

令和元年6月22日現在

機関番号：62603

研究種目：基盤研究(C) (一般)

研究期間：2015～2018

課題番号：15K05013

研究課題名(和文) 分割の確率モデルとベイズ的データ解析への展開

研究課題名(英文) Stochastic models for partitions and the developments in Bayesian data analysis

研究代表者

間野 修平 (Shuhei, Mano)

統計数理研究所・数理・推論研究系・准教授

研究者番号：20372948

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)：自然数分割の確率モデル、もしくはランダムヤング図形は、ベイズ統計や分類に関わるデータ解析の基礎になります。本研究では確率論と代数的枠組みに基づいて、モデルを探求し、計算機を援用したデータ解析の方法論を与えました。ある分割に双対な跳躍過程について推移確率密度の双直交展開を与えました。Bell多項式により特徴づけられる分割のクラスの統計的推測をGelfandらによるA超幾何系に着目して考察し、ポリトープの情報幾何に基づいて最尤推定を議論しました。サンプラーはデータ解析に欠かせません。微分作用素環のグレブナー基底を用いて分割を含む広いクラスの離散確率モデルについて直接抽出のアルゴリズムを与えました。

研究成果の学術的意義や社会的意義

自然数分割の確率モデルの研究では興味の重点が理論的性質にありましたが、本研究は計算機を援用したデータ解析の方法論を考察したことに特徴があります。特に、本研究では分割を含む広いクラスの離散確率モデルについて直接抽出のアルゴリズムを与えました。このことにより従来はマルコフ連鎖モンテカルロ法(MCMC)が唯一の選択肢と考えられてきたモデルからのサンプリングについても原理的には直接抽出が可能になり、MCMCが目的とする分布をマルコフ連鎖の定常分布として実現することに起因する原理的限界を直接抽出により打破できる可能性が示されました。

研究成果の概要(英文)：Stochastic models of partitions, or random Young diagrams, are bases of Bayesian data analysis and classification. Based on probabilistic and algebraic frameworks, I have investigated models and given methods for computer-aided data analysis. For a jump process which is a dual to a process of partition, a bi-orthogonal expansion of the transition density was provided. For a class of partitions associated with Bell polynomials, the statistical inference was considered with focusing on the A-hypergeometric system defined by Gelfand et al. The maximum likelihood estimation was discussed by using the information geometry of polytopes. Samplers are inevitable for data analysis. By using Groebner bases of a ring of differential operators, a direct sampling algorithm was provided for a broad class of discrete stochastic models including partitions.

研究分野：統計数学

キーワード：統計数学 応用数学 データ解析 組み合わせ確率過程 計算代数 代数統計 ベイズ統計

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

## 1. 研究開始当初の背景

自然数分割の確率モデル，もしくはランダムヤング図形は，ノンパラメトリックベイズ統計や分類に関わるデータ解析の基礎になります．よく知られている Dirichlet 過程は様々な特徴づけが可能です．サイズの分布は Poisson-Dirichlet 分布とよばれ，ある測度値拡散過程の定常分布です．その標本は Ewens 分割とよばれます．拡散過程の双対は自然数分割の確率過程である Kingman's Coalescent で，Ewens 分割はその定常分布です．半世紀以上の研究の蓄積があるのでモデルの研究の拡がりがありますが，主に確率論の視点から議論されていて，データ解析の視点が不足していました．より具体的には，興味の重点が閉じた表示を得ることにあり，Dirichlet 過程のような性質の良いものを除いて，計算はほとんど議論されてこなかったということです．

## 2. 研究の目的

確率論と代数的枠組みに基づいて，モデルを探求するとともに，計算機を援用したデータ解析の方法論を提案することを目指しました．

## 3. 研究の方法

モデルの探求については，拡散過程と分割の確率過程の双対性や確率分割の組み合わせ論的な特徴づけを手がかりとしました．データ解析については，モデルの探求により得られた知見を活かして，ベイズ的方法論，機械学習，計算代数を適用することで，閉じた表示を得ることにはこだわらず，実用的な方法論を目指しました．

## 4. 研究成果

前世紀末に Lambda-coalescent とよばれる自然数分割の確率過程のクラスが提案されました．上述の双対な拡散過程に関する古典的結果に直交多項式による推移確率密度の展開があります．本研究では，ある単純な Lambda-coalescent の双対の跳躍過程の推移確率密度の双直交展開を超関数により陽に与えました (Griffiths and Mano 2016)．Dirichlet 過程以外の Lambda-coalescent に関連した推移確率密度を陽に与えた初めての例と思います．

続いて，Bell 多項式により特徴付けられる自然数分割のクラスである Gibbs 分割の統計的推測について考察しました (Mano 2017 の 2 論文)．Gibbs 分割の良く知られた例に，二母数 Poisson-Dirichlet 過程からの標本があります．Bell 多項式が Gelfand らにより導入された A 超幾何多項式であることに着目したことが鍵になりました．分割については A 超幾何イデアルと Newton ポリトープの性質が簡明であることから，様々な顕著な性質を具体的に得ることができました．統計的推測は母数空間が複雑なので，十分統計量による条件付分布を考察しました．Newton ポリトープ上の情報幾何を用いて最尤推定量の非存在を示し，曲指数型分布族の最尤推定について，Paffian 系とよばれる超幾何関数の隣接関係から従う関係式を導き，ホロノミック勾配法とよばれる数値計算の手続きを得ました．

サンプラーは，仮説検定や事後分布からの抽出など，データ解析の多くの局面で欠かせません．本研究では，Gibbs 分割からのサンプリングのアルゴリズムについて，多項式環のグレブナー基底からマルコフ連鎖モンテカルロ法 (MCMC) の推移の基底を得る手続きと微分作用素環のグレブナー基底から直接抽出法を導く手続きを示し，A

超幾何系に着目することで、この類似が自然数分割のみならず極めて広いクラスの離散確率モデルに適用できることを指摘しました (Mano 2017 *Electron. J. Stat.*). その典型にトリークモデルとよばれる離散グラフィカルモデルを含むクラスがあります。自然数分割でいえば交換可能性のないモデル, グラフィカルモデルでいえば対立仮説や分解不可能なモデル, そのようなモデルのサンプリングは従来 MCMC が唯一の選択肢と考えられてきましたが, それらについても直接抽出が可能であるということです。この類似は理論的にも興味深いですが, MCMC には目的とする分布をマルコフ連鎖の定常分布として実現することに起因する, 定常に達していることを保証することが難しいこと, 自己相関の存在といった原理的に避けられないといった欠点があることを考えると, 従来のサンプラーの限界を直接抽出により打破できる可能性が示されたという意味で実用的に重要な意義があります。

本研究の最終年度である平成 30 年度は, それまでに得られた成果をモノグラフとして出版しました (著書)。この書籍は 5 章から成りますが, 4 章までが準備で, 5 章が本研究で得られたデータ解析の方法論に関する成果の紹介です。最後に, 雑誌論文・書籍としては未発表ですが, 本研究の成果として得られたことを簡単に述べます。上に述べた直接抽出法については, 計算代数の観点からの効率化を高山信毅氏と検討を重ねていて, 各種学会や数理解析研究所講究録で発表しました。上に述べていないこととしては, 公的データの開示に伴うリスクの評価や波形データの分類などへの応用についても考察し, 研究集会で発表しました。また, 不連続曲線のベイズ的適合への応用として曲線を性質が異なる曲線に分割する研究も行い, Jaeyong Lee 氏の学生であった Youngseon Lee 氏が国際会議で発表しました。

## 5. 主な発表論文等

### 〔雑誌論文〕(計 3 件)

1. Shuhe Mano. 2017. Partition structure and the A-hypergeometric distribution associated with the rational normal curve. *Electronic Journal of Statistics* 11: 4452-4487 査読あり
2. Shuhe Mano. 2017. Extreme sizes in the Gibbs-type exchangeable random partitions. *Annals of the Institute of Statistical Mathematics* 69: 1-37 査読あり
3. Robert Griffiths, Shuhe Mano. 2016. The star-shaped Lambda-coalescent and Fleming-Viot process. *Stochastic Models* 32: 606-631 査読あり

### 〔学会発表〕(計 29 件)

1. 高山信毅, 間野修平, 計算代数の direct sampler への応用, RIMS meeting: Computer algebra - Theory and its applications, Research Institute of Mathematical Sciences, Kyoto University, 2018 年 12 月 19 日
2. 間野修平, Ewens 分割の  $q$  類似とその性質, 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」統計数理研究所, 2018 年 12 月 14 日
3. 間野修平, Posterior sampling from some non-exchangeable priors, シンポジウム「融合する統計科学」金沢大学, 2018 年 11 月 30 日

4. Shuhe Mano, Nobuki Takayama. Direct Sequential Sampler for A-hypergeometric Distributions, IASC-ARS and CASC joint conference, Beijing Conference Center, November 11, 2018
5. Shuhe Mano, Sampling from random partitions via A-hypergeometric systems associated with monomial curves, Symposium: Stochastic Processes and Risk Analysis. Institute of Statistical Mathematics, October 19, 2018
6. 間野修平, 分割, 超幾何系, Dirichlet 過程と統計的推測, 日本数学会 2018 年度秋季総合分科会, 岡山大学, 2018 年 9 月 26 日
7. Shuhe Mano, Nobuki Takayama, N. A direct sampler from A-hypergeometric distributions, 2018 年度統計関連学会連合大会, 中央大学, September 12, 2018
8. Shuhe Mano, Sampling from non-exchangeable priors, EAC-ISBA 2018 Conference, Seoul National University, July 12, 2018
9. Shuhe Mano, Samplers for non-exchangeable prior processes and hypergeometric systems, The 5th IMS-APRM meeting, National University of Singapore, June 27, 2018
10. Shuhe Mano, S. Direct sampler from A-hypergeometric distribution for count data analyses, ISM Symposium on Environmental Statistics 2018, Institute of Statistical Mathematics, March 22, 2018
11. 間野修平, A 超幾何分布からの直接抽出とランダム Young 図形への応用, 日本数学会 2018 年度年会, 東京大学, 2018 年 3 月 19 日
12. Shuhe Mano, Partition structure and the A-hypergeometric distribution, Current Topics on Algebraic Statistics and Related Fields, Kobe University, March 8, 2018
13. 間野修平, 交換可能でない分割からの抽出, 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」統計数理研究所, 2017 年 12 月 14 日
14. Youngseon Lee, Shuhe Mano., Jaeyong Lee. Bayesian curve fitting for discontinuous functions using overcomplete system with multiple kernels, IASC-ARS/NZSA 2017 Conference, University of Auckland, December 11, 2017
15. 間野修平, A 超幾何系とランダム Young 図形からの正確な抽出, 確率・統計・行列ワークショップ松本 2017, 信州大学, 2017 年 11 月 9 日
16. 間野修平, カウントデータ解析のための A 超幾何分布からの正確な抽出法, 研究集会「環境・生態データと統計解析」統計数理研究所, 2017 年 10 月 6 日
17. 間野修平, ベイズ統計における交換可能性と超幾何系, 超幾何学校 2017 統計と超幾何の再会 in 小樽, 2017 年 9 月 5 日
18. Shuhe Mano, Prior processes and A-hypergeometric systems, 11th Conference on Bayesian Nonparametrics, Ecole Normal Superieure, Paris, June 27, 2017
19. 間野修平, 分割の可乗測度と有理正規曲線に付随する A 超幾何系, 日本数学会年会 2017 年度年会, 首都大学東京, 2017 年 3 月 26 日
20. Shuhe Mano, KAGRA collaboration, Diagnosis of gravitational wave interferometer by iKAGRA glitch data with Dirichlet process mixture model, The 5th annual symposium of the innovative area “New development in astrophysics through multi-messenger observations of gravitational wave sources”, Yukawa Institute of Theoretical Physics, Kyoto University, December 28, 2016

21. 間野修平, 分割の可乗測度と可換代数, 確率論シンポジウム, 京都大学数理解析研究所, 2016年12月20日
22. 間野修平, 開示制御への代数的接近 -分割の可乗測度と整数計画の緩和-, 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」統計数理研究所, 2016年12月8日
23. Shuhe Mano, Exchangeable combinatorial structures and their statistical inferences, Yonsei Statistical Science Institute Seminar, Department of Statistics, Yonsei University, November 3, 2016
24. 間野修平, 有理正規曲線に付随する A 超幾何系と統計的推測, 2016年度統計関連学会連合大会, 金沢大学, 2016年9月5日
25. Shuhe Mano, Exchangeable Random Structures and the A-Hypergeometric System associated with the Rational Normal Curve, Workshop: Algebraic Statistics and Symbolic Computation, Research Institute of Mathematical Sciences, Kyoto University, July 27, 2016
26. 間野修平, Robert Griffiths, The star-shaped Lambda-coalescent and Fleming-Viot process, 無限粒子系と確率場の諸問題 11, 奈良女子大学, 2015年12月26日
27. 間野修平, Gibbs 確率分割における統計的推測と A 超幾何系, 研究集会「官庁統計データの公開における諸問題の研究と他分野への応用」統計数理研究所, 2015年12月11日
28. 間野修平, Gibbs 確率分割における統計的推測への代数的接近, 統計関連学会連合大会. 岡山大学, 2015年9月7日
29. Robert Griffiths, Shuhe Mano. "Some properties of the Lambda= $\delta_1$ -Fleming-Viot process", 38th Conference on Stochastic Processes and Their Applications, University of Oxford, July 13-17, 2015

〔図書〕(計1件)

Shuhe Mano. 2018. Partitions, Hypergeometric Systems, and Dirichlet Processes in Statistics, JSS Research Series in Statistics, SpringerBriefs in Statistics, Springer, 135 pages

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

〔その他〕

ホームページ

[https://researchmap.jp/mano\\_shuhe](https://researchmap.jp/mano_shuhe)

## 6. 研究組織

(1) 研究分担者

なし

(2) 研究協力者

なし