Semi-Parametric Bayesian Method on Spherical Data

Thomas J. Jiang[1]

[1]Department of Mathematical Sciences, National Chengchi University, Wen-Shan, Taipei 11605, Taiwan

Abstract

In many contexts in the earth sciences, astrophysics and other fields, it often needs to analyze spherical data, which can be treated as having positions on spherical surface. Many parametric Bayesian methods are used to examine spherical data. We give semi-parametric Bayesian method to make statistical inferences. In particular, we show how to use our semi-parametric method to find the Bayesian predictive density and to test whether two samples are from two populations with the same mean. A von Mises-Fisher distribution is shown to be conjugate for the von Mises-Fisher distribution, which is often used in directional statistics. This makes the computations of our semi-parametric Bayesian method easier.

Key Words: spherical data, Bayesian predictive density, Bayesian two-sample test, von Mises-Fisher distribution

1. Introduction

In many contexts in the earth sciences, astrophysics and other fields, it often needs to analyze spherical data, which can be treated as having positions on spherical surface. Traditional statistical methods on analyzing spherical data can be seen in, among others, Mardia (1972), Fisher, et al. (1987), Jammalamadaka and SenGupta (2001) and references therein. Nunez-Antonio and Gutierrez-Pena (2005) use parametric Bayesian method to examine spherical data. Ghosh, Jammalamadaka, and Tiwari (2003) first study 2-dimenstional circular data with semi-parametric Bayesian techniques. The generalization of the semi-parametric Bayesian technique for 2-dimention to that for 3-dimension is not immediate. In this paper, we shall use semi-parametric method to analyze 3-dimensional spherical data.

The commonly used 3-dimensional distribution for spherical data is von Mises-Fisher distribution. The Ferguson-Dirichlet process is well known for its application to nonparametric Bayesian analysis. In section 2, we give the notations for the von Mises-Fisher distribution and the Ferguson-Dirichlet process. Using the Ferguson-Dirichlet process prior, we derive the posterior distribution for making statistical inference in section 3. In particular, von Mises-Fisher prior distribution is shown to be conjugate for von Mises-Fisher data. This would make the computation easier. In section 4, we give a semi-parametric method to find Bayesian predictive density of a new future datum. We also give a semi-parametric method to test whether two random samples are from two populations with the same mean in section 5. Finally, conclusions are given in section 6.

2. Von Mises-Fisher distribution and Ferguson-Dirichlet process

In directional statistics, the von Mises-Fisher distribution is often used on the (p-1)-dimensional sphere on R^p. Although we concentrate on the cases with p=3 in this paper, here we shall define its probability density for general p of the random unit vector $\mathbf{X}$ with mean $\boldsymbol{\lambda}$ and concentration parameter $\kappa \geq 0$, denoted by $\mathbf{X} \sim \mathrm{vMF}_p(\boldsymbol{\lambda}, \kappa)$, as

$$f_p(\mathbf{x}; \boldsymbol{\lambda}, \kappa) = C_p(\kappa) \exp(\kappa \boldsymbol{\lambda}' \mathbf{x}), \qquad \text{(A)}$$

where the length of the mean vector $\boldsymbol{\lambda}$ is 1 and the normalization constant $C_p(\kappa) = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)}$ with I_ν denoting the modified Bessel function of the first kind and order ν. In particular, $C_3(\kappa) = \frac{\kappa}{4\pi \sinh \kappa} = \frac{\kappa}{2\pi(e^{\kappa} - e^{-\kappa})}$. Using the spherical coordinate system, $\mathbf{x}' = (x_1, x_2, x_3) = \left(\sin\theta^{(1)} \cos\theta^{(2)}, \sin\theta^{(1)} \sin\theta^{(2)}, \cos\theta^{(1)}\right)$, where $\theta^{(1)}$ $(0 \leqq \theta^{(1)} \leqq \pi)$ is the colatitude (polar angle measured from the x_3 axis to $\mathbf{x}$) and $\theta^{(2)}$ $(0 \leqq \theta^{(2)} < 2\pi)$ is the longitude (azimuth angle measured from the x_1 axis to the orthogonal projection of $\mathbf{x}$ on the (x_1, x_2) reference plane). Similarly, $\boldsymbol{\lambda}$ can be expressed as $\boldsymbol{\lambda}' = (\lambda_1, \lambda_2, \lambda_3) = \left(\sin\alpha^{(1)} \cos\alpha^{(2)}, \sin\alpha^{(1)} \sin\alpha^{(2)}, \cos\alpha^{(1)}\right)$. Hence, the von Mises-Fisher distribution pdf of $\mathbf{x}$ (when p=3) in terms of the spherical coordinate system, denoted by $\left(\theta^{(1)}, \theta^{(2)}\right) \sim \mathrm{vMF}_s\left(\left(\alpha^{(1)}, \alpha^{(2)}\right), \kappa\right)$, can be expressed as

$g_3\left((\theta^{(1)},\theta^{(2)});(\alpha^{(1)},\alpha^{(2)}),\kappa\right) = \frac{\kappa}{4\pi\sinh\kappa}\exp\{\kappa\left[\sin\theta^{(1)}\sin\alpha^{(1)}\cos\left(\theta^{(2)}-\alpha^{(2)}\right)+\right.$

$\left.\cos\theta^{(1)}\cos\alpha^{(1)}\right]\}\sin\theta^{(1)}$. (B)

In this paper, we shall use Ferguson-Dirichlet process (Ferguson, 1973) as our nonparametric prior. We say G is a Ferguson-Dirichlet process over a set S with a probability measure G_0 on S and a positive measure c, denoted by $G\sim FDP(c, G_0)$, if $(G(B_1), G(B_2), \dots, G(B_n)) \sim$ Dirichlet $(cG_0(B_1), cG_0(B_2), \dots, cG_0(B_n))$, for any partition $\{B_i\}_{i=1,\dots,n}$ of S. With each λ_i follows prior $FDP(c, G_0)$, Blackwell and MacQueen (1973) show that the prior joint pde (probability density element) of $\boldsymbol{\lambda}' = (\lambda_1, \lambda_2, \dots, \lambda_n)$ has the following expression:

$$dF\left(\boldsymbol{\lambda}\middle|(c, G_0)\right) = \prod_{i=1}^{n} \frac{cG_0(d\lambda_i) + \sum_{j=1}^{i-1}\delta_{\lambda_j}(d\lambda_i)}{c+i-1},$$

where the Dirac delta function $\delta_{\lambda_j}(S) = \begin{cases}1 & \text{if } \lambda_j \in S,\\ 0 & \text{ow,}\end{cases}$. If a random sample **Y**=(Y_1, Y_2, …, Y_n), where each $Y_i|\lambda_i$ is iid, is further observed, then by Antoniak (1974) the posterior is a mixture of Ferguson-Dirichlet processes. In addition, it can be shown that, the posterior pde of $\boldsymbol{\lambda}' = (\lambda_1, \lambda_2, \dots, \lambda_n)$, after observing **Y**=(Y_1, Y_2, …, Y_n) with corresponding parameters $\boldsymbol{\lambda}' = (\lambda_1, \lambda_2, \dots, \lambda_n)$, is

$$dF\left(\boldsymbol{\lambda}\middle|\mathbf{Y},(c, G_0)\right) \propto \prod_{i=1}^{n} l(\lambda_i|Y_i)\frac{cG_0(d\lambda_i) + \sum_{j=1}^{i-1}\delta_{\lambda_j}(d\lambda_i)}{c+i-1}, \quad (2.1)$$

where $l(\lambda_i|Y_i)$ is the likelihood function of λ_i after observing Y_i. It is hard to estimate λ_i directly from the above expression in (2.1). To overcome such a problem, Escobar (1994) shows that expression (2.1) can also be expressed as the following conditional pde:

$$dF\left(\lambda_i\middle|\boldsymbol{\lambda}_{-i,}\mathbf{Y},(c, G_0)\right) \propto l(\lambda_i|Y_i)\, c\, G_0(d\lambda_i) + \sum_{j=1,j\neq i}^{n} l\left(\lambda_j\middle|Y_i\right)\delta_{\lambda_j}(d\lambda_i), \quad (2.2)$$

for i=1, 2, …, n, where $\boldsymbol{\lambda}'_{-i} = (\lambda_1, \dots, \lambda_{i-1}, \lambda_{i+1} \dots, \lambda_n)$. It can be further seen that expression (2.2) is equivalent to the following expression:

$$\lambda_i|\boldsymbol{\lambda}_{-i,}\mathbf{Y}\begin{cases}\sim h(\lambda_i|Y_i) & \text{with probability proportional to } c\, m(Y_i),\\ = \lambda_j & \text{with probability proportional to } l\left(\lambda_j\middle|Y_i\right), j\neq i,\end{cases} \quad (2.3)$$

for i=1, 2, …, n, where $h(\lambda_i|Y_i) \propto l(\lambda_i|Y_i)\, G_0(d\lambda_i)/d\lambda_i$, $m(Y_i) = \int l(\lambda_i|Y_i)\, G_0(d\lambda_i)$.

Hence, $h(\lambda_i|Y_i) = \frac{1}{m(Y_i)}\, l(\lambda_i|Y_i)\, G_0(d\lambda_i)/d\lambda_i$ is the posterior p.d.f. of λ_i when the prior is G_0. Notice that the conditional distribution of λ_i in (2.3) depend on $\mathbf{Y}$ only through Y_i. Therefore, expression $\lambda_i|\boldsymbol{\lambda}_{-i,}\mathbf{Y}$ in (2.3) can also be expressed as $\lambda_i|\boldsymbol{\lambda}_{-i,}Y_i$.

3. Semi-parametric Bayesian inference

In this paper, we assume to have a random sample of size n from the von Mises-Fisher distribution with dimension p=3, concentration parameter κ and mean vector $\boldsymbol{\lambda}$. We further assume that $\boldsymbol{\lambda}$ has a Ferguson-Dirichlet process prior with concentration parameter c and base probability distribution G_0, which is also a von Mises-Fisher distribution with concentration parameter ν and mean vector $\boldsymbol{\xi}$. Specifically, we have the following prior and sampling model:

$$\boldsymbol{\lambda}_i|G \sim G \qquad i = 1, 2, \dots, n,$$
$$G \sim FDP(c, G_0)$$
$$G_0 \sim vMF_3\,(\boldsymbol{\xi}, \nu\,),$$
$$\mathbf{Y}_i|\boldsymbol{\lambda}_i, \kappa_i \sim vMF_3\,(\boldsymbol{\lambda}_i, \kappa_i\,),\ \ i = 1, 2, \dots, n.$$

That is, $\mathbf{Y}_1|\boldsymbol{\lambda}_1, \kappa_1, \dots, \mathbf{Y}_n|\boldsymbol{\lambda}_n, \kappa_n$ are independent random variables, each $\mathbf{Y}_i|\boldsymbol{\lambda}_i, \kappa_i$ follows von Mises-Fisher distribution with concentration parameter κ_i and mean vector $\boldsymbol{\lambda}_i$; $\boldsymbol{\lambda}_1|G, \dots, \boldsymbol{\lambda}_n|G$ follow a distribution G, which is from a Ferguson-Dirichlet process with concentration parameter c and base distribution G_0; G_0 follows a von Mises-Fisher distribution with concentration parameter ν and mean vector $\boldsymbol{\xi}$.

The posterior p.d.e. of $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n$ can be expressed as

$$dF\big(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n\big|\mathbf{Y}, (\kappa_1, \dots, \kappa_n), (c, G_0)\big) \propto \prod_{i=1}^{n} f_3(\mathbf{Y}_i|\boldsymbol{\lambda}_i, \kappa_i)\frac{cG_0(d\boldsymbol{\lambda}_i) + \sum_{j=1}^{i-1}\delta_{\lambda_j}(d\boldsymbol{\lambda}_i)}{c + i - 1}, \qquad (3.1)$$

where $\mathbf{Y}' = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)$ and $f_3(\mathbf{Y}_i|\boldsymbol{\lambda}_i, \kappa_i)$ is the p.d.f. of $\mathbf{Y}_i$, which has a von Mises-Fisher distribution with parameters $\boldsymbol{\lambda}_i$ and κ_i. That is, $\mathbf{Y}_i|\boldsymbol{\lambda}_i, \kappa_i \sim vMF_3\,(\boldsymbol{\lambda}_i, \kappa_i\,)$.

As shown by Escobar (1994), (3.1) can also be expressed as

$$dF\big(\boldsymbol{\lambda}_i\big|\boldsymbol{\lambda}_{-i}, \mathbf{Y}, (\kappa_1, \dots, \kappa_n), (c, G_0)\big) \propto f_3(\mathbf{Y}_i|\boldsymbol{\lambda}_i, \kappa_i)c\,G_0(d\boldsymbol{\lambda}_i) + \sum_{j=1, j\neq i}^{n} f_3\big(\mathbf{Y}_i\big|\boldsymbol{\lambda}_j, \kappa_i\big)\,\delta_{\lambda_j}(d\boldsymbol{\lambda}_i). \quad (3.2)$$

Expression (3.2) is equivalent to the following expression:

$$\boldsymbol{\lambda}_i|\boldsymbol{\lambda}_{-i,}\mathbf{Y} \begin{cases} \sim h_3(\boldsymbol{\lambda}_i|\kappa_i, \mathbf{Y}_i) & \text{with probability proportional to } c\ m(\mathbf{Y}_i), \\ = \boldsymbol{\lambda}_j & \text{with probability proportional to } f_3\big(\mathbf{Y}_i\big|\boldsymbol{\lambda}_j, \kappa_i\big), j \neq i, \end{cases} \qquad (3.3)$$

for i=1, 2, …, n, where $h_3(\boldsymbol{\lambda}_i|\kappa_i,\mathbf{Y}_i) \propto f_3(\mathbf{Y}_i|\boldsymbol{\lambda}_i,\kappa_i)\, G_0(d\boldsymbol{\lambda}_i)/d\boldsymbol{\lambda}_i$, is the posterior p.d.f. of $\boldsymbol{\lambda}_i$ and $m(\mathbf{Y}_i) = \int f_3(\mathbf{Y}_i|\boldsymbol{\lambda}_i,\kappa_i)\, G_0(d\boldsymbol{\lambda}_i)$ is the normalized constant for the posterior p.d.f. of $\boldsymbol{\lambda}_i$.

In using polar coordinates, we let $\mathbf{Y}_i = \left(\sin\theta_i^{(1)}\cos\theta_i^{(2)}, \sin\theta_i^{(1)}\sin\theta_i^{(2)}, \cos\theta_i^{(1)}\right)$, $\boldsymbol{\lambda}_i = \left(\sin\alpha_i^{(1)}\cos\alpha_i^{(2)}, \sin\alpha_i^{(1)}\sin\alpha_i^{(2)}, \cos\alpha_i^{(1)}\right)$, and $\boldsymbol{\xi} = \left(\sin\beta^{(1)}\cos\beta^{(2)}, \sin\beta^{(1)}\sin\beta^{(2)}, \cos\beta^{(1)}\right)$. Then the posterior p.d.f. $h(\boldsymbol{\lambda}_i|\kappa_i,\mathbf{Y}_i)$ of $\boldsymbol{\lambda}_i$, in terms of polar coordinates, is

$$h_s\left(\left(\alpha_i^{(1)},\alpha_i^{(2)}\right)\middle|\left(\theta_i^{(1)},\theta_i^{(2)}\right),\kappa_i\right) \propto$$

$$f_s\left(\left(\theta_i^{(1)},\theta_i^{(2)}\right)\middle|\left(\alpha_i^{(1)},\alpha_i^{(2)}\right),\kappa_i\right) g_s\left(\left(\alpha_i^{(1)},\alpha_i^{(2)}\right)\middle|\left(\beta^{(1)},\beta^{(2)}\right),\nu\right) =$$

$$\frac{\kappa_i}{4\pi\sinh\kappa_i}\exp\{\kappa_i\left[\sin\theta_i^{(1)}\sin\alpha_i^{(1)}\cos(\theta_i^{(2)}-\alpha_i^{(2)}\right)+\cos\theta_i^{(1)}\cos\alpha_i^{(1)}]\}\sin\theta_i^{(1)}\ \cdot$$

$$\frac{\nu}{4\pi\sinh\nu}\exp\{\nu\left[\sin\alpha_i^{(1)}\sin\beta^{(1)}\cos(\alpha_i^{(2)}-\beta^{(2)}\right)+\cos\alpha_i^{(1)}\cos\beta^{(1)}]\}\sin\alpha_i^{(1)} =$$

$$\frac{\kappa_i\nu\sin\theta_i^{(1)}}{16\pi^2\sinh\kappa_i\sinh\nu}\exp\{\sin\alpha_i^{(1)}\left[\kappa_i\sin\theta_i^{(1)}\cos(\theta_i^{(2)}-\alpha_i^{(2)}\right)+\nu\sin\beta^{(1)}\cos(\alpha_i^{(2)}-\beta^{(2)})]+$$
$$\cos\alpha_i^{(1)}[\kappa_i\cos\theta_i^{(1)}+\nu\cos\beta^{(1)}]\}\sin\alpha_i^{(1)} =$$

$$\frac{\kappa_i\nu\sin\theta_i^{(1)}}{16\pi^2\sinh\kappa_i\sinh\nu}\exp\{\sin\alpha_i^{(1)}[\cos\alpha_i^{(2)}(\kappa_i\sin\theta_i^{(1)}\cos\theta_i^{(2)}+\nu\sin\beta^{(1)}\cos\beta^{(2)})+$$
$$\sin\alpha_i^{(2)}(\kappa_i\sin\theta_i^{(1)}\sin\theta_i^{(2)}+\nu\sin\beta^{(1)}\sin\beta^{(2)})]+$$
$$\cos\alpha_i^{(1)}(\kappa_i\cos\theta_i^{(1)}+\nu\cos\beta^{(1)})\}\sin\alpha_i^{(1)} =$$

$$\frac{\kappa_i\nu\sin\theta_i^{(1)}}{16\pi^2\sinh\kappa_i\sinh\nu}\exp\{\sin\alpha_i^{(1)}[\cos\alpha_i^{(2)}(\nu_i\sin\gamma_i^{(1)}\cos\gamma_i^{(2)})+$$
$$\sin\alpha_i^{(2)}(\nu_i\sin\gamma_i^{(1)}\sin\gamma_i^{(2)})]+\cos\alpha_i^{(1)}(\nu_i\cos\gamma_i^{(1)})\}\sin\alpha_i^{(1)},$$ where

$$\nu_i = \sqrt{\{\kappa_i^2+\nu^2+2\,\kappa_i\nu\left[\sin\theta_i^{(1)}\sin\beta^{(1)}\cos(\theta_i^{(2)}-\beta^{(2)}\right)+\cos\theta_i^{(1)}\cos\beta^{(1)}]\}},\ \cos\gamma_i^{(1)} =$$

$$\frac{1}{\nu_i}(\kappa_i\cos\theta_i^{(1)}+\nu\cos\beta^{(1)}),\ \sin\gamma_i^{(2)} = \frac{(\kappa_i\sin\theta_i^{(1)}\sin\theta_i^{(2)}+\nu\sin\beta^{(1)}\sin\beta^{(2)})}{\sqrt{\nu_i{}^2-(\kappa_i\cos\theta_i^{(1)}+\nu\cos\beta^{(1)})^2}},\ \text{and}\ \cos\gamma_i^{(2)} =$$

$$\frac{(\kappa_i\sin\theta_i^{(1)}\cos\theta_i^{(2)}+\nu\sin\beta^{(1)}\cos\beta^{(2)})}{\sqrt{\nu_i{}^2-(\kappa_i\cos\theta_i^{(1)}+\nu\cos\beta^{(1)})^2}}.$$

Hence, the posterior p.d.f. $h(\boldsymbol{\lambda}_i|\kappa_i, \mathbf{Y}_i)$ of $\boldsymbol{\lambda}_i$, in terms of polar coordinates, is

$$h_s\left(\left(\alpha_i^{(1)}, \alpha_i^{(2)}\right) \middle| \left(\theta_i^{(1)}, \theta_i^{(2)}\right), \kappa_i\right) =$$

$$\frac{\nu_i}{4\pi \sinh \nu_i} \exp\{\nu_i \left[\sin \alpha_i^{(1)} \sin \gamma_i^{(1)} \cos(\alpha_i^{(2)} - \gamma_i^{(2)}\right) + \cos \alpha_i^{(1)} \cos \gamma_i^{(1)}]\} \sin \alpha_i^{(1)},$$

where $\nu_i = \sqrt{\{\kappa_i^2 + \nu^2 + 2\,\kappa_i \nu \left[\sin \theta_i^{(1)} \sin \beta^{(1)} \cos(\theta_i^{(2)} - \beta^{(2)}\right) + \cos \theta_i^{(1)} \cos \beta^{(1)}]\}}$,

$\gamma_i^{(1)} =$
$\cos^{-1}[\frac{1}{\nu_i}\left(\kappa_i \cos \theta_i^{(1)} + \nu \cos \beta^{(1)}\right)]$, and $\gamma_i^{(2)} = \cos^{-1} \frac{(\kappa_i \sin \theta_i^{(1)} \cos \theta_i^{(2)} + \nu \sin \beta^{(1)} \cos \beta^{(2)})}{\sqrt{\nu_i{}^2 - (\kappa_i \cos \theta_i^{(1)} + \nu \cos \beta^{(1)})^2}}$.

Therefore, the posterior distribution of $\boldsymbol{\lambda}_i$ or $\left(\alpha_i^{(1)}, \alpha_i^{(2)}\right)'$ is still a von Mises-Fisher

distribution with updated parameters $\nu_i, \left(\gamma_i^{(1)}, \gamma_i^{(2)}\right)$. In short, if the prior

$\left(\alpha_i^{(1)}, \alpha_i^{(2)}\right) \sim \mathrm{vMF}\left(\beta^{(1)}, \beta^{(2)}; \nu\right)$ and data

$(\theta_i^{(1)}, \theta_i^{(2)})|(\alpha_i^{(1)}, \alpha_i^{(2)}; \kappa_i) \sim \mathrm{vMF}\left(\alpha_i^{(1)}, \alpha_i^{(2)}; \kappa_i\right)$, then the posterior

$\left(\alpha_i^{(1)}, \alpha_i^{(2)}\right) |(\theta_i^{(1)}, \theta_i^{(2)}) \sim \mathrm{vMF}\left(\gamma_i^{(1)}, \gamma_i^{(2)}; \nu_i\right)$.

In the next section, we give the predictive density based on the observed data.

4. Bayesian predictive density

Given the random sample $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$, the posterior predictive density of $\mathbf{Y}_{n+1}$ can be expressed as

$$f_3\left(\mathbf{y}_{n+1} \middle| (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)\right) =$$
$$\int f_3\left(\mathbf{y}_{n+1} \middle| (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n), (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)\right) h\left((\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n) \middle| (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)\right) d\boldsymbol{\lambda}_1 \dots d\boldsymbol{\lambda}_n,$$

where $h\left((\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n) \middle| (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)\right)$ is the posterior joint p.d.f. of $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n$ and $f_3\left(\mathbf{y}_{n+1} \middle| (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n), (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)\right)$ is the posterior predictive density function of $\mathbf{y}_{n+1}$ when $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n$ are also known. However,

$$f_3\left(\mathbf{y}_{n+1} \middle| (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n), (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)\right) =$$
$$\int f_3\left(\mathbf{y}_{n+1} \middle| \boldsymbol{\lambda}_{n+1}, (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n), (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)\right) dF\left(\boldsymbol{\lambda}_{n+1} \middle| (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n), (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)\right) =$$
$$\int f_3(\mathbf{y}_{n+1}|\boldsymbol{\lambda}_{n+1}) dF\left(\boldsymbol{\lambda}_{n+1} \middle| (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n)\right) = \int f_3(\mathbf{y}_{n+1}|\boldsymbol{\lambda}_{n+1})\{\frac{c}{c+n} G_0(d\boldsymbol{\lambda}_{n+1}) +$$

$\frac{1}{c+n}\sum_{j=1}^{n} \delta_{\boldsymbol{\lambda}_j}(d\boldsymbol{\lambda}_{n+1})$ }. By using polar coordinates with

$\mathbf{Y}_i = \left(\sin\theta_i^{(1)}\cos\theta_i^{(2)}, \sin\theta_i^{(1)}\sin\theta_i^{(2)}, \cos\theta_i^{(1)}\right), \boldsymbol{\lambda}_i =$

$\left(\sin\alpha_i^{(1)}\cos\alpha_i^{(2)}, \sin\alpha_i^{(1)}\sin\alpha_i^{(2)}, \cos\alpha_i^{(1)}\right)$, and $\boldsymbol{\xi} =$

$\left(\sin\beta^{(1)}\cos\beta^{(2)}, \sin\beta^{(1)}\sin\beta^{(2)}, \cos\beta^{(1)}\right)$,

$f_3\left(\mathbf{y}_{n+1}\middle|(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n), (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)\right) = \int f_3(\mathbf{y}_{n+1}|\boldsymbol{\lambda}_{n+1}) dF\left(\boldsymbol{\lambda}_{n+1}\middle|(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n)\right) =$

$\frac{c}{c+n}\int \frac{\kappa_{n+1}}{4\pi\sinh\kappa_{n+1}} \exp\{\kappa_{n+1}\left[\sin\theta_{n+1}^{(1)}\sin\alpha_{n+1}^{(1)}\cos(\theta_{n+1}^{(2)} - \alpha_{n+1}^{(2)}\right) +$

$\cos\theta_{n+1}^{(1)}\cos\alpha_{n+1}^{(1)}]\}\sin\theta_{n+1}^{(1)} \cdot \frac{\nu}{4\pi\sinh\nu}\exp\{\nu\left[\sin\alpha_{n+1}^{(1)}\sin\beta^{(1)}\cos(\alpha_{n+1}^{(2)} - \beta^{(2)}\right) +$

$\cos\alpha_{n+1}^{(1)}\cos\beta^{(1)}]\}\sin\alpha_{n+1}^{(1)}\ d\alpha_{n+1}^{(1)} d\,\alpha_{n+1}^{(2)} +$

$\frac{1}{c+n}\sum_{j=1}^{n} \frac{\kappa_{n+1}}{4\pi\sinh\kappa_{n+1}}\exp\{\kappa_{n+1}\left[\sin\theta_{n+1}^{(1)}\sin\alpha_j^{(1)}\cos(\theta_{n+1}^{(2)} - \alpha_j^{(2)}\right) +$

$\cos\theta_{n+1}^{(1)}\cos\alpha_j^{(1)}]\}\sin\theta_{n+1}^{(1)} =$

$\frac{c}{c+n}\ \frac{\kappa_{n+1}\sin\theta_{n+1}^{(1)}}{4\pi\sinh\kappa_{n+1}}\ \frac{\nu}{4\pi\sinh\nu}\ \frac{4\pi\sinh\nu_{n+1}}{\nu_{n+1}} + \frac{1}{c+n}\sum_{j=1}^{n} f_s\left((\theta_{n+1}^{(1)}, \theta_{n+1}^{(2)})\middle|\left(\alpha_j^{(1)}, \alpha_j^{(2)}; \kappa_{n+1}\right)\right) =$

$\frac{c}{c+n}\ \frac{\nu\,\kappa_{n+1}\sin\theta_{n+1}^{(1)}}{4\pi\ \ \nu_{n+1}\sinh\kappa_{n+1}}\ \frac{\sinh\nu_{n+1}}{\sinh\nu} + \frac{1}{c+n}\sum_{j=1}^{n} f_s\left((\theta_{n+1}^{(1)}, \theta_{n+1}^{(2)})\middle|\left(\alpha_j^{(1)}, \alpha_j^{(2)}; \kappa_{n+1}\right)\right)$, where

$\nu_{n+1} =$

$\sqrt{\{\kappa_{n+1}^2 + \nu^2 + 2\,\kappa_{n+1}\nu\left[\sin\theta_{n+1}^{(1)}\sin\beta^{(1)}\cos(\theta_{n+1}^{(2)} - \beta^{(2)}\right) + \cos\theta_{n+1}^{(1)}\cos\beta^{(1)}]\}}$, and

$f_s\left((\theta_{n+1}^{(1)}, \theta_{n+1}^{(2)})\middle|\left(\alpha_j^{(1)}, \alpha_j^{(2)}; \kappa_{n+1}\right)\right)$ is the von Mises-Fisher p.d.f. of $\mathbf{Y}_{n+1} =$

$\left(\sin\theta_{n+1}^{(1)}\cos\theta_{n+1}^{(2)}, \sin\theta_{n+1}^{(1)}\sin\theta_{n+1}^{(2)}, \cos\theta_{n+1}^{(1)}\right)$ with mean

$\boldsymbol{\lambda}_j = \left(\sin\alpha_j^{(1)}\cos\alpha_j^{(2)}, \sin\alpha_j^{(1)}\sin\alpha_j^{(2)}, \cos\alpha_j^{(1)}\right)$. Hence, the posterior predictive

density of $\mathbf{Y}_{n+1}$ can be expressed as $f_3\left(\mathbf{y}_{n+1}\middle|(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)\right) = \int \left[\frac{c}{c+n} f_3(\mathbf{y}_{n+1}) +\right.$

$\left.\frac{1}{c+n}\sum_{j=1}^{n} f_3\left(\mathbf{y}_{n+1}\middle|\boldsymbol{\lambda}_j; \kappa_{n+1}\right)\right] h\left((\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n)\middle|(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)\right) d\boldsymbol{\lambda}_1 \dots d\boldsymbol{\lambda}_n =$

$\frac{c}{c+n} f_3(\mathbf{y}_{n+1}) +$

$\frac{1}{c+n}\sum_{j=1}^{n}\int f_3\left(\mathbf{y}_{n+1}\middle|\boldsymbol{\lambda}_j;\kappa_{n+1}\right) h\left((\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\dots,\boldsymbol{\lambda}_n)\middle|(\mathbf{y}_1,\mathbf{y}_2,\dots,\mathbf{y}_n)\right) d\boldsymbol{\lambda}_1 \dots d\boldsymbol{\lambda}_n$, where

$f_3(\mathbf{y}_{n+1}) = \frac{\nu\,\kappa_{n+1}\sin\theta_{n+1}^{(1)}\ \sinh\nu_{n+1}}{4\pi\ \ \nu_{n+1}\sinh\kappa_{n+1}\ \sinh\nu}$ is the prior predictive density of $\mathbf{Y}_{n+1}$.

5. Bayesian two-sample test

Given that the first random sample $\mathbf{X}_1,\mathbf{X}_2,\dots,\mathbf{X}_n$ is from a von Mises-Fisher distribution with mean vector $\boldsymbol{\lambda}_1$ and concentration parameter κ_1 and the second random sample $\mathbf{Y}_1,\mathbf{Y}_2,\dots,\mathbf{Y}_m$ is from another von Mises-Fisher distribution with mean vector $\boldsymbol{\lambda}_2$ and concentration parameter κ_2, we want use our semi-parametric method to test whether $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2$. Formally, let $\mathbf{X}_i|\boldsymbol{\lambda}_1,\ \kappa_1 \sim \mathrm{vMF}_3(\boldsymbol{\lambda}_1,\ \kappa_1), \quad i = 1,\dots,n,$

$\mathbf{Y}_j|\boldsymbol{\lambda}_2,\ \kappa_2 \sim \mathrm{vMF}_3(\boldsymbol{\lambda}_2,\ \kappa_2), \quad j = 1,\dots,m,$

$\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2|G \sim G$

$G \sim \mathrm{FDP}(c, G_0)$

$G_0 \sim \mathrm{vMF}_3(\boldsymbol{\xi},\nu)$

$\boldsymbol{\lambda}_i = \left(\sin\alpha_i^{(1)}\cos\alpha_i^{(2)}, \sin\alpha_i^{(1)}\sin\alpha_i^{(2)}, \cos\alpha_i^{(1)}\right), \quad i = 1, 2,$ and

$\boldsymbol{\xi} = \left(\sin\beta^{(1)}\cos\beta^{(2)}, \sin\beta^{(1)}\sin\beta^{(2)}, \cos\beta^{(1)}\right),$

$$\Omega = \{\boldsymbol{\omega} = (\boldsymbol{\omega}_1;\ \boldsymbol{\omega}_2) = \left(\alpha_1^{(1)},\alpha_1^{(2)};\ \alpha_2^{(1)},\alpha_2^{(2)}\right) : 0 \leqq \alpha_1^{(1)},\alpha_2^{(1)} \leqq \pi, 0 \leqq \alpha_1^{(2)},\alpha_2^{(2)} \leqq 2\pi\}$$

$$\Omega_0 = \left\{\boldsymbol{\omega} \in \Omega : \left(\alpha_1^{(1)},\alpha_1^{(2)}\right) = \left(\alpha_2^{(1)},\alpha_2^{(2)}\right)\right\} \text{ and}$$

$$\Omega_1 = \left\{\boldsymbol{\omega} \in \Omega : \left(\alpha_1^{(1)},\alpha_1^{(2)}\right) \neq \left(\alpha_2^{(1)},\alpha_2^{(2)}\right)\right\}$$

We want to test $H_0: \boldsymbol{\omega} \in \Omega_0$ against the alternative $H_1: \boldsymbol{\omega} \in \Omega_1$ in this section.

Assume that $(\theta_i^{(1)},\theta_i^{(2)})$ and $(\phi_j^{(1)},\phi_j^{(2)})$ are the corresponding polar coordinates for $\mathbf{X}_i$ and $\mathbf{Y}_j$, respectively, $p_i = P(H_i)$ is the prior probability that H_i is true for i=1, 2, and $h_i(\boldsymbol{\omega})$ is the conditional probability density of $\boldsymbol{\omega}$ given H_i is true for i=0,1. The likelihood function of $\boldsymbol{\omega}$ is

$L(\boldsymbol{\omega}|\mathbf{X}_1,\mathbf{X}_2,\dots,\mathbf{X}_n,\mathbf{Y}_1,\mathbf{Y}_2,\dots,\mathbf{Y}_m) =$

$\prod_{i=1}^{n} f_s\left((\theta_i^{(1)}, \theta_i^{(2)}) \middle| \left(\alpha_1^{(1)}, \alpha_1^{(2)}\right), \kappa_1\right) \prod_{j=1}^{m} f_s\left((\phi_j^{(1)}, \phi_j^{(2)}) \middle| \left(\alpha_2^{(1)}, \alpha_2^{(2)}\right), \kappa_2\right)$. By Bayes theorem, the posterior probability that H_i is true is

$$P(H_i|\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n, \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m) = \frac{P(H_i)P(\mathbf{X}_1,\mathbf{X}_2,\dots,\mathbf{X}_n,\mathbf{Y}_1,\mathbf{Y}_2,\dots,\mathbf{Y}_m|H_i)}{\sum_{j=0}^{1} P(H_j)P(\mathbf{X}_1,\mathbf{X}_2,\dots,\mathbf{X}_n,\mathbf{Y}_1,\mathbf{Y}_2,\dots,\mathbf{Y}_m|H_j)}$$

$$= \frac{P(H_i) \int P(\mathbf{X}_1,\mathbf{X}_2,\dots,\mathbf{X}_n,\mathbf{Y}_1,\mathbf{Y}_2,\dots,\mathbf{Y}_m|H_i,\boldsymbol{\omega})p(\boldsymbol{\omega}|H_i)d\boldsymbol{\omega}}{\sum_{j=0}^{1} P(H_j) \int P(\mathbf{X}_1,\mathbf{X}_2,\dots,\mathbf{X}_n,\mathbf{Y}_1,\mathbf{Y}_2,\dots,\mathbf{Y}_m|H_j,\boldsymbol{\omega})p(\boldsymbol{\omega}|H_j)d\boldsymbol{\omega}}$$

$$= \frac{p_i \int_{\Omega_i} L(\boldsymbol{\omega}|\mathbf{X}_1,\mathbf{X}_2,\dots,\mathbf{X}_n,\mathbf{Y}_1,\mathbf{Y}_2,\dots,\mathbf{Y}_m)h_i(\boldsymbol{\omega})d\boldsymbol{\omega}}{\sum_{j=0}^{1} p_j \int_{\Omega_j} L(\boldsymbol{\omega}|\mathbf{X}_1,\mathbf{X}_2,\dots,\mathbf{X}_n,\mathbf{Y}_1,\mathbf{Y}_2,\dots,\mathbf{Y}_m)h_j(\boldsymbol{\omega})d\boldsymbol{\omega}}.$$

From the given Ferguson-Dirichlet prior distribution and Blackwell and MacQueen (1973), it can be seen that $p_0 = 1/(c+1)$, $p_1 = c/(c+1)$. Hence,

$$P(H_0|\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n, \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m) =$$

$$[(1/(c+1)) \int_{\Omega_0} \prod_{i=1}^{n} f_s\left(\left(\theta_i^{(1)}, \theta_i^{(2)}\right) \middle| \left(\alpha_1^{(1)}, \alpha_1^{(2)}\right), \kappa_1\right) \prod_{j=1}^{m} f_s\left(\left(\phi_j^{(1)}, \phi_j^{(2)}\right) \middle| \left(\alpha_2^{(1)}, \alpha_2^{(2)}\right), \kappa_2\right)$$

$$g_s\left(\left(\alpha_1^{(1)}, \alpha_1^{(2)}\right) \middle| (\beta^{(1)}, \beta^{(2)}), \nu\right) g_s\left(\left(\alpha_2^{(1)}, \alpha_2^{(2)}\right) \middle| (\beta^{(1)}, \beta^{(2)}), \nu\right) I_0^{-1} d\alpha_1^{(1)} d\alpha_1^{(2)} d\alpha_2^{(1)} d\alpha_2^{(2)}]$$

/

$$\{[(1/(c+1)) \int_{\Omega_0} \prod_{i=1}^{n} f_s\left(\left(\theta_i^{(1)}, \theta_i^{(2)}\right) \middle| \left(\alpha_1^{(1)}, \alpha_1^{(2)}\right), \kappa_1\right) \prod_{j=1}^{m} f_s\left(\left(\phi_j^{(1)}, \phi_j^{(2)}\right) \middle| \left(\alpha_2^{(1)}, \alpha_2^{(2)}\right), \kappa_2\right)$$

$$g_s\left(\left(\alpha_1^{(1)}, \alpha_1^{(2)}\right) \middle| (\beta^{(1)}, \beta^{(2)}), \nu\right) g_s\left(\left(\alpha_2^{(1)}, \alpha_2^{(2)}\right) \middle| (\beta^{(1)}, \beta^{(2)}), \nu\right) I_0^{-1} d\alpha_1^{(1)} d\alpha_1^{(2)} d\alpha_2^{(1)} d\alpha_2^{(2)}]$$

$$+ [(c/(c+1)) \int_{\Omega_1} \prod_{i=1}^{n} f_s\left(\left(\theta_i^{(1)}, \theta_i^{(2)}\right) \middle| \left(\alpha_1^{(1)}, \alpha_1^{(2)}\right), \kappa_1\right) \prod_{j=1}^{m} f_s\left(\left(\phi_j^{(1)}, \phi_j^{(2)}\right) \middle| \left(\alpha_2^{(1)}, \alpha_2^{(2)}\right), \kappa_2\right)$$

$$g_s\left(\left(\alpha_1^{(1)}, \alpha_1^{(2)}\right) \middle| (\beta^{(1)}, \beta^{(2)}), \nu\right) g_s\left(\left(\alpha_2^{(1)}, \alpha_2^{(2)}\right) \middle| (\beta^{(1)}, \beta^{(2)}), \nu\right) I_1^{-1} d\alpha_1^{(1)} d\alpha_1^{(2)} d\alpha_2^{(1)} d\alpha_2^{(2)}]\}$$

, where $I_i =$

$$\int_{\Omega_i} g_s\left(\left(\alpha_1^{(1)}, \alpha_1^{(2)}\right) \middle| (\beta^{(1)}, \beta^{(2)}), \nu\right) g_s\left(\left(\alpha_2^{(1)}, \alpha_2^{(2)}\right) \middle| (\beta^{(1)}, \beta^{(2)}), \nu\right) d\alpha_1^{(1)} d\alpha_1^{(2)} d\alpha_2^{(1)} d\alpha_2^{(2)},$$

for i=0,1. By the argument in Section 4, a von Mises-Fisher prior p.d.f. with a von Mises-Fisher data still has a von Mises-Fisher posterior p.d.f., it would not be hard to compute the posterior probability that H_i is true. In practice, we accept H_i if the posterior probability that H_i is true is more 1/2. Otherwise, we reject H_i.

6. Conclusions

In this paper, we first show that the von Mises-Fisher prior distribution is conjugate for the von Mises-Fisher sample data on sphere. We then give semi-parametric Bayesian methods to make statistical inferences. In particular, Bayesian predictive density and Bayesian two-sample test methods are given.

Acknowledgements

This research is partly supported by the Ministry of Science and Technology, Taiwan.

References

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *The Annals of Statistics*, 2(6), 1152-1174.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Polya urn schemes, *The Annals of Statistics*, 1, 353-355.

Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior, *Journal of the American Statistical Association*, 89, 268-277.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *The Annals of Statistics*, 1, 209-230.

Fisher, N. I., Lewis, T., and Embleton, B. J. J. (1987), *Statistical analysis of spherical data*, Cambridge University Press.

Ghosh, K., Jammalamadaka, S. R., and Tiwari, R. C. (2003), Semiparametric Bayesian techniques for problems in circular data, *Journal of Applied Statistics*, 30, 145-161.

Jammalamadaka, S. R. and SenGupta, A. (2001), *Topics in circular statistics*, World Scientific Press.

Mardia, K. V. (1972), *Statistics of directional data*, Academic Press.

Nunez-Antonio, G. and Gutierrez-Pena, E. (2005), A Bayesian analysis of directional data using the von Mises-Fisher distribution, *Communications in Statistics – Simulation and computation*, 34, 989-999.