

Effects of inclusion of auxiliary variables on bias reduction for NMAR missingness

大阪大・基礎工 狩野 裕
サノフィ・統計解析・プログラミング部 高木義治

Observed Likelihood (OL or FIML) による尤度法は欠測値データ解析の定番である。OL による推測は欠測メカニズムを特定する必要がなく識別性も満たされることが多い。しかし、欠測メカニズムが MAR であるという制限的な仮定が満たされなければならない。NMAR の下では OL による MLE はバイアスをもつ。そういったバイアスを減少させる目的で、直接分析目的には関係しないが欠測変数や欠測指標と関係する補助変数 (auxiliary variables) をモデルに組み込む方法がある。補助変数導入の有効性は数値実験で確かめられている (e.g., Ibrahim, et al. 2001; Graham 2003; O'Neill & Temple 2012; Collins et al. 2001)。補助変数導入がバイアスを減少させる理由として以下の3点が語られている。

1. MAR の仮定をより妥当なもの (more reasonable or plausible) にする
2. 欠測変数の代理変数 (proxy variable) となる
3. 欠測指標の不確定性 (uncertainty) を減少させる

一方で、補助変数の組込みは MLE のバイアスを増加させることがある、補助変数は欠測変数や欠測指標変数と非常識的に高い相関が求められ、それゆえその効果は限定的である、といった批判もある。本発表では上記三点の理論的背景を検討する。単純なモデルを考え、補助変数の具体的な役割を上述の3点と関連させて議論する。補助変数法に対する批判についても、それがどのような状況で生じるのかを理論的に検討する。

具体的には、NMAR の欠測が生じ得る観測変数 Y の期待値 μ_y の推定問題を考える。 X を補助変数とする。得られるデータを以下のように表す。

$$\begin{bmatrix} Y_1 \\ X_1 \end{bmatrix}, \dots, \begin{bmatrix} Y_m \\ X_m \end{bmatrix}, X_{m+1}, \dots, X_n$$

正規性の仮定の下で、補助変数を含めない μ_y の推定量と含めた場合の推定量はそれぞれ

$$\bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i, \quad \hat{\mu}_y = \bar{Y} + \frac{s_{xy}}{s_{xx}}(\hat{\mu}_x - \bar{X}) \quad (1)$$

で与えられる。記号の定義は当日示す。これらの推定量のバイアスを正確に評価することにより、上述の課題に答える。なお、NMAR たる欠測メカニズムは任意である。また、(1) の推定量の比較は Complete Case Analysis と OL による MLE との比較にもなっている。

Collins, Schafer and Kam (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol. Meth.*, 6, 330-51.

Graham(2003). Adding missing-data-relevant variables to FIML-based structural equation models. *SEM*, 10, 80-100.

Ibrahim, Lipsitz and Horton(2001). Using auxiliary data for parameter estimation with non-ignorably missing outcomes. *Appl. Statist.*, 50, 361-373.

O'Neill and Temple (2012). The prevention and treatment of missing data in clinical trials: an fda perspective on the importance of dealing with it. *Clinical Pharmacology and Therapeutics*, 91(3), 550-554.

Schafer & Graham (2002) Missing data: Our view of the state of the art. *Psychol. Meth.*, 7, 147-177.

狩野 裕 (2014). NMAR の下での尤度法. 日本統計学会誌. Vol.43(2), 359-377.