

ロバストな欠測データ解析手法

小川 光紀

データの欠測は現実のデータ解析でしばしば遭遇する問題であり，状況に応じて適切な解析方法を適用する必要がある [2]，欠測データ解析の手法には，尤度に基づく方法，推定方程式に基づく方法，補完に基づく方法など様々なものがあるが，いずれの方法でも使用する何かしらのモデルを正しく特定することが，妥当な推定量を与えるための基本的な仮定になっている．たとえば，多重補完における補完モデル，逆確率重み付け推定量における欠測確率のモデルなどがこれに相当する．

欠測データ解析におけるモデル誤特定に頑健な推定量として，二重ロバスト推定量がある [3]．二重ロバスト推定量には，欠測確率のモデルと条件付き期待値のモデルのいずれかが正しく特定されていれば一致性をもつという有用な性質がある．しかし，両方のモデルが誤特定されている状況では，誤特定の程度が大きくない場合であっても，性能が落ち得ることも知られている [1]．そのため，素朴な二重ロバスト推定量の欠点を克服するための様々な提案がされてきた．

本発表では，欠測データ解析におけるモデル誤特定の問題に焦点をあて，ロバストな欠測データ解析手法を提案したいくつかの先行研究について概観する．そして，密度比 [4] を利用した既存の欠測データ解析手法のロバスト化について検討する．

References

- [1] J. D. Y. Kang and J. L. Schafer. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.*, 22(4):523–539, 2007.
- [2] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2002.
- [3] J. M. Robins, A. Rotnitzky, and L. P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.*, 89(427):846–866, 1994.
- [4] M. Sugiyama, T. Suzuki, and T. Kanamori. *Density Ratio Estimation in Machine Learning*. Cambridge University Press, Cambridge, 2012. With a foreword by Thomas G. Dietterich.