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:

Introduction of Factor Analytic Models

Factor analysis is a statistical method which attempts to use fewer underlying factors to explain the correlation between a large set of observed variables. Suppose we have an observable random vector $y_i \in \mathbb{R}^p$ for the i-th subject with mean $\mathbb{E}(y_i) = \mu$ and variance $\mathbb{V}(y_i) = \Sigma$. Then a k-order factor analysis model for y_i can be given by

$$y_i = \mu + \Lambda f_i + \epsilon_i, \tag{1}$$

where $\Lambda \in \mathbb{R}^{p\times k}$, $f_i \in \mathbb{R}^k$ is and $\epsilon_i \in \mathbb{R}^p$ are called the *loading matrix*, *common factors* and *unique (or specific) factors*, respectively. The order k is usually much smaller than p. For simplicity, we can assume that $\mu = 0_p$. We also assume that the common factors and the unique factors are independent. We aim to estimate a *sparse* loading matrix as well as estimate the diagonal covariance matrix of the unique factor.

Current Work Results

I have completed the introduction to the factor analysis model and reviewed the traditional two-step procedure for obtaining the sparse loading matrix. However, this traditional method has been criticized for potentially causing overfitting and producing overly dense solutions. I conducted a small simulation to demonstrate these issues. Following this, I explored Expecation-Maximisation (EM) algorithm and penalised likelihood approaches to fitting factor analytic models, including a brief proof of the algorithm's correctness. I detailed extensively procedures for both the E-step and the M-step in the aforementioned algorithm. In the M-step, I employed a proximal gradient method. I implemented the algorithm (that is still in work-in-progress) in the statistical software, R.

Future Expectations

This semester, I will continue polishing my code, including bug fixes and stability enhancements. I will also focus on the simulation settings, exploring various interesting aspects such as initialization strategies, parameter tuning, and order selection. I plan to implement these elements using R and visualize the results in my thesis. Additionally, I will discuss the results and highlight the limitations of the model or algorithm. I intend to allocate at least a month to polish my notations, language, and typesetting.

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

Date: 2024-08-09