Lecture 10

Next week: No office hour Tuesday.
Office Hours on Friday: 10am–noon.

A2 on BB now.

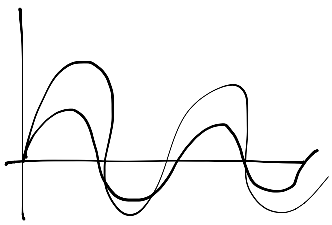
Andrew Curves (Assignment 2)

– simple multivariate visualization tool

Idea: Expression/represent each $\underset{\sim}{X_i} = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}$ as a function on $[0, 1]$

$$g_i(t) = \sum_{k=1}^{p} x_{ik} \phi_k(t)$$

$\uparrow$ some specified function.



Need to choose $\phi_1, \cdots, \phi_p$ so that $g_i(t) = g_j(t)$ for all $t \Rightarrow \underset{\sim}{X_i} = \underset{\sim}{X_j}$

$g_i(t)$ is close to $g_j(t)$ (e.g. $|g_i(t) - g_j(t)| < \delta$) then $\underset{\sim}{X_i}$ is close $\underset{\sim}{X_j}$.

Andrew curves: Take $\phi_1(t) = \frac{1}{\sqrt{2}}$ $(0 \le t \le 1)$

$\phi_2(t) = \sin(2\pi t)$

$\phi_3(t) = \cos(2\pi t)$

$\phi_4(t) = \sin(4\pi t)$

$\phi_5(t) = \cos(4\pi t)$

In general, $\phi_k(t) = \begin{cases} \sin(k\pi t) & \text{if } k \text{ is even} \\ \cos((k-1)\pi t) & \text{if } k \text{ is odd} \end{cases}$

– can also look at Andrew curves of PC scores etc.

Factor Analysis:

Idea: variable $X_1, \cdots, X_p$ ($\underset{\sim}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$) driven by $r < p$ latent (unobserved) factors (variables) $F_1, \cdots, F_r$

Data: $\underset{\sim}{X_1}, \cdots, \underset{\sim}{X_n}$ from some population with mean $\underset{\sim}{\mu}$, covariance C e.g. $N_p(\underset{\sim}{\mu}, C)$

Assumption. C has some special (lower dimensional) structure i.e. C is not an arbitrary non-negative definite matrix.

Examples: ① Graphical representation of dependency structure

$\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, C)$, $K = C^{-1}$ $\hookrightarrow$ Concentration matrix

$K = (K_{ij})$, $K_{ij} = 0 \Rightarrow$ no link between $X_i \neq X_j$

i.e. given remaining $p-2$ variables $\underset{\sim}{X_i}$ & $\underset{\sim}{X_j}$ are conditionally independent

② Low rank matrix approximation & PCA

$$X = \begin{pmatrix} \underset{\sim}{x_1}^T - \bar{\underset{\sim}{X}}^T \\ \vdots \\ \underset{\sim}{x_n}^T - \bar{\underset{\sim}{X}}^T \end{pmatrix} \xrightarrow{\text{SVD}} \underbrace{U}_{n\times p} \underbrace{D}_{p\times p} \underbrace{V^T}_{p\times p}$$

rank p

$$D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_p \end{pmatrix} \leftarrow \text{singular values}$$

$$X \doteq X^* = U D^* V^T \text{ where } D^* = \begin{pmatrix} d_1 & & & & 0 \\ & \ddots & & & \\ & & d_r & & \\ & & & \ddots & \\ 0 & & & & 0 \end{pmatrix}$$

$$\text{PCA}: \hat{C} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \underset{\sim}{\bar{X}})(\underset{\sim}{X_i} - \underset{\sim}{\bar{X}})^T = V \Lambda V^T \text{ where } \Lambda = \frac{D^2}{n-1}$$

Now approximate data using first r PCs where $\frac{\lambda_1 + \cdots + \lambda_r}{\lambda_1 + \cdots + \lambda_p}$ is close to 1

Use PCA to motivate Factor Analysis

Assume $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, C)$

Question: What happens if rank of C is less than p? $\text{rank}(C) = r < p$

Then, $C = V \Lambda V^T$ where $\Lambda = \begin{pmatrix} \lambda_1 & & & & 0 \\ & \ddots & & & \\ & & \lambda_r & & \\ & & & 0 & \\ 0 & & & & \ddots\ 0 \end{pmatrix}$ $\quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0.$

$V = (\underset{\sim}{V_1}, \cdots, \underset{\sim}{V_p})$

$$C = (\underbrace{\lambda_1^{\frac{1}{2}} \underset{\sim}{V_1}\, \lambda_2^{\frac{1}{2}} \underset{\sim}{V_2} \cdots \lambda_r^{\frac{1}{2}} \underset{\sim}{V_r}}_{r}\ \underbrace{\cancel{\underset{\sim}{0} \cdots \underset{\sim}{0}}}_{p-r})(\lambda_1^{\frac{1}{2}} \underset{\sim}{V_1} \cdots \lambda_r^{\frac{1}{2}} \underset{\sim}{V_r}\ \cancel{\underset{\sim}{0} \cdots \underset{\sim}{0}})^T$$

$$= \underset{p \times r}{L}\ \underset{r \times p}{L^T} \text{ where } L = (\lambda_1^{\frac{1}{2}} \underset{\sim}{V_1} \cdots \lambda_r^{\frac{1}{2}} \underset{\sim}{V_r}) = (\underset{\sim}{l_1}, \cdots, \underset{\sim}{l_r})$$

So if $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, C)$ then we can write

$$\underset{\sim}{X} = \underset{\sim}{\mu} + \underset{p \times r}{L}\ \underset{r \times 1}{\underset{\sim}{F}} \text{ where } \underset{\sim}{F} \sim N_r(\underset{\sim}{0}, I)$$

$$\underset{\sim}{X} = \underset{\sim}{\mu} + \sum_{k=1}^{r} \underset{\text{factors}}{\underset{\sim}{F_k}}\ \underset{\sim}{l_k} \leftarrow \text{loadings}$$

This just PCA assuming $\text{rank}(C) = r < p$

What's wrong with this model?

- assumption that $\text{rank}(C) < p$ not realistic

How to fix this?

① Write to keep (some of) the lower dimensional structure of C.

② Make C have full rank.

Naive (PCA) model

$$\underset{\sim}{X} = \underset{\sim}{\mu} + \underset{\text{loadings}}{L}\ \underset{\text{factors}}{\underset{\sim}{F}}, \quad \underset{\sim}{F} \sim N_r(\underset{\sim}{0}, I).$$

Factor analysis model

$$\underset{\sim}{X} = \underset{\sim}{\mu} + L \underset{\sim}{F} + \underset{\sim}{\varepsilon}$$

where ① $\underset{\sim}{F} \sim N_r(\underset{\sim}{0}, I)$

② $\underset{\sim}{\varepsilon} \sim N_p(\underset{\sim}{0}, \Psi)$ where $\Psi = \begin{pmatrix} \psi_1 & & 0 \\ & \ddots & \\ 0 & & \psi_p \end{pmatrix}$

③ Components of $\underset{\sim}{\varepsilon}$ are independent of $\underset{\sim}{F}$

Note: Can also specify model in terms of $\text{Cov}(\underset{\sim}{F})$, $\text{Cov}(\underset{\sim}{\varepsilon})$ and $\underset{\sim}{\varepsilon}, \underset{\sim}{F}$ uncorrelated.

$$\Rightarrow \underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \underbrace{LL^T + \Psi}_{C \text{ (rank } p)})$$

Notes:

⓪ Why factor analysis?

Unobserved factors explain relationships b/w variables (via loadings)

① If $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ $\quad Var(X_j) = \underbrace{\sum_{k=1}^{r} l_{jk}}_{\text{commonality}} + \Psi_j$ $\uparrow$ uniqueness $\quad \to$ loadings for X_r for factors $1, \dots, r$

② Factor analysis is invariant (equivalent?) to changes in scale for each variable.

$$\underset{\sim}{Y} = \underset{\downarrow \text{ diagonal}}{D}\underset{\sim}{X} + \underset{\sim}{a} \quad \text{then } \underset{\sim}{Y} = \underbrace{D\underset{\sim}{u} + \underset{\sim}{a}}_{\text{new mean vector}} + \underbrace{DL}_{\text{new loadings}}\overset{\text{same factors}}{\underset{\sim}{F}} + D\underset{\sim}{\varepsilon} \to Cov(D\underset{\sim}{\varepsilon}) = \underbrace{D\Psi D}_{\text{diagonal}}$$

③ Relationship to graphical models

Example: Single factor model.

$$\underset{\sim}{X} = \underset{\sim}{\mu} + F\underset{\sim}{l} + \underset{\sim}{\varepsilon} \sim N_p(\underset{\sim}{\mu}, \underbrace{\Psi + \underset{\sim}{l}\underset{\sim}{l}^T}) = C$$

$$K = (\Psi + \underset{\sim}{l}\underset{\sim}{l}^T)^{-1} = \Psi^{-1} - \frac{\Psi^{-1}\underset{\sim}{l}\underset{\sim}{l}^T\Psi^{-1}}{1 + \underset{\sim}{l}^T\Psi^{-1}\underset{\sim}{l}} \leftarrow \text{diagonal matrix}$$

$$= \begin{pmatrix} k_{11} & \cdots & k_{1p} \\ \vdots & \ddots & \\ k_{p1} & \cdots & k_{pp} \end{pmatrix}$$

$k_{ij} = 0$ if either $l_i = 0$ or $l_j = 0$

- not terribly interesting!

Now look at $\underset{\sim}{Y} = \begin{pmatrix} \underset{\sim}{X} \\ F \end{pmatrix} \sim N_{p+1}\left(\begin{pmatrix} \underset{\sim}{\mu} \\ 0 \end{pmatrix}, \underbrace{\begin{pmatrix} \Psi + \underset{\sim}{l}\underset{\sim}{l}^T & \underset{\sim}{l} \\ \underset{\sim}{l}^T & 1 \end{pmatrix}}_{C_+}\right)$

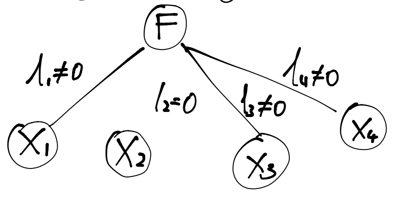
Concentration matrix $K_+ = C_+^{-1}$

$$= \begin{pmatrix} \Psi^{-1} & -\Psi^{-1}\underset{\sim}{l} \\ -\underset{\sim}{l}^T\Psi^{-1} & 1 + \underset{\sim}{l}^T\Psi^{-1}\underset{\sim}{l} \end{pmatrix}$$

$\to$ links between $\underset{\sim}{X}$ & F

$\leftarrow$ diagonal

From this get following dependency structure



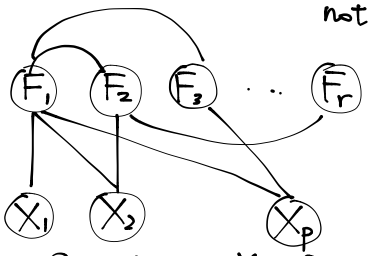
- link exists between F & X_j if $l_j \neq 0$
- no link if $l_j = 0$.

$\Rightarrow X_1, \dots, X_p$ are all conditionally independent given F, i.e. all dependency comes via F.

Extension to general factor model. $\quad \underset{\sim}{X} = \underset{\sim}{\mu} + L\underset{\sim}{F} + \underset{\sim}{\varepsilon}$

Define $\underset{\sim}{Y} = \begin{pmatrix} \underset{\sim}{X} \\ \underset{\sim}{F} \end{pmatrix} \begin{matrix} \to p \\ \to r \end{matrix} \sim N_{p+r}\left(\begin{pmatrix} \underset{\sim}{\mu} \\ \underset{\sim}{0} \end{pmatrix}, \underbrace{\begin{pmatrix} \Psi + LL^T & L \\ L^T & I \end{pmatrix}}_{C_+} \right)$

$$K_+ = C_+^{-1} = \begin{pmatrix} \Psi^{-1} & -\Psi^{-1}L \\ -L^T\Psi^{-1} & \underbrace{I + L^T\Psi^{-1}L}_{\text{not necessarily diagonal}} \end{pmatrix}$$



Again, $X_1, \cdots, X_p$ are conditional independent given $F_1, \cdots, F_r$

Estimation: Given $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_n$ from factor model

① How many factors?

② Given r, how to estimate loadings matrix L and uniqueness matrix Ψ?

- L is not uniquely determined!
- both a bug & a feature.