Asymptotic Theory for Common Principla Component Analysis

Ganchao Wei

October 13, 2021

Overview

1. Introduction
2. Asymptotic distribution of MLE
3. test for q hypothetical eigenvectors
4. inference for eigenvalues
5. Applications

Introduction

Common principle component analysis (CPCA): the $p \times p$ covariance matrices of k populations can be diagonalized by the same orthogonal transformation.

The hypothesis of common principal components (CPC's).

$\boldsymbol{H}_C : \boldsymbol{\beta}'\boldsymbol{\Sigma}_i\boldsymbol{\beta} = \boldsymbol{\Lambda}_i$, for $i = 1, \ldots, k$

We can further arrange $\boldsymbol{\beta}$ according to the first group, i.e.

$\boldsymbol{\beta}_1'\boldsymbol{\Sigma}_1\boldsymbol{\beta}_1 > \boldsymbol{\beta}_2'\boldsymbol{\Sigma}_1\boldsymbol{\beta}_2 > \ldots > \boldsymbol{\beta}_p'\boldsymbol{\Sigma}_1\boldsymbol{\beta}_p$

Notations & Assumptions:

- $\boldsymbol{\Sigma}_i$ are p.d.s.
- $\boldsymbol{\Lambda}_i = diag(\lambda_{i1}, \ldots, \lambda_{ip})$
- $n_i\boldsymbol{S}_i \sim W_p(n_i, \Sigma_i)$
- The ML estimates are $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_p)$ and $\hat{\boldsymbol{\Lambda}}_i = diag(\hat{\lambda}_{i1}, \ldots, \hat{\lambda}_{ip})$

Asymptotic distribution of MLE

The log-likelihood function of the k samples, up to an additive constant:

$$g(\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_k, \boldsymbol{\beta} | \mathbf{S}_1, \ldots, \mathbf{S}_k) = -\tfrac{1}{2} \sum_{i=1}^{k} n_i \left[\sum_{j=1}^{p} \left(\log \lambda_{ij} + \boldsymbol{\beta}_j' \mathbf{S}_i \boldsymbol{\beta}_j / \lambda_{ij} \right) \right]$$

Let $\lambda'_{(i)} = (\lambda_{i1}, \ldots, \lambda_{ip})$, $n = n_1 + \ldots + n_k$ and $r_i = n_i/n$. Then the information matrix is:

$$\begin{array}{c|cccc|c}
 & \boldsymbol{\lambda}'_{(1)} & \boldsymbol{\lambda}'_{(2)} & \cdots & \boldsymbol{\lambda}'_{(k)} & \boldsymbol{\beta}^{*\prime} \\
\hline
\boldsymbol{\lambda}_{(1)} & \frac{1}{2} n r_1 \boldsymbol{\Lambda}_1^{-2} & \mathbf{0} & \cdots & \mathbf{0} & \\
\boldsymbol{\lambda}_{(2)} & \mathbf{0} & \frac{1}{2} n r_2 \boldsymbol{\Lambda}_2^{-2} & \cdots & \mathbf{0} & \\
\vdots & \vdots & \vdots & & \vdots & \mathbf{G}' \\
\boldsymbol{\lambda}_{(k)} & \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{2} n r_k \boldsymbol{\Lambda}_k^{-2} & \\
\hline
\boldsymbol{\beta}^* & & \mathbf{G} & & & \mathbf{A}
\end{array}$$

, where $\boldsymbol{G}$ and $\boldsymbol{A}$ are not yet determined.

Asymptotic distribution of MLE (eigenvalues)

Use the asymptotic normality of $n_i \boldsymbol{S}_i$ to get the asymptotic univariate distribution of $\hat{\lambda}_{ij}$ as

$$\sqrt{n_i}(\hat{\lambda}_{ij} - \lambda_{ij}) \sim N(0, 2\lambda_{ij}^2)$$

From the Fisher information, the joint asymptotic distribution of $(\hat{\lambda}'_{(1)}, \ldots, \hat{\lambda}'_{(k)})'$ has covariance matrix

$$\frac{1}{n}\mathbf{V}_\lambda = \left[\begin{pmatrix} \frac{1}{2}nr_1\boldsymbol{\Lambda}_1^{-2} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{1}{2}nr_k\boldsymbol{\Lambda}_k^{-2} \end{pmatrix} - \mathbf{G}'\mathbf{A}^{-1}\mathbf{G}\right]^{-1}$$

By comparison, we can see that $\boldsymbol{G} = 0$. Therefore,

THEOREM 1. *The statistics* $\sqrt{n_i}(\hat{\lambda}_{ij} - \lambda_{ij})$ *are asymptotically* ($\min_{1 \le i \le k} n_i \to \infty$) *distributed as* $N(0, 2\lambda_{ij}^2)$, *independent of each other and independent of the* $\hat{\boldsymbol{\beta}}_j$.

Let $\boldsymbol{V}_i$ be the asymptotic covariance matrix of $\sqrt{n_i}vec(\hat{\beta}) = \sqrt{n}(\hat{\beta}'_1, \ldots, \hat{\beta}'_k)'$. Further, define:

$$g_{jh}^{(i)} = \frac{1}{r_i} \frac{\lambda_{ij}\lambda_{ih}}{(\lambda_{ij} - \lambda_{ih})^2} \qquad (h \neq j)$$

Then, we get

$$\begin{array}{c|cccc}
 & \hat{\boldsymbol{\beta}}'_1 & \hat{\boldsymbol{\beta}}'_2 & & \hat{\boldsymbol{\beta}}'_p \\
\hline
\hat{\boldsymbol{\beta}}_1 & \sum_{\substack{h=1 \\ h \neq 1}}^{p} g_{1h}^{(i)} \boldsymbol{\beta}_h \boldsymbol{\beta}'_h & -g_{12}^{(i)} \boldsymbol{\beta}_2 \boldsymbol{\beta}'_1 & \cdots & -g_{1p}^{(i)} \boldsymbol{\beta}_p \boldsymbol{\beta}'_1 \\
\hat{\boldsymbol{\beta}}_2 & -g_{21}^{(i)} \boldsymbol{\beta}_1 \boldsymbol{\beta}'_2 & \sum_{\substack{h=1 \\ h \neq 2}}^{p} g_{2h}^{(i)} \boldsymbol{\beta}_h \boldsymbol{\beta}'_h & \cdots & -g_{2p}^{(i)} \boldsymbol{\beta}_p \boldsymbol{\beta}'_2 \\
\vdots & \vdots & \vdots & & \vdots \\
\hat{\boldsymbol{\beta}}_p & -g_{p1}^{(i)} \boldsymbol{\beta}_1 \boldsymbol{\beta}'_p & -g_{p2}^{(i)} \boldsymbol{\beta}_2 \boldsymbol{\beta}'_p & \cdots & \sum_{\substack{h=1 \\ h \neq p}}^{p} g_{ph}^{(i)} \boldsymbol{\beta}_h \boldsymbol{\beta}'_h
\end{array} = \mathbf{V}_i$$

Asymptotic distribution of MLE (eigenvectors)

Since $\boldsymbol{V}_i$ are simultaneously diagonalizable, there exists an orthogonal matrix $\boldsymbol{H} = (\boldsymbol{H}_1, \boldsymbol{H}_2)$, where $\boldsymbol{H}_1 = (\boldsymbol{h}_1, \ldots, \boldsymbol{h}_s), s = p(p-1)/2$ s.t.

$$\mathbf{H}'\mathbf{V}_i\mathbf{H} = \begin{pmatrix} \mathbf{E}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \qquad (i = 1, \ldots, k),$$

Then we can get the information matrix for the transformed variable $\boldsymbol{u} = \boldsymbol{H}_1' vec\hat{\boldsymbol{\beta}}$ as

$$\begin{aligned} \mathbf{A}^* = \sum_{i=1}^{k} \mathbf{A}_i^* &= n \operatorname{diag}\left(\sum_{i=1}^{k} e_{i1}^{-1}, \ldots, \sum_{i=1}^{k} e_{is}^{-1} \right) \\ &= n \operatorname{diag}\left(e_1^{-1}, \ldots, e_s^{-1} \right), \end{aligned}$$

Asymptotic distribution of MLE (eigenvectors)

Then transform back and write out $\boldsymbol{H}_1$ explicitly, we can get

THEOREM 2. *The asymptotic distribution of* $\sqrt{n}\,\mathrm{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ *is normal with mean* $\mathbf{0}$ *and covariance matrix* $\mathbf{V}$ *given by*

$$
(2.13)\qquad
\begin{array}{c|cccc}
 & \boldsymbol{\beta}_1' & \boldsymbol{\beta}_2' & \cdots & \boldsymbol{\beta}_p' \\ \hline
\hat{\boldsymbol{\beta}}_1 & \sum\limits_{\substack{h=1\\ h\neq 1}}^{p} g_{1h}\boldsymbol{\beta}_h\boldsymbol{\beta}_h' & -g_{12}\boldsymbol{\beta}_2\boldsymbol{\beta}_1' & \cdots & -g_{1p}\boldsymbol{\beta}_p\boldsymbol{\beta}_1' \\
\hat{\boldsymbol{\beta}}_2 & -g_{21}\boldsymbol{\beta}_1\boldsymbol{\beta}_2' & \sum\limits_{\substack{h=1\\ h=2}}^{p} g_{2h}\boldsymbol{\beta}_h\boldsymbol{\beta}_h' & \cdots & -g_{2p}\boldsymbol{\beta}_p\boldsymbol{\beta}_2' \\
\vdots & \vdots & \vdots & & \vdots \\
\hat{\boldsymbol{\beta}}_p & -g_{p1}\boldsymbol{\beta}_1\boldsymbol{\beta}_p' & -g_{p2}\boldsymbol{\beta}_2\boldsymbol{\beta}_p' & \cdots & \sum\limits_{\substack{h=1\\ h\neq p}}^{p} g_{ph}\boldsymbol{\beta}_h\boldsymbol{\beta}_h'
\end{array}
= \mathbf{V},
$$

where the g_{jl} *are defined in* (2.5), *and the* $\boldsymbol{\beta}_j$ *are the* (*common*) *eigenvectors of the* k *matrices* $\boldsymbol{\Sigma}_i$.

An asymptotic test for q hypothetical eigenvectors

The null hypothesis is $H_q : (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_q) = (\boldsymbol{\beta}_1^0, \dots, \boldsymbol{\beta}_q^0)$, which is based on the submatrix of asymptotic covariance $\boldsymbol{V}$. Denote the submatrix as $\boldsymbol{V}(q)$. $\boldsymbol{V}(q)$ has following eigenstructure:

THEOREM 3. *The upper left $pq \times pq$ submatrix of* $\mathbf{V}$ *has the following eigenvectors and eigenvalues*:

1. $\binom{q}{2}$ *eigenvectors* (*one for each pair j, l with $1 \le j < l \le q$*) *have* $\boldsymbol{\beta}_l/\sqrt{2}$ *in position j and* $-\boldsymbol{\beta}_j/\sqrt{2}$ *in position l. All other positions are zero, and the associated roots are* $2g_{jl}$.
2. $(p-q)q$ *eigenvectors* (*one for each combination of indices j, l such that $1 \le j \le q < l \le p$*) *have* $\boldsymbol{\beta}_l$ *in position j and* $\mathbf{0}$ *in all other positions; the associated roots are* g_{jl}.
3. $\binom{q}{2}$ *eigenvectors* (*one for each pair of indices j, l such that $1 \le j \le l \le q$*) *can be chosen to have* $\boldsymbol{\beta}_l/\sqrt{2}$ *in position j,* $\boldsymbol{\beta}_j/\sqrt{2}$ *in position l, and zeros in all other positions. The associated roots are zero.*
4. *q eigenvectors* (*one for each j with $1 \le j \le q$*) *can be chosen to have* $\boldsymbol{\beta}_j$ *in position j and zeros elsewhere. The associated roots are zero.*

An asymptotic test for q hypothetical eigenvectors

Let $\mathbf{\Phi}$ be a diagonal matrix with diagonal elements equal to the nonzero roots of $\boldsymbol{V}(q)$, i.e. $\mathbf{\Phi} = diag(2g_{12}, \ldots, 2g_{q-1,1}, g_{1,q+1}, \ldots, g_{qp})$. Also let the columns of matrix Γ be given by the characteristic vectors associated with the nonzero roots. Then,

$$\mathbf{z}'\mathbf{z} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \vdots \\ \hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q \end{pmatrix}' \mathbf{\Gamma}\mathbf{\Phi}^{-1}\mathbf{\Gamma}' \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \vdots \\ \hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q \end{pmatrix}$$

$$= \frac{1}{4} \sum_{j=1}^{q-1} \sum_{l=j+1}^{q} g_{jl}^{-1} \left(\boldsymbol{\beta}_l' \hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j' \hat{\boldsymbol{\beta}}_l \right)^2$$

$$+ \sum_{j=1}^{q} \sum_{l=q+1}^{p} g_{jl}^{-1} \left(\boldsymbol{\beta}_l' \hat{\boldsymbol{\beta}}_j \right)^2.$$

follows $\chi^2(t)$ distribution.

An asymptotic test for q hypothetical eigenvectors

This leads to the distribution of the test statistics for $H_q : (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q) = (\boldsymbol{\beta}_1^0, \ldots, \boldsymbol{\beta}_q^0)$:

THEOREM 4. *Under H_q as defined in* (3.1), *the statistic*

$$X^2(H_q) = n\left[\frac{1}{4}\sum_{j=1}^{q-1}\sum_{l=j+1}^{q}\hat{g}_{jl}^{-1}\left(\hat{\boldsymbol{\beta}}_l'\boldsymbol{\beta}_j^0 - \hat{\boldsymbol{\beta}}_j'\boldsymbol{\beta}_l^0\right)^2 + \sum_{j=1}^{q}\sum_{l=q+1}^{p}\hat{g}_{jl}^{-1}\left(\hat{\boldsymbol{\beta}}_l'\boldsymbol{\beta}_j^0\right)^2\right] \tag{3.3}$$

is asymptotically distributed as chi square with $q(p-(q+1)/2)$ *degrees of freedom.*

The paper also illustrated 2 special cases: (1) $q = 1$ and (2) $q = p$.

Asymptotic inference for eigenvalues

Similar to regular PCA, we may need to discard CPC's with relatively small variances.

Let $c_i = \sum_{j=1}^{q} \lambda_{ij}$, $d_i = tr\Sigma_i - c_i$, $f_i = d_i / tr\Sigma_i$ (relative contribution to trace for the last $p-q$ CPC's) and f_0 be the pre-specified fraction. Then,

$$z_i = \frac{\sqrt{n_i}\left[(1-f_0)\hat{d}_i - f_0 \hat{c}_i\right]}{\left(2\left[f_0^2 \Sigma_{j=1}^{q} \hat{\lambda}_{ij}^2 + (1-f_0)^2 \Sigma_{j=q+1}^{p} \hat{\lambda}_{ij}^2\right]\right)^{1/2}} \sim N(0,1)$$

when $f_i = f_0$ (null hypothesis).

So for testing the H_1: all f_i are less than or equal to f_0, we can just reject the hypothesis if

$$\max_{1 \le i \le k} z_i > z_\beta \quad \text{with } \beta = 1 - (1-\alpha)^{1/k},$$

LRT for sphericity of p-q CPC's

In PCA, the motivation for testing equality of p-q characteristic roots stems from the model $\Sigma = \Psi + \sigma^2 \boldsymbol{I}_p$, where Ψ is p.s.d. of rank q.
In CPCA, we can consider similar model for each group, i.e.
$\Sigma_i = \Psi_i + \sigma_i^2 \boldsymbol{I}_p$, with Ψ_i be simultaneously diagonalizable and of rank q.
This is equivalent as the following test (**hypothesis of partial sphericity**):

$$H_S : \lambda_{i,q+1} = \ldots = \lambda_{ip}$$

Putting $H_S : \lambda_{i,q+1} = \ldots = \lambda_{ip} = \lambda_i^*$, we get

$$\begin{aligned} & -2g(\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_k, \boldsymbol{\beta} | \mathbf{S}_1, \ldots, \mathbf{S}_k) \\ &= \sum_{i=1}^{k} n_i \left[\sum_{j=1}^{q} \left(\log \lambda_{ij} + \boldsymbol{\beta}_j' \mathbf{S}_i \boldsymbol{\beta}_j / \lambda_{ij} \right) \right. \\ & \qquad \left. + (p-q) \log \lambda_i^* + \left(\sum_{j=q+1}^{p} \boldsymbol{\beta}_j' \mathbf{S}_i \boldsymbol{\beta}_j \right) \Big/ \lambda_i^* \right] \end{aligned}$$

LRT for sphericity of p-q CPC's

After some algebra, we can get the LRT test statistic:

$$X_S^2 = \sum_{i=1}^{k} n_i \log \frac{(\tilde{\lambda}_i^*)^{p-q} \prod_{j=1}^{q} \tilde{\lambda}_{ij}}{\prod_{j=1}^{p} \hat{\lambda}_{ij}}$$

The null distribution of X_S^2 is asymptotically $\chi^2((p-q-1)(p-q+2k)/2)$. The paper also provided the approximated statistic:

$$X_S^2(\text{approx}) = \sum_{i=1}^{k} n_i \log \frac{(\hat{\lambda}_i^*)^{p-q}}{\prod_{j=q+1}^{p} \hat{\lambda}_{ij}}$$

But we need to be careful: since $X_S^2(approx) \geq X_S^2$, the approximate statistic can be used to accept H_S, but not necessarily to reject it.

Applications

Data: 2 groups (24 males and 24 females), with 3 features.
We do log-transformation of data because of their relationship to allometry. The data and MLE's are shown in the table:

TABLE 1
Common principal component analysis of turtle carapace dimensions, transformed logarithmically.

(a) Sample covariance matrices[a]

males ($n_1 = 23$)

$$\mathbf{S}_1 = \begin{pmatrix} 1.1072 & 0.8019 & 0.8160 \\ 0.8019 & 0.6417 & 0.6005 \\ 0.8160 & 0.6005 & 0.6773 \end{pmatrix}$$

females ($n_2 = 23$)

$$\mathbf{S}_2 = \begin{pmatrix} 2.6391 & 2.0124 & 2.5443 \\ 2.0124 & 1.6190 & 1.9782 \\ 2.5443 & 1.9782 & 2.5899 \end{pmatrix}$$

(b) Variances of CPC's and eigenvalues of $\mathbf{S}_i$

males	$\hat{\lambda}_{1j}$	2.3148	0.0729	0.0385
	eigenvalues	2.3303	0.0599	0.0360
females	$\hat{\lambda}_{2j}$	6.7135	0.0807	0.0538
	eigenvalues	6.7200	0.0751	0.0530

(c) Coefficients of CPC's[b]

$$\hat{\boldsymbol{\beta}}_1 = \begin{pmatrix} 0.6406 \\ 0.4905 \\ 0.5907 \end{pmatrix} \begin{matrix} (0.013) \\ (0.015) \\ (0.016) \end{matrix} \qquad \hat{\boldsymbol{\beta}}_2 = \begin{pmatrix} -0.3839 \\ -0.4617 \\ -0.7997 \end{pmatrix} \begin{matrix} (0.182) \\ (0.201) \\ (0.032) \end{matrix} \qquad \hat{\boldsymbol{\beta}}_3 = \begin{pmatrix} -0.6650 \\ 0.7391 \\ 0.1075 \end{pmatrix} \begin{matrix} (0.105) \\ (0.126) \\ (0.218) \end{matrix}$$

[a]Multiplied by 10^2

Applications

Test 1: if allometric growth is true, then the first PC of log-data should be $\beta_1' = (1, \ldots, 1)/\sqrt{p}$. Therefore, the test is
$H_0 : \beta_1 = \beta_1^0 = (1, \ldots, 1)'/\sqrt{3}$ The test statistic $X^2(H_1) = 46.17$, which follows $\chi^2(2)$ under null. $\Rightarrow$ reject the null.
Test 2: test $H_S : \lambda_{i2} = \lambda_{i3}$ for $i = 1, 2$ (simultaneous sphericity of the second and third CPC's). The resulting statistic is $X_S^2(approx) = 3.24$, which (approximately) follows $\chi^2(3)$ under the null. $\Rightarrow$ fail to reject.
Taking into consideration the relative smallness of these 2 roots in both groups $\Rightarrow$

- the 3 shell dimensions are distributed about a single principal axis ("size") and 2 minor axes.
- the main axis having the same orientation in space for both groups.