Principal Component Models for Sparse Functional Data

Hyunsung Kim

August 9, 2019

Department of Statistics
Chung-Ang University

Outline

1. Reduced rank model

2. EM Algorithm

3. Applications

Reduced rank model

Reduced rank model

$$\mathbf{Y}_i = \mathbf{B}_i \boldsymbol{\theta}_\mu + \mathbf{B}_i \boldsymbol{\Theta} \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i$$

where $\boldsymbol{\theta}_\mu$ is $q \times 1$ vector, $\boldsymbol{\Theta}$ is $q \times k$ matrix, $\boldsymbol{\alpha}_i$ is $k \times 1$ vector,

$$\boldsymbol{\epsilon}_i \sim (\mathbf{0}, \sigma^2 \mathbf{I}),\ \boldsymbol{\alpha}_i \sim (\mathbf{0}, \mathbf{D})$$

subject to

$$\boldsymbol{\Theta}^T \boldsymbol{\Theta} = \mathbf{I},\ \int \mathbf{b}(t)^T \mathbf{b}(t) dt = 1,\ \int \int \mathbf{b}(t)^T \mathbf{b}(s) dt = 0$$

Reduced rank model

How to find functional PCs

- mean function
  $\mu(t) = \mathbf{b}(t)^T \boldsymbol{\theta}_\mu$
- PC function
  $f(t)^T = \mathbf{b}(t)^T \boldsymbol{\Theta}$

where $\mathbf{b}(t)$ is the orthonormal spline basis and it can be computed by Gram–Schmidt orthonormalization.(Zhou et al.(2008))

To find the PC curves, We should estimate $\boldsymbol{\theta}_\mu, \boldsymbol{\Theta}$.

EM Algorithm

EM Algorithm

Maximum likelihood (or Penalized least squares)

In the reduced rank model, we should minimized $L(\boldsymbol{\theta}_\mu, \boldsymbol{\Theta}, \mathbf{D}, \sigma^2)$,

$$\sum_{i=1}^{N} \left\{ (\mathbf{Y}_i - \mathbf{B}_i\boldsymbol{\theta}_\mu - \mathbf{B}_i\boldsymbol{\Theta}\boldsymbol{\alpha}_i)^T (\mathbf{Y}_i - \mathbf{B}_i\boldsymbol{\theta}_\mu - \mathbf{B}_i\boldsymbol{\Theta}\boldsymbol{\alpha}_i) + \sigma^2 \boldsymbol{\alpha}_i^T \mathbf{D}^{-1} \boldsymbol{\alpha}_i \right\}$$

We can minimize this equation using the `EM` algorithm.

EM Algorithm

Complete data

Let $\boldsymbol{\alpha}_i$ is the latent variable(unobserved), then we can define the complete data $\mathbf{Z} = (\mathbf{Y}, \boldsymbol{\alpha})$ and employ the `EM` algorithm.

Let $\boldsymbol{\Omega} = (\boldsymbol{\theta}_\mu, \boldsymbol{\Theta})$ and $L(\boldsymbol{\Omega}|\mathbf{Z}) = -L(\boldsymbol{\theta}_\mu, \boldsymbol{\Theta}, \mathbf{D}, \sigma^2)$, then the minimization problem will be equivalent to maximize $L(\boldsymbol{\Omega}|\mathbf{Z})$.

EM Algorithm

E-step

Compute the expectation of the objective function($L(\mathbf{\Omega}|\mathbf{Z})$) for complete data $\mathbf{Z}$,

$$Q(\mathbf{\Omega}|\mathbf{\Omega}^{(t)}) = E\left\{L(\mathbf{\Omega}|\mathbf{Z})|\mathbf{Y}, \mathbf{\Omega}^{(t)}\right\}$$

where $\mathbf{\Omega}^{(t)} = (\boldsymbol{\theta}_{\mu}^{(t)}, \mathbf{\Theta}^{(t)})$.
Also we can predict $\boldsymbol{\alpha}_i$ on the E-step.

EM Algorithm

M-step

$$\boldsymbol{\Omega}^{(t+1)} = \arg\max_{\boldsymbol{\Omega}} Q(\boldsymbol{\Omega}|\boldsymbol{\Omega}^{(t)})$$

Also we can estimate $\mathbf{D}, \sigma^2$ on the M-step.

With many `EM` iterations, it is known to converge to true parameters.

EM Algorithm

Result: The procedure to fit the reduced rank model

initialization $(\widehat{\boldsymbol{\theta}}_{\mu}^{(1)}, \widehat{\boldsymbol{\Theta}}^{(1)}, \widehat{\boldsymbol{\alpha}}_{i}^{(1)}, \widehat{\mathbf{D}}^{(1)}, \widehat{\sigma^2}^{(1)})$;

for $i = 2$ *to iterations* **do**

(M-step)

$$
\begin{aligned}
\widehat{\sigma^2}^{(t+1)} &\leftarrow (\widehat{\boldsymbol{\theta}}_{\mu}^{(t)}, \widehat{\boldsymbol{\Theta}}^{(t)}, \widehat{\boldsymbol{\alpha}}_{i}^{(t)}, \widehat{\mathbf{D}}^{(t)}, \widehat{\sigma^2}^{(t)}) \\
\widehat{\mathbf{D}}_{jj}^{(t+1)} &\leftarrow (\widehat{\boldsymbol{\Theta}}^{(t)}, \widehat{\boldsymbol{\alpha}}_{i}^{(t)}, \widehat{\mathbf{D}}^{(t)}, \widehat{\sigma^2}^{(t+1)}) \\
\widehat{\boldsymbol{\theta}}_{\mu}^{(t+1)} &\leftarrow (\widehat{\boldsymbol{\Theta}}^{(t)}, \widehat{\boldsymbol{\alpha}}_{i}^{(t)}) \\
\widehat{\boldsymbol{\Theta}}^{(t+1)} &\leftarrow (\widehat{\boldsymbol{\theta}}_{\mu}^{(t)}, \widehat{\boldsymbol{\Theta}}^{(t)}, \widehat{\boldsymbol{\alpha}}_{i}^{(t)}, \widehat{\mathbf{D}}^{(t+1)}, \widehat{\sigma^2}^{(t+1)})
\end{aligned}
$$

(E-step)

$$
\widehat{\boldsymbol{\alpha}}_{i}^{(t+1)} \leftarrow (\widehat{\boldsymbol{\Theta}}^{(t+1)}, \widehat{\boldsymbol{\alpha}}_{i}^{(t+1)}, \widehat{\mathbf{D}}^{(t+1)}, \widehat{\sigma^2}^{(t+1)})
$$

end

Applications

Bone Mineral Density data

- 48 white females
- 160 observations
- It was measured at the different time points and sparsely observed.

Applications

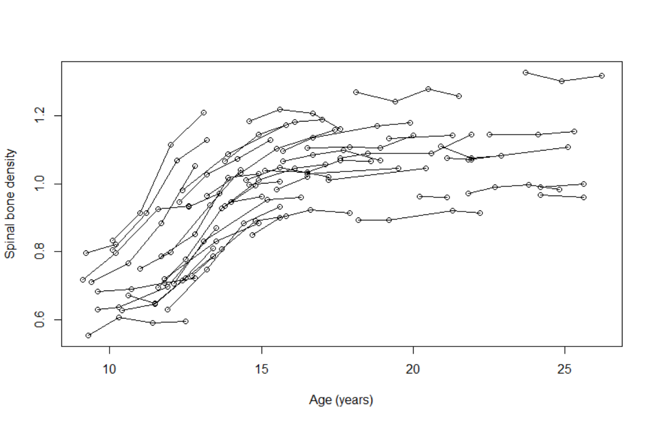


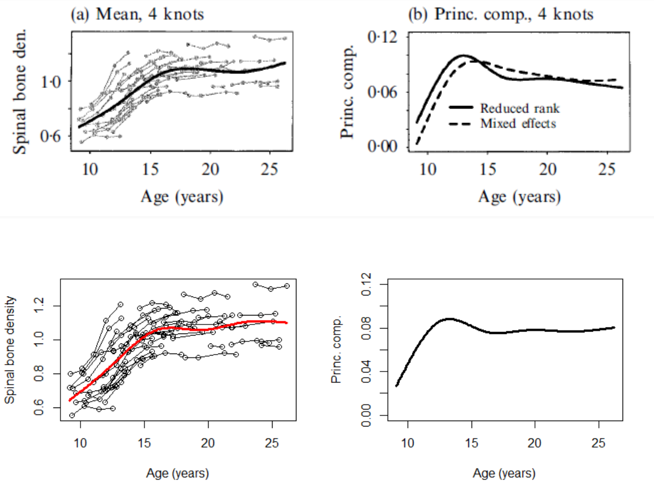
Figure 1: The bone mineral density of 48 females

Applications

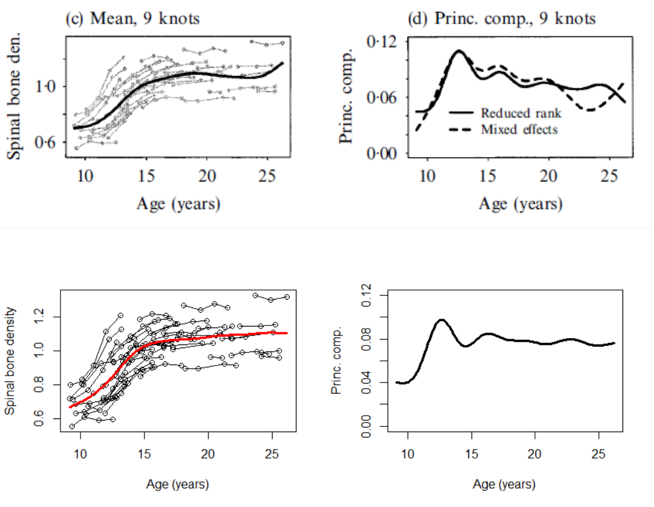
- Fit the reduced rank model using `EM` algorithm
- Initial values $= 0.1$ $(\boldsymbol{\theta}_\mu, \boldsymbol{\Theta}, \mathbf{D}, \sigma^2, \boldsymbol{\alpha}_i)$
- The number of PCs $= 2$
- 100 `EM` iterations

Applications

4 knots

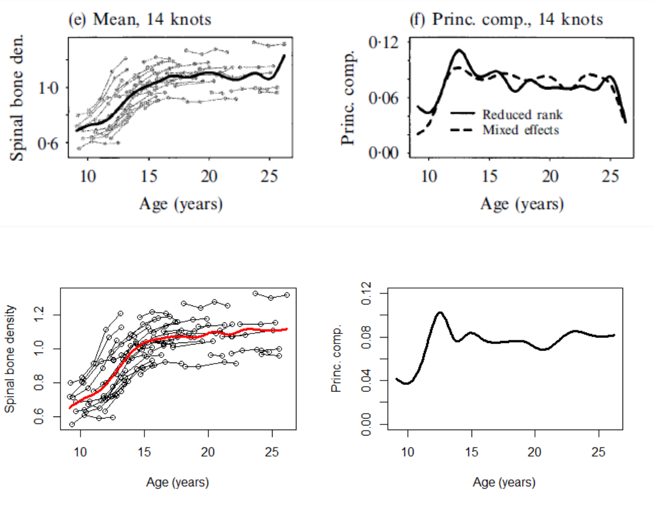


9 knots



(c) Mean, 9 knots
Spinal bone den.
1·0
0·6
10
15
20
25
Age (years)
(d) Princ. comp., 9 knots
Princ. comp.
0·12
0·06
0·00
Reduced rank
Mixed effects
Spinal bone density
0.6
0.8
1.0
1.2
Princ. comp.
0.00
0.04
0.08
0.12

14 knots


(e) Mean, 14 knots
Spinal bone den.
1·0
0·6
10
15
20
25
Age (years)
(f) Princ. comp., 14 knots
Princ. comp.
0·12
0·06
0·00
Reduced rank
Mixed effects
Spinal bone density
1.2
1.0
0.8
0.6
0.12
0.08
0.04
0.00

Loglikelihoods for the reduced rank fits

Number of knots	Paper	Coding
4	389.22	263.22
9	409.81	296.23
14	411.36	308.98

Reference

Reference

James G.M., Hastie T.J., Sugar C.A.
Principal component models for sparse functional data
Biometrika, 87(3):587–602, 2000.

J.O. Ramsay, B.W. Silverman.
Functional Data Analysis 2nd edition.
Springer, 2005.

Zhou L., Huang J.Z., Carroll R.J.
Joint modeling of paired sparse functional data using principal components
Biometrika, 95(3):601–619, 2008.