High-dimensional Gaussian graphical models on network-linked data (Li et al., JMLR 2020)

Hugues Van Assel

05/2021

Plan

Description of the model

Mean estimation error bound

Description of the model

Context

Networks provide **relational information** that most traditional multivariate analysis methods are not designed to use.

This paper develops an analog to the widely used **Gaussian graphical model** for network-linked data.

We consider the problem of estimating a graphical model with **heterogeneous mean vectors** when a network connecting the observations is available[1].

[1] Tianxi Li et al. "High-dimensional Gaussian graphical models on network-linked data.". In: *Journal of Machine Learning Research* 21.74 (2020), pp. 1–45.

Model

$$X_i \sim \mathcal{N}(\mu_i, \Sigma), \quad \mu_i \in \mathbb{R}^p \quad \Sigma \in \mathcal{S}_+^p \quad i = 1, 2, ..., n$$

In what follows : $M = (\mu_1, \mu_2, ..., \mu_n)^T \in \mathbb{R}^{n \times p}$ and $\Theta = \Sigma^{-1} \in \mathcal{S}_+^p$.

Log Likelihood:

$$l(M, \Theta) = \log\det(\Theta) - \frac{1}{n}(\Theta(X - M)^T(X - M))$$

Maximizing the above with respect to M leads to $\widehat{M} = X$.

A network connecting the observations is available.

Regularization : connected nodes should have similar mean vectors.

Network-linked data

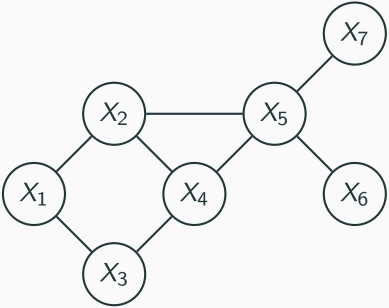
Given a network $\mathcal{G}$, we define:

$$A_{ij} = \begin{cases} 1 & \text{if} \quad i \underset{\mathcal{G}}{\sim} j \\ 0 & \text{otherwise.} \end{cases}$$

$$D = \mathrm{diag}(d_1, d_2, ..., d_n)$$

where $d_i = \sum_j^n A_{ij}$ is the degree of node i.

Example: nearest neighbor graph in spatial transcriptomics.



Graph Laplacian

The graph Laplacian is defined as follows :

$$L = D - A$$

For any vector $\mu \in \mathbb{R}^n$,

$$\begin{aligned}
\mu^T L\mu &= \sum_{i=1}^{n} \mu_i^2 D_{ii} - 2 \sum_{i,j=1}^{n} \mu_i \mu_j A_{ij} \\
&= \sum_{i=1}^{n} \mu_i^2 \sum_{j=1}^{n} A_{ij} - 2 \sum_{i=1}^{n} \sum_{j=1}^{n} \mu_i \mu_j A_{ij} \\
&= \sum_{i<j}^{n} (\mu_i - \mu_j)^2 A_{ij} = \sum_{i \underset{\mathcal{G}}{\sim} j} (\mu_i - \mu_j)^2
\end{aligned}$$

For $M = (\mu_1, \mu_2, ..., \mu_n)^T, \quad \mathrm{tr}(M^T LM) = \sum_{i \underset{\mathcal{G}}{\sim} j} \|\mu_i - \mu_j\|_2^2.$

Regularization on Manifold

The discrete graph approximates a low dimensional manifold $\mathcal{M}$[2].

$$\operatorname{tr}(M^T L M) \xrightarrow{n \to \infty} \int_{\mathcal{M}} \|\nabla \mu(x)\|^2 \, dx$$

This quantity is called the *Dirichlet Energy*.

Applications in semi-supervised learning[3].

---

[2] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. "Graph laplacians and their convergence on random neighborhood graphs.". In: *Journal of Machine Learning Research* 8.6 (2007).

[3] Andreas Argyriou, Mark Herbster, and Massimiliano Pontil. "Combining graph laplacians for semi-supervised learning". In: *NIPS*. vol. 5. Citeseer. 2005, pp. 67–74.

Fitting the model

Regularized Mean estimation problem:

$$\widehat{M}_\alpha = \underset{M}{\operatorname{argmin}} \|X - M\|_F^2 + \alpha \operatorname{tr}(M^T \mathcal{L}_{\mathcal{S}} M)$$

where $\mathcal{L}_{\mathcal{S}} = \frac{D-A}{\frac{1}{n}\sum_i d_i}$

The above problem has the closed form solution:

$$\widehat{M}_{\bullet j} = (I_n + \alpha \mathcal{L}_{\mathcal{S}})^{-1} X_{\bullet j}$$

Covariance estimation in high dimension

Maximizing the log likelihood:

$$\widehat{\Theta}_\alpha = \underset{\Theta \in \mathcal{S}_+^p}{\text{argmin}} \log\det(\Theta) - \text{tr}(\Theta \widehat{S})$$

where $\widehat{S}_\alpha = \frac{1}{n}(X - \widehat{M}_\alpha)^T(X - \widehat{M}_\alpha)$

Difficulty in high dimension:

Let us consider $M = 0$ and $\Sigma = I_p$, the above problem is solved for $\Theta^* = \frac{X^T X}{n}$

$$\mathbb{P}\Bigg(\underbrace{\frac{\gamma_{\max}(X)}{\sqrt{n}}}_{\sqrt{\sigma_{\max}(\frac{X^T X}{n})}} \geq 1 + \sqrt{\frac{p}{n}} + \delta \Bigg) \leq e^{-n\frac{\delta^2}{2}}$$

Covariance estimation in high dimension

Graphical lasso:[4]

$$\widehat{\Theta}_{\alpha,\lambda} = \underset{\Theta \in \mathcal{S}_+^p}{\operatorname{argmin}} \log\det(\Theta) - \operatorname{tr}(\Theta\widehat{S}) - \lambda \|\Theta\|_{1,off}$$

where $\widehat{S}_\alpha = \frac{1}{n}(X - \widehat{M})^T(X - \widehat{M})$

The above estimator is twice regularized.

Gaussian Graphical Model

$$X_i \perp\!\!\!\perp X_j \mid \mathbf{X}/\{X_i, X_j\} \iff \Theta_{ij} = 0$$

Hence we make the assumption of a sparse conditional dependence graph.

[4] Pradeep Ravikumar et al. "High-dimensional covariance estimation by minimizing ℓ1-penalized log-determinant divergence". In: *Electronic Journal of Statistics* 5 (2011), pp. 935–980.

Mean estimation error bound

Definition network-cohesive

$\mathcal{L}_{\mathcal{S}} = U \Lambda U^T$ where $\Lambda = \text{diag}(\tau_1, ..., \tau_n)$, $\tau_1 \geq ... \geq \tau_n = 0$

Given a network A and a vector v, let $v = \sum_{i=1}^n \beta_i u_i$ be the expansion of v in th basis of eigenvectors of L_S.

v is cohesive on A with rate $\delta > 0$ if for all i :

$$(1) \quad \frac{\tau_i^2 |\beta_i|^2}{\|\beta\|_2^2} \leq n^{-\frac{2(1+\delta)}{3} - 1}$$

which implies:

$$(2) \quad \frac{\|\mathcal{L}_{\mathcal{S}} v\|_2^2}{\|v\|_2^2} \leq n^{-\frac{2(1+\delta)}{3}}$$

A matrix M is cohesive on A if all of its columns are cohesive on A.

Assumption

The mean matrix M is cohesive over the network A with rate δ.
Moreover, $\|M_{\cdot j}\|_2^2 \leq b^2 n$ for every j for some positive constant b.

Our goal is to obtain a bound on the difference between M and $\widehat{M}$ under the above assumptions.

Error Bound Estimation of M

The optimization problem on M has a closed form solution:

$$\widehat{M}_{\bullet j} = (I_n + \alpha \mathcal{L}_{\mathcal{S}})^{-1} X_{\bullet j}$$

Let $\widehat{B} = U^T \widehat{M}$:

$$\widehat{B}_{\bullet j} = (I_n + \alpha \Lambda)^{-1} B_{\bullet j} + (I_n + \alpha \Lambda)^{-1} \underbrace{U^T E_{\bullet j}}_{\tilde{E}_{\bullet j}}$$

$\tilde{E}_{\bullet j}$ can be bounded in magnitude by $\mathcal{N}(0, \sigma^2 I)$ where $\sigma^2 = \max_j \Sigma_{jj}$.

Let us now write:

$$\widehat{B}_{ij} - B_{ij} = \underbrace{\frac{\alpha \tau_i}{1 + \alpha \tau_i} B_{ij}}_{Q_i^j} + \underbrace{\frac{1}{1 + \alpha \tau_i} \tilde{E}_{ij}}_{R_i^j}$$

and:

$$\left\| \widehat{B} - B \right\|_2^2 \leq \sum_j \| Q^j \|_2^2 + \sum_j \| R^j \|_2^2$$

Error Bound Estimation of M

$$\sum_j \|Q^j\|_2^2 = \sum_j \sum_i \frac{\alpha^2 \tau_i^2 |B_{ij}|^2}{(1+\alpha\tau_i)^2}$$

$$\leq \sum_j \left(\sum_{i \leq n-m_A} |B_{ij}|^2 + \sum_{i > n-m_A} \frac{\alpha^2 \tau_i^2}{(1+\alpha\tau_i)^2} |B_{ij}|^2 \right)$$

where $m_A = \inf\{m : 0 \leq m \leq n-1, \tau_{n-m} \geq \frac{1}{\sqrt{m}}\}$ is the **effective dimension**.

$$\leq b^2 \sum_j \left(\frac{n-m_A}{\tau_{n-m_A}^2} n^{-\frac{2(1+\delta)}{3}} + \sum_{i > n-m_A} \left(\frac{\alpha}{1+\alpha\tau_{n-1}} \right)^2 n^{-\frac{2(1+\delta)}{3}} \right)$$

$$\leq b^2 p \left((n-m_A) m_A n^{-\frac{2(1+\delta)}{3}} + \frac{m_A}{(1+\Delta)^2} + 1 \right)$$

where $\Delta = n^{\frac{1+\delta}{3}} \tau_{n-1}$ and $\alpha = n^{\frac{1+\delta}{3}}$.

Error Bound Estimation of M

$$
\begin{aligned}
\sum_j \left\| R^j \right\|_2^2 &\leq \sum_j \left(\sum_{i \leq n-m_A} \left(\frac{1}{1+\alpha\tau_i} \right)^2 |\tilde{E}_{ij}|^2 + \sum_{i > n-m_A} \left(\frac{1}{1+\alpha\tau_i} \right)^2 |\tilde{E}_{ij}|^2 \right) \\
&\leq \frac{1}{\tau_{n-m_A}^2} n^{-\frac{2(1+\delta)}{3}} \sum_{i \leq n-m_A} \sum_j |\tilde{E}_{ij}|^2 + \sum_{n-m_A < i < n} \sum_j \frac{|\tilde{E}_{ij}|^2}{(1+\Delta)^2} \\
&+ \sum_j |\tilde{E}_{nj}|^2 \\
&\leq m_A n^{-\frac{2(1+\delta)}{3}} \sum_{i \leq n-m_A} \left\| \tilde{E}_{i.} \right\|_2^2 + \sum_{n-m_A < i < n} \frac{\left\| \tilde{E}_{i.} \right\|_2^2}{(1+\Delta)^2} \\
&+ \left\| \tilde{E}_{n.} \right\|_2^2
\end{aligned}
$$

Concentration of norm multivariate Gaussian

For $x \sim \mathcal{N}(0, \Sigma)$, $\phi_{\mathsf{max}}(\Sigma)$ the largest eigenvalue of Σ, by applying a Lipschitz function of a sub-Gaussian random vector, we get:

$$\mathbb{P}\left(| \|x\|_2^2 - \mathrm{tr}(\Sigma)| > t \right) \leq 2 \exp\left(-c \frac{t}{\phi_{\mathsf{max}}(\Sigma)} \right)$$

and by applying Bernstein's inequality with $t = n\mathrm{tr}(\Sigma)$:

$$\mathbb{P}\left(\sum_i^n \|x_i\|_2^2 > 2n\mathrm{tr}(\Sigma) \right) \leq 2 \exp\left(-cn\mathrm{r}(\Sigma) \right)$$

where $\mathrm{r}(\Sigma) = \frac{\|\Sigma\|_F^2}{\|\Sigma\|_2^2}$ is the stable rank of Σ.

Concentration of norm multivariate Gaussian

$$\mathbb{P}\left(\sum_{i \leq n-m_A} \left\|\tilde{E}_{i.}\right\|_2^2 > 2(n-m_A)p\sigma^2\right) \leq 2\exp(-c(n-m_A)\mathrm{r}(\Sigma))$$

$$\mathbb{P}\left(\sum_{n-m_A < i < n} \left\|\tilde{E}_{i.}\right\|_2^2 > 2m_A p\sigma^2\right) \leq 2\exp(-cm_A\mathrm{r}(\Sigma))$$

$$\mathbb{P}\left(\left\|\tilde{E}_{n.}\right\|_2^2 > 2(n-m_A)p\sigma^2\right) \leq 2\exp\left(-c\frac{p\sigma^2}{\phi_{\max}(\Sigma)}\right)$$

Error bounds on M

$$\begin{aligned}\left\|\hat{M}-M\right\|_F^2 &= \left\|\hat{B}-B\right\|_F^2 \\ &\leq \sum_j \left\|Q^j\right\|_2^2 + \sum_j \left\|R^j\right\|_2^2 \\ &\leq (b^2+2\sigma^2)p\left((n-m_A)m_A n^{-\frac{2(1+\delta)}{3}} + \frac{m_A}{(1+\Delta)^2} + 1\right)\end{aligned}$$

with probability at least:
$1 - 2\exp(-c(n-m_A)\mathrm{r}(\Sigma)) - 2\exp(-cm_A\mathrm{r}(\Sigma)) - 2\exp\left(-c\frac{p\sigma^2}{\phi_{\max}(\Sigma)}\right)$

Therefore the mean estimation error is vanishing with high probability as long as $m_A = o(n^{-\frac{2(1+\delta)}{3}})$ (for a lattice $m_A \leq n^{\frac{2}{3}}$).

Error bounds on Σ

With the previous findings one can prove that, under classical hypothesis for graphical lasso, if m_A is bounded:

$$\left\|\widehat{\Theta} - \Theta\right\|_F^2 \leq\propto 2(1 + \frac{8}{\rho})\sqrt{\frac{\log p}{n}} n^{\frac{\max(1-2\delta)}{6}}$$

In particular, if $\delta \geq \frac{1}{2}$, $\widehat{\Theta}$ is consistent as long as $\log p = o(n)$.

Therefore we have consistent estimations of M and Θ in high dimension under controlled effective dimension of the network and network-cohesion of M.

References

- Argyriou, Andreas, Mark Herbster, and Massimiliano Pontil. “Combining graph laplacians for semi-supervised learning”. In: *NIPS*. Vol. 5. Citeseer. 2005, pp. 67–74.
- Hein, Matthias, Jean-Yves Audibert, and Ulrike von Luxburg. “Graph laplacians and their convergence on random neighborhood graphs.”. In: *Journal of Machine Learning Research* 8.6 (2007).
- Li, Tianxi et al. “High-dimensional Gaussian graphical models on network-linked data.”. In: *Journal of Machine Learning Research* 21.74 (2020), pp. 1–45.
- Ravikumar, Pradeep et al. “High-dimensional covariance estimation by minimizing $\ell 1$-penalized log-determinant divergence”. In: *Electronic Journal of Statistics* 5 (2011), pp. 935–980.