

傾向スコアモデルにおける変数選択手法

因果推論の分野において交絡を制御するための効果的なアプローチとして傾向スコアによる方法が広く用いられている。傾向スコアを用いて処置効果を推定する際は、傾向スコアモデルに含める共変量の選択が重要である。処置にのみ影響を与える変数を傾向スコアモデルに含めると平均処置効果の推定量の分散が大きくなり、バイアスの原因になること、結果変数のみ影響を与える変数を傾向スコアモデルに組み込むと平均処置効果の推定量の分散が小さくなるということがシミュレーションにより示唆されている [1][2]。有向非巡回グラフや背景知識を使用し、どの共変量を含めるかを決定することが望ましいが、事前知識が不足する場合や候補となる変数が非常に多い場合には、他の方法を検討する必要がある。

そのような状況の対処法の一つとして、データ駆動型アプローチがある。しかし、従来のデータ駆動型のアプローチであるステップワイズ法やスパース推定などの罰則付き回帰手法は因果推論の立場から妥当な変数を選択する方法にはなっておらず、そのまま適用すると適切な結果を導くことが懸念される。

本抄読会では、傾向スコアモデルにおける変数選択を行う手法として Outcome adaptive lasso (OAL)[3]を紹介する。従来の手法との比較や実データへの適用例を紹介する。さらに、最近提案された OAL の発展である Generalized outcome adaptive lasso[4]を紹介する。

参考文献

- [1] Xavier de Luna, Ingeborg Waernbaum, and Thomas S. Richardson. Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika*, 98(4):861–875, 2011.
- [2] M. Alan Brookhart, Sebastian Schneeweiss, Kenneth J. Rothman, Robert J. Glynn, Andrew M. Walker, and Timm Stürmer. Variable selection for propensity score models. *American Journal of Epidemiology*, 163(12):1149–1156, 2006.
- [3] Susan M. Shortreed and Ashkan Ertefaie. Outcome-adaptive lasso: variable selection for causal inference. *Biometrics*, 73(4):1111–1122, 2017.
- [4] Ismaila Baldé, Yi Archer Yang, and Geneviève Lefebvre. Reader reaction to “outcome-adaptive lasso: Variable selection for causal inference” by shortreed and ertefaie (2017). *Biometrics*, 79(1):514–520, 2023.