Statistica Sinica Preprint No: SS-2024-0276	
Title	Grade of Membership Analysis for Multi-Layer Ordinal Categorical Data
Manuscript ID	SS-2024-0276
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0276
Complete List of Authors	Huan Qing
Corresponding Authors	Huan Qing
E-mails	qinghuan@u.nus.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

Grade of membership analysis for multi-layer ordinal categorical data

Huan Qing

School of Economics and Finance, Chongqing University of Technology

Abstract: Consider a group of individuals (subjects) participating in the same psychological tests with numerous questions (items) at different times, where the choices of each item have an implicit ordering. The observed responses can be recorded in multiple response matrices over time, named multi-layer ordinal categorical data, where layers refer to time points. Assuming that each subject has a common mixed membership shared across all layers, enabling it to be affiliated with multiple latent classes with varying weights, the objective of the grade of membership (GoM) analysis is to estimate these mixed memberships from the data. When the test is conducted only once, the data becomes traditional single-layer ordinal categorical data. The GoM model is a popular choice for describing single-layer categorical data with a latent mixed membership structure. However, GoM cannot handle multi-layer ordinal categorical data. In this work, we propose a new model, multi-layer GoM, which extends GoM to multi-layer ordinal categorical data. To estimate the common mixed memberships, we propose a new approach, GoM-DSoG, based on a debiased sum of Gram matrices. We establish GoM-DSoG's per-subject convergence rate under the multi-layer GoM model. Our theoretical results suggest that fewer no-responses, more subjects, more items, and more layers are beneficial for GoM analysis. We also propose an approach to select the number of latent classes. Extensive experimental studies verify the theoretical findings and show GoM-DSoG's superiority over its competitors, as well as the accuracy of our method in determining the number of latent classes.

Key words and phrases: Multi-layer ordinal categorical data, multi-layer grade of membership model, debiased spectral clustering

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

1. Introduction

Ordinal categorical data is extensively collected in psychological tests, educational assessments, and political surveys (Sloane and Morgan, 1996; Agresti, 2012; Chen et al., 2019; Shang et al., 2021). Such data involves responses that follow a natural and logical order sequence, where the choices can be meaningfully ranked, though the intervals between them may not be precisely measurable or equal. Examples of ordinal categorical data include Likert scale items (strongly disagree/disagree/neutral /agree/strongly agree) or disagree/agree in psychological tests, education levels (high school/ bachelor's degree/master's degree/doctorate) in educational research, and medical assessment scales (mild/moderate/severe) in clinical studies.

Without confusion, we will use categorical data to represent ordinal categorical data throughout the remainder of this paper. This type of data typically comprises responses from subjects to a set of items and is often represented by an $N \times J$ observed response matrix R, where N denotes the number of subjects (individuals) participating in the test and J represents the number of items (questions). In this matrix, $R(i, j)$ signifies the response of subject i to item j. When each item has only two choices (e.g., yes/no choices or right/wrong answers), the data is referred to as categorical data with binary responses. In this case, $R \in \{0, 1\}^{N \times J}$, where 0 and 1 are used to indicate the binary choices. When each item has M ($M \geq 3$) choices, the data is known as categorical data with polytomous responses. For such data,we consider $R \in \{0, 1, \ldots, M\}^{N \times J}$, where integer m denotes the m-th ordered choice for $m = 0, 1, \ldots, M$, with 0 representing the lowest intensity response. In this work, we do not consider missing data. If the original dataset contains missing responses (typically coded as NaN), we recommend removing those subjects with missing values during

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

data preprocessing. It is evident that polytomous responses are more general than binary responses, and thus, this paper focuses on polytomous responses. A key characteristic of traditional categorical data is that it is often collected only once for each subject, resulting in a single-layer categorical dataset. For such data, learning the latent classes (subgroups) of subjects is often of great interest, as subjects with similar subgroup information typically exhibit similar response patterns. Specifically, the task of learning subjects' latent classes is known as latent class analysis (LCA) (Hagenaars and McCutcheon, 2002; Nylund-Gibson and Choi, 2018), where the latent classes can be different personalities, abilities, and political ideologies in psychological tests, educational assessments, and political surveys (Chen and Gu, 2024), respectively. Therefore, equipped with LCA, categorical data plays a key role in understanding human behaviors, preferences, and attitudes.

However, despite its prevalence, the single-layer structure of traditional categorical data inherently limits the depth of analysis that can be performed, as it captures only a snapshot of the subjects' responses at a single point in time. In many practical scenarios, the same group of subjects may be assessed multiple times over different time points using the same set of items. This repeated assessment generates multi-layer categorical data, with each layer representing a different time point. For such data, all layers share common subjects and items, creating a longitudinal dataset rich in information. The additional temporal dimension offered by multi-layer categorical data provides a more comprehensive view of subjects' responses, potentially revealing patterns and trends that are not discernible in single-layer categorical data. Figure 1 illustrates a simple example of multi-layer categorical data, where subjects' responses exhibit variation across layers.

For single-layer categorical data, a popular statistical model for describing its latent class

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

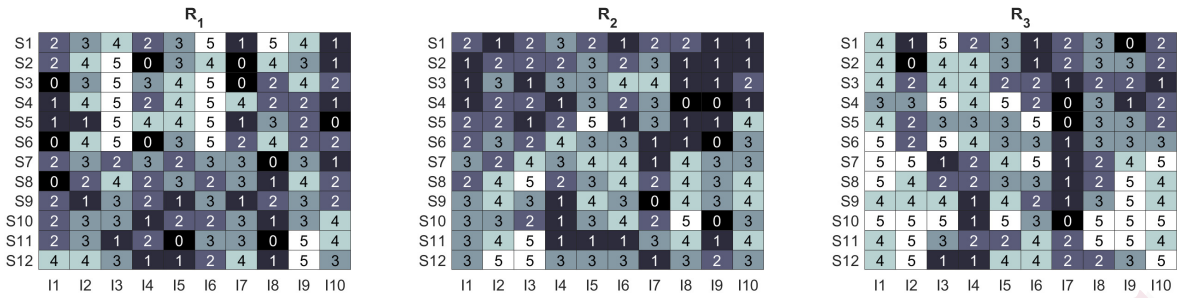


Figure 1: A toy example of multi-layer categorical data with 12 subjects, 10 items, 3 layers, and 6 choices per item, where we use Si, Ij, R_l to represent the i-th subject, the j-th item, and the l-th observed response matrix, respectively, for $i = 1, 2, \ldots, 12, j = 1, 2, \ldots, 10$, and $l = 1, 2, 3$.

structure is the latent class model (LCM) (Goodman, 1974), which operates under the assumption that each subject is assigned to a single latent class. However, the non-overlapping classes property of LCM can be restrictive, potentially failing to fully capture the complexity and diversity of subjects' response patterns. The grade of membership (GoM) model (Woodbury et al., 1978) overcomes LCM's limitation and allows each subject to belong to at least one latent class, providing a more powerful modeling capacity than LCM. Existing approaches for learning subjects' latent mixed memberships in single-layer categorical data modeled by the GoM encompass Bayesian inference utilizing MCMC (Erosheva et al., 2007; Gormley and Murphy, 2009; Gu et al., 2023), joint maximum likelihood algorithms (Robitzsch, 2023), and spectral methods (Chen and Gu, 2024; Qing, 2024a, 2025c).

However, all the above methods are inherently limited to single-layer categorical data and cannot be directly applied to learning the common latent classes of subjects in multi-layer categorical data. The additional temporal dimension in multi-layer categorical data introduces new challenges and opportunities for latent class analysis. Recently, to describe

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

the latent class structures in multi-layer categorical data, Qing (2024b) proposed a multi-layer LCM which assumes that each subject belongs to only one latent subgroup. Qing (2024b) also provides several spectral methods with theoretical guarantees to discover the latent classes of subjects. However, one main limitation of the multi-layer LCM is that it cannot model multi-layer categorical data with a latent mixed membership structure, where each subject can belong to different subgroups with varying weights.

To address the limitations of the aforementioned models and methods for multi-layer categorical data with a latent mixed membership structure, this work introduces the multi-layer grade of membership (multi-layer GoM) model, which extends the traditional GoM model to multi-layer categorical data. This extension enables us to capture the temporal dynamics of subject responses provided by multiple layers and allows subjects to belong to more than one latent class simultaneously. For the estimation of the common mixed memberships, we develop a novel approach, GoM-DSoG, based on a debiased sum of Gram matrices. This approach effectively handles the challenges posed by multi-layer data. We also establish the per-subject convergence rate of GoM-DSoG under the proposed model. Our theoretical and experimental results highlight the benefits of considering more layers than a single layer in the task of grade of membership analysis for categorical data. In particular, if there are too many zeros in the categorical data, more layers should be considered, and our GoM-DSoG is an efficient tool to estimate subjects' common mixed memberships for such multi-layer categorical data. Additionally, an approach for selecting the number of latent classes K in multi-layer categorical data generated from the proposed model is developed. By conducting substantial experimental studies, we empirically verify the theoretical results and show the superiority of the proposed GoM-DSoG method over its competitors and the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

high accuracy of our method in choosing K.

Notation: $I_{m\times m}$ and $[m]$ represent the $m \times m$ identity matrix and set $\{1, 2, \ldots, m\}$, respectively. For any matrix H, H' denotes its transpose, $H(s, :)$ denotes its submatrix for rows in the set s, $\|H\|_1$ is its the l_1 norm, $\|H\|_F$ represents its Frobenius norm, $\|H\|$ means its spectral norm, $\|H\|_\infty$ represents its maximum absolute row sum, $\|H\|_{2\to\infty}$ denotes its maximum row-wise l_2 norm, $\lambda_k(H)$ denotes its k-th largest eigenvalue in magnitude, and rank(H) denotes its rank. e_i is a vector satisfying $e_i(i) = 1$ and $e_i(j) = 0$ for $j \neq i$. $\mathbb{E}(\cdot)$ denotes expectation and $\mathbb{P}(\cdot)$ represents probability.

We organize the remainder of this paper as follows. We introduce our model in Section 2. Our method for fitting proposed model and its theoretical guarantees are presented in Section 3. Section 4 proposes an approach for determining K. Section 5 conducts experimental studies. Section 6 concludes the paper.

2. Multi-layer grade of membership model

2.1 Data preprocessing

Before introducing the multi-layer GoM model, we first describe the preprocessing steps applied to the original response matrices to ensure that the responses are appropriately encoded for analysis. Let R_l^{original} denote the l-th layer's original response matrix, which may contain missing values (e.g., NaN). We apply the following preprocessing steps:

1. Removal of subjects with missing responses: For each subject i, if any entry $R_l^{\text{original}}(i, j)$ is missing (e.g., NaN), we remove the entire subject from the analysis. This ensures that the resulting dataset contains only complete responses.

2. Recoding of response values: For each non-missing response $R_l^{\text{original}}(i, j)$, we recode

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

the values to ensure they lie in the set $\{0, 1, \ldots, M\}$, where 0 represents the lowest intensity of response. Specifically:

- If the original responses are in $\{0, 1, \ldots, M\}$, we set $R_l(i, j) = R_l^{\text{original}}(i, j)$. Here, 0 denotes the lowest intensity response.
- If the original responses are in $\{1, 2, \ldots, M+1\}$, we set $R_l(i, j) = R_l^{\text{original}}(i, j) - 1$. This transformation shifts the responses to $\{0, 1, \ldots, M\}$, preserving the ordinal nature of the data without loss of information.

After preprocessing, the resulting observed response matrix R_l for the l-th layer has elements in $\{0, 1, \ldots, M\}$, where 0 represents the lowest response intensity (e.g., "strongly disagree" on a Likert scale), and not a missing value. This encoding allows us to model the responses using a Binomial distribution as described in Definition 1 given later.

2.2 The model

Suppose that all subjects belong to K common latent classes, denoted as:

$$\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K. \tag{2.1}$$

We assume that K is known in this paper and we will also propose an approach for selecting K in Section 4. The memberships of subjects belonging to different latent classes can be characterized by a common membership matrix $\Pi \in [0, 1]^{N \times K}$ shared across all layers, where:

$$\Pi(i, k) \text{ is the "weight" for the partial membership of subject } i \text{ on } \mathcal{C}_k, \quad i \in [N], k \in [K]. \tag{2.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

By Π's definition, for $i \in [N]$, we have $\sum_{k\in[K]} \Pi(i,k) = 1$. Call subject i pure if one entry of $\Pi(i,:)$ equals 1, and mixed otherwise. In this paper, we assume that:

$$\text{There is at least one pure subject in each latent class.} \tag{2.3}$$

Since Equations (2.1)–(2.3) serve as the foundational assumptions of our model, we adopt them as default conditions throughout this paper. Combining Equations (2.2) and (2.3) implies that Π is full rank, and its rank is K since we assume $K \ll \min(N, J)$ in this paper. Let p_k denote the index of any pure subject in the k-th latent class $\mathcal{C}_k$ for $k \in [K]$. Define the pure subject index set $\mathcal{I}$ as $\mathcal{I} = \{p_1, p_2, \ldots, p_K\}$. Without loss of generality, similar to the analysis in Mao et al. (2021), reorder all subjects such that $\Pi(\mathcal{I}, :) = I_{K\times K}$.

For each layer $l \in [L]$, let $\Theta_l \in [0, M]^{J\times K}$ denote the item parameter matrix of the l-th layer categorical data. Our model, a multi-layer extension of the classical GoM model, is formally defined as follows.

Definition 1. *(Multi-layer grade of membership model) Let $R_l \in \{0, 1, \ldots, M\}^{N\times J}$ represent the observed response matrix of the l-th layer categorical data for $l \in [L]$. Our multi-layer grade of membership (multi-layer GoM) model for generating the L observed response matrices $\{R_l\}_{l=1}^{L}$ is defined as:*

$$\mathscr{R}_l := \Pi\Theta_l', \quad \textit{where } R_l(i,j) \sim \textit{Binomial}(M, \frac{\mathscr{R}_l(i,j)}{M}), \quad i \in [N], j \in [J], l \in [L]. \tag{2.4}$$

In our model, $\mathscr{R}_l(i,j)$ represents the expected choice of subject i to item j by the property of Binomial distribution, i.e., $\mathbb{E}(R_l(i,j)) = \mathscr{R}_l(i,j)$ under the multi-layer GoM. Equation (2.4) gives $\mathscr{R}_l(i,j) = \sum_{k\in[K]} \Pi(i,k)\Theta_l(j,k)$, which means that the response of subject i to item j tends to be larger for larger $\Theta_l(j,k)$. For any two distinct latent classes k_1 and k_2, if $\Theta_l(j,k_1) > \Theta_l(j,k_2)$, $\mathscr{R}_l(i,j) = \sum_{k\in[K]} \Pi(i,k)\Theta_l(j,k)$ implies that subjects

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

predominantly affiliated with latent class k_1 (i.e., those with higher $\Pi(i, k_1)$) are more likely to choose higher responses for item j at layer l, while subjects in latent class k_2 exhibit weaker or lower responses for the same item. Meanwhile, the choice of modeling categorical responses as Binomial random variables is motivated by the ordered nature of the data. Unlike unordered categorical data (e.g., multiple-choice questions with no inherent order), ordinal categorical data (e.g., Likert scales) can be meaningfully interpreted as counts of "successes" or levels of agreement. The Binomial distribution provides a flexible framework to model such ordered responses, where the probability of each response value is determined by the latent membership matrix Π and item parameter matrix Θ_l for the l-th layer. For convenience, call $\mathscr{R}_l$ the l-th population response matrix since $\mathbb{E}(R_l) = \mathscr{R}_l$ for $l \in [L]$ under the Binomial distribution. When all subjects are pure, our model degenerates the multi-layer LCM introduced in (Qing, 2024b). When the number of layers L equals 1, the multi-layer GoM defined by Equation (2.4) simplifies to the GoM model for single-layer categorical data considered in (Qing, 2024a). Furthermore, when both $L = 1$ and $M = 1$, the proposed model reduces to the GoM model studied in (Chen and Gu, 2024).

Equation (2.4) necessitates that the elements of Θ_l range within $[0, M]$ for $l \in [L]$ because $\frac{\mathscr{R}_l(i,j)}{M}$ represents a probability within $[0, 1]$, and Π satisfies Equation (2.2). By the definition of multi-layer GoM, at the l-th layer, the probability that subject i selects the m-th choice in item j equals

$$\mathbb{P}(R_l(i,j) = m) = \frac{M!}{m!(M-m)!}(\frac{\mathscr{R}_l(i,j)}{M})^m(1 - \frac{\mathscr{R}_l(i,j)}{M})^{M-m}, \qquad m = 0, 1, \ldots, M, \quad (2.5)$$

where $\mathbb{P}(R_l(i,j) = 0)$ represents the probability that subject i tends to have the lowest intensity response to item j. By Equation (2.4), it is clear that the membership matrix Π is shared by all layers while the probability that subject i selects the m-th choice in item

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

j changes between layers in our model. Such setting is common for the task of community detection in multi-layer networks (Paul and Chen, 2020; Lei et al., 2020; Lei and Lin, 2023; Xu et al., 2023; Su et al., 2024; Qing, 2025a,b). Using Equation (2.4), our multi-layer GoM can generate $\{R_l\}_{l=1}^L$ with a latent true membership matrix Π and item parameter matrices $\{\Theta_l\}_{l=1}^L$. This paper aims to recover Π and $\{\Theta_l\}_{l=1}^L$ given the L observed response matrices $\{R_l\}_{l=1}^L$.

2.3 Response intensity scaling

In real-world multi-layer categorical data with polytomous responses, some lowest-intensity responses (0s) exist. Specifically, for multi-layer categorical data with binary responses when $M = 1$, we have $R_l \in \{0,1\}^{N\times J}$ for $l \in [L]$, which may result in numerous zeros in the data. A large number of zeros in the dataset poses a significant challenge for the task of grade of membership analysis. Therefore, it is crucial to control the number of zeros in the multi-layer categorical data. Define $\rho = \max_{l\in[L]}\max_{j\in[J],k\in[K]}\Theta_l(j,k)$ and $B_l = \frac{\Theta_l}{\rho}$ for $l \in [L]$. Given that $\Theta_l \in [0,M]^{J\times K}$, it follows that $\rho \in (0,M]$, the maximum element of B_l is no larger than 1, and $\mathscr{R}_l(i,j) \le \rho$ for $i \in [N], j \in [J], l \in [L]$. According to Equation (2.5), we have:

$$\mathbb{P}(R_l(i,j)=0) = (1-\frac{\mathscr{R}_l(i,j)}{M})^M = (1-\frac{\rho\Pi(i,:)B_l'(j,:)}{M})^M,$$

which indicates that increasing ρ decreases the probability of generating the lowest-intensity responses across all layers. Consequently, ρ serves as a controlling factor for the overall response intensity of the data, thus justifying its designation as the response intensity parameter. To thoroughly examine the performance of the proposed method under varying levels of data's response intensity, we will permit ρ to approach zero and incorporate it into the error bound for further analysis.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

Example 1. *Consider a simple case where $M = 4$, and the expected response $\mathscr{R}_l(i,j)$ takes values from $\{1, 1.5, 2, 2.5, 3\}$. Table 1 illustrates how the probability of each response value changes based on $\mathscr{R}_l(i,j)$. The results show that as $\mathscr{R}_l(i,j)$ increases, the probability of a lowest intensity response (0) decreases monotonically, while the probability of higher intensity response (e.g., 3 or 4) increases. This pattern aligns with theoretical expectations, as $\mathscr{R}_l(i,j)$ represents the expected value of $R_l(i,j)$ under the proposed model. Intuitively, higher values of $\mathscr{R}_l(i,j)$ lead to fewer zeros and shift the response distribution toward higher-valued choices. Consequently, by adjusting the response intensity parameter ρ, we can effectively control the overall response intensity of the data.*

Table 1: Probability distribution of responses under varying population response values $\mathscr{R}_l(i,j)$ when $M = 4$.

Population response values	$\mathbb{P}(R_l(i,j)=0)$	$\mathbb{P}(R_l(i,j)=1)$	$\mathbb{P}(R_l(i,j)=2)$	$\mathbb{P}(R_l(i,j)=3)$	$\mathbb{P}(R_l(i,j)=4)$
$\mathscr{R}_l(i,j)=1$	0.3164	0.4219	0.2109	0.0469	0.0039
$\mathscr{R}_l(i,j)=1.5$	0.1526	0.3662	0.3296	0.1318	0.0198
$\mathscr{R}_l(i,j)=2$	0.0625	0.2500	0.3750	0.2500	0.0625
$\mathscr{R}_l(i,j)=2.5$	0.0198	0.1318	0.3296	0.3662	0.1526
$\mathscr{R}_l(i,j)=3$	0.0039	0.0469	0.2109	0.4219	0.3164

3. Debiased spectral method and its theoretical guarantees

We propose an efficient method for Π and $\{\Theta_l\}_{l=1}^L$ in this section. Subsequently, we establish its theoretical guarantees under the multi-layer GoM model.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

3.1 Debiased spectral method

Suppose that the population response matrices $\{\mathscr{R}_l\}_{l=1}^L$ are observed, and our objective is to accurately recover Π and $\{\Theta_l\}_{l=1}^L$ from $\{\mathscr{R}_l\}_{l=1}^L$. Define $\mathcal{S} = \sum_{l=1}^L \mathscr{R}_l\mathscr{R}_l'$, where $\mathcal{S}$ is a symmetric matrix and it represents an aggregation of L Gram matrices $\{\mathscr{R}_l\mathscr{R}_l'\}_{l=1}^L$. The following proposition guarantees the identifiability of our multi-layer GoM model.

Proposition 1. *(Identifiability). Assume that* $\text{rank}(\sum_{l=1}^L \Theta_l'\Theta_l) = K$. *Then, our multi-layer GoM model is identifiable: For any* $(\Pi, \{\Theta_l\}_{l=1}^L)$ *and* $(\tilde{\Pi}, \{\tilde{\Theta}_l\}_{l=1}^L)$, *if* $\Pi\Theta_l' = \tilde{\Pi}\tilde{\Theta}_l'$ *for* $l \in [L]$, *we have* $\tilde{\Pi} = \Pi\mathcal{P}$ *and* $\tilde{\Theta}_l = \Theta_l\mathcal{P}$ *for* $l \in [L]$, *where* $\mathcal{P}$ *is a permutation matrix.*

The identifiability proposed in Proposition 1 guarantees that our multi-layer GoM is well-defined since the model parameters $(\Pi, \{\Theta_l\}_{l=1}^L)$ can be uniquely recovered up to a permutation. The intuition of the proposed method comes from the following lemma.

Lemma 1. *When* $\text{rank}(\sum_{l=1}^L \Theta_l'\Theta_l) = K$, *let* $\mathcal{S} = U\Lambda U'$ *denote the leading* K *eigendecomposition of* $\mathcal{S}$, *where* $U'U = I_{K\times K}$. *Then, we have:*

- $U = \Pi U(\mathcal{I}, :)$,
- $\Theta_l = \mathscr{R}_l'\Pi(\Pi'\Pi)^{-1}$ *for* $l \in [L]$.

Note that Lemma 1 is directly linked to the pure subject assumption stated in Equation (2.3), as it necessitates the presence of the pure subject index set $\mathcal{I}$. Given that the summation of each row of Π equals 1 and Π satisfies Equation (2.3), the form $U = \Pi U(\mathcal{I}, :)$ is known as the Ideal Simplex (IS), where $U(\mathcal{I}, :)$'s K rows are the K vertexes in the simplex structure. Figure 2 shows the IS structure when $K = 3$. It is clear to observe that the rows corresponding to pure subjects form the vertices of the simplex, while those representing

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

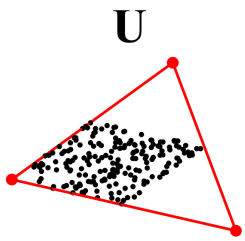


Figure 2: Illustration of the Ideal Simplex geometry of the eigenvector matrix U with $K = 3$. Each point denotes a row vector of U, red points denote rows corresponding to pure subjects (because $U(i,:) = U(j,:)$ for any two pure subjects i and j from the same latent class, so each red point may denote many pure subjects), each black points corresponds to mixed subjects, and the red triangle denotes the Ideal Simplex. For visualization, these points have been projected from $\mathbb{R}^3$ to $\mathbb{R}^2$.

mixed subjects lie within its interior. A similar Ideal Simplex is also discovered in the field of mixed membership community detection in network science (Mao et al., 2021; Jin et al., 2024; Qing and Wang, 2024), grade of membership analysis for single-layer categorical data (Chen and Gu, 2024; Qing, 2024a), and topic modeling (Klopp et al., 2023; Ke and Wang, 2024).

Since $U(\mathcal{I},:)$ is non-singular, we have $\Pi = UU^{-1}(\mathcal{I},:)$ by $U = \Pi U(\mathcal{I},:)$. This implies that Π can be exactly recovered if we can obtain $U(\mathcal{I},:)$ from the $N \times K$ eigenvector matrix U. Similar to previous works related to the IS structure, the *successive projection algorithm* (SPA) (Araújo et al., 2001; Gillis and Vavasis, 2013, 2015) can be employed to precisely identify the K vertices $U(\mathcal{I},:)$ within U, leveraging the simplex structure $U = \Pi U(\mathcal{I},:)$. SPA is a widely used technique for vertex hunting in simplex structure (Mao et al., 2021; Jin et al., 2024; Qing and Wang, 2024; Chen and Gu, 2024; Qing, 2024a; Agterberg and

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

Zhang, 2025; Klopp et al., 2023; Ke and Wang, 2024). Once we obtain Π, we can recover $\{\Theta_l\}_{l=1}^L$ using the second result of Lemma 1. To adapt this idea to the real case, we set $H = \max(0, UU^{-1}(\mathcal{I}, :)) \equiv \Pi$, ensuring that $\Pi(i,:) = \frac{H(i,:)}{\|H(i,:)\|_1}$ and $\Theta_l = \mathscr{R}_l'\Pi(\Pi'\Pi)^{-1}$ for $i \in [N], l \in [L]$.

For real data, the L response matrices $\{R_l\}_{l=1}^L$ are observed, while their expectations $\{\mathscr{R}_l\}_{l=1}^L$ are not available. For each $l \in [L]$, let D be an $N \times N$ diagonal matrix with $D_l(i,i) = \sum_{j\in[J]} R_l^2(i,j)$ for $i \in [N]$. Define $S = \sum_{l\in[L]}(R_l R_l' - D_l)$, where S is known as the debiased sum of Gram matrices (Su et al., 2024). Let $\hat{U}\hat{\Lambda}\hat{U}'$ be the leading K eigendecomposition of S, where $\hat{U}'\hat{U} = I_{K\times K}$. It is known that the $N \times N$ symmetric matrix S is a debiased estimator of $\mathcal{S}$, while $\sum_{l\in[L]} R_l R_l'$ is a biased estimator, following similar analysis as in (Lei and Lin, 2023; Su et al., 2024). Therefore, let $\hat{\mathcal{I}}$ represent the estimated pure subjects' index set found by employing SPA to $\hat{U}$ with K vertices. Define $\hat{H} = \max(0, \hat{U}\hat{U}^{-1}(\hat{\mathcal{I}}, :))$, let $\hat{\Pi}$ be an $N \times K$ matrix such that $\hat{\Pi}(i,:) = \frac{\hat{H}(i,:)}{\|\hat{H}(i,:)\|_1}$ for $i \in [N]$. Note that the elements of $R_l'\hat{\Pi}(\hat{\Pi}'\hat{\Pi})^{-1}$ may be negative or larger than M while $\Theta_l \in [0, M]^{J\times K}$, we define $\hat{\Theta}_l = \min(M, \max(0, R_l'\hat{\Pi}(\hat{\Pi}'\hat{\Pi})^{-1}))$ for $l \in [L]$. We see that $\hat{\Pi}$ and $\{\hat{\Theta}_l\}_{l=1}^L$ are good estimations of Π and $\{\Theta_l\}_{l=1}^L$, respectively. The above analysis is summarized in Algorithm 1. Recall that the membership scores remain constant across all layers. If we allow the membership scores to vary, Lemma 1 no longer holds. Consequently, Algorithm 1 becomes inapplicable for membership score estimation. For alternative approaches to this scenario, readers may consult methodologies developed in (Pensky and Zhang, 2019; Lin and Lei, 2024). This paper exclusively addresses the case where Π remains consistent across all layers while analyzing the time-varying membership scores is reserved for future research. Additionally, we employ spectral method rather than likelihood-based approaches for estimating membership scores,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

as spectral method exhibits superior computational efficiency (Lei and Lin, 2023; Chen and Gu, 2024).

Algorithm 1 **G**rade **o**f **M**embership analysis via the **D**ebiased **S**um **o**f **G**ram matrices (**GoM-DSoG**)

Require: The L observed response matrices $R_1, R_2, \ldots, R_L$ and the number of latent classes K, where $R_l \in \{0, 1, 2, \ldots, M\}^{N \times J}$ for $l \in [L]$.

Ensure: Estimated membership matrix $\hat{\Pi}$ and estimated item parameter matrices $\{\hat{\Theta}_l\}_{l=1}^{L}$.

1: Calculate $S = \sum_{l \in [L]} (R_l R_l' - D_l)$, where D_l is a diagonal matrix with $D_l(i,i) = \sum_{j \in [J]} R_l^2(i,j)$ for $i \in [N], l \in [L]$.

2: Compute $\hat{U}\hat{\Lambda}\hat{U}'$, the leading K eigendecomposition of S.

3: Apply SPA to $\hat{U}$ assuming that there are K vertices to get the estimated pure subjects' index set $\hat{\mathcal{I}}$.

4: Set $\hat{H} = \max(0, \hat{U}\hat{U}^{-1}(\hat{\mathcal{I}}, :))$.

5: For $i \in [N]$, set $\hat{\Pi}(i, :) = \frac{\hat{H}(i,:)}{\|\hat{H}(i,:)\|_1}$.

6: For $l \in [L]$, set $\hat{\Theta}_l = \min(M, \max(0, R_l' \hat{\Pi} (\hat{\Pi}' \hat{\Pi})^{-1}))$.

Remark 1. *In this paper, our multi-layer GoM model employs a Binomial distribution in Equation (2.4) for analyzing multi-layer ordinal categorical data. Notably, extending the community detection approaches for weighted networks proposed in (Qing and Wang, 2023, 2024), our framework allows $\mathcal{F}$ to represent any distribution whose first-moment exists and satisfies $\mathbb{E}(R_l(i,j)) = \mathscr{R}_l(i,j)$ with $\mathscr{R}_l := \Pi\Theta_l'$ for $i \in [N], j \in [J], l \in [L]$. This generalization enables the modeling of multi-layer weighted data. Below we present illustrative examples:*

- ***Normal distribution**: Let $\mathcal{F}$ be a Normal distribution such that $R_l(i,j) \sim \text{Normal}(\mathscr{R}_l(i,j), \sigma^2)$, where σ^2 denotes variance. For this distribution, $R_l(i,j)$ can be any real value, Θ_l's*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

elements can be negative, and $\rho \in (0, +\infty)$.

- ***Poisson distribution****: Let* $\mathcal{F}$ *be a Poisson distribution such that* $R_l(i,j) \sim \text{Poisson}(\mathscr{R}_l(i,j))$. *Here,* $R_l(i,j)$ *is a nonnegative integer,* Θ_l*'s elements should be nonnegative, and* $\rho \in (0, +\infty)$.

- ***Uniform distribution****: Let* $\mathcal{F}$ *be a Uniform distribution such that* $R_l(i,j) \sim \text{Uniform}(0, 2\mathscr{R}_l(i,j))$. *For this distribution,* $R_l(i,j)$ *is positive value that is no larger than* 2ρ*,* Θ_l*'s elements should be nonnegative, and* $\rho \in (0, +\infty)$.

- ***Discrete signed distribution****: Let* $\mathcal{F}$ *be a discrete distribution such that* $\mathbb{P}(R_l(i,j) = 1) = \frac{1+\mathscr{R}_l(i,j)}{2}$ *and* $\mathbb{P}(R_l(i,j) = -1) = \frac{1-\mathscr{R}_l(i,j)}{2}$. *For this case,* $R_l(i,j)$ *takes values in* $\{-1, 1\}$ *(i.e., signed responses),* Θ_l*'s elements should range in* $[-1, 1]$*, and* $\rho \in (0, 1]$.

$\mathcal{F}$ *can also be any other distribution as long as* $\mathbb{E}(R_l) = \mathscr{R}_l$ *holds under distribution* $\mathcal{F}$. *Crucially, Lemma 1 remains valid under these generalized distributions, enabling consistent estimation of* Π *and* $\{\Theta_l\}_{l=1}^L$ *using our GoM-DSoG algorithm. When different distributions* $\mathcal{F}$ *are considered, our multi-layer GoM can model multi-layer weighed categorical data because* R_l*'s elements have more choices now. While this extension broadens applicability to multi-layer weighted data through diverse response value domains, our current focus remains on multi-layer ordinal categorical data using the Binomial distribution. The theoretical investigation of GoM-DSoG's performance under different distributions* $\mathcal{F}$ *is of independent interest.*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

3.2 Consistency results

This subsection presents the theoretical findings exploring the advantages of multi-layer categorical data over single-layer categorical data for grade of membership analysis. We investigate the asymptotic consistency of GoM-DSoG under the proposed multi-layer GoM model across different intensity levels. Specifically, we assume that N, J, and L can approach infinity in the asymptotic setting, with no assumed relationship among their growth rates. Before presenting the main result, we outline some assumptions.

Assumption 1. $\rho^2 NJL \gg \log(N+J+L)$.

Since N, J, and L can grow to infinity in our analysis, Assumption 1 is mild, as it implies that ρ should be significantly larger than $\sqrt{\frac{\log(N+J+L)}{NJL}}$ (therefore, ρ can be sufficiently small for large values of N, J, and L). This assumption imposes a lower bound on ρ, indicating that there cannot be too many zeros in the multi-layer categorical data. It is worth noting that when $J = O(N)$, our Assumption 1 is in agreement with the sparsity requirement stated in Theorem 1 of Lei and Lin (2023). This alignment is expected since both our GoM-DSoG and the Bias-adjusted SoS algorithm provided in Lei and Lin (2023) are debiased spectral clustering methods.

Assumption 2. $|\lambda_K(\sum_{l\in[L]} B_l' B_l)| \geq c_1 JL$ *for some positive constant* c_1.

This assumption requires a linear growth of $|\lambda_K(\sum_{l\in[L]} B_l' B_l)|$ concerning the number of layers L and it serves a similar purpose in our theoretical analysis as the second point of Assumption 1 in Lei and Lin (2023), Assumption 2 in Su et al. (2024), and the last point of Condition 2 in Chen and Gu (2024). The rationality of this assumption can be illustrated by considering a simple case where $B_1 = B_2 = \ldots = B_L$. Under this scenario, it follows

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

that $|\lambda_K(\sum_{l\in[L]} B_l'B_l)| = L|\lambda_K(B_1'B_1)| = O(JL)$, which demonstrates the validity of the assumption in a straightforward manner.

Condition 1. *$K = O(1)$ and $\lambda_K(\Pi'\Pi) = O(\frac{N}{K})$.*

This condition serves a similar purpose to the first point of Assumption 1 in Lei and Lin (2023), Condition 2 in Chen and Gu (2024), Assumption 1 in Su et al. (2024), and so on. It is necessary to simplify our theoretical analysis.

Theorem 1 presents our main result for the task of grade membership analysis. It provides the per-node error rate of the proposed algorithm. To obtain Theorem 1, we first bound $\|S - \mathcal{S}\|_\infty$ using the Bernstein inequality provided in Tropp (2012). Subsequently, we utilize this bound to establish a bound for the row-wise eigenvector error $\|\hat{U}\hat{U}' - UU'\|_{2\to\infty}$ based on Theorem 4.2 of Cape et al. (2019).

Theorem 1. *Consider a multi-layer GoM parameterized by $\{\Pi, \{\Theta_l\}_{l=1}^L\}$. When Assumption 1, Assumption 2, and Condition 1 hold, with high probability, we have*

$$\max_{i\in[N]} \|e_i'(\hat{\Pi} - \Pi\mathcal{P})\|_1 = O(\sqrt{\frac{\log(N+J+L)}{\rho^2 NJL}}) + O(\frac{1}{N}).$$

The result presented in Theorem 1 indicates that enhancing the values of ρ, N, and J enhances the performance of GoM-DSoG in estimating Π, which aligns with our intuition. In particular, a higher intensity parameter ρ reduces the number of low-intensity answers, providing more informative and varied data. This makes the distinct response patterns of each latent class more pronounced and easier for the algorithm to distinguish, leading to more accurate parameter estimates. Moreover, it indicates that as L increases, GoM-DSoG demonstrates improved performance for the task of GoM analysis. This observation underscores the benefits of considering multi-layer categorical data over single-layer categorical

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

data, as the former contains more responses and, consequently, more information. Through the proof of Theorem 1, it is evident that the impact of the three model parameters ρ, N, J, and L cannot be analyzed without Assumption 1, Assumption 2, and Condition 1. This underscores that these assumptions are employed for theoretical purposes. Furthermore, Theorem 1 relies critically on the pure subject assumption in Equation (2.3), as this ensures the existence of the Ideal Simplex structure in Lemma 1. When Lemma 1 holds, applying SPA to U can exactly identify the K true pure subjects, implying that GoM-DSoG will exactly recover Π and $\{\Theta_l\}_{l=1}^L$ when we use $\mathcal{S}$ to replace S in GoM-DSoG. When Equation (2.3) holds, we can apply the SPA algorithm to $\hat{U}$ to estimate pure subjects when $\{R_l\}_{l=1}^L$ are generated from the proposed model with parameters Π and $\{\Theta_l\}_{l=1}^L$. Finally, we can bound the difference between Π and $\hat{\Pi}$. Otherwise, if Equation (2.3) does not hold (say, there is at least one latent class having no pure subjects), GoM-DSoG can not exactly recover Π and $\{\Theta_l\}_{l=1}^L$ even when we use $\mathcal{S}$ to replace S in GoM-DSoG, which leads to a result that we can not obtain Theorem 1 anymore.

For the task of estimating the item parameter matrices $\{\Theta_l\}_{l=1}^L$, the following lemma provides a bound on $\frac{\|\sum_{l\in[L]}(\hat{\Theta}_l-\Theta_l\mathcal{P})\|_F}{\|\sum_{l\in[L]}\Theta_l\|_F}$.

Lemma 2. *Consider a multi-layer GoM parameterized by* $\{\Pi, \{\Theta_l\}_{l=1}^L\}$. *Under the conditions that* $\rho L\max(N, J) \gg \log(N + J + L)$ *and* $|\lambda_1(\sum_{l\in[L]} B_l)| \geq c_2\sqrt{JL}$ *for some constant* $c_2 > 0$, *with high probability, we have*

$$\frac{\|\sum_{l\in[L]}(\hat{\Theta}_l - \Theta_l\mathcal{P})\|_F}{\|\sum_{l\in[L]}\Theta_l\|_F} = O(\sqrt{\frac{\max(N, J)\log(N + J + L)}{\rho NJL}}).$$

Note that the conditions specified in Lemma 2 differ from those of Theorem 1. This difference arises because bounding $\|\sum_{l\in[L]}(R_l - \mathscr{R}_l)\|$ is necessary to obtain the result presented in Lemma 2. Despite the distinct conditions, the result in Lemma 2 also indicates

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

that enhancing the values of ρ, N, J, and L leads to improved performance of the GoM-DSoG method.

4. Estimation of K

Selecting the appropriate number of latent classes K in multi-layer categorical data with a latent mixed membership structure poses a significant challenge. While Algorithm 1 assumes K is known, this is often not feasible in practical applications. Given L observed responses $\{R_l\}_{l=1}^{L}$, our GoM-DSoG algorithm can always produce an estimated mixed membership matrix $\hat{\Pi}_k \in [0,1]^{N\times k}$ for any number of latent classes k. When faced with multiple potential choices for K, such that the true K may be one of $\{1, 2, \ldots, K_c\}$, we can run our GoM-DSoG algorithm on the data with k latent classes for all $k \in [K_c]$, resulting in K_c estimated mixed membership matrices $\{\hat{\Pi}_k\}_{k=1}^{K_c}$. Naturally, this leads to the question of which estimated membership matrix indicates the best quality of mixed membership partition.

To address this question, we propose a metric called averaged fuzzy modularity, inspired by the fuzzy modularity introduced in (Qing, 2024a) for single-layer categorical data with a latent mixed membership structure. The averaged fuzzy modularity based on the estimated mixed membership matrix $\hat{\Pi}_k$ when we assume there are k latent classes is defined as follows:

$$Q(\hat{\Pi}_k) = \frac{1}{L}\sum_{l\in[L]}\sum_{i\in[N]}\sum_{j\in[N]}\frac{1}{\eta_l}(G_l(i,j) - \frac{g_l(i,i)g_l(j,j)}{\eta_l})\hat{\Pi}_k(i,:)\hat{\Pi}_k'(j,:), \tag{4.1}$$

where $G_l = R_l R_l'$, g_l is an $N \times N$ diagonal matrix with $g_l(i,i) = \sum_{j\in[N]} G_l(i,j)$, and $\eta_l = \sum_{j\in[N]} g_l(j,j)$ for $i \in [N], l \in [L]$. This metric has several notable properties:

- When there are no mixed subjects, our modularity reduces to the averaged modularity of (Paul and Chen, 2021; Qing, 2024b).

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

- When $L = 1$, it reduces to the fuzzy modularity of (Nepusz et al., 2008; Qing, 2024a).
- When $L = 1$ and there are no mixed subjects, it becomes the popular Newman-Girvan modularity (Newman and Girvan, 2004).

A larger value of Q indicates better quality of mixed membership partition (Newman, 2006; Nepusz et al., 2008; Paul and Chen, 2021). Therefore, similar to (Nepusz et al., 2008), we infer the optimal K by choosing the one that maximizes Q defined in Equation (4.1).

5. Experimental studies

5.1 Simulations

This part investigates the performance of the proposed GoM-DSoG method for the task of the grade of membership analysis and its accuracy in the estimation of K. To this end, we conduct five experiments to study the effects of various factors, including the number of subjects N, the number of layers L, the response intensity parameter ρ, the number of pure subjects in each latent class, and the number of latent classes K. In all experiments, the GoM-DSoG method is compared with the following two different algorithms:

- **GoM-Sum**: Set $R_{\text{sum}} = \sum_{l\in[L]} R_l$. Let $\hat{U}_{R_{\text{sum}}}\hat{\Sigma}_{R_{\text{sum}}}\hat{V}'_{R_{\text{sum}}}$ represent R_{sum}'s leading K singular value decomposition, where $\hat{U}'_{R_{\text{sum}}}\hat{U}_{R_{\text{sum}}} = I_{K\times K}$. This approach estimates the parameters Π and $\{\Theta_l\}_{l=1}^N$ by substituting R_{sum} for S and $\hat{U}_{R_{\text{sum}}}$ for $\hat{U}$ in Algorithm 1.
- **GoM-SoG**: Set $\tilde{S} = \sum_{l\in[L]} R_l R'_l$. This method estimates Π and $\{\Theta_l\}_{l=1}^N$ by replacing S with $\tilde{S}$ in Algorithm 1.

Remark 2. *Define $\mathscr{R}_{\text{sum}} = \sum_{l\in[L]} \mathscr{R}_l$. When the rank of the $J \times K$ aggregation matrix $\sum_{l\in[L]} B_l$ is K, let $U_{\mathscr{R}_{\text{sum}}}\Sigma_{\mathscr{R}_{\text{sum}}}V'_{\mathscr{R}_{\text{sum}}}$ denote the leading K singular value decomposition of*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

$\mathscr{R}_{\text{sum}}$ such that $U'_{\mathscr{R}_{\text{sum}}}U_{\mathscr{R}_{\text{sum}}} = I_{K\times K}$. The intuition behind GoM-Sum stems from the observation that $U_{\mathscr{R}_{\text{sum}}} = \Pi U_{\mathscr{R}_{\text{sum}}}(\mathcal{I},:)$ also exhibits a simplex structure. The intuition behind GoM-SoG is analogous to that of GoM-DSoG, as $\tilde{S}$ serves as a biased estimator of $\mathcal{S}$ (Lei and Lin, 2023; Su et al., 2024; Qing, 2024b). By following a similar analysis as in (Qing, 2024b), we can obtain theoretical guarantees for GoM-Sum and GoM-SoG. Furthermore, we can prove that our proposed GoM-DSoG consistently outperforms both GoM-Sum and GoM-SoG. For brevity, we omit the detailed theoretical analysis of GoM-Sum and GoM-SoG in this paper.

To assess the performance of all approaches in estimating Π, the **Relative l_1 error** defined as $\min_{\mathcal{P}\in\mathcal{X}}\frac{\|\hat{\Pi}-\Pi\mathcal{P}\|_1}{N}$ is used, where $\mathcal{X}$ collects all $K\times K$ permutation matrices. To measure their performances in estimating $\{\Theta_l\}_{l=1}^L$, we use the **Relative l_2 error**, defined as $\min_{\mathcal{P}\in\mathcal{X}}\frac{\|\sum_{l\in[L]}(\hat{\Theta}_l-\Theta_l\mathcal{P})\|_F}{\|\sum_{l\in[L]}\Theta_l\|_F}$. To evaluate the accuracy of all methods in estimating K, the **Accuracy rate** calculated by the proportion of times an approach accurately chooses the true K by maximizing Q is employed. The Relative l_1 error and Relative l_2 error metrics are such that smaller values are better, while the Accuracy rate metric is such that larger values are better.

For all experiments, we set $J = \frac{N}{5}$ and $M = 5$, so that $R_l \in \{0,1,2,3,4,5\}^{N\times\frac{N}{5}}$ for all $l \in [L]$. Let each latent class have N_0 pure subjects. For a mixed subject i, set $\Pi(i,k) = \frac{\text{rand}(1)}{K-1}$ for $k \in [K-1]$ and $\Pi(i,K) = 1-\sum_{k\in[K-1]}\Pi(i,k)$ when $K \geq 2$, where rand(1) denotes a random value from $(0,1)$. For $j \in [J], k \in [K], l \in [L]$, set $B_l(j,k) = \text{rand}(1)$. For all experiments, N, K, ρ, N_0, and L are set independently. For each parameter setting, we report the averaged metric over the 50 repetitions. For all approaches, K is given for estimating Π and $\{\Theta_l\}_{l=1}^L$ while K is not given and is estimated by maximizing the averaged

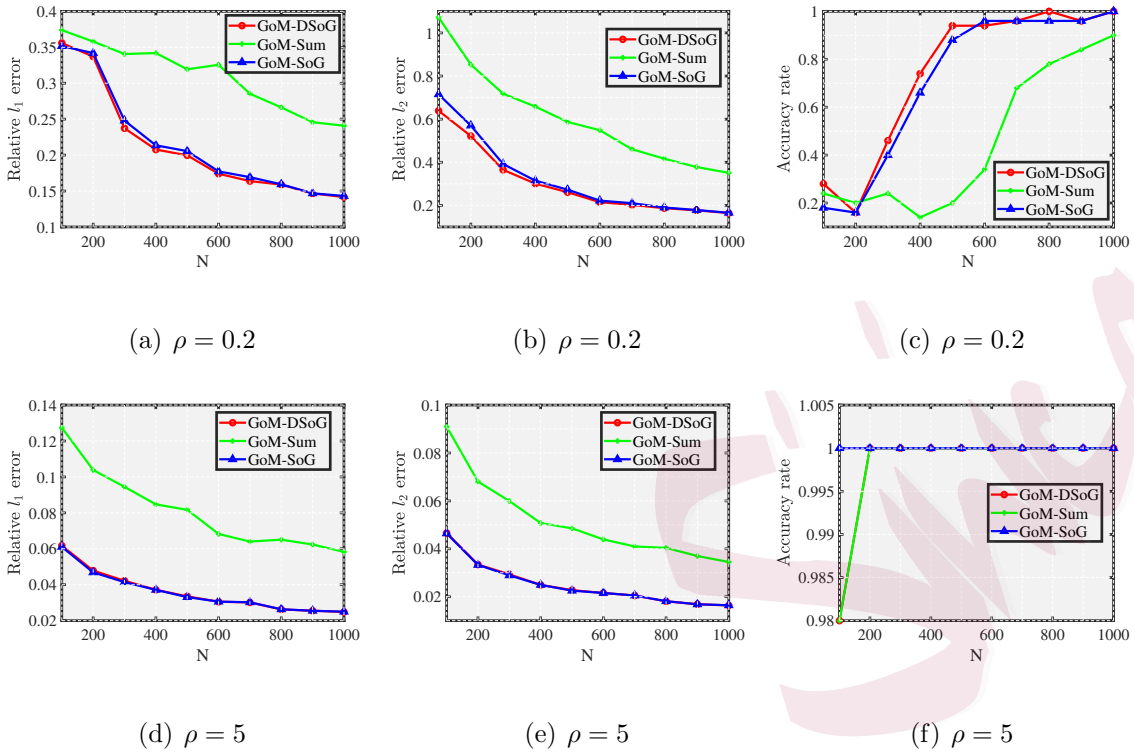
Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

fuzzy modularity value Q for the task of determining K. Meanwhile, for each experiment, there are two cases: the low-intensity case and the high-intensity case. In the low-intensity case, we let the response intensity parameter ρ be small, while in the high-intensity case, ρ is large. The low-intensity case is characterized by numerous zeros in the generated data, which aligns multi-layer categorical data with binary responses or the scenario where many participants have lowest responses to all items in each test. Conversely, the high-intensity case features only a few zeros, corresponding to real-world situations where only a minority of participants have lowest responses in multi-layer categorical data with polytomous responses.

Experiment 1: Effect of changing N. Set $L = 5$, $N_0 = N/5$, and $K = 3$. Let N vary in $\{100, 200, \ldots, 1000\}$. For the low-intensity case, set $\rho = 0.2$. For the high-intensity case, set $\rho = 5$. Results are presented in Figure 3, indicating that (a) all methods enjoy better performance as the number of subjects N grows; (b) our GoM-DSoG slightly outperforms GoM-SoG, and both methods significantly outperform GoM-Sum, especially in the low-intensity case; (c) all methods perform better in the high-intensity case compared to the low-intensity case for estimating Π, $\{\Theta_l\}_{l=1}^{L}$, and K. Notably, all methods demonstrate satisfactory performance in the high-intensity case, which is consistently observed in other experiments (thus, similar analyses are omitted for subsequent experiments).

Experiment 2: Effect of changing L. Set $N = 500$, $N_0 = 100$, and $K = 3$. Let L vary in $\{1, 2, \ldots, 10\}$. For the low-intensity case, set $\rho = 0.2$. For the high-intensity case, set $\rho = 5$. Results are reported in Figure 4. We observe that GoM-DSoG and GoM-SoG perform similarly, both methods behave better as L increases, and they consistently outperform GoM-Sum. The results of this experiment support the advantages of considering multiple tests (i.e., layers) at different times. Specifically, if a significant proportion of lowest

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276



(a) $\rho = 0.2$ (b) $\rho = 0.2$ (c) $\rho = 0.2$

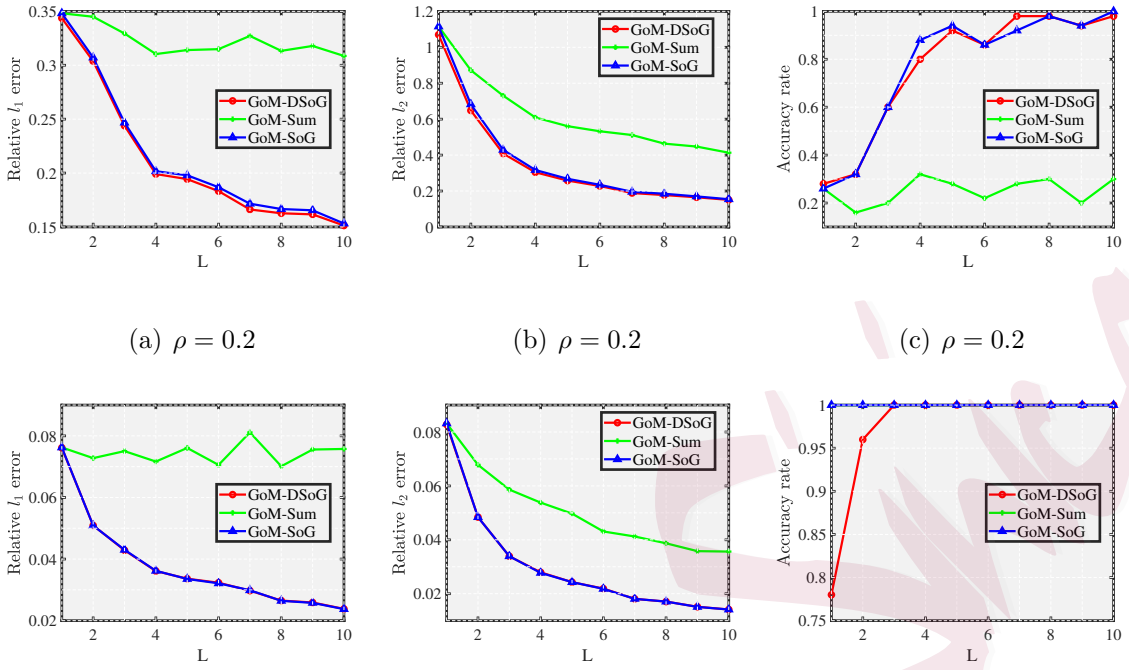
(d) $\rho = 5$ (e) $\rho = 5$ (f) $\rho = 5$

Figure 3: Experiment 1.

responses is present in the data, conducting additional tests at various times, i.e., increasing L, can be beneficial. This process results in multi-layer categorical data. However, for such data, simply summing all response matrices (i.e., using the GoM-Sum method) for the grade of membership analysis and estimating K is inefficient. Instead, our proposed GoM-DSoG method is more effective.

Experiment 3: Effect of changing ρ. Set $N = 500$, $L = 5$, $N_0 = 100$, and $K = 3$. For the low-intensity case, let ρ vary in $\{0.05, 0.1, \ldots, 0.5\}$. For the high-intensity case, let ρ vary in $\{0.5, 1, \ldots, 5\}$. Results are displayed in Figure 5. Observations indicate that all methods perform better as the response intensity parameter ρ increases, our GoM-DSoG slightly outperforms GoM-SoG, and both significantly outperform GoM-Sum. We also plot the estimated simplex structure returned by applying the SPA algorithm to the eigenvector

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276



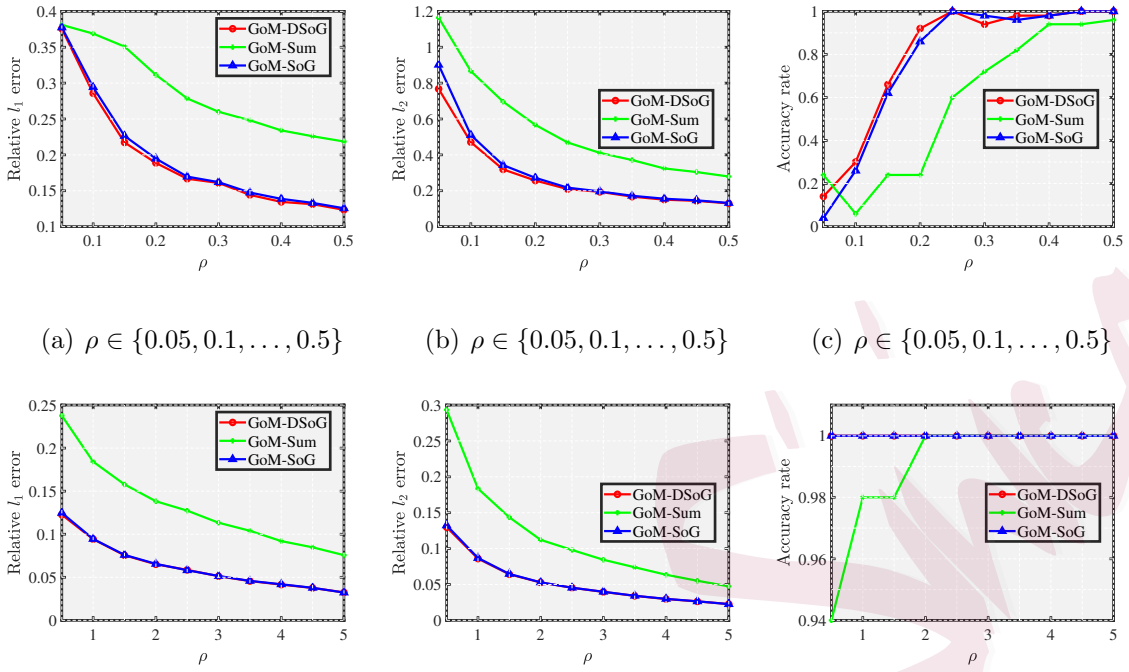
(a) $\rho = 0.2$ (b) $\rho = 0.2$ (c) $\rho = 0.2$

(d) $\rho = 5$ (e) $\rho = 5$ (f) $\rho = 5$

Figure 4: Experiment 2.

matrix $\hat{U}$ under varying response intensity parameters ρ in Figure 6. As ρ increases, the points representing pure subjects (denoted by red points) become increasingly concentrated near the vertices of the simplex structure, while the points representing mixed subjects (denoted by black points) tend to distribute more uniformly within the interior of the simplex. This observation aligns with the theoretical expectation that higher values of ρ reduce the number of zero entries in the data, thereby improving the separability of pure and mixed subjects and making it easier for SPA to accurately detect pure subjects. We further report the Accuracy rate of SPA in identifying pure subjects in Figure 7, where the Accuracy rate is defined as the proportion of times, over 50 repetitions, that applying SPA to $\hat{U}$ correctly identifies K pure subjects, each from a distinct latent class. From Figure 7, it is evident that the Accuracy rate of SPA increases as ρ increases. When ρ is small, the data has a significant

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276



(a) $\rho \in \{0.05, 0.1, \ldots, 0.5\}$ (b) $\rho \in \{0.05, 0.1, \ldots, 0.5\}$ (c) $\rho \in \{0.05, 0.1, \ldots, 0.5\}$

(d) $\rho \in \{0.5, 1, \ldots, 5\}$ (e) $\rho \in \{0.5, 1, \ldots, 5\}$ (f) $\rho \in \{0.5, 1, \ldots, 5\}$

Figure 5: Experiment 3.

number of zero entries. This intensity introduces noise and makes it challenging for SPA to distinguish pure subjects from mixed subjects. However, as ρ increases, the data has more higher responses, and the inherent structure of the simplex becomes more pronounced. This allows SPA to more accurately identify pure subjects. These results highlight SPA's vertex stability—its ability to consistently and reliably in hunting the vertices of the simplex structure, even under varying levels of data intensity controlled by the parameter ρ.

Experiment 4: Effect of changing N_0. Set $N = 600$, $L = 10$, and $K = 3$. Let N_0 vary in $\{20, 40, \ldots, 200\}$. For the low-intensity case, set $\rho = 0.2$. For the high-intensity case, set $\rho = 5$. Performances of the three methods are reported in Figure 8. Results suggest that all methods exhibit improved performance in estimating the item parameter matrices and K when N_0 grows, with GoM-DSoG performing slightly better than GoM-SoG and both

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

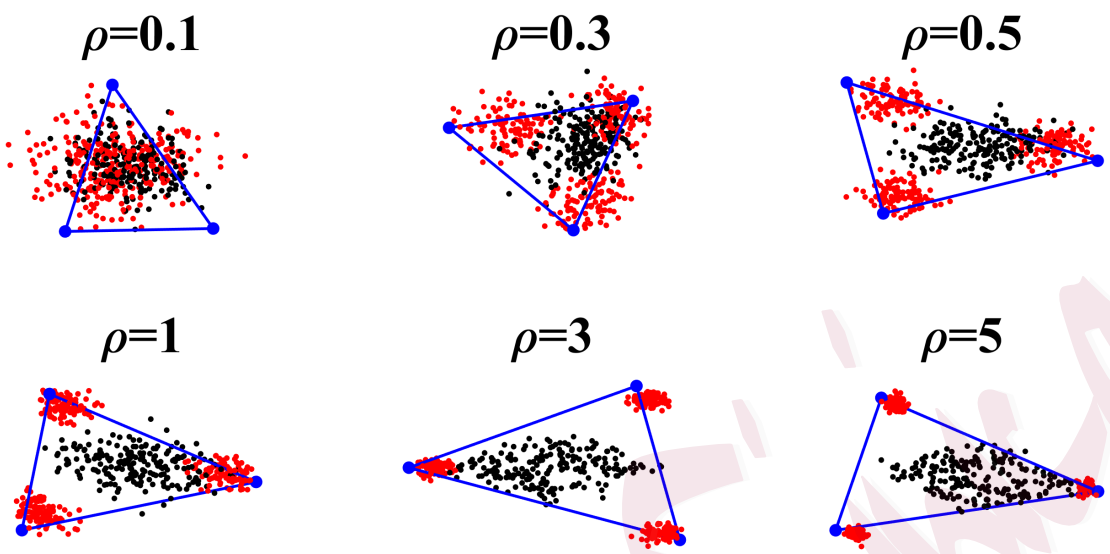


Figure 6: Illustration of the estimated simplex structure for different values of the response intensity parameter ρ. In each sub-figure, each point represents a row of $\hat{U}$ obtained from the debiased sum of Gram matrices S. Red points represent pure subjects, while black points represent mixed subjects. The three blue points denote the estimated pure subjects obtained by applying the SPA algorithm to $\hat{U}$, and the blue triangle represents the estimated simplex structure. For visualization, these points have been projected from $\mathbb{R}^3$ to $\mathbb{R}^2$.

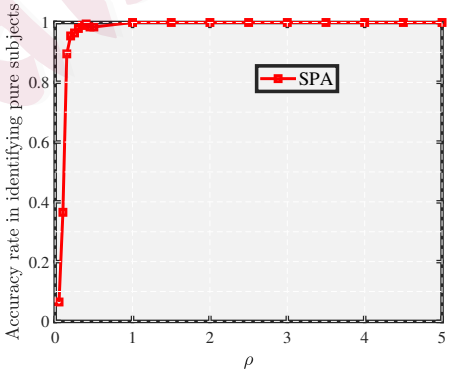
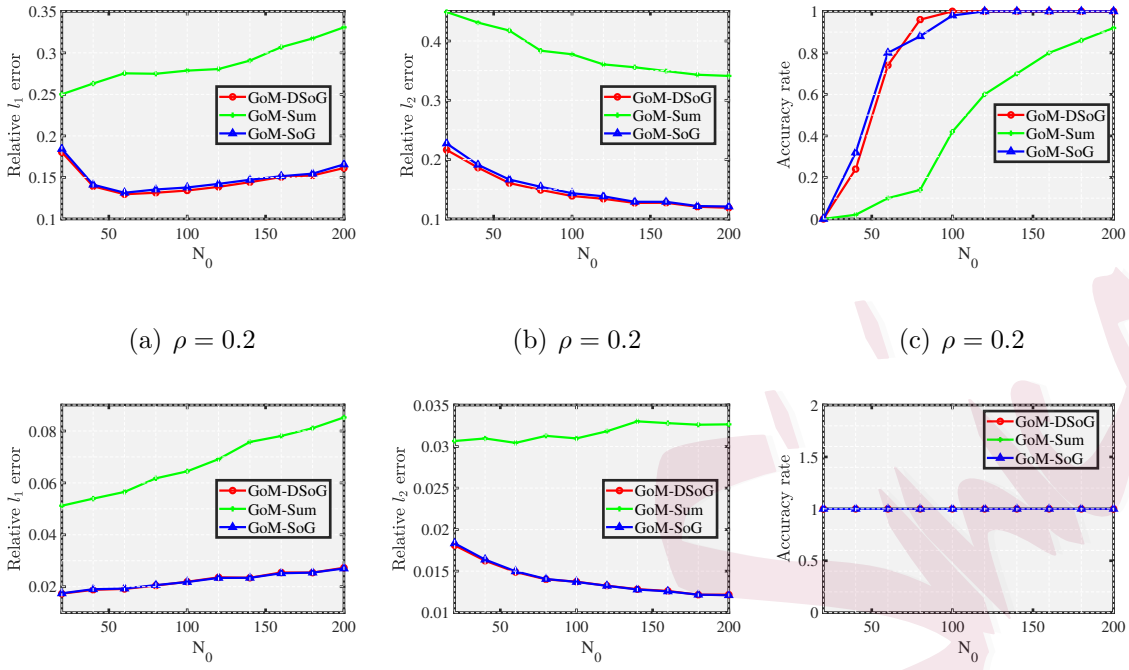


Figure 7: Accuracy rate of the SPA algorithm in identifying pure subjects when ρ takes values in $\{0.05, 0.1, \ldots, 0.5, 1, 1.5, 2, \ldots, 5\}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276



(a) $\rho = 0.2$ (b) $\rho = 0.2$ (c) $\rho = 0.2$

(d) $\rho = 5$ (e) $\rho = 5$ (f) $\rho = 5$

Figure 8: Experiment 4.

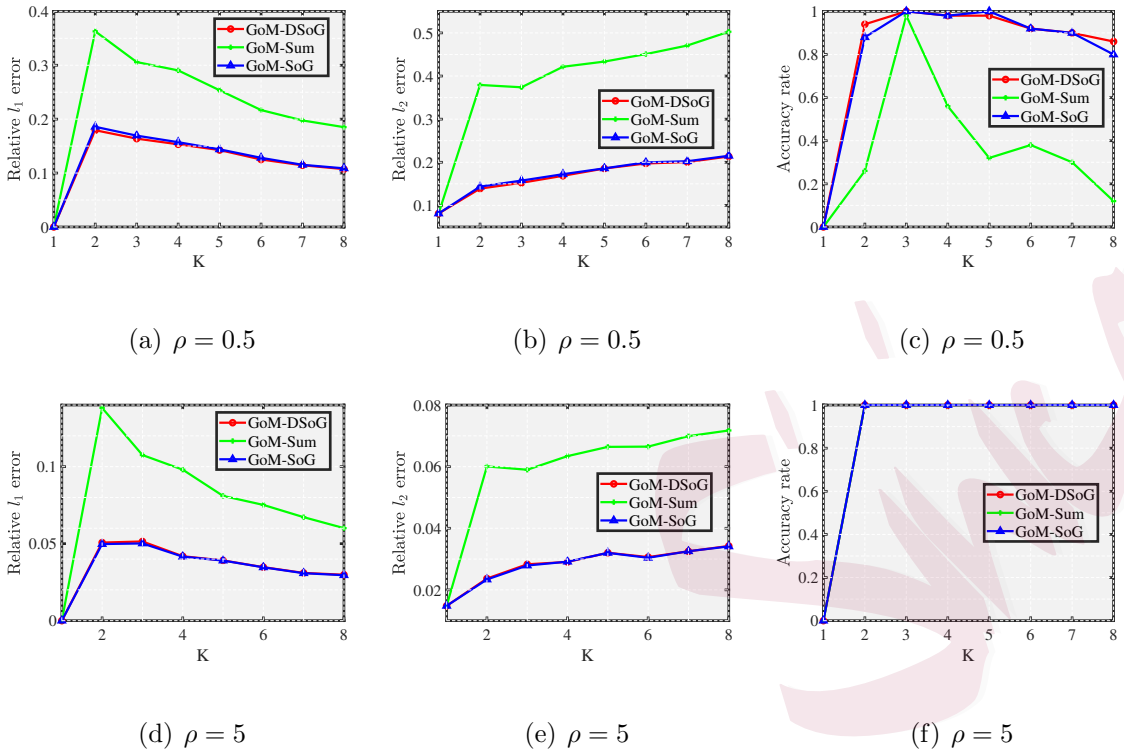
significantly outperforming GoM-Sum.

Experiment 5: Effect of changing K. Set $N = 100K, N_0 = N/K, L = 5$. Let K vary in $\{1, 2, \ldots, 8\}$. For the low-intensity case, set $\rho = 0.5$. For the high-intensity case, set $\rho = 5$. Results are summarized in Figure 9. We observe that (a) GoM-DSoG and GoM-SoG significantly outperform GoM-Sum in the dense case; (b) all methods fail to estimate K when the true K is 1. However, this is not a critical issue since we typically assume that there are at least 2 latent classes for the task of grade of membership analysis.

5.2 Real data

In real-world data applications, the scarcity of readily available multi-layer ordinal categorical datasets poses a challenge for evaluating the performance of our GoM-DSoG algo-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276



(a) $\rho = 0.5$ (b) $\rho = 0.5$ (c) $\rho = 0.5$

(d) $\rho = 5$ (e) $\rho = 5$ (f) $\rho = 5$

Figure 9: Experiment 5.

rithm. To address this, we utilize a single-layer ordinal categorical dataset derived from a survey, which we subsequently transform into a multi-layer structure through a randomized selection process. This dataset, sourced from a survey where participants were asked to generate random numbers within specified ranges corresponding to the Big Five Personality Test (hereafter referred to as BFPT), is accessible for download at `https://openpsychometrics.org/_rawdata/`. The BFPT dataset comprises 1369 subjects and 50 items, categorized into five personality traits: extraversion (E1-E10), neuroticism (N1-N10), agreeableness (A1-A10), conscientiousness (C1-C10), and openness (O1-O10). The details of each item can be found in Figure 11. For this data, responses 0, 1, 2, 3, 4, and 5 represent different levels of agreement with 0 denoting the lowest intensity and 5 denoting the highest intensity. Consequently, we have $R \in \{0, 1, 2, 3, 4, 5\}^{1369\times 50}$. BFPT contains a

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

total of 206 0s, constituting approximately 0.3% of the dataset ($\frac{206}{1369\times 50}$). To convert the original single-layer BFPT data into a multi-layer ordinal categorical format, we employ a probabilistic approach. For $l \in [L]$, we set $R_l(i,j) = R(i,j)$ with a probability of 0.5, and $R_l(i,j) = 0$ otherwise. In this way, we generate a synthetic multi-layer dataset that approximates real-world conditions. In this study, we set $L = 3$ layers for our analysis, and we continue to refer to the transformed dataset as BFPT.

We estimate the number of latent classes by maximizing the modularity defined in Equation (4.1) using our GoM-DSoG algorithm. The inferred value of K is 3, thus we have $\hat{\Pi} \in [0,1]^{1369\times 3}$. We call the three estimated latent classes as Class 1, Class 2, and Class 3. We find that $\sum_{i\in[N]} \hat{\Pi}(i,1) = 370.0706$, $\sum_{i\in[N]} \hat{\Pi}(i,2) = 323.1968$, and $\sum_{i\in[N]} \hat{\Pi}(i,3) = 675.7326$, implying that most subjects tend to belong to Class 3. To further analyze the estimated mixed membership matrix $\hat{\Pi}$, we call subject i strongly pure if $\max_{k\in[K]}\hat{\Pi}(i,k) \geq 0.9$, strongly mixed if $\max_{k\in[K]}\hat{\Pi}(i,k) < 0.5$, and moderate pure otherwise. We find that there are 19 strongly pure subjects, 322 strongly mixed subjects, and 1028 moderate pure subjects. The estimated mixed memberships of strongly pure subjects and a ternary visualization of the estimated mixed membership matrix $\hat{\Pi}$ are shown in Figure 10. In the right panel of Figure 10, the vertices of the triangle represent pure subjects. For any subject, a position closer to one of the triangle's vertices indicates a higher likelihood of being a pure subject. Conversely, a position closer to the center of the triangle suggests a more mixed membership. The ternary diagram clearly shows that most subjects tend to belong to Class 3.

Figure 11 displays the estimated item parameter matrices for the BFPT data. Below, we provide a comprehensive interpretation of these results across the five personality dimensions.

Extraversion dimension:

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

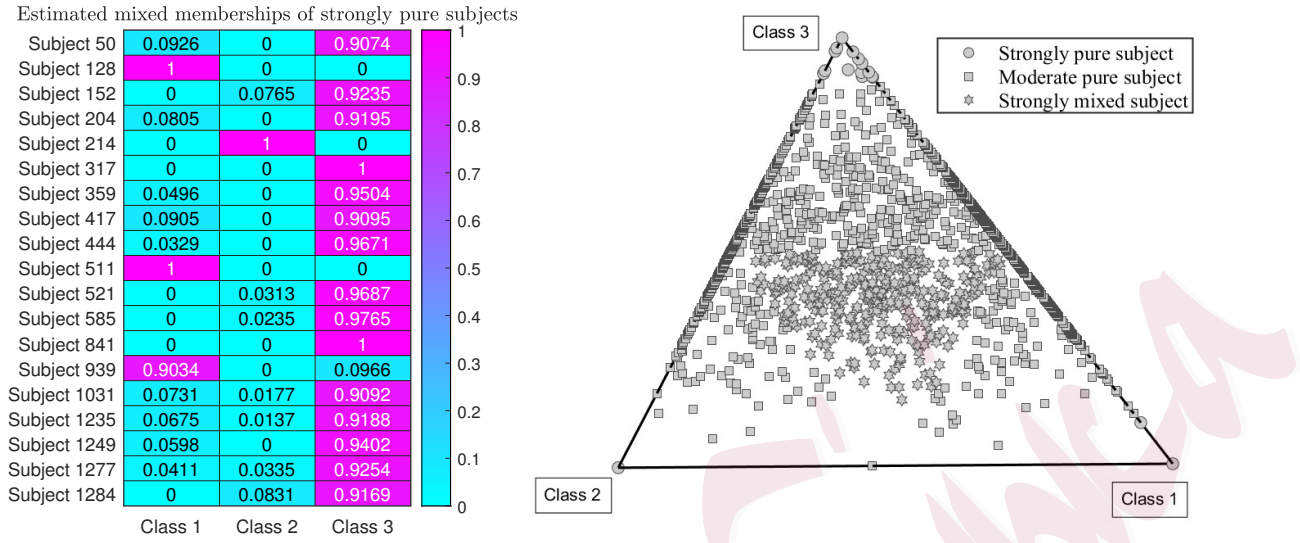


Figure 10: Left: estimated mixed memberships of strongly pure subjects for the BFPT data. Right: ternary diagram of $\hat{\Pi}$.

Figure 11: Estimated item parameter matrices for the three layers of the BFPT data.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

- Class 1: Individuals in this class exhibit moderate agreement with statements like "I feel comfortable around people" (E3) and "I talk to a lot of different people at parties" (E7). This reflects a balanced level of extroversion: they can interact with others but do not seek to dominate social interactions.

- Class 2: Individuals in this class are more likely to strongly agree with statements such as "I am the life of the party" (E1) and "I do not mind being the center of attention" (E9). This suggests they are more outgoing and proactive in social settings, and comfortable engaging with others even in unfamiliar environments.

- Class 3: Individuals in this class are less likely to strongly agree with statements such as "I am the life of the party" (E1) and "I start conversations" (E5). This indicates a lower tendency toward extroverted behaviors, suggesting they may prefer solitary activities or smaller social gatherings.

Neuroticism dimension:

- Class 1: Individuals in this class are more likely to strongly agree with statements like "I get stressed out easily" (N1) and "I often feel blue" (N10). This indicates a higher likelihood of experiencing stress and negative emotions, suggesting they may be more sensitive to stressors and prone to emotional fluctuations.

- Class 2: Individuals in this class are less likely to strongly agree with statements like "I am relaxed most of the time" (N2) and "I seldom feel blue" (N4). This indicates lower stress levels and negative emotions, suggesting they are more emotionally stable and better able to cope with stress.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

- Class 3: Individuals in this class exhibit moderate agreement with statements such as "I worry about things" (N3) and "I change my mood a lot" (N7). This suggests they experience some level of worry and mood variability but to a lesser extent than Class 1.

Agreeableness dimension:

- Class 1: Individuals in this class show moderate agreement with statements for items like "I am interested in people" (A2) and "I sympathize with the feelings of others" (A4). This indicates a moderate level of empathy and concern for others.

- Class 2: Individuals in this class may have relatively high response values for items such as "I make people feel at ease" (A10). This suggests they are skilled at making others feel comfortable and are more considerate of others' emotions.

- Class 3: Individuals in this class may exhibit relatively low response values for items such as "I sympathize with the feelings of others" (A4) and "I feel the emotions of others" (A9). This suggests they may be less considerate of others' feelings.

Conscientiousness dimension:

- Class 1: Individuals in this class may have relatively low response values for items like "I am always prepared" (C1) and "I follow a schedule" (C9). This suggests they may be less organized and less likely to plan ahead.

- Class 2: Individuals in this class may have relatively high response values for items like "I get chores done right away" (C5) and "I follow a schedule" (C9). This suggests they are more conscientious, tending to complete tasks promptly and to high standards.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

- Class 3: Individuals in this class exhibit moderate response values for items such as "I make a mess of things" (C4) and "I shirk my duties" (C8). This indicates a moderate level of attention to detail and organizational skills.

Openness dimension:

- Class 1: Individuals in this class show moderate response values. This indicates a moderate level of openness to new experiences and ideas.
- Class 2: Individuals in this class may have relatively low response values for items such as "I have difficulty understanding abstract ideas" (O2), "I am not interested in abstract ideas" (O4), and "I do not have a good imagination" (O6). This suggests they are more open to new experiences and ideas, with stronger curiosity and creativity.
- Class 3: Individuals in this class may have relatively low response values for items such as "I have a vivid imagination" (O3) and "I have excellent ideas" (O5). This suggests they may be less open to new experiences and ideas.

6. Conclusion

In this paper, we propose a novel model, the multi-layer GoM model, which extends the traditional GoM model to effectively describe multi-layer ordinal categorical data with a latent mixed membership structure. Our multi-layer GoM assumes that the response matrix for each layer is generated by the classical GoM model, sharing common mixed memberships but with varying item parameter matrices. To facilitate GoM analysis in multi-layer ordinal categorical data, we develop a new approach, GoM-DSoG, inspired by the recently popular technique of debiased spectral clustering in network analysis. We establish

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

the estimation consistency of our method and find that, compared to single-layer analysis, multi-layer analysis exhibits greater power for the task of grade of membership analysis. Furthermore, we propose an efficient approach for estimating the number of latent classes of multi-layer ordinal categorical data with latent mixed membership structures. Our extensive experimental studies demonstrate the effectiveness of our methods in estimating subjects' mixed memberships, the number of latent classes, and other model parameters. To our knowledge, we are the first to explore the grade of membership analysis in multi-layer ordinal categorical data. Our contributions advance the field of ordinal categorical data analysis and have practical implications, particularly in psychological testing and other applications involving multi-layer ordinal categorical data.

The multi-layer GoM framework and the proposed GoM-DSoG algorithm offer a robust and scalable approach for analyzing complex multi-layer ordinal categorical data. However, the current lack of publicly available data underscores a critical methodological gap that our work seeks to address. A promising direction for future work involves applying this methodology to real-world longitudinal studies where the same ordinal-scale instrument is administered repeatedly over time. We encourage researchers in fields such as educational measurement (e.g., tracking student attitudes across a semester with identical surveys), clinical psychology (e.g., monitoring patient symptom severity through repeated clinical assessments), and organizational behavior (e.g., measuring employee engagement across multiple annual surveys) to utilize our model and algorithm to further validate their practical utility and provide valuable insights into the latent mixed memberships of subjects in multi-layer ordinal categorical data.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

Supplementary Materials

All technical details and the MATLAB codes of GoM-DSoG can be found in the Supplementary Material.

Acknowledgements

H.Q. was supported by the Scientific Research Foundation of Chongqing University of Technology (Grant No. 2024ZDR003), and the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN202401168).

References

Agresti, A. (2012). *Categorical data analysis*, Volume 792. John Wiley & Sons.

Agterberg, J. and A. R. Zhang (2025). Estimating higher-order mixed memberships via the $\ell_{2,\infty}$ tensor perturbation bound. *Journal of the American Statistical Association 120*(550), 1214–1224.

Araújo, M. C. U., T. C. B. Saldanha, R. K. H. Galvao, T. Yoneyama, H. C. Chame, and V. Visani (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems 57*(2), 65–73.

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.

Chen, L. and Y. Gu (2024). A spectral method for identifiable grade of membership analysis with binary responses. *psychometrika 89*(2), 626–657.

Chen, Y., X. Li, and S. Zhang (2019). Joint maximum likelihood estimation for high-dimensional exploratory item factor analysis. *Psychometrika 84*, 124–146.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

REFERENCES

Erosheva, E. A., S. E. Fienberg, and C. Joutard (2007). Describing disability through individual-level mixture models for multivariate binary data. *Annals of Applied Statistics 1*(2), 346.

Gillis, N. and S. A. Vavasis (2013). Fast and robust recursive algorithmsfor separable nonnegative matrix factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence 36*(4), 698–714.

Gillis, N. and S. A. Vavasis (2015). Semidefinite programming based preconditioning for more robust near-separable nonnegative matrix factorization. *SIAM Journal on Optimization 25*(1), 677–698.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika 61*(2), 215–231.

Gormley, I. C. and T. B. Murphy (2009). A grade of membership model for rank data. *Bayesian Analysis 4*(2), 265–295.

Gu, Y., E. A. Erosheva, G. Xu, and D. B. Dunson (2023). Dimension-grouped mixed membership models for multivariate categorical data. *Journal of Machine Learning Research 24*(88), 1–49.

Hagenaars, J. A. and A. L. McCutcheon (2002). *Applied latent class analysis*. Cambridge University Press.

Jin, J., Z. T. Ke, and S. Luo (2024). Mixed membership estimation for social networks. *Journal of Econometrics 239*(2), 105369.

Ke, Z. T. and M. Wang (2024). Using svd for topic modeling. *Journal of the American Statistical Association 119*(545), 434–449.

Klopp, O., M. Panov, S. Sigalla, and A. B. Tsybakov (2023). Assigning topics to documents by successive projections. *Annals of Statistics 51*(5), 1989–2014.

Lei, J., K. Chen, and B. Lynch (2020). Consistent community detection in multi-layer network data. *Biometrika 107*(1), 61–73.

Lei, J. and K. Z. Lin (2023). Bias-adjusted spectral clustering in multi-layer stochastic block models. *Journal of the American Statistical Association 118*(544), 2433–2445.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

REFERENCES

Lin, K. Z. and J. Lei (2024). Dynamic clustering for heterophilic stochastic block models with time-varying node memberships. *arXiv preprint arXiv:2403.05654*.

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.

Nepusz, T., A. Petróczi, L. Négyessy, and F. Bazsó (2008). Fuzzy communities and the concept of bridgeness in complex networks. *Physical Review E 77*(1), 016107.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences 103*(23), 8577–8582.

Newman, M. E. and M. Girvan (2004). Finding and evaluating community structure in networks. *Physical Review E 69*(2), 026113.

Nylund-Gibson, K. and A. Y. Choi (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science 4*(4), 440.

Paul, S. and Y. Chen (2020). Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *Annals of Statistics 48*(1), 230–250.

Paul, S. and Y. Chen (2021). Null models and community detection in multi-layer networks. *Sankhya A*, 1–55.

Pensky, M. and T. Zhang (2019). Spectral clustering in the dynamic stochastic block model. *Electronic Journal of Statistics 13*(1), 678–709.

Qing, H. (2024a). Finding mixed memberships in categorical data. *Information Sciences 676*, 120785.

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

Qing, H. (2025a). Community detection by spectral methods in multi-layer networks. *Applied Soft Computing 171*, 112769.

Qing, H. (2025b). Discovering overlapping communities in multi-layer directed networks. *Chaos, Solitons & Frac-*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0276

tals 194, 116175.

Qing, H. (2025c). Mixed membership estimation for categorical data with weighted responses. *TEST 34*(3), 612–659.

Qing, H. and J. Wang (2023). Community detection for weighted bipartite networks. *Knowledge-Based Systems 274*, 110643.

Qing, H. and J. Wang (2024). Bipartite mixed membership distribution-free model. a novel model for community detection in overlapping bipartite weighted networks. *Expert Systems with Applications 235*, 121088.

Robitzsch, A. (2023). *sirt: Supplementary Item Response Theory Models*. R package version 3.13-228.

Shang, Z., E. A. Erosheva, and G. Xu (2021). Partial-mastery cognitive diagnosis models. *Annals of Applied Statistics 15*(3), 1529–1555.

Sloane, D. and S. P. Morgan (1996). An introduction to categorical data analysis. *Annual Review of Sociology 22*(1), 351–375.

Su, W., X. Guo, X. Chang, and Y. Yang (2024). Spectral co-clustering in multi-layer directed networks. *Computational Statistics & Data Analysis 198*, 107987.

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

Woodbury, M. A., J. Clive, and A. Garson Jr (1978). Mathematical typology: a grade of membership technique for obtaining disease definition. *Computers and Biomedical Research 11*(3), 277–298.

Xu, S., Y. Zhen, and J. Wang (2023). Covariate-assisted community detection in multi-layer networks. *Journal of Business & Economic Statistics 41*(3), 915–926.