# Homework 1
# Introduction of Graphical Models and Exact Inference [1]

10-708 Probabilistic Graphical Models (Spring 2022)
https://piazza.com/cmu/spring2022/10708/home

OUT: Jan. 26, 2022
DUE: Feb. 9, 2022 11:59 PM
TAs: Yuchen, Che-Ping

## START HERE: Instructions

**Summary** In this assignment, you will explore the distributions captured by directed graphical models (Bayesian Networks), undirected graphical Models (MRFs / CRFs), and factor graphs as well as variable elimination.

- **Collaboration policy:** The purpose of student collaboration is to facilitate learning, not to circumvent it. Studying the material in groups is strongly encouraged. It is also allowed to seek help from other students in understanding the material needed to solve a particular homework problem, provided no written notes (including code) are shared, or are taken at that time, and provided learning is facilitated, not circumvented. The actual solution must be done by each student alone. The presence or absence of any form of help or collaboration, whether given or received, must be explicitly stated and disclosed in full by all involved. See the Academic Integrity Section on the course site for more information: https://andrejristeski.github.io/10708-22/#:~:text=Academic%20Integrity%20Policies
- **Late Submission Policy:** See the late submission policy here: https://andrejristeski.github.io/10708-22/#:~:text=Grace%20Day/Late%20Homework%20Policy
- **Submitting your work to Gradescope:** We use Gradescope (https://www.gradescope.com/courses/349316/assignments) to collect PDF submissions of open-ended questions on the homework (e.g. mathematical derivations, plots, short answers). The course staff will manually grade your submission, and you'll receive personalized feedback explaining your final marks. The homework template must be used and can be completed in Latex or by hand. Handwritten submissions must be legible otherwise we will not be able to give credit to your solutions. No changes should be made to the template, boxes and choices **MUST** remain the same size and in the same locations between the template and your completed submission, the document has 23 pages so your submission must contain no more and no less than 23 pages.

For multiple choice or select all that apply questions, shade in the box or circle in the template document corresponding to the correct answer(s) for each of the questions. For LaTeX users, replace `\choice` with `\CorrectChoice` to obtain a shaded box/circle, and don't change anything else.

[1] Compiled on Wednesday 26th January, 2022 at 22:22

## Instructions for Specific Problem Types

For "Select One" questions, please fill in the appropriate bubble completely:

1. **Select One:** Who taught this course?

   ● Andrej Risteski and Hoda Heidari

   ○ Marie Curie

   ○ Noam Chomsky

If you need to change your answer, you may cross out the previous answer and bubble in the new answer:

2. **Select One:** Who taught this course?

   ● Andrej Risteski and Hoda Heidari

   ○ Marie Curie

   ~~●~~ Noam Chomsky

For "Select all that apply" questions, please fill in all appropriate squares completely:

3. **Select all that apply:** Which are scientists?

   ■ Stephen Hawking

   ■ Albert Einstein

   ■ Isaac Newton

   □ I don't know

Again, if you need to change your answer, you may cross out the previous answer(s) and bubble in the new answer(s):

4. **Select all that apply:** Which are scientists?

   ■ Stephen Hawking

   ■ Albert Einstein

   ■ Isaac Newton

   ~~■~~ I don't know

For questions where you must fill in a blank, please make sure your final answer is fully included in the given space. You may cross out answers or parts of answers, but the final answer must still be within the given space.

5. **Fill in the blank:** What is the course number?

   10-708      10-~~9~~708

# 1 Written Questions [100 pts]

Answer the following questions in the template provided. Then upload your solutions to Gradescope. You may use LaTeX or print the template and hand-write your answers then scan it in. Failure to use the template may result in a penalty. There are 100 points and 7 questions.

1. Consider an undirected graphical model shown in 1.1. This model structure looks as follows:

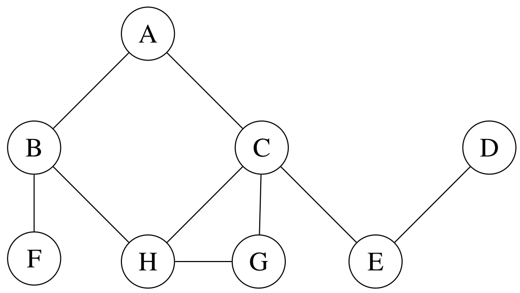



Figure 1.1: Undirected Graphical Model

For this model structure, answer the following questions:

(a) (2 points) Factorize the joint distribution $P(A, B, C, D, E, F, G, H)$ according to the undirected graph in Figure 1.1.

(b) (1 point) Is $C \perp D \mid E$?

◯ True

◯ False

(c) (1 point) Is $A \perp F \mid B$?

◯ True

◯ False

(d) (1 point) Is $A \perp G \mid B$?

◯ True

◯ False

(e) (1 point) Is $B \perp C \mid H$?

◯ True

◯ False

(f) (1 point) Which nodes are present in the Markov blanket of $B$?

(g) (1 point) Which nodes are present in the Markov blanket of $D$?

2. Consider the Bayesian Network described in Figure 1.2

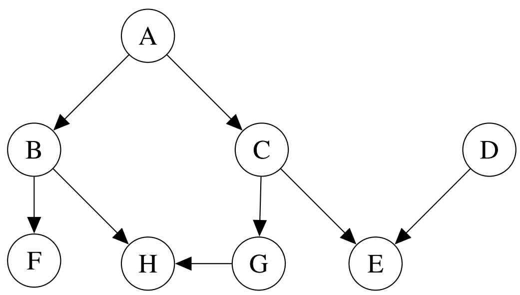



Figure 1.2: Bayesian Network Structure

Based on this network structure, answer the following questions:

(a) (1 point) Factorize the joint distribution $P(A, B, C, D, E, F, G, H)$ according to the directed graph in Figure 1.2.

(b) (1 point) Is $C \perp D \mid E$?

◯ True

◯ False

(c) (1 point) Is $A \perp F \mid B$?

◯ True

◯ False

(d) (1 point) Is $A \perp G \mid B$?

◯ True

◯ False

(e) (1 point) Is $P(B|H) = P(B|C, H)$?

◯ True

◯ False

(f) (1 point) Which nodes are present in the Markov blanket of $B$?

(g) (1 point) Which nodes are present in the Markov blanket of $D$?

(h) (2 points) Please draw the moralized graph of the Bayesian network.

3. **(Markov Properties for UGMs)** Prove the following properties, by using the following equivalence:

$$X \perp Y \mid Z \equiv P(X, Y, Z) = \phi_1(X, Z)\, \phi_2(Y, Z),$$

for some factors $\phi_1(\cdot), \phi_2(\cdot)$.

(a) (2 points) If $A \perp (B, D) \mid C$ then $A \perp B \mid C$.

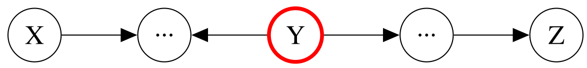
(b) (3 points) If $A \perp (B, D) \mid C$ then $A \perp B \mid (C, D)$ and $A \perp D \mid (B, C)$.

(c) (5 points) For strictly positive distributions, if $A \perp B \mid (C, D)$ and $A \perp C \mid (B, D)$ then $A \perp (B, C) \mid D$.

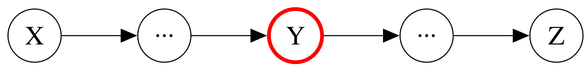
4. (7 points) **(D-separation)** In this question we will formally prove that d-separation implies conditional independence.

Recall the definition of d-separation: two disjoint sets of variables $X, Z$ are said to be *d-separated* by an evidence set $E$ if and only if every (undirected) path from $X$ to $Z$ is "blocked" by $E$. A path is "blocked" whenever any of the following hold:

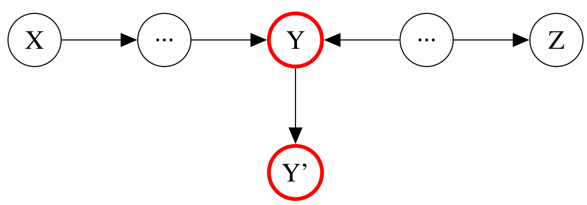
- $\exists Y \in E$ on the path and $Y$ is a common parent:




- $\exists Y \in E$ on the path and $Y$ is in a "cascade" (in either direction):




- $\exists Y$ on the path such that $Y$ is in a v-structure and $\forall \tilde{Y} \in \{Y\} \cup \text{descendants}(Y)$, $\tilde{Y} \notin E$ (**note the NOT**):




Suppose we have a Bayesian network described by a directed graph $G = (V, E)$. Consider a partition of $V$ into three disjoint sets $X, Y, Z$ such that $X \cup Y \cup Z = V$. Prove that

$$\text{dsep}(X, Z \mid Y) \implies X \perp Z \mid Y.$$

*Hint: think about how you could decompose the factors over $X, Y, Z$ into separate factors over $X, Y$ and $Z, Y$.*

5. **(Factor graphs)** Please answer the following problems:

   (a) (2 points) Convert the undirected graphical model in Figure 1.3 to a factor graph.

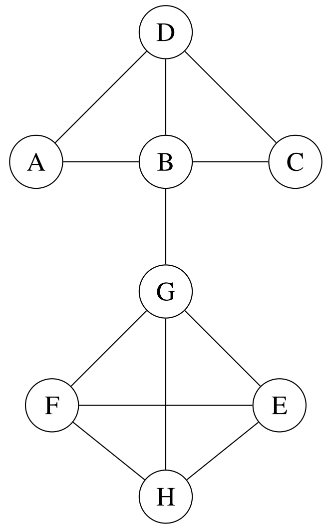



Figure 1.3: Undirected Graphical Model

(b) (2 points) Convert the directed graphical model in Figure 1.4 to a factor graph

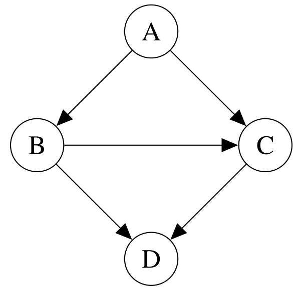



Figure 1.4: Directed Graphical Model

(c) (2 points) Convert the factor graph in Fig 1.5 to an undirected graphical model.

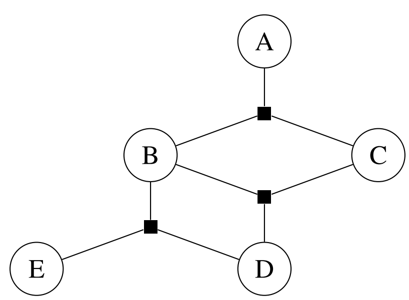



Figure 1.5: Factor Graph

(d) (2 points) Convert the factor graph in Fig 1.6 to a directed graphical model.

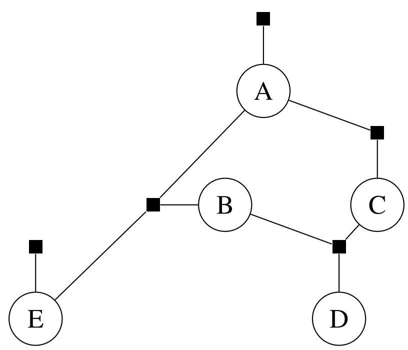
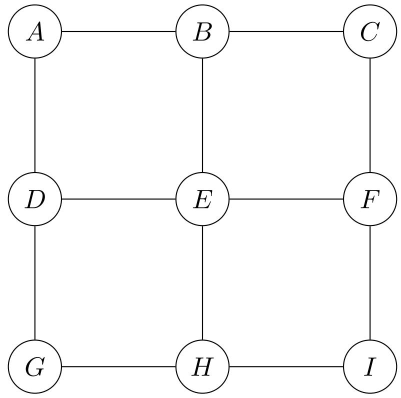



Figure 1.6: Factor Graph

6. Consider the following Markov network:




We are going to see how *treewidth*, a property of the graph, is related to the intrinsic complexity of variable elimination of a distribution.

(a) (2 points) Write down largest clique(s) for the elimination order $E, D, H, F, B, A, G, I, C$.

(b) (2 points) Write down largest clique(s) for the elimination order $A, G, I, C, D, H, F, B, E$.

(c) (2 points) Which of the above ordering is preferable? Explain briefly.

(d) (2 points) The *elimination width* of a given ordering is the size (i.e. number of variables) of the largest factor produced by variable elimination with that ordering. The *treewidth* is the minimum elimination width over all possible orderings, minus one. Give a reasonable ($\ll n^2$) upper bound on the tree-width of the $n \times n$ grid.

7. (Hardness and easiness of inference in graphical models). Classical results show that (both approximate and exact) inference in PGMs is computationally hard in general [1, 2].

   In part (a) of this question, we will guide you through two reductions, showing that the computational complexity of calculating partition functions and marginals is essentially the same. (More results of this type will be included in a course on computational complexity.)

   In part (b), you will see that for certain undirected models (namely, undirected Gaussian graphical models), inference is in fact computationally tractable. In fact, it can be reduced to performing simple linear algebraic computations!

   Combining both parts gives you a flavor of a recurring theme in this course: some tasks are computationally hard for general PGMs, but they are tractable for some particular classes of PGMs by utilizing their specific properties.

   (a) (20 points) The complexity of calculating partition functions and marginals in Ising models is comparable.

   For this part, consider the random variables $\mathbf{X} = \{\mathbf{X}_s \in \{-1, 1\} : s \in [n]\}$ following the distribution of an Ising model $G$ with parameters $\mathbf{J}$. Precisely, the joint distribution of the random variables is expressed as:

$$p(\mathbf{X} = \mathbf{x}) = \frac{1}{Z_G} \exp\big(\sum_{s \in [n]} \mathbf{J}_s \mathbf{x}_s + \sum_{s \neq t \in [n]} \mathbf{J}_{st} \mathbf{x}_s \mathbf{x}_t\big), \tag{1.1}$$

   in which $Z_G$ is the partition function.

   i. **[2 points]** Write down the partition function $Z_G$ for the above Ising model $G$.

ii. **[6 points]** Suppose you are given an oracle $O(H)$ which takes as input an Ising model $H$ and output its partition function $Z_H$ in $\mathcal{O}(1)$ time. Given access to such an oracle $O$, design a polynomial-time algorithm that calculates the marginal probability $p(\mathbf{X}_1 = 1)$.

iii. **[12 points]** Suppose you are given an oracle $O'_H(A, \mathbf{x}_A)$ which takes as input an Ising model $H$, a subset $A \subseteq [n]$ and values $\{\mathbf{x}_i, i \in A\}$ for the set of variables in $A$, and calculates its marginal probability $p(\mathbf{X}_A = \mathbf{x}_A)$ in $\mathcal{O}(1)$ time. Given access to such an oracle $O'$, design a polynomial-time algorithm that calculates the partition function $Z_G$ for our Ising model $G$.

**Remark.** The above problem shows that *if calculating marginals is easy, then calculating partition functions is easy* (and vice versa). To show hardness for both problems in general, one needs a bit more formalism from computational complexity (NP-hardness, #-P hardness, as well as the notion of a reduction). Those familiar with these concepts can look e.g. at [1, 2].

(b) (30 points) Easiness of inference in Gaussian Graphical model

Let $\mathbf{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$ be a vector of random variables which are jointly Gaussian. As a reminder, such a vector follows the distribution

$$p(\mathbf{x}) \propto \exp(-1/2(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)) \tag{1.2}$$

where $\mu$ is the mean of $\mathbf{X}$ and $\Sigma$ is the covariance matrix. The matrix $\Lambda := \Sigma^{-1}$ is called the precision matrix. Note that this can be viewed as an *undirected graphical model*, s.t. the interaction function between nodes $i, j$ is $\Lambda_{i,j} x_i x_j$.

We will see that all the tasks associated with inference, i.e. computing partition functions and marginals/conditional marginals can be reduced to (efficient) linear-algebraic operations.

i. **[5 points]** Show that the partition function of a univariate Gaussian (i.e. $d = 1$) with mean 0 and variance $\sigma^2$ is given by $\sigma\sqrt{2\pi}$.

ii. **[9 points]** Show that the partition function of a multivariate Gaussian with mean $\mu$ and variance $\Sigma$ is given by $\sqrt{\det(\Sigma)(2\pi)^d}$.
*Hint: It may be helpful to first work out the case when $\Sigma$ is diagonal, and reduce the general case to this one.*

iii. **[8 points]** Show that for any $A \subseteq [d]$, the random variables $X_A$ are jointly Gaussian. (In other words, the marginals for any set $A$ are also jointly Gaussian.) Express the mean and covariance of in terms of $\Sigma$ and $\mu$.

iv. **[8 points]** Show that for any subsets $A, B \subseteq [d]$, s.t. $A \cap B = \emptyset$, the random variables $\mathbf{X}_A | (\mathbf{X}_B = \mathbf{x}_B)$ for any $\mathbf{x}_B \in \mathbb{R}^{|B|}$ are jointly Gaussian. (In other words, the conditional distributions for any set $A$, conditioned on another set $B$ are also jointly Gaussian.) Express the mean and covariance of in terms of $\Sigma$ and $\mu$.

# 2 Collaboration Policy

After you have completed all other components of this assignment, report your answers to the collaboration policy questions detailed in the Academic Integrity Policies for this course.

1. Did you receive any help whatsoever from anyone in solving this assignment? If so, include full details including names of people who helped you and the exact nature of help you received.

2. Did you give any help whatsoever to anyone in solving this assignment? If so, include full details including names of people you helped and the exact nature of help you offered.

3. Did you find or come across code that implements any part of this assignment? If so, include full details including the source of the code and how you used it in the assignment.

## References

[1] Francisco Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241, 1982.

[2] Gregory F. Cooper. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence*, 42(2):393–405, 1990.