Bayesian Analysis (2017) **12**, Number 1, pp. 113–133

The General Projected Normal Distribution of Arbitrary Dimension: Modeling and Bayesian Inference

Daniel Hernandez-Stumpfhauser*, F. Jay Breidt†, and Mark J. van der Woerd‡

Abstract. The general projected normal distribution is a simple and intuitive model for directional data in any dimension: a multivariate normal random vector divided by its length is the projection of that vector onto the surface of the unit hypersphere. Observed data consist of the projections, but not the lengths. Inference for this model has been restricted to the two-dimensional (circular) case, using Bayesian methods with data augmentation to generate the latent lengths and a Metropolis-within-Gibbs algorithm to sample from the posterior. We describe a new parameterization of the general projected normal distribution that makes inference in any dimension tractable, including the important three-dimensional (spherical) case, which has not previously been considered. Under this new parameterization, the full conditionals of the unknown parameters have closed forms, and we propose a new slice sampler to draw the latent lengths without the need for rejection. Gibbs sampling with this new scheme is fast and easy, leading to improved Bayesian inference; for example, it is now feasible to conduct model selection among complex mixture and regression models for large data sets. Our parameterization also allows straightforward incorporation of covariates into the covariance matrix of the multivariate normal, increasing the ability of the model to explain directional data as a function of independent regressors. Circular and spherical cases are considered in detail and illustrated with scientific applications. For the circular case, seasonal variation in time-of-day departures of anglers from recreational fishing sites is modeled using covariates in both the mean vector and covariance matrix. For the spherical case, we consider paired angles that describe the relative positions of carbon atoms along the backbone chain of a protein. We fit mixtures of general projected normals to these data, with the best-fitting mixture accurately describing biologically meaningful structures including helices, β-sheets, and coils and turns. Finally, we show via simulation that our methodology has satisfactory performance in some 10-dimensional and 50-dimensional problems.

Keywords: circular data, directional data, Gibbs sampler, Markov chain Monte Carlo, protein structure analysis, spherical data.

*Department of Biostatistics, The University of North Carolina at Chapel Hill, danielhs@live.unc.edu

†Department of Statistics, Colorado State University, jbreidt@stat.colostate.edu

‡Department of Biochemistry & Molecular Biology, Colorado State University, Mark.van_der_Woerd@colostate.edu

 DOI: 10.1214/15-BA989

1 Introduction

1.1 Directional data

Directional data arise in many scientific disciplines. Circular data are encountered in studies of wind directions (Breckling, 1989; Nuñez-Antonio et al., 2015), geological layering (Oliveira et al., 2012) and animal navigation (Schmidt-Koenig, 1965; Batschelet, 1981; Ferreira et al., 2008; Oliveira et al., 2012). Temporal series with seasonality, such as monthly cases of acute lymphoblastic leukemia (Gao et al., 2007), may also be treated as circular data. Spherical data arise in applications such as earthquake epicenters, tectonic plate rotations (Chang, 1993), and high-energy cosmic ray sources on the celestial sphere (Mardia and Edwards, 1982; Ferreira et al., 2008), as well as in modeling of protein structures (Levitt, 1976; Oldfield and Hubbard, 1994). In information retrieval, each text document is represented by a vector of frequencies of word occurrences and "cosine similarity" between two documents is measured as the cosine of the angle between their respective vectors (the normalized inner product of the vectors). In "collaborative filtering", the similarities among users of an information database are computed as the cosine similarity of their preference vectors. Further, in microarray experiments, each gene is represented by a vector of expression values, with similarities between genes measured as the cosine similarity between their expression vectors. Treating the angles between such vectors as data leads to directional data of arbitrary dimension. See, for example, (Sarwar et al., 2001; Banerjee et al., 2005) and the references therein.

A circular observation can be regarded as a point on the unit circle or a unit vector in the plane. A spherical observation can be regarded as a point on the unit sphere or as a unit vector in space, and higher-dimensional directional data as points on the unit hypersphere or unit vectors in that space.

Many statistical models and methods exist for circular data; see the monographs (Fisher, 1995; Mardia and Jupp, 2000; Jammalamadaka and Sengupta, 2001; Pewsey et al., 2013), and (Wang and Gelfand, 2014) for recent spatio-temporal extensions. The options for spherical data are more sparse, but include the von Mises–Fisher family; see, for example, (Watson, 1983; Mardia and Jupp, 2000). Recent developments for directional data are often focused on flexible distributional modeling, including finite mixtures of von Mises distributions for hyperspheres of any dimension (Ferreira et al., 2008), Dirichlet process mixtures of von Mises distributions for circular data (Ghosh et al., 2003), Dirichlet process mixtures of triangular distributions for circular data (McVinish and Mengersen, 2008), log-spline distributions for hyperspheres of any dimension (Ferreira et al., 2008), Dirichlet process mixtures of normal distributions for circular data (Nuñez-Antonio et al., 2015), and projected normal distributions (Presnell et al., 1998; Nuñez-Antonio and Gutiérrez-Peña, 2005; Nuñez-Antonio et al., 2011; Hernandez-Stumpfhauser, 2012; Wang and Gelfand, 2013; Nuñez-Antonio et al., 2015) and their mixtures (Wang and Gelfand, 2014), the focus of this paper.

1.2 Projected normal distributions

Projected distributions are obtained by radial projection of distributions on the plane or space; most commonly, a multivariate normal random vector is divided by its length

to give the projection of that vector onto the surface of the unit hypersphere. For $k \geq 2$, the random unit vector $\boldsymbol{U} = \boldsymbol{X}/\|\boldsymbol{X}\|$ is distributed as $\mathcal{PN}_k(\boldsymbol{\mu}, \Sigma)$, a general projected normal distribution with parameters $\boldsymbol{\mu}$ and Σ, if $\boldsymbol{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$, a multivariate normal distribution. The unit vector $\boldsymbol{U}$ is interpreted as the directional random vector $\boldsymbol{\Theta}$ through appropriate choice of coordinate system; in the circular case, for example, the random direction Θ is obtained from $\boldsymbol{U} = (\cos\Theta, \sin\Theta)^T$ in polar coordinates. Projected normal distributions have been called off-set normal distributions by Mardia (1972), displaced normal by Kendall (1974) and angular normal by Watson (1983).

The distribution $\mathcal{PN}_k(\boldsymbol{\mu}, \Sigma)$ generalizes the projected normal distribution with identity covariance matrix, $\Sigma = \boldsymbol{I}$. The density $\mathcal{PN}_k(\boldsymbol{\mu}, \boldsymbol{I})$ is unimodal and symmetric about the direction of $\boldsymbol{\mu}$. It was used by Presnell et al. (1998) to introduce the Spherically Projected Multivariate Linear Model (SPMLM) for directional data, in which $\boldsymbol{\mu}$ is specified with a linear model. An EM algorithm was used to find maximum likelihood estimates. Bayesian inference for $\mathcal{PN}_k(\boldsymbol{\mu}, \boldsymbol{I})$ was considered by Nuñez-Antonio and Gutiérrez-Peña (2005), Nuñez-Antonio et al. (2011), Nuñez-Antonio et al. (2015), and Hernandez-Stumpfhauser (2012), showing in particular how to analyze various regression structures in $\boldsymbol{\mu}$. A key idea used in these papers is to augment the observed directional data $(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n)$ with the latent length variables $(r_1, r_2, \ldots, r_n)$. The joint density of $(r_i, \boldsymbol{\theta_i})$ can then be obtained using the Jacobian method and the multivariate normal density for $\boldsymbol{x}_i = r_i \boldsymbol{u}_i$.

Recently, Wang and Gelfand (2013) and Wang and Gelfand (2014) demonstrated how to handle the general projected normal, $\mathcal{PN}_2(\boldsymbol{\mu}, \Sigma)$ with $\Sigma \neq \boldsymbol{I}$, for the special case of circular data ($k = 2$). The density of Θ in $\boldsymbol{U} = (\cos\Theta, \sin\Theta)^T$ can be written explicitly as

$$p(\theta \mid \boldsymbol{\mu}, \Sigma) = \left(\frac{1}{2\pi A(\theta)}\right) |\Sigma|^{-\frac{1}{2}} \exp(C) \left\{1 + \frac{B(\theta)}{\sqrt{A(\theta)}} \frac{\Phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)}{\varphi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)}\right\} I_{[0,2\pi)}(\theta) \quad (1)$$

where $\boldsymbol{u}^T = (\cos\theta, \sin\theta)$, $A(\theta) = \boldsymbol{u}^T\Sigma^{-1}\boldsymbol{u}$, $B(\theta) = \boldsymbol{u}^T\Sigma^{-1}\boldsymbol{\mu}$ and $C = -\frac{1}{2}\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}$; $I_{(0,2\pi]}(\cdot)$ is an indicator function; and $\Phi(\cdot), \varphi(\cdot)$ are the standard normal distribution and density functions, respectively. Equation (1) is similar to an expression found in Pukkila and Rao (1988). It readily generalizes to higher dimensions.

The general projected density $\mathcal{PN}_k(\boldsymbol{\mu}, \Sigma)$ can be asymmetric or bimodal. If $\boldsymbol{\mu}$ is orthogonal to any of the eigenvectors of Σ then the distribution is symmetric. See the supplementary materials (Hernandez-Stumpfhauser et al., 2016) for details where we also provide, for the circular case, an analytic method to distinguish between symmetric unimodal and symmetric bimodal.

The distribution of $\boldsymbol{U} = \boldsymbol{X}/\|\boldsymbol{X}\|$ does not change if the random variable $\boldsymbol{X}$ is scaled by any $a > 0$, and so further constraints on the parameters are needed for identifiability. Without loss of generality Wang and Gelfand (2013) set

$$\Sigma = \begin{pmatrix} \tau^2 & \rho\tau \\ \rho\tau & 1 \end{pmatrix} \quad (2)$$

and $\boldsymbol{\mu} = (\mu_1, \mu_2)$. They place a bivariate normal prior $\mathcal{N}_2(\mathbf{0}, \lambda_0 \boldsymbol{I})$ on $\boldsymbol{\mu}$, a uniform prior $\mathcal{U}(-1, 1)$ on ρ and an inverse Gamma prior $\mathcal{IG}(a_\tau, b_\tau)$ on τ^2. Under this parameterization, no closed forms for the full conditional posteriors of τ^2 and ρ are available. Therefore, they draw τ^2 and ρ and the latent lengths $r_1, \ldots, r_n$ iteratively via Metropolis-Hastings within Gibbs.

2 General projected normal distribution $\mathcal{PN}_k(\boldsymbol{\mu}, \Sigma)$

We turn now to the general projected normal $\mathcal{PN}_k(\boldsymbol{\mu}, \Sigma)$ with arbitrary dimension k, which has not previously been considered. By using an alternative parameterization for the constrained covariance matrix, $\Sigma \neq \boldsymbol{I}$, we are able to obtain closed forms for all of the full conditionals of the parameters in the model. The new parameterization works in any dimension, including the important spherical case $k = 3$, and has the added benefit of allowing the incorporation of covariates into the covariance matrix. Further, we propose a new slice sampler to handle the drawing of the latent lengths.

2.1 Reparameterizing the general projected normal

As in the two-dimensional case in Wang and Gelfand (2013), we restrict one of the diagonal terms of the $k \times k$ covariance matrix Σ to be equal to one, without loss of generality, writing

$$\Sigma = \begin{pmatrix} \Gamma & \boldsymbol{\gamma} \\ \boldsymbol{\gamma}^T & 1 \end{pmatrix}$$

where $\boldsymbol{\gamma}$ is a $k-1$ vector of regression coefficients. Now Σ is non-negative definite (nnd) and hence a covariance matrix if and only if

$$\Sigma = \begin{pmatrix} \Sigma^* + \boldsymbol{\gamma}\boldsymbol{\gamma}^T & \boldsymbol{\gamma} \\ \boldsymbol{\gamma}^T & 1 \end{pmatrix} \tag{3}$$

where Σ^* is a $k-1$ by $k-1$ nnd covariance matrix. To see this, first assume that Σ is nnd and define the random vector $\boldsymbol{Z} = (\boldsymbol{I}, -\boldsymbol{\gamma})\boldsymbol{Y}$, where $\boldsymbol{Y}$ is a random vector with mean $\mathbf{0}$ and covariance matrix Σ. Then $\text{var}(\boldsymbol{Z}) = \Gamma - \boldsymbol{\gamma}\boldsymbol{\gamma}^T = \Sigma^*$ is nnd. Conversely, if Σ^* is nnd, let $(\boldsymbol{c}^T, d)^T \in I\!R^k$ be any non-zero vector with $\boldsymbol{c} \in I\!R^{k-1}$. Then

$$\begin{pmatrix} \boldsymbol{c}^T & d \end{pmatrix} \begin{pmatrix} \Sigma^* + \boldsymbol{\gamma}\boldsymbol{\gamma}^T & \boldsymbol{\gamma} \\ \boldsymbol{\gamma}^T & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{c} \\ d \end{pmatrix} = \boldsymbol{c}^T \Sigma^* \boldsymbol{c} + \left(\boldsymbol{\gamma}^T \boldsymbol{c} + d\right)^2 \geq 0,$$

hence Σ is nnd.

Write $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)^T$, $\boldsymbol{\mu}_{-k} = (\mu_1, \mu_2, \ldots, \mu_{k-1})^T$ and $(x_{1,i}, x_{2,i}, \ldots, x_{k,i})^T = r_i \boldsymbol{u}_i$, where $\|\boldsymbol{u}_i\| = 1$. Making use of spherical coordinates we have that associated with each $\boldsymbol{u}_i$ are $k-1$ angles $(\theta_{1,i}, \theta_{2,i}, \ldots, \theta_{k-1,i}) = \boldsymbol{\theta}_i$ where $\theta_{k-1,i} \in [0, 2\pi)$ and $\theta_{s,i} \in [0, \pi]$ for $s = 1, \ldots, k-2$. Hence we can write the joint density of $(r_i, \boldsymbol{\theta}_i)$ as

$$\begin{aligned} p\left(r_i, \boldsymbol{\theta}_i \mid \boldsymbol{\mu}, \Sigma\right) &= \left(\frac{1}{2\pi}\right)^{\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} r_i^{k-1} \exp\left\{-\frac{1}{2}\left(r_i \boldsymbol{u}_i - \boldsymbol{\mu}\right)^T \Sigma^{-1} \left(r_i \boldsymbol{u}_i - \boldsymbol{\mu}\right)\right\} \\ &= r_i^{k-1} p\left(x_{1,i}, \ldots, x_{k-1,i} \mid x_{k,i}, \boldsymbol{\mu}, \Sigma\right) p\left(x_{k,i} \mid \boldsymbol{\mu}, \Sigma\right) \end{aligned} \tag{4}$$

where $p(x_{1,i}, \ldots, x_{k-1,i} \mid x_{k,i}, \boldsymbol{\mu}, \Sigma)$ is a normal density with mean $\boldsymbol{\mu}_{-k} + \boldsymbol{\gamma}(x_{k,i} - \mu_k)$ and variance Σ^*. The density function $p(x_{k,i} \mid \boldsymbol{\mu}, \Sigma)$ is a normal density with mean μ_k and variance equal to 1. Closed forms for the full conditionals are obtained by placing normal priors on $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ and an inverse Wishart prior on Σ^*. From (4) we can obtain the projected normal density $p(\boldsymbol{\theta}_i \mid \boldsymbol{\mu}, \Sigma) = \int p(r_i, \boldsymbol{\theta}_i \mid \boldsymbol{\mu}, \Sigma) dr$. The area element on the unit sphere is $d\theta_{k-1,i} \prod_{s=1}^{k-2} [\sin(\theta_{s,i})]^{k-1-s} d\theta_{s,i}$.

Remark. It is also possible to reparameterize Σ to obtain closed forms for the full conditionals without the use of an inverse Wishart prior, by using the Cholesky decomposition approach to unconstrained parameterization of a covariance matrix (Pourahmadi, 1999) with a constraint on the variance of x_k equal to one. We omit the details.

2.2 Incorporating covariates in the general projected normal

When covariates are available for the ith case, they can be incorporated into $\boldsymbol{\mu}$, and they can also be incorporated into $\boldsymbol{\gamma}$, increasing the flexibility of the model. We use this flexibility in the model specification of Section 3.2.

Let $\boldsymbol{z}_{\mu,i}$ denote a p-dimensional vector of known covariates and $\boldsymbol{z}_{\gamma,i}$ denote a q-dimensional vector of known covariates. Then the general projected normal with covariates is $\boldsymbol{\Theta}_i \sim \mathcal{PN}_k(\boldsymbol{\mu}_i, \Sigma_i)$, where

$$\boldsymbol{\mu}_i^T = \boldsymbol{z}_{\mu,i}^T \left(\boldsymbol{\beta}_1 \quad \boldsymbol{\beta}_2 \quad \cdots \quad \boldsymbol{\beta}_k \right), \tag{5}$$

with each $\boldsymbol{\beta}_j$ a p-vector of unknown regression coefficients, and

$$\Sigma_i = \begin{pmatrix} \Sigma^* + \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T & \boldsymbol{\gamma}_i \\ \boldsymbol{\gamma}_i^T & 1 \end{pmatrix}, \quad \boldsymbol{\gamma}_i^T = \boldsymbol{z}_{\gamma,i}^T \left(\boldsymbol{\alpha}_1 \quad \boldsymbol{\alpha}_2 \quad \cdots \quad \boldsymbol{\alpha}_{k-1} \right), \tag{6}$$

with each $\boldsymbol{\alpha}_j$ a q-vector of unknown regression coefficients. Closed forms for the full conditionals are then obtained by placing an inverse Wishart prior on Σ^* and normal priors on the vectors $\boldsymbol{\beta}_j$ and $\boldsymbol{\alpha}_j$ instead of on $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$.

2.3 Slice sampling for latent lengths

To draw from the latent lengths $r_1, \ldots, r_n$ within the Gibbs sampler we make use of slice sampling (Neal, 2003). From (4) the full conditional distribution for each of the latent lengths is $p(r_i \mid \cdot) \propto r_i^{k-1} \exp\{-\frac{1}{2} A_i (r_i - \frac{B_i}{A_i})^2\}$ where $A_i = \boldsymbol{u}_i^T \Sigma^{-1} \boldsymbol{u}_i$ and $B_i = \boldsymbol{u}_i^T \Sigma^{-1} \boldsymbol{\mu}$. To sample from the latent length distribution we introduce a latent variable υ_i which has joint density with r_i given by

$$p\left(r_i, \upsilon_i \mid \boldsymbol{\mu}, \Sigma, \theta_i\right) \propto r_i^{k-1} I_{\left(0, \exp\left\{-\frac{1}{2} A_i \left(r_i - \frac{B_i}{A_i}\right)^2\right\}\right)}\left(\upsilon_i\right) I_{(0,\infty)}\left(r_i\right).$$

Then the full conditionals are

$$p\left(\upsilon_i \mid r_i, \boldsymbol{\mu}, \Sigma, \theta_i\right) \quad = \quad \mathcal{U}\left(0, \exp\left\{-\frac{1}{2} A_i \left(r_i - \frac{B_i}{A_i}\right)^2\right\}\right)$$

$$p\left(r_i \mid \upsilon_i, \boldsymbol{\mu}, \Sigma, \theta_i\right) \quad \propto \quad r_i^{k-1} I_{\left(\frac{B_i}{A_i}+\max\left\{-\frac{B_i}{A_i}, -\sqrt{\frac{-2\ln \upsilon_i}{A_i}}\right\}, \frac{B_i}{A_i}+\sqrt{\frac{-2\ln \upsilon_i}{A_i}}\right)}\left(r_i\right)$$

where it is easy to sample from the latter using the inverse cumulative distribution function technique. Thus we draw $\upsilon_i \sim \mathcal{U}(0, \exp\{-\frac{1}{2}A_i(r_i - \frac{B_i}{A_i})^2\})$ and independently $u_i \sim \mathcal{U}(0,1)$. Finally $r_i = \{(\varrho_{2i}^k - \varrho_{1i}^k)u_i + \varrho_{1i}^k\}^{1/k}$, where for $i = 1, 2, \ldots, n$,

$$\varrho_{1i} = \frac{B_i}{A_i} + \max\left\{-\frac{B_i}{A_i}, -\sqrt{\frac{-2\ln \upsilon_i}{A_i}}\right\}, \quad \varrho_{2i} = \frac{B_i}{A_i} + \sqrt{\frac{-2\ln \upsilon_i}{A_i}}.$$

The combination of closed forms for the full conditionals of the model parameters and slice sampling without rejection for the latent lengths makes Gibbs sampling fast and easy for the general projected normal. In Section 3, we describe the details for the circular case, illustrated by modeling time-of-day departures of anglers from recreational fishing sites using covariates in both the mean vector and covariance matrix. In Section 4, we describe analysis of spherical data, illustrated with an application to protein structure analysis.

3 Circular case

For the circular case, $x_{1,i} = r_i \cos\theta_i$, $x_{2,i} = r_i \sin\theta_i$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and (3) becomes

$$\Sigma = \begin{pmatrix} \sigma^2 + \gamma^2 & \gamma \\ \gamma & 1 \end{pmatrix}.$$

From (4), the joint density of (r_i, θ_i) is

$$\begin{aligned} & p\left(r_i, \theta_i \mid \boldsymbol{\mu}, \Sigma\right) \\ & \quad = \left(\frac{1}{2\pi\sigma}\right) r_i \exp\left\{-\frac{1}{2\sigma^2}\left(r_i\boldsymbol{u}_i - \boldsymbol{\mu}\right)^T \begin{pmatrix} 1 & -\gamma \\ -\gamma & \sigma^2+\gamma^2 \end{pmatrix}\left(r_i\boldsymbol{u}_i - \boldsymbol{\mu}\right)\right\} \\ & \quad = r_i p\left(x_{1,i} \mid x_{2,i}, \boldsymbol{\mu}, \Sigma\right) p\left(x_{2,i} \mid \boldsymbol{\mu}, \Sigma\right) \end{aligned}$$

where $p(x_{1,i} \mid x_{2,i}, \boldsymbol{\mu}, \Sigma)$ is a normal density with mean $\mu_1 + \gamma(x_{2,i} - \mu_2)$ and variance σ^2, and $p(x_{2,i} \mid \boldsymbol{\mu}, \Sigma)$ is a normal density with mean μ_2 and variance 1. Closed forms for the full conditionals are then obtained by placing normal priors on $\boldsymbol{\mu}$ and γ (or on $\boldsymbol{\beta}_j$'s and $\boldsymbol{\alpha}_j$'s in the regression case), and an inverse gamma prior on σ^2. Full conditional distributions for $\boldsymbol{\mu}$ (or $\boldsymbol{\beta}_1, \boldsymbol{\alpha}_1$ in the regression case) are the same as in Wang and Gelfand (2013). The full conditional distributions for γ and σ^2 are

$$\begin{aligned} & (\gamma \mid \cdot) \sim \\ & \quad \mathcal{N}\left(\frac{\sigma^{-2}\sum_{i=1}^n \left(x_{1,i} - \mu_1\right)\left(x_{2,i} - \mu_2\right) + \mu_\gamma \sigma_\gamma^{-2}}{\sigma^{-2}\sum_{i=1}^n \left(x_{2,i} - \mu_2\right)^2 + \sigma_\gamma^{-2}}, \frac{1}{\sigma^{-2}\sum_{i=1}^n \left(x_{2,i} - \mu_2\right)^2 + \sigma_\gamma^{-2}}\right) \\ & \left(\sigma^2 \mid \cdot\right) \sim \mathcal{IG}\left(\frac{n}{2} + a, \frac{1}{2}\sum_{i=1}^n \left(x_{1,i} - \left\{\mu_1 + \gamma\left(x_{2,i} - \mu_2\right)\right\}\right)^2 + b\right) \end{aligned}$$

where $\sigma_\gamma^2, \mu_\gamma$ are the prior variance and mean parameters of γ and a, b are the prior shape and rate parameters of σ^2.

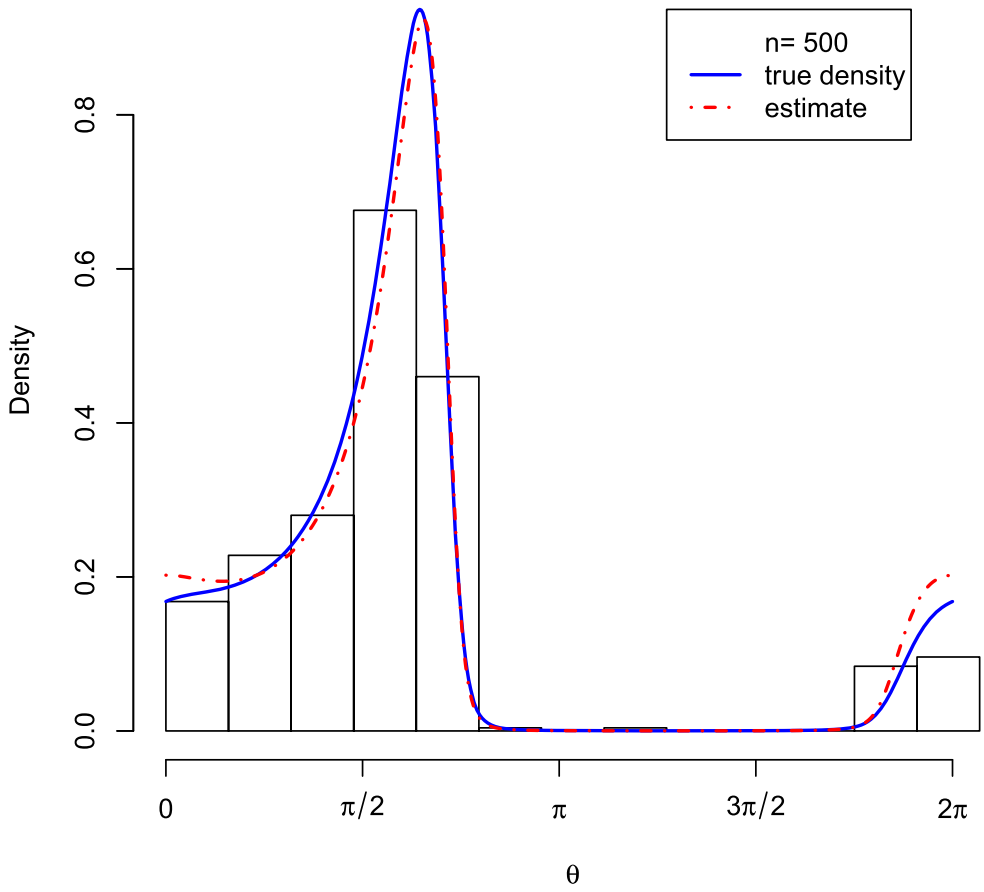

Figure 1: Histogram for a sample of $n = 500$ circular observations generated from a general projected normal distribution with parameters $\boldsymbol{\mu} = (-0.19, 2.09)^T$, $\sigma_1 = 1.58$, $\sigma_2 = 1.4$, and $\rho = -0.84$. The true probability density function for this model is computed from (1) and shown as a solid curve. The posterior predictive density estimate (projected normal density at each angle, averaged over posterior draws) is shown as the dashed curve.

3.1 Simulated circular data

Figure 1 shows a histogram for $n = 500$ circular observations simulated from a general projected normal distribution, along with the true density from (1). Parameters of the general projected normal distribution are $\boldsymbol{\mu} = (-0.19, 2.09)^T$ and $\sigma_1 = 1.58, \sigma_2 = 1.4, \rho = -0.84$ in Σ as in Wang and Gelfand (2013) (Section 4.1 of that paper incorrectly states that $\boldsymbol{\mu} = (-0.19, 1.25)^T$; we thank the authors for clarifying). We specified priors as $\boldsymbol{\mu} \sim \mathcal{N}_2(\mathbf{0}, 10^5\mathbf{I})$ and $\sigma^2 \sim \mathcal{IG}(0.01, 0.01)$, then ran the Gibbs sampler for 10,000 iterations and kept the last 5,000. Figure 1 shows the estimated posterior predictive

density $G^{-1}\sum_{g=1}^{G} p(\theta \mid \boldsymbol{\mu}^{(g)}, \Sigma^{(g)})$ with $p(\theta \mid \boldsymbol{\mu}^{(g)}, \Sigma^{(g)})$ as in (1) and with $\boldsymbol{\mu}^{(g)}, \Sigma^{(g)}$ the gth posterior draws.

3.2 Daily departure times of recreational saltwater anglers

We illustrate the general projected normal regression model for the circular case, including the novel use of covariates within Σ, using observations of the departure times of saltwater recreational anglers from shoreline fishing sites in the 14 states along the Atlantic coast of the United States, with Maine the northernmost state and Florida the southernmost state. The $N = 176,523$ time-of-day observations come from the Coastal Household Telephone Survey (CHTS), conducted by the National Marine Fisheries Service to measure recreational angling activity in saltwater. The CHTS sample is selected via random digit dialing of households, stratified by county within each of six two-month waves (January–February through November–December). Data are collected during the two weeks following completion of each wave, and each active angler is asked to recall all saltwater fishing trips during the two-month wave. We combine data for each wave across years 1990–2008 in the following analysis.

Figure 2 shows posterior predictive densities, to be described below, for departure times at twelve time points in a year, and for every state in the sample. We are specifically interested in modeling the seasonal pattern over the course of the year and the variation in the seasonal pattern across states. The analysis is part of an overall effort to improve recreational fisheries statistics, motivated by the critique in Sullivan et al. (2006).

We treat time-of-year as a circular covariate by transforming it to an angle $\delta = 2\pi(\text{day within year})/(\text{number of days in year})$ (e.g., Mardia and Jupp 2000, p. 257). We model the ith departure time Θ_i as a circular variable with a general projected normal distribution $\Theta_i \sim \mathcal{PN}_2(\boldsymbol{\mu}_i, \Sigma_i)$, where

$$\boldsymbol{\mu}_i^T = \boldsymbol{z}_{\mu,i}^T \begin{pmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 \end{pmatrix}, \quad \Sigma_i = \begin{pmatrix} \sigma^2 + \gamma_i^2 & \gamma_i \\ \gamma_i & 1 \end{pmatrix}, \quad \gamma_i = \boldsymbol{z}_{\gamma,i}^T \boldsymbol{\alpha} \tag{7}$$

following (5) and (6). In our full model, each state has its own intercept, coefficient of $\cos\delta_i$, and coefficient of $\sin\delta_i$ in $\boldsymbol{z}_{\mu,i}^T$, so $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ each have 42 parameters. Further, $\boldsymbol{z}_{\mu,i} = \boldsymbol{z}_{\gamma,i}$, so $\boldsymbol{\alpha}$ has an additional 42 parameters in the covariance matrix model. In our reduced model, there are no covariates in the covariance matrix, so $\gamma_i = \alpha$.

We placed independent normal priors on $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha} \overset{\text{ind}}{\sim} \mathcal{N}_{42}(\mathbf{0}, 10^5\mathbf{I})$ and inverse Gamma on $\sigma^2 \sim \mathcal{IG}(0.01, 0.01)$. All full conditionals are available in closed form. To compare the full model with covariates in Σ_i against the reduced model with constant covariance matrix, we set aside roughly 10% of the observations from each state, leaving a total of $m = 17,657$ observations to be used for out-of-sample model comparison based on the Predictive Log Scoring Loss (PLSL) (Gneiting and Raftery, 2007). Making use of posterior draws, the PLSL can be computed as

$$\text{PLSL} = -\frac{2}{G}\sum_{g=1}^{G}\sum_{j=1}^{m} \log p\left(\theta_j^\star \mid \boldsymbol{\mu}_j^{(g)}, \Sigma_j^{(g)}\right), \tag{8}$$

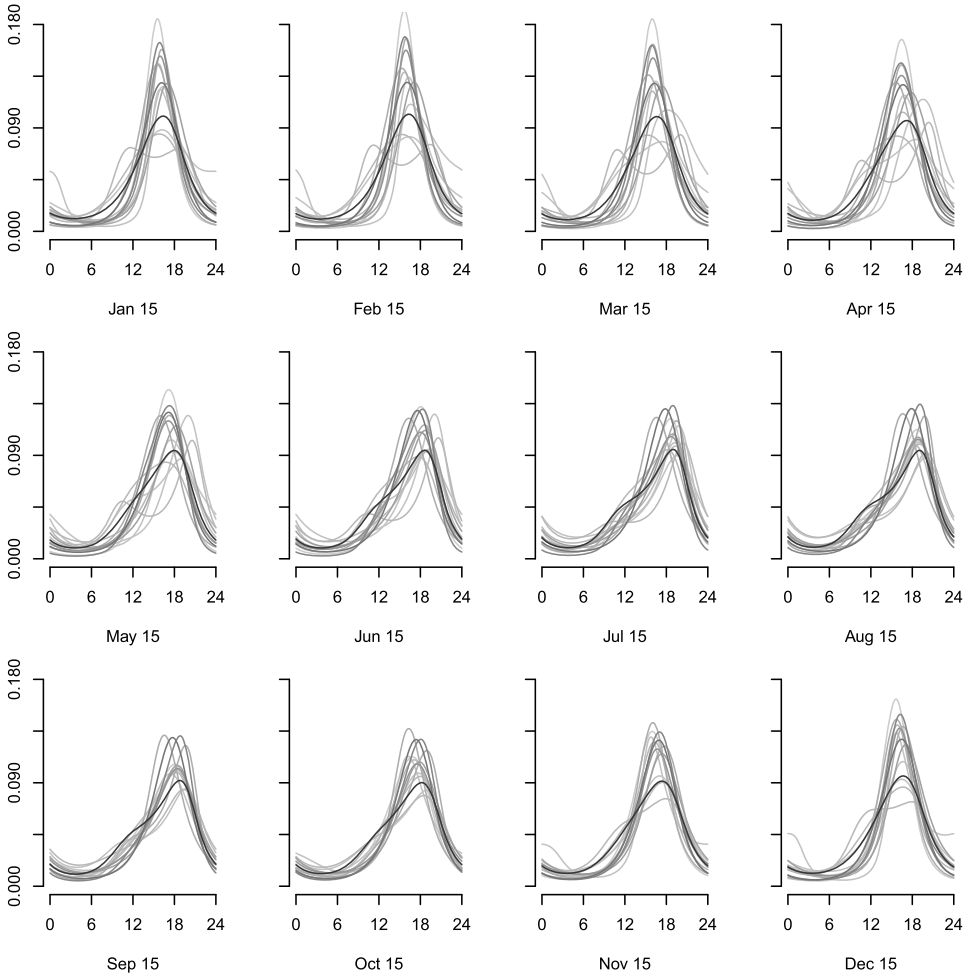

Figure 2: Estimated posterior predictive densities for time-of-day departure times at twelve time points in a year, and for every state in the sample. The darkest curves are the estimated posterior predictive densities at the lowest latitude (Florida) and the lightest curves are the ones at the highest latitude (Maine).

with $p(\theta_j^\star \mid \boldsymbol{\mu}_j^{(g)}, \Sigma_j^{(g)})$ as in (1), $\theta_j^\star, \boldsymbol{z}_{\mu,j}^\star, \boldsymbol{z}_{\gamma,j}^\star$ the jth holdout values and $\boldsymbol{\mu}_j^{(g)}, \Sigma_j^{(g)}$ as in 7 with $\boldsymbol{z}_{\mu,j}, \boldsymbol{z}_{\gamma,j}$ replaced by $\boldsymbol{z}_{\mu,j}^\star, \boldsymbol{z}_{\gamma,j}^\star$ and $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}$ replaced by the gth posterior draw $\boldsymbol{\beta}_1^{(g)}, \boldsymbol{\beta}_2^{(g)}, \boldsymbol{\alpha}^{(g)}$. The remaining $n = 158,866$ observations were used to fit the model. For both models, we ran the Gibbs sampler for 20,000 iterations and kept the last 10,000. The Gibbs sampler runs fast since no rejection methods are used within it.

The difference in PLSL between the full and reduced model was 134.81, favoring the full model with covariates in Σ. We also computed DIC (Spiegelhalter et al., 2002) and

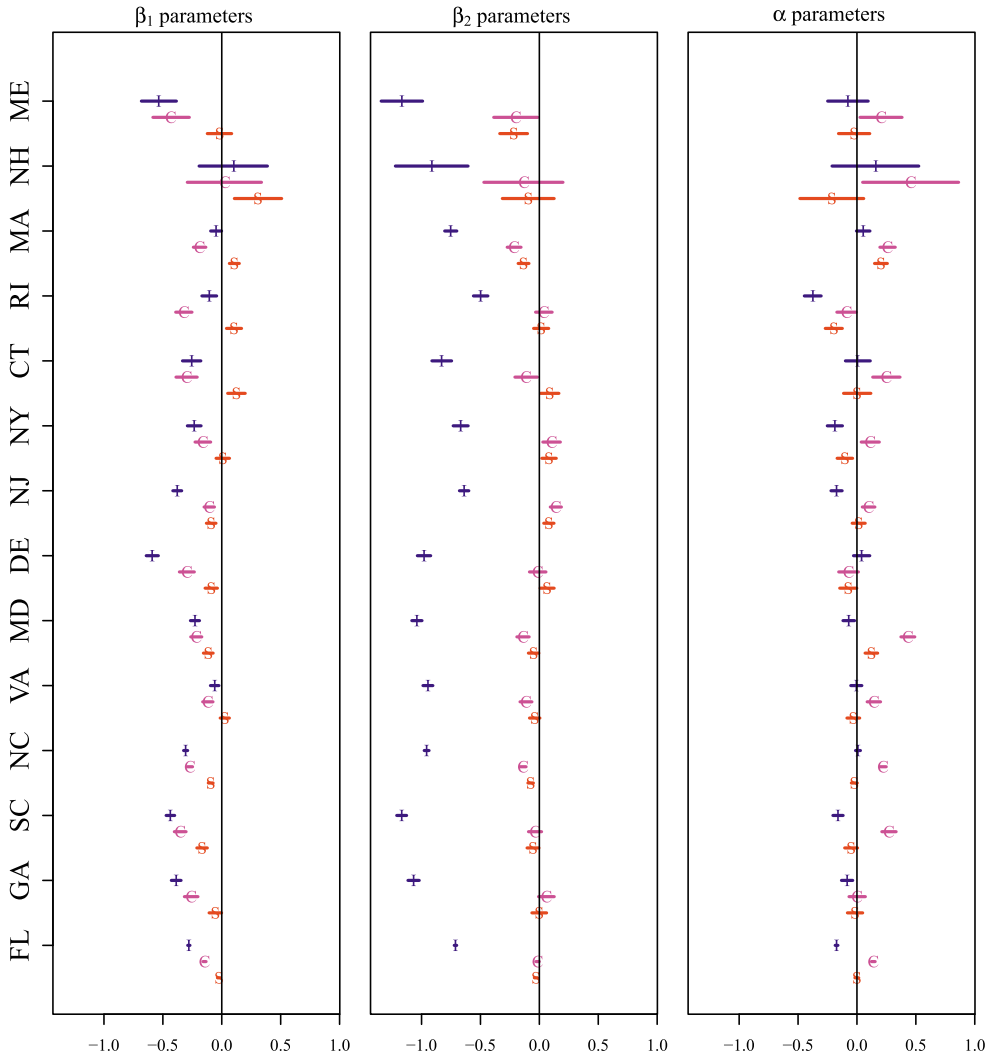

Figure 3: Posterior means and 95% credible intervals by state for the full model of departure time distributions, with 42 parameters in each of the two mean parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, and an additional 42 parameters in the covariance matrix parameter vector $\boldsymbol{\alpha}$. Parameters for each state in each of the three vectors are intercept ("I"), coefficient of $\cos\delta_i$ ("C"), and coefficient of $\sin\delta_i$ ("S"). States are ordered by latitude with Maine (ME) the northernmost state and Florida (FL) the southernmost.

found a difference of 1654.3, again favoring the full model. Several forms of DIC are available in the literature (Celeux et al., 2006). Here we used DIC $= 2\overline{D(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha})} - D(\bar{\boldsymbol{\beta}}_1, \bar{\boldsymbol{\beta}}_2, \bar{\boldsymbol{\alpha}})$ where $\bar{\boldsymbol{\beta}}_1, \bar{\boldsymbol{\beta}}_2, \bar{\boldsymbol{\alpha}}$ are the posterior means and the deviance $D(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}) = -2\log p(\boldsymbol{\theta}|\boldsymbol{\mu}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2), \Sigma(\boldsymbol{\alpha}))$ is minus twice the log-likelihood.

Figure 2 shows estimated posterior predictive densities for each one of the states at twelve time points in the year. The mode shifts gradually from around 4 PM in winter, to 6 PM in spring, to almost 8 PM in late summer, and back to 6 PM in the fall. The darkest curve in each plot corresponds to the lowest latitude (Florida) and the lightest curve corresponds to the highest latitude (Maine). There is a clear latitude effect in most months other than autumn, with a more sharply peaked departure time for the northernmost states, sharpest in mid-winter. This changing latitude effect over the year is captured in the interaction of the categorical covariates for states with the trigonometric terms, again appearing in both the mean and the covariance matrix. Figure 3 gives 95% posterior credible intervals for all parameters in the full model, showing support in the data for the model that varies across states and includes time-of-year not only in the mean but in the covariance. We investigated replacing categorical state effects by a continuous latitude variable, but the state effects were too complex, as suggested by the estimates in Figure 3.

4 Spherical case

For the spherical case, $\boldsymbol{\Theta} = (\Theta_1, \Theta_2)^T$ is obtained from the expression of the random unit vector $\boldsymbol{U}$ in spherical coordinates,

$$\boldsymbol{U} = (\cos\Theta_1 \sin\Theta_2, \sin\Theta_1 \sin\Theta_2, \cos\Theta_2)^T .$$

Equation (4) with $k = 3$ then gives the joint density in terms of (θ_1, θ_2) and the augmented length, r. Integrating r out of this expression (see the supplementary materials (Hernandez-Stumpfhauser et al., 2016) for details), we obtain the joint density for $\boldsymbol{\Theta}$,

$$p(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \Sigma) = |\Sigma|^{-\frac{1}{2}} \exp(C) \times \left(\frac{1}{2\pi A(\boldsymbol{\theta})}\right)^{\frac{3}{2}} \left(\left[1 + D(\boldsymbol{\theta}) \frac{\Phi\{D(\boldsymbol{\theta})\}}{\varphi\{D(\boldsymbol{\theta})\}}\right] D(\boldsymbol{\theta}) + \frac{\Phi\{D(\boldsymbol{\theta})\}}{\varphi\{D(\boldsymbol{\theta})\}}\right) I_{[0,2\pi)}(\theta_1)\, I_{[0,\pi]}(\theta_2) \quad (9)$$

where $D(\boldsymbol{\theta}) = B(\boldsymbol{\theta}) A^{-\frac{1}{2}}(\boldsymbol{\theta})$, $A(\boldsymbol{\theta}) = \boldsymbol{u}^T \Sigma^{-1} \boldsymbol{u}$, $B(\boldsymbol{\theta}) = \boldsymbol{u}^T \Sigma^{-1} \boldsymbol{\mu}$ and $C = -\frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}$. The area element on the sphere is equal to $\sin(\theta_2) d\theta_1 d\theta_2$.

4.1 Simulated spherical data

Figure 4 shows a sample of $n = 500$ spherical observations simulated from a general projected normal distribution with parameters $\boldsymbol{\mu} = (1.26, -0.62, 0.23)^T$, and

$$\Sigma = \mathrm{diag}(2.11, 2.46, 2.38) \begin{pmatrix} 1 & 0.32 & -0.03 \\ 0.32 & 1 & -0.12 \\ -0.03 & -0.12 & 1 \end{pmatrix} \mathrm{diag}(2.11, 2.46, 2.38).$$

We ran the Gibbs sampler with priors $\boldsymbol{\mu} \sim \mathcal{N}_3(\mathbf{0}, 10^5 \mathbf{I})$, $\boldsymbol{\gamma} \sim \mathcal{N}_2(\mathbf{0}, 10^5 \mathbf{I})$, $\Sigma^\star \sim \mathcal{IW}(\mathbf{I}, 4)$. Contours of the true probability density function from (9) are shown as solid curves in Figure 4, and contours of the posterior predictive density estimate, based on 5000 values retained from a Gibbs sampler with 10,000 total iterations, are shown as dashed curves. Both densities are multiplied by $\sin(\theta_2)$.

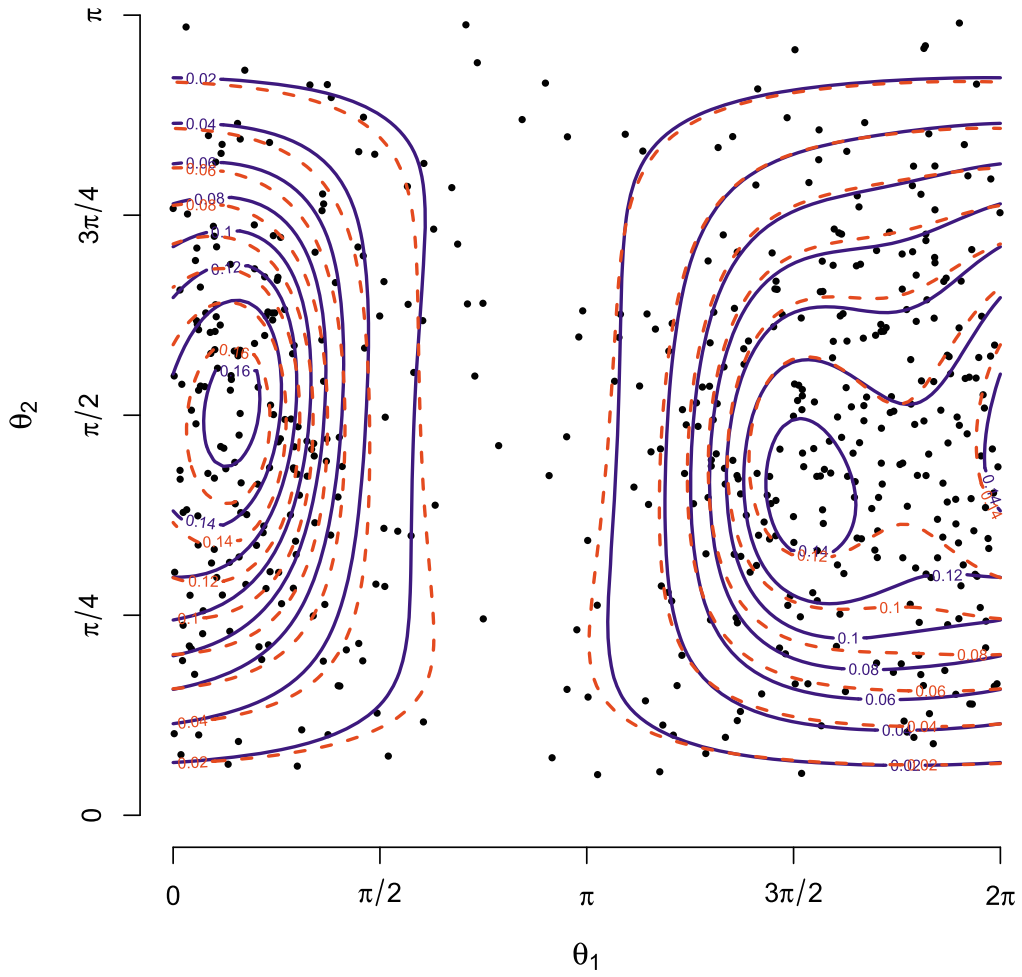

Figure 4: Scatterplot for a sample of $n = 500$ spherical observations generated from the general projected normal distribution of Section 4.1. Contours of the true probability density function from (9) are shown as solid curves and contours of the posterior predictive density estimate (projected normal density at each angle pair, averaged over posterior draws) are shown as dashed curves.

4.2 α-Carbon representation of protein backbone structure

Structural biology studies the structure of biologically active molecules, which are the building blocks of cellular organelles, cells, tissues, and organs. Investigating the relationship between structure and function of such molecules can ultimately lead to a better understanding of health and disease. Although molecules are too small to see under an ordinary light microscope, well-known methods to determine molecular struc-

tures are available; for example, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy.

Here we are concerned with protein structure. Proteins consist of sequences of the 21 natural amino acids, all of which have the same backbone and different side chains. There can be tens to hundreds of amino acids in the protein, and hundreds to hundreds of thousands of atoms. A complete model for the protein consists of coordinates for every atom, as shown in the top left panel of Figure 5 for halophilic malate dehydrogenase (hMDH) from the archaebacterium *Haloarcula marismortui* (Dym et al. 1995, Protein Data Bank entry `1HLP`). The top right panel of the figure shows the same structure, but in a simplified representation that color-codes three main types of structure: helices, β-sheets, and coils and turns.

Further simplifications are often useful. One class of simplified models is obtained by focusing only on the α-carbon sequence, consisting of the primary carbon atom in the backbone of each amino acid. The bottom panel of Figure 5 shows the α-carbon atoms as spheres, connected by pseudo-bonds to depict the protein backbone. The three-dimensional structure of the resulting α-carbon backbone can be described as a sequence of angle pairs, $\{(\tau_i, \theta_i)\}_{i=1}^{n}$ (Levitt, 1976; Oldfield and Hubbard, 1994), as shown in the inset to the bottom panel of Figure 5. The bond angle θ has a range $[0, \pi)$ and τ is a dihedral angle with range $[-\pi, \pi)$. These two angles lie naturally on a sphere. Hamelryck et al. (2006) used the distributions of these two angles to randomly generate plausible models for protein structures.

A related but different simplified description leads to the Ramachandran plot, which uses different pairs of angles to describe the backbone structure (Ramachandran et al., 1963). The Ramachandran angle pairs lie naturally on a torus rather than a sphere. Methods described in this paper can be adapted to the toroidal case, which is beyond the scope of the present work. Both the Ramachandran plot and our analysis summarize the protein structure in terms of the marginal joint density of the angle pairs, ignoring any dependence in the sequence. Time series methods for the angle-pair sequence are a natural extension of the work presented here, a point we return to briefly in Section 6.

Here we estimate the joint density of (τ, θ) for 1295 proteins by making use of mixtures of general projected normal distributions $\mathcal{PN}_3(\boldsymbol{\mu}, \Sigma)$. While mixtures of projected normals have been used in (Wang and Gelfand, 2014; Nuñez-Antonio et al., 2015), this is to the best of our knowledge the first mixture of general projected normals in the spherical case. The flexibility of the general projected normal distribution, which includes the possibility of two modes and asymmetry, can potentially describe two separate regions with a single mixture component making it more attractive for the use of mixture models than say von Mises-Fisher distributions (Banerjee et al., 2005) or Kent distributions (Peel et al., 2001), which are both unimodal distributions.

Mixtures of projected normal densities inherit excellent approximation properties from mixtures of normal densities. To see this, let $p(\boldsymbol{\theta})$ denote any density on the k-dimensional hypersphere, and let $p(r)$ denote an arbitrary one-dimensional density on $\mathbb{R}^+$. Then $p(\boldsymbol{\theta}, r) = p(\boldsymbol{\theta})p(r)$ is a density in $\mathbb{R}^k$; if it is well-approximated by a mixture of k-variate normal densities, then $p(\boldsymbol{\theta})$ is well-approximated by the corresponding mixture

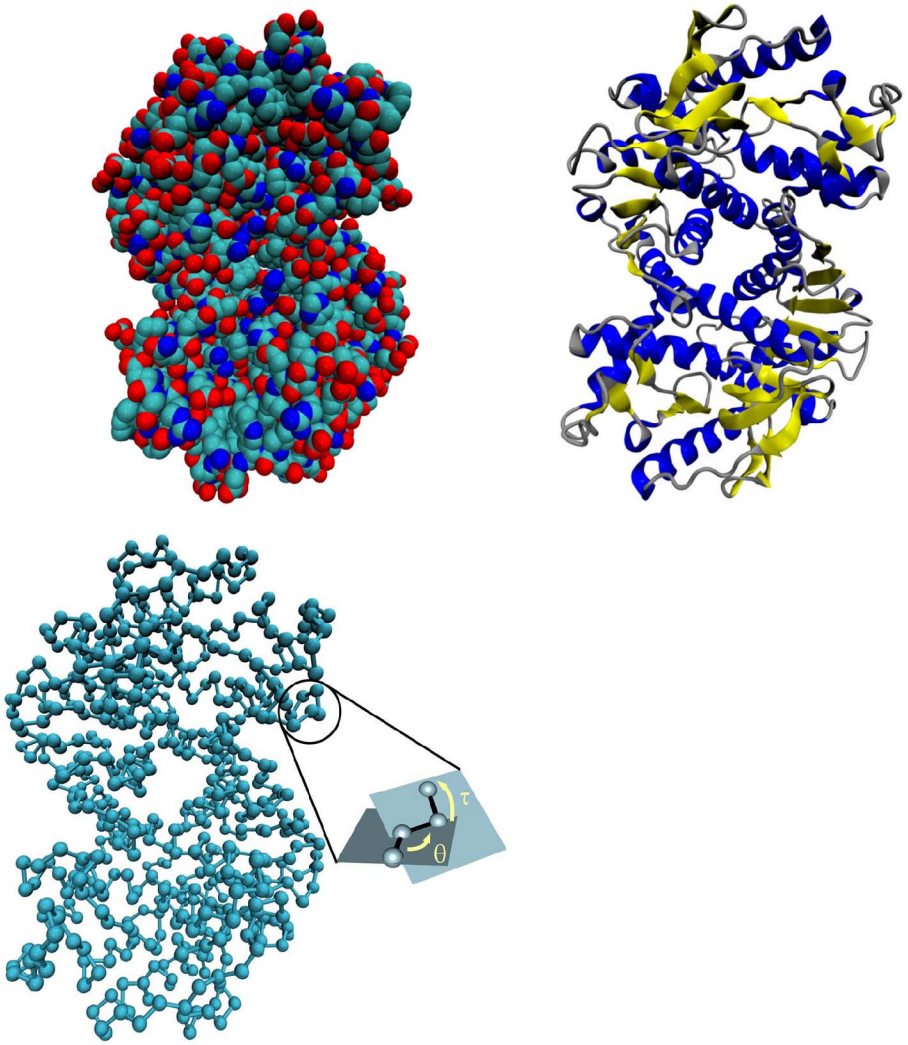

Figure 5: Representations of the molecular structure of malate dehydrogenase. In the top left view all atoms (coordinates derived from Protein Data Bank entry `1HLP` Dym et al. 1995, hydrogen atoms omitted) are depicted as spheres, with carbon, nitrogen, and oxygen in cyan, blue, and red, respectively. The top right panel shows the same structure, but in a simplified representation that color-codes three main types of secondary structure: helices in dark blue, β-sheets in yellow, and coils and turns in gray. The bottom panel shows a different simplified view of the structure (Levitt, 1976), in which the α-carbon atom for each amino acid is shown as a sphere, connected by pseudo-bonds to depict the protein backbone. The three-dimensional structure of the resulting α-carbon backbone can be described as a sequence of angle pairs, $\{(\tau_i, \theta_i)\}_{i=1}^{n}$. where the angles are defined in the inset to the bottom panel. (Images made with `VMD` Humphrey et al. 1996.)

of projected normal densities. Theorem 1 of Wang and Gelfand (2014) makes these statements precise for the circular case, showing that mixtures of circular projected normal densities $\mathcal{PN}_2(\boldsymbol{\mu}_j, \Sigma_j)$ can approximate arbitrarily well in L_1 norm a large class of densities on the circle. Their results extend immediately to k dimensions.

The 1295 proteins used in this example are a (non-random) subset of the 1424 proteins used in Hamelryck et al. (2006). We randomly sampled 15 angle pairs (τ_j, θ_j) from each of the 1295 proteins. We also randomly selected 295 proteins to set aside, so that $n = 15,000$ angle pairs from 1000 proteins were used to fit the mixture model, and $m = (15)(295) = 4425$ from 295 proteins were set aside for out-of-sample model comparison based on the PLSL.

We model the data as a mixture of J general projected normal distributions

$$\sum_{j=1}^{J} w_j \mathcal{PN}_3\left(\boldsymbol{\mu}_j, \Sigma_j\right), \quad \Sigma_j = \begin{pmatrix} \Sigma_j^* + \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T & \boldsymbol{\gamma}_j \\ \boldsymbol{\gamma}_j^T & 1 \end{pmatrix}.$$

For each $j = 1, \ldots, J$, we placed normal priors on $\boldsymbol{\mu}_j \sim \mathcal{N}_3(\boldsymbol{\mu}_{\boldsymbol{\mu}}, \sigma^2_{\boldsymbol{\mu}}\mathrm{I})$ and $\boldsymbol{\gamma}_j \sim \mathcal{N}_2(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \sigma^2_{\boldsymbol{\gamma}}\mathrm{I})$ and inverse Wishart priors on $\Sigma_j^* \sim \mathcal{IW}(\mathrm{I}, 4)$. We placed normal priors $\boldsymbol{\mu}_{\boldsymbol{\mu}} \sim \mathcal{N}_3(\mathbf{0}, 10^2\mathrm{I})$ and $\boldsymbol{\mu}_{\boldsymbol{\gamma}} \sim \mathcal{N}_2(\mathbf{0}, 10^2\mathrm{I})$. Finally, we placed a Dirichlet prior on the weights $\boldsymbol{w} = (w_1, \ldots, w_J)^T \sim \mathcal{D}(1, \ldots, 1)$.

We ran the Gibbs sampler for 70,000 iterations and kept the last 10,000 to estimate all parameters in the model. We tried numbers of mixtures $J = 3, 4, 5, 6, 7, 8$ and chose the model with $J = 7$ mixtures using PLSL, which was 57.4 less than that for $J = 6$ and 69.1 less than that for $J = 8$. The mixture model with $J = 3$ had the worst predictions with a PLSL 734.2 greater than that of $J = 7$.

Figure 6 shows a perspective plot of the estimated posterior predictive density and a contour plot of the log estimated posterior predictive density. The first most probable region, with mode $(\tau, \theta) = (0.86, 1.58)$, contains angle pairs primarily from helix structures; the second most probable region, with mode $(\tau, \theta) = (-2.81, 2.17)$, corresponds primarily to β-sheets; and the third most probable region, with mode $(\tau, \theta) = (-2.38, 1.60)$, corresponds typically to coils and turns.

In a related application to dihedral angles $\{\tau_i\}_{i=1}^n$, DeWitte and Shakhnovich (1994) developed an analysis of the probability density function for τ_i as a function of a categorical covariate, determined by the central pair of amino acids for angle i. Each amino acid was classified into one of three groups (Helix, Sheet and Turn formers), so that the pair of amino acids led to nine possible values of the categorical covariate. It is straightforward to conduct this analysis for circular data τ_i or to extend to spherical data (τ_i, θ_i) using the methods of this paper. A complete analysis of this model is beyond the scope of the present paper, but involves no new ideas.

5 Higher-dimensional cases

As a further test of our methods, we explored estimation of $\mathcal{PN}_k(\boldsymbol{\mu}_k, \Sigma_k)$ for $k = 10$ and $k = 50$. For each dimension, we placed a simple structure on

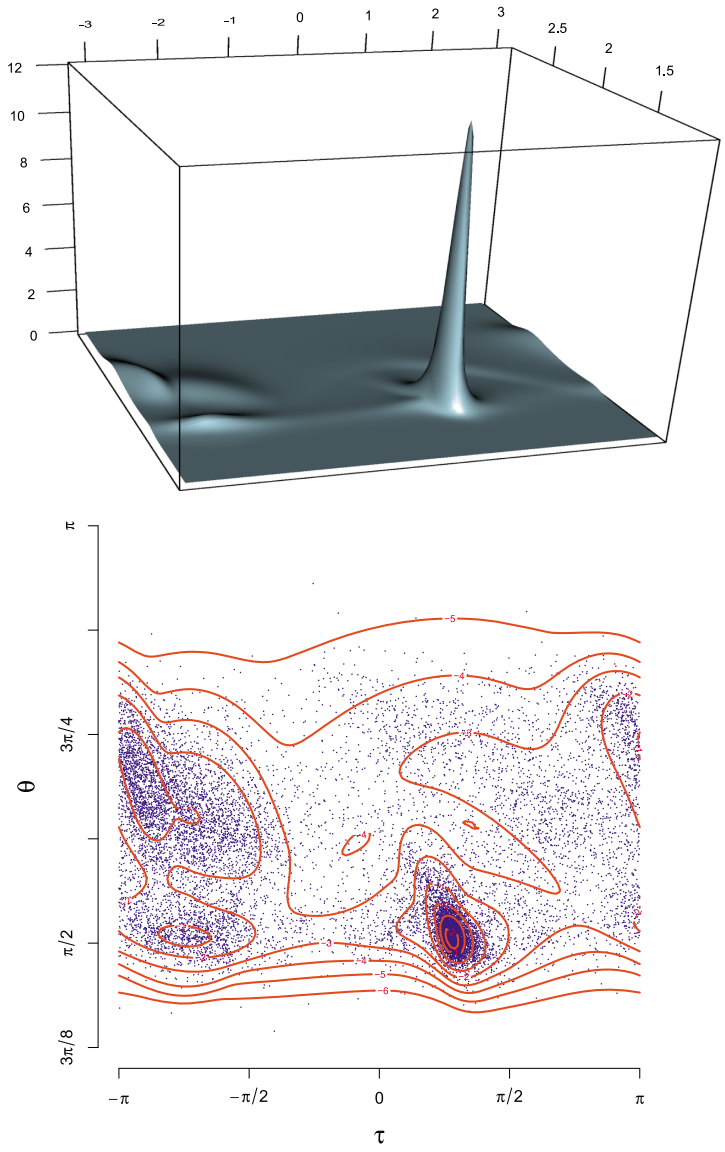

Figure 6: Bottom figure is a scatterplot of $n = 15,000$ angle pairs chosen as 15 angle pairs selected at random from each of 1000 proteins. The contour lines show the log of the estimated posterior predictive density from modeling the data as a mixture of seven general projected normal distributions $\mathcal{PN}_3(\boldsymbol{\mu}_j, \Sigma_j)$. The top figure shows the posterior predictive density in three dimensions. The dominant peak with mode $(\tau, \theta) = (0.86, 1.58)$ corresponds to angle pairs primarily from helical structures; the second most probable region, with mode $(\tau, \theta) = (-2.81, 2.17)$ ($\tau = -\pi$ wraps around to π), corresponds to β-sheets; and the third most probable region, with mode $(\tau, \theta) = (-2.38, 1.60)$, corresponds to coil and turn structures.

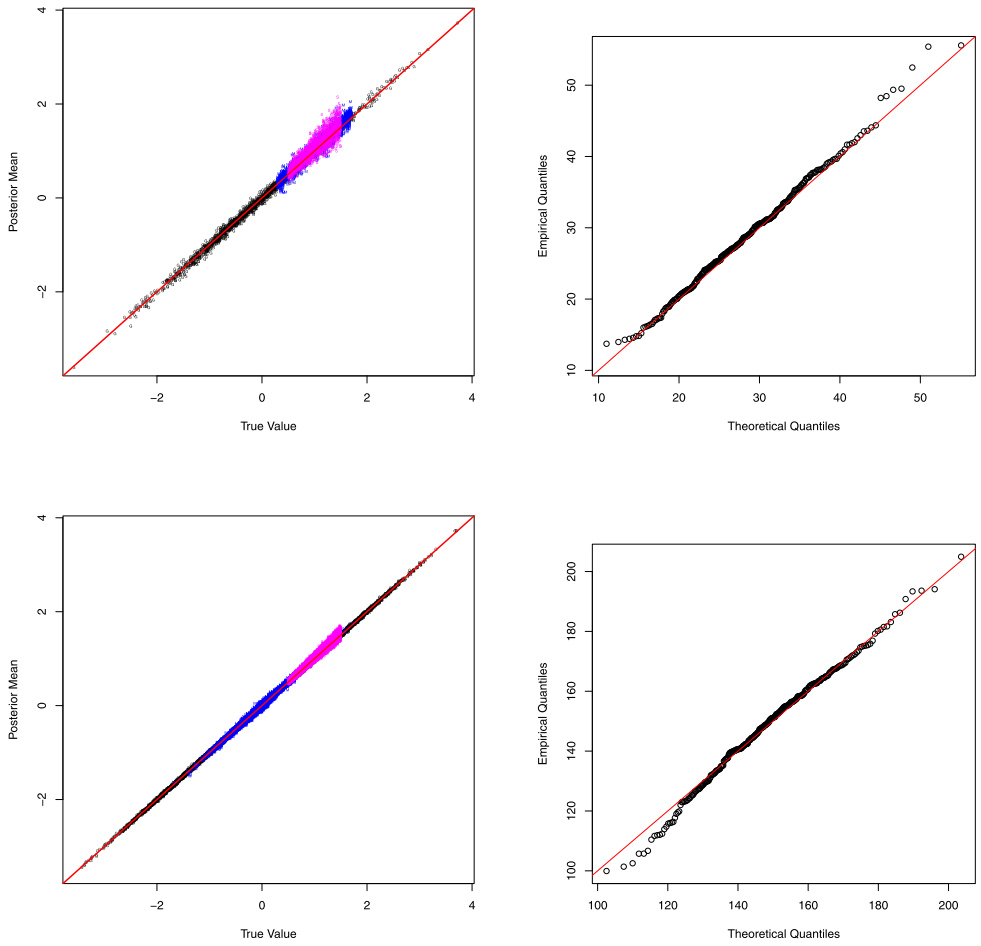

Figure 7: Assessment of estimation performance for $k = 10$ dimensional problem (top row) and $k = 50$ dimensional problem (bottom row). Left column: posterior means versus true values for $3k-2$ parameters in each of 300 different simulated k-dimensional cases, with $k \times 300$ μ values in blue, $(k-1) \times 300$ γ values in black, and $(k-1) \times 300$ σ^2 values in magenta. Right column: theoretical quantiles of asymptotic χ^2_{3k-2} distribution versus 300 empirical quantiles of the quadratic form $(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^T \boldsymbol{S}_{\boldsymbol{\beta}}^{-1} (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})$, where $\boldsymbol{\beta}$ is one draw per case from the $(3k-2)$-dimensional posterior, $\bar{\boldsymbol{\beta}}$ is the sample mean of the posterior draws, and $\boldsymbol{S}_{\boldsymbol{\beta}}$ is their sample covariance matrix.

$$\Sigma_k = \begin{pmatrix} \Sigma^*_{k-1} + \boldsymbol{\gamma}_{k-1}\boldsymbol{\gamma}^T_{k-1} & \boldsymbol{\gamma}_{k-1} \\ \boldsymbol{\gamma}^T_{k-1} & 1 \end{pmatrix}$$

by assuming $\Sigma^*_{k-1} = \text{diag}(\sigma^2_1, \ldots, \sigma^2_{k-1})$, with the diagonal elements strictly positive. This structure yields $3k-2$ model parameters, 28 for $k = 10$ and 148 for $k = 50$. We gen-

erated the true parameters at random, drawing σ_j^2 iid $\mathcal{U}(0.5, 1.5)$, $\boldsymbol{\gamma}_{k-1} \sim \mathcal{N}(\mathbf{0}_{k-1}, \mathbf{I}_{k-1})$, $\boldsymbol{u}$ from a uniform on the k-dimensional hypersphere, $\|\boldsymbol{\mu}\| \sim \mathcal{U}(0, 4)$, and setting $\boldsymbol{\mu}_k = \|\boldsymbol{\mu}\|\boldsymbol{u}$. For each random parameter draw, we then generated $n_{10} = 500$ or $n_{50} = 2500$ iid $\mathcal{PN}_k(\boldsymbol{\mu}_k, \Sigma_k)$ observations. The ratio of number of observations to number of parameters is roughly equal in each case, $148/28 \simeq 2500/500$.

We then conducted inference by placing normal priors on $\boldsymbol{\mu}_k \sim \mathcal{N}(\mathbf{0}, 10^5\mathbf{I}_k)$ and $\boldsymbol{\gamma}_{k-1} \sim \mathcal{N}(\mathbf{0}, 10^5\mathbf{I}_{k-1})$, and independent inverse gamma priors on the diagonal elements of Σ^*_{k-1}, $\sigma_j^2 \sim \mathcal{IG}(0.1, 0.1)$. This prior specification is not the same as the actual data generating mechanism. For each simulated data set, we ran the Gibbs sampler for 30,000 iterations, retaining the last 10,000. Trace plots indicated no problems with convergence of the chains and posterior means track true values well; see the plots in the left column of Figure 7. Further, let $\boldsymbol{\beta} = (\boldsymbol{\mu}_k^T, \boldsymbol{\gamma}_{k-1}^T, \sigma_1^2, \ldots, \sigma_{k-1}^2)^T$ denote one random draw of the $3k-2$ vector of parameters from the approximate posterior distribution and let $\bar{\boldsymbol{\beta}}, \boldsymbol{S}_{\boldsymbol{\beta}}$ denote the sample mean and sample covariance matrix of the posterior draws. Then, for sufficiently large sample size, the quadratic form $(\boldsymbol{\beta}-\bar{\boldsymbol{\beta}})^T \boldsymbol{S}_{\boldsymbol{\beta}}^{-1} (\boldsymbol{\beta}-\bar{\boldsymbol{\beta}})$ is asymptotically χ^2 with $3k-2$ degrees of freedom. The right column of Figure 7 shows empirical quantiles of the 300 quadratic forms versus theoretical quantiles of the χ^2_{3k-2} distribution. The plots suggest that these quadratic forms, which summarize vector-valued posterior draws, are at least reasonably consistent with their asymptotic χ^2 distributions.

6 Discussion

The empirical examples of Section 3.2 and Section 4.2 and the simulated examples of Section 5 illustrate the fact that our new parameterization of the general projected normal yields a fast sampler, allowing for the efficient exploration of high-dimensional parameter spaces. Our method makes it feasible to conduct model selection among complex mixture and regression models. In addition, our parameterization makes possible interesting extensions, such as building in mixture/regression or time series structures for spherical data. As noted above, the example of Section 4.2 estimates only the marginal density of (τ, θ), ignoring any dependence in the sequence of successive angle pairs along the α-carbon backbone. Using extensions of the methods described in this paper, we are currently developing vector autoregressive models for such time series of spherical data, including autoregressive models that switch parameterizations according to local structure in the protein (such as helices, beta-sheets, and coils).

Supplementary Material

Supplementary Material of "The General Projected Normal Distribution of Arbitrary Dimension: Modeling and Bayesian Inference" (DOI: 10.1214/15-BA989SUPP; .pdf).

References

Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S. (2005). "Clustering on the unit hypersphere using von Mises-Fisher distributions." *Journal of Machine Learning Research*, 6: 1345–1382. MR2249858. 114, 125

Batschelet, E. (1981). *Circular Statistics in Biology*. Academic Press. MR0659065. 114

Breckling, J. (1989). *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*. Springer-Verlag. MR1027836. doi: http://dx.doi.org/10.1007/978-1-4612-3688-7. 114

Celeux, G., Forbes, F., Robert, F., and Titterington, D. (2006). "Deviance information criteria for missing data models." *Bayesian Analysis*, 1: 651–706. MR2282197. doi: http://dx.doi.org/10.1214/06-BA122. 122

Chang, T. (1993). "Spherical regression and the statistics of tectonic plate reconstructions." *International Statistical Review*, 61: 299–316. 114

DeWitte, R. and Shakhnovich, E. (1994). "Pseudodihedrals: Simplified protein backbone representation with knowledge-based energy." *Protein Science*, 3: 1570–1581. 127

Dym, O., Mevarech, M., and Sussman, J. L. (1995). "Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium." *Science*, 267(5202): 1344–1346. 125, 126

Ferreira, J. T., Juárez, M. A., and Steel, M. F. (2008). "Directional log-spline distributions." *Bayesian Analysis*, 3(2): 297–316. MR2407428. doi: http://dx.doi.org/10.1214/08-BA311. 114

Fisher, N. I. (1995). *Statistical Analysis of Circular Data*. Cambridge University Press. MR1251957. doi: http://dx.doi.org/10.1017/CBO9780511564345. 114

Gao, F., Chia, K.-S., and Machin, D. (2007). "On the evidence for seasonal variation in the onset of acute lymphoblastic leukemia (ALL)." *Leukemia Research*, 31: 1327–1338. 114

Ghosh, K., Jammalamadaka, R., and Tiwari, R. (2003). "Semiparametric Bayesian techniques for problems in circular data." *Journal of Applied Statistics*, 30(2): 145–161. MR1956181. doi: http://dx.doi.org/10.1080/0266476022000023712. 114

Gneiting, T. and Raftery, A. (2007). "Strictly proper scoring rules." *Journal of the American Statistical Association*, 102: 359–378. MR2345548. doi: http://dx.doi.org/10.1198/016214506000001437. 120

Hamelryck, T., Kent, J., and Krogh, A. (2006). "Sampling realistic protein conformations using local structural bias." *PLoS Computational Biology*, 2: e131. 125, 127

Hernandez-Stumpfhauser, D. (2012). "Topics in design-based and Bayesian inference for surveys." Ph.D. thesis, Colorado State University. MR3152405. 114, 115

Hernandez-Stumpfhauser, D., Breidt, F. J., and van der Woerd, M. J. (2016). "Supplementary Material of The General Projected Normal Distribution of Arbitrary Dimension: Modeling and Bayesian Inference" *Bayesian Analysis*. doi: http://dx.doi.org/10.1214/15-BA989SUPP. 115, 123

Humphrey, W., Dalke, A., and Schulten, K. (1996). "VMD – Visual Molecular Dynamics." *Journal of Molecular Graphics*, 14: 33–38. 126

Jammalamadaka, S. R. and Sengupta, A. (2001). *Topics in Circular Statistics*, volume 5. World Scientific. MR1836122. doi: http://dx.doi.org/10.1142/9789812779267. 114

Kendall, D. G. (1974). "Pole-seeking Brownian motion and bird navigation." *Journal of the Royal Statistical Society*, 36: 261–294. MR0423561. 115

Levitt, M. (1976). "A simplified representation of protein conformation for rapid simulation of protein folding." *Journal of Molecular Biology*, 104: 59–107. 114, 125, 126

Mardia, K. and Edwards, R. (1982). "Weighted distributions and rotating caps." *Biometrika*, 69: 323–330. MR0671970. doi: http://dx.doi.org/10.1093/biomet/69.2.323. 114

Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press. MR0336854. 115

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Chichester, UK: Wiley. MR1828667. 114, 120

McVinish, R. and Mengersen, K. (2008). "Semiparametric Bayesian circular statistics." *Computational Statistics & Data Analysis*, 52(10): 4722–4730. MR2521617. doi: http://dx.doi.org/10.1016/j.csda.2008.03.016. 114

Neal, R. M. (2003). "Slice sampling." *The Annals of Statistics*, 31: 705–741. MR1994729. doi: http://dx.doi.org/10.1214/aos/1056562461. 117

Nuñez-Antonio, G., Ausin, M., and Wiper, M. (2015). "Nonparametric models of circular variables based on Dirichlet process mixtures of normal distributions." *Journal of Agricultural, Biological, and Environmental Statistics*, 20: 47–64. MR3334466. doi: http://dx.doi.org/10.1007/s13253-014-0193-y. 114, 115, 125

Nuñez-Antonio, G. and Gutiérrez-Peña, E. (2005). "A Bayesian analysis of directional data using the projected normal distribution." *Journal of Applied Statistics*, 32(10): 995–1001. MR2221902. doi: http://dx.doi.org/10.1080/02664760500164886. 114, 115

Nuñez-Antonio, G., Gutiérrez-Peña, E., and Escalera, G. (2011). "A Bayesian regression model for circular data based on the projected normal distribution." *Statistical Modeling*, 11: 185–201. MR2849683. doi: http://dx.doi.org/10.1177/1471082X1001100301. 114, 115

Oldfield, T. and Hubbard, R. (1994). "Analysis of C_α geometry in protein structures." *Proteins*, 18: 324–337. 114, 125

Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2012). "A plug-in rule for bandwidth selection in circular density estimation." *Computational Statistics & Data Analysis*, 56(12): 3898–3908. MR2957840. doi: http://dx.doi.org/10.1016/j.csda.2012.05.021. 114

Peel, D., Whiten, W. J., and McLachlan, G. J. (2001). "Fitting mixtures of Kent distributions to aid in joint set identification." *Journal of the American Statistical Association*, 96: 56–63. MR1973782. doi: http://dx.doi.org/10.1198/016214501750332974. 125

Pewsey, A., Neuhäuser, M., and Ruxton, G. D. (2013). *Circular Statistics in R*. Oxford University Press. MR3156170. 114

Pourahmadi, M. (1999). "Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation." *Biometrika*, 86: 677–690. MR1723786. doi: http://dx.doi.org/10.1093/biomet/86.3.677. 117

Presnell, B., Morrison, S. P., and Littell, R. C. (1998). "Projected multivariate linear models for directional data." *Journal of the American Statistical Association*, 93(443): 1068–1077. MR1649201. doi: http://dx.doi.org/10.2307/2669850. 114, 115

Pukkila, T. and Rao, C. (1988). "Pattern recognition based on scale invariant functions." *Information Sciences*, 45: 379–389. MR0952440. doi: http://dx.doi.org/10.1016/0020-0255(88)90012-6. 115

Ramachandran, G., Ramakrishnan, C., and Sasisekharan, V. (1963). "Stereochemistry of polypeptide chain configurations." *Journal of Molecular Biology*, 7: 95–99. 125

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). "Item-based collaborative filtering recommendation algorithms." In: *Proceedings of the 10th International Conference on World Wide Web*, 285–295. ACM. 114

Schmidt-Koenig, K. (1965). "Current problems in bird orientation." In: Lehrman, D. (ed.), *Advances in the Study of Behaviour, Volume I*, 217–278. Academic Press. 114

Spiegelhalter, D. J., Best, N., Carlin, B. P., and van der Linde, A. (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society, Series B*, 64: 583–639. MR1979380. doi: http://dx.doi.org/10.1111/1467-9868.00353. 121

Sullivan, P. J., Breidt, F. J., Ditton, R. B., Knuth, B. A., Leaman, B. M., O'Connell, V. M., Parsons, G. R., Pollock, K. H., Smith, S. J., and Stokes, S. L. (2006). *Review of Recreational Fisheries Survey Methods*. Washington, DC: National Academies Press. 120

Wang, F. and Gelfand, A. E. (2013). "Directional data analysis under the general projected normal distribution." *Statistical Methodology*, 10: 113–127. MR2974815. doi: http://dx.doi.org/10.1016/j.stamet.2012.07.005. 114, 115, 116, 118, 119

Wang, F. and Gelfand, A. E. (2014). "Modeling space and space-time directional data using projected Gaussian processes." *Journal of the American Statistical Association*, 109: 1565–1580. MR3293610. doi: http://dx.doi.org/10.1080/01621459.2014.934454. 114, 115, 125, 127

Watson, G. S. (1983). *Statistics on Spheres*. Wiley. MR0709262. 114, 115

Acknowledgments

This work is partially supported by the Joint NSF/NIGMS Initiative to Support Research in the Area of Mathematical Biology (R01GM096192), and partially by the NIEHS T32ES007018 training grant in environmental biostatistics. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. We appreciate the helpful comments of the associate editor and two anonymous referees.