

令和 5 年 5 月 14 日現在

機関番号：32660

研究種目：基盤研究(C) (一般)

研究期間：2018～2022

課題番号：18K11199

研究課題名(和文)自然保護のためのノンパラメトリック法の開発

研究課題名(英文)The development of nonparametric methods for environment protection

研究代表者

村上 秀俊(Murakami, Hidetoshi)

東京理科大学・理学部第一部応用数学科・准教授

研究者番号：60453677

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：Ranked Set Sampling (RSS) は、母集団からの標本抽出方法の一つであり、様々な分野で用いられるサンプリング方法である。本研究では1標本、2標本、多標本検定問題に対する新しいノンパラメトリック検定統計量を提案した。検定統計量の極限分布や精緻な近似分布の導出を行ない、数値実験を通して提案統計量の有用性を示した。また、漸近検出力や一致性、不偏性など対立仮説の下で重要な理論的性質について示した。さらに、多変量データに対する新しい順位決定方法を提案した。実際のRSSデータはどのような分布に従うか不明であるため、カーネル密度推定を用いて分布推定やパラメータ推定を行なった。

研究成果の学術的意義や社会的意義

一般的な仮説検定問題では、単純無作為抽出から得られたデータに対して分析を行なう。しかし、多種多様なデータが存在する現代社会においては、様々なサンプリング方法によって得られたデータが存在するため、既存の手法では対処できない問題が多々発生する。本研究成果は、その解決方法の一つとして、統計学のさらなる発展を担うものである。

研究成果の概要(英文)：Ranked set sampling (RSS) is one of random sampling methods from population. RSS is widely used in many scientific fields. In this research, we proposed one-, two- and multisample nonparametric test statistics in testing problem. We derived the limiting distributions and the approximate distributions of proposed test statistics. In addition, we showed theoretical properties such as the asymptotic power, the consistency and the unbiasedness. We also suggested how to determine the rank of vector value of data. Since, in practice, the distribution of RSS data is unknown, we used the kernel density estimation to estimate the population distribution and parameters.

研究分野：統計学

キーワード：ノンパラメトリック法 Ranked Set Sampling 近似分布 密度推定 検定統計量

### 1. 研究開始当初の背景

自然保護や環境問題への対策は、現代社会において必要不可欠な研究課題である。統計学は精度の良いデータを収集することで良い分析が可能となるが、Ranked Set Sampling というサンプリング手法を用いることで、データを単純無作為抽出するよりも分散が小さくなるデータの収集が可能である。しかし、多変量データに対する Ranked Set Sampling 法およびノンパラメトリック法の理論構成については、理論的難しさからあまり研究がなされていない。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、Ranked Set Sampling による多変量データの収集や多変量データに対するノンパラメトリック法の考案が必要不可欠である。また、Ranked Set Sampling によって収集したデータに対して、母集団分布を仮定する根拠が見いだせない状況が多くある。そこで、Ranked Set Sampling に基づくノンパラメトリック法の理論構成が重要な研究課題のひとつとなっている。特に、Ranked Set Sampling に基づく多変量データの順位決定とカーネル密度推定、カーネル分布推定は不十分な点や未解決な問題も多く、研究課題として残されている。

### 2. 研究の目的

上記の背景のもと、本研究では、単純無作為抽出によって得られたデータに対するノンパラメトリック検定統計量を考案するとともに、新しい Ranked Set Sampling の提案、単純無作為抽出によって得られた多変量データに対するノンパラメトリック検定統計量の提案、多変量 Ranked Set Sampling によって得られたデータに対するノンパラメトリック検定統計量の考案、Ranked Set Sampling に対するカーネル密度推定、実データへの適用に焦点を当てて、提案したノンパラメトリック検定統計量、もしくは既存のノンパラメトリック検定統計量を拡張することを目的とする。

(1) 単純無作為抽出によるデータに対して、1 標本、2 標本、多標本検定統計量を提案する。さらに、提案統計量の極限分布や近似分布を導出する。

(2) 多変量ノンパラメトリック法では、多変量データに対する順位決定が重要な役割を担う。多変量解析の手法を用いることで、多変量データに対する順位決定方法を提案する。

(3) Ranked Set Sampling によるデータは特殊な構造を持っているため、新しい Ranked Set Sampling を提案するとともに、Ranked Set Sampling によるデータに対して、カーネル密度推定による母集団密度の推定やパラメータ推定を行う。

(4) 検定統計量の漸近検出力、一致性、不偏性など対立仮説下での理論的性質について導出する。

### 3. 研究の方法

上述した 4 つの項目に対して、提案統計量の理論構築、シミュレーション実験による提案統計量の性質に対する考察、既存の手法との比較、実データへの応用を行なった。研究目的を達成するために以下のアプローチで研究を推進した。

(1) 学術論文を通して先行研究の結果を理解することに加え、類似研究に関する文献調査や理論の本質を整理し、現在直面している研究課題に適用可能か確認した。

(2) 理論的検討および数値解析ソフトを用いた数値検証をすることで理論的誤りがないか確認した。

(3) シミュレーション実験による検証を行なった。

(4) 学会発表や研究集会への参加を通して得られる他者からの意見や疑問点について検討した。

(5) 理論的検討が困難な場合、数値的アプローチを用いて検証した。

### 4. 研究成果

本研究は、2018 年度から 2020 年度の 3 年間で計画されていたが、新型コロナウイルスの影響に伴い、2 年間研究期間が延長されたため、2022 年度の 5 年間で実施された。研究期間内の主な研究成果は次の通りである。

一つ目の研究成果は「1 標本、2 標本、多標本検定統計量の提案」である。先行研究では、位置母数に対する検定や尺度母数に対する検定が多く提案されているが、実際のデータ解析にお

いては位置母数や尺度母数,もしくは別な母数が同時に異なっていることが多々存在する.そこで,様々なパラメータに対して差異を測れるような1標本,2標本,多標本検定統計量を複数提案した.また,大規模なシミュレーション実験によって既存の検定統計量との検出力比較を行なうことで提案検定統計量の有用性を示した.

二つ目の研究成果は「提案検定統計量の極限分布や近似分布の導出」である.統計的仮説検定を行う際,検定統計量の分布は重要な役割を果たしている.近年の計算機の発達によりシミュレーション実験が容易に行えるようになったことで,数値的に検定統計量の分布を推定することが多く見られるようになってきているのが現状である.しかし,標本サイズが大きくなった場合,数値実験では精度に問題が生じるため,本研究では検定統計量の極限分布や近似分布を理論的に導出した.近似分布の導出では検定統計量の積率が必要となるため,積率母関数を導出することで,検定統計量の積率が容易に計算できるようになった.また,既に提案されているが極限分布や近似分布が知られていない検定統計量も存在するため,その検定統計量の分布を導出した.

三つ目の研究成果は「検定統計量の漸近検出力,一致性,不偏性の導出」である.検定統計量を提案する上で帰無仮説の下での議論は先行研究でも多くなされている.しかし,実際のデータ解析においては対立仮説での議論が重要な役割を果たす.特にノンパラメトリック検定の不偏性については半世紀以上前に問題提起がされたが,今も尚研究されている課題である.そこで,既存の1標本および2標本ノンパラメトリック検定統計量の不偏性について証明した.さらに,対立仮説の下で重要な役割を果たす漸近検出力や検定統計量の一貫性について導出した.

四つ目は「Ranke Set Sampling データに対する推定」である.一般的に,Ranked Set Samplingに基づく推定量の分布は複雑な構造を持っているため,計算機指向型の手法を用いる必要がある.また,先行研究では数値比較のみによるRanke Set Samplingの提案が多くされており,理論的検証があまりなされていない.そこで,新たな平滑化ブートストラップ法を提案し,カーネル分布推定量の平均積分2乗誤差を漸近的に最小化するための平滑化パラメータや一貫性について示した.

最後に「多変量データに対する順位決定」である.多変量ノンパラメトリック法では,如何に多変量データの順序(順位)を決定するかが重要な問題となってくる.次元ごとに順序(順位)を決定すると,それらの間にある相関が無視され,適切な推定が行えなくなる.それゆえ,相関を考慮した上で順序(順位)を決定する必要がある.そこで,多変量解析で主要な役割を担っている主成分分析と独立成分分析に焦点を当て,主成分スコアもしくは独立成分スコアを用いて検定する手法を提案した.提案手法を既存の多標本検定統計量に応用し,その妥当性をシミュレーション実験により示した.

## 5. 主な発表論文等

〔雑誌論文〕 計20件（うち査読付論文 20件 / うち国際共著 3件 / うちオープンアクセス 0件）

1. 著者名 Yamaguchi Hikaru, Murakami Hidetoshi	4. 巻 33
2. 論文標題 Symmetric smoothed bootstrap methods for ranked set samples	5. 発行年 2021年
3. 雑誌名 Journal of Nonparametric Statistics	6. 最初と最後の頁 435 ~ 463
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/10485252.2021.1971226	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 Kitani Masato, Murakami Hidetoshi	4. 巻 92
2. 論文標題 One-sample location test based on the sign and Wilcoxon signed-rank tests	5. 発行年 2021年
3. 雑誌名 Journal of Statistical Computation and Simulation	6. 最初と最後の頁 610 ~ 622
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/00949655.2021.1968399	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 Mukherjee Amitava, Kossler Wolfgang, Murakami Hidetoshi	4. 巻 55
2. 論文標題 Two new distribution-free two-sample tests for versatile alternative	5. 発行年 2021年
3. 雑誌名 Statistics	6. 最初と最後の頁 1123 ~ 1153
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/02331888.2021.1998053	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 該当する
1. 著者名 Nishino Takuya, Murakami Hidetoshi	4. 巻 90
2. 論文標題 The generalized multisample Cucconi test statistic for the location and scale parameters	5. 発行年 2020年
3. 雑誌名 Journal of Statistical Computation and Simulation	6. 最初と最後の頁 2291 ~ 2305
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/00949655.2020.1774884	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 Toru Ogura, Hidetoshi Murakami	4. 巻 13
2. 論文標題 Canonical Correlation Analysis of Principal Component Scores for Multiple-set Random Vectors	5. 発行年 2020年
3. 雑誌名 Electronic Journal of Applied Statistical Analysis	6. 最初と最後の頁 47 ~ 74
掲載論文のDOI (デジタルオブジェクト識別子) 10.1285/i20705948v13n1p47	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 Hanyuda Bunto, Murakami Hidetoshi	4. 巻 50
2. 論文標題 Power comparisons of the unbiased Berk-Jones test and the unbiased reversed Berk-Jones test	5. 発行年 2019年
3. 雑誌名 Communications in Statistics - Simulation and Computation	6. 最初と最後の頁 1009 ~ 1024
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/03610918.2019.1571608	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 Mukherjee Amitava, Murakami Hidetoshi	4. 巻 64
2. 論文標題 Multivariate Kruskal-Wallis tests based on principal component score and latent source of independent component analysis	5. 発行年 2022年
3. 雑誌名 Australian & New Zealand Journal of Statistics	6. 最初と最後の頁 356 ~ 380
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/anzs.12371	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 該当する

1. 著者名 Yamaguchi Hikaru, Murakami Hidetoshi	4. 巻 -
2. 論文標題 Interpoint distance-based two-sample tests for functional data	5. 発行年 2022年
3. 雑誌名 Communications in Statistics - Theory and Methods	6. 最初と最後の頁 1 ~ 21
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/03610926.2022.2148475	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 Murakami Hidetoshi, Lee Seong Keon	4. 巻 -
2. 論文標題 On unbiasedness and biasedness of the Wilcoxon and some nonparametric tests	5. 発行年 2022年
3. 雑誌名 WIREs Computational Statistics	6. 最初と最後の頁 1~16
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/wics.1600	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 該当する

1. 著者名 Nishino Takuya, Murakami Hidetoshi	4. 巻 48
2. 論文標題 The generalized Cucconi test statistic for the two-sample problem	5. 発行年 2019年
3. 雑誌名 Journal of the Korean Statistical Society	6. 最初と最後の頁 593~612
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jkss.2019.05.004	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25件 (うち招待講演 2件 / うち国際学会 14件)

1. 発表者名 山口光、村上秀俊
2. 発表標題 Interpoint distance-based test for functional data
3. 学会等名 日本計算機統計学会 第34回シンポジウム
4. 発表年 2020年

1. 発表者名 Hikaru Yamaguchi, Hidetoshi Murakami
2. 発表標題 Multivariate Kernel Density Estimator From Ranked Set Samples
3. 学会等名 10th International Workshop on Simulation and Statistics (国際学会)
4. 発表年 2019年

1. 発表者名 山口 光, 村上 秀俊
2. 発表標題 Nonparametric tests for interval-valued data
3. 学会等名 日本計算機統計学会第33回大会
4. 発表年 2019年

1. 発表者名 Yamaguchi Hikaru, Murakami Hidetoshi
2. 発表標題 A new scale estimator based on the kernel density estimator with ranked set sampling
3. 学会等名 The 23rd International Conference on Computational Statistics (国際学会)
4. 発表年 2018年

1. 発表者名 羽生田文登、村上秀俊
2. 発表標題 Calibrated Kolmogorov-Smirnov検定の近似分布と不偏性
3. 学会等名 2018年度統計関連学会連合大会
4. 発表年 2018年

1. 発表者名 金澤優気、村上秀俊
2. 発表標題 順位変換に基づくカーネル分布推定
3. 学会等名 日本計算機統計学会第32回大会
4. 発表年 2018年

1. 発表者名 Matsuo Kazuki, Mukherjee Amitava, Murakami Hidetoshi
2. 発表標題 A Class of Quasi-Distribution-free Multivariate Multisample Lepage Tests and Their Applications in High-Dimensional Data Study
3. 学会等名 DSBFI 2023 (国際学会)
4. 発表年 2023年

1. 発表者名 Funato Mika, Murakami Hidetoshi, Mukherjee Amitava
2. 発表標題 An Adaptive Two-sample Test for Combined Location, Scale and Lehmann Alternatives
3. 学会等名 日本計算機統計学会第36回シンポジウム
4. 発表年 2022年

〔図書〕 計1件

1. 著者名 杉山 高一, 藤越 康祝 (監修), 村上秀俊他13名	4. 発行年 2020年
2. 出版社 勉誠出版	5. 総ページ数 253
3. 書名 R・Pythonによる 統計データ科学	

〔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



インド	XLRI- Xavier School of Management			
-----	-----------------------------------	--	--	--