

令和 3 年 5 月 25 日現在

機関番号：17102

研究種目：若手研究

研究期間：2018～2020

課題番号：18K13454

研究課題名（和文）多様性を表現する統計モデルの開発

研究課題名（英文）Developing statistical models representing diversity

研究代表者

佃 康司（Tsukuda, Koji）

九州大学・数理学研究院・准教授

研究者番号：30764972

交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：多様性の表現に用いられる確率分割モデルを主な対象として、既存モデルに対する検討と新しいモデルの開発を主目的に研究を遂行した。成果として、基本モデルであるユーエンス抽出公式およびピットマン抽出公式に従う確率分割に対する近似や漸近評価について知見を深めたほか、ディリクレ多項モデルを用いた判別分析法を提案した。また、研究課題に関連する確率過程や多変量の統計モデルに対する推測法についての研究も遂行した。具体的には、確率過程に対する適合度検定・変化点検定、二つの母集団の共分散行列に関する典型的な仮説に対する高次元検定の検討・提案を行った。

研究成果の学術的意義や社会的意義

ユーエンス抽出公式やピットマン抽出公式に従う確率分割の挙動をこれまでより正確に評価できるようになり、これらのモデルを応用する場合により詳細な議論が可能になると期待される。また、提案した判別分析法は多様性が高い集団を考える場合に特に有用である。関連するモデルについての研究で得られた成果は、それぞれが独自のアプローチで検定法を考えただけであり、他のモデルへの広がりも期待できる。

研究成果の概要（英文）：This project aims to study random partition models used for representing statistical diversities from the viewpoint of mathematical properties and develop a new random partition model. In particular, we investigated approximations and asymptotic evaluations associated with the Ewens sampling formula and the Pitman sampling formula. Moreover, we proposed a discrimination analysis method based on the Dirichlet-multinomial model.

Statistical inference for some related stochastic process models and multivariate models were also studied. We discussed and proposed goodness-of-fit tests and change-point tests for stochastic process models. We proposed high-dimensional tests for some typical models of covariance matrices.

研究分野：数理統計

キーワード：統計モデル 確率分割 確率過程 共分散行列 適合度検定 統計的漸近理論 無限次元空間における弱収束 ランダム組み合わせ構造

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

## 1. 研究開始当初の背景

離散的な現象を観察して得られるデータを用いたモデル化がよく行われる。典型的なのは母集団頻度に興味がある場合で、存在する全ての属性がわかっているならば標本を多項分布でモデル化して推測することが基本的である。それに対して、「存在するかわからない属性がある」状況もあり、このときは母集団の多様性に対する推測を経由するといったアプローチがある。また、「各属性の母集団頻度については興味がないけれども母集団の多様性(多様性の指標には例えば母集団頻度の平方和やエントロピーがある)を推測したい」という状況もある。

多様性の推測において、ランダムに起こる整数の分割のモデルである確率分割モデルが重要な役割を果たしている。確率分割の基本的なモデルとして、母集団の属性数が有限の場合に対応するモデルとして対称ディリクレ多項分布、無限の場合に対応するモデルとしてユーエンス抽出公式がある。また、ユーエンス抽出公式を拡張したモデルには、アセンブリやピットマン抽出公式といったモデルがある。多様性の推測を行うにあたってはこういったモデルの特徴を詳細に知っておくことが肝要であり、これまでもよく調べられてきた。しかしながら、既存の結果の多くは母集団の多様性が高い状況 - この状況では標本における属性数が多くなり、存在するかわからない属性の存在が示唆されることもある - を想定しておらず、母集団の多様性が高い状況における確率分割モデルの性質に関する知見が十分ではなかった。

## 2. 研究の目的

本研究の主な目的は、

- ・既存のモデルについて特に母集団の多様性が高い状況での性質を詳細に調べること
- ・多様性の推測のために役立つ新しいモデルを開発すること

である。また、よい着想があれば多様性の統計的推測に関連した問題について研究する。

## 3. 研究の方法

漸近理論および計算機によるシミュレーションを用いて研究する。漸近理論には、特に研究代表者が本研究課題の開始以前から進めてきた設定やそれに関連した設定のもとで議論する。シミュレーションでも漸近理論の設定に対応するような状況を考える。

## 4. 研究成果

### (1) 確率分割モデルについての研究

確率分割モデルについては、既存の確率分割モデルに対する検討と新しいモデルの開発の研究を行った。

既存の確率分割モデルに対する検討について、ユーエンス抽出公式およびピットマン抽出公式を主な対象として研究を進めた。以前より調査してきたユーエンス抽出公式に関するポアソン近似を調べる問題について研究を進展させ、論文が *The Annals of Applied Probability* 誌に掲載され、それまでに得た成果をまとめた内容についてシンポジウムで発表した。ユーエンス抽出公式に従う確率分割の長さを正規分布で近似したときの誤差に対する新しい評価を与える問題について、成果を学会・国際ワークショップ等で発表し、論文が国際ワークショップの会議録として出版された書籍 *Pioneering Works on Distribution Theory* に掲載された。以前より進めているピットマン ヨー過程(ピットマン抽出公式に関連するモデルの一つ)を事前過程として用いた場合の事後過程の性質を調べる問題について研究を進展させ、学会等で発表した。ピットマン抽出公式に従う確率分割の長さの積率に対する漸近評価を精密化する問題について研究を行い、成果を学会等で発表し、論文を投稿した。これまでに述べた研究は研究代表者が進めてきた漸近理論に基づいた考察を行って独自性が高いと考えており、今後調べるべき問題も多くあると考えている。また、以前より進めてきたディリクレ多項モデルを応用した判別分析法を提案する研究を進展させ、論文が *Statistical Applications in Genetics and Molecular Biology* 誌に掲載された。

新しいモデルの開発について、ユーエンス抽出公式にこれまでとは異なった形でパラメータを加えるモデルの検討を行い、その中の一案について成果を学会で発表した。この内容については研究期間中には十分な成果を得ることができなかったものの、この方向で研究を今後も進める余地はあると考えている。

(2)関連した問題についての研究

関連した問題については、確率過程モデルと多変量モデルに対する推測の研究を行った。

確率過程モデルの推測の研究として、適合度検定と変化点検定について研究を進めた。マーク付き経験過程に関する極限定理に基づいた非線形拡散過程と非線形自己回帰モデルに対する適合度検定法を開発し、学会・国際会議で発表し、論文が *Electronic Journal of Statistics* 誌に掲載された。また、以前に開発した拡散過程に対する変化点検定について国際会議で発表した。これらの研究では研究代表者が以前より研究してきたヒルベルト空間に値をとる確率変数に対する極限定理が重要な役割を果たしている。今後については、同様のアプローチを用いて他の問題に取り組みないか考えている。

多変量モデルの推測の研究として、高次元確率ベクトルの背後にある母集団の共分散行列に関する仮説検定について研究を行った。二つの高次元共分散行列の比例性の検定法を開発し、成果を学会等で発表し、論文が *Journal of Multivariate Analysis* 誌に掲載された。二つの高次元共分散行列が同時に対角化できるかを検定(共通主成分性の検定)する方法を研究し、成果を学会で発表し、論文を投稿した。今後については、共通主成分性の高次元検定について研究を進展させることが重要であるほか、母集団の共分散行列についてこれまで議論されていないような仮説を考える価値もあると考えている。

## 5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 Koji Tsukuda	4. 巻 -
2. 論文標題 Error Bounds for the Normal Approximation to the Length of a Ewens Partition	5. 発行年 2020年
3. 雑誌名 Pioneering Works on Distribution Theory	6. 最初と最後の頁 55-73
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-981-15-9663-6_3	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 Koji Tsukuda, Yoichi Nishiyama	4. 巻 14
2. 論文標題 Weak convergence of marked empirical processes in a Hilbert space and its applications	5. 発行年 2020年
3. 雑誌名 Electronic Journal of Statistics	6. 最初と最後の頁 3914-3938
掲載論文のDOI（デジタルオブジェクト識別子） 10.1214/20-EJS1761	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Koji Tsukuda, Shuhei Mano, Toshimichi Yamamoto	4. 巻 19
2. 論文標題 Bayesian approach to discriminant problems for count data with application to multilocus short tandem repeat dataset	5. 発行年 2020年
3. 雑誌名 Statistical Applications in Genetics and Molecular Biology	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1515/sagmb-2018-0044	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 Koji Tsukuda, Shun Matsuura	4. 巻 171
2. 論文標題 High-dimensional testing for proportional covariance matrices	5. 発行年 2019年
3. 雑誌名 Journal of Multivariate Analysis	6. 最初と最後の頁 412-420
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.jmva.2019.01.011	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Koji Tsukuda	4. 巻 29
2. 論文標題 On Poisson approximations for the Ewens sampling formula when the mutation parameter grows with the sample size	5. 発行年 2019年
3. 雑誌名 The Annals of Applied Probability	6. 最初と最後の頁 1188-1232
掲載論文のDOI (デジタルオブジェクト識別子) 10.1214/18-AAP1433	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

[学会発表] 計17件(うち招待講演 3件/うち国際学会 3件)

1. 発表者名 佃 康司
2. 発表標題 Pitman 分割の長さの積率に対する漸近評価
3. 学会等名 日本数学会2021年度年会
4. 発表年 2021年

1. 発表者名 佃 康司, 松浦 峻
2. 発表標題 四つのウィシャート行列の積のトレースに関する高次元極限定理とその応用
3. 学会等名 日本数学会2020年度秋季総合分科会
4. 発表年 2020年

1. 発表者名 佃 康司, 松浦 峻
2. 発表標題 ウィシャート行列に関する高次元極限定理を用いた共通主成分性の仮説検定
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 佃 康司
2. 発表標題 Pitman 分割の長さの積率に対する漸近評価
3. 学会等名 研究会「大規模データの公開におけるプライバシー保護の理論と応用」
4. 発表年 2020年

1. 発表者名 佃 康司
2. 発表標題 Pitman--Yor事前過程を用いた場合の事後過程の弱収束について
3. 学会等名 日本数学会2020年度年会
4. 発表年 2020年

1. 発表者名 Koji Tsukuda
2. 発表標題 A change detection procedure for an ergodic diffusion process
3. 学会等名 13th International Conference on Computational and Financial Econometrics (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 佃 康司
2. 発表標題 Ewens分割の長さに対する正規近似の誤差評価
3. 学会等名 日本数学会2019年度秋季総合分科会
4. 発表年 2019年

1. 発表者名 佃 康司
2. 発表標題 Ewens分割の長さに対する正規近似の誤差評価
3. 学会等名 2019年度統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 Koji Tsukuda, Yoichi Nishiyama
2. 発表標題 Goodness-of-fit tests for Markovian processes based on marked empirical processes
3. 学会等名 3rd International Conference on Econometrics and Statistics (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 佃 康司, 松浦 峻
2. 発表標題 高次元正規ベクトルの分散共分散行列の比例性に関する2標本検定
3. 学会等名 応用統計学会2019年度年会
4. 発表年 2019年

1. 発表者名 佃 康司
2. 発表標題 Pitman--Yor 事前過程を用いた場合の事後過程の弱収束について
3. 学会等名 研究集会「公的データの利用とプライバシー保護の理論」
4. 発表年 2019年

1. 発表者名 佃 康司, 西山 陽一
2. 発表標題 マーク付き経験過程に基づくマルコフ過程の適合度検定
3. 学会等名 日本数学会2019年度年会
4. 発表年 2019年

1. 発表者名 佃 康司, 間野 修平
2. 発表標題 Poisson-Dirichlet分布に従う母集団からの標本が保持する情報量について
3. 学会等名 日本数学会2018年度秋季総合分科会
4. 発表年 2018年

1. 発表者名 Koji Tsukuda
2. 発表標題 An error bound for the normal approximation to the length of a Ewens partition
3. 学会等名 国際ワークショップ「Pioneering Workshop on Extreme Value and Distribution Theories in Honor of Professor Masaaki Sibuya」 (招待講演)(国際学会)
4. 発表年 2019年

1. 発表者名 佃 康司
2. 発表標題 Ewens分割の長さの正規近似について
3. 学会等名 研究集会「公的大規模データの利用におけるプライバシー保護の理論と応用」
4. 発表年 2018年



1. 発表者名 佃 康司
2. 発表標題 Ewens抽出公式に対する正規近似とポアソン近似
3. 学会等名 科研費シンポジウム「融合する統計科学」
4. 発表年 2018年

1. 発表者名 佃 康司
2. 発表標題 Ewens--Pitman抽出公式に関連したいくつかの話題について
3. 学会等名 合同研究会「公的大規模データの利用におけるプライバシー保護の理論と応用」
4. 発表年 2018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
---------	---------