Homework 1 - Introduction to Probabilistic Graphical Models

kipngeno koech - bkoech

February 8, 2025

1 Bayesian Networks

1. Consider a simple Markov Chain structure $X \to Y \to Z$, where all variables are binary. You are required to:
   (a) Write a code (using your preferred programming language) that generates a distribution (not necessarily a valid BN one) over the 3 variables.

   [in the notebook]

   (b) Write a code that verifies whether a distribution is a valid BN distribution.

   [in the notebook]

   (c) Using your code, generate 10000 distributions and compute the fraction of distributions that are valid BN distributions.

   [in the notebook]

2. Given the following Bayesian Network

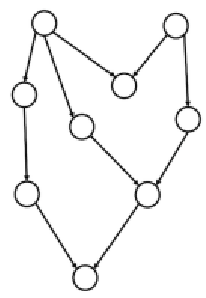

Figure 1: A Bayesian network.

Figure 1: Bayesian Network

   (a) Propose a topological ordering of this graph
   In Figure 2, the topological ordering is:
      i. $A \to B \to C \to D \to E \to F \to G \to H \to I$
      ii. $A \to B \to C \to E \to D \to F \to G \to I \to H$

   (b) Let **X** be a random vector that is Markov with respect to the graph. We assume that the random variables X, are binary. Write all the local conditional independence

X_A has no parents, so no independence condition applies here.

X_B has no parents, so no independence condition applies here.

X_C is conditionally independent of all other nodes given its parent X_A :

$$X_C \perp \{X_B, X_D, X_E, X_F, X_G, X_H, X_I\} \mid X_A$$

X_D is conditionally independent of all other nodes given its parent X_A :

$$X_D \perp \{X_B, X_C, X_E, X_F, X_G, X_H, X_I\} \mid X_A$$

X_E is conditionally independent of all other nodes given its parents X_C and X_D :

$$X_E \perp \{X_A, X_B, X_F, X_G, X_H, X_I\} \mid \{X_C, X_D\}$$

X_F is conditionally independent of all other nodes given its parent X_B :

$$X_F \perp \{X_A, X_C, X_D, X_E, X_G, X_H, X_I\} \mid X_B$$

X_G is conditionally independent of all other nodes given its parents X_E and X_F :

$$X_G \perp \{X_A, X_B, X_C, X_D, X_H, X_I\} \mid \{X_E, X_F\}$$

X_H is conditionally independent of all other nodes given its parent X_D :

$$X_H \perp \{X_A, X_B, X_C, X_E, X_F, X_G, X_I\} \mid X_D$$

X_I is conditionally independent of all other nodes given its parents X_G and X_H :

$$X_I \perp \{X_A, X_B, X_C, X_D, X_E, X_F\} \mid \{X_G, X_H\}$$

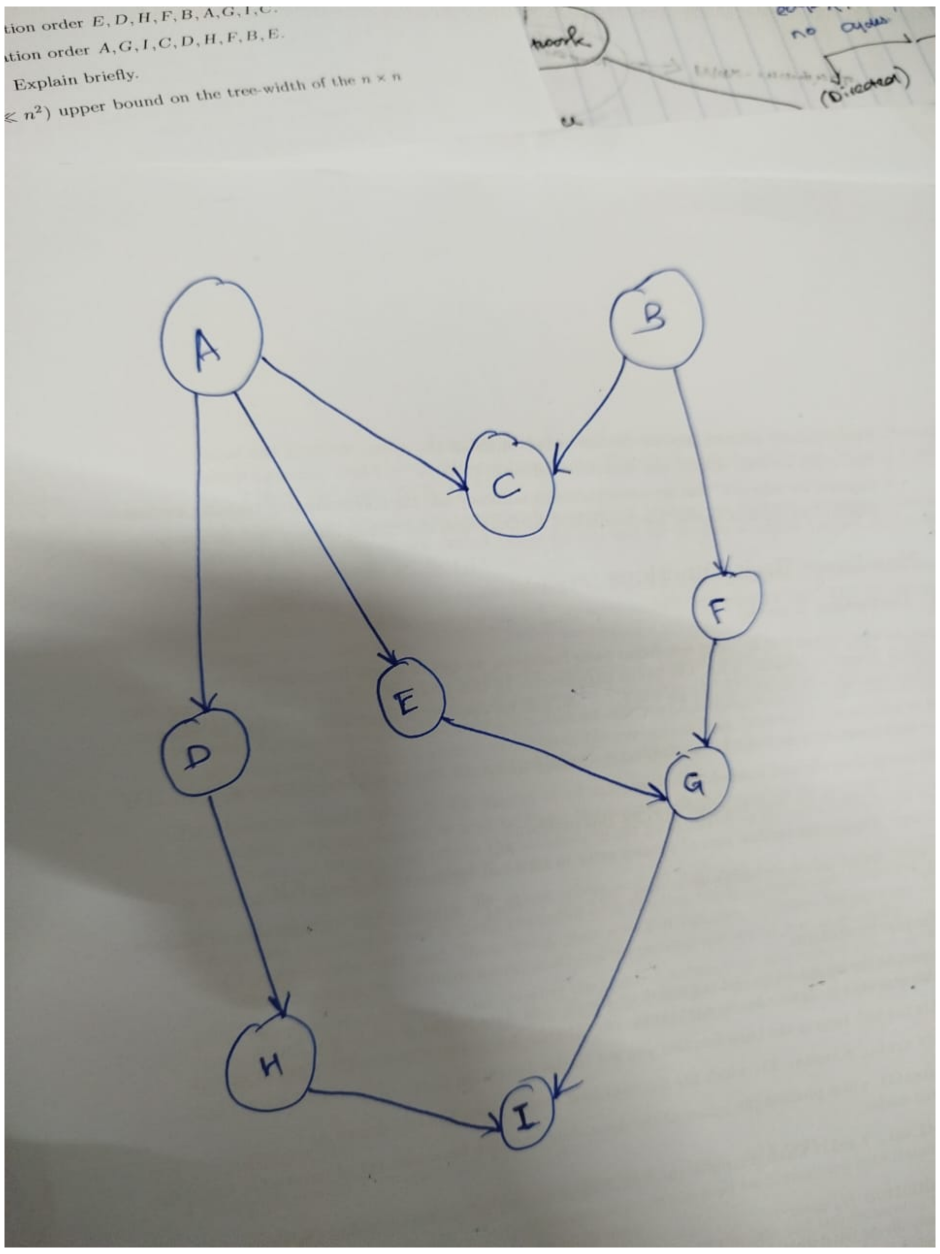

Figure 2: Bayesian Network

3. State True or False, and briefly justify your answer within 3 lines. The statements are either direct consequences of theorems in Koller and Friedman (2009, Ch. 3), or have a short proof. In the follows, P is a distribution and G is a BN structure.

   (a) If $A \perp B \mid C$ and $A \perp C \mid B$, then $A \perp B$ and $A \perp C$. (Suppose the joint distribution of A, B, C is positive.) (This is a general probability question not related to BNs.)

   - **False**. Conditional independence does not imply marginal independence. For example, A and B can be dependent but become independent given C.

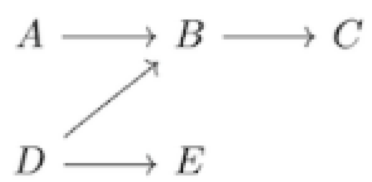

Figure 2: A Bayesian network.

Figure 3: Bayesian Network

   (b) In Figure 2, $E \perp C \mid B$

   (c) in Figure 2, $A \perp E \mid C$

   In figure 3, Recall the definitions of local and global independences of G and independences of P.

$$I_l(G) = \{(X \perp \text{NonDescendants}_G(X) \mid \text{Parents}_G(X))\} \tag{1}$$

$$I(G) = \{(X \perp Y \mid Z) : \text{d-separated}_G(X, Y \mid Z)\} \tag{2}$$

$$I(P) = \{(X \perp Y \mid Z) : P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)\} \tag{3}$$

   (d) In Figure 3, relation 1 is true.

   (e) In Figure 3, relation 2 is true.

   (f) In Figure 3, relation 3 is true.

   (g) If G is an I-map for P, then P may have extra conditional independencies than G.

   (h) Two BN structures G_1 and G_2 are I-equivalent if they have the same skeleton and the same set of v-structures.

$$\begin{tikzcd}
P \text{ factorizes over } \mathcal{G} \arrow[r, Rightarrow, "(1)"] & \mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}(P) \arrow[r, Rightarrow, "(2)"] & \mathcal{I}_\ell(\mathcal{G}) \subseteq \mathcal{I}(P) \arrow[ll, Rightarrow, bend left, "(3)"]
\end{tikzcd}$$

Figure 3: Some relations in Bayesian networks.

(i) If G_1 is an I-map of distribution P, and G_1 has fewer edges than G_2, then G_2 is not a minimal I-map of P.

(j) The P-map of a distribution, if it exists, is unique.

2 Markov Networks

Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The partial correlation matrix $\mathbf{R}$ of $\mathbf{X}$ is a $d \times d$ matrix where each entry $R_{ij} = \rho(X_i, X_j \mid \mathbf{X}_{-ij})$ is the partial correlation between X_i and X_j given the $d-2$ remaining variables $\mathbf{X}_{-ij}$. Let $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ be the inverse covariance matrix of $\mathbf{X}$.

We will prove the relation between $\mathbf{R}$ and $\boldsymbol{\Theta}$, and furthermore how $\boldsymbol{\Theta}$ characterizes conditional independence in Gaussian graphical models.

1. (10 points) Show that

$$\begin{pmatrix} \Theta_{ii} & \Theta_{ij} \\ \Theta_{ji} & \Theta_{jj} \end{pmatrix} = \begin{pmatrix} \mathrm{Var}[e_i] & \mathrm{Cov}[e_i, e_j] \\ \mathrm{Cov}[e_i, e_j] & \mathrm{Var}[e_j] \end{pmatrix}^{-1}$$

for any $i, j \in [d], i \neq j$. Here e_i is the residual resulting from the linear regression of X_{-ij} to X_i, and similarly e_j is the residual resulting from the linear regression of X_{-ij} to X_j.

Schur complement of a block matrix is used to find the inverse of the matrix. For a block matrix of the form:

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$$

the inverse is given by:

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{pmatrix}$$

In our case, we have:

$$\boldsymbol{\Theta} = \begin{pmatrix} \Theta_{ii} & \Theta_{ij} \\ \Theta_{ji} & \Theta_{jj} \end{pmatrix}$$

and

$$\mathbf{V} = \begin{pmatrix} \mathrm{Var}[e_i] & \mathrm{Cov}[e_i, e_j] \\ \mathrm{Cov}[e_i, e_j] & \mathrm{Var}[e_j] \end{pmatrix}$$

We need to show that:

$$\boldsymbol{\Theta} = \mathbf{V}^{-1}$$

Using the Schur complement, we find the inverse of $\mathbf{V}$:

$$\mathbf{V}^{-1} = \frac{1}{\mathrm{Var}[e_i]\mathrm{Var}[e_j] - \mathrm{Cov}[e_i, e_j]^2} \begin{pmatrix} \mathrm{Var}[e_j] & -\mathrm{Cov}[e_i, e_j] \\ -\mathrm{Cov}[e_i, e_j] & \mathrm{Var}[e_i] \end{pmatrix}$$

Thus, we have:

$$\Theta_{ii} = \frac{\mathrm{Var}[e_j]}{\mathrm{Var}[e_i]\mathrm{Var}[e_j] - \mathrm{Cov}[e_i, e_j]^2}$$

$$\Theta_{jj} = \frac{\mathrm{Var}[e_i]}{\mathrm{Var}[e_i]\mathrm{Var}[e_j] - \mathrm{Cov}[e_i, e_j]^2}$$

$$\Theta_{ij} = \Theta_{ji} = \frac{-\mathrm{Cov}[e_i, e_j]}{\mathrm{Var}[e_i]\mathrm{Var}[e_j] - \mathrm{Cov}[e_i, e_j]^2}$$

2. (10 points) Show that

$$R_{ij} = -\frac{\Theta_{ij}}{\sqrt{\Theta_{ii}\Theta_{jj}}}$$

3. (15 points) From the above result and the relation between independence and correlation, we know

$$\Theta_{ij} = 0 \iff R_{ij} = 0 \implies X_i \perp X_j \mid X_{-ij}$$

Note the last implication only holds in one direction. Now suppose $X \sim N(\mu, \Sigma)$ is jointly Gaussian. Show that $R_{ij} = 0 \implies X_i \perp X_j \mid X_{-ij}$.

3 Exact Inference - Variable Elimination

Reference materials for this problem:

- Jordan textbook Ch. 3, available at `https://people.eecs.berkeley.edu/~jordan/prelims/chapter3.pdf`
- Koller and Friedman (2009, Ch. 9 and Ch. 10)

3.1 Variable elimination on a grid [10 points]

Consider the following Markov network:

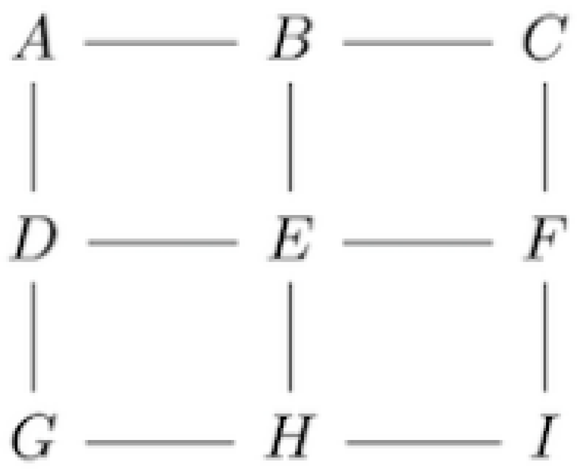

Figure 4: Markov Network

We are going to see how tree-width, a property of the graph, is related to the intrinsic complexity of variable elimination of a distribution

1. (5 points) Write down largest clique(s) for the elimination order $E, D, H, F, B, A, G, I, C$.
   We start by eliminating E, its neighbors are D, F, H, B, so the clique here is :$\{D, F, H, B\}$. size: **4**
   We then eliminate D, its neighbors are A, B, G, H, so the clique here is: $\{A, B, G, H\}$. size: **4**
   We then eliminate H, its neighbors are G, I, B, F, so the clique here is: $\{G, I, B, F\}$. size: **4**
   We then eliminate F, its neighbors are B, G, H, E, so the clique here is: $\{B, G, H, E\}$. size: **4**
2. (5 points) Write down largest clique(s) for the elimination order $A, G, I, C, D, H, F, B, E$.
3. (5 points) Which of the above ordering is preferable? Explain briefly.
4. (10 points) Using this intuition, give a reasonable ($\ll n^2$) upper bound on the tree-width of the $n \times n$ grid.