

## 科学研究費助成事業 研究成果報告書

平成 26 年 5 月 22 日現在

機関番号：14401

研究種目：基盤研究(B)

研究期間：2010～2013

課題番号：22300096

研究課題名(和文) 予測・因果・不完全データ解析とサイエンスの基礎

研究課題名(英文) Statistical prediction, causation, incomplete data analysis and foundation of science

研究代表者

狩野 裕 (KANO, YUTAKA)

大阪大学・基礎工学研究科・教授

研究者番号：20201436

交付決定額(研究期間全体)：(直接経費) 13,700,000円、(間接経費) 4,110,000円

研究成果の概要(和文)：不完全データの統計解析問題は理論と応用の両面において重要な課題であり、特に、ランダムでない欠測(NMAR)への対処方法が長年の課題となっている。加えて、欠測値問題は統計的因果推論を議論する際の重要なフレームワークを与えている。いくつかの顕著な研究成果は、2重中途打ち切りデータに対する新しい推測方法の開発、LiNGAMによる潜在交絡変数が存在する場合の因果推測法の提案、shared-parameterモデルを用いた新たな統計的推測方法の提案、制約のあるEMアルゴリズムの開発などである。

研究成果の概要(英文)：Analysis of incomplete data has been troublesome both theoretically and practically. In particular nonignorable missingness has been a serious issue in statistics. An alternative perspective of the theory of missing data analysis is to provide an insightful view of statistical causal inference. Some notable research outcomes include development of the analysis of doubly censored data, a new method of exploring causal structure for data with latent confounders via the LiNGAM approach, incomplete data analysis with a shared-parameter model and development of the EM algorithm with constraints.

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：shared-parameter モデル 統計的因果推論 無視できない欠測 NMAR LiNGAM 潜在交絡変数 2重中途打ち切り

### 1. 研究開始当初の背景

因果と予測は明確に区別されるべきである。適切な因果モデルの同定は重要な課題であり、また、それは予測精度の向上に寄与する。因果関係は理論的考察(演繹)と実証的研究(帰納)が相まって解明されていく。その中で、データに基づく統計学的な因果推論の方法は、諸科学における因果立証のための文法—方法論—を提供する。この意味できわめて重要な基礎的学問領域である。

### 2. 研究の目的

因果構造を発見(discovery)するための新しい方法として構造方程式モデリング、グラフィカルモデリング、ベイジアンネットワーク、ダイナミック因果モデル、LiNGAM等が提案されている。これらの手法の実質科学における各分野における有効性は早急に検討されなければならない。これらは方法論から実質科学への流れであるが、言うまでもなく、実質科学から方法論への流れも重要である。応用研究における因果分析がどのような問題に直面しているのか。また、具体的な応用研究には多くの制約が課されているが、方法論を提供する側からそれらの制約は見えないことが多い。これらの情報を共有することにより新たな方法を開発しそれらを用いることにより実質科学に貢献する。

### 3. 研究の方法

研究課題を遂行し学際融合研究を進めるため研究集会を開催する。研究集会は、(i)小規模の定期的研究会、(ii)中規模のシンポジウム、(iii)国際シンポジウムまたは国際会議での招待講演セッションの編成、の三種である。共同研究を推進するため、適宜、研究打合せを行う。

### 4. 研究成果

上述の研究の方法に基づき、小規模研究集会では延べ11回、計22件の講演を行った。(ii)と(iii)については、下記研究集会を開催し意見交換を行い共同研究を進めた。

- ◆ 2013.12.21 科学研究費シンポジウム「事象時間データ解析に関する理論と方法論およびその応用」於：弘前大学
- ◆ 2013.09.22-23 International Symposium at Osaka University 「Incomplete Data Analysis and Causal Inference」
- ◆ 2012.11.21-22 科学研究費シンポジウム「統計科学の基礎的理論とその応用」於：国際奈良学セミナーハウス
- ◆ 2012.07.03 APRM\_ism 2012 at Tsukuba. Invited-Paper Session 「Incomplete Data Analysis and Latent Variable Model」
- ◆ 2012.03.04 日本統計学会春季集会招待講演セッション「リスク、信頼性、そして確率統計：震災一年、統計学の役割を問う」於：一橋大学

- ◆ 2011.11.03-05 科学研究費シンポジウム「生命科学と統計学」於：大阪大学
- ◆ 2011.07.09 IMPS2011 at Hong Kong. Invited Symposium 「Analysis of Missing Data and Causal Inference」

主な研究成果を具体的に述べる。

(1) NMARの欠測に対処する潜在変数モデル：NMARの欠測メカニズムを記述するShared-parameterモデルが、潜在変数モデルと相性が良いことを示した。同モデルの下で、欠測メカニズムを明示的に用いない推測方法を提示し、その良さを数値実験により検証した。得られた結果はJournal of Multivariate Analysisに掲載された。

(2) 多重エンドポイントに基づく治療効果の評価方法：多重エンドポイントに基づき治療効果を評価する臨床試験において、群逐次デザインの枠組みのなかで、試験の有効中止に関する二つの意思決定の仕方を検討した。提案する意思決定の仕方では、どのように検出力と第1種の過誤確率が挙動するかを検討した。さらに、試験途中での条件付き検出力に基づく標本サイズの変更の方法を提案した。これらの研究成果は、Statistics in Medicineに掲載された。

(3) 2重中途打ち切りデータのノンパラメトリック推定量：右側中途打ち切りを超える不完全情報を伴う事象時間データの解析のために、生存分布の推測や回帰モデルの拡張法が提案されてきている。右側中途打ち切りデータでは、マルチンゲール接近法は推定量の漸近的性質や回帰モデルの診断法に広く利用されているが、これを超える不完全データのもとで、そのような接近法の開発はあまり進んでいない。本研究では、2重中途打ち切りデータに焦点をあて、ノンパラメトリック推定量(NPMLE)の推測のためのマルチンゲール接近法を議論している。セミパラメトリック・プロフィール尤度のスコア関数を用いて、NPMLEのマルチンゲール構造を定式化した。最終的に、NPMLEの漸近分布が無次元行列や作用素表現のない扱いやすい陽形式で導出された。大標本においても計算可能な形式の分散共分散公式が提示された。これらの研究成果はAnnals of the Institute of Statistical Mathematicsに掲載された。

(4) 潜在交絡変数がある場合の因果構造推定：観察データに基づいて因果構造を推定する場合の難題は、潜在交絡変数にどう対処するかである。本研究では、潜在交絡変数に頑健な因果構造推定法を提案した。提案手法の妥当性を、従来法よりも非常にゆるい条件下で数学的に証明した。また、脳活動測定データに基づく脳内ネットワーク推定への応用について考察した。研究成果

は国際学術雑誌 *Neural Computation* に掲載された。

(5) 2 標本一般化片側仮説の検定問題：2 標本正規モデルにおいて，母平均間の関係を示す仮説が，ある微分不可能な関数による不等式で表されている一般化片側仮説を考察した．研究成果としては，尤度比検定統計量の漸近分布が導かれ，関数の凸凹に応じて，intersection-union 検定と union-intersection 検定に帰着することが示された．研究成果は *Communications in Statistics: Theory and Methods* に掲載が決定している。

(6) 制約のある EM アルゴリズムの開発：EM アルゴリズムでは，パラメタに制約を課すことによってアルゴリズムの好ましい特徴である単調増加性や計算の単純さが失われてしまうことが多い．本研究では，ラグランジュ法と射影法を組み合わせることによって，制約があっても単調増加性と単純さが維持できることを示した．研究成果は，*Computational Statistics* に掲載された。

#### 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 30 件)

- ① Takagi, Y. (in press). Asymptotics of likelihood ratio tests for general one-sided hypotheses in the two-sample normal model. *Communications in Statistics - Theory and Methods*. (掲載決定) 査読有
- ② Katayama, S. and Kano, Y. (in press). A new test on high-dimensional mean vector without any assumption on population covariance matrix. *Communications in Statistics - Theory and Methods*, (accepted on July 23, 2012). 査読有
- ③ Hojo, S., Yamamoto, M. and Kano, Y. (in press). Effect of violation of the normal assumption on MI and ML estimators in the analysis of incomplete data. *Communications in Statistics - Theory and Methods*. (accepted on June 15, 2013). 査読有
- ④ Tashiro, T., Shimizu, S., Hyvarinen, A. and Washio, T. (2014). ParceLiNGAM: A causal ordering method robust against latent confounders. *Neural Computation*, 26(1): 57-83. 査読有 [DOI:10.1162/NECO\_a\_00533]
- ⑤ Asakura, K., Hamasaki, T., Sugimoto, T., Hayashi, K., Evans, S., Sozu T. (2014). Sample size determination in group-sequential clinical trials with two co-primary endpoints. *Statistics in Medicine*. (First published online: 27 March 2014). 査読有 [DOI: 10.1002/sim.6154]
- ⑥ Katayama, S., Kano, Y., and Srivastava, M. S. (2013). Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, Vol.116(Apr), 410-421. 査読有 [DOI: 10.1016/j.jmva.2013.01.008]
- ⑦ Mori, J., Kano, Y., Yoshizaki, M. and Fukinbara, S. (2013). Latent class models for medical diagnostic tests in multicenter trials. *Statistics in Medicine*, Vol.32(29), 5091-5105. 査読有 [DOI: 10.1002/sim.5962]
- ⑧ Srivastava, M. S., Katayama, S. and Kano, Y. (2013). A two sample test in high dimensional data. *Journal of Multivariate Analysis*, Vol.114, 349-358. 査読有 [DOI: 10.1016/j.jmva.2012.08.014]
- ⑨ Sugimoto, T. (2013). Asymptotic distribution of the nonparametric distribution estimator based on a martingale approach in doubly censored data. *Annals of the Institute of Statistical Mathematics*, Vol.65

(5), pp. 859-888. 査読有

[DOI 10.1007/s10463-012-0395-4]

- ⑩ Li, Y.-X., Kano, Y., Panc, J.-H. and Song, X.-Y. (2012). A criterion-based model comparison statistic for structural equation models with heterogeneous data. *Journal of Multivariate Analysis*, Vol.112(Nov), 92-107. 査読有  
[DOI:10.1016/j.jmva.2012.05.010]
- ⑪ Takai, K. (2012). Constrained EM algorithm with projection method. *Computational Statistics*, 27, 701-714. 査読有  
[DOI:10.1007/s00180-011-0285-x]
- ⑫ Kano, Y. and Takai, K. (2011). Analysis of NMAR missing data without specifying missing-data mechanisms in a linear latent variate model. *Journal of Multivariate Analysis*, Vol.102(Oct), 1241-1255. 査読有  
[DOI:10.1016/j.jmva.2011.04.007]

[学会発表] (計 16 件)

- ① Kano, Y. (2013/09/22). Approximate population bias and nonignorable missingness. *An International Symposium on Incomplete Data Analysis and Causal Inference*. Osaka, Japan.
- ② Mori, J. and Kano, Y. (2013/08/26). Latent class models for medical diagnostic tests in multicenter trials. *ISCB Conference 2013*. Munich, Germany.
- ③ Kano, Y. (2012/07/03). NMARness and approximate population bias of the direct MLE in the analysis of missing data. *IMS-APRM2012*. Tsukuba, Japan.
- ④ Kamada, A. and Kano, Y. (2012/07/04).

Statistical inference in structural equation modeling with a near singular covariance matrix. *IMS-APRM2012*. Tsukuba, Japan.

- ⑤ Takai, K. and Hayashi, K. (2011/12/18). Effect of unlabeled data and labeling strategy on error rate in discriminant analysis. *The Joint Meeting of 7th Conference of the Asian Regional Section of the International Association for Statistical Computing and 2011 Taipei International Statistical Symposium*. Taipei, Taiwan.

[その他]

ホームページ等

<http://www.sigmath.es.osaka-u.ac.jp/~kano/research/seminar/index.html>

## 6. 研究組織

### (1) 研究代表者

狩野 裕 (KANO, Yutaka)  
大阪大学・基礎工学研究科・教授  
研究者番号：20201436

### (2) 研究分担者

出口 康夫 (DEGUCHI, Yasuo)  
京都大学・文学研究科・准教授  
研究者番号：20314073

鷺尾 隆 (WASHIO, Takashi)  
大阪大学・産業科学研究所・教授  
研究者番号：00192815

濱崎 俊光 (HAMAZAKI, Toshimitsu)  
大阪大学・医学系研究科・准教授  
研究者番号：40379243

高木 祥司 (TAKAGI, Yoshiji)  
奈良教育大学・教育学部・准教授  
研究者番号：00231390

杉本 知之 (SUGIMOTO, Tomoyuki)  
弘前大学・理工学研究科・准教授  
研究者番号：70324892

(平成20年度-平成24年度までは  
連携研究者, 平成25年度から分担者)

高井 啓二 (TAKAI, Keiji)  
関西大学・商学部・准教授  
研究者番号：20572019  
(平成23年度から分担者,  
平成24年度-平成25年度は連携研究者)

(3) 連携研究者

内藤 貫太 (NAITO, Kanta)  
島根大学・総合理工学部・教授  
研究者番号：80304252

清水 昌平 (Shimizu, Shohei)  
大阪大学・産業科学研究所・准教授  
研究者番号：10509871

(4) 研究協力者

片山 翔太 (KATAYAMA, Shota)  
日本学術振興会特別研究員 DC1

山本 倫生 (YAMAMOTO, Michio)  
日本学術振興会特別研究員 DC2

Xinyuan Song  
The Chinese University of Hong Kong, HK,  
Professor

Mortaza Jamshidian  
California State University, Fullerton,  
USA  
Professor

Aapo Hyvarinen  
University of Helsinki, Finland  
Professor

Ke-Hai Yuan  
University of Notre Dame, USA  
Professor