

平成 29 年 6 月 19 日現在

機関番号：62603

研究種目：基盤研究(C) (一般)

研究期間：2013～2016

課題番号：25400218

研究課題名(和文) 角度の観測を含む多変量データのための統計解析法

研究課題名(英文) Statistical methods for multivariate data which include angular observations

研究代表者

加藤 昇吾 (Kato, Shogo)

統計数理研究所・数理・推論研究系・准教授

研究者番号：60468535

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：風向や分子のねじれ角は観測値が角度として表され、このような角度の観測は様々な学問分野において存在している。そして、角度の観測を含むデータの多くは、他の変数と共に観測される多変量データである。本研究では、角度の観測を含む多変量データのための統計的手法を研究した。具体的には、トーラス上の確率分布、角度データのための混合効果モデル、実数値と角度が混在するデータのためのベイジアンネットワーク、などを提案し、それらの統計的性質を得た。

研究成果の概要(英文)：In various academic fields, it is often required to handle measurements which can be expressed as angles. Examples of such measurements include wind directions in meteorology and torsional angles in molecular biology. In practice, it is common that angular measurements are recorded with other linear and/or circular measurements. In our study, we have studied statistical methods for multivariate data which include angular observations. Specifically, we have proposed the following three models, among others, and obtained their statistical properties: families of distributions on the general-dimensional torus, a random effect model for circular data, and a Bayesian network model for linear-circular data.

研究分野：統計数学

キーワード：統計数学 方向統計学 多変量解析

## 1. 研究開始当初の背景

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、北を0とし、時計回りを正の向きとすれば、東を  $\pi/2$ 、南を  $\pi$ 、西を  $3\pi/2$  のように角度で表すことができる。つまり、任意の風向は0以上  $2\pi$  未満の角度、もしくは円周上の点  $(\cos \theta, \sin \theta)$  として表現できる。他には、バイオインフォマティクスにおける分子のねじれ角も同様に、角度の観測と解釈することができる。その他、医学・地震学・動物行動学などの分野においても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、このようなデータの解析する上では、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。この問題は、角度(または円周)には周期性があり、その位相が実数の位相と異なっていることに起因している。

このような背景のもとで、現在に至るまで角度の1変量データの統計解析法に関する多くの研究がなされてきた。それらの研究は例えば、Mardia and Jupp (1999) や Jammalamadaka and SenGupta (2001) などの書籍にまとめられている。

## 2. 研究の目的

「1. 研究開始当初の背景」で述べたように、過去には角度の1変量データの統計解析法に関する多くの研究がなされてきた。

しかし、角度の1変量データの方法論のみでは現実には不十分であることが多い。なぜなら、実際に得られるデータは、角度のみの1変量データではなく、他の変数(実数値や角度など)と共に観測される多変量データがほとんどであるためである。特に近年では、観測機器などの発達により、多くの変数を持つデータが得られるようになり、多変量データを解析するための統計解析法が以前にも増して必要とされている。

このような現状にもかかわらず、角度の観測を含む多変量データのための統計的手法に関する研究は限られたものしかない。その中から統計モデルに関するいくつかの結果をここで挙げることにする。以下、 $p$ 次元トーラス上の点( $p$ 変量の角度)を  $(\theta_1, \dots, \theta_p)$ 、 $q$ 次元ユークリッド空間  $\mathbb{R}^q$  上の点を  $(x_1, \dots, x_q)$  とする。Johnson and Wehrly (1978) はシリンダー  $(\theta_1, x_1)$  上の確率分布を提案し、関連した回帰モデルを提案した。近年ではバイオインフォマティクスとの関連から、2次元トーラス  $(\theta_1, \theta_2)$  上の確率分布が、Mardia et al. (2007)、Kent et al. (2008)、Shieh et al. (2011) らによって提案されている。研

究代表者も2次元トーラス  $(\theta_1, \theta_2)$  上の確率分布を研究してきた(Kato, 2009; Kato and Pewsey, 研究開始時改訂中, 2015年に出版)。角度の観測を含む一般次元のデータのモデルは、研究代表者の知る限り、Mardia et al. (2008)による  $p$ 次元トーラス  $(\theta_1, \dots, \theta_p)$  上の分布しかない。

このように、角度の観測を含む多変量データのための統計的手法に関する研究は限られたものしかない。また、角度の観測を含む2変量データのための統計的手法については、ある程度の研究結果があるものの、既存のモデルではうまく当てはめることができないデータが多く存在することが知られている。

さらに、2つより多くの変数を持つ分布としては、Mardia et al. (2008)の  $p$ 次元トーラス  $(\theta_1, \dots, \theta_p)$  上の分布しか知られておらず、彼らの分布も、モデルの解釈・統計的性質・推測に関する考察が十分になされておらず、実用的なモデルとは言い難い。このような現実の中、実用性と数学的な扱いやすさを併せ持つモデルが望まれるところである。本研究ではこれらの問題を解決すべく、角度の観測を含む多変量データのための統計的手法の発展を目指した。

## 3. 研究の方法

本研究では、角度の観測を含む多変量データのための統計的手法の発展に取り組んだ。ここでは、3つの具体的な研究テーマについて、それぞれどのような方法で研究を行ったかを説明する。

1つめの研究テーマは、新たな2次元トーラス  $(\theta_1, \theta_2)$  上の分布の提案およびその統計的性質の考察である。提案した分布は、周辺分布が一様分布となる分布、つまりコピュラ、である。コピュラは任意の連続な1変量周辺分布を持つように変換することができ、柔軟な確率分布族を得るために用いられる。分布の提案においては、trigonometric moment による分布の特徴づけを用いた。周辺分布が一様分布となる trigonometric moment の必要十分条件を求めることにより、周辺分布が一様分布となる2次元トーラス  $(\theta_1, \theta_2)$  上の分布を導いた。

2つめの研究テーマは、新たな一般次元トーラス  $(\theta_1, \dots, \theta_p)$  上の分布に関する研究である。1つめの研究テーマと同様、今回の分布も全ての1変量周辺分布が円周上の一様分布となる性質を持つ分布(コピュラ)の提案を行った。分布を導く方法として平面上のランダムウォーク問題を考えることにより、周辺分布が一様分布となる分布が得られるよう工夫を行った。そして得られた分布について統計的性質を調べた。

最後に、3つめの研究テーマとして、シリンダー  $(\theta_1, \dots, \theta_p, x_1, \dots, x_q)$  上のデータの

ためのベイジアンネットワークを研究した。今回は、それぞれのノードに対して最大1つの親ノードを持つ、木構造のベイジアンネットワークモデルを対象とした。解釈が容易かつ解析的に扱いやすいモデルを得るため、親ノードと子ノードの同時分布として、2つの実数のペアの分布には2変量正規分布、2つの角度のペアの同時分布には Kato and Pewsey (2015)の分布、1つの角度と1つの実数のペアの分布には Johnson and Wehrly (1978)の分布の特別な場合、を仮定した。このベイジアンネットワークモデルについて、モデルの解釈、相互情報量の計算、統計的推測などを考察した。

その他、角度の観測を含む多変量データのための統計的手法に関する研究として、シリンダー( $\theta_1, \dots, \theta_p, x_1, \dots, x_q$ )上の確率分布および角度データのための混合効果モデルについても研究を行った。詳細は、「4. 研究成果」で述べる。

#### 4. 研究成果

以下、(1) 2013年度、(2) 2014年度、(3) 2015年度、(4) 2016年度の研究成果についてそれぞれ述べていく。

(1) 1年目にあたる2013年度には、一般次元トーラス( $\theta_1, \dots, \theta_p$ )上の確率分布の提案とその統計的性質・推測の考察を行った。この研究にあたっては、2次元トーラス( $\theta_1, \theta_2$ )上の分布を議論した Wehrly and Johnson (1980)および研究代表者の共著論文(Kato and Pewsey, 当時改訂中, 2015年に出版)を参考にした。本研究では、はじめに Wehrly and Johnson (1980)の2次元トーラス上の分布(以下、WJ分布)が特別な場合として周辺分布が一樣分布となる確率分布となる点、および、WJ分布は周辺分布が一樣となる特別な場合を確率変数変換することにより得られる点に着目した。そして、周辺分布が一樣分布となるWJ分布の特別な場合を一般次元へと拡張し、周辺分布が一樣分布となる一般次元トーラス( $\theta_1, \dots, \theta_p$ )上の分布を導いた。さらに、この分布を確率変数変換することにより、特定の周辺分布を持つ一般次元トーラス上の分布を得た。特に周辺分布が円周上のコーシー分布となる場合について考察を行い、この分布は Kato and Pewsey (当時改訂中)の2次元トーラス分布の一般次元への拡張であることが明らかになった。

(2) 2年目にあたる2014年度は、2013年度に議論した一般次元トーラス( $\theta_1, \dots, \theta_p$ )上の確率分布に関連した統計的手法を議論した。また、Kato and Pewsey (当時掲載予定, 2015年に出版)とは異なるアプローチによって導出される2次元トーラス( $\theta_1, \theta_2$ )上の確率分布を提案し、その統計的性質を考

察した。提案した分布は trigonometric moment を特徴づけることによって導かれ、周辺分布が共に一樣分布となる性質を持つ(つまり、コピュラとなる)。この分布は trigonometric moment の与え方によって、WJ分布等、いくつかの既存の分布を特別な場合として含むことを示した。また、提案した分布のいくつかのサブモデルについて、密度関数が陽な形で表されること、パラメータ解釈が容易であること、擬似乱数発生が容易であること、などを明らかにした。また、この分布の一般次元トーラス( $\theta_1, \dots, \theta_p$ )上の確率分布への拡張と関連した統計的手法を検討した。

(3) 3年目にあたる2015年度は、3つの研究テーマに取り組んだ。

1つは、2014年度に提案した2次元トーラス( $\theta_1, \theta_2$ )上の確率分布について、引き続き研究を行った。具体的には、この分布の新たなサブモデルを提案し、その統計的性質を考察した。そして、そのサブモデルについて、確率密度関数が陽な形で表されること、パラメータの解釈が容易であること、過去のサブモデルとは異なり非対称な確率密度関数を持つこと、などの性質を明らかにした。

2つめの研究テーマは、新たな一般次元トーラス( $\theta_1, \dots, \theta_p$ )上の分布の提案とその統計的性質に関する考察である。新たな分布は、平面上のランダムウォーク問題を考えることにより提案した。そしてこの分布が、周辺分布がすべて一樣分布となること、binding functionの設定により条件付分布をよく知られた分布になるようにできることを明らかにした。さらに、trigonometric moment が簡潔な形で表現できることを示した。そして、この分布の特別な場合として、条件付分布がフォンミーゼス分布となるサブモデルの統計的推測を議論した。具体的には、モーメント推定量が陽な形で表されること、EMアルゴリズムを用いて最尤推定値を求めることが可能であるなどの理論的結果を得た。また、シミュレーションにより、数値的にこれらの推定量を比較した。

最後に3つめの研究テーマとして、2015年度に取り組んだ上記2つの確率分布から導かれる一般次元シリンダー( $\theta_1, \dots, \theta_p, x_1, \dots, x_q$ )上の確率分布を提案した。上記の2つのテーマにおける2次元および一般次元トーラス上の分布は、すべての周辺分布が一樣分布となる性質を持つ(つまり、コピュラとなる)。コピュラ理論の結果を応用することにより、確率変数変換を通して2次元および一般次元シリンダー上の分布を得られることを示した。そして、この得られた2次元および一般次元シリンダー上の分布について統計的性質を考察した。

(4) 4年目にあたる2016年度は、3つの研究テーマに取り組んだ。

1つめの研究テーマは、2014、2015年度に組み込んだ2次元トーラス( $\theta_1, \theta_2$ )上の確率分布の新たなサブモデルの研究である。このサブモデルについて、trigonometric moment、密度関数が非負の値をとるためのパラメータの条件、密度関数のmodeとantimodeに関する結果を得た。また、提案したサブモデルは、パラメータの値を変えることにより、非対称性と不等分散性を表現できることを明らかにした。

2つめの研究として、一般次元シリンダー( $\theta_1, \dots, \theta_p, x_1, \dots, x_q$ )上のデータのためのベイジアンネットワークを研究した。角度と実数値の分布に既存の分布を仮定し、2つの変数間の関係を測る情報量として相互情報量を用いることにより、ベイジアンネットワークを構成することに成功した。そしてこのモデルは、パラメータ推定が容易であること、相互情報量の計算が容易であること、擬似乱数を効率良く発生できること、などの扱いやすい理論的性質を持つことを明らかにした。また、シミュレーションにより、これらの理論的性質を検証した。

最後に、角度データのための混合効果モデルに関する研究を行った。過去には、被説明変数が角度、説明変数が複数の角度となる回帰モデルは提案されていたが、混合効果を加えたモデルに関しては、研究代表者の知る限り提案されていなかったようである。この研究では、被説明変数が角度、説明変数が複数の角度または実数値となる混合効果モデルを提案し、その統計的性質に関する結果を得た。具体的には、パラメータの解釈が容易であること、observed Fisher informationが比較的簡潔に表せること、同一クラスター内の相関係数が陽に表せること、EMアルゴリズムを用いてパラメータを推定できること、などの統計的性質を明らかにした。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計10件)

- (1) Kato, S. and Eguchi, S., Robust estimation of location and concentration parameters for the von Mises-Fisher distribution. *Statistical Papers*, 査読有, 57, 205-234, 2016, doi: 10.1007/s00362-014-0648-9.
- (2) Pewsey, A. and Kato, S., Parametric bootstrap goodness-of-fit testing for Wehrly-Johnson bivariate circular distributions. *Statistics and Computing*, 査読有, 26, 1307-1317, 2016, doi: 10.1007/s11222-015-9605-2.
- (3) 加藤 昇吾, 円周上のコーシー分布と関

連した統計モデル. 査読有, 日本統計学会誌, 46, 85-111, 2016, URL: <http://www.terrapub.co.jp/journals/jjssj/frame/46.html>.

- (4) Kato, S. and Pewsey, A., A Möbius transformation-induced distribution on the torus. *Biometrika*, 査読有, 102, 359-370, 2015, doi:10.1093/biomet/asv003.
- (5) Kato, S. and Jones, M.C., A tractable and interpretable four-parameter family of unimodal distributions on the circle. *Biometrika*, 査読有, 102, 181-190, 2015, doi:10.1093/biomet/asu059.
- (6) Jones, M.C., Pewsey, A. and Kato, S., On a class of circular copulas for circular distributions. *Annals of the Institute of Statistical Mathematics*, 査読有, 67, 843-862, 2015, doi: 10.1007/s10463-014-0493-6.
- (7) Takahashi, A., Sugimoto, H., Kato, S., Shiroishi, T. and Koide, T., Mapping of genetic factors that elicit intermale aggressive behavior on Mouse Chromosome 15: intruder effects and the complex genetic basis. *PLOS ONE*, 査読有, 10(9), e0137764, 2015, doi:10.1371/journal.pone.0137764.
- (8) Kato, S., Ishii, A., Nishi, A., Kuriki, S. and Koide, T., Segregation of a QTL cluster for home-cage activity using a new mapping method based on regression analysis of congenic mouse strains. *Heredity*, 査読有, 113, 416-423, 2015, doi:10.1038/hdy.2014.42.
- (9) Kato, S. and McCullagh, P., Conformal mapping for multivariate Cauchy families. *arXiv*, 査読無, arXiv:1510.07679, 2015, URL: <https://arxiv.org/abs/1510.07679>.
- (10) Kato, S. and McCullagh, P., A characterization of a Cauchy family on the complex space. *arXiv*, 査読無, arXiv:1402.1905, 2014, URL: <https://arxiv.org/abs/1402.1905>.

[学会発表](計23件)

- (1) 加藤昇吾, 角度データのためのコピュラ. 統計数理セミナー, 2017年3月8日, 統計数理研究所, 立川.
- (2) Kato, S., The wrapped Cauchy distribution on the circle and its bivariate extension. Keio International Symposium "Statistical Analysis for High-Dimensional, Circular or Time Series Data", 2017年3月3日, Keio University, Yokohama, Japan.

- (3) Kato, S., Some properties of a family of distributions on the sphere related to the Möbius transformation. ISI-ISM-ISSAS Joint Meeting 2017, 2017年2月7日, Indian Statistical Institute, Delhi, India.
- (4) 加藤昇吾, 角度データのための統計モデル. 科研費基盤研究(A)シンポジウム「統計科学の現代的課題」, 2017年1月27日, 金沢大学, 金沢.
- (5) 加藤昇吾, A tractable and interpretable four-parameter family of unimodal distributions on the circle. 2016年度統計関連学会連合大会, 2016年9月5日, 金沢大学, 金沢.
- (6) Kato, S. and Jones, M.C., A tractable and interpretable four-parameter family of unimodal distributions on the circle. ISI-ISM-ISSAS Joint Conference 2016, 2016年2月2日, Academia Sinica, Taipei, Taiwan.
- (7) Kato, S., A family of distributions for bivariate circular data. ISM Symposium on Environmental Statistics 2016, 2016年1月15日, The Institute of Statistical Mathematics, Tachikawa, Japan.
- (8) 加藤昇吾, 角度データの統計解析法. データサイエンス特別講義, 2015年12月16日, 慶應義塾大学, 横浜.
- (9) 加藤昇吾, 角度データのための統計的手法. データ科学特論 special lecture, 2015年11月27日, 大阪大学, 大阪.
- (10) Kato, S., A tractable and interpretable four-parameter family of unimodal distributions on the circle. Waseda International Symposium "High Dimensional Statistical Analysis for Spatio-Temporal Processes & Quantile Analysis for Time Series", 2015年11月10日, Waseda University, Tokyo, Japan.
- (11) Kato, S., The wrapped Cauchy distribution on the circle and its bivariate extension. Seminar, 2015年9月16日, The Open University, Milton Keynes, UK.
- (12) 加藤昇吾, 円周上のコーシー分布と関連した統計モデル. 2015年度統計関連学会連合大会 企画セッション: 日本統計学会各賞受賞者記念講演, 2015年9月8日, 岡山大学, 岡山.
- (13) Kato, S., A family of copulas for circular data and its extension. International Symposium on Dependence and Copulas 2015, 2015年6月23日, The Institute of Statistical Mathematics, Tachikawa, Japan.
- (14) 加藤昇吾, 多変量コーシー分布と等角写像. 統計数理セミナー, 2015年4月8日, 統計数理研究所, 立川.
- (15) Kato, S., McCullagh, P., Some properties of a family of distributions on the sphere related to the Möbius transformation. The 7th International Conference of the ERCIM WG on Computational and Methodological Statistics (ERCIM 2014), 2014年12月7日, Conference Center, Pisa, Italy.
- (16) 加藤昇吾, The wrapped Cauchy distribution on the circle and its bivariate extension. 応用統計ワークショップ, 2014年10月24日, 東京大学, 東京.
- (17) Kato, S., A tractable and interpretable four-parameter family of unimodal distributions on the circle. Seminar, 2014年9月19日, University of Valladolid, Valladolid, Spain.
- (18) 加藤昇吾, 複素空間上のコーシー分布. 統計数理セミナー, 2014年6月4日, 統計数理研究所, 立川.
- (19) Kato, S. and Jones, M.C., A tractable and interpretable four-parameter family of unimodal distributions on the circle. Advances in Directional Statistics International Workshop, 2014年5月21日, Free University of Brussels and the Atomium, Brussels, Belgium.
- (20) Kato, S. and Pewsey, A., A Möbius transformation-induced distribution on the torus, Workshop "Advances and Applications in Distribution Theory", 2014年1月14日, The Institute of Statistical Mathematics, Tachikawa, Japan.
- (21) Kato, S. and Jones, M.C., A family of circular distributions related to wrapped Cauchy distributions via Brownian motion. 2013 Joint Statistical Meetings, 2013年8月8日, Montreal, Canada.
- (22) 加藤昇吾, 回帰分析を用いたマウス活動量データの遺伝解析. 回帰分析を用いたマウス活動量データの遺伝解析, 新領域融合センタープロジェクト「生命システム融合研究」プロジェクト会議, 2013年7月24日, 国立情報学研究所, 東京.
- (23) 加藤昇吾, 回帰分析を用いたマウス活動量データの遺伝解析. 統計数理セミナー, 2013年7月10日, 統計数理研究所, 立川.

〔その他〕

ホームページ等

(1) 加藤昇吾 個人ホームページ

<http://www.ism.ac.jp/~skato/>

(2) researchmap (加藤昇吾、日本語)

<http://researchmap.jp/read0138993/?lang=japanese>

(3) researchmap (加藤昇吾、英語)

<http://researchmap.jp/read0138993/?lang=english>

## 6. 研究組織

(1) 研究代表者

加藤 昇吾 (Kato, Shogo)

統計数理研究所・数理・推論研究系・准教授

研究者番号：60468535

(2) 研究分担者

なし

(3) 連携研究者

なし

(4) 研究協力者

なし