High-dimensional Mixed Graphical Model with Ordinal Data: Parameter Estimation and Statistical Inference

Huijie Feng
Cornell University

Yang Ning
Cornell University

Abstract

We consider parameter estimation and statistical inference of high-dimensional undirected graphical models for mixed data comprising both ordinal and continuous variables. We propose a flexible model called Latent Mixed Gaussian Copula Model that simultaneously deals with such mixed data by assuming that the observed ordinal variables are generated by latent variables. For parameter estimation, we introduce a convenient rank-based ensemble approach to estimate the latent correlation matrix, which can be subsequently applied to recover the latent graph structure. In addition, based on the ensemble estimator, we develop test statistics via a pseudo-likelihood approach to quantify the uncertainty associated with the low dimensional components of high-dimensional parameters. Our theoretical analysis shows the consistency of the estimator and asymptotic normality of the test statistic. Experiments on simulated and real gene expression data are conducted to validate our approach.

1 Introduction

Graphical models have become a fundamental tool to study the complex dependence structures over random variables. In a wide variety of applications such as social science, genetics, statistical physics and image processing, learning high dimensional undirected graphical models is of great interest. As a class of popular tools, Gaussian graphical models (GGMs) assume $\mathbf{X} \in \mathbb{R}^d$ follows a zero mean multivariate Gaussian distribution with covariance $\mathbf{\Sigma}$, in which the conditional independence structure of variables is encoded by the precision matrix $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$: X_i and X_j are conditionally independent given remaining variables if and only if $\Omega_{ji} = 0$ (Lauritzen, 1996). There has been a large number of methods proposed for learning GGMs in high dimension, including the neighborhood selection (Meinshausen and Bühlmann, 2006), the penalized likelihood approach (Yuan and Lin, 2007; Friedman et al., 2008; Banerjee et al., 2008; Rothman et al., 2008), CLIME (Cai et al., 2011, 2016), among others. At the same time, the statistical inferential problems for high-dimensional GGMs attract increasing attention (Liu et al., 2013; Ren et al., 2015; Jankova et al., 2015; Janková and van de Geer, 2017). We refer to Cai et al. (2016); Fan et al. (2016); Drton and Maathuis (2017) for a more complete list of references on graphical models and Zhang and Zhang (2014); Javanmard and Montanari (2014); Van de Geer et al. (2014); Gu et al. (2015); Barber and Kolar (2015); Ning et al. (2017); Cai et al. (2017) for the recent advance in high-dimensional inference.

GGMs are not appropriate for modeling mixed data (i.e., a combination of categorical and continuous data), yet such type of data is ubiquitous in practice, particularly the mix of ordinal and continuous data (referred as ordinal-mixed data) are . For instance, in genetics the data may contain both continuous gene expression values as well as ordinal disease stages and phenotypic effects. In food sensory analysis, continuous rating and ordinal scales are commonly combined to describe different aspects of a food product. Directly treating the ordinal levels as numerical values may not truly reflect the dependence structure between ordinal and continuous variables. Nevertheless, high dimensional graphical models tailored to ordinal-mixed data have attracted less attention. Moreover, how to perform statistical inference on this type of model is largely unknown.

In this paper we propose a unified framework for estimation and statistical inference of the graphical model named Latent Mixed Gaussian Copula Model, which unifies and extends the models in (Ashford and Sowden, 1970; Amemiya, 1974; Guo et al., 2015; Fan et al.,

Proceedings of the 22[nd] International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89.

2017; Suggala et al., 2017) for higher flexibility and generality, particularly for ordinal-mixed data. This model assumes that the observed ordinal data are generated by discretizing a set of latent continuous variables at a set of unknown cutoffs, and that the latent continuous variables together with the observed continuous variables jointly satisfy the nonparanormal distribution Liu et al. (2009, 2012); Xue et al. (2012).

Our main contributions can be summarized as follows. (1) We propose the Latent Mixed Gaussian Copula Model for modeling ordinal-mixed data, along with a convenient ensemble rank-based approach for parameter estimation. To our knowledge, such a simple but very practical procedure has not been proposed in the related literature. (2)Based upon the ensemble rank-based estimator, we propose a pseudo-score statistic for hypothesis testing and constructing confidence intervals. (3) Theoretically, we show rate of convergence of the ensemble rank-based estimator and prove the asymptotic normality of the proposed test statistics with a special U-statistic structure under mild assumptions.

Compared to existing work on learning and inference for ordinal and mixed graphical models, our method is different. For ordinal data, Guo et al. (2015) proposed a probit graphical model and an EM algorithm by maximizing a penalized log likelihood based on the observed data. For the same model, Suggala et al. (2017) proposed a direct estimation method that maximizes the marginal likelihood for each entry independently. In contrast, the latent variables in our model follow a more flexible semiparametric distribution. In another strand of research, the mixed graphical models in Chen et al. (2014); Yang et al. (2014a,b); Lee and Hastie (2015); Cheng et al. (2017) are generally based on modeling the conditional distribution of each node given all the others as an exponential family. In contrast, our model is generative and based on an ensemble estimator. As for the inferential problem, the methods in Liu et al. (2013); Ren et al. (2015); Jankova et al. (2015); Janková and van de Geer (2017) for GGMs and Gu et al. (2015); Barber and Kolar (2015) for transelliptical graphical models reply on their particular modeling structures, such as specifying Gaussian likelihood or pseudo-likelihood. Thus, these methods cannot be directly applied to the proposed model.

The rest of the paper is organized as follows. In section 2 we define the latent mixed Gaussian copula model for ordinal-mixed data. We introduce the methods for latent graph estimation and inference in section 3. Theoretical properties are analyzed in section 4. Simulation studies and a real data analysis are conducted in sections 5 and 6.

2 Background

2.1 Notation.

Let $\boldsymbol{A} = (A_{jk}) \in \mathbb{R}^{d\times d}$ and $\boldsymbol{v} = (v_1, \dots, v_d)^T \in \mathbb{R}^d$. We use $\boldsymbol{v}_S$ as the subvector of $\boldsymbol{v}$ with entries indexed by the set S. For a vector $\boldsymbol{v}$, we define $||\boldsymbol{v}||_1 = \sum_{i=1}^d v_i$, $||\boldsymbol{v}||_2 = (\sum_{i=1}^d v_i^2)^{1/2}$ and $||\boldsymbol{v}||_\infty = \max_{1\le i\le d}\{|v_d|\}$. For a matrix $\boldsymbol{A}$, we define the matrix ℓ_1 norm $||\boldsymbol{A}||_1 = \max_{1\le j\le d}\sum_{i=1}^d |A_{ij}|$ and the matrix elementwise maximum norm $||\boldsymbol{A}||_{\max} = \max\{|A|_{ij}\}$. We use $\boldsymbol{A} \succ 0$ to indicate that $\boldsymbol{A}$ is positive definite. We use $a_n \lesssim b_n$ if $a_n \le Cb_n$ for some constant $C > 0$. We use $\lambda_{\min}(\boldsymbol{A})$ and $\lambda_{\max}(\boldsymbol{A})$ to denote the smallest and largest eigenvalues of $\boldsymbol{A}$.

2.2 Latent Gaussian Copula Model

Given a d-dimensional continuous variable $\boldsymbol{Z}$, we say $\boldsymbol{Z}$ satisfies the Gaussian copula model (Liu et al., 2009, 2012; Xue et al., 2012), i.e., $\boldsymbol{Z} \sim \text{NPN}(0, \boldsymbol{\Sigma}, f)$, if $f(\boldsymbol{Z}) := (f_1(Z_1), \dots, f_d(Z_d))^T \sim N_d(0, \boldsymbol{\Sigma})$ with $\Sigma_{jj} = 1$ and some monotonic transformations $(f_1, ..., f_d)$, where $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate Gaussian distribution with dimension d, mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. To deal with ordinal-mixed data, we consider the following extension.

Definition 1. Latent Mixed Gaussian Copula Model Assume that $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$, where $\boldsymbol{X}_1$ represents d_1-dimensional ordinal variables and $\boldsymbol{X}_2$ represents d_2- dimensional continuous variables. Suppose $X_j \in \{0, \dots, N_j\}$ for $1 \le j \le d_1$. The random vector $\boldsymbol{X}$ satisfies the latent mixed Gaussian copula model, if there is a d_1-dimensional random vector $\boldsymbol{Z}_1 = (Z_1, \dots, Z_{d_1})^T$ such that $\boldsymbol{Z} := (\boldsymbol{Z}_1, \boldsymbol{X}_2) \sim \text{NPN}(0, \boldsymbol{\Sigma}, f)$ and

$$X_j = \sum_{k=1}^{N_j} \mathbb{1}(Z_j > C_j^k) \qquad \forall\, j = 1, \dots, d_1$$

where $C_j^1 < C_j^2 < \dots < C_j^{N_j}$ is a sequence of constants. Let $\mathcal{C} = (\boldsymbol{C}_1, \dots, \boldsymbol{C}_{d_1})$ where $\boldsymbol{C}_j = (C_j^1, \dots, C_j^{N_j})$ and $\mathcal{N} = (N_1, \dots, N_{d_1})$. We denote $\boldsymbol{X} \sim \text{LMNPN}(0, \boldsymbol{\Sigma}, f, \mathcal{C}, \mathcal{N})$.

The proposed model accounts for the ordinal-mixed data by assuming a latent continuous vector $\boldsymbol{Z}$ that generates the ordinal vector $\boldsymbol{X}_1$ via some unknown cutoffs $\mathcal{C}$. Let $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ be the precision matrix. The zero pattern of $\boldsymbol{\Omega}$ encodes the underlying conditional independence among latent variables $\boldsymbol{Z}$, even if the latent variables are not directly observable. Therefore it suffices to estimate the sparsity pattern of the latent precision matrix $\boldsymbol{\Omega}$. The proposed model is invariant to re-ordering the ordinal variables. Specifically, if $\boldsymbol{X}$ is a vector of ordinal variables and denote

$X_j^* = N_j - X_j$, the latent correlation matrix $\boldsymbol{\Sigma}$ is invariant in the sense that if $\boldsymbol{X} \sim \text{LMNPN}(0, \boldsymbol{\Sigma}, f, \mathcal{C}, \mathcal{N})$ then $\boldsymbol{X}^* \sim \text{LMNPN}(0, \boldsymbol{\Sigma}, f^*, \mathcal{C}^*, \mathcal{N})$ for some f^* and $\mathcal{C}^*$. In addition, the model parameters are not fully identifiable; only the transformed unknown cutoffs $f_j(C_j^p)$ rather than the original cutoffs C_j^p are identifiable for the ordinal entries. To ease notation, we define $\triangle_j^{(p)} = f_j(C_j^p)$.

We note that when the input only contains ordinal data, and the monotonic transformations are trivial in the sense that $f(x) = x$, the proposed model is equivalent to the probit graphical model (Ashford and Sowden, 1970; Amemiya, 1974; Guo et al., 2015; Suggala et al., 2017). In addition, when all ordinal variables only contain two levels, the proposed model reduces to the model for binary-mixed data in Fan et al. (2017). We can see that the family of latent mixed Gaussian copula models is strictly larger then the aforementioned ones.

3 Methodology

3.1 Estimate the Latent Correlation Matrix

Given n independent observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim \text{LMNPN}(0, \boldsymbol{\Sigma}, f, \mathcal{C}, \mathcal{N})$, in this section we introduce a procedure to estimate the latent covariance matrix $\boldsymbol{\Omega}$. Our main idea is to firstly estimate the latent correlation matrix, which is then plugged into a quadratic optimization framework to reconstruct the sparsity pattern of $\boldsymbol{\Omega}$.

Due to the existence of ordinal data, direct estimation of ordinal-ordinal and ordinal-continuous correlations become difficult. Under the framwork of GGMs, Chen et al. (2014); Yang et al. (2014a,b); Lee and Hastie (2015); Cheng et al. (2017) resolve this issue by imposing additional distributional assumptions. However, under the proposed Latent Mixed Gaussian Copula Models, such methods are infeasible. Rank-based approach, on the other hand can be applied under the proposed framework (Liu et al., 2012; Xue et al., 2012; Fan et al., 2017), where in brief, the authors connect the latent correlation that is hard to estimate to a rank correlation coefficient such as Kendall's τ or Spearman's ρ, which are convenient to estimate.

Nevertheless, constructing an explicit connection between the rank correlation coefficients and the parameter of interest is complicated for ordinal variables, especially when the number of levels within the ordinal variable becomes large. More importantly, the convergence rate of the estimator cannot be derived when the connection function becomes involved. We therefore propose an ensemble approach: we firstly binarize the ordinal variable at each level and construct a set of preliminary rank-based estimators. At the next stage, we combine the preliminary estimators into a single but strong estimator, of which the theoretical property and empirical advantage over a single preliminary estimator is confirmed in section 4 and 5.

Formally, given an ordinal variable X_{ij} for the ith observation, we consider the binary form with respect to level p as $X_{ij}^{(p)} = \mathbb{1}(X_{ij} \geq p)$, $p = 1, \ldots, N_j$. In the case where X_{ij} and X_{ik} are both ordinal, consider the Kendall's τ calculated from the binarized observed data

$$\hat{\tau}_{jk}^{(p,q)} = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < i' \leq n} \text{sign}\{(X_{ij}^{(p)} - X_{i'j}^{(p)})(X_{ik}^{(q)} - X_{i'k}^{(q)})\},$$

with $\text{sign}(0) = 0$.

Define $\Phi_2(u, v, s) = \int_{x_1 < u} \int_{x_2 < v} \phi_2(x_1, x_2; s) \mathrm{d}x_1 \mathrm{d}x_2$ as the cumulative distribution function of the standard bivariate Gaussian distribution with correlation s, and $\Phi(\cdot)$ as the cumulative distribution of the standard Gaussian distribution. To recover the (latent) correlation between variable j and k from $\hat{\tau}_{jk}^{(p,q)}$, it can be shown that

$$\tau_{jk}^{(p,q)} := \mathbb{E}[\hat{\tau}_{jk}^{(p,q)}] = F(\Sigma_{jk}; \triangle_j^{(p)}, \triangle_k^{(q)}),$$

where

$$F(s; \triangle_j, \triangle_k) = 2\{\Phi_2(\triangle_j, \triangle_k, s) - \Phi(\triangle_j)\Phi(\triangle_k)\}$$

is an invertible function with respect to s.

In practice, $\triangle_j^{(p)}$ and $\triangle_k^{(q)}$ are unknown. Using the first moment condition, $\triangle_j^{(p)}$ can be estimated by $\hat{\triangle}_j^{(p)} = \Phi^{-1}(1 - \frac{1}{n}\sum_{i=1}^n X_{ij}^{(p)})$, and similarly for $\triangle_k^{(q)}$. Thus, we define the base estimate of Σ_{jk} with respect to levels p and q as

$$\hat{R}_{jk}^{(p,q)} = F^{-1}(\hat{\tau}_{jk}^{(p,q)}; \hat{\triangle}_j^{(p)}, \hat{\triangle}_k^{(q)}). \tag{1}$$

Similarly, in the case where variable X_j is ordinal and variable X_k is continuous, the base estimate with respect to level p is given by

$$\hat{R}_{jk}^{(p)} = F^{-1}(\hat{\tau}_{jk}^{(p)}, \hat{\triangle}_j^{(p)}), \tag{2}$$

where

$$\hat{\tau}_{jk}^{(p)} = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < i' \leq n} \text{sign}\{(X_{ij}^{(p)} - X_{i'j}^{(p)})(X_{ik} - X_{i'k})\}$$

and

$$F(s; \triangle_j) = 4\Phi_2(\triangle_j, 0, s/\sqrt{2}) - 2\Phi(\triangle_j).$$

Finally, when both X_{ij} and X_{ik} are continuous, the estimator of Σ_{jk} is given by

$$\hat{R}_{jk} = \sin(\frac{\pi}{2}\hat{\tau}_{jk}), \tag{3}$$

where $\hat{\tau}_{jk}$ is the original Kendall's τ estimator Kendall (1948).

Apparently, the base estimates in (1) and (2) depend on the levels at which the data are binarized. When the ordinal data only contain two levels , the binarizing step is unnecessary and $\hat{R}_{jk}^{(p,q)}$ is independent of (p,q). However, for the ordinal data, one may construct multiple base estimators $\{\hat{R}_{jk}^{(p,q)} : p = 1, \dots, N_j, q = 1, \dots, N_k\}$ for the same target Σ_{jk}. The final estimator can be potentially more efficient when the base estimators are properly aggregated.

The ensemble estimator for the ordinal-by-ordinal entry can be generally defined as

$$\hat{R}_{jk} = \sum_{1 \le p \le N_j} \sum_{1 \le q \le N_k} w_{jk}^{(p,q)} \hat{R}_{jk}^{(p,q)},$$

where for each (j,k) the weight $w_{jk}^{(p,q)}$ is defined as

$$0 \le w_{jk}^{(p,q)} \le 1, \quad \text{and} \quad \sum_{1 \le p \le N_j} \sum_{1 \le q \le N_k} w_{jk}^{(p,q)} = 1. \quad (4)$$

Similarly, the ensemble estimator for the ordinal-by-continuous entry is defined as

$$\hat{R}_{jk} = \sum_{1 \le p \le N_j} w_j^{(p)} \hat{R}_{jk}^{(p)}.$$

In the following, we illustrate two concrete examples for aggregating the base estimators.

Example 1. Simple Average Ensemble A natural idea is to take simple average of base estimators for each entry. This is equivalent to the ensemble estimator with the uniform weight $w_{jk}^{(p,q)} = 1/(N_j N_k)$ and $w_j^{(p)} = 1/N_j$. Such simple approach is known to be efficient when the base estimators are calculated from independent data, see (Battey et al., 2015; Lee et al., 2015). However, our base estimators are possibly correlated and therefore the simple average estimator may not be optimal. As confirmed in section 5 ensemble estimators with a non-uniform weight may outperform the simple average method.

Example 2. Data Adaptive Ensemble Consider the following weighted ensemble rank-based estimator, where the weights are determined by the bivariate log likelihood with plug-in estimates of cutoffs. Intuitively, the base estimators associated with higher likelihood should enjoy higher weights, because the estimate is likely to be closer to the truth. Specifically, in the case where X_{ij} and X_{ik} are both ordinal, we define the weight as

$$w_{jk}^{(p,q)} = \frac{\exp\left\{l(\hat{R}_{jk}^{(p,q)})\right\}}{\sum_{\substack{1 \le m \le N_j \\ 1 \le n \le N_k}} \exp\left\{l(\hat{R}_{jk}^{(m,n)})\right\}},$$

where

$$l(\hat{R}_{jk}^{(p,q)}) = \sum_{i=1}^{n} \log\left(\int_{\hat{U}_{ij}} \int_{\hat{U}_{ik}} \phi_2(x_1, x_2; \hat{R}_{jk}^{(p,q)}) \mathrm{d}x_1 \mathrm{d}x_2\right)$$

is the bivariate log likelihood function and

$$\hat{U}_{ij} = [\hat{\triangle}_j^{(X_{ij})}, \hat{\triangle}_j^{(X_{ij}+1)}]$$

with $\hat{\triangle}_j^{(0)} = -\infty$ and $\hat{\triangle}_j^{(N_j+1)} = \infty$. The weight $w_{jk}^{(p,q)}$ is proportional to the bivariate likelihood function of (X_{ij}, X_{ik}) evaluated at the estimate $\hat{R}_{jk}^{(p,q)}$. Thus, unlike the simple average method, we can adaptively reweight the base estimator $\hat{R}_{jk}^{(p,q)}$ according to whether it provides a good fit to the data. Similarly, when X_{ij} is ordinal and X_{ik} is continuous we have

$$w_{jk}^{(p)} = \frac{\exp\left\{l(\hat{R}_{jk}^{(p)})\right\}}{\sum_{1 \le m \le N_j} \exp\left\{l(\hat{R}_{jk}^{(m)})\right\}},$$

where

$$l(\hat{R}_{jk}^{(p)}) = \sum_{i=1}^{n} \log\left(\int_{\hat{U}_{ij}} \phi_2(x_1, X_{ik}; \hat{R}_{jk}^{(p)}) \mathrm{d}x_1\right).$$

3.2 Learning and testing the latent graph structure

To learn the latent graph structure, it suffices to estimate the sparsity pattern of the precision matrix $\mathbf{\Omega}$. Write $\mathbf{\Omega} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_d)$ and denote the true value of $\mathbf{\Omega}$ by $\mathbf{\Omega}^* = (\boldsymbol{\beta}_1^*, \dots, \boldsymbol{\beta}_d^*)$. We estimate $\boldsymbol{\beta}_k$ by the following quadratic loss with ℓ_1 penalty

$$\hat{\boldsymbol{\beta}}_k = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\{\frac{1}{2}\boldsymbol{\beta}^T \tilde{\boldsymbol{R}} \boldsymbol{\beta} - \boldsymbol{e}_k^T \boldsymbol{\beta} + \lambda ||\boldsymbol{\beta}||_1\}, \quad (5)$$

where $\boldsymbol{e}_k$ is vector with a 1 in the kth coordinate and 0's elsewhere. λ is a tuning parameter and $\tilde{\boldsymbol{R}}$ is the projection of $\hat{\boldsymbol{R}}$ onto the cone of positive definite matrices, that is, $\tilde{\boldsymbol{R}} = \operatorname{argmin}_{\boldsymbol{R} \succ 0} ||\hat{\boldsymbol{R}} - \boldsymbol{R}||_{\max}$. Similar to Liu and Luo (2012); Cai et al. (2011), the final estimator $\hat{\mathbf{\Omega}}$ is obtained after a symmetrization step

$$\hat{\Omega}_{jk} = \hat{\Omega}_{kj} = \hat{\boldsymbol{\beta}}_{jk} \mathbb{1}(|\hat{\boldsymbol{\beta}}_{jk}| < |\hat{\boldsymbol{\beta}}_{kj}|) + \hat{\boldsymbol{\beta}}_{kj} \mathbb{1}(|\hat{\boldsymbol{\beta}}_{kj}| < |\hat{\boldsymbol{\beta}}_{jk}|).$$

We note that similar to the Gaussian copula model, we may modify many existing procedures by plugging $\hat{\boldsymbol{R}}$ into these methods , such as graphical lasso Friedman et al. (2008) and CLIME Cai et al. (2011), to learn the latent precision matrix. However these methods do not provide a good framework for statistical inference for our model. In contrast, the proposed method not only gives a consistent estimator but for testing and constructing confidence region, our approaches also build upon the estimator defined in (5).

A similar quadratic optimization approach to (5) with the sample covariance matrix is called SCIO (Liu and Luo, 2012). In our approach, we plug in the projected estimator $\tilde{\boldsymbol{R}}$ rather than the ensemble estimator $\hat{\boldsymbol{R}}$ itself, since $\hat{\boldsymbol{R}}$ may not be positive semidefinite. In practice, we find that the optimization (5) tends to be more robust than that with $\hat{\boldsymbol{R}}$. More importantly, with the projected estimator $\tilde{\boldsymbol{R}}$, one can establish a variety of error bounds for $\hat{\boldsymbol{\beta}}$, which are essential for the statistical inference. The theoretical properties of $\hat{\boldsymbol{\beta}}$ are established in section 4.

In the following, we consider how to quantify the estimation uncertainty of the conditional independence between two pre-specified variables of interest. In particular, we aim to test the null hypothesis $H_0 : \Omega^*_{jk} = 0$. The existing high-dimensional inference methods for Gaussian (or transelliptical) graphical models rely on the Gaussian (or elliptical) structures and are not directly applied to our problem. In general, the maximum likelihood estimation is the default approach for statistical inference. However, our likelihood function is highly intractable due to the presence of high-dimensional integral for the ordinal data, which makes the likelihood based inference infeasible. To tackle this challenge, we borrow the loss function (5) in the proposed estimation procedure and view it as a pseudo-likelihood, and extend the likelihood based inference using the score function along the Stein's least favorable direction (Ning et al., 2017) to our pseudo-likelihood.

To simplify notation, denote the kth column of $\boldsymbol{\Omega}$ as $\boldsymbol{\beta}$, which is further partitioned as $\boldsymbol{\beta} = (\theta, \boldsymbol{\gamma}^T)^T$. Consider a partitioning of $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

where Σ_{11} is scalar and $\boldsymbol{\Sigma}_{22}$ is $(d-1)\times(d-1)$. It can be shown that $\boldsymbol{v}^* = (1, -\boldsymbol{w}^{*T})^T$ is the least favorable direction for the pseudo-likelihood (5), where $\boldsymbol{w}^* = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. The gradient of the pseudo-likelihood (i.e, pseudo-score) is simply $\tilde{\boldsymbol{R}}\boldsymbol{\beta} - \boldsymbol{e}_k$.

To construct the test statistic, one has to analyze the stochastic structure of the estimator $\tilde{\boldsymbol{R}}$. However, this analysis seems mathematically intractable due to the projection of the estimator. To get around this issue, when defining the test statistic based on the pseudo-score in (5), we replace $\tilde{\boldsymbol{R}}$ in the loss function by the ensemble estimator $\hat{\boldsymbol{R}}$. With a nice U-statistics structure of $\hat{\boldsymbol{R}}$, the pseudo-score function with $\hat{\boldsymbol{R}}$ is more convenient.

We define the pseudo-score function in the direction of $\boldsymbol{v}^*$ as

$$S(\boldsymbol{\beta}) = \boldsymbol{v}^{*T}(\hat{\boldsymbol{R}}\boldsymbol{\beta} - \boldsymbol{e}_k).$$

By the definition of $\boldsymbol{v}^*$, one can estimate $\boldsymbol{v}^*$ by $\hat{\boldsymbol{v}} = (1, -\hat{\boldsymbol{w}}^T)^T$, where

$$\hat{\boldsymbol{w}} = \text{argmin}\, ||\boldsymbol{w}||_1 \quad s.t.\ ||\tilde{\boldsymbol{R}}_{12} - \boldsymbol{w}^T\tilde{\boldsymbol{R}}_{22}||_\infty \le \lambda', \quad (6)$$

where λ' is a tuning parameter. Since $\boldsymbol{w}$ is a sparse vector by $||\boldsymbol{w}^*||_0 = ||\boldsymbol{\Omega}^*_{21}||_0$, the estimator $\hat{\boldsymbol{w}}$ in (6) is a sparse approximation of the true parameter $\boldsymbol{w}^*$. Denote $\boldsymbol{\Phi}^* = \lim_{n\to\infty} \text{Var}_{\boldsymbol{\beta}^*}(n^{1/2}(\hat{\boldsymbol{R}}\boldsymbol{\beta}^* - \boldsymbol{e}_k))$ and $\sigma^* = \boldsymbol{v}^{*T}\boldsymbol{\Phi}^*\boldsymbol{v}^*$. Let $\hat{\boldsymbol{\beta}}_0 = (0, \hat{\boldsymbol{\gamma}}^T)^T$ be the estimate under the null hypothesis and $\hat{\sigma}$ be a consistent estimator of σ^*. Our score test statistic is defined as

$$T_n = \sqrt{n}\hat{S}(\hat{\boldsymbol{\beta}}_0)/\hat{\sigma}^{1/2},$$

where $\hat{S}(\boldsymbol{\beta}) = \hat{\boldsymbol{v}}^T(\hat{\boldsymbol{R}}\boldsymbol{\beta} - \boldsymbol{e}_k)$. In the next section, we will show that $T_n \to_d N(0,1)$ under H_0, and thus it can be used to evaluate the validity of the null hypothesis. The test statistic T_n can be equivalently used to construct confidence interval for θ; due to space limit please see the supplementary materials for details.

4 Theoretical properties[1]

In this section we analyze the theoretical properties of the ensemble rank-based estimator of $\boldsymbol{\Sigma}$ and pseudo-score test statistic. Here we focus on the simple average estimator, and the theory can be easily generalized to the weighted case. We consider the following assumptions:

Assumption 1. $|\Sigma^*_{jk}| \le 1-\delta$, $\forall 1 \le j < k \le d_1$ for some constant $\delta > 0$.

Assumption 2. $|\triangle_j^{(p)}| \le M$, $\forall j = 1,\dots,d_1, p = 1,\dots,N_j$ for some constant M.

Assumption 3. $||\boldsymbol{\Omega}^*||_1 \le M_n$.

Assumption 4. $\frac{1}{\tau} \le \lambda_{\min}(\boldsymbol{\Sigma}^*)$ for some constant $\tau > 0$.

Assumptions 1 and 2 are mild technical assumptions. Assumption 1 makes sure that there is no perfect collinearity for any pair of $f_j(Z_j)$ and $f_k(Z_k)$. Assumption 2 is mainly used to control the variation of $F^{-1}(\tau, \triangle_j, \triangle_k)$ with respect to $(\tau, \triangle_j, \triangle_k)$. It can be shown that under these assumptions $F^{-1}(\tau, \triangle_j, \triangle_k)$ is Lipschitz in τ uniformly (Fan et al., 2017). Assumption 3 controls the magnitude of $\boldsymbol{\Omega}^*$, which is commonly used for the precision matrix estimation (Cai et al., 2011). Note that we allow M_n to increase with n. Assumption 4 implies the standard restricted eigenvalue (RE) condition.

Proposition 1. Under Assumption 1 and 2, it holds that

$$\mathbb{P}(||\hat{\boldsymbol{R}} - \boldsymbol{\Sigma}^*||_{\max} \le C\sqrt{\frac{\log d}{n}}) \ge 1 - d^{-1},$$

[1]The proofs are given in the supplementary material.

where C is a constant independent of n and d.

Theorem 1. Let s^* be the maximum degree of the graph and $\lambda \asymp M_n\sqrt{\frac{\log d}{n}}$. Under Assumption 1 - 4, with probability as least $1 - d^{-1}$, we have

$$||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*||_1 \lesssim M_n s^* \sqrt{\frac{\log d}{n}}, \tag{7}$$

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T \tilde{\boldsymbol{R}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \lesssim M_n^2 \frac{s^* \log d}{n}. \tag{8}$$

Theorem 1 shows that our estimator $\hat{\boldsymbol{\beta}}$ achieves the same convergence rate as the SCIO estimator Liu and Luo (2012) and CLIME Cai et al. (2011) for GGMs.

To establish the asymptotic distribution of the test statistic T_n, we additionally make the following assumption:

Assumption 5. $\lambda_{\max}(\boldsymbol{\Sigma}^*) \leq \kappa$ for some constant $\kappa > 0$.

This assumption has been made in Ren et al. (2015); Jankova et al. (2015); Janková and van de Geer (2017) to study the testing problems for GGMs. It is natural for many important classes of covariance matrices, e.g., bandable, Toeplitz, and sparse covariance matrices. The following theorem validates the asymptotic normality of the test statistic.

Theorem 2. Suppose Assumption 1 - 5 hold. If $\lambda \asymp \lambda' \asymp M_n\sqrt{\frac{\log d}{n}}$, $M_n^2 s^* \log d/\sqrt{n} = o(1)$ and $\sigma^* \geq K$ for some constant K, then under the null hypothesis we have

$$T_n \to_d N(0, 1).$$

When M_n is a constant, our condition $M_n^2 s^* \log d/\sqrt{n} = o(1)$ is identical to the best possible scaling under GGMs established by Ren et al. (2015). In general, our condition appears to be stronger, as M_n may increase with n. This can be viewed as the price paid for estimating the mixed graphical model, due to the sophisticated U-statistic structure of $\hat{\boldsymbol{R}}$. In particular, we show in the proof that the remainder term in the expansion of T_n scales with $M_n^2 s^* \log d/\sqrt{n}$, which needs to be small enough to prove the normal approximation result.

5 Simulation study

In this section we evaluate the empirical performance of our methods on a set of synthetic datasets. Our data generating procedure is similar to that in (Liu et al., 2012). Specifically, for the precision matrix $\boldsymbol{\Omega}$, we set $\Omega_{jj} = 1$, and $\Omega_{jk} = t a_{jk}$, where t is chosen to ensure the positive definiteness of $\boldsymbol{\Omega}$ and a_{jk} is a Bernoulli random variable with success probability $p_{jk} = (2\pi)^{-1/2}\exp\{||z_j - z_k||_2/2s\}$. Here $z_j = (z_j^{(1)}, z_j^{(2)})$ is sampled from a bivariate uniform distribution on $[0, 1]$, and s is chosen to control the sparsity level of $\boldsymbol{\Omega}$. In our case we choose s to make sure there are about 200 edges in each latent graph. After the latent precision matrix $\boldsymbol{\Omega}$ is created, we take the inverse and rescale it to obtain the correlation matrix $\boldsymbol{\Sigma}$, where the diagonal elements are 1. To generate the observed ordinal data, for X_j with $N_j + 1$ levels, we sample a sequence of cutoffs $C_j^{(k)} \sim \text{Uniform}[\Phi^{-1}((k - 0.5)\frac{1}{N_j+1}), \Phi^{-1}((k + 0.5)\frac{1}{N_j+1})]$ for $k = 1, \ldots, N_j$. This procedure ensures the randomness of the simulated cutoffs and guarantees that the difference of the number of samples at each level is not too large. In the following we fix the sample size $n = 200$, choose $t = 0.15$ and repeat the simulation 50 times for each scenario in low dimensional case $d = 50$ and high dimensional case $d = 250$:

(a) Simulate $\boldsymbol{X} = (X_1, \ldots, X_d)$, where $X_j = \sum_{i=1}^{2} \mathbb{1}(Z_j > C_j^{(i)})$ and $\boldsymbol{Z} \sim N(0, \boldsymbol{\Sigma})$.

(b) Simulate $\boldsymbol{X} = (X_1, \ldots, X_d)$, where $X_j = \sum_{i=1}^{4} \mathbb{1}(Z_j > C_j^{(i)})$ and $\boldsymbol{Z} \sim N(0, \boldsymbol{\Sigma})$.

(c) Simulate $\boldsymbol{X} = (X_1, \ldots, X_d)$, where $X_j = \sum_{i=1}^{2} \mathbb{1}(Z_j > C_j^{(i)})$ for $j = 1, \ldots, d/2$, $\boldsymbol{Z} \sim N(0, \boldsymbol{\Sigma})$, and $X_j = Z_j$ for $j = d/2 + 1, \ldots, d$.

(d) Simulate $\boldsymbol{X} = (X_1, \ldots, X_d)$, where $X_j = \sum_{i=1}^{2} \mathbb{1}(Z_j > C_j^{(i)})$ for $j = 1, \ldots, d/2$, $\boldsymbol{Z} \sim NPN(0, \boldsymbol{\Sigma}, f)$, where $f_j(x) = x^3$ for $j = d/2 + 1, \ldots, d$.

To investigate the empirical estimation error of the latent correlation and precision matrix, we compare our method LMNPN[2] with (1) the rank-based estimator RBE (Fan et al., 2017) and (2) the elementwise maximum likelihood approach EMLE with plug-in estimates of cutoffs (Suggala et al., 2017). Since both method can not be directly applied to ordinal-mixed data, we adapt both methods to this case with simple extension. Due to space limit please refer to the supplementary material for more details.

Notice that the precision matrix estimation problem for mixed data is not considered in the class of nodewise regression methods (Lee and Hastie, 2015; Chen et al., 2014; Cheng et al., 2017; Yang et al., 2014a). We will only compare this class of methods for the graph recovery properties .

[2] In Table 1 and 2 LMNPN_1 stands for the LMNPN method with simple average ensemble and LMNPN_2 stands for data adaptive ensemble. The numbers in the parentheses are the simulation standard errors.

Table 1: Average estimation error of latent correlation matrix

d	Scenario	Estimation errors			
		LMNPN_1	LMNPN_2	RBE	EMLE
50	(a)	4.58	4.45	5.81	10.21
		(0.08)	(0.10)	(0.11)	(0.48)
	(b)	4.02	3.87	6.37	7.01
		(0.09)	(0.11)	(0.17)	(0.26)
	(c)	4.15	4.16	4.80	5.54
		(0.10)	(0.09)	(0.12)	(0.23)
	(d)	4.10	4.18	4.78	5.14
		(0.11)	(0.10)	(0.18)	(0.14)
250	(a)	23.16	21.82	29.46	50.60
		(0.06)	(0.09)	(0.15)	(0.87)
	(b)	20.00	19.38	32.48	72.01
		(0.10)	(0.05)	(0.15)	(0.62)
	(c)	20.49	19.10	23.92	21.06
		(0.08)	(0.11)	(0.03)	(0.34)
	(d)	20.45	19.39	23.89	21.12
		(0.04)	(0.07)	(0.03)	(0.36)

Table 2: Average estimation error of latent precision matrix

d	Scenario	Estimation errors			
		LMNPN_1	LMNPN_2	RBE	EMLE
50	(a)	2.66	2.65	2.81	3.04
		(0.09)	(0.11)	(0.09)	(0.09)
	(b)	2.43	2.41	2.78	2.98
		(0.11)	(0.14)	(0.12)	(0.14)
	(c)	2.50	2.51	2.65	2.76
		(0.14)	(0.11)	(0.14)	(0.12)
	(d)	2.55	2.55	2.72	2.86
		(0.09)	(0.10)	(0.09)	(0.11)
250	(a)	2.77	2.77	2.81	2.84
		(0.10)	(0.11)	(0.10)	(0.10)
	(b)	2.03	2.04	2.08	2.09
		(0.07)	(0.05)	(0.07)	(0.07)
	(c)	1.96	1.94	1.99	1.98
		(0.14)	(0.15)	(0.15)	(0.17)
	(d)	2.05	2.04	2.08	2.08
		(0.05)	(0.05)	(0.06)	(0.06)

Tables 1 and 2 report the mean estimation error in terms of the Frobenius norm. It suggests that the ensemble rank-based estimator suffers less loss of information than the RBE estimator and the EMLE estimator in terms of both correlation and precision matrix estimation. The improvement upon the RBE estimator to the proposed one is more transparent when the number of levels for the ordinal data grows from three to five, comparing case (a) and case (b). For the proposed method, the data adaptive ensemble performs generally better than simple average ensemble for latent correlation matrix estimation and achieves similar performance for latent precision matrix.

5.1 Graph Structure Recovery

To investigate the graph recovery performance, we define the number of false positive $\text{FP}(\lambda)$ and true positive $\text{TP}(\lambda)$ given an estimator $\hat{\boldsymbol{\Omega}}^{\lambda}$ as

$$\text{FP}(\lambda) := |\{(j,k):\ \Omega_{jk} = 0,\ \hat{\Omega}^{\lambda}_{jk} \neq 0\}|,$$
$$\text{TP}(\lambda) := |\{(j,k):\ \Omega^{\lambda}_{jk} \neq 0,\ \hat{\Omega}_{jk} \neq 0\}|$$

We further define the false positive rate $\text{FPR}(\lambda)$ and true positive rate $\text{TPR}(\lambda)$ as $\text{FPR}(\lambda) = \frac{\text{FP}(\lambda)}{d(d-1)/2-|\mathcal{E}|}$ and $\text{TPR}(\lambda) = \frac{\text{TPR}(\lambda)}{|\mathcal{E}|}$. In addition to the RBE and EMLE method, we compare the proposed LMNPN estimator via simple average ensemble with other approaches including: the weighted ℓ_1 penalized nodewise regression WNR (Cheng et al., 2017) and pairwise exponential family nodewise regression PNR (Chen et al., 2014). According to the ROC curves in Figure 1, the proposed method generally outperforms competing methods.

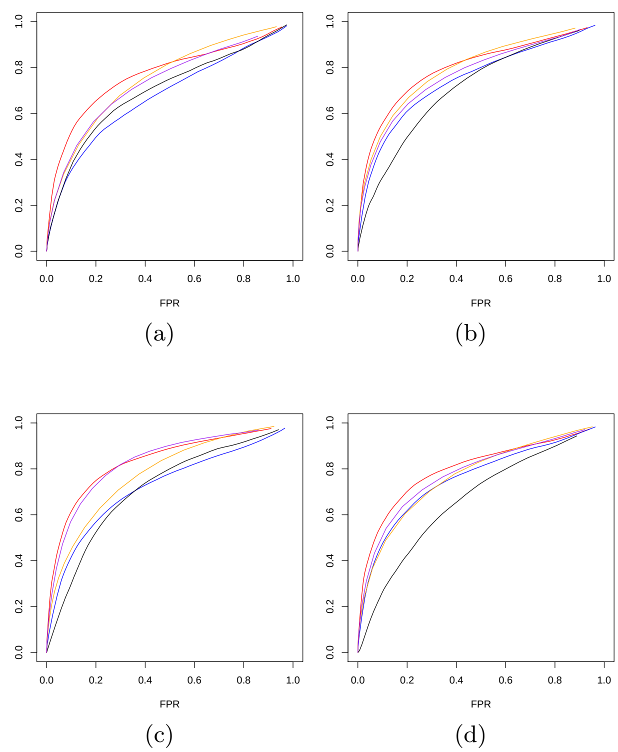
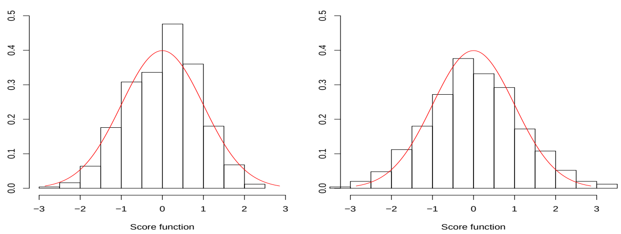


Figure 1: Left (a) - (d): ROC curves for graph recovery of the proposed method (———), RBE (———),WNR (———),PNR (———) and EMLE (———) when d = 50 for scenario (a),(b),(c) and (d)

5.2 Asymptotic Normality

Finally, we empirically validate the asymptotic normality of the score test statistic. For simplicity, we adopt scenario (c) to generate the mixed data. We consider the score test statistic T_n for the null hypothesis $H_0 : \Omega^*_{jk} = 0$ for some pre-specified (j, k). Two cases are studied: case (1) X_{ij} and X_{ik} are both ordinal and case (2) X_{ij} is ordinal and X_{ik} are continuous. The tuning parameters are chosen by the cross-validation. The results in Figure 2 implies that the proposed test statistic agrees well with the standard normal distribution, which validates our Theorem 2.



(a) (b)

Figure 2: Histograms for the score test statistic for (a): an ordinal-by-ordinal entry and (b): ordinal-by-continuous entry with the standard normal density superimposed over 500 iterations.

6 Real data application

We analyze the NKI breast cancer dataset (http://compbio.dfci.harvard.edu/) that contains 24481 gene expression and multiple clinical information of 337 patients with primary breast carcinomas (Van't Veer et al., 2002). The proposed method is advantageous when researchers want to simultaneously study the association among genes, clinical measures, and phenotype-genotype associations. For the purpose of our analysis, we filtered the data by removing all patients, gene expressions and clinical information with missing entries, and select the top 500 expressed loci. We then apply the proposed method on the filtered dataset with a tuning parameter that controls the sparsity level of the latent graph at around two percent.

According to Figure 3, the latent graph indicates associations between two ordinal clinical features (oestrogen receptor (ER) and tumor grade (Grade)) and several genes. In particular, ER has been shown to be a key driven factor of breast cancers (Carroll, 2016), and thus the strong associations between ER and multiple highly expressed gene loci in breast cancer patients can be justified. In addition, our latent graph also shows association between the breast tumor grade and locus NM_020974 (Homo sapiens signal peptide, CUB domain and EGF like domain containing 2 (SCUBE2), transcript variant 1, mRNA). According to Poorhosseini et al. (2016), over-expression of SCUBE2 has

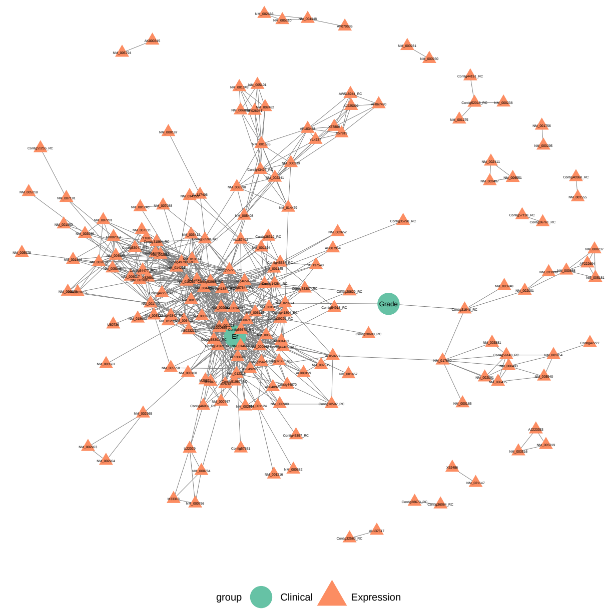


Figure 3: The estimated latent graph for the mixed breast cancer dataset using the proposed method. Nodes (genes) without any edge are omitted.

been shown to be associated with breast cancer recurrence . In another study, SCUBE2 has been shown to suppresses proliferation in breast tumor cell and confers a favorable prognosis in invasive breast cancer (Cheng et al., 2009). These results confirm the validity of the proposed method, and we expect other associations unveiled by the latent graph can be helpful for future research.

7 Conclusion

In this work we present and analyze a general framework for estimating and inferring a graphical model, called Latent Mixed Gaussian Copula Model. The proposed framework is shown to be effective both theoretically and empirically. Since most methods did not consider graphical models tailored to ordinal-mixed data and the fact that such mixed data are ubiquitous in practice, we believe our work fill this gap by providing a convenient and efficient modeling framework with theoretical guarantees for the community.

References

Takeshi Amemiya. Bivariate probit analysis: Minimum chi-square methods. *Journal of the American Statistical Association*, 69(348):940–944, 1974.

JR Ashford and RR Sowden. Multi-variate probit analysis. *Biometrics*, pages 535–546, 1970.

Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine learning research*, 9(Mar):485–516, 2008.

Rina Foygel Barber and Mladen Kolar. Rocket: Robust confidence intervals via kendall's tau for transelliptical graphical models. *arXiv preprint arXiv:1502.07641*, 2015.

Heather Battey, Jianqing Fan, Han Liu, Junwei Lu, and Ziwei Zhu. Distributed estimation and inference with statistical guarantees. *arXiv preprint arXiv:1509.05457*, 2015.

T Tony Cai, Weidong Liu, and Harrison H Zhou. Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *The Annals of Statistics*, 44(2):455–488, 2016.

T Tony Cai, Zijian Guo, et al. Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *The Annals of statistics*, 45(2):615–646, 2017.

Tony Cai, Weidong Liu, and Xi Luo. A constrained l1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

Jason Carroll. Eje prize 2016: Mechanisms of estrogen receptor (er) gene regulation in breast cancer. *European journal of endocrinology*, pages EJE–16, 2016.

Shizhe Chen, Daniela M Witten, and Ali Shojaie. Selection and estimation for mixed graphical models. *Biometrika*, 102(1):47–64, 2014.

Chien-Jui Cheng, Yuh-Charn Lin, Ming-Tzu Tsai, Ching-Shyang Chen, Mao-Chih Hsieh, Chi-Long Chen, and Ruey-Bing Yang. Scube2 suppresses breast tumor cell proliferation and confers a favorable prognosis in invasive breast cancer. *Cancer Research*, 69(8):3634–3641, 2009.

Jie Cheng, Tianxi Li, Elizaveta Levina, and Ji Zhu. High-dimensional mixed graphical models. *Journal of Computational and Graphical Statistics*, 26 (2):367–378, 2017.

Mathias Drton and Marloes H Maathuis. Structure learning in graphical modeling. *Annual Review of Statistics and Its Application*, 4:365–393, 2017.

Jianqing Fan, Yuan Liao, and Han Liu. An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal*, 19(1), 2016.

Jianqing Fan, Han Liu, Yang Ning, and Hui Zou. High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):405–421, 2017.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

Quanquan Gu, Yuan Cao, Yang Ning, and Han Liu. Local and global inference for high dimensional nonparanormal graphical models. *arXiv preprint arXiv:1502.02347*, 2015.

Jian Guo, Elizaveta Levina, George Michailidis, and Ji Zhu. Graphical models for ordinal data. *Journal of Computational and Graphical Statistics*, 24 (1):183–204, 2015.

Jana Janková and Sara van de Geer. Honest confidence regions and optimality in high-dimensional precision matrix estimation. *Test*, 26(1):143–162, 2017.

Jana Jankova, Sara Van De Geer, et al. Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, 9(1):1205–1229, 2015.

Adel Javanmard and Andrea Montanari. Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909, 2014.

Maurice George Kendall. Rank correlation methods. 1948.

Steffen L Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

Jason D Lee and Trevor J Hastie. Learning the structure of mixed graphical models. *Journal of Computational and Graphical Statistics*, 24(1):230–253, 2015.

Jason D Lee, Yuekai Sun, Qiang Liu, and Jonathan E Taylor. Communication-efficient sparse regression: a one-shot approach. *arXiv preprint arXiv:1503.04337*, 2015.

Han Liu, John Lafferty, and Larry Wasserman. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, 10(Oct):2295–2328, 2009.

Han Liu, Fang Han, Ming Yuan, John Lafferty, Larry Wasserman, et al. High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326, 2012.

Weidong Liu and Xi Luo. High-dimensional sparse precision matrix estimation via sparse column inverse operator. *arXiv preprint arXiv:1203.3896*, 2012.

Weidong Liu et al. Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics*, 41(6):2948–2978, 2013.

Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. *The annals of statistics*, pages 1436–1462, 2006.

Yang Ning, Han Liu, et al. A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics*, 45(1): 158–195, 2017.

SM Poorhosseini, M Hashemi, N Olyaei Alipour, A Izadi, E Moslemi, Z Ravesh, F Hashemi-Gorji, HR Kheiri, and VR Yassaee. New gene profiling in determination of breast cancer recurrence and prognosis in iranian women. *Asian Pacific journal of cancer prevention: APJCP*, 17(S3):155–160, 2016.

Zhao Ren, Tingni Sun, Cun-Hui Zhang, Harrison H Zhou, et al. Asymptotic normality and optimalities in estimation of large gaussian graphical models. *The Annals of Statistics*, 43(3):991–1026, 2015.

Adam J Rothman, Peter J Bickel, Elizaveta Levina, Ji Zhu, et al. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2: 494–515, 2008.

Arun Sai Suggala, Eunho Yang, and Pradeep Ravikumar. Ordinal graphical models: A tale of two approaches. In *International Conference on Machine Learning*, pages 3260–3269, 2017.

Sara Van de Geer, Peter Bühlmann, Ya'acov Ritov, Ruben Dezeure, et al. On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202, 2014.

Laura J Van't Veer, Hongyue Dai, Marc J Van De Vijver, Yudong D He, Augustinus AM Hart, Mao Mao, Hans L Peterse, Karin Van Der Kooy, Matthew J Marton, Anke T Witteveen, et al. Gene expression profiling predicts clinical outcome of breast cancer. *nature*, 415(6871):530, 2002.

Lingzhou Xue, Hui Zou, et al. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, 40 (5):2541–2571, 2012.

Eunho Yang, Yulia Baker, Pradeep Ravikumar, Genevera Allen, and Zhandong Liu. Mixed Graphical Models via Exponential Families. In Samuel Kaski and Jukka Corander, editors, *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pages 1042–1050, Reykjavik, Iceland, 22–25 Apr 2014a. PMLR.

Zhuoran Yang, Yang Ning, and Han Liu. On semiparametric exponential family graphical models. *arXiv preprint arXiv:1412.8697*, 2014b.

Ming Yuan and Yi Lin. Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1): 19–35, 2007.

Cun-Hui Zhang and Stephanie S Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242, 2014.