Final Project - Bayesian Inference for Gaussian Graphical Models

Gaussian Graphical Models (GMM) Gaussian Graphical Models (GGMs) are powerful probabilistic tools for modeling the conditional dependencies among variables in multivariate Gaussian distributions. These dependencies are encoded using an **undirected graph**, where each node represents a random variable, and the absence of an edge between two nodes indicates *conditional independence* between the corresponding variables, given all others. This graphical structure makes GGMs particularly valuable for analyzing complex relationships in high-dimensional datasets, such as protein expression profiles. For instance, in the context of protein regulatory networks, GGMs can be applied to model interactions among proteins based on expression data from patients. In this setup, each protein is represented as a node, and the absence of an edge indicates no direct regulatory relationship between the proteins. The strength of an edge, if present, quantifies the magnitude of the regulatory relationship between the proteins.

Formally, a **Gaussian graphical model** for a random p-dimensional vector $\boldsymbol{Y} = (Y^1, \ldots, Y^p)$, where superscripts denote features and subscripts (introduced later) will index samples, is defined by a tuple

$$\mathcal{G}_{\boldsymbol{Y}} = \{G, \mathcal{P}(\boldsymbol{Y})\},$$

where $G = (V, E)$ is an undirected graph representing the structure of conditional independence, and $\mathcal{P}(\boldsymbol{Y})$ is the associated probability distribution. The node set is defined as $V = \{1, 2, \ldots, p\}$, and the edge set $E \subseteq V \times V$ determines which variable pairs are conditionally dependent. An edge $(i, j) \in E$ exists if and only if variables i and j are not conditionally independent given all other variables. The distribution $\mathcal{P}(\boldsymbol{Y})$ is assumed to be multivariate normal with mean $\boldsymbol{\mu} \in \mathbb{R}^p$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$, i.e.,

$$\boldsymbol{Y}_n \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}), \quad n = 1, \ldots, N,$$

where $\boldsymbol{\Omega} = (\omega^{ij})$ is the **precision matrix**, which is the inverse of the covariance matrix. If the entry $\omega^{ij} = 0$, then the variables i and j are conditionally independent given all other variables. This implies that there is no edge between nodes i and j in the graph G. Therefore, the conditional independence structure of G can be inferred by identifying the zero pattern in the precision matrix $\boldsymbol{\Omega}$. In the graph, the strength of the edge between nodes i and j is quantified by the partial correlation $\rho^{ij} \neq 0$ between the random variables Y^i and Y^j, where $\rho^{ij} = -\frac{\omega^{ij}}{\sqrt{\omega^{ii}\omega^{jj}}}$. For simplicity, in this project, we assume $\boldsymbol{\mu} = 0$ and focus solely on the estimation of $\boldsymbol{\Omega}$.

Pseudo Likelihood [3] A straightforward method to estimate $\boldsymbol{\Omega}$ is through node-wise regression, where each ω^{ij} is estimated by regressing Y^i against all other variables. When Y^i $(1 \leq i \leq p)$ is expressed in linear regression form, we have

$$Y^i = \sum_{j \neq i} \theta^{ij} Y^j + \epsilon^i, \quad i = 1, \ldots, p$$

$$\theta^{ij} = -\frac{\omega^{ij}}{\omega^{ii}}, \quad \epsilon^i \sim \mathcal{N}\left(0, \frac{1}{\omega^{ii}}\right)$$

Finally, a pseudo-likelihood function for all parameters can be derived in terms of Y^i, which is governed by a Gaussian distribution. This pseudo-likelihood also defines the full conditional distribution of Y^i:

$$Y^i | \boldsymbol{Y^{-i}}, \boldsymbol{\omega^{i,-i}}, \omega^{ii} \sim \mathcal{N}\left(-\frac{\sum_{j \neq i}^{p} \omega^{ij} y^j}{\omega^{ii}}, \frac{1}{\omega^{ii}}\right)$$

Joint regression [6] One limitation of node-wise regression is that the estimated values of ω^{ij} and ω^{ji} may differ, even though they should be equal due to the symmetry of the precision matrix $\boldsymbol{\Omega}$. To resolve this inconsistency, a joint regression approach has been proposed within the Bayesian framework. This method enforces the constraint $\omega^{ij} = \omega^{ji}$ by jointly modeling the regressions for nodes i and j during the sampling process. Specifically, the full conditional distribution of ω^{ij} is derived by combining the two pseudo-likelihoods corresponding to Y^i and Y^j, leading to:

$$f(\omega^{ij} \mid \cdot) \propto f(Y^i \mid \boldsymbol{Y}^{-i}, \omega^{i,-(i,j)}, \omega^{ij}, \omega^{ii}) \times f(Y^j \mid \boldsymbol{Y}^{-j}, \omega^{j,-(i,j)}, \omega^{ij}, \omega^{jj}) \times \text{prior}(\omega^{ij})$$

This formulation ensures symmetry is preserved in the estimation of the precision matrix.

Project Content Download the dataset from `http://bioinformatics.mdanderson.org/Supplements/Kornblau-AML-RPPA/aml-rppa.xls`, which is provided as supplementary material to [2]. The dataset contains measurements from 256 newly diagnosed AML patients, each profiled for 51 total and phosphoprotein epitopes involved in signal transduction, apoptosis, and cell cycle regulatory pathways. This can be interpreted as a data matrix with $n = 256$ samples and $p = 51$ features (proteins).

Your task is to implement and explore Bayesian inference methods to construct Gaussian Graphical Models (GGMs) that capture the protein regulatory network among these features. Please follow the steps outlined below to complete this task.

1. Download the dataset and load it into `R/Python`. Normalize the data to ensure the assumption $\boldsymbol{\mu} = 0$ holds.

2. Based on the joint regression model, derive the full conditional distributions for each off-diagonal element ω^{ij} and diagonal element ω^{ii} of the precision matrix. Be sure to specify appropriate prior distributions as part of your derivation.

3. To reflect the sparsity assumption of the underlying graph (i.e., some edges are present and others absent), incorporate sparsity-inducing priors into the model. Two common options are the spike-and-slab prior [4] and adaptive shrinkage priors (e.g., the normal-gamma prior [1], horseshoe prior [5]). Review the relevant literature and select one of these priors to use in deriving the full conditional distribution of ω^{ij}.

4. Implement a Gibbs sampling algorithm (or a Gibbs-within-Metropolis-Hastings sampler, depending on your prior choice) based on the full conditional distributions derived earlier.

5. After inference, construct the final graph structure. *Hint*: For shrinkage priors, you may apply a post-processing step to threshold posterior samples. For instance, retain only the edges for which the estimated partial correlations ρ^{ij} exceed a threshold (e.g., $\rho^{ij} > 0.1$).

6. Perform standard MCMC diagnostics to assess convergence and mixing of the sampling algorithm. Consider using trace plots, autocorrelation plots, effective sample size, the Gelman–Rubin statistic and other relevant diagnostic measures.

7. Evaluate whether Variational Inference can be applied under the chosen prior, and derive and implement the Variational Inference steps if it is feasible.

8. Generate visual representations of the constructed graphs and clearly present the resulting network structures.

References

[1] Jim E Griffin and Philip J Brown. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188, 2010.

[2] Steven M Kornblau, Raoul Tibes, Yi Hua Qiu, Wenjing Chen, Hagop M Kantarjian, Michael Andreeff, Kevin R Coombes, and Gordon B Mills. Functional proteomic profiling of aml predicts response and survival. *Blood, The Journal of the American Society of Hematology*, 113(1):154–164, 2009.

[3] NICOLAI MEINSHAUSEN and PETER BÜHLMANN. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462, 2006.

[4] Toby J Mitchell and John J Beauchamp. Bayesian variable selection in linear regression. *Journal of the american statistical association*, 83(404):1023–1032, 1988.

[5] Juho Piironen and Aki Vehtari. On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. In *Artificial intelligence and statistics*, pages 905–913. PMLR, 2017.

[6] Zeya Wang, Veerabhadran Baladandayuthapani, Ahmed O Kaseb, Hesham M Amin, Manal M Hassan, Wenyi Wang, and Jeffrey S Morris. Bayesian edge regression in undirected graphical models to characterize interpatient heterogeneity in cancer. *Journal of the American Statistical Association*, 117(538):533–546, 2022.