

氏 名 富田 裕章

学位(専攻分野) 博士(統計科学)

学位記番号 総研大甲第 1997 号

学位授与の日付 平成 30 年 3 月 23 日

学位授与の要件 複合科学研究科 統計科学専攻
学位規則第6条第1項該当

学位論文題目 多重代入法におけるバイアス補正推定量に関する研究

論文審査委員 主 査 教授 藤澤 洋徳
准教授 逸見 昌之
准教授 野間 久史
教授 服部 聡 大阪大学 大学院
医学系研究科
教授 金森 敬文 東京工業大学 情報理工
学院

論文の要旨

Summary (Abstract) of doctoral thesis contents

本論文では、欠測値が含まれたデータに対し、代表的な解析手法の 1 つである多重代入法を用いた時に得られる新たな推定量を提案した。

この推定量は多重代入法において生じる「欠測値に代入される補完値の生成方法が妥当ではなかった場合に、解析結果として得られる推定量には一致性が存在しない」という問題に対して、重み付き最尤推定法を用いて推定量を導出することで対処したものである。

本論文は全 7 章から構成されており、78 ページからなる。すべて日本語で執筆されている。

第 1 章では序論として欠測値に関する統計手法について概観し、問題提起を行うとともに、本論文の構成について記されている。

第 2 章、第 3 章は先行研究のレビューである。第 2 章では欠測に関する定義について述べられている。欠測の発生メカニズムや欠測の無視可能性の概念の定義を行い、多重代入法を除いた欠測を含んだデータの伝統的な解析手法について述べられている。例えば完全ケース解析法や単一代入法について触れられている。

第 3 章では、多重代入法について解説がなされている。Rubin によって提唱されたオリジナルの多重代入法の定式化と、それ以後の発展について述べられている。また、多重代入法を実際の解析で用いるための計算手法についてもまとめられている。計算手法は、Rubin によって定式化された多重代入法がベイズ統計学の手法であることから、2 つのパラメトリック法をはじめに取り上げている。一つは単調性のあるデータに対して用いられる逐次回帰法、もう一つは、欠測値を含むすべての変数について同時分布をモデル化し、同時分布から代入を行う条件付き分布を生成する同時分布モデリング法である。その後、多重代入法の発展を踏まえ、連鎖式による多重代入法について述べている。この手法は、同時分布について仮定をおかない代わりに、欠測しうる変数 1 つ 1 つに対して、他の変数からの条件付き分布をモデル化し、Gibbs Sampling 同様に逐次的に補完値を生成することが特徴である。本手法は実用上広く用いられているほか、特に第 6 章で扱う、回帰分析における説明変数が欠測した場合の補完手法として開発された「独立モデルに融和的な完全条件付き分布の指定法」に発展するため詳述している。その後ノンパラメトリック法による代入など、そのほかの補完法について触れ、次章以降につながる問題提起を行っている。

第 4 章では、重み付き最尤推定法の枠組みから、一変数が欠測し、なおかつ欠測値を含む変数に対し、他の変数を条件つけた場合の真の確率密度が既知であるとする、補完値を発生させた代入モデルの密度で真の密度を除いたものを、欠測値が存在する部分の尤度に掛け合わせた重み付き最尤推定によって、漸近的に一致性のある統合推定量が得られることを KL-divergence を用いて示している。この統合推定量は Rubin が提唱した、多重代入法において広く用いられている Rubin のルールとは異なる新しいものである。なお、こうした一般的な方法と異なる統合推定量は過去にも検定の文脈で導出されたことがあり、また推定量の一致性を考える上で最尤推定を用いるほうが容易であるために、こうした手法の提案は有意義なことと考えられる。その後、実際の解析においては真の条件付き密度

(別紙様式 2)
(Separate Form 2)

は一般的に不明であることから、条件付き密度推定法を用いることによって重みを推定する手法を提案した。この手法を「バイアス補正多重代入法」と呼んでいる。

第 5 章では、説明変数が欠測した場合の本手法の適用について検討がなされている。この状況下では一般に完全ケース解析が一致性を持たず、また代入モデルが誤っている場合には通常の多重代入法では一致性をもたないことが知られている。本章では数値実験によってバイアス補正多重代入法が、完全ケース解析や通常の多重代入法に比べて偏りが少ない推定量を得ることが可能であることを示した。あわせて、冠動脈疾患に関する実データ解析に提案手法を利用した結果について検討をおこなった。この解析の結果から、実データ解析においてもバイアス補正多重代入法が有用であることが示されている。

第 6 章では、目的変数が欠測した回帰分析において、多重代入法を利用した新たな予測手法について提案をおこなっている。この提案にあたっては、第 5 章までに提案した手法に交差検証法によって定めるチューニングパラメータを加えている。このパラメータの追加によって、本状況下で一致性を持つことが知られている完全ケース解析などを含んだ手法から、代入の誤り具合によらずにバランスをとった予測を行うことができる手法の提案となっている。実際に数値実験をおこなうことにより、提案手法が代入の誤り具合に応じて、チューニングパラメータを変化させることによって、平均的に予測二乗誤差が小さくなるような予測手法となっていることを確認している。

第 7 章ではまとめと今後の課題について述べられている。

博士論文審査結果の要旨

Summary of the results of the doctoral thesis screening

(論文審査結果) [2018年 2月 7日実施]

富田裕章氏の博士論文本審査を、10時00分から2時間に渡って本人と5名の委員全員の出席の下に行い、公開の博士論文発表会および本審査のための会議を行った。その結果、富田氏の博士論文は学位を授与するのに十分な内容であると判断した。

[論文の概要について]

本論文は全7章78ページから成り、全て日本語で執筆されている。論文の主な内容は、欠測が生じているデータへの多重代入法の適用において、欠測の補完値を生成するモデルが妥当ではなかった場合に、解析段階で重み付き最尤法による調整を行うことによって、一致性を有する推定量が得られることを主張するものである。推定量の理論的な導出の他、シミュレーションによる性能評価と実データへの適用についても議論がなされている。

第1章では、序論として欠測値を含むデータに対する統計手法について概観され、問題提起を行うとともに、本論文の構成について記されている。

第2章、第3章は先行研究のレビューである。第2章では、欠測の発生メカニズムや欠測の無視可能性等の欠測に関する諸概念の定義が行われ、欠測を含んだデータに対する、多重代入法以外の基本的な解析手法について述べられている。第3章では、Rubinによって提唱されたオリジナルの多重代入法の手順と、それ以後の発展について述べられている。また、多重代入法を実際の解析で用いるための欠測値補完手法についてもまとめられている。補完手法については、単調性のあるデータに対して用いられる「逐次回帰法」や「データ拡大法」といった古典的な手法の他に、実用上広く用いられている、「連鎖式による多重代入法」についても述べられている。この手法は、同時分布について仮定を置かない代わりに、欠測が生じている個々の変数に対して他の変数からの条件付き分布をモデル化し、ギブスサンプリングと同様に逐次的に補完値を生成することが特徴である。さらに、ノンパラメトリック法による欠測値補完についても言及されている。

第4章では、欠測値を補完するためのモデルが誤特定されている場合に、データの解析段階で調整を行うことによって、漸近的に一致性のある統合推定量が得られることを重み付き最尤法の枠組みで示している。調整に用いる重みは、欠測値補完で用いる条件付き密度関数と真の条件付き密度関数との比として与えられるが、真の条件付き密度は一般に未知であるため、「最小二乗条件付き密度推定」と呼ばれる機械学習の分野で提案されている推定法を用いることによって重みを推定することを提案した。本論文では、以上の手順によって行う多重代入法を「バイアス補正多重代入法 (BCMI)」と呼んでいる。

第5章では、回帰分析において説明変数に欠測が生じた場合への、本手法の適用について検討がなされている。この状況下では一般に完全ケース解析によって得られる回帰係数の推定量が一致性を持たず、また補完モデルが誤っている場合には、通常の方法では一致性を持つ推定量が得られない。本章では、数値実験によってバイアス補正多重代入法が、完全ケース解析や通常の方法に比べて偏りが少ない推定量を与えることを示している。また、冠動脈疾患に関する実データ解析に提案手法を適用した結果についても

(別紙様式 3)

(Separate Form 3)

検討がなされており、実データ解析においても有用であることが示されている。

第 6 章では、目的変数に欠測が生じた場合への本手法の適用について検討がなされている。この場合は、(欠測のメカニズムがMARの下では) 一般に完全ケース解析による推定量が一致性をもつ状況であるので、予測精度の向上に提案手法を用いることができるか否かを検討し、新たにハイパーパラメータを用いた交差検証を追加することで、完全ケース解析と多重代入法 (BCMI) のうち、補完モデルの外れ具合に応じて適切な解析手法を選ぶことができる方法の提案を行っている。この方法と完全ケース解析、通常の多重代入法、(交差検証なしの) BCMI を数値実験で比較し、ここで提案している方法の有用性が示されている。

第 7 章では、本論文で得られた結果のまとめと今後の課題について述べられている。

[論文の評価について]

本論文では、欠測データ解析手法の 1 つである多重代入法において、欠測の補完モデルが誤特定されていても、解析段階で重み付け最尤法による調整を行うことによって、一致性を有する統合推定量が得られることを示し、その有用性が数値実験や実データ解析を通して検証されている。欠測の補完モデルとしてパラメトリックモデルを仮定せずノンパラメトリックに行えば、補完モデルの誤特定の問題を気にする必要はないので、その場合と比べて提案手法が常に優っている (例えば推定精度が良くなる) というわけではないが、補完モデルとして補完を行いやすいものを設定できるというのは 1 つの利点であり、事後的に補正を行うことで一致性を有する統合推定量を得るという本提案手法は、実際の計算手続きの負担を軽くするという面も有する。一方、多重代入法に関するこれまでの研究では、多重代入法で得られる統合推定量の分散について議論するものが多く、また、欠測値の補完方法の提案も多くあるが、補完モデルの誤特定によって生じるバイアスや多重代入によって得られる推定量の統合方法についての研究は少ない。その意味でも、本論文の内容は、多重代入法の研究において新しい切り口を提案しているものと考えられる。

なお、本論文の第 4 章と第 6 章の内容は、査読付き学術誌「応用統計学」に採択済みであり、また、第 4 章と第 5 章の内容は、査読付き国際学術誌「Statistics in Medicine」に投稿し、現在改訂中であるが、いずれも申請者が筆頭著者となっている。以上から、博士論文審査委員会は、申請者の博士論文が博士 (統計科学) に十分値する水準であると全員一致で判断した。