共通主成分分析に基づく構造変化点探索とその応用
Structural change detection via common principal component analysis and its application

数学専攻　松川 達也
MATSUKAWA，Tatsuya

1　はじめに

共通主成分分析 (Common Principal Component Analysis; CPCA) は，複数年のデータ集合 (グループ) の分散共分散行列を同時に対角化することで共通の主成分軸を導出して，視覚的に情報・パターン抽出を行う手法として Flury (1984, 1988) によって提唱された．この解析手法の利点は，個々に行う主成分分析に比べ推定するモデルのパラメータ数が少なく，より小さい標準偏差をもつ係数が求められるという意味で安定的であり，また共通の主成分軸で相違や類似度を測れる点などが挙げられる．経済データ，社会構造データなどでは経年データの情報やパターン抽出を考察するための有用な手法として用いられている (勝浦, 1988)．

このように複数年のグループを分析する際，データ構造が大規模災害や経済危機によって変化する点が存在すると考えられる．この変化点を客観的に特定することは実社会において過去に発生した類似する出来事による影響について観察することによって，将来どんなことが起こるかを予測し，それに対処することが可能となる面で重要となる．

本研究では客観的に年度データにおける変化点を特定するために，共分散構造に着目した共通主成分モデルの下での逸脱度（CPCA deviance）に基づく変化点探索法を提案する．変化点前後では共分散構造が異なり，2 つの分散共分散行列に分類されると考える．この 2 つの分散共分散行列に分けることで，境目となるところを変化点として検出することができる．この考えに基づき，共通主成分分析にモデル評価の指標の 1 つである逸脱度を導入することで共通主成分モデル下の逸脱度とフルモデル下での逸脱度の差から構造変化点探索を議論する．

2　共通主成分分析

k 個のグループが存在し，各グループのデータ数を $N_1, N_2, \ldots, N_k$ とする．この時，共通主成分分析はグループの構造を考慮に入れ，一括して主成分分析を行うことでグループ間の類似度や相違を明らかにし共通の測度での比較を可能にする手法である．$k = 1$ のとき，通常の主成分分析に一致する．数理的には，各グループの分散共分散行列 $\Sigma_1, \ldots, \Sigma_k$ を同時対角化する直交行列 W を考える．すなわち，共通主成分分析は，以下の仮説の下で分析される．

$$H_c : W^T \Sigma_i W = \Lambda_i \quad (i = 1, 2, \ldots, k). \tag{2.1}$$

ただし，$\Lambda_1, \ldots, \Lambda_k$ は対角行列であるが，共通ではないことに注意する．

次に共通主成分仮説における直交行列 W，対角行列 Λ_i を最尤法を用いて推定する．まず，第 i グループで観測された p 次元データは次の分布に従うと仮定する．

$$\boldsymbol{x}_{ij} \overset{iid}{\sim} N_p(\boldsymbol{\mu}_i, \Sigma_i) \quad (i = 1, 2, \ldots, k,\ j = 1, 2, \ldots, N_i).$$

ここで $\boldsymbol{\mu}_i \in \mathbb{R}^p$ で Σ_iは $p \times p$ 正定値対称行列である．簡単のために，$n_i = N_i - 1$ とおいて考える．グループ

ごとの不偏標本分散共分散行列を $S_i = \dfrac{1}{n_i}\displaystyle\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)^T$ とおく．このとき，行列 n_iS_i は互いに独立に自由度 n_i のウィシャート分布に従うことから，確率密度関数は次の式で与えられる (Muirhead, 1982).

$$f(n_iS_i) = \frac{1}{2^{\frac{pn_i}{2}}|\Sigma_i|^{\frac{n_i}{2}}\Gamma_p\left(\dfrac{n_i}{2}\right)}|n_iS_i|^{\frac{n_i-p-1}{2}}\exp\left(-\frac{n_i}{2}\operatorname{tr}(\Sigma_i^{-1}S_i)\right).$$

この確率密度関数を用いると，$\Sigma_1, \Sigma_2, \ldots, \Sigma_k$ に関する尤度関数は

$$\begin{aligned} L(\Sigma_1, \Sigma_2, \ldots, \Sigma_k) &= \prod_{i=1}^{k}\frac{1}{2^{\frac{pn_i}{2}}|\Sigma_i|^{\frac{n_i}{2}}\Gamma_p\left(\dfrac{n_i}{2}\right)}|n_iS_i|^{\frac{n_i-p-1}{2}}\exp\left(-\frac{n_i}{2}\operatorname{tr}(\Sigma_i^{-1}S_i)\right) \\ &= C \times \prod_{i=1}^{k}\exp\left[\operatorname{tr}\left(-\frac{n_i}{2}\Sigma_i^{-1}S_i\right)\right]|\Sigma_i|^{-\frac{n_i}{2}} \end{aligned} \tag{2.2}$$

となる．ただし，C は Σ_i に依存しない定数である．式 (2.2) を最大化するかわりに，次の式を最小化して，$\Sigma_1, \Sigma_2, \ldots, \Sigma_k$ の最尤推定量を求める．

$$\begin{aligned} g(\Sigma_1, \Sigma_2, \ldots, \Sigma_k) &= -2\log L(\Sigma_1, \Sigma_2, \ldots, \Sigma_k) + 2\log C \\ &= \sum_{i=1}^{k} n_i\left[\log|\Sigma_i| + \operatorname{tr}(\Sigma_i^{-1}S_i)\right]. \end{aligned} \tag{2.3}$$

ここで，共通主成分分析を行うために共通主成分仮説 (2.1) を仮定する．この仮説が成立するという条件下での最尤推定量は

$$\boldsymbol{w}_l^T\left(\sum_{i=1}^{k} n_i\frac{\lambda_{il} - \lambda_{ij}}{\lambda_{il}\lambda_{ij}}S_i\right)\boldsymbol{w}_j = 0 \quad (l, j = 1, 2, \ldots, p \ \ (l \neq j)) \tag{2.4}$$

という $\dfrac{p(p-1)}{2}$ 個の方程式の解となる．ただし，$\boldsymbol{w}_l$ は直交行列 W の第 l 列ベクトル，λ_{il} は対角行列 Λ_i の第 l 対角要素である．式 (2.4) の方程式の解として求められた最尤推定量を

$$\begin{aligned} &\hat{W} = (\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_p), \\ &\hat{\Lambda}_i = \operatorname{diag}(\hat{\lambda}_{ij}) = \begin{pmatrix} \hat{\lambda}_{i1} & 0 & \cdots & 0 \\ 0 & \hat{\lambda}_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\lambda}_{ip} \end{pmatrix} \end{aligned}$$

とおく．$\hat{W}$ は共通主成分係数，$\hat{\Lambda}_i$ は各グループの共通主成分ごとの分散の推定量である．これより，共通主成分仮説における Σ_i の最尤推定量は，次の式で表せる．

$$\hat{\Sigma}_i = \hat{W}\hat{\Lambda}_i\hat{W}^T. \tag{2.5}$$

3 CPCA deviance

年度データにおける構造変化地点の特定を共分散構造に着目した共通主成分モデルの下での逸脱度に基づいて行う方法について議論する (松川他 (2019)，Matsukawa *et al.* (2019))．社会構造や経済構造は 2011 年の東日

本大震災といった大規模自然災害や 2007 年の世界金融危機といった経済危機が起こった際に変化が起こると思われる．上記と同様の出来事が発生した時に備えるには，過去に発生した出来事による影響を観察し，将来どんなことが起こるか考察することが大切であり，そのためには，変化点を特定することが重要となる．本研究では変化点前後での年度データにおいてグループごとのデータの散らばり具合，すなわち分散共分散行列が大きく異なると仮定する．その考えに基づき，データの構造を分散共分散行列でとらえ，共通主成分分析によって変化点の特定を行う．

そのための尺度として逸脱度を導入する．逸脱度とは，最大対数尤度に -2 を乗じた式で定義され，モデルの当てはまりの程度を表す指標である．構造変化点を特定するための過程として，まず複数のグループに対して 2 つに分割し，分割した各グループに対してそれぞれ共通主成分分析を行い，分散共分散行列の最尤推定量を求める．その最尤推定量を逸脱度に代入し，共通主成分モデルに基づく逸脱度を求める．一方で，不偏標本分散共分散行列を最尤推定量とするフルモデル (共通主成分仮説を仮定しないモデル) に基づく逸脱度を求める．そして，共通主成分モデルに基づく逸脱度からフルモデルに基づく逸脱度を引いた数値 (CPCA deviance) を用いて各分割パターンで比較していく．CPCA deviance が最小となるパターンを見つけた場合，それはモデルの当てはまりが最も良いものであると言えることから，正しく 2 つの分散共分散行列に分けることができたとみなせる．従って，そのパターンの境界を共分散構造変化点とみなすことにする．詳しいステップは以下のとおりである．

Step 1　k グループを k_1個 と k_2 個 $(1 \leq k_1 < k,\ k_2 = k - k_1)$ の 2 つに分割する．この時，k_1 個のグループを G_1 とおき，k_2 個のグループを G_2 とおく．

Step 2　G_1 における分散共分散行列 $\Sigma_1, \ldots, \Sigma_{k_1}$ と，G_2 における分散共分散行列 $\Sigma_{k_1+1}, \ldots, \Sigma_{k_1+k_2}$ に対して，それぞれ共通主成分分析を適用し，最尤推定量 $\hat{\Sigma}_1^{(1)}, \ldots, \hat{\Sigma}_{k_1}^{(1)}, \hat{\Sigma}_{k_1+1}^{(2)}, \ldots, \hat{\Sigma}_{k_1+k_2}^{(2)}$ を求める．($k_1 = 1$ または $k_2 = 1$ のとき，G_1 または G_2 のグループ数が 1 になり，共通主成分分析を用いて推定できないため，不偏標本分散共分散行列 S_1 または $S_{k_1+k_2}$ を最尤推定量として考えることにする．)

Step 3　求めた最尤推定量 $\hat{\Sigma}_1^{(1)}, \ldots, \hat{\Sigma}_{k_1}^{(1)}, \hat{\Sigma}_{k_1+1}^{(2)}, \ldots, \hat{\Sigma}_{k_1+k_2}^{(2)}$ を逸脱度に代入し，共通主成分モデルに基づく逸脱度を求める．

Step 4　不偏標本分散共分散行列 $S_1, \ldots, S_{k_1}, S_{k_1+1}, \ldots, S_{k_1+k_2}$ を用いてフルモデルに基づく逸脱度を導出する．そして Step3 で求めた逸脱度からフルモデルに基づく逸脱度を引いた数値 (CPCA deviance) を求める．

Step 5　$k_1 = 1, 2, \ldots, k-1$ の $(k-1)$ パターンで Step1 から Step4 を行い，CPCA deviance が最小となるパターンの境界を構造変化点として考える．

共通主成分仮説に従うモデルの最尤推定量を表す式 (2.5) を用いることで共通主成分モデルに基づく逸脱度は次の式で与えられる．

$$\begin{aligned} &-2\log L(\hat{\Sigma}_1^{(1)}, \ldots, \hat{\Sigma}_{k_1}^{(1)}, \hat{\Sigma}_{k_1+1}^{(2)}, \ldots, \hat{\Sigma}_{k_1+k_2}^{(2)}) \\ &\qquad = -2\log C + \sum_{i=1}^{k_1+k_2} n_i p + \sum_{i=1}^{k_1} n_i \log|\hat{\Sigma}_i^{(1)}| + \sum_{i=k_1+1}^{k_1+k_2} n_i \log|\hat{\Sigma}_i^{(2)}|. \end{aligned} \tag{3.1}$$

一方でフルモデルを正規分布に従うモデルとすると，その下での最尤推定量 S_i を表す逸脱度は次の式で与えられる．

$$-2\log L(S_1, \ldots, S_{k_1+k_2}) = -2\log C + \sum_{i=1}^{k_1+k_2} n_i p + \sum_{i=1}^{k_1+k_2} \log|S_i|. \tag{3.2}$$

共通主成分仮説に基づくモデルの逸脱度 (3.1) とフルモデル下での逸脱度 (3.2) の差は

$$-2\log L(\hat{\Sigma}_1^{(1)},\ldots,\hat{\Sigma}_{k_1}^{(1)},\hat{\Sigma}_{k_1+1}^{(2)},\ldots,\hat{\Sigma}_{k_1+k_2}^{(2)}) - (-2\log L(S_1,\ldots,S_{k_1+k_2})) \\ = \sum_{i=1}^{k_1} n_i \log\frac{|\hat{\Sigma}_i^{(1)}|}{|S_i|} + \sum_{i=k_1+1}^{k_1+k_2} n_i \log\frac{|\hat{\Sigma}_i^{(2)}|}{|S_i|} \tag{3.3}$$

で与えられる．これより，仮説下とフルモデルの逸脱度の差を表す式 (3.3) (CPCA deviance) が最小となるものが最適なグループの区分であるとみなす．

修士論文では提案した共分散構造変化点探索法を 2001 年から 2016 年の 16 年分の 47 都道府県別エネルギー消費データへの分析およびそのデータを参考にしたモンテカルロシミュレーションに適用して，その有用性を検証した．

4 終わりに

本研究では，主成分分析の一般形である共通主成分分析の基本的な考え方および仮説における同時対角化するための直交行列 W を推定する方法と関連する推測論の定式化を行った．そして，年度データにおける変化点を客観的に探索するために共分散構造に着目した共通主成分モデルの下での逸脱度 (CPCA deviance) に基づく変化点探索法を提案した．

今後取り組むべき課題としては，共通主成分分析では多変量観測データをグループごとに多変量正規分布に従うと仮定して行なっており，いくつかのグループに外れ値がある場合では，推測の精度が落ちるという問題点があげられる．これに関して外れ値を考慮した推測法の修正を行えば，共通主成分分析の理論に当てはまりやすくなるだけでなく，共分散構造の変化を検出しやすくなると考えている．このことから，CPCA deviance に M 推定，最小共分散行列推定法などのロバスト推定を導入した Robust CPCA deviance の導出を検討したい．また，Chen and Gupta (2012) の考えを参考にし，ベイズ型モデル評価基準 BIC を用いて複数個の変化点を探索し，変数選択することで変化点の要因を特定する方法を検討したい．

参考文献

[1] Chen, J. and Gupta, A.K. (2012): *Parametric Statistical Change Point Analysis*. Birkhaüser, New York.

[2] Flury, B.N. (1984): Common principal components in K groups. *Journal of the American Statistical Association.* **79**, 892–898.

[3] Flury, B.N. (1988): *Common Principal Components and Related Multivariate Models*. John Wiley & Sons, New York.

[4] 勝浦正樹 (1988): 共通主成分分析による景気動向指数採用系列の分析. 経済学研究年報 (早大). **28**, 47–62.

[5] 松川達也, 三角俊裕, 小西貞則, 前園宜彦 (2019): 共通主成分分析に基づく構造変化点探索とその応用. 第 24 回情報・統計科学シンポジウム講演.

[6] Matsukawa, T., Misumi, T., Maesono, Y. and Konishi, S. (2019): Structural change detection via common principal component analysis. The 12th International Conference of the ERCIM WG on Computational and Methodological Statistics.

[7] Muirhead, R.J. (1982): *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, New York.