多重代入法におけるバイアス補正推定量について

総合研究大学院大学　富田　裕章
統計数理研究所　藤澤　洋徳　　統計数理研究所　逸見　昌之

多重代入法 (Multiple Imputation, MI) とは Rubin(1987) によって提唱された欠測値を含むデータにたいする解析手法の一つである．これは欠測値に対して補完値を独立に複数回生成し，複数の擬似完全データを作成したのち，それぞれに対して推定量を計算し，それらを統合して最終的な推定量（統合推定量）を得る手法である．

MI において問題となる事柄の一つに，補完の妥当性がある．また本報告では代入を行うことで生成される各完全データセットで計算される推定量の統合方法について検討する．

補完方法としてよく用いられるものに，欠測値に対し回帰モデルを仮定して補完する手法がある．特に多変量の欠測が存在した際の手法では van Buuren ら (2006) による連鎖式を用いた多重代入（Multiple Imputation with Chained Equation： MICE）がある．これは欠測する成分１つ１つに対して他の変数を説明変数とした回帰モデルを仮定し補完する手法である．この方法は利便性が高い反面，得られる統合推定量が妥当であるためには，補完するモデルがなんらかの意味で正しい補完モデルを利用する必要がある．例えば，補完モデルが真の条件付き密度分布を含んでいれば一致性のある統合推定量を得ることができる．

さて我々は，補完モデルの妥当性には拘らず，欠測値の補完モデルが正しくなかったとしても，良い推定値を得ることができる **バイアス補正多重代入法（BCMI）** を提案した．この手法は，補完値に対する真の条件付き密度を機械学習を用いて推定した条件付き密度と仮定した補完モデルの条件付き密度との密度比を重みとして利用する，**重み付き最尤推定法**である．

特に，説明変数のうち１変数に欠測が生じている状況下で回帰分析を行う際の BCMI を以下に述べる．目的変数を Y，欠測しうる説明変数を X，欠測しない説明変数を Z をする．欠測した X について M 回補完を行い，それぞれ補完値を x^m $(m = 1, \ldots, M)$ と表す．回帰モデルを $f(y|x, z; \theta)$，重み付け関数を $w(x, y, z)$，推定した条件付き密度を $\hat{g}(x|y, z)$，補完モデルを $h(x|y, z)$ とすると，BCMI による推定量 $\hat{\theta}$ は

$$\hat{\theta} = \arg\max_{\theta} \left[\sum_{x\text{ が観測}} \log f(y|x, z; \theta) + \sum_{x\text{ が欠測}} \frac{1}{M} \sum_{m=1}^{M} w(x^m, y, z) \log f(y|x^m, z; \theta) \right]$$

として与えられる．ただし，

$$w(x, y, z) = \frac{\hat{g}(x|y, z)}{h(x|y, z)}$$

である．本日の報告では，BCMI をさらに発展させる方向性についても述べたい．