MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Properties and Applications of Fisher Distribution on the Rotation Group

Tomonari SEI, Hiroki SHIBATA, Akimichi TAKEMURA, Katsuyoshi OHARA and Nobuki TAKAYAMA

METR 2011–32 October 2011

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

Properties and applications of Fisher distribution on the rotation group

Tomonari Sei[*†], Hiroki Shibata[‡], Akimichi Takemura[†‡], Katsuyoshi Ohara[§], Nobuki Takayama[¶†]

October 2011

Abstract

We study properties of Fisher distribution (von Mises-Fisher distribution, matrix Langevin distribution) on the rotation group $SO(3)$. In particular we apply the holonomic gradient descent, introduced by Nakayama et al. [2011], and a method of series expansion for evaluating the normalizing constant of the distribution and for computing the maximum likelihood estimate. The rotation group can be identified with the Stiefel manifold of two orthonormal vectors. Therefore from the viewpoint of statistical modeling, it is of interest to compare Fisher distributions on these manifolds. We illustrate the difference with an example of near-earth objects data.

Keywords: algebraic statistics; directional statistics; holonomic gradient descent; maximum likelihood estimation; rotation group.

1 Introduction

In this paper we apply the holonomic gradient descent (HGD) introduced in Nakayama et al. [2011] and a method of series expansion for evaluating the normalizing constant of Fisher distribution on the rotation group and on Stiefel manifolds and for obtaining the maximum likelihood estimate. Fisher distribution is the most basic exponential family model for these manifolds.

The general theory of exponential family is well established (e.g. Barndorff-Nielsen [1978]). In nice "textbook" cases, the normalizing constant of the exponential family (i.e. its cumulant generating function) can be explicitly evaluated and then the calculation of

[*]Department of Mathematics, Keio University

[†]JST, CREST

[‡]Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo

[§]Faculty of Mathematics and Physics, Kanazawa University

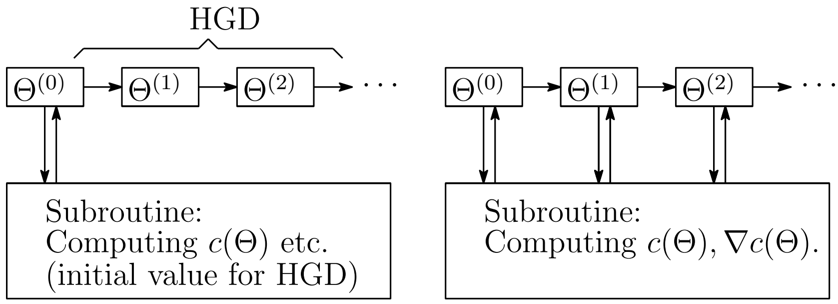
[¶]Department of Mathematics, Kobe University

maximum likelihood estimate is also simple. However in general, the integral defining the normalizing constant of an exponential family can not be explicitly evaluated. Various techniques, such as infinite series expansion, numerical integration, Markov chain Monte Carlo methods, iterative proportional scaling, have been applied for these cases.

Recently, we introduced a very novel approach, the holonomic gradient descent, for evaluation of the normalizing constant and solving the likelihood equation (Nakayama et al. [2011]). Our approach provides a systematic methodology for these tasks. Note that the normalizing constant is a definite integral over the sample space, where the integrand contains the parameter of the family of distributions. The likelihood equation involves differentiation of the normalizing constant with respect to the parameter. In the holonomic gradient descent, the theory of D-modules is used to derive a set of partial differential equations satisfied by the normalizing constant.

In this paper we apply the holonomic gradient descent and a method of series expansion to Fisher distribution on the rotation group $SO(3)$ and on the Stiefel manifold $V_2(\mathbb{R}^3)$ of two orthonormal vectors. The Fisher distribution on Stiefel manifolds and orthogonal groups has been studied by number of authors. However only a few papers (Prentice [1986], Wood [1993]) study the Fisher distribution on the special orthogonal group $SO(p)$.

The holonomic gradient descent needs the initial value for the differential equation. To evaluate this value, we develop an explicit formula of the infinite series expansion for $SO(3)$. An alternative method is a one-dimensional integration formula proposed by Wood [1993]. In Figure 1, we illustrate a diagram that clarifies the difference between HGD and direct use of gradient descent method. These variations will make the numerical evaluation of the maximum likelihood estimator more flexible.



(a) holonomic gradient descent. (b) usual gradient descent.

Figure 1: The difference between HGD and the usual gradient descent. In the diagram (a), a subroutine that computes the normalizing constant $c(\Theta)$ (and related values) is called only once. In the diagram (b), the subroutine is called every time the parameter Θ is updated.

The organization of the paper is as follows. For the rest of this section we set up notation and summarize preliminary facts on special orthogonal groups and Stiefel manifolds. In Section 2 we derive some properties of Fisher distribution on special orthogonal groups and Stiefel manifolds. In Section 3 we derive the set of partial differential equations satisfied by the normalizing constant (Section 3.1). We also give an infinite series expansion for the normalizing constant (Section 3.2). In Section 4 we apply the results of previous sections to the data on orbits of near-earth objects.

1.1 Notation and preliminary facts

Here we set up notation of this paper and summarize some preliminary facts. Although we are primarily interested in 3×3 matrices for practical and computational reasons, we set up our notation for general dimension. Let

$$V_r(\mathbb{R}^p) = \{A \in \mathbb{R}^{p\times r} \mid A^\top A = I_r\} \quad (0 < r \le p)$$

denote the Stiefel manifold of $p \times r$ real matrices with orthonormal columns, where $\mathbb{R}^{p\times r}$ denotes the set of $p \times r$ real matrices and $A^\top$ denote the transpose of A. In particular for $r = p$,

$$V_r(\mathbb{R}^p) = O(p)$$

is the set of $p \times p$ orthogonal matrices.

$$SO(p) = \{X \in O(p) \mid \det X = 1\}$$

denotes the special orthogonal group.

The total volume of $V_r(\mathbb{R}^p)$ is given as

$$\mathrm{Vol}(V_r(\mathbb{R}^p)) = \frac{2^r \pi^{rp/2}}{\Gamma_r(p/2)},$$

where

$$\Gamma_r(a) = \pi^{r(r-1)/4} \prod_{i=1}^{r} \Gamma[a - \frac{1}{2}(i-1)].$$

See Theorem 2.1.12 of Muirhead [1982].

Let Vol denote the invariant measure (volume element) on $V_r(\mathbb{R}^p)$ and let

$$\mu(\cdot) = \frac{1}{\mathrm{Vol}(V_r(\mathbb{R}^p))}\mathrm{Vol}(\cdot)$$

denote the invariant probability measure on $V_r(\mathbb{R}^p)$. Similarly for $SO(p)$, by $\mu(\cdot) = \mathrm{Vol}(\cdot)/\mathrm{Vol}(SO(p))$ we denote the invariant measure with

$$\mathrm{Vol}(SO(p)) = \frac{1}{2}\mathrm{Vol}(O(p)).$$

For a $p \times r$ matrix $\Theta \in \mathbb{R}^{p\times r}$, $r \le p$, let

$$\Theta = QDR, \quad Q \in V_r(\mathbb{R}^p),\ D\colon \text{diagonal},\ R \in \mathcal{O}(r)$$

denote its singular value decomposition (SVD). In this usual SVD, the diagonal elements of D are taken to be non-negative. Now let $\Theta \in \mathbb{R}^{p\times p}$ be a square matrix and restrict Q, R to be in $SO(p)$. Then the sign of $\det\Theta$ has to be equal to the sign of $\det D$. Let $\rho_1, \dots, \rho_p \geq 0$ denote the singular values of Θ. For non-singular Θ, the sign of $\det D$ can be adjusted by multiplying ρ_1 by $\epsilon = \pm 1$. Therefore we can write

$$\Theta = QDR, \quad Q, R \in SO(p),\ D = \mathrm{diag}(\epsilon\rho_1, \rho_2, \dots, \rho_p),\ \epsilon = \mathrm{sgn}\det\Theta. \tag{1}$$

We call this decomposition the *sign-preserving SVD* of Θ with respect to $SO(p)$. We also call $\phi_1 = \epsilon\rho_1$, $\phi_i = \rho_i$, $i \geq 2$, the *sign-preserving singular values* of Θ. The decomposition is also used in Prentice [1986] and Wood [1993].

2 Fisher distributions on $V_r(\mathbb{R}^p)$ and $SO(p)$

In this section we consider Fisher distribution on $V_r(\mathbb{R}^p)$ and $SO(p)$. In particular we clarify the difference between Fisher distributions on $V_{p-1}(\mathbb{R}^p)$ and $SO(p)$. Basic facts on Fisher distribution on $V_r(\mathbb{R}^p)$ is summarized in Chapter 13 of Mardia and Jupp [2000].

Let $\mathcal{X}$ denote either $V_r(\mathbb{R}^p)$ or $SO(p)$. The density of the Fisher distribution on $\mathcal{X}$ with respect to the uniform probability measure μ is given by

$$f(X;\Theta) = \frac{1}{c(\Theta)}\,\mathrm{etr}(\Theta^\top X), \quad X \in \mathcal{X},$$

where $\Theta = (\theta_{ij}) \in \mathbb{R}^{p\times r}$ is the parameter matrix, $\mathrm{etr}(\cdot) = \exp(\mathrm{tr}(\cdot))$, and

$$c(\Theta) = \int_{\mathcal{X}} \mathrm{etr}(\Theta^\top X)\mu(dX) \tag{2}$$

is the normalizing constant. For $V_r(\mathbb{R}^p)$ it is well known (e.g. Khatri and Mardia [1977], Muirhead [1982], Chikuse [2003]) that $c(\Theta)$ is a matrix-valued hypergeometric function $c(\Theta) = {}_0F_1(p/2, Y)$, where $Y = \Theta^\top\Theta/4$. However properties of $c(\Theta)$ for the special orthogonal group $\mathcal{X} = SO(p)$ are not studied in detail. For the case of $SO(3)$, following the approach in Prentice [1986], Wood [1993] used the correspondence between the Fisher distribution on $SO(3)$ and the Bingham distribution on the unit sphere S^3 in $\mathbb{R}^4$ and showed that $c(\Theta)$ can be written as a one-dimensional integral involving the modified Bessel function of degree zero. In Section 3 we derive differential equations and an infinite series expansion of $c(\Theta)$ for $SO(3)$.

Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p$ be the columns of $X \in SO(p)$. Since $\boldsymbol{x}_p$ is uniquely determined from $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{p-1}$, we can identify $SO(p)$ with $V_{p-1}(\mathbb{R}^p)$ by

$$(\boldsymbol{x}_1, \dots, \boldsymbol{x}_p) \in SO(p) \ \leftrightarrow\ (\boldsymbol{x}_1, \dots, \boldsymbol{x}_{p-1}) \in V_{p-1}(\mathbb{R}^p) \tag{3}$$

This leads to the question of differences of Fisher distributions on $SO(p)$ and those on $V_{p-1}(\mathbb{R}^p)$. Let $\Theta = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p) \in \mathbb{R}^{p\times p}$ be a parameter matrix for Fisher distribution on $SO(p)$. By setting $\boldsymbol{\theta}_p = 0$, we clearly obtain a Fisher distribution on $V_{p-1}(\mathbb{R}^p)$. Hence the family of Fisher distributions on $V_{p-1}(\mathbb{R}^p)$ is a submodel of the family of Fisher distributions on $SO(p)$. It can be easily seen that for $p = 2$, $\boldsymbol{\theta}_2$ is redundant and these two families are the same. However for $p \geq 3$, the family of Fisher distributions on $V_{p-1}(\mathbb{R}^p)$ is a strict submodel of that on $SO(p)$. We state this as a lemma.

Lemma 1. *For $p \geq 3$, the family of Fisher distributions on $V_{p-1}(\mathbb{R}^p)$ is a strict submodel of that on $SO(p)$.*

Proof. In general, let K be a positive integer and consider a K-dimensional exponential family

$$p(x|\theta) = \frac{1}{C(\theta)} \exp\left(\theta^\top x\right) \quad (x \in S), \quad C(\theta) = \int_S \exp(\theta^\top x)\nu(dx),$$

where θ is a K-dimensional vector, S is a smooth submanifold of $\mathbb{R}^K$ and ν is a measure on S. Assume that $C(\theta)$ exists in some open neighborhood of the origin $\theta = 0$. The parameter θ is estimable if and only if

$$\text{Affine}(\text{support}(\nu)) \;=\; \mathbb{R}^K$$

(Corollary 9.6 of Barndorff-Nielsen [1978]), where support(ν) is the support of ν and Affine(U), $U \subset \mathbb{R}^K$, denotes the affine hull of U.

We now show that if $p \geq 3$ then Affine$(SO(p)) = \mathbb{R}^{p\times p}$ and the distribution $p(X|\Theta) \propto \text{etr}(\Theta^\top X)$ is estimable, which is sufficient to prove the lemma. Let $L = \text{Affine}(SO(p))$. We first see that the zero matrix 0 belongs to L. Let $\epsilon_i \in \{-1, 1\}$ for $1 \leq i \leq p$. Then the average of 2^{p-1} matrices $\text{diag}(\epsilon_1, \ldots, \epsilon_{p-1}, \prod_{i=1}^{p-1} \epsilon_i)$ in $SO(p)$ is zero. Hence $0 \in L$. We now show that $e_i e_j^\top$ belongs to L $(\forall i, j)$, where $e_i = (0, \ldots, 1, \ldots, 0)^\top$ is the standard basis vector with 1 as the i-th element. Then together with $0 \in L$ it follows that $L = \mathbb{R}^{p\times p}$. Take matrices $P_i \in SO(p)$ $(i = 1, \ldots, p)$ such that $P_i e_i = e_1$. For example, let $P_1 = I_p$ and $P_i = e_1 e_i^\top - e_i e_1^\top + \sum_{j\neq 1,i} e_j e_j^\top$ for $i \neq 1$. Then $e_i e_j^\top \in L$ if and only if $e_1 e_1^\top = P_i e_i e_j^\top P_j^\top \in L$. Now it suffices to show that $e_1 e_1^\top$ belongs to L. Take the average of 2^{p-2} matrices $\text{diag}(1, \epsilon_2, \ldots, \epsilon_{p-1}, \prod_{i=2}^{p-1} \epsilon_i)$. Then $e_1 e_1^\top \in L$. □

For $\mathcal{X} = V_r(\mathbb{R}^p)$ the maximum likelihood estimate (MLE) of the Fisher distribution is obtained by the following procedure (Khatri and Mardia [1977]). Let $X^{(1)}, \ldots, X^{(N)}$ be a data set on $V_r(\mathbb{R}^p)$. Let $\bar{X} = N^{-1}\sum_{t=1}^N X^{(t)}$ be the sample mean matrix and let $\bar{X} = Q\,\text{diag}(g_1, \ldots, g_r)R$ be the SVD of $\bar{X}$, where $Q \in V_r(\mathbb{R}^p)$, $R \in O(r)$ and $g_1 \geq \cdots \geq g_r \geq 0$. Then the maximum likelihood estimate $\hat{\Theta}$ is given by $\hat{\Theta} = Q\,\text{diag}(\hat{\phi}_1, \ldots, \hat{\phi}_r)R$, where $\hat{\phi}_i$ is the solution of

$$\frac{\int_{V_r(\mathbb{R}^p)} x_{ii} \exp(\sum_{k=1}^r \hat{\phi}_k x_{kk})\mu(dX)}{\int_{V_r(\mathbb{R}^p)} \exp(\sum_{k=1}^r \hat{\phi}_k x_{kk})\mu(dX)} = g_i, \quad i = 1, \ldots, r.$$

This procedure is also valid for $SO(p)$ if we use the sign-preserving SVD in (1). We give the fact as a lemma since it is not explicitly proved in the literature. Remark that for $SO(p)$ the normalizing constant $c(\Theta)$ in (2) is invariant under a transformation $\Theta \mapsto Q\Theta R$ for any $Q, R \in SO(p)$.

Lemma 2. *Let $X^{(1)}, \dots, X^{(N)}$ be a data set on $SO(p)$. Let $\bar{X} = N^{-1}\sum_{t=1}^{N} X^{(t)}$ be the sample mean matrix and $\bar{X} = Q\,\mathrm{diag}(g_1, \dots, g_p)R$ be the sign-preserving SVD of $\bar{X}$, where $Q, R \in SO(p)$ and $|g_1| \geq g_2 \cdots \geq g_p \geq 0$. Then the maximum likelihood estimate of the Fisher distribution on $SO(p)$ is $\hat{\Theta} = Q\,\mathrm{diag}(\hat{\phi}_1, \dots, \hat{\phi}_r)R$, where $\hat{\phi}_i$ is the maximizer of the function*

$$(\phi_k)_{k=1}^{p} \mapsto \sum_{k=1}^{p} \phi_k g_k - \log c(\mathrm{diag}(\phi_1, \dots, \phi_r)), \tag{4}$$

or equivalently, the solution of

$$\frac{\int_{SO(p)} x_{ii} \exp(\sum_{k=1}^{p} \hat{\phi}_k x_{kk})\mu(dX)}{\int_{SO(p)} \exp(\sum_{k=1}^{p} \hat{\phi}_k x_{kk})\mu(dX)} = g_i, \quad i = 1, \dots, p. \tag{5}$$

Proof. We change the parameter variable from Θ to $\Phi = (\phi_{ij})_{i,j=1}^{p} = Q^\top \Theta R^\top$. Then the ($1/N$ times) log likelihood function is written as

$$\mathrm{tr}(\Theta^\top \bar{X}) - \log c(\Theta) = \mathrm{tr}(\Phi^\top G) - \log c(\Phi), \tag{6}$$

where $G = \mathrm{diag}(g_1, \dots, g_p)$. Since (6) is strictly convex in Φ, the unique maximizer makes its first-order derivatives zero. Note that the first term on the right hand side of (6) does not depend on the off-diagonal elements of Φ. Therefore the condition for maximization of (6) with respect to an off-diagonal element is written as

$$0 = \frac{\partial}{\partial \phi_{ij}} \log c(\Phi), \quad (i \neq j). \tag{7}$$

We now fix $i \neq j$ and evaluate $(\partial/\partial\phi_{ij}) \log c(\Phi)$ at $(\phi_{i'j'})_{i' \neq j'} = 0$. Then we have

$$\left.\frac{\partial}{\partial \phi_{ij}} \log c(\Phi)\right|_{\phi_{i'j'}=0\ \forall i' \neq j'} = \frac{\int_{SO(p)} x_{ij} \exp(\sum_{k=1}^{p} \phi_{kk} x_{kk})\mu(dX)}{\int_{SO(p)} \exp(\sum_{k=1}^{p} \phi_{kk} x_{kk})\mu(dX)}.$$

However

$$\int_{SO(p)} x_{ij} \exp(\sum_{k=1}^{p} \phi_{kk} x_{kk})\mu(dX) = \int_{SO(p)} (-x_{ij}) \exp(\sum_{k=1}^{p} \phi_{kk} x_{kk})\mu(dX) = 0$$

because the uniform distribution μ on $SO(p)$ is invariant with respect to multiplication of -1 to the i-th row and the i-th column of X. Therefore any diagonal matrix Φ satisfies (7). The log-likelihood function of the diagonal matrix is (4) and the maximizer satisfies (5). □

When $\det \bar{X} < 0$, it is not correct to use the ordinary singular values of $\bar{X}$ on the right-hand side of (5).

Remark 1. *The determinant of the sample mean matrix* $\bar{X}$ *is not necessarily positive even if all* $X^{(t)}$, $t = 1, \ldots, N$, *are in* $SO(p)$. *Indeed for the case of uniform distribution on* $SO(p)$ *we prove*

$$P(\det \bar{X} < 0) \to \frac{1}{2}, \qquad (N \to \infty),$$

as long as $p \geq 3$. *By the central limit theorem* $\sqrt{N}(\bar{X} - E(X))$ *converges to a Gaussian random matrix* Z *with the same covariances as* X. *We will show* $E(X) = 0$ *and the covariances of* X *are diagonal when* $p \geq 3$. *Then* Z *and any sign change of a column of* Z *have the same probability distribution and therefore the probability of* $\det(Z) < 0$ *is* $1/2$. *Hence the probability of* $\det(\bar{X}) < 0$ *converges to* $1/2$.

To prove that the mean is zero and the covariance is diagonal, it is sufficient to consider $E(X_{11})$ *and* $E(X_{a1}X_{b2})$ $(1 \leq a, b \leq p)$ *by symmetry. Define a random matrix* Y *by* $Y_{ij} = X_{ij}$ *for* $j \neq 1, 3$ *and* $Y_{ij} = -X_{ij}$ *for* $j = 1, 3$. *Since both* X *and* Y *have the uniform distribution on* $SO(p)$, *we deduce that* $E(X_{11}) = E(-X_{11}) = 0$, $E(X_{a1}X_{b2}) = E(-X_{a1}X_{b2}) = 0$.

Remark 2. *Even if* $\det \bar{X} > 0$, *the determinant of the estimated parameter* $\hat{\Theta}$ *may be negative. Indeed, let the sign-preserving singular values of* $\bar{X}$ *and* $\hat{\Theta}$ *be* $g = (g_1, g_2, g_3)$ *and* $\hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3)$, *respectively. We prove that* $g_1 g_2 g_3$ *and* $\hat{\phi}_1\hat{\phi}_2\hat{\phi}_3$ *can have the opposite signs. To see this, we first consider the case* $\hat{\phi}_1 = 0$, $\hat{\phi}_2 > 0$ *and* $\hat{\phi}_3 > 0$. *Then, by using the Taylor expansion formula developed in Subsection 3.2, we deduce that* g_1, g_2 *and* g_3 *are strictly positive. By continuity, there exist some* $\hat{\phi}_1 < 0$, $\hat{\phi}_2 > 0$ *and* $\hat{\phi}_3 > 0$ *while all* g_i*'s are positive.*

3 Computation of the normalizing constant and its derivatives

For computing the maximum likelihood estimate of Fisher distribution we need numerical evaluation of the normalizing constant $c(\Theta)$ of (2) and its derivatives. In this section we study two methods for this purpose. The first method is the holonomic gradient descent. In the second method, we use series expansion of $\mathrm{etr}(\Theta^\top X)$. The second method is also used to compute the initial value of HGD (see Figure 1 (a)).

3.1 The holonomic gradient descent for Stiefel manifolds and special orthogonal group

Let us briefly describe the holonomic gradient descent. As to details, we refer to Nakayama et al. [2011]. An algebraic computation is the first step; we construct linear ODE's (ordinary differential equations) satisfied by $c(\Theta)$ with respect to each θ_{ij} by Gröbner bases of a set of partial differential equations satisfied by c. Variables other than θ_{ij} appear as parameters in the ODE. The rank of ODE's is called the holonomic rank. The ODE's give a dynamical system for the function $c(\Theta)\,\mathrm{etr}(-\Theta^\top \bar{X})$, the reciprocal of the likelihood.

The gradient of the function can also be expressed in terms of derivatives of the reciprocal standing for the standard monomials and $\bar{X}$. The second step is a numerical procedure; a point in the dynamical system moves toward the maximum likelihood estimate along the gradient direction, simultaneously updating the values of $c(\Theta)$ and its derivatives.

For the holonomic gradient descent, we study differential operators A annihilating $c(\Theta)$, that is, $A \cdot c(\Theta) = 0$. Denote the differential operator $\partial/\partial\theta_{ij}$ by ∂_{ij}. We first study the special orthogonal group and then study the Stiefel manifold.

3.1.1 The case of special orthogonal group

Let $\Theta \in \mathbb{R}^{p\times p}$. We consider the following three types of differential operators:

$$\begin{aligned}
A_{ij}^{(1)} &= \sum_{k=1}^{p} \partial_{ik}\partial_{jk} - \delta_{ij}, \quad \tilde{A}_{ij}^{(1)} = \sum_{k=1}^{p} \partial_{ki}\partial_{kj} - \delta_{ij} \quad (i \le j), \\
A^{(2)} &= \det(\partial_{ij}) - 1, \\
A_{ij}^{(3)} &= \sum_{k=1}^{p} \left(-\theta_{jk}\partial_{ik} + \theta_{ik}\partial_{jk}\right), \quad \tilde{A}_{ij}^{(3)} = \sum_{k=1}^{p} \left(-\theta_{kj}\partial_{ki} + \theta_{ki}\partial_{kj}\right) \quad (i < j),
\end{aligned}$$

where δ_{ij} is the Kronecker's delta. The following lemma is an analogy of Theorem 2 of Nakayama et al. [2011].

Lemma 3. *The above differential operators annihilate $c(\Theta)$ of $SO(p)$.*

Proof. We first prove that the operators $A_{ij}^{(1)}$, $\tilde{A}_{ij}^{(1)}$ and $A^{(2)}$ annihilate $\mathrm{etr}(\Theta^\top X)$ for any $X \in SO(p)$. Then they also annihilates $c(\Theta)$ because $A \cdot c(\Theta) = \int_{SO(p)} A \cdot \mathrm{etr}(\Theta^\top X)\mu(dX)$ for any operator A. Since $\partial_{ij} \cdot \mathrm{etr}(\Theta^\top X) = x_{ij}\,\mathrm{etr}(\Theta^\top X)$ and $XX^\top = I$, we have

$$A_{ij}^{(1)} \cdot \mathrm{etr}(\Theta^\top X) = \left(\sum_{k=1}^{p} x_{ik}x_{jk} - \delta_{ij}\right) \mathrm{etr}(\Theta^\top X) = 0.$$

Similarly, we obtain $\tilde{A}_{ij}^{(1)} \cdot \mathrm{etr}(\Theta^\top X) = 0$ from $X^\top X = I$ and $A^{(2)} \cdot \mathrm{etr}(\Theta^\top X) = 0$ from $\det(X) = 1$. Next consider $A_{ij}^{(3)}$ and $\tilde{A}_{ij}^{(3)}$. We note $c(\Theta) = c(Q\Theta) = c(\Theta Q)$ for any $Q \in SO(p)$. For any fixed $i < j$, define a rotation matrix $Q = Q(\epsilon)$ by

$$Q = (\cos\epsilon)(E_{ii} + E_{jj}) + (\sin\epsilon)(-E_{ij} + E_{ji}) + \sum_{k \ne i,j} E_{kk},$$

where E_{kl} is the matrix whose (i,j)-th component is 1 if $k = i$ and $l = j$ and 0 otherwise. Then

$$\begin{aligned}
0 &= c(Q\Theta) - c(\Theta) \\
&= c\left(\Theta - \epsilon\sum_{k}\theta_{jk}E_{ik} + \epsilon\sum_{k}\theta_{ik}E_{jk} + \mathrm{o}(\epsilon)\right) - c(\Theta) \\
&= \epsilon\sum_{k=1}^{p}(-\theta_{jk}\partial_{ik} + \theta_{ik}\partial_{jk}) \cdot c(\Theta) + \mathrm{o}(\epsilon),
\end{aligned}$$

as $\epsilon \to 0$. Hence we have $A_{ij}^{(3)} \cdot c(\Theta) = 0$. Similarly we obtain $\tilde{A}_{ij}^{(3)} \cdot c(\Theta) = 0$ from $c(\Theta Q) = c(\Theta)$. $\square$

Let D be the ring of differential operators with polynomial coefficients in θ_{ij} and let I denote the ideal generated by the above differential operators $A_{ij}^{(1)}, \ldots, \tilde{A}_{ij}^{(3)}$ in D. Also let I_{diag} denote I restricted to diagonal matrices $\Theta = \mathrm{diag}(\theta_{11}, \ldots, \theta_{pp})$. $I \cdot f(\Theta) = 0$ implies $I_{\mathrm{diag}} \cdot f(\mathrm{diag}(\Theta)) = 0$. We denote by R_p the ring of differential operators with rational function coefficients in θ_{ij}, $1 \leq i, j \leq p$.

The following proposition is necessary for the holonomic gradient descent. We refer to Nakayama et al. [2011] for the definition of holonomic ideals in D and zero-dimensional ideals in R_p. Once zero-dimensionality of $R_p I$ is proved and a Gröbner basis is constructed, we can find ODE's and apply the holonomic gradient descent for the maximum likelihood estimate.

Proposition 1. *If $p = 2$, then the ideal I is holonomic. In particular, the ideal $R_2 I$ is zero-dimensional. The holonomic rank is equal to* 2.

The proposition is proved by Macaulay2 (Grayson and Stillman) and the yang package on Risa/Asir (RisaAsir developing team) by utilizing Gröbner basis computations in rings of differential operators. Also the set of generators of I is obtained by nk_restriction function of asir from the integral representation of $c(\Theta)$ as

$$\begin{aligned}
&g_1 = -\partial_{12} - \partial_{21}, \quad g_2 = -\partial_{11} + \partial_{22}, \quad g_3 = \partial_{21}^2 + \partial_{22}^2 - 1, \\
&g_4 = (\theta_{22} + \theta_{11})\partial_{21} + (-\theta_{21} + \theta_{12})\partial_{22}, \\
&g_5 = (\theta_{21} - \theta_{12})\partial_{22}\partial_{21} + (\theta_{22} + \theta_{11})\partial_{22}^2 + \partial_{22} - \theta_{22} - \theta_{11}, \\
&g_6 = (-\theta_{21} + \theta_{12})\partial_{21} + (\theta_{21}^2 - 2\theta_{12}\theta_{21} + \theta_{22}^2 + 2\theta_{11}\theta_{22} + \theta_{11}^2 + \theta_{12}^2)\partial_{22}^2 \\
&\quad + (\theta_{22} + \theta_{11})\partial_{22} - \theta_{22}^2 - 2\theta_{11}\theta_{22} - \theta_{11}^2.
\end{aligned}$$

Furthermore the set of generators of I_{diag} is given as

$$h_1 = (-\theta_{22} - \theta_{11})\partial_{11}^2 - \partial_{11} + \theta_{22} + \theta_{11}, \quad h_2 = -\partial_{11} + \partial_{22}.$$

Proposition 2. *If $p = 3$, then the ideal $R_3 I$ is zero-dimensional. The holonomic rank is less than or equal to* 4. *$R_3/(R_3 I)$ is spanned by $1, \partial_{31}, \partial_{32}, \partial_{33}$ as a vector space over the field of rational functions.*

The proposition is proved by a large scale computation on Risa/Asir with Gröbner bases. The algorithm for it is explained in, e.g., Nakayama et al. [2011]. Programs and obtained data are at the website OpenXM/Math (OpenXM Mathematics Repository). We conjecture that I is holonomic and consequently $R_p I$ is zero-dimensional for any p in the case of $SO(p)$.

3.1.2 The case of Stiefel manifold

Let $\Theta \in \mathbb{R}^{p\times r}$ $(r \le p)$. Consider the following differential operators:

$$
\begin{aligned}
A_{ij}^{(1)} &= \sum_{k=1}^{p} \partial_{ki}\partial_{kj} - \delta_{ij} \quad (1 \le i \le j \le r),\\
A_{ij}^{(2)} &= \sum_{k=1}^{r} (-\theta_{jk}\partial_{ik} + \theta_{ik}\partial_{jk}) \quad (1 \le i < j \le p),\\
\tilde{A}_{ij}^{(2)} &= \sum_{k=1}^{p} (-\theta_{kj}\partial_{ki} + \theta_{ki}\partial_{kj}) \quad (1 \le i < j \le r).
\end{aligned}
$$

Lemma 4. *The above operators annihilate $c(\Theta)$ of $V_r(\mathbb{R}^p)$.*

Proof. The proof is similar to that of Lemma 3. The operator $A_{ij}^{(1)}$ annihilates $\mathrm{etr}(\Theta^\top X)$ if $X \in V_r(\mathbb{R}^p)$. Since $c(\Theta) = c(Q\Theta) = c(\Theta R)$ for any $Q \in O(p)$ and $R \in O(r)$, we have $A_{ij}^{(2)} \cdot c(\Theta) = 0$ and $\tilde{A}^{(2)} \cdot c(\Theta) = 0$, respectively. □

Let I denote the ideal generated by the above operators and let I_{diag} denote its restriction to diagonal matrices $\Theta = \mathrm{diag}(\theta_{11}, \ldots, \theta_{rr}) \in \mathbb{R}^{p\times r}$. We denote by $R_{r,p}$ the ring of differential operators with rational function coefficients in θ_{ij}, $1 \le i \le p$, $1 \le j \le r$.

Proposition 3. *If $r = 2$, $p = 3$, then the ideal $R_{2,3}I$ is zero-dimensional. The holonomic rank is equal to 4. $R_{2,3}/(R_{2,3}I)$ is spanned by $1, \partial_{11}, \partial_{12}, \partial_{11}^2$ over the field of rational functions.*

This proposition is also proved by a computation on Risa/Asir. Programs to verify the proposition are at the website OpenXM/Math (OpenXM Mathematics Repository). We conjecture that I is holonomic and consequently $R_{r,p}I$ is zero-dimensional for any r and p in the case of $V_r(\mathbb{R}^p)$.

We close this subsection with some notes on our result and a study of hypergeometric functions. For the matrix-valued hypergeometric function $c(\Theta) = {}_0F_1(p/2, Y)$, $Y = \Theta^\top\Theta/4$, the following partial differential equation is well known (Muirhead [1970], [Muirhead, 1982, Thm.7.5.5]). Let $y_1, \ldots, y_r$ denote the eigenvalues of Y. F satisfies the following partial differential equations:

$$
y_i\partial_i^2 F + \left\{ \frac{p}{2} - \frac{r-1}{2} + \frac{1}{2}\sum_{j=1, j\ne i}^{r} \frac{y_i}{y_i - y_j} \right\} \partial_i F - \frac{1}{2}\sum_{j=1, j\ne i}^{r} \frac{y_j}{y_i - y_j}\partial_j F \;=\; F, \quad i = 1, \ldots, p. \tag{8}
$$

Muirhead [1970] obtained these partial differential equations from the partial differential equations satisfied by zonal polynomials (James [1968], [Takemura, 1984, Sec.4.5]). In Appendix A we check that for low dimensional cases these equations are also derived from the differential operators in Lemma 4.

3.1.3 Practice of HGD

Although the HGD is a general method which can be applied to broad problems, we need a good guess (oracle) of a starting point to search the optimal point (MLE). We explain why we need a good guess of a starting point with an example of $V_2(\mathbb{R}^3)$. Let Θ be the optimal point for a given data and $\frac{\partial Q}{\partial \theta_{ij}} = P_{ij}(\theta)Q$ be the Pfaffian system to apply for the HGD. The denominator of the coefficient matrix P_{ij} is a polynomial in θ. The Figure 2 shows the zero set in the θ_{11}, θ_{12} space of the product of the polynomials standing for P_{11} and P_{12} when $\theta_{ij}, i = 2, 3$ is restricted to the constant $(\hat{\Theta}_{1:2})_{ij}$, which is the MLE for the comets data (Section 4.2).

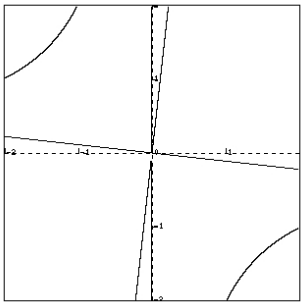


Figure 2: Singular locus in θ_{11}, θ_{12} space for $V_2(\mathbb{R}^3)$

Similarly, in the case of $SO(3)$, the Figure 3 shows the zero set in the θ_{11}, θ_{12} space of the product of the polynomials standing for the Pfaffian system when $\{\theta_{ij} \mid (i,j) \neq (1,1), (1,2)\}$ is restricted to the MLE for the comets data (Section 4.2).

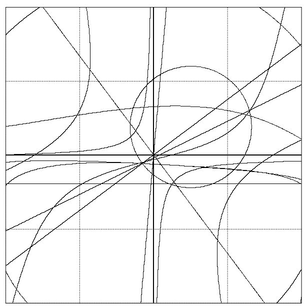


Figure 3: Singular locus in θ_{11}, θ_{12} space for $SO(3)$

The numerical integration procedure of the Pfaffian system becomes unstable near the zero set of the product of the polynomials, which is called the singular locus of the Pfaffian system. Therefore, the starting point must be in the same component with the optimal point in the semi-algebraic set defined by the zero set. In our current implementation of HGD, we find the starting point by preparing a table of the values of the normalization constant (integral) at grids and making the exhaustive search of the optimal point on the grids or by using an approximate MLE obtained by an other method. Note that the table can be used not only for specific data but also for general data.

3.2 Series expansion approach for $SO(3)$ and $V_2(\mathbb{R}^3)$

We describe a method to compute the maximum likelihood estimate by an infinite series expansion of $c(\Theta)$. By Lemma 2, computation of the maximum likelihood estimate for $SO(p)$ is reduced to computation of $c(\mathrm{diag}(\phi_1,\dots,\phi_p))$ and its derivatives with respect to ϕ_i's, together with the usual gradient method. In this subsection we give an explicit series expansion of $c(\mathrm{diag}(\phi_1,\phi_2,\phi_3))$ when $p=3$. Note that $c(\Theta)$ for any $\Theta \in \mathbb{R}^{3\times 3}$ is also obtained via sign-preserving SVD due to the rotational invariance of $c(\Theta)$. By using the expansion formula we also clarify the difference between the normalizing constants for the orthogonal group $O(3)$ and the special orthogonal group $SO(3)$. The series expansion approach for $V_2(\mathbb{R}^3)$ is also discussed.

From mathematical viewpoint, the holonomic descent and the infinite series expansion is related as follows. In the general recipe of the holonomic gradient descent and holonomic systems, we can construct series expansion of the normalization constant $c(\Theta)$ for any p up to any degree modulo finite constants in an algorithmic method from a holonomic system of differential equations satisfied by $c(\Theta)$, which is obtained in the previous subsection. The existence of finite recurrence relations for coefficients of the series is proved by the holonomicity. This is a multi-variable generalization of the fact that coefficients of series solutions of linear ODE satisfy a finite recurrence relation. Since this computation requires huge computational resources, constructing the series expansion in a more efficient way is preferable to using the general algorithm. Here we derive an infinite series expansion for $SO(3)$ with an analysis of integrals.

Let $E[\cdot]$ denote the expectation with respect to the uniform distribution on $SO(3)$. Let ϕ_1,ϕ_2,ϕ_3 be the sign-preserving singular values of Θ. By the rotational invariance, the expansion of $c(\Theta)$ is

$$\begin{aligned} c(\Theta) &= \sum_{h=0}^{\infty} \frac{1}{h!} E[(\mathrm{tr}\,\Theta^\top X)^h] = \sum_{h=0}^{\infty} \frac{1}{h!} E[(\phi_1 x_{11} + \phi_2 x_{22} + \phi_3 x_{33})^h] \\ &= \sum_{k,l,m=0}^{\infty} \frac{1}{k!\,l!\,m!} \phi_1^k \phi_2^l \phi_3^m E[x_{11}^k x_{22}^l x_{33}^m] \end{aligned} \tag{9}$$

and the problem is reduced to the evaluation of

$$E(k,l,m) = E[x_{11}^k x_{22}^l x_{33}^m].$$

Again by the rotational invariance we can simultaneously change the sign of any two of x_{11}, x_{22}, x_{33}. From this it is easily seen that $E(k,l,m) = 0$ unless k,l,m are all even or k,l,m are all odd.

Note that for $O(3)$ we can individually change the signs of x_{11}, x_{22}, x_{33}. Hence for $O(3)$ $E(k,l,m) = 0$ unless k,l,m are all even and $c(\Theta)$ is indeed a function of the eigenvalues of $Y = \Theta^\top\Theta/4$. Therefore the difference between $c(\Theta)$ for $SO(3)$ and $c(\Theta)$ for $O(3)$ comes from terms $E[k,l,m] = 0$ with k,l,m all odd.

We now express $X = (x_{ij}) \in SO(3)$ by the Euler angles θ, ϕ, ψ.

$$X = \begin{pmatrix} \sin\theta\sin\phi & \cos\phi\sin\psi + \cos\theta\sin\phi\cos\psi & -\cos\phi\cos\psi + \cos\theta\sin\phi\sin\psi \\ \sin\theta\cos\phi & -\sin\phi\sin\psi + \cos\theta\cos\phi\cos\psi & \sin\phi\cos\psi + \cos\theta\cos\phi\sin\psi \\ \cos\theta & -\sin\theta\cos\psi & -\sin\theta\sin\psi \end{pmatrix}.$$

The Jacobian of the above transformation is $\sin\theta$ and the range of variables is

$$0 \le \theta \le \pi, \quad 0 \le \phi \le 2\pi, \quad 0 \le \psi \le 2\pi.$$

Hence the integral of f over $SO(3)$ with respect to the uniform probability measure is expressed as

$$\int_{SO(3)} f(X) d\mu(X) = \frac{1}{8\pi^2} \int_0^\pi d\theta \int_0^{2\pi} d\phi \int_0^{2\pi} d\psi \; f(X(\theta,\phi,\psi)) \sin\theta.$$

For

$$f = x_{11}^k x_{22}^l x_{33}^m = (\sin\theta\sin\phi)^k(-\sin\phi\sin\psi + \cos\theta\cos\phi\cos\psi)^l(-\sin\theta\sin\psi)^m$$

we have

$$\begin{aligned} f \cdot \sin\theta &= (-1)^m \sin^{k+m+1}\theta \sin^k\phi \sin^m\psi \\ &\qquad \cdot \sum_{n=0}^{l} \binom{l}{n} (-1)^n \sin^n\phi \sin^n\psi \cos^{l-n}\theta \cos^{l-n}\phi \cos^{l-n}\psi \\ &= \sum_{n=0}^{l} \binom{l}{n} (-1)^{m+n} \sin^{k+m+1}\theta \cos^{l-n}\theta \sin^{k+n}\phi \cos^{l-n}\phi \sin^{m+n}\psi \cos^{l-n}\psi. \end{aligned}$$

Define

$$I[m,n] = \frac{(m-1)!!(n-1)!!}{(m+n)!!},$$

where $(2a)!! = \prod_{j=1}^a (2j)$ and $(2a-1)!! = \prod_{j=1}^a (2j-1)$ for each non-negative integer a. Then from well-known results on the definite integrals of trigonometric functions we have

$$E(k,l,m) = \sum_{\substack{0 \le n \le l \\ l-n:\text{ even}}} \binom{l}{n} I[k+m+1, l-n] \cdot I[k+n, l-n] \cdot I[m+n, l-n]. \tag{10}$$

By numerical experiments we found that (10) can be computed easily and we can evaluate $c(\Theta)$ by the right-hand side of (9) to a desired accuracy. For large k,l,m the value of $E(k,l,m)$ can be approximated by Laplace's method. Laplace approximation to $E(k,l,m)$ is given in Appendix B.

We now consider the maximization of (4) with respect to $\{\phi_i\}_{i=1}^3$ when we adopt direct use of the gradient descent as in Figure 1 (b). The gradient method uses the first derivatives of (4). The Hessian matrix is also needed if one uses the Newton method. Since the first term of (4) is linear, it is sufficient to give the series expansion of the derivatives of $c(\mathrm{diag}(\phi_1, \phi_2, \phi_3))$. They are easily obtained from the expansion of $c(\Theta)$. For example the derivative with respect to ϕ_1 is

$$\frac{\partial c(\mathrm{diag}(\phi_1, \phi_2, \phi_3))}{\partial \phi_1} = \sum_{k,l,m=0}^{\infty} \frac{1}{k!\, l!\, m!} \phi_1^k \phi_2^l \phi_3^m E(k+1, l, m).$$

Similarly,

$$\frac{\partial^2 c(\mathrm{diag}(\phi_1, \phi_2, \phi_3))}{\partial \phi_1^2} = \sum_{k,l,m=0}^{\infty} \frac{1}{k!\, l!\, m!} \phi_1^k \phi_2^l \phi_3^m E(k+2, l, m),$$
$$\frac{\partial^2 c(\mathrm{diag}(\phi_1, \phi_2, \phi_3))}{\partial \phi_1 \partial \phi_2} = \sum_{k,l,m=0}^{\infty} \frac{1}{k!\, l!\, m!} \phi_1^k \phi_2^l \phi_3^m E(k+1, l+1, m).$$

Finally we note that the series expansion of $c(\Theta)$ for $SO(3)$ is directly used for the maximum likelihood estimate of $V_2(\mathbb{R}^3)$. Let $\bar{X}_{1:2}$ be the first two columns of the averaged data matrix $\bar{X} \in \mathbb{R}^{3\times 3}$. Let $\bar{X}_{1:2} = Q\,\mathrm{diag}(g_1, g_2)R$ be the (usual) SVD. Then, as stated before Lemma 2, the maximum likelihood estimator for $V_2(\mathbb{R}^3)$ is given by $\hat{\Theta} = Q\,\mathrm{diag}(\hat{\phi}_1, \hat{\phi}_2)R$, where $(\hat{\phi}_i)$ is the maximizer of

$$\sum_{k=1}^{2} \phi_k g_k - \log\left(\int_{V_2(\mathbb{R}^3)} \exp(\sum_{k=1}^{2} \phi_k x_{kk}) \mu(dX)\right) = \sum_{k=1}^{2} \phi_k g_k - \log c(\mathrm{diag}(\phi_1, \phi_2, 0))$$

in terms of $c(\Theta)$ for $SO(3)$. Then the MLE is obtained via the series expansion of $c(\Theta)$.

4 Application to data on orbits of near-earth objects

In this section as an illustration of the above discussion, we fit Fisher distributions of $SO(3)$ and $V_2(\mathbb{R}^3)$ to data of orbits of near-earth objects. We obtained the data from the web page of Near Earth Object Program of National Aeronautics and Space Administration (cf. `http://neo.jpl.nasa.gov/cgi-bin/neo_elem`). Near-earth objects are comets and asteroids around the Earth. Jupp and Mardia [1979] fitted Fisher distribution on $V_2(\mathbb{R}^3)$ to data of comets from Marsden [1972], but did not consider Fisher distribution on $SO(3)$. See also Mardia [1975] for analysis of data of perihelion direction.

The near-earth objects have ellipsoidal orbits with the Sun as their focus. The orbits are characterized by the following two directions:

1. the perihelion direction $\boldsymbol{x}_1$, which is the direction of the closest point on the orbit from the Sun.

2. the normal direction $\boldsymbol{x}_2$ to the orbit, which is determined by the right-hand rule for the rotation of the object.

The pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is an element of $V_2(\mathbb{R}^3)$. We can also define $\boldsymbol{x}_3 = \boldsymbol{x}_1 \times \boldsymbol{x}_2$ such that $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ is an element of $\mathcal{SO}(3)$.

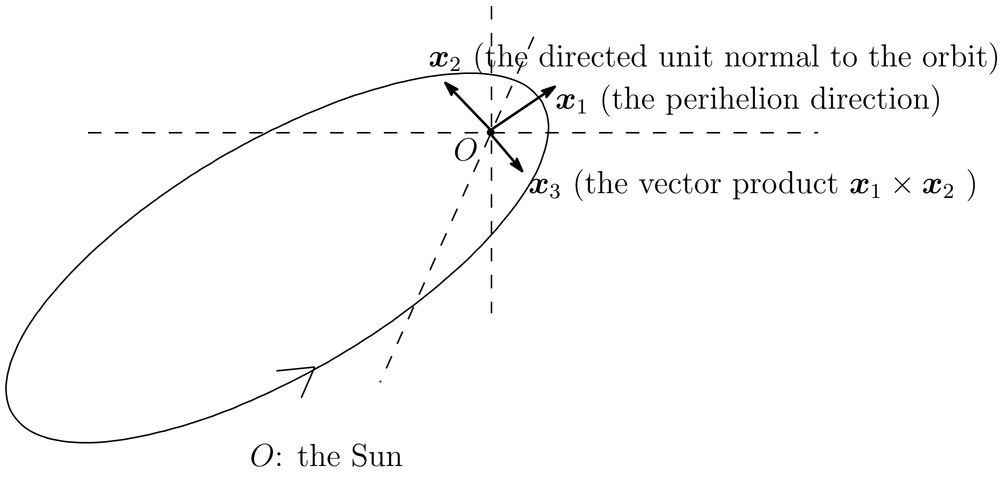


Figure 4: Orbits of near-earth objects

We analyzed 151 comets and 6496 asteroids separately. To obtain a meaningful result, we identified a tight cluster of 67 similar comets, which we treated as one comet, and therefore the actual sample size of comets is $N = 85$. Parts of the data are shown in Table 1 and Table 2. As discussed in Section 2 we can analyze the data either on $V_2(\mathbb{R}^3)$ or on $SO(3)$.

Table 1: Part of data on 85 comets around the Earth $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$.

index	object	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$
1	1P/Halley	$(\ 0.527, -0.304,\ 0.794)$	$(\ 0.010, -0.931, -0.363)$	$(\ 0.849,\ 0.199, -0.488)$
2	2P/Encke	$(\ 0.901,\ 0.431,\ 0.048)$	$(-0.001,\ 0.113, -0.994)$	$(-0.434,\ 0.895,\ 0.102)$
3	3D/Biela	$(-0.341,\ 0.700,\ 0.628)$	$(-0.010,\ 0.665, -0.747)$	$(-0.940, -0.261, -0.220)$
4	5D/Brorsen	$(-0.235,\ 0.939,\ 0.250)$	$(\ 0.003, -0.257,\ 0.966)$	$(\ 0.972,\ 0.227,\ 0.058)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
85	P/2009 L2 (Yang–Gao)	$(-0.164, -0.961,\ 0.221)$	$(-0.005,\ 0.225,\ 0.974)$	$(-0.986,\ 0.159, -0.042)$
mean		$(\ 0.115,\ 0.113,\ 0.022)$	$(\ 0.001, -0.102, 0.038)$	$(\ 0.140, -0.233, -0.091)$

4.1 The test of uniformity based on Rayleigh's statistic

As a preliminary analysis we test whether the orbits of the comets and asteroids are uniformly distributed over $V_2(\mathbb{R}^3)$ or $SO(3)$.

We first recall the Rayleigh's statistic for Stiefel manifolds. Let $\bar{\boldsymbol{x}}_{1:r}$ be the sample mean matrix of a data set on $V_r(\mathbb{R}^p)$ and N be the sample size. Under the null hypothesis of uniformity over $V_r(\mathbb{R}^p)$ the Rayleigh statistic

$$S_{1:r} = pN \cdot \mathrm{tr}(\bar{\boldsymbol{x}}_{1:r}^T \bar{\boldsymbol{x}}_{1:r}) \tag{11}$$

Table 2: Part of data on 6496 asteroids $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$.

index	object	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$
1	433 Eros	$(-0.548,\ 0.837,\ 0.004)$	$(-0.155, -0.110,\ 0.982)$	$(\ 0.822,\ 0.538,\ 0.187)$
2	719 Albert	$(\ 0.939, -0.340,\ 0.082)$	$(-0.014,\ 0.201,\ 0.980)$	$(-0.340, -0.920, -0.183)$
3	887 Alinda	$(-0.191,\ 0.981, -0.030)$	$(\ 0.153,\ 0.057,\ 0.987)$	$(\ 0.970,\ 0.184, -0.160)$
4	1036 Ganymed	$(\ 0.933, -0.140,\ 0.331)$	$(-0.262,\ 0.365,\ 0.893)$	$(-0.250, -0.920,\ 0.304)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
6496	(6344 P–L)	$(\ 0.536,\ 0.842, -0.070)$	$(-0.005,\ 0.082,\ 0.997)$	$(\ 0.844, -0.530,\ 0.048)$
mean		$(\ 0.074,\ 0.018, -0.000)$	$(\ 0.012,\ 0.003,\ 0.949)$	$(\ 0.016, -0.070,\ 0.002)$

is asymptotically distributed according to the chi-square distribution with rp degrees of freedom. Similarly we can define the Rayleigh statistic for the special orthogonal group. Let $\bar{\boldsymbol{x}}$ be the sample mean matrix of a data set on $SO(p)$ and N be the sample size. Under the null hypothesis of uniformity over $SO(p)$, the Rayleigh statistic

$$S = pN \cdot \mathrm{tr}(\bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}}) \tag{12}$$

is asymptotically distributed according to the chi-square distribution with p^2 degrees of freedom (see Remark 1).

From Table 1, the sample mean matrix of comets' data is calculated as

$$\bar{\boldsymbol{x}} = \begin{pmatrix} 0.257 & 0.044 & 0.189 \\ 0.158 & -0.052 & -0.146 \\ 0.079 & 0.765 & 0.004 \end{pmatrix}.$$

Since the $(3, 2)$ element of $\bar{\boldsymbol{x}}$ is large, the orbital plane of the comets are typically close to that of the Earth. Let $\bar{\boldsymbol{x}}_{1:2}$ be the first two columns of $\bar{\boldsymbol{x}}$. The Rayleigh statistic (11) for $V_2(\mathbb{R}^3)$ is

$$S_{1:2} = 3 \cdot 85 \cdot \mathrm{tr}(\bar{\boldsymbol{x}}_{1:2}^\top \bar{\boldsymbol{x}}_{1:2}) = 175.2$$

with the p-value almost zero. The Rayleigh statistic (12) for $SO(3)$ is

$$S = 3 \cdot 85 \cdot \mathrm{tr}(\bar{\boldsymbol{x}}^\top \bar{\boldsymbol{x}}) = 189.8$$

with the p-value almost zero.

Similarly for asteroids data in Table 2 the sample mean matrix is given as

$$\bar{\boldsymbol{x}} = \begin{pmatrix} 0.074 & 0.012 & 0.016 \\ 0.018 & 0.003 & -0.070 \\ -0.000 & 0.949 & 0.002 \end{pmatrix}$$

and the null hypothesis of uniformity is rejected by both

$$S_{1:2} = 3 \cdot 6496 \cdot \mathrm{tr}(\bar{\boldsymbol{x}}_{1:2}^\top \bar{\boldsymbol{x}}_{1:2}) = 1.77 \times 10^4$$

and

$$S = 3 \cdot 6496 \cdot \mathrm{tr}(\bar{\boldsymbol{x}}^\top \bar{\boldsymbol{x}}) = 1.78 \times 10^4.$$

The p-values are almost zero.

4.2 Maximum likelihood estimate of Fisher distributions

We compute the MLE (maximum likelihood estimate) of the Fisher distribution on $V_2(\mathbb{R}^3)$ and $SO(3)$ by using the two methods described in Section 3. For clarity we denote the parameter of the Fisher distribution on $V_2(\mathbb{R}^3)$ and $SO(3)$ by $\Theta_{1:2}$ and Θ, respectively.

First we compute the MLE by the holonomic gradient descent with solving numerically the associated dynamical system along gradient directions. We add a superscript (h) as $\hat{\Theta}^{(\mathrm{h})}$ for values computed by the holonomic gradient descent. For the comets' data the MLE of the Fisher distribution on $V_2(\mathbb{R}^3)$ is

$$\hat{\Theta}^{(\mathrm{h})}_{1:2} = \begin{pmatrix} 0.689 & 0.341 \\ 0.394 & -0.229 \\ 0.496 & 4.273 \end{pmatrix}$$

The starting point is $\begin{pmatrix} 0.1 & 0.1 \\ 0.1 & -0.1 \\ 0.1 & 5.1 \end{pmatrix}$ which is found by the exhaustive search of the optimal point on the grids with $\theta_{ij} = \pm 0.1, \pm 5.1$ (Section 3.1.3). We apply the HGD and obtain an optimal point and apply it again to correct numerical errors and obtain the MLE stated first. The search domain of the HGD is $\begin{pmatrix} (0.1, 1.0) & (0.1, 1.0) \\ (0.1, 1.0) & (-1, -0.1) \\ (0.1, 1.0) & (0.1, 10.0) \end{pmatrix}$. Here the $(1,1)$-entry $(0.1, 1.0)$ of the search domain means that the variable θ_{11} is confined in the interval $(0.1, 1.0)$ during the gradient descent. Other entries stand for corresponding variables.

The MLE of the Fisher distribution on $SO(3)$ is

$$\hat{\Theta}^{(\mathrm{h})} = \begin{pmatrix} 2.953 & 0.200 & 0.871 \\ -0.423 & -0.317 & 2.390 \\ 0.378 & 5.566 & 0.251 \end{pmatrix}$$

Since the MLE is near to the singular locus of the Pfaffian system in case of $SO(3)$, the grid method (Section 3.1.3) to find a starting point does not work well. We use the MLE obtained by the series expansion method below as the starting point. In this case, the HGD is used to confirm and make the answer by an other method more precise.

For the asteroid data, the difference scheme of the dynamical system is instable and we cannot find the MLE even when we start from the MLE obtained by the series expansion.

We next compute the MLE (maximum likelihood estimate) of the Fisher distribution on $V_2(\mathbb{R}^3)$ and $SO(3)$ by using the series expansion approach. We add a superscript (s) as $\hat{\Theta}^{(\mathrm{s})}$ for values computed by this method. For the comets' data the MLE of the Fisher distribution on $V_2(\mathbb{R}^3)$ and its SVD are

$$\hat{\Theta}^{(s)}_{1:2} = \begin{pmatrix} 0.689 & 0.341 \\ 0.394 & -0.229 \\ 0.496 & 4.273 \end{pmatrix}$$
$$= \begin{pmatrix} 0.098 & 0.835 \\ -0.041 & 0.547 \\ 0.994 & -0.060 \end{pmatrix} \begin{pmatrix} 4.326 & 0 \\ 0 & 0.767 \end{pmatrix} \begin{pmatrix} 0.126 & 0.992 \\ 0.992 & -0.126 \end{pmatrix}$$

The MLE of the Fisher distribution on $SO(3)$ and its sign-preserving SVD are

$$\hat{\Theta}^{(s)} = \begin{pmatrix} 2.953 & 0.200 & 0.871 \\ -0.423 & -0.317 & 2.390 \\ 0.378 & 5.566 & 0.251 \end{pmatrix}$$
$$= \begin{pmatrix} -0.109 & 0.964 & 0.242 \\ 0.048 & 0.248 & -0.968 \\ -0.993 & -0.093 & -0.073 \end{pmatrix} \begin{pmatrix} -5.614 & 0 & 0 \\ 0 & 3.079 & 0 \\ 0 & 0 & 2.387 \end{pmatrix} \begin{pmatrix} 0.128 & 0.991 & 0.041 \\ 0.879 & -0.132 & 0.458 \\ 0.459 & -0.023 & -0.888 \end{pmatrix}$$

Note that $\det \bar{\boldsymbol{x}} > 0$ but $\det \hat{\Theta} < 0$.

For the asteroid data the MLEs are

$$\hat{\Theta}^{(s)}_{1:2} = \begin{pmatrix} 0.157 & 0.254 \\ 0.038 & 0.060 \\ 0.005 & 19.568 \end{pmatrix}$$
$$= \begin{pmatrix} 0.013 & 0.972 \\ 0.003 & 0.235 \\ 1.000 & -0.013 \end{pmatrix} \begin{pmatrix} 19.570 & 0 \\ 0 & 0.161 \end{pmatrix} \begin{pmatrix} 0.000 & 1.000 \\ 1.000 & -0.000 \end{pmatrix}$$

and

$$\hat{\Theta}^{(s)} = \begin{pmatrix} 0.291 & 0.257 & -0.781 \\ 0.817 & 0.056 & 0.134 \\ 0.001 & 19.601 & 0.056 \end{pmatrix}$$
$$= \begin{pmatrix} 0.013 & -0.721 & 0.693 \\ 0.003 & -0.693 & -0.721 \\ 1.000 & 0.011 & -0.007 \end{pmatrix} \begin{pmatrix} 19.603 & 0.000 & 0.000 \\ 0.000 & 0.908 & 0.000 \\ 0.000 & 0.000 & 0.747 \end{pmatrix} \begin{pmatrix} 0.000 & 1.000 & 0.002 \\ -0.855 & -0.001 & 0.518 \\ -0.518 & 0.002 & -0.855 \end{pmatrix}.$$

The AIC values are given in Table 3. For each data, AIC was minimized by the Fisher distribution on $SO(3)$.

Table 3: AIC of each data and each model.

	comets			asteroids		
	Uniform	$V_2(\mathbb{R}^3)$	$SO(3)$	Uniform	$V_2(\mathbb{R}^3)$	$SO(3)$
AIC	0	-207.0	-219.7	0	-3.47×10^4	-3.48×10^4

A Partial differential equation for ${}_0F_1(p/2, Y)$

If $\Theta = \mathrm{diag}(\theta_{ii})$ is diagonal, then $y_i = \theta_{ii}^2/4$. By change of variables from (8) we have

$$\begin{aligned} & y_i\partial_i^2 + \left\{\frac{p}{2} - \frac{r-1}{2} + \frac{1}{2}\sum_{j=1,j\neq i}^{r}\frac{y_i}{y_i-y_j}\right\}\partial_i - \frac{1}{2}\sum_{j=1,j\neq i}^{r}\frac{y_j}{y_i-y_j}\partial_j - 1 \\ = \; & \partial_{ii}^2 + (p-r)\frac{1}{\theta_{ii}}\partial_{ii} + \sum_{j\neq i}\frac{1}{\theta_{ii}^2-\theta_{jj}^2}\{\theta_{ii}\partial_{ii} - \theta_{jj}\partial_{jj}\} - 1. \end{aligned} \qquad (13)$$

We now show that the numerator of (13) belongs I_{diag} for small dimensions.

For $p = r = 2$ (i.e. for $\mathcal{O}(2)$) by Macaulay2 we checked that the above I is holonomic. Also by asir (nk_restriction), a set of generators of I_{diag} is given as

$$\begin{aligned} h_1 &= (-\theta_{22}^2+\theta_{11}^2)\partial_{11}^4 + 6\theta_{11}\partial_{11}^3 + (2\theta_{22}^2 - 2\theta_{11}^2 + 6)\partial_{11}^2 - 6\theta_{11}\partial_{11} - \theta_{22}^2 + \theta_{11}^2 - 3, \\ h_2 &= (\theta_{22}^2 - \theta_{11}^2)\partial_{11}^2 - \theta_{11}\partial_{11} + \theta_{22}\partial_{22} - \theta_{22}^2 + \theta_{11}^2, \\ h_3 &= \theta_{22}\partial_{11}^4 + \theta_{11}\partial_{22}\partial_{11}^3 + (3\partial_{22} - \theta_{22})\partial_{11}^2 - \theta_{11}\partial_{22}\partial_{11} - 2\partial_{22}, \\ h_4 &= \theta_{11}\theta_{22}\partial_{11}^3 + (\theta_{11}^2\partial_{22} - \theta_{22})\partial_{11}^2 + (-\theta_{11}^2 - 1)\partial_{22} + \theta_{22}, \\ h_5 &= -\partial_{11}^2 + \partial_{22}^2. \end{aligned}$$

Looking at h_2 and h_5 we have

$$\begin{aligned} h_2 &= (\theta_{22}^2 - \theta_{11}^2)\left\{\partial_{11}^2 + \frac{\theta_{11}\partial_{11} - \theta_{22}\partial_{22}}{\theta_{11}^2 - \theta_{22}^2} - 1\right\}, \\ \frac{h_2}{\theta_{22}^2 - \theta_{11}^2} + h_5 &= \left\{\partial_{22}^2 + \frac{\theta_{22}\partial_{22} - \theta_{11}\partial_{11}}{\theta_{22}^2 - \theta_{11}^2} - 1\right\}. \end{aligned}$$

These are the same as (13) for $p = r = 2$.

For the case of $V_2(\mathbb{R}^3)$ ($p = 3$, $r = 2$) by Macaulay2 we have checked that I is holonomic By asir (nk_restriction) I_{diag} has the set of generators:

$$\begin{aligned} h_1 &= -\theta_{11}\partial_{22}\partial_{11}^2 + (-\theta_{22}\partial_{22}^2 - 3\partial_{22} + \theta_{22})\partial_{11} + \theta_{11}\partial_{22}, \\ h_2 &= \theta_{11}\theta_{22}\partial_{11}^2 + \theta_{22}\partial_{11} - \theta_{11}\theta_{22}\partial_{22}^2 - \theta_{11}\partial_{22}, \\ h_3 &= \theta_{11}^2\partial_{11}^2 + 2\theta_{11}\partial_{11} - \theta_{22}^2\partial_{22}^2 - 2\theta_{22}\partial_{22} + \theta_{22}^2 - \theta_{11}^2, \\ h_4 &= -\theta_{11}\partial_{11}^2 + (\theta_{22}^2\partial_{22}^2 + 2\theta_{22}\partial_{22} - \theta_{22}^2 - 1)\partial_{11} + \theta_{11}\theta_{22}\partial_{22}^3 + 2\theta_{11}\partial_{22}^2 - \theta_{11}\theta_{22}\partial_{22}, \\ h_5 &= (-\theta_{11}\theta_{22}\partial_{22}^2 - \theta_{11}\partial_{22} + \theta_{11}\theta_{22})\partial_{11} - \theta_{22}^2\partial_{22}^3 - 4\theta_{22}\partial_{22}^2 + (\theta_{22}^2 - 2)\partial_{22} + 2\theta_{22}, \\ h_6 &= -\theta_{11}\theta_{22}\partial_{11} + (\theta_{22}^3 - \theta_{11}^2\theta_{22})\partial_{22}^2 + (2\theta_{22}^2 - \theta_{11}^2)\partial_{22} - \theta_{22}^3 + \theta_{11}^2\theta_{22}. \end{aligned}$$

Looking at h_6

$$\begin{aligned} h_6 &= -\theta_{11}\theta_{22}\partial_{11} + (\theta_{22}^3 - \theta_{11}^2\theta_{22})\partial_{22}^2 + (2\theta_{22}^2 - \theta_{11}^2)\partial_{22} - \theta_{22}^3 + \theta_{11}^2\theta_{22} \\ &= (\theta_{22}^2 - \theta_{11}^2)\theta_{22}\left\{\partial_{22}^2 + \frac{2\theta_{22}^2 - \theta_{11}^2}{(\theta_{22}^2 - \theta_{11}^2)\theta_{22}}\partial_{22} - \frac{\theta_{11}}{\theta_{22}^2 - \theta_{11}^2}\partial_{11} - 1\right\} \\ &= (\theta_{22}^2 - \theta_{11}^2)\theta_{22}\left\{\partial_{22}^2 + \frac{1}{\theta_{22}}\partial_{22} + \frac{\theta_{22}\partial_{22} - \theta_{11}\partial_{11}}{\theta_{22}^2 - \theta_{11}^2} - 1\right\} \end{aligned}$$

we see that it coincides with the case of $p = 3, r = 2, i = 2$ in (13).

B Asymptotic evaluation of $E(k, l, m)$

We derive an asymptotic form of $E(k, l, m)$ when k, l, m simultaneously go to infinity. Let $k = n\alpha$, $l = n\beta$ and $m = n\gamma$ where α, β and γ are fixed positive numbers. We use Laplace's method to show

$$E(k, l, m) \sim \sqrt{\frac{2}{\pi}}((k + l)(l + m)(k + m))^{-1/2} \tag{14}$$

as $n \to \infty$. The integrand $x_{11}^k x_{22}^l x_{33}^m$ of $E(k, l, m)$ is maximized at four points $(x_{11}, x_{22}, x_{33}) = (1, 1, 1)$, $(-1, -1, 1)$, $(-1, 1, -1)$ and $(1, -1, -1)$ as long as k, l, m are all even or all odd. By symmetry it is sufficient to consider the neighborhood of $\mathrm{diag}(1, 1, 1)$, where X is approximated by

$$X = \begin{pmatrix} (1 - \epsilon_1^2 - \epsilon_2^2)^{1/2} & -\epsilon_1 & -\epsilon_2 \\ \epsilon_1 & (1 - \epsilon_1^2 - \epsilon_3^2)^{1/2} & -\epsilon_3 \\ \epsilon_2 & \epsilon_3 & (1 - \epsilon_2^2 - \epsilon_3^2)^{1/2} \end{pmatrix}$$

with sufficiently small numbers $\epsilon_1, \epsilon_2, \epsilon_3$. The density of $(\epsilon_1, \epsilon_2, \epsilon_3)$ with respect to the Lebesgue measure $d\epsilon_1 d\epsilon_2 d\epsilon_3$ is $1/\mathrm{Vol}(SO(3)) = 1/(8\pi^2)$. Hence we obtain

$$\begin{aligned} E(k, l, m) &\sim 4\int (1 - \epsilon_1^2 - \epsilon_2^2)^{k/2}(1 - \epsilon_1^2 - \epsilon_3^2)^{l/2}(1 - \epsilon_2^2 - \epsilon_3^2)^{m/2}\frac{1}{8\pi^2}d\epsilon_1 d\epsilon_2 d\epsilon_3 \\ &\sim 4\int \mathrm{e}^{-(k+l)\epsilon_1^2/2 - (k+m)\epsilon_2^2/2 - (l+m)\epsilon_3^2/2}\frac{1}{8\pi^2}d\epsilon_1 d\epsilon_2 d\epsilon_3 \\ &= \sqrt{\frac{2}{\pi}}((k + l)(l + m)(k + m))^{-1/2}. \end{aligned}$$

We have checked that the right-hand side gives a good approximation to the exact value of $E(k, l, m)$ for $k + l + m \geq 100$.

The same argument shows that for $SO(p)$

$$E\left[\prod_{i=1}^{p} x_{ii}^{k_i}\right] \sim \frac{p(p-1)}{2}\frac{1}{\mathrm{Vol}(SO(p))}\left(\prod_{i<j}(k_i + k_j)\right)^{-1/2}$$

as $n \to \infty$ when $k_i = n\alpha_i$, $\alpha_i > 0$, and k_i's are are all even or all odd.

References

Ole Barndorff-Nielsen. *Information and Exponential Families.* John Wiley & Sons Inc., New York, 1978. Wiley Series in Probability and Mathematical Statistics.

Yasuko Chikuse. *Statistics on Special Manifolds*, volume 174 of *Lecture Notes in Statistics.* Springer-Verlag, New York, 2003. ISBN 0-387-00160-3.

Daniel R. Grayson and Michael E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/.

A. T. James. Calculation of zonal polynomial coefficients by use of the Laplace-Beltrami operator. *Ann. Math. Statist*, 39:1711–1718, 1968. ISSN 0003-4851.

P. E. Jupp and K. V. Mardia. Maximum likelihood estimators for the matrix von Mises–Fisher and Bingham distributions. *The Annals of Statistics*, 7:599–606, 1979.

C. G. Khatri and K. V. Mardia. The von Mises–Fisher distribution in orentation statistics. *Journal of the Royal Statistical Society: Series B*, 39:95–106, 1977.

K. V. Mardia. Statistics of directional data. *Journal of the Royal Statistical Society: Series B*, 37:349–393, 1975.

Kanti V. Mardia and Peter E. Jupp. *Directional Statistics.* Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2000. ISBN 0-471-95333-4.

Brian G. Marsden. *Catalogue of Cometary Orbits.* Smithonian Astrophysical Observatory, Cambridge, MA, 1972.

R. J. Muirhead. Systems of partial differential equations for hypergeometric functions of matrix argument. *Ann. Math. Statist.*, 41:991–1001, 1970. ISSN 0003-4851.

Robb J. Muirhead. *Aspects of Multivariate Statistical Theory.* John Wiley & Sons Inc., New York, 1982. ISBN 0-471-09442-0. Wiley Series in Probability and Mathematical Statistics.

Hiromasa Nakayama, Kenta Nishiyama, Masayuki Noro, Katsuyoshi Ohara, Tomonari Sei, Nobuki Takayama, and Akimichi Takemura. Holonomic gradient descent and its application to the Fisher-Bingham integral. *Advances in Applied Mathematics*, 47: 639–658, 2011. doi: 10.1016/j.aam.2011.03.001.

OpenXM Mathematics Repository. Data for the special orthogonal group and the stiefel manifold. http://www.math.kobe-u.ac.jp/OpenXM/Math/Fisher-Bingham/SO3.

Michael J. Prentice. Orientation statistics without parametric assumptions. *J. Roy. Statist. Soc. Ser. B*, 48(2):214–222, 1986. ISSN 0035-9246.

RisaAsir developing team. Risa/asir, a computer algebra system. Available at http://www.math.kobe-u.ac.jp/Asir/.

Akimichi Takemura. *Zonal Polynomials.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 4. Institute of Mathematical Statistics, Hayward, CA, 1984. ISBN 0-940600-05-6.

Andrew T. A. Wood. Estimation of the concentration parameters of the Fisher matrix distribution on SO(3) and the Bingham distribution on $S_q,\ q \geq 2$. *Austral. J. Statist.*, 35(1):69–79, 1993. ISSN 0004-9581.