

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): Your project title

Summary:

Notation. In this article, we will denote:

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

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 et al., 1995). It provides a useful tool for exploring the covariance structure among observable variables (Hirose, 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 et al., 1995).

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

$$(1) \quad \mathbf{y} = \Lambda \mathbf{f} + \mu + \epsilon,$$

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

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

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

$$(2) \quad \mathbb{V}[\mathbf{y}] = \Lambda \Lambda^T + \Psi$$

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

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

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

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

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 $\beta := \{\Lambda, \Psi\}$. 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 $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$ from a p -dimensional multi-variate normal distribution $N_p(\mu, \Sigma)$ and by our hypothesis, we have $\Sigma = \Lambda \Lambda^T + \Psi$. Then the likelihood function is given by

$$L(\Lambda, \Psi) = \prod_{i=1}^N [(2\pi)^{-\frac{p}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp(-\frac{1}{2}(x_i - \mu)^T \Sigma^{-1} (x_i - \mu))]$$

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, 2015).

PENALIZED LIKELIHOOD METHOD

The biggest challenge traditional two-step method facing is unsuitability(Hirose, 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, 2015) Our research later will focus on penalized likelihood method such like LASSO penalty(Ning,2011 and Choi, 2010) and may include some non-convex penalty like MC+(Zhang, 2010).

Reference. Hirose, Kei, and Michio Yamamoto. “Sparse estimation via nonconcave penalized likelihood in factor analysis model.” *Statistics and Computing* 25 (2015): 863-875.

Ning, Lipeng, and Tryphon T. Georgiou. “Sparse factor analysis via likelihood and 1-regularization.” 2011 50th IEEE conference on decision and control and European control conference. IEEE, 2011.

Choi, Jang, Gary Oehlert, and Hui Zou. “A penalized maximum likelihood approach to sparse factor analysis.” *Statistics and its Interface* 3.4 (2010): 429-436.

Zhang, Cun-Hui. “Nearly unbiased variable selection under minimax concave penalty.” (2010): 894-942.

Mardia et al.”Multivariate Analysis” Chapter 9

Student’s signature

Supervisor’s signature

Date: 2024-03-06