Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Multivariate Gaussian network structure learning

Xingqi Du, Subhashis Ghosal *

Department of Statistics, North Carolina State University, 5109 SAS Hall, Campus Box 8203, Raleigh, NC 27695, USA

ARTICLE INFO

Article history:
Received 29 August 2017
Received in revised form 16 May 2018
Accepted 28 July 2018
Available online xxxx

Keywords:
Graphical model
Group penalty
Multivariate normal
Rate of convergence

ABSTRACT

We consider a graphical model where a multivariate normal vector is associated with each node of the underlying graph and estimate the graphical structure. We minimize a loss function obtained by regressing the vector at each node on those at the remaining ones under a group penalty. We show that the proposed estimator can be computed by a fast convex optimization algorithm. We show that as the sample size increases, the estimated regression coefficients and the correct graphical structure are correctly estimated with probability tending to one. By extensive simulations, we show the superiority of the proposed method over comparable procedures. We apply the technique on two real datasets. The first one is to identify gene and protein networks showing up in cancer cell lines, and the second one is to reveal the connections among different industries in the US.

1. Introduction

Finding structural relations in a network of random variables $(X_i : i \in V)$ is a problem of significant interest in modern statistics. The intrinsic dependence between variables in a network is appropriately described by a graphical model, where two nodes $i, j \in V$ are connected by an edge if and only if the two corresponding variables X_i and X_j are conditionally dependent given all other variables. If the joint distribution of all variables is multivariate normal with precision matrix $\Omega = ((\omega_{ij}))$, the conditional independence between the variable located at node i and that located at node j is equivalent of having zero at the (i, j)th entry of Ω. In a relatively large network of variables, generally conditional independence is abundant, meaning that in the corresponding graph edges are sparsely present. Thus in a Gaussian graphical model, the structural relation can be learned from a sparse estimate of Ω, which can be naturally obtained by a regularization method with a lasso-type penalty. Friedman et al. (2008) and Banerjee et al. (2008) proposed the graphical lasso (`glasso`) estimator by minimizing the sum of the negative log-likelihood and the ℓ_1-norm of Ω, and its convergence property was studied by Rothman et al. (2008). A closely related method was proposed by Yuan and Lin (2007). An alternative to the graphical lasso is an approach based on regression of each variable on others, since ω_{ij} is zero if and only if the regression coefficient β_{ij} of X_j in regressing X_i on other variables is zero. Equivalently this can be described as using a pseudo-likelihood obtained by multiplying one-dimensional conditional densities of X_i given $(X_j, j \neq i)$ for all $i \in V$ instead of using the actual likelihood obtained from joint normality of $(X_i, i \in V)$. The approach is better scalable with dimension since the optimization problem is split into several optimization problems in lower dimensions. The approach was pioneered by Meinshausen and Bühlmann (2006), who imposed a lasso-type penalty on each regression problem to obtain sparse estimates of the regression coefficients, and showed that the correct edges are selected with probability tending to one. However, a major drawback of their approach is that the estimator of β_{ij} and that of β_{ji} may not be simultaneously zero (or non-zero), and hence may lead to logical inconsistency while selecting edges based on the estimated values. Peng et al. (2009) proposed the Sparse PArtial

* Corresponding author.
E-mail addresses: xdu8@ncsu.edu (X. Du), sghosal@ncsu.edu (S. Ghosal).

https://doi.org/10.1016/j.jspi.2018.07.009
0378-3758/© 2018 Elsevier B.V. All rights reserved.

Correlation Estimation (space) by taking symmetry of the precision matrix into account. The method is shown to lead to convergence and correct edge selection with high probability, but it may be computationally challenging. A weighted version of space was considered by Khare et al. (2015), who showed that a specific choice of weights guarantees convergence of the iterative algorithm due to the convexity of the objective function in its arguments. Khare et al. (2015) named their estimator the CONvex CORrelation selection methoD (concord), and proved that the estimator inherits the theoretical convergence properties of space. By extensive simulation and numerical illustrations, they showed that concord has good accuracy for reasonable sample sizes and can be computed very efficiently.

However, in many situations, such as if multiple characteristics are measured, the variables X_i at different nodes $i \in V$ may be multivariate. The methods described above apply only in the context when all variables are univariate. Even if the above methods are applied by treating each component of these variables as separate one-dimensional variables, ignoring their group structure may be undesirable, since all component variables refer to the same subject. For example, we may be interested in the connections among different industries in the US, and may like to see if the GDP of one industry has some effect on that of other industries. The data is available for 8 regions, and we want to take regions into consideration, since significant difference in relations may exist because of regional characteristics, which is not possible to capture using only national data. It seems that the only paper which addresses multi-dimensional variables in a graphical model context is Kolar et al. (2014), who pursued a likelihood based approach.

In this article, we propose a method based on a pseudo-likelihood obtained from multivariate regression on other variables. We formulate a multivariate analog of concord, to be called mconcord, because of the computational advantages of concord in univariate situations. Our regression based approach appears to be more scalable than the likelihood based approach of Kolar et al. (2014). Moreover, we provide theoretical justification by studying large sample convergence properties of our proposed method, while such properties have not been established for the procedure introduced by Kolar et al. (2014).

The paper is organized as follows. Section 2 introduces the mconcord method and describes its computational algorithm. Asymptotic properties of mconcord are presented in Section 3. Section 4 illustrates the performance of mconcord, compared with other methods mentioned above. In Section 5, the proposed method is applied to two real datasets on gene/protein profiles and GDP respectively. Proofs are presented in Section 6 and in the Appendix.

2. Method description

2.1. Model and estimation procedure

Consider a graph with p nodes, where at the ith node there is an associated K_i-dimensional random variable $Y_i = (Y_{i1}, \ldots, Y_{iK_i})^T$, $i = 1, \ldots, p$. Let $Y = (Y_1^T, \ldots, Y_p^T)^T$. Assume that Y has multivariate normal distribution with zero mean and covariance matrix $\Sigma = ((\sigma_{ijkl}))$, where $\sigma_{ijkl} = \operatorname{cov}(Y_{ik}, Y_{jl})$, $k = 1, \ldots, K_i$, $l = 1, \ldots, K_j$, $i, j = 1, \ldots, p$. Let the precision matrix Σ^{-1} be denoted by $\Omega = ((\omega_{ijkl}))$, which can also be written as a block-matrix $((\Omega_{ij}))$. The primary interest is in the graph which describes the conditional dependence (or independence) between Y_i and Y_j given the remaining variables. We are typically interested in the situation where p is relatively large and the graph is sparse, that is, most pairs Y_i and Y_j, $i \neq j$, $i, j = 1, \ldots, p$, are conditionally independent given all other variables. When Y_i and Y_j are conditionally independent given other variables, there will be no edge connecting i and j in the underlying graph; otherwise there will be an edge. Under the assumed multivariate normality of Y, it follows that there is an edge between i and j if and only if Ω_{ij} is a non-zero matrix. Therefore the problem of identifying the underlying graphical structure reduces to estimating the matrix Ω under the sparsity constraint that most off-diagonal blocks Ω_{ij} in the grand precision matrix Ω are zero.

Suppose that we observe n independent and identically distributed (i.i.d.) samples from the graphical model, which are collectively denoted by $\mathbf{Y}$, while $\mathbf{Y}_i$ stands for the sample of n many K_i-variate observations at node i and $\mathbf{Y}_{ik}$ stands for the vector of observations of the kth component at node i, $k = 1, \ldots, K_i$, $i = 1, \ldots, p$. Following the estimation strategies used in univariate Gaussian graphical models, we may propose a sparse estimator for Ω by minimizing a loss function obtained from the conditional densities of Y_i given Y_j, $j \neq i$, for each i and a penalty term. However, since sparsity refers to off-diagonal blocks rather than individual elements, the lasso-type penalty used in univariate methods like space or concord should be replaced by a group-lasso type penalty, involving the sum of the Frobenius-norms of each off-diagonal block Ω_{ij}. A multivariate analog of the loss used in a weighted version of space is given by

$$L_n(\omega, \sigma, \mathbf{Y}) = \frac{1}{2} \sum_{i=1}^{p} \sum_{k=1}^{K_i} \Big(-\log \sigma^{ik} + \frac{w_{ik}}{n} \Big\| \mathbf{Y}_{ik} + \sum_{j \neq i} \sum_{l=1}^{K_j} \frac{\omega_{ijkl}}{\sigma^{ik}} \mathbf{Y}_{jl} \Big\|_2^2 \Big), \tag{1}$$

where $\sigma^{ik} = \omega_{iikk}$, $\mathbf{w} = (w_{11}, \ldots, w_{pK_p})$ are nonnegative weights and $\omega_{ijkl} = \omega_{jilk}$ due to the symmetry of precision matrix. Writing the quadratic term in the above expression as

$$w_{ik} \Big\| \mathbf{Y}_{ik} + \sum_{j \neq i} \sum_{l=1}^{K_j} \frac{\omega_{ijkl}}{\sigma^{ik}} \mathbf{Y}_{jl} \Big\|_2^2 = \frac{w_{ik}}{(\sigma^{ik})^2} \Big\| \sigma^{ik} \mathbf{Y}_{ik} + \sum_{j \neq i} \sum_{l=1}^{K_j} \omega_{ijkl} \mathbf{Y}_{jl} \Big\|_2^2,$$

Please cite this article in press as: Du X., Ghosal S., Multivariate Gaussian network structure learning. J. Statist. Plann. Inference (2018), https://doi.org/10.1016/j.jspi.2018.07.009.