

令和 3 年 5 月 12 日現在

機関番号：23803

研究種目：若手研究

研究期間：2018～2020

課題番号：18K13459

研究課題名（和文）生態・環境科学分野における角度データに着目した解析のための統計モデリング

研究課題名（英文）Statistical modeling for circular data in ecological and environmental science

研究代表者

井本 智明（Imoto, Tomoaki）

静岡県立大学・経営情報学部・助教

研究者番号：20749296

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本研究では、角度や時間のように周期性のある現象に着目したデータに対して幅広い統計解析を可能とするための確率モデルを構成した。具体的には、現象の発生傾向に柔軟に対応できる円周上分布、大きな観測値の発生を想定に入れた分析が可能なシリンダー上分布、計算コストの少ないトーラス上分布の構成法を提案した。これらの確率モデルを、地震発生データや2組のバクテリアゲノムの共通遺伝子位置に関するデータへ当てはめることで、地震発生傾向の推測やバクテリアの共通性に関する考察を得ることができた。

研究成果の学術的意義や社会的意義

動物の移動方向や風向などの周期性のある現象に関するデータは、正規分布やポアソン分布のように直線上で定義される確率分布ではなく、円周上で定義される確率分布を用いて解析を行う必要がある。生態・環境科学分野ではこのようなデータが多くみられる。本研究で提案した柔軟かつ情報損失の少ない円周上分布、トーラス上分布、シリンダー上分布を用いることで、適切な生態動向推定や自然環境予測が可能となることが期待される。

研究成果の概要（英文）：In this research project, we constructed a probabilistic model to enable a flexible statistical analysis for data focusing on periodic phenomena such as angle and time. Specifically, we proposed a method for constructing circular distributions that can give flexible fitting to various circular data, a cylinder distribution that can model long-tailed for its linear part, and torus distributions that can be used with a low computational cost. Applications to seismic data and shared genes position data of the bacterial genomes show the tendency of big earthquakes and commonality of bacteria which has the same ancestor, respectively,

研究分野：統計科学

キーワード：方向統計学 確率分布 数理統計学

1. 研究開始当初の背景

生態・環境科学分野での因果関係の推定や将来の予測、そしてリスクの評価は自然災害や疫病による被害を軽減し、また効率的な環境保全による利益をもたらす重要な学問である。しかし、地震や異常気象の例でみられるように、関心のある現象を解析するためのデータサンプル数は必ずしも多いとは限らず、大標本を基にしたノンパラメトリック手法による解析では信頼性を持った結果を導けない恐れがある。また、このようなデータの測定に用いる機器は自然の中に野晒し状態にあるため、計測誤差が生じやすくなる問題もある。そのため、角度データならば 5°刻みにするなど、観測値を丸めたデータを用いた解析を行うことも多い。このように、生態・環境科学分野における推測は精密でなければならない一方で、データのサンプル数や観測精度に限りがあることから、モデル選択や統計的推測に信頼性を持たせることが困難となる。これらの問題を解決するための統一的な手法を、本研究では角度データに着目した統計モデリングを基に確立した。

2. 研究の目的

角度データは生態・環境科学分野では頻繁にみられ、特に、渡り鳥の移動方向と移動距離、風向と風速、地震発生日点間の角度と距離のように、周期性を持つ変量(角度変量)と実数値全体を取る変量(直線上変量)が混在したデータとして現れる。このようなデータを用いて、角度変量と直線上変量の関係性を表すモデルの構成や類似度の比較などを通じた将来予測やリスク評価を行うことで、環境保護や災害対策に役立たせることができる。角度変量を含んだデータを解析する際、その周期性を反映させるために円周上やトーラス上で定義される確率モデルを用いる必要があり、さらに直線上変量と同時に解析するためにはシリンダー上で定義される確率モデルを用いる必要がある。本研究では、角度変量を含んだデータに対し、柔軟かつ精密な統計的解析を行うための確率モデルの構成と統計的推測法の提案を目指した。

3. 研究の方法

本研究では、角度を含んだデータ分析のためのモデルの提案を行った。そのために、(1) 柔軟な当てはめが可能な円周上分布の構成法についての研究、(2) 丸めた角度観測データに適用すべき円周上分布の構成法とその推定についての研究、(3) 裾の重い直線上変量を考慮に入れたシリンダー上分布についての研究、(4) 計算コストの少ないトーラス上分布の構成法についての研究、の4つを重点的に行った。以下がそれぞれの具体的な研究方法である。

- (1) 現象の発生頻度について様々な特徴を捉えることが可能な円周上確率分布を構成することで、情報の少ない現象に対しても柔軟な解析が可能となる。そのために、角度変量の要約統計量である平均方向とその集中度、そして分布の非対称性を表す円周歪度とピークの鋭さを表す円周尖度を自由に設定できる円周上分布を構成した。これらの4つの指標は角度変量の特性関数を用いて定義され、また特性関数は複素平面上で閉曲線を描く性質がある。そこで、複素平面上の閉曲線をまず構成し、その図形が表す関数を特性関数としてもつ円周上分布を構成した。
- (2) 観測が 5°刻みになっているような丸めた観測値からなるデータの分析に利用できるモデルを構成するため、円周上分布の離散化法を提案した。そのために正弦関数や余弦関数の周期性を利用した構成法を考えることで、複雑な積分や和による表現を必要としない確率分布の離散化法を提案した。この方法によって、古くから利用されている対称な円周上分布や、歪度や尖度に柔軟な円周上分布から構成される離散化円周上分布を導出した。また導出した分布を通じたモンテカルロシミュレーションによって、丸められた円周上データに対して用いるべき分布やその適用法を提案した。
- (3) 直線上変量が大きな値を取り得ることを想定した分析を可能とするため、裾の重い直線上分布と円周上分布の同時確率分布を構成した。そのために、指数分布やその一般化分布の混合分布を考えることで裾の重い確率分布が構成可能であることを利用し、先行研究にある Abe and Ley (2017) で提案されたワイブル分布を直線上分布としてもつシリンダー上分布に対する一般化ガンマ混合分布を考えることで、目的となる確率分布を導出した。
- (4) 多くの変量を持ったデータに対しても高速で分析可能なモデルを考案するために、分布関数のような計算コストのかかる関数を利用せず、周辺分布を特定の円周上分布として指定できるトーラス上分布の構成法を提案した。そのために、角度変量に関する同時分布 $f(t_1, t_2) = c(t_1, t_2) * g(t_1) * g(t_2)$ は、 $c(t_1, t_2)$ が奇関数のような性質をもつ関数であるとき、その周辺分布が $g(t)$ となることを利用した構成法を提案した。

4. 研究成果

[2018 年度]

1 年目には研究の方法(3)にあるシリンダー上分布の構成についての研究を実施した。ここで提案した分布は、直線上変量に対する周辺分布が裾の重いデータにも対応できるパレート分布となる。その結果、提案モデルによる分析は、少数の極端なデータのみに影響されない頑健なものとなる。例えば、図1では2011年東日本大震災発生までの72時間に起きたマグニチュード4以上の地震と連続した地震発生地点間のなす角度についての散布図と、Abe and Ley (2017)による分布と提案分布による当てはめ結果を示している。図をみてわかるように、既存分布では1つのサンプルに大きく影響を受けた当てはめを与えており、角度の観測に対して非対称性がみられる結果となっている。一方、提案分布では180°の角度を中心とした対称性がみられる結果となっている。そのため、提案分布による分析からは連続する大きな地震の発生地点は正反対の位置で起こりやすくなるということがわかる。

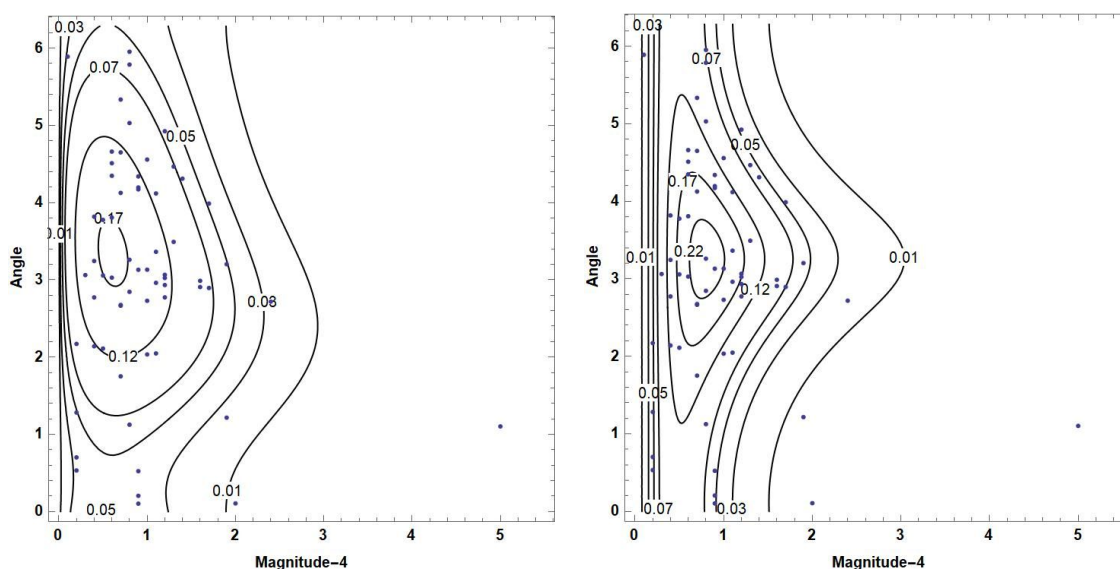


図1. 2011年東日本大震災発生までの72時間におきた地震に関する散布図と

既存分布による当てはめ結果(左図)と提案分布による当てはめ結果(右図)

その他にも、提案分布のモーメントや分布関数は初等関数で表現されるため、直線上変量と角度変量の関係性についての分析として、平均に関する回帰や、外れ値に影響されづらい中央値に関する回帰が容易に実行できる。また、提案分布に含まれるパラメータはそれぞれが解釈可能となっているため、各パラメータに対する回帰分析によって有用な情報の抽出が可能となる。

[2019 年度]

2 年目には、まず研究の方法(1)にある柔軟な円周上分布の構成についての研究を実施した。そこでは複素平面上的閉曲線に対して回転・縮小・移動の操作を行い、得られた図形が表す関数を特性関数として持つ分布の構成法を提案した。この方法で構成される円周上分布の平均・集中度・円周歪度・円周尖度は先に述べた回転・縮小・移動の操作に関連した構造で表される。また、特性関数自体がシンプルな関数で表されるため、その性質を利用したモデル推定法も提案している。この提案法による確率分布を用いることで、平均に関するモデリングだけでなく、集中度・円周歪度・円周尖度に影響する要因分析のためのモデルの実装が可能となることが期待される。

次に研究の方法(2)にある円周上分布の離散化法に関する研究を実施した。提案した手法を用いて、カージオイド分布、巻き込みコーシー分布、フォンミーゼス分布のような対称円周上分布や、Fernandez-Duran (2004)の分布やKato and Jones (2015)の分布のような非対称な円周上分布の離散化を考え、モンテカルロシミュレーションを通して、それらの離散化分布は最尤法よりも最小カイ二乗法を用いる方がモデルの推定が効率的であることを示し、同時にデータが18分割のときまでは元の連続円周上分布よりも離散化分布の方が当てはめのパフォーマンスがよいことを示している。また、提案法による Fernandez-Duran (2004)の分布の離散化は、モーメントを保存する性質を持っていることから、情報の損失を失うことなくデータの分析を行うことができる。この性質を利用することで観測が荒い場合に得られるようなデータの分析にも、情報を十分に利用した分析ができるようになることが期待される。

[2020 年度]

3 年目には研究の方法(4)のシンプルなトラス上分布の構成法についての研究を実施した。ここでの提案法によって構成されたトラス上分布は、任意の円周上分布を周辺分布としてもつことができ、かつ中程度までの相関を持つデータに対してよい当てはめを与えることができる。また、複雑な関数を用いることのないシンプルなモデル構成となっていることから、統計的推定や検定を容易に行うことができ、変量間の相関をコントロールするパラメータが他の平均や分散パラメータとは独立になっているため、情報損失の少ない分析が可能である。ここで提案した確率分布を2組のバクテリアゲノムの遺伝子共通位置に関するデータへ当てはめ、その相関を推定することで、バクテリアの共通祖先の推測に役立つであろうことを示した。この構成法をシリンダー上分布に対する手法として拡張することで、2つ以上の直線上変量と円周上変量からなるデータへの応用が可能な多変量シリンダー上分布の構成もできる。この拡張法によって構成されるモデルを用いることで、直線上変量と円周上変量のペアデータに対する時系列分析も可能となることが期待され、風速と風向、地震の発生地点間の距離と角度のようなデータに応用することで、自然エネルギーの活用や災害のリスク評価につなげることが可能となる。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Imoto Tomoaki, S. Shieh Grace, Shimizu Kunio	4. 巻 123
2. 論文標題 Discrete Circular Distributions with Applications to Shared Orthologs of Paired Circular Genomes	5. 発行年 2020年
3. 雑誌名 Computer Modeling in Engineering & Sciences	6. 最初と最後の頁 1131 ~ 1149
掲載論文のDOI (デジタルオブジェクト識別子) 10.32604/cmescs.2020.08466	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Tomoaki Imoto, Kunio Shimizu, Toshihiro Abe	4. 巻 2
2. 論文標題 A cylindrical distribution with heavy-tailed linear part	5. 発行年 2019年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 129-154
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-019-00031-5	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計9件（うち招待講演 0件 / うち国際学会 7件）

1. 発表者名 井本智明
2. 発表標題 トーラス分布の新しい構成法とその性質
3. 学会等名 2020年度統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 Imoto Tomoaki
2. 発表標題 Bivariate GIT_{3,1} distribution
3. 学会等名 The 3rd International Conference on Statistical Distributions and Applications (国際学会)
4. 発表年 2019年

1. 発表者名 Imoto Tomoaki
2. 発表標題 Sine-correlated distribution on the torus
3. 学会等名 62nd ISI World Statistics Congress 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Imoto Tomoaki
2. 発表標題 A method for constructing a circular distributions from a single valued analytic function
3. 学会等名 32nd Edition of the European Meeting of Statisticians (国際学会)
4. 発表年 2019年

1. 発表者名 Imoto Tomoaki
2. 発表標題 Discretized Circular Distribution
3. 学会等名 The 3rd International Conference on Econometrics and Statistics (国際学会)
4. 発表年 2019年

1. 発表者名 Tomoaki Imoto
2. 発表標題 A cylindrical distribution whose linear part is heavy-tailed
3. 学会等名 The 2nd International Conference on Econometrics and Statistics (EcoSta 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Tomoaki Imoto
2. 発表標題 Discrete circular distributions
3. 学会等名 The 4th ISM International Statistical Conference (ISM-IV) (国際学会)
4. 発表年 2018年

1. 発表者名 井本智明
2. 発表標題 解析関数を基にした角度分布の構成
3. 学会等名 Various Studies of Statistical Analysis for Asymptotic Theory, Circular or Time Series
4. 発表年 2018年

1. 発表者名 Tomoaki Imoto
2. 発表標題 Construction of circular distributions from closed curves in the plane
3. 学会等名 Pioneering Workshop on Extreme Value and Distribution Theories In Honor of Professor Masaaki Sibuya (国際学会)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------