SBM Laplace prior for concentration matrix

April 3, 2019

1 Model

Data: X, $i \in \{0, ..., n\}$ observations. $X \sim \mathcal{N}(0, \Sigma^{-1} = \Omega)$, Ω concentration matrix. Assumption: Q hidden clusters in the data. For each observation i, $Z_{iq} = \mathbb{1}_{i \in q}$, $q \in \{1, ..., Q\}$. Z_{iq} = independant latent variables.

- $\alpha_q = \mathbb{P}(i \in q)$ (Probability that observation i belongs to cluster q);
- $\sum_q \alpha_q = 1$;
- $\underline{Z_i} \sim \mathcal{M}(1, \underline{\alpha})$ with $\underline{\alpha} = (\alpha_1, ..., \alpha_Q)$ and $\underline{Z_i} = (Z_{i1}, ..., Z_{iQ})$

Assumption:

$$\Omega_{ij} | \{Z_{iq} Z_{jl}\} \sim Laplace(0, \lambda_{ql}), \quad (i,j) \in \{1, ..., n\}^2, (q,l) \in \{1, ..., Q\}^2$$

Laplace distribution:

$$\forall x \in \mathbb{R}, \quad f_{ql}(x) = \frac{1}{2\lambda_{ql}} \exp - \frac{|x|}{\lambda_{ql}} \;\; \text{if} \;\; q \neq l \;\; \text{and} \;\; f_0 x = \frac{1}{2\lambda_0} \exp - \frac{|x|}{\lambda_0} \;\; \text{otherwise}$$

where $\lambda_{ql}, \lambda_0 > 0$ are scaling parameters and $\lambda_{ql} = \lambda_{lq}$. Λ matrix of parameters λ_{ql}, $(q,l) \in \{1, ..., Q\}^2$.

2 Complete likelihood

Assumptions reminder:

$$X | \Omega \sim \mathcal{N}(O, \Omega^{-1}), \quad \Omega | Z \sim Laplace(0, \Lambda), \quad Z \sim \mathcal{M}(1, \underline{\alpha})$$

Complete likelihood decomposition:

$$\begin{aligned} L_c(X, \Omega, Z) &= \mathbb{P}(X, \Omega, Z) \\ &= \mathbb{P}(X | \{\Omega, Z\}) \mathbb{P}(\Omega | Z) \mathbb{P}(Z) \;\; \text{conditional probabilities definition/formula} \\ &= \mathbb{P}(X | \Omega) \mathbb{P}(\Omega | Z) \mathbb{P}(Z) \end{aligned}$$

$\mathbb{P}(X | \{\Omega, Z\}) = \mathbb{P}(X | \Omega)$, known distribution (knowing Z is equivalent to know Ω?).
Complete log-likelihood formula: (when $i \neq j$, otherwise replace λ_{ql} with λ_0)

$$\begin{aligned} \log L_c(X, \Omega, Z) &= \log \mathbb{P}(X, \Omega, Z) \\ &= \underbrace{\log \mathbb{P}(X|\Omega)}_{(1)} + \underbrace{\log \mathbb{P}(\Omega|Z)}_{(2)} + \underbrace{\log \mathbb{P}(Z)}_{(3)} \\ &= \frac{n}{2} \left(\log(|\Omega|) - tr(S\Omega) - p \log(2\pi) \right) + \sum_{q,l,i,j,i \neq j} Z_{iq} Z_{jl} \left(- \log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}} \right) \\ &+ \sum_{i=1}^{n} \sum_{q=1}^{Q} Z_{iq} \log \alpha_q \end{aligned}$$

Penalty term (lasso approach) in the article: $\sum_{q,l,i,j,i \neq j} Z_{iq} Z_{jl} \frac{|\Omega_{ij}|}{\lambda_{ql}}$. In the paper: notation $\|\rho_Z(\Omega)\|_{l_1}$.

(1): $X|\Omega \sim \mathcal{N}(0, \Omega^{-1})$
Notation: $|\Omega| = det(\Omega) = 1/|\Omega^{-1}|$

$$\begin{aligned}
\log \mathbb{P}(X|\Omega) &= \sum_{i=1}^{n} \log\left(\frac{1}{(2\pi)^{\frac{p}{2}}|\Omega^{-1}|}\exp(\frac{1}{2}x_i^T\Omega x_i)\right) \\
&= -\frac{n}{2}\left(p\log(2\pi) + \log(|\Omega^{-1}|\right) - \frac{1}{2}tr(X^TX\Omega) \\
&= -\frac{n}{2}\left(p\log(2\pi) - \log(|\Omega|)\right) - \frac{n}{2}tr(S\Omega) \text{ En notant } S = \frac{1}{n}X^TX \\
&= \frac{n}{2}\left(-p\log(2\pi) + \log(|\Omega|)\right) - \frac{n}{2}tr(S\Omega) \\
&= \frac{n}{2}\left(\log(|\Omega|) - tr(S\Omega) - p\log(2\pi)\right)
\end{aligned}$$

(2): $\Omega|Z \sim Laplace(0, \Lambda)$
Sophie : ici il faut changer tous les n en p = taille de Ω. If $i \neq j$:

$$\begin{aligned}
\log \mathbb{P}(\Omega|Z) &= \log \prod_{i,j,q,l} \mathbb{P}(\Omega_{ij}|\{Z_{iq}Z_{jl}\})^{Z_{iq}Z_{jl}} \text{ Astuce termes puissance 0 ne participent pas au produit} \\
&= \sum_{\substack{q,l,i,j \\ i\neq j}} Z_{iq}Z_{jl}\left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}}\right) \\
&= -\underbrace{\sum_{\substack{q,l,i,j \\ i\neq j}} Z_{iq}Z_{jl}\log(2\lambda_{ql})}_{\sum_{q,l} n_{ql}\log(2\lambda_{ql})} - \underbrace{\sum_{\substack{q,l,i,j \\ i\neq j}} Z_{iq}Z_{jl}\frac{|\Omega_{ij}|}{\lambda_{ql}}}_{\|\rho_Z(\Omega)\|_1}
\end{aligned}$$

Otherwise:

$$\log \mathbb{P}(\Omega|Z) = \sum_i \left(-\log(2\lambda_0) - \frac{|\Omega_{ii}|}{\lambda_0}\right)$$

(3): $Z \sim \mathcal{M}(1, \underline{\alpha})$
Sophie : ici il faut changer tous les n en p = taille de Ω.

$$\begin{aligned}
\log \mathbb{P}(Z) &= \log \prod_{i=1}^{n} \mathbb{P}(Z_i) = \log \prod_{i=1}^{n}\prod_{q=1}^{Q} \alpha_q^{Z_{iq}} \\
&= \sum_{i=1}^{n}\sum_{q=1}^{Q} \log \alpha_q^{Z_{iq}} = \sum_{i=1}^{n}\sum_{q=1}^{Q} Z_{iq}\log\alpha_q
\end{aligned}$$

3 EM algorithm and variational estimation

Steps

1. Expectation (E): Calculate the expected value of the likelihood under the current parameters. Estimate Z_i, knowing Ω from previous step.
2. Maximization (M): Find parameters that maximizes the likelihood. Compute parameters knowing Z_i.

3.1 Definitions

Let $\mathcal{Z}$ the space of all possibilities.
Definitions give us:

$$\log \mathbb{P}(Z|X, \Omega) = \log \mathbb{P}(X, \Omega, Z) - \log \mathbb{P}(X, \Omega)$$

3.2 E step: impossibility of direct compuation and variational approach

In this part, Ω is known (as $\Omega^{(t)}$) from a previous M step.

Considering $\mathbb{P}(X, \Omega)$ being constant with respect to Z:

$$\begin{aligned}
\log \mathbb{P}(X, \Omega) &= \log \mathbb{P}(X, \Omega, Z) - \log \mathbb{P}(Z|X, \Omega) \\
&= \mathbb{E}_{Z|\Omega^{(t)}}[\log \mathbb{P}(X, \Omega, Z)] - \mathbb{E}_{Z|\Omega^{(t)}}[\log \mathbb{P}(Z|X, \Omega)] \\
&= \sum_{z \in \mathcal{Z}} \mathbb{P}(Z = z|\Omega^{(t)}) \log \mathbb{P}(X, \Omega, Z) - \sum_{z \in \mathcal{Z}} \mathbb{P}(Z = z|\Omega^{(t)}) \log \mathbb{P}(Z|X, \Omega) \\
&= \mathbb{E}_{Z|\Omega^{(t)}}[\log \mathbb{P}(X, \Omega, Z)] - \mathcal{H}(Z|X)
\end{aligned}$$

The expected complete likelihood under the current parameters is:

$$\mathcal{Q}(\Omega|\Omega^{(t)}) = \mathbb{E}_{Z|\Omega^{(t)}}[\log \mathbb{P}(X, \Omega, Z)] = \sum_{z \in \mathcal{Z}} \mathbb{P}(Z = z|\Omega^{(t)}) \log \mathbb{P}(X, \Omega, Z)$$

$\mathbb{P}(Z_{iq}Z_{jl} = 1|\Omega^{(t)}_{ij})$ is unknown as Z_{iq} and Z_{jl} are not independant. Variational approach: take an approximation for $\mathbb{P}(Z|\Omega^t) := R_t(Z)$ for E step.

Now we consider the lower bond $\mathcal{J}$ of $\mathbb{P}(X, \Omega)$:

$$\mathcal{J}_\tau(X, \Omega, R(Z)) := \log \mathbb{P}(X, \Omega) - D_{KL}\{R(Z)||\mathbb{P}(Z|\Omega)\},$$

where $D_{KL}\{R(Z)||\mathbb{P}(Z|\Omega)\}$ is the Küllback-Leibler divergence (i.e. distance between these two distributions). We take the following distribution for $R(.)$[1]

$$R_\tau(Z) = \prod_{i=1}^{n} h_{\underline{\tau_i}}(Z_i),$$

where $h_{\underline{\tau_i}}$ is the density of the multinomial probability distribution $\mathcal{M}(1, \underline{\tau_i})$ and $\underline{\tau_i} = (\tau_{i1}, ..., \tau_{iQ})$ is a random vector containing the parmeters to optimize in the variational approach. Approximation of the probability that vertex i belongs to cluster q, τ_{iq} estimates $\mathbb{P}(Z_{iq} = 1|\Omega)$, under the constraint $\sum_q \tau_{iq} = 1$.

Küllback-Leibler divergence:

$$\begin{aligned}
D_{KL}\{R_\tau(Z)||\mathbb{P}(Z|\Omega)\} &= \sum_{Z \in \mathcal{Z}} R_\tau(Z) \log \frac{R_\tau(Z)}{\mathbb{P}(Z|\Omega)} \\
&= \sum_{Z \in \mathcal{Z}} R_\tau(Z) \left(\log R_\tau(Z) - \log \mathbb{P}(Z|\Omega)\right) \\
&= -\mathcal{H}(R_\tau(Z)) - \sum_{Z \in \mathcal{Z}} R_\tau(Z) \log \mathbb{P}(Z|\Omega)
\end{aligned}$$

New formula for the bound[2]:

$$\begin{aligned}
\mathcal{J}_\tau(X, \Omega, R(Z)) &= \log \mathbb{P}(X, \Omega) - D_{KL}\{R(Z)||\mathbb{P}(Z|\Omega)\} \\
&= \log \mathbb{P}(X, \Omega) + \mathcal{H}(R_\tau(Z)) + \sum_{Z \in \mathcal{Z}} R_\tau(Z) \log \mathbb{P}(Z|\Omega) \\
&= \mathcal{H}(R_\tau(Z)) + \sum_{Z \in \mathcal{Z}} R_\tau(Z) \left(\log \mathbb{P}(Z|\Omega) + \log \mathbb{P}(X, \Omega)\right) \\
&= \mathcal{H}(R_\tau(Z)) + \underbrace{\sum_{Z \in \mathcal{Z}} R_\tau(Z) \log L_c(X, \Omega, Z)}_{\hat{Q}_\tau(\Omega) = \mathbb{E}_{\mathcal{R}_\tau}[\log \mathbb{P}(X, \Omega, Z)]}
\end{aligned}$$

Objective: maximization of $\mathcal{J}(X, \Omega, R(Z))$.

[1] cf. Mariadassous and Robin, Uncovering latent structure in valued graphs: a variational approach, Technical Report 10, Statistics for Systems Biology, 2007

[2] En considérant que $\mathbb{P}(Z|\Omega) = \mathbb{P}(Z|\Omega, X)$ et en remarquant que $\log \mathbb{P}(X, \Omega)$ est indep de Z

Expression for $\hat{Q}_\tau(\Omega)$ Assume $R_\tau(Z) = \prod_{i=1}^n h_{\underline{\tau_i}}(Z_i)$

$$\hat{Q}_\tau(\Omega) = \sum_{Z \in \mathcal{Z}} R_\tau(Z) \log L_c(X, \Omega, Z) = \mathbb{E}_{R_\tau}[\log L_c(X, \Omega, Z)]$$
$$= \frac{n}{2}\left(\log(|\Omega|) - tr(S\Omega) - p\log(2\pi)\right) + \sum_{i,j=1, j\neq i}^{n} \sum_{q,l=1}^{Q} \tau_{iq}\tau_{jl}\left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}}\right) + \sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log\alpha_q$$

Expression for $\mathcal{H}(R_\tau(Z))$ Assume $R_\tau(Z) = \prod_{i=1}^n h_{\underline{\tau_i}}(Z_i)$

$$\mathcal{H}(R_\tau(Z)) = -\sum_{i=1}^{n}\sum_{q=1}^{Q}\mathbb{P}(Z_{iq}=1)\log\mathbb{P}(Z_{iq}=1)$$
$$= -\sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log\tau_{iq}$$

Complete expression of $\mathcal{J}_\tau(X, \Omega, R(Z))$

$$\mathcal{J}_\tau(X,\Omega,R(Z)) = \mathcal{H}(R_\tau(Z)) + \hat{Q}_\tau(\Omega)$$
$$= \underbrace{\frac{n}{2}\left(\log(|\Omega|) - tr(S\Omega) - p\log(2\pi)\right)}_{\text{Constant par rapport à Z, et autres, désigné par c dans l'article}}$$
$$+ \sum_{\substack{i,j=1\\ j\neq i}}^{n}\sum_{q,l=1}^{Q}\tau_{iq}\tau_{jl}\left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}}\right) + \sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log\alpha_q - \sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log\tau_{iq}$$

Similar to the one in the complete likelihood expression, the penalty term is: $\sum_{i,j=1,j\neq i}^{n}\sum_{q,l=1}^{Q}\tau_{iq}\tau_{jl}\frac{|\Omega_{ij}|}{\lambda_{ql}}$.

3.3 E step: parameters estimation

Strategy: estimate τ_{iq} with fixed α_q and λ_{lq}, then estimate α_q and λ_{lq} considering $\hat{\tau}_{iq}$.

Estimating $\hat{\tau}_{iq}$: Introducing the constraint using Lagrange multiplier method. Constraint: $\sum_q \tau_{iq} = 1$.

$$\frac{\partial \mathcal{J}_\tau(X,\Omega,Z) - \lambda(\sum_{q=1}^{Q}\tau_{iq} - 1)}{\partial\tau_{iq}} = \sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{l=1}^{Q}\tau_{jl}\left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}}\right) + \log\alpha_q - 1 - \log\tau_{iq} - \lambda$$

$$\tau_{iq} = \exp\left[\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{l=1}^{Q}\tau_{jl}\left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}}\right) + \log\alpha_q - 1 - \lambda\right] = \exp(-1)\exp(-\lambda)\alpha_q\exp\left[\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{l=1}^{Q}\tau_{jl}\left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}}\right)\right]$$
$$= \exp(-1)\exp(-\lambda)\alpha_q\prod_{\substack{j=1\\ j\neq i}}^{n}\prod_{l=1}^{Q}\exp\left[\tau_{jl}\left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}}\right)\right] = \exp(-1)\exp(-\lambda)\alpha_q\prod_{\substack{j=1\\ j\neq i}}^{n}\prod_{l=1}^{Q}\left(\frac{1}{2\lambda_{ql}}\exp\left(\frac{-|\Omega_{ij}|}{\lambda_{ql}}\right)\right)^{\tau_{jl}}$$

Using the constraint to find λ:

$$\sum_{q=1}^{Q}\tau_{iq} = 1 = \exp(-1)\exp(-\lambda)\alpha_q \prod_{\substack{j=1\\ j\neq i}}^{n}\prod_{l=1}^{Q}\left(\frac{1}{2\lambda_{ql}}\exp\left(\frac{-|\Omega_{ij}|}{\lambda_{ql}}\right)\right)^{\tau_{jl}}$$

$$\Rightarrow \exp(\lambda) = \left[\sum_{q=1}^{Q}\exp(-1)\alpha_q \prod_{\substack{j=1\\ j\neq i}}^{n}\prod_{l=1}^{Q}\left(\frac{1}{2\lambda_{ql}}\exp\left(\frac{-|\Omega_{ij}|}{\lambda_{ql}}\right)\right)^{\tau_{jl}}\right]^{-1}$$

$$\boxed{\hat{\tau}_{iq} = \frac{\exp(-1)\alpha_q \prod_{\substack{j=1\\ j\neq i}}^{n}\prod_{l=1}^{Q}\left(\frac{1}{2\lambda_{ql}}\exp\left(\frac{-|\Omega_{ij}|}{\lambda_{ql}}\right)\right)^{\tau_{jl}}}{\sum_{q=1}^{Q}\exp(-1)\alpha_q \prod_{\substack{j=1\\ j\neq i}}^{n}\prod_{l=1}^{Q}\left(\frac{1}{2\lambda_{ql}}\exp\left(\frac{-|\Omega_{ij}|}{\lambda_{ql}}\right)\right)^{\tau_{jl}}}}$$

In the paper:

$$\boxed{\hat{\tau}_{iq} \propto \alpha_q \prod_{\substack{j=1\\ j\neq i}}^{n}\prod_{l=1}^{Q}\left(\frac{1}{2\lambda_{ql}}\exp\left(\frac{-|\Omega_{ij}|}{\lambda_{ql}}\right)\right)^{\tau_{jl}}}$$

Estimating $\hat{\alpha}_q$: With constraint $\sum_q \alpha_q = 1$.

$$\left.\frac{\partial \mathcal{J}_\tau(X,\Omega,Z)}{\partial \alpha_q}\right|_{\tau_{iq}} = \sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\frac{1}{\alpha_q}$$

The value if α_q will be derived from the following expression:

$$\left.\frac{\partial \mathcal{J}_\tau(X,\Omega,Z) - \eta(\sum_q \alpha_q - 1)}{\partial \alpha_q}\right|_{\tau_{iq}} = \sum_{i=1}^{n}\tau_{iq}\frac{1}{\alpha_q} - \eta$$

$$0 = \sum_{i=1}^{n}\tau_{iq}\frac{1}{\alpha_q} - \eta \Leftrightarrow \alpha_q = \frac{\sum_i \tau_{iq}}{\eta}$$

$$\Leftrightarrow \sum_q \alpha_q = \frac{\sum_q\sum_i \tau_{iq}}{\eta} \Leftrightarrow 1 = \frac{n}{\eta} \Leftrightarrow \boxed{\eta = n}$$

$$0 = \sum_{i=1}^{n}\tau_{iq}\frac{1}{\alpha_q} - \eta \Leftrightarrow \alpha_q = \frac{\sum_i^n \tau_{iq}}{\eta}$$

$$\boxed{\hat{\alpha}_q = \frac{\sum_i^n \tau_{iq}}{n}}$$

Estimating $\hat{\lambda}_{lq}$:

$$\left.\frac{\partial \mathcal{J}_\tau(X,\Omega,Z)}{\partial \lambda_{ql}}\right|_{\tau_{iq}} = \sum_{\substack{j=1\\ j\neq i}}^{p}\tau_{iq}\tau_{jl}\left(-\frac{1}{2\lambda_{ql}} + \frac{|\Omega_{ij}|}{\lambda_{ql}^2}\right)$$

$$
\begin{aligned}
0 &= \sum_{\substack{j=1\\j\neq i}}^{p} \tau_{iq}\tau_{jl}\left(-\frac{1}{2\lambda_{ql}}+\frac{|\Omega_{ij}|}{\lambda_{ql}^2}\right)\\
\Rightarrow 0 &= \sum_{\substack{j=1\\j\neq i}}^{p} \tau_{iq}\tau_{jl}\left(-\frac{1}{2}+\frac{|\Omega_{ij}|}{\lambda_{ql}}\right)
\end{aligned}
$$

$$
\boxed{\hat{\lambda}_{ql} = \frac{\sum_{\substack{j=1\\j\neq i}}^{n} \tau_{iq}\tau_{jl}|\Omega_{ij}|}{\sum_{\substack{j=1\\j\neq i}}^{n} \tau_{iq}\tau_{jl}}}
$$

4 Likelihood penalized point of view

Maybe talking of "model" like in Section 1 is not a good idea. We should consider the penalized likelihood approach.

Inference of graphical model with prior point of view Originally, to estimate Ω, it is standard to consider the lasso penalization

$$
\begin{aligned}
&= \log \mathbb{P}(X|\Omega) - \lambda^{-1}\sum_{i<j}|\Omega_{ij}| - \lambda_0^{-1}\sum_{i=1}^{p}|\Omega_{ii}|\\
\widehat{\Omega}_\lambda &= \arg\max_{\Omega} \mathcal{S}(X,\Omega,\lambda)
\end{aligned}
$$

We see that :

$$
\begin{aligned}
\lambda^{-1}\sum_{i<j}|\Omega_{ij}| &= \sum_{i<j}\log\exp\left(-\frac{\|\Omega\|}{\lambda}\right)\\
&= \sum_{i<j}\log\left(\frac{1}{2\lambda}\exp\left(-\frac{|\Omega_{ij}|}{\lambda}\right)\right) + \sum_{i<j}\log(2\lambda)\\
&= \log\mathbb{P}(\Omega;\lambda) + Cste(\lambda) + \frac{p(p-1)}{2}\log(\lambda)
\end{aligned}
$$

where $\mathbb{P}(\Omega;\lambda)$ is the prior such that the

$$
\Omega_{ij} \sim_{i.i.d} Laplace(0,\lambda), \quad i<j, \text{ and } \quad \Omega_{ii} \sim_{i.i.d} Laplace(0,\lambda_0),
$$

normalized over the precision matrices that are positive definite (meaning that we set a null prior on the non positive definite matrices Ω). More precisely, let $\mathcal{D}$ be the positive definite matrices, we set:

$$
\mathbb{P}(\Omega;\lambda) = \frac{\mathbf{1}_{\Omega\in\mathcal{D}}\prod_{i<j}\lambda e^{-\lambda|\Omega_{ij}|}\prod_i \lambda_0 e^{-\lambda_0|\Omega_{ii}|}}{\int \mathbf{1}_{\Omega\in\mathcal{D}}\prod_{i<j}\lambda e^{-\lambda|\Omega_{ij}|}\prod_i \lambda_0 e^{-\lambda_0|\Omega_{ii}|}d\Omega}
$$

If λ is known, then the denominator of $\mathbb{P}(\Omega;\lambda)$ does not depend on Ω and so :

$$
\begin{aligned}
\widehat{\Omega}_\lambda &= \arg\max_{\Omega} \mathcal{S}(X,\Omega,\lambda)\\
&= \arg\max_{\Omega}\ \log\mathbb{P}(X|\Omega) + \log\mathbb{P}(\Omega;\lambda) + Cste(\lambda)\\
&= \arg\max_{\Omega}\ \log\mathbb{P}(X|\Omega) + \log\mathbb{P}(\Omega;\lambda)\\
&= \arg\max_{\Omega} \mathbb{P}_\lambda(\Omega|X)
\end{aligned}
$$

where $\mathbb{P}_\lambda(\Omega|X)$ is the posterior distribution corresponding to the prior distribution $\mathbb{P}(\Omega;\lambda)$

A different prior distribution on Ω Now we propose to change the prior distribution :

We set a SBM prior on Ω. We fix K the number of clusters

$$\begin{array}{rcl} Z_i & \sim & i.i.d. \mathcal{M}(1,\pi) \quad \forall i \in \{1,\dots,p\} \\ \Omega_{ij}|Z_i = k, Z_j = \ell & \sim & Laplace(\lambda_{k\ell}) \quad \forall i < j \in \{1,\dots,p\} \\ \Omega_{ii} & \sim & Laplace(\lambda_0) \quad \forall i \in \{1,\dots,p\} \end{array}$$

But, to consider the "prior" point of view, we have to restrict this prior over the definite positive matrices

$$\mathbb{P}_{\pi,\lambda,K,\mathcal{D}}(\Omega) \propto \mathbf{1}_{\Omega\in\mathcal{D}} \sum_{\mathbf{Z}} P(\Omega|\mathbf{Z};\lambda)P(\mathbf{Z};\pi) = \frac{\mathbf{1}_{\Omega\in\mathcal{D}} \sum_{\mathbf{Z}} P(\Omega|\mathbf{Z};\lambda)P(\mathbf{Z};\pi)}{\int \mathcal{D} \sum_{\mathbf{Z}} P(\Omega|\mathbf{Z};\lambda)P(\mathbf{Z};\pi)d\Omega}$$

If K, λ and π are know then we have to optimize:

$$\widehat{\Omega}_{\pi,\lambda,K} = \arg\max_{\Omega} \mathcal{S}(X,\Omega,\lambda,\pi,K)$$
$$\mathcal{S}(X,\Omega,\lambda,\pi,K) = \log \mathbb{P}(X|\Omega) + \log \mathbb{P}_{\pi,\lambda,K,\mathcal{D}}(\Omega;\theta)$$

However, that way, we have to optimize Ω under the constraint...

An other strategy similar to the one proposed by [1] or [2] is to remove the constraint from the prior and maximise a pseudo-likelihood. More precisely :

$$\widehat{\Omega}_{\lambda} = \arg\max_{\Omega} \widetilde{\mathcal{S}}(X,\Omega,\lambda)$$

with

$$\widetilde{\mathcal{S}}(X,\Omega,\lambda,\pi,K) = \log \widetilde{\mathcal{S}}(X|\Omega) + \log \mathbb{P}_{\pi,\lambda,K}(\Omega)$$

where

$$\mathbb{P}_{\pi,\lambda,K}(\Omega) = \sum_{\mathbf{Z}} P(\Omega|\mathbf{Z};\lambda)P(\mathbf{Z};\pi)$$

and $\log \widetilde{\mathcal{S}}(X|\Omega)$ is the linear regression version of the Ω_{ij}.

New strategy An other strategy could be to put a SBM prior distribution on the Ω_{ij} $(i<j)$ and set $\Omega_{ii} = \sum_{j\neq i} \Omega_{ij}$ In that case, the matrix Ω automatically becomes positive definite. It could be any other transformation making the matrix inversible.

So let us define : Ω_T the upper triangular coefficients of Ω. We set a prior SBM distribution on Ω_T. Ω is deduced deterministically from Ω_T. We denote by Φ this deterministic function.

We set :

$$\widetilde{\widetilde{\mathcal{S}}}(X,\Omega_T,\lambda,\Omega,\lambda,\pi,K) = \log \mathbb{P}(X|\Phi(\Omega_T)) + \log \mathbb{P}_{\pi,\lambda,K}(\Omega_T)$$

We want to optimize it with respect to Ω_T and π,λ (for a fixed K). We won't be able to do it because $\log \mathbb{P}_{\pi,\lambda,K}(\Omega_T)$ has no explicit expressions. We introduced a lower bound of $\widetilde{\widetilde{\mathcal{S}}}(X,\Omega_T,\lambda,\Omega,\lambda,\pi,K)$, namely $\mathcal{J}(X,\Omega_T,\lambda,\Omega,\lambda,\pi,K,\tau)$

$$\begin{array}{rcl} \mathcal{J}_K(X;\Omega_T,\lambda,\pi,\tau) & = & \log \mathbb{P}(X|\Phi(\Omega_T)) + \log \mathbb{P}_{\pi,\lambda,K}(\Omega_T) - \mathcal{K}(\mathcal{R}_\tau, P(Z|\Omega)) \\ & = & \log \mathbb{P}(X|\Phi(\Omega_T)) + \mathbb{E}_{\mathcal{R}_\tau}\left[\log P(\Omega_T,\mathbf{Z};\lambda,\pi)\right] + \mathcal{H}(\mathcal{R}_\tau(\cdot)) \end{array}$$

We can optimize it iteratively. At iteration t.

1. $$(\pi^{(t)},\lambda^{(t)}) = \arg\max_{\pi,\lambda} \mathcal{J}_K(X;\Omega_T^{(t-1)},\lambda,\pi,\tau^{(t-1)})$$

   It is the M-step of the VEM.

2.

$$\tau^{(t)} = \arg\max_{\tau} \mathcal{J}_K(X; \Omega_T^{(t-1)}, \lambda^{(t)}, \pi^{(t)}, \tau)$$

It the E-step of the VEM. Resulting into a fixed point equation. We could also use a lower bound of this preventing from resolving a fixed point equation (GEM).

3.

$$\begin{aligned}
\Omega_T^{(t)} &= \arg\max_{\Omega_T} \log \mathbb{P}(X|\Phi(\Omega_T)) + \mathbb{E}_{\mathcal{R}_{\tau^{(t)}}}\left[\log P(\Omega_T, \mathbf{Z}; \lambda^{(t)}, \pi^{(t)})\right] + \mathcal{H}(\mathcal{R}_{\tau^{(t)}}(\cdot)) \\
&= \arg\max_{\Omega_T} \left[\log \mathbb{P}(X|\Phi(\Omega_T)) + \sum_{i<j}\sum_{k,l} \tau_{ik}^{(t)}\tau_{j\ell}^{(t)}(\log \lambda_{k,\ell}^{(t)} - \lambda_{k\ell}^{(t)}|\Omega_{ij}|)\right] \\
&= \arg\max_{\Omega_T} \left[\log \mathbb{P}(X|\Phi(\Omega_T)) - \sum_{i<j}\sum_{k,l} \tau_{ik}^{(t)}\tau_{j\ell}^{(t)}\lambda_{k\ell}^{(t)}|\Omega_{ij}|\right]
\end{aligned}$$

I state that it is equivalent to estimating Ω in its complete version (upper and diagonal) under the constraint $\Omega_{ii} = \sum_{i\neq j} \Omega_{ij}$ which is a linear constraint :

$$\Omega_{ii} - \sum_{i\neq j} \Omega_{ij} = 0$$

Using a Lagrange approach, we get:

$$\Omega^{(t)} = \arg\max_{\Omega} \left[\log \mathbb{P}(X|\Omega) - \sum_{i<j}\sum_{k,l} \tau_{ik}^{(t)}\tau_{j\ell}^{(t)}\lambda_{k\ell}^{(t)}|\Omega_{ij}|\right] + \sum_{i=1}^{p} \nu_i \left(\Omega_{ii} - \sum_{i\neq j} \Omega_{ij}\right)$$

And $\Omega_T^{(t)}$ is the upper triangular part of $\Omega^{(t)}$ Is it feasible Julien ????

References

[1] Benjamin M. Marlin and Kevin P. Murphy. Sparse Gaussian graphical models with unknown block structure. pages 1–8, 2009.

[2] Siqi Sun and Hai Wang. Inferring Block Structure of Graphical Models in Exponential Families. *Proceedings of the 18th International Conference on Artifical Intelligence and Statistics (AISTATS) International Conference on Machine Learning - ICML '09*, 38, 2015.