Network estimation of multi-dimensional binary variables with application to divorce data

Yihe Yang, Renwen Luo, Bing Guo, Yingting Luo and Jianxin Pan

College of Mathematics, Sichuan University, Chengdu, 610063, China

Corresponding author’s e-mail: harry.y.yang@foxmail.com

Abstract. The cross-integration of statistics with social and scientific applications is one of the most popular topics in the past decade. Motivated by divorce data collected from the rural areas of Sichuan Province, China, we propose a new method to estimate the network of multiple binary variables, which specifies the dependence structures of multiple binary variables through the Gaussian copula model. Method of moments is employed to estimate the latent correlation matrix of the multiple binary variables. Alternating direction method of multipliers algorithm is then used to estimate the corresponding latent Gaussian network from the empirical latent correlation matrix. This method modifies the traditional estimation of latent Gaussian network from the perspectives of computational efficiency and positive definite guarantee. Analysis of the divorce data is conducted for illustration.

1. Introduction

The cross-integration of statistics with social and scientific applications is one of the most popular topics in the past decade. For example, many universities and institutions around the world are interested in big data collected from legal cases. This paper is motivated by a divorce data collected from rural area of Sichuan Province, China, which consists of 445 legal divorce cases and 27 variables. Every variable like whether the husband and wife live separately or whether the defendant approves to divorce, is binary variable. In this paper, we propose a new statistical method to estimate and visualize the dependence structure of binary variables by using the Gaussian copula model.

One of the most popular statistics to describe the pairwise dependence of two variables is the Pearson correlation coefficient. However, for discrete variables such as binomial variables and Poisson variables, the Pearson correlation coefficient may vary as the means change, and therefore is not appropriate to describe the pairwise dependence. In practice, Ising model [1] is one of the most popular statistical models to specify the dependence structure of multiple binary variables. Under the Ising model, the probability distribution function of multivariate binary variable $X$ takes the form

$$p(x) \propto \exp(\sum_{j} \beta_j x_j + \sum_{(j,k)} \theta_{jk}^* x_j x_k).$$

(1)

Especially, $E(X_j) = \beta_j^*$, and $X_i \perp X_j$ if $\theta_{jk}^* = 0$. Using the lasso [2], we can estimate and select the non-zero parameters $\theta_{jk}$, $1 \leq j < k \leq p$, and then recover a sparse network of the binary variables $X_1, ..., X_p$. Direct estimations of $\beta_1^*, ..., \beta_p^*$ and $\theta_{12}^*, ..., \theta_{p-1p}^*$ are very difficult because the value of $p(x)$ may be computationally intractable in real application. Hence, Ravikumar et al. [3] and Xue et al. [4] considered to minimize the logarithm conditional likelihood of the Ising model.
Latent Gaussian copula model [5] provides a more interpretable way to describe the dependence structure of multi-dimensional binary variables than the Ising model. Generally, the Gaussian copula model assumes the observed discrete variables $X_i, X_j$ are indeed determined by latent continuous variable $Z_i, Z_j$. Under Gaussian copula model, i.e., the pair $(Z_i, Z_j)$ follows the bivariate Gaussian distribution, and determines the means and covariance of $X_i, X_j$ by

$$
\mu_i = \Pr(Z_i \leq \Delta_i) = \Phi(\Delta_i), \quad \mu_j = \Pr(Z_j \leq \Delta_j) = \Phi(\Delta_j), \quad (2)
$$

$$
\Pr(X_iX_j = 1) = \Pr(Z_i \leq \Delta_i, Z_j \leq \Delta_j) = \Phi_2(\Delta_i, \Delta_j | \mathcal{C}_{ij}), \quad (3)
$$

where $\Delta_i$ and $\Delta_j$ are fixed but unknown cutoffs, $\Phi(\cdot)$ is standard normal cumulative distribution function (CDF), $\Phi_2(\cdot, | C)$ is standard bivariate normal CDF with the correlation coefficient $C$. Chaganty and Joe [6] proposed to specify the working correlation matrix of generalized estimating equation by Gaussian copula model when the response is binary distributed. Oman and Zucker [7], Huang and Pan [8] considered an alternative latent variable model to generate correlated binary variables. Fan et al. [9] provided an empirical method to estimate the latent correlation coefficient $\mathcal{C}_{ij}$ by using the Kendall-$\tau$ correlation coefficient of $X_i$ and $X_j$.

Although Fan et al. [9] have provided a method to estimate the sparse network of multiple binary variables, their method suffers from many computational problems. For example, their method uses the Kendall-$\tau$ correlation coefficient to yield the empirical latent correlation matrix. Since the computational cost of Kendall-$\tau$ coefficient is $O(n^2)$, their method is time-consuming when the sample size is large. Besides, the Fan’s method utilizes the ordinary graphical lasso (glasso) algorithm [10] to estimate the latent network, which neither is computationally efficient nor can ensure the latent network estimator to be positive definite. To solve the first problem, we consider using the linear covariance rather than the Kendall-$\tau$ correlation coefficient to yield the empirical latent correlation matrix, because the computational cost to estimate the linear covariance through the method of moments is only $O(n)$. As for the latter problem, we propose a new algorithm base on the alternating direction method of multipliers (ADMM) [11] method, which not only is faster than the ordinary graphical lasso algorithm but also can ensure the estimator of the latent network to be positive definite.

The rest of this paper is organized as follows. In section 2, we introduce the latent Gaussian copula model and show how to estimate the empirical latent correlation matrix by using the method of moments. In section 3, we propose the ADMM-based glasso to estimate the latent Gaussian network from the empirical latent correlation matrix. In section 4, we apply the new method to analyze the divorce data. In section 5, we show the simulation study of our method with comparison to the Fan’s method. In section 6, we give the concluding remarks.

2. Latent Gaussian copula model

2.1. Settings

Throughout this paper we use the following notations. For a vector $\mathbf{a} = (a_j)_{p \times 1}$, let $||\mathbf{a}||_2^2 = \sum_j |a_j|^2$. For a matrix $\mathbf{A} = (A_{ij})_{p \times p}$, suppose $\sigma_{\max}(\mathbf{A})$ is the maximum eigenvalue of $\mathbf{A}$, $\sigma_{\min}(\mathbf{A})$ is the minimum eigenvalue of $\mathbf{A}$, $\mathbf{A}^+$ is the generalized inverse of $\mathbf{A}$, $||\mathbf{A}||_2 = \text{tr}(\mathbf{A}^T \mathbf{A})$, and $||\mathbf{A}||_2 = \sqrt{\sigma_{\max}(\mathbf{A}^T \mathbf{A})}$. Denote $\mathbf{1}_p$ be the $p \times p$ identity matrix, $\text{diag}(\mathbf{A})$ be a diagonal matrix with diagonal elements are diagonal elements of matrix $\mathbf{A}$, $\mathbf{1}(\cdot)$ be the indicator function. Suppose the eigenvalue decomposition: $\mathbf{A} = \mathbf{U} \text{diag}(\mathbf{D}_1, ..., \mathbf{D}_p) \mathbf{U}^T$, where the columns of $\mathbf{U}$ are eigenvectors of $\mathbf{A}$ and $\mathbf{D}_1, ..., \mathbf{D}_p$ are the eigenvalues of $\mathbf{A}$. For a smooth function $f(\cdot), f(\mathbf{A}) = \mathbf{U} \text{diag}(f(\mathbf{D}_1), ..., f(\mathbf{D}_p)) \mathbf{U}^T$.

2.2. Latent Gaussian copula network for binary data

Primarily, we present the definition of latent Gaussian model for multi-dimensional binary variable:
Definition 1. A \( p \)-dimensional binary vector \( X = (X_1, ..., X_p)^T \) is generated from a latent Gaussian copula model \( LGC(0, \mathbf{C}, \Delta) \), if there is a \( p \)-dimensional Gaussian vector \( Z = (Z_1, ..., Z_p)^T \sim \mathcal{N}(0, \mathbf{C}) \) such that \( X_j = 1(Z_j < \Delta_j) \) for \( j \in \{1, ..., p\} \). Note that \( \mathbf{C} \) is a symmetric positive definite matrix with \( \text{diag}(\mathbf{C}) = 1_p \) and \( \Delta = (\Delta_1, ..., \Delta_p)^T \) is a fixed vector of cutoffs.

The latent precision matrix \( \mathbf{\Theta} = \mathbf{C}^{-1} \) is a well-defined specification of the network of the multivariate binary variable \( X \). Specifically, an ordinary network \( G \), also known as undirected graph, is composed of a set of vertices \( V = \{1, ..., p\} \) and a set of edges \( E \subset V \times V \). Since \( Z \) follows a multivariate Gaussian distribution, a pair \((i, j)\) is included in the set of edges \( E \) if and only if \( Z_i \) is conditionally dependent on \( Z_j \), in the presence of \( Z_{-ij} = (Z_1, ..., Z_{i-1}, Z_{i+1}, ..., Z_{j-1}, Z_{j+1}, ..., Z_p)^T \).

Similarly, the pairs of components which are not included in \( E \) are all conditionally independent in the present of \( Z_{-ij} \), and correspond to the zero entries in the precision matrix. In other words,

\[
Z_i \perp Z_j | Z_{-ij} \Leftrightarrow \theta_{ij} = 0.
\]

Hence, it defines whether the pair \((i, j)\) is in \( E \) by verifying whether \( \theta_{ij} \) is not zero. Since the Gaussian variable \( Z \) determines the binary variable \( X \), it is feasible to specify the network of \( X \) by \( \mathbf{\Theta} \).

2.3. Empirical correlation matrix estimator
The latent correlation matrix is linked to the joint probability \( \text{Pr}(X_i, X_j = 1) \) through the equation:

\[
\text{Pr}(X_i, X_j = 1) = \Phi_2(\delta_i, \delta_j | \mathbf{C}_{ij}).
\]

For binary variables, the Kendall-\( \tau \) coefficient is equal to \( \tau_{ij} = 2\text{Pr}(X_i, X_j = 1) - 2\mu_i\mu_j \), while the linear covariance of \( X_i, X_j \) is equal to \( E((X_i - \mu_i)(X_j - \mu_j)) = \text{Pr}(X_i, X_j = 1) - \mu_i\mu_j \). Based on this connection, Fan et al. [9] proposed a two-steps estimation, which first yield an empirical estimator of the latent correlation matrix \( \mathbf{C} \) from the empirical Kendall-\( \tau \) coefficients and then estimates \( \mathbf{\Theta} \) from the empirical estimator of \( \mathbf{C} \). We observe that both the linear covariance and Kendall-\( \tau \) coefficient can link to the latent correlation coefficient \( \mathbf{C}_{ij} \), especially, the former computational cost is much less than the latter cost. Hence, we propose to estimate \( \mathbf{C}_{ij} \) from the empirical linear covariance.

Suppose that there are \( n \) samples of the variable \( X \), in the form of matrix, the \( n \) samples are \( \mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_p)^T \), where \( \mathbf{X}_i = (X_{i1}, ..., X_{in})^T, \quad j \in \{1, ..., p\} \). Let \( \bar{\mathbf{X}}_i = \sum_{s=1}^{n} X_{is} / n, \bar{\mathbf{X}}_j = \sum_{s=1}^{n} X_{sj} / n, \hat{\delta}_i = \Phi^{-1}(\bar{\mathbf{X}}_i), \hat{\delta}_j = \Phi^{-1}(\bar{\mathbf{X}}_j), \) and

\[
\Phi_2(\hat{\delta}_i, \hat{\delta}_j | \bar{\mathbf{C}}_{ij}) = (\mathbf{X}_j - \bar{\mathbf{X}}_j)^T(\mathbf{X}_j - \bar{\mathbf{X}}_j) / n + \bar{\mathbf{X}}_j \bar{\mathbf{X}}_j.
\]

We then solve the equation to find the solution \( \hat{\mathbf{C}}_{ij} \). Note that the latent covariance matrix estimator \( \hat{\mathbf{C}} = (\hat{\mathbf{C}}_{ij}) \) has multiple negative eigenvalues in general. Hence, we apply the projection technique to find an alternative estimator in set \{\mathbf{C} : \mathbf{C} > 0\}, otherwise it is not possible to recover a consistent network estimator in practice. Define \( \hat{\mathbf{C}} = \mathbf{U} \text{diag}(D_1, ..., D_p) \mathbf{U}^T \). Then the projection of \( \hat{\mathbf{C}} \) onto the set \{\mathbf{C} : \mathbf{C} > 0\} is \( \check{\mathbf{C}} = \mathbf{U} \text{diag} \left( \max(D_1, \delta), ..., \max(D_p, \delta) \right) \mathbf{U}^T \), where \( \delta \) is a very small constant, such as \( \delta = 10^{-10} \).

3. Graphical lasso with eigenvalue constraint
The glasso is a standard technique to estimate a sparse inverse matrix of \( \hat{\mathbf{C}} \). This method is based on a coordinate descent algorithm, which needs to solve a \((p - 1)\)-dimensional linear regression in each step. Consequently, if the dimension of the precision matrix \( \mathbf{\Theta} \) is large, the glasso algorithm is extremely computationally expensive. On the other hand, the ordinary glasso algorithm cannot ensure the outputting precision matrix estimator \( \hat{\mathbf{\Theta}} \) to be positive definite. Hence, we modify the traditional glasso algorithm by the technique of ADMM algorithm. Consider the following modified glasso:
\[ \Theta = \arg\min_{\Theta} \{ \text{tr}(C\Theta) - \log\det(\Theta) + \lambda \sum_{i=1}^{p} \sum_{j>i}^{p} |\Theta_{ij}| \}, \tag{7} \]

subject to \( \Theta > \delta I \), where \( \lambda \) is the tuning parameter of the lasso, and \( \delta \) is another small constant. The ADMM algorithm converts (11) into the following function with three constraints:

\[ Q(\Theta, \Omega, \Gamma) = \text{tr}(C\Theta) - \log|\Theta| + \lambda \sum_{i=1}^{p} \sum_{j>i}^{p} |\Omega_{ij}|, \tag{8} \]

subject to \( \Theta = \Omega, \Theta = \Gamma, \Gamma > \delta I \). The corresponding Lagrange augmented function of (12) is

\[ Q(\Theta, \Omega, \Gamma, \Lambda_1, \Lambda_2) = \text{tr}(C\Theta) - \log|\Theta| + \lambda \sum_{i=1}^{p} \sum_{j>i}^{p} |\Omega_{ij}| + \text{tr}(\Lambda_1(\Theta - \Gamma)) + \frac{\rho}{2} ||\Theta - \Omega||_F^2 \]

\[ + \text{tr}(\Lambda_2(\Theta - \Gamma)) + \frac{\rho}{2} ||\Theta - \Gamma||_F^2, \tag{9} \]

subject to \( \Gamma > \delta I \). Here \( \Lambda_1 \) and \( \Lambda_2 \) are the Lagrange multipliers corresponding to constraints \( \Theta = \Omega \) and \( \Theta = \Gamma \), respectively, and the quadratic terms \( \frac{\rho}{2} ||\Theta - \Omega||_F^2 \) and \( \frac{\rho}{2} ||\Theta - \Gamma||_F^2 \), which involve a parameter \( \rho \), are imposed such that the constraints are in smoother fashions.

The update of \( \Theta \) is

\[ \Theta^{(t+1)} = \arg\min_{\Theta} Q(\Theta, \Omega^{(t)}, \Gamma^{(t)}, \Lambda_1^{(t)}, \Lambda_2^{(t)}). \tag{10} \]

By some algebra computation, the solution of \( \Theta \) is \( \Theta^{(t+1)} = (-Q + (Q^2 + 8\rho I)^{1/2})/4\rho \), where \( Q = \bar{C} + \Lambda_1^{(t)} + \Lambda_2^{(t)} - \rho \Omega^{(t)} - \rho \Gamma^{(t)} \). As well, the update of \( \Omega \) is

\[ \Omega^{(t+1)} = \arg\min_{\Omega} Q(\Theta^{(t+1)}, \Omega, \Gamma^{(t)}, \Lambda_1^{(t)}, \Lambda_2^{(t)}). \tag{11} \]

The score equation of \( \Omega_{ij} \) is \( \lambda s_{ij} - \Lambda_{ij}^{(t)} + \rho \left( \Omega_{ij} - \Theta_{ij}^{(t+1)} \right) = 0 \) where \( s_{ij} \) is the subgradient of \( |\Omega_{ij}| \).

The solution of this equation is also explicit: \( \Omega_{ij}^{t+1} = \text{sign}(\Theta_{ij}^{(t+1)} + \Lambda_{ij}^{(t)}/\rho) \max(|\Theta_{ij}^{(t+1)} + \Lambda_{ij}^{(t)}/\rho| - \lambda/\rho, 0) \). Next, the update of \( \Gamma \) is

\[ \Gamma^{t+1} = \arg\min_{\Gamma > \delta I} (2^{-1} \rho \text{tr}(\Gamma^2) - \text{tr}(\Gamma(\rho\Theta^{(t+1)} + \Lambda_2^{(t)}))). \tag{12} \]

Similarly, the \( \Gamma \) has an explicit update as \( \Gamma^{t+1} = [\Theta^{(t+1)} + \Lambda_2^{(t)}/\rho, \delta I]_+ \). The updates of \( \Lambda_1^{k} \) and \( \Lambda_2^{k} \) become trivial, which are given by

\[ \Lambda_1^{(t+1)} = \Lambda_1^{(t)} + \rho(\Theta^{(t+1)} - \Omega^{(t+1)}), \quad \Lambda_2^{(t+1)} = \Lambda_2^{(t)} + \rho(\Theta^{(t+1)} - \Gamma^{(t+1)}). \tag{13} \]

The solutions in (10)-(13) are iterated until the minimizer \( \Omega^{(\infty)} \) is found. The sparse precision matrix estimator \( \hat{\Theta} \) is defined as \( \Omega^{(\infty)} \).

4. Real data analysis

In this section, we estimate the network of a divorce data collected from rural area of Sichuan Province, China. The divorce data consists of 445 legal divorce cases and 27 variables in terms of legal basis for judgements or judgement result. Part of important variables include: whether the judge decides to divorce (divorce), whether the couple lives separately (separ), whether the couple lives separately for more than two years (spe2year), whether the couple lives separately for more than one year in the case of rejection to divorce in the last judgment (ref1year), whether the defendant agree to divorce (agreement), and whether there is a divorce agreement (div.agr). These variables are emphasized because they are usually legal basis for judgement or judgement result.
We solve the glasso problem with the ADMM algorithm. Optimal tuning parameter is determined by the cross-validation. Specifically, we first subsample rows of the data matrix $\mathbf{X}$ into two disjoint submatrices $\mathbf{X}_{\text{train}}$ and $\mathbf{X}_{\text{test}}$. Then, we estimate the corresponding precision matrix $\hat{\mathbf{\Theta}}_{\text{train}}$ from $\mathbf{X}_{\text{train}}$ and calculate the latent correlation matrix $\hat{\mathbf{C}}_{\text{test}}$. Next, we measure cross-validation error (CVE) by the quadratic loss between $\hat{\mathbf{\Theta}}_{\text{train}}$ and $\hat{\mathbf{C}}_{\text{test}}$, i.e., $\text{CVE} = \| \hat{\mathbf{\Theta}}_{\text{train}} - \mathbf{I} \|_F^2$. The number of rows of $\mathbf{X}_{\text{train}}$ is $[0.6n]$ and the number of rows of $\mathbf{X}_{\text{test}}$ is $[0.4n]$. We replicate the procedure of cross-validation by 100 times and determine the lambda with the smallest average of CVE. Moreover, $\delta$ is set as 0.1.

![Figure 1. The latent network of divorce data.](image)

Figure 1 shows the network of the 27 legal variables. The variables $X_i, X_j$ are linked by a red line or a grey line if the sign of $\hat{\theta}_{ij}$ is negative or positive, respectively. Note that the sign of $\hat{\theta}_{ij}$ is opposite to the sign of correlation of two variables $X_i$ and $X_j$. Hence, the red line links two positively correlated legal variables. It is easy to see from Figure 1 that the legal variables that are linked to divorce directly are whether the couple lives separately (separation), whether the couple lives separately for more than two years (spe2year), and whether the couple lives separately for more than one year in the case of rejection to divorce in the last judgment (ref1year). According to the Marriage Law of the People’s Republic of China, it should be decided to divorce if the couple has been separated for more than two years or the couple has been separated for more than one year in the case of rejection to divorce in the last judgment. The network also reflects many facts of rural area of Sichuan Province. For example, the gamble is a common problem that causes a quarrel between husband and wife. Besides, it is very common for divorced couples to form a combined family with their own children, and the emotions of divorced families are generally fragile.

5. Simulation study
We conduct a simulation study to compare our proposed method and the Fan’s method [9] in terms of the estimation accuracy and computational efficiency. We randomly generate $p$-dimensional latent
variable \( Z_i \) from \( \mathcal{N}(0, \mathbf{C}) \). Here, we consider two specific structures of \( \mathbf{C} \). For structure-1, \( \mathbf{C} \) is a block diagonal correlation matrix with two submatrices \( \mathbf{C}_1 \) and \( \mathbf{C}_2 \). The first \( 20 \times 20 \) submatrix is \( \mathbf{C}_1 \):

\[
\mathbf{C}_1 = \begin{pmatrix}
\text{cs}_5(0.6) & 0.411^T & 0.211^T \\
0.411^T & \text{cs}_5(0.4) & 0.211^T \\
0.211^T & 0.111^T & \text{cs}_{20}(0.2)
\end{pmatrix},
\]

where \( \text{cs}_d(\rho) \) denotes the \( d \)-dimensional correlation matrix that is of compound symmetry structure with correlation coefficients \( \rho \). The other submatrix \( \mathbf{C}_2 \) is \( \mathbf{I}_{p-20} \) where \( p \) is the dimension of \( \mathbf{C} \). For structure-2, we set \( \mathbf{C} \) to be the AR(1) correlation matrix with correlation coefficient 0.75. The cutoffs \( \Delta_1, \ldots, \Delta_p \) for \( Z_i \) is randomly generated from normal distribution \( \mathcal{N}(0,1) \). Without losing generality, we fix the sample size \( n = 400 \) and let the dimension \( p \) increase from 20 to 60 with lag 5.

Figure 2. The computing time of joint probability, computing time of glasso algorithm, and the estimation accuracy of the outputting network.

Figure 2 shows the computing time of Kendall-\( \tau \) coefficient/linear covariance, the computing time of the standard/ADMM-based glasso, and the entropy loss of outputting estimator: 

\[
L_E = \text{tr}(\mathbf{C} \hat{\Theta}) - \log |\mathbf{C} \hat{\Theta}| - p,
\]

where \( \hat{\Theta} \) is the outputting estimator of the standard or ADMM-based glasso. It is easy to see that our new method outperforms the Fan’s method in terms of computational efficiency. The traditional glasso algorithm is much slower than our new method because the former needs to solve \( p \)-dimensional lasso regression in each step. On the contrary, our ADMM-based glasso does not have this problem: the updates of all terms are explicit. As for estimation accuracy, our method is still better than the tradition because it can ensure the outputting network to be positive definite.

6. Concluding remarks

In this paper, our aim is to model the network of multiple binary variables and analyze the divorce data collected from the rural areas of Sichuan Province, China. We employ the latent Gaussian copula model to specify the dependence structure of multiple binary variables, and the inverse of the latent correlation matrix is a well-defined network for multiple binary variables. The traditional method proposed by Fan et al. [9] is shown to be slow to estimate a high-dimensional latent network, and we modify this method from two aspects. First, we propose to estimate the correlation matrix \( \mathbf{C} \) by using moment estimation, which is computationally intensive. Second, we modify the traditional glasso
algorithm by ADMM algorithm with a sparse penalty and an eigenvalue constraint. As a result, the outputting network estimator is not only sparse but also positive definite. In real data analysis, our method successfully recovers the latent network of the legal variables. In the simulation study, this method still outperforms the traditional Fan’s method.

Acknowledgments
This work was supported by the National Key R&D Program of China (2018YFC0831900), Sichuan University Post-Doctor Research Project, the National Natural Science Foundation of China (11871357), and SCU LAIW.

References
[1] Besag J 1974 Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B*, 36(2): 192-225.
[2] Tibshirani R 1996 Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1): 267-288.
[3] Ravikumar P, Wainwright M J, Lafferty J D 2010 High-dimensional Ising model selection using l1-regularized logistic regression. *The Annals of Statistics*, 38(3): 1287-1319.
[4] Xue L, Zou H, Cai T 2012 Nonconcave penalized composite conditional likelihood estimation of sparse Ising models. *The Annals of Statistics*, 40(3): 1403-1429.
[5] Durante F, Sempi C 2016 *Principles of copula theory*. Boca Raton, FL: CRC press.
[6] Rao Chaganty N, Joe H 2004 Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Statistical Society: Series B*, 66(4): 851-860.
[7] Oman S D, Zucker D M 2001 Modelling and generating correlated binary variables. *Biometrika*, 88(1): 287-290.
[8] Huang Y, Pan J 2021 Joint generalized estimating equations for longitudinal binary data. *Computational Statistics & Data Analysis*, 155: 107110.
[9] Fan J, Liu H, Ning Y, et al. 2017 High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society: Series B*, 79(2): 405-421.
[10] Friedman J, Hastie T, Tibshirani R 2008 Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3): 432-441.
[11] Boyd S, Parikh N, Chu E, et al. 2010 Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Machine Learning*, 3(1): 1-122.