3.2 Conditional independence graphs

Def^n X, Y, Z with joint density p_{XYZ}. Say X is conditionally independent of Y given Z and write $X \perp\!\!\!\perp Y \mid Z$ if

$$f_{XY|Z}(x, y \mid z) = f_{X|Z}(x \mid z)\, f_{Y|Z}(y \mid z) \qquad \forall x, y, z \text{ s.t. } f_Z(z) > 0 \}$$

or equivalently if

$$f_{X|Y,Z}(x \mid y, z) = f_{X,Z}(x \mid z).$$

Here graph $\mathcal{G}$ will be an undirected graph.

Reminder of notation $\begin{matrix}-k \\ -jk\end{matrix}$ in subscripts means $\begin{matrix}\{1, \dots, p\} \setminus \{k\} \\ \{1, \dots, p\} \setminus \{j, k\}\end{matrix}$

Let P be the distribution of $Z = (Z_1, \dots, Z_p)^T$.

Def^n Say P satisfies the pairwise Markov property w.r.t $\mathcal{G}$ if for any $j, k \in V$ with $j \neq k$ and $(j, k), (k, j) \notin E$, we have

$$Z_k \perp\!\!\!\perp Z_j \mid Z_{-jk}$$

The minimal graph satisfying the pairwise Markov property is called the conditional independence graph (CIG) for P.

Def^n P satisfies the global Markov property w.r.t $\mathcal{G}$ if for any triple (A, B, S) of disjoint subsets of V s.t. S separates A from B, then

$$Z_A \perp\!\!\!\perp Z_B \mid Z_S.$$

Prop 29 If P has a positive density then if it satisfies the pairwise Markov property w.r.t. $\mathcal{G}$ then it also satisfies the global Markov property w.r.t $\mathcal{G}$.

3.3.1 Normal conditionals

Let $Z \sim N_p(\mu, \Sigma)$ with Σ p.d. (NB $\Sigma_{A,A}$ is p.d. $\forall A$)

Notation: if $M \in \mathbb{R}^{p \times p}$, $A, B \subseteq \{1, \dots, p\}$ then $M_{A,B}$ will be the $|A| \times |B|$ submatrix of M taking those rows and cols of M indexed by A and B respectively.

Prop 30 Let A and B be disjoint

$$Z_A \mid Z_B = z_B \sim N_{|A|}\Big(\mu_A + \Sigma_{A,B}\Sigma_{B,B}^{-1}(z_B - \mu_B),$$
$$\Sigma_{A,A} - \Sigma_{A,B}\Sigma_{B,B}^{-1}\Sigma_{B,A}\Big)$$

Proof: Idea: Write $Z_A = MZ_B + (Z_A - MZ_B)$ where $M \in \mathbb{R}^{|A|\times|B|}$ is s.t.

$Z_A - MZ_B \perp\!\!\!\perp Z_B$ i.e. s.t.

$$\mathrm{Cov}(Z_A - MZ_B, Z_B) = \Sigma_{A,B} - M\Sigma_{B,B} = 0$$

So take $M = \Sigma_{A,B}\Sigma_{B,B}^{-1}$. Then $Z_A - MZ_B \mid Z_B \overset{d}{=} Z_A - MZ_B$.

Now $\mathbb{E}(Z_A - MZ_B) = \mu_A - \Sigma_{A,B}\Sigma_{B,B}^{-1}\mu_B$

$$\mathrm{Var}(Z_A - MZ_B) = \Sigma_{A,A} + \Sigma_{A,B}\Sigma_{B,B}^{-1}\Sigma_{B,B}\Sigma_{B,B}^{-1}\Sigma_{B,A} - 2\Sigma_{A,B}\Sigma_{B,B}^{-1}\Sigma_{B,A}$$
$$= \Sigma_{A,A} - \Sigma_{A,B}\Sigma_{B,B}^{-1}\Sigma_{B,A}$$

□

3.3.2 Neighbourhood regression

Specialising $A = \{k\}$, $B = A^c$, conditioning on Z_{-k}

$$Z_k = m_k + Z_{-k}^T \Sigma_{-k,-k}^{-1}\Sigma_{-k,k} + \varepsilon_k$$

where $m_k = \mu_k - \mu_{-k}^T \Sigma_{-k,-k}^{-1}\Sigma_{-k,k}$

$$\varepsilon_k \mid Z_{-k} \sim N(0, \Sigma_{k,k} - \Sigma_{k,-k}\Sigma_{-k,-k}^{-1}\Sigma_{-k,k})$$

If j^{th} component of $\Sigma_{-k,-k}^{-1}\Sigma_{-k,k}$ is 0, then the distribution of $Z_k \mid Z_{-k}$ will not depend on j^{th} component of Z_{-k}, say $Z_{j'}$.

thus $Z_k \mid Z_{-k} \overset{d}{=} Z_k \mid Z_{-kj'}$

$$\iff Z_k \perp\!\!\!\perp Z_{j'} \mid Z_{-j'k}$$

Given $x_1, \dots, x_n \overset{iid}{\sim} Z$ and $X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}$, may estimate $\Sigma_{-k,-k}^{-1}\Sigma_{-k,k}$ by regressing X_k on X_{-k} with an intercept included, using the Lasso.

Neighbourhood selection (Meinshausen, ... 2006) does this for each variable k to obtain selected sets $\hat{S}_k$.

Two ways of estimating CIG.

OR rule : add an edge $(k,j), (j,k)$ if $k \in \hat{S}_j$ or $j \in \hat{S}_k$

AND rule : ... and ...

3.3.3 The precision matrix and conditional independence

Prop 31 Let $M \in \mathbb{R}^{p \times p}$ symmetric, p.d. and write

$$M = \begin{pmatrix} P & Q^T \\ Q & R \end{pmatrix} \quad \text{where } P \text{ and } R \text{ are square.}$$

The Schur complement of R is $P - Q^T R^{-1} Q = S$

- S is p.d.
- $M^{-1} = \begin{pmatrix} S^{-1} & -S^{-1}Q^T R^{-1} \\ -R^{-1}QS^{-1} & R^{-1} + R^{-1}QS^{-1}Q^T R^{-1} \end{pmatrix}$
- $\det M = \det S \det R$