Mixtures of factor-analytic models with covariates for modeling multiply censored dependent variables

具共變數之混合因子分析模型對於多重設限相依變數之建模

Student: Wan-Chen Hsieh Advisor: Tsung-I Lin

Institute of Statistics, National Chung Hsing University, Taichung, Taiwan

Abstract

Censored data arise frequently in diverse applications in which observations to be measured may be subject to some upper and lower detection limits due to the restriction of experimental apparatus such that they are not exactly quantifiable. Mixtures of factor analyzers with censored data (MFAC) have been recently proposed for model-based density estimation and clustering of high-dimensional data under the presence of censored observations. In this paper, we consider an extended version of MFAC with covariates to accommodate multiply censored dependent variables and develop two analytically feasible EM-type algorithm for computing maximum likelihood estimates of the parameters with closed-form expressions. Moreover, we provide an information-based method to compute asymptotic standard errors of mixing proportions and regression coefficients. The utility and performance of the proposed methodology is illustrated through two real data examples.

Introduction

An extension of the MFAC approach (Wang et al. 2019) is introduced by including regression covariates. The proposed model clearly encompasses the MFAC approach with advantages including the flexibility in situations involving related independent variables within *g* classes. A description of finding ML estimates of model parameters through two EM-based algorithms in which the E-step relies on calculating the first-two moments of the truncated multivariate-normal distribution (Lin, 2009) is provided. The computational procedures we developed are rather stable and efficient for estimating the proposed model. In addition, we offer an approximation of the empirical Hessian matrix via the Louis' method (Louis, 1982), so that the standard errors of mixture regression coefficients and mixing proportions can be obtained. The utility of our methodology is illustrated through the analysis of two real-life datasets related to educational assessment and water quality.

Model Formulation

MFA with linear regression model

$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{\beta}_i + \boldsymbol{B}_i\boldsymbol{u}_{ij} + \boldsymbol{\varepsilon}_{ij}$ with probability π_i (for i=1,…,g).

$\boldsymbol{u}_{ij} \sim N_q(\boldsymbol{0}, \boldsymbol{I}_q)$ and $\boldsymbol{\varepsilon}_{ij} \sim N_p(\boldsymbol{0}, \boldsymbol{D}_i)$ are mutually independent for all i and j.

Censoring information

Define $\boldsymbol{v}_j = (v_{j1}, \ldots, v_{jp})^{\mathsf{T}}$ as the vector of uncensored reading or censoring level, and $\boldsymbol{c}_j = (c_{j1}, \ldots, c_{jp})^{\mathsf{T}}$ be the vector of censoring indicators. We shall focus the case of left censoring:

$$\begin{cases} y_{jk} \le v_{jk}, & \text{if} \quad c_{jk} = 1 \\ y_{jk} = v_{jk}, & \text{if} \quad c_{jk} = 0 \end{cases}$$

Incorporating the censoring information into the model, we have the distribution, $\boldsymbol{y}_j \mid (\boldsymbol{v}_j, \boldsymbol{c}_j, z_{ij} = 1) \sim TN_p(\boldsymbol{X}_j\boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i; \mathbb{A}_j)$ where $TN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbb{A}_j)$ denotes a *p*-variate truncated normal distribution restricted within a right-truncated hyperplane $\mathbb{A}_j = \mathbb{A}_j^{\mathrm{o}} \times \mathbb{A}_j^{\mathrm{c}}$, where $\mathbb{A}_j^{\mathrm{o}} = \{-\infty < y_{jk} < \infty, k = 1, \ldots, p_j^{\mathrm{o}}\}$ and $\mathbb{A}_j^{\mathrm{c}} = \{y_{jk} \le v_{jk},\ k = 1, \ldots, p - p_j^{\mathrm{o}}\}$.

Therefore, the MFAC admits a two-level hierarchical representation:

$$\begin{aligned} \boldsymbol{v}_j \mid (\boldsymbol{c}_j, z_{ij} = 1) &\sim f_{ij}(\boldsymbol{v}_j \mid \boldsymbol{c}_j, \boldsymbol{\theta}_i), \\ \boldsymbol{z}_j &\sim \mathcal{M}(1; \pi_1, \cdots, \pi_g). \end{aligned}$$

where $f_{ij}(\boldsymbol{v}_j \mid \boldsymbol{c}_j, \boldsymbol{\theta}_i) = f(\boldsymbol{y}_j^c \le \boldsymbol{v}_j^c \mid \boldsymbol{y}_j^o = \boldsymbol{v}_j^o, \boldsymbol{\theta}_i) f(\boldsymbol{y}_j^o = \boldsymbol{v}_j^o \mid \boldsymbol{\theta}_i) = \Phi_{p-p_j^o}(\boldsymbol{v}_j^c \mid \boldsymbol{\mu}_{ij}^{c\cdot o}, \boldsymbol{\Sigma}_{ij}^{cc\cdot o}) \phi_{p_j^o}(\boldsymbol{v}_j^o \mid \boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo})$

and $\boldsymbol{\mu}_{ij}^{\mathrm{o}} = \boldsymbol{O}_j\boldsymbol{X}_j\boldsymbol{\beta}_i$, $\boldsymbol{\Sigma}_{ij}^{\mathrm{oo}} = \boldsymbol{O}_j\boldsymbol{\Sigma}_i\boldsymbol{O}_j^{\top}$, $\boldsymbol{\mu}_{ij}^{\mathrm{c\cdot o}} = \boldsymbol{C}_j\boldsymbol{X}_j\boldsymbol{\beta}_i + \boldsymbol{C}_j\boldsymbol{\Sigma}_i\boldsymbol{S}_{ij}^{\mathrm{oo}}(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}_i)$,

$$\boldsymbol{S}_{ij}^{\mathrm{oo}} = \boldsymbol{O}_j^{\top}(\boldsymbol{O}_j\boldsymbol{\Sigma}_i\boldsymbol{O}_j^{\top})^{-1}\boldsymbol{O}_j \quad \text{and} \quad \boldsymbol{\Sigma}_{ij}^{\mathrm{cc\cdot o}} = \boldsymbol{C}_j(\boldsymbol{I}_p - \boldsymbol{\Sigma}_i\boldsymbol{S}_{ij}^{\mathrm{oo}})\boldsymbol{\Sigma}_i\boldsymbol{C}_j^{\top}.$$

Given the hierarchical structures, we have the following conditional distributions

$$\begin{aligned} \boldsymbol{z}_j \mid (\boldsymbol{v}_j, \boldsymbol{c}_j) &\sim \mathcal{M}(1; \tilde{\pi}_{1j}, \ldots, \tilde{\pi}_{gj}) \\ \boldsymbol{y}_j^{\mathrm{o}} \mid (z_{ij} = 1) &\sim N_{p_j^o}(\boldsymbol{\mu}_{ij}^{\mathrm{o}}, \boldsymbol{\Sigma}_{ij}^{\mathrm{oo}}) \\ \boldsymbol{y}_j^{\mathrm{c}} \mid (\boldsymbol{y}_j^{\mathrm{o}}, z_{ij} = 1) &\sim N_{p-p_j^o}(\boldsymbol{\mu}_{ij}^{\mathrm{c\cdot o}}, \boldsymbol{\Sigma}_{ij}^{\mathrm{cc\cdot o}}) \\ \boldsymbol{y}_j^{\mathrm{c}} \mid (\boldsymbol{y}_j^{\mathrm{o}}, \boldsymbol{v}_j, \boldsymbol{c}_j, z_{ij} = 1) &\sim TN_{p-p_j^o}(\boldsymbol{\mu}_{ij}^{\mathrm{c\cdot o}}, \boldsymbol{\Sigma}_{ij}^{\mathrm{cc\cdot o}}; \boldsymbol{A}_j^{\mathrm{c}}) \\ \boldsymbol{u}_{ij} \mid (\boldsymbol{y}_j, z_{ij} = 1) &\sim N_q(\boldsymbol{\Gamma}_i^{\top}(\boldsymbol{y}_j - \boldsymbol{\mu}_i), \boldsymbol{\Omega}_i), \ \boldsymbol{\Gamma}_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{B}_i,\ \boldsymbol{\Omega}_i = \boldsymbol{I}_q - \boldsymbol{B}_i^{\top}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{B}_i, \end{aligned}$$

$$\tilde{\pi}_{ij} = P(z_{ij} = 1 \mid \boldsymbol{v}_j, \boldsymbol{c}_j) = \frac{\pi_i f_{ij}(\boldsymbol{v}_j \mid \boldsymbol{c}_j, \boldsymbol{\theta}_i)}{\sum_{t=1}^{g} \pi_t f_{tj}(\boldsymbol{v}_j \mid \boldsymbol{c}_j, \boldsymbol{\theta}_t)}$$

ML Estimation

The EM algorithm

Let $\boldsymbol{\Theta} = (\boldsymbol{\pi_i}, \{\boldsymbol{\beta}_i, \boldsymbol{B}_i, \boldsymbol{D}_i\}_{i=1}^{g})$ be the entire unknown parameters in the model.

- E-step: $Q(\Theta \mid \hat{\Theta}^{(k)}) = E[\ell_c(\Theta \mid \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{z}) \mid \boldsymbol{v}_j, \boldsymbol{c}_j, \hat{\Theta}^{(k)}]$
- M-step: Find $\hat{\Theta}^{(k+1)}$ by maximizing *Q*-function, yielding

$$\begin{aligned} \hat{\pi}_i^{(k+1)} &= \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}{n}, \\ \hat{\boldsymbol{\beta}}_i^{(k+1)} &= \left(\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \boldsymbol{X}_j^{\top}\boldsymbol{X}_j\right)^{-1} \left(\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \boldsymbol{X}_j^{\top}(\hat{\boldsymbol{y}}_{ij}^{(k)} - \boldsymbol{B}_i\hat{\boldsymbol{u}}_{ij}^{(k)})\right), \\ \hat{\boldsymbol{B}}_i^{(k+1)} &= \left(\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} (\widehat{\boldsymbol{y}\boldsymbol{u}}_{ij}^{(k)} - \boldsymbol{X}_j\hat{\boldsymbol{\beta}}_i^{(k+1)}\hat{\boldsymbol{u}}_{ij}^{(k)\top})\right) \left(\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{\boldsymbol{u}}_{ij}^{2\,(k)}\right)^{-1}, \\ \hat{\boldsymbol{D}}_i^{(k+1)} &= \frac{\mathrm{Diag}(\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{\boldsymbol{\Psi}}_{ij}^{(k+1)})}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}, \end{aligned}$$

where

$$\begin{aligned} \hat{\boldsymbol{\Psi}}_{ij}^{(k+1)} = {} & \hat{\boldsymbol{y}}_{ij}^{2\,(k)} - (\hat{\boldsymbol{y}}_{ij}^{(k)} - \hat{\boldsymbol{B}}_i^{(k+1)}\hat{\boldsymbol{u}}_{ij}^{(k)})\hat{\boldsymbol{\beta}}_i^{(k+1)\top}\boldsymbol{X}_j^{\top} - \boldsymbol{X}_j\hat{\boldsymbol{\beta}}_i^{(k+1)} \times \\ & (\boldsymbol{y}_{ij}^{(k)\top} - \hat{\boldsymbol{B}}_i^{(k+1)}\hat{\boldsymbol{u}}_{ij}^{(k)})^{\top} - \widehat{\boldsymbol{y}\boldsymbol{u}}_{ij}^{(k)}\hat{\boldsymbol{B}}_i^{(k+1)\top} + \\ & \boldsymbol{X}_j\hat{\boldsymbol{\beta}}_i^{(k+1)}\hat{\boldsymbol{\beta}}_i^{(k+1)\top}\boldsymbol{X}_j^{\top} - \hat{\boldsymbol{B}}_i^{(k+1)}\widehat{\boldsymbol{y}\boldsymbol{u}}_{ij}^{(k)\top} + \\ & \hat{\boldsymbol{B}}_i^{(k+1)}\hat{\boldsymbol{u}}_{ij}^{2\,(k)}\hat{\boldsymbol{B}}_i^{(k+1)\top}. \end{aligned}$$

The AECM algorithm

Partition the unknown parameters into two subjects: $\Theta_1 = \{\pi_i, \boldsymbol{\beta}_i\}_{i=1}^{g}$ and $\Theta_2 = \{\boldsymbol{B}_i, \boldsymbol{D}_i\}_{i=1}^{g}$.

- The 1st cycle:
  - E-step: $Q^{[1]}(\Theta_1 \mid \hat{\Theta}^{(k)}) = E[\ell_c^{[1]}(\Theta_1 \mid \boldsymbol{y}, \boldsymbol{z}) \mid \boldsymbol{v}_j, \boldsymbol{c}_j, \hat{\Theta}^{(k+1)}]$
  - CM-step: Update $\hat{\Theta}_1^{(k+1)}$ by maximizing $Q^{[1]}$ function, yielding

$$\begin{aligned} \hat{\boldsymbol{\pi}}_i^{(k+1)} &= \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}{n}, \\ \hat{\boldsymbol{\beta}}_i^{(k+1)} &= \left(\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \boldsymbol{X}_j^{\top}\boldsymbol{X}_j\right)^{-1} \left(\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \boldsymbol{X}_j^{\top}\hat{\boldsymbol{Y}}_{ij}^{(k)}\right). \end{aligned}$$

- The 2nd cycle:
  - E-step: $Q^{[2]}(\Theta_2 \mid \hat{\Theta}^{(k)}) = E[\ell_c^{[2]}(\Theta_1 \mid \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{z}) \mid \boldsymbol{v}_j, \boldsymbol{c}_j, \hat{\Theta}^{(k+1)}]$
  - CM-step: Update $\hat{\Theta}_2^{(k+1)}$ by maximizing $Q^{[2]}$ function, yielding

$$\begin{aligned} \hat{\boldsymbol{B}}_i^{(k+1)} &= \boldsymbol{S}_i^{(k)}\hat{\boldsymbol{\Gamma}}_i^{(k)}(\hat{\boldsymbol{\Gamma}}_i^{(k)\top}\boldsymbol{S}_i^{(k)}\hat{\boldsymbol{\Gamma}}_i^{(k)} + \hat{\boldsymbol{\Omega}}_i^{(k)})^{-1}, \\ \hat{\boldsymbol{D}}_i^{(k+1)} &= \mathrm{Diag}\left\{\boldsymbol{S}_i^{(k)} - \hat{\boldsymbol{B}}_i^{(k)}\hat{\boldsymbol{\Gamma}}_i^{(k)\top}\boldsymbol{S}_i^{(k)}\right\}, \text{ where} \\ \boldsymbol{S}_i^{(k)} &= \frac{1}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}} \sum_{j=1}^{n} \hat{z}_{ij}^{(k)} [(\hat{\boldsymbol{Y}}_{ij}^{(k)} - \boldsymbol{X}_j\hat{\boldsymbol{\beta}}_i^{(k+1)}) \times \\ & \quad (\hat{\boldsymbol{Y}}_{ij}^{(k)} - \boldsymbol{X}_j\hat{\boldsymbol{\beta}}_i^{(k+1)})^{\top} + \hat{\boldsymbol{\Lambda}}_{ij}^{(k)}]. \end{aligned}$$

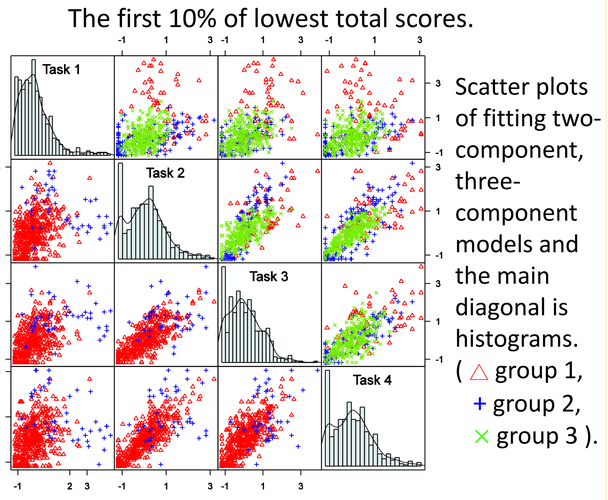
Real Examples

EGRA data

We analyze the dataset presented by Costa et al. (2014) for 502 Peruvian students in 2007, who were taken by four of ten EGRA tasks**,** the four tasks in the dataset are:
(i) recognizing letters of the alphabet,
(ii) recognizing simple words,
(iii) simple decoding meaningless words, and
(iv) reading of a passage.
The design matrix $\boldsymbol{X}_j$ of dimension 4×4 for each student *j* corresponds to fixed effects vector $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{i4})$, for $i = 1, \ldots, g$.
The covariates in $\boldsymbol{X}_j$ related to $\boldsymbol{\beta}_i$ include **gender** (0=F, 1=M), **grade** (0=2nd, 1=3rd), **residence zone** (0 = Rural, 1 = Urban) and the adjusted **age** by subtracting the mean of students' ages.

- Censoring rate:

The first 10% of lowest total scores.



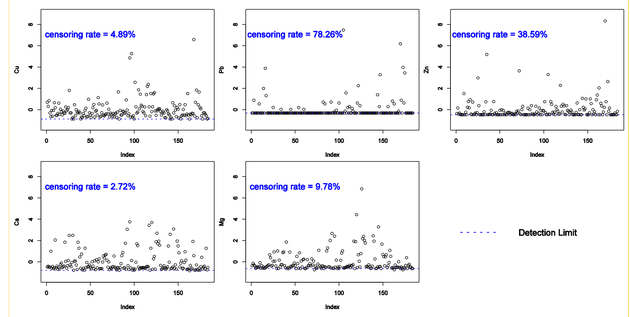
Scatter plots of fitting two-component, three-component models and the main diagonal is histograms. (△ group 1, + group 2, × group 3).

	$g=1$	$g=2$		$g=3$		
Parameters	group 1	group 1	group 2	group 1	group 2	group 3
π	-	0.099(0.017)	-	0.135(0.022)	0.595(0.043)	-
β_{i1}	-0.235(0.067)	-0.141(0.153)	-0.216(0.062)	-0.052(0.189)	-0.153(0.080)	-0.461(0.152)
β_{i2}	0.513(0.086)	1.012(0.167)	0.458(0.081)	0.857(0.205)	0.532(0.097)	0.197(0.199)
β_{i3}	0.027(0.057)	1.171(0.112)	-0.113(0.054)	1.102(0.125)	-0.050(0.074)	-0.408(0.118)
β_{i4}	-0.121(0.043)	-0.201(0.107)	-0.129(0.075)	-0.212(0.127)	-0.234(0.060)	-0.083(0.090)
B	-0.519	1.375	-0.411	-0.527	-0.179	0.529
	-0.889	-0.452	-0.851	0.712	-0.685	0.834
	-0.852	-0.120	-0.782	0.454	-0.567	0.879
	-0.855	-0.587	-0.828	0.901	-0.634	0.922
D	0.807	0.065	0.464	1.571	0.471	0.218
	0.157	0.725	0.106	0.489	0.129	0.016
	0.226	0.579	0.191	0.486	0.239	0.055
	0.204	0.888	0.137	0.389	0.169	0.056
AIC	4307.835	4152.751			**4113.519**	
BIC	4358.458	**4258.216**			4273.826	

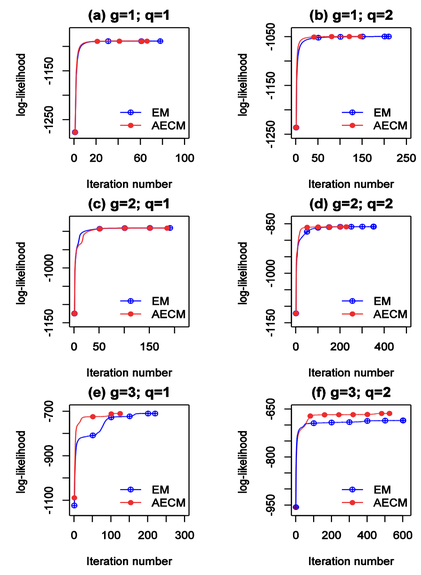
- The regression coefficients suggest that third-grade girls from urban zone in general have significantly better performance in speaking fluent Spanish than the other students.
- The two model selection criteria show that the data can be better described by fitting models with more than one component.

Water resources data

Our 2nd example concerns the water resources data (Hoffman and Johnson, 2015) that collected the trace metal concentration levels of certain dissolved trace metals in the freshwater. There are p=5 (trace metals) concentration levels from n=184 independent randomly selected sites in freshwater streams across Virginia.



- The figure shows the sequential points of five trace metals together with their detection limits and censoring proportions.
- In this example, we are interested in comparing the convergence behavior of the EM and AECM algorithms for fitting the proposed model.



- The CPU time of EM and AECM algorithms are (1.8 min, 36 min) and (1.5 min, 35 min).
- All plots indicate the EM algorithm is worse than AECM in terms of the log-likelihood values and required iterations.
- The CPU time for the AECM algorithm is substantially less than those spent by EM, showing a dramatic advantage of using AECM for acceleration.

Estimation of standard errors

The score vector and Hessian matrix

Let $\tilde{\Theta} = (\tilde{\boldsymbol{\theta}}_1, \ldots \tilde{\boldsymbol{\theta}}_g)$, where $\tilde{\boldsymbol{\theta}}_i = (\pi_i, \boldsymbol{\beta}_i)$ represent the unknown parameters in the ith component. The individual score vector and Hessian matrix are defined as

$$\boldsymbol{s}(\tilde{\Theta} \mid \boldsymbol{y}, \boldsymbol{z}) = \frac{\partial \ell_c(\tilde{\Theta} \mid \boldsymbol{y}, \boldsymbol{z})}{\partial \tilde{\Theta}} = (\boldsymbol{s}_{\tilde{\boldsymbol{\theta}}_1}^{\top}, \ldots, \boldsymbol{s}_{\tilde{\boldsymbol{\theta}}_g}^{\top})^{\top} \text{ and } \boldsymbol{H}(\tilde{\Theta} \mid \boldsymbol{y}, \boldsymbol{z}) = \frac{\partial^2 \ell_c(\tilde{\Theta} \mid \boldsymbol{y}, \boldsymbol{z})}{\partial \tilde{\Theta} \partial \tilde{\Theta}^{\top}} = \mathrm{Diag}\{\boldsymbol{H}_{\tilde{\theta}_1\tilde{\theta}_1}, \ldots, \boldsymbol{H}_{\tilde{\theta}_g\tilde{\theta}_g}\}, \text{ where}$$

$$\boldsymbol{s}_{\tilde{\theta}_i} = \sum_{j=1}^{n} z_{ij}\boldsymbol{s}_{\tilde{\boldsymbol{\theta}}_i}^{(j)} = \sum_{j=1}^{n} z_{ij}[\boldsymbol{s}_{\pi_i}^{(j)\top}, \boldsymbol{s}_{\boldsymbol{\beta}_i}^{(j)\top}]^{\top},\ \boldsymbol{s}_{\pi_i}^{(j)} = \frac{\partial \tilde{\ell}_{ij}}{\partial \pi_i} = \frac{1}{\pi_i},\ \boldsymbol{s}_{\boldsymbol{\beta}_i}^{(j)} = \frac{\partial \tilde{\ell}_{ij}}{\partial \boldsymbol{\beta}_i} = \boldsymbol{X}_j^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_j - \boldsymbol{X}_j\boldsymbol{\beta}_i),$$

$$\boldsymbol{H}_{\tilde{\theta}_i\tilde{\theta}_i} = \sum_{j=1}^{n} z_{ij}\boldsymbol{H}_{\tilde{\boldsymbol{\theta}}_i\tilde{\boldsymbol{\theta}}_i}^{(j)} = \sum_{j=1}^{n} z_{ij} \begin{bmatrix} \boldsymbol{H}_{\pi_i\pi_i}^{(j)} & \boldsymbol{H}_{\pi_i\boldsymbol{\beta}_i}^{(j)} \\ \boldsymbol{H}_{\boldsymbol{\beta}_i\pi_i}^{(j)} & \boldsymbol{H}_{\boldsymbol{\beta}_i\boldsymbol{\beta}_i}^{(j)} \end{bmatrix}. \text{ Furthermore, it can be verified that } \boldsymbol{H}_{\boldsymbol{\beta}_i\pi_i}^{(j)} = \boldsymbol{H}_{\pi_i\boldsymbol{\beta}_i}^{(j)\top} = 0,\ \boldsymbol{H}_{\pi_i\pi_i}^{(j)} = \frac{\partial^2 \tilde{\ell}_{ij}}{\partial \pi_i^2} = -\pi_i^{-2},$$

and $\boldsymbol{H}_{\boldsymbol{\beta}_i\boldsymbol{\beta}_i}^{(j)} = \frac{\partial^2 \tilde{\ell}_{ij}}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_i^{\top}} = -\boldsymbol{X}_j^{\top}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_j.$

Louis's formula(Louis, 1982): The Fisher information matrix of $\tilde{\Theta}$ can be approximated by

$$\boldsymbol{I}(\tilde{\Theta} \mid \boldsymbol{y}) = -E\{\boldsymbol{H}(\tilde{\Theta} \mid \boldsymbol{y}, \boldsymbol{z}) \mid \boldsymbol{v}, \boldsymbol{c}\} - \mathrm{cov}\{\boldsymbol{s}(\tilde{\Theta} \mid \boldsymbol{y}, \boldsymbol{z}) \mid \boldsymbol{v}, \boldsymbol{c}\}$$

Conclusion & Future Work

- We propose two feasible EM-type algorithms without resorting to direct evaluation of the intractable observed likelihood function for estimating the MFAC model with covariates.
- Numerical results suggest the AECM algorithm with less amount of missing information has typically better convergence behavior than EM.
- Future research would be interested in extending the current approach to a more flexible framework that can accommodate missing values and censored responses occurring simultaneously (Lin et al. 2018).

References

- Costa, D.R., Lachos, V.H., Bazan, J.L., Azevedo, C.L.N., (2014) Estimation methods for multivariate Tobit confirmatory factor analysis. Comput. Statist. Data Anal. 79, 248–260
- Hoffman, H., Johnson, R., (2015) Pseudo-likelihood estimation of multivariate normal parameters in the presence of left-censored data. J. Agric. Biol. Environ. Stat. 20,156–171
- Louis, T.A., (1982) Finding the observed information matrix when using the EM algorithm. J. Roy. Statist. Soc. B 44, 226–233
- Lin, T.I. (2009) Maximum likelihood estimation for multivariate skew normal mixture models. Journal of Multivariate Analysis 100, 257-265
- Lin, TI, Lachos, VH and Wang, WL (2018) Multivariate longitudinal data analysis with censored and intermittent missing responses. Statistics in Medicine 37, 2822-2835
- Wang, W.L., Castro, L.M., Lachos, V.H., Lin, T.I. (2019) Model-based clustering of censored data via mixtures of factor analyzers, revised for Computational Statistics and Data Analysis