STA414 & STA2104 MIDTERM

Monday Section - Winter 2022

University of Toronto

Exam duration: **90 minutes**

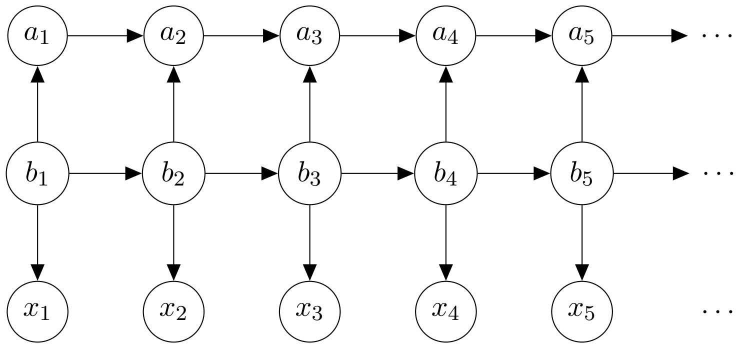
- Start: 2:15 p.m.
- End: 3:45 p.m.
- There will be an additional 10 mins to upload any remaining work. Late submissions after 3:55 will receive 2 points per minute late penalty. Crowdmark time will be taken as reference when determining submission times.

The total possible number of points is **50**.

Read the following instructions carefully:

1. The exam is open-book, but you musn't communicate with other students about the midterm while writing it.
2. If a question asks you to do some calculations or derivations, you must **show your work** for full credit.
3. After completing each question, we recommend you scan and upload your work to crowdmark, to avoid last-minute submission problems.
4. If you run into technical difficulties submitting, send an email to `sta414-instructors@cs.toronto.edu` attaching your solutions.
5. Feel free to ask for clarifications on the zoom chat, but do not discuss potential answers.
6. For questions that ask you to explain your reasoning, we are not expecting an airtight, detailed answer. We just want a rephrasing or intuition behind your answer, in order to make sure you didn't just copy a formula without understanding it.
7. Do not share the exam with anyone, even after the midterm is over.
8. Good luck!

1. Hidden Markov Models. Given the following directed acyclic graphical model:



(a) **[4 points]** Write the factorized joint distribution implied by this DAG. Don't be afraid to add extra brackets or parentheses to avoid ambiguity.

$$p(a_1, a_2, \ldots, a_T, b_1, b_2, \ldots, b_T, x_1, x_2, \ldots, x_T) =$$

(b) **[2 points]** Given the elimination order: $a_1, b_1, a_2, b_2, a_3, b_3, \ldots, a_T, b_T$, what is the time complexity of exactly computing $p(x_1, x_2, \ldots, x_T)$ using variable elimination, as a function of T? Explain your answer in 1 sentence.

- $\mathcal{O}(1)$
- $\mathcal{O}(T)$
- $\mathcal{O}(T^2)$
- $\mathcal{O}(T^3)$

(c) **[1 point]** Is $x_1 \perp x_2$?
(d) **[1 point]** Is $x_1 \perp x_2 | b_1$?
(e) **[1 point]** Is $x_1 \perp x_2 | b_2$?
(f) **[1 point]** Is $a_1 \perp a_3 | a_2$?
(g) **[1 point]** Is $b_1 \perp b_3 | b_2$?
(h) **[1 point]** Is $b_1 \perp b_3 | a_2, b_2$?

2. Simple Monte Carlo. Imagine we have a rain prediction model that outputs samples of

$$p(R_1, R_2, \ldots, R_T | \text{measurements})$$

where each R_i is a Bernoulli random variable indicating whether it rains or not on the ith day ahead. $R = 1$ means rain, and $R = 0$ means no rain.

You are given a set of N i.i.d. samples from this joint predictive distribution:

$$\begin{aligned}
r_1^{(1)}, r_2^{(1)}, \ldots, r_T^{(1)} &\sim p(R_1, R_2, \ldots, R_T | \text{measurements}) \\
r_1^{(2)}, r_2^{(2)}, \ldots, r_T^{(2)} &\sim p(R_1, R_2, \ldots, R_T | \text{measurements}) \\
&\vdots \\
r_1^{(N)}, r_2^{(N)}, \ldots, r_T^{(N)} &\sim p(R_1, R_2, \ldots, R_T | \text{measurements})
\end{aligned}$$

A note on notation: In the questions below, it's not necessary to use the indicator function, you can simply use the random variables themselves to describe the events. For example, $r_1^{(1)} = 1$ means that it rains on the first day, therefore the probability that it rains on the first day can be estimated by $\frac{1}{N}\sum_{i=1}^{N} r_1^{(i)}$, and similarly the probability that it does not rain on the first day can be estimated by $\frac{1}{N}\sum_{i=1}^{N}(1 - r_1^{(i)})$. Also, for the first set of samples, the event that it rains on the first day but it doesn't rain on the second day translates to $r_1^{(1)}(1 - r_2^{(1)}) = 1$.

(a) (5 pts) Write an unbiased estimator for the probability that it rains on at least one of days 1, 2, and 3. Explain your answer in one sentence. As usual, you must write your answer in terms of mathematical functions such as sum, add multiply, divide, and **not** simply write e.g. "at least one is true". Hint: You might want to use the `min` function that returns the minimum of its arguments.

(b) (3 pts) For each of the following estimators for the probability of rain on day 1, say whether it is unbiased.

- Estimator 1: $\frac{1}{N}\sum_{i=1}^{N} r_1^{(i)}$
- Estimator 2: $\frac{1}{N}\sum_{i=1}^{N} r_1^{(n)}$
- Estimator 3: $r_1^{(1)}$

(c) (4 pts) When $N > 1$, and $0 < p(R_1 | \text{measurements}) < 1$ which of the above estimators has the lowest variance? Explain your answer in one sentence.

(a) Estimator 1
(b) Estimator 2
(c) Estimator 3
(d) All have the same variance.

3. Graphical model notation.

(a) **[2 points]** Draw the DAG corresponding to the following factorization of a joint distribution:

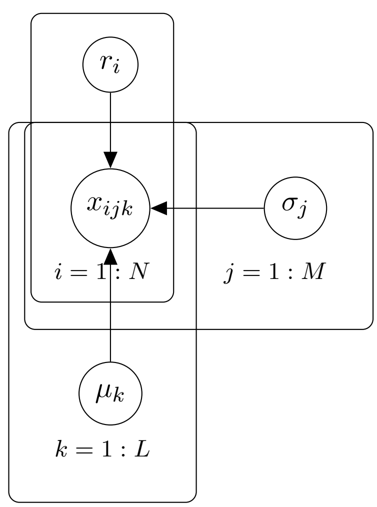
$$p(A, B, C, D, E) = P(A)P(B|A)P(C|A)P(D|C)P(E|A, B, D)$$

(b) If all variables are categorical, and variable A can take one of K_A states, variable B can take on of K_B states, etc.:

- **[3 points]** How many states can this set of variables take on? Explain your answer in one sentence.

- **[4 points]** How many parameters are required to parameterize the joint distribution $p(A, B, C, D, E)$, again assuming the factorization given by the DAG above? Explain your reasoning.

(c) **[2 points]** Write the factorized joint distribution implied by the following graphical model:



4. Decision Theory - 6 points. Imagine you are writing a quiz that has a true or false section. To discourage random guessing, the quiz awards 3 points for a correct answer, -4 points for a wrong answer, and 0 points for no answer.

You think you know the correct answer with probability θ. How high must θ be for the expected number of points to be higher for guessing what you think is the correct answer than for leaving the question blank? Explain your reasoning.

5. Maximum Likelihood (9 Points). The probability density function of a continuous random variable x distributed according to a uniform distribution with parameter $\theta > 0$ is:

$$p(x|\theta) = \frac{1}{\theta} \text{ for } 0 \leq x \leq \theta, 0 \text{ otherwise}$$

Assume that we observed $x_1, x_2, ..., x_n$ i.i.d. draws from a uniform distribution with unknown parameter θ. Find the maximum likelihood estimator for θ. Hint: Calculus won't help you with this one. Try plotting $p(x|\theta)$. What happens if even a single datapoint is outside of the range 0 to θ?

Remember to show your work. We will award part marks for geting partway through the derivation.