

平成 22 年 4 月 1 日現在

研究種目：基盤研究（B）

研究期間：2006～2009

課題番号：18300094

研究課題名（和文） 統計的因果推論の総合的研究

研究課題名（英文） Comprehensive study on statistical causal inference

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

研究成果の概要（和文）：少人数の定例研究会の開催，中規模研究集会の計画と開催，そして国際研究集会の編成と開催を通じて，統計科学，情報科学，社会科学（特に科学哲学を含む）において情報交換を積極的に行った．具体的な研究成果として，共分散選択のロバスト推定法の開発，非正規性の利用による因果の方向の同定， $2 \times 2$  分割表と潜在変数モデルにおける無視不可能な欠測に対する新たな分析方法の開発，条件付き確率と科学的証拠の関係の明確化などが得られた．

研究成果の概要（英文）：Active academic exchanges among researchers in statistical, informatics and social sciences, including philosophy of science in particular, have been made through hosting colloquiums of a few speakers regularly, symposiums of 10 to 15 talks several times a year, and international meetings. Some of our fruits obtained with the help of the governmental scientific research grant are i) a new robust methodology of the covariance selection, ii) a new method for determining direction of causation using nonnormality, ii) new estimation methods for data with nonignorable missing values in a  $2 \times 2$  contingency table and a latent-variable model, and iii) clarification between conditional probability and scientific evidence.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2006年度	3,500,000	0	3,500,000
2007年度	4,000,000	1,200,000	5,200,000
2008年度	3,200,000	960,000	4,160,000
2009年度	4,200,000	1,260,000	5,460,000
総計	14,900,000	3,420,000	18,320,000

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：無視可能な欠測・欠損，潜在変数モデル，条件付き確率の理解と認識，

仮説と観察の歴史的経緯，傾向スコア，共分散選択，非正規性の利用，因果の方向

## 1. 研究開始当初の背景

20世紀の100年間で、学問は細かい専門分野へと分化し、対象を絞った多くの学会が誕生、そして、その中で専門性を深めた。いま伝統的学問の多くが行き詰まり、魅力的なブレークスルーを求めらる中で、この細分化に反省が見られる。残念ながら、統計科学、情報科学、社会科学も例外ではなく、各分野の研究者が互いに交流する場は多くはなく、互いの問題共有もほとんど無い状況であった。

たとえば、因果関係の哲学的な定義、一回しか起こり得ない事象の確率の定義、原因の濾過、尤度と確率の違い、ベイズ主義の効用等、統計科学的にも重要な問題が哲学で議論されているが、多くの統計学者は哲学での議論を認識していない。一方、統計科学においては、推定量の可測性、可積性などの数学的基礎概念、また、漸近有効性、漸近十分性などの推測の最適性等はしばしば議論されるわけであるが、それらの哲学的意義は問われない。

## 2. 研究の目的

科学の個別分野で研究されている因果推論の諸問題を共有し学際的に研究することによりそれらの解決を図る。境界領域での研究を活性化させる。学際研究を遂行するため国際研究集会を開催する。本研究では、統計科学、情報科学、社会科学を主な研究対象分野とする。

## 3. 研究の方法

社会科学と統計・情報科学との共同研究を進めるため各種の研究集会を開催する。具体的には以下の3種である。

(1) 少人数の講演者による定例研究会。一件当たりの研究発表と研究討論は約2時間。年に10回程度で水曜または金曜に開催する。

話題提供者は研究分担者・連携研究者を中心に依頼するが、分担者・連携研究者以外の研究者を招聘することもありえる。

(2) 中規模の研究集会。数日かけて10~15件程度の研究発表を編成する。年に2件程度開催する。講演者は国内外を問わず広く公募する。

(3) 国際研究集会の編成と実行。科学研究費研究集会として単独で開催する場合と既存の国際大会のセッションとして開催する場合がある。

## 4. 研究成果

### (1) 少人数の講演者による定例研究会

4年間で延べ23回の定例研究会を開催し、外国人研究者による6件の講演を含む44件の講演を実施した。詳細は以下のURLを参照されたい。

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

### (2) 中規模の研究集会

研究集会の概要を年度ごとにまとめる。  
平成18年度実施実績

① 9月27,28日に東京外国語大学において「多変量解析の理論と応用における最近の話題：因果推論、構造方程式モデリング、因子分析を巡って」の研究集会を開催した。9件の講演と40名の出席者を得、活発な議論を展開した。理工系研究者が中心の集会であった。

② 12月16日に京都大学文学研究科において「統計学の哲学と推論」の研究集会を開催した。4件の講演と35名の出席を得た。オーガナイザは研究分担者である出口康夫准教授（京都大学）と研究代表者の狩野である。文学系の研究者が中心の集会であったが、理工系の研究者からも活発な意見が出され、特に、最後の総合

論議は進行役が取りまとめに苦慮する一面もあった。

#### 平成 19 年度実施実績

- ① 9 月 22 日～24 日に九州大学において数理統計学と数理言語学に関する研究集会・研究打合せを行った。
- ② 12 月 20 日に東京大学において研究集会「哲学 - 統計学 - 心理学のクロストーク：多分野交流による方法論の創出」を開催した。本集会において、統計的因果推論について哲学、心理学そして統計学の立場から議論を行い、それぞれの問題意識や概念の違いを明確化した。

#### 平成 20 年度実施実績

- ① 9月29日～31日に大阪大学にて多変量解析に関する研究集会「多変量解析における最近の話題」を開催した。講演件数は16件であった。研究代表者は大学院生と共著で無視不可能な欠測を含む離散データにもとづく因果推論について講演した。35名の参加者を得た。

#### 平成 21 年度実施実績

- ① 4月25日～26日に京都にて開催された応用哲学会年次研究大会において、「統計学の哲学の逆襲」というタイトルのワークショップを開催し、因果推論における哲学と統計学について議論した。主担当は連携研究者の出口康夫准教授であった。
- ② 10月21日～23日に島根大学にてノンパラメトリック法と機械学習による因果推論に関するシンポジウム「統計科学の方法論と応用の新展開」を開催した。主担当は研究分担者の内藤貫太准教授であった。40名の参加者を得た。
- ③ 12月4日に高根芳雄教授 (McGill University) をキーノートスピーカーとして招聘した国際シンポジウム「多変量

データ解析法への計算機統計学・行動計量学的アプローチの新展開」を開催した。合計7件の研究報告があり、活発な意見交換がなされた。主担当は研究分担者の足立浩平教授であった。

#### (3) 国際研究集会の編成と実施

- ① 平成18年12月23日に大阪大学教育実践センターにおいて「Propensity score: Theory and applications」の国際研究集会を開催した。アリゾナ大学の Keisuke Hirano 教授をお招きして基調講演をお願いした。講演総数は4件とトピックを絞った小さな集会であったにもかかわらず45名の出席者を得た。大学の研究者だけでなく生物統計実務家も多く出席した。
- ② 平成19年7月に東京で開催された計量心理学会国際大会 (IMPS2007) において、国内外の著名な研究者を多数招聘し、招待講演セッションを2件開催、真摯な議論を交わした。具体的なセッション名は「Statistical causal inference in observational studies」と「Recent developments in multivariate analysis」であった。
- ③ 平成20年6月にトルコ共和国のカイセリで開催された国際研究集会HDM2008において招待講演セッション「High Dimensional Multivariate Models and Related Topics」を編成し、高次元データにもとづく統計的因果推論について議論した。なお、研究代表者はICAにもとづく因果推論について最新の研究結果を発表した。
- ④ 平成20年12月に横浜で開催された国際学会 IASC2008 において、4名の国際的に著名な研究者を招聘し、統計的因果推論に関連する基調講演をしていただいた。研究代表者は構造方程式モデリングにおけ

る結果指標と原因指標に関する最新の研究結果を発表した。

(4) 個別研究成果

- ① データセットへの外れ値の混入は多くの統計的手法を無意味化する。連続変数間の因果関係を同定する方法に共分散選択がある。Miyamura and Kano (2006)では共分散選択の推定問題においてJダイバージェンスのアイデアを用い、外れ値に対して頑健な方法論を提案し、その方法の有用性を数値実験と数値例を用いて実証した。
- ② 統計的因果推論の最も基本的な姿は二つの確率変数でどちらが原因でどちらが結果かを同定することであろう。しかし、一般に2変数だけから因果の同定を行うことは不可能であるとされてきた。Shimizu and Kano (2008)では高次モーメントの情報を活用することで因果同定が可能となることを示した。数値実験によって有効性を確認した。
- ③  $2 \times 2$  分割表データの解析は因果関係同定の最も基本的な状況である。このような単純な場合でもデータに欠測値が含まれると推測に困難が生じる。Takai and Kano (2008)では、欠測メカニズムをグラフィカルモデルによってモデル化し、無視不可能な欠測を含むデータセットに対する新しい分析方法を提示した。
- ④ 統計的因果推論にバイアスを生じさせる要因として、未観測共変量・サンプルセレクション・測定誤差がある。潜在変数モデルは測定誤差を分離できないという意味で統計的因果推論に貢献する。Kano and Takai (2009)では、まず、欠測メカニズムが潜在変数に依存する場合、欠測は無視できないことを確認し、次に、

そのような場合における新たな統計的推測方法を提案した。

- ⑤ 条件付き確率は時間を遡及した推測に用いられることがある。原因変数が確定しその後に結果変数が確定するのが通常であるから、結果変数を与えた下での原因変数の条件付き確率は何を表すのか。この基本的な問題について頻度論からの説明を試みた。

また、ある変数がある仮説に対する科学的証拠であるためには歴史的経緯が重要であるという主張があり、一方、歴史的経緯に無関係であるべきであるという考えもある。この矛盾した主張に対して、証拠変数が仮説の原因変数となるか結果変数であるかの区別が重要であることを指摘した。

5. 主な発表論文等

〔雑誌論文〕(計26件)

- ① 狩野 裕 (2009). 不完全データの解析. 生産と技術. Vol.61, No. 1, 71-76, 生産技術振興協会.(査読無)
- ② Takai, K. and Kano, Y. (2009). Simple computation of maximum likelihood estimates in latent class model with equality and constant constraints. Communications in Statistics - Simulation and Computation. Vol. 38, Issue 3, 654 - 665. (査読有)
- ③ Takai, K. and Kano, Y. (2008). Test of independence in a  $2 \times 2$  contingency table with nonignorable nonresponse via constrained EM algorithm. Computational Statistics and Data Analysis. Vol.52, No.11, 5229-5241. (査読有)
- ④ Shimizu, S. and Kano, Y. (2008). Use

of non-normality in structural equation modeling: Application to direction of causation. Journal of Statistical Planning and Inference. Vol.138, No.11, 3483-3491. (査読有)

- ⑤ Miyamura, M. and Kano, Y. (2006). Robust Gaussian graphical modeling. Journal of Multivariate Analysis, Vol.97, No.7, 1525-1550. (査読有)
- ⑥ Shimizu, S., Hyvärinen, A., Hoyer, P. O. and Kano, Y. (2006). Finding a causal ordering via independent component analysis. Computational Statistics and Data Analysis. Vol.50, No.11, 3278-3293. (査読有)

[学会発表](計13件)

- ① Kano, Y. and Takai, K. (2009/July, Cambridge). Should complete-case analysis always be avoided? -- SEM for incomplete data with nonignorable missing. IMPS2009, Abstracts, p.77.
- ② Kano, Y. (2008/Dec, Yokohama). Scaling with SEM: The Role of Effect and Causal Indicators. IASC2008. [invited talk]
- ③ Kano, Y. (2008/Jun, Kayseri). Separability of noisy ICA for high dimensional data. HDM2008. [invited talk]
- ④ Kano, Y. and Murayama, K. (2007/July, Tokyo). On the identifiability issue of noisy independent component analysis and factor score indeterminacy. IMPS2007. [Invited talk]

[図書](計8件)

Kano, Y. (2007). Selection of manifest variables. In Handbook of Latent

Variable and Related Models, pp.65-86. Springer.

Yanai, H. and Ichikawa, M. (2007). Factor Analysis. In C. R. Rao and S. Sinharay (Eds.), Handbook of Statistics, pp. 257-296. Amsterdam: Elsevier.

## 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

柳本 武美 (YANAGIMOTO TAKEMI)

中央大学・理工学部・教授

研究者番号: 40000195

(H19 H20: 連携研究者)

山本 英二 (YAMAMOTO EIJI)

岡山理科大学・総合情報学部・教授

研究者番号: 50068920

(H19 H20: 連携研究者)

佐藤 俊哉 (SATO TOSHIYA)

京都大学・医学(系)研究科・教授

研究者番号: 50187217

(H19 H20: 連携研究者)

熊谷 悦生 (KUMAGAI ETSUO)

大阪大学・基礎工学研究科・講師

研究者番号: 20273617

(H19 H20: 連携研究者)

山口 和範 (YAMAGUCHI KAZUNORI)

立教大学・経営学部・教授

研究者番号: 60230348

(H19 H20: 連携研究者)

渡辺 美智子 (WATANABE MICHIKO)

東洋大学・経済学部・教授

研究者番号: 50150397

(H19 H20: 連携研究者)

宮川 雅巳(MIYAKAWA MASAMI)  
東京工業大学・社会理工学研究科・教授  
研究者番号：90157595  
(H19 H20: 連携研究者)  
黒木 学(KUROKI MANABU)  
大阪大学・基礎工学研究科・准教授  
研究者番号：60334512  
(H19 H20: 連携研究者)  
繁榊 算男(SHIGEMASU KAZUO)  
帝京大学・文学部・教授  
研究者番号：90091701  
(H19 H20: 連携研究者)  
植野 真臣(UENO MAOMI)  
電気通信大学・情報システム学研究科・准教授  
研究者番号：50262316  
(H19 H20: 連携研究者)  
本村 陽一(MOTOMURA YOICHI)  
産業技術総合研究所・デジタルヒューマン  
研究センター・主任研究員  
研究者番号：30358171  
(H19 H20: 連携研究者)  
戸田山 和久(TODAYAMA KAZUHISA)  
名古屋大学・情報科学研究科・教授  
研究者番号：90217513  
(H19 H20: 連携研究者)  
一ノ瀬 正樹(ICHINOSE MASAKI)  
東京大学・人文社会系研究科・教授  
研究者番号：20232407  
(H19 H20: 連携研究者)  
出口 康夫(DEGUCHI YASUO)  
京都大学・文学研究科・准教授  
研究者番号：20314073  
(H19 H20: 連携研究者)  
足立 浩平(ADACHI KOHEI)  
大阪大学・人間科学研究科・教授  
研究者番号：60299055  
(H19 H20: 連携研究者)

(H20 H21: 研究分担者)  
唐沢 かおり(KARASAWA KAORI)  
東京大学・人文社会系研究科・准教授  
研究者番号：50249348  
(H19 H20: 連携研究者)  
南風原 朝和(HAEBARA TOMOKAZU)  
東京大学・教育学研究科・教授  
研究者番号：50156246  
(H19 H20: 連携研究者)  
乾 敏郎(INUI TOSHIO)  
京都大学・情報学研究科・教授  
研究者番号：30107015  
(H19 H20: 連携研究者)  
盛山 和夫(SEIYAMA KAZUO)  
東京大学・人文社会系研究科・教授  
研究者番号：50113577  
(H19 H20: 連携研究者)  
清水 泰隆(SHIMIZU YASUTAKA)  
大阪大学・基礎工学研究科・助教  
研究者番号：70423085  
(H19 H20: 連携研究者)  
宮本 友介(MIYAMOTO YUSUKE)  
大阪大学・人間科学研究科・助教  
研究者番号：50379228  
(H19 H20: 連携研究者)  
市川 雅教(ICHIKAWA MASANORI)  
東京外国語大学・外国語学部・教授  
研究者番号：20168313  
(H19 H20: 連携研究者)  
柳原 宏和(YANAGIHARA HIROKAZU)  
広島大学・理学研究科・准教授  
研究者番号：70342615  
(H19; H19 H20: 連携研究者)  
内藤 貫太(NAITO KANTA)  
島根大学・総合理工学部・准教授  
研究者番号：80304252  
(H21)  
(3)連携研究者