

傾向スコア解析のための二重頑健規準および差分の差法への展開

二宮 嘉行

統計数理研究所 数理・推論研究系

総合研究大学院大学 統計科学専攻

因果推論において、傾向スコアを用いて逆確率重み付け推定や二重頑健推定をおこなうセミパラメトリックアプローチは今や標準であり、様々な方向で急速に拡張・一般化が進んでいる。一方、統計解析においてモデル選択は不可欠であるにもかかわらず、因果効果に影響を与える変数を選択するための情報量規準は開発途上の段階にある。本講演では、情報量規準の基本的な考え方にに基づき、AIC 型のそれを導出する。情報量規準の礎となるリスク関数としては Kullback-Leibler ダイバージェンスに基づくものを考え、線形モデルとして表されるタイプのものとは限らない一般の因果推論モデルを扱う。推定する因果効果としては、コントロール群平均処置効果やケース群平均処置効果など、一般の母集団におけるものを想定する。また、割り当て変数のモデルと結果変数のモデルのどちらか一方は誤っていてもよい、という二重頑健性が推定で重用されていることを鑑み、どちらか一方は誤っていても妥当な基準になるという、情報量規準自体の二重頑健性をもたせるようにする。数値実験では、導いた基準と形式的な議論で導かれた既存の基準を比較し、前者の性能が上回ることを確認する。具体的には、推定された構造と真の構造の差がいずれのシミュレーション設定においてもはっきりと小さくなること、および真の構造を含むモデルあるいはそれに近いモデルを選択する確率がはっきりと高くなることを確認する。ここまでは Baba and Ninomiya ('22 arXiv) の内容であり、本講演では、このようなアプローチをセミパラメトリック差分の差法 (Abadie '05 The Review of Economic Studies) や共変量バランシング (Imai and Ratkovic '14 JRSSB) に適用することも考える。無視できる割り当て条件ではなく条件付き平行トレンド条件を課し、僅かに近似に頼らなければならないものの、性能的には問題のない情報量規準を導く。本内容は馬場崇充氏 (塩野義製薬, 総合研究大学院大学) との共同研究に基づく。