

平成 30 年 6 月 4 日現在

機関番号：14401
研究種目：挑戦的萌芽研究
研究期間：2016～2017
課題番号：16K12402
研究課題名(和文) How serious is nonignorable missingness?

研究課題名(英文) How serious is nonignorable missingness?

研究代表者

狩野 裕 (Kano, Yutaka)

大阪大学・大学院基礎工学研究科・教授

研究者番号：20201436

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：無視不可能な欠測(NMAR)が存在するとき、欠測メカニズムを組み入れない尤度(観測尤度)に基づく最尤推定量(MLE)にはバイアスが生じる。本研究では、そのバイアスを理論的に評価する方法論を新規に開発した。その結果に基づき、適切な補助変数を導入することがMLEのバイアスを減少させること、および不適切な補助変数の導入はバイアスを増加させることを証明した。
欠測メカニズムと観測変数間の因果関係とが独立にバイアス評価に関与する状況の数学的本質を解明した。
欠測メカニズムの特定においてセミパラメトリック法を採用し、欠測メカニズムに要求される強い仮定を緩和した下での統計的推測方法を発展させた。

研究成果の概要(英文)：Under NMAR missingness, the observed likelihood, without a missing-data mechanism, leads to a biased MLE. In this research, we developed a new methodology to express the bias of the MLE due to the missingness in closed form. Using the formula, we provided several mathematical conditions under which inclusion of auxiliary variables reduces or inflates the bias.

The formula described above holds for any missing-data mechanism. This strong consequence can be proved because a shared-parameter model is taken for missingness.

A final contribution of this research to be reported is to take a semi-parametric way to relax the strong condition required for the conventional missing-data analysis.

研究分野：統計科学

キーワード：無視不可能な欠測 shared-parameter model セミパラメトリック法 NMAR バイアス 補助変数

1. 研究開始当初の背景

欠測値を含むデータを適切に統計分析することは意外に難しい。最も典型的なりストライズ削除を行い完全データ化し、得られた完全データを、元から完全データであったかのように分析したならば、無視できないバイアスが生じることが分かっている。そこで、バイアスが発生しないための条件(MAR, MCAR)や推測方法(欠測メカニズム)が提案されてきた。しかし、これらの条件が満たされないときや別の推測方法を採用したとき、どの程度深刻なバイアスが生じるのか、バイアスを定量的に評価する方法論が無かった。つまり、バイアスが落ちる(0となる)ための条件の研究は進んでいたが、バイアスの大きさの評価については全く手が付けられていなかった。すなわち、「How serious is nonignorable missingness」についての研究が存在していなかった。

2. 研究の目的

本研究では、欠測が生じることによって起こり得る推定量のバイアスを理論的に評価する方法論を新規に展開し、欠測値データ解析の理論体系を再構築する。様々な新規の統計的推測問題へ応用することを目的とした。

欠測メカニズムがMARであるとき観測データに基づくMLE (based on observed likelihood)は漸近的に不偏な推定量となる。しかし、MARの仮定が崩れたとき、どのようなバイアスが生じるのだろうか。バイアスを明示的に表現できればMARの条件を緩和できる可能性もある。また、バイアス削減の有力な方法とされる補助変数の追加が、いつも推定量のバイアスを減じるのであろうか。そうでないならば、バイアスを減じるための数学的な条件の導出が必要になる。

3. 研究の方法

統計的因果推論は欠測値データ解析に帰着する。統計的因果推論における最近の成果に二重頑健推定法(Robins等)がある。狩野(2015)は二重頑健推定量のバイアスを明示的に表現することにより二重頑健性の別証明を与えた。また、そのバイアスから様々な考察を導いた。具体的には、

	真	モデル
傾向スコア	$E(R X)$	$e(X)$
回帰	$E(Y X)$	$m(X)$

とし、 $E(Y)$ を推定したいとする。このとき、強い意味の無視可能性条件の下で、二重頑健推定量のバイアスは

$$\hat{\mu}_y^{(DR)} - \mu_y = E_X \left[\frac{\{E(R|X) - e(X)\} \{E(Y|X) - m(X)\}}{e(X)} \right] + o_p(1)$$

と表現される。この結果から容易に、傾向スコアが正しく特定化されるか、もしくは、回帰モデルが正しく特定されれば、不偏推定量

が構成できることが分かる。また、その条件は十分条件に過ぎないことも直ぐに分かる。

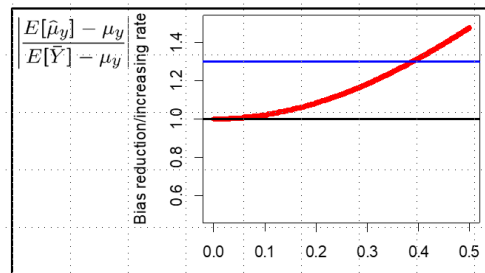
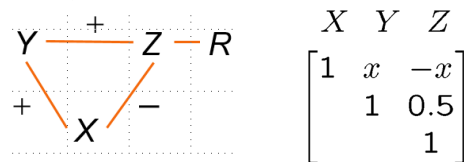
このように推定量のバイアスを明示的に表現することには意味があると思われる。狩野のアプローチを発展させることにより、一般のMLEにおいて欠測が誘導するバイアスを明示的に表現し、その応用を展開する。

研究代表者の狩野は本課題に関係する基礎的な研究成果を国際会議IMPS2015@Beijingにおいてkeynote speakerとして講演している。

このような研究を進めるため、小規模および中規模の研究集会を各年に数回開催した。主な講演者は、Ke-Hai Yuan氏(University of Notre Dame)、服部 聡氏(大阪大学 大学院医学系研究科)、平野敏弘氏(日本電気株式会社 中央研究所)、Vladimir Ulyanov氏(Moscow State University)らであった。

4. 研究成果

- (1) 無視不可能な欠測(NMAR)の下での欠測メカニズムを組み入れない尤度(観測尤度)に基づく最尤推定量(MLE)にはバイアスが生じる。本研究では、このバイアスを理論的に評価する方法論を新規に開発し、適切な補助変数を導入することがMLEのバイアスを減少させること、および不適切な補助変数の導入はバイアスを増加させることを証明した。補助変数の追加がバイアスを増加させるための条件を明確化したことは重要な結果である。以下のグラフにその状況を示す。



上のグラフは、補助変数を追加した推定量 $\hat{\mu}_y$ のバイアスの、補助変数を追加しなお推定量のバイアスに対する比を表しており、この状況ではその比が常に1以上である、すなわち、補助変数の追加がバイアスを悪化させることが分かる。また、

$$x = \rho_{yx} = -\rho_{zx}$$

が大きくなるにつれてその非が大きくなることが示されている。この結果を用いて、サロゲートエンドポイントを用いたモデルにおいて同様の結果を得てい

る。論文として公開。

- (2) (1)の状況で、欠測メカニズムと観測変数間の因果関係とが独立にバイアス評価に關与するという事実は極めて強いものである。その数学的本質 shared-parameter model の構造にあることを証明した。論文として公開。
- (3) 経時データで脱落がある状況で、最尤推定法が適用可能であるための識別性条件を導いた。それを応用し、識別可能な意味のあるモデルを2件、構築した。具体的には、以下の図に示すような、割り付け外生変数 X を導入したモデルと2次の自己回帰モデルである。論文として公開。

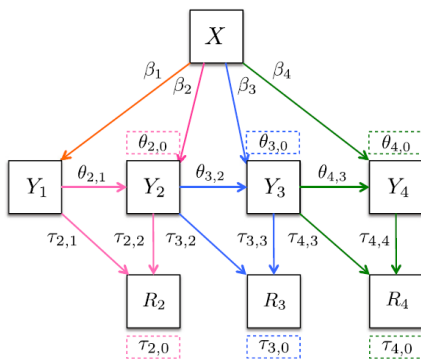


Fig. 3. AR(1) model with one covariate.

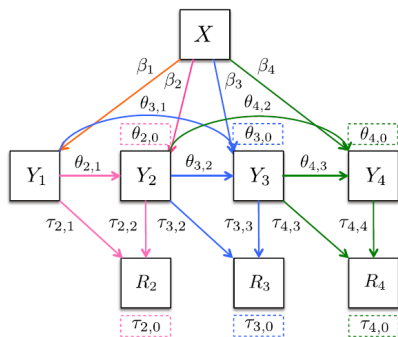


Fig. 4. AR(2) model with one covariate.

- (4) 欠測値問題における欠測メカニズムの特定においてセミパラメトリック法を採用し、欠測メカニズムに要求される強い仮定を緩和した下での推測方法を開発し、その漸近的性質を導出した。論文として公開。
- (5) 論文は、電気情報通信学会の100周年記念特集号において欠測値問題に関する論文の寄稿を求められたものである。欠測に関する研究の最前線とバイアス評価の重要性について解説した。
- (6) 実験参加者が割り当てられた処置どおりに行動しないことをノンコンプライアンスといい、ノンライアンスが生じると処置効果の推定に偏りをもたらす可能性が高い。ノンコンプライアンスは不完全データの種類である。本研究では、

見かけ上のノンコンプライアナーが存在する場合での処置効果の推定に関する感度分析の新しい方法を提案した。論文として公開。

- (7) 因子分析モデルおよび因子得点の回帰モデルの推定においてスパース性を導入し大規模モデルの推定を可能にした。ここでは潜在変数である因子得点を欠測値と捉え EM アルゴリズムを適用するが、観測変数の欠測については EM の対象としないところにアイデアがあり、その結果、推定の高速化に成功した。論文として公開。
- (8) 因果推論で困難な問題として残されている双方向因果モデルの統計的性質を明らかにした。同モデルは無限回の潜在的な双方向因果効果の集積として現在のデータが得られたと考える欠測値問題である。特に、ある種の非線形因果が推定可能であることを証明した。学会発表として口頭発表。

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9件)

Takagi, Y. and Kano, Y. (2018). Bias reduction using surrogate endpoints as auxiliary variables. *Annals of the Institute of Statistical Mathematics*. Published online on 17 May 2018. 査読有 . [DOI: 10.1007/s10463-018-0667-8]

Morikawa, K. and Kano, Y. (2018). Identification problem of transition models for repeated measurement data with nonignorable missing. *Journal of Multivariate Analysis*. Vol.165, 216-230. 査読有 . [DOI: 10.1016/j.jmva.2017.12.007]

Hirose, K. and Imada, M. (2018). Sparse factor regression via penalized maximum likelihood estimation. *Statistical Papers*, Vol.59(2), 633-662.

Yuan, K.-H., Jamshidian, M. and Kano, Y. (2018). Missing data mechanisms and homogeneity of means and variances-covariances. *Psychometrika*, Vol.83(2), 425-442. 査読有 . [DOI: 10.1007/s11336-018-9609-x]

Morikawa, K., Kim, Jae-Kwang and Kano, Y. (2017). Semiparametric maximum likelihood estimation under not missing at random. *Canadian Journal of Statistics*, Vol.44, 393-409. 査読有 . [DOI: 10.1002/cjs.11340]

Takahashi, M., Iwasaki, M., and Tsubaki, H. (2017). Imputing the mean of a heteroscedastic log-normal missing variable: A unified approach to ratio imputation. *Statistical Journal of the IAOS*, Vol.33, 763-776. 査読有.

Hirose K., Fujisawa, H. and Sese, J. (2017). Robust sparse Gaussian graphical modeling. *Journal of Multivariate Analysis*, Vol.161, 172-190. 査読有.

狩野 裕・今田美幸(2017)データの欠測. *電気情報通信学会誌*, Vol.100, 1274-1278. 査読なし.

佐野文哉・岩崎 学 (2017). 見かけ上のノンコンプライアンスが存在する場合の因果効果の感度分析の一例. *行動計量学*, Vol.44, 199-207. 査読有.

〔学会発表〕(計6件)

Nagase, M. and Kano, Y. (2017). Cyclic nonlinear structural equation models. Joint Statistical Meeting. Baltimore, Maryland, USA.

Nagase, M. and Kano, Y. (2017). Causal analysis with cyclic structural equation models. IFCS2017, Tokai University (Tokyo, Japan).

高木義治・狩野 裕(2017). Bias reduction using auxiliary variables in clinical trial. 2017 年度統計関連学会連合大会, 南山大学(愛知県・名古屋市).

高木義治・狩野 裕 (2016). NMAR のもとでサロゲートエンドポイントを補助変数に用いて得られる推定量のバイアスへの影響について. 2016 年度統計関連学会連合大会, 金沢大学 (石川県・金沢市).

岩崎 学・高野海斗・戸松真太郎(2016). 説明変数に欠測を含む回帰モデルによる予測. 日本行動計量学会, 札幌学院大学(北海道・札幌市).

Morikawa, K., Imori, S. and Kano, Y. (2016). A new information criterion for ignorable missing data. IC-SMHD-2016, Cappadocia, Turkey.

〔図書〕(計1件)

高井啓二, 星野崇宏, 野間久史(2016). 欠測データの統計科学. 岩波書店. Pp.1-240.

〔その他〕

ホームページ(小規模セミナーの一覧)
<http://www.sigmath.es.osaka-u.ac.jp/~kano/research/seminar/tokubetsu/abstract.html>

6. 研究組織

(1)研究代表者

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

(2)研究分担者

岩崎 学 (IWASAKI, Manabu)
横浜市立大学・データサイエンス学部・教授
研究者番号: 40255948

(3)連携研究者

高井 啓二 (TAKAI, Keiji)
関西大学・商学部・准教授
研究者番号: 20572019

大津 起夫 (OTSU, Tatsuo)
(独)大学入試センター・研究開発部・教授
研究者番号: 10203829

廣瀬 慧
九州大学・マス・フォア・インダストリ
研究所・准教授
研究者番号: 40609806

菊地 賢一
東邦大学・理学部・教授
研究者番号: 50270426

伊森 晋平
広島大学・大学院理学研究科・助教
研究者番号: 80747345

(4)研究協力者

森川 耕輔 (MORIKAWA, Kosuke)
東京大学・地震研究所・特任研究員
研究者番号: 40824305

今田 美幸 (IMADA, Miyuki)
日本電信電話(株)NTT未来ねっと研究所

高木 義治 (TAKAGI, Yoshiharu)
サノフィ(株)・統計解析・プログラミング部・開発職 / 大阪大学・大学院基礎工学研究科・博士後期課程

長瀬 真利雄 (NAGASE, Mario)
フリーランス / 大阪大学・大学院基礎工学研究科・博士後期課程

(5)海外研究協力者

Jae-Kwang Kim, Professor
Iowa State University
Department of Statistics

Ke-Hai Yuan, Professor
University of Notre Dame
Department of Psychology

Mortaza (Mori) Jamshidian, Professor
California State University, Fullerton,
Department of Mathematics

以上