Let $\Omega = \Sigma^{-1}$ be the precision matrix. Observe that

$$\underbrace{\Sigma_{k,k}}_{P} - \underbrace{\Sigma_{k,-k}}_{Q^T} \underbrace{\Sigma_{-k,-k}^{-1}}_{R^{-1}} \underbrace{\Sigma_{-k,k}}_{Q} = \Omega_{k,k}^{-1}$$

(More generally, $\mathrm{Var}(Z_A \mid Z_{A^c}) = \Omega_{A,A}^{-1}$)

$$\Sigma_{-k,-k}^{-1} \Sigma_{-k,k} = -\Omega_{-k,k} \Omega_{k,k}^{-1}$$

Thus $(\Sigma_{-k,-k}^{-1} \Sigma_{-k,k})_j = 0 \iff \begin{cases} \Omega_{jk} & \text{if } j<k \\ \Omega_{j+1,k} & \text{if } j \geq k \end{cases}$

So $Z_k \perp\!\!\!\perp Z_j \mid Z_{-jk} \iff \Omega_{jk} = 0$

3.3.4 The Graphical Lasso

Recall the density of $N_p(\mu, \Sigma)$ is

$$f(z) = \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(z-\mu)^T \Sigma^{-1}(z-\mu)\right)$$

Thus the log-likelihood based on iid sample $x_1, \ldots, x_n \sim N_p(\mu, \Sigma)$ is

$$\ell(\mu, \Omega) = \frac{n}{2} \log\det\Omega - \frac{1}{2}\sum_{i=1}^n (x_i-\mu)^T \Omega (x_i - \mu)$$

Let $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, $S = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$

$$\sum_{i=1}^n (x_i-\mu)^T \Omega (x_i-\mu) = \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^T \Omega (x_i - \bar{x} + \bar{x} - \mu)$$

$$= \sum_{i=1}^n (x_i - \bar{x})^T \Omega (x_i - \bar{x}) + n(\bar{x}-\mu)^T \Omega (\bar{x} - \mu)$$

Also $\sum_{i=1}^n (x_i - \bar{x})^T \Omega (x_i - \bar{x}) = \sum_{i=1}^n \mathrm{tr}\left((x_i - \bar{x})(x_i - \bar{x})^T \Omega\right)$

$$= n\,\mathrm{tr}(S\Omega)$$

Thus, $\ell(\mu, \Omega) = -\frac{n}{2}\left\{\mathrm{tr}(S\Omega) - \log\det\Omega + (\bar{x}-\mu)^T \Omega (\bar{x}-\mu)\right\}$

$$\max_{\mu \in \mathbb{R}^p} \ell(\mu, \Omega) = -\frac{n}{2}\left\{\mathrm{tr}(S\Omega) - \log\det\Omega\right\}$$

MLE of Ω, $\hat{\Omega}^{ML}$ can be found by minimising

$$\min_{\Omega : \underset{\text{p.d.}}{\Omega \succ 0}} \left\{\mathrm{tr}(S\Omega) - \log\det\Omega\right\}$$

Let $Q : \Omega \to \begin{cases} \operatorname{tr}(S\Omega) - \log \det \Omega & : \Omega \text{ is p.d} \\ \infty & : \text{o/w} \end{cases}$

It can be shown that Q is convex.

If $\Omega \in \cancel{\text{int}}(\operatorname{dom} Q)$ i.e. Ω is p.d then

$$\frac{\partial}{\partial \Omega_{jk}} \log \det(\Omega) = (\Omega^{-1})_{jk}$$

$$\frac{\partial}{\partial \Omega_{jk}} \operatorname{tr}(S\Omega) = S_{jk}$$

If S is p.d. then $\hat{\Omega}^{ML} = S^{-1}$.

The graphical Lasso solves

$$\min_{\Omega : \Omega \succ 0} \{ \operatorname{tr}(S\Omega) - \log \det \Omega + \lambda \|\Omega\|_1 \} \qquad \text{where } \|\Omega\|_1 = \textstyle\sum_{jk} |\Omega_{jk}|$$

This gives estimate $\hat{\Omega}_\lambda$ from which we can form an estimated CIG.

3.4 Structural Equation Models

Now by a graph we mean any graph.

<u>Def</u>

Defⁿ A structural equation model (SEM) $\mathcal{S}$ for random vector $Z \in \mathbb{R}^p$ is a collection of p equations

$$Z_k = h_k(Z_{P_k}, \varepsilon_k) \qquad k = 1, \dots, p$$

where

- $\varepsilon_1, \dots, \varepsilon_p$ are all independent random variables
- $P_k \subseteq \{1, \dots, p\} \setminus \{k\}$ are such that the graph with $pa(k) = P_k$ is a DAG.

Note that an SEM does determine the distribution of Z (with distributions of the ε_k given). Indeed, take a topological ordering π for the DAG; we can write Z_k as a function of $\varepsilon_{\pi^{-1}(1)}, \varepsilon_{\pi^{-1}(2)}, \dots, \varepsilon_{\pi^{-1}(\pi(k))}$.

<u>Example</u>

Taking this course $Z_1 = 1$
Catch-up lectures $Z_2 = 1$
Heard about ML $Z_3 = 1$

$Z_3 = \varepsilon_3 \sim \text{Bern}(\tfrac{1}{4})$

$Z_2 = \mathbb{1}_{\{\varepsilon_2(1+Z_3) > \frac{1}{2}\}}, \ \varepsilon_2 \sim U[0,1] \qquad Z_1 = \mathbb{1}_{\{\varepsilon_1(Z_2+Z_3) > \frac{1}{2}\}}, \ \varepsilon_3 \sim U[0,1]$

3.5 Interventions

We can modify an SEM by setting e.g. $Z_k = a$; this is called a <u>perfect intervention</u>.
The new SEM gives us a new distribution for Z.