

Probabilistic Graphical Models: Homework Module 1

Svante Linusson, Liam Solus
KTH Royal Institute of Technology

Due Date: 10 October 2023 (23:59)

Code of conduct(Hederskodex):

- You are allowed to discuss the problems with other students in small groups. If you do, you must write the names of those people that you collaborated with at the beginning of your solution set.
- **You must write down the solutions to each problem based on your own understanding of the problem following the group discussions, and it must be written in your own words.**
- You cannot use resources other than the course notes and the course literature when solving the problems. If you violate this rule, you must report the resources used. You will not receive credit for the portions of the problems where you used other resources.

You should upload a **pdf file** of your solutions to canvas (**do not upload image files that are not converted to a pdf!**). If you choose to scan handwritten solutions into a pdf make sure that the scanned copy is legible. If it is not possible to upload your solutions to canvas, you should submit them to Liam Solus, Institutionen för Matematik, KTH, SE-10044 Stockholm.

1. Let Σ denote an $m \times m$ positive definite matrix. Given a random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ where $\mathbf{X}_i = [X_{1,i}, \dots, X_{m,i}]^T \sim N(\mathbf{0}, \Sigma)$ with sample covariance matrix

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T, \quad \text{where} \quad \bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i,$$

one can apply the following level 0.05 hypothesis test for conditional independence:

$$\text{Reject } H_0 : X_i \perp\!\!\!\perp X_j \mid \mathbf{X}_{[m] \setminus \{i,j\}} \quad \text{when} \quad (\sqrt{n-m-3})|z(\rho_{i,j})| > 1.96,$$

where

$$\hat{\Sigma}^{-1} = [r_{i,j}]_{i,j=1}^m, \quad \rho_{i,j} = -\frac{r_{i,j}}{\sqrt{r_{i,i}r_{j,j}}}, \quad \text{and} \quad z(t) = \frac{1}{2} \log \left(\frac{1+t}{1-t} \right).$$

Suppose we observe a sample with $n = 100000$ that has the sample covariance matrix

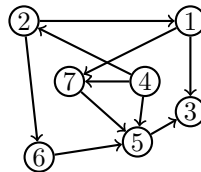
$$\hat{\Sigma} = \begin{bmatrix} 0.176 & -0.03 & 0.004 & 0.004 & 0.005 & -0.031 \\ -0.03 & 0.178 & -0.03 & 0.005 & -0.002 & 0.005 \\ 0.004 & -0.03 & 0.178 & -0.031 & 0.004 & 0.003 \\ 0.004 & 0.005 & -0.031 & 0.18 & -0.025 & -0.026 \\ 0.005 & -0.002 & 0.004 & -0.025 & 0.175 & -0.026 \\ -0.031 & 0.005 & 0.003 & -0.026 & -0.026 & 0.18 \end{bmatrix}$$

from the joint distribution $\mathbf{X} = [X_1, X_2, X_3, X_4, X_5, X_6]^T \sim N(0, \Sigma)$, where Σ is an unknown 6×6 positive definite matrix.

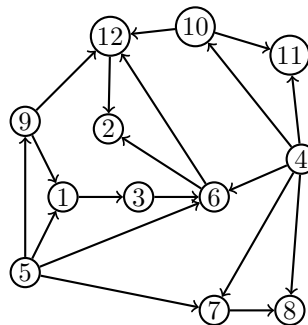
- (2p) Estimate an undirected graph $G = ([6], E)$ with respect to which the data-generating distribution \mathbf{X} factors. Draw the graph G . Be sure to explain why your method of estimation results in a graph G such that \mathbf{X} factors with respect to G .
- (4p) Using the graph G from part (a), give two conditional independence relations $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$ where $|A|, |B|, |C| > 1$ that hold in the data-generating distribution \mathbf{X} . Be sure to explain how you used the graph to deduce that these independence relations hold in \mathbf{X} .

- (c) (4p) We estimate the data-generating distribution as $\mathbf{X} \sim N(\mathbf{0}, \hat{\Sigma})$. Give an expression for the potential functions $\psi_C(\mathbf{x}_C)$ used in the factorization of $f_{\mathbf{X}}(x_1, \dots, x_6)$ according to the estimated graph G for each maximal clique C in G . (Hint: Here, you may need to compute $\hat{\Sigma}^{-1}$. When you do this, choose the number of decimal places you round to so that you have a zero entry whenever $\{i, j\}$ is not an edge in the graph G estimated in part (a) (Why? See Practice Sheet 1, Problem 3.). Two decimal places should be enough when using `numpy`, for instance.)

2. (5pts) Let \mathcal{G} be the DAG depicted below:



- (a) Draw the moral graph \mathcal{G}^m of \mathcal{G} .
(b) Is \mathcal{G}^m chordal? Explain why or why not.
(c) Suppose that the joint distribution $\mathbf{X} = [X_1, X_2, X_3, X_4, X_5, X_6, X_7]^T$ factors according to \mathcal{G} . Is it true that $X_4 \perp\!\!\!\perp X_1 \mid X_2$? Explain why or why not.
(d) Is it true that 4 and 1 are separated given 2 in $(\mathcal{G}|_{\text{An}(1,2,4)})^m$ (i.e., the moral graph of the induced subgraph of \mathcal{G} on $\text{An}(1,2,4)$)? What about in \mathcal{G}^m ? Explain why or why not.
3. (8pts) Consider the DAG $\mathcal{G} = ([12], E)$ depicted below:



- (a) What are the v-structures in this DAG?
(b) Is there a path $\langle i, j, k \rangle$ in \mathcal{G} for which $i \rightarrow j, k \rightarrow j \in E$ that is not a v-structure? If so, give an example.
(c) Draw the essential graph of \mathcal{G} .
(d) How many elements are in the Markov equivalence class of \mathcal{G} ?
4. (4p) Construct a mixed graph that violates all the 4 conditions for being an essential graph in the theorem by Andersson, Madigan, Perlman (don't forget to motivate).
5. (5p) Suppose the following CI relations hold among the variables in the joint distribution $\mathbf{X} = [A, B, C, D, E, F]^T$:

$$\begin{aligned} \text{CI} \supseteq \{ & C \perp\!\!\!\perp F \mid \emptyset, \quad A \perp\!\!\!\perp F \mid \emptyset, \quad B \perp\!\!\!\perp F \mid \emptyset, \quad D \perp\!\!\!\perp F \mid \{B, E\}, \\ & A \perp\!\!\!\perp E \mid \{B, C\}, \quad A \perp\!\!\!\perp D \mid \{B, C\}, \quad B \perp\!\!\!\perp E \mid C, \quad D \perp\!\!\!\perp C \mid \{B, E\} \} \end{aligned}$$

Assume that the joint distribution \mathbf{X} is faithful to an unknown DAG \mathcal{G} on node set $\{A, B, C, D, E, F\}$. Run the PC-algorithm to determine

- (a) the skeleton of \mathcal{G} ,
(b) the v-structures in \mathcal{G} , and
(c) any other edges that are directed in the essential graph of \mathcal{G} .

6. Consider the graph $G_n = (V, E)$ where

$$V = \{(i, j) \in \mathbb{Z}_{\geq 0}^2 : 0 \leq i \leq n, 0 \leq j \leq n\},$$

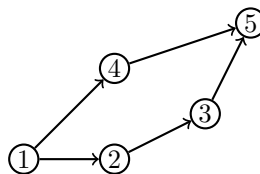
$$E = \{\{(i, j), (i+1, j)\} : 0 \leq i < n, 0 \leq j \leq n\} \cup \{\{(i, j), (i, j+1)\} : 0 \leq i \leq n, 0 \leq j < n\}.$$

(Here, $\mathbb{Z}_{\geq 0}^2$ denotes the points (x, y) in \mathbb{R}^2 where x and y are nonnegative integers.) The undirected graph G_n is often used when modeling correlations between intensities in adjacent pixels in image data as well as the spread of diseases between crops in a field. In the crop-disease model, we may think of each node (i, j) as representing a binary random variable that is equal to 1 if the plant in position (i, j) is infected and 0 otherwise. When treated as an undirected graphical model, we may wish to compute the marginal probability that a certain crop in a certain location in the field is infected. One way to do this is to apply Variable Elimination (VE). However, VE may be computationally expensive for large n , in which case we would opt for approximate inference methods. To determine when we can reasonably use VE for exact inference in a graphical model for G_n , we can compute an upper bound on the treewidth for G_n .

- (a) (8pt) Show that the tree-width of G_n less than or equal to $n+1$. (In fact it can be shown that the tree-width of G_n is exactly equal to $n+1$. This would require a bit more graph theory...)
- (b) (2pt) Discuss whether or not you think it is reasonable to use the exact inference method VE in the crop-disease model where all variables are binary. Are approximate methods preferable in this context? Motivate your decision based on the treewidth of G_n .
7. (8p) The heights of trees can be impacted by the heights of the trees with which the share the surrounding forest area. For instance, the heights of five different species of trees in a certain forest were measured and classified as either short or tall, resulting in observations from the joint distribution $\mathbf{X} = [X_1, X_2, X_3, X_4, X_5]^T$ where

$$X_i = \begin{cases} 1 & \text{if the tree of species } i \text{ is classified as "tall",} \\ 0 & \text{if the tree of species } i \text{ is classified as "short".} \end{cases}$$

Based on observed data, it was estimated that the distribution \mathbf{X} factors according to the DAG



The following conditional probabilities were also estimated:

$$\begin{array}{ll} P(X_1 = 1) = 0.75 & P(X_3 = 1 | X_2 = 0) = 0.75 \\ P(X_2 = 1 | X_1 = 0) = 0.7 & P(X_3 = 1 | X_2 = 1) = 0.65 \\ P(X_2 = 1 | X_1 = 1) = 0.5 & P(X_5 = 1 | X_2 = 0, X_4 = 0) = 0.4 \\ P(X_4 = 1 | X_1 = 0) = 0.8 & P(X_5 = 1 | X_2 = 0, X_4 = 1) = 0.2 \\ P(X_4 = 1 | X_1 = 1) = 0.65 & P(X_5 = 1 | X_2 = 1, X_4 = 0) = 0.3 \\ & P(X_5 = 1 | X_2 = 1, X_4 = 1) = 0.1 \end{array}$$

We are interested in the expected heights of each species in the forest; i.e., we would like to compute $E[X_i]$ for $i = 1, 2, 3, 4, 5$. Use the clique-tree algorithm to compute these marginal expectations. You may implement the algorithm in either Python or R. If you do so, be sure to submit a script or Jupyter notebook showing your implementation.

(Hint: An implementation can be done relatively easy in Python using `numpy` arrays. Here, we represent a potential function $\psi_{1,2,3}(x_1, x_2, x_3)$ as a $2 \times 2 \times 2$ array where entry $[i, j, k]$ in the array is the value $\psi_{1,2,3}(i, j, k)$. A good starting point is to make an array for each conditional probability distribution $P(X_i | \mathbf{X}_{\text{pa}_{\mathcal{G}}(i)})$ where \mathcal{G} is the above DAG.)