1: Conditional independence properties

(15 Pts)

1.1 For the sets of four random variables $X, Y, Z$ and $W$ prove the following statements:

(a) $(X \perp W | Z, Y)$ and $(X \perp Y | Z) \Rightarrow (X \perp Y, W | Z)$
(b) $(X \perp Z | Y)$ and $(X \perp Y | Z) \Rightarrow (X \perp Z, Y)$
(c) For positive distributions: $(X \perp Y | Z, W)$ and $(X \perp W | Z, Y) \Rightarrow (X \perp Y, W | Z)$

2: d-separation, I-map and P-map

(30 Pts)

2.1 (10 Pts) I-map
Assume the probability distribution $P(x_1, x_2, x_3, x_4)$ is defined as follows:

$$P(x_1, x_2, x_3, x_4) = \begin{cases} 
0.5 & \text{if } x_1 = x_2 = x_3 = x_4 = 0 \\
0.5 & \text{if } x_1 = x_2 = x_3 = x_4 = 1 
\end{cases}$$

(a) Prove that the undirected graphical model shown in Figure 1 is a minimal I-map for $P(x)$.
(b) Give two more graphs that are minimal I-map for this distribution.

2.2 (10 Pts) Directed model and d-separation
Consider the directed graphical model given in Figure 2 and answer the following questions.

![Directed graphical model](image)

Figure 1: Directed graphical model for question 2.1
(a) Write the joint distribution of variables according to the model.

(b) What is the Markov blanket of $e$.

(c) Which of the following statements are true. (Briefly justify your answers using d-separation.)
   i. $a \perp f$
   ii. $a \perp g$
   iii. $b \perp i | f$
   iv. $d \perp j | g, h$
   v. $i \perp b | h$
   vi. $j \perp d$
   vii. $i \perp c | f, h$

2.3 (10 Pts) P-map
Construct a directed graph that is a perfect map for four random variables $a, b, c$ and $d$
that satisfies the following independence assumptions.

\[ a \perp b \text{ and } a \perp d | b, c \]

3: Directed graphical models

(15 Pts)
Constructing directed model
Construct the corresponding Bayesian network for each of the following cases.

(a) (10 Pts) We want to model a mixture of Gaussian distributions for a given input data
\[ \{X_1, X_2, X_3, \ldots, X_N\} \] that the assignment of each $X_n$ to a cluster is given by a variable $Z_n \in \{1, \ldots, K\}$. Assume that the parameters of the distributions in the $k$-th cluster are denoted as $\mu_k, \Sigma_k$.

Considering the following prior distributions for each of the variables, draw the Bayesian
network.
\[ Z_n \sim \text{Categorical}(\beta), \mu_k, \Sigma_k \sim \text{Normal} - \text{invWishart}(\mu_0, \lambda, \Phi, \nu) \]

(b) (5 Pts) Logistic regression assigns a binary output \( Y_n \) to the input vector \( X_n \). The output distribution is \( Y_n = \text{Bernoulli}(\sigma(X_n W)) \). In Bayesian logistic regression \( \beta \) has a prior given by \( W \sim \text{Normal}(\mu, \Sigma) \). Draw the Bayesian network for this model.

4: Undirected graphical models

(25 Pts)

4.1 Markov properties (5 Pts)
Consider the undirected graphical model in Figure 3. Write local, global and pairwise Markov properties according to this model.

4.2 Pairwise \( \Rightarrow \) global Markov properties (10 Pts)
Prove that pairwise Markov property \( \Rightarrow \) global Markov property, if the followings are true for all disjoint subsets \( A, B, C \) and \( D \).
if \( X_A \perp X_D | \{X_C, X_B\}, X_A \perp X_B | \{X_C, X_D\} \) then \( X_A \perp \{X_C, X_B\} | X_D \)

4.3 Ising model (10 Pts)
Recall the Ising model from undirected models lecture. This model takes the form
\[ P(X_1, ..., X_N | \theta) \propto \prod_{(i,j) \in E} \exp(\theta_{ij} X_i X_j) \prod_{i \in V} \exp(\theta_i X_i) \]
where \( \theta = \{\theta_{ij}, \theta_i\} \) is the parameters set.

(a) (5 Pts) Drive an expression for the following conditional form.
\[ P(X_i | X_{-i}; \theta) \]

(b) (5 Pts) Show that this model can be rewritten in a form similar to the form of the Gaussian graphical model.
The latent Dirichlet allocation (LDA) model [1] is a useful type of hierarchical Bayes model. Please read the referred paper and answer the following questions.

(a) The graphical model representation of LDA is shown in Figure 1. Describe the random variables, the conditional probability distributions, the parameters, and the structure of the graphical model in this figure.

(b) Based on your description from Figure 1 explain what is the main goal of the LDA model.

(c) What is the main question that we can answer about a new document after learning an LDA model on a corpus?

(d) Figure 3 shows three simpler models for text modeling. Describe each of these models and also the parameters that must be learn in them.

(e) Compare the LDA model with the three models shown in Figure 3 and specify the advantages and disadvantages of each model.