

# 共変量の影響を除外した成分抽出法

立教大・社会 酒折 文武

## 1 概要

二集団の特徴を比較する際、背後にある共変量が偏っていることにより各種統計量や分布を単純に比較できない場合がある。このようなとき、共変量の偏りを考慮に入れて分析する方法として、傾向スコアを用いて共変量のバランスをとる方法がある。しかしながら過去の研究では、バランスをとることによって結果変数をひとつひとつ比較するに過ぎなかった。

本発表では、共変量のバランスをとったうえで多変量の結果変数の特徴を抽出する新しい手法を提案する。

## 2 手法

比較したい二群(処理群と対照群)への割り当て変数を  $Z$  とし、 $Z = 1$  のとき第一群、 $Z = 2$  のとき第二群へ割り当てるとする。共変量を  $X$  とする。特徴を調べたい変数(結果変数)を  $Y$  とする。

傾向スコアは、あたえられた共変量に対して第一群へ割り当てられる確率

$$e(X) = \Pr(Z = 1|X)$$

のことである。この確率は一般に未知のため、ロジスティック回帰モデル等を用いて推定することになる。通常は、傾向スコアの値が近いもので層別して群間比較をおこなったり、傾向スコアの値によって重み付けをして加重平均を求めることで共変量のバランスをとった推定値を求める。

さて、観測された共変量と結果変数をそれぞれ  $x_{jk}, y_{jk}, j = 1, 2, k = 1, 2, \dots, N_j$  とする。また  $x_{jk}$  より得られた傾向スコアの推定値  $e_{jk}$  とする。このとき、結果変数の線形変換のなかで、 $e_{jk}$  と無相関であり、割り付け変数  $z_{ji}$  との相関比

$$\nu^2 = \frac{S_b}{S_t}$$

を最大とするようなものを考える。ここで  $S_t$  は総平方和、 $S_b$  は群間平方和である。この線形変換により得られる成分は、共変量のバランスを二群でとったうえで、第一群と第二群との違いを抽出するようなものになっているはずである。以降、得られた成分とも無相関になるような制限をつけて相関比の最大化を繰り返していくと、特徴をあらわす成分が順次得られる。

なお、この条件を満たす相関比  $\nu^2$  は固有値問題に帰着しない。したがって  $\nu^2$  の値と係数ベクトルを理論的に求めるのは難しい。そこで Quasi-Newton 法などを用いて数値解析的に求めることになる。

本手法の詳細や数値例などは当日報告する。

## 参考文献

- [1] Rosenbaum, P.R. (1987). Model-Based Direct Adjustment, *J. Amer. Statist. Assoc.*, **82**, 387-394.
- [2] Rosenbaum, P.R. (2002). *Observational Studies*, 2nd ed., Springer Verlag.