A computational method to estimate sparce multiple Gaussian graphical models

Rossella Onorati, Luigi Augugliaro, Angelo M. Mineo

Abstract In recent years several researchers have proposed the use of the Gaussian graphical model defined on a high dimensional setting to explore the dependence relationships between random variables. Standard methods, usually proposed in literature, are based on the use of a specific penalty function, such as the L1-penalty function. In this paper our aim is to estimate and compare two or more Gaussian graphical models defined in a high dimensional setting. In order to accomplish our aim, we propose a new computational method, based on glasso method, which let us to extend the notion of p-value.

Key words: Gaussian graphical models, glasso, model selection

1 Introduction

In recent years many researchers have developed new methods to estimate a genetic network which can bring useful results to elucidate the biological process of interest. From a pure statistical point of view, the problems that we have when we want to estimate a genetic network are concerned with the study of an high-dimensional data set, namely a data set in which the number of variables is larger than the sample size. Under this setting the dominant paradigm postulates that "what makes high dimensional statistical inference possible is the assumption that the parameter vector is sparse" [1].

Aim of this paper is to propose a new computational method to estimate and compare the topological structure of two Gaussian Graphical Models (GGMs) defined in a high dimensional setting. Inside the theory of this kind of graphical models, the assumption of sparsity is translated on the assumption that the concentration matrix, denoted by $\Omega = \{\omega_{ij}\}$, is a sparse matrix. To make this paper clearer, we briefly

Dipartimento di Scienze Statistiche e Matematiche "S. Vianelli", University of Palermo, e-mail: rossella.onorati@unipa.it; luigi.augugliaro@unipa.it; angelo.mineo@unipa.it

review the theory underlying the GGMs. A *graph* is defined as a pair $\mathscr{G} = \{V, E\}$, where the *vertex set* V is a finite set of p *vertices* and $E \subseteq V \times V$ is the set of ordered pairs of distinct vertices. A given *edge* $(i, j) \in E$ is called *undirected* if and only if the pair (j, i) belongs to the set E, otherwise it is called *directed*. When any edge belonging to E is undirected, the corresponding graph $\mathscr{G}$ is said *undirected*. The GGM is the most simple undirected graphical model where the vertices represent the variables and the edges represent the partial correlation between two variables given all the other variables. Formally two given random variables, say X_i and X_j, are conditional independent given all the other variables when the corresponding *partial correlation coefficient*, $\rho_{ij|V\setminus(ij)} = \omega_{ij}/\sqrt{\omega_{ii}\omega_{jj}}$ is equal to zero. To make parameter estimation and model selection in the GGM, [2] propose the following L_1-penalized likelihood estimator

$$\hat{\Omega}(\lambda) = \max_{\Omega} \ell(\Omega; S) - \lambda \|\Omega\|_1, \tag{1}$$

where S is the empirical variance/covariance matrix, $\ell(\Omega; S) = \log\det\Omega - tr(S\Omega)$ is the log-likelihood function and $\|\Omega\|_1 = \sum_{i\neq j} |\omega_{ij}|$. Then, the corresponding estimator is called graphical lasso (glasso). The optimal value of the turning parameter, denoted by $\hat{\lambda}_{opt}$, is chosen by means of the Bayesian Information Criterion (BIC). Although the estimated sparse concentration matrix, $\hat{\Omega}(\hat{\lambda}_{opt})$, can be useful, for example, to elucidate a specific biological process, there is a considerable interest in understanding if penalized partial correlation coefficients change going from an experimental condition to another one. In this paper we propose a permutation-based method to test the null hypothesis that there is no change in penalized partial correlation coefficients.

The remaining part of this paper is structured as follows. In the section 2 we explain the proposed permutation method and in the section 3 we evaluate the behavior of the proposed method through a simulation study. Conclusions are also given.

2 A permutation method

In this section we shall assume, without loosing in generality, that we are working with only two experimental conditions. Let $\mathbf{y}$ be a n-dimensional vector of labels identifying the two conditions, i.e. using y_i^+ to identify the ith unit belonging to the first condition and y_j^- to identify the jth unit belonging to the second condition, we can write:

$$\mathbf{y} = \left(y_1^+, \ldots, y_{n^+}^+, y_{n^++1}^-, \ldots, y_{n^++n^-}^-\right)'$$

where n^+ is the number of units belonging to the first condition, n^- is the number of units belonging to the other condition. Then, the sample size n is equal to $n^+ + n^-$. Let $\mathbf{X}$ be a $n \times p$ random matrix. Using the vector $\mathbf{y}$, the random matrix $\mathbf{X}$ can be defined joining the two random matrices $\mathbf{X}^+$ and $\mathbf{X}^-$ corresponding to the first and second condition, respectively. To formalize our method, we shall assume that the rows of the matrices $\mathbf{X}^+$ and $\mathbf{X}^-$ are normally distributed with $\mathbf{0}$ expected value and

concentration matrices Ω^+ and Ω^-. Let $\hat{\omega}_{ij}^+(\hat{\lambda}_{opt}^+)$ be the estimated L_1-penalized partial correlation coefficients corresponding to the vertices i and j estimated using $\mathbf{X}^+$. Similar notation is used to identify the quantities related to $\mathbf{X}^-$. A first simple method to compare $\hat{\omega}_{ij}^+(\hat{\lambda}_{opt}^+)$ and $\hat{\omega}_{ij}^-(\hat{\lambda}_{opt}^-)$ is by using the observed Euclidean distance, namely

$$d(i,j) = \left| \hat{\omega}_{ij}^+(\hat{\lambda}_{opt}^+) - \hat{\omega}_{ij}^-(\hat{\lambda}_{opt}^-) \right|. \tag{2}$$

To assess the statistical significance of the change in the penalized partial correlation coefficients, the following method to approximate the null distribution of (2) is proposed.

Let $\mathbf{y}_b$ be the b-th permutation of the vector $\mathbf{y}$ and let $\mathbf{X}_b^+$ and $\mathbf{X}_b^-$ the two corresponding permuted matrices. Using the *Bayesian Information Criterio* (BIC) we can estimate the two permuted sparse concentration matrices, denoted by $\widehat{\Omega}_b^+(\hat{\lambda}_{opt}^+)$ and $\widehat{\Omega}_b^-(\hat{\lambda}_{opt}^-)$, and then we can compute the permutation version of (2), denoted by $d_b(i,j)$. The permutation p-value can be easily computed repeating the previous step B times as shown in Table 1. Since the L_1-penalized Gaussian graphical model is a consistent variable selection method, in other words the estimated model contains the true model with probability tending to one, we can set $\hat{\omega}_{ij,b}^+$ and $\hat{\omega}_{ij,b}^-$ equal to zero if $\hat{\omega}_{ij}^+ = 0$ and $\hat{\omega}_{ij}^- = 0$. In this way we can significantly reduce the total amount of hypothesis that we have to test.

Table 1 Pseudo-code of the L_1-penalized differential sparse network analysis

Steps	*Algorithm*
1	for $b = 1$ to B
2	compute the permutation vector $\mathbf{y}_b$
3	use $\mathbf{y}_b$ to identify the matrix $\mathbf{X}_b^+$
4	use $\mathbf{X}_b^+$, the glasso and the BIC to compute $\widehat{\Omega}_b^+(\hat{\lambda}_{opt,b}^+)$
5	use $\mathbf{y}_b$ to identify the matrix $\mathbf{X}_b^-$
6	use $\mathbf{X}_b^-$, the glasso and the BIC to compute $\widehat{\Omega}_b^-(\hat{\lambda}_{opt,b}^-)$
7	compute $d_b(i,j) = \left\| \hat{\omega}_{ij,b}^+(\hat{\lambda}_{opt,b}^+) - \hat{\omega}_{ij,b}^-(\hat{\lambda}_{opt,b}^-) \right\|$
8	for the edge (i,j), the permutation p-value is defined by $p_{ij} = \frac{\sum_{b=1}^{B} I(d_b(i,j) > d(i,j))}{B}$ with $I(\cdot)$ equal to one if the argument is true, zero otherwise

3 Simulation study and conclusions

In order to evaluate the proposed method, a simulation study has been carried out. We run 50 simulations with $p = 150$ and $n^+ = n^- = 100$. Each row of the matrix $\mathbf{X}^+$ was drawn from $N_p(0; \Omega^{-1})$ where $\omega_{i,i} = 1$ and $\omega_{i,i+1} = 0.5$. Each row of $\mathbf{X}^-$ was drawn from $N_p(0; I)$, where I is the identity matrix. We set $B = 1000$. Under

the considered scenario, Fig. 1 seems to suggest that, under the null-hypothesis, the p-values are not distributed as a uniform distribution.

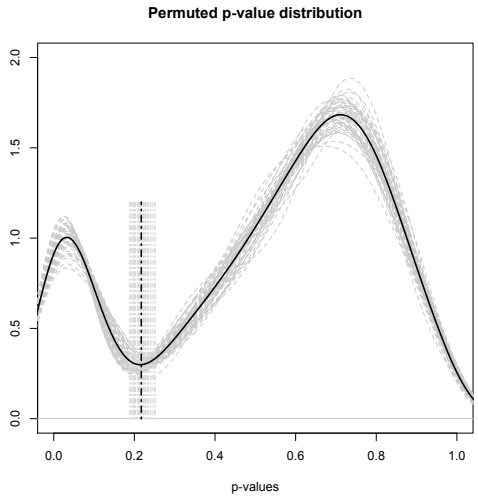

Fig. 1 Gray dot curves are corresponding to the distribution of the permuted p-values. Black curve identify the mean curve. Vertical gray dot lines correspond to the thresholds used to identify which connections are significantly different. Black dot line corresponds to the mean value of the thresholds.

In order to identify which connections are significant differently, Fig. 1 points out a threshold equal to the minimum of each curve indicated by grey dotted vertical lines. In correspondence with the just defined threshold values, Table 2 shows the main summary measures used to evaluate the behaviour of the proposed permutation method. These results seem to emphasize that for the considered scenario, our procedure gives satisfactory results. Aim of our future work will be to define a mixture model to estimate the FDR.

Table 2 Results from the simulation study. We reported the median of the cardinality of the active set $|\mathscr{A}|$, of the false negative (FN) and of the false positive (FP). The mean values of the False Discovery Rate (FDR), of the Sensitivity (Se.) and of the Specificity (Sp.) are also reported

$\lvert\mathscr{A}\rvert$	FN	FP	FDR	Se.	Sp.
309.5	3	13	0.01	0.99	0.99

References

[1] Fan, J. and Lv, J.: A selective overview of variable selection in high dimensional feature space. Statist. Sinica, **20**, 101–148 (2010)
[2] Yuan, M. and Lin, Y.: Model selection and estimation in the Gaussian graphical model. Biometrika. **94**, 19–35 (2007)