Journal of Multivariate Analysis 124 (2014) 345–352

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

On the Bingham distribution with large dimension

A. Kume *, S.G. Walker

Institute of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, UK
Department of Mathematics, and Division of Statistics and Scientific Computation, University of Texas at Austin, TX, United States

ARTICLE INFO

Article history:
Received 15 March 2013
Available online 11 November 2013

AMS subject classifications:
60E05
62H11

Keywords:
Bingham distribution
Dirichlet distribution
Mixture model
Series expansion

ABSTRACT

In this paper, we investigate the Bingham distribution when the dimension p is large. Our approach is to use a series expansion of the distribution from which truncation points can be determined yielding particular errors. A point of comparison with the approach of Dryden (2005) is highlighted.

1. Introduction

The Bingham distribution is constructed by constraining multivariate normal distributions to lie on the surface of a sphere of unit radius. Hence, if $x = (x_0, x_1, \ldots, x_p)$ are distributed according to such a distribution then the norm of x is 1. Hence, $x^2 = (x_0^2, x_1^2, \ldots, x_p^2)$ lies on the simplex and we will write $s_i = x_i^2$ for each i.

To elaborate, the Bingham distribution is obtained via a multivariate normal random vector in $\mathbf{R}^{p+1}$ constrained to lie on the unit sphere $\mathcal{S}^p$. In particular, for a given matrix Σ of dimension $(p+1) \times (p+1)$ the density with respect to the uniform measure $\mathrm{d}_{\mathcal{S}^p}(x)$ in $\mathcal{S}^p$ is given as

$$f(x;\, \Sigma) = \mathcal{C}(\Sigma)^{-1} \exp(x^\top \Sigma x)$$

where $\mathcal{C}(\Sigma)$ is the corresponding normalizing constant and $x \in \mathbf{R}^{p+1}$ such that $x^\top x = 1$. Note that the distribution $f(x;\, \Sigma)$ is the conditional distribution of some multivariate distributed vector in $\mathbf{R}^{p+1}$ with mean zero and covariance matrix $-\Sigma^{-1/2}$. The uniform measure in $\mathcal{S}^p$ is invariant with the orthogonal transformations; it can be easily shown that if X has density $f(x;\, \Sigma)$ then for each orthogonal matrix $V \in O(p+1)$, $Y = XV$ has density $f(y;\, V^\top \Sigma V)$. A special case from this family of distribution is that of the complex Bingham which can be applied to shape analysis of objects in 2 dimensions (see [3]). The inferential issues in such cases are possible via the likelihood approach.

For the general cases Kume and Wood [6] show that the saddle point approximation approach is highly accurate and robust for various parameter values.

However, our aim in this paper is to study the class of Bingham distributions for large p. The practical importance of these cases is closely related to the statistical inferential problems in images where the number of points of interest p is large and the invariance on location, rescale and possibly rotation is essential for inference. Recent work done in this direction is that of Dryden [2] who undertakes inference by considering the connection between the uniform measure on the infinite dimensional sphere with the Wiener processes and its generalizations. The key result which links these processes with

* Corresponding author at: Institute of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, UK.
E-mail addresses: a.kume@kent.ac.uk (A. Kume), s.g.walker@math.utexas.edu (S.G. Walker).

0047-259X/$ – see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jmva.2013.10.023

the sphere is the fact that if X is a multivariate normal distribution of dimension p with zero mean and variance–covariance matrix I_p/p then $|X| = 1+O_p(p^{-1/2})$ for large p. Therefore, in probability $|X|^2$ goes to 1 and therefore X approaches the sphere of large dimension p. Hence, under certain conditions, [2] removes the constraint on the multivariate normal distribution.

Our approach is based on a series expansion of the Bingham distribution in terms of $1/p$ and then by studying the limiting behaviour if p is large by primarily making sure that the vector of study X remains rescaled so that it remains on the sphere and then simply observing the corresponding limits as the sphere increases in dimension. Hence, our paper is to complement the findings of Dryden [2]. The aim is to find the density for large p correct to the order p^{-2} and we show this results in a standard mixture of Dirichlet distributions and hence inference is straightforward.

In Section 2 we describe the approach based on a series expansion using Dirichlet distributions. This gives us that under certain conditions we need a determinable finite mixture to achieve a certain level of approximation when p is large. With this information, in Section 3, we use a matrix type expansion and take the terms discovered in Section 2. This gives us directly a value for $\mathrm{E}(XX^\top)$ which forms a basis for estimation of Σ. We can at this point make a comparison with the approach of Dryden [2].

2. Expansion of the Bingham distribution

We will start with the Bingham distribution of dimension $p + 1$; so we have

$$\begin{aligned} f(x) &\propto \exp(x^\top \Sigma x)\mathbf{1}(x^\top x = 1) \\ &\propto \exp\left\{\sum_{i=0}^{p} \Sigma_{ii}x_i^2 - 2\sum_{i<j} \Sigma_{ij}x_i x_j\right\}\mathbf{1}(x^\top x = 1). \end{aligned}$$

Letting Σ_{00} be the largest diagonal element of Σ, we can write

$$x^\top \Sigma x = \Sigma_{00} - \sum_{i=1}^{p} a_i x_i^2 + 2\sum_{i<j} \Sigma_{ij}x_i x_j,$$

when $x^\top x = 1$, and where $a_i = \Sigma_{00} - \Sigma_{ii} \geq 0$. If we now transform to $(w_0, \ldots, w_p, s_0, \ldots, s_p)$, where $s_i = x_i^2$ and $w_i = x_i/|x_i|$, then

$$f(w, s) \propto \exp\left\{\sum_{i=1}^{p} a_i s_i - 2\sum_{i<j} \sqrt{b_{ij}s_i s_j}\, w_i w_j\right\}(s_0 \ldots s_p)^{-1/2},$$

where $b_{ij} = \Sigma_{ij}^2$. Hence,

$$f(s) \propto \exp\left\{\sum_{i=1}^{p} a_i s_i\right\}\sum_{w_0\ldots w_p\in\{-1,+1\}} \exp\left\{-2\sum_{i<j}\sqrt{b_{ij}s_i s_j}\, w_i w_j\right\}(s_0 \ldots s_p)^{-1/2}.$$

This expression for $f(w, s)$ has appeared in [5] in which it is used as a starting point for a Gibbs sampler in order to sample from $f(x)$. One can see that the form of $f(w_i|s)$ is not difficult to obtain. Hence, for what follows we will concentrate on expanding $f(s)$. We start by assuming Σ is diagonal.

2.1. Zero off-diagonal elements

Let us first deal with the case when $b_{ij} = 0$ for $i \neq j$. From now on, we will write the normalizing constant $\mathcal{C}(\Sigma)$ as $\mathcal{C}_p(\Sigma)$. We have the following Lemma:

Lemma 1. *For large p, and with $b_{ij} = 0$ for all $i \neq j$ and*

$$\sum_{i=1}^{p} a_i = O(p),$$

then it is that

$$f(s) = \mathcal{C}_p(\Sigma)\left\{\mathrm{Dir}(s;\, 1/2, 1/2, \ldots, 1/2) + (p+1)^{-1}\sum_{i=1}^{p} a_i \mathrm{Dir}(s;\, 1/2, 1/2+\delta_{i,1}, \ldots, 1/2+\delta_{i,p}) + O(p^{-2})\right\},$$

where $\delta_{i,k} = 1$ if $i = k$ and $\delta_{i,k} = 0$ if $i \neq k$.

Here $\mathcal{C}_p(\Sigma)$ is the normalizing constant and $\mathcal{C}_p(\Sigma) \to 1$ as $p \to +\infty$. Also, $\mathrm{Dir}(s;\, \alpha_0, \alpha_1, \ldots, \alpha_p)$ is the density function of the Dirichlet distribution with parameters $(\alpha_0, \alpha_1, \ldots, \alpha_p)$ see e.g. Chapter 49 of [4].

The proof to this and all lemmas and theorems are deferred to the Appendix.