

機関番号：17102

研究種目：基盤研究（B）

研究期間：2008～2010

課題番号：20340021

研究課題名（和文）計算代数統計学的手法による高次元配列データ解析法の研究とその画像統計への応用

研究課題名（英文）A study of analysis of high dimensional array data through computational algebraic statistical methods and its application to statistical image analysis

研究代表者

坂田 年男（SAKATA TOSHIO）

九州大学・芸術工学研究院・教授

研究者番号：20117352

研究成果の概要（和文）：

高次元分割表に対する逐次条件付き検定における推論の枠組みである周辺和一定の分割表集合をマルコフ基底により時系列的に構築することができることを近傍定理を証明することにより示した。また高次元テンソルデータの最大階数決定問題に対して、アトキンソンらの主張の部分証明、絶対正則テンソルの定義を与え、特徴づけ、探索および構築法などを明らかにした。また絶対正則なテンソル間の非同値性の検証法を提案しその有効性を明らかにした。加えてランダム行列の最大固有値分布論および高次元データを扱う地理・空間統計学の応用として森林面積減少の解明や地理・空間画像データの分類などに成果を挙げた。

研究成果の概要（英文）：

In the sequential exact conditional test for three-way contingency tables lifting problem is studied. The problem studied is how to construct the inferential frame (the set of all contingency tables with the same marginals as a given datum) at the time t from that of the time $t-1$. We made clear by r -neighborhood theorem that the frame at the time t is constructible from the frame at the time $t-1$ by using Markov basis. On the other hand for the real valued three dimensional datum, that is, 3-tensor, we studied the rank and the maximal rank. Especially, we proved Atkinson's claim for the complex number fields with no condition and proved it over the real number field with some condition. We called tensors, which does not satisfy the condition, as absolutely nonsingular tensors. For studying absolutely nonsingular tensors we devised the determinant polynomial of tensors and made clear the link between the absolutely non-singularity and the positivity of the determinant polynomial. We obtained methods how to find and how to construct absolutely nonsingular tensor. Also, we proposed methods to detect non equivalence between them by using differential geometric invariants and the integrations over the orthogonal or the unitary group. In addition, some results were obtained in the distribution theory of the largest eigenvalue of a random matrix and in the deforestation modeling and the image classification, based on geo-spatial data.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	2,800,000	840,000	3,640,000
2009年度	2,100,000	630,000	2,730,000
2010年度	2,200,000	660,000	2,860,000
総計	7,100,000	2,130,000	9,230,000

研究分野：数理統計学

科研費の分科・細目：数学・数学一般（含確率論・統計数学）

キーワード：計算代数統計学、テンソル型データ解析、テンソルの最大階数、絶対正則テンソ

1. 研究開始当初の背景

(1) 分割表の逐次検定においてマルコフ連鎖モンテカルロ法を使用する場合、ある時点での推測のフレーム（周辺和一定の分割表の全体）と次の時点での推測のフレームは同じではなくどんどん変化していく。ある時点のフレームを一時点前のフレームから構築する問題をリフティング問題と名づけ、近傍定理を考案し、マルコフ基底から構築することができることを示す研究を開始していた。

(2) 高次元配列データであるテンソルデータの複雑性を階数およびモデルの複雑性である最大階数を定めるための様々な研究がある中で、2層のテンソルの最大階数問題および3層のテンソルの最大階数問題、および応用を含めてテンソルデータ解析の諸問題に対する総合的な研究を開始していた。

2. 研究の目的

(1) テンソルデータの基礎的な問題である階数および最大階数決定問題及び関連する問題に対して解決を図りテンソルデータ解析の数学的基礎付けを行うこと。また関連する分布論や高次元データの応用研究を行うこと。

(2) 分割表の逐次条件付正確検定におけるリフティング問題にたいして近傍定理に証明を与えプログラムにより有効性を検証すること。

3. 研究の方法

共同研究者間のセミナーを重ね、また、諸外国の研究者を招聘し情報交換を行うとともに、国際会議国内のシンポジウム等で積極的に成果を発表し、かつ、情報収集を行った。方法論的には分割表の検定、積分による統計量の統計的検定論、正値多項式理論、微分幾何学的不変量理論、数値計算、組みあわせ球面デザイン論などを用いて総合的に問題の解決に当たった。

4. 研究成果

(1) 高次分割表の条件付き正確検定での p 値計算は通常MCMC法で近似的に行う。ここでは、より複雑な「逐次」条件付検定における p 値計算をMCMCではなく、実際に、「すべての周辺和一定の分割表全体」 Ω のなかの表に対して、 p 値への寄与を計算するための手法を研究した。我々は $t-1$ 時点の分割表の集合 $\Omega(t-1)$ を用いて、 t 時点での分割表の集合 $\Omega(t)$ をマルコフ基底を用いて構成することができることを、一定のタイプの3層の分割表に対して、はじめ数値的に、次に理論的に証明した。

(2) ①アトキンソン等の主張「 $n \times n \times 3$ テン

ソルの最大階数は $2n-1$ 以下である」に対して複素数体上の証明を与え、条件付きで実数体上の証明を与えた。その条件とは「 $T=(A;B;C)$ とするとある零ベクトルでない (x, y, z) に対して $xA+yB+zC$ が退化する」であった。

②アトキンソン等の主張の証明から外れた例外テンソルを絶対正則テンソルと名づけた。即ち、「 $T=(A;B;C)$ とすると零ベクトルでないどのようなベクトル (x, y, z) に対しても $xA+yB+zC$ が退化しない」テンソルを絶対正則と名づけた。テンソルに対して行列式多項式 $f(x, y, z) = \det(xA+yB+zC)$ を導入した。

$f(x, y, z)$ は次数 n の斉次多項式となる。絶対正則であることは「行列多項式がゼロを取らないこと」と特徴づけることができる。これは一般性を失うことなく正定値と仮定できる。斉次多項式の正値性はヒルベルトの17問題と呼ばれる分野で研究されており、その中のポリヤの定理を使って $4 \times 4 \times 3$ の場合に絶対正則テンソルを探索するプログラムを作成し多数の例を発見した。統計的問題が行列多項式を導入することで、代数学におけるヒルベルトの17問題と関連することを示したことは意義が高くその後の研究成果に大いに寄与した。

③テンソルの同値性はテンソル $T=(A;B;C)$ への左右からの正則行列 P, Q による変換 $T'=(PAQ;PBQ;PCQ)$ および3層の行列 A, B, C の線形和による正則な変換である。同値性はテンソルの階数を保存するので絶対正則なテンソルの間の同値性を明らかにすることは階数決定に意義がある。同値性を証明することは一般に容易ではなく、本研究では「非同値性」を検出する手法を提案した。行列式多項式の定数曲面 $\{(x, y, z) \mid f(x, y, z)=1\}$ の微分幾何的不変量による非同値性の検証法を提案し、数値積分によりその数値が安定的に得られることを示した。合わせてその数値積分を球面デザインを使用して簡便に計算する計算法を提案した。微分幾何学的不変量とテンソルの非同値性の検証問題を関連させた本研究の成果は斬新で今後のテンソルデータ解析研究の新たな一面の端緒となりうる成果である。

④変換群を $GL_n \times GL_n \times O(3)$ に限定した場合と $GL_n \times GL_n \times U(3)$ に限定した場合に、直交群およびユニタリー群上の積分の数値積分の与える推定量による非同値性の検証法を提案し、有効性を示した。これは変換群を制約して考えた場合には数値積分を用いて簡便に非同値性が統計的検定問題の枠組みで議論できることを示した成果であり、ランダムな直交行列およびユニタリー行列の成分からなる多項式の積分公式理論とテンソル解析を結

びつけたテンソルデータ解析の一つの新たな研究の方向性を示したものである。

⑤絶対正則の存在は自明ではなく、証明すべきことである。これに対して、様々な絶対正則テンソルの構成法を提案した。具体的には帰納的にサイズの小さな絶対正則なテンソルからよりサイズの大きなテンソルを構成する複数の手法を提案した。また、絶対正則なテンソルには絶対正則なテンソルだけからなる開近傍の存在することを示した。

⑥絶対正則なテンソルの存在と典型ランクとの関係を示した。

⑦絶対正則性の概念を矩形型テンソルに拡張し、特徴付けを与えた。

⑧絶対正則テンソルの存在と典型テンソルの階数との関連性を示した。

(3) 積分幾何学にもとづく確率分布理論(チューブ法)を用いて、ランダム行列の最大固有値や探索的射影追跡指標の分布を導出した。非心ウィシャート分布のモーメントの組合せ構造を明らかにし、そのグラフによる表現を与えた。またラグール多項式の係数の組合せ論的な解釈を与えた。

(4) 森林面積比率に対する非線形回帰モデルを導出し、森林減少メカニズムを説明した。一方、ARXモデルについて、予測域の重要度に重みをつけたモデル選択基準を導出した。また、これに基づく特許出願の書類を準備中である。

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 25 件)

① Toshio Sumi and Toshio Sakata(2011), 2-neighborhood theorem for $3 \times 3 \times 3$ contingency tables, Journal of Indian Society of Probability and Statistics, 2011, to appear, (査読有り).

② Toshio Sakata, Kazumitsu Maehara, Toshio Sumi and Mitsuhiro Miyazaki(2010), An Enlarging Method of Obtaining Absolutely Nonsingular Tensors of Larger Size and Open Neighborhood Theorem, The proceedings of ICOQSIA2010, pp. 123-130, (査読あり).

③ Toshio Sumi and Toshio Sakata(2010), The Set of $3 \times 4 \times 4$ Contingency Tables has 3-Neighborhood Property, Proceedings of COMPSTAT' 2010, Electronic Supplementary Material (e-book), pp. 1629-1636, (査読有り).

④ Toshio Sumi, Mitsuhiro Miyazaki and Toshio Sakata(2010), About the maximal rank of 3-tensors over the real and the complex number field, Ann. Inst. Math.,

62, pp. 807-822, (査読有り).

⑤ Satoshi Kuriki, Yasuhide Numata(2010), Graph presentations for moments of noncentral Wishart distributions and their applications, Annals of the Institute of Statistical Mathematics, Vol. 62, No. 4, 645-672, (査読有).

⑥ Q. Pan, R. Nishii, T. Nakagawa and T. Nakamoto(2010), ARX models for time-varying systems estimated by recursive penalized weighted least squares method, Journal of Math-for-Industry, Vol. 2, 109-114, (査読有り).

⑦ Mitsuhiro Miyazaki, Toshio Sumi, and Toshio Sakata(2009), Tensor rank determination problem, International conference Non Linear Theory and its Applications 2009, Proceedings C.D. of NOLTA 2009, pp. 391-394, (査読有).

⑧ Toshio Sumi, Mitsuhiro Miyazaki and Toshio Sakata(2009), Rank of 3-tensors with 2 slices and Kronecker canonical forms, Linear Algebra and its Application, 431, pp. 1858-1868, (査読有).

⑨ S. Tanaka and R. Nishii. (2009), Non-linear regression models to identify functional forms of deforestation in East Asia. IEEE Transactions on Geoscience and Remote Sensing. Vol. 47(8), 2617--2626, (査読有).

⑩ Toshio Sakata and Toshio Sumi(2008), Lifting Between the Sets of Three-Way Contingency Tables and R-Neighbourhood Property, pp87-94, Proceedings CD of Compstat2008, (査読有).

⑪ Satoshi Kuriki, Akimichi Takemura(2008), Euler characteristic heuristic for approximating the distribution of the largest eigenvalue of an orthogonally invariant random matrix, Journal of Statistical Planning and Inference, Vol. 138, 3357-3378, (査読有).

[学会発表] (計 65 件)

① Toshio Sakata, A challenge of fundamental problems of tensor rank from multiple mathematical disciplines Mathematics of Date, Alahabad, India, 2011. 1. 3.

② Toshio Sakata, Stochastic tests of orthogonal equivalence of 3-tensors ASC2010, Perth, Australia, 2010. 12. 7.

③ Toshio Sakata, An Enlarging Method of Obtaining Absolutely Nonsingular Tensors of Larger Size and Open Neighborhood Theorem, ICOQSIA, Penang,

Malaysia, 2010, 11. 8.

④ Toshio Sakata, Implementation of Moment Formula of Unitary Matrix elements by statistical soft R and its application, COMPSTAT2010, Paris, France, 2010. 8. 24.

⑤ Toshio Sumi, The Set of $3 \times 4 \times 4$ Contingency Tables has 3-Neighborhood Property, COMPSTAT2010, Paris, France, 2010. 8. 23

⑥ Toshio Sakata, Exceptional tensors of $3 \times 4 \times 4$ tensors and Hilbert 17-th problem, Statistics, Probability, Operation Research, Computer Science and allied Areas, 2010. 1. 8.

⑦ Toshio Sakata, Cross-Entropy Minimization of functions on the Orthogonal Group and Its Applications (A comparison of Von-Mises Distribution and Cayley Distribution), International Workshop on recent advances in mathematical algorithms and data analysis, 2010. 01. 02.

⑧ Mitsuhiro Miyazaki, Tensor Rank Determination Problem NOLTA2009, Sapporo, Japan, 2009. 10. 29

⑨ Toshio Sakata, Exceptional tensors with three slices and the positivity of its determinant polynomial, South Africa ISI2009, Durban, South Africa, 2009. 8. 20.

⑩ Toshio Sakata, Lifting Between the Sets of Three-Way Contingency Tables and R-Neighbourhood Property, COMPSTAT2008, Port, Portugui, 2008. 8. 26.

⑪ Toshio Sumi, ABOUT THE MAXIMAL RANK OF 3-TENSORS OVER THE REAL AND THE COMPLEX NUMBER FIELD, Casta2008, 2008. 12. 7

⑫ Toshio Sakata, The Evaluation of the Maximal Rank of Tensors Simply by Row and Column Operations and Symmetrization, IASC2008, Yokohama, Japan, 2008. 12. 6

[図書] (計 1 件)

栗木哲, QTL 解析の統計モデルと検定の多重性調整, 「21 世紀の統計科学 II」(小西貞則, 国友直人編) 東京大学出版会 (2008), 315-356.

6. 研究組織

(1) 研究代表者

坂田 年男 (SAKATA TOSHIO)

九州大学・芸術工学研究院・教授

研究者番号 : 20117352

(2) 研究分担者

角 俊雄 (SUMI TOSHIO)

九州大学・芸術工学研究院・准教授

研究者番号 : 50258513

研究分担者

宮崎 充弘 (MIYAZAKI MITSUHIRO)

京都教育大学・教育学部・准教授

研究者番号 : 90219767

研究分担者

西井 龍映 (NISHII RYUEI)

九州大学・数理学府・教授

研究者番号 : 40127684

研究分担者

栗木 哲 (KURIKI SATOSHI)

統計数理研究所・大学共同利用機関などの部局・教授

研究者番号 : 90195545

(3) 研究連携者

