

経験尤度法を用いた NMAR データに対する セミパラメトリック多重代入法

大阪大・基礎工

Iowa State U., Dept. Statist.

森川 耕輔

Jae Kwang Kim

1. はじめに

データの“欠測”は疫学，経済学，標本調査といったあらゆる分野で重大な問題となっている。そのような問題を扱う際，適当なモーメント条件を満たす推定方程式をベースにした推定法 (e.g. inverse probability weighted method, IPW 法); 欠測値を補完する方法 (e.g. 多重代入法)，の2つの方法が主流である。いずれの方法にしる，データが欠測する原因である“欠測メカニズム”の特定が重要となる。特に，欠測メカニズムが観測されるデータのみで説明可能な場合，その欠測メカニズムは MAR (missing at random) と呼ばれる。また，そうでない場合，NMAR (not missing at random) と呼ばれる。本研究では，NMAR データに対しても有効な多重代入法を提案する。

NMAR データの解析には欠測メカニズムの特定だけでなく，欠測する変数に関するモデリングの特定も必要となる。例えば，欠測し得る応答変数 Y ，欠測の無い説明変数 X が観測される，単純な無回答データを考える。この場合， Y が観測（欠測）している場合 $1(0)$ を表す指示変数 δ を用意すれば，欠測メカニズムは $P(\delta = 1 | x, y)$ ，欠測変数に関するモデルとは $Y | X$ の分布を表す。応用上，これらの2つのモデルを正しく特定することは非常に難しい。

2. 提案手法

そこで，本研究では欠測する変数に関する特定を不要とするセミパラメトリックな多重代入法を提案する。欠測データを補完する際，その欠測する変数に関する代入モデル（無回答データの場合， $Y | (x, \delta = 0)$ ）の分布をうまく推定することが必要であるが，本研究によりその分布をデータの分布に何も制約を必要としない，欠測メカニズムに関する仮定のみで推定することが可能となる。本研究は以下の2つのステップからなる：

- (i) 経験尤度法を用いた NMAR データに対する（頻度論的）推定法の提案;
- (ii) (i) を利用した多重代入法の提案。

経験尤度法を用いた推定量は一般化推定方程式ベースの推定量に比べて，推定量の信頼性検定において大きなアドバンテージがある (Qin and Lawless, 1994)。通常仮説検定を行う際，パラメータの漸近分散を求めねばならず，(特にセミパラメトリック推定量の) 漸近分散は非常に複雑な形をしていることも多いが，経験尤度法を用いた推定量では漸近分散を用いず推定量の仮説検定が行える。それにも関わらず，これまで経験尤度法を用いた推定法は提案されている (e.g., Qin et al., 2002) が，有効な推定法はない。そのため，(i) の研究で，経験尤度法を用いた自然なセミパラメトリック漸近有効推定量を提案する。また (i) で推定された経験尤度を尤度と見做すことで，パラメータの事後分布を経験的に求めることができる (Lazar, 2003)。これらを利用することで，観測されているデータを最大限生かしたセミパラメトリックな欠測値の代入法を提案する。

3. 参考文献

- [1] Lazar, N. A. (2003). *Biometrika*, **2**, 319–326.
- [2] Qin, J., Leung, D. and Shao, J. (2002). *J. Amer. Statist. Assoc.*, **97**, 193-200.
- [3] Qin, J. and Lawless, J. (1994). *Ann. Statist.*, **22**, 300-325.