Journal of Statistical Planning and Inference 143 (2013) 992–999

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

On conjugate families and Jeffreys priors for von Mises–Fisher distributions

Kurt Hornik [a], Bettina Grün [b,*]

[a] Institute for Statistics and Mathematics, WU Wirtschaftsuniversität Wien, Augasse 2–6, 1090 Vienna, Austria
[b] Department of Applied Statistics, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria

ARTICLE INFO

Article history:
Received 27 September 2011
Received in revised form
7 August 2012
Accepted 6 November 2012
Available online 23 November 2012

Keywords:
Bayesian inference
Conjugate prior
Jeffreys prior
von Mises–Fisher distribution

ABSTRACT

This paper discusses characteristics of standard conjugate priors and their induced posteriors in Bayesian inference for von Mises–Fisher distributions, using either the canonical natural exponential family or the more commonly employed polar coordinate parameterizations. We analyze when standard conjugate priors as well as posteriors are proper, and investigate the Jeffreys prior for the von Mises–Fisher family. Finally, we characterize the proper distributions in the standard conjugate family of the (matrix-valued) von Mises–Fisher distributions on Stiefel manifolds.

1. Introduction

A random unit length vector in $\mathbb{R}^d$ has a von Mises–Fisher (or Langevin, short: vMF) distribution with parameter $\theta \in \mathbb{R}^d$ if its density with respect to the uniform distribution on the unit hypersphere $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ is given by

$$f(x|\theta) = e^{\theta' x} / {}_0F_1(; d/2; \|\theta\|^2/4),$$

where, using the rising factorial $(\nu)_n = \Gamma(\nu+n)/\Gamma(\nu)$,

$${}_0F_1(; \nu; z) = \sum_{n=0}^{\infty} \frac{1}{(\nu)_n} \frac{z^n}{n!} = \sum_{n=0}^{\infty} \frac{\Gamma(\nu)}{\Gamma(\nu+n)} \frac{z^n}{n!}$$

is a generalized hypergeometric series and related to the modified Bessel function of the first kind I_ν via

$${}_0F_1(; \nu+1; \kappa^2/4) = \frac{I_\nu(\kappa)\Gamma(\nu+1)}{(\kappa/2)^\nu}$$

(e.g., Mardia and Jupp, 1999, p. 168).

We note that the vMF distribution is commonly parameterized using polar coordinates, i.e., $\theta = \kappa\mu$, where $\kappa = \|\theta\|$ and $\mu \in \mathbb{S}^{d-1}$ are the concentration and mean direction parameters, respectively (if $\theta \neq 0$, μ is uniquely determined as $\theta/\|\theta\|$). Using θ as the parameter, the family $\mathcal{F}$ of vMF distributions on $\mathbb{S}^{d-1}$ becomes a natural exponential family through the

* Corresponding author.
E-mail addresses: Kurt.Hornik@wu.ac.at (K. Hornik), Bettina.Gruen@jku.at (B. Grün).

0378-3758/$ - see front matter
http://dx.doi.org/10.1016/j.jspi.2012.11.003

uniform distribution U on $\mathbb{S}^{d-1}$, commonly written as

$$f(x|\theta) = e^{\theta' x - M(\theta)},$$

where in the vMF case, the cumulant transform $M(\theta)$ of U is given by

$$e^{M(\theta)} = \int_{\mathbb{S}^{d-1}} e^{\theta' x} dU(x) = {}_0F_1(; d/2; \|\theta\|^2/4).$$

Bayesian inference for the vMF distribution is first discussed in Mardia and El-Atoum (1976), who give conjugate priors for μ when κ is known, and derive the Jeffreys prior for the polar coordinates (μ, κ) parameterization. Guttorp and Lockhart (1988) introduce a Bayesian approach for finding the direction of a signal based on developing *standard* (e.g., Gutiérrez-Peña and Smith, 1997, Definition 3.1) conjugate priors for the von Mises (vM) distribution (i.e., for $d=2$) using the canonical (θ) parameterization. Damien and Walker (1999) present a full Bayesian analysis of circular data using the vM distribution by employing standard conjugate priors for the polar coordinates (μ, κ) parameterization, and developing a Gibbs sampler for this family of distributions. Nuñez-Antonio and Gutiérrez-Peña (2005) provide a full Bayesian analysis of directional (i.e., $d \geq 2$) data using the vMF distribution, again using standard (μ, κ) conjugate priors and obtaining samples from the posterior using a sampling-importance-resampling method found to outperform Gibbs sampling. Bangert et al. (2010) construct (possibly infinite) mixtures of vMF distributions using standard conjugate priors for the (μ, κ) parameterization and Dirichlet (process) priors for the mixing probabilities.

Interestingly, none of these references explicitly discuss when the employed priors (and respective posteriors) are actually proper, or whether the conjugate families obtained using the θ or (μ, κ) parameterizations are the same. In this paper, we settle these open issues, and also discuss Jeffreys priors for the general ($d \geq 2$) vMF family (Section 2). We also provide results for (matrix-valued) vMF distributions on Stiefel manifolds (Section 3).

2. Results

2.1. Propriety of priors from the standard conjugate family

In what follows, it will be convenient to write

$$C_d(\kappa) = 1/{}_0F_1(; d/2; \kappa^2/4),$$

so that $e^{-M(\theta)} = C_d(\|\theta\|)$. Let $\theta = \theta(\lambda)$ be a suitable parameterization of θ. For a sample $x_1, \ldots, x_n$ of independent, identically distributed (i.i.d.) observations from the vMF family $\mathcal{F}$, the likelihood function for λ is given by

$$L(\lambda|s,n) = e^{s'\theta(\lambda) - nM(\theta(\lambda))} = C_d(\|\theta(\lambda)\|)^n e^{s'\theta(\lambda)},$$

where $s = x_1 + \cdots + x_n$ is the resultant of the sample. Following Gutiérrez-Peña and Smith (1997, Definition 3.1), the *standard conjugate family* for $\mathcal{F}$ relative to λ, denoted by $\mathcal{C}_\lambda(\mathcal{F})$, has densities

$$\pi(\lambda|s,v) \propto L(\lambda|s,v).$$

Using such a prior with parameters s_0 and v_0 will result in a posterior with parameters $s_n = s_0 + x_1 + \cdots + x_n$ and $v_n = v_0 + n$. As clearly

$$\frac{s_n}{v_n} = \frac{v_0}{v_0+n}\frac{s_0}{v_0} + \frac{n}{v_0+n}\bar{x},$$

v_0 can be interpreted as the prior sample size, and s_n/v_n as a sample size weighted average of the "prior mean" s_0/v_0 and the sample mean $\bar{x}$.

The standard conjugate family $\mathcal{C}_\theta(\mathcal{F})$ relative to the canonical parameter θ has several important properties, in particular the linear relationship between the posterior mean and the sample mean (Diaconis and Ylvisaker, 1979).

We note that the densities $\pi(\lambda|s,v)$ are usually taken relative to the Lebesgue measure, which does not quite fit the needs of the commonly used polar coordinates (μ, κ) parameterization of the vMF family $\mathcal{F}$. Let us generally write η for the reference measure employed. Previous work using the (μ, κ) parameterization seem to take η as the product of the Lebesgue measure on $[0,\infty)$ (for κ) and the uniform distribution U on $\mathbb{S}^{d-1}$ (for μ), i.e., $d\eta \propto d\kappa\, dU(\mu)$. As for $\theta = \kappa\mu$ we have $d\theta = a_d \kappa^{d-1}\, d\kappa\, dU(\mu)$ (where a_d is the area of the unit hypersphere). The latter may be more natural as reference measure, turning the standard conjugate family relative to the polar coordinates (μ, κ) parameterization into the (obvious generalization) of what Gutiérrez-Peña and Smith (1997) call the DY-conjugate family for $\mathcal{F}$ relative to the parameterization.

Let $\mathcal{H}_{\lambda;\eta}$ denote the set of all hyperparameters s and v for which $\pi(\lambda|s,v)$ is a proper distribution on the employed parameter space Λ (using η as reference measure), i.e.,

$$\mathcal{H}_{\lambda;\eta} = \left\{ (s,v) : \int_\Lambda C_d(\|\theta(\lambda)\|)^v e^{s'\theta(\lambda)} d\eta(\lambda) < \infty \right\},$$