Project Results

Isaac Ray, Renat Sergazinov, Gözde Sert

Background and Introduction

Undirected graphical models enables to model multivariate distributions. Suppose we observe a p-dimensional sample $x = (x_1, \dots, x_p)$ from a multivariate Gaussian distribution with a non-singular covariance matrix. Then the conditional independence structure of the distribution can be represented with a graph G. The graph $G = (V, E)$ is characterized by a node set $V = (1, \dots, p)$ corresponding to the p variables, and an edge set E such that $(i, j) \in E$ if and only if x_i and x_j are conditionally dependent given all other variable. The goal is to estimate the underlying graph G from given n idd observations $x_1, \dots, x_p$. Several methods developed under this assumption however, in practice, the observations may not be identically distributed. In this paper they suppose the variability in the graph structure across observations depending on additional covariate information.

Current Literature and Previous Methods

There are several approaches to model heterogeneous graphs. Here we mention some of them.

- Without using covariate information: These methods depend on the criteria of first splitting the data into homogeneous groups and sharing information withing groups
- Adding the covariates into the mean structure of Gaussian graphical models as multiple linear regressions such that the mean is a continuous function of covariates. This approaches studied from a Bayesian perspective and a frequentist perspective. For this approach still uses the homogeneous graph structure for all observation which we do not want.
- Modeling the underlying covariance matrix as a function of the covariates. The main difficulty of this approach is to enforce sparsity in the precision matrix while being positive definite, as the sparsity in the covariance matrix does not normally carry to the precision matrix through matrix inversion.

Proposed formulation

Let $X \in \mathbb{R}^{n\times p}$ stand for the data matrix corresponding to n individuals on p variables. We denote the rows $X_i \in \mathbb{R}^p$ corresponding the observation for individual i and the columns $x_j \in \mathbb{R}^n$.\ The main goal of this paper is to learn the graph structure G from a collection of p-variate independent samples X_i, as a function of some extraneous covariates z_i corresponding to the samples. The only assumption on the dependence structure is that the graph parameters vary smoothly with respect to the covariates, that is, if z_i and z_j are similar, then the graph structure corresponding to X_i and X_j will be similar. In this method, a weighted pseudo-likelihood (W-PL) function to obtain a posterior distribution for the graph structure for a fixed individual, with the weights defined as a function of the covariates.

A weighted pseudo-likelihood (W-PL) approach

First we begin with introducing the pseudo-likelihood approach. Suppose there are n individuals, indexed $i = 1, .., n$. Let the i-th observation in the data set X be denoted as $X_i = (x_{i,1}, ..., x_{i,p})$, which corresponds to the i-th individual. Let $x_{i,-j} \in \mathbb{R}^{p-1}$ denote the vector of the i-th observation including all variables except $x_{i,j}$. This approach tries to model the conditional distribution of each of the x_j's given all other variables, denoted by $X_{-j} \in \mathbb{R}^{n\times(p-1)}$. Let the $p-1$-dimensional vector β_j indicate the regression effect on X_{-j} on x_j. Then the conditional likelihood of x_j denoted by $L(j)$ can be written as

$$L(j) = p(x_j|X_{-j}, \beta_j) \sim \prod_{i=1}^{n} \exp\left\{-(x_{i,j} - x_{i,-j}^T\beta_j)^2/2\sigma^2\right\}, \tag{1}$$

with a possibly sparse coefficient vector β_j. Then for a fixed graph G the pseudo-likelihood can be calculated as

$$L(G) = \prod_{j=1}^{n} L(j) = \prod_{j=1}^{n} p(x_j|X_{-j}, \beta_j). \tag{2}$$

In this paper different from the previous methods, we define a weighted version of this conditional likelihood for each individual. They assume that the underlying graph structure is a function of extraneous covariates z. Thus, we allow the coefficient vector β_j's to be different for different individuals,depending on the extraneous covariates. $\beta_j^l \in \mathbb{R}^{p-1}$ denotes the coefficient vector corresponding to the regression of the variable x_j on the remaining variables for individual l. Let z_i denote the covariate vector associated with the i-th individual and define $\mathbf{z} = (\mathbf{z}_1, ..., \mathbf{z}_n)$. Next, relative to the covariate z, we assign weights $w(z, \mathbf{z}_i) = \phi_\tau(\|z - \mathbf{z}_l\|)$ to every individual where ϕ_τ is the Gaussian density with mean 0 and variance τ^2. When $z = z_l$ corresponds to the l-th individual in the study, we use the notation $w_l(\mathbf{z}_i) = w(\mathbf{z}_l, \mathbf{z}_i)$ to denote the weight associated with the i-th individual. \ Proposed conditional working model: for $i = 1, ..., n$,

$$x_{ij}|x_{i,-j}, \mathbf{z}, z \sim N(x_{i,-j}^T \beta_j(z), \sigma^2 / w(z, \mathbf{z}_i))$$

Next let $W(z, \mathbf{z})$ denote the diagonal matrix $\text{Diag}(w(z, \mathbf{z}_1), \ldots, w(z, \mathbf{z}_n))$. The weighted conditional distribution function can be calculated as

$$p^w(x_j|\beta(\mathbf{z}, X_{-j}, \mathbf{z}, z)) = \left(\prod_{i=1}^{n} \sqrt{\frac{w(z, \mathbf{z}_i)}{2\pi\sigma_*^2}} \right) \exp\left\{ -\frac{(x_j - X_{-j}\beta_j(z))^T W(z, \mathbf{z})(x_j - X_{-j}\beta_j(z))}{2\sigma_*^2} \right\} \tag{3}$$

Problem

Despite this model demonstrating superior sensitivity to true dependence relations than competing methods, it suffers from lower specificity. Compared with competing methods from the `mgm` (Haslbeck and Waldorp 2020) and `JGL` (Danaher 2018)packages, the specificity gets substantially worse as the number of features increases. In the case of analyzing gene expression data, this could lead to worse outcomes than a less sensitive and more specific model since the cost of carrying out experiments which show a lack of a predicted relation may be very expensive. Further, validating a true but relatively weak relationship may not be desirable considering the cost. Ideally we want to **increase the specificity of the model without substantially hurting the model's sensitivity and speed**.

Attempted Solutions

We'll look at a variety of different modifications to the algorithm in order of least to most extensive changes necessary, and use the `covdepGE` package(Helwig et al. 2022) in order to generate data for simulation studies. For some of these solutions, the package's functions will be used without modification and changes will occur outside of the inference algorithm. Otherwise, any changes to the functions will be explicitly noted.

In general, we expect that the underlying cause of the specificity issue is due to having a common prior inclusion probability π across every spike-and-slab regression being performed despite the varying values of Z (and potentially X). We may expect that for certain values of Z, we have a different belief about whether variables in X are related. We'll approach this from 2 angles; first by trying to modify our variables and covariates in such a way as to make a common π a more appropriate choice, and then by modifying the algorithm to allow for multiple π values to be specified either a priori or as a function $\pi(X, Z)$ through something like clustering.

Feature Scaling Changes

The first approach will be to use a different or additional approach to feature scaling on X and/or Z in order to try and make a singular prior inclusion probability more appropriate.

Existing Feature Scaling through Normalization

Currently, the default behavior in the `covdepGE` function is to perform a columnwise Z-score Normalization on Z and a columnwise 0 centering on X. For brevity we'll denote this procedure by "normalization". The baseline performance under this scheme is given below. All experiments were run under 4 different setups each having different values for p and n, and data simulated using the `generateData` function. To assess sensitivity and specificity, we'll examine the number of false positives per sample and number of false negatives per sample across 100 replications of each simulation setup. So, in all cases lower numbers are desirable. First, we'll look at the baseline performance of the existing function with no changes to the default behavior.

p	n	Median	Mean
5	90	0.03	0.34
15	90	0.62	0.82
25	150	1.08	1.29
50	150	4.37	4.36

Table 1: False **positives** per sample - Normalized Z, Centered X

p	n	Median	Mean
5	90	1.03	0.99
15	90	1.32	1.41
25	150	0.92	0.91
50	150	1.13	1.19

Table 2: False **negatives** per sample - Normalized Z, Centered X

Max-Min Feature Scaling

Max-Min + Normalization Feature Scaling

Multiple Prior Inclusion Probabilities

Oracle or Informative Prior

Prior Inclusion through Covariate Clustering

Prior Inclusion at the Individual Level

Danaher, Patrick. 2018. "JGL: Performs the Joint Graphical Lasso for Sparse Inverse Covariance Estimation on Multiple Classes." https://CRAN.R-project.org/package=JGL.

Haslbeck, Jonas M. B., and Lourens J. Waldorp. 2020. "{Mgm}: Estimating Time-Varying Mixed Graphical Models in High-Dimensional Data" 93. https://doi.org/10.18637/jss.v093.i08.

Helwig, Jacob, Sutanoy Dasgupta, Peng Zhao, Bani Mallick, and Debdeep Pati. 2022. "covdepGE: Covariate Dependent Graph Estimation." https://CRAN.R-project.org/package=covdepGE.