

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 16 日現在

機関番号：82723

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700349

研究課題名(和文)生態統計学におけるノンパラメトリック法の適用

研究課題名(英文)Application of nonparametric methods in ecological statistics

研究代表者

村上 秀俊(Murakami, Hidetoshi)

防衛大学校(総合教育学群、人文社会科学群、応用科学群、電気情報学群及びシステム工・その他部局等・講師)

研究者番号：60453677

交付決定額(研究期間全体)：(直接経費) 2,400,000円、(間接経費) 0円

研究成果の概要(和文)：環境問題への対策として、生態統計学が注目を集めているが、生態統計学ではノンパラメトリック法が多用されている。しかしながら、多変量データに対するノンパラメトリック法の構成については、理論的難しさからあまり研究がなされてきていないのが現実であった。そこで、統計学に不可欠な検定問題と変化点の推定問題に焦点を当てて理論の構成を行う。変化点問題では、新たな検定統計量を提案した。1変量2標本検定問題では、最大値型検定統計量および修正型Wilcoxon検定統計量を提案した。また、多変量2標本検定問題では、様々な検定統計量を提案し、既存の統計量より高い検出力を得ることを示した。

研究成果の概要(英文)：In an environmental problem, the ecological statistics attracts attention in Statistics. In ecological statistics, the nonparametric statistical methods are often used to analyze the ecological data. However, by theoretical difficulty, the nonparametric methods for multivariate data had not been studied by many researchers. Therefore, we focused on the testing hypothesis for the univariate and multivariate two-sample problems and the detection of a change-point. For a change-point problem, we proposed a nonparametric test and analyzed various real data. By a simulation study, the suggested statistic was more accurate than the other statistics. For univariate two-sample problem, we suggested the maxed-type nonparametric test and the modified Wilcoxon rank sum test. For multivariate two-sample problem, we considered various nonparametric rank sum tests based on the Jureckova-Kalin's ranks of distances. The proposed statistics were more powerful than the other nonparametric statistics.

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：ノンパラメトリック法 検定統計量 多変量データ 変化点問題

1. 研究開始当初の背景

環境問題への対策は、現代社会における重要な研究課題のひとつである。統計学の分野では、生態統計学 (Ecological Statistics) が注目を集めている。生態統計学においては、ノンパラメトリック法の様々な手法を如何に上手くデータに適用するかが重要な問題であり、ノンパラメトリック法が多用されている。特に、近年におけるコンピュータの発達により、統計学において容易に多変量データを扱うことが可能となった。しかしながら、多変量データに対するノンパラメトリック法の構成については、理論的難しさからあまり研究がなされてきていないのが現実である。そのため、多変量データを各次元で、つまり1次元データとして分析することも多くある。その場合、変量間の相関を考慮しないことになってしまうため、重要な特性を見いだせない可能性がある。そこで、生態学データに対応するノンパラメトリック法の理論の構築が必要であった。特に、統計学に不可欠な検定問題と変化点の推定問題に焦点を当てて理論の構成を行う必要があった。また、生態学、医学、疫学等においては、多くのデータを得ることは非常に困難であり、分布を仮定する根拠が見いだせない状況が多くある。そのような状況において、小さい標本サイズで高い検出力を得ることの出来る検定統計量の考案や変化点の推定問題が必要となっている。

2. 研究の目的

本研究では、生態学データにおいて特に重要な役割を果たす変化点推定問題へ、既存のノンパラメトリック検定統計量の拡張を行なう。また、新しい1次元ノンパラメトリック検定統計量を考案すると共に、提案した検定統計量を多変量ノンパラメトリック検定統計量へと拡張し、生態学データへの適用について研究を行なう。

● ノンパラメトリック検定統計量の変化点推定問題への拡張

では、既存のノンパラメトリック統計量を変化点問題へと拡張し、推定精度の比較により統計量の妥当性を示す。変化点問題ではデータに特定の分布を仮定することは困難であるため、位置母数や尺度母数、特定のパラメータのみを検定する検定統計量を扱うのではなく、分布の同等性を検定できる検定統計量に焦点を当てる必要がある。分析をする上で、提案統計量の分布および推定精度が重要となってくるので、提案した検定統計量の極限分布を導出すると共に、シミュレーション実験により変化点の推定精度を既存の統計量と比較する。また、実

データを用いて既存の統計量との比較を行なうことで、様々なデータに対しても適用可能であることを示していく。

● 新しいノンパラメトリック検定統計量の考案と多変量統計量への拡張

では、位置母数や尺度母数、または特定のパラメータのみについて検定出来るノンパラメトリック検定統計量を考案するのではなく、分布の歪みや対称性などを含めた分布の同等性を検定できる統計量、つまり、より広い範囲で有効とされるノンパラメトリック検定統計量の考案を試みる。特に、生態統計学やデータを多く得ることが出来ない分野においては、データに分布を仮定することは非常に困難であるため、様々な分布において高い検出力を得ることが出来る1次元2標本ノンパラメトリック検定統計量の考案は重要な研究課題となっている。

また、多変量統計量への拡張として、多次元データを有益な情報のみにし、次元を下げる事が可能な主成分分析を用いる。各次元において分析を行なうと、変数間に存在する相関関係を考慮しないことになるため、適切な分析を行なうことが出来なくなる。そこで、主成分分析において重要な役割を果たす主成分スコアを用いて提案された検定法に応用することで新しい展開を試みる。

検定をおこなう際、最も重要となってくるのが検定統計量の分布の導出である。しかしながら、精密分布の導出は困難であるため、精密分布により近い近似分布を導出する必要がある。Edgeworth展開、鞍点近似、正規近似をはじめ、様々な近似法が提案されているので、それらの手法を用いることで、どの近似方法が提案統計量に対して良いか精度比較を行う。

● 生態学データへの適用

では、提案された手法を実際のデータへ応用し、提案法をはじめ、様々なノンパラメトリック手法との分析結果の比較を行う。

なお、本研究では検定統計量の考案や漸近分布の導出が中心となるが、シミュレーションによる数値実験や多数の実データにおける解析も視野において研究を行なうことを研究目的とする。

3. 研究の方法

検定問題や変化点の推定問題に関するノンパラメトリック検定統計量の構成に向けて、具体的な検定統計量での研究を進める。研究目的に述べた“ノンパラメトリック検定統計量の変化点問題への拡張”では、非常に高い検出力が得られる Lepage

統計量や Baumgartner 統計量をはじめ、有用なノンパラメトリック統計量を変化点問題へと拡張する。また、変化点問題で最も重要となるのが変化点の推定精度である。そこで、提案統計量の良さを示すため、様々な分布において乱数を発生させてシミュレーション実験を行い、既存の検定統計量と変化点推定精度の比較を行なう。

拡張した検定統計量の極限分布の導出など理論的展開を行なう。また、鞍点近似、Edgeworth 近似などにより近似分布を導出し、有限標本のもとでの分布の導出および近似精度の比較を行なう。

提案した検定統計量を実際のデータに適用し、既存の統計量と違った知見を得ることができるか検証する。

について研究を行なう。理論的研究が進まない場合は、Wilcoxon 統計量や Mood 統計量といった基本的なノンパラメトリック検定統計量を取り扱い、数値検証に重点をおきながら問題解決のための糸口を探る。また、国内の専門家を訪問し、最新の情報やコメントを得ながら研究を行う。特に、これらの問題に関して共同で研究を行なっている研究者との積極的な議論を進める。国内での学会や研究集会において、研究成果の発表を行う。ノンパラメトリック検定による変化点の推定問題に関して、海外でも関心をもって同様な課題に取り組んでいる研究者がおり、海外の研究者との情報交換は研究の促進に重要である。そこで研究代表者は外国における研究状況把握のため、海外での学会や研究集会へ参加する。

“新たな検定統計量の考案と多変量統計量への拡張”では、

高い検出力が得られる Baumgartner 統計量や Lepage 統計量の特徴を生かし、様々なパラメータの変化の違いを検出できるノンパラメトリック検定統計量を提案する。

多次元データにおける順位の方法について考える。また、順位を用いた様々な検定統計量に適用させ、多変量順位の方法が妥当であるかを検証する。特に、データの次元間の相関が強い場合と弱い場合の順位決定やその時の棄却点の導出を行なう。

様々な分布に従う乱数において主成分スコアを求め、パーミュテーションテストにより、有意水準の確認と検出力の動向を調べる。

検定問題において、検定統計量の分布の導出は最も重要な研究課題のひとつである。しかしながら、精密分布の導出は困難であるため、正規近似、鞍点近似、Edgeworth 展開を行なうことで、より正確な近似分布の導出を行なう。

について研究を行なう。理論的研究が進まない場合は、基本的なノンパラメトリック

検定統計量における理論の拡張を行なうと共に、シミュレーション実験等の数値検証に重点をおきながら問題解決のための糸口を探る。本研究で取り扱おうとしている統計量は、Neuhauser 教授も研究を行なっており、研究代表者とは定期的に連絡をとりあっているため、最新の情報やコメントを得ながら研究を行う。

“生態学データへの適用”では、提案された手法を実際のデータへ応用し、提案法をはじめ、様々なノンパラメトリック手法との分析結果の比較を行う。また、変化点推定の統計量を、多変量検定統計量と組み合わせることで、多変量データにおける変化点推定問題へと拡張する。

4. 研究成果

“ノンパラメトリック検定統計量の変化点問題への拡張”や“生態学データへの適用”に対する研究成果として、Lepage 型検定統計量を変化点推定問題へと拡張した。提案統計量の極限分布を導出し、極限分布を用いることが可能となる標本サイズをシミュレーション実験により推定した。また、提案統計量による変化点の推定精度を既存の統計量と比較し、シミュレーション実験により既存の統計量より推定精度が良いことを示した。また、シミュレーションアルゴリズムを示すと共に、実データにおいて変化点の推定を行なった。

“新たな検定統計量の考案と多変量統計量への拡張”や“生態学データへの適用”に対する研究成果として、1次元データに対しては、最大値型 Baumgartner 統計量を提案し、その極限分布を導出した。また、シミュレーション実験では様々な分布に従う乱数を発生させ、既存のノンパラメトリック検定統計量と検出力の比較を行ない、提案統計量の良さを示した。さらに、修正型 Wilcoxon 検定統計量の不偏性・非不偏性を証明した。修正型 Wilcoxon 検定統計量の漸近効率・極限分布・近似分布を導出し、その効率の良さ及び近似精度の比較を行なった。シミュレーションにより既存の統計量より検出力が高くなることを示し、実データにおいて分析を行うことで、提案統計量の妥当性を示した。多変量検定統計量として、Jureckova-Kalina の順位距離を用いることにより、最大値型多変量 2 標本 Baumgartner 統計量、無作為化多変量 2 標本 Baumgartner 統計量、多変量 Wilcoxon 型検定統計量を提案し、シミュレーションアルゴリズムを示すと共に、シミュレーション実験により、様々な分布において既存の検定統計量より高い検出力を得ることができることを示した。

5. 主な発表論文等

〔雑誌論文〕(計6件)

MURAKAMI, Hidetoshi. A nonparametric location-scale statistic for detecting a change-point. The International Journal of Advanced Manufacturing Technology. 査読有. 82巻. 2012. 711-728.

MURAKAMI, Hidetoshi. The max-type multivariate two-sample rank statistic based on the Baumgartner statistic. A Springer Company: Proceedings in Computational Statistics. 査読有. 617-627.

MURAKAMI, Hidetoshi. A max-type Baumgartner statistic for the two-sample problem and its power comparison. 査読有. Journal of the Japanese Society of Computational Statistics. 25巻. 39-49.

MURAKAMI, Hidetoshi. A randomized Baumgartner statistic for multivariate two-sample testing hypothesis. 査読有. Journal of Statistical Computation and Simulation.

DOI:10.1080/00949655.2013.809087.

MURAKAMI, Hidetoshi. Power comparison of multivariate Wilcoxon-type tests based on the Jureckova-Kalina's ranks of distances. 査読有. Communications in Statistics - Simulation and Computation. DOI:10.1080/03610918.2013.854905.

MURAKAMI, Hidetoshi. The power of the modified Wilcoxon rank sum test for the one-sided alternative. 査読有. Statistics.

DOI:10.1080/02331888.2014.913049.

〔学会発表〕(計3件)

MURAKAMI, Hidetoshi. The max-type multivariate two-sample rank statistic based on the Baumgartner Statistic. COMPSTAT2012

村上 秀俊. ノンパラメトリック統計量における鞍点近似および不偏性について. 日本数学会 特別講演.

MURAKAMI, Hidetoshi. A randomized nonparametric statistic for multivariate multisample testing hypothesis. The 59th Session of the International Statistical Institute.

6. 研究組織

(1)研究代表者

村上 秀俊 (MURAKAMI, Hidetoshi)

研究者番号: 60453677