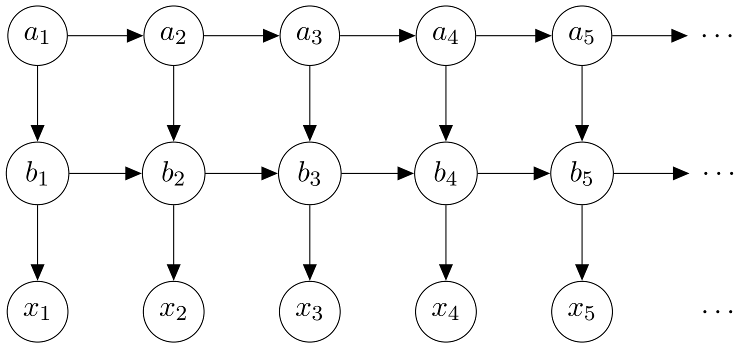
PRACTICE MIDTERM VERSION 1.5 (UPDATED FEB 25)

STA414 - Winter 2022

University of Toronto

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



1. **[2 points]** Write the factorized joint distribution implied by this DAG. Don't be afraid to add extra brackets or parentheses to avoid ambiguity.

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

2. If each variable a_i can take one of K_a states, each variable b_i can take one of K_b states, and each variable x_i can take one of K_x states:

   - **[2 points]** How many states can this set of variables take on?

   - **[3 points]** How many parameters are required to parameterize the joint distribution $p(a_1, a_2, \ldots, a_T, b_1, b_2, \ldots, b_T, x_1, x_2, \ldots, x_T)$, again assuming the factorization given by the DAG above? Note that this factorization does not imply that the factors at each time share any parameters. Also recall that for a categorical variable with K settings, only $K - 1$ parameters are required.

3. **[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?

4. **[1 point]** Is $x_1 \perp x_2$?
5. **[1 point]** Is $x_1 \perp x_2 | b_1$?
6. **[1 point]** Is $x_1 \perp x_2 | b_2$?
7. **[1 point]** Is $a_1 \perp a_3 | a_2$?
8. **[1 point]** Is $b_1 \perp b_3 | b_2$?
9. **[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.

Given a set of N i.i.d. samples from this joint predictive distribution:

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

1. **[2 points]** Write an unbiased estimator for the predicted probability that it rains every day for the next T days. You might want to use the notation $I(\text{statement})$ which takes value 1 if the statement is true, and 0 if it is false.

2. **[3 points]** How does the variance of this estimator change as a function of N?

3. **[1 point]** Write an unbiased estimator for the probability that it rains on day 3.

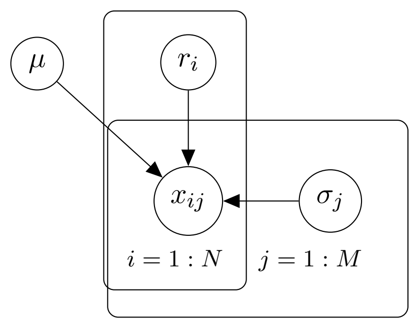
4. **[4 points]** Write an unbiased estimator for the probability that it rains on day 3 given that it rained on day 4. This one is a little tricky.

3. Graphical model notation.

1. **[2 points]** Draw the DAG corresponding to the following factorization of a joint distribution:

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

2. **[2 points]** Write the factorized joint distribution implied by the following graphical model with plate notation:



4. Decision Theory. Imagine we are running a nuclear power plant that is undergoing a malfunction. We have two options: A) Vent the core, and B) do nothing.

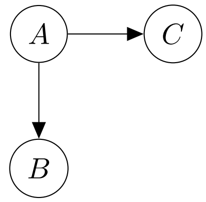
Our current beliefs are that the amount of radiation in the core is uniform between 10 and 20 units, i.e.

$$R|\text{vent} \sim U(10, 20)$$

If we do nothing, there is a $X\%$ chance that no radiation will be released, and a $(1 - X)\%$ that 100 units of radiation will be released.

1. [**2 points**] For what range of probabilities X would venting the core release less radiation in expectation?

5. Ancestral Sampling.



1. **[2 points]** Describe how you would generate a sample (A, B, C) from the DAG above using ancestral sampling.

2. **[2 points]** Describe how you would generate a sample (A, B, C) from the DAG above using ancestral sampling and rejection sampling given that $C = 5$.

3. **[1 point]** What is the joint probability of (A=0, B=0, C=5) based on the DAG and conditional probability tables below.

a	0	1	2	3
P(A=a)	0.3	0.3	0.3	0.1

$P(B = b \mid A = a)$	a= 0	a=1	a=2	a=3
b=-1	0.7	0.55	0.5	0.25
b=0	0.25	0.3	0.2	0.25
b=1	0.05	0.15	0.3	0.5

$P(C = c \mid A = a)$	a = 0	a = 1	a = 2	a = 3
c = 5	0.9	0.75	0.2	0.15
c = 10	0.05	0.1	0.7	0.15
c = 15	0.05	0.15	0.1	0.7

6. Maximum Likelihood. The probability density function of a random variable x distributed according to an exponential distribution with parameter $\theta > 0$ is:

$$p(x) = \theta e^{-\theta x} \quad \text{for} \quad x \geq 0.$$

(a) (6 pts) Assume that we observed $x_1, x_2, ..., x_n$ i.i.d. draws from an exponential distribution with unknown parameter θ. Find the maximum likelihood estimator for θ.