２変量角度データのための確率分布

加藤　昇吾　　数理・推論研究系　助教

研究の目的

２変量角度データ

２変量角度データとは，２つの角度のペア (θ_1, θ_2) $(-\pi \le \theta_1, \theta_2 < \pi)$ として表される観測の集合のことをいう．

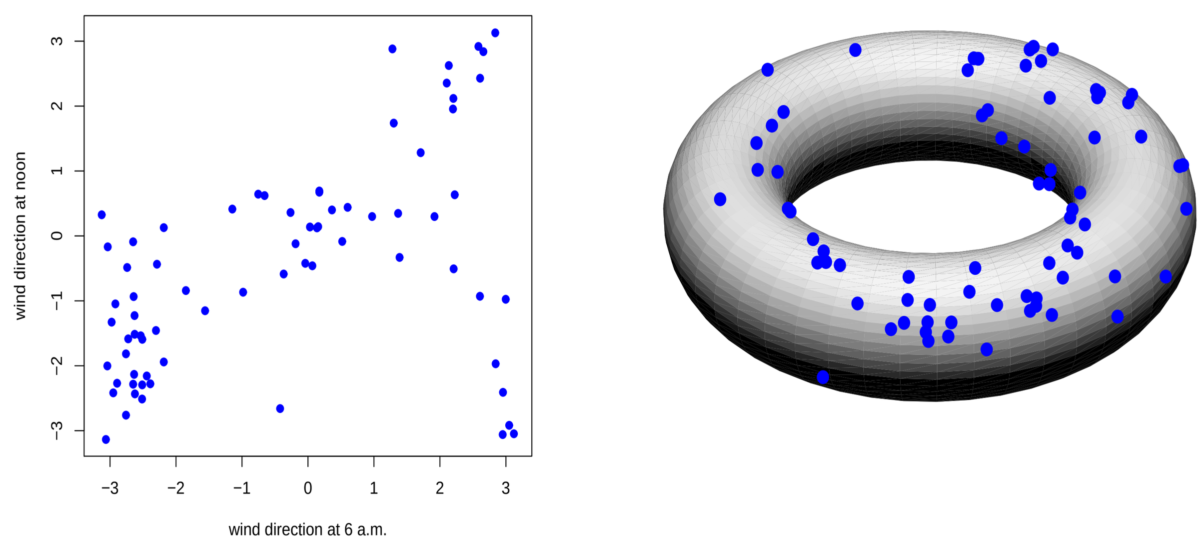

図１．ヒューストン（アメリカ）の気象観測所にて観測された午前６時と正午の風向データ．個々の観測が，左図では $[-\pi,\pi)^2$ 上の点として，右図ではトーラス上の点として表わされている．

２変量角度データのための確率分布

角度には周期性があるため，２変量角度データに（周期性を持たない）２変量確率分布をそのまま当てはめても，満足な当てはめは期待できない．

このような背景の下で，Mardia *et al.* (2007)，Shieh *et al.* (2011) らが２変量角度データのための確率分布を提案．

$\Longrightarrow$ しかし，モデルの解釈・解析的な扱いやすさなどに問題．

研究の目的

本研究では，以下の性質を持つ確率分布の提案を目指す：

i. 確率密度関数が陽な形で表わされる，

ii. ５つのパラメータを持ち，それぞれの解釈が明確である，

iii. 周辺分布・条件付分布がよく知られた分布になる，

iv. 変量間のdependenceを幅広く調節することができる．

なお，本研究はArthur Pewsey (University of Extremadura) との共同研究である．

２変量角度データのための確率分布

以後便利のため，角度 θ を複素平面における単位円周上の点 $\cos\theta + i\sin\theta$ として表すことにする．ここで，i は虚数をあらわす（$i^2=-1$）．

分布の定義

Ω：複素平面上の単位円，

$U_1, U_2 \sim i.i.d.$ Ω上の一様分布，

とする．このとき，新たな確率分布を

$$(Z_1, Z_2) = \left(\eta_1 \frac{V_1 + \xi_1}{\xi_1 V_1 + 1}, \eta_2 \frac{V_2 + \xi_2}{\xi_2 V_2 + 1} \right), \qquad (1)$$

で定義する．ここに，$V_1 = U_1$，$V_2 = (U_2 + r)/(\overline{r} U_2 + 1)$，$r = |\rho| U_1^{\mathrm{sgn}(\rho)}$，$\eta_1, \eta_2 \in \Omega$，$0 \le \xi_1, \xi_2 < 1$，$-1 < \rho < 1$，$\mathrm{sgn}(x) = 1$ $(x \ge 0)$，$\mathrm{sgn}(x) = -1$ $(x < 0)$．

確率密度関数

確率ベクトル (Z_1, Z_2) の確率密度関数は，

$$f(z_1, z_2) = \frac{1}{4\pi^2} \frac{(1-\rho^2)(1-\xi_1^2)(1-\xi_2^2)}{|a_{11}(\overline{z}_1\eta_1)^q\, z_2\overline{\eta}_2 + a_{12}(\overline{z}_1\eta_1)^q + a_{21} z_2\overline{\eta}_2 + a_{22}|^2}, \quad z_1, z_2 \in \Omega,$$

で与えられる．ここで，$a_{11} = \xi_1\xi_2 - \psi$，$a_{12} = \psi\xi_2 - \xi_1$，$a_{21} = \psi\xi_1 - \xi_2$，$a_{22} = 1 - \psi\xi_1\xi_2$，$\psi = |\rho|$，$q = \mathrm{sgn}(\rho)$．

パラメータの解釈

５つのパラメータのうち，η_j は Z_j の位置を調節し，ξ_j は Z_j の集中度を調節し，ρ は Z_1 と Z_2 のdependenceを調節する役割を持つ（$j = 1, 2$）．

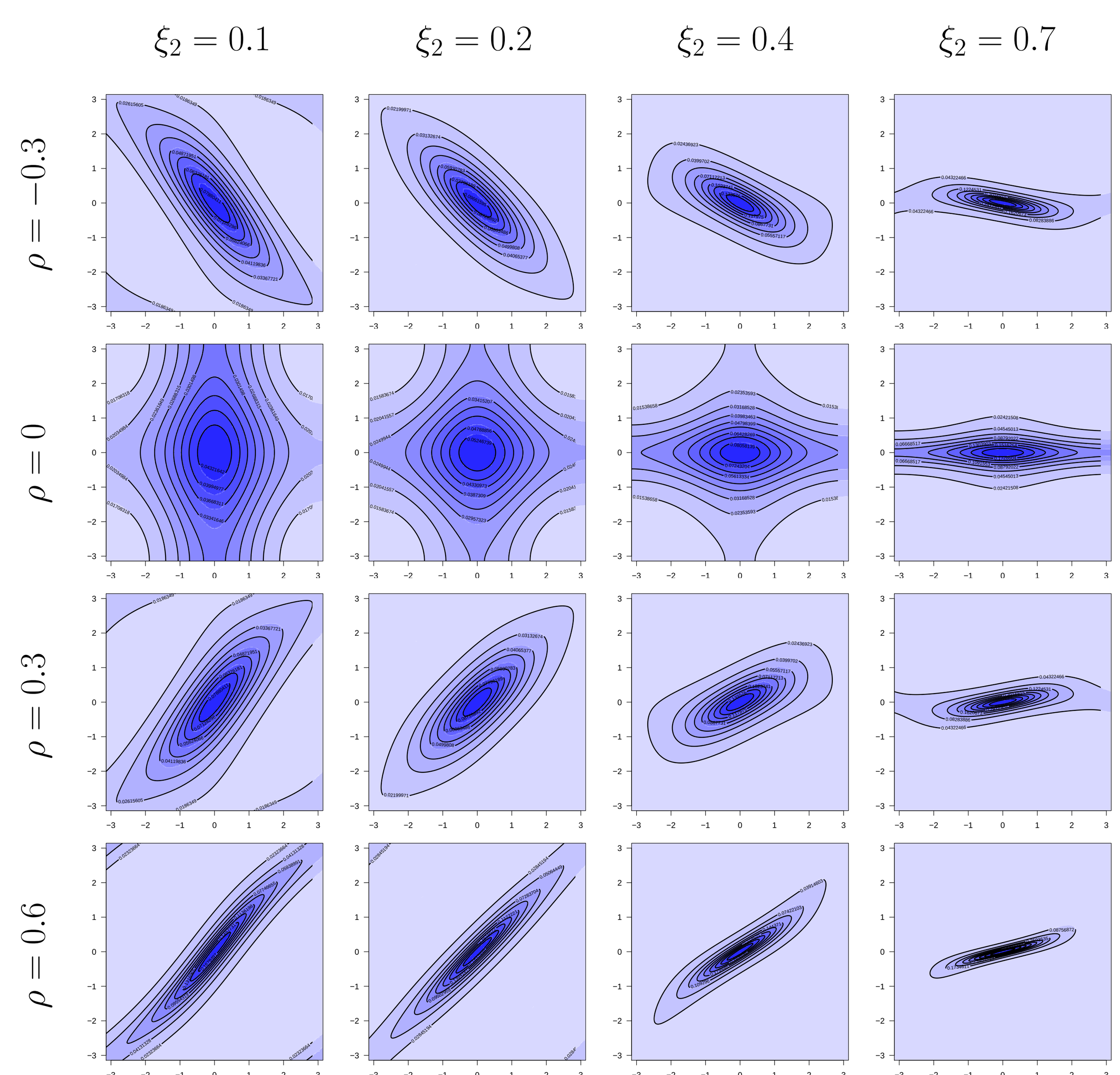

図２．確率密度関数 (2)（$\arg(\eta_1) = \arg(\eta_2) = 0$，$\xi_1 = 0.2$）のプロット．

周辺分布と条件付分布

確率ベクトル (Z_1, Z_2) が，分布 (2) に従っているとする．このとき，

$$Z_1 \sim C^*(\xi_1\eta_1), \quad Z_2 \sim C^*(\xi_2\eta_2),$$

$$Z_1 \,|\, z_2 \sim C^* \left(-\eta_1 \frac{a_{11}\tilde{z}_2 + a_{12}}{a_{21}\tilde{z}_2 + a_{22}} \right), \quad Z_2 \,|\, z_1 \sim C^* \left(-\eta_2 \frac{a_{11}\tilde{z}_1 + a_{21}}{a_{12}\tilde{z}_1 + a_{22}} \right),$$

が成り立つ．ここに，$\tilde{z}_j = (z_j\overline{\eta}_j)^q$，$j = 1, 2$. また，$C^*(\phi)$ は以下の確率密度関数を持つwrapped Cauchy分布を表す．

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

References

[1] 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.

[2] Shieh, G.S., Zheng, S., Johnson, R.A., Chang, Y.-F., Shimizu, K., Wang, C.-C. & Tang, S.-L. (2011). Modeling and comparing the organization of circular genomes. *Bioinformatics* **27**, 912–918.