

欠損値を含むアンケートデータに対して 頑健な多分因子分析法

甲斐田 将[†] 中田 洋平[†]

[†] 明治大学総合数理学部ネットワークデザイン学科

1. はじめに

本研究では、回答の割り当て方を考慮したアンケートにより、被験者一人当たり設問数を減らすことで被験者の負担を軽減しつつ、統計的正当性の高い分析を可能とするため、欠損値に頑健な多分因子分析法[1]を提案する。

2. 提案手法の概要

多分因子分析法とは、アンケートなどから得られたような順序尺度で表されるデータに対し、統計的正当性の高い因子分析法である[1]。典型的には、多分相関行列を用いて因子分析法を実施することで、実行される。本研究の提案手法も、基本的に類似の構造を持つ。以下に提案手法の手順を示す。

I. 初期化:

従来手法により、閾値の推定を行う。また多分相関行列を単位行列などで初期化する。

II. 多分相関行列の推定:

- i. **潜在変数の事後サンプル生成:** ギブスサンプリング法を用いて、潜在変数の事後サンプルを生成する。なお、潜在変数の事後サンプルは、成分ごとに切断正規分布から生成する。
- ii. **統計量の更新:** 確率的近似 EM 法[2][3]を用いて、統計量を算出する。
- iii. **パラメータの更新:** 算出した統計量を用いて、多分相関行列パラメータを更新する。

III. 因子分析:

推定した多分相関行列パラメータを用いて、因子分析を行う。

3. 検証実験

本研究では、2つの検証実験を実施した。1つ目は、0%、10%、...、50%毎の欠損値比率を変化させた模擬アンケートデータを作成し、従来手法と提案手法での多分相関行列の推定誤差を検証した。また、2つ目の検証実験では、より実際のアンケート収集の状況を想定し、順序尺度の背後に3つの因子が存在する数値例を用いて、因子スコアの推定を行った。その際に、想定するアンケートへの回答人数や回答の割り当て方なども変化した実験を実施し、因子スコアの推定誤差を比

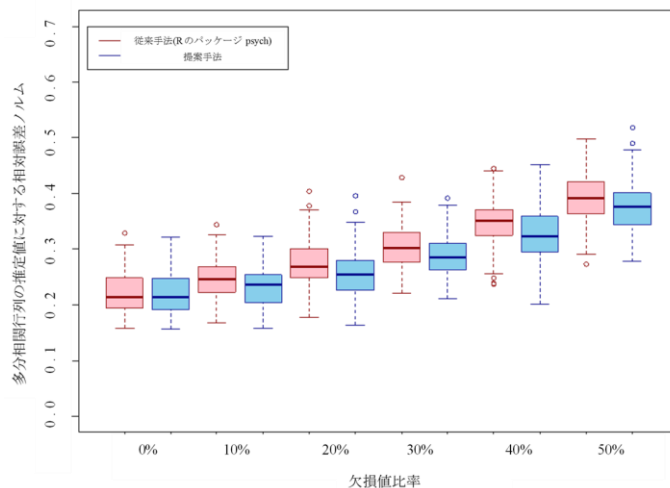


図1. 多分相関行列の推定誤差

較した。図1には、1つ目の検証実験での結果を示す。図示されるように、欠損値が存在する場合、従来手法に比べて、提案手法の誤差の方が小さくなっていることが確認できる。

4. まとめと今後の課題

本研究では、欠損値に頑健な多分因子分析法を提案し、2つの検証実験を実施した。その結果、提案手法の有効性の一端が確認された。今後の課題としては、実際のアンケート調査への適用実験が考えられる。その際には、実際に設問数の増加が被験者にどれだけ負担があり、提案手法によりどれだけ負担を軽減できたのかという観点も加え、主観的指標、および、客観的指標を用いて検証していく。

参考文献

- [1] F. P. Holgado-Tello, S. Chacón-Moscoso, I. Barbero-García, E. Vila-Abad, "Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables," *Quality & Quantity*, Vol. 44, No. 1, 153, 2010.
- [2] B. Delyon, M. Lavielle, and E. Moulines, "Convergence of a stochastic approximation version of the EM algorithm," *Annals of Statistics*, Vol. 27, No. 1, pp. 94-128, 1999.
- [3] K. Estelle, and M. Lavielle, "Coupling a stochastic approximation version of EM with an MCMC procedure," *ESAIM: Probability and Statistics*, Vol. 8, pp. 115-131, 2004.