

セミパラメトリックモデルにもとづく共変量調整と欠測データ解析

ランダム化研究における治療効果のセミパラメトリック推定について考える。以下のような 2 群比較の単純な状況設定で話をすすめる。ランダム化前のアウトカム(治療前値 Y_1)、ベースライン共変量(X_1)、ランダム化後の時間依存性共変量(X_2)、治療後のアウトカム(Y_2)、割付群 ($Z=1$:試験群、 $Z=0$:対照群)。ただし、 Y_2 の値が欠測する対象者が存在する。ターゲットパラメータである因果治療効果は以下の式で定義される周辺治療効果 β である。

$$\begin{aligned} & \{E(Y_2 | Z = 1) - \mu_1\} - \{E(Y_2 | Z = 0) - \mu_1\} \\ & = E(Y_2 | Z = 1) - E(Y_2 | Z = 0) = \beta \end{aligned}$$

上記の治療効果 β を推定する方法にはいくつかのものがあるが、どの方法がよいかの一般的なコンセンサスは存在しない。

セミパラメトリックモデルのもとでは、 (Y_1, Y_2) が正規分布に従わなくても、また Y_1 と Y_2 が線型関係でなくても、 t 検定推定量や ANCOVA などの多くのポピュラーな β の推定量は、CAN (consistent and asymptotically normal) 推定量であることが知られているものの、現実のデータでは、非線型関係が多くみられ(効率ロスにつながる)、 Y_2 の欠測メカニズムに対してMCARを仮定することは非現実的である(バイアスにつながる)。

本抄読会では、 β の推定量のクラスとして、RAL (regular asymptotically linear) 推定量を考え、治療後の時間依存性共変量 X_2 の情報も考慮した欠測メカニズム MAR (missing at random) のもとでのセミパラメトリック有効推定量を紹介する。また、セミパラメトリック推定量を得る際に必要な(欠測確率、 Y_2 の予測値に対する)モデル構築の際の注意点を与えるとともに、ANCOVA、IPCW などのポピュラーな推定量と性能比較したシミュレーション実験の結果を示す。