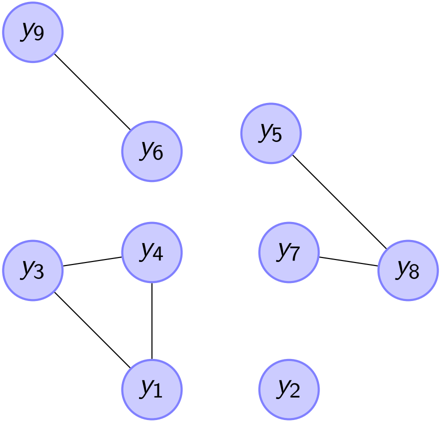
Estimating bayesian mutliple graphical models with EM

Luca Bracone

December 8, 2022

Undirected bayesian graph



There is an edge (i, j) if and only if $y_i \not\perp\!\!\!\perp y_j | y_{-i,-j}$

Setting and assumptions

We observe p parameters across n samples. That is we have n observations $y^1, \ldots, y^n \in \mathbb{R}^p$ of a p dimensional vector. We assume that $y^1, \ldots, y^n \overset{i.i.d.}{\sim} \mathcal{N}_p(0, \Sigma)$ have a multivariate normal distribution.

- To estimate the graph we would like to know whether $y_i \not\perp\!\!\!\perp y_j | y_{-i,-j}$.
- Idea: We know that in the *precision matrix*, $\Omega = \Sigma^{-1}$ the (i, j)-th entry is zero only if $y_i \perp\!\!\!\perp y_j | y_{-i,-j}$
- How do we find the zero entries of Ω ?

Why should we use the Bayesian method ?

- Flexible hierarchical modeling
- Ease of interpretation of results
- Use prior knowledge

If we consider Ω as being drawn from a prior distribution $p(\Omega)$ we can obtain a *posterior distrbution* $p(\Omega|X)$ of which the maximum is the *maximum à-posteriori* estimate Ω.

Spike and slab prior

The spike and slab prior for Ω helps us differentiate between zero and non-zero entries of Ω

$$
\begin{aligned}
y|\Omega &\sim N_p(0, \Omega^{-1}), \\
\omega_{ij}|\delta_{ij} &\sim \delta_{ij} N(0, v_1^2) + (1 - \delta_{ij}) N(0, v_0^2), i \neq j, \\
\omega_{ii} &\sim Exp(\lambda/2), \\
\delta_{ij}|\pi &\sim Bern(\pi), \\
\pi &\sim Beta(a, b).
\end{aligned}
$$

Posterior joint distribution

After a few manipulations we find that

$$p(\Omega, \delta, \pi | y) \propto p(y|\Omega)p(\Omega|\delta)p(\delta|\pi)p(\pi)$$

$$\log(p(\Omega, \delta, \pi | Y)) =$$

$$\sum_{j<k} - \log(v_0^2(1-\delta_{jk}) + v_1^2\delta_{jk}) - \frac{\omega_{jk}^2}{2}\frac{1}{v_0^2(1-\delta_{jk}) + v_1^2\delta_{jk}} - \sum_j \frac{\lambda}{2}\omega_{jj}$$

$$+ \sum_{j<k} \delta_{jk} \log\left(\frac{\pi}{1-\pi}\right) + \log(1-\pi)$$

$$+ (a-1)\log(\pi) + (b-1)\log(1-\pi)$$

$$+ \frac{n}{2}\log\det(\Omega) - \frac{1}{2}\mathrm{tr}(Y^tY\Omega) + \text{constants.}$$

Taking expectations

$$
\begin{aligned}
&Q(\Omega, \pi | \Omega^{(l)}, \pi^{(l)}) = E_{\delta | Y, \Omega^{(l)}, \pi^{(l)}}(\log(p(\Omega, \delta, \pi | Y) | Y, \Omega^{(l)}, \pi^{(l)}) = \\
&- \sum_{j<k} \frac{\omega_{jk}^2}{2} E_{\delta_{jk}|\cdot} \left(\frac{1}{v_0^2(1-\delta_{jk}) + v_1^2 \delta_{jk}} \right) - \sum_{j} \frac{\lambda}{2} \omega_{jj} \\
&+ \sum_{j<k} E_{\delta_{jk}|\cdot}(\delta_{jk}) \log \left(\frac{\pi}{1-\pi} \right) + \log(1-\pi) \\
&+ (a-1)\log(\pi) + (b-1)\log(1-\pi) \\
&+ \frac{n}{2} \log\det(\Omega) - \frac{1}{2} \operatorname{tr}(Y^t Y \Omega) + \text{constants.}
\end{aligned}
$$

Computing expectation terms

$$q_{jk} := E_{\delta_{jk}|\cdot}(\delta_{jk}) = \frac{\pi p(\omega_{jk}|\delta = 1)}{\pi p(\omega_{jk}|\delta = 1) + (1-\pi)p(\omega_{jk}|\delta = 0)}.$$

And

$$d_{jk} := E_{\delta_{jk}|\cdot}\left(\frac{1}{v_0^2(1-\delta_{jk}) + v_1^2\delta_{jk}}\right) = \frac{1-q_{jk}}{v_0^2} + \frac{q_{jk}}{v_1^2}$$

Maximising π

Taking the derivative and setting to zero we find that

$$\pi^{(l+1)} = \frac{a - 1 + \sum_{j<k} q_{jk}}{a + b - 2 + \frac{p(p-1)}{2}}$$

Maximising Ω

If we partition

$$\Omega = \begin{pmatrix} \Omega_{11} & \omega_{12} \\ \omega_{12}^t & \omega_{22} \end{pmatrix} \quad X^t X = \begin{pmatrix} S_{11} & s_{12} \\ s_{12}^t & s_{22} \end{pmatrix}$$

we find that

$$\omega_{12} \sim N(-C^{-1} s_{12}, C) \quad C = (s_{22} + \lambda)\Omega_{11}^{-1} + \mathrm{diag}(v_{12}^{-1})$$

and that

$$\omega_{22} \sim Gamma\left(\frac{n}{2} + 1, \frac{s_{22} + \lambda}{2}\right) + \omega_{12}^t \Omega_{11}^{-1} \omega_{12}.$$

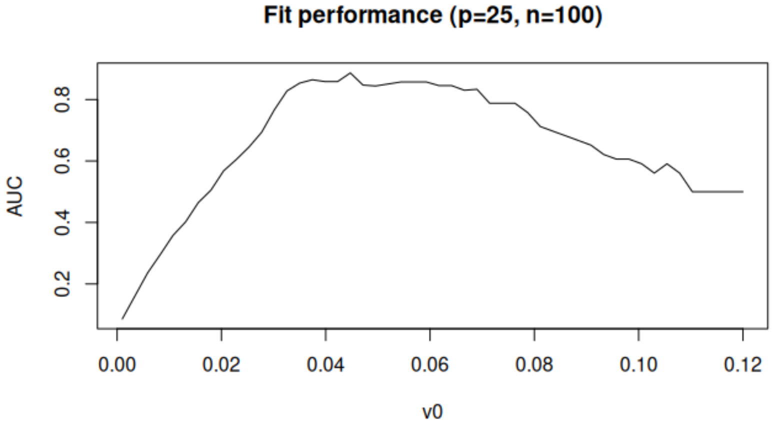
Taking the modes we find the update steps

$$\omega_{12}^{(l+1)} = -((s_{22} + \lambda)\Omega_{11}^{-1} + \mathrm{diag}(d_{12}))^{-1} s_{12}$$

$$\omega_{22}^{(l+1)} = \frac{n}{s_{22} + \lambda} + (\omega_{12}^{(l+1)})^t \Omega_{11}^{-1} \omega_{12}^{(l+1)}$$

Choosing hyperparameters

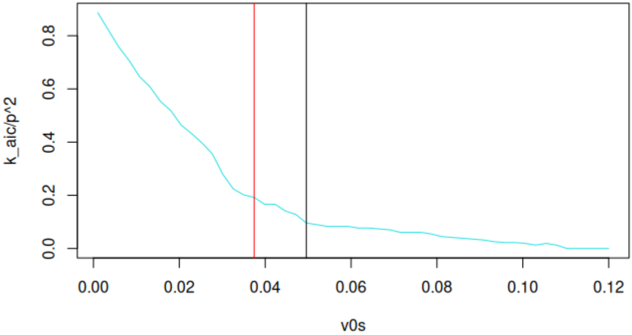
Big influence on performance



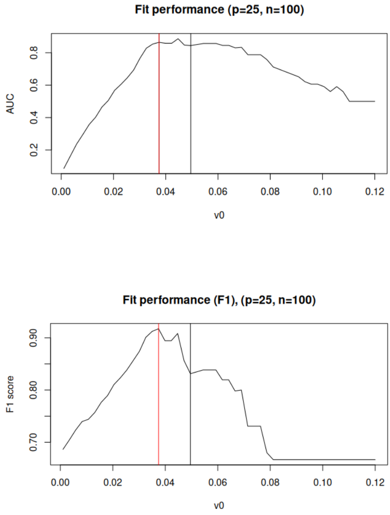
Try: fixing sparsity

Say I know I want about $s = 10\%$ sparsity.

- Then fix a, b so that $a/(a+b) = s$
- Try many v_0, pick fit with closest sparsity.



Effect on performance

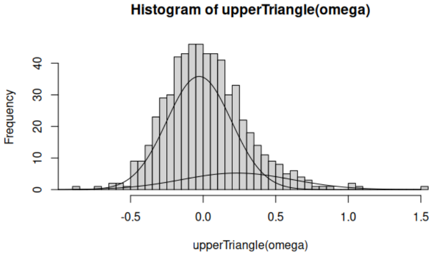


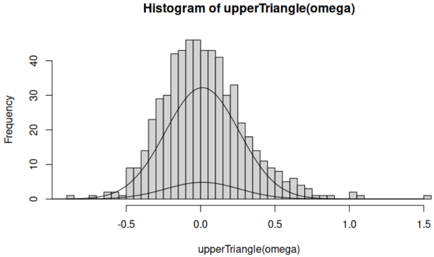
Method	EMGS	MB
p=25	0.86	0.45
p=35	0.78	0.54
p=50	0.72	0.52

Table: AUC values for n = 100, p=25, 35, 50

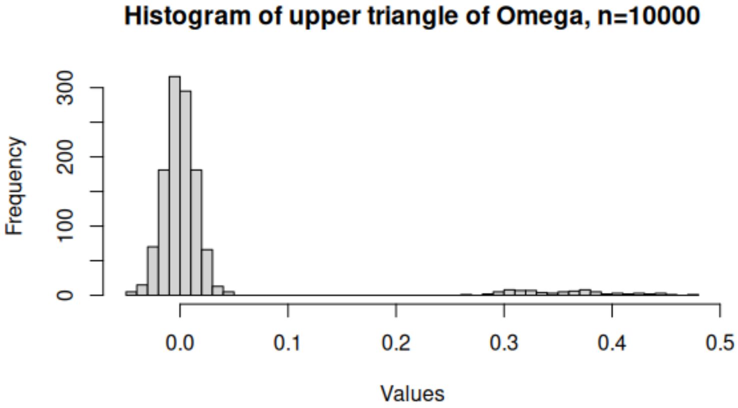
How to choose sparsity?

Fitting gaussian mixture model, not very convincing.





It will depend on generation method



Future work

The original aim of the project: multiple graphs. We now have a hierarchical model

$$
\begin{aligned}
y|\Omega_k &\sim N_p(0, \Omega_k^{-1}), \\
\omega_{ijk}|\delta_{ijk} &\sim \delta_{ijk} N(0, v_1^2) + (1-\delta_{ijk}) N(0, v_0^2) \text{for } i \neq j, \\
\omega_{iik} &\sim Exp(\lambda_k/2), \\
\delta_{ijk}|\theta_{ijk} &\sim Bern(\Phi(\theta_{ijk})), \\
\theta_{ij} &\sim N_K(0, \Sigma).
\end{aligned}
$$

The parameter Σ is a shared parameter across graphs.