

多次元トーラス上のコピュラモデル

加藤 昇吾 数理・推論研究系 准教授

はじめに

トーラス上のデータ

d 次元トーラス上のデータとは、 d 個の角度から成るベクトル $(u_1, u_2, \dots, u_d) (\in [0, 2\pi)^d)$ として表される観測の集合のことをいう。

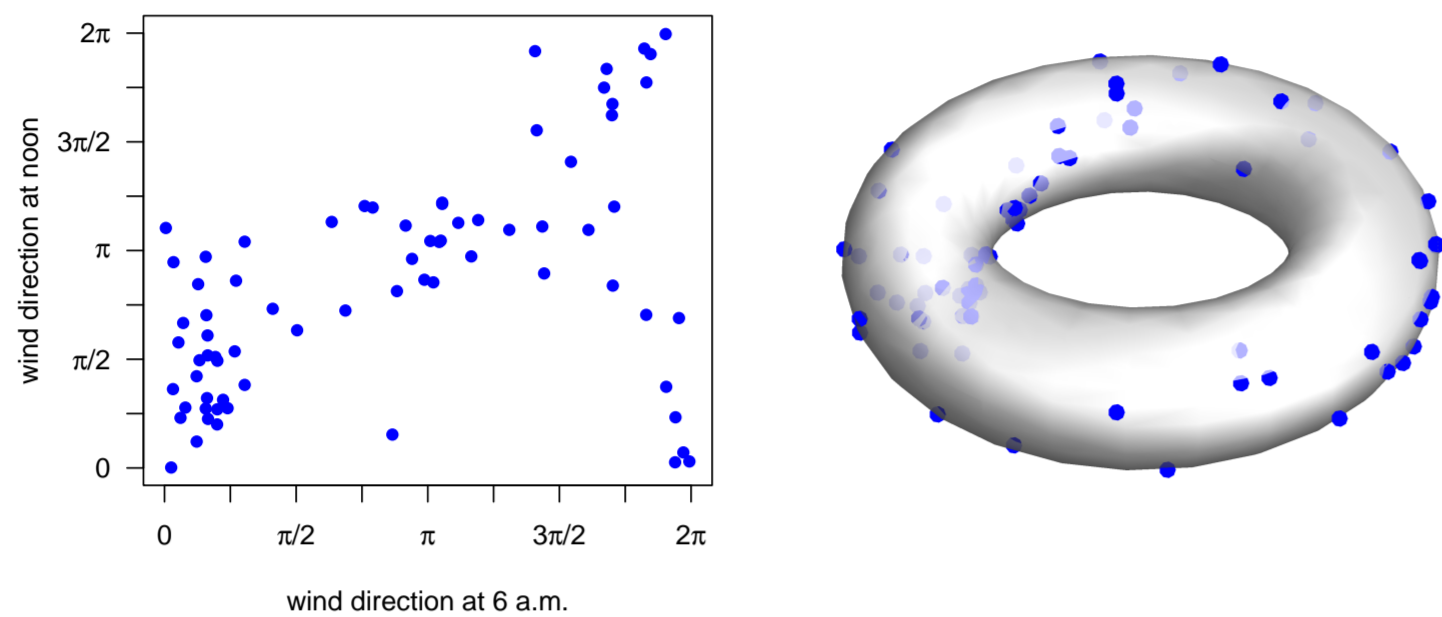


図1. ヒューストン（アメリカ）の気象観測所にて観測された午前6時と正午の風向データ。個々の観測が、左図では $[0, 2\pi)^2$ 上の点として、右図では2次元トーラス上の点として表されている。

コピュラ

d 次元コピュラとは、それぞれの1次元周辺分布が $[0, 1]$ 上の一様分布となる d 次元分布関数のことをいう。

d 次元コピュラを変換することにより、任意の1次元周辺分布関数を持つ d 次元分布関数を得ることが可能（スクラーの定理）。

⇒ コピュラにより、柔軟なモデリングが可能となる！

研究の目的

2次元トーラス上のコピュラとしては、Wehrly & Johnson (1980) のコピュラがよく知られているが、その一般次元トーラス上への拡張は知られていないようである。

そこで本研究では、Wehrly & Johnson (1980) によるコピュラを拡張し、多次元トーラス上のコピュラを提案する。

なお本研究は、Christophe Ley 准教授 (University of Luxembourg, Luxembourg) との共同研究である。

3次元トーラス上のコピュラモデル

本研究では、 d 次元トーラス上のコピュラを議論するが、はじめに3次元トーラス上のコピュラを考える。

定義

以下の確率密度関数を持つ3次元トーラス上の確率分布を提案する：

$$c(u_1, u_2, u_3) = c_2 \left[c_1 + 2 \{ \rho_{12} \cos(u_1 - u_2) + \rho_{13} \cos(u_1 - u_3) + \rho_{23} \cos(u_2 - u_3) \} \right]^{-1}, \quad (u_1, u_2, u_3) \in [0, 2\pi)^3. \quad (1)$$

ただし、

$$c_1 = \frac{\rho_{12}\rho_{13}}{\rho_{23}} + \frac{\rho_{12}\rho_{23}}{\rho_{13}} + \frac{\rho_{13}\rho_{23}}{\rho_{12}},$$

$$c_2 = \frac{1}{(2\pi)^3} \left\{ \left(\frac{\rho_{12}\rho_{13}}{\rho_{23}} \right)^2 + \left(\frac{\rho_{12}\rho_{23}}{\rho_{13}} \right)^2 + \left(\frac{\rho_{13}\rho_{23}}{\rho_{12}} \right)^2 - 2\rho_{12}^2 - 2\rho_{13}^2 - 2\rho_{23}^2 \right\}^{1/2}.$$

また、パラメータの条件として、 $\rho_{12}, \rho_{13}, \rho_{23} \in \mathbb{R} \setminus \{0\}$, $\rho_{12}\rho_{13}\rho_{23} > 0$, そして、ある (i, j, k) に対して $|\rho_{kl}| < |\rho_{jk}\rho_{je}| / (|\rho_{jk}| + |\rho_{je}|)$ を仮定する。

分布(1)の性質

modes と antimodes 確率密度関数(1)が最大となるのは、

- (i) $u_j = u_k = u_\ell$ if $\rho_{kl} < \rho_{jk}\rho_{je}/|\rho_{jk} + \rho_{je}|$, $\rho_{jk} + \rho_{je} < 0$,
- (ii) $u_j = u_k + \pi = u_\ell + \pi$ if $\rho_{kl} < \rho_{jk}\rho_{je}/|\rho_{jk} + \rho_{je}|$, $\rho_{jk} + \rho_{je} > 0$,
- (iii) $u_j = u_k + \pi = u_\ell$ if $\rho_{kl} > \rho_{jk}\rho_{je}/(\rho_{jk} - \rho_{je})$.

のときである。また、最小となるのは、 $u_j = u_k = u_\ell + \pi$ のときである。

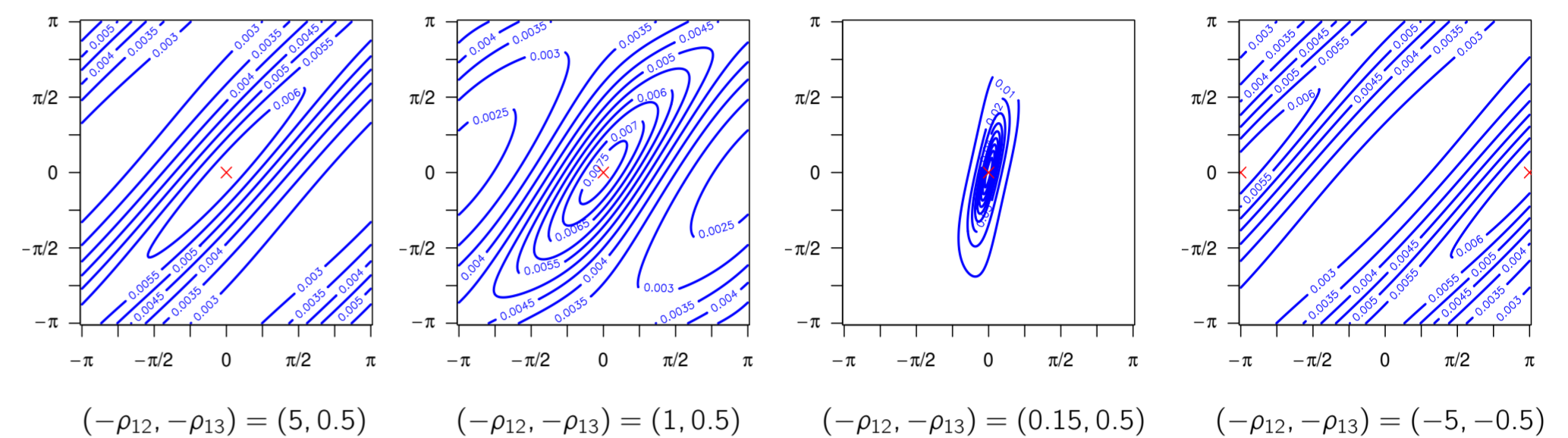


図2. 確率密度関数(1)の等高線プロット。ただし、 $(u_3, \rho_{23}) = (0, 0.1)$ であり、 x 軸、 y 軸はそれぞれ u_1, u_2 を表す。また、'x' は mode を表す。

周辺分布・条件付き分布

$(U_1, U_2, U_3) \sim$ 分布(1) とする。このとき、以下が成り立つ：

- U_j の周辺分布： $[0, 2\pi)$ 上の一様分布。

$$c(u_j) = \frac{1}{2\pi}, \quad 0 \leq u_j < 2\pi.$$

- (U_j, U_k) の周辺分布： Wehrly-Johnson コピュラのサブモデル。

$$c(u_j, u_k) = \frac{1}{4\pi^2} \frac{|1 - \delta_{jk}^2|}{1 + \delta_{jk}^2 - 2\delta_{jk} \cos(u_j - u_k - \eta_{jk})}, \quad 0 \leq u_j, u_k < 2\pi.$$

ここで、 $\delta_{jk} = |\phi_{jk}|$, $\eta_{jk} = \arg(\phi_{jk}) \in \{0, \pi\}$, $\phi_{jk} = \{ \rho_{je}\rho_{ke}/\rho_{jk} - \rho_{jk}\rho_{je}/\rho_{ke} - \rho_{jk}\rho_{ke}/\rho_{je} - (2\pi)^3 c_2 \} / (2\rho_{jk})$.

- 条件付き分布についても、下記のように知られた分布となる：

$(U_j, U_k) | U_\ell = u_\ell \sim$ Kato & Pewsey (2015) の分布,

$U_j | U_k = u_k \sim$ wrapped Cauchy 分布,

$U_j | (U_k, U_\ell) = (u_k, u_\ell) \sim$ wrapped Cauchy 分布.

上述の結果を用いることにより、3つの独立な $(0, 1)$ 上の一様乱数から、提案した分布(1)に従う確率ベクトルを棄却なく生成することが可能。

d 次元トーラス上のコピュラモデル

定義 d 次元トーラス上の分布を、以下の確率密度関数で定義する：

$$c(u_1, \dots, u_d) \propto \left\{ c_1 + 2 \sum_{1 \leq j < k \leq d} \rho_{jk} \cos(u_j - u_k) \right\}^{-1}, \quad 0 \leq u_j < 2\pi. \quad (2)$$

ここで、 $\rho_{jk} \in \mathbb{R}$, $c_1 \geq 2 \sum_{1 \leq j < k \leq d} |\rho_{jk}|$.

性質 $(U_1, \dots, U_d) \sim$ 分布(2) とする。このとき、以下が成立する：

- U_j の周辺分布： $[0, 2\pi)$ 上の一様分布。
- $U_j | (U_1, \dots, U_{j-1}, U_{j+1}, \dots, U_d) = (u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_d)$ の条件付き分布： wrapped Cauchy 分布。

参考文献

1. Wehrly, T.E. & Johnson, R.A. (1980). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, **67**, 255–256.
2. Kato, S. & Pewsey, A. (2015). A Möbius transformation-induced distribution on the torus. *Biometrika*, **102**, 359–370.