

角度データのための統計モデル

加藤 昇吾

統計数理研究所

2017年1月27日

本報告では、以下のレビュー論文の内容に基づき、角度データのための統計モデルを議論する：

加藤昇吾 (2017). 「円周上のコーシー分布と関連した統計モデル」. 『日本統計学会誌』, 46, 頁未定.

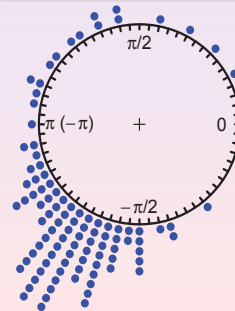
発表の流れ

- 1 角度データ
 - 角度データとは
 - 統計解析における問題点
- 2 角度データのための統計手法の基礎
 - 要約統計量
 - 確率分布
- 3 円周上のコーシー分布とその拡張
 - 円周上のコーシー分布
 - 2変量分布への拡張
- 4 まとめ

角度データとは

角度データ (円周上のデータ)

個々の観測が $[-\pi, \pi)$ の角度として表されるデータ.

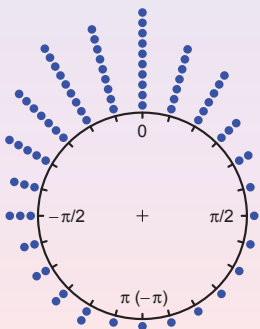


角度データの例

- 風向データ,
- 渡り鳥の移動方向のデータ.

(図) 1987年秋に記録された渡り鳥の移動方向を記録したデータ.

角度データの例

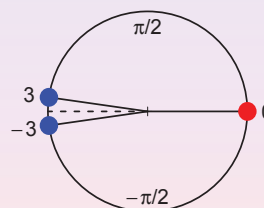


(図) 1987年から1998年に、ピッツバーグ(米国)で銃犯罪が発生した時刻のプロット. (Gill & Hangartner, 2010).

- 24時間を $[-\pi, \pi)$ の角度として表している. 例えば, 0は0時, π は正午に対応している.
- 1つの点は200件の犯罪を表す.

角度データの解析における問題点

角度データの統計解析においては、実数値データのための統計的手法をそのまま用いることができない場合がある.



【例1】標本平均

2つの角度 $(-3$ と $3)$ の実数値データにおける標本平均は

$$\frac{-3 + 3}{2} = 0$$

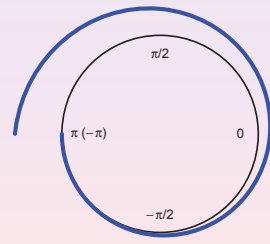
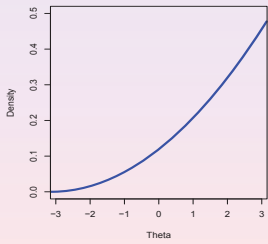
と計算できるが、0を“角度の平均”とするのは自然な定義とは言えない.

同様に、分散、歪度、尖度などの要約統計量も、そのままでは不自然な定義となってしまう.

【例2】確率分布

$(-\pi, \pi)$ 上のベータ分布を、角度データのための確率分布とみなすと...

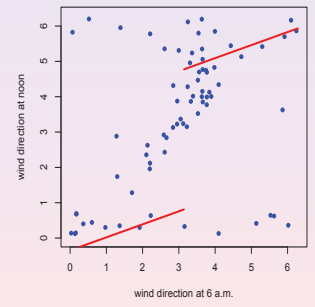
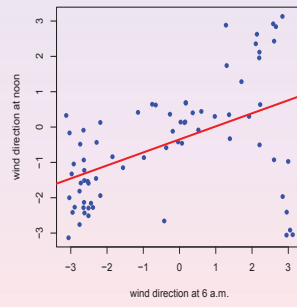
$$f(\theta) \propto (\pi + \theta)^{\alpha-1} (\pi - \theta)^{\beta-1}, \quad -\pi < \theta < \pi; \quad \alpha, \beta > 0.$$



(図) ベータ分布の密度関数 $f(\theta)$ のプロット ($\alpha = 3, \beta = 1$)。同じ密度関数が、(左) $[-\pi, \pi)$ 上、(右) 円周上、でプロットされている。

【例3】角度のペア $(\Theta_{1j}, \Theta_{2j})$ に以下の線形回帰モデルを当てはめる：

$$\Theta_{2j} = \beta_0 + \beta_1 \Theta_{1j} + \varepsilon_j, \quad j = 1, \dots, n.$$



(図) 午前6時 (θ_{1j}) と正午 (θ_{2j}) の風向のプロットと推定された回帰直線。角度のペアは、(左) $[-\pi, \pi)^2$ 、(右) $[0, 2\pi)^2$ 、で表されている。

角度データには、実数値データのための多くの統計的手法がそのままは使えない。

では、どのように統計解析すれば良いのか？

発表の流れ

- 1 角度データ
 - 角度データとは
 - 統計解析における問題点
- 2 角度データのための統計手法の基礎
 - 要約統計量
 - 確率分布
- 3 円周上のコーシー分布とその拡張
 - 円周上のコーシー分布
 - 2変量分布への拡張
- 4 まとめ

要約統計量

mean direction

$\Theta_1, \Theta_2, \dots, \Theta_n$: 角度の標本とする。

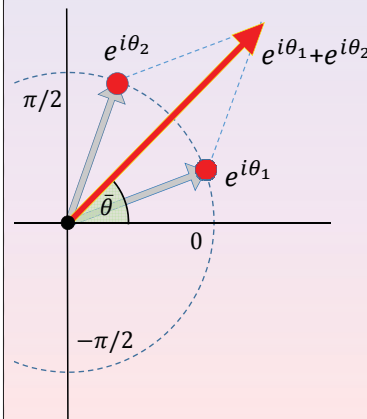
このとき、(sample) mean direction ((標本)平均方向) $\bar{\theta}$ は、

$$\bar{\theta} = \text{合成ベクトル} \left(\sum_{j=1}^n \cos \Theta_j, \sum_{j=1}^n \sin \Theta_j \right) \text{ の角度}$$

として定義される。また、同等な表現として、

$$\bar{\theta} = \arg \left(\sum_{j=1}^n e^{i\Theta_j} \right)$$

と表すこともできる。ここで、 i は虚数を表す。



mean direction

$$\bar{\theta} = \arg \left(\sum_{j=1}^n e^{i\Theta_j} \right).$$

(図) 2つの角度 θ_1, θ_2 の mean direction $\bar{\theta}$.

標本の集中度を表す要約統計量として、mean resultant length がしばしば用いられる。

mean resultant length

$\Theta_1, \Theta_2, \dots, \Theta_n$: 角度の標本とする。

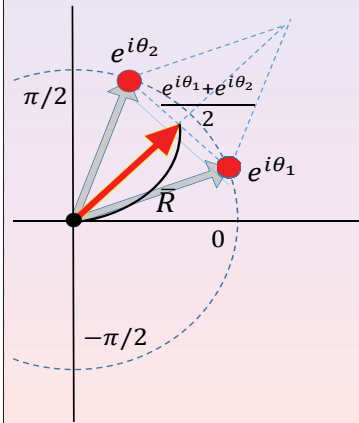
このとき、mean resultant length は、

$$\bar{R} = \frac{1}{n} \left| \sum_{j=1}^n e^{i\Theta_j} \right|$$

で定義される。

mean resultant length

$$\bar{R} = \frac{1}{n} \left| \sum_{j=1}^n e^{i\Theta_j} \right|$$



(図) 2つの角度 θ_1, θ_2 の mean resultant length \bar{R} .

その他の基本的な要約統計量をまとめた本に、Mardia & Jupp (1999) や Jammalamadaka & SenGupta (2001) がある。

角度データのための確率分布 (円周上の確率分布)

この発表では、連続型の円周上の確率分布を議論する。

確率密度関数

円周上の確率分布の密度関数 f を、以下を満たすものとして定義する:

- f は $[-\pi, \pi)$ 上の関数,
- $f(\theta) \geq 0, \quad -\pi \leq \theta < \pi,$
- $\int_{-\pi}^{\pi} f(\theta) d\theta = 1.$

円周上の密度関数 f としては、 $f(-\pi) = \lim_{\theta \rightarrow \pi} f(\theta)$ を満たすものを用いることが多い。

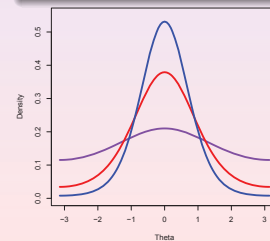
フォン・ミーゼス分布

フォン・ミーゼス分布, $\text{VM}(\mu, \kappa)$, は確率密度関数

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp \{ \kappa \cos(\theta - \mu) \}, \quad -\pi \leq \theta < \pi,$$

で定義される。ここで、 $-\pi \leq \mu < \pi, \quad \kappa \geq 0,$

$I_j(\cdot)$: 位数 j の第一種変形ベッセル関数。



基本的な性質

- 一山型,
- $\theta = \mu$ に関して対称,
- $\theta = \mu$ ($\mu + \pi$) で最大(最小).

(図) フォン・ミーゼス分布 $\text{VM}(0, \kappa)$ の密度関数. $\kappa = 0.3, \kappa = 1.2, \kappa = 2.1.$

フォン・ミーゼス分布は、'円周上の正規分布' と呼ばれ、角度データの (パラメトリックな) 統計解析において中心的な役割を果たしてきた。

長所

- 指数型分布族に属する,
- 最尤推定量が陽な形で表現できる,
- 統計的推測に関して成熟した理論が存在する.

短所

- 正規化定数が特殊関数を含む形である,
- よく知られた '尺度' の変換 ('scale' change) に閉じていない,
- 再生性が成り立たない.

wrapped 分布

X を実数値確率変数とする。

また、確率変数 Θ を以下のように定義する:

$$X \equiv \Theta \pmod{2\pi}.$$

このとき、 Θ の従う分布を、分布 X の wrapped (巻き込み) 分布と呼ぶ。

特に、確率変数 X が密度関数 $g(x)$ を持つとき、 Θ の密度関数は、

$$f(\theta) = \sum_{m=-\infty}^{\infty} g(\theta + 2\pi m), \quad -\pi \leq \theta < \pi,$$

で与えられる。

wrapped 分布の長所と短所

- 実数上の分布から直接的に、円周上の分布を得ることが可能。
- X が一山型（対称）の分布に従うとき、 X の wrapped 分布も一山型（対称）の分布となる。
- 一般に wrapped 分布の密度関数は、無限級数の形であらわされる。

例えば、 X が正規分布 $N(\mu, \sigma^2)$ に従うとき、 X の wrapped 分布は

$$f(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{m=-\infty}^{\infty} \exp\left\{-\frac{(\theta + 2\pi m - \mu)^2}{2\sigma^2}\right\}, \quad -\pi \leq \theta < \pi,$$

となる（wrapped 正規分布）。ここに、 $-\pi \leq \mu < \pi, \sigma > 0$ 。

trigonometric moments と関連した要約統計量

trigonometric moments

円周上に値をとる確率変数 Θ が、密度関数 $f(\theta)$ を持つとする。このとき、 Θ の p 次 trigonometric moment を

$$\varphi_{\Theta}(p) = E(e^{ip\Theta}) = \int_{-\pi}^{\pi} e^{ip\theta} f(\theta) d\theta, \quad p = 0, \pm 1, \dots,$$

と定義する。

特に、 Θ の mean direction と mean resultant length を、

$$\arg\{\varphi_{\Theta}(1)\} = \arg\{E(e^{i\Theta})\} \quad \text{and} \quad |\varphi_{\Theta}(1)| = |E(e^{i\Theta})|$$

とそれぞれ定義する。

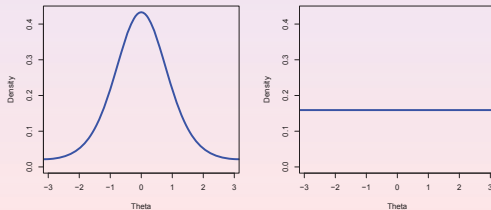
【和の分布に関する問題】 以下を仮定する：

$\Theta_1 \sim$ フォン・ミーゼス $vM(\mu, \kappa)$, $\Theta_2 \sim$ 円周上の一様分布。

また、 Θ_1 と Θ_2 は独立とする。このとき、

$$\Theta_1 + \Theta_2 \pmod{2\pi}$$

の分布はどうか？



(左) フォン・ミーゼス分布 $vM(0, 1.5)$, (右) 円周上の一様分布。

発表の流れ

- 1 角度データ
 - 角度データとは
 - 統計解析における問題点
- 2 角度データのための統計手法の基礎
 - 要約統計量
 - 確率分布
- 3 円周上のコーシー分布とその拡張
 - 円周上のコーシー分布
 - 2変量分布への拡張
- 4 まとめ

円周上のコーシー分布

この節では、以下の円周上の確率分布を議論する。

確率密度関数

円周上のコーシー分布（CC 分布）は、確率密度関数

$$f(\theta) = \frac{1}{2\pi} \frac{1 - \xi^2}{1 + \xi^2 - 2\xi \cos(\theta - \mu)}, \quad -\pi \leq \theta < \pi, \quad (1)$$

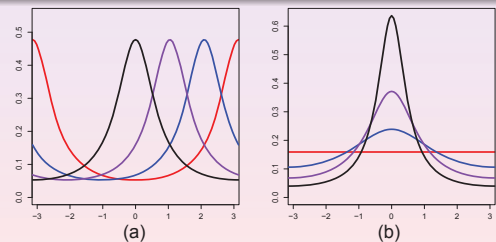
で定義される。ここに、 $-\pi \leq \mu < \pi, 0 \leq \xi < 1$ 。

以後、確率変数 Θ が密度関数 (1) を持つ分布に従うとき、 $\Theta \sim CC(\mu, \xi)$ とあらわすことにする。

$$f(\theta) = \frac{1}{2\pi} \frac{1 - \xi^2}{1 + \xi^2 - 2\xi \cos(\theta - \mu)}, \quad -\pi \leq \theta, \mu < \pi; 0 \leq \xi < 1.$$

密度関数の性質

- 一山型,
- 対称分布,
- μ : 位置パラメータ,
- ξ : 集中度パラメータ.



(図) CC 分布の密度関数のプロット。パラメータは以下の通り：

(a) $\xi = 0.5$ とし、 $\mu = 0, \pi/3, 2\pi/3, \pi$, (b) $\mu = 0$ とし、 $\xi = 0, 0.2, 0.4, 0.6$ 。

trigonometric moments

$\Theta \sim \text{CC}(\mu, \xi)$ とする. このとき,

$$E(e^{ip\Theta}) = (\xi e^{i\mu})^p, \quad p = 0, \pm 1, \dots$$

特に, Θ の mean direction は,

$$\begin{aligned} \arg\{E(e^{i\Theta})\} &= \arg\{E(\cos \Theta) + iE(\sin \Theta)\} \\ &= \mu, \end{aligned}$$

となり, Θ の mean resultant length は,

$$\begin{aligned} |E(e^{i\Theta})| &= \sqrt{\{E(\cos \Theta)\}^2 + \{E(\sin \Theta)\}^2} \\ &= \xi, \end{aligned}$$

で与えられる.

CC 分布の導出法

導出法 1 (wrapped コーシー分布)

確率変数 X が実数上のコーシー分布に従うとする.

$$f(x) = \frac{1}{\pi} \frac{\sigma}{\sigma^2 + (x - \eta)^2}, \quad -\infty < x, \eta < \infty, \sigma > 0.$$

また, Θ を以下のように定義する.

$$X \equiv \Theta \pmod{2\pi}.$$

このとき, $\Theta \sim \text{CC}(\mu, \xi)$ が成り立つ. ここに, $\eta \equiv \mu \pmod{2\pi}$, $\xi = e^{-\sigma}$.

前節で述べたように, 一般には wrapped 分布の密度関数は無限級数の形で表される.

導出法 2 (ブラウン運動)

$\{B_t; t \geq 0\}$: \mathbb{R}^2 値ブラウン運動,

$$B_0 = \xi(\cos \mu, \sin \mu)'$$

$$(-\pi \leq \mu < \pi, 0 < \xi < 1),$$

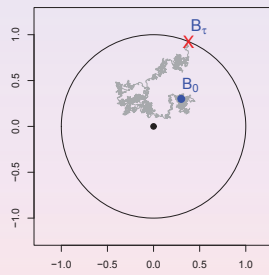
$$\tau = \inf\{t \geq 0; \|B_t\| = 1\}.$$

確率変数 Θ を次のように定義する.

$$\begin{pmatrix} \cos \Theta \\ \sin \Theta \end{pmatrix} = B_\tau.$$

このとき, 以下が成り立つ.

$$\Theta \sim \text{CC}(\mu, \xi).$$



(図) ブラウン運動.

パラメータ推定

$\Theta_1, \dots, \Theta_n \sim i.i.d. \text{CC}(\mu, \xi)$.

モーメント法による推定

1 次の trigonometric moment に基づくモーメント推定量:

$$\hat{\mu} = \arg\left(\sum_j e^{i\Theta_j}\right), \quad \hat{\xi} = \frac{1}{n} \left| \sum_j e^{i\Theta_j} \right|.$$

最尤法による推定

- $n \leq 4$ となるときは, 最尤推定量は陽な形であらわすことができる (McCullagh, 1996).
- 一般の n については, 最尤推定値は Kent & Tyler (1988) によるアルゴリズムを用いて数値的に求めることが可能である.

CC 分布のさらなる性質を調べる上で, 次のように確率変数とパラメータの変換を行うと便利である.

複素数を用いた表現 (McCullagh, 1996)

$\Theta \sim \text{CC}(\mu, \xi)$ とする. ここで,

$$Z = e^{i\Theta} (= \cos \Theta + i \sin \Theta), \quad \phi = \xi e^{i\mu},$$

とおく. このとき, Z の密度関数は次のようにあらわすことができる.

$$f(z) = \frac{1}{2\pi} \frac{1 - |\phi|^2}{|z - \phi|^2}, \quad |z| = 1; \quad |\phi| < 1. \quad (2)$$

以後, 確率変数 Z が密度関数 (2) を持つとき, $Z \sim C^*(\phi)$ と書くことにする.

CC 分布について, 以下の closure property が成り立つ:

1. 確率変数の積,

$$Z_1 \sim C^*(\phi_1), Z_2 \sim C^*(\phi_2), Z_1 \perp Z_2 \implies Z_1 Z_2 \sim C^*(\phi_1 \phi_2),$$

2. べき乗,

$$Z \sim C^*(\phi) \implies Z^n \sim C^*(\phi^n), \quad n = 1, 2, \dots,$$

3. メビウス変換,

$$Z \sim C^*(\phi) \implies \alpha \frac{Z + \beta}{\beta Z + 1} \sim C^*\left(\alpha \frac{\phi + \beta}{\beta \phi + 1}\right), \quad |\alpha| = 1, |\beta| < 1.$$

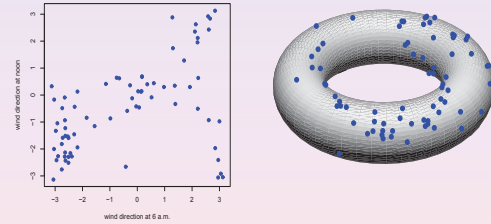
ちなみにこれらは, フォン・ミーゼス分布には成り立たない性質である.

発表の流れ

- 1 角度データ
 - 角度データとは
 - 統計解析における問題点
- 2 角度データのための統計手法の基礎
 - 要約統計量
 - 確率分布
- 3 円周上のコーシー分布とその拡張
 - 円周上のコーシー分布
 - 2変量分布への拡張
- 4 まとめ

2変量角度データ (トラス上のデータ)

2変量角度データ ... 2つの角度のペア $(\theta_{1j}, \theta_{2j})$ の集合 $(j = 1, \dots, n)$.



(図) ある観測所における午前6時と正午の風向を記録したデータ。
同じデータを、(左) $[-\pi, \pi]^2$ 上、(右) トラス上、でプロット。

トラスの持つ特殊な位相構造により、(従来の) 有界領域 $[a, b]^2$ 上の確率分布をそのまま当てはめるのは不自然である。

2変量角度データのモデル化のため、今までに主に以下の2つ確率分布の研究がなされてきた:

既存の確率分布

- 1 Mardia (1975) による分布
... 指数型分布族に属し、あるモーメントの条件の下でシャノンエントロピーを最大にする分布として導かれる。
[Submodels]:
Singh et al. (2002), Mardia et al. (2007), Kent et al. (2008).
- 2 Wehrly & Johnson (1980) による分布
... 特定の周辺分布を持つように特徴付けられた確率分布。
[Submodels]:
Shieh & Johnson (2005), Fernández-Durán (2007), Shieh et al. (2011).

研究の目的

本研究では、次のような性質を持つ2変量確率分布の提案を目指す:

- (i) 密度関数を陽な形であらわすことができる,
- (ii) 5つのパラメータがあり、それぞれの解釈が容易である,
- (iii) 周辺分布・条件付分布が共に扱いやすい確率分布となる,
- (iv) 2変量間の dependence の強さを幅広く調節することができる,
- (v) その他の統計的に扱いやすい性質を持っている.

CC分布の2変量分布への拡張

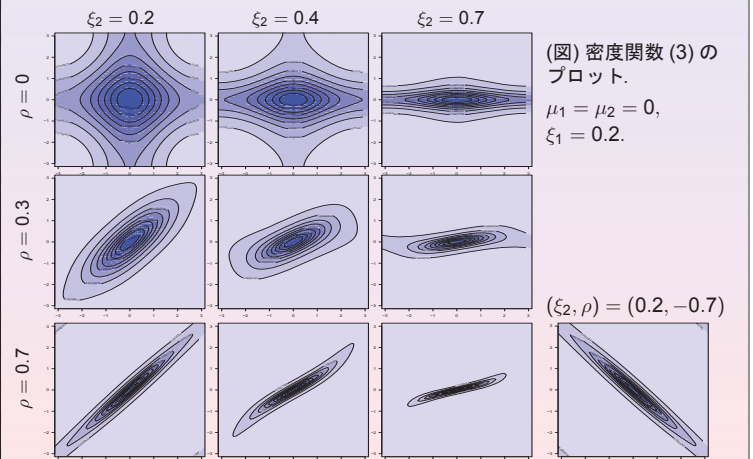
確率密度関数 (Kato & Pewsey, 2015)

新しいトラス上の確率分布を以下の確率密度関数によって定義する.

$$f(\theta_1, \theta_2) = \frac{(1-\rho^2)(1-\xi_1^2)(1-\xi_2^2)}{4\pi^2} \left\{ c_0 + c_1 \cos(\theta_1 - \mu_1) + c_2 \cos(\theta_2 - \mu_2) + c_3 \cos(\theta_1 - \mu_1) \cos(\theta_2 - \mu_2) + c_4 \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2) \right\}^{-1},$$

$$-\pi \leq \theta_1, \theta_2, \mu_1, \mu_2 < \pi, \quad 0 \leq \xi_1, \xi_2 < 1, \quad -1 < \rho < 1, \quad (3)$$

ここで、 c_0, \dots, c_4 は ξ_1, ξ_2, ρ を用いて陽にあらわされる.



(図) 密度関数 (3) のプロット。
 $\mu_1 = \mu_2 = 0,$
 $\xi_1 = 0.2.$

$(\xi_2, \rho) = (0.2, -0.7)$

周辺分布と条件付分布

$(\Theta_1, \Theta_2) \sim$ 密度関数 (3) とする。このとき、以下が成り立つ。

$$e^{i\Theta_1} \sim C^*(\xi_1 e^{i\mu_1}), \quad e^{i\Theta_2} \sim C^*(\xi_2 e^{i\mu_2}),$$

$$e^{i\Theta_1} | e^{i\Theta_2} \sim C^* \begin{pmatrix} -e^{i\mu_1} a_{11} e^{i\theta_2} + a_{12} \\ a_{21} e^{i\theta_2} + a_{22} \end{pmatrix},$$

$$e^{i\Theta_2} | e^{i\Theta_1} \sim C^* \begin{pmatrix} -e^{i\mu_2} a_{11} e^{i\theta_1} + a_{21} \\ a_{12} e^{i\theta_1} + a_{22} \end{pmatrix}.$$

ここに、 $C^*(\phi)$: パラメータ ϕ の CC 分布,

$$\tilde{\theta}_j = q(\theta_j - \mu_j), \quad q = \text{sgn}(\rho), \quad j = 1, 2,$$

$$a_{11} = \xi_1 \xi_2 - |\rho|, \quad a_{12} = |\rho| \xi_2 - \xi_1,$$

$$a_{21} = |\rho| \xi_1 - \xi_2, \quad a_{22} = 1 - |\rho| \xi_1 \xi_2.$$

擬似乱数の発生法

アルゴリズム

Step 1: $U_1, U_2 \sim i.i.d. U(0, 1)$.

Step 2: 以下の確率ベクトルを記録する。

$$(\Theta_1, \Theta_2) = (\mathcal{M}(2\pi U_1; \mu_1, \xi_1), \mathcal{M}\{\mathcal{M}(2\pi U_2; 2\pi q U_1, |\rho|); \mu_2, \xi_2\}),$$

ここで、 $q = \text{sgn}(\rho)$,

$$\mathcal{M}(\theta; \mu, \xi) = \mu + 2 \text{atan} \left(\frac{1 - \xi \tan \frac{\theta}{2}}{1 + \xi \tan \frac{\theta}{2}} \right).$$

このとき、 $(\Theta_1, \Theta_2) \sim$ 提案された分布 (3).

trigonometric moments

$(\Theta_1, \Theta_2) \sim$ 密度関数 (3) とする。このとき、

$$E(e^{i(p_1 \theta_1 + p_2 \theta_2)}) = \begin{cases} (\xi_1^q \eta_1)^{p_1} (\xi_2 \eta_2)^{p_2}, & qp_1, p_2 > 0, \\ \{\xi_1^{qp_1} \xi_2^{p_2} + a(qp_1, -p_2)\} \eta_1^{p_1} \eta_2^{p_2}, & qp_1 > 0, p_2 < 0, \\ \{\xi_1^{-qp_1} \xi_2^{-p_2} + a(-qp_1, p_2)\} \eta_1^{p_1} \eta_2^{p_2}, & qp_1 < 0, p_2 > 0, \\ (\xi_1^{-q} \eta_1)^{p_1} (\xi_2 \eta_2)^{p_2}, & qp_1, p_2 < 0, \\ \xi_1^{|p_1|} \eta_1^{|p_1|} \xi_2^{|p_2|} \eta_2^{p_2}, & p_1 = 0 \text{ or } p_2 = 0. \end{cases}$$

ここで、 $p_j \in \mathbb{Z}$, $\eta_j = e^{i\mu_j}$, $q = \text{sgn}(\rho)$, $j = 1, 2$,

$$a(p_1, p_2) = \frac{1 - \xi_2^2}{(p_2 - 1)!} \lim_{z \rightarrow 0} \frac{d^{p_2-1}}{dz^{p_2-1}} \left\{ \left(\frac{-a_{11}z - a_{12}}{a_{21}z + a_{22}} \right)^{p_1} \frac{1}{(z - \xi_2)(1 - \xi_2 z)} \right\}.$$

相関係数

Fisher & Lee (1983) による相関係数は、次のように定義される。

$$\rho_{FL} = \frac{\det\{E(X_1 X_1')\}}{[\det\{E(X_1 X_1')\} \det\{E(X_2 X_2')\}]^{1/2}}.$$

ここで、 $X_j = (\cos \Theta_j, \sin \Theta_j)'$, $j = 1, 2$.

提案した分布 (3) について ρ_{FL} を計算すると以下を得る。

$$\rho_{FL} = \frac{\rho}{1 - |\rho| \xi_1 \xi_2} \sqrt{\frac{1 - \xi_1^2}{1 + \xi_1^2} \cdot \frac{1 - \xi_2^2}{1 + \xi_2^2} \left\{ \frac{|\rho|(1 - \xi_1^2)(1 - \xi_2^2)}{1 - |\rho| \xi_1 \xi_2} + 2\xi_1 \xi_2 \right\}}.$$

同様に、Jupp & Mardia (1980) の相関係数や Johnson & Wehrly (1977) の相関係数も陽な形であらわすことが可能である。

パラメータ推定

$(\Theta_{11}, \Theta_{21}), \dots, (\Theta_{1n}, \Theta_{2n}) \sim i.i.d.$ 分布 (3).

モーメント推定

trigonometric moments に基づく推定を行うと、全てのパラメータの推定量を陽な形で表すことができる。

最尤法による推定

- CC 分布と同様、最尤推定値は数値的に求める必要がある。
- 私たちの simulation study によれば、
 - モーメント推定値は初期値として有用である。
 - 最尤推定のアルゴリズムは早く収束する。

まとめ

角度データの問題点

- 角度データに対しては、実数値データのための統計手法をそのまま使えないことがある。

角度データのための統計手法の基礎

- 基本的な要約統計量として、mean direction と mean resultant length を紹介した。
- よく知られた確率分布として、フォン・ミーゼス分布、wrapped 分布、CC 分布について概説した。

CC 分布の拡張

- 講演者の研究として、CC 分布の2変量拡張に関する結果を紹介。
- 提案した分布は、既存の分布にはない扱いやすい性質を持つ。

今後の発展性

(一般次元) 球面上の観測を含む多変量データ

$$S^{p_1} \times \dots \times S^{p_m} \times \mathbb{R}^q$$

の統計解析法. (S^p は p 次元球面.)

ご清聴ありがとうございました.

Appendix: 分布 (3) の導出法

提案された 2 変量分布 (3) は, 次のように導くことができる.

導出法 (Kato & Pewsey, 2015)

以下を仮定する:

$$(\Theta_U, \Theta_V) \sim BC(\rho) : \text{Kato (2009) による分布, } -1 < \rho < 1.$$

このとき, 次の確率ベクトルの分布は, 分布 (3) と等しくなる.

$$(\Theta_1, \Theta_2) = \left(\mu_1 + 2 \operatorname{atan} \left(\frac{1 - \xi_1}{1 + \xi_1} \tan \frac{\Theta_U}{2} \right), \mu_2 + 2 \operatorname{atan} \left(\frac{1 - \xi_2}{1 + \xi_2} \tan \frac{\Theta_V}{2} \right) \right).$$

ここに, $-\pi \leq \mu_1, \mu_2 < \pi, 0 \leq \xi_1, \xi_2 < 1$.

Kato (2009) による分布

Kato (2009) による分布は, 特別な場合 $BC(\rho)$ に以下の密度関数を持つ.

$$g(\theta_U, \theta_V) = \frac{1}{4\pi^2} \frac{1 - \rho^2}{1 + \rho^2 - 2|\rho| \cos(q\theta_U - \theta_V)}, \quad -\pi \leq \theta_U, \theta_V < \pi.$$

ここで, $q = \operatorname{sgn}(\rho), -1 < \rho < 1$.

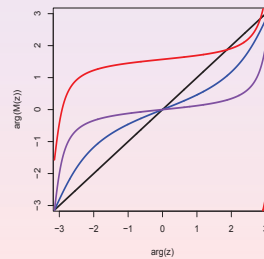
性質

- 周辺分布: 円周上の一様分布.
- 条件付分布: CC 分布.
- パラメータ ρ は, Θ_U と Θ_V の dependence の強さを調節.

メビウス変換

メビウス変換 $\mathcal{M}: [-\pi, \pi) \rightarrow [-\pi, \pi)$ は, 次のように定義される.

$$\mathcal{M}(\theta) = \mu + 2 \operatorname{atan} \left(\frac{1 - \xi}{1 + \xi} \tan \frac{\theta}{2} \right), \quad -\pi \leq \theta, \mu < \pi, 0 \leq \xi < 1.$$



パラメータ μ の解釈

μ は, 回転パラメータである.

パラメータ ξ の解釈

円周上の点は, $\theta = 0$ の方向に, ξ の強さで引き付けられる.

(図) メビウス変換 $\mathcal{M}(\theta)$ のプロット.

$$(\mu, \xi) =: (0, 0), (0, 0.4), (0, 0.8), (0.5\pi, 0.8).$$

複素数を用いると, メビウス変換は次のように表現可能である.

$$\mathcal{M}_c(e^{i\theta}) = e^{i\mu} \frac{e^{i\theta} + \xi}{\xi e^{i\theta} + 1}, \quad -\pi \leq \theta, \mu < \pi, 0 \leq \xi < 1.$$

CC 分布との関係

前に述べたように, 以下が成り立つ.

$$e^{i\theta} \sim C^*(\psi e^{i\eta}) \implies e^{i\mu} \frac{e^{i\theta} + \xi}{\xi e^{i\theta} + 1} \sim C^* \left(e^{i\mu} \frac{\psi e^{i\eta} + \xi}{\xi \psi e^{i\eta} + 1} \right).$$

ここで, $0 \leq \psi < 1, -\pi \leq \eta < \pi$. 特に $\psi = 0$ のときは,

$$e^{i\theta} \sim \text{円周上の一様分布} \implies e^{i\mu} \frac{e^{i\theta} + \xi}{\xi e^{i\theta} + 1} \sim C^*(\xi e^{i\mu}).$$

【再訪】 提案された 2 変量分布 (3) は, 次のように導くことができる.

導出法 (Kato & Pewsey, 2015)

以下を仮定する:

$$(\Theta_U, \Theta_V) \sim BC(\rho) : \text{Kato (2009) による分布, } -1 < \rho < 1.$$

このとき, 次の確率ベクトルの分布は, 分布 (3) と等しくなる.

$$(\Theta_1, \Theta_2) = \left(\mu_1 + 2 \operatorname{atan} \left(\frac{1 - \xi_1}{1 + \xi_1} \tan \frac{\Theta_U}{2} \right), \mu_2 + 2 \operatorname{atan} \left(\frac{1 - \xi_2}{1 + \xi_2} \tan \frac{\Theta_V}{2} \right) \right).$$

ここに, $-\pi \leq \mu_1, \mu_2 < \pi, 0 \leq \xi_1, \xi_2 < 1$.

参考文献 1

- [1] FERNÁNDEZ-DURÁN, J.J. (2007). Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics*, **63**, 579–585.
- [2] FISHER, N.I. & LEE, A.J. (1983). A correlation coefficient for circular data. *Biometrika*, **70**, 327–332.
- [3] GILL, J. & HANGARTNER, D. (2010). Circular data in political science and how to handle it. *Political Analysis*, **18**, 316–336.
- [4] JAMMALAMADAKA, S.R. & SENGUPTA, A. (2001), *Topics in Circular Statistics*, Singapore: World Scientific.

参考文献 2

- [5] JOHNSON, R.A. & WEHRLY, T.E. (1977). Measures and models for angular correlation and angular-linear correlation. *Journal of the Royal Statistical Society, Series B*, **39**, 222–229.
- [6] JUPP, P.E. & MARDIA, K.V. (1980). A general correlation coefficient for directional data and related regression problems. *Biometrika*, **67**, 163–173.
- [7] KATO, S. (2009). A distribution for a pair of unit vectors generated by Brownian motion. *Bernoulli*, **15**, 898–921.
- [8] 加藤昇吾 (2017). 「円周上のコーシー分布と関連した統計モデル」. 『日本統計学会誌』, **46**, 頁未定.

参考文献 3

- [9] KATO, S. & PEWSEY, A. (2015). A Möbius transformation-induced distribution on the torus. *Biometrika*, **102**, 181–190.
- [10] KENT, J.T., MARDIA, K.V. & TAYLOR, C.C. (2008). Modelling strategies for bivariate circular data. *LASR 2008 – The Art and Science of Statistical Bioinformatics*. Leeds University Press, Leeds.
- [11] KENT, J.T. & TYLER, D.E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics*, **15**, 247–254.
- [12] MARDIA, K.V. (1975). Statistics of directional data (with discussion). *Journal of the Royal Statistical Society, Series B*, **37**, 349–393.

参考文献 4

- [13] MARDIA, K.V. & JUPP, P.E. (1999), *Directional Statistics*, Chichester: Wiley.
- [14] MARDIA, K.V., TAYLOR, C.C. & SUBRAMANIAM, G.K. (2007). Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics*, **63**, 505–512.
- [15] MCCULLAGH, P. (1996). Möbius transformation and Cauchy parameter estimation. *The Annals of Statistics*, **24**, 787–808.
- [16] SHIEH, G.S. & JOHNSON, R.A. (2005). Inference based on a bivariate distribution with von Mises marginals. *Annals of the Institute of Statistical Mathematics*, **57**, 789–802.

参考文献 5

- [17] SHIEH, G.S., ZHENG, S., JOHNSON, R.A., CHENG Y.-F., SHIMIZU, K., WANG, C.-C. & TANG, S.-L. (2011). Modeling and comparing the organization of circular genomes. *Bioinformatics*, **27**, 912–918.
- [18] SINGH, H., HNIZDO, V. & DEMCHUK, E. (2002). Probabilistic model for two dependent circular variables. *Biometrika*, **89**, 719–723.
- [19] WEHRLY, T.E. & JOHNSON, R.A. (1980). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, **67**, 255–256.