

多重補完法の再考

東京大学 大学院情報学環

小川 光紀

多重補完法 (Rubin (1987)) は、欠測データ解析における補完に基づく方法として代表的な手法の一つである。多重補完法は欠測値を補完する際の枠組みとしてベイズ予測分布を用いるものと頻度論的方法に基づくものに大別され、最終的な推定量の統計的性質については Wang and Robins (1998) および Robins and Wang (2000) で調べられている。

多重補完法に基づき妥当な推定量を得るためには、適切な補完モデルを用いることが重要になる。補完モデルの誤特定は結果のバイアスにつながることから、たとえば van Buuren (2018) などでも補完モデルとして柔軟なモデルを用いることが推奨されている。一方、ノンパラメトリックな補完モデルを安易に用いると、補完モデルからの乱数生成にコストがかかることがあり、そのような困難を回避する工夫についても議論がなされている (Tomita et al. (2018))。

本抄読会では、多重補完法に関するいくつかの先行研究をレビューするとともに、補完モデルの誤特定について柔軟な補完モデルを用いることとは別の視点からの対処法について検討する。

参考文献

- Robins, J. M. and Wang, N. “Inference for imputation estimators.” *Biometrika*, 87(1):113–124 (2000).
- Rubin, D. B. *Multiple Imputation for Nonresponse in Surveys*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York (1987).
- Tomita, H., Fujisawa, H., and Henmi, M. “A bias-corrected estimator in multiple imputation for missing data.” *Stat. Med.*, 37(23):3373–3386 (2018).
- van Buuren, S. *Flexible Imputation of Missing Data*. Chapman and Hall/CRC, New York, second edition (2018).
- Wang, N. and Robins, J. M. “Large-sample theory for parametric multiple imputation procedures.” *Biometrika*, 85(4):935–948 (1998).