On the Fisher–Bingham Distribution

BY A. Kume and S.G Walker
Institute of Mathematics, Statistics and Actuarial Science, University of Kent
Canterbury, CT2 7NF, UK
A.Kume@kent.ac.uk and S.G.Walker@kent.ac.uk

Abstract

This paper primarily is concerned with the sampling of the Fisher–Bingham distribution and we describe a slice sampling algorithm for doing this. A by-product of this task gave us an infinite mixture representation of the Fisher–Bingham distribution; the mixing distributions being based on the Dirichlet distribution. Finite numerical approximations are considered and a sampling algorithm based on a finite mixture approximation is compared with the slice sampling algorithm.

Keywords: Directional statistics; Fisher–Bingham distribution; Gibbs sampling.

1 Introduction.

The Bingham, and more generally the Fisher–Bingham distribution, are constructed by constraining multivariate normal distributions to lie on the surface of a sphere of unit radius. They are used in modeling spherical data which usually represent directions but in some cases they can also be used in shape analysis.
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. The contribution of this paper is to transform x to (ω, s) and to study the marginal and conditional distributions of ω and s. Here $s_i = x_i^2$ and $\omega_i = x_i/|x_i|$, so that $\omega_i \in \{-1, +1\}$.
Clearly the Lebesgue measure in $I\!R^{p+1}$ induces the uniform measure on the unit sphere $\mathcal{S}^p$. Hence the Fisher–Bingham distribution which is obtained in $\mathcal{S}^p$ via the constrained multivariate normal random vector in with covariance Σ and mean $\mu \in I\!R^{p+1}$ has the density with respect to uniform measure $\mathrm{d}_{\mathcal{S}^p}(x)$ in $\mathcal{S}^p$

$$f(x|\mu, \Sigma) = \mathcal{C}(\mu, \Sigma)^{-1} \exp\{-(x-\mu)^t \Sigma (x-\mu)\}$$

where $\mathcal{C}(\mu, \Sigma)$ is the corresponding normalizing constant and $x \in \mathrm{R}^{p+1}$ such that $x^t x = 1$. This distribution is an extension of the Bingham distribution which consist of μ being zero.
The uniform measure in $\mathcal{S}^p$ is invariant of the orthogonal transformations, it can be easily shown that if X has $\mathrm{FB}(\mu, \Sigma)$ then for each orthogonal matrix $V \in O(p+1)$, $Y = XV$ has $\mathrm{FB}(\mu V, V^t \Sigma V)$. So, without loss of generality, we assume that the covariance matrix is diagonal i.e. $\Sigma = \Lambda = \mathrm{diag}(\lambda_0, \lambda_1, \ldots, \lambda_p)$.

With $x^t x = 1$ we have in terms of $(\omega_0, \dots, \omega_p, s_1, \dots, s_p)$

$$\begin{aligned}(x-\mu)^t\Sigma(x-\mu) &= x^t\Sigma x - 2x^t\Sigma\mu + \mu^t\Sigma\mu \\ &= \sum_{i=0}^{p}\lambda_i s_i - 2\sum_{i=0}^{p}\lambda_i\omega_i\mu_i\sqrt{s_i} + \mu^t\Sigma\mu \\ &= \sum_{i=1}^{p}(-a_i s_i - b_i\omega_i\sqrt{s_i}) - b_0\omega_0\sqrt{1-s} + c\end{aligned}$$

where, for $i = 1, \dots, p$, $a_i = \lambda_0 - \lambda_i$, and for $i = 0, \dots, p$, $b_i = 2\lambda_i\mu_i$, with $c = \mu^t\Sigma\mu + \lambda_0$ and $s = 1 - s_0 = s_1 + \cdots + s_p$. Implementing the transformation from x to $(\omega_0, \dots, \omega_p, s_1, \dots, s_p)$ yields the joint density of interest given by

$$f(\boldsymbol{\omega}, \mathbf{s}) \propto \exp\left\{\sum_{i=1}^{p}(a_i s_i + b_i\omega_i\sqrt{s_i})\right\}\exp\left\{b_0\omega_0\sqrt{1-s}\right\}\prod_{i=1}^{p}s_i^{-1/2}\,(1-s)^{-1/2}\,\mathbf{1}(s \le 1). \qquad (1)$$

An important point which will come in useful later is to take, without loss of generality, λ_0 to be the largest of the λ's and so $a_i \ge 0$ for all $i = 1, \dots, p$.
Another important point to make is that the Bingham density (i.e. when $\mu = 0$) remains unchanged following any addition of a constant, say ξ, to all the diagonal elements of Σ. This follows since, with $x^t x = 1$,

$$x^t(\Sigma + \xi\, I)x = x^t\Sigma x + \xi.$$

In this case we can not expect to be able to estimate p λ_i's, rather it is only the differences, such as $\lambda_i - \lambda_{i'}$, that we can estimate. Note that this scenario also holds whether Σ is diagonal or not.
The function in (**??**) is the joint density of interest and from it we will make contributions about the Fisher–Bingham distribution by providing a mixture representation and simulating samples using Gibbs sampling from it.
In Section 2 we provide the mixture representation of the Fisher–Bingham distribution and demonstrate a method for truncating this to a finite mixture with known error. In some parameter cases a workable approximation can be obtained. In Section 3 we provide a Gibbs sampling approach to sampling the Fisher–Bingham distribution and hence compare distributions obtained from our mixture representation and the Gibbs sampling. This problem of sampling the Fisher–Bingham distribution was raised by a referee of the paper by Kume and Walker (2006).

2 Mixture representation of Fisher–Bingham distribution.

From the joint density of $(\boldsymbol{\omega}, \mathbf{s})$, it is clear that, for $i = 0, \dots, p$, we have

$$\mathrm{P}(\omega_i = 1|s_i) = \frac{\exp(b_i\sqrt{s_i})}{\exp(b_i\sqrt{s_i}) + \exp(-b_i\sqrt{s_i})}$$

and the ω_i are independent given the s_i. Hence this is easy to understand and so for the rest of this section we will concentrate on the marginal density of $\mathbf{s}$. This is given up to a constant of proportionality, and with respect to the Lebesgue measure $\mathrm{d}s_1 \dots \mathrm{d}s_p$, by

$$f(\mathbf{s}) \propto g(\mathbf{s}) = \left\{ \prod_{i=1}^{p} \exp(a_i s_i) \cosh(b_i \sqrt{s_i}) s_i^{-1/2} \right\} \cosh(b_0 \sqrt{1-s}) \, (1-s)^{-1/2} \mathbf{1} \, (s \leq 1) \, .$$

Note that $h(\mathbf{s}) = \prod_{i=1}^{p} s_i^{-1/2} (1-s)^{-1/2}$ is proportional to $\mathrm{Dir}(\mathbf{s}; 1/2, \ldots, 1/2)$, the pdf at $(s_1, \ldots, s_p)$ of the Dirichlet distribution with its $p+1$ parameters equal to 1/2.
Now both $\exp(\cdot)$ and $\cosh(\cdot)$ can be expanded in powers; so

$$\exp(a_i s_i) = \sum_{l=0}^{\infty} a_i^l s_i^l / l!$$

and

$$\cosh(b_i \sqrt{s_i}) = \sum_{m=0}^{\infty} b_i^{2m} s_i^m / (2m)!$$

leading to

$$g(\mathbf{s}) = \sum_{l_1=0}^{\infty} \cdots \sum_{l_p=0}^{\infty} \sum_{m_0=0}^{\infty} \cdots \sum_{m_p=0}^{\infty} w(l,m) \mathrm{Dir}(\mathbf{s}; l_1 + m_1 + 1/2, \ldots, l_p + m_p + 1/2, m_0 + 1/2),$$

where

$$w(l,m) = \frac{b_0^{2m_0} \Gamma(m_0 + 1/2)}{(2m_0)! \Gamma\left(\sum_{i=1}^{p} (l_i + m_i + 1/2) + m_0 + 1/2\right)} \prod_{i=1}^{p} \frac{b_i^{2m_i} a_i^{l_i} \Gamma(l_i + m_i + 1/2)}{l_i! (2m_i)!}.$$

Hence, $f(\mathbf{s})$ is an infinite mixture of Dirichlet distributions.

2.1 Finite approximation.

The idea here is to truncate the infinite mixture to a finite number of terms and to compute the error in such a procedure. To this end, let us define

$$w(l,n,\mathbf{s},\boldsymbol{\omega}) = \prod_{i=1}^{p} \left\{ l_i!^{-1} (a_i s_i)^{l_i} \, n_i!^{-1} (\omega_i b_i \sqrt{s_i})^{n_i} \right\} \, n_0!^{-1} (\omega_0 b_0 \sqrt{1-s})^{n_0} \, h(\mathbf{s}).$$

So $w(l,m)$ is the integral of $w(l, n = 2m, \mathbf{s}, \boldsymbol{\omega})$ with respect to $d\mathbf{s}$ and $d\boldsymbol{\omega}$, the uniform measure on $\{-1,1\}^{p+1}$. We now consider a direct expansion of the expression on the right-hand side of (**??**)

$$f(\boldsymbol{\omega}, \mathbf{s}) \propto \sum_{k=0}^{\infty} k!^{-1} \left\{ \sum_{i=1}^{p} (a_i s_i + b_i \omega_i \sqrt{s_i}) + b_0 \omega_0 \sqrt{1-s} \right\}^k h(\mathbf{s})$$

and see that these terms are related to $w(l,n,\mathbf{s},\boldsymbol{\omega})$ as

$$\sum_{l_1 + l_2 + \cdots + l_p + n_0 + \cdots + n_p = k} w(l,n,\mathbf{s},\boldsymbol{\omega}) = k!^{-1} \left\{ \sum_{i=1}^{p} (a_i s_i + b_i \omega_i \sqrt{s_i}) + b_0 \omega_0 \sqrt{1-s} \right\}^k h(\mathbf{s})$$

where the summation on the left is made for all possible integer partitions of k (including zeros) such that $l_1 + l_2 + \cdots + l_p + n_0 + \cdots + n_p = k$. We notice that

$$\left\{\sum_{i=1}^{p}(a_i s_i + b_i \omega_i \sqrt{s_i}) + b_0 \omega_0 \sqrt{1-s}\right\} \leq \sum_{i=1}^{p}(a_i s_i + |b_i|\sqrt{s_i}) + |b_0|\sqrt{1-s} \leq M\left(\sqrt{1-s} + 1 - s + \sum_{i=1}^{p}\sqrt{s_i}\right),$$

where $M = \max\{a_i; i = 1, \ldots, p,\ |b_i|; i = 0, \ldots, p\}$. Also

$$\sum_{i=0}^{p}\sqrt{s_i} \leq \sqrt{p+1}$$

and so

$$\sum_{l_1+l_2+\cdots+l_p+n_0+\cdots+n_p=k} w(l, n, \mathbf{s}, \boldsymbol{\omega}) \leq k!^{-1} M^k (1 + \sqrt{p+1})^k\, h(\mathbf{s}).$$

Note that, if any of n_i's is odd then the integral of $w(l, n, \mathbf{s}, \boldsymbol{\omega})$ with respect to $d\boldsymbol{\omega}$ is zero. Therefore, collecting only the non-zero terms left the after integrating over $d\boldsymbol{\omega}$ and d$\mathbf{s}$ in the summation above, we obtain

$$\sum_{l_1+l_2+\cdots+l_p+2m_0+\cdots+2m_p=k} w(l, m) \leq k!^{-1} M^k (1 + \sqrt{p+1})^k\, \frac{\Gamma(1/2)^{p+1}}{\Gamma(p/2 + 1/2)}.$$

Hence,

$$T_N = \sum_{l_1+l_2+\cdots+l_p+2m_0+\cdots+2m_p \geq N} w(l, m) \leq \sum_{k=N}^{\infty} \tau(p)\, M^k \frac{(1 + \sqrt{p+1})^k}{k!},$$

where $\tau(p) = \Gamma(1/2)^{p+1}/\Gamma(p/2 + 1/2)$, leading to

$$T_N \leq \phi(N) = \frac{\tau(p)\, M^N (1 + \sqrt{p+1})^N}{N!} e^{M(1+\sqrt{p+1})}.$$

We can find a value of N, which will form the basis of our truncation, based on the decreasing speed of $\phi(N)$. For a given error ϵ we find N such that $T_N \leq \phi(N) \leq \epsilon$ and so an approximation to the Fisher–Bingham distribution has density function for $\mathbf{s} = (s_1, \ldots, s_p)$ given by

$$f_N(s) = \sum_{l_1=0}^{N-1} \cdots \sum_{l_p=0}^{N-1} \sum_{m_0=0}^{[N/2]} \cdots \sum_{m_p=0}^{[N/2]} q_N(l, m) \mathrm{Dir}(\mathbf{s}; l_1 + m_1 + 1/2, \ldots, l_p + m_p + 1/2, m_0 + 1/2),$$

where

$$q_N(l, m) = \frac{w(l, m)}{\sum_{l_1=0}^{N-1} \cdots \sum_{l_p=0}^{N-1} \sum_{m_0=0}^{[N/2]} \cdots \sum_{m_p=0}^{[N/2]} w(l, m)}.$$

This works since

$$\{(l, m) : 1 \leq l_i, 2m_i < N\}^c \subset \left\{(l, m) : \sum_i l_i + 2\sum_i m_i \geq N\right\}.$$

With this approximation statistical inference becomes highly feasible based on a mixture distribution for which estimating methods are well documented. For example, Titterington et al. (1988). Problems of working directly with the Fisher–Bingham distribution are that the general expression for the normalising constant is unknown and hence the exact likelihood function is unavailable. The summation of $w(l, m)$ is in fact the normalising constant for the Fisher–Bingham distribution. Previous attempts at computing the normalizing constant have been given in Kent (1987), who dealt with the Bingham distribution, and Kume and Wood (2005). Our approach is clearly computationally intensive in some instances when N needs to be large; though highly accurate estimates can be found in many cases with a small value of N. Additionally, we can in such cases know the accuracy of the normalising constant and hence use this to consider the performance of alternative approximations. In particular, in Table 1, we compare estimates of the normalising constant with those obtained by Kume and Wood (2005), who use saddle–point approximations. It is also possible to simulate approximately from the Fisher–Bingham distribution via sampling the weights q_N. However, an exact method is possible via Markov chain Monte Carlo methods; in particular using the Gibbs sampling with slice sampling (Damien et al., 1999). This is described in Section 3.

2.2 Large p.

When p is large, and for this we will assume that $p = 50$, the approximation works well with small N. Recall we are interested in finding N such that

$$\phi(N) = \frac{\tau(p)\, M^N (1+\sqrt{p})^N}{N!} e^{M(1+\sqrt{p+1})} < \epsilon.$$

Now, assuming $M = 1$,

$$\phi(N) = \frac{\tau(51)\,(1+\sqrt{51})^N}{N!} e^{(1+\sqrt{51})} = \frac{8.7 \times 10^{-13} \times 8.1^N}{N!} \times 3200$$

and clearly for $N = 2$ this turns out to be very small. In this case, the approximation is well served by only considering terms for which

$$l_1 + \cdots + l_p + m_0 + \cdots + m_p \leq 1$$

in which case our approximation becomes

$$\widehat{f}(\mathbf{s}) \propto w_0 \,\mathrm{Dir}(\mathbf{s}; 1/2, \ldots, 1/2) + \sum_{k=1}^{2p+1} w_k \mathrm{Dir}(\mathbf{s}; 1/2 + \delta_{1,k}, \ldots, 1/2 + \delta_{p,k}, 1/2 + \delta_{p+1,k}),$$

where $w_0 = w(0, 0)$ and w_k is $w(l, m)$ with the kth element of $(l_1, \ldots, l_p, m_0, \ldots, m_p)$ set to 1 and the rest set to 0. Also, for $j = 1, \ldots, p$,

$$\delta_{j,k} = \mathbf{1}(k = j) + \mathbf{1}(k = j + p)$$

and $\delta_{p+1,k} = \mathbf{1}(k = 2p + 1)$. It is quite easy to see that one can take out a common term $w_0 = \Gamma(1/2)^{p+1}/\Gamma((p + 1)/2)$ from all the w's leaving the terms to decay as $1/p$. That is, $w(l, m)$ behaves as $w(0, 0)/p^n$ where

$$n = l_1 + \cdots + l_p + m_0 + \cdots + m_p.$$

Inference for this is quite straightforward. This would be an alternative approach to dealing with large p for the Fisher–Bingham distribution to that given by Dryden (2005).

3 Gibbs sampling of Fisher–Bingham distribution.

In this section we describe a slice sampling algorithm for sampling from the Fisher–Bingham distribution. This is a natural yet non-trivial extension from the sampling algorithm which appears in Kume and Walker (2006) which was developed to sample the Bingham distribution. The Bingham distribution arises by taking $b_i = 0$ for all i in the Fisher-Bingham distribution. In the Bingham case it was not necessary to find the joint for $(\boldsymbol{\omega}, \mathbf{s})$, we simply worked with $\mathbf{s}$, since the sign of the original variables $\{x_0, \dots, x_p\}$ is independent of their squares.

Similar to what was shown in Kume and Walker (2006), by thinning the output of the Gibbs method shown here we can obtain practically independent and identically distributed samples from Fisher–Bingham.

Let us consider the density in (**??**) which is to be sampled. We introduce 3 latent variables (u, v, w) and construct the joint density with $(\boldsymbol{\omega}, \mathbf{s})$ given by

$$f(\boldsymbol{\omega}, \mathbf{s}, u, v, w) \propto \mathbf{1}\{u < \exp(L)\}\, \mathbf{1}\{v < \exp(b_0 \omega_0 \sqrt{1-s})\}\, \mathbf{1}\{w < (1-s)^{-1/2}\} \prod_{i=1}^{p} s_i^{-1/2}\, \mathbf{1}(s \leq 1),$$

where

$$L = \sum_{i=1}^{p} (a_i s_i + b_i \omega_i \sqrt{s_i}).$$

Clearly, the full conditional densities for u, v and w are easily seen to be uniform and are therefore easy to sample. The full conditional mass function for ω_i, $i = 0, \dots, p$, is also easy to obtain, and

$$\mathrm{P}(\omega_i = +1 | \cdots) = \frac{\exp(b_i \sqrt{s_i})}{\exp(-b_i \sqrt{s_i}) + \exp(b_i \sqrt{s_i})}$$

noting that $s_0 = 1 - s_1 - \cdots - s_p$.

The more complicated full conditional densities to sample are those for s_i, $i = 1, \dots, p$. Without loss of generality we will consider the full conditional density for s_1; the others will follow by switching indices in a cyclic fashion.

It is that

$$f(s_1 | \cdots) \propto \mathbf{1}\{A_u \cap A_v \cap A_w \cap A\}\, s_1^{-1/2},$$

where A_y is a set formed by inverting the inequality involving $y = (u, v, w)$ and

$$A = \left\{ 0 < s_1 < 1 - \sum_{2 \leq i \leq p} s_i \right\}.$$

So,

$$A_w = \{s_1 : (1-s)^{-1/2} > w\} = \left\{ s_1 > 1 - \sum_{2 \leq i \leq p} s_i - w^{-2} \right\}.$$

Also,

$$A_u = \{s_1 : u < \exp(L)\} = \{s_1 : a_1 s_1 + b_1 \omega_1 \sqrt{s_1} > d\},$$

where

$$d = \log u - \sum_{i=2}^{p}(a_i s_i + b_i \omega_i \sqrt{s_i}).$$

Therefore, since $a_1 > 0$

$$A_u = \{s_1 : l_u < \sqrt{s_1}\} \cup \{s_1 : \sqrt{s_1} < t_u\},$$

where

$$l_u = \left\{\sqrt{d/a_1 + (b_1/(2a_1))^2} - b_1\omega_1/(2a_1)\right\}$$

and

$$t_u = \left\{-\sqrt{d/a_1 + (b_1/(2a_1))^2} - b_1\omega_1/(2a_1)\right\}.$$

If $d/a_1 + (b_1/(2a_1))^2 < 0$ then $A_u = (0, 1)$ and clearly, if $a_1 = 0$ then $A_u = \{s_1 : b_1\omega_1\sqrt{s_1} > d\}$. Finally, we have

$$A_v = \left\{s_1 : v < \exp(b_0\omega_0\sqrt{1-s})\right\}.$$

There are two scenarios here;

(i) $b_0\omega_0 < 0$: in this case we have

$$A_v^+ = \left\{s_1 > 1 - \sum_{2\leq i\leq p} s_i - ((-\log v)/b_0)^2\right\}.$$

(ii) $b_0\omega_0 > 0$: care is needed here. If $v < 1$ then there is no constraint from this. If $v > 1$ then we have

$$A_v^- = \left\{s_1 < 1 - \sum_{2\leq i\leq p} s_i - ((\log v)/b_0)^2\right\}.$$

Given that we will intersect this with the set A, we may as well take

$$A_v^- = \left\{s_1 < 1 - \sum_{2\leq i\leq p} s_i - \mathbf{1}(v > 1)((\log v)/b_0)^2\right\}.$$

Hence, $A_v = A_v^+$ if $b_0\omega_0 > 0$ and $A_v = A_v^-$ if $b_0\omega_0 < 0$.
Writing the intersection of the sets as (ξ_l, ξ_u) then

$$f(s_1|\cdots) \propto \mathbf{1}\{s_1 \in (\xi_l, \xi_u)\}s_1^{-1/2}.$$

It is therefore possible to sample this density by taking

$$s_1 = \left\{\tau(\xi_u^{1/2} - \xi_l^{1/2}) + \xi_l^{1/2}\right\}^2,$$

where τ is a uniform r.v. from the interval $(0, 1)$.

The aim here is to compare densities obtained for s_1 from both approaches; namely using samples from the Gibbs sampler and the approximation based on a mixture representation. Marginally, it is clear that s_1 will be a mixture of beta distributions and that $f(s_1) \propto g(s_1)$ and

$$g(s_1) = \sum_{l_1=0}^{\infty} \cdots \sum_{l_p=0}^{\infty} \sum_{m_0=0}^{\infty} \cdots \sum_{m_p=0}^{\infty} w(l,m) \mathrm{Be}\left(s_1;\, l_1 + m_1 + 1/2,\, m_0 + 1/2 + \sum_{i \neq 1} (l_i + m_i + 1/2) \right).$$

The normalising constant will be the reciprocal of

$$\sum_{l_1=0}^{\infty} \cdots \sum_{l_p=0}^{\infty} \sum_{m_0=0}^{\infty} \cdots \sum_{m_p=0}^{\infty} w(l,m).$$

With the approximation outlined in Section 2, we take $N = 5$ with $p = 3$ and take $a_1 = 0.1$, $a_2 = 0.3$, $a_3 = 2.0$, $b_0 = 0$, $b_1 = 1.0$, $b_2 = 0.9$ and $b_3 = 0.3$. The numerical construction of the marginal density of s_1 is given in Figure 2 together with the histogram of a random sample of size 10^5 from s_1 generated using our Gibbs method. Note it is easy to see that $f(0) = \infty$ due to the term of the mixture with $l_1 = 0$ and $m_1 = 0$.

4 Discussion

The key to this paper is the introduction of the variable ω into the Fisher–Bingham distribution. It serves two purposes. The first is that we can use it to facilitate a Gibbs sampling approach to sample from the Fisher–Bingham distribution. The second is that it allows a mixture representation of the Fisher–Bingham distribution.

For the former, it is clear from equation (**??**) that each s_i has a non-central chi–squared distribution with 1 degree of freedom, and therefore the simulation method described could be easily adopted to generate samples from convolutions of such distributions constrained to have sum 1. Such a distribution would generalise those considered by Fearnhead and Meligkotsidou (2004).

For the latter, the mixture representation of the Fisher–Bingham distribution permits a finite approximation based on evaluation of the error. Once a finite mixture is constructed, inference becomes quite straightforward; whereas for the Fisher–Bingham distribution the normalising constant does not always have a closed form. Interestingly, as p gets large, the value of N required to get useful approximations decreases and so the terms in the mixture are manageable.

Our mixture representation suggests a number of generalizations to constructing distributions on the sphere. One is to allow the weights to be arbitrary, rather then based on the $w(l,m)$ construction. Another is to assign a probability distribution on N.

We could also attempt to approximate sampling from the Fisher–Bingham distribution by sampling the weights $w(l,m)$, though this would probably end up being quite messy.

References

Damien, P., Wakefield, J.C. & Walker, S.G. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *J. R. Statist. Soc. B* **61**, 331–344.

Dryden, I.L. (2005). Statistical analysis on high–dimensional spheres and shape spaces. *Ann. Statist.* **33**, 1643–1665.

De Waal, D.J. (1979). On the normalizing constant for the Bingham-von Mises-Fisher matrix distribution. *South African Statist.*, **13**, *103–112.*

Fearnhead, P. & Meligkotsidou, L. (2004). Exact filtering for partially observed continuous time models. J. R. Statist. Soc. B **66**, *771–790.*

Kent, J.T., Constable, P.D.L. & Er, F. (2004). Simulation for the complex Bingham distribution. Statistics and Computing **14**, *53–57.*

Kent, J.T. (1987) Asymtotic expansion for the Bingham distribution, Appl. Statistics, ***36***, *139–44.*

Kume, A. & Walker, S.G. (2006). Sampling from compositional and directional distributions. Statistics and Computing **16**, *261 – 265.*

Kume, A. & Wood, A. T.A. (2005). Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. Biometrika, ***92***,*465–476.*

Titterington, D.M., Smith, A.F.M. & Makov, U.E. (1988). Statistical Analysis of Finite Mixture Distributions. Wiley, New York.

	p=2		p=3		p=4	
c	Ex	Sp	Ex	Sp	Ex	Sp
3	3.85	3.84	6.57	6.55	8.80	8.78
4	4.34	4.34	7.70	7.69	10.72	10.70
5	4.67	4.67	8.47	8.47	12.08	12.07

Table 1: Exact (correct to 3 decimal places) and saddle point approximations for a range of choices of p and c where $a = 1/c, 2/c, \ldots, p/c$ with $p = 2, 3, 4$ and $c = 3, 4, 5$

Figure 1: The numerical approximate marginal density function of s_1 with $N = 5$.