

環境科学における方向統計学の利用

清水 邦夫^{1,2}・王 敏真³

(受付 2013年2月25日; 改訂 6月13日; 採択 6月14日)

要 旨

風向は環境科学における典型的で重要な角度観測値であり、また風向と共に風速やオゾン濃度などの別の観測値を伴う場合がある。本稿では、角度を含む観測値をモデル化・解析する学問である方向統計学における最新の結果を含んで総合的に報告を行うことを目的としている。

最初に、角度データに対する基本統計量としての標本平均方向および標本平均ベクトル長の計算法およびそれらの量の持つ意味と性質が線形な観測値の標本平均および標本分散と比較して述べられる。また、国立環境研究所が提供する霞ヶ浦における風向データを用いてデータのプロットと基本統計量の計算の例示を行う。Pewseyの分布対称性検定法の説明の後に、JonesとPewseyによる柔軟な対称角度分布および正弦関数摂動法による対称分布の非対称化が紹介され、風向データを用いて当てはめ例が示される。さらに、対称/非対称性とは異なる性質として、分布の頂上付近で平坦/急傾斜の性質を表すPapakonstantinou分布とBatschelet分布が紹介される。説明変数に角度を含む場合の線形回帰がシリンダー上の分布から自然に導かれることを述べると共に、角度変数間の回帰モデルおよび構造モデルに関する最近の結果も紹介する。

キーワード：角度回帰モデル，角度構造モデル，角度分布の当てはめ，シリンダー上の分布，トラス上の分布。

1. はじめに

大気中のオゾンや粒子状物質が風によって運ばれて特定の地域に影響を与えることがあるなどのように、気象要素の一つの風向が環境科学において欠くことのできない情報を与えることがあるのは周知の通りである。風向は典型的な角度の観測値であるので、角度変量をどのようにモデルに取り入れるかを研究することは環境科学研究の一つとして意義があると思われる。本稿では、より一般に角度を含む観測値を統計的に扱う学問(方向統計学)について、最新の結果を含めて報告することを目的としている。なお、方向統計学について従来の研究をまとめた本としてはBatschelet(1981), Fisher(1993), Mardia and Jupp(2000), Jammalamadaka and SenGupta(2001), Zar(2010)などが知られている。

角度が線形変数とは著しく異なる性質を持つことは、単純に平均を取る操作においても現れる。例えば、方位を表すのに北を 0° とし、時計回りに一周(360°)することにより、北(0°)-東

¹ 慶應義塾大学 理工学部：〒223-8522 神奈川県横浜市港北区日吉 3-14-1

² 統計数理研究所 客員

³ 慶應義塾大学 理工学研究科：〒223-8522 神奈川県横浜市港北区日吉 3-14-1 (現 統計数理研究所：〒190-8562 東京都立川市緑町 10-3)

(90°)-南(180°)-西(270°)-北(360°を法として0°)と表すことにしよう. 北北東(22.5°)と北北西(337.5°)の「平均」は直観的には北であり, これを南 $((22.5 + 337.5)/2 = 180^\circ)$ とするのは不自然である. しかしながら, このような不合理性は, 方位を(単位)円周上のベクトルと解釈し, 「平均」を「ベクトルの平均」とすることにより解決される. こうして, 角度の統計学は(単位)円周上のベクトルの統計学とすることができる. 環境科学では, 1変量の角度観測値だけでなく, 風向・風速や風向・汚染物質濃度のように角度・線形変量の形式の2変量の観測値や2地点において同時に風向を観測するときのように2変量の角度観測値がよく現れる. したがって, 方向統計学は環境科学を研究する上で極めて有用な方法の一つと考えられる.

本稿の第2節では, 角度観測値が得られたときの基本統計量の計算の仕方と統計量の意味・性質について述べ, また, いくつかの角度分布の紹介と共にデータへの角度分布の当てはめについて紹介する. 第3節では角度変数を含む回帰モデルについて述べる. 第4節では, 回帰とは異なるモデル化として, 角度変数間の構造モデルについて現在知られている結果について紹介する. 第5節はまとめと課題である.

2. 基本統計量の計算と分布の当てはめ

得られた角度データを $\theta_1, \theta_2, \dots, \theta_n$ とする. 本節では, 角度データ θ_j ($j=1, 2, \dots, n$)からどのような統計量を算出するとデータに内在するどのような情報を抽出することができるのか, データの分布状況をモデルで表すにはどのような分布を当てはめればよいのか, について記述する.

2.1 基本統計量

数学においては, 平面上の横軸の正方向の角度を0(ラジアン)とし反時計回りに1周して 2π まで, 区間で表せば $[0, 2\pi)$ で角度(の主値)を表現するのが普通である. しかしながら, 諸科学技術においては角度を区間 $[-\pi, \pi)$ で表したり, 縦軸の正方向から時計回りに計測することもある. しかし表し方がたとえ異なっていたとしても方向としての実体は変わらないはずであるので, それぞれの表現の仕方に即して適切な解釈が可能となるように角度データの統計理論を組み立てることが必要である. そのために, 角度 θ を単位円周上の点 $P(\cos\theta, \sin\theta)$ と同一視(図1(a))し, 点 P を大きさ(長さ)が1のベクトル $(\cos\theta, \sin\theta)'$ と捉える. なお, 'は転置を表す記号である. 角度データ θ_j ($j=1, 2, \dots, n$)を点 $(\cos\theta_j, \sin\theta_j)$ もしくはベクトル $(\cos\theta_j, \sin\theta_j)'$ と同一視し,

$$C = \sum_{j=1}^n \cos\theta_j, \quad S = \sum_{j=1}^n \sin\theta_j$$

とおくとき, ベクトル $(C, S)'$ は n 個のベクトル $(\cos\theta_j, \sin\theta_j)'$ の合成ベクトルを表し, また, $(C/n, S/n)'$ は平均合成ベクトルを表す. 図1(b)は $n=2$ のとき, 2つの点もしくはベクトル P_1 と P_2 から平均合成ベクトル \bar{P} を得る様子を示している. 合成ベクトル $(C, S)'$ もしくは平均合成ベクトル $(C/n, S/n)'$ の方向 $\bar{\theta}$ は標本平均方向(sample mean direction)と呼ばれ, 合成ベクトル $(C, S)'$ の大きさ $R = \sqrt{C^2 + S^2}$ の値が0でないとき

$$\cos\bar{\theta} = \frac{C}{R}, \quad \sin\bar{\theta} = \frac{S}{R},$$

すなわち $\tan\bar{\theta} = S/C$ を満たす $\bar{\theta}$ によって定義される. なお, $\bar{\theta}$ は次のように陽に表すことができる:

$$\bar{\theta} = \begin{cases} \tan^{-1}(S/C), & C > 0, S \geq 0, \\ \pi/2, & C = 0, S > 0, \\ \tan^{-1}(S/C) + \pi, & C < 0, \\ 3\pi/2, & C = 0, S < 0, \\ \tan^{-1}(S/C) + 2\pi, & C > 0, S < 0. \end{cases}$$

ここで、 $\tan^{-1}(\bullet) \in (-\pi/2, \pi/2)$ である。 R の値が 0 のとき平均方向は定義されない。 平均合成ベクトル $(C/n, S/n)'$ の大きさ $\bar{R} = \sqrt{C^2 + S^2}/n$ は標本平均合成ベクトル長 (sample mean resultant length) と呼ばれ、 $0 \leq \bar{R} \leq 1$ を満たす。 \bar{R} の値が 1 となるのは、すべての θ_j が同じ角度を表すときでそのときに限る。 このことから $\nu = 1 - \bar{R}$ を分布の分散 (拡がりを表す量) として採用するのは自然である。 実軸上のデータのときと異なり、角度データの分散 ν は $0 \leq \nu \leq 1$ と上に有界である。 2つの角度はどんなに離れても原点に関して対称な点までしか離れようがないことから、角度データの分散が有界な量として定義されるのは妥当と言える。 なお、 \bar{R} は $\bar{R} = n^{-1} \sum_{j=1}^n \cos(\theta_j - \bar{\theta})$ と表すことができる。

平均方向は実軸上のデータの場合の算術平均に類似の性質を持つ。 実際、実軸上のデータ x_1, x_2, \dots, x_n の算術平均 $\bar{x} = \sum_{j=1}^n x_j/n$ に対しては偏差 $x_j - \bar{x}$ の和をとると $\sum_{j=1}^n (x_j - \bar{x}) = 0$ が成立し、一方、平均方向 $\bar{\theta}$ に対しては $\sum_{j=1}^n \sin(\theta_j - \bar{\theta}) = 0$ が成立する。 また、角度分散は実軸上のデータの分散 $S^2 = n^{-1} \sum_{j=1}^n (x_j - \bar{x})^2$ と類似の性質を持つことを説明しよう。 実数 c に対して $U(c) = n^{-1} \sum_{j=1}^n (x_j - c)^2$ を最小にするには $U(c) = n^{-1} \sum_{j=1}^n (x_j - \bar{x})^2 + (\bar{x} - c)^2$ であるから $c = \bar{x}$ とすればよいことが分る。 そのときの値 $U(\bar{x})$ は分散 S^2 となる。 一方、単位円周上の 2 点の角度を θ, α とすると、2 点間の弦の長さの 2 乗が $2\{1 - \cos(\theta - \alpha)\}$ と表されることから、 $V(\alpha) = 1 - n^{-1} \sum_{j=1}^n \cos(\theta_j - \alpha)$ とおくと、 $V(\alpha) = \nu + 2\bar{R} \sin^2\{(\bar{\theta} - \alpha)/2\}$ を得るので、 $V(\alpha)$ は $\alpha = \bar{\theta}$ のときに最小となる。 そのときの最小値 $V(\bar{\theta})$ は角度分散 ν となる。 なお、 $t_j = (\cos \theta_j, \sin \theta_j)'$, $j = 1, \dots, n$, とし、 $t = (C, S)'$ とすると、恒等式

$$2(1 - \bar{R}) = \frac{1}{n} \sum_{j=1}^n \left\| t_j - \frac{t}{\|t\|} \right\|^2$$

が成り立つ。 ただし、 $\|\bullet\|$ は $a = (a_1, a_2)'$ のときノルム $\|a\| = \sqrt{a_1^2 + a_2^2}$ を表す。 この恒等式の右辺は、大きさ 1 のベクトル t_j ($j = 1, 2, \dots, n$) と大きさ 1 のベクトル $t/\|t\|$ (標本平均方向を表す合成ベクトル t の基準化ベクトル) の差の 2 乗ノルムの平均であることを示している。

2.2 プロット

環境科学において典型的な角度データである風向データを用いて例示する。 データとして国立環境研究所が提供する霞ヶ浦臨湖実験施設気象データサイト <http://www.cger.nies.go.jp/db/kasumi/index12.html> に掲載されている 2012 年 4 月 1 日 ~ 5 月 31 日の毎日朝 6 時と正午 12 時に本館

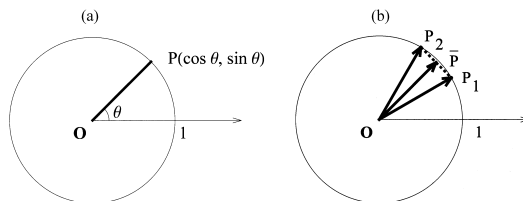


図 1. (a) 角度と単位円周上の点の対応, (b) $n = 2$ のときの平均合成ベクトル \bar{P} .

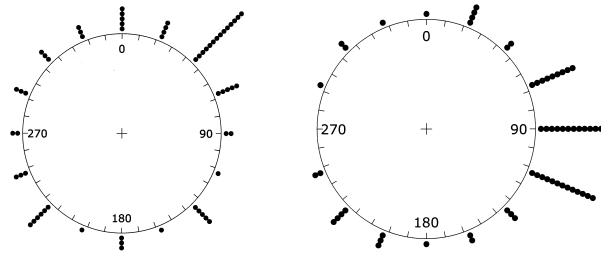


図2. 霞ヶ浦朝6時(左)と正午12時(右)の風向データ(2012年4月1日～5月31日)の円周プロット.

(地上)において観測された風向風速データ(時間平均データ)中の最多風向を用いる. 風向は16方位を用いて表されている. 図2は, 北の風を 0° とし時計回りに方位を配置(北-東-南-西-北)したときの円周プロットを表している. 標本平均方向 $\bar{\theta}$ の値は, 朝6時の風向データに対して 0.541 (ラジアン) $=31.03^\circ$ で, 正午12時に対して 1.702 (ラジアン) $=97.51^\circ$ である. また, 標本平均合成ベクトル長 \bar{R} の値は, 朝6時に対して 0.237 , 正午12時に対して 0.506 となる. したがって, 標本角度分散はそれぞれ 0.763 と 0.494 であり, 朝6時の方が分布の拡がり大きい. 言い換えれば, 正午12時の分布の方が一方向に集中する傾向がある.

2.3 対称分布の当てはめ

2.3.1 対称性の検定

Θ を角度確率変数とするとき, Θ の p 次三角モーメントは $E(e^{ip\Theta})$ ($i=\sqrt{-1}$)で定義され, 1次の三角モーメント $E(e^{i\Theta})=\rho e^{i\mu}$ ($\rho>0$)における μ は Θ の平均方向(mean direction), ρ は Θ の平均合成ベクトル長(mean resultant length)と呼ばれる. $\rho=0$ のとき平均方向は定義されない. μ の周りの2次正弦モーメント(second sine moment)は $\bar{\beta}_2=E\{\sin 2(\Theta-\mu)\}$ で定義され, 分布の歪みを量る歪度(skewness)は $\bar{\beta}_2$ と ρ を使って $\bar{\beta}_2/(1-\rho)^{3/2}$ で定義される. Pewsey(2002)は角度データ $\theta_1, \theta_2, \dots, \theta_n$ から「対称性」の仮説 $H_0: \bar{\beta}_2=0$ をノンパラメトリック検定する方法を以下のように与えている.

標本平均方向を $\bar{\theta}$ とし, $\bar{\theta}$ 周りの標本2次正弦モーメントを $\bar{b}_2=n^{-1}\sum_{j=1}^n \sin 2(\theta_j-\bar{\theta})$ とする. また, \bar{b}_2 の分散の推定値を代入法により

$$\hat{\text{var}}(\bar{b}_2)=\frac{1}{n}\left[\frac{1-\bar{a}_4}{2}-2\bar{a}_2+\frac{2\bar{a}_2}{\bar{R}}\left\{\bar{a}_3+\frac{\bar{a}_2(1-\bar{a}_2)}{\bar{R}}\right\}\right]$$

とする. ここで,

$$\bar{a}_p=\frac{1}{n}\sum_{j=1}^n \cos p(\theta_j-\bar{\theta}) \quad (p=2,3,4), \quad \bar{R}=\frac{1}{n}\sqrt{\left(\sum_{j=1}^n \cos \theta_j\right)^2+\left(\sum_{j=1}^n \sin \theta_j\right)^2}$$

を表す. 検定統計量を

$$Z=\frac{\bar{b}_2}{\sqrt{\hat{\text{var}}(\bar{b}_2)}}$$

とすると, Z は帰無仮説の下で近似的に標準正規分布に従うので, データから計算される Z の値 z に対し片側 P 値 $P(U>|z|)$ を計算する. U は標準正規分布に従う確率変数を表す. 対称性の仮説の採択/棄却を P 値によって判断する. 仮説 H_0 の有意水準 α の検定では, u_α を標準正規分布の上側 $100\alpha\%$ 点として $|z|>u_\alpha$ のとき H_0 を棄却する.

2.3.2 Jones と Pewsey の分布

データへの分布の当てはめが可能であるように「データのサイズは十分に大きい」ものとしておく。柔軟性を持つ対称分布の一つに Jones and Pewsey (2005) の分布があり、確率密度関数は次式で与えられる：

$$(2.1) \quad f_\psi(\theta) = \frac{\{\cosh(\kappa\psi) + \sinh(\kappa\psi) \cos(\theta - \mu)\}^{1/\psi}}{2\pi P_{1/\psi}(\cosh(\kappa\psi))}, \quad 0 \leq \theta < 2\pi.$$

ここで、 $\kappa \geq 0$, $0 \leq \mu < 2\pi$, $-\infty < \psi < \infty$ (極限の意味で $\psi = 0$ も含む) であり、 $P_{1/\psi}$ は Legendre の陪関数

$$\int_0^\pi (z + \sqrt{z^2 - 1} \cos x)^{1/\psi} dx = \pi P_{1/\psi}(z), \quad \psi > 0,$$

$$\int_0^\pi \frac{1}{(z + \sqrt{z^2 - 1} \cos x)^{-1/\psi}} dx = \pi P_{-1/\psi-1}(z) = \pi P_{1/\psi}(z), \quad \psi < 0$$

を表す。分布族 (2.1) は特別な場合として以下の分布を含むことが知られている。

- (1) 一様分布 ($\kappa = 0; \psi \rightarrow \pm\infty$, κ 有限) :

$$f(\theta) = \frac{1}{2\pi}.$$

- (2) von Mises 分布 $VM(\mu, \kappa)$ ($\psi \rightarrow 0$) :

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}, \quad I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos \theta} d\theta.$$

I_0 は 0 位の第 1 種変形 Bessel 関数を表す。

- (3) 巻き込み Cauchy 分布 $WC(\mu, \rho)$ ($\psi = -1, \rho = \tanh(\kappa/2)$) :

$$f(\theta) = \frac{1}{2\pi \cosh(\kappa) \{1 - \tanh(\kappa) \cos(\theta - \mu)\}} = \frac{1 - \rho^2}{2\pi \{1 + \rho^2 - 2\rho \cos(\theta - \mu)\}}.$$

- (4) ハート型分布 $C(\mu, \rho)$ ($\psi = 1, \rho = \tanh(\kappa/2)$) :

$$f(\theta) = \frac{1}{2\pi} \{1 + 2\rho \cos(\theta - \mu)\}.$$

- (5) Cartwright (1963) の分布 ($\psi = 1/m, \kappa \rightarrow \infty, m$ は負でない整数) :

$$f(\theta) = \frac{\Gamma(m+1)}{2^{m+1} \sqrt{\pi} \Gamma(m+1/2)} \{1 + \cos(\theta - \mu)\}^m.$$

- (6) 自由度 $d (> 0)$ の円周 t 分布 ($-1 < \psi < 0$) (Shimizu and Iida, 2002) :

$$f(\theta) = \frac{\{1 + \tanh(\kappa\psi) \cos(\theta - \mu)\}^{1/\psi}}{2\pi {}_2F_1(-1/(2\psi), -1/(2\psi) + 1/2; 1; \tanh^2(\kappa\psi))}, \quad \frac{1}{\psi} = -\frac{d}{2} - \frac{1}{2}.$$

${}_2F_1$ は Gauss の超幾何関数を表す。

図 3 に、 $\mu = \pi, \kappa = 1.5$ のとき、 $\psi = -1$ (巻き込み Cauchy 分布), 0 (von Mises 分布), 1 (ハート型分布) の値に対して Jones と Pewsey の分布の確率密度関数 (2.1) のプロットを与える。

データ $\theta_1, \theta_2, \dots, \theta_n$ から分布族 (2.1) のパラメータ μ, κ, ψ を最尤推定するには、例えば最適化の手法により、尤度

$$L(\mu, \kappa, \psi) = \prod_{j=1}^n f_\psi(\theta_j)$$

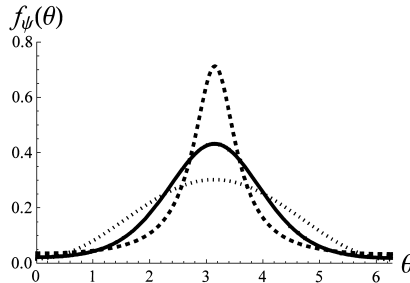


図3. Jones と Pewsey の分布の確率密度関数 (2.1) のプロット： $\mu = \pi$, $\kappa = 1.5$; $\psi = -1$
(巻き込み Cauchy：破線), 0 (von Mises：実線), 1 (ハート型：点線).

を μ, κ, ψ に関して最大化する. モデル選択のためには, 分布族 (2.1) をフルモデル, (1)~(6) を (2.1) のサブモデルとして AIC ($= -2 \times (\text{最大対数尤度}) + 2 \times (\text{自由パラメータ数})$) や BIC ($= -2 \times (\text{最大対数尤度}) + \log n \times (\text{自由パラメータ数})$) の値を最小とするモデルを採用すればよい. n は標本の大きさを表す. また, 選択されたモデルの適合の良さを調べるには, χ^2 適合度統計量の P 値を計算する.

Abe et al. (2010) は Jones and Pewsey (2005) の分布の拡張形を与え, 例として Jones and Pewsey (2005) で取り上げられた Mardia and Jupp (2000, Table 1) の渡り鳥の飛翔方向データ ($n = 714$) を用いて解析を行っている. しかし, 渡り鳥の飛翔方向データへの分布の適合に関しては僅かに改善が見られる程度であった.

2.3.3 Papakonstantinou 分布と Batschelet 分布

分布の頂上付近が平坦であったり, 反対に急傾斜であったりすることが観察されるときには, Papakonstantinou 分布 (Papakonstantinou, 1979) や Batschelet 分布 (Batschelet, 1981) の利用が考えられる.

Papakonstantinou 分布の確率密度関数は

$$(2.2) \quad f(\theta) = \frac{1}{2\pi\{1 - \kappa J_1(\nu)\}} [1 + \kappa \cos\{(\theta - \mu) + \nu \sin(\theta - \mu)\}], \quad -\pi \leq \theta < \pi$$

で与えられる. J_r は r 位の第 1 種 Bessel 関数

$$J_r(z) = \frac{1}{\pi} \int_0^\pi \cos(r\theta - z \sin\theta) d\theta, \quad r = 0, 1, 2, \dots$$

を表す. μ ($-\pi \leq \mu < \pi$) は平均方向, κ ($0 \leq \kappa \leq 1$) は分布の集中度を表すパラメータで, ν ($-\infty < \nu < \infty$) は分布の形と単峰・多峰を制御するパラメータである. $\kappa \neq 0$ (一様分布でない) のとき, 分布が単峰であるための必要十分条件は $|\nu| \leq 1$ である. $\nu = 0$ であれば, (2.2) がハート型分布に帰着することは明らかであろう. $-1 \leq \nu < 0$ ならばハート型分布より平坦であり, $0 < \nu \leq 1$ ならば, より急傾斜となる. Papakonstantinou 分布の性質や適用例は Abe et al. (2009) に見ることができる. 図 4 に, $\mu = \pi$, $\kappa = 0.75$, $\nu = -1$, 0 (ハート型分布), 1 の組合せに対して Papakonstantinou 分布の確率密度関数のプロットを与える.

Batschelet 分布の確率密度関数は

$$(2.3) \quad f(\theta) = C^{-1} \exp[\kappa \cos\{(\theta - \mu) + \nu \sin(\theta - \mu)\}], \quad -\pi \leq \theta < \pi$$

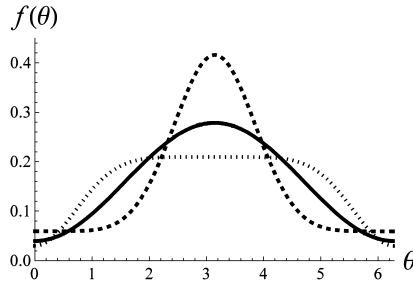


図 4. Papakonstantinou 分布の確率密度関数 (2.2) のプロット： $\mu = \pi$, $\kappa = 0.75$; $\nu = -1$ (点線), 0 (ハート型: 実線), 1 (破線).

である。ここで、 $-\pi \leq \mu < \pi$, $\kappa \geq 0$, $-\infty < \nu < \infty$ であり、 C は積分

$$C = \int_{-\pi}^{\pi} \exp\{\kappa \cos(\theta + \nu \sin \theta)\} d\theta$$

で与えられる正規化定数を表す。 $\nu = 0$ であれば、(2.3) は von Mises 分布に帰着する。Batschelet 分布の単峰性の必要十分条件は Papakonstantinou 分布の場合と同じ $|\nu| \leq 1$ であり、 $-1 \leq \nu < 0$ ならば von Mises 分布より平坦であり、 $0 < \nu \leq 1$ ならば、より急傾斜となる。Batschelet 分布のパラメータ推定と適用例は Pewsey et al. (2011) に与えられている。Batschelet 分布の確率密度関数の形状は Papakonstantinou 分布と似たものになる。

2.4 非対称分布の当てはめ

第 2.3.1 節で対称性の検定を述べ、第 2.3.2 節で柔軟性を持つ対称分布の当てはめについて紹介し、第 2.3.3 節では別の形の対称分布の利用可能性について簡単に触れた。対称性の検定によって対称性が棄却された場合は非対称分布を当てはめることになるであろうし、また対称性に拘ることなくデータに対して良い適合を与える分布に関心が持たれる場合は、初めから非対称分布の適合を考えるかも知れない。本節では、第 2.3.2 節および第 2.3.3 節の分布の非対称化について述べる。得られる分布はそれぞれの対称分布を特別な場合として含んでいる。

2.4.1 Jones と Pewsey の分布の正弦関数摂動

Abe and Pewsey (2011) において Jones and Pewsey (2005) の分布の非対称化が与えられている。非対称分布構成のアイデアは単純明快であり、(2.1) の $f_{\psi}(\theta)$ を正弦 (sin) 関数によって

$$(2.4) \quad f(\theta) = \{1 + \lambda \sin(\theta - \mu)\} f_{\psi}(\theta), \quad -\pi \leq \theta < \pi, \quad -\pi \leq \mu < \pi, \quad -1 \leq \lambda \leq 1$$

と摂動するものである。確率密度関数 (2.4) は一般に非対称 ($\lambda = 0$ ならば対称) であり、 $\kappa > 0$ および $-1 \leq \lambda \leq 1$ に対して $4 - 3\sqrt{3} \leq \psi \leq -0.5$ ならば単峰である。Jones と Pewsey の分布が特別な場合として von Mises, 巻き込み Cauchy, ハート型の各分布を含んでいることから、Abe と Pewsey の分布は正弦関数摂動 von Mises, 巻き込み Cauchy, ハート型の各分布を特別な場合として持つことは明らかである。図 5 に、 $\mu = \pi$, $\kappa = 1$, $\lambda = 0$ (von Mises 分布), 0.5, 1 の組合せに対して von Mises 分布の正弦関数摂動確率密度関数のプロットを与える。

(2.4) の分布をフルモデルとし、正弦関数摂動 von Mises, 巻き込み Cauchy, ハート型の各分布および Jones と Pewsey の対称分布をサブモデルとして AIC, BIC 最小化法により最適モデルを選択する方法が考えられる。Abe et al. (2012b) では、北西フィンランド Pallas-Ylläs National Park と北西ロシア Dvina-Pinega 域における *Picea abies* 倒木方向データに対して上記の対称お

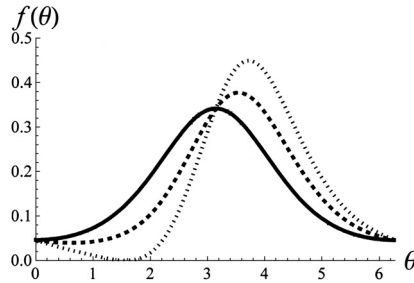


図 5. von Mises 分布の正弦関数摂動確率密度関数のプロット： $\mu = \pi$, $\kappa = 1$; $\lambda = 0$ (von Mises: 実線), 0.5 (破線), 1 (点線).

表 1. 霞ヶ浦風向データへの正弦関数摂動分布の当てはめ (上段: 朝 6 時, 下段: 正午 12 時).

正弦関数摂動分布	$\hat{\mu}$	$\hat{\kappa}(\hat{\rho})$	$\hat{\psi}$	$\hat{\lambda}$	MLL	AIC	BIC
Jones-Pewsey	0.786	0.969	-3.097	-0.153	-101.243	210.486	218.930
	1.574	1.337	-2.243	0.198	-89.464	186.928	195.372
von Mises	1.143	0.399	($\psi \rightarrow 0$)	-0.316	-108.437	222.874	229.206
	1.163	0.966		0.637	-94.489	194.977	201.310
巻き込み Cauchy	0.743	0.619	($\psi = -1$)	-0.127	-107.544	221.088	227.420
	1.602	1.279		0.155	-91.000	188.000	194.333
ハート型	6.114	0.394	($\psi = 1$)	0.290	-108.624	223.248	229.581
	2.371	1.008		-0.5534	-95.909	197.817	204.150

よび非対称 von Mises と巻き込み Cauchy の分布を当てはめて対称 von Mises 分布を最適な分布として選択している. また, 北西ロシアの Kazkim 川流域のデータでは正弦関数摂動巻き込み Cauchy 分布を選択している.

2.4.2 風向データへの当てはめ例

第 2.2 節の霞ヶ浦風向データに対する Pewsey (2002) の対称性の検定 (第 2.3.1 節) の P 値は, 朝 6 時の風向データに対して 0.013 で, 正午 12 時に対して 0.008 であった. この結果から, データに非対称分布を当てはめることとし, 第 2.4.1 節の方法を適用してみよう. 表 1 に, 正弦関数摂動 Jones-Pewsey 分布をフルモデルとし, 正弦関数摂動 von Mises, 巻き込み Cauchy, ハート型分布をサブモデルとして霞ヶ浦朝 6 時と正午 12 時の風向データへの当てはめを行ったときの最尤推定値, 最大対数尤度の値, AIC と BIC の値を与える. 6 時と 12 時の双方とも AIC 最小の意味ではフルモデルが選択される. また, 図 6 は, 円周プロット上で分布の当てはめを示している.

2.4.3 非対称 Papakonstantinou 分布と非対称 Batschelet 分布

第 2.3.3 節の Papakonstantinou 分布と Batschelet 分布の確率密度関数における cos-sin の組合せを sin-sin もしくは本質的に同等であるが cos-cos とするだけで非対称性を表すようにできる. 具体的な確率密度関数は次の通りである:

$$(2.5) \quad f(\theta) = \frac{1}{2\pi} [1 + \kappa \sin\{(\theta - \mu) + \nu \sin(\theta - \mu)\}],$$

$$(2.6) \quad f(\theta) = C^{-1} \exp[\kappa \sin\{(\theta - \mu) + \nu \sin(\theta - \mu)\}].$$

非対称 Papakonstantinou 分布では正規化定数に特殊関数を含まないが, 非対称 Batschelet 分

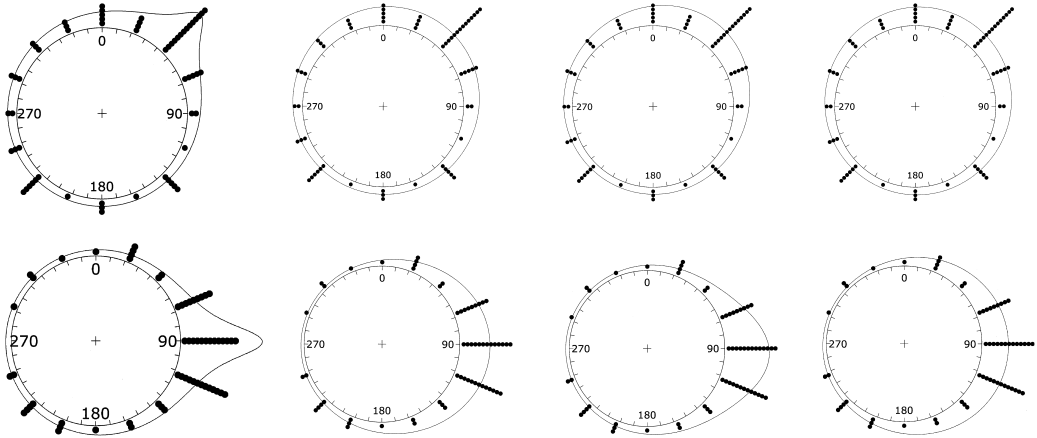


図 6. 朝 6 時(上段)と正午 12 時(下段)の風向データに対して、左から正弦関数摂動 Jones-Pewsey 分布, von Mises 分布, 巻き込み Cauchy 分布, ハート型分布の当てはめ。

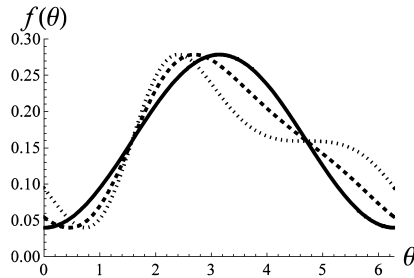


図 7. 非対称 Papakonstantinou 分布の確率密度関数のプロット： $\mu = \pi/2$, $\kappa = 0.75$; $\nu = 0$ (ハート型：実線), 0.5 (破線), 1 (点線)。

布の正規化定数では一般に積分

$$C = \int_{-\pi}^{\pi} \exp\{\kappa \sin(\theta + \nu \sin \theta)\} d\theta$$

で表現されることになる。 $\nu = 0$ であれば、(2.5)と(2.6)はそれぞれ明らかにハート型と von Mises 分布に帰着する。 図 7 に、 $\mu = \pi/2$, $\kappa = 0.75$, $\nu = 0$ (ハート型分布), 0.5, 1 の組合せに対して非対称 Papakonstantinou 分布の確率密度関数のプロットを与える。 非対称 Batschelet 分布の確率密度関数の形状は非対称 Papakonstantinou 分布の場合に似たものになる。

3. 角度変数を含む回帰モデル

3.1 角度／線形説明変数・線形目的変数の回帰モデル

説明変数に三角関数を含み、目的変数が線形であるような回帰モデルは Johnson and Wehrly (1978) によって与えられている。 モデルを導くための基礎となるのは次の事実である。

角度ベクトルを $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_p)'$ とし,

$$H(\Theta) = \begin{pmatrix} \cos \Theta_1 \cdots \cos n\Theta_1 & \sin \Theta_1 \cdots \sin n\Theta_1 \\ \vdots & \vdots \\ \cos \Theta_p \cdots \cos n\Theta_p & \sin \Theta_p \cdots \sin n\Theta_p \end{pmatrix}$$

とおき, 線形ベクトルを $\mathbf{X} = (X_1, X_2, \dots, X_q)'$ とする. Θ と \mathbf{X} は結合確率密度関数

$$(3.1) \quad f(\boldsymbol{\theta}, \mathbf{x}) = c^{-1} \exp \left\{ -\frac{1}{2} \mathbf{x}' \Sigma^{-1} \mathbf{x} + \boldsymbol{\lambda}' \Sigma^{-1} \mathbf{x} + \mathbf{a}(\boldsymbol{\theta})' \Sigma^{-1} \mathbf{x} \right\}$$

を持つとする. ここで, c は正規化定数で, $\boldsymbol{\theta} \in [0, 2\pi)^p$, $\mathbf{x} \in \mathbb{R}^q$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_q)' \in \mathbb{R}^q$, $\mathbf{a}(\boldsymbol{\theta}) = (a_1(\boldsymbol{\theta}), a_2(\boldsymbol{\theta}), \dots, a_q(\boldsymbol{\theta}))' \in \mathbb{R}^q$,

$$\begin{aligned} a_i(\boldsymbol{\theta}) &= \sum_{j=1}^p \sum_{k=1}^n a_{ijk} \cos\{k(\theta_j - \mu_{ijk})\} \\ &= \sum_{j=1}^p \sum_{k=1}^n \{\alpha_{ijk} \cos(k\theta_j) + \beta_{ijk} \sin(k\theta_j)\}, \quad i=1, \dots, q \end{aligned}$$

である. α_{ijk} と β_{ijk} は $\alpha_{ijk} = a_{ijk} \cos(k\mu_{ijk})$, $\beta_{ijk} = a_{ijk} \sin(k\mu_{ijk})$ を表す. また, Σ^{-1} は $q \times q$ 正定値行列である. (3.1) の結合確率密度関数 $f(\boldsymbol{\theta}, \mathbf{x})$ は $E(\mathbf{X}\mathbf{X}')$, $E(\mathbf{X})$, $E\{\mathbf{X} \otimes H(\Theta)\}$ が指定された値を持つという制約条件の下に $\int f(\boldsymbol{\theta}, \mathbf{x}) \log f(\boldsymbol{\theta}, \mathbf{x}) d\boldsymbol{\theta} d\mathbf{x}$ を最大化する確率密度関数として特徴付けられる. なお, \otimes はクロネッカー積を表す.

いま, $\mathbf{X} = (\mathbf{X}'_1; \mathbf{X}'_2)'$ と分割 ($\mathbf{X}_1: r \times 1, \mathbf{X}_2: (q-r) \times 1$) し, 対応して $\boldsymbol{\lambda} = (\boldsymbol{\lambda}'_1; \boldsymbol{\lambda}'_2)'$, $\mathbf{a}(\boldsymbol{\theta}) = (\mathbf{a}_1(\boldsymbol{\theta})'; \mathbf{a}_2(\boldsymbol{\theta})')'$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ と分割すると, $\mathbf{X}_2 = \mathbf{x}_2$ と $\Theta = \boldsymbol{\theta}$ が所与のときの \mathbf{X}_1 の条件付き分布は平均ベクトル $\boldsymbol{\lambda}_1 + \mathbf{a}_1(\boldsymbol{\theta}) + \Sigma_{12}\Sigma_{22}^{-1}\{\mathbf{x}_2 - (\boldsymbol{\lambda}_2 + \mathbf{a}_2(\boldsymbol{\theta}))\}$ で分散共分散行列 $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ の r 次元正規分布に従うことが分る. 特に, 各成分 X_i ($i=1, 2, \dots, r$) の平均は

$$\nu_0 + \sum_{i=r+1}^q \nu_i x_i + \sum_{i=r+1}^q \sum_{j=1}^p \sum_{k=1}^n \{\gamma_{ijk} \cos(k\theta_j) + \delta_{ijk} \sin(k\theta_j)\}$$

の形となる. 上の事実より, 角度変数 $\theta_1, \theta_2, \dots, \theta_p$ に対して正弦・余弦変換を施して, 角度変数ベクトル $\boldsymbol{\theta}$ と線形変数ベクトル \mathbf{x}_2 を説明変数とする線形回帰モデルを構成できたことになる. したがって, 通常回帰モデル分析の手続きを適用することにより, モデルのパラメータを推定することができる.

特別な場合として, 角度変数 Θ と線形変数 X の次の結合確率密度関数

$$(3.2) \quad f(\theta, x) = c^{-1} \exp \left\{ -\frac{x^2}{2\sigma^2} + \frac{\lambda x}{\sigma^2} + \frac{\kappa x}{\sigma^2} \cos(\theta - \mu) \right\}, \quad 0 \leq \theta < 2\pi, \quad -\infty < x < \infty$$

を持つシリンダー上の分布 (Johnson and Wehrly, 1978) を考えてみる. ここで, c は正規化定数であり, $\mu, \sigma^2, \lambda, \kappa$ はパラメータで $0 \leq \mu < 2\pi$, $\sigma > 0$, $-\infty < \lambda < \infty$, $\kappa > 0$ である. 周辺分布はよく知られた形ではないが, $X = x$ が所与のときの Θ の条件付き分布は von Mises 分布 $VM(\mu, \kappa x/\sigma^2)$ となり, また $\Theta = \theta$ が所与のときの X の条件付き分布は平均 $\lambda + \kappa \cos(\theta - \mu)$, 分散 σ^2 の正規分布となる. したがって, 説明変数 \mathbf{x} (線形ベクトル) と θ (角度), 目的変数 y (線形) に対して

$$y = \mathbf{x}'\boldsymbol{\beta} + \kappa \cos(\theta - \mu)$$

の形の回帰モデルの利用が自然となることが分る. ここで, $\boldsymbol{\beta}$ はパラメータベクトル, κ と μ はパラメータである. なお, Johnson and Wehrly (1978) では, 解析例として大気質指標 x_1 を

温度 x_2 と風向 θ で説明する回帰式 $x_1 = a + bx_2 + c_1 \cos \theta + c_2 \sin \theta$ が与えられている。ここにおいて a, b, c_1, c_2 はパラメータを表す。

3.2 角度／線形説明変数・線形目的変数の変形回帰モデル

対称 Papakonstantinou 分布もしくは対称 Batschelet 分布の確率密度関数に現れる形を参考にして、前節の (3.2) 式を

$$f(\theta, x) = c^{-1} \exp \left[-\frac{x^2}{2\sigma^2} + \frac{\lambda x}{\sigma^2} + \frac{\kappa x}{\sigma^2} \cos\{\theta - \mu + \nu \sin(\theta - \mu)\} \right],$$

$$0 \leq \theta < 2\pi, \quad -\infty < x < \infty$$

と修正してみよう。そうすると、 $\Theta = \theta$ が所与のときの X の条件付き分布は平均 $\lambda + \kappa \cos\{\theta - \mu + \nu \sin(\theta - \mu)\}$ 、分散 σ^2 の正規分布となることが分る。同様に非対称 Papakonstantinou 分布もしくは非対称 Batschelet 分布の確率密度関数に現れる形を参考に \cos - \sin の組合せを \sin - \sin もしくは \cos - \cos の組合せに修正することができる。こうして、SenGupta and Ugwuowo (2006) は、説明変数 x (線形ベクトル) と θ (角度)、目的変数 y (線形) に対して次の形の回帰式の使用を提案している：

$$y = \mathbf{x}'\beta_1 + a_1 \cos\{\theta - \mu_1 + \nu_1 \sin(\theta - \mu_1)\},$$

$$y = \mathbf{x}'\beta_2 + a_2 \cos\{\theta - \mu_2 + \nu_2 \cos(\theta - \mu_2)\}.$$

ここで、 $(\beta_1, a_1, \mu_1, \nu_1)$ と $(\beta_2, a_2, \mu_2, \nu_2)$ は、それぞれのモデルのパラメータである。誤差が正規分布 $N(0, \sigma^2)$ に従うと仮定するときには、パラメータ $\beta_1, a_1, \mu_1, \nu_1, \sigma^2$ と $\beta_2, a_2, \mu_2, \nu_2, \sigma^2$ の最尤推定値を最適化の手法で求めることができる。

3.3 線形説明変数・角度目的変数の回帰モデル

Fisher and Lee (1992) と Gill and Hangartner (2010) は、von Mises 分布の平均方向パラメータが共変数の一次結合として表されると仮定するモデル化について考察を与えている。角度観測値 $\theta_1, \theta_2, \dots, \theta_n$ は von Mises 分布 $VM(\mu_j, \kappa)$ からのものとし、各平均方向 μ_j は

$$\mu_j = \mu + g(\beta' \mathbf{x}_j), \quad g(z) = 2 \tan^{-1} z$$

と表されると仮定する。集中度を表すパラメータ κ は j に依らず一定と仮定している。ここで、 \mathbf{x}_j は p 次元共変数ベクトルであり、 β は p 次元パラメータベクトルを表す。データ $\theta_1, \theta_2, \dots, \theta_n$ に基づいてパラメータ μ, κ, β を最尤推定するためには、尤度関数

$$L(\mu, \kappa, \beta | \theta_1, \theta_2, \dots, \theta_n; \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = -n \log\{2\pi I_0(\kappa)\} + \kappa \sum_{j=1}^n \cos\{\theta_j - \mu - g(\beta' \mathbf{x}_j)\}$$

を μ, κ, β に関して最大化する。

いま、

$$u_j = \sin\{\theta_j - \mu - g(\beta' \mathbf{x}_j)\}, \quad \mathbf{u} = (u_1, u_2, \dots, u_n)', \quad \mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)',$$

$$G = \text{diag} \left(\left. \frac{dg(t)}{dt} \right|_{t=\beta' \mathbf{x}_1}, \left. \frac{dg(t)}{dt} \right|_{t=\beta' \mathbf{x}_2}, \dots, \left. \frac{dg(t)}{dt} \right|_{t=\beta' \mathbf{x}_n} \right),$$

$$S = \frac{1}{n} \sum_{j=1}^n \sin\{\theta_j - g(\beta' \mathbf{x}_j)\}, \quad C = \frac{1}{n} \sum_{j=1}^n \cos\{\theta_j - g(\beta' \mathbf{x}_j)\}, \quad R = (S^2 + C^2)^{1/2}$$

とおくとき、最尤推定方程式は

$$\mathbf{X}'G\mathbf{u} = \mathbf{0}, \quad R \sin \hat{\mu} = S, \quad R \cos \hat{\mu} = C, \quad A(\hat{\kappa}) = R$$

与えられる. ここで, $A(\kappa)$ は, $I_r(\kappa)$ を r 位の第 1 種変形 Bessel 関数 $I_r(\kappa) = (2\pi)^{-1} \int_0^{2\pi} \cos(r\theta) e^{\kappa \cos \theta} d\theta$ として, $A(\kappa) = I_1(\kappa)/I_0(\kappa)$ を表す. Fisher and Lee (1992) は, 最尤推定方程式の数値解を求めるために反復再重み付け最小二乗法を用いている.

3.4 角度説明変数・角度目的変数の回帰モデル

説明変数と目的変数が共に角度であるときの回帰モデルは Downs and Mardia (2002), Kato et al. (2008), Kato and Jones (2010) で取り扱われている. 回帰モデルを表現するために Möbius 変換を用いるので, 複素数を持ち出すと数学的扱いが簡潔になる部分がある. 以下では, 通常の角度表現による解釈も与える.

複素平面上の単位円周を $\Omega = \{z \in \mathbb{C}; |z| = 1\}$ とし, $j = 1, 2, \dots, n$ に対し Ω 上の共変数を $x_j = e^{i\xi_j}$, 独立同一分布する Ω 上の誤差を $\varepsilon_j = e^{i\Phi_j}$ と複素数表示する. i は虚数単位 $i = \sqrt{-1}$ を表す. また, Ω 上にない複素定数 $\alpha = \rho_\alpha e^{i\mu_\alpha} \in \mathbb{C}$ ($\rho_\alpha \geq 0$, $\rho_\alpha \neq 1$, $-\pi \leq \mu_\alpha < \pi$) と Ω 上の複素定数 $\beta = e^{i\mu_\beta} \in \Omega$ ($-\pi \leq \mu_\beta < \pi$) を取る. このとき, Ω 上の目的変数を Y_j とする回帰モデル

$$(3.3) \quad Y_j = \beta \frac{x_j + \alpha}{1 + \bar{\alpha}x_j} \varepsilon_j$$

を考察することができる. ここで, $\bar{\alpha}$ は α の共役複素数を表す. 変換 $M(z) = (z + \alpha)/(1 + \bar{\alpha}z)$ は, 単位円周上の点を単位円周上に写す Möbius 変換として知られている. (3.3) の表現を用いれば複素関数論の知識を使って単位円周上の回帰モデル理論を展開することが可能となるが, ここでは (3.3) の角度表現

$$(3.4) \quad \Theta_j \equiv \arg(Y_j) = \nu(\mu_\alpha, \mu_\beta, \rho_\alpha, \xi_j) + \Phi_j,$$

$$\nu(\mu_\alpha, \mu_\beta, \rho_\alpha, \xi_j) = 2 \arctan \left\{ \left(\frac{1 - \rho_\alpha}{1 + \rho_\alpha} \right) \tan \frac{1}{2}(\xi_j - \mu_\alpha) \right\} + (\mu_\alpha + \mu_\beta)$$

を用いることにする. (3.4) は, 単位円 (点 (0, 1) を除く) と実軸の 1 対 1 対応を表す立体射影 (図 8) を用いて次のように解釈することができる. 角度 $\xi_j - \mu_\alpha$ を立体射影により実軸に写した点を $(1 - \rho_\alpha)/(1 + \rho_\alpha)$ 倍して尺度を変え, 次に逆立体射影により実軸から単位円に写した後に角度 $\mu_\alpha + \mu_\beta$ 分を回転することにより角度が ν となる.

Downs and Mardia (2002) は Φ_j に von Mises 分布 $VM(0, \kappa)$ を仮定し, データ (θ_j, ξ_j) から対数尤度関数

$$L(\mu_\alpha, \mu_\beta, \rho_\alpha, \kappa) = -n \log I_0(\kappa) + \kappa \sum_{j=1}^n \cos\{\theta_j - \nu(\mu_\alpha, \mu_\beta, \rho_\alpha, \xi_j)\} + \text{const.}$$

を最大化し, パラメータ $\mu_\alpha, \mu_\beta, \rho_\alpha, \kappa$ を最尤推定する方法について研究した. また, Kato et al. (2008) は Φ_j に巻き込み Cauchy 分布 $WC(0, \rho)$ を仮定することを提案している. 数値計算の手

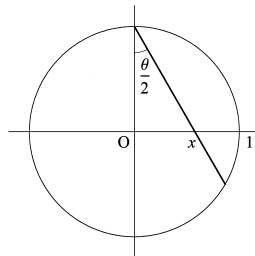


図 8. 立体射影 $x = \tan(\theta/2)$, $-\pi < \theta < \pi$.

間としては誤差分布として VM を仮定しても WC を仮定しても大差はないが、WC では WC 変数の Möbius 変換は再び WC 分布に従い、WC 分布は再生性を持ち、また、 ξ_j が所与のときの Θ_j の条件付き分布は WC 分布となることが知られている。これらのように、誤差分布として WC を仮定すると Möbius 変換と相性が良いので理論展開がスムーズに進むというメリットがある。

さらに、Kato and Jones (2010) は VM 変数を Möbius 変換した分布を誤差分布として採用することを提案した。その誤差分布は一般に非対称、2 峰性であり、VM と WC 分布を特別な場合として含むという大きなメリットを持つ。少し複雑な分布ではあるが、角度変数間回帰モデルによるデータ解析においてはこの分布を誤差分布として採用することが推奨される。

4. 角度変数間の構造モデル

線形変数に関する構造モデルに類似して、角度変数間の構造モデル (Wang and Shimizu, 2012)

$$\theta = \xi + \delta, \quad \phi = \eta + \varepsilon$$

を考えよう。 ξ と η は角度確率変数であり、関係 $\eta = M(\xi)$ を持つとする。また、 δ と ε は観測誤差を表す独立な角度確率変数で、 ξ とともに独立とする。簡単のために、 ξ はハート型分布 $C(\mu, \rho)$ に従い、 δ と ε はそれぞれ平均方向 0 のハート型分布 $C(0, \rho_1)$ と $C(0, \rho_2)$ に従うと仮定する。関係 $\eta = M(\xi)$ は角度と角度の関係であるので、(3.4) の右辺で現れている Möbius 変換を用いることにする。

確率ベクトル $(\theta, \phi)'$ の分布は 2 変量角度分布 (トーラス上の分布) を表す。この分布を $MC_2(\mu, \rho, \mu_\alpha, \rho_\alpha, \mu_\beta, \rho_\beta, \rho_1, \rho_2)$ と表記することにしよう。 θ の周辺分布はハート型 $C(\mu, \rho\rho_1)$ であり、 ϕ の周辺分布はハート型 $C(\mu_\phi, \rho_\phi)$ である。ここで、

$$\begin{aligned} \mu_\phi &= \mu_\alpha + \mu_\beta + \tan^{-1} \left\{ \frac{\rho(1 - \rho_\alpha^2) \sin(\mu - \mu_\alpha)}{\rho_\alpha + \rho|1 - \rho_\alpha^2| \cos(\mu - \mu_\alpha)} \right\}, \\ \rho_\phi &= \rho_2 \left\{ \rho^2(1 - \rho_\alpha^2)^2 + \rho_\alpha^2 + 2\rho\rho_\alpha(1 - \rho_\alpha^2) \cos(\mu - \mu_\alpha) \right\}^{1/2} \end{aligned}$$

を表す。 $MC_2(\mu, \rho, \mu_\alpha, \rho_\alpha, \mu_\beta, \rho_\beta, \rho_1, \rho_2)$ に従う $(\theta, \phi)'$ の結合確率密度関数は幾分面倒な形であるが、陽に表すことができる。最初に仮定した ξ, δ, ε の分布はハート型と限定されるわけではなく von Mises や巻き込み Cauchy 分布でも構わないが、ハート型以外のときにモデルの統計的性質を導くことは現在のところ成功していない。

Wang and Shimizu (2012) の方法に倣い、本稿では例として第 2.2 節の霞ヶ浦風向データ中、朝 6 時と正午 12 時の風向データの解析例を以下に示す。表 2 に、(f) $MC_2(\mu, \rho, \mu_\alpha, \rho_\alpha, \mu_\beta, \rho_\beta, \rho_1, \rho_2)$ をフルモデルとし、(a) 恒等変換 $M(\xi) = \xi$ 、(b) 角度変換 $M(\xi) = \xi + \mu_\beta$ 、2 つの独立モデル (c) $MC_2(\mu, \rho, \mu_\alpha, \rho_\alpha, \mu_\beta, 0, \rho_2)$ と (d) $MC_2(\mu, \rho, -, -, -, \rho_1, 0)$ 、(e) モデル $MC_2(\mu, \rho, \mu_\alpha, \rho_\alpha, \mu_\beta, \rho_1, \rho_2)$ 、 $\rho_1 = \rho_2$ をサブモデルとするときの最尤推定値、最大対数尤度の値、AIC と BIC の値を与える。AIC と BIC 最小の意味では (e) $MC_2(\mu, \rho, \mu_\alpha, \rho_\alpha, \mu_\beta, \rho_1, \rho_2)$ 、 $\rho_1 = \rho_2$ のモデルが選択される。図 9 は選択されたモデルの等高線と構造曲線の図と 3 次元プロットを示している。 μ_β の値は 0.933 と推定されている。この値は標本平均方向の差 1.161 ($= 1.702 - 0.541$) に近い。しかしながら、選択されたモデルは朝 6 時と正午 12 時の間に平均方向のズレだけでなく、点 $\alpha = \rho_\alpha e^{i\mu_\alpha}$ を含む変換を通じた構造の存在を示唆している。なお、選択されたモデルでは $\hat{\rho}_1 = \hat{\rho}_2 = 0.5$ であり、ハート型分布 $C(\mu, \rho)$ の ρ の範囲 $0 \leq \rho \leq 1/2$ の上限に達している。このことはハート型よりも急傾斜な性質を持つ分布が採用されるべきであることを示唆しているので、 ξ, δ, ε の分布がハート型以外のときの理論構築が望まれる。

表 2. (a) 恒等変換 $M(\xi) = \xi$, (b) 角度変換 $M(\xi) = \xi + \mu_\beta$, 独立モデル (c) $MC_2(\mu, \rho, \mu_\alpha, \rho_\alpha, \mu_\beta, 0, \rho_2)$ と (d) $MC_2(\mu, \rho, -, -, -, \rho_1, 0)$, (e) モデル $MC_2(\mu, \rho, \mu_\alpha, \rho_\alpha, \mu_\beta, \rho_1, \rho_2)$, $\rho_1 = \rho_2$, (f) フルモデル $MC_2(\mu, \rho, \mu_\alpha, \rho_\alpha, \mu_\beta, \rho_1, \rho_2)$ におけるモデル選択.

モデル	$\hat{\mu}$	$\hat{\rho}$	$\hat{\mu}_\alpha$	$\hat{\rho}_\alpha$	$\hat{\mu}_\beta$	$\hat{\rho}_1$	$\hat{\rho}_2$	MLL	AIC	BIC
(a)	1.616	0.5	—	—	—	0.304	0.5	-210.625	429.25	437.693
(b)	0.709	0.5	—	—	1.070	0.434	0.5	-206.855	423.71	434.264
(c)	6.045 (2.903)	0.292	4.048 (0.906)	0.908	3.994 (0.852)	—	0.469	-210.311	432.622	445.288
(d)	0.424	0.427	—	—	—	0.461	—	-221.348	448.696	455.029
(e)	0.557	0.335	1.018	0.534	0.933	0.5	0.5	-203.714	419.429	432.094
(f)	0.557	0.335	1.018	0.534	0.933	0.5	0.5	-203.714	421.429	436.205

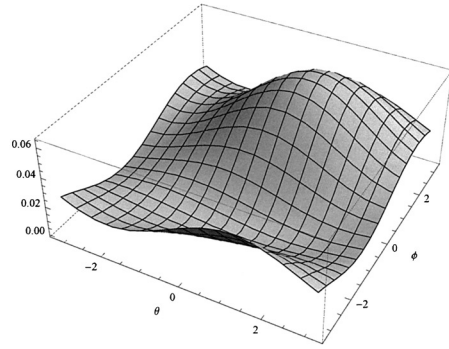
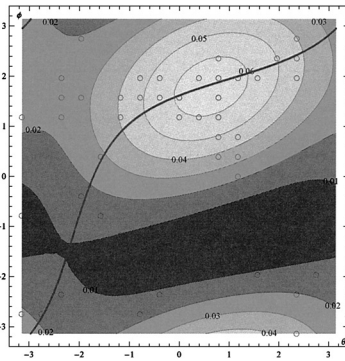


図 9. 霞ヶ浦朝 6 時と正午 12 時の風向データに対する構造モデルの推定 (左: 推定された等高線と構造曲線, 右: 3 次元プロット).

5. おわりに

本稿では、環境科学においてしばしば現れる角度データに対して、データに含まれる情報を抽出するためにはどのような統計量を算出すればよいのか、また計算された統計量はどのような意味・意義を持つのかから始めて、角度分布とその当てはめ、シリンダー上の分布から導かれる角度回帰モデル、角度構造モデルとトーラス上の分布について述べた。これらは、角度を含むデータを扱う方向統計学の一部をなしている。とは言え、環境統計学において現れる角度を含むデータは本稿で述べたものに限られるわけでないことは明らかである。他にも興味のあるデータの型はさまざまな場面で現れる。例えば、Abe et al. (2012a) は倒木データの向きを軸データとして捉えたときの解析方法とデータ解析例を示している。関心のある読者のために、軸分布に関する文献として Arnold and SenGupta (2006, 2011) を挙げておこう。また Abe et al. (2012b) は、第 2.4.1 節にあるように、通常角度分布を用いて倒木データを解析している。島谷編 (2011) は統計数理研究所で行われた共同利用研究集会「生物群集の多様性と統計数理」の講演のテープを起して作成された共同研究レポートである。その中に阿部と加藤のそれぞれの講演が第 4 章に方向統計学として文章化されており、倒木データの角度統計解析と第 3.4 節の Kato and Jones (2010) の角度回帰モデルが分かりやすく解説されているので、環境統計学・生態学への方向統計学の利用に関心を持つ読者に一読を薦めたい。また、島谷 (2012) の第 8 章に鳥の飛行方向データへ角度分布を適用した例が紹介されている。

環境科学では典型的な(風向, 風速)のモデル化にはシリンダー上の分布を欠くことができないが, 未だ研究途上のものが多いため本稿では多くを述べることをしなかった. ディスク上の分布の理論と実際についてや, 外れ値もしくはモデルに強く影響を与える観測値の検出に有効な影響診断法の開発等についても機会があれば稿を改めて紹介できればと考えている.

謝 辞

原稿は, 2名の査読者による有益なコメントに基づいて改訂されました. ここに両査読者に感謝します.

参 考 文 献

- Abe, T. and Pewsey, A. (2011). Sine-skewed circular distributions, *Statistical Papers*, **52**, 683–707.
- Abe, T., Pewsey, A. and Shimizu, K. (2009). On Papakonstantinou's extension of the cardioid distribution, *Statistics & Probability Letters*, **79**, 2138–2147.
- Abe, T., Shimizu, K. and Pewsey, A. (2010). Symmetric unimodal models for directional data motivated by inverse stereographic projection, *Journal of the Japan Statistical Society*, **40**, 45–61.
- Abe, T., Shimizu, K., Kuuluvainen, T. and Aakala, T. (2012a). Sine-skewed axial distributions with an application for fallen tree data, *Environmental and Ecological Statistics*, **19**, 295–307.
- Abe, T., Kubota, Y., Shimatani, K., Aakara, T. and Kuuluvainen, T. (2012b). Circular distributions of fallen logs as an indicator of forest disturbance regimes, *Ecological Indicators*, **18**, 559–566.
- Arnold, B. C. and SenGupta, A. (2006). Probability distributions and statistical inference for axial data, *Environmental and Ecological Statistics*, **13**, 271–285.
- Arnold, B. C. and SenGupta, A. (2011). Models for axial data, *Advances in Directional and Linear Statistics* (eds. M. T. Wells and A. SenGupta), 1–9, Springer-Verlag, Berlin.
- Batschelet, E. (1981). *Circular Statistics in Biology*, Academic Press, London.
- Cartwright, D. E. (1963). The use of directional spectra in studying the output of a wave recorder on a moving ship, *Ocean Wave Spectra*, 203–218, Prentice Hall, Englewood Cliffs.
- Downs, T. D. and Mardia, K. V. (2002). Circular regression, *Biometrika*, **89**, 683–697.
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*, Cambridge University Press, Cambridge.
- Fisher, N. I. and Lee, A. J. (1992). Regression models for an angular response, *Biometrics*, **48**, 665–677.
- Gill, J. and Hangartner, D. (2010). Circular data in political science and how to handle it, *Political Analysis*, **18**, 316–336.
- Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*, World Scientific, Singapore.
- Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models, *Journal of the American Statistical Association*, **73**, 602–606.
- Jones, M. C. and Pewsey, A. (2005). A family of symmetric distributions on the circle, *Journal of the American Statistical Association*, **100**, 1422–1428.
- Kato, S. and Jones, M. C. (2010). A family of distributions on the circle with links to, and applications arising from, Möbius transformation, *Journal of the American Statistical Association*, **105**, 249–262.
- Kato, S., Shimizu, K. and Shieh, G. S. (2008). A circular-circular regression model, *Statistica Sinica*, **18**, 633–645.
- Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*, Wiley, Chichester.

- Papakonstantinou, V. (1979). Beiträge zur zirkulären Statistik, Ph.D. Dissertation, University of Zurich, Switzerland.
- Pewsey, A. (2002). Testing circular symmetry, *Canadian Journal of Statistics*, **30**, 591–600.
- Pewsey, A., Shimizu, K. and de la Cruz, R. (2011). On an extension of the von Mises distribution due to Batschelet, *Journal of Applied Statistics*, **38**, 1073–1085.
- SenGupta, A. and Ugwuowo, F. I. (2006). Asymmetric circular-linear multivariate regression models with applications to environmental data, *Environmental and Ecological Statistics*, **13**, 299–309.
- 島谷健一郎 編 (2011). 生物群集の多様性と統計数理, 統計数理研究所共同研究レポート, No. 263, 177–211.
- 島谷健一郎 (2012). 『フィールドデータによる統計モデリングと AIC, ISM シリーズ: 進化する統計数理 2』, 近代科学社, 東京.
- Shimizu, K. and Iida, K. (2002). Pearson type VII distributions on spheres, *Communications in Statistics—Theory and Methods*, **31**, 513–526.
- Wang, M.-Z. and Shimizu, K. (2012). On applying Möbius transformation to cardioid random variables, *Statistical Methodology*, **9**, 604–614.
- Zar, J. H. (2010). *Biostatistical Analysis*, fifth ed., Prentice Hall, Upper Saddle River.

Use of Directional Statistics in Environmental Science

Kunio Shimizu^{1,2} and Minzhen Wang³

¹Department of Mathematics, Keio University

²Visiting Professor, The Institute of Statistical Mathematics

³School of Fundamental Science and Technology, Keio University;
Now at The Institute of Statistical Mathematics

Wind direction is a typical and important angular variable in environmental science. Sometimes observations include wind speed and ozone concentration together with wind direction. This article is aimed at obtaining fundamental results in directional statistics that deal with modeling and analysis of data, including angular observations and a review of recent studies in this area.

The difference between the sample mean direction for angular data and the sample mean for linear data, and that between the mean resultant length and the variance, are emphasized. We use wind direction data available at the site of the National Institute for Environmental Studies to illustrate circular plots and calculations of basic statistics. Pewsey's test for symmetry, a flexible symmetric angular distribution by Jones and Pewsey, and the method of sine-skewing are presented, and an illustration of fitting angular distributions using the wind direction data is shown. Papakonstantinou and Batschelet distributions, which have the flat-topped and sharply-peaked properties, are introduced. The article also discusses circular-linear regression models derived from distributions on the cylinder, circular-circular regression models, and circular-circular structural models.