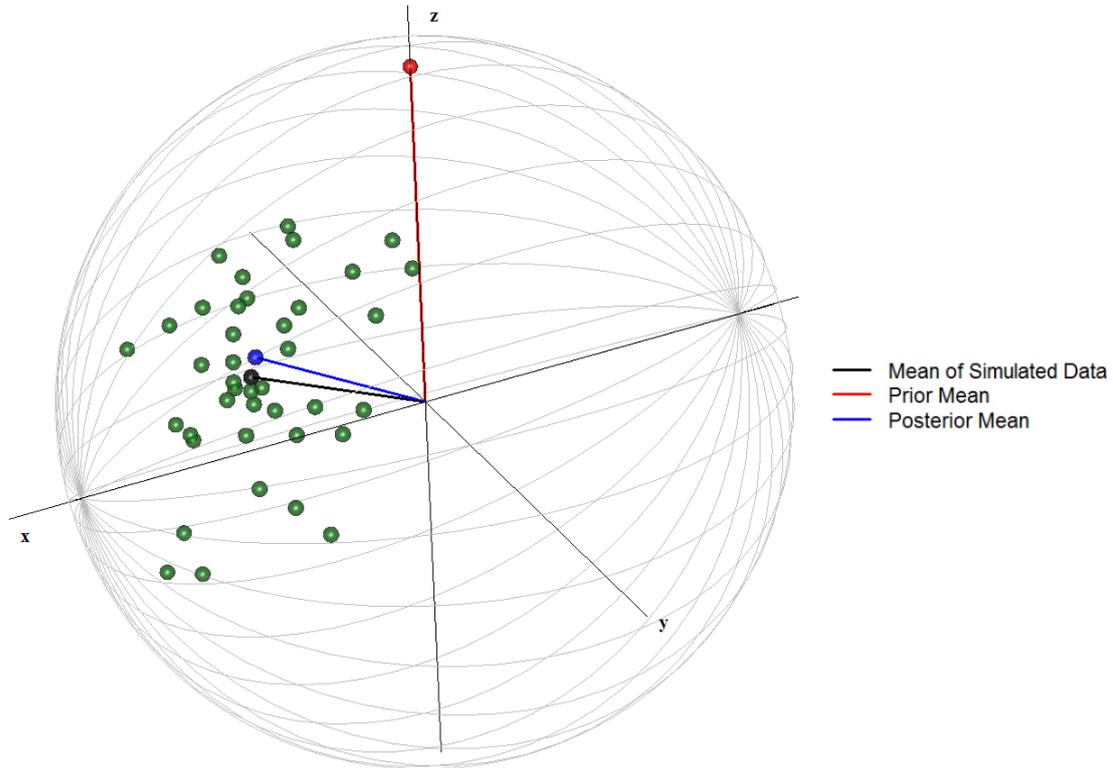


Figure 3.5.4: Comparison of the data, prior and posterior means of μ

Figure 3.5.4 shows that the posterior of μ has been dominated by the data rather than the prior. We calculate the Euclidean distances in \mathcal{S}_p of the data and prior means to the posterior mean as

$$\|E(\mu|\mathbf{x}) - \bar{\mathbf{x}}\| = 0.008, \quad \|E(\mu|\mathbf{x}) - E(\mu)\| = 0.971,$$

demonstrating that the posterior has been formed with heavy weighting from data information, and less from the prior.

Confidence regions for the prior and the posterior have proved difficult to plot on the surface of the sphere; however, an accurate measure of spread can be calculated. Taking an equi-spaced confidence region (CR) around the mean of the density concerned, such that $100\alpha\%$ of realizations lie within this region, we define a measure of spread r to be the radius from the mean to one point on the boundary of this region. Thus, densities with high concentration will have a lower r , and vice versa. We will take $\alpha = 0.95$ so 95% CRs will be measured.

The prior and posterior densities have r values calculated as

$$r_{prior} = 0.347, \quad r_{posterior} = 0.083$$

respectively. Thus, our prior 95% CR covers a larger area than the posterior equivalent, again suggesting posterior certainty has been heavily influenced by the data.

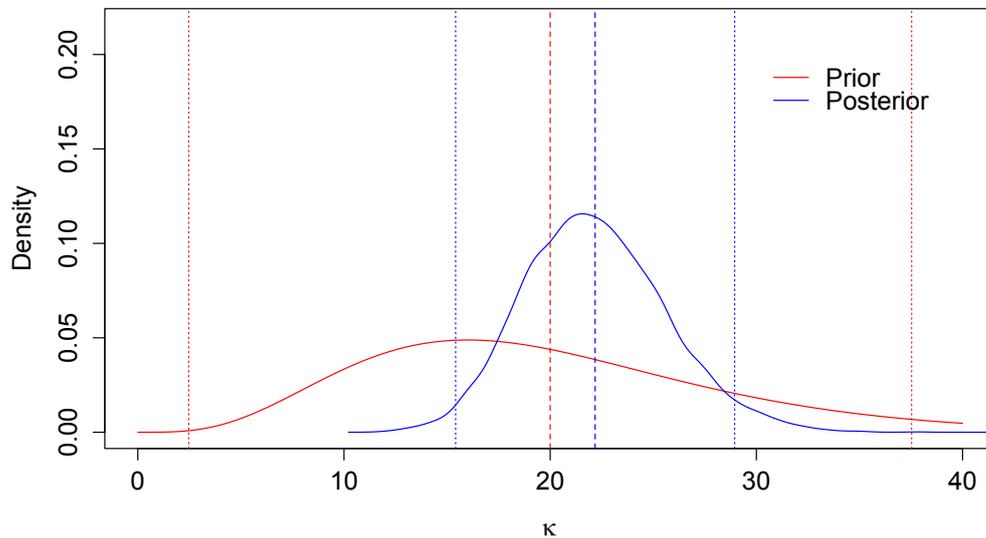


Figure 3.5.5: Prior and posterior κ with 95% CIs

The summary statistics for κ *a posteriori* are

$$E(\kappa|\mathbf{x}) = 22.64, \quad 95\%CI = (15.993, 29.296),$$

showing that our posterior mean is similar to our prior mean of 20.

As with the concentration in two dimensions (see Figures 2.8.3 and 2.9.3), the posterior density of κ appears to be almost symmetric (with a slight tendency towards positive skewness) around the posterior mean, suggesting that a Gamma or Log-normal density might be a valid approximation for the full conditional distribution in Equation (3.4.5).

3.6 Cosmic rays

Littlefield (1979) states that primary cosmic rays consist of highly charged protons or electrons that travel at velocities almost near the speed of light, with origins somewhere out in space. As cosmic rays enter the Earth's magnetic field, they collide with atmospheric molecules (principally Oxygen and Nitrogen) and produce a large number of secondary particles in showers. One of these secondary particles is the μ -meson, or muon, whose location on arrival on the surface of Earth can be measured by the impact it has on distilled water.

Although cosmic rays can have an energy as high as 10^{19} eV, the data collected by Toyoda *et al.*(1965) focuses on cosmic rays in the energy region 10^{14} to 10^{17} eV and we shall analyse this data set; henceforth, these are referred to as the low-level muons. The original data is given in terms of declination and right-ascension in degrees, and so an appropriate transformation to our preferred coordinate system is required. The arrival locations are presented in Figure 3.6.1.

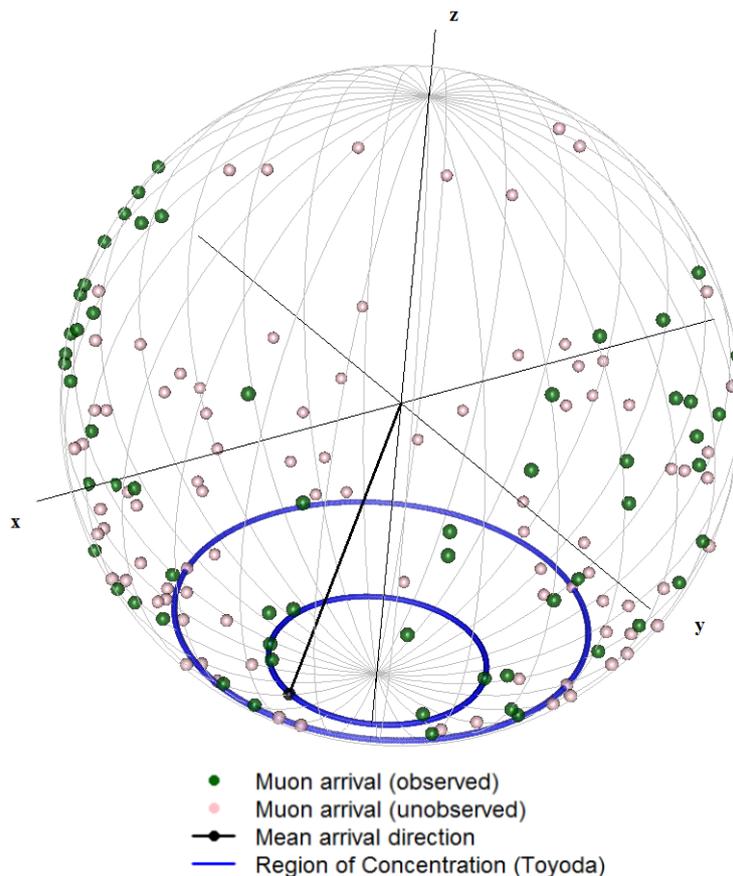


Figure 3.6.1: Arrival locations of low-level muons

Graphical assessment of Figure 3.6.1 suggests that low-level muons can arrive from all directions, although there does appear to a clustering around the lower southern hemisphere and a relative scarcity of arrivals near the North pole. Rudimentary analysis by Toyoda *et al.* suggests that there is a concentration of points with right ascension measurements in the range 200° to 220° , although variability in terms of declination is not discussed. This concentration region is also provided in Figure 3.6.1, and the mean direction of the arrival points is contained within this band, making the conclusions reached by Toyoda plausible.



Figure 3.6.2: Explanation of galactic cosmic ray movements

Appropriate prior distributions were elicited through discussions with Dr Andrew Fletcher. After providing a brief overview of the movements of cosmic rays within the galaxy (provided in Figure 3.6.2), Dr Fletcher was kind enough to calculate the specific direction in which he believes cosmic rays would be likely to arrive, given as a declination value of -29° and a right ascension value of 256° . Regarding concentration of the points, Dr Fletcher stated that the ratio of points on the hemisphere centered on the mean direction and the points centered on the antinode is approximately 30 to 1. After graphically exploring von Mises-Fisher distributions with a variety of concentration parameters, his beliefs best fit a distribution with concentration parameter 10.

Weighting the prior distributions with appropriate levels of certainty is also required. Dr Fletcher stated that his certainty in the concentration parameter κ could be represented by a 5% chance of a ratio greater than 100 to 1. Using the same deduction method as described above, this ratio appears equivalent to a von Mises-Fisher distribution with concentration parameter 40. Therefore, our prior distribution for κ satisfies the following properties:

$$E(\kappa) = 10, \quad Pr(\kappa > 40) = 0.05.$$

Through manipulation of the gamma distribution, this suggests prior parameters of $a = 0.031$ and $b = 0.0031$. This implies a high prior variance of 3225.8, suggesting that information for the posterior distribution for κ will be dominated by the data.

In summary, the prior distribution with independent components is

$$\boldsymbol{\mu} \sim vMF_3 \left(\boldsymbol{m} = (0.4848, -0.8486, 0.2116)^T, c = 15 \right), \quad \kappa \sim Ga(0.031, 0.0031).$$

Initial testing of our algorithm produced results (diagnostically) comparable to those in section 3.5 (omitted). Final analysis is subsequently produced by running for 10,000 iterations with thinning of 500, and a κ tuning parameter of 0.05 (as used previously). Note that initial values of $\boldsymbol{\mu}$ and κ are given as the prior means, i.e. $\boldsymbol{\mu}_0 = \boldsymbol{m}$ and $\kappa_0 = 10$.

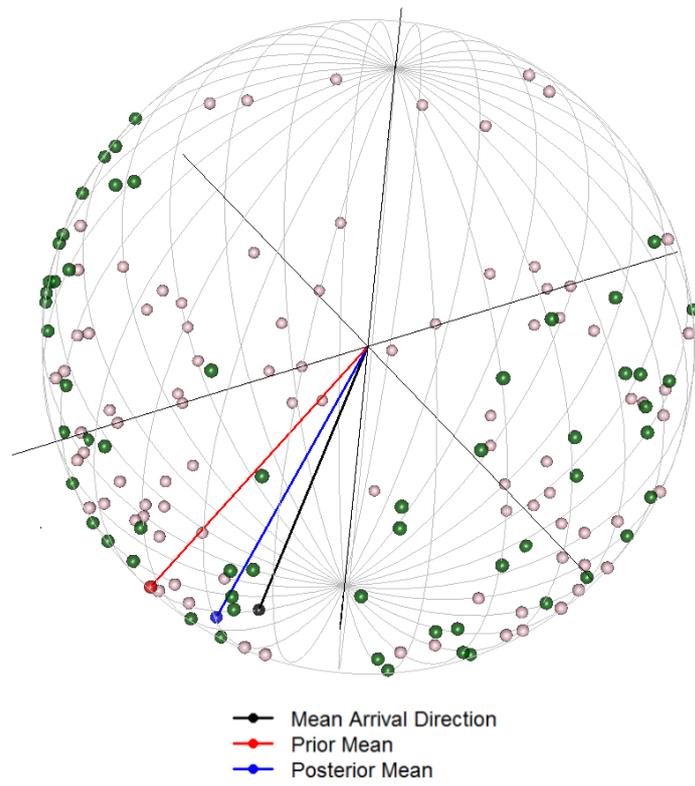


Figure 3.6.3: Comparison of data, prior and posterior

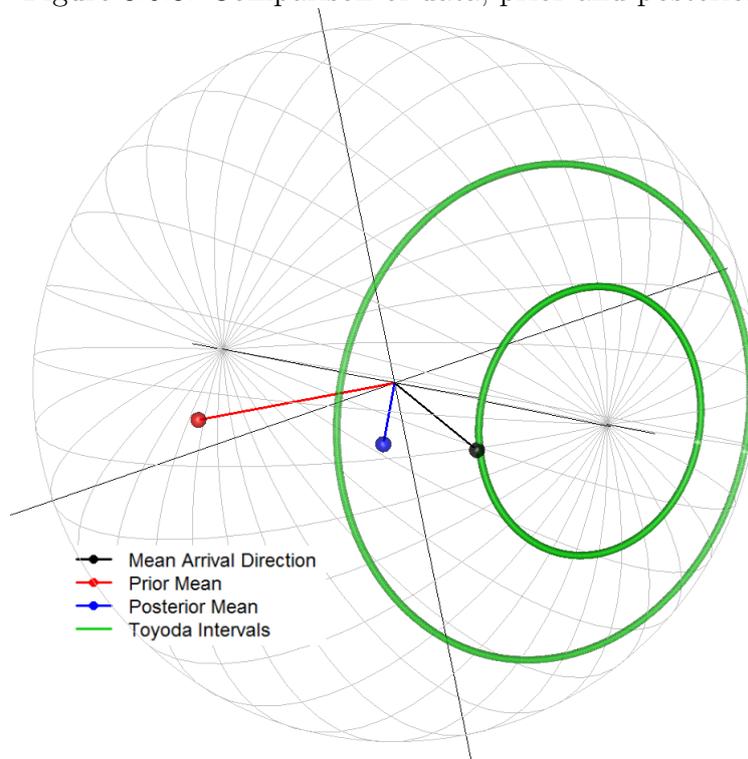


Figure 3.6.4: Comparison of data, prior and posterior

Figure 3.6.3 demonstrates the association of the mean, prior and posterior, and Figure 3.6.4 an alternative view of the same plot with the data removed and the right ascension confidence interval provided by Toyoda included.

Both figures suggest that the posterior mean direction is an almost equal weighted average of the prior mean and the data. Using the method for measuring spread discussed in Section 3.5, CR radius values for the prior and the posterior are

$$r_{prior} = 0.631, \quad r_{posterior} = 0.387,$$

respectively. This demonstrates that our uncertainty about μ has decreased from prior to posterior analysis, and so the data has been informative; however, the reduction in r (38.6%) is by significantly less than for our simulated test data in Section 3.5 (76.1%) due to the increased spread of the cosmic ray data as compared to the test data (Figure 3.5.1).

Figure 3.6.4 suggests that the posterior mean lies comfortably within the confidence interval provided by Toyoda.

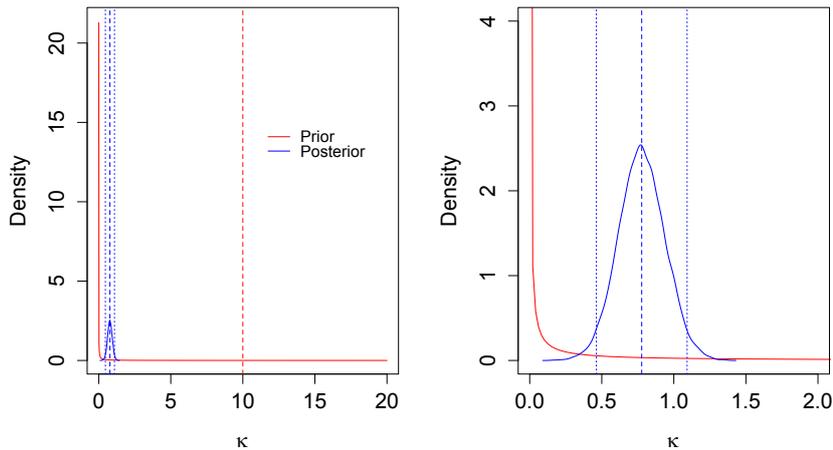


Figure 3.6.5: Prior and posterior κ density estimates, with 95% CIs

Figure 3.6.5 provides a kernel density estimate of our posterior for κ compared with the prior. As our prior has a significantly higher variance than our posterior, the original plot of comparison is not appropriate for graphical analysis of the posterior, and so a close-up has also been provided.

Again, we can see the distinctive “almost” symmetric posterior for κ . Summary statistics for κ *a posteriori* are

$$E(\kappa|\mathbf{x}) = 0.777, \quad 95\%CI = (0.462, 1.092).$$

Due to a large prior variance, the posterior for κ is dominated by information from the data. A posterior mean of 0.777 implies that there is little concentration in the arrival direction of muons.

Chapter 4

Conclusions and Further Work

MCMC methods have been successfully implemented on the von Mises-Fisher distribution in both two and three dimensions, and the algorithm in Section 3.5 can be applied in p dimensions. This suggests that Bayesian analysis on data for geometrical objects in \mathcal{S}_p is perfectly plausible and straightforward to implement.

Informal testing for the von Mises distribution was discussed in Section 2.4. However, no assessment of the goodness-of-fit is performed on the spherical distribution. Although frequentist methods are available, they are more complicated than for the case in two dimensions. Such methods are not explored within this report, as it focuses rather on the Bayesian applications to such data sets. However, for completeness in genuine applications, data should be at least graphically assessed for goodness-of-fit.

Many of spherical data sets considered for use in this project may benefit from modelling using a mixture of distributions. It appears to be a common trait that realistic data contains two or more sub-populations, and investigation into label switching and possible hidden Markov models would be a desirable avenue of exploration.

The investigation into the arrival directions of cosmic rays suggests that, for such celestial interactions as those described in Figures 3.5 and 3.6, distributions on a particular “ring” of values may be more suited for modelling than distributions focused on a particular direction. The von Mises-Fisher distribution provides an inadequate model for such distributions, and further research into the area is required.

References, Data and Acknowledgements

References

- Best, D.J and Fisher, N.I (1979) Efficient Simulation of the von Mises Distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 28, No. 2, pp 152 - 157.
- Embleton, B.J.J, Fisher, N.I and Lewis, T (1993) Statistical analysis of spherical data, 2nd Edition, *Cambridge University Press*, p280.
- Gamerman, D and Lopes, H.F (2006) Markov chain Monte Carlo: stochastic simulation for Bayesian inference, 2nd Edition, *Boca Raton : Taylor & Francis*, pp 211-214.
- Grun, B and Hornik, K (2014) *movMF: Mixtures of von Mises-Fisher Distributions*, version 0.2-0.
- Littlefield, T.A (1979) Atomic and Nuclear Physics - An introduction, 3rd Edition, *New York : Van Nostrand Reinhold*, p346.
- Agostinelli, C and Lund, U (2013) *circular: Circular Statistics*, version 0.4-7.
- El-Atoum, S.A.M and Mardia, K.V (1976) Bayesian Inference for the Von Mises-Fisher Distribution. *Biometrika*, Vol. 63, No. 1, pp. 203-206.
- Jupp, P.E and Mardia, K.V (2000) Directional statistics, 2nd Edition, *John Wiley & Sons Ltd*.
- Schmidt-Koenig, K (1975) Migration and Homing in Animals, *Zoophysiology and Ecology Volume 6*, *Springer-Verlag Berlin Heidelberg*, p88
- Ulrich, G (1984) Computer Generation of Distributions on the m-Sphere. *Journal of the Royal Statistical Society, Applied Statistics*, Vol. 33, No. 2, pp. 158-163
- Wood, A.T.A (1994) Simulation of the von Mises-Fisher Distribution. *Communications in Statistics - Simulation and Computation*, Vol. 23, No. 1, pp. 157-164

Data

Vanishing Directions of Homing Pigeons

The vanishing directions of 15 homing pigeons, used in Section 2.9. Measurements in degrees clockwise from North.

085	135	135	140	145	150	150	150
160	285	200	210	220	225	270	

Arrival Directions of Cosmic Rays

The right ascension (RA) and declination angles (Dec) of 148 muon arrival locations. These measurements are taken within the framework of the widely used equatorial coordinate system.

<i>RA</i> ^o	<i>Dec</i> ^o						
315	-66	203	-37	063	-26	234	07
198	-66	216	-41	077	-28	324	16
099	-63	252	-41	167	-21	311	18
050	-63	279	-41	176	-23	293	24
086	-61	288	-41	185	-26	275	20
221	-61	311	-37	194	-23	266	22
000	-52	320	-30	221	-17	207	13
014	-50	342	-23	216	-15	203	16
063	-50	356	-17	230	-15	212	20
176	-52	347	-19	234	-17	198	22
185	-52	347	-12	248	-10	140	13
207	-52	342	-15	279	-17	117	07
221	-52	338	-12	297	-12	117	11
230	-50	329	-17	234	24	086	11
243	-52	293	-26	234	-03	068	11
347	-48	284	-32	185	-03	032	13
342	-43	261	-35	162	-08	014	16
211	-48	252	-21	144	-15	059	16
293	-48	257	-19	144	-19	068	16
284	-48	261	-23	108	-19	068	22
216	-48	261	-28	104	-19	072	24
207	-46	252	-39	086	-19	086	18
144	-48	230	-35	077	-17	104	20
149	-46	230	-32	068	-12	212	18
077	-46	216	-28	063	-06	153	27
041	-46	212	-30	027	04	216	27
027	-46	207	-28	095	00	342	36
014	-46	194	-30	099	-03	338	40
005	-43	158	-35	122	-08	009	-55
041	-32	117	-30	140	-06	230	38
036	-41	099	-32	144	-08	221	36
050	-39	095	-37	171	04	203	38
063	-35	014	-21	176	02	198	42
077	-41	018	-19	185	07	099	31
113	-41	027	-15	198	09	086	38
153	-41	023	-12	216	02	068	36
176	-43	050	-15	216	04	045	36