

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24日現在

機関番号：12501

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500251

研究課題名（和文） 因果推論のための新しい代数統計学の構築

研究課題名（英文） Construction of a New Algebraic Statistical Theory for Statistical Causal Inference

研究代表者

汪金芳（Jinfang Wang）

千葉大学・大学院理学研究科・教授

研究者番号：10270414

研究成果の概要（和文）：

条件付き独立性の概念は、統計モデルを特徴付ける本質的に重要な概念である。時系列モデルやベイジモデルなどを始め、すべての統計モデルがある種の条件付き独立性の仮定の下で構築されている。条件付き独立性は確率変数間の因果関係を特徴づける。本研究の主な目的は公理的な立場から条件付き独立性の性質を明らかにすることであるが、確率密度関数の初等的性質から条件付き独立性と密接に関連する汎代数構造を誘導できることを示した。この汎代数体系をケーインとよび、ケーインは graphoid や separoid の一般化である。また、いくつかの条件付き独立性から別の条件付き独立性を導くためのアルゴリズムについて研究も行った。

研究成果の概要（英文）：

Conditional independence is of fundamental importance in statistical sciences. All statistical models, such as time series models or Bayesian models, are based on certain assumptions on conditional independence. Statistical conditional independence also characterizes the causal relations of a set of random variables. The main purpose of this research is to study the conditional independence using an axiomatic approach. We constructed a new statistical algebraic system called cain for studying statistical causal inference. This system is based on the basic properties of probability density functions which are essential for arguing for conditional independence. We further established an algebraic algorithm for deriving a set of relations concerning some conditional independence from other set of such conditions.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,200,000	360,000	1,560,000
2011年度	1,000,000	300,000	1,300,000
2012年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：数理統計学、代数統計学

科研費の分科・細目：情報学・統計科学

キーワード：Algebraic statistics, conditional independence, statistical causal inference

1. 研究開始当初の背景

統計的因果推論を行うためには、条件付き独立性の概念が本質的に重要である。それまで

には、Pearl のグラフィカルモデルや、Rubin の反事実的モデルが存在するが、これらの枠組みの中では、条件付き独立性を形式的に計算することができない。Dawid (2001)は、条件付き独立性の公理化に関する研究を行っているが、条件付き独立性を形式的に計算することはできない。

2. 研究の目的

本研究の主な目的は、条件付き独立性の概念を代数的に定式化し、条件付き独立性に関する関係のある種の多項式と見なし、形式的に計算することにより、条件付き独立性（因果関係）を導く方法を開発する。

3. 研究の方法

Dawid (2001)と同様に、因果関係にあるものの集まりを 1 つの束 (lattice) と見なす。この束の各要素は古典的な確率変数と対応している。次に確率密度関数の代数的抽象化であるコインを定義し、条件付き独立性と関わる確率密度関数の性質の代数的抽象化を行った。

4. 研究成果

確率密度関数の初等的性質から条件付き独立性和密接に関連する汎代数構造を誘導できることを示した。この汎代数構造をケーインとよび、ケーインでは、全ての条件付き独立性に関する記述は方程式の形で記述されている。ケーインは特に確率推論のための公理体系である graphoid や separoid を満たすことを示した。また、いくつかの条件付き独立性から別の条件付き独立性を導くためのアルゴリズムについて研究も行った。この問題は次の 2 つのステップを通して解決した。まず、ケーイン多項式と呼ばれる新しいタイプの多項式を導入し、条件付き独立性に関する命題をケーイン多項式で表現を与える。次に、目的とされるケーイン多項式を、与えられた条件付き独立性の関係と同型なケーイン多項式の線形結合で表す。

以下では年度ごとに得られた主な成果を簡潔に述べる。

22 年度に得られた主な研究成果

- (1) 確率密度関数の初等的性質から条件付き独立性の研究に相応しい汎代数構造 (universal algebraic structure) を誘導できることを示した。この汎代数構造をケーイン

(cain) とよび、ケーインでは、全ての条件付き独立性に関する記述は方程式の形で記述されている。ケーインは特に確率推論のための公理体系である graphoid や separoid を満たすことを示した。得られた研究成果は Wang (2010) (Annals of the Institute of Statistical Mathematics, 62, 747-773) に纏められている。

- (2) いくつかの条件付き独立性から別の条件付き独立性を導くためのアルゴリズムについて研究を行った。この問題は次の 2 つのステップを通して解決した。まず、ケーイン多項式 (cain polynomial) と呼ばれる新しいタイプの多項式を導入し、条件付き独立性に関する命題をケーイン多項式で表現を与える。次に、目的とされるケーイン多項式を、与えられた条件付き独立性の関係と同型なケーイン多項式の線形結合で表す。得られた研究成果は Wang (2011) (Far East Journal of Theoretical Statistics, 34, 41-72) に纏められている。

- (3) そのまま開示できないマイクロデータを不確実区間に置き換え、個票開示できる方法を提案した。ある種の条件の下で不確実区間を計算するアルゴリズムについて研究し、更に Barnard (1965) や Kempthorne (1966) らによって提案された群尤度法 (grouped likelihood) 法に基づいて、不確実区間データを解析するための方法も提案した。

23 年度に得られた主な研究成果

- (1) Kolmogorov-Smirnov 検定を用いて、structome 解析に適用し、得られた成果を次のところに発表した。Masashi Yamaguchi, et al. (2011). Journal of Electron Microscopy, 60(5), 337-351.
- (2) 平均治療効果を推定するための新たな方法として、subclassification matching 法を提案し、既存と方法との比較を行った。得られた成果を次のところに発表した。Ping Jing, et al. (2011). Journal of the Japanese Society of Computational

Statistics. 24, 67-84.

- (3) 線形混合モデルを用いたときの新たなモデル選択法としてのブートストラップ法を提案し、得られた成果を次のところに発表した。Y. Ping and Wang, J. (2011). Proceedings of the 58th World Statistics Congress of the International Statistical Institute.
- (4) 複数の評価者による対応のあるクラスターデータが得られている場合、2群の割合の差の検定法を提案した。得られた成果を次のところに発表した。佐伯浩之，丹後俊郎，汪金芳 (2012)，京都大学数理解析研究所，短期共同研究「A New Perspective to Statistical Models and Related Topics」.

24年度に得られた主な研究成果

複雑なデータを解析する際に、データを生成するメカニズムを数学的に便宜なパラメトリックモデル仮定し、尤度推論を進めることがしばしば困難である。離散データを解析する際に過分散と呼ばれる現象がしばしば起き、このようなデータを解析するときに安易に既存のアルゴリズムを用いることは危険である。例えば、過分散が伴う比例データの場合、安易に一般化線型モデルを適用し、通常のlogistic 回帰分析を行うと誤ってしまう恐れがある。このようなデータに対して、確率分布を仮定せずに、平均と分散のみについての制約を仮定する擬似尤度の考え方がある。弱い条件の下で、擬似尤度推定量は一致性を持つことが広く知られている。しかし、擬似尤度に基づくモデル選択に関する研究は殆どなされていない。今年度はこの問題を焦点に研究を進めてきた。得られた研究成果の一部は論文として纏めている最中である。擬似尤度に基づくモデル選択は、従来のAICを導いた考え方を本質的に拡張する必要がある。平均と分散についての一般的な仮定の下では、擬似スコアを積分することはできないため、AICの導出を支える重要な概念としての予測分布の概念を拡張し、予測空間の構築が必要となる。この時に、最適な予測分布を導くことは、数学的にも統計的にもいくつかの課題を克服する必要があった。平成24年度では、これらの基本的な課題を整理し、必ずしも積分で

きるとは限らない擬似スコアに基づく「セミパラメトリック情報量規準」を導き、また実際のデータ解析やシミュレーションなどでその有効性を確認した。

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7件)

- ① 佐伯浩之，丹後俊郎，汪金芳 (2012). 複数の評価者による対応のあるクラスターデータの割合の差の検定，数理解析研究所講義録，1804，「A New Perspective to Statistical Models and Related Topics」，査読無，30-39.
- ② Yamaguchi, M., Namiki, Y., Okada, H., Mori, Y., Furukawa, H., Wang, J., Ohkusu, M. and Kawamoto, S. (2011). Structome of *Saccharomyces cerevisiae* determined by freeze-substitution and serial ultrathin sectioning electron microscopy, *Journal of Electron Microscopy*, 査読有，60(5), 337-351.
- ③ Jing, P., Zhang, L., Tang, Y. and Wang, J. (2011). Subclassification matching method for average treatment effect and a numerical comparison of related methods. *Journal of the Japanese Society of Computational Statistics*, 査読有，24, 67-84.
- ④ Ping, Y. and Wang, J. (2011). Bootstrap information criterion for linear mixed models. Proceedings of the 58th World Statistics Congress of the International Statistical Institute. 査読有. 6ページ(in CD-ROM).
- ⑤ Wang, J. (2010). A universal algebraic approach for conditional independence, *Annals of the Institute of Statistical Mathematics*, 査読有，62, 747-773.
- ⑥ Wang, J. (2010). Computation of conditional independence using cain polynomials, *Far East Journal of Theoretical Statistics*, 査読有，34, 41-72.
- ⑦ Wang, J. (2010). Rubin's Model for Causal Inference: a review, 数理解析

研究所講究録, 査読無, 138-150.

[学会発表] (計 7 件)

- ① 汪金芳 (2013). 英国王立統計学会と日本統計学会との共同認定による「国際資格」について, 下関市立大学, ワークショップ「効果的な統計及び数学教材の開発と利用」. 2013.1.12.
- ② Saeki, H., Tango, T. and Wang, J. (2012). A test for the difference in correlated proportions of clustered data based on multiple raters, IBC2012, Kobe International Conference Center, Kobe, Japan, August 26-31, 2012.
- ③ Wang, J. (2012). Statistical Disclosure Control Using the epsilon-uncertainty intervals and the grouped likelihood method, 2nd Institute of Mathematical Statistics Asia Pacific Rim Meeting (ims-APRM2012), Tsukuba International Congress Center, Tsukuba, Japan, July 1-4, 2012.
- ④ Wang, J. (2012). Computation of probabilistic conditional independence, The 2012 Taipei International Statistics Workshop, Taipei, Taiwan, 1-3 May 2012.
- ⑤ 佐伯浩之, 丹後俊郎, 汪金芳 (2012). 複数の評価者による対応のあるクラスターデータの割合の差の検定, 京都大学数理解析研究所, 短期共同研究「A New Perspective to Statistical Models and Related Topics」. 2012.3.5-7.
- ⑥ Ping, Y. and Wang, J. (2011). Bootstrap information criterion for linear mixed models, 58th World Statistics Congress of the International Statistical Institute, 2011. 8. 25. Dublin, Ireland.
- ⑦ Wang, J. (2010). Statistical disclosure control using the epsilon-uncertainty intervals and the grouped likelihood method, 19th International Conference on Computational Statistics, Paris, France. 8.22-27.

[図書] (計 2 件)

- ① 汪金芳 (2011). 統計応用の百科事典, 700 ページ, 114-115, 「ジャックナイフ

法」, 丸善.

- ② 汪金芳・桜井 裕仁 (2011). ブートストラップ入門, 236 ページ, 共立.

[その他]
ホームページ等

<http://www.math.s.chiba-u.ac.jp/~wang/>

6. 研究組織

(1) 研究代表者
汪 金芳 (Jinfang Wang)
千葉大学・大学院理学研究科・教授
研究者番号: 10270414

(2) 研究分担者
()

研究者番号:

(3) 連携研究者
()

研究者番号: