

方向統計学における推測理論とホロノミック勾配法

東京大学・情報理工学系研究科 清 智也

Tomonari Sei

School of Information Science and Technology,
The University of Tokyo.

概要

球面や多様体上のデータを扱う統計学を方向統計学という。方向統計学で用いられる統計モデルにはしばしば正規化定数（高次元の積分）が現れる。本稿ではこのような正規化定数に対処するための方法を概観する。特にホロノミック勾配法を重点的に紹介する。

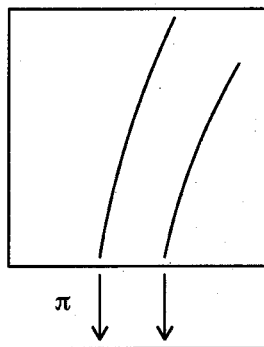
キーワード 積分計算, 方向統計学, ホロノミック勾配法.

1 はじめに

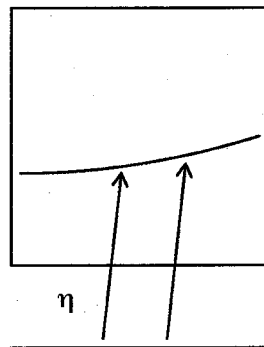
球面や多様体上のデータを扱う統計学を**方向統計学** (directional statistics) という。方向統計学で用いられる確率分布としては、次の2つが典型的である。

- 沈め込み (商空間) によって現れる周辺分布
- 埋め込み (部分空間) によって現れる条件付き分布

ここで、全空間上の確率分布としては多変量正規分布などのよく知られた分布を想定している。それぞれのイメージを図1に示す。



沈め込み (周辺分布)



埋め込み (条件付き分布)

図 1: 方向統計学で典型的に現れる確率分布.

まず球面上の分布の例をいくつか示す。詳細は [21] や [12] を参照されたい。

球面 $S^{p-1} = \{x \in \mathbb{R}^p \mid \|x\|^2 = \sum_{i=1}^p x_i^2 = 1\}$ を考える。球面はユークリッド空間に関する沈め込みとも埋め込みとも見なせる。すなわち、沈め込みは射影 $\pi(x) = x/\|x\|$ で与えられ、埋め込みは包含写像 $\eta(x) = x$ で与えられる。

例 1. 沈め込みの例として、射影正規分布

$$f(x) = \int_0^\infty \phi(rx|\mu, \Sigma) dr, \quad x \in S^{p-1},$$

がある。ただし

$$\phi(x|\mu, \Sigma) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}$$

は平均ベクトル μ 、共分散行列 Σ の多変量正規分布の密度関数を表す。射影正規分布に従う乱数は、多変量正規分布に従う乱数を規格化することによって容易に生成できる。特に $\mu = 0$ の場合、 $f(x)$ は**中心射影正規分布** (angular central Gaussian distribution) と呼ばれ、

$$f(x) = \frac{1}{|\Sigma|^{1/2} (x^\top \Sigma^{-1} x)^{p/2}}, \quad x \in S^{p-1},$$

と陽に表すことができる。

例 2. 埋め込みの例として **Fisher-Bingham 分布**

$$f(x) = \frac{\phi(x|\mu, \Sigma)}{\int_{S^{p-1}} \phi(\bar{x}|\mu, \Sigma) d\bar{x}}, \quad x \in S^{p-1},$$

がある。これは指数型分布族になっている。一般に全空間 (この例では \mathbb{R}^p) 上で指数型分布族であれば、それを制限して得られる分布族も指数型である。Fisher-Bingham 分布の特別な場合として、von Mises-Fisher 分布 ($\Sigma = I$)、Bingham 分布 ($\mu = 0$) がある。

円周 S^1 の場合、「巻き込み」と呼ばれる別の沈め込みがある。巻き込みとは

$$\pi(\theta) = (\cos \theta, \sin \theta), \quad \theta \in \mathbb{R},$$

によって与えられる写像 $\pi: \mathbb{R} \rightarrow S^1$ のことである。

例 3. $\theta \in [0, 2\pi)$ に対して

$$\begin{aligned} f(\theta) &= \sum_{k=-\infty}^{\infty} \frac{a}{\pi(a^2 + (\theta + 2k\pi)^2)} \\ &= \frac{1 - \rho^2}{2\pi(1 + \rho^2 - 2\rho \cos \theta)} \end{aligned}$$

は**巻き込みコーシー分布**と呼ばれる [21]. ただし $a > 0$, $\rho = e^{-a}$ はパラメータである. 2つ目の等号はフーリエ級数に関するポアソンの和公式

$$\sum_{k=-\infty}^{\infty} g(\theta + 2k\pi) = \sum_{l=-\infty}^{\infty} \hat{g}(l)e^{2\pi il\theta}, \quad \hat{g}(l) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(\theta)e^{-2\pi il\theta} d\theta,$$

を使って示される. 巻き込みコーシー分布は円周上の一様分布をメビウス変換したのになっており [13], 従って乱数生成が容易である.

次に, 形状空間 (shape space) を考える. 参考文献として [1] や [5] がある.

形状空間とは, ユークリッド空間上の点列 (landmarks と呼ばれる) に対し, 平行移動・回転・拡大縮小による違いを同一視することによって得られる空間である. より正確には次のように定義される. m 次元空間の, 番号の付いた k 点の座標を並べて得られる行列 $\mathbf{X} \in \mathbb{R}^{k \times m}$ を配置という. 2つの配置 \mathbf{X} と \mathbf{Y} が同値であるとは, ある $c > 0$, $v \in \mathbb{R}^m$, $Q \in SO(m)$ が存在して

$$\mathbf{Y} = c(1_k v^T + \mathbf{X})Q$$

が成り立つこととする. ただし $1_k = (1, \dots, 1)^T \in \mathbb{R}^k$ とした. この同値関係に関する商空間を**形状空間** (shape space) と呼び, Σ_m^k で表す. また各同値類 $[\mathbf{X}]$ を形状と呼ぶ.

図2は $m = 2$, $k = 7$ の場合の例である. なお $m = 2$ の場合, \mathbb{R}^2 と \mathbb{C} を同一視することにより, Σ_2^k は複素射影空間 $\mathbb{P}C^k$ と同一視できる. ただし退化した点を除く.

例 4. 配置 \mathbf{X} が多変量正規分布に従っていると仮定したときの $[\mathbf{X}]$ の分布によって, 形状空間 Σ_m^k 上の確率分布が定義される. これを**オフセット正規分布**という.

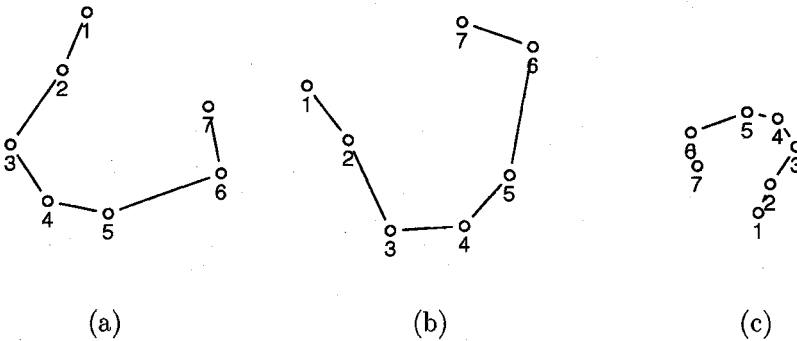


図2: 形状の例. (a)–(c) の点列は, 平行移動・回転・拡大縮小の変換によって同一視される.

2 積分計算の必要性

方向統計学で用いられる分布には、しばしば解析的に求まらない積分が含まれる。典型的な例が、例2の Fisher-Bingham 分布である。

積分を含む確率分布のパラメータを推定したいとき、いくつかの対処法がある。

1. 工夫して積分計算する。(級数法, 鞍点法, モンテカルロ法, ホロノミック勾配法など.)
2. 積分計算を避け, 別の推定法を用いる。(スコアマッチング推定, ベイズ推定 + MCMC など.)
3. 積分計算を避け, 別のモデルを用いる。(変数変換モデルなど.)

それぞれ説明するため, 共通の設定を確認しておく。 \mathcal{X} を多様体とする。主に $\mathcal{X} = S^{p-1}$ である。方向データは $x_1, \dots, x_n \in \mathcal{X}$ と表される。 dx を \mathcal{X} の基準測度とする。確率密度関数の集合 $\{f(x|\theta) \mid \theta \in \Theta\}$ を統計モデルという。ここで $\theta \in \Theta \subset \mathbb{R}^d$ をパラメータという。 $x_1, \dots, x_n \stackrel{iid}{\sim} f(x|\theta_0)$ と仮定し, θ_0 は未知とする。 θ_0 の推定量は $\hat{\theta}: \mathcal{X}^n \rightarrow \Theta$ と表される。本稿ではパラメータ推定のみ考え, 推定リスクは漸近分散で測るものとする: $\lim_{n \rightarrow \infty} nE[\|\hat{\theta} - \theta_0\|^2]$.

2.1 積分計算する場合

最尤法は対数尤度 $\sum_{i=1}^n \log f(x_i|\theta)$ を最大化することによってパラメータを推定する方法である。よく知られているように最尤推定量の漸近分散は漸近不偏推定量の中で最小である。しかし例えば指数型分布族

$$f(x|\theta) = \frac{e^{\theta^\top t(x)}}{Z(\theta)}, \quad Z(\theta) = \int_{\mathcal{X}} e^{\theta^\top t(x)} dx$$

を考えると, 正規化定数 $Z(\theta)$ の計算が厄介な場合がある。そのため様々な近似法が提案されてきた。

2.1.1 べき級数法

指数型分布族の正規化定数 $Z(\theta)$ は

$$Z(\theta) = \sum_{k \in \mathbb{N}_0^d} \frac{\theta^k}{k!} \int_{\mathcal{X}} t(x)^k dx$$

とテイラー展開される。ただし $\theta^k = \prod_{i=1}^d \theta_i^{k_i}$ などの多重添字記法を用いている。もしモーメント $\int_{\mathcal{X}} t(x)^k dx$ が解析的に計算できる場合、十分小さい θ に対しては無限級数を有限和で近似することにより、高速な近似計算が可能となる。

Fisher-Bingham 分布に対するべき級数表示は [9], [6] で与えられている。特に後者では打ち切り誤差も評価されている。

2.1.2 鞍点法

正規化定数 $Z(\theta)$ が周辺分布となっている場合、すなわち条件付き密度

$$f(x|u, \theta) = \frac{f(x, u|\theta)}{f(u|\theta)}, \quad x \in \mathcal{X}, \quad u \in \mathbb{R},$$

の分母が $Z(\theta) = f(u|\theta)$ となっている場合がある。もし $f(u|\theta)$ のモーメント母関数 $M(t|\theta) = \int f(u|\theta) e^{tu} du$ が解析的に得られる場合には、反転公式

$$Z(\theta) = f(u|\theta) = \frac{1}{2\pi i} \int M(t|\theta) e^{-tu} dt$$

を鞍点近似することによって、 $Z(\theta)$ の近似値が得られる。

Fisher-Bingham 分布に対して具体的に計算しているのが文献 [11] である。

2.1.3 モンテカルロ法

モンテカルロ法では、提案分布 $b(x)dx$ に従う乱数 ξ_1, \dots, ξ_N を生成し、

$$Z(\theta) = \int_{\mathcal{X}} \frac{e^{\theta^T t(x)}}{b(x)} b(x) dx \simeq \frac{1}{N} \sum_{i=1}^N \frac{e^{\theta^T t(\xi_i)}}{b(\xi_i)}$$

と近似する（重点サンプリング）。提案分布の選び方は工夫を要する。

文献 [2] では、提案分布を例 1 の中心射影正規分布として、Bingham 分布の重点サンプリングを行っている。なお、いまの目的とは異なるが、Fisher-Bingham 分布からの乱数生成法として Dirichlet 分布の混合表現による方法が提案されている [9]。

2.1.4 ホロノミック勾配法

ホロノミック勾配法は 3 節で詳しく述べる。

2.2 積分計算を避け、別の推定法を用いる方法

2.2.1 スコアマッチング法

ここでは簡単のため $\mathcal{X} = S^1 \simeq [0, 2\pi)$ (円周) に限定する. $x_1, \dots, x_n \in [0, 2\pi)$ とする. スコアマッチング推定量 [4] は

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^n \left\{ \partial_x^2 \log f(x_i|\theta) + \frac{1}{2} (\partial_x \log f(x_i|\theta))^2 \right\}$$

を最小化する θ として定義される. 特に指数型分布 $f(x|\theta) = e^{\theta^\top t(x)}/Z(\theta)$ の場合, $\partial_x \log f(x|\theta) = \theta^\top t'(x)$ となり, $S_n(\theta)$ は $Z(\theta)$ を含まない式となる. しかも θ の二次関数であるから陽に最適解が得られる.

スコアマッチング推定量が一致性を持つことは次の関係式から分かる: 真のパラメータが θ_0 のとき, 大数の法則と部分積分から

$$\begin{aligned} S_n(\theta) &\rightarrow \int_0^{2\pi} f(x|\theta_0) \left\{ \partial_x^2 \log f(x|\theta) + \frac{1}{2} (\partial_x \log f(x|\theta))^2 \right\} dx \\ &= \frac{1}{2} \int_0^{2\pi} f(x|\theta_0) \{ \partial_x \log f(x|\theta_0) - \partial_x \log f(x|\theta) \}^2 dx + (\text{const.}) \end{aligned}$$

となり, これは $\theta = \theta_0$ のとき最小となる.

スコアマッチング推定量の漸近分散は一般に最適ではない. また, 推定は可能であっても, 得られる分布 $f(x|\hat{\theta})$ は別途計算する必要があることにも注意する. より一般的な枠組みは local scoring rules として [16] で与えられている.

2.2.2 ベイズ法 + MCMC

一般に, パラメータに事前分布 $\pi(\theta)$ を仮定した場合の (二乗損失に対する) ベイズ推定量は

$$\hat{\theta} = \int \theta \pi(\theta|x_1, \dots, x_n) d\theta$$

で与えられる. 再び指数型分布族 $f(x|\theta) = e^{\theta^\top t(x)}/Z(\theta)$ を考える. この場合, マルコフ連鎖モンテカルロ法 (MCMC) を用いるにしても, そのままでは積分が残る. 実際, Metropolis アルゴリズムを用いる場合, θ から $\tilde{\theta}$ へ遷移するときの採択率は

$$r = \min \left(1, \frac{f(x|\tilde{\theta})\pi(\tilde{\theta})}{f(x|\theta)\pi(\theta)} \right) = \min \left(1, \frac{e^{\tilde{\theta}^\top t(x)} Z(\theta) \pi(\tilde{\theta})}{e^{\theta^\top t(x)} Z(\tilde{\theta}) \pi(\theta)} \right)$$

となり, $Z(\theta)$ が残ってしまう. exchange algorithm という方法を用いると, 採択率から $Z(\theta)$ を消すことが可能であることが [14], [2] で報告されている.

2.2.3 EM アルゴリズム

周辺分布 $f(x|\theta) = \int f(x, u|\theta) du$ に含まれるパラメータ θ を最尤推定したいとする。EM アルゴリズムでは、適当な初期値 θ から出発して

$$\theta \leftarrow \operatorname{argmax}_{\tilde{\theta}} \sum_{i=1}^n E[\log f(x_i, u|\tilde{\theta}) | x_i, \theta]$$

を繰り返し計算する。この方法は、右辺の条件付き期待値が容易に求まるときに有用である。形状空間のオフセット正規分布の場合は [10] で考察されている。

2.3 積分計算を避け、別のモデルを用いる方法

積分計算を避ける方法として、モデルを変えるというアプローチもあり得る。他に計算可能な手段がない場合には利用する価値はあるだろう。また、乱数生成が容易な場合にはモンテカルロ法の提案分布として活用できる。

積分計算が不要な典型例は変数変換である。すなわち、簡単な確率分布を変数変換すれば、密度関数の変換公式により、必要な計算はヤコビアンのみとなる。例 3 の巻き込みコーシー分布はその一例である。

球面の場合、最適輸送写像を用いて複雑な分布を作ることもできる [17]。

3 ホロノミック勾配法

3.1 ホロノミック関数

$\theta \in \mathbb{C}^d$ の多項式全体を $\mathbb{C}[\theta]$ と表す。解析関数 $h(\theta)$ が**ホロノミック関数** (holonomic function) であるとは、次の形の d 個の線形偏微分方程式が成り立つことである：

$$\sum_{j=0}^{r_i} a_{ij}(\theta) \partial_i^j h(\theta) = 0, \quad a_{ij}(\theta) \in \mathbb{C}[\theta], \quad i = 1, \dots, d. \quad (1)$$

ただし $\partial_i = \partial/\partial\theta_i$ であり、 $a_{ir_i}(\theta) \neq 0$ とする。式 (1) は、 $\theta_1, \dots, \theta_d$ のいずれの方向についても線形常微分方程式が成り立つことを意味する。

例 5. $(\theta, x) \in \mathbb{R}^2$ の関数 $h(\theta, x) = (1/2) \exp(-\theta x/2 - \theta/(2x))$ はホロノミックである。実際、次の線形偏微分方程式が満たされる：

$$\partial_\theta h = \left(-\frac{x}{2} - \frac{1}{2x}\right) h, \quad \partial_x h = \left(-\frac{\theta}{2} + \frac{\theta}{2x^2}\right) h. \quad (2)$$

ただし $\partial_\theta = \partial/\partial\theta$, $\partial_x = \partial/\partial x$ と略記した。

次の事実が強力である。

定理 1 (Zeilberger 1990). $h(\theta_1, \dots, \theta_d)$ がホロノミックならば, 積分

$$H(\theta_1, \dots, \theta_{d-1}) = \int_{\Gamma} h(\theta_1, \dots, \theta_d) d\theta_d$$

もホロノミックである. ここで $\Gamma \subset \mathbb{C}$ は適当な条件を満たすパスとする.

さらに, D -加群のアルゴリズムによって, この積分が満たす偏微分方程式を導くことが知られている (例えば [22]). その詳細には立ち入らない. 応用上, このアルゴリズムではなく人間が工夫して導出することも少なくない.

例 6. 例 5 の $h(\theta, x)$ に対し, 積分 $Z(\theta) = \int_0^\infty h(\theta, x) dx$ を考える. これは GIG 分布 [24] に関連しており, 第 3 種変形ベッセル関数で表される. 式 (2) を組み合わせると,

$$\theta^2 \partial_\theta^2 h + \theta \partial_\theta h - (\theta^2 + 1)h = \partial_x \left(-\frac{\theta}{2} x^2 h - xh + \frac{\theta}{2} h \right)$$

を示すことができる. この両辺を $x \in (0, \infty)$ で積分すると

$$\theta^2 \partial_\theta^2 Z + \theta \partial_\theta Z - (\theta^2 + 1)Z = 0$$

を得る. よって Z はホロノミック関数である.

3.2 Pfaffian system

関数 $Z(\theta)$ がホロノミックであるとき, ある有限個の導関数からなるベクトル

$$g(\theta) = (\partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d} Z)_{\alpha \in A}, \quad \emptyset \neq A \subset \mathbb{N}_0^d, \quad |A| < \infty, \quad 0 \in A,$$

が存在して, 次の形の偏微分方程式系を満たす:

$$\partial_i g(\theta) = P_i(\theta) g(\theta), \quad i = 1, \dots, d. \quad (3)$$

ここで $P_i(\theta)$ は有理関数からなる行列であり, 次の積分可能条件を満たすものとする:

$$\partial_i P_j(\theta) + P_j(\theta) P_i(\theta) = \partial_j P_i(\theta) + P_i(\theta) P_j(\theta). \quad (4)$$

偏微分方程式系 (3) を **Pfaffian system** という. また $g(\theta)$ を基底関数と呼ぶことにする. ベクトル $g(\theta)$ の次元 $|A|$ は偏微分方程式系 (1) から一意に定まり, **ホロノミックランク** と呼ばれる. 式 (1) から式 (3) の導出は代数的に行うことができる.

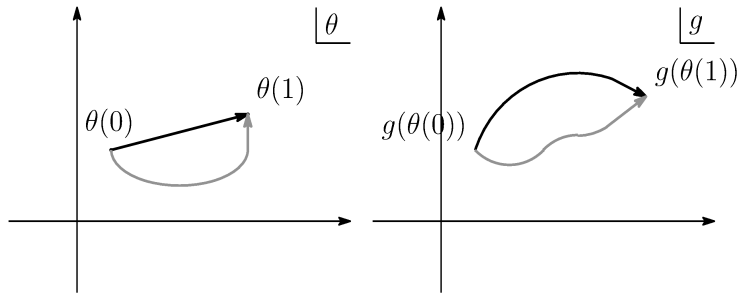


図 3: Pfaffian system の解はパスに依存しない。

3.3 ホロノミック勾配法

ホロノミック勾配法 (holonomic gradient method; **HGM**) は, $Z(\theta)$ の計算を助けるための手法である. Fisher-Bingham 分布の最尤推定を行う目的で, [15] により提案された.

HGM は以下の 3 ステップからなる. 与えられた点 $\theta(1) \in \Theta$ における $Z(\theta)$ の値を計算したいと仮定する.

1. 正規化定数 $Z(\theta)$ に関する Pfaffian system を導出する.
2. 適当な初期点 $\theta(0) \in \Theta$ で, 基底関数 $g(\theta(0))$ の値を (何らかの方法で) 計算する.
3. Pfaffian system を $\theta(0)$ から $\theta(1)$ にいたるパスに沿って数値的に解く.

ステップ 1 とステップ 2 は $\theta(1)$ が与えられていない段階でも実行できる. つまり「オフライン処理」が可能である.

ステップ 2 では計算しやすい $\theta(0)$ を選ばばよい. 例えばベキ級数を用いて初期値を求めるときには, 収束しやすい原点近傍を選ばばよい.

ステップ 3 において, 常微分方程式のソルバーとしてはルンゲクッタ法などを用いる. また積分可能条件 (4) により, $\theta(0)$ から $\theta(1)$ へいたるパスは任意に選んでよい (図 3). 実は積分可能条件が成り立たなくても, 初期値が正確である限り解はパスによらない. しかし初期値に誤差が含まれる場合にはパスの選び方によって結果が大きく変わる恐れがある.

3.4 ホロノミック勾配法の利点と欠点

ホロノミック勾配法の利点としては以下が挙げられる: (1) 精度を常微分方程式のソルバーによってコントロールできる. (2) ベキ級数のように, 原点からの距離に

よって計算時間が大きく変わるようなことは（あまり）ない。(3) 複数の θ に対する $g(\theta)$ を計算する場合でも、初期値 $g(\theta(0))$ の計算は1回で済む。しかも初期値計算はオフライン計算でよい。(4) いったん基底関数 $g(\theta)$ が求まれば、 $Z(\theta)$ の任意階の導関数は有理数演算だけで実行できる。特に、漸近分散の逆行列であるフィッシャー情報量の計算も容易となる。

欠点としては以下が挙げられる：(1) 初期値計算は問題ごとに工夫する必要がある。(2) ホロノミックランクはパラメータ空間の次元とともに指数的に増大することが多い。(3) 微分方程式に特異点がある場合は何らかの対処が必要である。

4 おわりに

ホロノミック勾配法について、これまで得られている結果については [25] にまとめられているので参照されたい。代表的なものを列挙すると、Fisher-Bingham 分布とその部分族 [15, 6, 18], 特殊直交群上の Fisher 分布 [19], Wishart 分布の最大固有値の分布関数 [3], 多変量正規分布の象限確率 [8] などがある。

ソフトウェアとしては R の HGM パッケージがある [7]。また文献 [22, 23] はこの分野の入門書として適切であろう。

謝辞

本研究は JSPS 科研費 26540013, 26108003 の助成を受けたものである。また、講究録の作成にあたり、狩野修平氏、田中冬彦氏から頂いた意見をそれぞれ参考にした。ここに感謝申し上げたい。

参考文献

- [1] Dryden, I. L., and Mardia, K. V. (1998). *Statistical Shape Analysis*. Wiley.
- [2] Fallaize, C.J. and Kypraios, T. (2014). Exact Bayesian inference for the Bingham distribution, *Stat. Comput.*, online first.
- [3] Hashiguchi, H., Numata, Y., Takayama, and N., Takemura, A. (2013). Holonomic gradient method for the distribution function of the largest root of a Wishart matrix, *J. Multivariate Anal.*, 117, (2013) 296–312.

- [4] Hyvärinen, A. (2005). Estimation of non-normalized statistical models by score matching, *J. Machine Learning Research*, **6**, 695–709.
- [5] Kendall, D.G., Barden, D., Carne, T.K., and Le, H. (1999). *Shape and Shape Theory*, Wiley.
- [6] Koyama, T., Nakayama, H., Nishiyama, K., and Takayama, N. (2014). Holonomic gradient descent for the Fisher-Bingham distribution on the d -dimensional sphere, *Computational Statistics*, **29** (3), 661–683.
- [7] Koyama, T., Nakayama, H., Ohara, K., Sei, T., and Takayama, N. (2014). Software Packages for Holonomic Gradient Method, *Mathematical Software — ICMS 2014, 4th International Conference, Proceedings*. Edited by Hoon Hong and Chee Yap, *Springer lecture notes in computer science* 8592, 706–712.
- [8] Koyama, T. and Takemura, A. (2015). Calculation of orthant probabilities by the holonomic gradient method, *Japan J. Indust. Appl. Math.*, **32** (1), 187–204.
- [9] Kume, A. and Walker, S. G. (2009). On the Fisher-Bingham distribution, *Stat. Comput.*, **19**, 167–172.
- [10] Kume, A. and Welling, M. (2010). Maximum likelihood estimation for the offset-normal shape distributions using EM, *J. Comput. Graphical Statist.*, **19** (3), 702–723.
- [11] Kume, A. and Wood, A. T. A. (2005). Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. *Biometrika*, **92**, 465–476.
- [12] Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Wiley.
- [13] McCullagh, P. (1996). Möbius transformation and Cauchy parameter estimation, *Ann. Statist.*, **24** (2), 787–808.
- [14] Murray, I., Ghahramani, Z., and MacKay, D.J.C. (2006). MCMC for doubly-intractable distributions, *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI2006)*, arxiv:1206.6848v1.
- [15] 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. *Adv. Appl. Math.*, **47**, 639–658.

- [16] Parry, M., Dawid, A.P., and Lauritzen, S. (2012). Proper local scoring rules, *Ann. Statist.* **40** (1), 561–592.
- [17] Sei, T. (2013). A Jacobian inequality for gradient maps on the sphere and its application to directional statistics, *Comm. Statist. Theory and Methods*, **42** (14), 2525–2542.
- [18] Sei, T. and Kume, A. (2015). Calculating the normalising constant of the Bingham distribution on the sphere using the holonomic gradient method, *Stat. Comput.*, **25**, 321–332.
- [19] Sei, T., Shibata, H., Takemura, A., Ohara, K., and Takayama, N. (2013). Properties and applications of Fisher distribution on the rotation group, *J. Multivariate Anal.*, 116 (2013), 440–455.
- [20] Zeilberger, D. (1990). A holonomic systems approach to special functions identities, *J. Comput. Appl. Math.*, **32**, 321–368.
- [21] 清水 邦夫 (2008). 方向統計学の最近の発展, 「計算機統計学」, **19** (2), 127–150.
- [22] JST CREST 日比チーム (2011). 「グレブナー道場」, 共立出版.
- [23] 竹村 彰通, 日比 孝之, 原 尚幸, 東谷 章弘, 清 智也, 『グレブナー道場』著者一同 (2015). 「グレブナー教室」, 共立出版.
- [24] 増田 弘毅 (2002). GIG 分布と GH 分布に関する解析, *統計数理*, **50** (2), 165–199.
- [25] ホロノミック勾配法関連の文献リスト : References for the Holonomic Gradient Method (HGM) and the Holonomic Gradient Descent Method (HGD), <http://www.math.kobe-u.ac.jp/OpenXM/Math/hgm/ref-hgm.html>