Mod 6: Gaussian Graphical Models

Ruben Zamar
Department of Statistics
UBC

November 15, 2018

Graphical model:

$$\mathcal{G} = (V, E)$$

$$V = \{1, 2, ..., m\} \qquad \text{set of nodes}$$

$$E \subseteq V \times V \qquad \text{set of edges}$$

Adjacent nodes and neighborhoods:

$$i \text{ and } j \text{ are neighbors} \quad \Leftrightarrow \quad (i,j) \in E$$

Undirected graph:

$$(i,j) \quad \in \quad E \Leftrightarrow (j,i) \in E$$

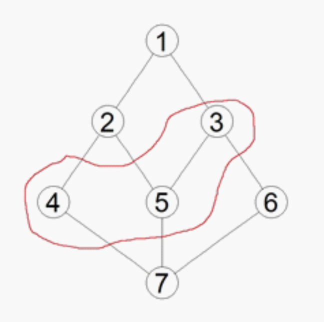
Path from i to j:

$$\text{ordered set } \{i_1, i_2, ..., i_l\} \subseteq V$$

$$\begin{aligned} i_1 &= i \quad , \quad i_l = j \\ (i_k, i_{k+1}) &\in E, \quad k = 1, ..., l-1 \end{aligned}$$

Let A, B, C, be disjoint subsets of V.

C separates A and B if all path from A to B goes through some element of C.



Gaussian graphical model:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} \sim N(\boldsymbol{\mu}, \Sigma)$$

$$(i, j) \in E \Leftrightarrow Corr\,(X_i, X_j \mid \text{all other variables}) \neq 0$$

Conditional Independence (denoted by $\perp$)

$$X_1 \perp X_2 \mid X_3 \Leftrightarrow f\left(x_1, x_2 \mid x_3\right) = f\left(x_1 \mid x_3\right) f\left(x_2 \mid x_3\right)$$

Notation: Let $A = \{i_1, ..., i_l\} \subseteq V$

$$\mathbf{X}_A = \begin{pmatrix} X_{i_1} \\ X_{i_2} \\ \vdots \\ X_{i_l} \end{pmatrix}$$

Suppose that A, B and C are disjoint. Then

$$\mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_C \Leftrightarrow f\left(\mathbf{x}_A, \mathbf{x}_B \mid \mathbf{x}_C\right) = f\left(\mathbf{x}_A \mid \mathbf{x}_C\right) f\left(\mathbf{x}_B \mid \mathbf{x}_C\right)$$

Some general properties of $\perp$

1. $\mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_C \Rightarrow \mathbf{X}_B \perp \mathbf{X}_A \mid \mathbf{X}_C$
2. $\mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_C$ and $\mathbf{U} = h(\mathbf{X}_A) \Rightarrow \mathbf{U} \perp \mathbf{X}_B \mid \mathbf{X}_C$
3. $\mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_C$ and $\mathbf{U} = h(\mathbf{X}_A) \Rightarrow \mathbf{X}_\mathbf{A} \perp \mathbf{X}_B \mid (\mathbf{X}_C, \mathbf{U})$
4. $\mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_C$ and $\mathbf{X}_A \perp \mathbf{W} \mid (\mathbf{X}_B, \mathbf{X}_C) \Rightarrow \mathbf{X}_\mathbf{A} \perp (\mathbf{W}, \mathbf{X}_B) \mid \mathbf{X}_C$

Node neighborhood:

$$\mathcal{A}_i = \{j \neq i : (i,j) \in E\}$$

Markov-Like Properties

Local Markov property: for all i,

$$X_i \quad \perp \quad \mathbf{X}_{V \backslash \overline{\mathcal{A}}_i} \mid \mathbf{X}_{\mathcal{A}_i}, \qquad \text{with} \qquad \overline{\mathcal{A}}_i = \mathcal{A}_i \cup \{i\}$$

Global Markov property: let A, B, C, be disjoint subsets of V

$$\text{if } C \text{ separates } A \text{ and } B \text{ then } \quad \mathbf{X}_A \quad \perp \quad \mathbf{X}_B \mid \mathbf{X}_C$$

Some important notes:

- If the distribution of $\mathbf{X}$ has a positive and continuous density (w.r.l.m.), then these Markov properties are equivalent.
- Moreover, in this case, both Markov properties hold when E is defined by the condition

$$\boxed{\begin{array}{c} \text{“}(j,l) \in E \quad \text{iff} \\[1em] X_j \text{ and } X_l \;\text{ are dependent given} \\[1em] \text{given all the other variables ”} \end{array}}$$

Parametrization of the GGM

1) In terms of the conditional correlations:

$$\boxed{E = \left\{ (j,l) \in V^2 : Corr\left(X_j, X_l | X_{V \setminus \{j,l\}} \right) \neq 0 \right\}}$$

2) In terms of the entries of the precision matrix $(\omega_{pq}) = \Omega = \Sigma^{-1}$

$$\boxed{E = \left\{ (j, l) \in V^2 : \omega_{jl} \neq 0 \right\}}$$

Note: it can be shown that $Corr\left(X_j, X_l | X_{V \setminus \{j,l\}}\right) = 0 \Longleftrightarrow \omega_{jl} = 0$ (see "Reading")

3) In terms of the Pearson correlation of the regression residuals

a) Let $e_j = X_j - E\left(X_j | X_{V\setminus\{j,l\}}\right)$ and $e_l = X_l - E\left(X_l | X_{V\setminus\{j,l\}}\right)$

b) Let $\beta_{jl} = Corr\left(e_j, e_l\right)$

$$\boxed{E = \left\{(j,l) \in V^2 : \beta_{jl} \neq 0\right\}}$$

Note: it can be shown that $Corr\left(X_j, X_l | X_{V\setminus\{j,l\}}\right) = \beta_{jl}$ (see "Reading")

GLASSO proposed by Friedman, Hastie and Tibshirani (2008):

$$\widehat{\Omega} = \arg\min_{\Omega \succ 0} \left\{ tr\left(S\Omega\right) - \log\det\left(\Omega\right) + \lambda \sum_{i \neq j} |\omega_{ij}| \right\}$$

where

- S is the sample covariance matrix, and
- $\Omega = \Sigma^{-1}$ is the precision matrix
  See the package "glasso" in CRAN

Stepwise GGM

Set the entry threshold $\overline{\beta}$ (a number between 0 and 1)

- Given a system of neighborhoods $\left\{\mathcal{A}_j^{(k)}\right\}$ compute the regression residuals $\mathbf{e}_j$ by regressing X_j on $\mathbf{X}_{\mathcal{A}_j^{(k)}}$
- For each $l \notin \mathcal{A}_j^{(k)}$ (and therefore $j \notin \mathcal{A}_l^{(k)}$) compute the Pearson correlation

$$\widehat{\beta}_{jl} = \frac{\mathbf{e}_j'\mathbf{e}_l}{\sqrt{\mathbf{e}_j'\mathbf{e}_j}\sqrt{\mathbf{e}_l'\mathbf{e}_l}}$$

- Let

$$\left|\widehat{\beta}_{j_0 l_0}\right| = \max_{l \notin \mathcal{A}_j^{(k)}, j \in V} \left|\widehat{\beta}_{jl}\right|$$

- If $\left|\widehat{\beta}_{j_0 l_0}\right| \geq \overline{\beta}$, set $\mathcal{A}_{j_0}^{(k+1)} = \mathcal{A}_{j_0}^{(k)} \cup \{l_0\}$, $\mathcal{A}_{l_0}^{(k+1)} = \mathcal{A}_{l_0}^{(k)} \cup \{j_0\}$, $\mathcal{A}_j^{(k+1)} = \mathcal{A}_j^{(k)}$ for $j \neq j_0, l_0$ and continue.

- If $\left|\widehat{\beta}_{j_0 l_0}\right| < \overline{\beta}$, stop