Bayesian structure learning in graphical models

Seung Bong Jung

February 13, 2021

Contents

Abstract

In this note, we introduce the method of estimating of a sparse precision matrix of a multivariate Gaussian distribution, where dimension p is very large, from Bayesian perspective based on the article "Bayesian structure learning in graphical models" written by Sayantan Banerjee and Subhashis Ghosal. There are some well-known frequentist methods for the estimation. These methods include Graphical Lasso(denoted by 'GL'), which minimizes the penalized log-likelihood of data with a Lasso type penalty on the elements of the precision matrix. Its Bayesian version also has been developed by H.Wang. However, there are some shortcomings. In Bayesian version of Graphical lasso, we put Exponential priors on the diagonal elements of the precision matrix, while Laplace priors on the off-diagonal elements. But in doing so, we do not introduce any sparsity in the graphical structure due to the absence of point mass at zero in Laplace priors on the off-diagonal elements. To overcome such shortcomings, we introduce point masses to the prior of off-diagonal elements. Since it is difficult to compute posterior distribution with point masses introduced in the priors, we approximate posterior distribution around the posterior mode through Laplace approximation and calculate posterior convergence rate in terms of Frobenius norm. Also, we show we can ignore the non-regular model under specific conditions and estimate the error in the Laplace approximation for regular models. Finally, we provide some simulation results and the Graphical model for real data using the method suggested in this note.

1 Notations and preliminaries

- **Notation 1:** Let $\mathbf{V}$ be the set of indices of p vertices of p-dimensional random vector $\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{\Omega^{-1}})$ and suppose an undirected graph $\mathbf{G}$ comprises with set $\mathbf{V}$. Then we denote the set of edges in the graph by $\mathbf{E} \subset \{(i,j) \in \mathbf{V} \times \mathbf{V} : i < j\}$ and let $\bar{\mathbf{E}} = \mathbf{E} \cup \{(i,i) : 1 \leq i \leq p\}$. Also, let $\mathbf{\Omega} = ((\omega_{ij}))$ We may write **G=(V,E)**, meaning graphical model $\mathbf{G}$ has nodes $\mathbf{V}$ with edges $\mathbf{E}$.

Theorem 1.1. *Suppose $\boldsymbol{X} \sim N_p(\boldsymbol{0}, \mathbf{\Omega^{-1}})$. Then $\omega_{ij} = \omega_{ji} = 0$ implies conditional independence of X_i and X_j given $(X_r : r \neq i, j)$ with $i \neq j$.*

Proof. w.l.o.g. assume $i = 1$, $j = 2$. Let $\mathbf{\Sigma} = \mathbf{\Omega^{-1}} = ((\sigma_{ij}))$. Consider $\mathbf{\Sigma}$ as following :

$$\mathbf{\Sigma} = \left[\begin{array}{c|c} \mathbf{\Sigma_{11}} & \mathbf{\Sigma_{12}} \\ \hline \mathbf{\Sigma_{21}} & \mathbf{\Sigma_{22}} \end{array}\right]$$

where $\mathbf{\Sigma_{11}}$ is 2×2 matrix, $\mathbf{\Sigma_{22}}$ is $(p-2) \times (p-2)$ matrix, and $\mathbf{\Sigma_{12}}$ is $2 \times (p-2)$ matrix, $\mathbf{\Sigma_{21}} = \mathbf{\Sigma_{12}^t}$.

Then $X_1, X_2 | X_{r:r\neq 1,2} \sim N_2(\mathbf{0}, \mathbf{\Sigma_{11.2}^{-1}})$, where $\mathbf{\Sigma_{11.2}} = \mathbf{\Sigma_{11}} - \mathbf{\Sigma_{12}}\mathbf{\Sigma_{22}^{-1}}\mathbf{\Sigma_{21}} = \mathbf{\Omega_{11}^{-1}}$. Here $\mathbf{\Omega_{11}}$ follows block representation for $\mathbf{\Sigma}$ similarly. Thus, $\mathbf{Cov(X_1, X_2|X_{r:r\neq 1,2})} = (\mathbf{\Omega_{11}^{-1}})_{12}$. Since X_i and X_j given $(X_{r:r\neq 1,2})$ are independent if and only if their conditional covariance given other covariates is 0, with symmetry of $\mathbf{\Omega_{11}}$, we have $(\mathbf{\Omega_{11}^{-1}})_{\mathbf{12}} = (\mathbf{\Omega_{11}^{-1}})_{\mathbf{21}} = 0$, which implies 2×2 matrix $(\mathbf{\Omega_{11}})^{\mathbf{-1}}$ is a diagonal matrix. Hence, $\mathbf{\Omega_{11}}$ is also a diagonal matrix and so $\omega_{\mathbf{12}} = \omega_{\mathbf{21}} = \mathbf{0}$. Thus, we get the desired result. □

From theorem 1.1, we can see that Gaussian Graphical Structure is indeed useful for introducing sparsity structure in the precision matrix because conditional independence given other nodes being 0 between two nodes implies there is no edge between them.

- **Notation 2:** Let $\mathcal{M}$ denote the linear space of symmetric matrices of order p and $\mathcal{M}^+ \subset \mathcal{M}$ the cone of positive definite matrices of order p. If $\mathbf{A} \in \mathcal{M}^+$, we say $\mathbf{A} > \mathbf{0}$(or $\mathbf{0} < \mathbf{A}$) and denote is unique positive definite square root by $\mathbf{A}^{1/2}$, where $\mathbf{0}$ is $p \times p$ zero matrix. Let denote $p \times p$ identity matrix in $\mathcal{M}$ by $\mathbf{I_p}$. Also, let $\mathcal{P}_G \subset \mathcal{M}^+$ be the cone of positive definite matrices of p with zero entries corresponding to each missing edge in $\mathbf{E}$.

Definition 1.1. Suppose two numerical sequences r_n and s_n are given. If $\frac{r_n}{s_n}$ is bounded as $n \to \infty$, we say $r_n = O(s_n)$ or equivalently $r_n \lesssim s_n$(or $s_n \gtrsim r_n$). If $r_n \lesssim s_n$ and $s_n \lesssim r_n$ both hold, we denote this by $r_n \asymp s_n$. In case $\frac{r_n}{s_n} \to 0$ as $n \to \infty$, we denote this by $r_n = o(s_n)$ or equivalently $r_n \ll s_n$(or $s_n \gg r_n$). If $\frac{r_n}{s_m} \to 1$ as $n \to \infty$, denote this by $r_n \sim s_n$.

Definition 1.2. Suppose random sequence X_n is given. We say $X_n = O_p(\delta_n)$, if ${}^\exists M$ such that $\mathbf{P}(|X_n| \leqslant M\delta_n) \to 1$ as $n \to \infty$. Similarly, we say $X_n = o_p(\delta_n)$, if ${}^\forall \epsilon > 0$, $\mathbf{P}(|X_n| < \epsilon\delta_n) \to 1$ as $n \to \infty$.

- **Notation 3:** Suppose $\mathbf{A} = ((a_{ij}))$ is $p \times p$ matrix and $\mathbf{x} \in \mathbb{R}^p$. Then, for $1 \leq r < \infty$, $||\mathbf{A}||_r = (\sum_{i=1}^{p}\sum_{j=1}^{p}|a_{ij}|^r)^{1/r}$(respectively, $||\mathbf{x}||_r = (\sum_{i=1}^{p}|x_i|^r)^{1/r}$) and $||\mathbf{A}||_\infty = \max_{i,j}|a_{ij}|$(respectively, $||\mathbf{x}||_\infty = (\max_i |x_i|)$ If $\mathbf{A}$ is symmetric, let $\text{eig}_i(\mathbf{A})$ denote ith biggest eigenvalue of $\mathbf{A}$.

- **Notation 4:** Suppose $\mathbf{A} = ((a_{ij}))$ is $r \times s$ matrix. Then, $||\mathbf{A}||_{(r,s)} = \sup(||\mathbf{Ax}||_s : ||\mathbf{x}||_r = 1, \mathbf{x} \in \mathbb{R}^\mathbf{r})$

Using the notations in the **Notation 3** and **4**, if $\mathbf{A}$ is $p \times p$ symmetric matrix, clearly one can see $||\mathbf{A}||_2 = \sqrt{\mathbf{tr(A^tA)}}$, which is Frobenius norm, and $||\mathbf{A}||_{(2,2)} = \sqrt{\max_i eig_i(\mathbf{A^tA})} = \max_i |eig_i(\mathbf{A})|$.

Theorem 1.2. *Let $\mathbf{A} = ((a_{ij})), \mathbf{B} = ((b_{ij})) \in \mathcal{M}$. Then followings hold:*

$$\begin{gathered} ||\mathbf{A}||_\infty \leq ||\mathbf{A}||_{\mathbf{(2,2)}} \leq ||\mathbf{A}||_\mathbf{2} \leq \mathbf{p}||\mathbf{A}||_\infty, \\ ||\mathbf{AB}||_\mathbf{2} \leq \min\{||\mathbf{A}||_{\mathbf{(2,2)}}||\mathbf{B}||_\mathbf{2}, ||\mathbf{B}||_{\mathbf{(2,2)}}||\mathbf{A}||_\mathbf{2}\} \end{gathered} \tag{1}$$

Proof. We first prove the first inequality. Let $\mathbf{A} = \mathbf{P\Lambda P^t}$ be spectral decomposition of $\mathbf{A}$, where $\mathbf{\Lambda}$ is a diagonal matrix with eigenvalues λ_i on diagonal entries in biggest order and $\mathbf{P}$ is orthogonal matrix with corresponding eigenvector on columns. Suppose $||\mathbf{A}||_\infty = |a_{ij}|$ and let $\mathbf{e_j} \in \mathbb{R}^p$ with 1 on jth entry and 0 elsewhere. Then $||\mathbf{A}||_{(2,2)} = \sup(||\mathbf{Ax}||_\mathbf{2} : ||\mathbf{x}||_\mathbf{2} = \mathbf{1}) \geq ||\mathbf{Ae_j}||_\mathbf{2} = (\sum_{k=1}^{p} a_{kj}^2)^{1/2} \geq (a_{ij}^2)^{1/2} = (||\mathbf{A}||_\infty^\mathbf{2})^{\mathbf{1/2}} = ||\mathbf{A}||_\infty$ $(\because ||\mathbf{e_j}||_\mathbf{2} = \mathbf{1})$. Also,

$$\begin{aligned} ||\mathbf{A}||_\mathbf{2} &= \sqrt{\mathbf{tr(A^tA)}} = \sqrt{\mathbf{tr(P\Lambda P^tP\Lambda P^t)}} = \sqrt{\mathbf{tr(P\Lambda^2P^t)}} \\ &= \sqrt{\mathbf{tr(P^tP\Lambda^2)}} = \sqrt{\mathbf{tr(\Lambda^2)}} = \sqrt{\sum_{\mathbf{k=1}}^{\mathbf{p}} \lambda_\mathbf{k}^\mathbf{2}} \geq \sqrt{||\mathbf{A}||_{\mathbf{(2,2)}}^\mathbf{2}} = ||\mathbf{A}||_{\mathbf{(2,2)}}. \end{aligned}$$

Finally, $||\mathbf{A}||_\mathbf{2} = \sqrt{\sum_{\mathbf{i,j=1}}^{\mathbf{p}} \mathbf{a_{ij}^2}} \leq \sqrt{\sum_{\mathbf{i,j=1}}^{\mathbf{p}} ||\mathbf{A}||_\infty^\mathbf{2}} = \sqrt{\mathbf{p^2}||\mathbf{A}||_\infty^\mathbf{2}} = \mathbf{p}||\mathbf{A}||_\infty$.

Now, it remains to show the second inequality.

$$||\mathbf{AB}||_\mathbf{2} = \sqrt{\mathbf{tr}((\mathbf{AB})^\mathbf{t}\mathbf{AB})} = \sqrt{\mathbf{tr}(\mathbf{B^tA^tAB})} = \sqrt{\mathbf{tr}(\mathbf{B^tP\Lambda^2P^tB})} = \sqrt{\sum_{\mathbf{i,j,k=1}}^{\mathbf{p}} (\mathbf{B^tP})_\mathbf{ij}\mathbf{\Lambda}^\mathbf{2}_\mathbf{jk}(\mathbf{P^tB})_\mathbf{ki}}$$

$$= \sqrt{\sum_{\mathbf{i,j=1}}^{\mathbf{p}} (\mathbf{B^tP})_\mathbf{ij}\lambda^\mathbf{2}_\mathbf{j}(\mathbf{P^tB})_\mathbf{ji}} = \sqrt{\sum_{\mathbf{i,j=1}}^{\mathbf{p}} (\mathbf{B^tP})_\mathbf{ij}\lambda^\mathbf{2}_\mathbf{j}(\mathbf{B^tP})_\mathbf{ij}} = \sqrt{\sum_{\mathbf{i,j=1}}^{\mathbf{p}} (\mathbf{B^tP})^\mathbf{2}_\mathbf{ij}\lambda^\mathbf{2}_\mathbf{j}} \le \sqrt{||\mathbf{A}||^\mathbf{2}_{(\mathbf{2,2})}\sum_{\mathbf{i,j=1}}^{\mathbf{p}} (\mathbf{B^tP})^\mathbf{2}_\mathbf{ij}}$$

$$= ||\mathbf{A}||_{(\mathbf{2,2})}\sqrt{\sum_{\mathbf{i,j=1}}^{\mathbf{p}} (\mathbf{B^tP})^\mathbf{2}_\mathbf{ij}} = ||\mathbf{A}||_{(\mathbf{2,2})}||\mathbf{B^tP}||_\mathbf{2} = ||\mathbf{A}||_{(\mathbf{2,2})}\sqrt{\mathbf{tr}((\mathbf{B^tP})^\mathbf{t}\mathbf{B^tP})} = ||\mathbf{A}||_{(\mathbf{2,2})}\sqrt{\mathbf{tr}(\mathbf{P^tBB^tP})}$$

$$= ||\mathbf{A}||_{(\mathbf{2,2})}\sqrt{\mathbf{tr}(\mathbf{B^tPP^tB})} = ||\mathbf{A}||_{(\mathbf{2,2})}\sqrt{\mathbf{tr}(\mathbf{B^tB})} = ||\mathbf{A}||_{(\mathbf{2,2})}||\mathbf{B}||_\mathbf{2}$$

As $||\mathbf{AB}||_\mathbf{2} = \sqrt{\mathbf{tr}(\mathbf{B^tA^tAB})} = \sqrt{\mathbf{tr}(\mathbf{ABB^tA^t})}$, do spectral decomposition to $\mathbf{B}$ and the same job in the above to $\sqrt{\mathbf{tr}(\mathbf{ABB^tA^t})}$. Then we obtain $||\mathbf{AB}||_\mathbf{2} \le ||\mathbf{B}||_{(\mathbf{2,2})}||\mathbf{A}||_\mathbf{2}$. Therefore, combining this inequality and the inequality in the above, we have the second inequality. $\square$

- **Notation 5:** The cardinality of a set $\mathbf{T}$ is denoted by $\#\mathbf{T}$. Let $\mathbb{1}$ be indicator function. Define the symmetric matrix $\mathbf{E}_{(ij)} = ((\mathbb{1}_{\{(i,j),(j,i)\}}(l,m)))$
- **Notation 6:** For a subset $\mathbf{A}$ of a metric space $(\mathbf{S},\mathbf{d})$, the minimum number of $\mathbf{d}$-balls of size ϵ in $\mathbf{S}$ needed to cover $\mathbf{A}$ is denoted by $\mathbf{N}(\epsilon,\mathbf{A},\mathbf{d})$

Definition 1.3. The Hellinger distance between two probability densities $\mathbf{q_1}$ and $\mathbf{q_2}$ is defined by $\mathbf{h}(\mathbf{q_1},\mathbf{q_2}) = ||\sqrt{\mathbf{q_1}} - \sqrt{\mathbf{q_2}}||_\mathbf{2} = (\int(\sqrt{\mathbf{q_1(x)}} - \sqrt{\mathbf{q_2(x)}})^\mathbf{2}\mathbf{dx})^{\mathbf{1/2}}$

In this note, we may use Hellinger distance as the measure of how close two probability densities are.

2 Prior settings and posterior concentration

Let $\mathbf{X_1},\mathbf{X_2},\dots,\mathbf{X_n} \overset{\mathbf{i.i.d.}}{\sim} \mathbf{N_p}(\mathbf{0},\mathbf{\Sigma})$, where $\mathbf{\Sigma} \in \mathcal{M}^+$ and the precision matrix $\mathbf{\Omega} = \mathbf{\Sigma^{-1}}$. Then we clearly see that $\hat{\mathbf{\Sigma}} = \frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{X_iX_i^t}$ is MLE of Σ. Then the graphical lasso solves the following optimization problem with penalty parameter $\lambda \ge 0$ and constraint $\mathbf{\Omega} \in \mathcal{M}^+$:

$$\text{minimize} \quad \underbrace{-\log\det(\mathbf{\Omega}) + \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega})}_{\text{negative data log-likelihood}} + \underbrace{\frac{\lambda}{\mathbf{n}}||\mathbf{\Omega}||_\mathbf{1}}_{\text{Lasso type penalty}} \tag{2}$$

Let $\mathcal{U}(\epsilon_0,s) = \{\mathbf{\Omega} \in \mathcal{M}^+ : \#\{(i,j) : 1 \le i < j \le p, \omega_{ij} \ne 0\} \le s, 0 < \epsilon_0 \le \mathbf{eig}_1(\mathbf{\Omega}) \le \mathbf{eig}_p(\mathbf{\Omega}) \le \epsilon_0^{-1} < \infty\}$ Denote the solution of (2) by $\mathbf{\Omega}^*$, which is graphical lasso estimate of true precision matrix $\mathbf{\Omega_0}$. Then, by Theorem 1 in article "Sparse permutation invariant covariance estimation" written by Rothman, $||\mathbf{\Omega}^* - \mathbf{\Omega_0}||_\mathbf{2} = \mathbf{O_p}(\sqrt{\frac{(p+s)\log p}{n}})$. Hence $||\mathbf{\Omega}^* - \mathbf{\Omega_0}||_\mathbf{2} \xrightarrow{P} 0$ as $n^{-1}(p+s)\log p \to 0$.

Theorem 2.1. *Suppose conditions stated in the above hold. Then* $||\mathbf{\Omega}^*||_{(\mathbf{2,2})} = \mathbf{O_p(1)}$ *and* $||\mathbf{\Omega}^{*-\mathbf{1}}||_{(\mathbf{2,2})} = \mathbf{O_p(1)}$

Proof. For the first equation, by triangle inequality,

$$||\mathbf{\Omega}^*||_{(\mathbf{2,2})} \le ||\mathbf{\Omega}^* - \mathbf{\Omega_0}||_{(\mathbf{2,2})} + ||\mathbf{\Omega_0}||_{(\mathbf{2,2})} \le ||\mathbf{\Omega}^* - \mathbf{\Omega_0}||_{(\mathbf{2,2})} + ||\mathbf{\Omega_0}||_\mathbf{2}$$

Now as $||\mathbf{\Omega}^* - \mathbf{\Omega_0}||_{(\mathbf{2,2})} = o_p(1)$ under the stated conditions and $||\mathbf{\Omega_0}||_\mathbf{2}$ is constant, we have $||\mathbf{\Omega}^*||_{(\mathbf{2,2})} = \mathbf{O_p(1)}$.

To show the second equation, again by triangle inequality,

$$||\mathbf{\Omega^{*-1}}||_{(\mathbf{2},\mathbf{2})} \leq ||\mathbf{\Omega^{*-1}} - \mathbf{\Omega_0^{-1}}||_{(\mathbf{2},\mathbf{2})} + ||\mathbf{\Omega_0^{-1}}||_{(\mathbf{2},\mathbf{2})} = ||\mathbf{\Omega^{*-1}}||_{(\mathbf{2},\mathbf{2})}||\mathbf{\Omega^*} - \mathbf{\Omega_0}||_{(\mathbf{2},\mathbf{2})}||\mathbf{\Omega_0^{-1}}||_{(\mathbf{2},\mathbf{2})} + ||\mathbf{\Omega_0^{-1}}||_{(\mathbf{2},\mathbf{2})}$$

Therefore, we have $(\mathbf{1} - ||\mathbf{\Omega_0^{-1}}||_{(\mathbf{2},\mathbf{2})}||\mathbf{\Omega^*} - \mathbf{\Omega_0}||_{(\mathbf{2},\mathbf{2})})||\mathbf{\Omega^{*-1}}||_{(\mathbf{2},\mathbf{2})} \leq ||\Omega_0^{-1}||_{(2,2)}$ or equivalently $||\mathbf{\Omega^{*-1}}||_{(\mathbf{2},\mathbf{2})} \leq \frac{||\mathbf{\Omega_0^{-1}}||_{(\mathbf{2},\mathbf{2})}}{\mathbf{1}-||\mathbf{\Omega_0^{-1}}||_{(\mathbf{2},\mathbf{2})}||\mathbf{\Omega^*}-\mathbf{\Omega_0}||_{(\mathbf{2},\mathbf{2})}}$. Now as $||\mathbf{\Omega^*} - \mathbf{\Omega_0}||_{(\mathbf{2},\mathbf{2})} \leq ||\mathbf{\Omega^*} - \mathbf{\Omega_0}||_{\mathbf{2}}$, $||\mathbf{\Omega^{*-1}}||_{(\mathbf{2},\mathbf{2})} \leq \frac{||\mathbf{\Omega_0^{-1}}||_{(\mathbf{2},\mathbf{2})}}{\mathbf{1}-||\mathbf{\Omega_0^{-1}}||_{(\mathbf{2},\mathbf{2})}||\mathbf{\Omega^*}-\mathbf{\Omega_0}||_{\mathbf{2}}}$.

Thus we obtain $||\mathbf{\Omega^{*-1}}||_{(\mathbf{2},\mathbf{2})} = \mathbf{O_p}(\mathbf{1})$. □

Bayesian version of graphical lasso was introduced by Wang. Wang used $\lambda \exp(-\lambda x)$ as prior on diagonal elements of Ω, while $\frac{\lambda}{2}\exp(-\lambda|x|)$ on off-diagonal elements. But as priors are absolutely continuous, for probability of the event $\{\omega_{ij} = 0\}$ in prior is 0, the posterior probability of the event is also 0. Thus it is hard to introduce sparsity to precision matrix in Bayesian context of graphical lasso suggested by Wang.

To overcome such drawbacks, Banerjee and Ghosal put point-mass prior on the events $\{\omega_{ij} = 0\}$. By doing so, it becomes able to make posterior inference about the sparse structure of the underlying graphical model. Let $\mathbf{\Gamma} = (\gamma_{ij} : 1 \leq i < j \leq p)$ be a $\binom{p}{2}$-dimensional vector of edge-inclusion vector, where $\gamma_{ij} = \mathbb{1}\{(i,j) \in \mathbf{E}\}, \quad 1 \leq i < j \leq p$. Identifying element of $\mathbf{\Gamma}$ with the set of indices $\{(i,j) : \gamma_{ij} = 1\}$, let $\#\mathbf{\Gamma}$ denote the number of non-zero elements in $\mathbf{\Gamma}$, i.e. the number of edges included in the model.

As the same with priors on precision matrix suggested by Wang, we put Exponential density on diagonal elements and Laplace density on off-diagonal elements. To establish convergence rate, we need to restriction on eigenvalues of precision matrix. Let $\mathcal{M}_0^+$ be the subset of $\mathcal{M}^+$,consisting of positive definite matrices of order p, whose eigenvalues are bounded by two fixed positive numbers.

We consider two different priors on $\mathbf{\Gamma}$. Suppose $\gamma_{ij} \overset{i.i.d.}{\sim} \text{Ber}(q)$, $1 \leq i < j \leq p$, where q is fixed. In order to introduce sparsity to precision matrix, it may be reasonable to put restriction on $\#\Gamma$, namely the size of model. In case fixed positive number $\bar{r}$ is imposed as the restriction on the size of model, we have prior $\mathbf{p}(\mathbf{\Gamma}) \propto q^{\#\mathbf{\Gamma}}(1-q)^{\binom{p}{2}-\#\mathbf{\Gamma}}\mathbb{1}(\#\mathbf{\Gamma} \leq \bar{r})$. As another prior, we assume $\bar{r}$ is not fixed value, but random variable $\bar{\mathbf{R}}$. Choose the prior distribution on $\bar{\mathbf{R}}$ satisfying $\mathbf{P}(\bar{\mathbf{R}} > a_1 m) \leq e^{-a_2 m \log m}$ for some $a_1, a_2 > 0$ and ${}^\forall m \in \mathbb{N}$. In this case, we have another prior which is given by $\mathbf{p}(\mathbf{\Gamma}|\bar{\mathbf{R}}) \propto q^{\#\mathbf{\Gamma}}(1-q)^{\binom{p}{2}-\#\mathbf{\Gamma}}\mathbb{1}(\#\mathbf{\Gamma} \leq \bar{\bar{R}})$ leading to $\mathbf{p}(\mathbf{\Gamma}) \propto q^{\#\mathbf{\Gamma}}(1-q)^{\binom{p}{2}-\#\mathbf{\Gamma}}\mathbf{P}(\bar{\mathbf{R}} \geq \#\mathbf{\Gamma})$. Note that bigger q implies that the model is likely to allow more edges. To sum up, we have following priors and corresponding posteriors.

- **Prior 1:**

$$\mathbf{p}(\mathbf{\Omega}|\mathbf{\Gamma}) \propto \prod_{r_{ij}=1} \exp(-\lambda|\omega_{ij}|) \prod_{i=1,p} \exp(-\lambda\omega_{ii}/2)\mathbb{1}_{\mathcal{M}_0^+}(\Omega), \mathbf{p}(\mathbf{\Gamma}) \propto q^{\#\mathbf{\Gamma}}(1-q)^{\binom{p}{2}-\#\mathbf{\Gamma}}\mathbf{P}(\bar{\mathbf{R}} \geq \#\mathbf{\Gamma}) \quad (3)$$

- **Prior 2:**

$$\mathbf{p}(\mathbf{\Omega}|\mathbf{\Gamma}) \propto \prod_{r_{ij}=1} \exp(-\lambda|\omega_{ij}|) \prod_{i=1,p} \exp(-\lambda\omega_{ii}/2)\mathbb{1}_{\mathcal{M}_0^+}(\Omega), \mathbf{p}(\mathbf{\Gamma}) \propto q^{\#\mathbf{\Gamma}}(1-q)^{\binom{p}{2}-\#\mathbf{\Gamma}}\mathbb{1}(\#\mathbf{\Gamma} \leq \bar{\mathbf{r}}) \quad (4)$$

- **Posterior:**

$$\begin{aligned}\mathbf{p}(\mathbf{\Omega}, \mathbf{\Gamma}|\mathbf{X^{(n)}}) &\propto \mathbf{p}(\mathbf{X^{(n)}}|\mathbf{\Omega}, \mathbf{\Gamma})\mathbf{p}(\mathbf{\Omega}|\mathbf{\Gamma})\mathbf{p}(\mathbf{\Gamma}) = (2\pi)^{-np/2}(\det(\Omega))^{n/2}\exp(-n\mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega})/2) \\ &\times \prod_{\gamma_{ij}=1}\{\lambda\exp(-\lambda|\omega_{ij}|)/2\}\prod_{i=1}^{p}\{\lambda\exp(-\lambda\omega_{ii}/2)/2\} \times \mathbf{p}(\mathbf{\Gamma})\mathbb{1}_{\mathcal{M}_r^+}(\Omega) = \mathbf{C_\Gamma}\mathbf{Q}(\mathbf{\Omega}, \mathbf{\Gamma}|\mathbf{X^{(n)}}) \quad (5)\end{aligned}$$

where

$$\mathbf{C_\Gamma} = (2\pi)^{-np/2}q^{\#\mathbf{\Gamma}}(1-q)^{\binom{p}{2}-\#\mathbf{\Gamma}}(\lambda/2)^{p+\#\mathbf{\Gamma}}\beta(\mathbf{\Gamma}), \beta(\mathbf{\Gamma}) = \mathbf{p}(\mathbf{\Gamma})$$

$$\mathbf{Q}(\mathbf{\Omega}, \mathbf{\Gamma}|\mathbf{X^{(n)}}) = (\det(\mathbf{\Omega}))^{\mathbf{n/2}}\exp(-\mathbf{ntr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega})/\mathbf{2}) \prod_{\gamma_{ij}=1}\{\lambda\exp(-\lambda|\omega_{ij}|)/2\}\prod_{i=1}^{p}\{\lambda\exp(-\lambda\omega_{ii}/2)\}\mathbb{1}_{\mathcal{M}_r^+}(\mathbf{\Omega}) \quad (6)$$

The posterior is very intractable due to positive definite constraint on $\mathbf{\Omega}$. Hence we use Laplace approximation to approximate posterior around posterior mode. Specific method will be proposed on the next section. However, to understand the motivation for approximation more clearly, we may study RJMCMC(reversible jump Markov Chain Monte Carlo) algorithm, which is often used to evaluate the posterior probabilities throughout models of varying dimensions.

Suppose we have the model set $\{\mathcal{M}_k : k \in \mathbb{N}\}$ and let θ_k be parameter in model $\mathcal{M}_k$ with k being the index of model. Then posterior distribution of θ_k is given by following with data y :

$$p(\theta_k|k,y) \propto p(y|k,\theta_k)p(\theta_k|k) \tag{7}$$

Here $p(y|k,\theta_k)$ and $p(\theta_k|k)$ represent the probability model and the prior distribution of the parameters of model k. Also,

$$p(\theta_k,k|y) \propto p(k)p(\theta_k|k,y) \tag{8}$$

.

Basically, RJMCMC is generalized MCMC. It involves Metropolis-Hastings type algorithms that move a simulation analysis between models defined by (θ_k,k) and $(\theta_{k'},k')$. The resulting Markov chain simulation jumps between two different models and form sample from θ_k,k The algorithms are made reversible so as to maintain detailed balance f a irreducible and a periodic chain that converges to the correct target measure. With jump probability $J(k \to k')$, the algorithm is given by as following:

Step1: Visit to Model $\mathcal{M}_{k'}$ with probability $J(k \to k')$.

Step2: Sample u from a proposal density $q(u|\theta_k,k,k')$

Step3: Set $(\theta_{k'},u') = g_{k,k'}(\theta_k,u)$, where $g_{k,k'}$ is a bijection.

Step4: The acceptance probability of the new model, $(\theta_{k'},k')$ can be calculated between 1 and

$$\underbrace{\frac{p(y|\theta_{k'},k')p(\theta_{k'}|k')p(k')}{p(y|\theta_k,k)p(\theta_k|k)p(k)}}_{\text{model ratio}}\underbrace{\frac{J(k'\to k)q(u|\theta_{k'},k',k)}{J(k\to k')q(u|\theta_k,k,k')}\left|\frac{\partial g_{k,k'}(\theta_k,u)}{\partial(\theta_k,u)}\right|}_{\text{proposal ratio}} \tag{9}$$

Iterating step1 step4, we have sample $\{k_l : l = 1,2,\dots,L\}$ and estimate $p(k|y)$ by $\hat{p}(k|y) = \frac{1}{L}\sum_{l=1}^{L}\mathbb{1}_k(k_l)$.
In our setting, we have $2^{\binom{p}{2}}$ models. Since too many jumps occur as we can see in the algorithm of RJMCMC, it is very time-consuming to estimate posterior. Because of too many possible models, posteriors are very unreliable. Therefore, we have to resort to other method.

Before we calculate posterior, we first establish posterior convergence rate as $n \to \infty$ under some conditions. We assume that the true $\mathbf{\Omega_0} \in \mathcal{U}(\epsilon_0,s)$.

Claim 2.1. *Suppose* $\mathbf{A},\mathbf{B} \in \mathcal{M}$. *Then,* $tr(\{\mathbf{AB}\}^{\mathbf{2}}) \le tr(\mathbf{A^2B^2})$.

Proof. If $\mathbf{A},\mathbf{B} \in \mathcal{M}$, as $(\mathbf{AB}-\mathbf{BA})^{\mathbf{t}} = \mathbf{B^tA^t} - \mathbf{A^tB^t} = \mathbf{BA}-\mathbf{AB} = -(\mathbf{AB}-\mathbf{BA})$, $\mathbf{AB}-\mathbf{BA}$ is skew-hermitian. Thus,

$$\begin{aligned}
2\mathrm{tr}(\{\mathbf{AB}\}^{\mathbf{2}}) - 2\mathrm{tr}(\mathbf{A^2B^2}) &= \mathrm{tr}(\mathbf{ABAB}) + \mathrm{tr}(\mathbf{ABAB}) - \mathrm{tr}(\mathbf{A^2B^2}) - \mathrm{tr}(\mathbf{A^2B^2}) \\
&= \mathrm{tr}(\mathbf{ABAB}) + \mathrm{tr}(\mathbf{BABA}) - \mathrm{tr}(\mathbf{AB^2A}) - \mathrm{tr}(\mathbf{BA^2B}) \\
&= \mathrm{tr}(\mathbf{ABAB}+\mathbf{BABA}-\mathbf{AB^2A}-\mathbf{BA^2B}) \\
&= \mathrm{tr}(\mathbf{AB}-\mathbf{BA})^{\mathbf{2}} \\
&= \sum_{\mathbf{i,j=1}}^{\mathbf{p}}(\mathbf{AB}-\mathbf{BA})_{\mathbf{ij}}(\mathbf{AB}-\mathbf{BA})_{\mathbf{ji}} = -\sum_{\mathbf{i,j=1}}^{\mathbf{p}}(\mathbf{AB}-\mathbf{BA})^{\mathbf{2}}_{\mathbf{ij}} \le 0
\end{aligned}$$

Note the second equality in the last line holds by definition of skew-hermitian. Therefore, we have $\mathrm{tr}(\{\mathbf{AB}\}^{\mathbf{2}}) \leq \mathrm{tr}(\mathbf{A^2B^2})$. $\square$

Lemma 2.1. *If $\mathbf{p_{\Omega_k}}$ is the density of $\mathbf{N_p}(\mathbf{0}, \mathbf{\Omega_k^{-1}})$, $k = 1, 2$, then for all $\mathbf{\Omega_k} \in \mathcal{M}_0^+$, $k = 1, 2$, and $d_i, i = 1, \ldots, p$, eigenvalues of $\mathbf{A} = \mathbf{\Omega_1^{-1/2}\Omega_2\Omega_1^{-1/2}}$, we have that for sufficiently small $\delta > 0$ and constant $c_0 > 0$.*

(i) $c_0^{-1}||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_{\mathbf{2}}^{\mathbf{2}} \leq \sum_{i=1}^{p} |d_i - 1|^2 \leq c_0||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_{\mathbf{2}}^{\mathbf{2}}$

(ii)$\mathbf{h}(\mathbf{p_{\Omega_1}}, \mathbf{p_{\Omega_2}}) < \delta$ *implies* $\max_i |d_i - 1| < 1$ *and* $||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_{\mathbf{2}}^{\mathbf{2}} \leq c_0\mathbf{h^2}(\mathbf{p_{\Omega_1}}, \mathbf{p_{\Omega_2}})$,

(iii)$\mathbf{h^2}(\mathbf{p_{\Omega_1}}, \mathbf{p_{\Omega_2}}) \leq c_0||\mathbf{\Omega_1} - \mathbf{\Omega}||_{\mathbf{2}}^{\mathbf{2}}$.

Proof. By definition of $\mathcal{M}_0^+$, $||\mathbf{\Omega_1}||_{\mathbf{2}}$ and $||\mathbf{\Omega^{-1}}||_{\mathbf{2}}$ are bounded by some constants, say $\sqrt{c_0}$, where $c_0 \geq 1$. Let $\mathbf{P\Lambda P^t}$ be spectral decomposition of $\mathbf{A}$, where $\mathbf{\Lambda}$ is a diagonal matrix with d_i's on diagonal entries and $\mathbf{P}$ is an orthogonal matrix with corresponding eigenvectors on columns. Then as $\det(\mathbf{I_p} - \mathbf{A})) = \det(\mathbf{P}(\mathbf{I_p} - \mathbf{\Lambda})\mathbf{P^t}) = \det((\mathbf{I_p} - \mathbf{\Lambda PP^t})) = \det(\mathbf{I_p} - \mathbf{\Lambda}) = \prod_{\mathbf{i=1}}^{p}(1 - d_i)$, eigenvalues of $\mathbf{I_p} - \mathbf{A}$ are $1 - d_i$ for $i = 1, \ldots, p$.
Hence, we have

$$\begin{aligned}
||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_{\mathbf{2}}^{\mathbf{2}} &= ||\mathbf{\Omega_1^{1/2}}(\mathbf{I_p} - \mathbf{\Omega_1^{-1/2}\Omega_2\Omega_1^{-1/2}})\mathbf{\Omega_1^{1/2}}||_{\mathbf{2}}^{\mathbf{2}} = ||\mathbf{\Omega_1^{1/2}}(\mathbf{I_p} - \mathbf{A})\mathbf{\Omega_1^{1/2}}||_{\mathbf{2}}^{\mathbf{2}} \\
&= \mathrm{tr}(\{\mathbf{\Omega_1^{1/2}}(\mathbf{I_p} - \mathbf{A})\mathbf{\Omega_1^{1/2}}\}^{\mathbf{2}}) = \mathrm{tr}(\mathbf{\Omega_1^{1/2}}(\mathbf{I_p} - \mathbf{A})\mathbf{\Omega_1}(\mathbf{I_p} - \mathbf{A})\mathbf{\Omega_1^{1/2}}) \\
&= \mathrm{tr}(\mathbf{\Omega_1}(\mathbf{I_p} - \mathbf{A})\mathbf{\Omega_1}(\mathbf{I_p} - \mathbf{A})) = \mathrm{tr}(\{\mathbf{\Omega_1}(\mathbf{I_p} - \mathbf{A})\}^{\mathbf{2}}) \\
&\leq \mathrm{tr}(\mathbf{\Omega_1^2}(\mathbf{I_p} - \mathbf{A})^{\mathbf{2}}) \leq ||\mathbf{\Omega_1^2}||_{(\mathbf{2},\mathbf{2})}||\mathbf{I_p} - \mathbf{A}||_{\mathbf{2}}^{\mathbf{2}} \\
&= ||\mathbf{\Omega_1}||_{(\mathbf{2},\mathbf{2})}^{\mathbf{2}}||\mathbf{I_p} - \mathbf{A}||_{\mathbf{2}}^{\mathbf{2}} = ||\mathbf{\Omega_1}||_{(\mathbf{2},\mathbf{2})}^{\mathbf{2}}\mathrm{tr}(\mathbf{I_p} - \mathbf{A})^{\mathbf{2}} \\
&= ||\mathbf{\Omega_1}||_{(\mathbf{2},\mathbf{2})}^{\mathbf{2}}\mathrm{tr}(\mathbf{P}(\mathbf{I_p} - \mathbf{\Lambda})^{\mathbf{2}}\mathbf{P^t}) = ||\mathbf{\Omega_1}||_{(\mathbf{2},\mathbf{2})}^{\mathbf{2}}\mathrm{tr}(\mathbf{P^tP}(\mathbf{I_p} - \mathbf{\Lambda})^{\mathbf{2}}) \\
&= ||\mathbf{\Omega_1}||_{(\mathbf{2},\mathbf{2})}^{\mathbf{2}}\mathrm{tr}((\mathbf{I_p} - \mathbf{A})^{\mathbf{2}}) = ||\mathbf{\Omega_1}||_{(\mathbf{2},\mathbf{2})}^{\mathbf{2}}\sum_{\mathbf{i=1}}^{\mathbf{p}}|\mathbf{d_i} - \mathbf{1}|^{\mathbf{2}} \\
&\leq ||\mathbf{\Omega_1}||_{\mathbf{2}}^{\mathbf{2}}\sum_{\mathbf{i=1}}^{\mathbf{p}}|\mathbf{d_i} - \mathbf{1}|^{\mathbf{2}} \leq c_0\sum_{\mathbf{i=1}}^{\mathbf{p}}|\mathbf{d_i} - \mathbf{1}|^{\mathbf{2}}
\end{aligned}$$

.

Here, the first inequality in the fourth line holds by claim 2.1. and the last inequality in the same line holds by theorem 1.2. This establishes the inequality $c_0^{-1}||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_{\mathbf{2}}^{\mathbf{2}} \leq \sum_{\mathbf{i=1}}^{\mathbf{p}}|\mathbf{d_i} - \mathbf{1}|^{\mathbf{2}}$. Using similar techniques in the above, we can obtain

$$\sum_{\mathbf{i=1}}^{\mathbf{p}}|\mathbf{d_i} - \mathbf{1}|^{\mathbf{2}} = ||\mathbf{I_p} - \mathbf{A}||_{\mathbf{2}}^{\mathbf{2}} = ||\mathbf{\Omega_1^{-1/2}}(\mathbf{\Omega_1} - \mathbf{\Omega_2})\mathbf{\Omega_1^{-1/2}}||_{\mathbf{2}}^{\mathbf{2}} \leq ||\mathbf{\Omega_1^{-1}}||_{(\mathbf{2},\mathbf{2})}^{\mathbf{2}}||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_{\mathbf{2}}^{\mathbf{2}} \leq c_0||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_{\mathbf{2}}^{\mathbf{2}}.$$

which is the second inequality in (i). So we have established (i). To show (ii), we need to calculate $\mathbf{h^2}(\mathbf{p_{\Omega_1}}, \mathbf{p_{\Omega_2}})$ first.

$$\begin{aligned}
\mathbf{h^2}(\mathbf{p_{\Omega_1}}, \mathbf{p_{\Omega_2}}) &= \int_{\mathbb{R}^p}(\sqrt{\mathbf{p_{\Omega_1}(x)}} - \sqrt{\mathbf{p_{\Omega_2}(x)}})^{\mathbf{2}}\mathrm{d}\mathbf{x} = \int_{\mathbb{R}^p}\mathbf{p_{\Omega_1}(x)} + \mathbf{p_{\Omega_2}(x)} - \mathbf{2}\sqrt{\mathbf{p_{\Omega_1}(x)p_{\Omega_2}(x)}}\mathrm{d}\mathbf{x} \\
&= 2 - 2\int_{\mathbb{R}^p}\sqrt{\mathbf{p_{\Omega_1}(x)p_{\Omega_2}(x)}}\mathrm{d}\mathbf{x}
\end{aligned}$$

$$
\begin{aligned}
\int_{\mathbb{R}^p}\sqrt{\mathbf{p}_{\boldsymbol{\Omega}_1}(\mathbf{x})\mathbf{p}_{\boldsymbol{\Omega}_2}(\mathbf{x})}d\mathbf{x} &= \int_{\mathbb{R}^p}\sqrt{\frac{\mathbf{1}}{|\mathbf{2}\pi\boldsymbol{\Omega}_1^{-1}|^{1/2}|\mathbf{2}\pi\boldsymbol{\Omega}_2^{-1}|^{1/2}}\exp\left(-\frac{\mathbf{1}}{\mathbf{2}}\mathbf{x^t}(\boldsymbol{\Omega}_1+\boldsymbol{\Omega}_2)\mathbf{x}\right)}\mathrm{d}\mathbf{x}\\
&= \int_{\mathbb{R}^p}\frac{\mathbf{1}}{|\mathbf{2}\pi\boldsymbol{\Omega}_1^{-1}|^{1/4}|\mathbf{2}\pi\boldsymbol{\Omega}_2^{-1}|^{1/4}}\exp\left(-\frac{\mathbf{1}}{\mathbf{2}}\mathbf{x^t}\left(\frac{\boldsymbol{\Omega}_1+\boldsymbol{\Omega}_2}{\mathbf{2}}\right)\mathbf{x}\right)\mathrm{d}\mathbf{x}\\
&= \frac{|\mathbf{2}\pi(\frac{\boldsymbol{\Omega}_1+\boldsymbol{\Omega}_2}{\mathbf{2}})^{-1}|^{1/2}}{|\mathbf{2}\pi\boldsymbol{\Omega}_1^{-1}|^{1/4}|\mathbf{2}\pi\boldsymbol{\Omega}_2^{-1}|^{1/4}} = \frac{|(\frac{\boldsymbol{\Omega}_1+\boldsymbol{\Omega}_2}{\mathbf{2}})^{-1}|^{1/2}}{|\boldsymbol{\Omega}_1^{-1}|^{1/4}|\boldsymbol{\Omega}_2^{-1}|^{1/4}}\\
&= \frac{|(\frac{\boldsymbol{\Omega}_1^{1/2}(\mathbf{I_p}+\mathbf{A})\boldsymbol{\Omega}_1^{1/2}}{\mathbf{2}})^{-1}|^{1/2}}{|\boldsymbol{\Omega}_1^{-1}|^{1/4}|\boldsymbol{\Omega}_2^{-1}|^{1/4}} = \frac{|\boldsymbol{\Omega}_1^{-1/2}(\frac{(\mathbf{I_p}+\mathbf{A})}{\mathbf{2}})^{-1}\boldsymbol{\Omega}_1^{-1/2}|^{1/2}}{|\boldsymbol{\Omega}_1^{-1}|^{1/4}|\boldsymbol{\Omega}_2^{-1}|^{1/4}}\\
&= \frac{|\boldsymbol{\Omega}_1^{-1/2}|^{1/2}|(\frac{(\mathbf{I_p}+\mathbf{A})}{\mathbf{2}})^{-1}|^{1/2}|\boldsymbol{\Omega}_1^{-1/2}|^{1/2}}{|\boldsymbol{\Omega}_1^{-1}|^{1/4}|\boldsymbol{\Omega}_2^{-1}|^{1/4}}\\
&= \frac{|\boldsymbol{\Omega}_1^{-1}|^{1/4}|(\frac{(\mathbf{I_p}+\mathbf{A})}{\mathbf{2}})^{-1}|^{1/2}|\boldsymbol{\Omega}_1^{-1/2}|^{1/2}}{|\boldsymbol{\Omega}_1^{-1}|^{1/4}|\boldsymbol{\Omega}_2^{-1}|^{1/4}}\\
&= |(\frac{\mathbf{I_p}+\mathbf{A}}{\mathbf{2}})^{-1}|^{1/2}||\boldsymbol{\Omega}_1^{-1/2}|^{1/4}|\boldsymbol{\Omega}_2|^{1/4}|\boldsymbol{\Omega}_1^{-1/2}|^{1/4}\\
&= |(\frac{\mathbf{I_p}+\mathbf{A}}{\mathbf{2}})^{-1}|^{1/2}|\boldsymbol{\Omega}_1^{-1/2}\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1^{-1/2}|^{1/4} = |(\frac{\mathbf{I_p}+\mathbf{A}}{\mathbf{2}})^{-1}|^{1/2}|\mathbf{A}|^{1/4}\\
&= |(\frac{\mathbf{I_p}+\mathbf{A}}{\mathbf{2}})^{-1}|^{1/2}|\mathbf{A}^{1/2}|^{1/2} = |(\frac{\mathbf{I_p}+\mathbf{A}}{\mathbf{2}})^{-1}\mathbf{A}^{1/2}|^{1/2}\\
&= |(\frac{\mathbf{A}^{1/2}+\mathbf{A}^{-1/2}}{\mathbf{2}})^{-1}|^{1/2} = |\frac{\mathbf{A}^{1/2}+\mathbf{A}^{-1/2}}{\mathbf{2}}|^{-1/2}\\
&= |\frac{\mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P^t}+\mathbf{P}\boldsymbol{\Lambda}^{-1/2}\mathbf{P^t}}{\mathbf{2}}|^{-1/2} = |\frac{\mathbf{P}(\boldsymbol{\Lambda}^{1/2}+\boldsymbol{\Lambda}^{-1/2})\mathbf{P^t}}{\mathbf{2}}|^{-1/2}\\
&= |\mathbf{P^t}\mathbf{P}(\frac{\boldsymbol{\Lambda}^{1/2}+\boldsymbol{\Lambda}^{-1/2})}{\mathbf{2}}|^{-1/2} = |\frac{\boldsymbol{\Lambda}^{1/2}+\boldsymbol{\Lambda}^{-1/2}}{\mathbf{2}}|^{-1/2}\\
&= \{\prod_{i=1}^{p}\frac{1}{2}(d_i^{1/2}+d_i^{-1/2})\}^{-1/2}
\end{aligned}
$$

Therefore, we have

$$
\mathbf{h^2}(\mathbf{p}_{\boldsymbol{\Omega}_1},\mathbf{p}_{\boldsymbol{\Omega}_2}) = 2 - 2\{\prod_{i=1}^{p}\frac{1}{2}(d_i^{1/2}+d_i^{-1/2})\}^{-1/2} \tag{10}
$$

,or equivalently,

$$
\frac{1}{2}\mathbf{h^2}(\mathbf{p}_{\boldsymbol{\Omega}_1},\mathbf{p}_{\boldsymbol{\Omega}_2}) = 1 - \{\prod_{i=1}^{p}\frac{1}{2}(d_i^{1/2}+d_i^{-1/2})\}^{-1/2} \tag{11}
$$

Note that by arithmetic-geometric mean inequality $1 = 2\frac{1}{2}\sqrt{d_i^{1/2}d_i^{-1/2}} \le \frac{1}{2}(d_i^{1/2}+d_i^{-1/2})$. Denote $\mathbf{h}(\mathbf{p}_{\boldsymbol{\Omega}_1},\mathbf{p}_{\boldsymbol{\Omega}_2})$ by $\mathbf{h}$ for convenience. Suppose $\mathbf{h}(\mathbf{p}_{\boldsymbol{\Omega}_1},\mathbf{p}_{\boldsymbol{\Omega}_2}) < \delta$. Then, by (11), $1 - \{\prod_{i=1}^{p}\frac{1}{2}(d_i^{1/2}+d_i^{-1/2})\}^{-1/2} \le \delta^2/2$,$1-\delta^2/2 \le \{\prod_{i=1}^{p}\frac{1}{2}(d_i^{1/2}+d_i^{-1/2})\}^{-1/2}$, or $\prod_{i=1}^{p}\frac{1}{2}(d_i^{1/2}+d_i^{-1/2}) \le (1-\delta^2/2)^{-2} = 1+\eta$ for some $\eta \ge 0$.

Since each term in the product is not less than 1, we have $\max_i \frac{1}{2}(d_i^{1/2}+d_i^{-1/2}) \le 1+\eta$ or equivalently $\frac{1}{2}(d_i^{1/2}+d_i^{-1/2}) \le 1+\eta$. Let $\alpha = d_i^{1/2} > 0$. Then we have $\frac{1}{2}(\alpha+\alpha^{-1}) \le 1+\eta, \alpha+\alpha^{-1} \le 2(1+\eta), \alpha^2 - 2(1+\eta)\alpha+1 \le 0$. Solving this inequality, we obtain $1+\eta-\sqrt{\eta^2+2\eta} \le \alpha = d_i^{1/2} \le 1+\eta+\sqrt{\eta^2+2\eta}$. Note that the lower bound is positive. Upon squaring this inequality, $2(\eta^2+2\eta)-2(1+\eta)\sqrt{\eta^2+2\eta}+1 \le d_i \le 2(\eta^2+2\eta)+2(1+\eta)\sqrt{\eta^2+2\eta}+1$, $\beta_1 \le d_i - 1 \le \beta_2$, where $\beta_i = 2(\eta^2+2\eta)+(-1)^i2(1+\eta)\sqrt{\eta^2+2\eta}, i = 1,2$. Note $\beta_i \to 0$, so $\beta_i^2 \to 0$ as $\eta \to 0$ for i=1,2. Hence $(d_i-1)^2 \le \max\{\beta_1^2,\beta_2^2\}$.Taking sufficiently small $\delta > 0$, η

can be made sufficiently small so that $\max\{\beta_1^2, \beta_2^2\} < 1$. Therefore, with such δ, $(d_i - 1)^2 < 1$ or $|d_i - 1| < 1$ for all $i = 1, 2, \ldots, p$. So $\max_i |d_i - 1| < 1$.

Rearranging terms in equation (11), $1 - \frac{1}{2}\mathbf{h^2} = \{\prod_{i=1}^{p} \frac{1}{2}(d_i^{1/2} + d_i^{-1/2})\}^{-1/2}$, $(1 - \frac{1}{2}\mathbf{h^2})^{-2} = \prod_{i=1}^{p} \frac{1}{2}(d_i^{1/2} + d_i^{-1/2})$, $\prod_{i=1}^{p}(d_i^{1/2} + d_i^{-1/2}) = 2^p(1 - \frac{1}{2}\mathbf{h^2})^{-2}$, $2^{-p}\prod_{i=1}^{p}(d_i^{1/2} + d_i^{-1/2}) = (1 - \frac{1}{2}\mathbf{h^2})^{-2}$. By Taylor's expansion, $(1-\frac{1}{2}\mathbf{h^2})^{-2} = 1+\mathbf{h^2}+o(h^3)$. If $\delta > 0$ is chosen sufficiently small,$\max_i |d_i-1| < 1$ for some constants $c_1, c_2 > 0$ and so $[1 + c_1 \sum_{i=1}^{p}(d_i - 1)^2] \leq 2^{-p}\prod_{i=1}^{p}(d_i^{1/2} + d_i^{-1/2}) \leq [1 + c_2 \sum_{i-1}^{p}(d_i - 1)^2]$. Thus, $1 + c_1 \sum_{i=1}^{p}(d_i - 1)^2 \leq 2^{-p}\prod_{i=1}^{p}(d_i^{1/2} + d_i^{-1/2}) \sim 1 + \mathbf{h^2}$ and so $c_1 \sum_{i=1}^{p}(d_i - 1)^2 \leq \mathbf{h^2}$. $||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_\mathbf{2}^\mathbf{2} = ||\mathbf{\Omega_1}||_{\mathbf{(2,2)}}^\mathbf{2} \sum_{\mathbf{i=1}}^{\mathbf{p}}(\mathbf{d_i} - \mathbf{1})^\mathbf{2} \leq \frac{||\mathbf{\Omega_1}||_{\mathbf{(2,2)}}^\mathbf{2}}{c_1}\mathbf{h^2}$. Now set $c_0 = \frac{||\mathbf{\Omega_1}||_{\mathbf{(2,2)}}^\mathbf{2}}{c_1}$. Thus $||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_\mathbf{2}^\mathbf{2} \leq c_0\mathbf{h^2}$, which proves (ii).

To prove (iii), note that in equation (10), one can clearly see that $\mathbf{h^2} \leq 2$. So Hellinger distance is bounded. Therefore it suffices to consider the case only when $||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_\mathbf{2} < \delta$ where $\delta > 0$ is chosen to be sufficiently small. By the second inequality in (i), one can see that $\sum_{i=1}^{p} |d_i - 1|^2 \leq 1$, if $||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_\mathbf{2}$ is sufficiently small. Thus $\max_i |d_i - 1| < 1$ and so similar to proof of (ii)' with Taylor's expansion, $^\exists c_2 > 0 \text{s.t.} 1 + \mathbf{h^2} \sim (1 - \frac{1}{2}\mathbf{h^2})^{-2=}2^{-p}\prod_{i=1}^{p}(d_i^{1/2} + d_i^{-1/2}) \leq 1 + c_2\sum_{i=1}^{p}(d_i - 1)^2$. Consequently,

$$\mathbf{h^2} \leq c_2 \sum_{i=1}^{p} |d_i - 1|^2 \leq c_2c_0||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_\mathbf{2}^\mathbf{2}$$

Here c_0 is constant chosen in (i). Set $\mathrm{C} = c_2c_0$ and we have $\mathbf{h^2} \leq \mathrm{C}||\mathbf{\Omega_1} - \mathbf{\Omega_2}||_\mathbf{2}^\mathbf{2}$, which establishes (iii). This ends the proof of Lemma 2.1. $\square$

With results of Lemma 2.1., we may finally obtain posterior convergence rate as $n \to \infty$. Note that we assumed true precision matrix $\mathbf{\Omega_0}$ to be sparse and its eigenvalues are bounded away from 0. Also, recall that for two probability densities p_0, p_1, Kullback-Leibler divergence is given by $\mathrm{K}(p_0, p_1) = \int p_0 \log(\frac{p_0}{p_1})$, which we may refer to as 'KL divergence'. Let $\mathrm{V}(p_0, p_1) = \int p_0 \log^2(\frac{p_0}{p_1})$.

Claim 2.2. *Suppose p_0, p_1 are two probability densities. Then $K(p_0, p_1) \geq \mathbf{h^2}(p_0, p_1)$.*

Proof. For $^\forall x < 1$, we know that $-\log(1 - x) \geq x$. So,

$$\begin{aligned}\mathrm{K}(p_0, p_1) &= \int p_0 \log(\frac{p_0}{p_1}) \geq \int -2p_0 \log(\sqrt{\frac{p_1}{p_0}}) = \int -2p_0 \log(1 - (1 - \sqrt{\frac{p_1}{p_0}}))\\ &= \int 2p_0(1 - \sqrt{\frac{p_1}{p_0}}) = 2(\int p_0 - \int \sqrt{p_0p_1}) = 2 - 2\int \sqrt{p_0p_1} = \mathbf{h^2}(p_0, p_1).\end{aligned}$$

$\square$

Claim 2.3. *Let $p_{\mathbf{\Omega}_i}$ be density of $\mathbf{N_p}(\mathbf{0}, \mathbf{\Omega_i^{-1}})$, $i = 0, 1$ and $d_i, i = 1, 2, \ldots p$ be eigenvalues of $\mathbf{\Omega_0^{-1/2}\Omega_1\Omega_0^{-1/2}}$. Then followings hold:*

$$\begin{aligned}K(p_{\mathbf{\Omega}_0}, p_{\mathbf{\Omega}_1}) &= -\frac{1}{2}\sum_{i=1}^{p} \log d_i - \frac{1}{2}\sum_{i=1}^{p}(1 - d_i)\\ V(p_{\mathbf{\Omega}_0}, p_{\mathbf{\Omega}_1}) &= \frac{1}{4}(\sum_{i=1}^{p}(\log d_i + 1 - d_i))^2 + \frac{1}{2}\sum_{i=1}^{p}(1 - d_i)^2\end{aligned} \tag{12}$$

Proof. Suppose $\mathbf{Z} \sim \mathbf{N_p}(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{A} \in \mathcal{M}$. Assume that $\mathbf{\Sigma} = \mathbf{I_p}$. Let $\mathbf{P\Lambda P^t}$ be spectral decomposition of $\mathbf{A}$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ is a diagonal matrix with eigenvalues λ_i's on diagonal entries and $\mathbf{P} = (\mathbf{P_1}|\cdots|\mathbf{P_p})$ is an orthogonal matrix with corresponding eigenvectors as columns. We know that $\mathbf{Cov}(\mathbf{P_i^t Z}, \mathbf{P_j^t Z}) = \mathbf{P_i^t Var(Z) P_j} = \mathbf{P_i^t P_j} = \mathbf{0}$, which implies $\mathbf{P_i^t Z}, \mathbf{P_j^t Z}$ are independent, for $i \neq j$. Also as $\mathbf{P_i^t Z} \sim \mathbf{N}(\mathbf{0}, \mathbf{P_i^t P_i}) = \mathbf{N}(\mathbf{0}, \mathbf{1})$, so $(\mathbf{P_i^t Z})^\mathbf{2} \sim \chi^2(1)$. Further recall that variance of chi-square distribution with df=1 is 2. Then we have

$$\mathbf{E}(\mathbf{Z^t A Z}) = \mathbf{E}(\mathbf{tr}(\mathbf{Z^t A Z})) = \mathbf{E}(\mathbf{tr}(\mathbf{AZZ^t})) = \mathbf{tr}(\mathbf{E}(\mathbf{AZZ^t})) = \mathbf{tr}(\mathbf{AE}(\mathbf{ZZ^t})) = \mathbf{tr}(\mathbf{A Var(Z)}) = \mathbf{tr}(\mathbf{AI_p}) = \mathbf{tr}(\mathbf{A})$$

$$\begin{aligned}\mathbf{Var}(\mathbf{Z^t A Z}) &= \mathbf{Var}((\mathbf{P^t Z})^\mathbf{t} \mathbf{\Lambda} (\mathbf{P^t Z})) = \mathbf{Var}(\sum_{i=1}^p \lambda_i (\mathbf{P_i^t Z})^2) = \sum_{i=1}^p \lambda_i^2 \mathbf{Var}(\mathbf{P_i^t Z}) = 2\sum_{i=1}^p \lambda_i^2 = 2\mathbf{tr}(\mathbf{\Lambda^2}) = 2\mathbf{tr}(\mathbf{P^t P \Lambda^2}) \\ &= 2\mathbf{tr}(\mathbf{P\Lambda^2 P^t}) = 2\mathbf{tr}(\mathbf{P\Lambda P^t P \Lambda P^t}) = 2\mathbf{tr}(\mathbf{AA}).\end{aligned}$$

For general $\mathbf{\Sigma}$, consider that $\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{Z} \sim \mathbf{N_p}(\mathbf{0}, \mathbf{I_p})$. From the results above, we may obtain

$$\begin{aligned}\mathbf{E}(\mathbf{Z^t A Z}) &= \mathbf{E}((\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{Z})^\mathbf{t} \mathbf{\Sigma}^{\frac{1}{2}} \mathbf{A} \mathbf{\Sigma}^{\frac{1}{2}} (\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{Z})) = \mathbf{tr}(\mathbf{\Sigma}^{\frac{1}{2}} \mathbf{A} \mathbf{\Sigma}^{\frac{1}{2}}) = \mathbf{tr}(\mathbf{A\Sigma}) \\ \mathbf{Var}(\mathbf{Z^t A Z}) &= \mathbf{Var}((\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{Z})^\mathbf{t} \mathbf{\Sigma}^{\frac{1}{2}} \mathbf{A} \mathbf{\Sigma}^{\frac{1}{2}} (\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{Z})) = 2\mathbf{tr}(\mathbf{\Sigma}^{\frac{1}{2}} \mathbf{A} \mathbf{\Sigma}^{\frac{1}{2}} \mathbf{\Sigma}^{\frac{1}{2}} \mathbf{A} \mathbf{\Sigma}^{\frac{1}{2}}) \\ &= 2\mathbf{tr}(\mathbf{\Sigma}^{\frac{1}{2}} \mathbf{A\Sigma A} \mathbf{\Sigma}^{\frac{1}{2}}) = 2\mathbf{tr}(\mathbf{A\Sigma A\Sigma})\end{aligned}$$

Using theses facts,

$$\begin{aligned}\mathrm{K}(p_{\mathbf{\Omega}_0}, p_{\mathbf{\Omega}_1}) &= \int_{\mathbb{R}^p} p_{\mathbf{\Omega}_0}(\mathbf{x}) \log(\frac{p_{\mathbf{\Omega}_0}(\mathbf{x})}{p_{\mathbf{\Omega}_1}(\mathbf{x})})\mathrm{d}\mathbf{x} = \int_{\mathbb{R}^p} p_{\mathbf{\Omega}_0}(\mathbf{x}) \log(\frac{|2\pi\mathbf{\Omega_1^{-1}}|^{\frac{1}{2}}}{|2\pi\mathbf{\Omega_0}^{-1}|^{\frac{1}{2}}} \exp(-\frac{1}{2}\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x})\mathrm{d}\mathbf{x} \\ &= \int_{\mathbb{R}^p} p_{\mathbf{\Omega}_0}(\mathbf{x})(\log(\frac{|\mathbf{\Omega_0}|^{\frac{1}{2}}}{|\mathbf{\Omega_1}|^{\frac{1}{2}}}) - \frac{1}{2}\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x})\mathrm{d}\mathbf{x} \\ &= \int_{\mathbb{R}^p} \log(\frac{|\mathbf{\Omega_0}|^{\frac{1}{2}}}{|\mathbf{\Omega_1}|^{\frac{1}{2}}}) p_{\mathbf{\Omega_0}}(\mathbf{x})\mathrm{d}\mathbf{x} - \frac{1}{2}\int_{\mathbb{R}^p} \mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x} p_{\mathbf{\Omega_0}}(\mathbf{x})\mathrm{d}\mathbf{x} \\ &= \log(\frac{|\mathbf{\Omega_0}|^{\frac{1}{2}}}{|\mathbf{\Omega_1}|^{\frac{1}{2}}}) - \frac{1}{2}\mathbf{E_0}(\mathbf{X^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{X}) = -\frac{1}{2}\log(\frac{|\mathbf{\Omega_1}|}{|\mathbf{\Omega_0}|}) - \frac{1}{2}\mathbf{tr}((\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{\Omega_0^{-1}}) \\ &= -\frac{1}{2}\log(|\mathbf{\Omega_0}|^{-\frac{1}{2}}|\mathbf{\Omega_1}||\mathbf{\Omega_0}|^{-\frac{1}{2}}) - \frac{1}{2}\mathbf{tr}(\mathbf{\Omega_0}^{-\frac{1}{2}}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{\Omega_0}^{-\frac{1}{2}}) \\ &= -\frac{1}{2}\log(|\mathbf{\Omega_0}^{-\frac{1}{2}}||\mathbf{\Omega_1}||\mathbf{\Omega_0}^{-\frac{1}{2}}|) - \frac{1}{2}\mathbf{tr}(\mathbf{I_p} - \mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}}) \\ &= -\frac{1}{2}\log(|\mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}}|) - \frac{1}{2}\mathbf{tr}(\mathbf{I_p} - \mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}})\end{aligned}$$

Let $\mathbf{QDQ^t}$ be spectral decomposition of $\mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}}$, where $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_p)$ and $\mathbf{Q}$ is an orthogonal matrix with corresponding eigenvectors on columns. Then, $\mathbf{I_p} - \mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}} = \mathbf{Q}(\mathbf{I_p} - \mathbf{D})\mathbf{Q^t}$, hence eigenvalues of $\mathbf{I_p} - \mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}}$ are $1 - d_i$, $i = 1, 2, \ldots, p$. As a consequence, we have

$$\mathrm{K}(p_{\mathbf{\Omega_0}}, p_{\mathbf{\Omega_1}}) = -\frac{1}{2}\log(\prod_{i=1}^p d_i) - \frac{1}{2}\sum_{i=1}^p (1 - d_i) = -\frac{1}{2}\sum_{i=1}^p \log d_i - \frac{1}{2}\sum_{i=1}^p (1 - d_i)$$

This proves the first equation in (12). For the second equation,

$$\begin{aligned}\mathrm{V}(p_{\mathbf{\Omega_0}}, p_{\mathbf{\Omega_1}}) &= \int_{\mathbb{R}^p} p_{\mathbf{\Omega}_0}(\mathbf{x}) \log^2(\frac{p_{\mathbf{\Omega}_0}(\mathbf{x})}{p_{\mathbf{\Omega}_1}(\mathbf{x})})\mathrm{d}\mathbf{x} = \int_{\mathbb{R}^p} p_{\mathbf{\Omega_0}}(\mathbf{x})(-\frac{1}{2}\log(|\mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}}|) = \frac{1}{2}\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x})^2\mathrm{d}\mathbf{x} \\ &= \frac{1}{4}\int_{\mathbb{R}^p} p_{\mathbf{\Omega_0}}(\mathbf{x})((\log(|\mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}}|))^2 + 2\log(|\mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}}|)(\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x}) + (\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x})^2)\mathrm{d}\mathbf{x} \\ &= \frac{1}{4}((\log(|\mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}}|))^2 + 2\log(|\mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}}|)\mathbf{E_0}(\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x}) + \mathbf{E_0}((\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x})^\mathbf{2})) \\ &= \frac{1}{4}((\log(|\mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}}|))^2 + 2\log(|\mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega_1}\mathbf{\Omega_0}^{-\frac{1}{2}}|)\mathbf{E_0}(\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x}) + (\mathbf{E_0}(\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x}))^\mathbf{2} \\ &\quad + \mathbf{Var_0}(\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x}))\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{4}((\log(|\mathbf{\Omega_0^{-\frac{1}{2}}\Omega_1\Omega_0^{-\frac{1}{2}}}|) + \mathbf{E_0}(\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x}))^2 + \mathbf{Var_0}(\mathbf{x^t}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{x}))\\
&= \frac{1}{4}((\sum_{i=1}^{p}(\log d_i - 1 + d_i))^2 + 2\mathbf{tr}((\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{\Omega_0^{-1}}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{\Omega_0^{-1}}))\\
&= \frac{1}{4}((\sum_{i=1}^{p}(\log d_i - 1 + d_i))^2 + 2\mathbf{tr}(\mathbf{\Omega_0^{-\frac{1}{2}}}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{\Omega_0^{-\frac{1}{2}}}\mathbf{\Omega_0^{-\frac{1}{2}}}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{\Omega_0^{-\frac{1}{2}}}))\\
&= \frac{1}{4}((\sum_{i=1}^{p}(\log d_i - 1 + d_i))^2 + 2\mathbf{tr}([\mathbf{\Omega_0^{-\frac{1}{2}}}(\mathbf{\Omega_0} - \mathbf{\Omega_1})\mathbf{\Omega_0^{-\frac{1}{2}}}]^{\mathbf{2}})) = \frac{1}{4}((\sum_{i=1}^{p}(\log d_i - 1 + d_i))^2 + 2\mathbf{tr}([\mathbf{I_p} - \mathbf{\Omega_0^{-\frac{1}{2}}\Omega_1\Omega_0^{-\frac{1}{2}}}]^{\mathbf{2}}))\\
&= \frac{1}{4}(\sum_{i=1}^{p}(\log d_i - 1 + d_i))^2 + \frac{1}{2}\mathbf{tr}([\mathbf{Q}(\mathbf{I_p} - \mathbf{D})\mathbf{Q^t}]^{\mathbf{2}}) = \frac{1}{4}(\sum_{i=1}^{p}(\log d_i - 1 + d_i))^2 + \frac{1}{2}\mathbf{tr}(\mathbf{Q}(\mathbf{I_p} - \mathbf{D})^{\mathbf{2}}\mathbf{Q^t})\\
&= \frac{1}{4}(\sum_{i=1}^{p}(\log d_i - 1 + d_i))^2 + \frac{1}{2}\mathbf{tr}(\mathbf{Q^t Q}(\mathbf{I_p} - \mathbf{D})^{\mathbf{2}}) = \frac{1}{4}(\sum_{i=1}^{p}(\log d_i - 1 + d_i))^2 + \frac{1}{2}\mathbf{tr}((\mathbf{I_p} - \mathbf{D})^{\mathbf{2}})\\
&= \frac{1}{4}(\sum_{i=1}^{p}(\log d_i - 1 + d_i))^2 + \frac{1}{2}\sum_{i=1}^{p}(1 - d_i)^2
\end{aligned}
$$

□

Claim 2.4. *Let $p \in \mathbb{N}$ and choose $\bar{r} \in \mathbb{N}$ such that $\bar{r} < \binom{p}{2}/2$. If $j \leq \bar{r}$,*

$$\binom{\binom{p}{2}}{j} \leq \binom{p + \binom{p}{2}}{\bar{r}}.$$

Proof. Suppose $n \in \mathbb{N}$. $k = \underset{i=1,2,\ldots,n}{\arg\max} \binom{n}{i}$ can be found by finding the maximum k such that $\binom{n}{k-1} \leq \binom{n}{k}$.

$$\binom{n}{k-1} \leq \binom{n}{k} \Leftrightarrow \frac{n!}{(k-1)!(n-k+1)!} \leq \frac{n!}{k!(n-k)!} \Leftrightarrow k \leq n - k + 1 \Leftrightarrow k \leq \frac{n+1}{2}$$

Hence, naively one can see that the desired k is $\frac{n+1}{2}$. Also, one can deduce that $\binom{n}{k}$ is monotone increasing sequence provided that $k \leq \frac{n+1}{2}$...(#)

Take $\alpha \in \mathbb{N}$ such that $\alpha \leq \min\{p, \bar{r} - j\}$. Suppose there a box with p red balls and $\binom{p}{2}$ blue balls and these balls are numbered from 1 to $p + \binom{p}{2}$. Assume one chooses $j + \alpha$ balls at once in the box. Then there are possible $\binom{p+\binom{p}{2}}{j+\alpha}$ outcomes. If we suppose further that among $j + \alpha$ balls, j balls are blue balls and the rest of them are red balls, there are possible $\binom{\binom{p}{2}}{j}\binom{p}{\alpha}$ outcomes. Because of additional condition, one can clearly see that $\binom{\binom{p}{2}}{j}\binom{p}{\alpha} \leq \binom{p+\binom{p}{2}}{j+\alpha}$. By our choice of α, $j + \alpha \leq \bar{r}$. Also we chose $\bar{r}$ to be smaller than $\binom{p}{2}/2 < (p + \binom{p}{2})/2$. Therefore, using the fact #,

$$\binom{\binom{p}{2}}{j} \leq \binom{\binom{p}{2}}{j}\binom{p}{\alpha} \leq \binom{p + \binom{p}{2}}{j + \alpha} \leq \binom{p + \binom{p}{2}}{\bar{r}}$$

□

Theorem 2.2. *Let $\mathbf{X^{(n)}} = (\mathbf{X_1}, \ldots, \mathbf{X_n}) \sim \mathbf{N_p}(\mathbf{0}, \mathbf{\Omega_0^{-1}}$, where $\mathbf{\Omega_0} \in \mathcal{U}(\epsilon_0, s)$ for some $0 < \epsilon_0 < \infty$ and $0 \leq s \leq p(p-1)/2$. Also assume that priors are given in (1) or (2) with $q < 1/2$ and the range of eigenvalues of matrices in $\mathcal{M}_0^+$ is sufficiently broad to contain $[\epsilon_0, \epsilon_0^{-1}]$. Then the posterior distribution of $\mathbf{\Omega}$ satisfies*

$$\mathbf{E_0}[\mathbf{P}\{||\mathbf{\Omega} - \mathbf{\Omega_0}||_\mathbf{2} > \mathbf{M}\epsilon_\mathbf{n}|\mathbf{X^{(n)}}\}] \to 0, \tag{13}$$

for $\epsilon_n = n^{-1/2}(p + s)^{1/2}(\log n)^{1/2}$ and a sufficiently large $\mathbf{M} > 0$.

Proof. Take sufficiently large n so that ϵ_n is sufficiently small. Let $\mathrm{B}(p_{\mathbf{\Omega_0}}, \epsilon_n) = \{p_{\mathbf{\Omega}} : \mathrm{K}(p_{\mathbf{\Omega_0}}, p_{\mathbf{\Omega}}) \leq \epsilon_n^2,\ \mathrm{V}(p_{\mathbf{\Omega_0}}, p_{\mathbf{\Omega}}) \leq \epsilon_n^2\}$, which is Kullback-Leibler neighborhood of $p_{\mathbf{\Omega_0}}$ with distance ϵ_n. Assume $p_{\mathbf{\Omega}} \in \mathrm{B}(p_{\mathbf{\Omega_0}}, \epsilon_n)$. By claim 2.2., $\mathrm{K}(p_{\mathbf{\Omega_0}}, p_{\mathbf{\Omega}}) \geq \mathbf{h^2}(p_{\mathbf{\Omega_0}}, p_{\mathbf{\Omega}})$. Then $\mathbf{h}(\mathbf{p_{\Omega_0}}, \mathbf{p_{\Omega}}) \leq \epsilon_n$. Since ϵ_n is sufficiently small, by lemma 2.1. (ii), we have $\max_i |d_i - 1| < 1$.

Let $d_i, i = 1, 2, \ldots, p$ be eigenvalues of $\mathbf{\Omega_0}^{-\frac{1}{2}}\mathbf{\Omega}\mathbf{\Omega_0}^{-\frac{1}{2}}$. By the result of claim 2.3., $\mathrm{K}(p_{\mathbf{\Omega_0}}, p_{\mathbf{\Omega}}) = -\frac{1}{2}\sum_{i=1}^{p} \log d_i - \frac{1}{2}\sum_{i=1}^{p}(1 - d_i)$. By Taylor's expansion $\log d_i = \log(1 - (1 - d_i)) \sim -(1 - d_i) - \frac{1}{2}(1 - d_i)^2$. Note that this approximation is vaild only when $|d_i - 1| < 1$. Hence $\mathrm{K}(p_{\mathbf{\Omega_0}}, p_{\mathbf{\Omega}}) = -\frac{1}{2}\sum_{i=1}^{p}(\log d_i + 1 - d_i) \sim -\frac{1}{2}\sum_{i=1}^{p}(-(1 - d_i) - \frac{1}{2}(1 - d_i)^2 + 1 - d_i) = \frac{1}{4}\sum_{i=1}^{p}(1 - d_i)^2$. Again, by lemma 2.1. (i), $^{\exists}c > 0$ s.t. $\sum_{i=1}^{p}(1 - d_i)^2 \leq 4c^{-2}||\mathbf{\Omega} - \mathbf{\Omega_0}||_{\mathbf{2}}^{\mathbf{2}}$. By the result of theorem 1.2., $||\mathbf{\Omega} - \mathbf{\Omega_0}||_{\mathbf{2}}^{\mathbf{2}} \leq \mathbf{p^2}||\mathbf{\Omega} - \mathbf{\Omega_0}||_{\infty}^{\mathbf{2}}$ and so

$$\mathrm{K}(p_{\mathbf{\Omega_0}}, p_{\mathbf{\Omega}}) \sim \frac{1}{4}\sum_{i=1}^{p}(1 - d_i)^2 \leq c^{-2}\mathbf{p^2}||\mathbf{\Omega} - \mathbf{\Omega_0}||_{\infty}^{\mathbf{2}}$$

Thus, if $c^{-2}\mathbf{p^2}||\mathbf{\Omega} - \mathbf{\Omega_0}||_{\infty}^{\mathbf{2}} \leq \epsilon_n^2$, then the above approximation of $\mathrm{K}(p_{\mathbf{\Omega_0}}, p_{\mathbf{\Omega}})$ is valid and $\mathcal{P} = \{p_{\mathbf{\Omega}} : ||\mathbf{\Omega} - \mathbf{\Omega_0}||_{\infty} \leq c\epsilon_n/p\} \subset \mathrm{B}(p_{\mathbf{\Omega_0}}, \epsilon_n)$. We may estimate prior concentration of $\mathrm{B}(p_{\mathbf{\Omega_0}}, \epsilon_n)$. To do this, it suffices to get a lower estimate of the prior probability of $\mathcal{P}$. Because the prior for $\mathbf{\Omega}$ is truncated to $\mathcal{M}_0^+$, the components of $\mathbf{\Omega}$ are not independent. However, a small neighborhood of $\mathbf{\Omega_0} \in \mathcal{U}(\epsilon_0, s)$ lies within $\mathcal{M}_0^+$, which was the assumption. Hence, the truncation can only increase prior concentration of $\mathrm{B}(p_{\mathbf{\Omega_0}}, \epsilon_n)$, as the prior probability of $\mathcal{P}$ tends to increase. So in order to obtain lower bound for the prior probability of $\mathcal{P}$, we may assume independence on components of $\mathbf{\Omega}$. Let $\bar{\mathbf{E}} = \mathbf{E} \cup \{(i, i) : i = 1, 2, \ldots, p\}$, where $\mathbf{E}$ is edge set corresponding to $\mathbf{\Omega}$. Also, let $\mathbf{\Omega} = ((\omega_{ij}))$ and $\mathbf{\Omega_0} = ((\omega_{ij}^0))$. Naively, consider the case only when $(\omega_{ij}^0 - c\epsilon_n/p, \omega_{ij}^0 + c\epsilon_n/p) \subset (0, \infty)$ for $^{\forall}(i, j) \in \bar{\mathbf{E}}$. Note that by our assumption, $\#\mathbf{E} = s$. Denoting the prior of $\mathbf{\Omega}$ by Π,

$$\begin{aligned}
\Pi(||\mathbf{\Omega_0} - \mathbf{\Omega}||_{\infty} \leq c\epsilon_n/p) &= \Pi(|\omega_{ij}^0 - \omega_{ij}| \leq c\epsilon_n/p, (i,j) \in \bar{\mathbf{E}}) = \Pi(\omega_{ij}^0 - c\epsilon_n/p < \omega_{ij} < \omega_{ij}^0 + c\epsilon_n/p, (i,j) \in \bar{\mathbf{E}}) \\
&= \prod_{i=1}^{p}\int_{\omega_{ii}^0 - c\epsilon_n/p < \omega_{ii} < \omega_{ii}^0 + c\epsilon_n/p} \lambda\exp(-\lambda\omega_{ii})\mathrm{d}\omega_{ii} \prod_{(i,j)\in\mathbf{E}}\int_{\omega_{ij}^0 - c\epsilon_n/p < \omega_{ij} < \omega_{ij}^0 + c\epsilon_n/p} \frac{\lambda}{2}\exp(-\lambda|\omega_{ij}|)\mathrm{d}\omega_{ij} \\
&= \prod_{i=1}^{p}\exp(-2\lambda c\epsilon_n/p)\prod_{(i,j)\in\mathbf{E}}\frac{1}{2}\exp(-2\lambda\epsilon_n/p) = \frac{1}{2^s}\exp(-2\lambda c\epsilon_n(p+s)/p)
\end{aligned}$$

$$\frac{(c\epsilon_n/p)^{p+s}}{\frac{1}{2^s}\exp(-2\lambda c\epsilon_n(p_s)/p)} \to 0 \text{ as } n \to \infty \ (\because \epsilon_n \to 0 \text{ as } n \to \infty) \tag{14}$$

Therefore, (14) gives the estimate of lower bound for $\Pi(||\mathbf{\Omega_0} - \mathbf{\Omega}||_{\infty} \leq c\epsilon_n/p) \gtrsim (c\epsilon_n/p)^{p+s}$.

$$\begin{aligned}
\frac{-\log([c\epsilon_n/p]^{p+s})}{n\epsilon_n^2} &= \frac{(p+s)(-\log\epsilon_n - \log c + \log p)}{n\epsilon_n^2} = \frac{-\log c + \log p + \frac{1}{2}(\log\frac{n}{\log n}(p+s))}{\log n} \\
&= \frac{-\log c + \log p + \frac{1}{2}(\log(p+s) + \log n + \log\log n)}{\log n} \to \frac{1}{2} \text{ as } n \to \infty
\end{aligned} \tag{15}$$

By (15), $-\log([c\epsilon_n]^{p+s}) \lesssim n\epsilon_n^2$. Similarly, $n\epsilon_n^2 \lesssim -\log([c\epsilon_n]^{p+s})$. Therefore, we have $-\log([c\epsilon_n]^{p+s}) \asymp n\epsilon_n^2$. Equivalently, $(c\epsilon_n)^{p+s} \asymp \exp(-n\epsilon_n^2)$. To sum up,

$$\Pi(||\mathbf{\Omega_0} - \mathbf{\Omega}||_{\infty} \leq c\epsilon_n/p) \gtrsim \exp(-n\epsilon_n^2) \tag{16}$$

Let $\mathcal{P}_n$ be the set of all densities $p_{\mathbf{\Omega}}$ such that the graph induced by $\mathbf{\Omega}$ has maximum number of edges $\bar{r} < \binom{p}{2}/2$ and each entry of $\mathbf{\Omega}$ is at most L in absolute value, where $\bar{r}$ and L depend only on n and to be determined. Let $\mathbf{d}$ be metric with respect to Frobenius distance. To find $\mathbf{N}(\epsilon_n, \mathcal{P}_n, \mathbf{d})$, suppose the graph induced by $\mathbf{\Omega}$ has j edges($j \leq \bar{r}$). Clearly, the minimum number of $\mathbf{d}$-balls of size ϵ_n is achieved when the balls cover only the nonzero entries of $\mathbf{\Omega}$, where the number of such entries is j. Because each entry of $\mathbf{\Omega}$ is at most L in absolute value, the number of $\mathbf{d}$-balls of size ϵ_n to cover each entry $\frac{\mathrm{L}}{\epsilon_n}$. For there are j entries, we need $(\frac{\mathrm{L}}{\epsilon_n})^j$ balls. Also the number of graphs with j is $\binom{\binom{p}{2}}{j}$. Thus, we need total $(\frac{\mathrm{L}}{\epsilon_n})^j \binom{\binom{p}{2}}{j}$ balls to cover $\mathbf{\Omega}$ with j edges. Hence, $\mathbf{N}(\epsilon_n, \mathcal{P}_n, \mathbf{d}) = \sum_{j=1}^{\bar{r}} (\frac{\mathrm{L}}{\epsilon_n})^j \binom{\binom{p}{2}}{j}$. Taking sufficiently large n, ϵ_n can be made sufficiently large so that $\frac{\mathrm{L}}{\epsilon_n} > 1$. with such n, by the result of claim 2.4.,

$$
\begin{aligned}
\log\{\sum_{i=1}^{\bar{r}} (\frac{\mathrm{L}}{\epsilon_n})^j \binom{\binom{p}{2}}{j}\} &\leq \log\{\sum_{i=1}^{\bar{r}} (\frac{\mathrm{L}}{\epsilon_n})^{\bar{r}} \binom{p+\binom{p}{2}}{\bar{r}}\} = \log\{\bar{r}(\frac{\mathrm{L}}{\epsilon_n})^{\bar{r}} \binom{p+\binom{p}{2}}{\bar{r}}\} \\
&= \log \bar{r} + \bar{r} \log \mathrm{L} + \bar{r} \log \epsilon_n^{-1} + \log \binom{p+\binom{p}{2}}{\bar{r}} \\
&\lesssim \log \bar{r} + \bar{r} \log \mathrm{L} + \bar{r} \log \epsilon_n^{-1} + \bar{r} \log p
\end{aligned}
$$

Choose sufficiently large constants $b_1, b_2 > 0$ such that $\bar{r} \sim b_1 n\epsilon_n^2 / \log n = b_1(p+s)$ and $\mathrm{L} \sim b_2 n\epsilon_n^2 = b_2(p+s)\log n$.

$$
\begin{aligned}
&\frac{\log \bar{r} + \bar{r} \log \mathrm{L} + \bar{r} \log \epsilon_n^{-1} + \bar{r} \log p}{n\epsilon_n^2} \sim \frac{\log b_1(p+s) + b_1(p+s) \log \frac{b_2(p+s)p\log n}{n^{-\frac{1}{2}}(p+s)^{\frac{1}{2}}(\log n)^{\frac{1}{2}}}}{(p+s)\log n} \\
&= \frac{\log b_1(p+s) + b_1(p+s)\frac{1}{2}\log n + b_1(p+s)\log b_2 p(p+s)^{\frac{1}{2}}(\log n)^{\frac{1}{2}}}{(p+s)\log n} \to \frac{b_1}{2} \text{ as } n \to \infty
\end{aligned}
$$

Thus, $\log \bar{r} + \bar{r} \log \mathrm{L} + \bar{r} \log \epsilon_n^{-1} + \bar{r} \log p \lesssim n\epsilon_n^2$ and similarly, one can see that $n\epsilon_n^2 \lesssim \log \bar{r} + \bar{r} \log \mathrm{L} + \bar{r} \log \epsilon_n^{-1} + \bar{r} \log p$. So $\log \bar{r} + \bar{r} \log \mathrm{L} + \bar{r} \log \epsilon_n^{-1} + \bar{r} \log p \approx n\epsilon_n^2$.

$$
\mathbf{N}(\epsilon_n, \mathcal{P}_n, \mathbf{d}) = \log\{\sum_{i=1}^{\bar{r}} (\frac{\mathrm{L}}{\epsilon_n})^j \binom{\binom{p}{2}}{j}\} \lesssim n\epsilon_n^2 \tag{17}
$$

Note that under the condition $n\epsilon_n^2 / \log n << \binom{p}{2}$, the requirement $\bar{r} < \binom{p}{2}/2$ is satisifed as $n \to \infty$.

Suppose $p_{\mathbf{\Omega}}$ does not belong to $\mathcal{P}_n$. By definition of $\mathcal{P}_n$, there exists an entry of $\mathbf{\Omega}$ that exceeds L in magnitude or the number of edges in the graph induced by $\mathbf{\Omega}$ exceeds $\bar{r}$. The probability of latter event is $P(\bar{R} > \bar{r}) \leq \exp(-a_2' b_1 n\epsilon_n^2)$. Note b_1 is chosen to be sufficiently large and this holds by the prior for $\bar{R}$. To bound the probability of the formal event, if we naively say that prior on all entries of $\mathbf{\Omega}$ is given by Exponential prior, the probability of each entry exceeds L is $\int_{\mathrm{L}}^{\infty} \lambda \exp(-\lambda\omega_{ij}) \mathrm{d}\omega_{ij} = \exp(-\lambda \mathrm{L})$. For there are total $p + \binom{p}{2}$ distinct entries in $\mathbf{\Omega}$, assuming independence on entries for convenience, we have the probability of former event $(p + \binom{p}{2})\exp(-\lambda \mathrm{L}) \lesssim \binom{p}{2}\exp(-\mathrm{L})$. Hence to sum up, there exists a constant $\mathrm{C} > 0$ such that

$$
\Pi(\mathcal{P}_n^c) \leq \exp(-n\epsilon_n^2(\mathrm{C}+4)) \tag{18}
$$

- **Remark:** Suppose that for a sequence ϵ_n with $\epsilon_n \to 0$ and $n\epsilon_n^2 \to \infty$, a constant $\mathrm{C} > 0$ and sets $\mathcal{P}_n \subset \mathcal{P}$, where $\mathcal{P}$ is a space of densities, we have

  (Condition 1) $\log \mathbf{N}(\epsilon_n, \mathcal{P}_n, \mathbf{d}) \leq n\epsilon_n^2$

  (Condition 2) $\Pi(\mathcal{P}_n^c) \leq \exp(-n\epsilon_n^2(\mathrm{C}+4))$

(Condition 3) $\Pi(\mathrm{B}(p_{\Omega_0}, \epsilon_n)) \geq \exp(-n\epsilon_n^2 \mathrm{C})$

Then for sufficiently large M, $\mathbf{E_0[\Pi(d(p, p_0)} \geq \mathrm{M}\epsilon_\mathbf{n})|\mathbf{X^{(n)}}] \to 0$.

The theorem in remark is theorem 2.1 in article "Convergence Rate of Posterior Distributions", written by Ghosal, Ghosh, and Van Der Vaart, which is general theory of posterior convergence rate. (16), (17) and (18) verifies Condition 3, Condition1, and Condition2 in remark, respectively. Thus, by general theory of posterior cdonvergence rate, we see that

$$\mathbf{E_0[P\{||\Omega - \Omega_0||_2 > M\epsilon_n|X^{(n)}\}]} \to 0 \tag{19}$$

as $n \to \infty$. This ends the proof. $\square$

By theorem 2.2, we establish posterior convergence rate $\mathrm{O}_p(\epsilon_n)$. Thus $||\mathbf{\Omega} - \mathbf{\Omega_0}||_\mathbf{2} = \mathrm{O}(\epsilon_\mathbf{n})$ with posterior probability tending to one in probability and using theorem1 in article "Sparse permutation invariant covariance estimation" written by Rothmann, $||\mathbf{\Omega}^* - \mathbf{\Omega_0}||_\mathbf{2} = \mathbf{O_p}(\epsilon_\mathbf{n})$, where $\mathbf{\Omega}^*$ is the graphical lasso. Note the second inequality hodls because the assumption of theorem 2.3 satisfies that of theorem 1 in the just mentioned article. Resorting to triangle inequality, we have $||\mathbf{\Omega} - \mathbf{\Omega}^*||_\mathbf{2} = \mathrm{O}(\epsilon_\mathbf{n})$ with posterior probability tending to one in probability. This says, for a bounded and positive measurable function $f(\cdot)$ of $\mathbf{\Omega}$,

$$\frac{\int_{||\mathbf{\Omega}-\mathbf{\Omega}^*||_2 \leq \epsilon_n} f(\mathbf{\Omega}) \prod_{(i,j)\in\bar{\mathbf{E}}} \mathrm{d}\omega_{ij}}{\int_{\mathbf{\Omega}\in\mathcal{M}_0^+} f(\mathbf{\Omega}) \prod_{(i,j)\in\bar{\mathbf{E}}} \mathrm{d}\omega_{ij}} \to 1 \tag{20}$$

If the model is restricted to $\mathbf{\Gamma}$, then the posterior and the graphical lasso will concentrate around the projection of the true precision matrix on the model at rate ϵ_n, so that the posterior probability of an ϵ_n-Frobenius neighborhood around the graphical lasso in the model $\mathbf{\Gamma}$ will go to 1.

3 Posterior computation

Let $\mathbf{\Omega_\Gamma} = ((\omega_{\mathbf{\Gamma},ij}))$ be the precision matrix in model $\mathbf{\Gamma}$. In (6), $\mathbf{Q(\Omega, \Gamma|X^{(n)})}$ can be reformulated as

$$\mathbf{Q(\Omega,\Gamma|X^{(n)})} = \exp(\frac{n}{2}\log\det(\mathbf{\Omega}) - \frac{n}{2}\mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega}) - \lambda\sum_{\gamma_{ij}=1}|\omega_{\mathbf{\Gamma},ij}| - \frac{\lambda}{2}\sum_{i=1}^{p}\omega_{\mathbf{\Gamma},ii})\mathbb{1}_{\mathcal{M}_0^+}(\mathbf{\Omega_\Gamma}) = \exp(-\frac{n}{2}h_\mathbf{\Gamma}(\Omega_\mathbf{\Gamma}))\mathbb{1}_{\mathcal{M}_0^+}(\mathbf{\Omega_\Gamma})$$

where $h_\mathbf{\Gamma}(\mathbf{\Omega_\Gamma}) = -\log\det(\mathbf{\Omega_\Gamma}) + \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega_\Gamma}) + \dfrac{2\lambda}{n}\sum_{\gamma_{ij}=1}|\omega_{\mathbf{\Gamma},ij}| + \dfrac{\lambda}{n}\sum_{i=1}^{p}\omega_{\mathbf{\Gamma},ii}$

Let $\bar{\mathbf{E}}_\mathbf{\Gamma} = \{(i,j) \in \mathbf{V} \times \mathbf{V} : i = j, \text{ or } \gamma_{ij} = 1 i \neq j\}$. The marginal posterior density of the graphical structure indicator Γ is given by

$$\begin{aligned}\mathbf{p(\Gamma|X^{(n)})} &\propto \int \mathbf{p(\Gamma,\Omega|X^{(n)})} \prod_{(i,j)\in\bar{\mathbf{E}}_\mathbf{\Gamma}} \mathrm{d}\omega_{\mathbf{\Gamma},ij} = \mathbf{C_\Gamma}\int \mathbf{Q(\Gamma,\Omega|X^{(n)})} \prod_{(i,j)\in\bar{\mathbf{E}}_\mathbf{\Gamma}} \mathrm{d}\omega_{\mathbf{\Gamma},ij} \\ &= \mathbf{C_\Gamma}\int_{\mathbf{\Omega_\Gamma}\in\mathcal{M}_0^+} \exp(-\frac{n}{2}h_\mathbf{\Gamma}(\mathbf{\Omega_\Gamma})) \prod_{(i,j)\in\bar{\mathbf{E}}_\mathbf{\Gamma}} \mathrm{d}\omega_{\mathbf{\Gamma},ij}\end{aligned} \tag{21}$$

Considering the difficulty of calculating the integral in (21) in explicit form, the posterior of $\mathbf{\Gamma}$ is very intractable. Thus, we may use Laplace approximation to approximate $\mathbf{p(\Gamma|X^{(n)})}$.

3.1 Approximation of posterior

Suppose $h_\mathbf{\Gamma}(\mathbf{\Omega_\Gamma})$ is uniquely minimized at $\mathbf{\Omega_\Gamma}^*$, which is graphical lasso. Define $\Delta_\Gamma = \mathbf{\Omega_\Gamma} - \mathbf{\Omega}^*_\mathbf{\Gamma} = ((u_{\mathbf{\Gamma},ij}))$ Then, (21) can be reformulated as :

$$\begin{aligned}\mathbf{p}(\mathbf{\Gamma}|\mathbf{X^{(n)}}) &\propto \mathbf{C_\Gamma}\int_{\mathbf{\Omega_\Gamma}\in\mathcal{M}_0^+}\exp(-\frac{n}{2}h_{\mathbf{\Gamma}}(\mathbf{\Omega_\Gamma}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}}\mathrm{d}\omega_{\mathbf{\Gamma},ij} \\ &= \mathbf{C_\Gamma}\int_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+}\exp(-\frac{n}{2}h_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}+\mathbf{\Omega}^*_{\mathbf{\Gamma}}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}}\mathrm{d}u_{\mathbf{\Gamma},ij}.\end{aligned} \tag{22}$$

Let $g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}) = -\log\det(\mathbf{\Delta_\Gamma}+\mathbf{\Omega}^*_{\mathbf{\Gamma}}) + \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Delta_\Gamma}) + \frac{2\lambda}{n}\sum_{\gamma_{ij}=1}(|u_{\mathbf{\Gamma},ij}+\omega^*_{\mathbf{\Gamma},ij}| - |\omega^*_{\mathbf{\Gamma},ij}|) + \frac{\lambda}{n}\sum_{i=1}^{p}u_{\mathbf{\Gamma},ii}$.

$$\begin{aligned}h_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}+\mathbf{\Omega_\Gamma}^*) &= -\log\det(\mathbf{\Delta_\Gamma}+\mathbf{\Omega}^*_{\mathbf{\Gamma}}) + \mathbf{tr}(\hat{\mathbf{\Sigma}}(\mathbf{\Delta_\Gamma}+\mathbf{\Omega}^*_{\mathbf{\Gamma}})) + \frac{2\lambda}{n}\sum_{\gamma_{ij}=1}|u_{\mathbf{\Gamma},ij}+\omega^*_{\mathbf{\Gamma},ij}| + \frac{\lambda}{n}\sum_{i=1}^{p}(u_{\mathbf{\Gamma},ii}+\omega^*_{\mathbf{\Gamma},ii}) \\ &= g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}) + \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega}^*_{\mathbf{\Gamma}}) + \frac{2\lambda}{n}\sum_{\gamma_{ij}=1}|\omega^*_{\mathbf{\Gamma},\mathbf{ij}}| + \frac{\lambda}{n}\sum_{i=1}^{p}\omega^*_{\mathbf{\Gamma},ii} \\ &= g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}) - \log\det(\mathbf{\Omega_\Gamma}^*) + \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega}^*_{\mathbf{\Gamma}}) + \frac{2\lambda}{n}\sum_{\gamma_{ij}=1}|\omega^*_{\mathbf{\Gamma},\mathbf{ij}}| + \frac{\lambda}{n}\sum_{i=1}^{p}\omega^*_{\mathbf{\Gamma},ii} + \log\det(\mathbf{\Omega_\Gamma}^*) \\ &= g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}) + h_{\mathbf{\Gamma}}(\mathbf{\Omega}^*_{\mathbf{\Gamma}}) + \log\det(\mathbf{\Omega_\Gamma}^*)\end{aligned} \tag{23}$$

Plugging (23) into (22),

$$\begin{aligned}\mathbf{p}(\mathbf{\Gamma}|\mathbf{X^{(n)}}) &\propto \mathbf{C_\Gamma}\int_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+}\exp(-\frac{n}{2}(g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}) + h_{\mathbf{\Gamma}}(\mathbf{\Omega}^*_{\mathbf{\Gamma}}) + \log\det(\mathbf{\Omega_\Gamma}^*)))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}}\mathrm{d}u_{\mathbf{\Gamma},ij} \\ &= \mathbf{C_\Gamma}\exp(-\frac{n}{2}(h_{\mathbf{\Gamma}}(\mathbf{\Omega}^*_{\mathbf{\Gamma}}) + \log\det(\mathbf{\Omega_\Gamma}^*)))\int_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+}\exp(-\frac{n}{2}g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}}\mathrm{d}u_{\mathbf{\Gamma},ij} \\ &= \mathbf{C_\Gamma}\exp(-\frac{n}{2}h_{\mathbf{\Gamma}}(\mathbf{\Omega_\Gamma}^*))[\det(\mathbf{\Omega_\Gamma}^*)]^{-\frac{n}{2}}\int_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+}\exp(-\frac{n}{2}g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}}\mathrm{d}u_{\mathbf{\Gamma},ij}\end{aligned} \tag{24}$$

From equation (23), since $\mathbf{\Omega}^*_{\mathbf{\Gamma}}$ is an unique minimizer of $h_{\mathbf{\Gamma}}(\mathbf{\Omega_\Gamma})$, $g_{\mathbf{\Gamma}}$ is uniquely minimized at $\mathbf{0}$. Suppose that $g_{\mathbf{\Gamma}}$ is differentiable at $\mathbf{0}$ and so the derivative of $g_{\mathbf{\Gamma}}$ at $\mathbf{0}$ vanishes. Recall that Laplace approximation for twice differentiable function h on $\mathbb{R}^d$ with unique minimizer $\hat{\mathbf{x}}$ and $M > 0$,

$$\int e^{-Mh(\mathbf{x})}\mathrm{d}\mathbf{x} \approx e^{-Mh(\hat{\mathbf{x}})}(2\pi)^{\frac{d}{2}}M^{-\frac{d}{2}}[\det(D^2h(\mathbf{x}))|_{\mathbf{x}=\hat{\mathbf{x}}}]^{-\frac{1}{2}} \tag{25}$$

Define the matrix $\mathbf{H_B} = ((h_{\mathbf{B}}\{(i,j),(l,m)\}))$, where $h_{\mathbf{B}}\{(i,j),(l,m)\} = \mathbf{tr}(\mathbf{B^{-1}}\mathbf{E}_{(i,j)}\mathbf{B}^{-1}\mathbf{E}_{(l,m)})$. Suppose $\mathbf{X} \in \mathcal{M}^+$ and $\mathbf{A} \in \mathcal{M}$. Then we know that $\frac{\partial \mathbf{tr}(\mathbf{AX})}{\partial \mathbf{X}} = 2\mathbf{A} - \mathrm{diag}(\mathbf{A})$. So the partial derivative of $\mathbf{tr}(\mathbf{AX})$ with respect to any coordinate of $\mathbf{X}$ is constant and hence the second partial derivative with respect to any one or two coordinates of $\mathbf{X}$ will yield 0. In this view, $\mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Delta_\Gamma})$ will be vanished if it is twice differentiated with respect to any one or two coordiantes. Assuming that $g_{\mathbf{\Gamma}}$ is differentiable, clearly, the term $\frac{2\lambda}{n}\sum_{\gamma_{ij}=1}(|u_{\mathbf{\Gamma},ij}+\omega^*_{\mathbf{\Gamma},ij}| - |\omega^*_{\Gamma,ij}|) + \frac{\lambda}{n}\sum_{i=1}^{p}u_{\mathbf{\Gamma},ii}$ will be also vanished with twice differentiation. Thus, to obtain Hessian matrix of $g_{\mathbf{\Gamma}}$, it suffices to consider that of $-\log\det(\mathbf{\Delta_\Gamma}+\mathbf{\Omega}^*_{\mathbf{\Gamma}})$. Recall that if $p \times p$ matrix $\mathbf{F}$ is a matrix as a smooth function of vector $\mathbf{x} \in \mathrm{U}$ for some open set $\mathrm{U} \subset \mathbb{R}^s$ and $\det(\mathbf{F}(\mathbf{x})) > 0$, then

$$\frac{\partial^2 \log\det\mathbf{F}(\mathbf{x})}{\partial x_i \partial x_j} = \mathbf{tr}(\mathbf{F^{-1}}\frac{\partial^2 \mathbf{F}}{\partial \mathbf{x_i}\partial \mathbf{x_j}}) - \mathbf{tr}(\mathbf{F^{-1}}\frac{\partial \mathbf{F}}{\partial \mathbf{x_i}}\mathbf{F^{-1}}\frac{\partial \mathbf{F}}{\partial \mathbf{x_j}}) \tag{26}$$

Viewing $\mathbf{\Delta_\Gamma}+\mathbf{\Omega}^*_{\mathbf{\Gamma}}$ as a function of $\#\bar{\mathbf{E}}_{\mathbf{\Gamma}}$-vector $(u_{\mathbf{\Gamma},ij})$,

$$\frac{\partial(\mathbf{\Delta_\Gamma}+\mathbf{\Omega}^*_{\mathbf{\Gamma}})}{\partial u_{\mathbf{\Gamma},ij}} = \mathbf{E}_{(i,j)},\;\; \frac{\partial^2(\mathbf{\Delta_\Gamma}+\mathbf{\Omega}^*_{\mathbf{\Gamma}})}{\partial u_{\mathbf{\Gamma},ij}\partial u_{\mathbf{\Gamma},lm}} = \mathbf{0} \tag{27}$$

Therefore, by equation (26),

$$-\frac{\partial^2 \log\det(\mathbf{\Delta_\Gamma}+\mathbf{\Omega^*_\Gamma})}{\partial u_{\mathbf{\Gamma},ij}\partial u_{\mathbf{\Gamma},lm}} = \mathbf{tr}((\mathbf{\Delta_\Gamma}+\mathbf{\Omega^*_\Gamma})^{\mathbf{-1}}\mathbf{E}_{(i,j)}(\mathbf{\Delta_\Gamma}+\mathbf{\Omega^*_\Gamma})^{-1}\mathbf{E}_{(l,m)}) \tag{28}$$

Denoting approximation of $\mathbf{p}(\mathbf{\Gamma}|\mathbf{X^{(n)}})$ by $\mathbf{p}^*(\mathbf{\Gamma}|\mathbf{X^{(n)}})$, by (24), (25), and (28),

$$\begin{aligned}
\mathbf{p}^*(\mathbf{\Gamma}|\mathbf{X^{(n)}}) &\propto \mathbf{C_\Gamma}\exp(-\frac{n}{2}h_{\mathbf{\Gamma}}(\mathbf{\Omega^*_\Gamma}))[\det(\mathbf{\Omega^*_\Gamma})]^{-\frac{n}{2}}e^{-\frac{n}{2}g_{\mathbf{\Gamma}}(\mathbf{0})}(2\pi)^{\frac{\#\bar{\mathbf{E}}}{2}}(\frac{n}{2})^{-\frac{\#\bar{\mathbf{E}}}{2}}[\det(D^2 g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma})|_{\mathbf{\Delta_\Gamma}=\mathbf{0}}]^{-\frac{1}{2}} \\
&= \mathbf{C_\Gamma}\exp(-\frac{n}{2}h_{\mathbf{\Gamma}}(\mathbf{\Omega^*_\Gamma}))[\det(\mathbf{\Omega^*_\Gamma})]^{-\frac{n}{2}}e^{\frac{n}{2}\log\det(\mathbf{\Omega_\Gamma}^*)}(2\pi)^{\frac{\#\bar{\mathbf{E}}}{2}}(\frac{n}{2})^{-\frac{\#\bar{\mathbf{E}}}{2}}[\det(\mathbf{H}^*_{\mathbf{\Omega_\Gamma}})]^{-\frac{1}{2}} \\
&= \mathbf{C_\Gamma}\exp(-\frac{n}{2}h_{\Gamma}(\Omega^*_{\Gamma}))[\det(\mathbf{\Omega^*_\Gamma})]^{-\frac{n}{2}}[\det(\mathbf{\Omega^*_\Gamma})]^{\frac{n}{2}}(\frac{4\pi}{n})^{\frac{\#\bar{\mathbf{E}}}{2}}[\det(\mathbf{H}^*_{\mathbf{\Omega_\Gamma}})]^{-\frac{1}{2}} \\
&= \mathbf{C_\Gamma}\exp(-\frac{n}{2}h_{\Gamma}(\Omega^*_{\Gamma}))(\frac{4\pi}{n})^{\frac{\#\bar{\mathbf{E}}}{2}}[\det(\mathbf{H}^*_{\mathbf{\Omega_\Gamma}})]^{-\frac{1}{2}}
\end{aligned} \tag{29}$$

With respect to prior 2 in (4), we set $\bar{\mathbf{r}}$ by # of edges in graphical lasso of optimization problem (2). But still we have difficulty in implementation. Though we put restriction on the model size, we still have $\sum_{k=1}^{\bar{\mathbf{r}}}\binom{\binom{p}{2}}{k}$ models to search over. In case $p=30, \bar{\mathbf{r}}=29$, $\binom{\binom{p}{2}}{\bar{\mathbf{r}}} \geq 10^{45}$. Hence, in practice, it is impossible to search over all these models. Thus the idea to find maximizer $\mathbf{\Gamma}$ of $\mathbf{p}^*(\mathbf{\Gamma}|\mathbf{X^{(n)}})$ may be in fact bad. In section 4, we introduce Metropolis-Hastings algorithm so that we obtain MCMC samples of $\mathbf{\Gamma}$. Specific scheme will be discussed in section 4.

In our approximation, we assumed $g_{\mathbf{\Gamma}}$ to be differentiable at $\mathbf{0}$, which led to valid approximation. However, it is possible that $g_{\mathbf{\Gamma}}$ is not differentiable at $\mathbf{0}$ so that the derivative does not exist. Further, we actually need to verify the determinant of $\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}$ is bounded away from 0, because of the term $[\det(\mathbf{H}^*_{\mathbf{\Omega_\Gamma}})]^{-\frac{1}{2}}$ in the approximation. We may refer to the model with valid approximation as regular model, otherwise non-regular model. But in the next section, we show that we can ignore non-regular models and so we may use the approximation in (29) without considering non-regular models. Before discussing ignorability of non-regular models, we first show that $\det(\mathbf{H}^*_{\mathbf{\Omega_\Gamma}})$ is indeed bounded away from 0. To show this, it sufficies to show that the minimum eigenvalue of $\mathbf{H}^*_{\mathbf{\Omega_\Gamma}}$ is bounded away from 0 as Hessian matrix is symmetric.

Theorem 3.1. *Given a model $\mathbf{\Gamma}$, the minimum eigenvalue of the Hessian $\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}$ is bounded away from 0.*

Proof. Clearly one can see that the Hessian $\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}$ evaluated at the graphical lasso $\mathbf{\Omega^*_\Gamma}$ corresponding to the model $\mathbf{\Gamma}$ has the form $\mathbf{T^t}\mathbf{A}^*_{\mathbf{\Omega_\Gamma}}\mathbf{T}$, where the $(p+2\#\mathbf{\Gamma})$-dimensional matrix $\mathbf{A}^*_{\mathbf{\Omega_\Gamma}}$ is a principal minor of the $p^2 \times p^2$ matrix $(\mathbf{\Omega^*_\Gamma})^{\mathbf{-1}}\bigotimes(\mathbf{\Omega^*_\Gamma})^{\mathbf{-1}}$, and $\mathbf{T}$ is a $(p+2\#\Gamma)\times(p+\#\Gamma)$ matrix of 0s and 1s having full column rank. The determinant of Hessian $\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}$ is not 0 if and only if it is full rank, which holds if and only $\mathbf{T^t}\mathbf{A}^*_{\mathbf{\Omega_\Gamma}}\mathbf{T}$ has full rank, again which is true only when $\text{eig}_1((\mathbf{\Omega^*_\Gamma})^{\mathbf{-1}}\bigotimes(\mathbf{\Omega^*_\Gamma})^{\mathbf{-1}})$ is bounded away from 0. It is trivial that $\text{eig}_1((\mathbf{\Omega^*_\Gamma})^{\mathbf{-1}}\bigotimes(\mathbf{\Omega^*_\Gamma})^{\mathbf{-1}}) = [\text{eig}_1((\mathbf{\Omega^*_\Gamma})^{\mathbf{-1}})]^2$. We assumed precision matrices to be belonged to $\mathcal{M}_0^+$, which insists the minimum and maximum eigenvalues each element in the element are bounded by fixed constants. Thus the eigenvalues of $\mathbf{\Omega^*_\Gamma}$ are bounded away from 0. Therefore, $[\text{eig}_1((\mathbf{\Omega^*_\Gamma})^{\mathbf{-1}})]^2 = \frac{1}{||\mathbf{\Omega^*_\Gamma}||_{\mathbf{2}}^{\mathbf{2}}} > 0$. This establishes theorem 3.1. □

3.2 Ignorability of non-regular models

For convenience, suppose that the vector $\mathbf{\Gamma}$ has t 1s on its first components and the rest of them are 0. The model $\mathbf{\Gamma}$ is non-regular when the corresponding graphical lasso to the model $\mathbf{\Gamma}$ has at least one 0 entry on the edge which is presented in the model $\mathbf{\Gamma}$. So among those t 1s, say the last r of them have corresponding graphical solution equal to 0. In this context, we obtain submodel $\mathbf{\Gamma}'$ of $\mathbf{\Gamma}$ with first $(t-r)$ 1s and rest 0s. We argue that the graphical lasso corresponding $\mathbf{\Gamma}'$ and $\mathbf{\Gamma}$ are the same. Thus, $\mathbf{\Gamma}'$ is a regular submodel of non-regular model $\mathbf{\Gamma}$.

Lemma 3.1. *For the corresponding regular submodel $\mathbf{\Gamma}'$ of $\mathbf{\Gamma}$, the graphical lasso solutions corresponding to models $\mathbf{\Gamma}$ and $\mathbf{\Gamma}'$ are identical.*

Proof. Let us first recall the Karush-Kuhn-Tucker(KKT) condition.

- **Remark :** Consider the convex optimization problem:

$$\begin{aligned} &\text{minimize} && f_0(x) \\ &\text{subject to} && f_i(x) \leq 0, i = 1, 2, \ldots, m \\ & && h_i(x) = 0, i = 1, 2, \ldots, p \end{aligned}$$

  where f_i's are convex functions and h_i's are affine functions. If its corresponding primal problem is also convex, the sufficiency for $(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ to be primal and dual optimal, which we refer such condition as to KKT condition, is given by:

$$\begin{aligned} f_i(\tilde{x}) \leq 0, i =, 1, 2, \ldots, m \\ h_i(\tilde{x}) = 0, i = 1, 2, \ldots, p \\ \tilde{\lambda}_i \geq 0, i = 1, 2, \ldots, m \\ \tilde{\lambda}_i f_i(\tilde{x}) = 0, i = 1, 2, \ldots, m \\ \nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{x}) = 0 \end{aligned}$$

Our convex optimization problem is to minimize $-\log\det(\mathbf{\Omega_\Gamma}) + \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega_\Gamma}) + \frac{2\lambda}{n}\sum_{\gamma_{ij}=1} |\omega_{\mathbf{\Gamma},ij}| + \frac{\lambda}{n}\sum_{i=1}^{p} \omega_{\mathbf{\Gamma},ii}$. For no constraints f_i, h_j, $i = 1, 2, \ldots, m, j = 1, 2, \ldots, p$ are given, f_i's and h_i's are identically 0. Clearly one can see that for any square matrix $\mathbf{X}$, $a(\mathbf{X}) = 2\mathbf{X} - \operatorname{diag}(\mathbf{X})$ is one-to-one function, so $a(\mathbf{X}) = \mathbf{0}$ only when $\mathbf{X} = \mathbf{0}$. Therefore, by the last condition in KKT condtion,

$$\begin{aligned} &\frac{\partial(-\log\det(\mathbf{\Omega_\Gamma}) + \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega_\Gamma}) + \frac{2\lambda}{n}\sum_{\gamma_{ij}=1}|\omega_{\mathbf{\Gamma},\mathbf{ij}}| + \frac{\lambda}{n}\sum_{i=1}^{p}\omega_{\mathbf{\Gamma},ii})}{\partial\mathbf{\Omega_\Gamma}} \\ &= -(2(\mathbf{\Omega_\Gamma})^{-1} - \operatorname{diag}((\mathbf{\Omega_\Gamma})^{-1})) + 2\hat{\mathbf{\Sigma}} - \operatorname{diag}(\hat{\mathbf{\Sigma}}) + \frac{2\lambda}{n}(\mathbf{G} - \mathbf{I_p}) + \frac{\lambda}{n}I_p \\ &= -(2(\mathbf{\Omega_\Gamma})^{-1} - \operatorname{diag}((\mathbf{\Omega_\Gamma})^{-1})) + 2\hat{\mathbf{\Sigma}} - \operatorname{diag}(\hat{\mathbf{\Sigma}}) + \frac{2\lambda}{n}\mathbf{G} - \frac{\lambda}{n}I_p \\ &= -(2(\mathbf{\Omega_\Gamma})^{-1} - \operatorname{diag}((\mathbf{\Omega_\Gamma})^{-1})) + 2\hat{\mathbf{\Sigma}} - \operatorname{diag}(\hat{\mathbf{\Sigma}}) + \frac{2\lambda}{n}\mathbf{G} - \frac{\lambda}{n}\operatorname{diag}(\mathbf{G}) \\ &= -a(\mathbf{\Omega_\Gamma^{-1}} - \hat{\mathbf{\Sigma}} - \frac{\lambda}{\mathbf{n}}\mathbf{G}) = 0 \Leftrightarrow \mathbf{\Omega_\Gamma^{-1}} - \hat{\mathbf{\Sigma}} - \frac{\lambda}{\mathbf{n}}\mathbf{G} = \mathbf{0} \end{aligned} \tag{30}$$

, where $\mathbf{G} = ((g_{ij}))$ is a matrix with if $i = j$, $g_{ij} = 1$ and otherwise $g_{ij} = \frac{\omega_{\mathbf{\Gamma},ij}}{|\omega_{\mathbf{\Gamma},ij}|}$ for $\omega_{\mathbf{\Gamma},ij} \neq 0$ and $|g_{ij}| \leq 1$ for $\omega_{\mathbf{\Gamma},ij} = 0$. Since $\mathbf{\Omega}^*_{\mathbf{\Gamma}}$ is graphical lasso, by (30), $\mathbf{\Omega}^{*\mathbf{-1}}_{\mathbf{\Gamma}} - \hat{\mathbf{\Sigma}} - \frac{\lambda}{\mathbf{n}}\mathbf{G} = \mathbf{0}$. Let (i, j) be a pair such that $\gamma_{ij} = 1$ but $\omega^*_{\mathbf{\Gamma},ij} = 0$. By the definition of regular submodel $\mathbf{\Gamma}'$, $\Omega^*_{\mathbf{\Gamma}'}$ also satisfies $\mathbf{\Omega_\Gamma^{-1}} - \hat{\mathbf{\Sigma}} - \frac{\lambda}{\mathbf{n}}\mathbf{G} = \mathbf{0}$ because $\omega^*_{\mathbf{\Gamma},ij} = 0$ for any $\gamma'_{ij} = 0$. $\Omega^*_{\mathbf{\Gamma}'} = \Omega^*_{\mathbf{\Gamma}}$ follows from the uniqueness of solution of the given optimization problem. $\square$

Lemma 3.2. *Consider a non-regular model $\mathbf{\Gamma}$ and let $\mathbf{\Gamma'}$ be its corresponding regular submodel with their common graphical lasso $\mathbf{\Omega}^*_{\mathbf{\Gamma}}$. Let $\mathbf{\Delta_\Gamma} = \mathbf{\Omega_\Gamma} - \mathbf{\Omega}^*_{\mathbf{\Gamma}} = ((u_{\mathbf{\Gamma},ij}))$ and $\mathbf{\Delta_{\Gamma'}} = ((u_{\mathbf{\Gamma'},ij}))$ such that $u_{\mathbf{\Gamma'},\mathbf{ij}} = u_{\mathbf{\Gamma},\mathbf{ij}}$ if $i=j$ or $\gamma_{ij} = \gamma'_{ij} = 1$ and $u_{\mathbf{\Gamma'},ij} = 0$ for pairs (i,j) with $\gamma'_{ij} = 0$. Then fixed values of $u_{\mathbf{\Gamma'},ij}$ for $\gamma'_{ij} = 1$, we have,*

$$\log\det(\mathbf{\Delta_\Gamma} + \mathbf{\Omega}^*_{\mathbf{\Gamma}}) - \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Delta_\Gamma}) \leq \log\det(\mathbf{\Delta_{\Gamma'}} + \mathbf{\Omega}^*_{\mathbf{\Gamma}}) - \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Delta_{\Gamma'}}) \tag{31}$$

Proof. Suppose that $\mathbf{F}$ is $p \times p$ matrix as a differentiable function with respect to $\mathbf{x}$. If $\det \mathbf{F}(\mathbf{x}) > 0$ and $\mathbf{A} \in \mathcal{M}$ is given, then we have

$$\frac{\partial \log\det \mathbf{F}}{\partial \mathbf{x}_i} = \mathbf{tr}(\mathbf{F^{-1}}\frac{\partial \mathbf{F}}{\partial \mathbf{x}_i}), \quad \frac{\partial \mathbf{tr}(\mathbf{AF})}{\partial \mathbf{x}_i} = \mathbf{tr}(\mathbf{A}\frac{\partial \mathbf{F}}{\partial \mathbf{x}_i}) \tag{32}$$

Consider the maximization of the function $f(\mathbf{\Delta_\Gamma}) = \log\det(\mathbf{\Delta_\Gamma} + \mathbf{\Omega}^*_{\mathbf{\Gamma}}) - \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Delta_\Omega})$ with respect to the elements $u_{\mathbf{\Gamma},ij}$ where $(i,j) \in \{(i,j) : \gamma_{ij} = 1, \gamma'_{ij} = 0\}$. By (22), he maximizer $\hat{u_{\mathbf{\Gamma},ij}}$ of f satisfies,

$$\frac{\partial f(\mathbf{\Delta_\Gamma})}{\partial u_{\mathbf{\Gamma},ij}} = \mathbf{tr}((\mathbf{\Delta_\Gamma} + \mathbf{\Omega}^*_{\mathbf{\Gamma}})^{\mathbf{-1}}\mathbf{E}_{(i,j)}) - \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{E}_{(i,j)}) = \mathbf{0} \tag{33}$$

Also, for $g_{\mathbf{\Gamma}}$ defined in Section 3.1, we saw that $\mathbf{0}$ is an unique minimizer of $g_{\mathbf{\Gamma}}$. Since $g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}) = -f(\mathbf{\Delta_\Gamma}) + \frac{2\lambda}{n}\sum_{\gamma_{ij}=1}(|u_{\mathbf{\Gamma},ij} + \omega^*_{\mathbf{\Gamma},ij}| - |\omega^*_{\mathbf{\Gamma},ij}|) + \frac{\lambda}{n}\sum_{i=1}^{p} u_{\mathbf{\Gamma},ii}$,

$$\begin{aligned}
\frac{\partial g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma})}{\partial u_{\mathbf{\Gamma},ij}}\bigg|_{u_{\mathbf{\Gamma},ij}=0^+, u_{\mathbf{\Gamma},lm}=0, \forall(l,m)\neq(i,j)} &= \frac{2\lambda}{n} \geq 0 \\
\frac{\partial g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma})}{\partial u_{\mathbf{\Gamma},ij}}\bigg|_{u_{\mathbf{\Gamma},ij}=0^-, u_{\mathbf{\Gamma},lm}=0, \forall(l,m)\neq(i,j)} &= -\frac{2\lambda}{n} \leq 0
\end{aligned} \tag{34}$$

Note that the equalities in (34) follow from our choice of (i,j) to satisfy $\omega^*_{\mathbf{\Gamma},ij} = 0$. Thus, we have

$$\begin{aligned}
\frac{2\lambda}{n} &\geq 0 = \mathbf{tr}((\mathbf{\Delta_\Gamma} + \mathbf{\Omega}^*_{\mathbf{\Gamma}})^{\mathbf{-1}}\mathbf{E}_{(i,j)}) - \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{E}_{(i,j)})\bigg|_{u_{\mathbf{\Gamma},ij}=0^+, u_{\mathbf{\Gamma},lm}=0, \forall(l,m)\neq(i,j)} \\
-\frac{2\lambda}{n} &\leq 0 = \mathbf{tr}((\mathbf{\Delta_\Gamma} + \mathbf{\Omega}^*_{\mathbf{\Gamma}})^{\mathbf{-1}}\mathbf{E}_{(i,j)}) - \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{E}_{(i,j)})\bigg|_{u_{\mathbf{\Gamma},ij}=0^+, u_{\mathbf{\Gamma},lm}=0, \forall(l,m)\neq(i,j)}
\end{aligned} \tag{35}$$

Because the derivative of f is continuous at $\mathbf{0}$, we have $\hat{u}_{\mathbf{\Gamma},\mathbf{ij}} = 0$. This shows that f is maximized when $\mathbf{\Delta_\Gamma} = \mathbf{\Delta_{\Gamma'}}$, which establishes inequality (31). ☐

Considering $h_\Gamma, \bar{\mathbf{E}}_{\mathbf{\Gamma}}$ defined in Section 3 and (20), we have that

$$\frac{\mathbf{p}(\mathbf{\Gamma}|\mathbf{X^{(n)}})}{\mathbf{p}(\mathbf{\Gamma'}|\mathbf{X^{(n)}})} = \frac{\mathbf{C_\Gamma}\int_{||\mathbf{\Delta_\Gamma}||_2 \leq \epsilon_n} \exp(-\frac{n}{2}h_{\mathbf{\Gamma}}(\mathbf{\Omega_\Gamma}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}} \mathrm{d}\omega_{\mathbf{\Gamma},ij} + \mathrm{o}(1)}{\mathbf{C_{\Gamma'}}\int_{||\mathbf{\Delta_{\Gamma'}}||_2 \leq \epsilon_n} \exp(-\frac{n}{2}h_{\mathbf{\Gamma'}}(\mathbf{\Omega_{\Gamma'}}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma'}}} \mathrm{d}\omega_{\mathbf{\Gamma'},ij} + \mathrm{o}(1)} \tag{36}$$

where $\epsilon_n = n^{-\frac{1}{2}}(p+s)^{\frac{1}{2}}(\log n)^{\frac{1}{2}}$

Theorem 3.2. *Consider the prior on $\mathbf{\Gamma}$ as given in (3) or (4) with $q < \frac{1}{2}$. The posterior probability of a non-regular model $\mathbf{\Gamma}$ is always less than that of the corresponding regular submodel $\mathbf{\Gamma'}$*

Proof. It is clear that $\{u_{\mathbf{\Gamma},ij} : ||\mathbf{\Delta_\Gamma}||_2 \leq \epsilon_n\} \subset \{u_{\mathbf{\Gamma'},ij} : ||\mathbf{\Delta_{\Gamma'}}||_2 \leq \epsilon_n\} \cdots (*)$. Because the integrands in nominator and denominator in equation (36) are positive, similar to (24), one can deduce that

$$\begin{aligned}
\frac{\mathbf{p}(\mathbf{\Gamma}|\mathbf{X^{(n)}})}{\mathbf{p}(\mathbf{\Gamma'}|\mathbf{X^{(n)}})} &\approx \frac{\mathbf{C_\Gamma}\int_{||\mathbf{\Delta_\Gamma}||_2 \leq \epsilon_n} \exp(-\frac{n}{2}h_{\mathbf{\Gamma}}(\mathbf{\Omega_\Gamma}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}} \mathrm{d}\omega_{\mathbf{\Gamma},ij}}{\mathbf{C_{\Gamma'}}\int_{||\mathbf{\Delta_{\Gamma'}}||_2 \leq \epsilon_n} \exp(-\frac{n}{2}h_{\mathbf{\Gamma'}}(\mathbf{\Omega_{\Gamma'}}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma'}}} \mathrm{d}\omega_{\mathbf{\Gamma'},ij}} \\
&= \frac{\mathbf{C_\Gamma}\exp(-\frac{n}{2}h_{\mathbf{\Gamma}}(\mathbf{\Omega_\Gamma}^*))[\det(\mathbf{\Omega_\Gamma}^*)]^{-\frac{n}{2}}\int_{||\mathbf{\Delta_\Gamma}||_2 \leq \epsilon_n} \exp(-\frac{n}{2}g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}} \mathrm{d}u_{\mathbf{\Gamma},ij}}{\mathbf{C_{\Gamma'}}\exp(-\frac{n}{2}h_{\mathbf{\Gamma'}}(\mathbf{\Omega_{\Gamma'}}^*))[\det(\mathbf{\Omega_{\Gamma'}}^*)]^{-\frac{n}{2}}\int_{||\mathbf{\Delta_{\Gamma'}}||_2 \leq \epsilon_n} \exp(-\frac{n}{2}g_{\mathbf{\Gamma'}}(\mathbf{\Delta_{\Gamma'}}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma'}}} \mathrm{d}u_{\mathbf{\Gamma'},ij}} \\
&= \frac{\mathbf{C_\Gamma}\int_{||\mathbf{\Delta_\Gamma}||_2 \leq \epsilon_n} \exp(-\frac{n}{2}g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}} \mathrm{d}u_{\mathbf{\Gamma},ij}}{\mathbf{C_{\Gamma'}}\int_{||\mathbf{\Delta_{\Gamma'}}||_2 \leq \epsilon_n} \exp(-\frac{n}{2}g_{\mathbf{\Gamma'}}(\mathbf{\Delta_{\Gamma'}}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma'}}} \mathrm{d}u_{\mathbf{\Gamma'},ij}} \\
&\leq \frac{\mathbf{C_\Gamma}\int_{||\mathbf{\Delta_\Gamma}||_2 \leq \epsilon_n} \exp(-\frac{n}{2}g_{\mathbf{\Gamma}}(\mathbf{\Delta_\Gamma}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}} \mathrm{d}u_{\mathbf{\Gamma},ij}}{\mathbf{C_{\Gamma'}}\int_{||\mathbf{\Delta_\Gamma}||_2 \leq \epsilon_n} \exp(-\frac{n}{2}g_{\mathbf{\Gamma'}}(\mathbf{\Delta_{\Gamma'}}))\prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma'}}} \mathrm{d}u_{\mathbf{\Gamma'},ij}}
\end{aligned} \tag{37}$$

Here the last equation in (37) holds by the result of Lemma 3.1, which says $\mathbf{\Omega}^*_{\mathbf{\Gamma}} = \mathbf{\Omega}^*_{\mathbf{\Gamma}'}$, and the last inequality holds by (*). Considering the term $-\log\det(\mathbf{\Delta}_{\mathbf{\Gamma}} + \mathbf{\Omega}^*_{\mathbf{\Gamma}}) + \mathbf{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Delta}_{\mathbf{\Gamma}})$ in g_Γ, by the result of Lemma 3.2 and the substitution of $\mathbf{C}_{\mathbf{\Gamma}}$ in (6),

$$
\begin{aligned}
&\frac{\mathbf{C}_{\mathbf{\Gamma}} \int_{||\mathbf{\Delta}_{\mathbf{\Gamma}}||_2 \le \epsilon_n} \exp(-\frac{n}{2} g_{\mathbf{\Gamma}}(\mathbf{\Delta}_{\mathbf{\Gamma}})) \prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}} \mathrm{d}u_{\mathbf{\Gamma},ij}}{\mathbf{C}_{\mathbf{\Gamma}'} \int_{||\mathbf{\Delta}_{\mathbf{\Gamma}}||_2 \le \epsilon_n} \exp(-\frac{n}{2} g_{\mathbf{\Gamma}'}(\mathbf{\Delta}_{\mathbf{\Gamma}'})) \prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}'}} \mathrm{d}u_{\mathbf{\Gamma}',ij}} \\
&= \frac{\mathbf{C}_{\mathbf{\Gamma}}}{\mathbf{C}_{\mathbf{\Gamma}'}} \int_{||\mathbf{\Delta}_{\mathbf{\Gamma}}||_2 \le \epsilon_n} \exp(-\frac{n}{2}\frac{2\lambda}{n} \sum_{\gamma_{ij}=1,\gamma'_{ij}=0} |u_{\mathbf{\Gamma},ij}|) \prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}\cap\bar{\mathbf{E}}^c_{\mathbf{\Gamma}'}} \mathrm{d}u_{\mathbf{\Gamma},ij} \\
&= \frac{\mathbf{C}_{\mathbf{\Gamma}}}{\mathbf{C}_{\mathbf{\Gamma}'}} \int_{||\mathbf{\Delta}_{\mathbf{\Gamma}}||_2 \le \epsilon_n} \exp(-\lambda \sum_{\gamma_{ij}=1,\gamma'_{ij}=0} |u_{\mathbf{\Gamma},ij}|) \prod^{c}_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}\cap\bar{\mathbf{E}}_{\mathbf{\Gamma}'}} \mathrm{d}u_{\mathbf{\Gamma},ij} \\
&\le \frac{\mathbf{C}_{\mathbf{\Gamma}}}{\mathbf{C}_{\mathbf{\Gamma}'}} \int \exp(-\lambda \sum_{\gamma_{ij}=1,\gamma'_{ij}=0} |u_{\mathbf{\Gamma},ij}|) \prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}\cap\bar{\mathbf{E}}_{\mathbf{\Gamma}'}} \mathrm{d}u_{\mathbf{\Gamma},ij} \\
&= \frac{\mathbf{C}_{\mathbf{\Gamma}}}{\mathbf{C}_{\mathbf{\Gamma}'}} (\int \exp(-\lambda|u_{\mathbf{\Gamma},ij}|)\mathrm{d}u_{\mathbf{\Gamma},ij})^{\#\Gamma-\#\Gamma'} = \frac{\mathbf{C}_{\mathbf{\Gamma}}}{\mathbf{C}_{\mathbf{\Gamma}'}} (\frac{2}{\lambda})^{\#\Gamma-\#\Gamma'} \\
&= (\frac{q}{1-q})^{\#\mathbf{\Gamma}-\#\mathbf{\Gamma}'} (\frac{\lambda}{2})^{\#\mathbf{\Gamma}-\#\mathbf{\Gamma}'} (\frac{2}{\lambda})^{\#\Gamma-\#\Gamma'} \frac{\beta(\mathbf{\Gamma})}{\beta(\mathbf{\Gamma}')} \\
&= (\frac{q}{1-q})^{r} \frac{\beta(\mathbf{\Gamma})}{\beta(\mathbf{\Gamma}')}
\end{aligned}
\tag{38}
$$

where $r = \#\mathbf{\Gamma} - \#\mathbf{\Gamma}' > 0$. If prior for Γ follows (3), because $\mathbf{P}(\bar{\mathbf{R}} \le \#\mathbf{\Gamma}) > \mathbf{P}(\bar{\mathbf{R}} \le \#\mathbf{\Gamma}')$ for $r > 0$, so $\frac{\beta(\mathbf{\Gamma})}{\beta(\mathbf{\Gamma}')} \le 1$ Otherwise prior for Γ follows (4) and $\frac{\beta(\mathbf{\Gamma})}{\beta(\mathbf{\Gamma}')} = 1$. Therefore, combining (37) and (38),

$$
\frac{\mathbf{p}(\mathbf{\Gamma}|\mathbf{X}^{(\mathbf{n})})}{\mathbf{p}(\mathbf{\Gamma}'|\mathbf{X}^{(\mathbf{n})})} \le (\frac{q}{1-q})^r \tag{39}
$$

If $q < \frac{1}{2}$, $\frac{q}{1-q} < 1$ and so from (38), we have $\mathbf{p}(\mathbf{\Gamma}|\mathbf{X}^{(\mathbf{n})}) \le \mathbf{p}(\mathbf{\Gamma}'|\mathbf{X}^{(\mathbf{n})})$. Hence the posterior probability of regular submodel $\mathbf{\Gamma}'$ is larger than that of non-regular model $\mathbf{\Gamma}$. Thus regular submodel is more likely than non-regular model.

□

As a consequence of theorem 3.2., if we choose $q < \frac{1}{2}$, then we can focus on regular models only, ignoring non-regular models.

3.3 Error in Laplace approximation

In this section, we estimate error in approximation of posterior probability of $\mathbf{\Gamma}$. Following the notation in section 3.1, let $\mathbf{\Delta}_{\mathbf{\Gamma}} = \mathbf{\Omega}_{\mathbf{\Gamma}} - \mathbf{\Omega}^*_{\mathbf{\Gamma}}$ and $\mathrm{vec}(\mathbf{\Delta}_{\mathbf{\Gamma}})$ denote the vectorized version of $\mathbf{\Delta}_{\mathbf{\Gamma}}$, but excluding the entries set to zero by the model $\mathbf{\Gamma}$. As a consequence of section 3.2, we only consider regular model $\mathbf{\Gamma}$. Note $\mathrm{vec}(\mathbf{\Delta}_{\mathbf{\Gamma}})$ is a vector of dimension at most $p + \#\mathbf{\Gamma}$ and the approximation of posterior probability of model $\mathbf{\Gamma}$ was based on the Taylor's expansion of $h_{\mathbf{\Gamma}}$ around the minimizer, which was $\mathbf{\Omega_\Gamma}^*$ and so the derivative of $h_{\mathbf{\Gamma}}$ at $\mathbf{\Omega_\Gamma}^*$ is $\mathbf{0} \cdots (**)$

Lemma 3.3. *For any regular model $\mathbf{\Gamma}$, with probability tending to one, the remainder term of function $h_{\mathbf{\Gamma}}(\mathbf{\Omega}_{\mathbf{\Gamma}})$, defined in the beginning of section (3), is bounded by $(p+\#\mathbf{\Gamma})||\mathbf{\Delta}_{\mathbf{\Gamma}}||_2^2(C_1||\mathbf{\Delta}_{\mathbf{\Gamma}}||_2 + C_2||\mathbf{\Delta}_{\mathbf{\Gamma}}||_2^2)/2$ for some constants $C_1, C_2 > 0$.*

Proof. Referring to (**), by Taylor's expansion, we see that

$$
\begin{aligned}
h_{\mathbf{\Gamma}}(\mathbf{\Omega}_{\mathbf{\Gamma}}) &= h_{\mathbf{\Gamma}}(\mathbf{\Omega_\Gamma}^*) + \frac{1}{2}\mathrm{vec}(\mathbf{\Omega}_{\mathbf{\Gamma}} - \mathbf{\Omega}^*_{\mathbf{\Gamma}})^t \mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}} \mathrm{vec}(\mathbf{\Omega}_{\mathbf{\Gamma}} - \mathbf{\Omega}^*_{\mathbf{\Gamma}}) + \mathrm{R}_n \\
&= h_{\mathbf{\Gamma}}(\mathbf{\Omega_\Gamma}^*) + \frac{1}{2}\mathrm{vec}(\mathbf{\Delta}_{\mathbf{\Gamma}})^t \mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}} \mathrm{vec}(\mathbf{\Delta}_{\mathbf{\Gamma}}) + \mathrm{R}_n
\end{aligned}
\tag{40}
$$

where R_n is the remainder term in the expansion. Recall that we've shown the Hessian matrix of $h_{\mathbf{\Gamma}}$ at $\mathbf{\Omega}^*_{\mathbf{\Gamma}}$ is $\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}$ in the section 3.1. Want to show $|\mathrm{R}_n| \leq (p+\#\mathbf{\Gamma})||\mathbf{\Delta_\Gamma}||_2^2(\mathrm{C}_1||\mathbf{\Delta_\Gamma}||_2 + \mathrm{C}_2||\mathbf{\Delta_\Gamma}||_2^2)/2$. Using the integral form of the remainder, we have

$$h_{\mathbf{\Gamma}}(\mathbf{\Omega_\Gamma}) = h_{\mathbf{\Gamma}}(\mathbf{\Omega_\Gamma}^*) + \mathrm{vec}(\mathbf{\Delta_\Gamma})^t[\int_0^1 (1-v)\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}}\mathrm{d}v]\mathrm{vec}(\mathbf{\Delta_\Gamma}) \tag{41}$$

Subtracting equation (41) from equation (40),

$$\begin{aligned}
\mathrm{R}_n &= \mathrm{vec}(\mathbf{\Delta_\Gamma})^t[\int_0^1 (1-v)\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}}\mathrm{d}v]\mathrm{vec}(\mathbf{\Delta_\Gamma}) - \frac{1}{2}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) \\
&= \mathrm{vec}(\mathbf{\Delta_\Gamma})^t[\int_0^1 (1-v)\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}}\mathrm{d}v]\mathrm{vec}(\mathbf{\Delta_\Gamma}) - [\int_0^1 (1-v)\mathrm{d}v]\,\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) \\
&= \mathrm{vec}(\mathbf{\Delta_\Gamma})^t[\int_0^1 (1-v)\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}}\mathrm{d}v]\mathrm{vec}(\mathbf{\Delta_\Gamma}) - \mathrm{vec}(\mathbf{\Delta_\Gamma})^t[\int_0^1 (1-v)\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}\mathrm{d}v]\mathrm{vec}(\mathbf{\Delta_\Gamma}) \\
&= \mathrm{vec}(\mathbf{\Delta_\Gamma})^t[\int_0^1 (1-v)(\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}} - \mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}})\mathrm{d}v]\mathrm{vec}(\mathbf{\Delta_\Gamma})
\end{aligned} \tag{42}$$

By Cauchy-Shwartz inequality and theorem 1.2,

$$\begin{aligned}
|\mathrm{R}_n| &\leq ||\mathbf{\Delta_\Gamma}||_2||[\int_0^1 (1-v)(\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}} - \mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}})\mathrm{d}v]\mathrm{vec}(\mathbf{\Delta_\Gamma})]||_2 \\
&\leq ||\mathbf{\Delta_\Gamma}||_2||\int_0^1 (1-v)(\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}} - \mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}})\mathrm{d}v||_{(2,2)}||\mathbf{\Delta_\Gamma}||_2 \\
&= ||\mathbf{\Delta_\Gamma}||_2^2||\int_0^1 (1-v)(\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}} - \mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}})\mathrm{d}v||_{(2,2)} \\
&= ||\mathbf{\Delta_\Gamma}||_2^2\int_0^1 (1-v)||\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}} - \mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}||_{(2,2)}\mathrm{d}v \\
&= [\int_0^1 (1-v)\mathrm{d}v]||\mathbf{\Delta_\Gamma}||_2^2 \max_{0\leq v\leq 1} ||\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}} - \mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}||_{(2,2)} \\
&= \frac{1}{2}||\mathbf{\Delta_\Gamma}||_2^2 \max_{0\leq v\leq 1} ||\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}} - \mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}||_{(2,2)} \\
&\leq \frac{1}{2}||\mathbf{\Delta_\Gamma}||_2^2(p+\#\mathbf{\Gamma}) \max_{0\leq v\leq 1} ||\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{v}\mathbf{\Delta_\Gamma}} - \mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}||_\infty
\end{aligned} \tag{43}$$

The last inequality holds because Hessian matrix of $h_{\mathbf{\Gamma}}$ is of order $(p+\#\mathbf{\Gamma})\times(p+\#\mathbf{\Gamma})$. Woodbury's formula says that if matrix $\mathbf{A}+\mathbf{UCV^t}$ and $\mathbf{C}$ are invertible, then

$$(\mathbf{A}+\mathbf{UCV^t})^{\mathbf{-1}} = \mathbf{A^{-1}} - \mathbf{A^{-1}U}(\mathbf{C^{-1}} + \mathbf{V^tA^{-1}U})^{\mathbf{-1}}\mathbf{V^tA^{-1}} \tag{44}$$

Set $\mathbf{A} = \mathbf{\Omega}^*_{\mathbf{\Gamma}}, \mathbf{U} = v\mathbf{\Delta_\Gamma}$ and $\mathbf{C}, \mathbf{V} = \mathbf{I}$. Then we have

$$(\mathbf{\Omega_\Gamma}^* + v\mathbf{\Delta_\Gamma})^{-1} = \mathbf{\Omega_\Gamma}^{*-1} - v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma}(\mathbf{I} + v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma})^{-1}\mathbf{\Omega_\Gamma}^{*-1} \tag{45}$$

Again applying Cauchy-Shwartz inequality, theorem 1.2 and substituting equation (45), with probability tending to 1,

$$\begin{aligned}
||(\mathbf{\Omega_\Gamma}^* + v\mathbf{\Delta_\Gamma})^{-1} - \mathbf{\Omega_\Gamma}^{*-1}||_\infty &\leq ||(\mathbf{\Omega_\Gamma}^* + v\mathbf{\Delta_\Gamma})^{-1} - \mathbf{\Omega_\Gamma}^{*-1}||_{(2,2)} \\
&= ||\mathbf{\Omega_\Gamma}^{*-1} - v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma}(\mathbf{I} + v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma})^{-1}\mathbf{\Omega_\Gamma}^{*-1} - \mathbf{\Omega_\Gamma}^{*-1}||_{(2,2)} \\
&= ||-v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma}(\mathbf{I} + v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma})^{-1}\mathbf{\Omega_\Gamma}^{*-1}||_{(2,2)} \\
&= ||v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma}(\mathbf{I} + v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma})^{-1}\mathbf{\Omega_\Gamma}^{*-1}||_{(2,2)} \\
&\leq v||\mathbf{\Omega_\Gamma}^{*-1}||_{(2,2)}||\mathbf{\Delta_\Gamma}||_{(2,2)}||(\mathbf{I} + v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma})^{-1}||_{(2,2)}||\mathbf{\Omega_\Gamma}^{*-1}||_{(2,2)} \\
&= v||\mathbf{\Omega_\Gamma}^{*-1}||_{(2,2)}^2||\mathbf{\Delta_\Gamma}||_{(2,2)}||(\mathbf{I} + v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma})^{-1}||_{(2,2)}
\end{aligned} \tag{46}$$

To bound $||(\mathbf{I}+v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma})^{-1}||_{(2,2)}$, $||(\mathbf{I}+v\mathbf{\Omega_\Gamma}^{*-1}(\mathbf{\Omega_\Gamma}-\mathbf{\Omega^*_\Gamma}))^{-1}||_{(2,2)} = ||((1+v)\mathbf{I}+v\mathbf{\Omega^{*-1}_\Gamma}\mathbf{\Omega_\Gamma})^{-1}||_{(2,2)}$.

$$\begin{aligned}||((1+v)\mathbf{I}+v\mathbf{\Omega^{*-1}_\Gamma}\mathbf{\Omega_\Gamma})^{-1}||_{(2,2)} &= ||((1+v)\mathbf{\Omega_\Gamma}+v\mathbf{\Omega_\Gamma}\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Omega_\Gamma})^{-1}\mathbf{\Omega_\Gamma}||_{(2,2)} \\ &\le ||((1+v)\mathbf{\Omega_\Gamma}+v\mathbf{\Omega_\Gamma}\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Omega_\Gamma})^{-1}||_{(2,2)}||\mathbf{\Omega_\Gamma}||_{(2,2)} \\ &= \frac{||\mathbf{\Omega_\Gamma}||_{(2,2)}}{||(1+v)\mathbf{\Omega_\Gamma}+v\mathbf{\Omega_\Gamma}\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Omega_\Gamma}||_{(2,2)}}\end{aligned} \tag{47}$$

Note $\mathbf{\Omega_\Gamma}, \mathbf{\Omega_\Gamma}\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Omega_\Gamma}$ are positive definite. Recall that if $\mathbf{A}\in\mathcal{M}$, $\text{eig}_1(\mathbf{A}) = \min_{||\mathbf{x}||_2=1}\mathbf{x^t}\mathbf{A}\mathbf{x}$, $\text{eig}_p(\mathbf{A}) = \max_{||\mathbf{x}||_2=1}\mathbf{x^t}\mathbf{A}\mathbf{x}$ and clearly $\mathbf{x^t}\mathbf{A}\mathbf{x}\ge 0$ for any $\mathbf{x}\in\mathbb{R}^p$. Thus, $^\forall\mathbf{x}$ such that $||\mathbf{x}||_2 = 1$,

$$(1+v)\mathbf{x^t}\mathbf{\Omega_\Gamma}\mathbf{x} \le= \mathbf{x^t}((1+v)\mathbf{\Omega_\Gamma}+v\mathbf{\Omega_\Gamma}\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Omega_\Gamma})\mathbf{x} = (1+v)\mathbf{x^t}\mathbf{\Omega_\Gamma}\mathbf{x}+v\mathbf{x^t}(\mathbf{\Omega_\Gamma}\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Omega_\Gamma})\mathbf{x} \tag{48}$$

Consequently, $\text{eig}_p(\mathbf{\Omega_\Gamma}) = ||(1+v)\mathbf{\Omega_\Gamma}||_{(2,2)} \le ||(1+v)\mathbf{\Omega_\Gamma}+v\mathbf{\Omega_\Gamma}\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Omega_\Gamma}||_{(2,2)} = \text{eig}_p((1+v)\mathbf{\Omega_\Gamma}+v\mathbf{\Omega_\Gamma}\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Omega_\Gamma})$. Using this inequality, from (47),

$$\begin{aligned}||((1+v)\mathbf{I}+v\mathbf{\Omega^{*-1}_\Gamma}\mathbf{\Omega_\Gamma})^{-1}||_{(2,2)} &\le \frac{||\mathbf{\Omega_\Gamma}||_{(2,2)}}{||(1+v)\mathbf{\Omega_\Gamma}+v\mathbf{\Omega_\Gamma}\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Omega_\Gamma}||_{(2,2)}} \\ &\le \frac{||\mathbf{\Omega_\Gamma}||_{(2,2)}}{||(1+v)\mathbf{\Omega_\Gamma}||_{(2,2)}} = \frac{1}{1+v} \le 1\ (\because 0\le v\le 1)\end{aligned} \tag{49}$$

Therefore, applying this result to (46),

$$\begin{aligned}||(\mathbf{\Omega_\Gamma}^*+v\mathbf{\Delta_\Gamma})^{-1}-\mathbf{\Omega_\Gamma}^{*-1}||_\infty &\le v||\mathbf{\Omega_\Gamma}^{*-1}||^2_{(2,2)}||\mathbf{\Delta_\Gamma}||_{(2,2)}||(\mathbf{I}+v\mathbf{\Omega_\Gamma}^{*-1}\mathbf{\Delta_\Gamma})^{-1}||_{(2,2)} \\ &\le v||\mathbf{\Omega_\Gamma}^{*-1}||^2_{(2,2)}||\mathbf{\Delta_\Gamma}||_{(2,2)} \\ &\le \mathrm{K}||\mathbf{\Delta_\Gamma}||_{(2,2)} \\ &\le \mathrm{K}||\mathbf{\Delta_\Gamma}||_2\end{aligned} \tag{50}$$

Here K is some constant. $v||\mathbf{\Omega_\Gamma}^{*-1}||^2_{(2,2)}$ can be bounded by some constant K because $v\in[0,1]$ and we assumed $||\mathbf{\Omega_\Gamma}^{*-1}||_{(2,2)}$ to be bounded by some constants.

For any symmetric matrix $\mathbf{A} = ((a_{ij}))\in\mathcal{M}$, because $\mathbf{tr}(\mathbf{A}\mathbf{E}_{(i,j)}\mathbf{A}\mathbf{E}_{(l,m)}) = \sum_{c,d,e,f=1}^{p} a_{cd}(\mathbf{E}_{(i,j)})_{de}a_{ef}(\mathbf{E}_{(l,m)})_{fc}$,

$$\mathbf{tr}(\mathbf{A}\mathbf{E}_{(i,j)}\mathbf{A}\mathbf{E}_{(l,m)}) = \begin{cases} a_{mi}a_{jl}+a_{li}a_{jm}+a_{mj}a_{il}+a_{lj}a_{im} = 2a_{il}a_{jm}+2a_{im}a_{jl}, & i\ne j, l\ne m \\ a_{mi}a_{il}+a_{li}a_{im} = 2a_{im}a_{lm}, & i=j, l\ne m \\ 2a_{li}a_{lj} & i\ne j, l=m \\ a_{il}a_{li} = a^2_{il}, & i=j, l=m\end{cases} \tag{51}$$

Simply, let $(\mathbf{\Omega_\Gamma}^*)^{-1} = ((a_{ij}))$ and $(\mathbf{\Omega_\Gamma}^*+v\mathbf{\Delta_\Gamma})^{-1} = ((b_{ij}))$. From equation (51) and the definition of $\mathbf{H_B}$, the entries of $\mathbf{H}_{\mathbf{\Omega^*_\Gamma}+\mathbf{v\Delta_\Gamma}}-\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}$ have the respective forms $2b_{il}b_{jm}+2b_{im}b_{jl}-2a_{il}a_{jm}-2a_{im}a_{jl}$, $2b_{im}b_{lm}-2a_{im}a_{lm}$, $2b_{li}b_{lj}-2a_{li}a_{lj}$, and $b^2_{il}-a^2_{il}$. In (50), we have $|b_{ij}-a_{ij}|\le||(\mathbf{\Omega_\Gamma}^*+v\mathbf{\Delta_\Gamma})^{-1}-\mathbf{\Omega_\Gamma}^{*-1}||_\infty\le \mathrm{K}||\mathbf{\Delta_\Gamma}||_2$. Considering form of entries of $\mathbf{H}_{\mathbf{\Omega^*_\Gamma}+\mathbf{v\Delta_\Gamma}}-\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}$, with probability tending to 1, for all $((i,j),(l,m))$, there exist some positive constants $\mathrm{C}_1, \mathrm{C}_2$ such that

$$|\sum b_{il}b_{jm}-\sum a_{il}a_{jm}|\le \mathrm{C}_1||\mathbf{\Delta_\Gamma}||_2+\mathrm{C}_2||\mathbf{\Delta_\Gamma}||^2_2 \tag{52}$$

Since this holds for all $((i,j),(l,m))$, $||\mathbf{H}_{\mathbf{\Omega^*_\Gamma}+\mathbf{v\Delta_\Gamma}}-\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}||_\infty\le \mathrm{C}_1||\mathbf{\Delta_\Gamma}||_2+\mathrm{C}_2||\mathbf{\Delta_\Gamma}||^2_2$
Combining the results of inequality (52) and inequality (43),

$$|\mathrm{R}_n|\le\frac{1}{2}(p+\#\mathbf{\Gamma})||\mathbf{\Delta_\Gamma}||^2_2(\mathrm{C}_1||\mathbf{\Delta_\Gamma}||_2+\mathrm{C}_2||\mathbf{\Delta_\Gamma}||^2_2) \tag{53}$$

,which is the result of lemma 3.3. □

Using the result of lemma 3.3., under some condition, we achieves the goal of this section, which is to show the error in Laplace approximation of posterior probability of regular model $\mathbf{\Gamma}$ is very small.

Theorem 3.3. *The error in the Laplace approximation tends to 0 in probability if* $(p + \#\mathbf{\Gamma})^2\epsilon_n = n^{-\frac{1}{2}}(p + \#\mathbf{\Gamma})^{\frac{5}{2}}(\log p)^{\frac{1}{2}} \to 0$ *in probability, hence asymptotically negligible, where* $\epsilon_n = n^{-\frac{1}{2}}(p + \#\mathbf{\Gamma})^{\frac{1}{2}}(\log p)^{\frac{1}{2}}$*, which is posterior convergence rate established in theorem 2.3.*

Proof. Using the Taylor's expansion of $h_{\mathbf{\Gamma}}$ in equation (40), equation (22) can be written as

$$\mathbf{p}(\mathbf{\Gamma}|\mathbf{X^{(n)}}) \propto \mathbf{C_\Gamma}\int_{\mathbf{\Omega^*_\Gamma}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+} \exp(-\frac{n}{2}(h_{\mathbf{\Gamma}}(\mathbf{\Omega^*_\Gamma}) + \frac{1}{2}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) + \mathrm{R}_n)) \prod_{(i,j)\in\bar{\mathbf{E}}_{\mathbf{\Gamma}}} \mathrm{d}u_{\mathbf{\Gamma},ij} \tag{54}$$

For notational simplicity, denote $\prod_{(i,j)\in\bar{\mathbf{E}}} \mathrm{d}u_{\mathbf{\Gamma},ij}$ by $\mathrm{d}\mathbf{\Delta_\Gamma}$. Want to show:

$$\begin{aligned}
&\frac{\int_{\mathbf{\Omega^*_\Gamma}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+} \exp(-\frac{n}{2}(h_{\mathbf{\Gamma}}(\mathbf{\Omega^*_\Gamma}) + \frac{1}{2}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) + \mathrm{R}_n))\mathrm{d}u_{\mathbf{\Gamma},ij}}{\int_{\mathbf{\Omega^*_\Gamma}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+} \exp(-\frac{n}{2}(h_{\mathbf{\Gamma}}(\mathbf{\Omega^*_\Gamma}) + \frac{1}{2}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma})))\mathrm{d}u_{\mathbf{\Gamma},ij}} \\
&\quad = \frac{\int_{\mathbf{\Omega^*_\Gamma}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) - \frac{1}{2}\mathrm{R}_n)\mathrm{d}u_{\mathbf{\Gamma},ij}}{\int_{\mathbf{\Omega^*_\Gamma}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}))\mathrm{d}u_{\mathbf{\Gamma},ij}} \to 1
\end{aligned} \tag{55}$$

By equation (20),

$$\begin{aligned}
&\frac{\int_{||\mathbf{\Delta_\Gamma}||_2\le\epsilon_n} \exp(-\frac{n}{2}(h_{\mathbf{\Gamma}}(\mathbf{\Omega^*_\Gamma}) + \frac{1}{2}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) + \mathrm{R}_n))\mathrm{d}u_{\mathbf{\Gamma},ij}}{\int_{\mathbf{\Omega^*_\Gamma}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+} \exp(-\frac{n}{2}(h_{\mathbf{\Gamma}}(\mathbf{\Omega^*_\Gamma}) + \frac{1}{2}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) + \mathrm{R}_n))\mathrm{d}u_{\mathbf{\Gamma},ij}} \\
&\quad = \frac{\int_{||\mathbf{\Delta_\Gamma}||_2\le\epsilon_n} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) - \frac{n}{2}\mathrm{R}_n)\mathrm{d}u_{\mathbf{\Gamma},ij}}{\int_{\mathbf{\Omega^*_\Gamma}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) - \frac{n}{2}\mathrm{R}_n)\mathrm{d}u_{\mathbf{\Gamma},ij}} \to 1
\end{aligned} \tag{56}$$

So we have $\int_{\mathbf{\Omega^*_\Gamma}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) - \frac{n}{2}\mathrm{R}_n)\mathrm{d}u_{\mathbf{\Gamma},ij} = \int_{||\mathbf{\Delta_\Gamma}||_2\le\epsilon_n} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) - \frac{n}{2}\mathrm{R}_n)\mathrm{d}u_{\mathbf{\Gamma},ij} + \mathrm{o}(1)$. Now we give estimates for $\int_{||\mathbf{\Delta_\Gamma}||_2\le\epsilon_n} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) - \frac{n}{2}\mathrm{R}_n)\mathrm{d}u_{\mathbf{\Gamma},ij}$. Take sufficiently large n so that ϵ_n becomes sufficiently small. If $||\mathbf{\Delta_\Gamma}||_2 \le \epsilon_n$, because of sufficiently small ϵ_n and the result of lemma 3.3, there exists a constant $\mathrm{C} > 0$ such that $|\mathrm{R}_n| \le \frac{1}{2}\mathrm{C}(p + \#\mathbf{\Gamma})||\mathbf{\Delta_\Gamma}||_2^2\epsilon_n$. Thus the upper and lower bounds of the integral $\int_{||\mathbf{\Delta_\Gamma}||_2\le\epsilon_n} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) - \frac{n}{2}\mathrm{R}_n)\mathrm{d}u_{\mathbf{\Gamma},ij}$ are given by

$$\begin{aligned}
&\int_{||\mathbf{\Delta_\Gamma}||_2\le\epsilon_n} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) \pm \frac{n}{4}\mathrm{C}(p + \#\Gamma)||\mathbf{\Delta_\Gamma}||_2^2\epsilon_n)\mathrm{d}u_{\mathbf{\Gamma},ij} \\
\Leftrightarrow &\int_{||\mathbf{\Delta_\Gamma}||_2\le\epsilon_n} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}\mathrm{vec}(\mathbf{\Delta_\Gamma}) \pm \frac{n}{4}\mathrm{C}(p + \#\Gamma)\epsilon_n\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathrm{vec}(\mathbf{\Delta_\Gamma}))\mathrm{d}u_{\mathbf{\Gamma},ij} \\
\Leftrightarrow &\int_{||\mathbf{\Delta_\Gamma}||_2\le\epsilon_n} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t(\mathbf{H}_{\mathbf{\Omega^*_\Gamma}} \mp \mathrm{C}(\mathbf{p} + \#\mathbf{\Gamma})\epsilon_\mathbf{n}\mathbf{I})\mathrm{vec}(\mathbf{\Delta_\Gamma}))\mathrm{d}u_{\mathbf{\Gamma},ij}
\end{aligned} \tag{57}$$

Suppose $(p+\#\mathbf{\Gamma})\epsilon_n \to 0$. Since the minimum eigenvalue of $\mathbf{H}_{\mathbf{\Omega^*_\Gamma}}$ is bounded away from 0, as a consequence of theorem 3.1, on the domain $\{\mathbf{\Delta_\Gamma} : ||\mathbf{\Delta_\Gamma}||_2 > 0\}$, $\exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t(\mathbf{H}_{\mathbf{\Omega^*_\Gamma}} \mp \mathrm{C}(\mathbf{p} + \#\mathbf{\Gamma})\epsilon_\mathbf{n}\mathbf{I})\mathrm{vec}(\mathbf{\Delta_\Gamma}))$ uniformly converges to 0. Therefore,

$$\int_{||\mathbf{\Delta_\Gamma}||_2>\epsilon_n} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t(\mathbf{H}_{\mathbf{\Omega^*_\Gamma}} \mp \mathrm{C}(\mathbf{p} + \#\mathbf{\Gamma})\epsilon_\mathbf{n}\mathbf{I})\mathrm{vec}(\mathbf{\Delta_\Gamma}))\mathrm{d}u_{\mathbf{\Gamma},ij} \to 0. \tag{58}$$

So $\int_{||\mathbf{\Delta_\Gamma}||_2\le\epsilon_n} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t(\mathbf{H}_{\mathbf{\Omega^*_\Gamma}} \mp \mathrm{C}(\mathbf{p} + \#\mathbf{\Gamma})\epsilon_\mathbf{n}\mathbf{I})\mathrm{vec}(\mathbf{\Delta_\Gamma}))\mathrm{d}u_{\mathbf{\Gamma},ij}$ can be simplified into $\int_{\mathbf{\Delta_\Gamma}+\mathbf{\Omega^*_\Gamma}\in\mathcal{M}_0^+} \exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t(\mathbf{H}_{\mathbf{\Omega^*_\Gamma}} \mp \mathrm{C}(\mathbf{p} + \#\mathbf{\Gamma})\epsilon_\mathbf{n}\mathbf{I})\mathrm{vec}(\mathbf{\Delta_\Gamma}))\mathrm{d}u_{\mathbf{\Gamma},ij}$. Hence, the ratio in equation (55) can be approximately bounded by

$$\frac{\int_{\mathbf{\Delta_\Gamma}+\mathbf{\Omega}^*_{\mathbf{\Gamma}}\in\mathcal{M}_0^+}\exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t(\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}\mp \mathrm{C}(\mathbf{p}+\#\mathbf{\Gamma})\epsilon_{\mathbf{n}}\mathbf{I})\mathrm{vec}(\mathbf{\Delta_\Gamma}))du_{\mathbf{\Gamma},ij}}{\int_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}+\mathbf{\Delta_\Gamma}\in\mathcal{M}_0^+}\exp(-\frac{n}{4}\mathrm{vec}(\mathbf{\Delta_\Gamma})^t\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}\mathrm{vec}(\mathbf{\Delta_\Gamma}))du_{\mathbf{\Gamma},ij}} = \frac{[\det(2\pi\frac{4}{n}(\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}\mp \mathrm{C}(\mathbf{p}+\#\mathbf{\Gamma})\epsilon_{\mathbf{n}}\mathbf{I})^{-1})]^{\frac{1}{2}}}{[\det(2\pi\frac{4}{n}(\mathbf{H}^*_{\mathbf{\Omega_\Gamma}})^{-1})]^{\frac{1}{2}}}$$

$$= [\frac{\det(\mathbf{H}_{\mathbf{\Omega}^*_{\mathbf{\Gamma}}}\mp \mathrm{C}(\mathbf{p}+\#\mathbf{\Gamma})\epsilon_{\mathbf{n}}\mathbf{I}))}{\det(\mathbf{H}^*_{\mathbf{\Omega_\Gamma}})}]^{-\frac{1}{2}}$$

$$= [\det(\mathbf{I}\mp \mathrm{C}(p+\#\mathbf{\Gamma})\epsilon_n(\mathbf{H}_{\mathbf{\Omega_\Gamma}}{}^*)^{-1}]^{-\frac{1}{2}} \tag{59}$$

The equation (59) must lie between $[1\mp \mathrm{C}(p+\#\mathbf{\Gamma})\epsilon_n\{\mathrm{eig}_1(\mathbf{H}_{\mathbf{\Omega_\Gamma}}{}^*)\}^{-1}]^{-\frac{p+\#\Gamma}{2}}$, as the dimension of $\mathbf{H}_{\mathbf{\Omega_\Gamma}}{}^*$ is $p+\#\Gamma$. Again by theorem 3.1, the minimum eigenvalue of $\mathbf{H}_{\mathbf{\Omega_\Gamma}}{}^*$ is bounded away from 0. So if $(p+\#\mathbf{\Gamma})^2\epsilon_n \to 0$, the above bound on the ratio goes to 1. Thus, we conclude that the error in Laplace approximation is asymptotically marginal.

□

4 Simulation results

In this section, we simulate Bayesian graphical lasso. We provide specific scheme in the below. We use models with sparse precision matrix for simulation. Note the mean of models are $\mathbf{0}$. The models with precision matrix $\mathbf{\Omega} = ((\omega_{ij}))$ or covariance matrix $\mathbf{\Sigma} = ((\sigma_{ij}))$ used in the simulation are followings:

- **Model1 :** AR(1) model, $\sigma_{ij} = 0.7^{|i-j|}$
- **Model2 :** AR(2) model, $\omega_{ii} = 1, \omega_{i-1,i} = \omega_{i,i-1} - 0.5, \omega_{i,i-2} = \omega_{i-2,i} = 0.25$
- **Model3 :** Star model, $\omega_{ii} = 1, \omega_{1,i} = \omega_{i,1} = 0.1$ and $\omega_{ij} = 0$, otherwise.
- **Model4:** Circle model, $\omega_{ii} = 2, \omega_{i-1,i} = \omega_{i,i-1} = 1, \omega_{1,p} = \omega_{p,1} = 0.9$

We use specificity(SP), sensitivity(SE), and Matthews Correlation Coefficient(MCC) to measure the performance of the algorithm which will be described in the below. Here, specificity and sensitivity are defined in the context of edge.

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}},\ \mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}},\ \mathrm{MCC} = \frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}} \tag{60}$$

where TP,TN,FP, and FN respectively denote the true positives(edges included which are present in the true model), true negatives(edges excluded which are absent in the true model), false positives(edges included which are absent in the true model), and false negatives(edges excluded which are present in the true model). Note that for estimated $\hat{\mathbf{\Omega}}$, we say there is no edge between node i and j if $|\hat{\omega}_{ij}| \le 0.1^3$. To compare Bayesian graphical lasso described in this article with frequentist graphical lasso, we also do simulation for frequentist graphical lasso with the same models.

Corresponding to each model, we generate samples of size $n = 100, 200$ and dimension $p = 30, 50, 100$. The penalty parameter λ for the graphical lasso algorithm is chosen such that $\frac{\lambda}{n} = 0.5$ for model 1,2,and 4, and $\frac{\lambda}{n} = 0.2$ for model 3. By theorem 3.2, we can concentrate only on regular models if $q < \frac{1}{2}$. Therefore, we set $q = 0.4$. We may choose large q as possible, because since we allow less edges to enter the model as q gets lower, it is likely that sensitivity is very low. We run 50 replications of each model and calculate SP, SE, and MCC in each replication. We average each measure over the replications and calculate standard deviation of each measure also. We also show ROC curves corresponding to the various models with values of the penalty parameter ranging between 0 1, plotting Sensitivity against False Positive Rate in case $n = 100$ and $p = 30, 50$.

For Bayesian graphical lasso, we shall deal with the problem discussed in the end of section 3.1. Note that we have set the restriction of model size $\bar{\mathbf{r}}$ by that of frequentist graphical lasso with the same penalty parameter. When λ is chosen so that $\frac{\lambda}{n} = 0.5$, we have $\bar{\mathbf{r}} = 29$ in model1 with dimension $p = 30$. As discussed in the end of section 3.1, we saw that there are at least 10^{45} models to search over, which is impossible in practice. Hence we suggest Metropolis-Hastings algorithm so that we obtain MCMC samples of model indicator $\mathbf{\Gamma}$. In each replication of model, we choose final model by Median Probability Model(denoted by 'MPP'). To implement Metropolis-Hastings algorithm, we may suggest symmetric proposal distribution.

Suppose there are two p-dimension models $G = (V, E)$ and $G' = (V, E')$ with model indicators $\mathbf{\Gamma}, \mathbf{\Gamma}'$ respectively. Considering $\mathbf{\Gamma}$ as vector, we denote $\mathbf{\Gamma}(i,j) = 1$ if there is edge between node i and j and $\mathbf{\Gamma}(i,j) = 0$ otherwise. Let $\pi(\mathbf{\Gamma}|\mathbf{\Gamma}')$ symmetric proposal distribution, which samples G' uniformly from the graphs that differ from G in one position. To be more specific, we set $\mathbf{\Gamma}'(i,j) = 1$ for one pair (i,j) uniformly sampled from $V \times V \setminus E$, i.e., $\mathbf{\Gamma}(i,j) = 0$ and $\mathbf{\Gamma}'(i,j) = 0$ for one pair (i,j) uniformly sampled from E, i.e., $\mathbf{\Gamma}(i,j) = 1$. So we have equal probability to set $\mathbf{\Gamma}'(i,j) = 1$ or $\mathbf{\Gamma}'(i,j) = 0$. Set $\mathbf{\Gamma}'(-i,-j) = \mathbf{\Gamma}(i,j)$, i.e. set Γ' as the same as $\mathbf{\Gamma}$ except for pair (i,j). Using this symmetric proposal distribution, we have following algorithm for MCMC sampling of $\mathbf{\Gamma}$.

Algorithm 1 Metropolis-Hastings algorithm for sampling $\mathbf{\Gamma}$

1: Initial model indicator : $\mathbf{\Gamma}^{(0)}$, Given data : $\mathbf{X}^{(\mathbf{n})}$
2: **for** i=0,1,...,k **do**
3: $\quad \mathbf{\Gamma}^{\text{temp}} \sim \pi(\mathbf{\Gamma}|\mathbf{\Gamma}^{(i)})$
4: $\quad$ Check regularity of $\mathbf{\Gamma}^{\text{temp}}$
5: $\quad$ If $\mathbf{\Gamma}^{\text{temp}}$ is regular, $\mathbf{\Gamma}^{\text{cand}} = \mathbf{\Gamma}^{\text{temp}}$. Else, repeat $3 \sim 4$.
6: $\quad \alpha_i = \min\{1, \frac{\mathbf{p}^*(\mathbf{\Gamma}^{\text{cand}}|\mathbf{X}^{(\mathbf{n})})}{\mathbf{p}^*(\mathbf{\Gamma}^{(\mathbf{i})}|\mathbf{X}^{(\mathbf{n})})}\}$
7: $\quad U_i \sim U(0,1)$
8: $\quad$ If $U_i \leq \alpha_i$, $\mathbf{\Gamma}^{(i+1)} = \mathbf{\Gamma}^{\text{cand}}$. Else, $\mathbf{\Gamma}^{(i+1)} = \mathbf{\Gamma}^{(i)}$.
9: **end for**

We replicate this algorithm for 50 times. In each replication, we samples 500 $\mathbf{\Gamma}$ with 100 burn-in in the algorithm and select final model by MPP, which we've explained in the above. Though this number of iteration is not sufficient to get desired effective sample size, due to difficulty of implementation in practice, we samples only 500. The result of simulation is given in Table 1 and ROC curves are given in Figure 1,2. Note we denote frequentist graphical lasso by 'GL' and Bayesian graphical lasso by 'MPP'.

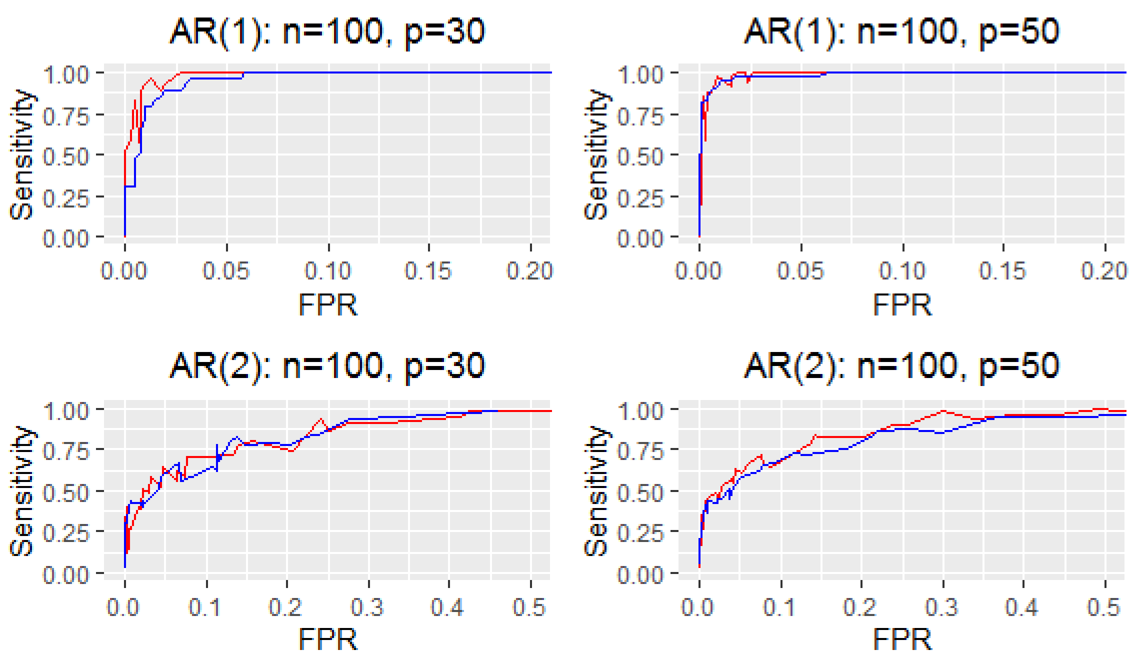

Figure 1: ROC curves for AR(1) and AR(2) structures. Red line : GL, Blue line: MPP

Model	p	n=100						n=200					
		MPP			GL			MPP			GL		
		SP	SE	MCC	SP	SE	MCC	SP	SE	MCC	SP	SE	MCC
AR(1)	30	0.977	0.919	0.813	0.977	0.970	0.844	0.980	0.972	0.861	0.982	0.992	0.886
		(0.009)	(0.054)	(0.048)	(0.011)	(0.045)	(0.055)	(0.007)	(0.038)	(0.041)	(0.007)	(0.012)	(0.035)
	50	0.986	0.893	0.802	0.985	0.969	0.839	0.990	0.962	0.869	0.990	0.992	0.888
		(0.004)	(0.044)	(0.038)	(0.005)	(0.038)	(0.041)	(0.003)	(0.032)	(0.032)	(0.004)	(0.011)	(0.035)
	100	0.994	0.872	0.804	0.993	0.957	0.837	0.995	0.959	0.878	0.995	0.994	0.888
		(0.001)	(0.038)	(0.030)	(0.002)	(0.029)	(0.029)	(0.001)	(0.027)	(0.028)	(0.001)	(0.010)	(0.022)
AR(2)	30	0.983	0.434	0.546	0.978	0.475	0.562	0.988	0.485	0.613	0.988	0.488	0.614
		(0.007)	(0.050)	(0.046)	(0.010)	(0.044)	(0.038)	(0.007)	(0.031)	(0.029)	(0.006)	(0.029)	(0.029)
	50	0.987	0.409	0.518	0.983	0.483	0.559	0.993	0.464	0.609	0.993	0.485	0.621
		(0.004)	(0.034)	(0.032)	(0.005)	(0.032)	(0.026)	(0.003)	(0.024)	(0.026)	(0.003)	(0.023)	(0.027)
	100	0.991	0.400	0.499	0.989	0.475	0.536	0.997	0.442	0.607	0.996	0.486	0.630
		(0.002)	(0.023)	(0.022)	(0.002)	(0.026)	(0.022)	(0.001)	(0.019)	(0.017)	(0.001)	(0.017)	(0.013)
Star	30	0.965	0.229	0.225	0.954	0.297	0.256	0.990	0.167	0.263	0.994	0.237	0.384
		(0.006)	(0.077)	(0.080)	(0.013)	(0.104)	(0.098)	(0.006)	(0.093)	(0.122)	(0.004)	(0.084)	(0.109)
	50	0.965	0.319	0.264	0.948	0.479	0.331	0.994	0.387	0.516	0.994	0.498	0.606
		(0.007)	(0.074)	(0.061)	(0.009)	(0.104)	(0.067)	(0.002)	(0.078)	(0.074)	(0.003)	(0.088)	(0.061)
	100	0.940	0.992	0.484	0.940	1.000	0.489	0.988	1.000	0.793	0.988	1.000	0.787
		(0.005)	(0.010)	(0.018)	(0.004)	(0.000)	(0.013)	(0.002)	(0.000)	(0.019)	(0.002)	(0.000)	(0.023)
Circle	30	0.784	1.000	0.448	0.746	1.000	0.411	0.754	1.000	0.418	0.724	1.000	0.392
		(0.021)	(0.000)	(0.022)	(0.015)	(0.000)	(0.013)	(0.013)	(0.000)	(0.012)	(0.018)	(0.000)	(0.015)
	50	0.828	0.976	0.395	0.830	1.000	0.408	0.836	0.996	0.413	0.834	1.000	0.412
		(0.010)	(0.019)	(0.016)	(0.009)	(0.000)	(0.011)	(0.007)	(0.009)	(0.010)	(0.006)	(0.000)	(0.008)
	100	0.892	0.984	0.371	0.891	1.000	0.376	0.905	0.997	0.402	0.904	1.000	0.400
		(0.003)	(0.012)	(0.008)	(0.004)	(0.000)	(0.006)	(0.003)	(0.006)	(0.007)	(0.003)	(0.000)	(0.006)

Table 1: Simulation results. Figures in parentheses indicate standard errors.

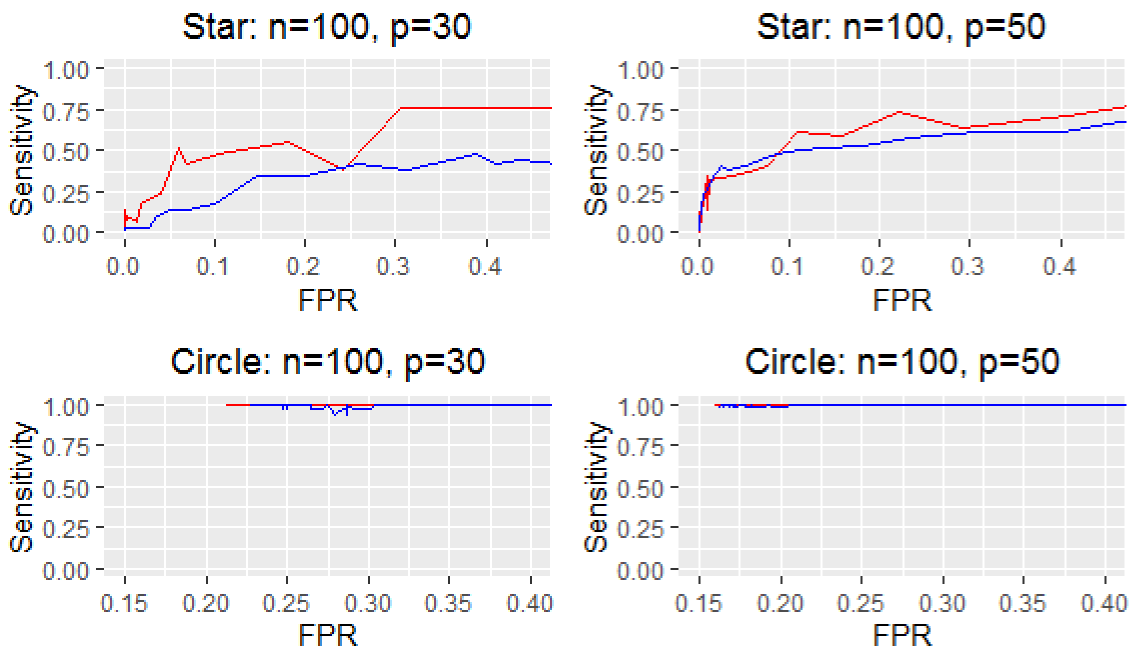

Figure 2: ROC curves for Star and Circle. Red line : GL, Blue line: MPP

From simulation results in table 1, one can see that Bayesian graphical lasso performs slightly better than the frequentist graphical lasso in terms of specificity, but suffers in sensitivity. This is because by choosing $q < \frac{1}{2}$, we focused only on regular models to make Laplace approximation valid and the smaller q tends to allow less edges to enter the model. Performance of Bayesian graphical lasso gets worse as the dimension p gets larger in all models. The sensitivity was not good for both frequentist GL and Bayesian GL in AR(2) and Star models. Also, in ROC curves, as the penalty parameter λ tends to get larger, sensitivity grows at the cost of higher false positive rate.

To mention one of drawbacks of Bayesian graphical lasso, it is very slower than frequentist graphical lasso. There are two possible reasons. First, in each iteration, we check regularity condition and sample $\mathbf{\Gamma}$ again until we have regular model. Also, in estimating precision matrix $\mathbf{\Omega_\Gamma}$ corresponding to regular model $\mathbf{\Gamma}$, we use frequentist graphical lasso. The algorithm of it requires very long time as the dimension grows especially in sparse models. Hence the algorithm of Bayesian graphical lasso is not efficient. Note that it took at least 12 hours to sample total 25000 model indicators $\mathbf{\Gamma}$ in all models with $p = 100$.

To verify that Bayesian graphical lasso indeed performs better than frequentist graphical lasso in terms of specificity, but suffers in sensitivity, we may also simulate with real data. We used protein data from Rpackage sparsebn. This data consists of n=7566 observations of p=11 continuous variables corresponding to different proteins and phospholipids in human immune system cells and each observation indicates the measured level of each biomolecule in a single cell under different experimental interventions. We may compare graphs obtained from Bayesian graphical lasso, frequentist graphical lasso with true network. Here we set penalty parameter $\lambda = 0.4$ and acceptance probability of edges q=0.4. For Bayesian GL, we obtain 10000 mcmc samples with 2000 burn-in. The performance of each method and estimated graphs are given as following :

Method	SP	SE	MCC
MPP	0.636	0.364	0.000
GL	0.606	0.455	0.0602

Table 2: Simulation result for protein data

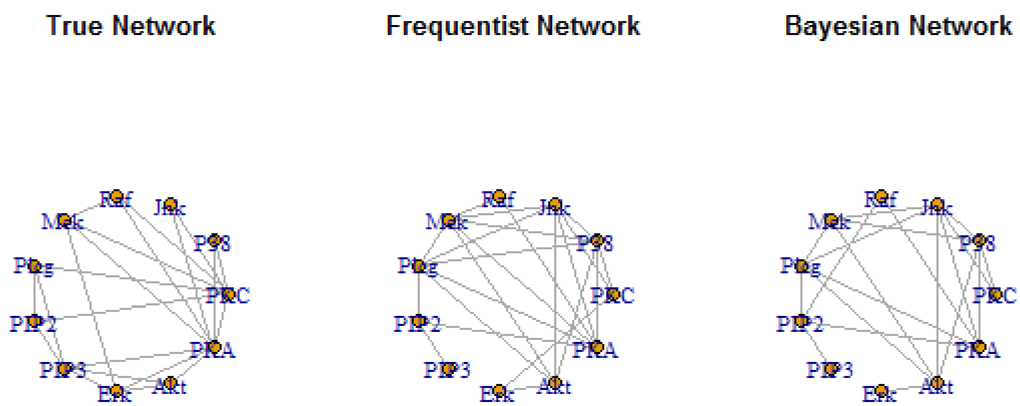

Figure 3: True network, estimated networks for protein data

From the graphs above, the model estimated by Bayesian graphical lasso shows less edges, hence suffers in sensitivity. However, one cannot say frequentist graphical lasso performs good in that specificity is low compared to Bayesian graphical lasso and more edges are connected to Jnk, AKt in the model estimated by frequentist graphical lasso than true network. Along with table 2, one can see what we wanted to verify with real data.

The method suggested in this article seems very brilliant in that it introduced sparsity to models, which was shortcoming of existing Bayesian methods in estimating graphical model, and it significantly reduced the number of models to search over. However, this was possible at cost of sensitivity and time. One may hope this method to be improved in sensitivity and efficiency.