

楕円状の等高線を持つ 球面上の確率密度関数について

高崎経済大学・経済学部* 宮田 庸一

Yoichi Miyata

Faculty of Economics,

Takasaki City University of Economics

概要

球面上の統計モデルとしてよく知られているのは von Mises-Fisher 分布であるが, その密度関数の等高線を球面に描くと, 平均方向を表すベクトルを中心にして同心円状に広がっていくことが知られている. 本稿では, von Mises-Fisher 分布では表現が難しい球面上に楕円状の等高線を持つ統計モデルについての紹介を行い, そのモデルの持つ問題点について解説する.

1 はじめに

方向統計学とは, 方向という観測値を扱うために開発された統計手法の分野である. 風向であれば, 真北の方向を 0 ラジアンとして基準となる点を決めておくと, そこから計測される角度 θ ($-\pi \leq \theta < \pi$) が方向データとなる. また, このようなデータは, $\mathbf{x} = (\cos \theta, \sin \theta)^\top$ とおくことで, $\|\mathbf{x}\| = 1$ となる単位円周 S^1 上のデータとみなすこともできる. この考え方は, 一般の単位超球面 S^{p-1} ($p \geq 2$) 上データに拡張することが可能である. このような超球面上のデータに対する統計解析としては, 古くは Fisher (1953) によりアイスランドの溶岩流 (歴史時代および先史時代) の残留磁化の S^2 上の方向データ*¹に対する統計解析がなされている. また, 球面上データに対する統計解析は, タンパク質の二面角に関するクラスタリング (Dowe et al., 1996), Kent 分布の有限混合モデル

* 〒370-0801 高崎市上並榎町 1300

*¹ このデータは, Mardia and Jupp (2000) の Table 9.1 で見ることができる.

を使用した岩塊の不連続面群のモデリング (Peel et al., 2001) など様々な分野で使用されている。また近年では、文書における p 個の単語の出現頻度を $\mathbf{n} = (n_1, \dots, n_p)$ とおくと、それを正規化したベクトル $\mathbf{x} = (n_1/\|\mathbf{n}\|, \dots, n_p/\|\mathbf{n}\|)$ は、 $\mathbb{S}^{p-1}$ 上の観測値であるとみなせるため、これを用いたクラスタリングが Banerjee et al. (2003, 2005) および Gopal and Yang (2014) によりなされている。超球面 $\mathbb{S}^{p-1}$ 上のデータに対して最尤推定量を導出するためには、統計モデルを定める必要がある。超球面上のモデルとして最もポピュラーなものは、von Mises-Fisher 分布、もしくは Langevin(ランジュバン) 分布と呼ばれる確率分布で、以下の確率密度関数を持つ:

$$f(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \left(\frac{\kappa}{2}\right)^{p/2-1} \frac{1}{2\pi^{p/2} I_{p/2-1}(\kappa)} \exp(\kappa \boldsymbol{\mu}^\top \mathbf{x}), \quad \mathbf{x} \in \mathbb{S}^{p-1} \quad (1)$$

ただし $I_{p/2-1}(\cdot)$ は $p/2 - 1$ 次の第 1 種変形ベッセル関数とする。ここで $\boldsymbol{\mu} \in \mathbb{S}^{p-1}$ は平均方向を表し、 $\kappa \geq 0$ は分布の集中度を制御するパラメーターとなる。この分布は、 $\mathbf{x} = \boldsymbol{\mu}$ の点で唯一の最大値を取ることから、単峰性 (Unimodality) を持つ。なお Fisher (1953) では $p = 3$, 即ち $\mathbb{S}^2$ 上の分布を提案している。この分布は様々な理論的に良い性質を持つことが知られているが、式 (1) からわかるように、密度関数の等高線を球面上に描くと、図 1 のように平均方向 $\boldsymbol{\mu}$ を中心にして同心円状に広がる。

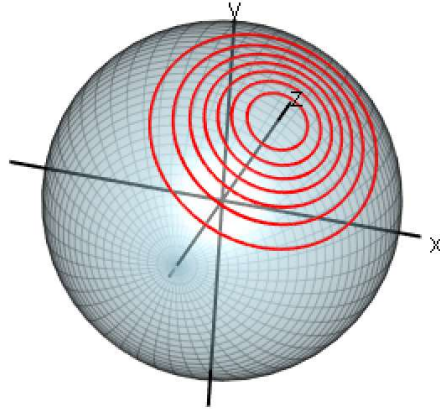


図 1 $\boldsymbol{\mu} = (0, 0, 1)^\top$, $\kappa = 10$.

von Mises-Fisher 分布では表現が難しい楕円状の等高線を持つ統計モデルとして、Fisher-Bingham 分布 (Mardia and Jupp, 2000, p.174), Kent 分布 (Kent, 1982), 射影正規分布 (Mardia and Jupp, 2000, p.178) などが知られている。また最近のトピックとしては、射影正規分布に対して適切な制約をいれることにより、実用上の取り扱いを容易に

した楕円対称角度ガウス分布が, Paine et al. (2018) により提案されている. 本報告においては, 特に Kent 分布および楕円対称角度ガウス分布に焦点をあて, それぞれの長所および短所について解説を行う.

2 Fisher-Bingham 分布

確率ベクトル $\mathbf{X} = (X_1, \dots, X_p)^\top \in \mathbb{S}^{p-1}$ が確率密度関数

$$f(\mathbf{x}; \boldsymbol{\mu}, \kappa, \mathbf{A}) = \frac{1}{a(\kappa, \mathbf{A})} \exp(\kappa \boldsymbol{\mu}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{A} \mathbf{x}), \quad \mathbf{x} \in \mathbb{S}^{p-1}, \quad (2)$$

を持つとき, Fisher-Bingham 分布に従うという. ただし $\boldsymbol{\mu}$ は $\|\boldsymbol{\mu}\| = 1$ を満たすベクトルとし, $\mathbf{A}$ は $p \times p$ の実対称行列とする. これは $\mathbf{x}^\top \mathbf{x} = 1$ の性質を用いると, 一般性を失わず $\text{tr}(\mathbf{A}) = 0$ とできる. これにより, 推定すべきパラメーター数を減らすことができる. また $d\sigma(\mathbf{x})$ は $\mathbb{S}^{p-1}$ 上の面積要素とすると, 正規化定数は

$$a(\kappa, \mathbf{A}) = \int_{\mathbb{S}^{p-1}} \exp(\kappa \boldsymbol{\mu}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{A} \mathbf{x}) d\sigma(\mathbf{x}) \quad (3)$$

として与えられる. Fisher-Bingham 分布は $\mathbf{A} = \mathbf{O}$ のとき, von Mises-Fisher 分布になり, $\kappa = 0$ のときは, Bingham 分布 (Bingham, 1964) となる.

2.1 Kent 分布

Fisher-Bingham 分布はパラメーターが $p(p+1)/2 + p - 1$ 個あり, かなり大きなモデルであるため, 実用上はよりパラメーター数を少なくしたモデルが好ましいことがしばしばある. Kent (1982) は, Fisher-Bingham 分布 (2) の対称行列 $\mathbf{A}$ に対して, 条件 $\text{tr}(\mathbf{A}) = 0$ に加えて, 制約

$$\mathbf{A}\boldsymbol{\mu} = \mathbf{0}_p \quad (4)$$

を入れた分布を提案した. ただし $\mathbf{0}_p$ は p 次元のゼロベクトルである. 制約 (4) は, $\boldsymbol{\mu}$ が $\mathbf{A}$ の固有ベクトルになり, 0 が $\mathbf{A}$ の固有値になることを課している. これらの制約を入れることで, Fisher-Bingham 分布は, より解釈のしやすい Parametrization を持つ統計モデルとして表すことができる.

ここでは $p = 3$ の場合を考える. β_1, β_2 を $\mathbf{A}$ の 0 以外の固有値とし, $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ は, 固有値 β_1, β_2 に対応する大きさ 1 の固有ベクトルとし, $\boldsymbol{\mu}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ はそれぞれ直交するものとす

る. ここで, $\beta = \beta_1 = -\beta_2$ とさらなる制約を入れることで, Fisher-Bingham 分布 (2) の密度関数は,

$$f_K(\mathbf{x}; \boldsymbol{\mu}, \kappa, \beta) = \frac{1}{a(\kappa, \beta)} \exp \{ \kappa \boldsymbol{\mu}^\top \mathbf{x} + \beta ((\boldsymbol{\xi}_1^\top \mathbf{x})^2 - (\boldsymbol{\xi}_2^\top \mathbf{x})^2) \}, \quad \mathbf{x} \in \mathbb{S}^2, \quad (5)$$

$$a(\kappa, \beta) = 2\pi \sum_{j=0}^{\infty} \frac{\Gamma(j+1/2)}{\Gamma(j+1)} \beta^{2j} \left(\frac{\kappa}{2} \right)^{-2j-(1/2)} I_{2j+(1/2)}(\kappa)$$

と, 比較的簡素な形で表すことができる. ただし $\Gamma(\cdot)$ はガンマ関数とする. ここで $\boldsymbol{\mu} \in \mathbb{S}^2$ は平均方向を表す単位ベクトルとなり, $\kappa \geq 0$ は分布の集中度を表し, $\beta \geq 0$ は楕円の形状を制御するパラメータとなる. 図 2, 3 からわかるように, この分布は, $\boldsymbol{\xi}_1$ を長軸の方向, $\boldsymbol{\xi}_2$ を短軸の方向とする楕円上の等高線を与えることができる.

$$\boldsymbol{\mu} = (0, 0, 1)^\top, \boldsymbol{\xi}_1 = (1, 0, 0)^\top, \boldsymbol{\xi}_2 = (0, 1, 0)^\top$$

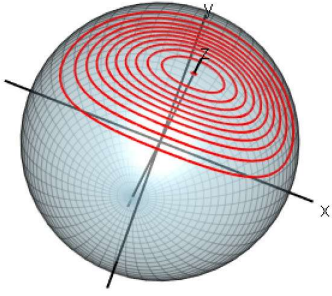


図 2 $\kappa = 10, \beta = 3$

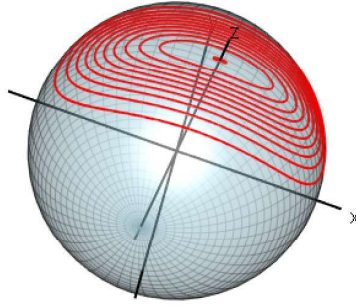


図 3 $\kappa = 10, \beta = 4.5$

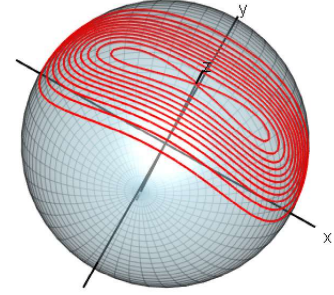


図 4 $\kappa = 10, \beta = 5.5$

$\boldsymbol{\mu}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ は本質的に 3 つの独立なパラメータを用いて表現できるので, Kent 分布 (5) は, 5 パラメータ Fisher-Bingham 分布とも呼ばれる. 分布の単峰性については, 以下の結果が知られている:

定理 1 (Kent, 1982) $\kappa > 2\beta \implies$ Kent 分布 (5) は単峰

一般の超球面 $\mathbb{S}^{p-1}$ ($p \geq 3$) 上の Kent 分布に対する単峰性に関する結果は, Kent et al. (2018, p.294) により, 証明なしで与えられている. また一般の超球面 $\mathbb{S}^{p-1}$ ($p \geq 3$) 上の Kent 分布に従う乱数は, 棄却法を用いて生成できることが Kent et al. (2013) により示されている.

2.2 Fisher-Bingham 分布, Kent 分布における問題点

Fisher-Bingham 分布, Kent 分布においては, 正規化定数 (3) を陽に表すことが一般的にはできないため, 何らかの方法により解決する必要がある. 一つ目の方法は, 正規化定数に対して何らかの近似式を与えることである. その方法としては, 鞍点法 (Kume and Wood, 2005), 連続型オイラー変換 (Chen and Tanaka, 2021), モンテカルロ法, ベキ級数法, ホロノミック勾配法 (Kume and Sei, 2018; Nakayama et al., 2011) などが知られている. ここでは, 鞍点法について解説する. 密度関数 (2) において, $\kappa\boldsymbol{\mu} = (\gamma_1, \dots, \gamma_p)^\top$ とおき, また $\mathbf{A}$ の固有値 $\lambda_1, \dots, \lambda_p$ と適当な直交行列 $\mathbf{Q}$ を用いて $\mathbf{A} = \mathbf{Q}\text{diag}\{\lambda_1, \dots, \lambda_p\}\mathbf{Q}^\top$ と表す. また W_j ($j = 1, \dots, p$) を独立に正規分布 $N(\gamma_j/(-2\lambda_j), 1/(-2\lambda_j))$ に従う確率変数とし, $R^2 = \sum_{j=1}^p W_j^2$ の密度関数を $f_{\lambda, \gamma}(r^2)$ とする. このとき, 正規化定数 (3) は

$$a(\kappa, \mathbf{A}) = 2\pi^{p/2} \prod_{j=1}^p (-\lambda_j)^{-1/2} f_{\lambda, \gamma}(1) \exp \left\{ \frac{1}{4} \sum_{j=1}^p \frac{\gamma_j^2}{-\lambda_j} \right\}$$

と表すことができる. ここで, 近似すべき箇所は R^2 の密度関数を $R^2 = 1$ で評価した $f_{\lambda, \gamma}(1)$ である. これより, $f_{\lambda, \gamma}(1)$ を鞍点法で近似したものが Kume and Wood (2005) であり, 連続型オイラー変換を用いて近似したものが Chen and Tanaka (2021) の結果となる.

2 番目の手法として考えられるのは, 正規化定数を用いない推定方法を用いることである. これはスコアマッチング法と呼ばれる, 正規化定数に依存しない, しかしモデルの密度関数から構成される損失関数を最小化させる方法であり, これは Hyvärinen and Dayan (2005) により与えられている.

3 番目の手法としては, 複雑な正規化定数を含まない, 別の統計モデルを用いることである. 球面上の統計モデルとしては, 次の節で紹介する角度正規分布がそれに該当する.

3 射影正規分布, 角度正規分布

確率ベクトル $\mathbf{X} = (X_1, \dots, X_p)^\top \in \mathbb{S}^{p-1}$ が以下の確率密度関数を持つとき, 射影正規分布もしくは角度正規 (Angular Gaussian) 分布に従うという

$$f_{\text{AG}}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{V}) = \frac{1}{(2\pi)^{(p-1)/2} |\mathbf{V}|^{1/2} (\mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x})^{p/2}} \exp \left[\frac{1}{2} \left\{ \frac{(\mathbf{x}^\top \mathbf{V}^{-1} \boldsymbol{\mu})^2}{\mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x}} - \boldsymbol{\mu}^\top \mathbf{V}^{-1} \boldsymbol{\mu} \right\} \right] \times \mathcal{M}_{p-1} \left(\frac{\mathbf{x}^\top \mathbf{V}^{-1} \boldsymbol{\mu}}{(\mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x})^{1/2}} \right), \quad \mathbf{x} \in \mathbb{S}^{p-1}, \quad (6)$$

ただし $\mathbf{V}$ は $p \times p$ の正値定符号の正則な実対称行列とし、

$$\mathcal{M}_{p-1}(\alpha) = \int_0^\infty u^{p-1} \frac{1}{(2\pi)^{1/2}} \exp \left\{ -\frac{1}{2}(u - \alpha)^2 \right\} du \quad (7)$$

とする. $\phi(z)$, $\Phi(z)$ はそれぞれ標準正規分布 $N(0, 1)$ の密度関数と累積分布関数とするとき, $\mathcal{M}_0(\alpha) = \Phi(\alpha)$, $\mathcal{M}_1(\alpha) = \alpha\Phi(\alpha) + \phi(\alpha)$, $\mathcal{M}_2(\alpha) = (1 + \alpha^2)\Phi(\alpha) + \alpha\phi(\alpha)$ と表せるため, 計算は容易に行うことができる. また, 式 (6) は, $\mathbf{Y}$ を p 変量正規分布 $N_p(\boldsymbol{\mu}, \mathbf{V})$ に従う確率ベクトルとするときに, $\mathbf{X} = \mathbf{Y}/\|\mathbf{Y}\|$ の分布の密度関数として導出されるため, 乱数生成は極めて容易である^{*2} しかしながら, 角度正規分布は, 任意の $c > 0$ に対して, $f_{\text{AG}}(\mathbf{x}; c\boldsymbol{\mu}, c^2\mathbf{V}) = f_{\text{AG}}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{V})$ が成り立つため, パラメーターに関して識別不可能なモデルであることがわかる. Paine et al. (2018) は, 制約のないパラメーターの個数を減らすために, 式 (6) に対して, 以下の制約を入れた:

$$\mathbf{V}\boldsymbol{\mu} = \boldsymbol{\mu} \quad (1 \text{ が固有値}), \quad (8)$$

$$|\mathbf{V}| = 1, \quad (9)$$

ただし $|\cdot|$ は行列式を表す記号とする. これにより, 以下の形の密度関数をえることができる:

$$\begin{aligned} f_{\text{ESAG}}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{V}) &= \frac{1}{(2\pi)^{(p-1)/2} (\mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x})^{p/2}} \exp \left[\frac{1}{2} \left\{ \frac{(\mathbf{x}^\top \boldsymbol{\mu})^2}{\mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x}} - \boldsymbol{\mu}^\top \boldsymbol{\mu} \right\} \right] \\ &\times \mathcal{M}_{p-1} \left(\frac{\mathbf{x}^\top \boldsymbol{\mu}}{(\mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x})^{1/2}} \right), \quad \mathbf{x} \in \mathbb{S}^{p-1}. \end{aligned} \quad (10)$$

ここでベクトル $\boldsymbol{\mu}$ は, Kent 分布とは異なり, $\mathbb{R}^3$ 上のベクトルであることに注意されたい. この分布は楕円対称角度ガウス (Elliptically symmetric angular Gaussian, ESAG) 分布と呼ばれる.

パラメーター $\boldsymbol{\mu}$, $\mathbf{V}$ は, 2つの制約 (8), (9) が課されているため, 最尤法のようにパラメーターに対して最適化を行うには, 取り扱いのしづらい形になっている. このため, ここでは Paine et al. (2018) に従って, $p = 3$ の場合, 即ち $\mathbb{S}^2$ 上の密度関数 (10) に対する Reparametrization を行う. まず, $\rho_1 \geq \rho_2 > 0$ を $\mathbf{V}$ の固有値とし, $\boldsymbol{\xi}_j$ ($j = 1, 2$) を ρ_j に対応する固有ベクトルとする. 式 (8) より 1 も $\mathbf{V}$ の固有値であり, その固有ベクトルは $\boldsymbol{\xi}_3 = \boldsymbol{\mu}/\|\boldsymbol{\mu}\|$ となっている. $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^\top \in \mathbb{R}^3$, $\mu_0 = \sqrt{\mu_2^2 + \mu_3^2}$ として, 以下の $\mathbb{R}^3$

^{*2} 即ち, $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$ として, $\mathbf{X} = \mathbf{Y}/\|\mathbf{Y}\|$ とおけば, 角度正規分布からの乱数となる.

における正規直交基底 $\tilde{\xi}_1, \tilde{\xi}_2, \xi_3$ を作る:

$$\begin{aligned}\xi_3 &= \frac{\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|}, \\ \tilde{\xi}_1 &= \frac{1}{\mu_2 \|\boldsymbol{\mu}\|} \begin{pmatrix} -\mu_0^2 \\ \mu_1 \mu_2 \\ \mu_1 \mu_3 \end{pmatrix} = \frac{1}{\sqrt{\mu_2^2 + \mu_3^2} \|\boldsymbol{\mu}\|} \begin{pmatrix} -\mu_2^2 - \mu_3^2 \\ \mu_1 \mu_2 \\ \mu_1 \mu_3 \end{pmatrix} \\ \tilde{\xi}_2 &= \frac{1}{\mu_0} \begin{pmatrix} 0 \\ -\mu_3 \\ \mu_2 \end{pmatrix}\end{aligned}$$

次に $\tilde{\xi}_1, \tilde{\xi}_2$ で張られる部分空間において, $\tilde{\xi}_1, \tilde{\xi}_2$ をそれぞれ $\psi \in (0, \pi]$ 回転させたものを ξ_1, ξ_2 とおく:

$$\begin{aligned}\xi_1 &= \cos \psi \tilde{\xi}_1 + \sin \psi \tilde{\xi}_2 \\ \xi_2 &= -\sin \psi \tilde{\xi}_1 + \cos \psi \tilde{\xi}_2\end{aligned}$$

ここで $\rho_1 = \rho, \rho_2 = 1/\rho$ ($\rho \in (0, 1]$) とおき, パラメーターに制約を入れると,

$$\begin{aligned}V^{-1} &= \xi_3 \xi_3^\top + \frac{1}{\rho_1} \xi_1 \xi_1^\top + \frac{1}{\rho_2} \xi_2 \xi_2^\top \\ &= \left(\frac{1}{\rho} \cos^2 \psi + \rho \sin^2 \psi \right) \tilde{\xi}_1 \tilde{\xi}_1^\top + \left(\frac{1}{\rho} \sin^2 \psi + \rho \cos^2 \psi \right) \tilde{\xi}_2 \tilde{\xi}_2^\top \\ &\quad + \frac{1}{2} \left(\frac{1}{\rho} - \rho \right) \sin 2\psi (\tilde{\xi}_1 \tilde{\xi}_2^\top + \tilde{\xi}_2 \tilde{\xi}_1^\top) + \tilde{\xi}_3 \tilde{\xi}_3^\top\end{aligned}$$

となる. 最後に $\mathbf{I}_3$ を 3 次の単位行列とし, $\gamma_1 = (1/2)(\rho^{-1} - \rho) \cos 2\psi$, $\gamma_2 = 2^{-1}(\rho^{-1} - \rho) \sin 2\psi$ とおくと,

$$\begin{aligned}V^{-1} &= \mathbf{I}_3 + \gamma_1 (\xi_1 \xi_1^\top - \xi_2 \xi_2^\top) + \gamma_2 (\xi_1 \xi_2^\top + \xi_2 \xi_1^\top) \\ &\quad + \left\{ \sqrt{\gamma_1^2 + \gamma_2^2 + 1} - 1 \right\} (\xi_1 \xi_1^\top + \xi_2 \xi_2^\top)\end{aligned}$$

と表すことができる. このようにすることで, $(\mu_1, \mu_2, \mu_3, \gamma_1, \gamma_2)^\top$ のパラメーター空間 $\{(\mu_1, \mu_2, \mu_3, \gamma_1, \gamma_2) | (\mu_2, \mu_3) \neq (0, 0)\}$ は, ほぼ制約のない形で表される. 以下では, ここで与えられたパラメーター $\boldsymbol{\mu}, \gamma_1, \gamma_2$ が, ESAG 分布にどのように作用するのかを見ていく. 以下の図 5, 6, 7 からわかるように, $\boldsymbol{\mu}/\|\boldsymbol{\mu}\|$ は分布の平均方向を表し, $\|\boldsymbol{\mu}\|$ が分布の集中度, γ_1, γ_2 は楕円の長軸および短軸の方向に影響を与えていることがわかる.

なお, Paine et al. (2018) で提案された Parametrization は, $p = 3$, 即ち $\mathbb{S}^2$ 上の ESAG 分布に対してのみ適用できる. Yu and Huang (2024) では, 一般の $\mathbb{S}^{p-1}$ 上の

ここでは, $\boldsymbol{\mu}^* = (0, 0, 1)^\top$, $\boldsymbol{\mu} = k\boldsymbol{\mu}^*$ とおく.

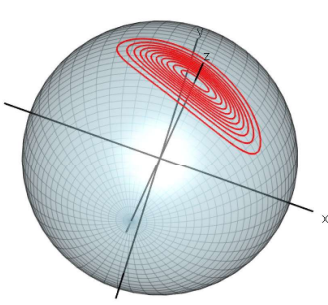


図 5 $k = 6, \gamma = (-1, 1)^\top$

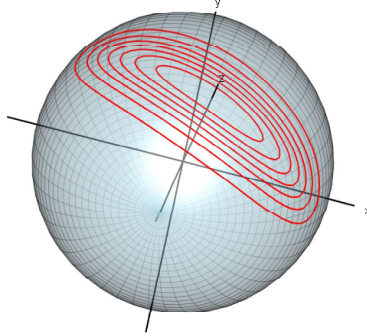


図 6 $k = 3, \gamma = (-1, 1)^\top$

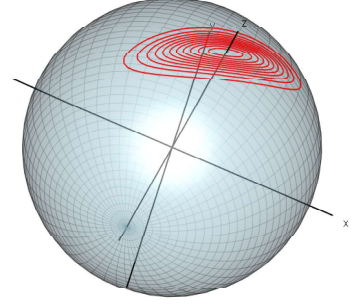


図 7 $k = 6, \gamma = (-1, -1)^\top$

ESAG 分布に対する Parametrization を与えている. しかし Yu and Huang (2024) の Parametrization は Paine et al. (2018) の拡張とはなっておらず, $\mathbb{S}^2$ 上では, Paine et al. (2018) の方法の方が, よりシンプルな Parametrization を与えている.

最後に, ESAG 分布の単峰性に関する結果を紹介する.

定理 2 (Paine et al., 2018, Proposition 1) $0 \leq \rho_1 \leq \dots \leq \rho_{p-1}, 1$ を $\mathbf{V}$ の固有値とし,

$$\mathcal{H}_p(\alpha) = 1 + \{\alpha^2 + (p-1)\alpha\mathcal{M}_{p-2}(\alpha)/\mathcal{M}_{p-1}(\alpha)\}$$

とする. このとき, 以下のことが成り立つ.

1. ESAG 分布 (10) は, $\mathbf{x} = \boldsymbol{\mu}/\|\boldsymbol{\mu}\|$ で常に停留点 (stationary point) を持つ.
2. $\rho_{p-1} \leq \mathcal{H}_p(\|\boldsymbol{\mu}\|)$ のとき, $\mathbf{x} = \boldsymbol{\mu}/\|\boldsymbol{\mu}\|$ での停留点は局所最大となる. また $\rho_{p-1} > \mathcal{H}_p(\|\boldsymbol{\mu}\|)$ のとき, $\mathbf{x} = \boldsymbol{\mu}/\|\boldsymbol{\mu}\|$ での停留点は局所最小となる.

大域的な単峰性についてはわからないが, Paine et al. (2018) は $\rho_{p-1} \leq \mathcal{H}_p(\|\boldsymbol{\mu}\|)$ の条件の下で, ESAG 分布 (10) が数値的に単峰であることを確かめている.

4 R パッケージ

2025 年 3 月の時点で, Kent 分布および ESAG 分布からの乱数の生成, 最尤推定を実行できる R のパッケージは, 筆者の知る限り Directional のみである. ここではこのパッケージを用いて, Kent 分布 (5) から大きさ $n = 30, 60, 100$ の乱数をそれぞれ生成し, それに基づいて最尤推定を求め, その誤差を計算するというシミュレーションを $N = 1000$ 回繰り返した. 最尤推定は, 関数 `kent.mle` により行う. この関数において, 正規化定数は

Kume and Wood (2005) の鞍点法を用いて近似を行っている.

ここで, Kent 分布の $\boldsymbol{\mu}$ は方向を表す単位ベクトルであるため, Paine et al. (2018) と同様に, 以下の誤差を用いて評価を行う:

$$\text{Error}(\hat{\boldsymbol{\mu}}) = \sqrt{2 \left\{ 1 - \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\mu}}_{(i)}^\top \boldsymbol{\mu} \right\}}, \quad (11)$$

ただし, $\hat{\boldsymbol{\mu}}_{(i)}$ は i 番目のシミュレーションにより得られた $\boldsymbol{\mu}$ に対して推定されたベクトルとする. それ以外のパラメーターに対しては, 通常のバイアスおよび平均二乗誤差の平方根 (RMSE) で評価を行う. 表 1 は, Kent 分布を用いたときの推定結果である. 標本のサイズ n を大きくするにつれ, 誤差 (11) および RMSE が小さくなることが確認できる.

n	Error($\hat{\boldsymbol{\mu}}$)	Bias($\hat{\kappa}$)	RMSE($\hat{\kappa}$)	n	Error($\hat{\boldsymbol{\mu}}$)	Bias($\hat{\kappa}$)	RMSE($\hat{\kappa}$)
30	0.092	1.409	2.828	30	0.049	4.827	8.825
60	0.064	0.597	1.609	60	0.034	2.363	5.181
100	0.050	0.395	1.154	100	0.026	1.275	3.500

表 1 左表: $\boldsymbol{\mu} = (0, 1, 0)^\top$, $\kappa = 10$, $\beta = 2$, 右表: $\boldsymbol{\mu} = (0, 1, 0)^\top$, $\kappa = 30$, $\beta = 2$

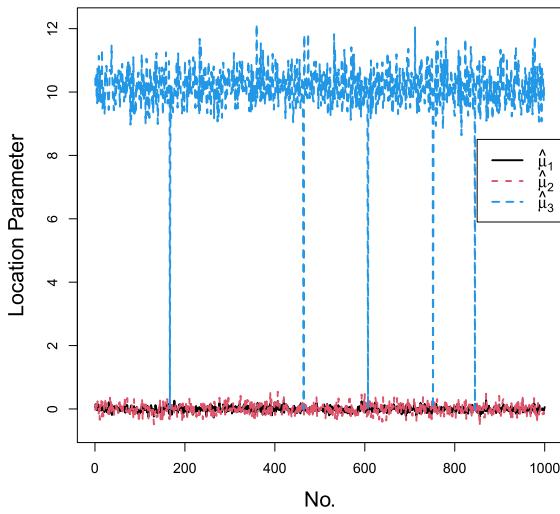
次に ESAG 分布 (10) から大きさ $n = 30, 60, 100$ の乱数をそれぞれ生成し, 関数 `esag.mle` を用いて最尤推定を求め, その誤差を計算するというシミュレーションを 1000 回繰り返した. ただし, ESAG 分布においては, $\boldsymbol{\mu}$ および $\hat{\boldsymbol{\mu}}$ は大きさ 1 のベクトルではないため, 誤差 (11) を用いるときには, $\hat{\boldsymbol{\mu}}_{(i)}$ を $\hat{\boldsymbol{\mu}}_{(i)} / \|\hat{\boldsymbol{\mu}}_{(i)}\|$, $\boldsymbol{\mu}$ を $\boldsymbol{\mu} / \|\boldsymbol{\mu}\|$ に置き換えたものを使用した. また $\hat{k} = \|\hat{\boldsymbol{\mu}}\|$ とする.

n	Error($\hat{\boldsymbol{\mu}}$)	Bias($\hat{k}$)	RMSE($\hat{k}$)	n	Error($\hat{\boldsymbol{\mu}}$)	Bias($\hat{k}$)	RMSE($\hat{k}$)
30	0.079	0.198	0.472	30	0.033	0.500	1.124
60	0.055	0.080	0.296	60	0.024	0.246	0.752
100	0.044	0.050	0.226	100	0.101	0.132	0.907

表 2 左表: $\boldsymbol{\mu}^* = (0, 0, 1)^\top$, $k = 4$, $\boldsymbol{\mu} = k\boldsymbol{\mu}^*$, $\boldsymbol{\gamma} = (1, -1)^\top$, 右表: $\boldsymbol{\mu}^* = (0, 0, 1)^\top$, $k = 10$, $\boldsymbol{\mu} = k\boldsymbol{\mu}^*$, $\boldsymbol{\gamma} = (1, -1)^\top$

上記の結果から分かるように, $k = 10$ の場合, n を大きくしても, 真の値とは大きく異なる推定をしてしまう場合があることがわかる. この原因を探るために, 図 8 をプロットした. これは表 2 の右側と同じパラメーター設定の下で, 横軸はシミュレーションの番号,

縦軸には、各シミュレーションごとの推定量 $\hat{\mu}$ の値をプロットしたものである。いくつかの場所で、 $\hat{\mu}_3$ の値が、真値とはかなり異なり、かなり小さな値を取っていることがわかる。右の表では、 $k = \|\mu\|$ と、標本のサイズ n を変更し、それぞれの設定の下でシミュレーションを $N = 1000$ 回繰り返した時に、図 8 のような異常な*3振る舞いをした推定量の割合を表している。 $k = \|\mu\|$ を大きくしたときに、このような問題の起こる割合が高くなることが予想されるが、現時点で明確な理由はわかっていない。



$n \backslash k$	4	10	20	30
30	0	0	0.007	0
60	0	0	0	0.007
100	0	0.002	0.002	0.004

$\|\hat{\mu}\| < k/2$ となる割合. ただし
 $\gamma = (1, -1)^\top$ とした.

図 8 $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3)^\top$ のプロット

5 まとめ

本報告においては、楕円状の等高線を持つ分布として Kent 分布と ESAG 分布を紹介した。分布が単峰になる条件については、Kent 分布は明確に与えられるのに対して、ESAG 分布は数値的に確認したのにとどまっている。一方で、正規化定数については、Kent 分布は何らかの近似が必要なのに対して、ESAG 分布は標準正規分布の分布関数というよく知られた関数の計算すれば求められる。乱数生成に関しては、Kent 分布は棄却法が必要なのに対して、ESAG 分布は非常に簡単な手続きで導出することができる。R パッケージの `Directional` を用いることで、Kent 分布および ESAG 分布に基づく最尤推定を行うこ

*3 ここでは $\|\hat{\mu}\| < k/2$ ならば異常であると判断したが、 $k/2$ という値は複数のグラフから判断して決めたものであり、理論的な妥当性はない。

とができる。しかし、集中度の高い球面上のデータに対しては, `esag.mle` に基づく ESAG 分布の最尤推定は, 不安定になることがある。

謝辞

本研究は, 研究費補助金基盤研究 (C) (19K11863) の助成を受けている。

参考文献

- Banerjee, A., Dhillon, I., Ghosh, J., and Sra, S. (2003). Generative model-based clustering of directional data. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 19–28.
- Banerjee, A., Dhillon, I. S., Ghosh, J., Sra, S., and Ridgeway, G. (2005). Clustering on the unit hypersphere using von mises-fisher distributions. *Journal of Machine Learning Research*, 6(9).
- Bingham, C. (1964). *Distributions on the sphere and on the projective plane*, PhD thesis. Yale University.
- Chen, Y. and Tanaka, K. (2021). Maximum likelihood estimation of the fisher-bingham distribution via efficient calculation of its normalizing constant. *Statistics and Computing*, 31:1–12.
- Dowe, D. L., Allison, L., Dix, T. I., Hunter, L., Wallace, C. S., and Edgoose, T. (1996). Circular clustering of protein dihedral angles by minimum message length. In *Pacific symposium on biocomputing*, volume 96, pages 242–255. World Scientific.
- Fisher, R. A. (1953). Dispersion on a sphere. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 217(1130):295–305.
- Gopal, S. and Yang, Y. (2014). Von mises-fisher clustering models. In *International Conference on Machine Learning*, pages 154–162. PMLR.
- Hyvärinen, A. and Dayan, P. (2005). Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4).
- Kent, J. T. (1982). The fisher-bingham distribution on the sphere. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(1):71–80.
- Kent, J. T., Ganeiber, A. M., and Mardia, K. V. (2013). A new method to simulate the bingham and related distributions in directional data analysis with applications.

arXiv preprint arXiv:1310.8110.

- Kent, J. T., Ganeiber, A. M., and Mardia, K. V. (2018). A new unified approach for the simulation of a wide class of directional distributions. *Journal of Computational and Graphical Statistics*, 27(2):291–301.
- Kume, A. and Sei, T. (2018). On the exact maximum likelihood inference of fisher–bingham distributions using an adjusted holonomic gradient method. *Statistics and Computing*, 28:835–847.
- Kume, A. and Wood, A. T. (2005). Saddlepoint approximations for the bingham and fisher–bingham normalising constants. *Biometrika*, 92(2):465–476.
- Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics*. John Wiley & Sons, Ltd., Chichester.
- Nakayama, H., Nishiyama, K., Noro, M., Ohara, K., Sei, T., Takayama, N., and Takemura, A. (2011). Holonomic gradient descent and its application to the fisher–bingham integral. *Advances in Applied Mathematics*, 47(3):639–658.
- Paine, P. J., Preston, S. P., Tsagris, M., and Wood, A. T. (2018). An elliptically symmetric angular gaussian distribution. *Statistics and Computing*, 28:689–697.
- Peel, D., Whiten, W. J., and McLachlan, G. J. (2001). Fitting mixtures of kent distributions to aid in joint set identification. *Journal of the American Statistical Association*, 96(453):56–63.
- Yu, Z. and Huang, X. (2024). A new parameterization for elliptically symmetric angular gaussian distributions of arbitrary dimension. *Electronic Journal of Statistics*, 18(1):301 – 334.