Grade of membership analysis for multi-layer ordinal categorical data

Huan Qing

School of Economics and Finance, Chongqing University of Technology

Supplementary Material

In this material, we provide the detailed proofs of the theoretical results in the main context and the MATLAB codes of our GoM-DSoG algorithm.

S1 Proof of theoretical results

S1.1 Proof of Proposition 1

Proof. Since $\mathcal{S}=\sum_{l=1}^{L}\Pi\Theta_l'\Theta_l\Pi'=\Pi(\sum_{l=1}^{L}\Theta_l'\Theta_l)\Pi'$, and given that $\mathrm{rank}(\sum_{l=1}^{L}\Theta_l'\Theta_l)=K$ and Π satisfies the condition stated in Equation (2.2) in the main context, by the first result in Theorem 2.1 (Mao et al., 2021), Π is identifiable up to a permutation, i.e., $\tilde{\Pi}=\Pi\mathcal{P}$. Furthermore, based on the second bullet of Lemma 1 and the fact that $\tilde{\Pi}=\Pi\mathcal{P}$, we have:

$$\Theta_l=\mathscr{R}_l'\Pi(\Pi'\Pi)^{-1},$$

$$\tilde{\Theta}_l=\mathscr{R}_l'\tilde{\Pi}(\tilde{\Pi}'\tilde{\Pi})^{-1}=\mathscr{R}_l'\Pi\mathcal{P}(\mathcal{P}'\Pi'\Pi\mathcal{P})^{-1}=\mathscr{R}_l'\Pi\mathcal{P}\mathcal{P}^{-1}(\Pi'\Pi)^{-1}(\mathcal{P}')^{-1}=\Theta_l\mathcal{P}.$$

Thus, for all $l \in [L]$, Θ_l is also identifiable up to the same permutation as Π. $\square$

S1.2 Proof of Lemma 1

Proof. $\mathcal{S} = \sum_{l=1}^{L} \Pi\Theta'_l\Theta_l\Pi' = \Pi(\sum_{l=1}^{L}\Theta'_l\Theta_l)\Pi' = U\Lambda U'$ gives $U = \Pi(\sum_{l=1}^{L}\Theta'_l\Theta_l)\Pi'U\Lambda^{-1}$. Let $X = (\sum_{l=1}^{L}\Theta'_l\Theta_l)\Pi'U\Lambda^{-1}$. Then, $U = \Pi X$ gives $U(\mathcal{I},:) = (\Pi X)(\mathcal{I},:) = \Pi(\mathcal{I},:)X = X$, which leads to $U = \Pi U(\mathcal{I},:)$. Additionally, since $\mathscr{R}_l = \Pi\Theta'_l$, we have $\Theta_l = \mathscr{R}'_l\Pi(\Pi'\Pi)^{-1}$ for $l \in [L]$. $\square$

S1.3 Proof of Theorem 1

Proof. First, we prove the following lemma.

Lemma 1. *If Assumption 1 is satisfied, with probability* $1 - o(\frac{1}{(N+J+L)^3})$,

$$\|S - \mathcal{S}\|_\infty = O(\sqrt{\rho^2 NJL\log(N+J+L)}) + \rho^2 JL.$$

Proof. Under the proposed model, we get

$$
\begin{aligned}
\|S-\mathcal{S}\|_\infty &= \max_{i\in[N]}\sum_{h\in[N]}|S(i,h)-\mathcal{S}(i,h)| = \max_{i\in[N]}\sum_{h\in[N]}|(\sum_{l\in[L]}R_lR_l'-D_l)(i,h)-(\sum_{l\in[L]}\mathscr{R}_l\mathscr{R}_l')(i,h)| \\
&= \max_{i\in[N]}\sum_{h\in[N]}|\sum_{l\in[L]}\sum_{j\in[J]}R_l(i,j)R_l(h,j)-\sum_{l\in[L]}\sum_{j\in[J]}\mathscr{R}_l(i,j)\mathscr{R}_l(h,j)-\sum_{l\in[L]}D_l(i,h)| \\
&= \max_{i\in[N]}\sum_{h\neq i,h\in[N]}|\sum_{l\in[L]}\sum_{j\in[J]}R_l(i,j)R_l(h,j)-\sum_{l\in[L]}\sum_{j\in[J]}\mathscr{R}_l(i,j)\mathscr{R}_l(h,j)| \\
&+ \max_{i\in[N]}|\sum_{l\in[L]}\sum_{j\in[J]}R_l^2(i,j)-\sum_{l\in[L]}\sum_{j\in[J]}\mathscr{R}_l^2(i,j)-\sum_{l\in[L]}D_l(i,i)| \\
&= \max_{i\in[N]}\sum_{h\neq i,h\in[N]}|\sum_{l\in[L]}\sum_{j\in[J]}(R_l(i,j)R_l(h,j)-\mathscr{R}_l(i,j)\mathscr{R}_l(h,j))| + \max_{i\in[N]}\sum_{l\in[L]}\sum_{j\in[J]}\mathscr{R}_l^2(i,j) \\
&\leq \max_{i\in[N]}\sum_{h\neq i,h\in[N]}|\sum_{l\in[L]}\sum_{j\in[J]}(R_l(i,j)R_l(h,j)-\mathscr{R}_l(i,j)\mathscr{R}_l(h,j))| + \sum_{l\in[L]}\sum_{j\in[J]}\rho^2 \\
&= \max_{i\in[N]}\sum_{h\neq i,h\in[N]}|\sum_{l\in[L]}\sum_{j\in[J]}(R_l(i,j)R_l(h,j)-\mathscr{R}_l(i,j)\mathscr{R}_l(h,j))| + \rho^2 JL
\end{aligned}
$$

Define v as any $(N-1)$-by-1 vector. For $i\in[N]$ and $h\neq i, h\in[N]$, let $z_{(ih)}=\sum_{l\in[L]}\sum_{j\in[J]}(R_l(i,j)R_l(h,j)-\mathscr{R}_l(i,j)\mathscr{R}_l(h,j))$. Then, for $i\in[N]$, define $F_{(i)}=\sum_{h\neq i,h\in[N]}z_{(ih)}v(h)$. Let $\tau=\max_{i\in[N],h\neq i,h\in[N]}|z_{(ih)}|$. Given that $\mathbb{E}(z_{(ih)})=0$, to simplify our analysis, we assume τ is no larger than a constant C. For $i\in[N]$ and $h\neq i, h\in[N]$, the following conclusions hold:

- $\mathbb{E}(z_{(ih)}v(h))=0$ because $h\neq i$.
- $|z_{(ih)}v(h)|\leq C\|v\|_\infty$.

- Set $\sigma^2 = \sum_{h\neq i, h\in[N]} \mathbb{E}(z^2_{(ih)}v^2(h))$. Under our multi-layer GoM model, we have

$$\begin{aligned}
\sum_{h\neq i,h\in[H]}\mathbb{E}[z^2_{(ih)}v^2(h)] &= \sum_{h\neq i,h\in[N]} v^2(h)\mathbb{E}[z^2_{(ih)}] \\
&= \sum_{h\neq i,h\in[N]} v^2(h)\sum_{l\in[L]}\sum_{j\in[J]}\mathbb{E}((R_l(i,j)R_l(h,j)-\mathscr{R}_l(i,j)\mathscr{R}_l(h,j))^2) \\
&= \sum_{h\neq i,h\in[N]} v^2(h)\sum_{l\in[L]}\sum_{j\in[J]}(\mathbb{E}(R^2_l(i,j))\mathbb{E}(R^2_l(h,j))-\mathscr{R}^2_l(i,j)\mathscr{R}^2_l(h,j)) \\
&= \sum_{h\neq i,h\in[N]} v^2(h)\sum_{l\in[L]}\sum_{j\in[J]}((\mathrm{Var}(R_l(i,j))+\mathscr{R}^2_l(i,j))(\mathrm{Var}(R_l(h,j)) \\
&\quad +\mathscr{R}^2_l(h,j))-\mathscr{R}^2_l(i,j)\mathscr{R}^2_l(h,j)) \\
&= \sum_{h\neq i,h\in[N]} v^2(h)\sum_{l\in[L]}\sum_{j\in[J]}(\mathrm{Var}(R_l(i,j))\mathrm{Var}(R_l(h,j))+\mathscr{R}^2_l(i,j)\mathrm{Var}(R_l(h,j)) \\
&\quad +\mathscr{R}^2_l(h,j)\mathrm{Var}(R_l(i,j))) \\
&\leq \sum_{h\neq i,h\in[N]} v^2(h)\sum_{l\in[L]}\sum_{j\in[J]}(\mathrm{Var}(R_l(i,j))\mathrm{Var}(R_l(h,j))+\rho^2\mathrm{Var}(R_l(h,j)) \\
&\quad +\rho^2\mathrm{Var}(R_l(i,j))) \\
&= \sum_{h\neq i,h\in[N]} v^2(h)\sum_{l\in[L]}\sum_{j\in[J]}(M\frac{\mathscr{R}_l(i,j)}{M}(1-\frac{\mathscr{R}_l(i,j)}{M})M\frac{\mathscr{R}_l(h,j)}{M}(1-\frac{\mathscr{R}_l(h,j)}{M}) \\
&\quad +\rho^2M\frac{\mathscr{R}_l(h,j)}{M}(1-\frac{\mathscr{R}_l(h,j)}{M})+\rho^2M\frac{\mathscr{R}_l(i,j)}{M}(1-\frac{\mathscr{R}_l(i,j)}{M})) \\
&\leq \sum_{h\neq i,h\in[N]} v^2(h)\sum_{l\in[L]}\sum_{j\in[J]}(\rho^2+2\rho^3) = (\rho^2+2\rho^3)JL\|v\|^2_F.
\end{aligned}$$

According to the Bernstein inequality in Theorem 1.4 Tropp (2012), for any

$t > 0$, we obtain

$$\mathbb{P}(|F_{(i)}| \geq t) \leq \exp(\frac{-t^2/2}{(\rho^2 + 2\rho^3)\|v\|_F^2 JL + \frac{C\|v\|_\infty t}{3}}).$$

Set t as $\sqrt{(\rho^2 + 2\rho^3\|v\|_F^2 JL\log(N+J+L)} \times \frac{\sqrt{\alpha+1}+\sqrt{\alpha+19}}{3}\sqrt{\alpha+1}$ for any $\alpha \geq 0$. Assuming that $(\rho^2 + 2\rho^2)\|v\|_F^2 JL \geq C^2\|v\|_\infty^2 \log(N+J+L)$ is satisfied, we get:

$$\mathbb{P}(|F_{(i)}| \geq t) \leq \exp(-(\alpha+1)\log(N+J+L)\frac{1}{\frac{18}{(\sqrt{\alpha+1}+\sqrt{\alpha+19})^2} + \frac{2\sqrt{\alpha+1}}{\sqrt{\alpha+1}+\sqrt{\alpha+19}}\sqrt{\frac{C^2\|V\|_\infty^2\log(N+J+L)}{(\rho^2+2\rho^3)\|v\|_F^2 JL}}})$$

$$\leq \frac{1}{(N+J+L)^{\alpha+1}}.$$

Setting $v \in \{-1,1\}^{(N-1)\times 1}$ and $\alpha = 3$ gives: when $(\rho^2 + 2\rho^3)(N-1)JL \gg \log(N+J+L)$, which is equivalent to $\rho^2 NJL \gg \log(N+J+L)$ since we allow ρ to approach zero and $N - 1 \approx N$ for large N, with probability $1 - o(\frac{1}{(N+J+L)^3})$,

$$\max_{i\in[N]}|F_{(i)}| = O(\sqrt{\rho^2 NJL\log(N+J+L)}).$$

Hence, we have $\|S - \mathcal{S}\|_\infty = O(\sqrt{\rho^2 NJL\log(N+J+L)}) + \rho^2 JL$. □

Theorem 4.2 (Cape et al., 2019) says that when $|\lambda_K(\mathcal{S})| \geq 4\|S - \mathcal{S}\|_\infty$ is satisfied,

$$\|\hat{U} - U\mathcal{O}\|_{2\to\infty} \leq 14\frac{\|S - \mathcal{S}\|_\infty \|U\|_{2\to\infty}}{|\lambda_K(\mathcal{S})|},$$

where $\mathcal{O}$ is an orthogonal matrix. Setting $\varpi := \|\hat{U}\hat{U}' - UU'\|_{2\to\infty}$ gives

$$\varpi \leq 2\|\hat{U} - U\mathcal{O}\|_{2\to\infty} \leq 14\frac{\|S - \mathcal{S}\|_\infty \|U\|_{2\to\infty}}{|\lambda_K(\mathcal{S})|}.$$

Combining Condition 1 with Lemma 3.1 (Mao et al., 2021) gives $\|U\|_{2\to\infty} = O(\sqrt{\frac{1}{N}})$. Thus,

$$\varpi = O(\frac{\|S - \mathcal{S}\|_\infty}{|\lambda_K(\mathcal{S})|\sqrt{N}}).$$

For $|\lambda_K(\mathcal{S})|$, we have $|\lambda_K(\mathcal{S})| \geq \rho^2 \lambda_K(\Pi'\Pi)|\lambda_K(\sum_{l\in[L]} B_l' B_l)|$, which gives

$$\varpi = O(\frac{\|S - \mathcal{S}\|_\infty}{\rho^2\sqrt{N}\lambda_K(\Pi'\Pi)|\lambda_K(\sum_{l\in[L]} B_l' B_l)|}).$$

By comparing the steps of our GoM-DSoG algorithm with that of the Algorithm 1 (Mao et al., 2021) in the task of estimating the mixed membership matrix Π, we find that all of their steps are the same except that the eigenvector matrix $\hat{U}$ is obtained from S in GoM-DSoG while it is from the adjacency matrix in Algorithm 1 in (Mao et al., 2021), where we do not consider the prune step in (Mao et al., 2021). Therefore, by Equation (3) of 3.2 (Mao et al., 2021), Condition 1, Assumption 2, and Lemma 1, we get

$$\begin{aligned}\max_{i\in[N]}\|e_i'(\hat{\Pi} - \Pi\mathcal{P})\|_1 &= O(\varpi\kappa(\Pi'\Pi)\sqrt{\lambda_1(\Pi'\Pi)}) = O(\frac{\|S - \tilde{S}\|_\infty}{\rho^2 NJL}) \\ &= O(\sqrt{\frac{\log(N+J+L)}{\rho^2 NJL}}) + O(\frac{1}{N}),\end{aligned}$$

where $\kappa(\cdot)$ denotes the condition number. □

S1.4 Proof of Lemma 2

Proof. Let $\Pi(\Pi'\Pi)^{-1}=G$ and $\hat{\Pi}(\hat{\Pi}'\hat{\Pi})^{-1}=\hat{G}$. By Condition 1, we get

$$\begin{aligned}\|\sum_{l\in[L]}(\hat{\Theta}_l-\Theta_l\mathcal{P})\| &= \|(\sum_{l\in[L]}R_l)'\hat{G}-(\sum_{l\in[L]}\mathscr{R}_l)'G\mathcal{P}\| \leq \|(\sum_{l\in[L]}(R_l-\mathscr{R}_l))'\hat{G}\|+\|(\sum_{l\in[L]}\mathscr{R}_l)'(\hat{G}-G\mathcal{P})\| \\ &\leq \|\sum_{l\in[L]}(R_l-\mathscr{R}_l)\|\|\hat{G}\|+\rho\|\Pi\|\|\sum_{l\in[L]}B_l\|\|\hat{G}-G\mathcal{P}\| \\ &= O(\frac{\|\sum_{l\in[L]}(R_l-\mathscr{R}_l)\|}{\sqrt{N}})+\rho|\lambda_1(\sum_{l\in[L]}B_l)|.\end{aligned}$$

According to the first statement of Lemma 2 (Qing, 2024), assume that $\rho L\max(N,J)\gg\log(N+J+L)$, with probability $1-o(\frac{1}{(N+J+L)^3})$, we have $\|\sum_{l\in[L]}(R_l-\mathscr{R}_l)\|=O(\sqrt{\rho L\max(N,J)\log(N+J+L)})$. When $|\lambda_1(\sum_{l\in[L]}B_l)|\geq c_2\sqrt{J}L$, we have

$$\begin{aligned}\frac{\|\sum_{l\in[L]}(\hat{\Theta}_l-\Theta_l\mathcal{P})\|_F}{\|\sum_{l\in[L]}\Theta_l\|_F} &\leq \frac{\sqrt{K}\|\sum_{l\in[L]}(\hat{\Theta}_l-\Theta_l\mathcal{P})\|}{\rho|\lambda_1(\sum_{l\in[L]}B_l)|}=O(\frac{\|\sum_{l\in[L]}(R_l-\mathscr{R}_l)\|}{\rho L\sqrt{NJ}}) \\ &= O(\sqrt{\frac{\max(N,J)\log(N+J+L)}{\rho NJL}}).\end{aligned}$$

□

S2 MATLAB codes of GoM-DSoG

The MATLAB codes of GoM-DSoG are provided below:

```
function [Pi_hat,Theta_all_hat]=DSoG(R_all,K)
% An implementation of GoM-DSoG algorithm
```

```
% Inputs:
%        R_all: a tensor with R_all (:,:, l) being the N*J matrix of layer
% l, where N and J denote the number subjects and items, respectively
%        K: the number of latent classes
% Outputs:
%        Pi_hat: the N*K estimated membership matrix
%        Theta_all_hat: a tensor with Theta_all_hat(:,:,l) being the J*K
% estimated item parameter matrix.
N=size(R_all,1);J=size(R_all,2);L=size(R_all,3);M=max(max(max(R_all)));
S=zeros(N,N);
for l=1:L
    Rl=R_all(:,:,l);Dl=sum(Rl.^2,2);Dl=diag(Dl);S=S+Rl*Rl'-Dl;
end
[U,~]=eigs(S,K);[ pure ]=SPA(U,K);C=U(pure,:);Pi_hat=U/C;
Pi_hat=max(0,Pi_hat);
for i=1:N
    Pi_hat(i,:)=Pi_hat(i,:)/sum(Pi_hat(i,:));
end
Theta_all_hat=zeros(J,K,L);
for l=1:L
```

```
    Rl=R_all(:,:,l);
    theta_l_hat=min(M,max(0,Rl'*Pi_hat*inv(Pi_hat'*Pi_hat)));
    Theta_all_hat(:,:,l)=theta_l_hat;
end


function [ pure ]=SPA(X,K)
% An implementation of SPA algorithm
% Input: X: n by K data matrix (eigenvectors in our setting),where
% n is the number of nodes, K is the vertices
% Output: pure: set of pure subjects indices
pure = [];
row_norm = vecnorm(X').^2';
for i = 1:K
    [~,idx_tmp] = max(row_norm);pure = [pure idx_tmp];
    U(i,:) = X(idx_tmp,:);
    for j = 1 : i-1
        U(i,:) = U(i,:) - U(i,:)*(U(j,:)'*U(j,:));
    end
    U(i,:) = U(i,:)/norm(U(i,:)); row_norm = row_norm - (X*U(i,:)').^2;
end
```

end

References

Cape, J., M. Tang, and C. E. Priebe (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Annals of Statistics 47*(5), 2405–2439.

Mao, X., P. Sarkar, and D. Chakrabarti (2021). Estimating mixed memberships with sharp eigenvector deviations. *Journal of the American Statistical Association 116*(536), 1928–1940.

Qing, H. (2024). Latent class analysis for multi-layer categorical data. *arXiv preprint arXiv:2408.05535*.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics 12*, 389–434.