

円周分布の歪対称化と 関連する多様体上の分布への拡張について

高崎経済大学・経済学部* 宮田 庸一

Yoichi Miyata

Faculty of Economics, Takasaki City University of Economics

南山大学・理工学部† 塩濱 敬之

Takayuki Shiohama

Department of Data Science, Nanzan University

法政大学・経済学部‡ 阿部 俊弘

Toshihiro Abe

Faculty of Economics, Hosei University

概要

正弦関数摂動円周分布族は、歪対称な密度関数の形状を有し、かつパラメータのとり方では対称な分布を含むことが知られている。正弦関数摂動法の関数型を柔軟に歪対称化を可能とするような拡張を考えることで、既存のものよりも強い歪みを与えることができる円周上の統計モデルに関する研究成果を紹介する。また実データ解析によりそのパフォーマンスを示す。さらに、ここで紹介する手法が、シリンダー上、球面上、トーラス上への統計モデルに対しても拡張が可能であることを示す。

* 〒370-0801 群馬県高崎市上並榎町 1300

† 〒466-8673 愛知県名古屋市昭和区山里町 18

‡ 〒194-0298 東京都町田市相原町 4342

1 はじめに

風向きや生物の移動方向など角度で表されるデータは、生物学、地理学、医学、天文学など、様々な分野で現れる。このようなデータにおいては0度と360度は同一視されるため、実軸上の数値をとる標本と同様の統計解析を行うと、平均や分散、共分散や相関係数等の計算に不具合が生じることが知られている。このため、単位円周 S^1 に対応する角度を表す確率変数 Θ の確率密度関数 $f(\theta)$ を通じて、統計解析を行う必要が生じる。この場合、確率密度関数 $f(\theta)$ は以下の3つの条件を満たす必要がある:

- (i) $f(\theta) \geq 0$ a.e. on $(-\infty, \infty)$.
- (ii) $f(\theta + 2\pi) = f(\theta)$ a.e. on $(-\infty, \infty)$.
- (iii) $\int_k^{k+2\pi} f(\theta)d\theta = 1$ for $\forall k \in \mathbb{R}$.

これらの条件を満たす代表的な確率分布としては、 $f_{VM}(\theta; \mu, \kappa) = \exp\{\kappa \cos(\theta - \mu)\} / (2\pi I_0(\kappa))$ を確率密度関数に持つフォン・ミーゼス分布が知られている。

実際の角度データにおいては、明らかに歪対称なパターンを示すものが少なくない。このため、Umbach and Jammalamadaka (2009), Abe and Pewsey (2011), Ley and Verdebout (2017), Kato and Jones (2010) など、多くの研究者により歪対称円周モデル (Skew-symmetric circular models) が提案されている。円周上の対称な分布を歪対称化を可能とするような方法の1つとして、正弦関数摂動法が知られており、Abe and Pewsey (2011) では、正弦関数摂動 (sine-skewed) フォン・ミーゼス分布、正弦関数摂動巻き込みコーシー分布など様々な歪対称な分布の特徴が詳細に説明されている。

Ley et al. (2021) では、よい統計モデルの指標として、1. 汎用性 (Versatility), 2. 確率 (密度) 関数の扱いやすさ (Tractability), 3. モデルの解釈のしやすさ (Interpretability), 4. データ生成機構 (Data generating mechanism) の容易さ, 5. シンプルなパラメータ推定 (Straightforward parameter estimation) を挙げている。

1. の汎用性とは、様々な歪度、尖度などを特徴にもつ分布にフィットできる性質 (多用途性) を意味している。2. の扱いやすさとは、確率密度関数を陽に表すことができる、正規化定数が無限和を含まないなど、計算が複雑でないことを表している。モーメントを簡潔に表すことができる統計モデルにおいては、モーメント推定量を容易に得ることができ、最尤推定量をアルゴリズムで求める場合の初期値として利用できる (Ley et al., 2021, p.5)。さらに、モーメントを陽に表せる場合、モーメントを通じて表現される分布の特徴を容易に得ることができる。このため、陽に表すことができるモーメントを持つ統計モデルは、上

記の性質の 3. および 5. に該当するものとみなせる. 分布族における識別可能性は上記の 3. および 5. に該当する. もし分布族が識別可能でない場合は, 異なる 2 つのパラメータにより同じ確率モデルを表現できてしまうため, モデルの解釈を行うことが困難になる. また, そのようなモデルにおいては, 尤度関数が複数のピークを持つことから, 最尤推定を求めるための最適化が容易でなくなるため, 5. のシンプルなパラメータ推定に必要な条件を満たさなくなる可能性がある.

正弦関数摂動円周分布族は, 三角モーメントを陽に表すことができ, パラメータに対する識別可能性 (Miyata et al., 2022) を持ち, 乱数生成が容易であるなど, 多くの望ましい性質を有し, Ley et al. (2021) で与えられた 2-5 を満たしていると考えられる. その一方で, この分布族は, 歪対称化させることができる度合いがそれほど大きくないことが知られており, Ley et al. (2021) で議論されている 1. 汎用性については改善の余地がある. このため, 歪み関数 (skewing function) を拡張することにより, 正弦関数摂動円周分布の理論的な良さを保持しながら, なおかつ強い歪対称性を持つことができる分布を紹介するのが本稿での目的となる. なお, 今回の解説は, Miyata et al. (2023a,b), Abe et al. (2022), および Ley and Verdebout (2017) に基づいたものである.

2 正弦関数摂動円周分布の拡張

はじめに, Ley and Verdebout (2017) により提案された超球面 \mathbb{S}^p ($p \geq 2$) 上の歪対称な統計モデルについて説明する. ここでは $p = 1$ とする. 関数 ψ を $\psi(\mu) = (\cos \mu, \sin \mu)^T$ のように定め, $c_\rho h_\rho(\psi(\theta)^T \psi(\mu))$ を円周上の対称な確率密度関数^{*1}とし, ρ をこの密度関数のベクトル値, もしくはスカラー値の形状パラメータとする. また c_ρ は正規化定数とし, $h_\rho : [-1, 1] \rightarrow \mathbb{R}^+$ は, ベクトル値のパラメータに依存することもできる連続関数とする. Ley and Verdebout (2017) は, 以下の形の歪対称分布を提案した:

$$f_{\text{SRS}}(\theta; \mu, \rho, \lambda) = 2c_\rho h_\rho(\psi(\theta)^T \psi(\mu)) G \left\{ \lambda \psi(\mu)^T \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \psi(\theta) \right\}, \quad (1)$$

ただし $G : \mathbb{R} \rightarrow [0, 1]$ は連続な単調増加関数とし, 任意の $x \in \mathbb{R}$ に対して $G(-x) + G(x) = 1$ を満たすものとする. 関数 $G(x)$ はこの条件を満たすものであれば, どのようなものであってもよい. また $\lambda = 0$ となるとき, (1) 式は, 対称な確率密度関数 $c_\rho h_\rho(\psi(\theta)^T \psi(\mu))$ となる. 以後, この密度関数のことを基本密度 (base density) と呼ぶ. ここでは, 関数

^{*1} $\psi(\theta)^T \psi(\mu) = \cos(\theta - \mu)$ となることより μ は位置パラメータであることがわかる.

$G(x)$ として、以下のものを考える。まず最初に、区間 $[-1, 1]$ 上の確率密度関数

$$g_m(x) = \frac{\Gamma(2(m+1))}{2^{2m+1}\Gamma(m+1)^2}(1-x^2)^m \quad (-1 \leq x \leq 1)$$

を考える。ただし $m \geq 0$ は事前に定めておく定数で、ここでは“次数”と呼ぶことにする。そして、その分布関数を $G_m(x) = \int_{-1}^x g_m(t)dt$ と置く。Miyata et al. (2023a) では、この $G_m(x)$ を用いて、確率変数 Θ に対する以下の新たな確率密度関数を提案している：

$$f_{\text{ESS}}^{(m)}(\theta; \mu, \rho, \lambda) = 2f_0(\theta - \mu; \rho)G_m(\lambda \sin(\theta - \mu)), \quad (2)$$

ただし $f_0(\theta; \rho)$ は、 $\theta = 0$ に関して対称な基本密度とする。実際には、フォン・ミーゼス分布、巻き込みコーシー分布のように三角モーメントを陽に表すことができる分布を採用する。 μ は位置パラメータに対応しており、 λ は歪度パラメータ、 ρ は集中度等の分布の形状を制御するパラメータとなる。ここで提案された確率密度関数 (2) を持つ確率分布のことを、次数 m の Extended sine-skewed (ESS) 円周分布と呼ぶことにする。次数 m に $0, 1, 2, 3, 4$ を代入して得られる歪み関数は、以下の形で表すことができる

$$\begin{aligned} G_0(x) &= \frac{x+1}{2}, & G_1(x) &= \frac{1}{4}(-x^3 + 3x + 2), \\ G_2(x) &= \frac{1}{16}(3x^5 - 10x^3 + 15x + 8), \\ G_3(x) &= \frac{1}{32}(-5x^7 + 21x^5 - 35x^3 + 35x + 16), \\ G_4(x) &= \frac{1}{256}(35x^9 - 180x^7 + 378x^5 - 420x^3 + 315x + 128). \end{aligned}$$

$m = 0$ の場合、正弦関数摂動円周分布 (Abe and Pewsey, 2011) になる。ESS 円周分布 (2) は任意の非負の $m \geq 0$ に対して定義することができるが、 m が非負の整数の場合、簡潔に表すことができる。実際、 $C_m = \frac{\Gamma\{2(m+1)\}}{2^{2m+1}\Gamma(m+1)^2}$ とおくと、 $m = 0, 1, 2, \dots$ に対して、

$$G_m(x) = \int_{-1}^x g_m(t)dt = C_m \sum_{\ell=0}^m \binom{m}{\ell} \frac{1}{2\ell+1} (-1)^\ell (x^{2\ell+1} + 1) \quad (3)$$

と表すことができる。図 1 では、3つの異なる次数に対する ESS 円周密度の曲線をプロットした。ここでの基本密度はいずれも $\mu = \pi/2$, $\kappa = 8$ とするフォン・ミーゼス分布とする。左上の図は $m = 0$, 右上の図は $m = 1$, 左下の図は $m = 5$ であることから、次数 m を大きくすると、分布の歪対称の度合いが大きくなることがわかる。

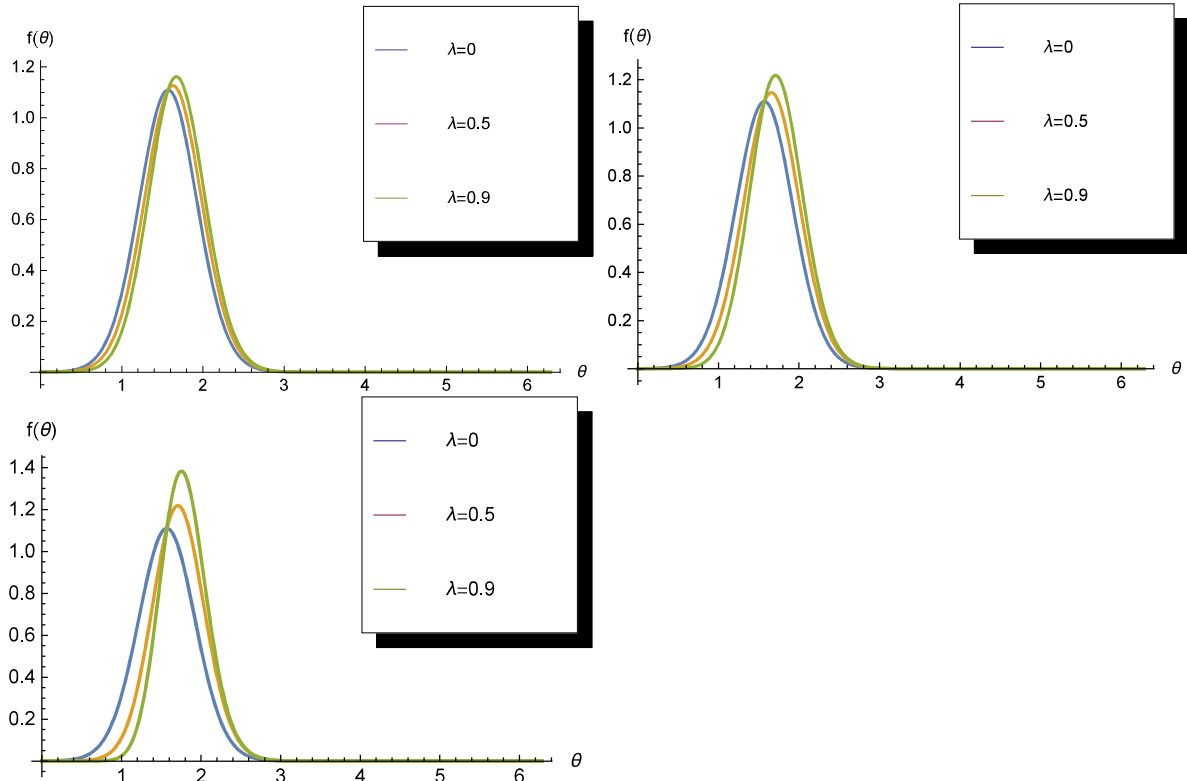


図1 次数 $m = 0, 1, 5$ に対する ESS フォン・ミーゼス分布の密度関数のプロット.

2.1 三角モーメント

ここでは統計モデル (2) が, 正弦関数摂動円周分布族と同様に理論的な良さを持つことを示す. $\alpha_{0,p} := E_{0,\mu=0} \{\cos(p\Theta)\}$ を, $\mu = 0$ とおいた基本密度 $f_0(\theta; \rho)$ における p 次の余弦モーメントとする. このとき, $f_{\text{ESS}}^{(m)}(\theta; \mu, \rho, \lambda)|_{\mu=0}$ における p 次の余弦モーメントは

$$\alpha_p := E_{\mu=0} \{\cos(p\Theta)\} = \alpha_{0,p}$$

となることが示せる. この結果を用いると, 一般の位置パラメータ μ を持つ $f_{\text{ESS}}^{(m)}(\theta; \mu, \rho, \lambda)$ における p 次の余弦モーメントと正弦モーメントは

$$\alpha_{p,\mu} = E \{\cos(p\Theta)\} = \cos(p\mu)\alpha_p - \sin(p\mu)\beta_p,$$

$$\beta_{p,\mu} = E \{\sin(p\Theta)\} = \cos(p\mu)\beta_p + \sin(p\mu)\alpha_p$$

となる. ただし,

$$\beta_p = 2C_m \sum_{\ell=0}^m \binom{m}{\ell} \frac{(-1)^\ell}{2\ell+1} \int_{-\pi}^{\pi} (\lambda \sin \theta)^{2\ell+1} \sin(p\theta) f_0(\theta; \rho) d\theta$$

となる.

2.2 乱数生成

ここでは, ESS 円周分布 (2) から, 以下の手順で容易に乱数を生成することを述べる.

1. 基本密度 $f_0(\theta; \boldsymbol{\rho})$ から乱数 Φ を生成する.
2. Φ とは独立に, 区間 $[0, 1]$ 上の一様分布 $U[0, 1]$ から乱数 U を生成する.
3. Θ_0 を以下のように定める:

$$\Theta_0 = \begin{cases} \Phi & \text{if } U < G_m(\lambda \sin \Phi) \\ -\Phi & \text{if } U \geq G_m(\lambda \sin \Phi) \end{cases}$$

4. $\Theta = (\Theta_0 + \mu + \pi)(\text{mod}2\pi) - \pi$ は $f_{\text{ESS}}^{(m)}(\theta; \mu, \boldsymbol{\rho}, \lambda)$ からの乱数となる.

2.3 識別可能性

ここでは, 基本密度をフォン・ミーゼス分布および巻き込みコーシー分布としたパラメトリック分布族を考える. まずは, 以下の密度関数を持つ次数 m の ESS-フォン・ミーゼス分布を考える.

$$f_{\text{ESSvM}}^{(m)}(\theta; \boldsymbol{\eta}) = \frac{1}{\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\} G_m(\lambda \sin(\theta - \mu)),$$

ただし, $\boldsymbol{\eta} = (\mu, \kappa, \lambda)^T$ とし, そのパラメータ空間を $\mathbf{H} = \{\boldsymbol{\eta} | \mu \in [-\pi, \pi), \kappa > 0, \lambda \in [-1, 1]\}$ とする. $\mathcal{F}_{\text{vM}}^{(m)} := \{f_{\text{ESSvM}}^{(m)}(\theta; \boldsymbol{\eta}); \boldsymbol{\eta} \in \mathbf{H}\}$ を ESS-フォン・ミーゼス分布の分布族とすると, 以下の結果が成り立つ.

命題 1 任意の整数 $m \geq 0$ に対して, 分布族 $\mathcal{F}_{\text{vM}}^{(m)}$ は識別可能.

証明は, Miyata et al. (2023a) で与えられているため, ここでは省略する.

次に, 次数 m の ESS-巻き込みコーシー分布を考える.

$$f_{\text{ESSWC}}^{(m)}(\theta; \boldsymbol{\eta}) = \frac{1 - \rho^2}{\pi(1 + \rho^2 - 2\rho \cos(\theta - \mu))} G_m(\lambda \sin(\theta - \mu)),$$

ただし $\boldsymbol{\eta} = (\mu, \rho, \lambda)^T$, $-\pi \leq \mu < \pi$, $\rho > 0$, $-1 \leq \lambda \leq 1$ とする. $\mathbf{H} = \{\boldsymbol{\eta} | \mu \in [-\pi, \pi), \rho \in (0, 1), \lambda \in [-1, 1]\}$ をパラメータ空間とし, $\mathcal{F}_{\text{WC}}^{(m)} := \{f_{\text{ESSWC}}^{(m)}(\theta; \boldsymbol{\eta}); \boldsymbol{\eta} \in \mathbf{H}\}$ をその分布族とする. このとき, 以下の結果が成り立つ.

命題 2 任意の整数 $m \geq 0$ に対して, 分布族 $\mathcal{F}_{\text{WC}}^{(m)}$ は識別可能.

証明は, Miyata et al. (2023a) で与えられているため, ここでは省略する.

3 データ解析

角度データに対する提案分布の当てはめを考える. パラメータベクトル $\boldsymbol{\eta}$ のパラメータ空間を $\mathbf{H} = M \times \mathbf{R} \times \Lambda$ と表す. ここで, $\delta_\lambda > 0$ とし, $M \subseteq [-\pi, \pi)$, \mathbf{R} , $\Lambda := [-1 + \delta_\lambda, 1 - \delta_\lambda]$ は, μ , $\boldsymbol{\rho}$, λ のパラメータ空間である.

今, $\boldsymbol{\theta}_n = (\theta_1, \dots, \theta_n)$ を次数 m の ESS 円周分布 $f_{\text{ESS}}^{(m)}(\boldsymbol{\theta}; \mu, \boldsymbol{\rho}, \lambda)$ からの無作為標本とする. このとき対数尤度関数は

$$\ell_m(\boldsymbol{\eta}) = \frac{1}{n} \sum_{i=1}^n \{\log(f_0(\theta_i - \mu; \boldsymbol{\rho})) + \log\{2G_m(\lambda \sin(\theta_i - \mu))\}\}$$

と表される. また, $\boldsymbol{\eta}$ の最尤推定量は

$$\hat{\boldsymbol{\eta}}_n = \underset{\boldsymbol{\eta}}{\operatorname{argmax}} \ell_m(\boldsymbol{\eta})$$

と定義する. このとき, 正則条件の下で最尤推定量の一致性および漸近正規性が成り立つ.

一般に, 次数 m は未知の超パラメータであるため, 情報量規準を用いた m の推定を考える. 上述の最尤推定量は予め指定した m に対するパラメータ推定方法であるため, 各 $m \in \{0, 1, 2, 3, 4\}$ に対して, パラメータ推定を行い, 情報量規準を計算する. 異なる m に対して, 分布のパラメータ数は不変であるため, 赤池情報量規準 (AIC) を比較することは最大対数尤度を比較することと同等である. そこで, 竹内情報量規準 (TIC) によるモデル選択を考える. 今, F と F' をデータ生成モデルと観測データの真の生成モデルとする. このとき AIC と TIC は次のように定義される.

$$\text{AIC} = -2\ell_m(\hat{\boldsymbol{\eta}}_n) + 2p$$

また

$$\text{TIC} = -2\ell_m(\hat{\boldsymbol{\eta}}_n) + 2\operatorname{tr}\{\hat{\mathbf{J}}^{-1}(F')\hat{\mathbf{I}}(F')\},$$

である. ここで, $\hat{\mathbf{J}}(F')$ と $\hat{\mathbf{I}}(F')$ は, それぞれ

$$\mathbf{J}(F') = -E_{F'} \left[\frac{\partial^2 \log f_{\text{ESS}}^{(m)}(\boldsymbol{\Theta}; \boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} \right], \quad \mathbf{I}(F') = E_{F'} \left[\frac{\partial \log f_{\text{ESS}}^{(m)}(\boldsymbol{\Theta}; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \frac{\partial \log f_{\text{ESS}}^{(m)}(\boldsymbol{\Theta}; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}^T} \right]$$

の一致推定量である。ただし、期待値はデータ生成モデルの下で評価する。罰則項における $\text{tr}\{\hat{\mathbf{J}}^{-1}(F')\hat{\mathbf{I}}(F')\}$ は、一般にバイアスと呼ばれる。 $\mathbf{I}(F)$ はフィッシャー情報量行列であり、データ生成モデルと真のモデルが一致するという条件 $F' = F$ の下で、 $\mathbf{J}(F) = \mathbf{I}(F)$ となるからバイアスはモデルのパラメータ数と一致し、AIC と TIC は一致する。

ESS 円周分布のデータ分析における柔軟性 (flexibility) を示すために、この章では、2 つのデータセットを用いて異なる次数 m に対する推定結果を紹介する。最初のデータセットは、2014 年から 2022 年における台風発生日のデータで標本数は $n = 231$ である*²。データは台風発生日をその年の 7 月 1 日を 0 に基準化して、 $[-\pi, \pi)$ 上に変換した。

2 つ目のデータセットは、直近 10 年間の全国の落雷発生月データを母集団分布と仮定し、その分布からランダムに標本数 $n = 100$ の離散円周データを抽出したデータを用いた*³。このデータも 7 月を 0 に規準化し、 $[-\pi, \pi)$ 上に値を取るよう変換した。表 1 に想定した月別落雷数の母集団分布を示した。

表 1 月別落雷数の母集団分布

月	1	2	3	4	5	6	7	8	9	10	11	12
確率	0.008	0.007	0.029	0.051	0.080	0.132	0.212	0.241	0.160	0.036	0.032	0.012

図 2 に使用したデータセット 1 および 2 のヒストグラムを示した。台風発生日のデータの標本平均方向 $\bar{\mu}$ は 1.125 ラジアンで、これはおおよそ、台風発生日の確率分布の平均が 9 月 5 日頃に相当する。標本合成ベクトル長 \bar{r} は 0.550 であり、これより円周分散を計算すると $1 - \bar{r} = 0.450$ となる。また、円周歪度を求めると -0.142 となり、分布は反時計回りの方向に歪んでいることを示している。これは、夏から秋にかけて台風が集中することを表している。

同様に、落雷発生月の円周分布の平均方向を求めると、 $\bar{\mu} = 0.205$ であり、これは 7 月上旬に相当すること、また標本合成ベクトル長 \bar{r} は 0.596 であり、これより円周分散を計算すると $1 - \bar{r} = 0.404$ となる。集中度の程度は台風発生日の分布と同様にあまり大きくない。また、円周歪度を求めると 0.315 となり、分布は時計回りの方向に歪んでいることを示している。

基本統計量の分析から、これらの 2 つのデータセットに対して、歪対称分布の当てはめが妥当であることがわかる。そこで、ESS-フォン・ミーゼス分布と ESS-巻き込みコーシー

*² データは国立情報学研究所、デジタル台風: 台風画像と台風情報のページから入手した。 <http://agora.ex.nii.ac.jp/digital-typhoon/>

*³ データは、株式会社フランクリン・ジャパン: 雷統計データのウェブページから入手した。 <https://www.franklinjapan.jp/raiburari/data/1381/>

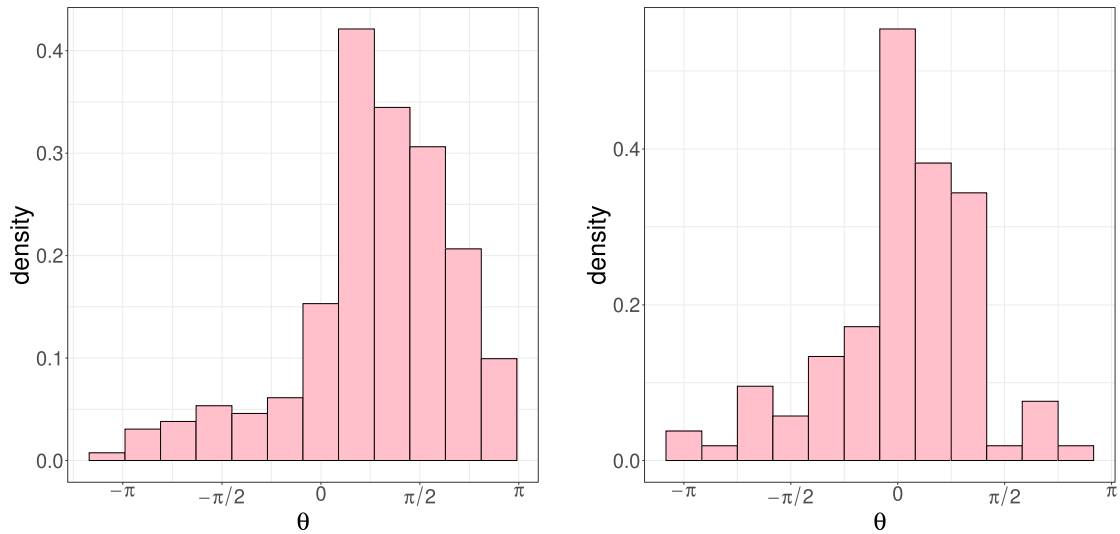


図2 台風発生日のヒストグラム (左) と月別落雷件数の分布からサイズ $n = 100$ の無作為抽出したデータのヒストグラム (右)

分布の最尤法によるパラメータ推定を試みた. 表2と表3にパラメータの推定値および, AIC と TIC の値をまとめた. この表より以下のことが確認できる.

第1に, 歪対称性のパラメータ λ の推定値の符号を調べることで, 分布の歪対称性の方向が分かる. すなわち, 台風発生日のデータに対しては λ の推定値は負の値を示しているし, 落雷発生月のデータに対しては負の値を示す. これらの λ の符号の正負は, 円周データの標本歪度の符号とは整合しない.

第2に, どちらのデータセットに対しても, m を大きくすると, 歪度パラメータ λ の推定値が小さな値を取ることが確認できる. これは, ESS 円周分布モデルにおいて, 適切な m を用意することで, 小さな λ の値を用いることで, 同程度の歪対称性を表すことができるを示している. λ のパラメータ空間が $[-1, 1]$ に制限されることを考えると, m を適切に設定することで, どのような歪対称性を表すデータに対しても分布の当てはめが可能になることを意味する.

第3に, どちらのデータセットに対しても AIC, TIC は m を変えてもほとんど変化しないことから, モデルの複雑性を考慮すると, これらのデータに対しては $m = 0$ および $m = 1$ モデルが十分であることを意味する. 一方, データセット1に対しては, ESS-フォン・ミーゼス分布の方が ESS-巻き込みコーシー分布よりも当てはまりの程度が良い. データの分布の特徴に応じて, 基本密度を適切に選択することの重要性を示している. データセット2に対しては, 基本密度の選択による当てはまりの良さの違いは明確に判断できない.

表2 データセット1のパラメータ推定結果

m	ESS-フォン・ミーゼス分布					ESS-巻き込みコーシー分布				
	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\lambda}$	AIC	TIC	$\hat{\mu}$	$\hat{\rho}$	$\hat{\lambda}$	AIC	TIC
0	0.3168	1.1883	0.6542	615.03	614.24	0.3086	0.4682	0.6684	619.21	619.16
1	0.3090	1.1744	0.4667	615.29	614.75	0.3035	0.4657	0.4745	619.45	619.46
2	0.3063	1.1694	0.3815	615.36	614.96	0.3021	0.4649	0.3865	619.52	619.56
3	0.3049	1.1669	0.3305	615.39	615.06	0.3014	0.4646	0.3342	619.55	619.61
4	0.3040	1.1654	0.2956	615.41	615.13	0.3010	0.4644	0.2985	619.57	619.65

表3 データセット2のパラメータ推定結果

m	ESS-フォン・ミーゼス分布					ESS-巻き込みコーシー分布				
	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\lambda}$	AIC	TIC	$\hat{\mu}$	$\hat{\rho}$	$\hat{\lambda}$	AIC	TIC
0	0.3517	1.2364	-0.7082	292.06	291.35	0.2485	0.5349	-0.4898	290.84	294.75
1	0.3767	1.1879	-0.5517	291.60	291.02	0.2594	0.5311	-0.3588	290.81	295.15
2	0.3833	1.1747	-0.4620	291.47	290.92	0.2627	0.5299	-0.2953	290.80	295.20
3	0.3862	1.1688	-0.4045	291.41	290.87	0.2643	0.5293	-0.2565	290.80	295.22
4	0.3879	1.1654	-0.3640	291.37	290.85	0.2652	0.5289	-0.2298	290.80	295.25

4 他の多様体上の確率モデルへの応用

(3) で提案された歪み関数は、他の多様体上の統計モデルへの拡張が可能である。

4.1 シリンダー上の統計モデルへの応用

シリンダー $[-\pi, \pi) \times \mathbb{R}_+$ 上の値をとる確率変数 (Θ, X) に対する統計モデルの1つとして、以下のものが知られている、

$$f_{AL}(\theta, x) = \frac{\alpha\beta^\alpha}{\pi \cosh(\kappa)} (1 + \lambda \sin(\theta - \mu)) x^{\alpha-1} \times \exp\{-(\beta x)^\alpha (1 - \tanh(\kappa) \cos(\theta - \mu))\}, \quad (4)$$

ただし $(x, \theta) \in [0, \infty) \times [-\pi, \pi)$, $\cosh(\kappa) = \{\exp(\kappa) + \exp(-\kappa)\}/2$ とし, $\tanh(\kappa) = \{\exp(\kappa) - \exp(-\kappa)\}/\{\exp(\kappa) + \exp(-\kappa)\}$, $\alpha > 0$, $\beta > 0$, $-\pi \leq \mu < \pi$, $\kappa > 0$, $-1 \leq \lambda \leq 1$ はパラメータとする。このとき、 Θ の周辺分布は正弦関数摂動巻き込みコーシー分布 (Abe and Pewsey, 2011) になる。このモデルは Abe and Ley (2017) により提案されたものである。この場合、(4) における $(1 + \lambda \sin(\theta - \mu))$ を $G_m(\lambda \sin(\theta - \mu))$ に

置き換えることで以下の密度関数を持つモデルへ拡張することができる:

$$f_{\text{WeiESSVM}}^{(m)}(\theta, x; \boldsymbol{\eta}) = \frac{\alpha\beta^\alpha}{\pi \cosh(\kappa)} G_m(\lambda \sin(\theta - \mu)) x^{\alpha-1} \\ \times \exp\{-(\beta x)^\alpha (1 - \tanh(\kappa) \cos(\theta - \mu))\}.$$

モーメント, 識別可能性, 乱数生成の詳細については, Miyata et al. (2023b) を参照のこと.

4.2 超球面 \mathbb{S}^{k-1} ($k \geq 2$) 上の統計モデルへの応用

これまで, 円周上の統計モデルを用いて説明を行ってきたが, これは超球面 \mathbb{S}^{k-1} ($k \geq 2$) へ拡張することができる. ここでは, 確率ベクトル $\mathbf{X} \in \mathbb{R}^k$ は \mathbb{S}^{k-1} 上のベクトルとする^{*4}. $f_k(\mathbf{x}; \boldsymbol{\mu}) := c_{f,k} f(\mathbf{x}^T \boldsymbol{\mu})$ を $\boldsymbol{\mu} \in \mathbb{S}^{k-1} \subseteq \mathbb{R}^k$ に関して, 回転対称 (rotationally symmetric) な \mathbf{X} の確率密度関数とする. Paindaveine and Verdebout (2015) において, 回転対称の概念は以下のように定義されている.

定義 3 確率ベクトル $\mathbf{X} \in \mathbb{S}^{k-1}$ はベクトル $\boldsymbol{\mu} \in \mathbb{S}^{k-1}$ に対して回転対称 $\stackrel{\text{def}}{\iff} \mathbf{O}\boldsymbol{\mu} = \boldsymbol{\mu}$ を満たす任意の直交行列 \mathbf{O} に対して, $\mathbf{O}\mathbf{X}$ と \mathbf{X} は同じ確率分布となる.

回転対称な分布は, \mathbf{x} と $\boldsymbol{\mu}$ の内積の単調非減少関数として表すことができる. また $f(\cdot)$ を $f: [-1, 1] \rightarrow \mathbb{R}_+$ である絶対連続な非負値関数とし, $c_{f,k}$ は正規化定数とする.

行列 $\boldsymbol{\Upsilon}_\boldsymbol{\mu}$ は,

$$\boldsymbol{\Upsilon}_\boldsymbol{\mu} \boldsymbol{\Upsilon}_\boldsymbol{\mu}^T = \mathbf{I}_k - \boldsymbol{\mu}\boldsymbol{\mu}^T, \quad \text{かつ} \quad \boldsymbol{\Upsilon}_\boldsymbol{\mu}^T \boldsymbol{\Upsilon}_\boldsymbol{\mu} = \mathbf{I}_{k-1}$$

を満たす $k \times (k-1)$ 実行列とする. ただし \mathbf{I}_k は $k \times k$ の単位行列とする. このとき, 歪み関数 G_m を用いた歪対称な確率密度関数は以下の形で与えられる:

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\lambda}) = 2c_{f,k} f(\mathbf{x}^T \boldsymbol{\mu}) G_m(\boldsymbol{\lambda}^T \boldsymbol{\Upsilon}_\boldsymbol{\mu}^T \mathbf{x}),$$

ただし $\boldsymbol{\lambda} \in \mathbb{R}^{k-1}$ が, 分布の歪対称性を制御するパラメータベクトルとなる. この分布の導出に関しては Ley and Verdebout (2017) を参照されたい.

4.3 トーラス ($\mathbb{S} \times \mathbb{S}$) 上の統計モデルへの応用

円周上の 2 つの確率変数の同時分布をトーラス上の確率分布という. ここでは, Abe et al. (2022) で紹介されたトーラス上の確率分布の構成手法にならい, 関数 $G_m(\cdot)$ を用い

^{*4} すなわち, $\|\mathbf{X}\| = 1$.

たトーラス上の同時密度関数を構成する. $f_1(\theta_1)$ と $f_2(\theta_2)$ を 2 つの円周上の対称な基本密度関数とする. このとき, 歪み関数 $G_m(\cdot)$ を用いたトーラス上の同時確率密度関数は次式で与えられる.

$$f_{\text{ESS}}^{(m)}(\theta_1, \theta_2) = 2G_m(\lambda \sin \theta_1 \sin \theta_2) f_1(\theta_1) f_2(\theta_2), \quad (5)$$

ここでパラメータ $\lambda \in [-1, 1]$ は, 2 つの確率変数 Θ_1 と Θ_2 の相関の程度を表す. この同時分布は, 周辺分布指定型の同時確率密度の構成方法になっていることから, $f_{\text{ESS}}^{(m)}(\theta_1, \theta_2)$ の周辺分布は $f_1(\theta_1)$ と $f_2(\theta_2)$ である. 図 3 には, 相関パラメータは, $\lambda = 0.8$ とし, $m = 0$ と $m = 4$ について, 周辺分布に共通の巻き込みコーシー分布を指定した場合と, 共通のフォン・ミーゼス分布を指定した場合のトーラス上の同時確率密度関数の等高線図をプロットした. この図から, m を大きくすることで正の関連性の程度が大きくなることが確認できる.

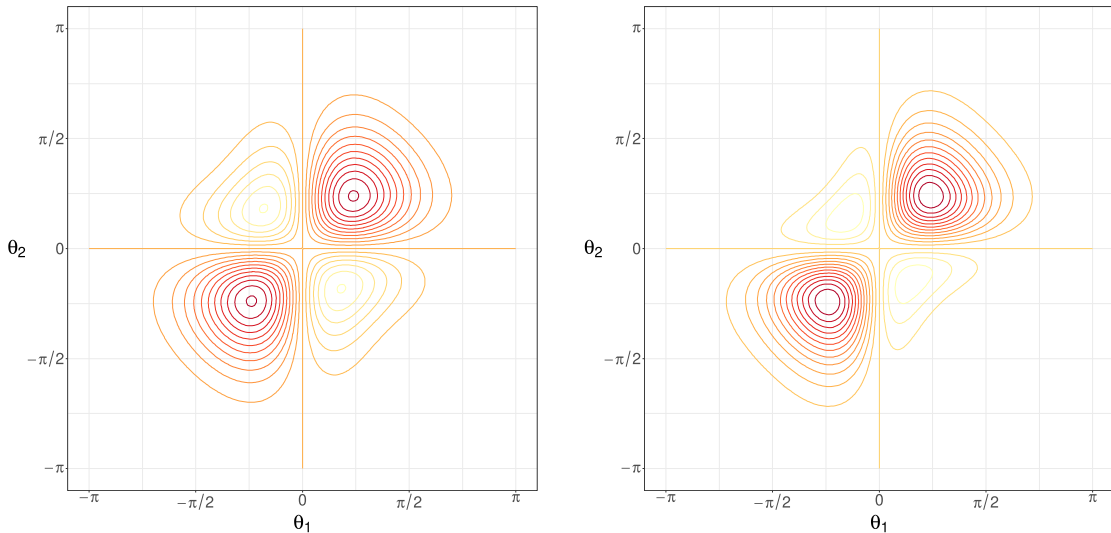


図 3 周辺分布をフォン・ミーゼス分布に指定した場合の $m = 0$ (左) と $m = 4$ (右) の拡張したトーラス分布の等高線プロット. パラメータは, $\mu_1 = \mu_2 = 0, \kappa_1 = \kappa_2 = 2, \lambda = 0.8$ とした.

Abe et al. (2022) や Imoto and Abe (2021) では, $m = 0$ の場合のトーラス上の同時分布を提案していたが, 提案した同時分布の問題点として, λ を 1 や -1 に近づけても角度変数の相関係数の絶対値が大きくなることを指摘していた. そこで, m の与え方に対して, 次の角度変数の相関係数の振る舞いを調べる.

$$\rho_{\text{JS}} = \frac{E[\sin \Theta_1 \sin \Theta_2]}{\sqrt{E[\sin^2 \Theta_1] E[\sin^2 \Theta_2]}}.$$

ただし、角度変数 Θ_1 と Θ_2 は平均方向がともに $\mu_1 = \mu_2 = 0$ としている。この相関係数の定義は、例えば、(清水, 2018, p.163) に定義が述べられている。図 4 には、 $f_1(\theta_1)$ と $f_2(\theta_2)$ を同一の巻き込みコーシー分布および、フォン・ミーゼス分布にした場合の角度データの相関係数をプロットした。ただし、巻き込みコーシー分布については、集中度のパラメータを $\rho = \rho_1 = \rho_2 = 0.8$ とし、フォン・ミーゼス分布については $\kappa = \kappa_1 = \kappa_2 = 2$ とした。この図から、ESS トーラス分布は、 m を大きく取ることによって既存手法よりもより関連性の強い 2 変数角度データの同時分布を記述できることが分かる。

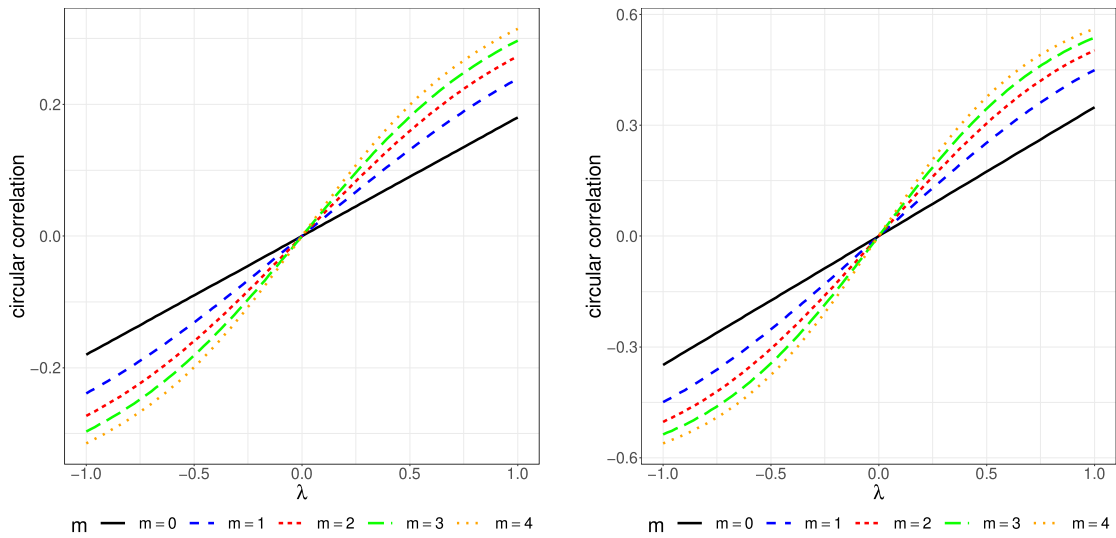


図 4 λ に対する角度変数の相関係数のプロット、周辺分布に巻き込みコーシー分布を指定した場合 (左) と、フォン・ミーゼス分布を指定した場合 (右)。

トーラス上の密度関数 (5) に従う乱数の生成は次の手順に従う。確率分布関数 $G_m(x)$ に従う確率変数 X を生成する。次に、周辺密度 $f(\theta_1)$ と $f(\theta_2)$ に従う確率変数 Θ'_1 と Θ'_2 を生成する。これを用いて確率変数の組 (Θ_1, Θ_2) を次の様に生成すると

$$(\Theta_1, \Theta_2) = \begin{cases} (\Theta'_1, \Theta'_2) & \text{if } X < \lambda \sin(\Theta'_1) \sin(\Theta'_2) \\ (-\Theta'_1, \Theta'_2) & \text{if } X \geq \lambda \sin(\Theta'_1) \sin(\Theta'_2) \end{cases}$$

(Θ_1, Θ_2) は密度関数 (5) に従う。これは $(\Theta_1, \Theta_2) = (\theta_1, \theta_2)$ における密度関数は、以下の様に表すことができるからである

$$G_m(\lambda \sin \theta_1 \sin \theta_2) f_1(\theta_1) f_2(\theta_2) + (1 - G_m(\lambda \sin(-\theta_1) \sin \theta_2)) f_1(-\theta_1) f_2(\theta_2).$$

5 まとめ

ここでは, Miyata et al. (2023a) により与えられた歪み関数を用いた円周分布を中心に特徴, および理論的な性質を紹介した. この歪み関数 $G_m(x)$ は, m が非負の整数の場合には多項式で表されるため, 容易に確率密度関数を評価することができる. 特に, ESS 円周分布は, 通常の変動円周分布よりも位置パラメータの周辺で強く歪ませることができるため, Ley et al. (2021) で述べられている汎用性 (Versatility) に関しても改善されたといえる. また乱数生成の容易さ, 識別可能性などの性質も保持しているため, 1章で述べた, よい統計モデルの指標 2-5 も維持しているといえる. その一方で, 基本密度を巻き込みコーシー分布とした正弦関数変動円周分布は単峰型であることが Abe and Pewsey (2011) により示されているが, 基本密度を巻き込みコーシー分布とした ESS 円周分布においては, 単峰型であるかどうかはわかっていない. また次数 1 の ESS 円周分布と次数 2 の円周分布は識別可能であるかについても明らかになっていない. このため, 提案手法の有用さを示すデータセットに加えて, 理論的な性質を明らかにすることも重要な課題となる.

謝辞

本研究集会にて講演の機会を与えていただきました日本大学商学部の小池健一先生に厚くお礼申し上げます. また本研究は, 科学研究費 (課題番号:19K11869, 19KK0287, 19K11863, 22K11944), 2023 年度南山大学パッへ研究奨励金の助成を受けている.

参考文献

- Abe, T., Imoto, T., Shiohama, T., and Miyata, Y. (2022). On some flexible models for circular, toroidal, and cylindrical data. In *Directional Statistics for Innovative Applications: A Bicentennial Tribute to Florence Nightingale*, pages 229–243. Springer.
- Abe, T. and Ley, C. (2017). A tractable, parsimonious and flexible model for cylindrical data, with applications. *Econom. Stat.*, 4:91–104.
- Abe, T. and Pewsey, A. (2011). Sine-skewed circular distributions. *Statist. Papers*, 52(3):683–707.

- Imoto, T. and Abe, T. (2021). Simple construction of a toroidal distribution from independent circular distributions. *J. Multivariate Anal.*, 186:104799.
- Kato, S. and Jones, M. C. (2010). A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *J. Amer. Statist. Assoc.*, 105(489):249–262.
- Ley, C., Babić, S., and Craens, D. (2021). Flexible models for complex data with applications. *Annu. Rev. Stat. Appl.*, 8:369–391.
- Ley, C. and Verdebout, T. (2017). Skew-rotationally-symmetric distributions and related efficient inferential procedures. *J. Multivariate Anal.*, 159:67–81.
- Miyata, Y., Shiohama, T., and Abe, T. (2022). Identifiability of asymmetric circular and cylindrical distributions. *Sankhya A*, <https://doi.org/10.1007/s13171-022-00294-3>.
- Miyata, Y., Shiohama, T., and Abe, T. (2023a). An extension of sine-skewed circular distributions. *preprint*.
- Miyata, Y., Shiohama, T., and Abe, T. (2023b). On a flexible cylindrical model with a skewing function. *preprint*.
- Paindaveine, D. and Verdebout, T. (2015). Optimal rank-based tests for the location parameter of a rotationally symmetric distribution on the hypersphere. In *Mathematical Statistics and Limit Theorems: Festschrift in Honour of Paul Deheuvels*, pages 249–269. Springer.
- Umbach, D. and Jammalamadaka, S. R. (2009). Building asymmetry into circular distributions. *Statist. Probab. Lett.*, 79(5):659–663.
- 清水邦夫 (2018). 角度データのモデリング (ISM シリーズ:進化する統計数理). 近代科学社.