Selected Topics in Mathematics of Learning

High-Dimensional Statistics

Lecturer: Marius Yamakou

Winter Semester 2024/25
Department of Data Science, FAU

December 10, 2024

Part V: Large Inverse Covariance Matrices continued ...

1. Precision matrix: Putting the elements from Lecture 8 together

Let $X = (X_1, X_2, X_3, \dots, X_p)^\top \sim \mathcal{N}(\mu, \Sigma)$ with $\Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$ and

partition the multivariate normal distribution $Y_a = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $Y_b = \begin{pmatrix} X_3 \\ \vdots \\ X_p \end{pmatrix}$

Why does a zero entry in the precision matrix imply conditional independence?

- Separate precision matrix as follows

$$\Theta = \Sigma^{-1} = (\theta_{ij})_{i,j=1,\dots,p} = \begin{bmatrix} \Theta_{aa} & \Theta_{ab} \\ \Theta_{ba} & \Theta_{bb} \end{bmatrix} \quad \text{with} \quad \Theta_{aa} = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix}$$

- Suppose $\theta_{12} = 0$ such that $\Theta_{aa} = \begin{bmatrix} \theta_{11} & 0 \\ 0 & \theta_{22} \end{bmatrix}$, i.e., Θ_{aa} is diagonal.

1. Precision matrix: Putting the elements from Lecture 8 together

Let $X = (X_1, X_2, X_3, \dots, X_p)^\top \sim \mathcal{N}(\mu, \Sigma)$ with $\Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$ and

partition the multivariate normal distribution $Y_a = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $Y_b = \begin{pmatrix} X_3 \\ \vdots \\ X_p \end{pmatrix}$

Why does a zero entry in the precision matrix imply conditional independence?

- Separate precision matrix as follows

$$\Theta = \Sigma^{-1} = (\theta_{ij})_{i,j=1,\dots,p} = \begin{bmatrix} \Theta_{aa} & \Theta_{ab} \\ \Theta_{ba} & \Theta_{bb} \end{bmatrix} \quad \text{with} \quad \Theta_{aa} = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix}$$

- Suppose $\theta_{12} = 0$ such that $\Theta_{aa} = \begin{bmatrix} \theta_{11} & 0 \\ 0 & \theta_{22} \end{bmatrix}$, i.e., Θ_{aa} is diagonal.
- Then by the results we have obtained already, if $X \sim \mathcal{N}(\mu, \Sigma)$, then $Y_a \mid Y_b$ is multivariate normal with covariance matrix
  $\mathrm{Cov}(Y_a|Y_b = y_b) = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$ and
  $\Theta_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$

1. Precision matrix: Putting the elements from Lecture 8 together

- The inverse can also be written as

$$\Theta = \Sigma^{-1} = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}^{-1} = \begin{bmatrix} \Theta_{aa} & \Theta_{ab} \\ \Theta_{ba} & \Theta_{bb} \end{bmatrix} = \begin{pmatrix} (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1} & \Theta_{ab} \\ \Theta_{ba} & \Theta_{bb} \end{pmatrix}.$$

- Θ_{aa} is diagonal, i.e., $\Theta_{aa} = \begin{bmatrix} \theta_{11} & 0 \\ 0 & \theta_{22} \end{bmatrix}$

1. Precision matrix: Putting the elements from Lecture 8 together

- The inverse can also be written as

$$\Theta = \Sigma^{-1} = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}^{-1} = \begin{bmatrix} \Theta_{aa} & \Theta_{ab} \\ \Theta_{ba} & \Theta_{bb} \end{bmatrix} = \begin{pmatrix} (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1} & \Theta_{ab} \\ \Theta_{ba} & \Theta_{bb} \end{pmatrix}.$$

- Θ_{aa} is diagonal, i.e., $\Theta_{aa} = \begin{bmatrix} \theta_{11} & 0 \\ 0 & \theta_{22} \end{bmatrix}$
- $\Theta_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$ is diagonal.
- $\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$ is diagonal, i.e., $\mathrm{Cov}(Y_a|Y_b = y_b) = \begin{bmatrix} (\theta_{11})^{-1} & 0 \\ 0 & (\theta_{22})^{-1} \end{bmatrix}$

  Where $Y_a = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, $Y_b = \begin{pmatrix} X_3 \\ \vdots \\ X_p \end{pmatrix}$ and $y_b = (x_3, .., x_p)$

1. Precision matrix: Putting the elements from Lecture 8 together

- The inverse can also be written as

$$\Theta = \Sigma^{-1} = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}^{-1} = \begin{bmatrix} \Theta_{aa} & \Theta_{ab} \\ \Theta_{ba} & \Theta_{bb} \end{bmatrix} = \begin{pmatrix} (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1} & \Theta_{ab} \\ \Theta_{ba} & \Theta_{bb} \end{pmatrix}.$$

- Θ_{aa} is diagonal, i.e., $\Theta_{aa} = \begin{bmatrix} \theta_{11} & 0 \\ 0 & \theta_{22} \end{bmatrix}$
- $\Theta_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$ is diagonal.
- $\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$ is diagonal, i.e., $\text{Cov}(Y_a|Y_b = y_b) = \begin{bmatrix} (\theta_{11})^{-1} & 0 \\ 0 & (\theta_{22})^{-1} \end{bmatrix}$

  Where $Y_a = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, $Y_b = \begin{pmatrix} X_3 \\ \vdots \\ X_p \end{pmatrix}$ and $y_b = (x_3, .., x_p)$
- Thus, if any off-diagonal element of Θ is zero, then the corresponding variables are conditionally independent given the remaining variables, i.e., X_1, X_2 are uncorrelated given Y_b.

1. Precision matrix: Putting the elements from Lecture 8 together

Let $X = (X_1, X_2, X_3, \dots, X_p)^\top \sim \mathcal{N}(0, \Sigma)$ with $Y_a = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $Y_b = \begin{pmatrix} X_3 \\ \vdots \\ X_p \end{pmatrix}$

1. Precision matrix: Putting the elements from Lecture 8 together

Let $X = (X_1, X_2, X_3, \ldots, X_p)^\top \sim \mathcal{N}(0, \Sigma)$ with $Y_a = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $Y_b = \begin{pmatrix} X_3 \\ \vdots \\ X_p \end{pmatrix}$

(i) The covariance matrix : $\Sigma = \mathbb{E}[XX'] = \mathbb{E}\Bigg[\begin{pmatrix} Y_a \\ Y_b \end{pmatrix} \begin{pmatrix} Y_a' & Y_b' \end{pmatrix} \Bigg] = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$,

where $\Sigma \in \mathbb{R}^{p\times p}$, $\quad \Sigma_{aa} \in \mathbb{R}^{2\times 2}$, $\quad \Sigma_{ab} \in \mathbb{R}^{2\times (p-2)}$, $\quad \Sigma_{bb} \in \mathbb{R}^{(p-2)\times(p-2)}$.
Conditional distribution:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \mid Y_b = y_b = Y_a \mid Y_b = y_b \sim \mathcal{N}\Big(\Sigma_{ab}\Sigma_{bb}^{-1} y_b, \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \Big).$$

1. Precision matrix: Putting the elements from Lecture 8 together

Let $X = (X_1, X_2, X_3, \dots, X_p)^\top \sim \mathcal{N}(0, \Sigma)$ with $Y_a = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $Y_b = \begin{pmatrix} X_3 \\ \vdots \\ X_p \end{pmatrix}$

(i) The covariance matrix : $\Sigma = \mathbb{E}[XX'] = \mathbb{E}\left[\begin{pmatrix} Y_a \\ Y_b \end{pmatrix} \begin{pmatrix} Y_a' & Y_b' \end{pmatrix} \right] = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$,

where $\Sigma \in \mathbb{R}^{p\times p}$, $\quad \Sigma_{aa} \in \mathbb{R}^{2\times 2}$, $\quad \Sigma_{ab} \in \mathbb{R}^{2\times (p-2)}$, $\quad \Sigma_{bb} \in \mathbb{R}^{(p-2)\times(p-2)}$.
Conditional distribution:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \mid Y_b = y_b = Y_a \mid Y_b = y_b \sim \mathcal{N}\Big(\Sigma_{ab}\Sigma_{bb}^{-1} y_b, \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \Big).$$

(ii) The precision matrix Θ:

$$\Theta = \Sigma^{-1} = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}^{-1} = \begin{pmatrix} (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1} & * \\ * & * \end{pmatrix},$$

where $(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1} = \begin{bmatrix} \theta_{11} & 0 \\ 0 & \theta_{22} \end{bmatrix}$

1. Precision matrix: Putting the elements from Lecture 8 together

Let $X = (X_1, X_2, X_3, \ldots, X_p)^\top \sim \mathcal{N}(0, \Sigma)$ with $Y_a = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $Y_b = \begin{pmatrix} X_3 \\ \vdots \\ X_p \end{pmatrix}$

(i) The covariance matrix : $\Sigma = \mathbb{E}[XX'] = \mathbb{E}\left[\begin{pmatrix} Y_a \\ Y_b \end{pmatrix} \begin{pmatrix} Y_a' & Y_b' \end{pmatrix} \right] = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$,

where $\Sigma \in \mathbb{R}^{p\times p}$, $\quad \Sigma_{aa} \in \mathbb{R}^{2\times 2}$, $\quad \Sigma_{ab} \in \mathbb{R}^{2\times(p-2)}$, $\quad \Sigma_{bb} \in \mathbb{R}^{(p-2)\times(p-2)}$.
Conditional distribution:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \mid Y_b = y_b = Y_a \mid Y_b = y_b \sim \mathcal{N}\Big(\Sigma_{ab}\Sigma_{bb}^{-1} y_b, \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}\Big).$$

(ii) The precision matrix Θ:

$$\Theta = \Sigma^{-1} = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}^{-1} = \begin{pmatrix} (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1} & * \\ * & * \end{pmatrix},$$

where $(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1} = \begin{bmatrix} \theta_{11} & 0 \\ 0 & \theta_{22} \end{bmatrix}$

Combining (i) and (ii): $X_1 \perp\!\!\!\perp X_2 \mid Y_b = y_b \iff \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$ is diagonal
$\iff (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$ is diagonal
$\iff$ zero entries in precision matrix $\implies \Theta_{12} = 0 = \Theta_{21}$.

1. Precision matrix: Putting the elements from Lecture 8 together

Example 1: Zero entry in the precision matrix implies conditional independence.

$$(X, Y, Z, R)^\top \sim \mathcal{N}(0, \Sigma)$$

$$\Sigma = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \quad , \quad \Theta = \Sigma^{-1} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ -1 & -1 & -1 & 4 \end{pmatrix}.$$

1. Precision matrix: Putting the elements from Lecture 8 together

Example 1: Zero entry in the precision matrix implies conditional independence.

$$(X, Y, Z, R)^\top \sim \mathcal{N}(0, \Sigma)$$

$$\Sigma = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \quad , \quad \Theta = \Sigma^{-1} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ -1 & -1 & -1 & 4 \end{pmatrix}.$$

(i) $X \perp\!\!\!\perp Z$ False!

(ii) $X \perp\!\!\!\perp Z \mid Y, R$ **True**

1. Precision matrix: Putting the elements from Lecture 8 together

Just to clarify a little bit more the example above, let's discuss conditional Independence: $X \perp\!\!\!\perp Z \mid Y, R$.

Goal: To understand more pricisly why conditioning on Y and R removes the dependency between X and Z.

We will break this into five steps:

- Step 1: What conditioning means
- Step 2: Dependencies before conditioning
- Step 3: Effect of conditioning on dependencies
- Step 4: Why conditioning removes the dependency

1. Precision matrix: Putting the elements from Lecture 8 together

Step 1: What does conditioning mean?

Conditioning on a variable means:

- We restrict our analysis to cases where the value of the conditioning variable is **fixed**.
- When conditioning on Y and R, we ask: **"What is the relationship between X and Z given that we know Y and R?"**
- This removes the influence of Y and R on X and Z to focus on their residual variation.

1. Precision matrix: Putting the elements from Lecture 8 together

Step 2: Dependencies before conditioning

From the precision matrix:

$$\Theta = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ -1 & -1 & -1 & 4 \end{pmatrix},$$

- $\Theta_{XR} = -1$: X and R are directly connected.
- $\Theta_{ZR} = -1$: Z and R are directly connected.
- $\Theta_{XZ} = 0$: X and Z are not directly connected.

- X and Z appear dependent because they both interact with R (indirect dependency).

Step 3: Effect of conditioning on dependencies

Conditioning on Y and R:

- Fixes the values of Y and R, effectively "controlling for" their influence on X and Z.
- Removes the indirect dependency between X and Z mediated by R.
- After conditioning:

  The apparent dependency between X and Z disappears.

- Conditioning eliminates dependencies mediated through Y and R, leaving only direct relationships.

1. Precision matrix: Putting the elements from Lecture 8 together

Step 4: Why does conditioning remove dependency?

Conceptually:

- Before conditioning:

$$X \text{ and } Z \text{ share an indirect dependency via } R.$$

- After conditioning:
  - The value of R is fixed, so its influence is "accounted for."
  - No statistical relationship between X and Z remains.

1. Precision matrix: Putting the elements from Lecture 8 together

Step 4: Why does conditioning remove dependency?

Conceptually:

- Before conditioning:

$$X \text{ and } Z \text{ share an indirect dependency via } R.$$

- After conditioning:
  - The value of R is fixed, so its influence is "accounted for."
  - No statistical relationship between X and Z remains.

Conclusion:

- $\Theta_{XZ} = 0$ guarantees no direct relationship between X and Z.
- Conditioning on R removes indirect effects, making $X \perp\!\!\!\perp Z \mid Y, R$ true.

1. Precision matrix: Putting the elements from Lecture 8 together

Step 4: Why does conditioning remove dependency?

Conceptually:

- Before conditioning:

X and Z share an indirect dependency via R.

- After conditioning:
  - The value of R is fixed, so its influence is "accounted for."
  - No statistical relationship between X and Z remains.

Conclusion:

- $\Theta_{XZ} = 0$ guarantees no direct relationship between X and Z.
- Conditioning on R removes indirect effects, making $X \perp\!\!\!\perp Z \mid Y, R$ true.

In the Precision matrix of our example, we also have that:

(iii) $Y \perp\!\!\!\perp Z \mid X, R$ **True**

(iv) $X \perp\!\!\!\perp Y \mid Z, R$ **True**

1. Precision matrix: Putting the elements from Lecture 8 together

How does the graph structure come into play?

1. Precision matrix: Putting the elements from Lecture 8 together

How does the graph structure come into play?

- Suppose that we have a graph G whose vertex set V represents a set of random variables having joint distribution $\mathbb{P}$

1. Precision matrix: Putting the elements from Lecture 8 together

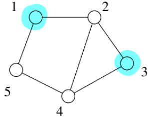
How does the graph structure come into play?

- Suppose that we have a graph G whose vertex set V represents a set of random variables having joint distribution $\mathbb{P}$
- the absence of an edge implies that the corresponding random variables are conditionally independent given the variables at the other vertices.

1. Precision matrix: Putting the elements from Lecture 8 together

How does the graph structure come into play?

- Suppose that we have a graph G whose vertex set V represents a set of random variables having joint distribution $\mathbb{P}$
- the absence of an edge implies that the corresponding random variables are conditionally independent given the variables at the other vertices.
- No edge joining X and Y if and only if $X \perp\!\!\!\perp Y|$rest.



Example: $X_1 \perp\!\!\!\perp X_3 \mid X_2, X_4, X_5$ (True)

1. Precision matrix: Putting the elements from Lecture 8 together

Let $(X_1, \ldots, X_p)' \sim \mathcal{N}(0, \Sigma)$

$\tilde{V} = \{1, \ldots, p\} \backslash \{i, j\}, \quad X_{\tilde{V}} = \{X_1, .., X_{i-1}, X_{i+1}, .., X_{j-1}, X_{j+1}, .., X_p\}, \quad Q \subseteq \tilde{V}.$

1. Precision matrix: Putting the elements from Lecture 8 together

Let $(X_1, \dots, X_p)' \sim \mathcal{N}(0, \Sigma)$

$\tilde{V} = \{1, \dots, p\} \backslash \{i, j\}, \quad X_{\tilde{V}} = \{X_1, .., X_{i-1}, X_{i+1}, .., X_{j-1}, X_{j+1}, .., X_p\}, \quad Q \subseteq \tilde{V}.$

What can and what can we not infer?

1. Precision matrix: Putting the elements from Lecture 8 together

Let $(X_1, \dots, X_p)' \sim \mathcal{N}(0, \Sigma)$

$\tilde{V} = \{1, \dots, p\} \backslash \{i, j\}, \quad X_{\tilde{V}} = \{X_1, .., X_{i-1}, X_{i+1}, .., X_{j-1}, X_{j+1}, .., X_p\}, \quad Q \subseteq \tilde{V}.$

What can and what can we not infer?

Two Perspectives:

1. **Precision Matrix:**
    1. $\Theta_{ij} = 0 \implies X_i \perp\!\!\!\perp X_j \mid X_{\tilde{V}}$
        - Explanation: In a multivariate Gaussian distribution, the precision matrix ($\Theta = \Sigma^{-1}$) encodes conditional independence.
        - If $\Theta_{ij} = 0$, X_i and X_j are conditionally independent given all other variables ($X_{\tilde{V}}$).

1. Precision matrix: Putting the elements from Lecture 8 together

Let $(X_1, \ldots, X_p)' \sim \mathcal{N}(0, \Sigma)$

$\tilde{V} = \{1, \ldots, p\} \backslash \{i, j\}, \quad X_{\tilde{V}} = \{X_1, .., X_{i-1}, X_{i+1}, .., X_{j-1}, X_{j+1}, .., X_p\}, \quad Q \subseteq \tilde{V}.$

What can and what can we not infer?

Two Perspectives:

1. **Precision Matrix:**
   1. $\Theta_{ij} = 0 \implies X_i \perp\!\!\!\perp X_j \mid X_{\tilde{V}}$
      - Explanation: In a multivariate Gaussian distribution, the precision matrix ($\Theta = \Sigma^{-1}$) encodes conditional independence.
      - If $\Theta_{ij} = 0$, X_i and X_j are conditionally independent given all other variables ($X_{\tilde{V}}$).
   2. $\Theta_{ij} = 0 \not\Longrightarrow X_i \perp\!\!\!\perp X_j \mid X_Q$ (for $Q \subseteq \tilde{V}$)
      - Explanation: If we condition on a subset $Q \subseteq \tilde{V}$, $\Theta_{ij} = 0$ does not guarantee conditional independence.
      - Conditional independence is only guaranteed when conditioning on $X_{\tilde{V}}$, not arbitrary subsets Q.

1. Precision matrix: Putting the elements from Lecture 8 together

2. Graph Perspective
   1. X_Q separates X_i and $X_j \implies X_i \perp\!\!\!\perp X_j \mid X_Q$
      - Explanation: In a graph, if the set of nodes X_Q separates X_i and X_j, there is no path connecting them when conditioning on X_Q.
      - This follows from the rules of d-separation in graphical models.

2. Graph Perspective
   1. X_Q separates X_i and $X_j \implies X_i \perp\!\!\!\perp X_j \mid X_Q$
      - Explanation: In a graph, if the set of nodes X_Q separates X_i and X_j, there is no path connecting them when conditioning on X_Q.
      - This follows from the rules of d-separation in graphical models.
   2. X_Q does not separate X_i and $X_j \not\Longrightarrow X_i \not\perp\!\!\!\perp X_j \mid X_Q$
      - Explanation: Lack of separation does not guarantee dependence; X_i and X_j may still be conditionally independent due to specific statistical relationships.
      - Conditional dependence requires both structural paths and statistical relationships.

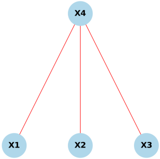
1. Precision matrix: Putting the elements from Lecture 8 together

Note:

- "Separation" refers to blocking all paths between two variables (X_i and X_j) using a set of nodes (X_Q).
- X_Q separates X_i and X_j if all paths between X_i and X_j are blocked by X_Q.
- If X_Q does not separate X_i and X_j, there is at least one unblocked path between X_i and X_j that bypasses X_Q.
- However, lack of separation does not necessarily imply conditional dependence, as specific statistical relationships might still result in X_i and X_j being independent.

1. Precision matrix: Putting the elements from Lecture 8 together

$$\Theta = \Sigma^{-1} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ -1 & -1 & -1 & 4 \end{pmatrix}$$



$X_1 \perp\!\!\!\perp X_2 \mid X_3, X_4$ (True)

1. Precision matrix: Putting the elements from Lecture 8 together

$$\Sigma = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \quad , \quad \overline{\Sigma} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

$$\overline{\Theta} = \overline{\Sigma}^{-1} = \begin{pmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{pmatrix}$$

$X_1 \perp\!\!\!\perp X_2 \mid X_3$ (False)

Summary

Independence and conditional Independence in Gaussians

- Independence in Gaussians is determined by sparsity pattern of the covariance Σ.
  - Sparsity pattern: "where the non-zeroes are".
  - $X_i \perp\!\!\!\perp X_j$ iff $\Sigma_{ij} = 0$.

Summary

Independence and conditional Independence in Gaussians

- Independence in Gaussians is determined by sparsity pattern of the covariance Σ.
    - Sparsity pattern: "where the non-zeroes are".
    - $X_i \perp\!\!\!\perp X_j$ iff $\Sigma_{ij} = 0$.
- Gaussians' conditional independence: sparsity of the precision matrix, $\Theta = \Sigma^{-1}$.
    - $X_i \perp\!\!\!\perp X_j \mid \{X_k | k \notin \{i, j\}\}$ iff $\Theta_{ij} = 0$.

Summary

Independence and conditional Independence in Gaussians

- Independence in Gaussians is determined by sparsity pattern of the covariance Σ.
  - Sparsity pattern: "where the non-zeroes are".
  - $X_i \perp\!\!\!\perp X_j$ iff $\Sigma_{ij} = 0$.
- Gaussians' conditional independence: sparsity of the precision matrix, $\Theta = \Sigma^{-1}$.
  - $X_i \perp\!\!\!\perp X_j \mid \{X_k | k \notin \{i, j\}\}$ iff $\Theta_{ij} = 0$.
- We use the sparsity pattern of Θ to define a graph.
  - Each node in the graph corresponds to a variable $j \in \{1, 2, \ldots, p\}$.
  - Each edge in the graph corresponds to a non-zero Θ_{ij}.

Summary

Independence and conditional Independence in Gaussians

- Independence in Gaussians is determined by sparsity pattern of the covariance Σ.
  - Sparsity pattern: "where the non-zeroes are".
  - $X_i \perp\!\!\!\perp X_j$ iff $\Sigma_{ij} = 0$.
- Gaussians' conditional independence: sparsity of the precision matrix, $\Theta = \Sigma^{-1}$.
  - $X_i \perp\!\!\!\perp X_j \mid \{X_k | k \notin \{i,j\}\}$ iff $\Theta_{ij} = 0$.
- We use the sparsity pattern of Θ to define a graph.
  - Each node in the graph corresponds to a variable $j \in \{1, 2, \ldots, p\}$.
  - Each edge in the graph corresponds to a non-zero Θ_{ij}.
- Checking independence and conditional independence using the graph:
  - Independence: $X_i \perp\!\!\!\perp X_j$ if no path exists between X_i and X_j in the graph of Σ.
  - Conditional Independence: $X_i \perp\!\!\!\perp X_j \mid X_k$ if X_k blocks all paths from X_i to X_j in the graph of Θ.

Summary

Independence and conditional Independence in Gaussians

- If Σ is diagonal then Θ is diagonal.
    - This gives a disconnected graph: all variables are independent.

Summary

Independence and conditional Independence in Gaussians

- If Σ is diagonal then Θ is diagonal.
  - This gives a disconnected graph: all variables are independent.

- If Θ is a full matrix, the graph does not imply any conditional independences.
  - “Everything depends on everything, no matter how many of the X_j you know."

- If Σ is diagonal then Θ is diagonal.
  - This gives a disconnected graph: all variables are independent.

- If Θ is a full matrix, the graph does not imply any conditional independences.
  - "Everything depends on everything, no matter how many of the X_j you know."

- Dependencies can exist if $\Theta_{ij} = 0$ due to correlations with other variables.
  - Only independent if all paths that correlation could go across are blocked.