統計数理研究所 75 周年記念事業（2019）
©2019 統計数理研究所

角度の観測を含むデータのための統計モデル

Statistical Models for Data Which Include Angular Ones

数理・推論研究系　加藤 昇吾（Shogo Kato）

1. はじめに

様々な学問分野において，角度として表される観測値が得られることがある．例えば，気象学における風向の観測はその一例である．風向は，西を $-\pi$ とし，反時計回りを正の向きとすれば，南を $-\pi/2$，東を 0，北を $\pi/2$ のように角度で表すことができる．つまり，任意の風向は $-\pi$ 以上 π 未満の角度 θ，もしくは円周上の点 $(\cos\theta, \sin\theta)$，として表現できる．他にも，医学・地震学・動物行動学・生命情報学など，多くの分野で角度の観測が存在している．

角度の観測を含むデータには，統計解析をする上で大きな問題がある．それは，このようなデータを解析する上では，統計学が主に対象としている実数値データのための解析手法をそのまま使うことができないという問題である．例えば，平均や分散などの要約統計量や，実数値データのための確率分布・回帰モデル・時系列モデルなどの統計モデルは，角度データに直接応用すると，しばしば不自然な解析結果を与えることにつながってしまう．この問題を解決するため，角度の観測を含むデータの統計的手法を考えることが統計学における重要な研究テーマとなっている．本報告では，このテーマにおける研究の背景と著者の研究結果を概観する．

2. 角度データのための確率分布

方向統計学における中心的なアプローチは，データに何らかの確率分布を仮定したパラメトリックな統計解析法である．著者は，角度データのための確率分布（円周上の確率分布）の 1 つとして知られる「円周上のコーシー分布」に着目し，これに関連したパラメトリックな統計モデルを研究してきた．円周上のコーシー分布は確率密度関数

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

で定義される確率分布である．ここで，μ は分布の位置を定めるパラメータ，ρ は分布の集中度を調節するパラメータである．

方向統計学においては，円周上のコーシー分布は古くから知られていたもののほとんど注目されていない分布であった．この分野で長きにわたり研究の中心だったのは，「フォン・ミーゼス分布」とよばれる円周上の確率分布およびそれを応用したパラメトリックな統計手法である．フォン・ミーゼス分布は，実数上の正規分布からの自然なアナロジーにより導かれることから，「円周上の正規分布」とよばれることもある．一方，正規分布が持ついくつかの扱いやすい性質が成立しない問題点も指摘されており，そのことがフォン・ミーゼス分布およびそれを応用した統計手法の理論的性質を導くことを困難にしている一面もあった．そのような中，McCullagh (1996) などによって開拓された円周上のコーシー分布に著者は興味を持ち，この分布に関連した統計モデルの研究を行うようになった．

3．角度の観測を含むデータのための統計モデル

著者は，円周上のコーシー分布に関連した統計モデルの論文を今までに複数篇執筆してきたが，ここでは，その中から 2 篇についてある程度詳しく説明し，4 篇について簡潔に紹介する．

Kato (2010) では，円周上のコーシー分布を誤差分布として用いた新たな円周上のマルコフ過程を提案した．この研究は Fisher and Lee (1994) によるマルコフ過程と関連がある．彼らは，回帰曲線としてメビウス変換，誤差分布としてフォン・ミーゼス分布を仮定した円周上のマルコフ過程を提案した．それに対し，Kato (2010) では，誤差分布としてフォン・ミーゼス分布の代わりに円周上のコーシー分布を用いることにより，Fisher and Lee (1994) のモデルでは得られなかった多くの挙動に関する性質を得ることに成功した．例えば，任意時点 t の角度を所与としたときの $t+h$ 時点における角度の条件付分布が円周上のコーシー分布となることや，その条件付分布のパラメータも複素数を用いれば簡潔に表現することができることを示した（t と h は自然数をあらわす）．

Kato and Pewsey (2015) では，円周上のコーシー分布の拡張として，2 次元トーラス $[-\pi,\pi)^2$ 上の分布を提案した．2 次元トーラス上の分布の既存研究としては，フォン・ミーゼス分布を拡張した 2 変量フォン・ミーゼス分布がよく知られている．この分布は，条件付分布がフォン・ミーゼス分布となる利点があるが，正規化定数が特殊関数の無限和で表されることや，パラメータの解釈が困難であること，周辺分布がよく知れていない分布になること，などの問題点がある．それに対して，Kato and Pewsey (2015) で提案した 2 次元トーラス上の分布ではこれらの問題がすべて解決されたモデルとなっている．

その他の研究結果について簡潔に紹介する．Kato *et al.* (2008) では説明変数・被説明変数が共に角度となる回帰モデルを提案し，誤差分布として円周上のコーシー分布を用いることで回帰モデルのいくつかの扱いやすい性質を得られることを示した．Kato and Jones (2010, 2015) では，円周上のコーシー分布を特別な場合として含む柔軟な円周上の確率分布を提案した．Kato and McCullagh (2018) では，円周上のコーシー分布を球面上の分布へと拡張し，メビウス変換に関して閉じていることやパラメータ推定が容易にできることを示した．

参 考 文 献

Fisher, N. I. and Lee, A. J. (1994). Time series analysis of directional data, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **56**, 327–339.

Kato, S. (2010). A Markov process for circular data, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72**, 655–672.

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*, **102**, 249–262.

Kato, S. and Jones, M. C. (2015). A tractable and interpretable four-parameter family of unimodal distributions on the circle, *Biometrika*, **102**, 181–190.

Kato, S. and McCullagh, P. (2018). Möbius transformation and a Cauchy family on the sphere, *arXiv:1510.07679v2*, 1–30.

Kato, S. and Pewsey, A. (2015). A Möbius transformation-induced distribution on the torus, *Biometrika*, **102**, 359–370.

Kato, S., Shimizu, K. and Shieh, G. S. (2008). A circular–circular regression model, *Statistica Sinica*, **18**, 633–645.

McCullagh, P. (1996). Möbius transformation and Cauchy parameter estimation, *The Annals of Statistics*, **24**, 787–808.