Masters Project Milestone Report

Name of the student: Zhining Wang

Name of the supervisor(s): Dr. Emi Tanaka and Dr. Qinian Jin (co-supervisor)

Title of the project (may be tentative): Model Selection in Factor Analysis

Summary: Notation is presented in Table 1.

What is factor analysis and why is it important?

Factor analysis is a mathematical model which tries to use fewer underlying factors to explain the correlation between a large set of observed variables(Mardia, Kent, and Bibby 1979). It provides a useful tool for exploring the covariance structure among observable variables (Hirose and Yamamoto 2015). One of the major assumptions that factor analytic model stands on is that it is impossible for us to observe those underlying factors directly. This assumption is especially suited to subjects like psychology where we cannot observe exactly some concept like how intelligent our subjects are(Mardia, Kent, and Bibby 1979).

Suppose we have a observable random vector $\mathbf{y} \in \mathbb{R}^p$ with mean $\mathbb{E}[\mathbf{y}] = \mu$ and variance $\mathbb{V}[\mathbf{y}] = \Sigma$. Then a d-order factor analysis model for $\mathbf{y}$ can be given by

$$\mathbf{y} = \Lambda\mathbf{f} + \mu + \epsilon, \tag{1}$$

where $\Lambda \in \mathbb{R}^{p\times d}$ is called *loading matrix*, we call $\mathbf{f} \in \mathbb{R}^d$ as *common vectors* and $\epsilon \in \mathbb{R}^p$ is *unique factors*. To make the model well-defined, we may assume

$$\mathbb{E}[\mathbf{f}] = \mathbf{0}_d, \mathbb{V}[\mathbf{f}] = \mathbf{I}_{d\times d}, \mathbb{E}[\epsilon] = \mathbf{0}_p, \mathbb{V}[\epsilon] =: \Psi = \mathrm{diag}(\psi_{11}, \dots, \psi_{pp})$$

and also the independence between any elements from $\mathbf{f}$ and ϵ separately, i.e.

$$\mathbf{Cov}[\mathbf{f}_i, \epsilon_j] = 0, \text{for all } i \in \{1, 2, \dots, d\} \text{ and } j \in \{1, 2, \dots, p\}$$

Straightforwardly, the covariance of observable vector $\mathbf{y}$ can be modelled by

$$\mathbb{V}[\mathbf{y}] = \Lambda\Lambda^\top + \Psi \tag{2}$$

Indeterminacy of the loading matrix

One can easily see that if our factor analytic model is given by (1), then it can also be modelled as

$$\mathbf{y} = (\Lambda\mathbf{M})(\mathbf{M}^\top\mathbf{f}) + \mu + \epsilon$$

where the matrix $\mathbf{M}$ is orthogonal and simultaneously the variance of $\mathbf{y}$ given by (2) still holds, since

$$\mathbb{V}[\mathbf{y}] = (\Lambda\mathbf{M}\mathbf{M}^\top)\mathbb{V}[\mathbf{f}](\Lambda\mathbf{M}\mathbf{M}^\top)^\top + \Psi = \Lambda\Lambda^\top + \Psi.$$

Therefore a rotated loading matrix $\Lambda\mathbf{M}$ is still a valid loading matrix for a factor analytic model.

Traditional Estimation of parameters in factor analytic models

We denote the set of parameters by $\beta := \{\text{vec}(\Lambda), \text{vec}(\Psi)\}$ where $\text{vec}(\cdot)$ is the vectorisation of the input. Traditionally, a two-step procedure is used to construct a factor analytic models: estimated parameters by Maximum likelihood estimation (aka, MLE) and then use rotation techniques to find an interpretable model.

Maximum Likelihood Estimation. Suppose we have N independent and identically distributed observations $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ from a p-dimensional multi-variate normal distribution $N_p(\mu, \Sigma)$ and by our hypothesis, we have $\Sigma = \Lambda\Lambda^\top + \Psi$. Then the likelihood function is given by

$$L(\Lambda, \Psi) = \Pi_{i=1}^N \left[(2\pi)^{-\frac{p}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathbf{y}_i - \mu)^\top \Sigma^{-1}(\mathbf{y}_i - \mu)) \right].$$

Many algorithms are developed to get the root of the equation

$$\frac{\partial l}{\partial \Lambda} = 0 \text{ and } \frac{\partial l}{\partial \Psi} = 0$$

where l is the log-likelihood given by $l(\Lambda, \Psi) := \log(L(\Lambda, \Psi))$.

Rotation Techniques. After estimation, we want to rotate the loading matrix to possess a sparse matrix in order to interpret the observable variables by underlying factors better. Also there are many method to achieve rotation as well such as the varimax method and the promax method (Hirose and Yamamoto 2015).

Penalized Likelihood method

The biggest challenge traditional two-step method facing is unsuitability(Hirose and Yamamoto 2015). Just like in regression models, MLE may lead to overfitness and rotatin techniques may not produce sparse enough loading matrix. In some simulation studies, the two-step method generated bad loading matrix even if the true loading matrix is sparse.(Hirose and Yamamoto 2015) Our research later will focus on penalized likelihood method such like LASSO penalty(Ning and Georgiou 2011) (Choi, Oehlert, and Zou 2010) and may include some non-convex penalty like MC+(Zhang 2010).

Plan

My plan is to

1. Conduct a detailed examination of the traditional two-step approach, with a particular focus on rotation techniques.
2. Explore how does the penalized likelihood method work for order-selection and overcome the defects traditional two-step approach are facing.
3. Explore the implementation of the LASSO penalty in a factor analytic model to enhance model selection and sparsity.
4. Simulation study
5. If possible, I may explore linear latent variable models, i.e. the model given by (1) will additionally have a linear term βX to fit the mean μ.

Appendix.

TABLE 1. List of notations used in this report.

Notation	Description
$A^\top$	the transpose of the matrix (or vector) A
$\mathbb{R}^p$	the space of all p-dimensional real column vectors like $[a_1, a_2, \dots, a_p]^\top$
$\mathbb{R}^{p\times q}$	the space of all real matrices with size $p \times q$
$\mathbb{E}[\cdot]$	the expectation, or mean of a random variable
$\mathbb{V}[\cdot]$	the variance, or covariance of a random variable
$\mathbf{Cov}[\cdot,\cdot]$	the covariance of two random variables
$\mathbf{0}_p$ and $\mathbf{0}_{p\times q}$	the p-dimensional 0 vector or 0 matrix in size $p \times q$ respectively
$\mathbf{I}_p$	the identity matrix in size $p \times p$

REFERENCES

Choi, Jang, Gary Oehlert, and Hui Zou. 2010. "A Penalized Maximum Likelihood Approach to Sparse Factor Analysis." *Statistics and Its Interface* 3 (4): 429–36.

Hirose, Kei, and Michio Yamamoto. 2015. "Sparse Estimation via Nonconcave Penalized Likelihood in Factor Analysis Model." *Statistics and Computing* 25 (5): 863–75. https://doi.org/10.1007/s11222-014-9458-0.

Mardia, K. V., J. T. Kent, and J. M. Bibby. 1979. *Multivariate Analysis*. Probability and Mathematical Statistics. London ; New York: Academic Press.

Ning, Lipeng, and Tryphon T. Georgiou. 2011. "Sparse Factor Analysis via Likelihood and ℓ_1-Regularization." In *2011 50th IEEE Conference on Decision and Control and European Control Conference*. IEEE.

Zhang, Cun-Hui. 2010. "Nearly Unbiased Variable Selection Under Minimax Concave Penalty." *Journal Name Here*, 894–942.

Student's signature **Supervisor's signature**

Date: 2024-03-07