

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 5月16日現在

機関番号：13101

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500252

研究課題名（和文）計算代数的手法を用いた大規模多変量推測の理論及び実用アルゴリズムに関する研究

研究課題名（英文）Theory and practice for inference of high dimensional statistical model with computational algebraic geometry.

研究代表者

原 尚幸 (HARA HISAYUKI)

新潟大学・人文社会・教育科学系・准教授

研究者番号：40312988

研究成果の概要（和文）：本研究では離散マルコフ連鎖モデル、Split モデルなどのいくつかの実用的な分割表モデルのマルコフ基底の理論的に導出し、それを用いた正確検定アルゴリズムを提案した。一般には離散指数型分布族のマルコフ基底の理論的な導出は困難とされるが、本研究ではマルコフ基底が得られない場合にも、計算が容易なマルコフ基底の部分集合や、格子基底を用いることにより、マルコフ基底を用いた場合と同等の精度の正確検定の実装が可能になることを示した。

研究成果の概要（英文）：

In this study we derived Markov bases of some practical models for contingency tables such as homogeneous Markov chain model and split model and proposed practical algorithms of exact tests by using these Markov bases. In general, however, it is not easy to obtain an exact list of a Markov basis for a discrete exponential family model. We gave some practical algorithms of exact test also for such models by using some practical subsets of Markov bases or lattice bases and checked that the performances of proposed algorithms are comparable to the ones by using Markov bases.

交付決定額

（金額単位：円）

	直接経費	間接経費	合 計
2010 年度	1,000,000	300,000	1,300,000
2011 年度	700,000	210,000	910,000
2012 年度	700,000	210,000	910,000
総 計	2,400,000	720,000	3,120,000

研究分野：情報学

科研費の分科・細目：情報学基礎

キーワード：分割表、グラフィカルモデル

1. 研究開始当初の背景

Diaconis and Sturmfels (1998) によるマルコフ基底を用いた正確検定アルゴリズムの提案と、マルコフ基底の計算代数的考察以降、さまざまな分割表モデルのマルコフ基底の構造解析に関する研究が統計学、代数学、さらには応用の視点から行われてきた。しかしながら、本研究開始当初までに得られていた

理論的成果は、構造が簡単なごく一部のモデルや、サイズの小さい分割表のモデルの場合に限られていた。正確検定はモデルの自由度が標本サイズに比して大きい状況で本来威力を発揮することを考えれば、こうした状況はやや不満足なものであった。

2. 研究の目的

本研究では、より一般のモデル、特に高次元

分割表のモデルへ既存のマルコフ基底の基礎理論を一般化し、実用的な正確検定アルゴリズムを提案することを目的とした。

3. 研究の方法

上記のような研究背景を踏まえ、本研究では主として以下の3つのアプローチで研究を行った。

- (1) マルコフ基底を理論的に導出する従来型のアプローチ。
- (2) 特定のファイバーだけを連結に結ぶマルコフ部分基底を求めるアプローチ。
- (3) モデルの move が張る整数格子の格子基底を用いるアプローチ。
- (4) セル頻度に制約のあるファイバーを連結に結ぶ move の集合を導出するアプローチ

4. 研究成果

上記の各アプローチについて以下のような成果を得た。

- (1) 有限状態定常マルコフ連鎖モデル、split モデルなどの実用的な分割表モデルのマルコフ基底を理論的に導出するとともに、move の適応的生成アルゴリズムを提案し、それを用いた正確検定アルゴリズムを提案した。また実データに提案手法を適用し、その有用性を確認した。
- (2) 離散ロジスティック回帰モデルのマルコフ基底は高次の複雑な構造の move を含み、また要素数も水準数に対し、指数的に増加するため、理論的な導出が極めて困難であることが知られている。本研究では、説明変数が2以下の場合、説明変数が3で3つともダミー変数(水準数が2)の場合に、各説明変数の組に対する試行回数が正という制約を満たすファイバーに限れば、単純な構造の4次の move のみで連結に結ぶことが可能であることを理論的に証明し、それを用いた正確検定アルゴリズムの提案を行った。また実データに提案手法を適用し、その有用性の確認も行った。
- (3) 任意モデルには move が張る整数格子に、計画行列の核の次元数の格子基底が存在することが知られている。そのとき任意の move は、格子基底の要素の整係数線形結合で表すことが可能である。このことは、格子基底の要素の任意の整係数線形結合が正の確率で発生するように move の生成を行えば、任意のファイバーを連結に結ぶことが可能になることを意味する。一方、格子基底は計画行列の整数核なので、どんなモデルで

あっても容易に計算が可能である。本研究では特にマルコフ基底の導出が困難とされる、無三因子交互作用モデル、離散多項ロジットモデルなどにおいて、格子基底を用いた正確検定アルゴリズムを提案し、実用的であることを数値実験を通して示した。

- (4) 項目反応理論のモデルや、社会ネットワークのモデルは、セルの頻度が0 or 1 に制約された分割表のモデルであると考えることができる。このような分割表を0-1分割表と呼ぶことにしよう。0-1分割表をmoveの足し引きによる状態遷移を用いてサンプリングする場合には、当然、頻度が0 or 1 であるという制約を満たしながら状態遷移させる必要がある。しかし一般に、特定の0-1分割表と同一ファイバーに含まれる分割表の中には、2以上の頻度のセルを有する表も含まれる。このことは0-1分割表のサンプリングにはマルコフ基底だけでは不十分であることを示している。本研究では、グレーバー基底と呼ばれる基底の要素の中で0 or 1 or -1しか含まない move のみの集合がこうした制約付きのファイバーを連結に結ぶことを理論的に証明するとともに、正確検定のアルゴリズムの提案を行った。さらにRaschモデルやp1モデルなどの実用的なモデルに対して、この議論を適用し、その有用性を確認した。

これらのアプローチに加え、以下のような成果を得ることもできた。

- (5) 分割表の階層モデルを一般化し、より柔軟な交互作用の構造を取り入れた階層部分空間モデルを提案した。またこのモデルの分解と推論の局所化について議論を行い、通常の階層モデルの持つ条件付き独立関係を表すグラフに基づくモデルの分解を一般化し、最尤推定量の計算やマルコフ基底の計算における実用的な局所化アルゴリズムを提案した。
- (6) 分割表の階層モデル・グラフィカルモデルの最尤推定量の計算は、条件付き独立構造を表すグラフの分離・分解を用いてアルゴリズムが局所化され、計算の効率化が可能であることが知られている。しかしガウスグラフィカルモデルではこうした議論がなされてこなかった。本研究ではガウスグラフィカルモデルの共分散行列の最尤推定量の計算の比例反復法の更新則を、条件付き独立構造の

グラフのコーダル拡張の分解を用いた局所化によって効率化できることを示し、改良型アルゴリズムを提案し、従来型アルゴリズムでは行列サイズの3乗のオーダーだった計算量を、疎なモデルにおいては線形にまで計算オーダーを軽減することに成功した。

- (7) これまでの本課題の研究成果を英文研究書、Markov bases in algebraic statistics にまとめ、昨年7月に刊行された。

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12件)

- (1) Hara, H., Sei, T. and Takemura, A., Hierarchical subspace models for contingency tables, Journal of Multivariate Analysis, 査読有, vol.103, 2012, 19-34.
- (2) Hara, H. and Takemura, A. A Markov basis for two-state toric homogeneous Markov chain model without initial parameters. Journal of the Japan Statistical Society, 査読有, vol. 41, 2011, 33-49.
- (3) Hara, H., Takemura, A. and Yoshida, R., On connectivity of fibers with positive marginals in multiple logistic regression, Journal of Multivariate Analysis, 査読有, vol.101, 2010, 909-925
- (4) Hara, H., Aoki, S. and Takemura, A. Minimal and minimal invariant Markov bases of decomposable models for contingency tables, Bernoulli, 査読有, vol.16, 2010, 208-233
- (5) Hara, H. and Takemura, A. A localization approach to improve iterative proportional scaling in Gaussian graphical models, Communications in Statistics - Theory and Methods, vol. 39, 2010, 1643-1654.

〔学会発表〕(計17件)

- (1) Hisayuki Hara, Markov chain Monte Carlo test of homogeneity of Markov chains. 2nd IMS Asia Pacific Rim Meeting, 2012年7月2日, つくば国際会議場
- (2) 原 尚幸, Running Markov chain without Markov basis, 2011年度統

計学関連学会連合大会、2011年9月5日、九州大学

- (3) 原 尚幸, マルコフ基底と分割表解析への応用について、IBIS2010、2010年11月3日、東京大学(招待講演)
- (4) Hisayuki Hara, Markov subbases for discrete logistic regression models with positive margins. 10th China-Japan Symposium on Statistics, 2010年10月16日 2010, Southwestern University of Finance and Economics, Chengdu, China.
- (5) Hara, H., Hierarchical subspace models for contingency tables, 73rd IMS Annual Meeting, 2010年8月10日, Chalmers University of Technology, Goteborg, Sweden.

〔図書〕(計1件)

- (1) Aoki, S., Hara, H. and Takemura, A. Springer, Markov bases in Algebraic Statistics, 2012, 300 pages.

〔産業財産権〕

○出願状況(計0件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

○取得状況(計0件)

名称:
発明者:
権利者:
種類:
番号:
取得年月日:
国内外の別:

〔その他〕

ホームページ等

6. 研究組織

(1)研究代表者

原 尚幸 (HARA HISAYUKI)

新潟大学・人文社会・教育科学系・准教授

研究者番号: 40312988

(2)研究分担者

なし

(3)連携研究者

竹村 彰通 (TAKEMURA AKIMICHI)

東京大学・情報理工学系研究科・教授

研究者番号：1 0 1 7 1 6 7 0

栗木 哲 (KURIKI SATOSHI)

統計数理研究所・数理推論研究系・教授

研究者番号：9 0 1 9 5 5 4 5

清 智也 (SEI TOMONARI)

慶應義塾大学・理工学研究科・講師

研究者番号：2 0 4 0 1 2 4 2