

令和4年度 未来研究ラボシステム 研究成果報告書

研究種目：個人研究

研究期間：令和4年10月～令和5年9月

研究課題名：無視できない欠測値データ解析におけるセミパラメトリック漸近有効推定量の提案—二重機械学習(double machine learning)を用いた最適な推定量の構築—

ラボ長

所属：基礎工学研究科システム創成専攻・社会システム数理領域

氏名：森川 耕輔

研究成果

有限母集団(finite population)の特性を調査する際、時間的・経済的な制約のため母集団全体からデータを収集することは難しい。そのため標本調査では、コストを最小限にしつつ情報量を落とさないように母集団全体から一部のデータ (sampled data)のみを抽出することを考える。図1でその概要を示す。

有限母集団はある無限母集団からの独立同一標本と考えられるが、最終的に得られる観測データは包含確率という調査者が決定する確率値に依存して抽出されるため、偏った標本となる。例えば図1では値が極端に小さいデータは優先的に抽出されるようにデザインされた標本である。この包含確率の情報を用いれば、得られた標本が偏っていたとしてもそのバイアスを補正し興味のある無限母集団のパラメータを推定することが可能となる。古典的な方法では、Horvitz and Thompson (1952)による重み付き推定量がある。しかし、この推定量は情報を無駄にしており、有効推定量ではない。

本研究では、セミパラメトリック推定の理論を用いることで、標本調査における種々の問題に対する効率的な推定量を2つ提案した。1つはパラメトリック作業用モデルを用いる方法で、もう1つは二重機械学習によるノンパラメトリック作業用モデルを用いる方法である。前者の方法は、用意した作業用モデルが正しい場合はセミパラメトリック漸近下限に達する推定量であるが、後者は常に下限に達する推定量である。図2では推定対象を回帰モデルと条件付き密度関数とした場合の回帰モデルの傾きの推定量の箱ひげ図である。上図の設定では、既存手法(CC, HT, CML)より提案手法(Eff)の推定効率が圧倒的に良いことが示されている。また、下図の設定では推定手法に二重機械学習による推定方法も示している。作業用モデルの特定が難しいような状況でも、提案手法による二重機械学習を用いることで効率的な推定量を構成可能となる。

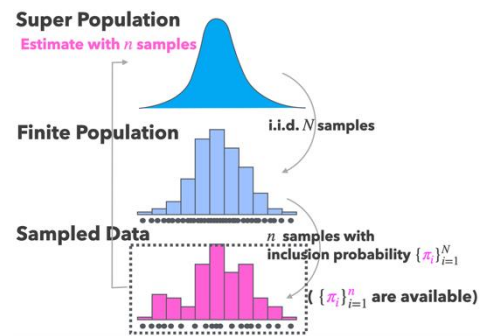


図2. 標本調査の概要。

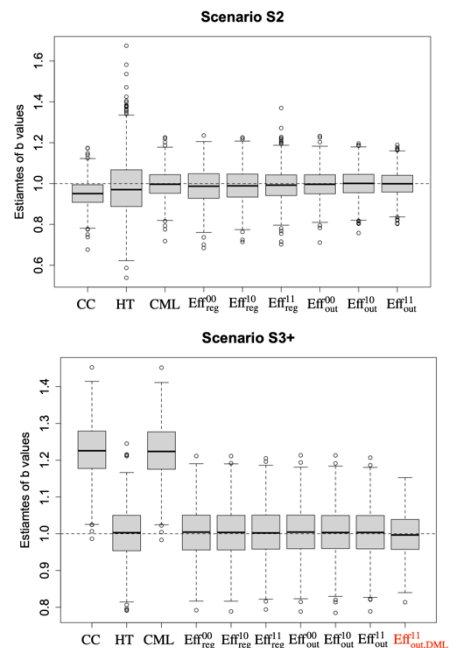


図1. 2つの設定下での提案推定量の箱ひげ図。Effが提案手法。reg, outはそれぞれ推定対象が回帰モデルと条件付き密度の場合の提案推定量。赤色の推定量が二重機械学習による推定量で、それ以外のEffはパラメトリック作業用モデルを用いた提案推定量。点線は真値。

キーワード：

数理統計学，欠測値データ解析，標本調査，生存時間解析，セミパラメトリック推測

研究経費（R4年度）の内訳

備品費	消耗品費	旅費	謝金	その他	合計
48,304 円	121,696 円	0 円	0 円	0 円	170,000 円

共同研究者等

(1) 共同研究者（氏名・所属）

Jae Kwang Kim・アイオワ州立大学統計学部

寺田吉壺・システム創成専攻数理科学領域

(2) 研究協力者（氏名・所属・学年（学生の場合））

発表論文等（令和5年3月31日現在）

[雑誌論文]

Morikawa, K. and Kim, J. K. (2022). Semiparametric adaptive estimation under informative sampling. arXiv:2208.06039. submitted to *Annals of Statistics*.

[学会発表]

Morikawa, K. and Kim, J. K. (2022.10). **NMAR** 欠測値データに対する最適なセミパラメトリック推測法と標本調査への応用. 第49回大阪医学統計セミナー.

森川 耕輔. (2022.10). 不完全データに対する統計解析と統計地震学への応用. 第20回情報計測オンラインセミナー.

Morikawa, K. and Kim, J. K. (2023.03). Semiparametric adaptive estimation under informative sampling. CSSM seminar at Iowa State University.

外部資金獲得状況・申請状況

令和5(2023)年度 基盤研究(A)「情報統計が拓く地震モデリング数理基盤」. 研究分担者.

参考となるHP等

<https://sites.google.com/site/kosukemorikawa/>