

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 6 月 12 日現在

機関番号： 12401
 研究種目： 基盤研究 (C)
 研究期間： 2009～2012
 課題番号： 21500271
 研究課題名 (和文) 離散確率空間上の MCMC 法の大域化による精度向上と最適値への収束性
 研究課題名 (英文) Global accurate or optimal computation for MCMC method
 over Discrete random space

研究代表者

橋口 博樹 (HASHIGUCHI HIROKI)
 埼玉大学・理工学研究科・准教授
 研究者番号： 50266920

研究成果の概要 (和文)：

マルコフチェーンモンテカルロ法 (MCMC 法) の大域化の方法を議論し、主に二つの問題に適用した。一つ目は分割表解析における確率計算、もう一つは過飽和実験計画の構成問題である。大域化によってモンテカルロ計算の精度が上がったことが確認され、また、過飽和実験計画の構成問題においても最適計画を構成できることが確認された。以上により、大域化の方法が MCMC 法に有効であることが、二つの問題に対する数値実験により検証できた。

研究成果の概要 (英文)：

We discussed the global accurate or optimal computation for the Markov chain Monte Carlo method and applied two practical problems. One is for calculating approximate probability in the analysis of contingency table, the other for constructing supersaturated designs. We confirmed that our global computation method worked well for the above two problems through the numerical experiments.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	1,500,000	450,000	1,950,000
2010 年度	500,000	150,000	650,000
2011 年度	600,000	180,000	780,000
2012 年度	700,000	210,000	910,000
年度			
総計	3,300,000	990,000	4,290,000

研究分野：複合領域

科研費の分科・細目：情報学・統計科学

キーワード：マルコフ連鎖モンテカルロ法, 過飽和実験計画, 分割表解析, 固有値分布

1. 研究開始当初の背景

マルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo method: MCMC 法) は、条件付分布からのサンプリングによる乱数を用いたモンテカルロ法の一つであり、近年多くの分野で注目を集めている。MCMC 法の適用分野は、計量経済、マーケティング、空間統計、画像処理など多岐にわたる。

本研究で考慮する MCMC 法の性質は二つある。一つは分布を近似することであり、もう一つは最適化問題の最適解を見つけるために使うことである。本研究で構築する大域化のアイデアは、分布の近似と最適化問題の両方に適用できる汎用性の高い方法論である。実用性を検証するために、以下の二つの問題を扱う。分布近似の問題としては、分割表解析における p 値計算を扱う。

幾何学的な構造が単純な2元分割表での実験と、より幾何学構造が複雑なグラフィカルモデルでの実験を行う。最適化問題の例としては、実験計画の構成問題を扱う。実験計画の古典的な方法論は代数的、組合せ論的な方法により既知であるので、今なお最適性に未知の問題を抱えている過飽和実験計画の構成を取り上げる。

2. 研究の目的

研究課題として、確率空間を離散確率の空間へ絞り、MCMCの大域化とその現実問題への適用を掲げ、次の4つに取り組んだ。

(1) 大域化理論の構築

MCMC法の大域化理論という新しい理論を提案する。この離散確率空間からの一様サンプリングと古典的なメトロポリス法を融合させる。

(2) 大域化理論の分布計算への応用

分割表(多次元の整数格子点)上をメトロポリスウォークする確率分布の精度検証を行う。

(3) MCMCの過飽和実験計画の構成問題への適用。MCMCによる過飽和実験計画の構成法を提案し、既存の構成ソフトウェアGendexとの比較を行う。大域化によってより良い解が得られることを数値的に検証する。

(4) 大域化理論とネットワークの固有値解析。大域化の理論的な解析には、ランダムウォークの移動によって作るネットワークの解析が必要となる。2,3の応用例では、離散確率空間の状態数が莫大であるため、遷移状態をグラフやネットワーク図で表すことは困難である。さらに、全ての状態数に関する遷移確率を決めることも困難である。しかし、状態数が数十程度で遷移確率を求め、固有値解析により大域化した場合の性質を明らかにしておくことは必要であると考えられる。第2固有値の値が1に近ければ収束が速いことが知られているので、大域化によりこの性質が生まれることを検証する。

3. 研究の方法

本研究は4つのテーマ[1] 大域化理論の構築, [2] 大域化理論の分布計算への応用, [3] 大域化理論の最適化問題への応用, [4] 大域化理論とネットワークの固有値解析から成る。代表者橋口が研究総括を行うとともに、研究体制にあるように役割分担を明確にして研究を遂行した。

○ 橋口博樹(埼玉大学) 研究代表者
役割: 研究総括。大域化理論の構築とその理論的評価[テーマ1], [4]。

○ 中川重和(倉敷芸術科学大学) 分担者
役割: 数値実験の検証[テーマ2], [3] と遷移行列の固有値解析[テーマ4]

○ 黒田正博(岡山理科大学) 連携研究者
役割: 分割表とグラフィカルモデルでのp値計算[テーマ2]に関するアドバイザー。

○ 山田秀(筑波大学) 連携研究者
役割: 過飽和実験計画の構成と解析方法に関するアドバイザー。

また、各種プログラムの実装は、研究協力者(5名から8名/年)が行った。

4. 研究成果

2009年度は、MCMC法の大域化の基本的アイデアをまとめるとともに、次年度以降の数値実験のためのプログラミング環境を整えた。具体的には、分布計算のためには、サンプリングの局所性により自己相関が高いことで近似が悪い場合や、最適化法として利用した場合に局所的な解に落ち込むことなどを、比較的小規模な例題を使って検証し、そのためのプロトタイプを実装した。過飽和実験計画構成のためのプログラミング環境を整え、C++言語でのクラスを設計した。

2009年度に数値検証のためのクラス設計などしたことを受けて、2010年度は、実験計画の構成を例とした最適化問題への大域化の方法論を提案した。実験計画の構築問題では、第一ステップとして、列間の非直交度を最小化するように乱択によって列追加を行うことで初期計画を生成し、第二ステップとして、得られた初期計画の列の要素を入れ替えることで直交性の改善を行うという二段階最適化法を提案した。第一ステップは、マルコフ連鎖モンテカルロ法を基本アイデアとして条件付きで独立な列ベクトルをサンプリングすることに対応する。さらに、第二ステップが大域性を担保し、局所解からより良い解へ改善するためのサンプリングに対応する。この方法の有効性を検証するために、複数の論文で有用性が示されている既存の構成法との比較を行った。比較実験した全ての場合において、提案法は既存の構成法と同等かもしくはより良い計画が得られることを示した。この成果を日本計算機統計学会の和文誌に投稿し、論文として採択された。また、ISBIS-2010で招待講演として報告した。

2011年度は、log-liner modelでの大域化の検証実験を行い、既存の方法と比較して精度が向上することを確認した。マルコフ連鎖が既約であること、および、既約性を保証した上で離散確率空間の範囲を広げるといふことの大域化の方法が、精度向上に結びつく

ことを確認した。この成果を論文投稿し、Computational Statistics で採択された。また、実験計画の構築問題を多角的に検討するため、固有値解析の観点での定式化や、有限体を利用する代数的方法の検討も前年度からの引継として研究した。実験計画の構築問題では、有限体を利用する方法を検討し、最適実験計画の構成法を提案し最適性の証明を行なった。さらに、2010年度に大域化の提案として、二段階最適化法を論文投稿し採択されていたが、二段階最適化法で構成したいくつか最適になることも、有限体での構成法との比較で確認することができた。構成した最適混合水準過飽和計画をコンピュータシミュレーション実験でのスクリーニングに適用し、有効性も確認した。

2011年度までの成果では、数値実験を通して本理論の有効性を示してきた。2011年度は、固有値分布論の観点から有効性を示す理論的な研究に取り組んだ。固有値分布論の統計学における一般的な議論を行ない、その成果を国際誌に投稿し採択された。固有値分布の統計学における一般的な議論では、最大固有値分布の正確な計算に成果や、最小固有値分布の近似計算、複素正規分布論への応用などの成果があり、それぞれ発表を行った。特に、最大固有値の正確計算では新しい方法を提案することができた。また近似分布について、精度が非常に良いことが分かった。マルコフ連鎖モンテカルロ法の収束の議論は2番目に大きい固有値が関係しており、前述の固有値分布に関する成果を、マルコフ連鎖モンテカルロ法の理論解析に適用するまでには至らず、今後の課題として残ってしまった。しかし、個々の問題を詳細に検討した結果、関連するテーマにおいて多くの成果が生まれた。例えば、これまでの研究成果を理論的にも検討する観点から、過飽和実験計画の構成では、最適性が保証できるパラメータを理論的に探し、そのパラメータでの構築を行い、大域化によるマルコフ連鎖で構成した乱数による実験計画も最適になっていることも確かめられた。

以上、結論と今後の課題は次のようになる。

(1) 大域化のための統一的な理論を提案できなかったものの、当初予定していた分割表解析や実験計画の構成では、大域化の方法がうまく機能することを示した。

(2) 大域化の方法の数値的な評価を行うために、2つの具体的な問題で検証し、有効性を確認した。理論的な評価を固有値分布と関連づけて行おうとしたが、今後の課題として残った。しかし、固有値分布論の統計学の枠組みでの新たな展開を示すことはできた。

(3) 大域化の方法において、最適化問題への適用を実験計画の構成を例に取り組んだ。最適解を乱数に頼らない理論的に構築する

方法を開発し、大域化の方法でも最適解が得られることを確認した。

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 8 件)

- ① Kuroda, M., Hashiguchi, H., Nakagawa, S. and Geng, Z., MCMC using Markov bases for computing p-values in decomposable log-linear models. Computational Statistics, 2013, Volume 28, Issue 2, pp 831-850 [査読有]
- ② Hiroki Hashiguchi, Yasuhide Numata, Nobuki Takayama, Akimichi Takemura, The holonomic gradient method for the distribution function of the largest root of a Wishart matrix, Journal of Multivariate Analysis, 117, 2013, 296-312. [査読有]
- ③ Shigekazu Nakagawa, Hiroki Hashiguchi, Naoto Niki, Improved omnibus test statistic for normality, Computational Statistics, Comput Stat (2012) 27:299-317. [査読有]
- ④ 大野貴慎, 橋口博樹, 山田秀, 二段階最適化法による混合水準過飽和計画の構成, 日本計算機統計学会, 計算機統計学 24(1), pp.23-35, 2011-10-01, 2011 [査読有]
- ⑤ N. Niki, M. Iwata, H. Hashiguchi, S. Yamada, Optimal selection and ordering of columns in supersaturated designs, Journal of Statistical Planning and Inference, Volume 141, Issue 7, 2011, Pages 2449-2462. [査読有]
- ⑥ Masahiro Kuroda, Hiroki Hashiguchi, Shigekazu Nakagawa. Computing p-values in conditional independence models for a contingency table. Computational Statistics, Volume 25, pp. 57-70, 2010 [査読有]
- ⑦ Y. Kakinuma, K. Hiraoka, H. Hashiguchi, Y. Kuwajima, T. Shigehara, Algorithm for computing Kronecker basis, JSIAM Letters 1, 2009, pp.60-63 [査読有]
- ⑧ 橋口博樹, ゴーナル多項式の数式処理と最小固有値分布の計算, 数式処理, Volume 16, Number 2, 23-31, 2009 [査読無]

[学会発表] (計 28 件)

- ① Shigekazu Nakagawa, Hiroki Hashiguchi,

Naoto Niki, On a bivariate density function for the distribution of the skewness and kurtosis statistics, *ims-APRM 2012*, Tsukuba July 2--4, 2012 (7/3 発表), *ims-APRM 2012*, Abstract, p. 125.

- ② Hiroki Hashiguchi, Takakazu Sugiyama, Approximate χ^2 distributions for the distribution of the smallest latent root of a Wishart matrix, *ims-APRM 2012*, Tsukuba July 2--4, 2012 (7/3 発表), *ims-APRM 2012*, Abstract, pp. 126--127.
- ③ Hasegawa, M., Kato, H. and Hashiguchi, H., Approximation for eigenvalue distributions of a real or complex Wishart matrix, *ims-APRM 2012*, Tsukuba July 2--4, 2012 (7/3 発表), *ims-APRM 2012*, Abstract, p. 288--289. (Poster Presentation)
- ④ 長谷川将義, 橋口博樹, 複素ウィシャート行列における最小固有値分布のガンマ分布近似, 計算機統計学会 5/8 発表, 函館福祉センター, 2011.
- ⑤ 加藤仁志, 橋口博樹, 特異ウィシャート行列の最小固有値分布の近似について, 計算機統計学会 (5/8 発表), 計算機統計学会第25回大会講演論文集 117--118, 函館福祉センター, 2011.
- ⑥ Hiroki Hashiguchi, Random generation for constructing supersaturated designs, ISBIS-2010, July 5 - 9, 2010, Portoroz, Slovenia, 7/7 発表
- ⑦ 橋口博樹, 固有値分布論における対称式の数式処理, 応用統計学会, 5/20 発表, 統計数理研究所, 2010.
- ⑧ 角昌竜, 橋口博樹, 山田秀, 有限体に基づく過飽和計画の構成について, 日本計算機統計学会 第24回シンポジウム 11/11-12(11/11 発表), 箕面市立メイプルホール, 2010.
- ⑨ 藤島徹, 橋口博樹, 山田秀, 最適計画の結合による混合水準過飽和計画の構成について, 日本計算機統計学会 第24回シンポジウム, 11/11-12(11/11 発表), 計算機統計学会第25回大会講演論文集 113-116, 箕面市立メイプルホール, 2010.

[その他]

ホームページ等

<http://www.tus.ac.jp/ridai/doc/ji/RIJIA>

01User.php?act=nam&kin=pap&diu=1f6e

6. 研究組織

(1) 研究代表者

橋口 博樹 (HASHIGUCHI HIROKI)
埼玉大学・理工学研究科・准教授
研究者番号：50266920

(2) 研究分担者

中川 重和 (NAKAGAWA SHIGEKAZU)
倉敷芸術科学大学・産業科学技術学部・教授
研究者番号：90248203

(3) 連携研究者

黒田 正博 (KURODA MASAHIRO)
岡山理科大学・総合情報学部・准教授
研究者番号：90279042

山田 秀 (YAMADA SHU)

筑波大学・ビジネス科学研究科・教授
研究者番号：60260965