Topic 16: Graphical Network Inference

by Sai Zhang

Key points:

Disclaimer: *The note is built on Prof. Jinchi Lv's lectures of the course at USC, DSO 607, High-Dimensional Statistics and Big Data Problems.*

16.1   Motivation

Consider a classic question: For N observations of dimension p, how can we capture the statistical relationships between the variables of interest? Consider the example of the multivariate Gaussian distribution:

> **Example 16.1.1: Multivariate Gaussian Distribution**
>
> Suppose we have N observations of dimension p, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. let $\mathbf{S}$ be the empirical covariance matrix. Then the probability density
>
> $$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \det(\mathbf{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$
>
> define the **inverse covariance matrix** or **precision matrix** as $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$, then we have
>
> $$f_{\boldsymbol{\mu}, \mathbf{\Omega}} = \exp\left\{\boldsymbol{\mu}'\mathbf{\Omega}\boldsymbol{x} - \left\langle\mathbf{\Omega}, \frac{1}{2}\mathbf{x}\mathbf{x}'\right\rangle - \frac{p}{2}\log(2\pi) + \frac{1}{2}\log\det(\mathbf{\Omega}) - \frac{1}{2}\boldsymbol{\mu}'\mathbf{\Omega}\boldsymbol{\mu}\right\}$$

In this example, we know that **every** multivariate Gaussian distribution can be represented by a pairwise **Gaussian Markov Random Field (GMRF)**, which

- represents the collection of variables $\mathbf{x}$ by a vertex set $\mathcal{V} = \{1, \cdots, p\}$
- encodes

an **undirected graph** $G = (V, E)$ with vertex set $V = 1, \cdots, p$ conrresponding to the random variable edge set $E = \{(i, j) \in V \mid i \neq j, \Omega_{ij} \neq 0\}$