

2014年2月5日

修士1年 奥井佑

欠測過程が無視できないデータの解析方法

反応変数の欠測確率が未測定の変数値に影響を受けることを、欠測過程が無視できない(Nonignorable)、または Missing not at random(MNAR)という。欠測過程が無視できない場合には、欠測確率と反応変数・共変量との間にモデルを仮定することで共変量の反応変数への効果(パラメータ)が推定され、モデル化と推定の方法について数多くの方法が考案されている。MNAR なデータのパラメータ推定方法の主なものとして尤度関数を用いた方法があり、補完(Imputation), Pseudo likelihood を用いて推定が行われる。

補完による方法では、反応変数と共変量との関連と欠測過程の確率関数がモデルにより完全に明示することで数値計算を用いて複数のパラメータを推定することが行われる。欠測値をパラメータとみなして EM algorithm や Markov Chain Monte Carlo を適用する方法が主であるが、平均スコア関数の近似により推定を行う Fractional Imputation を用いた方法も提案されている。

Pseudo-likelihood による方法は、経時データなど複数の反応変数、欠測変数間の関連性のモデル化やパラメータの推定が困難である場合に、擬似的な尤度関数を代わりに用いることで推定値の不偏性を保つ方法である。Troxel et al.により時点間で観測値は独立と仮定したモデルが最初に考えられ、その後時点間の相関を考慮した拡張がなされ、最近では反応変数、共変量、欠測変数間に Marginal model を取り入れた柔軟なモデルが提案されている。

本抄読会では補完法と Pseudo - likelihood による MNAR なデータの解析方法について、統計理論を中心に説明する。

参考文献

- Kim JK, Yu CL. A semi-parametric approach to fractional imputation for nonignorable missing data. *JSM section on survey research methods*. 2009.
- Kim JK. Fractional imputation using regression imputation model. *JSM section on survey research methods*. 2003.
- Troxel AB, Lipsitz SR, Harrington DP. Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data. *Biometrika* .1998; **85**: 661-72.
- Sinha SK, Troxel AB, Lipsitz SR, Sinha D, Fitzmaurice GM, Molenberghs G, et al. A bivariate pseudolikelihood for incomplete longitudinal binary data with nonignorable nonmonotone missingness. *Biometrics*. 2011; **67**: 1119-26.
- Li H, Yi G. Estimation methods for marginal and association parameters for longitudinal binary data with nonignorable missing observations. *Stat Med*. 2012; **32**: 833-48.