

多変量関数データに対する正準相関分析の定式化と解の存在性について

岡山大学大学院, 理化学研究所 AIP 山本 倫生

大阪大学大学院, 理化学研究所 AIP 寺田 吉壺

1. はじめに 多群正準相関分析 (MCCA) とは, 三つ以上の変数群間での共通変動を抽出する多変量解析法の一つであり, Hwang et al. (2012) によって, 関数データに対する方法 (関数多群正準相関分析; FMCCA) へと拡張されている. Hwang et al. (2012) では, 複数の被験者に対して経時的に測定された fMRI データを対象として, 被験者間で共通の変動を抽出することにより, 実験要因に関連する脳部位の特定に成功している. FMCCA は関数データ解析法として有用な方法であるが, いくつか問題点がある. 最も重要な問題点は, FMCCA は本来無限次元である関数データが有限個の基底関数で表現できることを仮定し, さらに, データを確率的に扱わない記述的な解析方法として定式化されている点である. そこで本研究では, 無限次元空間上でのランダムな関数に対して FMCCA を定式化し, 母集団での提案方法の理論的性質について検討する.

2. 提案方法 閉区間 $[0, T]$ 上の実数値二乗可積分関数の集合を L_2 とし, $x, y \in L_2$ に対して内積を $\langle x, y \rangle = \int_0^T x(t)y(t)dt$ と定義する. このとき, P 個の L_2 からなる直積集合を L_2^P とし, $\mathbf{X} = (X_1, \dots, X_P)^\top$ を L_2^P 上でのランダムな関数とする. また, L_2^P 上での内積を $\langle \mathbf{X}, \mathbf{Y} \rangle_P = \sum_{p=1}^P \langle X_p, Y_p \rangle$ と定義する. さらに, 各 L_2 での X_p ($p = 1, \dots, P$) の期待値を EX_p と表し, X_p は予め中心化されているとする. ここで, 変数 p に対する第 k 正準変数を $\eta_{kp} = \langle X_p, \alpha_k \rangle$ と定義する. なお, $\alpha_k = (\alpha_{k1}, \dots, \alpha_{kP})^\top \in L_2^P$ を第 k 正準変数に対応する重み関数とよぶ. このとき, $\ell(\alpha_k) = E[(\sum_{p=1}^P \eta_{kp})^2]$ という基準を考える. ここで, $\eta_k = \sum_{p=1}^P \eta_{kp}$ とすると, 正の整数 $l \leq k$ に対して, $E(\eta_k \eta_l) = \delta_{kl}$ という制約のもとで $\ell(\alpha_k)$ を最大化することにより, 重み関数 α_k を推定する. この方法を一般化関数正準相関分析 (GFCCA) とよぶこととする. いま, $\mathcal{V}$ を $\mathbf{X}$ に対する共分散作用素とし, $\mathbf{h} \in L_2^P$ に対して積分作用素 $\mathcal{W}$ を $\mathcal{W}\mathbf{h}(t) = \int_0^T D(t, s)\mathbf{h}(s)ds$ と定義する. ただし, $D(t, s)$ は $E[X_p(t)X_p(s)]$ を第 p 対角要素に持つ対角行列の値をとる関数である. ここで $\tilde{\alpha}_k = \mathcal{W}^{\frac{1}{2}}\alpha_k$ とすると, 基準 ℓ の α_k についての最大化問題は, 以下の基準 ℓ' の $\tilde{\alpha}_k$ についての最大化問題と等価である:

$$\ell'(\tilde{\alpha}_k) = \left\langle \tilde{\alpha}_k, \mathcal{W}^{-\frac{1}{2}}\mathcal{V}\mathcal{W}^{-\frac{1}{2}}\tilde{\alpha}_k \right\rangle_P.$$

ただし, $\mathcal{W}^{-\frac{1}{2}}$ は $\mathcal{W}^{\frac{1}{2}}$ の逆作用素を表す. 直感的には作用素 $\mathcal{U} = \mathcal{W}^{-\frac{1}{2}}\mathcal{V}\mathcal{W}^{-\frac{1}{2}}$ の固有値問題に帰着する. しかし, 一般に $\mathcal{W}^{\frac{1}{2}}$ が可逆でないため, L_2^P 全体をデータ空間と考えた場合には GFCCA を well-defined に定義することができない. そこで, He et al. (2003) の関数正準相関分析に対するアプローチを参考に, GFCCA が well-defined に定義可能なデータの存在範囲について検討し, GFCCA が定義可能となるための十分条件を与える. さらに, 有限次元の場合に MCCA と等価な方法である等質性分析を拡張した関数等質性分析 (FHA) を定式化し, 無限次元空間における GFCCA との同等性について議論する.

参考文献

- He, G., Müller, H.-G., and Wang, J.-L. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis*, 85, 54–77.
- Hwang, H., Jung, K., and Takane, Y. (2012). Functional multiple-set canonical correlation analysis. *Psychometrika*, 77, 48–64.