10-601: Homework 7

Due: 17 November 2014 11:59pm (Autolab)
TAs: Kuo Liu, Harry Gifford

Name: Dawei Wang
Andrew ID: daweiwan@andrew.cmu.edu

Please answer to the point, and do not spend time/space giving irrelevant details. Please state any additional assumptions you make while answering the questions. For Questions 1 to 5, 6(b) and 6(c), you need to submit your answers in a single PDF file on autolab, either a scanned handwritten version or a LaTeXpdf file. Please make sure you write legibly for grading. For Question 6(a), submit your m-files on autolab.

You can work in groups. However, no written notes can be shared, or taken during group discussions. You may ask clarifying questions on Piazza. However, under no circumstances should you reveal any part of the answer publicly on Piazza or any other public website. The intention of this policy is to facilitate learning, not circumvent it. Any incidents of plagiarism will be handled in accordance with CMU's Policy on Academic Integrity.

$\star$: Code of Conduct Declaration

- Did you receive any help whatsoever from anyone in solving this assignment? No.
- Did you give any help whatsoever to anyone in solving this assignment? No.

Dawei Wang
daweiwan@andrew.cmu.edu

1: Warmup (TA:- Either)

Which of the following independence statements are true in the following directed graphical model? Explain your answer. $A \perp B \mid C$ can be read as "A is independent of B given C".

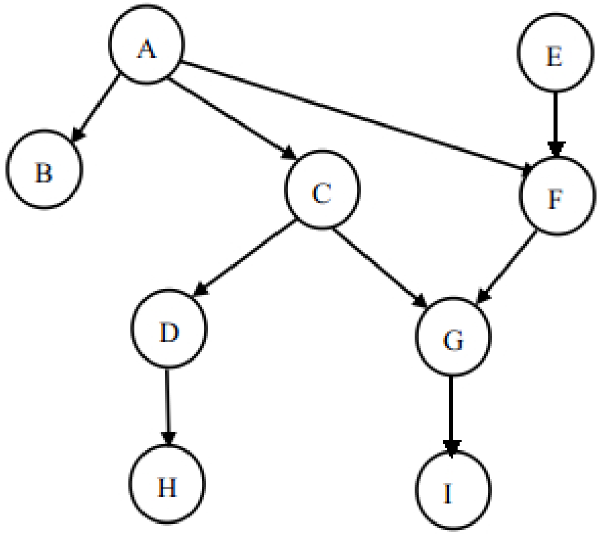


Figure 1: A simple DAG

(a) $A \perp E$ *[3 points]*

TRUE. There doesn't exist a path between A and E that doesn't contain a collider.

(b) $A \perp E \mid G$ *[3 points]*

FALSE. G is a collider, and there exists a path A-C-G-F-E that contains it.

(c) $C \perp F \mid \{A, G\}$ *[3 points]*

FALSE. G is a collider, and there exists a path C-G-F that contains it.

(d) $B \perp F \mid \{C, E\}$ *[3 points]*

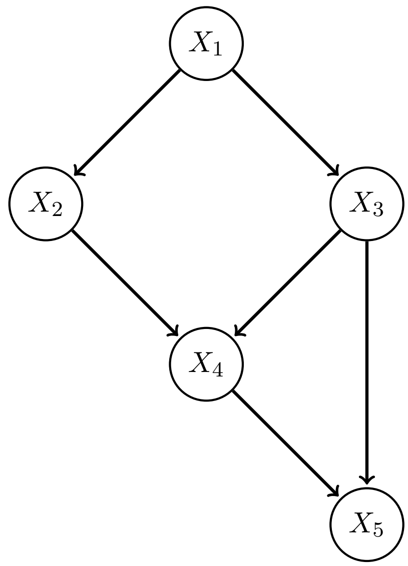
FALSE. There exists a path B-A-F that doesn't contain a collider nor C and E.

(e) $A \perp D \mid C$ *[3 points]*

TRUE. C is not a collider, and there doens't exist a path that doesn't contain C.

2: Bayesian Networks (TA:- Harry Gifford)

(a) Prove that the any Bayesian network represents a valid probability distribution. Your proof should be general enough to apply to any graphical model, but to avoid clunky notation you may just prove it for the following graphical model. Specifically, you should show $\forall i, \forall X_i, \Pr(X_1, ..., X_n) \geq 0$ and $\sum_{X_1} \sum_{X_2} ... \sum_{X_n} \Pr(X_1, ..., X_n) = 1$.



[*10 points*]

$$\begin{aligned}\Pr(X_1, \ldots, X_n) &= \prod_{X_i} \Pr\left(X_i \middle| \text{Parents}\left(X_i\right)\right) \\ &= \Pr(X_1)\Pr(X_2|X_1)\Pr(X_3|X_1)\Pr(X_4|X_2,X_3)\Pr(X_5|X_4,X_3) \geq 0\end{aligned}$$

These conditional probabilities are equal or greater than zero according to the probability axioms.

$$\begin{aligned}&\sum_{X_1}\sum_{X_2}\cdots\sum_{X_n}\Pr(X_1, \ldots, X_n) \\ =&\sum_{X_1}\sum_{X_2}\cdots\sum_{X_5}\Pr(X_1)\Pr(X_2|X_1)\Pr(X_3|X_1)\Pr(X_4|X_2,X_3)\Pr(X_5|X_4,X_3) \\ =&\sum_{X_1}\Pr(X_1)\sum_{X_2}Pr(X_2|X_1)\sum_{X_3}\Pr(X_3|X_1)\sum_{X_4}\Pr(X_4|X_2,X_3)\sum_{X_5}\Pr(X_5|X_4,X_3) \\ =&\sum_{X_1}\Pr(X_1)\sum_{X_2}Pr(X_2|X_1)\sum_{X_3}\Pr(X_3|X_1)\sum_{X_4}\Pr(X_4|X_2,X_3)\cdot 1 = \cdots = 1\end{aligned}$$

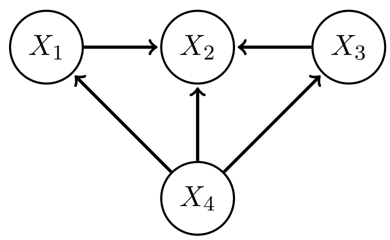
(b) Now we will explore the idea of equivalence for Bayesian networks. First some definitions. We say that two graphical models G_1 and G_2 are *Markov equivalent* if every independence statement in G_1 is also expressed in G_2 and likewise for G_2 into G_1. We define a *V-configuration*, $\langle i, j, k\rangle$ as a subgraph with three vertices and two edges connecting i, j and j, k. We say a V-configuration $\langle i, j, k\rangle$ is *shielded* if i is connected to k or k is connected to i. Two graphs have the same *skeleton* if the graphs obtained from removing the directions from the edges are the same.

Finally, we say two graphs G_1 and G_2 are Markov Equivalent iff

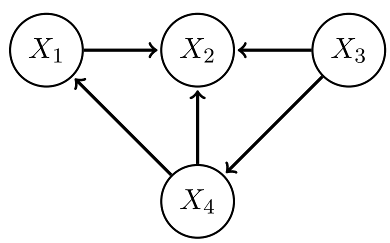
1. G_1 and G_2 have the same skeleton.
2. G_1 and G_2 have the same unshielded collider (i.e. $i \rightarrow j \leftarrow k$) V-configurations.

For each pair of graphs below (i.e. (G_1, G_2), (G_1, G_3), (G_2, G_3)) state whether they are Markov equivalent or not. Explain your answer.

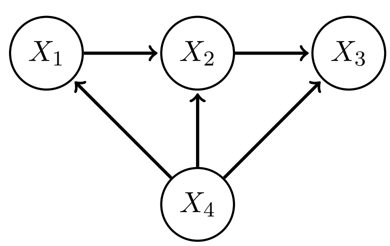
G_1



G_2



G_3



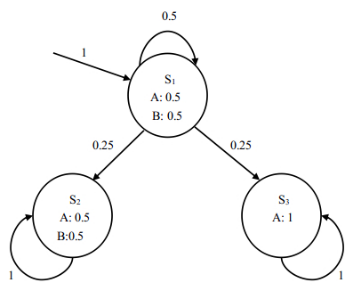
[5 points]

G_1 and G_2 are equivalent. First of all, all three graphs apparently have the same skeleton; G_1 has one unshielded collider X_1-X_2-X_3 V-configuration, so does G_2. However, G_3 doesn't have any such unshielded collider V-configurations. Therefore G_1 and G_2 are equivalent.

3: HMM I (TA:- Kuo Liu)

You have already learned forward method and backward method to compute the probability for a given observed sequence: $P(O_1...O_T)$

In this problem, we want to give you a different perspective of view to do this job and will use this new way to compute some probabilities for the following example HMM.



(a) Let $v_i^t = p(O_1...O_T|q_t = s_i)$. Write a formula for $P(O_1...O_T)$ using **only** v_i^t and $p_t(i)$. [we define $p_t(i)$=$p(q_t = s_i)$] *[3 points]*

$$\Pr(O_1, \ldots, O_T) = \sum_i P(O_1, \ldots, O_T|q_t = s_i)p(q_t = s_i)$$

(b) Compute $p(O_1 = B, ..., O_{200} = B)$ (observing 200 B's in the sequence) *[6 points]*

Observation: If the model is in S_1 at the i-th iteration, it can never reach any other states before that iteration - otherwise it won't be able to go back to S_1; if the model ends up in state S_2, at some point $t = t_0$, the model will go from S_1 to S_2, with a probability of 0.25. $q_{200} \neq S_3$ since the model in S_3 doesn't emit B. Also, as long as the model is in state S_1 or S_2, B emits with $\Pr = 0.5$.

$$\begin{aligned}
\Pr(O_1 = B, \ldots, O_{200} = B) =& \Pr(O_1 = B, \ldots, O_{200} = B|q_{200} = S_1) \Pr(q_{200} = S_1) \\
&+ \Pr(O_1 = B, \ldots, O_{200} = B|q_{200} = S_2) \Pr(q_{200} = S_2) \\
=& \left(0.5^{200}\right)\left(1 \times 0.5^{199}\right) + \left(0.5^{200}\right)\left(\textstyle\sum_{t_0=2}^{200} 1 \cdot 0.5^{t_0-2} \cdot 0.25 \cdot 1^{200-t_0}\right) \\
=& 0.5^{400} + 0.5^{201}
\end{aligned}$$

(c) Compute $p(O_1 = A, ..., O_{200} = A)$ (observing 200 A's in the sequence) *[6 points]*

Observation: Use similar idea with (b), but note that here the model can reach S_3 at some point, and the probabilities for emitting A are different between S_3 and S_1.

$$\begin{aligned}
\Pr(O_1 = A, \ldots, O_{200} = A) =& \Pr(O_1 = A, \ldots, O_{200} = A|q_{200} = S_1) \Pr(q_{200} = S_1) \\
&+ \Pr(O_1 = A, \ldots, O_{200} = A|q_{200} = S_2) \Pr(q_{200} = S_2) \\
&+ \textstyle\sum_{t_0=2}^{200} \left[\Pr(O_1 = A, \ldots, O_{200} = A|q_{200} = S_3, t = t_0) \Pr(q_{200} = S_3, t = t_0)\right] \\
=& \left(0.5^{200}\right)\left(1 \times 0.5^{199}\right) + \left(0.5^{200}\right)\left(\textstyle\sum_{t_0=2}^{200} 1 \cdot 0.5^{t_0-2} \cdot 0.25 \cdot 1^{200-t_0}\right) \\
&+ \left(\textstyle\sum_{t_0=2}^{200} 0.5^{t_0-1} 1^{200-(t_0-1)} \cdot 1 \cdot 0.5^{t_0-2} \cdot 0.25 \cdot 1^{200-t_0}\right) \\
=& \frac{1}{6} + \frac{1}{3} \cdot 0.5^{400} + 0.5^{201}
\end{aligned}$$

4: HMM II (TA:- Kuo Liu)

Consider the HMM defined by the transition and emission probabilities in the table below. This HMM has six states (plus a start and end states) and an alphabet with four symbols (A,C,G and T). Thus, the probability of transitioning from state S_1 to state S_2 is 1, and the probability of emitting A while in state $S1$ is 0.5.

	0	S_1	S_2	S_3	S_4	S_5	S_6	f	A	C	G	T
0	0	1	0	0	0	0	0	0				
S_1	0	0	1	0	0	0	0	0	0.5	0.3	0	0.2
S_2	0	0	0	0.3	0	0.7	0	0	0.1	0.1	0.2	0.6
S_3	0	0	0	0	1	0	0	0	0.2	0	0.1	0.7
S_4	0	0	0	0	0	0	0	1	0.1	0.3	0.4	0.2
S_5	0	0	0	0	0	0	1	0	0.1	0.3	0.3	0.3
S_6	0	0	0	0	0	0	0	1	0.2	0.3	0	0.5

State whether the following are true or false and explain your answer.

(a) $P(O_1 = A, O_2 = C, O_3 = T, O_4 = A, q_1 = S_1, q_2 = S_2) =$
$P(O_1 = A, O_2 = C, O_3 = T, O_4 = A | q_1 = S_1, q_2 = S_2)$ [*4 points*]

TRUE. Since $p(q_1 = S_1, q_2 = S_2) = 1$.

(b) $P(O_1 = A, O_2 = C, O_3 = T, O_4 = A, q_3 = S_3, q_4 = S_4) >$
$P(O_1 = A, O_2 = C, O_3 = T, O_4 = A | q_3 = S_3, q_4 = S_4)$ [*4 points*]

FALSE. Since $p(q_3 = S_3, q_4 = S_4) = p(q_4 = S_4 | q_3 = S_3)p(q_3 = S_3) = p(q_3 = S_3) < 1$.

(c) $P(O_1 = A, O_2 = C, O_3 = T, O_4 = A, q_3 = S_3, q_4 = S_4) <$
$P(O_1 = A, O_2 = C, O_3 = T, O_4 = A | q_3 = S_5, q_4 = S_6)$ [*4 points*]

TRUE. Since LHS $= 1 \cdot 1 \cdot 0.3 \cdot 1 \cdot 0.7 \cdot 0.1 < 1 \cdot 1 \cdot 0.3 \cdot 0.2 =$ RHS.

(d) $P(O_1 = A, O_2 = C, O_3 = T, O_4 = A) > P(O_1 = A, O_2 = C, O_3 = T, O_4 = A, q_3 = S_3, q_4 = S_4)$ [*4 points*]

TRUE. Since LHS $= \Pr(\dots, q_3 = S_3, q_4 = S_4) + \Pr(\dots, q_3 = S_5, q_4 = S_6) >$ RHS.

(e) $P(O_1 = A, O_2 = C, O_3 = T, O_4 = A) > P(O_1 = A, O_2 = C, O_3 = T, O_4 = A | q_3 = S_3, q_4 = S_4)$ [*4 points*]

FALSE. Since $\Pr(\dots, q_3 = S_3, q_4 = S_4) = 0.7 \cdot 0.1 > 0.3 \cdot 0.2 = \Pr(\dots, q_3 = S_5, q_4 = S_6)$, so conditioning on $q_3 = S_3, q_4 = S_4$ actually results in a probability higher than average (LHS).

Dawei Wang
daweiwan@andrew.cmu.edu

5: HMM II (TA:- Kuo Liu)

Consider the following two state HMM, answer the following questions. You can either get the answer by hand calculation or write a program to get the final answer.

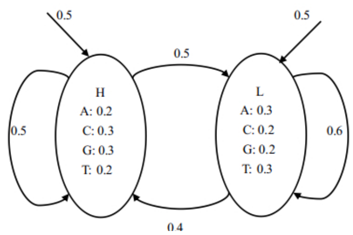


Figure 2: Figure of Q5

(a) What is the probability for you to get an output sequence like **GGCA** [*7 points*]

$\Pr = 0.0013767 + 0.0024665 = 0.0038432$. See `po.m` for the code that gave this result.

(b) What is the most likely hidden status sequence for the output sequence **GGCACTGAA** [*8 points*]

See the table below for $\log_2 \delta_t(j)$. Probabilities on the most probable path are italicized. The most likely hidden status sequence is therefore `HHHLLLLLL`. See `pqao.m` for the code.

	G	G	C	A	C	T	G	A	A
H	*-2.7370*	*-5.4739*	*-8.2109*	-11.5328	-14.0068	-17.3287	-19.5396	-22.8615	-25.6574
L	-3.3219	-6.0589	-8.7959	*-10.9479*	*-14.0068*	*-16.4807*	*-19.5396*	*-22.0135*	*-24.4874*

po.m

```

a = [0.5, 0.5; 0.4, 0.6]; p = [0.5; 0.5];
b = [0.2, 0.3, 0.3, 0.2; 0.3, 0.2, 0.2, 0.3];
o = [3, 3, 2, 1];

alpha(:, 1) = b(:, o(1)) .* p;
for i = 2: 4
  alpha(:, i) = b(:, o(i)) .* (a' * alpha(:, i - 1));
end
```

pqao.m

```

a = [0.5, 0.5; 0.4, 0.6]; p = [0.5; 0.5];
b = [0.2, 0.3, 0.3, 0.2; 0.3, 0.2, 0.2, 0.3];
o = [3, 3, 2, 1, 2, 4, 3, 1, 1];

warning('off');
d(:, 1) = b(:, o(1)) .* p;
for i = 2: 9
  [d(:, i), im]  = max(b(:, o(i))' .* a .* d(:, i - 1));
end
```