

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 1 日現在

機関番号：32660

研究種目：若手研究(B)

研究期間：2014～2016

課題番号：26730025

研究課題名(和文) 多変量ノンパラメトリック検定統計量の開発と環境データへの応用

研究課題名(英文) The development of the multivariate nonparametric test for the environment data

研究代表者

村上 秀俊 (Murakami, Hidetoshi)

東京理科大学・理学部第一部数理情報科学科・講師

研究者番号：60453677

交付決定額(研究期間全体)：(直接経費) 2,000,000円

研究成果の概要(和文)：様々な検定統計量を多変量検定統計量へと拡張した。生態統計学や生物統計学で多用されている多重比較法への拡張、管理図における異常検出のための検定統計量への拡張を行ない、シミュレーション実験により、既存の検定統計量と検出力の比較を行い、提案統計量の妥当性を示した。また、様々な検定統計量に対して積率母関数を導出し、より精度の良い近似分布を導くと共に、検定問題において重要な役割を果たす検定統計量の不偏性を証明した。データ分析をする上で、一様分布やガンマ分布の和の分布の導出が必要であるが、和の分布には無限級数和が含まれることから応用が難しいため、鞍点近似や積率多項式近似によって和の分布の近似分布を導出した。

研究成果の概要(英文)：In ecological statistics, the nonparametric statistical methods are often used to analyze the ecological data. We extended and proposed the nonparametric statistics for the multivariate data. In addition, we extended various statistics to the multiple comparison procedure in one-way layout in biometry. By a simulation study, the suggested statistics were more powerful than the other nonparametric statistics. Additionally, we derived the moment generating functions of statistics to approximate the exact distribution of the statistics. Moreover, we proved the unbiasedness of the test statistics. In data analysis, it is important to derive the distribution of the sum of independent non-identically gamma or uniform random variables. However, since the distributions include the sum of infinite series, it is difficult to apply for the real data analysis. Then, we derived the approximate distribution based on the saddlepoint approximation and the moment polynomial approximation.

研究分野：統計科学

キーワード：ノンパラメトリック法 多変量データ 仮説検定

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。統計学における検定方法の代表的なものとして、分布を仮定しないノンパラメトリック検定がある。ノンパラメトリック検定は、特に環境学、生態学、医学、薬学等をはじめとする広い分野で多用されており、データ分析において重要な役割を果たしてきた。また、環境データの分析では、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、国内外の研究者によって理論面および応用面において研究がなされており、数々の素晴らしい研究成果が発表されている。多用されているノンパラメトリック法ではあるが、環境データによっては更なる理論構築が必要である。特に、測定技術の発展により様々な状況において情報収集が容易になったこともあり、多変量データの分析にも関心が集まっているが、環境学、生態学、医学、薬学等では多数のデータを収集することが困難であり、多変量ノンパラメトリック法の考案が必要不可欠である。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築については、その理論的難しさからあまり研究がなされていない状況である。特に、多変量データに対する順位決定、すなわち変数間の相関を考慮して順位を決定する必要があるが、ノンパラメトリック法の枠組みでの順位決定は十分な研究がなされているとは言えなかった。そのため、統計学に不可欠な検定問題と変化点の推定問題(異常値の検出)に焦点を当てて理論の構成を行う必要があった。

2. 研究の目的

本研究では、ノンパラメトリック検定統計量を考案するとともに、提案した、もしくは既存の検定統計量の極限分布および近似分布の導出を行なう。また、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として提案し、変化点の推定問題(異常値の検出)へと拡張することを目的とする。

“ノンパラメトリック検定統計量の考案”では、位置母数や尺度母数、もしくは位置-尺度母数の違いを検出できる検定統計量の考案を試みる。環境統計学、生態統計学、医学・薬学データなど多くのデータを得ることが分野やデータに特定の分布を仮定する根拠が見いだせない分野では、高い検出力を得ることが出来る検定統計量の提案は重要な研究課題であるため、より有用な検定統計量の提案は不可欠である。同時に、既存の検定統計量については、漸近的性質の観点から検定統計量の妥当性を確かめることによって、理論的に妥当性を示すことで、必要性を提案することを目的とする。

“多変量統計量への拡張”では、多変量データに対する順位決定が不可欠である。多変量データについて、次元毎に分析・検定を行なうと、各変数の間に存在する相関関係を考慮しないことになるため、正確な分析結果、検定結果を導き出すことが出来なくなる。そのため、多変量データの情報を保つ変換が必要である。そこで、位置母数に対して不変な変換となる Jureckova-Kalina の順位距離および多変量データを重要な情報のみにし、データの次元縮約を行なうことが出来る主成分分析、特に、主成分スコアを用いて、提案した検定に応用して新しい展開を行なうことを目的とする。

“検定統計量の極限分布・近似分布の導出”では、検定統計量の分布に対する近似法の精度比較を行なうことを試みる。統計的仮説検定において、検定統計量の棄却点は必要不可欠である。すなわち、検定統計量の精密分布を導出する必要があるが、標本サイズが大きくなると精密分布の導出は困難となる。そのため、より精度の良い近似分布の導出が必要であるため、鞍点近似、積率多項式近似、エッジワース展開、正規近似、カイ2乗近似をはじめ、様々な近似法によって検定統計量の分布に対する近似分布を導出することを目的とする。

“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、既存のノンパラメトリック検定統計量を変化点の推定問題(異常値の検出問題)へと拡張し、シミュレーション実験による数値比較によって提案統計量の妥当性を示すことを試みる。データに特定の分布を仮定することは困難なため、ノンパラメトリック法に基づく検定統計量の構成が必要である。また、位置母数や尺度母数のみの違いによって変化点を推定するのではなく、位置-尺度母数の違いによって変化点および異常値を推定する統計量の導出を試みる。また、帰無仮説および対立仮説の下での極限分布を導出することで、統計量の理論的性質も導出することを目的とする。

“環境データへの応用”では、提案された手法を実データ分析へと応用し、既存の様々なノンパラメトリック法との比較を行う。

なお、本研究では検定統計量の考案や検定統計量の近似分布、極限分布の導出、その理論的性質の妥当性、漸近的性質の検証が研究の中心となるが、シミュレーション実験による数値比較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

本研究では、ノンパラメトリックな検定統計量の提案、検定統計量の極限分布および近

似分布の導出，変化点の推定問題における検定統計量の構成について，具体的な検定統計量で研究を進める．

“ノンパラメトリック検定統計量の考案”において：

- ・ 順序統計量の期待に基づく線形順位和検定は，ある対立仮説に対して局所最強力検定となる特性を用いて，位置母数や尺度母数のみの違いを検出するだけでなく，形状母数や対称性の違いも検出できるノンパラメトリックな検定統計量を提案する．
- ・ 検出力は，検定統計量の良さを表すために最もよく用いられる指標である．そこで，様々な分布に従う乱数を発生させ，有意水準の確認と検出力の比較を行なう．
- ・ 検定統計量が良くなることを示す概念のひとつに局所最強力検定がある．提案した検定統計量が，局所最強力検定となる分布を，微分方程式を解くことによって導出する．
- ・ 正しい検定結果，信頼できる検定結果として結論付けるためには，検定統計量が不偏となることが重要である．既存の検定統計量を含め，その不偏性/非不偏性を検証する．

について研究を行なう．

“検定統計量の極限分布・近似分布の導出”において：

- ・ 統計的仮説検定では，検定統計量の精密分布の導出は最も重要な研究課題であるが，標本サイズが大きくなると導出は困難である．そこで，エッジワース展開，鞍点近似，積率多項式近似などを適用することで，より正確な近似分布の導出を行なう．
- ・ 近似分布を導出するためには，検定統計量の高次積率が必要である．整数の分割公式を適応することで，検定統計量の積率母関数を導出する．

について研究を行なう．

“多変量検定統計量への拡張”において：

- ・ 位置母数に対して不変な変換となる Jureckova-Kalina の順位距離に基づき，多変量データを順位データに変換し，ノンパラメトリック検定に適用する．
- ・ 多変量データを重要な情報を保持しつつ，次元の縮約をすることが可能な主成分分析を用いる．特に，主成分スコアを用いた固有値の同等性検定を分布の同等性検定へ拡張する．

について研究を行なう．

“変化点問題への検定統計量の拡張”において：

- ・ データに特定の分布を仮定することが困

難であるため，位置母数や尺度母数のみを検定する統計量を扱うのではなく，位置母数 - 尺度母数を同時に検定することができる修正型 Baumgartner 検定を，異常値を検出するための管理図へ応用する．

- ・ Average Run Length が不偏となることは，統計量の妥当性を示すことに繋がるため，鞍点近似を用いて不偏性を検証する．
- ・ 検定統計量の妥当性を示すため，対立仮説の下での統計量の極限分布を導出し，理論値として検出力を求める．また，検定統計量の非心度を導くことで，検定統計量の効率を導出する．

について研究する．

“環境データへの応用”において：

- ・ 提案された手法を実際のデータへ応用し，提案手法と様々な既存のノンパラメトリック法との分析結果を比較する．

理論的研究が進まない場合は，基本的な検定統計量に関して理論の拡張を行ない，シミュレーション実験による数値検証に重点を置きながら問題解決のための糸口を探る．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を訪問し，積極的な討論を進めると共に，最新の情報を収集する．

4. 研究成果

“ノンパラメトリック検定統計量の考案”や“検定統計量の極限分布・近似分布の導出”に対する研究成果として，修正型 Makeham 分布の順序統計量に基づく検定統計量を提案した．さらに，精密分布への近似分布を導出し，シミュレーション実験により既存の検定統計量と検出力の比較を行い，提案統計量の妥当性を示した．また，既存の検定統計量に対して，その漸近効率の導出，局所最強力検定となる分布の導出，より精度の良い近似分布の導出，一般化した検定統計量の積率母関数を導出した．さらに，生態統計学や生物統計学で多用されている対比較における新たな検定統計量を提案し，検出力の比較によって提案統計量の良さを示した．検定統計量が不偏検定となることは，検定問題において重要な役割を果たす．研究を推進するにあたり，不偏性・非不偏性の証明は不可欠であるため，一般の多標本検定問題において，幾つかの検定統計量の不偏性・非不偏性を証明した．

“多変量検定統計量への拡張”に対する研究成果として，Jureckova-Kalina の順位距離を多標本問題へと拡張することにより，多変量 Kruskal-Wallis 型検定（すなわち，多元配置分散分析）を提案し，シミュレーションアルゴリズムを示すとともに，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出来ることをシミュレーション実験によって示した．また，1次元問題において，Kruskal-Wallis 検定が非不偏とな

ることを，反例を挙げることによって証明した。

“変化点問題への検定統計量の拡張”に対する研究成果として，修正型 Baumgartner 検定を管理図へ適用できるように拡張した。より正確に管理限界を超える/超えないことを導出するための方法を提案するとともに，そのアルゴリズムを示した。また，提案統計量の漸近的な不偏性を示し，極限分布への収束を速める連続補正を提案し，その妥当性を示した。

“環境データへの応用”に対する研究成果として，提案手法によってデータ分析を行い，既存の手法と結果の比較を行なった。それ以上に，環境データの分析をする上で，データ分布に特定の分布を仮定する必要がある場合がある。特に，一様分布の和の分布やガンマ分布の和の分布の導出が必要であるが，和の分布には無限級数和が含まれており，標本サイズが有限である実際のデータには応用することが出来ない。そこで，鞍点近似や積率多項式近似によって，和の分布の近似分布を導出し，その妥当性を示した。

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)

MURAKAMI, Hidetoshi and LEE, Seong Keon. (2017). 査読有. The unbiased nonparametric test and its moment generating function for the ordered alternative. *Statistics & Probability Letters*. 126, 193-197.

DOI: 10.1016/j.spl.2017.03.009

MURAKAMI, Hidetoshi. (2017). 査読有. The Multivariate Multisample Nonparametric Rank Statistics for the Location Alternatives. *Austrian Journal of Statistics*. 46, 3-13.

DOI: 10.17713/ajs.v46i1.130

MURAKAMI, Hidetoshi. (2016). 査読有. A moment generating function of a combination of linear rank tests and its asymptotic efficiency. *TEST*. 25, 674-691.

DOI: 10.1007/s11749-016-0490-6

MURAKAMI, Hidetoshi. (2016). 査読有. All-pairs multiple comparisons based on the Cucconi test. *AStA Advances in Statistical Analysis*. 100, 355-368.

DOI: 10.1007/s10182-016-0268-x

MURAKAMI, Hidetoshi. (2015). 査読有. Approximations to the distribution of sum of independent non-identically gamma random variables. *Mathematical Sciences*. 9, 205-213.

DOI: 10.1007/s40096-015-0169-2

MURAKAMI, Hidetoshi and HA, Hyung Tae.

(2015). 査読有. The modified Mood test for the scale alternative and its numerical comparisons. *Journal of the Korean Statistical Society*. 44, 592-605.

DOI: 10.1016/j.jkss.2015.03.005

MURAKAMI, Hidetoshi and LEE, Seong Keon. (2015). 査読有. Unbiasedness and biasedness of the Jonckheere-Terpstra and the Kruskal-Wallis tests. *Journal of the Korean Statistical Society*. 44, 342-351.

DOI: 10.1016/j.jkss.2014.10.001

MURAKAMI, Hidetoshi. (2014). 査読有. Asymptotic efficiency and small sample power of a locally most powerful linear rank test for the log-logistic distribution. *Mathematical Sciences*. 8, 109-115.

DOI: 10.1007/s40096-014-0135-4

MURAKAMI, Hidetoshi. (2014). 査読有. A saddlepoint approximation to the distribution of the sum of independent non-identically uniform random variables. *Statistica Neerlandica*. 68, 267-275.

DOI: 10.1111/stan.12032

OGURA, Toru and MURAKAMI, Hidetoshi. (2014). 査読有. A rank test based on the moments of order statistics of the modified Makeham distribution. *Computational Statistics*. 29, 1691-1711.

DOI: 10.1007/s00180-014-0513-2

[学会発表](計9件)

NISHINO, Takuya and MURAKAMI, Hidetoshi. (2016). The limiting distribution of a rank test based on the multisample Cucconi test. *COMPSTAT 2016. オビエド(スペイン)*

MURAKAMI, Hidetoshi. (2016). A test based on kernel density estimation for the eigenvalues in two-sample problem. *COMPSTAT 2016. オビエド(スペイン)*
MURAKAMI, Hidetoshi. (2015). Power Comparison of the Multivariate Kruskal-Wallis-Type Test. *Eighth International Workshop on Simulation. ウィーン(オーストリア)*

YAMAMOTO, Kouji and MURAKAMI, Hidetoshi. (2014). Performance of the skew normal distribution type symmetry model for analyzing square ordinal tables. *COMPSTAT 2014. ジュネーブ(スイス)*

MURAKAMI, Hidetoshi. (2014). The moment-based approximation with a skew-normal polynomial: a numerical comparison. *COMPSTAT 2014. ジュネー*

ブ(スイス)

白井 純平, 村上 秀俊. (2016). Median Ranked Set Sampling におけるカーネル密度推定. 日本計算機統計学会第 30 回シンポジウム. プラサ ヴェルデ(静岡県沼津市)

西野 拓哉, 村上 秀俊. (2016). The improvement of the Cucconi Test. 日本計算機統計学会第 30 回シンポジウム. プラサ ヴェルデ(静岡県沼津市)

霜田 小百合, 村上 秀俊. (2016). A Baumgartner Control Chart for Joint Monitoring of Unknown Location and Scale. 日本計算機統計学会第 30 回シンポジウム. プラサ ヴェルデ(静岡県沼津市)

村上 秀俊. (2016). ノンパラメトリック検定の不偏性と近似分布. 統計関連学会連合大会. 金沢大学(石川県金沢市)

〔図書〕(計 1 件)

村上 秀俊. (2015). シリーズ: 統計解析スタンダード「ノンパラメトリック法」朝倉書店. (総ページ数: 181 ページ)

6. 研究組織

(1) 研究代表者

村上 秀俊 (MURAKAMI, Hidetoshi)
東京理科大学・理学部第一部数理情報科学科・講師
研究者番号: 60453677