Report

Introduction

The precision matrix, $\Theta^\star = (\Sigma^\star)^{-1}$, has a specific mathematical interpretation. The elements in contains information about the partial correlation between variables. That is, the covariance between pairs i and j, conditioned on all other variables. One can think of this as a measure of how correlated two vectors are given the inuence of a set of other variables has been considered. An important result concerns if the vectors X_i and X_j are normally distributed, in which case a partial correlation of 0 implies conditional independence. The precision matrix is closely related to the Gaussian graphical models, which is a framework for representing the structure of the conditional dependencies between Gaussian random variables in a complex system. Graphical models provide an easily understood way of describing and visualizing the relationships between multiple random variables. A Gaussian graphical model is an undirected graph $G = (V, E)$ where the vertices(V) represent the Gaussian variables and the edges(E) represent the dependencies between the variables. Estimating a Gaussian graphical model is equivalent to estimating a precision matrix. The estimation of inverse covariance matrices in high dimensions plays a major role in a wide variety of application domains, such as graphical modeling of brain connectivity based on FMRI brain analysis, gene regulatory network discovery, nancial data processing, social network analysis and climate data analysis. A number of researchers worked on the problem of inverse covariance estimation in high-dimensional settings by using thresholding which leads to estimators whose asymptotic distribution largely depends on the underlying unknown parameter and makes it challenging to establish any results for statistical inference.

Main results

Friedman et al. (2008) discussed the graphical Lasso estimator $\widehat{\Theta}$ that is the solution to the optimization problem

$$\widehat{\Theta} := \arg \min_{\Theta \in S_{++}^P} \left\{ \text{trace}\left(\Theta^T \widehat{\Sigma}\right) - \log \det(\Theta) + \lambda \|\Theta\|_{1,\text{off}} \right\}, \tag{1}$$

where $\widehat{\Sigma}$ is the sample covariance matrix and $\|.\|_{1,\text{off}}$ is the ℓ_1 off-diagonal penalty, $\|\Theta\|_{1,\text{off}} = \sum_{i \neq j} |\Theta_{ij}|$. In this study, Jankova and Van De Geer (2014) proposed a de-sparsied estimator based on the graphical Lasso which builds on inverting Karush-KuhnTucker conditions for the optimization problem (1) that is defined as

$$\widehat{T} := 2\widehat{\Theta} - \widehat{\Theta}\widehat{\Sigma}\widehat{\Theta} \tag{2}$$

In Theorem 1, They demonstrated that each element $\widehat{T}_{ij}$ of de-sparsied estimator have asymptotic normal distribution as :

$$\sqrt{n}\big(\widehat{T}_{ij} - \Theta_{ij}^\star\big)/\sigma_{ij} \xrightarrow{D} N(0,1), \tag{3}$$

where $\xrightarrow{D}$ denotes convergence in distribution. The asymptotic variance σ_{ij} is typically unknown, so to calculate confidence intervals we need to find a consistent estimator $\widehat{\sigma}_{ij}$ for σ_{ij}. In Lemma 2, Jankova and Van De Geer (2014) calculated the estimate theoretical variance that is given by

$$\widehat{\sigma}_{ij} := \widehat{\Theta}_{ii}\widehat{\Theta}_{jj} + \widehat{\Theta}_{ij}^2. \tag{4}$$

Therefore, the $100(1-\alpha)\%$ asymptotic confidence interval for $\Theta_{ij}^\star$ is given by

$$I_{ij} \equiv I_{ij}\left(\widehat{T}_{ij}, \alpha, n\right) := \left[\widehat{T}_{ij} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{\widehat{\sigma}_{ij}}{\sqrt{n}}, \widehat{T}_{ij} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{\widehat{\sigma}_{ij}}{\sqrt{n}}\right]. \tag{5}$$

They also compared the performance of the graphical Lasso-based confidence intervals with three other methods based on a simulation study. It is observed from Table 1 that the de-sparsified graphical Lasso performs better than three other methods in terms of average length.

Simulation Study

In this part, we conduct a simulation study to assess the performance of the de-sparsified graphical Lasso-based confidence intervals suggested by Jankova and Van De Geer (2014). we generated $n = 100$ observations from a multivariate normal distribution with mean 0 and variance covariance matrix given by $\Sigma^0_{ij} = \rho^{|i-j|}$ for some $\rho \neq 0$. We consider combinations of five different values of $\rho, 0.1, 0.3, 0.5, 0.7, 0.9$ and three values of the dimensionality, $p = 50, 100$ or 150. The process of the simulation comprises $N = 1000$ iterations. In each iteration, we calculate de-sparsified graphical Lasso-based confidence interval for $\Theta^\star_{ij}, (i, j) \in \{(2, 5), (2, 7), (2, 9)\}$. The criteria used for the evaluation of the de-sparsified graphical Lasso-based confidence intervals are the average width (AW) and coverage probability (CP) and results are given in Table 1.

Table 1: The AWs and CPs of the de-sparsified graphical Lasso-based confidence intervals.

ρ	p	$\Theta^\star_{25}$		$\Theta^\star_{27}$		$\Theta^\star_{29}$	
		AW	CP	AW	CP	AW	CP
0.1	50	0.3330	0.977	0.3337	0.979	0.3331	0.980
	100	0.3282	0.988	0.3274	0.978	0.3277	0.976
	150	0.3270	0.983	0.3268	0.990	0.3263	0.984
0.5	50	0.3770	0.986	0.3784	0.994	0.3780	0.997
	100	0.3685	0.990	0.3680	0.996	0.3678	0.991
	150	0.3626	0.995	0.3620	0.998	0.3620	0.996
0.9	50	0.6492	0.983	0.6481	1	0.6477	1
	100	0.6090	0.975	0.6084	1	0.6081	1
	150	0.5900	0.979	0.5887	1	0.5885	0.999

From Table 1, we observe that the AWs of de-sparsified graphical Lasso-based confidence intervals are decreasing w.r.t. the dimension of the parameter. Figure 1 shows the histograms of $\sqrt{n}(\widehat{T}_{ij} - \Theta^\star_{ij})/\widehat{\sigma}_{ij}$ for $(i, j) \in \{(2, 5), (2, 7), (2, 9)\}$ with the density of $N(0, 1)$ superimposed.

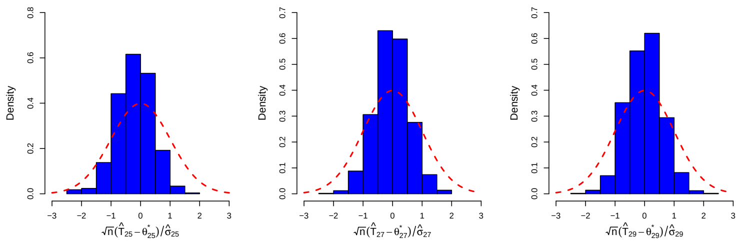

Figure 1: Histograms for $\sqrt{n}(\widehat{T}_{ij} - \Theta^\star_{ij})/\widehat{\sigma}_{ij}, (i, j) \in \{(2, 5), (2, 7), (2, 9)\}$.The sample size was $n = 100$ and the number of parameters $p = 150$. The de-sparsified graphical Lasso estimator was calculated 1000 times. The model was the chain graph with $\rho = 0.5$.

Suggestions for further work

There are also some other topics that are related to the one of this paper.

1- Jankova and Van De Geer (2014) suggest the approach presented in their study for other initial estimators of the precision matrix than the graphical Lasso (such as precision matrix estimation via linear programming and Bootstrapping Lasso estimators).

2- Jankova and Van De Geer (2014) suggest the problem of finding estimator of the precision matrix based on the nodewise regression approach. Asymptotic normality of the estimator may then be obtained under bounded eigenvalues of the true precision matrix, row sparsity of $\Theta^\star$ of small order $\sqrt{n}/\log p$ and assuming fourth-order moment conditions on the X_i's.

3- We could consider the problem of finding high-dimensional Ds and T optimal designs for discriminating between linear and nonlinear models.

References

[1] Jankova, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9(3)** 432 441.

[2] Friedman, J. and Van De Geer, S. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics* **9(1)** 1205-1229.