On the explicit Pffafian of the Fisher-Bingham normalizing constant

Alfred Kume & Tomonari Sei

Kent University, UK & Keio University, Japan

May 20-22, 2014. ADISTA 2014

Introduction

The Fisher-Bingham (FB) distribution (Mardia 1975) is defined on $S^{p-1} = \{x \in \mathbb{R}^p | x^\top x = 1\}$ as

$$f(x|\mu, \Sigma) \propto e^{\frac{(x-\mu)^\top \Sigma^{-1}(x-\mu)}{2}} \mathbf{1}(x^\top x = 1)$$
$$= \frac{e^{-x^\top \frac{\Sigma^{-1}}{2} x + x^\top \Sigma^{-1}\mu}}{\mathcal{C}(\frac{\Sigma^{-1}}{2}, \Sigma^{-1}\mu)} \mathbf{1}(x \in S^{p-1}),$$

where $\frac{\Sigma^{-1}}{2} = O^\top \Delta O \in \mathbb{R}^{p\times p}, \Sigma^{-1}\mu = O^\top \gamma \in \mathbb{R}^p$ and

$$\mathcal{C}(\frac{\Sigma^{-1}}{2}, \Sigma^{-1}\mu) = \int_{\mathcal{S}^{p-1}} e^{-x^\top \frac{\Sigma^{-1}}{2} x + x^\top \Sigma^{-1}\mu} d_{\mathcal{S}^{p-1}}(x) = \mathcal{C}(\Delta, \gamma)$$

Introduction: Applications of F-B distributions

- Directional Statistics: wind direction (2-dim); magnetism (3-dim); locations of stars (3-dim) (see Mardia & Jupp, 2000)
- Shape analysis (general p) Complex Bingham (Kent 1995) Size-and-shape distributions (Goodall and Mardia 1992); matching problems in Bioinformatics Mardia and Green 2006, EM on MLE (Kume and Dryden 2014).
- Compositional data square-root transformation of (high-dim) e.g. Scealy and Welsh (2011)

Introduction

- Log-likelihood function for observed data $X = (x_1, x_2, ..., x_n)$

$$\log L(\frac{\Sigma^{-1}}{2}, \Sigma^{-1}\mu, X) = -n \log \mathcal{C}(\frac{\Sigma^{-1}}{2}, \Sigma^{-1}\mu) - \sum_{i=1}^{n} x_i^\top \frac{\Sigma^{-1}}{2} x_i + x_i^\top \Sigma^{-1}\mu$$
$$= -n \log \mathcal{C}(\Delta, \gamma) - \sum_{i=1}^{n} (Ox_i)^\top \Delta O x_i + (Ox_i)^\top \gamma$$
$$= \log L(\Delta, \gamma, OX)$$

- MLE estimation with parametrization $\Delta = diag(\boldsymbol{\theta})$ is

$$(\hat{\boldsymbol{\theta}}, \hat{\gamma}, \hat{O}) = \underset{(\theta,\gamma,O)}{\operatorname{argmax}} \log L(diag(\boldsymbol{\theta}), \gamma, OX).$$

•Problem: The normalizing constant

$$\mathcal{C}(\theta,\gamma)=\int_{S^{p-1}} e^{-x^\top diag(\theta)x+x^\top\gamma} d_{\mathcal{S}^{p-1}}(x)=\sum_{k=0}^{\infty}\frac{E_x(-x^\top diag(\theta)x+x^\top\gamma)^k}{k!}$$

has no closed form and $\mathcal{C}(\theta,\gamma)e^c=\mathcal{C}(\theta-c,\gamma)$

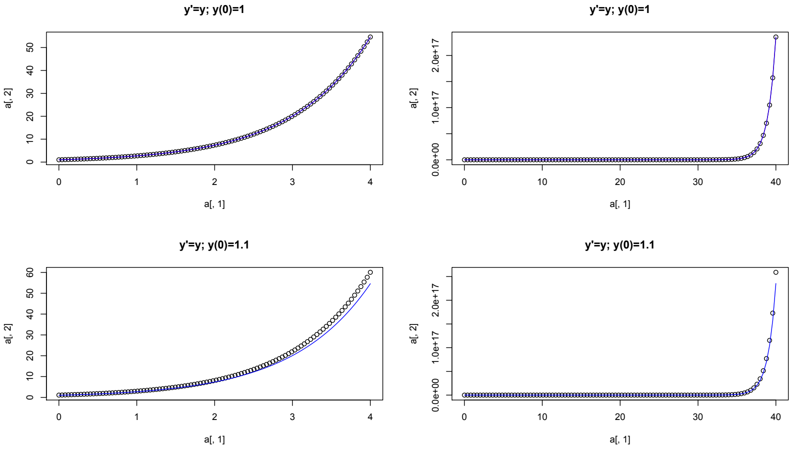
- Saddle point approximation by Kume and Wood (2005). It is shown that $\mathcal{C}(\theta,\gamma)$ is related to the density of a random variable defined as a linear combination of the p independent noncentral χ_1^2 random variables.
- Dirichlet mixture representation by Kume and Walker (2009).

Special cases of Fisher-Bingam distributions

- Bingham distribution, is generated if γ is set to zero.
- Fisher-Watson if $\theta_2 = \theta_3 = \cdots \theta_p$
- Kent distributions if $\gamma_2 = \gamma_3 = \cdots \gamma_p = 0$
- von Mises-Fisher if $\theta_1 = \theta_2 = \cdots = \theta_p$
- Bingham-Mardia if $\theta_2 = \theta_3 = \cdots \theta_p$ $\gamma_2 = \gamma_3 = \cdots \gamma_p = 0$
- Watson if $\theta_2 = \theta_3 = \cdots \theta_p$ and $\gamma_2 = \gamma_3 = \cdots \gamma_p = 0$

(see Mardia & Jupp, 2000, Table 9.2)

Taylor expansion vs ODE with *correct* starting point



Definition

A function $\mathcal{C}(\boldsymbol{\alpha})$ of $\boldsymbol{\alpha} \in \mathbb{R}^d$ is called holonomic if there is a finite-dim vector $\mathbf{g}(\boldsymbol{\alpha})$ of derivatives of $\mathcal{C}(\boldsymbol{\alpha})$ that satisfies

$$\frac{\partial}{\partial \alpha_i}\mathbf{g} = \mathbf{P}_i(\boldsymbol{\alpha})\mathbf{g}, \quad i = 1, \ldots, d,$$

where $\mathbf{P}_i(\boldsymbol{\alpha})$ is a matrix of rational functions (satisfying integrable conditions). This equation is called the Pfaffian system.

- Nakayama et al. (2011) showed that $\mathcal{C}(A, b)$ is holonomic, and derived the Pfaffian system for $p = 2, 3$.
- Koyama et al. (2013) gave the Pfaffian system for general p in a sophisticated way: the system is given in an implicit symbolic form and solved in numerical computation.

The HG algorithm

- Input: $\boldsymbol{\alpha}^{(0)}$, $\mathbf{g}(\boldsymbol{\alpha}^{(0)})$ and $\boldsymbol{\alpha}^{(1)}$.
- Output: $\mathbf{g}(\boldsymbol{\alpha}^{(1)})$.

Algorithm (HG)

- *Let* $\bar{\boldsymbol{\alpha}}(\tau) = (1-\tau)\boldsymbol{\alpha}^{(0)} + \tau\boldsymbol{\alpha}^{(1)}$ *and* $\bar{\mathbf{g}}(\tau) = \mathbf{g}(\bar{\boldsymbol{\alpha}}(\tau))$.
- *Solve the initial value problem*

$$\frac{d}{d\tau}\bar{\mathbf{g}}(\tau) = \sum_{i=1}^{d} \frac{d\bar{\alpha}_i(\tau)}{d\tau} \underbrace{\mathbf{P}_i(\bar{\boldsymbol{\alpha}}(\tau))\bar{\mathbf{g}}(\tau)}_{(\partial_i \mathbf{g})(\bar{\boldsymbol{\alpha}}(\tau))}, \quad \bar{\mathbf{g}}(0) = \mathbf{g}(\boldsymbol{\alpha}^{(0)}),$$

by standard numerical routines, and then return $\bar{\mathbf{g}}(1)$.

The HG algorithm

- For the HG algorithm, we need to compute initial values $\mathbf{g}(\boldsymbol{\alpha}^{(0)})$ at an appropriate initial point $\boldsymbol{\alpha}^{(0)}$.
- If $\boldsymbol{\alpha}^{(0)}$ is small, then $\mathbf{g}(\boldsymbol{\alpha}^{(0)})$ is efficiently calculated by the power series expansion with an appropriate truncation.
- For the general FB family $\mathcal{C}(\boldsymbol{\theta}, \boldsymbol{\gamma})$, the power series and its truncation formula are obtained by Koyama et al. (2013). This formula is applicable to our case.
- Note: the number of terms in the expansion is reduced if we put $\boldsymbol{\gamma}^{(0)} = 0$ i.e. the starting point is the Bingham normalizing constant.

Relevant results for the Fisher-Bingham normalizing constant

From Kume and Wood (2005) we derive

$$\mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma}) = \int_{\mathcal{S}^{p-1}} e^{-\sum_{i=1}^{p}(\theta_i x_i^2+\gamma_i x_i)} d_{\mathcal{S}^{p-1}}(\mathbf{x}) = \int_{\mathrm{i}\mathbb{R}+t_0} \mathcal{A}(\boldsymbol{\gamma},\boldsymbol{\theta}) e^{-t} dt \tag{1}$$

$$\mathcal{A}(\boldsymbol{\gamma},\boldsymbol{\theta}) = \frac{2\pi^{p/2}}{2\pi i} \prod_{i=1}^{p} \frac{e^{\frac{\gamma_i^2}{4(\theta_i - t)}}}{\sqrt{\theta_i - t}}$$

$$\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ...\gamma_p), \boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)$$

From (1), we can now derive

- the Pfaffian system of $\mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})$ for general $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$.
- the Pfaffian system of $\mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})$ for degenerate cases when some entries of $\boldsymbol{\theta}$ coincide and/or some entries of $\boldsymbol{\gamma}$ are zero.

Explicit Pffafian: general case

Define now
$\mathbf{g} = \left(\mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma}), \frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial \theta_1}, \cdots, \frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial \theta_p}, \frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial \gamma_1}, \cdots, \frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial \gamma_p}\right)$. Pfaffian system expresses of the second order derivatives of $\mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})$ in terms of those of $\mathbf{g}$.

$$
\begin{aligned}
\sum_{i=1}^{p} \frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial \theta_i} &= -\mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma}) \\
\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial \gamma_i \partial \theta_i} &= -\sum_{i \neq j=1}^{p} \frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial \theta_j \partial \gamma_i} - \frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial \gamma_i} \\
\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial^2 \theta_i} &= -\sum_{i \neq j=1}^{p} \frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial \theta_i \partial \theta_j} - \frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial \theta_i} \\
\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial^2 \gamma_i} &= \frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial \theta_i}
\end{aligned}
$$

If $i \neq j$

$$\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_i\partial\gamma_j} = \frac{\gamma_j}{2(\theta_j-\theta_i)}\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_i} - \frac{\gamma_i}{2(\theta_j-\theta_i)}\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_j} \tag{2}$$

$$\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\theta_i\partial\theta_j} = -\left(\frac{1}{2(\theta_j-\theta_i)} + \frac{\gamma_j^2}{4(\theta_j-\theta_i)^2}\right)\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\theta_i} - \left(\frac{1}{2(\theta_i-\theta_j)} + \frac{\gamma_i^2}{4(\theta_i-\theta_j)^2}\right)\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\theta_j} \tag{3}$$

$$- \left(\frac{\gamma_i}{4(\theta_j-\theta_i)^2} + \frac{\gamma_i\gamma_j^2}{4(\theta_j-\theta_i)^3}\right)\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_i} - \left(\frac{\gamma_j}{4(\theta_i-\theta_j)^2} + \frac{\gamma_i^2\gamma_j}{4(\theta_i-\theta_j)^3}\right)\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_j} \tag{4}$$

$$\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_i\partial\theta_j} = \frac{\gamma_i}{2(\theta_i-\theta_j)}\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\theta_j} - \left(\frac{1}{2(\theta_j-\theta_i)} + \frac{\gamma_j^2}{4(\theta_j-\theta_i)^2}\right)\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_i} + \frac{\gamma_i\gamma_j}{4(\theta_i-\theta_j)^2}\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_j} \tag{5}$$

Remark: This Pfaffian system is a smooth function at $\boldsymbol{\gamma} = 0$. Therefore the corresponding ODE is not degenerate at this point.

Main idea of proving the second part

The key to the proof is $\theta_i \neq \theta_j$ and $\gamma_i \neq 0 \neq \gamma_j$ then the second order derivatives are in this form

$$\int_{i\mathbb{R}+t_0} \left(\frac{A}{\theta_i - t} + \frac{B}{(\theta_i - t)^2}\right)\left(\frac{C}{\theta_j - t} + \frac{D}{(\theta_j - t)^2}\right) \mathcal{A}(\boldsymbol{\gamma}, \boldsymbol{\theta}) e^{-t} dt$$

$$= \int_{i\mathbb{R}+t_0} \left(\frac{a}{(\theta_i - t)} + \frac{b}{(\theta_i - t)^2} + \frac{c}{(\theta_j - t)} + \frac{d}{(\theta_i - t)^2}\right) \mathcal{A}(\boldsymbol{\gamma}, \boldsymbol{\theta}) e^{-t} dt$$

where a, b, c, d are as in the partial fractions identities and the fact that

$$\frac{\partial \mathcal{C}(\boldsymbol{\theta}, \boldsymbol{\gamma})}{\partial \theta_i} = -\int_{i\mathbb{R}+t_0} \left(\frac{1}{2(\theta_i - t)} + \frac{\gamma_i^2}{4(\theta_i - t)^2}\right) \mathcal{A}(\boldsymbol{\gamma}, \boldsymbol{\theta}) e^{-t} dt$$

$$\frac{\partial \mathcal{C}(\boldsymbol{\theta}, \boldsymbol{\gamma})}{\partial \gamma_i} = \int_{i\mathbb{R}+t_0} \frac{\gamma_i}{2(\theta_i - t)} \mathcal{A}(\boldsymbol{\gamma}, \boldsymbol{\theta}) e^{-t} dt$$

If $\theta_i = \theta_j$ for some $i \neq j$, then the Pfaffian becomes singular.

Degenerate case

Let assume that there are only q distinct values such that

$$\boldsymbol{\theta} = (\underbrace{\theta_1, \ldots, \theta_1}_{d_1}, \ldots, \underbrace{\theta_q, \ldots, \theta_q}_{d_q}) \quad d_1 + d_2 + \ldots + d_q = p$$

$$\mathcal{C}(\boldsymbol{\theta}, \gamma) = \frac{2\pi^{p/2}}{2\pi i} \int_{\mathrm{i}\mathbb{R}+t_0} \prod_{i=1}^{q} \frac{e^{\frac{\sum_{r=1}^{d_i} \gamma_{r,i}^2}{4(\theta_i - t)}}}{(\theta_i - t)^{d_i/2}} e^{-t} dt$$

with $\boldsymbol{\gamma} = (\underbrace{\gamma_1, 0, \ldots, 0}_{d_1}, \ldots, \underbrace{\gamma_q, 0 \ldots, 0}_{d_q})$

- In general theory of holonomic functions, this singularity can be removed by the restriction algorithm.
- For our case, this is done with minimal effort
- $dim(\mathbf{g}) = 2q$.

If $i \neq j$

$$\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_i\partial\gamma_j} = \frac{\gamma_j}{2(\theta_j-\theta_i)}\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_i} - \frac{\gamma_i}{2(\theta_j-\theta_i)}\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_j} \tag{6}$$

$$\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\theta_i\partial\theta_j} = -\left(\frac{d_j}{2(\theta_j-\theta_i)}+\frac{\gamma_j^2}{4(\theta_j-\theta_i)^2}\right)\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\theta_i} - \left(\frac{d_i}{2(\theta_i-\theta_j)}+\frac{\gamma_i^2}{4(\theta_i-\theta_j)^2}\right)\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\theta_j} \tag{7}$$

$$- \left(\frac{d_j\gamma_i}{4(\theta_j-\theta_i)^2}+\frac{\gamma_i\gamma_j^2}{4(\theta_j-\theta_i)^3}\right)\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_i} - \left(\frac{d_i\gamma_j}{4(\theta_i-\theta_j)^2}+\frac{\gamma_i^2\gamma_j}{4(\theta_i-\theta_j)^3}\right)\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_j} \tag{8}$$

$$\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_i\partial\theta_j} = \frac{\gamma_i}{2(\theta_i-\theta_j)}\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\theta_j} - \left(\frac{d_j}{2(\theta_j-\theta_i)}+\frac{\gamma_j^2}{4(\theta_j-\theta_i)^2}\right)\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_i} + \frac{\gamma_i\gamma_j}{4(\theta_i-\theta_j)^2}\frac{\partial \mathcal{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}{\partial\gamma_j} \tag{9}$$

Expressions for $\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\gamma)}{\partial^2\theta_i}$ and $\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\gamma)}{\partial\theta_i\partial\gamma_i}$ are the same as in the general case except

$$\frac{\partial^2 \mathcal{C}(\boldsymbol{\theta},\gamma)}{\partial^2\gamma_i} = -\frac{\partial \mathcal{C}(\boldsymbol{\theta},\gamma)}{\partial\theta_i} - \frac{d_i-1}{\gamma_i}\frac{\partial \mathcal{C}(\boldsymbol{\theta},\gamma)}{\partial\gamma_i}$$

Remark: This Pfaffian system is a not smooth at $\boldsymbol{\gamma} = 0$.

Bingham Distribution

Table : Computational time [sec] of the PS (power series) and HG algorithms. The parameter values examined are $\theta = (a(p-i)^b)_{i=1}^p$.

p	a	b	$\|\theta\|_1$	$C(\theta)/C(0)$	PS	HG
5	1/20	1	1/2	1.105961	0.1	0.3
5	1/10	1	1	1.224897	0.2	0.3
5	1	1	10	9.769432	17.1	0.3
5	10	1	100	3.824×10^{14}	NA	0.3
5	1/60	2	1/2	1.106713	0.1	0.3
5	1	2	30	5.253880×10^{4}	48.6	0.3
10	1/90	1	1/2	1.051360	14.0	14.8
10	1/45	1	1	1.105546	49.7	14.7
10	2/45	1	2	1.223062	386.2	14.6
10	1	1	45	1.757059×10^{2}	NA	14.6
10	1/570	2	1/2	1.051466	13.9	14.1
10	1	2	285	3.802×10^{28}	NA	15.2

Numerical evidence

Table : compares the saddle point approximation (SPA) and HG. The parameter θ is $(0, -1, -2, -\kappa)$.

κ	SPA	HG
5	4.237006	4.238950
10	2.982628	2.985576
30	1.708766	1.711919
50	1.321178	1.323994
100	0.932895	0.935094
200	0.659185	0.660814

Degenerate case: Complex Bingham

- As a corollary of the theorem, we obtain the Pfaffian system for the complex Bingham distribution.
- This is a special case of the Bingham distribution with multiplicities $d_1 = \cdots = d_q = 2$.
- It is known that $\mathcal{C}(\boldsymbol{\theta},))$ has a closed expression (Kent 1994).

Table : SPA, exact value and HG for $\theta = (0, 0, -1, -1, -2, -2, -\kappa, -\kappa)$.

κ	SPA	exact	HG
5	5.942975	5.936835	5.936835
10	3.429004	3.425468	3.425468
30	1.248280	1.246421	1.246421
50	0.761347	0.760180	0.760180
100	0.385272	0.384675	0.384675
200	0.193779	0.193477	0.193477

HG and Procrustes algorithm in perturbation models

$X_i = \mathbf{R}_i \Delta_i \mathbf{O}_i$ generated from $\mathcal{N}(\mu, \sigma^2 \mathbf{I})$ and only $\Delta_i \mathbf{O}_i$ observed.
We use EM such that if $\Delta \mathbf{O} \mu^t / \sigma^2 = \mathbf{U}_1 \Phi \mathbf{U}_2^t$

$$\int \mathbf{X} dF(\mathbf{X} | \Delta \mathbf{O}, \mu, \sigma^2) = \mathbf{U}_2 diag \left(\nabla_\Phi \log \int_{SO(m)} e^{tr(\mathbf{R}\Phi)} d\mathbf{R} \right) \mathbf{U}_1^t \Delta \mathbf{O}$$

$m = 2$ or $m = 3$

1. $$\mu_{r+1} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{U}_{2i} \frac{\int_{SO(m)} \mathbf{R} e^{\frac{tr(\mathbf{R}\phi_i)}{\sigma_r^2}} d\mathbf{R}}{\int_{SO(m)} e^{\frac{tr(\mathbf{R}\phi_i)}{\sigma_r^2}} d\mathbf{R}} \mathbf{U}_{1i}^T \Delta_i \mathbf{O}_i \quad \Delta_i \mathbf{O}_i \mu_r^t = \mathbf{U}_{1i} \phi_i \mathbf{U}_{2i}^T$$

2. $$\sigma_{r+1}^2 = \frac{1}{mk} \sum_{i=1}^{n} \frac{tr(\Delta_i^2)}{n} - tr(\mu_{r+1} \mu_{r+1}^t)$$

Cases of practical importance

m=2 $\int_{SO(2)} e^{tr(\mathbf{R}\Phi)} d\mathbf{R} = I_0(\phi_1 + \phi_2)$

$$\frac{\int\limits_{\mathbf{R}\in\mathcal{SO}(2)} \mathbf{R} e^{tr\mathbf{R}\Phi} d\mathbf{R}}{\int\limits_{\mathbf{R}\in\mathcal{SO}(2)} e^{tr\Phi} d\mathbf{R}} = diag \nabla_\Phi \log I_0(\phi_1 + \phi_2) = \frac{\mathcal{I}_1(\lambda_1 + \lambda_2)}{\mathcal{I}_0(\lambda_1 + \lambda_2)} \mathbf{I}_2$$

m=3 $\int\limits_{SO(3)} e^{tr(\mathbf{R}\Phi)} d\mathbf{R} = \int\limits_{vv^t=1; v\in\mathbb{R}^4} e^{-v^t diag(\xi_1,\xi_2,\xi_3,\xi_4) v} d_{\mathcal{S}^3}(v)$

$$\xi_4 = \phi_1 + \phi_2 + \phi_3 \quad \text{and} \quad \xi_i = 2\phi_i - \xi_4 \quad i = 1, 2, 3 \qquad (10)$$

$$diag\left(\nabla_\Phi \log \int_{SO(m)} e^{tr(\mathbf{R}\Phi)} d\mathbf{R}\right) = \mathbf{I}_3 - \begin{pmatrix} \frac{\mathcal{C}_6(A_2)+\mathcal{C}_6(A_3)}{\pi\mathcal{C}_4(A)} & 0 & 0 \\ 0 & \frac{\mathcal{C}_6(A_1)+\mathcal{C}_6(A_3)}{\pi\mathcal{C}_4(A)} & 0 \\ 0 & 0 & \frac{\mathcal{C}_6(A_1)+\mathcal{C}_6(A_2)}{\pi\mathcal{C}_4(A)} \end{pmatrix} \quad (11)$$

Simulation evidence

	$\mu_1 = diag(2,1,10)\mathbf{0}_{3\times 1}$						$\mu_2 = diag(2,1,10)\mathbf{0}_{3\times 2}$					
n	$\rho(\hat{\mu}_{proc}, \mu_1)$	$d(\hat{\mu}_{proc}, \mu_1)$	$\hat{\sigma}_{proc}$	$\rho(\hat{\mu}_{mle}, \mu_1)$	$d(\hat{\mu}_{mle}, \mu_1)$	$\hat{\sigma}_{mle}$	$\rho(\hat{\mu}_{proc}, \mu_2)$	$d(\hat{\mu}_{proc}, \mu_2)$	$\hat{\sigma}_{proc}$	$\rho(\hat{\mu}_{mle}, \mu_2)$	$d(\hat{\mu}_{mle}, \mu_2)$	$\hat{\sigma}_{mle}$
200	0.437	1.146	0.572	0.334	0.341	0.799	0.638	1.303	0.652	0.366	0.388	0.776
500	0.372	1.101	0.568	0.070	0.100	0.792	0.456	1.158	0.680	0.466	2.078	0.799
1000	0.410	1.087	0.571	0.073	0.091	0.800	0.661	2.883	0.674	0.380	2.057	0.801
2000	0.403	1.120	0.570	0.044	0.049	0.800	0.457	1.163	0.670	0.142	0.144	0.800
2500	0.413	1.128	0.573	0.034	0.034	0.805	0.467	1.169	0.670	0.103	0.105	0.799
200	0.490	1.524	0.695	0.536	2.105	0.967	0.640	3.291	0.799	0.399	0.604	0.949
500	0.741	3.248	0.701	0.065	0.074	0.983	0.909	3.241	0.820	0.389	0.442	0.978
1000	0.435	1.536	0.699	0.185	0.251	0.986	0.643	1.691	0.828	0.244	0.251	0.996
2000	0.441	1.517	0.700	0.074	0.083	0.995	0.527	1.604	0.835	0.197	0.215	1.000
2500	0.473	1.539	0.702	0.110	0.112	1.004	0.749	3.266	0.832	0.107	0.107	1.001
200	0.834	3.590	0.773	0.568	2.244	1.098	0.788	2.055	0.898	0.636	0.831	1.069
500	0.474	1.727	0.754	0.271	0.341	1.068	0.842	3.453	0.900	0.645	2.171	1.073
1000	0.454	1.724	0.760	0.120	0.196	1.078	0.602	1.873	0.907	0.342	0.421	1.087
2000	0.485	1.733	0.764	0.082	0.090	1.098	0.603	1.861	0.916	0.230	0.249	1.101
2500	0.477	1.715	0.765	0.058	0.058	1.092	0.560	1.826	0.910	0.237	0.266	1.094

Table : Comparison of Procrustes with mle estimators for sample size values $n = 200, 500, 1000, 2000, 2500$ and $\sigma = 0.8, 1, 1.1$ from $N(\frac{\mu_1}{\|\mu_1\|}, \sigma^2)$ and $N(\frac{\mu_2}{\|\mu_2\|}, \sigma^2)$.

Summary

- We derived the explicit Pffafian for the Fisher-Bingham family.
- We also show how the Pffafian can be easily adopted in the degenerate cases of multiplicities and zeros in the parameter values.
- In HG we can allow the starting value to be at $\gamma = \mathbf{0}$, i.e, we can use the Bingham normalizing constant.
- Show that for the Bingham distributions HG is accurate for many cases, including the MLE implementation for the size-and-shape models.

Sei and Kume (2013). arXiv:1304.7973

Kume and Sei (2014). To appear

Thank you!

References I

- Dryden, I.L. and Mardia, K.V. (1998). *Statistical Shape Analysis*. John Wiley, Chichester.
- Hashiguchi, H., Numata, Y., Takayama, N. and Takemura, A.: Holonomic gradient method for the distribution function of the largest root of a wishart matrix. *Journal Multivariate Analysis*, 117:296-312,2013.
- Goodall, C. and Mardia, K.V. (1992).
  The Non-central Bartlett Decomposition and Shape Densities. *J. Mult. Analysis* B **40**, 94-108.
- Green, P.J. & Mardia, K.V. (2006). Bayesian alignment using hierarchical models, with applications in protein bioinformatics. *Biometrika*, **93**, 235-254.
- Kent, J.T.(1982) The Fisher-Bingham distribution on the sphere *J.R. Statisct. Soc.* B 44:71-80.
- Kent, J.T., (1994) *The complex Bingham distribution and shape analysis*, *J.R. Statist. Soc* 56: 285-289.
- Koyama, T., Nakayama, H., Nishiyama, K. and Takayama, N.(2013): Holonomic gradient descent for the Fisher-Bingham distribution on the n-dimensional sphere. *Computational Statistics*.

References II

- Kume, A. and Wood, A. T. A.(2005): Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. *Biometrika*, 92:465–476.
- Kume, A and Dryden, I.L. (2014) : Shape inference based on multivariate normal matrix distributions, *To appear*
- Kume, A. and Sei, T. (2014). On the explicit form of the Pfaffian of the Fisher-Bingham integral. *To appear*
- Kume, A and Walker, S.G. On the FisherBingham distribution, *Statistics and Computing*, 19:167-172, 2009.
- Mardia, K. V. (1975) Statistics of Directional Data (with Discussion) *J.R. Statisct. Soc.* B 37:449-93
- Mardia, K. V. and Jupp, P. E.(2000): *Directional Statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester. ISBN 0-471-95333-4.
- Nakayama, H., Nishiyama, K., Noro, M., Ohara, K., Sei, T., Takayama, N. and Takemura, A.: Holonomic gradient descent and its application to the Fisher-Bingham integral. *Advances in Applied Mathematics*, 47:639–658, 2011.
- Scealy,J. L. and A. H. Welsh, A.H. (2011) Regression for compositional data by using distributions defined on the hypersphere, *JRSS, B*, 73: 371-375.

References III

- Sei, T. and Kume, A. (2013). Calculating the normalising constant of the Bingham distribution on the sphere using the holonomic gradient method. *Statistics and Computing*,
- Sei, T., Shibata, H., Takemura, A., Ohara, K. and Takayama, N.: Properties and applications of Fisher distribution on the rotation group. *Journal Multivariate Analysis*,116:440-455, 2013.