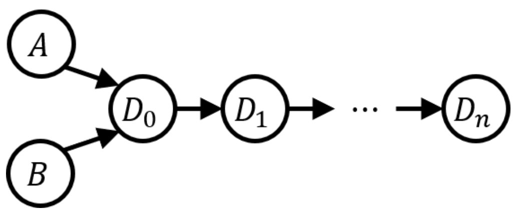
COMS W4701: Artificial Intelligence, Spring 2024

Homework 5

Instructions: Compile all written solutions for this assignment in a single, typed PDF file. If you use Python or any other computational tool, append the code to your PDF file. Turn in your submission on Gradescope, and **tag all pages**. Please be mindful of the deadline and late policy, as well as our policies on citations and academic honesty.

Problem 1: Descendants of Colliders (20 points)

Here we will investigate the absence of conditional independence guarantees between two random variables in a Bayes net when an arbitrary common descendant is observed. We will consider the case of a simple chain of descendants, as shown below.



Suppose that all random variables are binary. The marginal distributions of A and B are both uniform $(0.5, 0.5)$, and the CPTs of the collider D_0 and its descendants are as follows:

A	B	$\Pr(+d_0 \mid A, B)$
$+a$	$+b$	0.5
$+a$	$-b$	0.0
$-a$	$+b$	0.0
$-a$	$-b$	0.5

D_{i-1}	$\Pr(+d_i \mid D_{i-1})$
$+d_{i-1}$	0.0
$-d_{i-1}$	1.0

1. Compute the distributions $\Pr(A, B)$ and $\Pr(A, B | -d_0)$. Show whether A is conditionally independent of B given D_0.

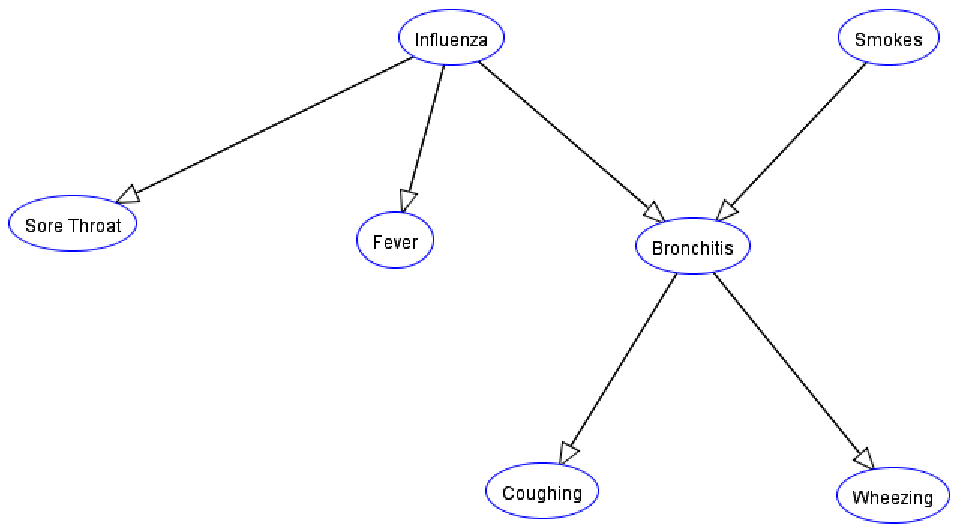
2. Find an analytical expression for the joint probabilities $\Pr(A, B, d_i)$, where $0 \leq i \leq n$. The expression should only contain terms in the Bayes net CPTs and involve chain rule and/or marginalization operations.

3. Now using the values from the given CPTs, numerically compute the joint conditional distribution $\Pr(A, B | -d_i)$, where i is odd. You can give a brief explanation for how you simplify the analytical expression. Show whether A is conditionally independent of B given D_i.

4. Finally, numerically compute the joint conditional distribution $\Pr(A, B | -d_0, +d_i)$, where i is odd. Again, you may explain how you can do so by simplifying the relevant analytical expression. Do **not** refer to the local independence rule of chains from class. Show how we can conclude that A and B are conditionally independent of D_i given D_0.

Problem 2: Diagnostic Bayes Net (24 points)

The following Bayes net is the "Simple Diagnostic Example" from the Sample Problems of the Belief and Decision Networks tool on AIspace. All variables are binary.



1. We are interested in computing the distribution Pr(Coughing | Fever = True). Find a minimal analytical expression for this distribution (or its unnormalized version) in terms of the Bayes net CPTs, and then rewrite each term as a factor explicitly showing the variable dependencies. What is the maximum size of the intermediate factor if all marginalization is done at the end?

2. We employ variable elimination to solve the query above. Consider the following two variable orderings (you may only need to sum over a subset of each):

   (a) influenza, smokes, sore throat, fever, bronchitis, coughing, wheezing

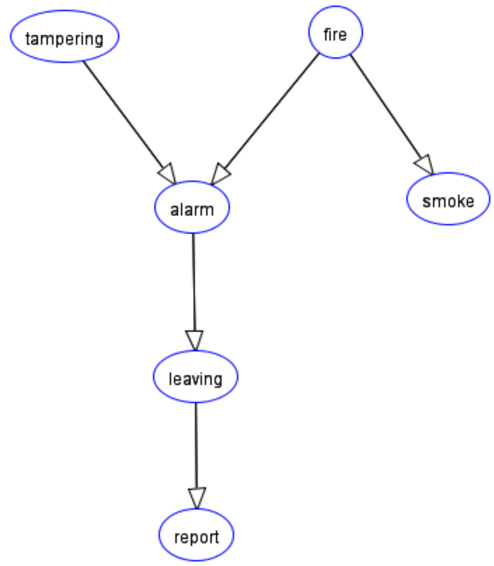
   (b) wheezing, coughing, bronchitis, fever, sore throat, smokes, influenza

   For each of these orderings, rewrite your analytical expression by splitting up the sum over each variable, and factor out the maximal number of terms from each summation. Remember that sums are evaluated from right to left, or from inside out. What is the size of the largest intermediate factor in each?

3. Follow the more efficient ordering from above to numerically compute Pr(Coughing | Fever = True) using the applet parameters. You can use the applet to verify that this comes out to $(0.45, 0.55)$, but you should carry out the computations yourself, either manually or using Python (please attach your notebook/code in your PDF if the latter). In addition to the final result, show each intermediate factor resulting from each variable summation step.

Problem 3: Fire Alarm Bayes Net (20 points)

The following Bayes net is the "Fire Alarm Belief Network" from the Sample Problems of the Belief and Decision Networks tool on AIspace. All variables are binary.



1. Consider using likelihood weighting to solve for a posterior distribution conditioned on the following sets of evidence: i) Tampering and Fire, ii) Smoke and Report. For each case, explain whether the generated samples reflect the prior or posterior distribution. Also explain the impact of the likelihood weights in helping us to estimate the posterior in each case.

2. Inspect the individual CPTs of this Bayes net in the applet. Identify all unique evidence events that are impossible to occur. (You don't need to list any events that contain those you already listed as subsets. For example, if you identify $\{+t, +f\}$ to be such an event, you do not need to write $\{+t, +f, +a\}$, $\{+t, +f, +s\}$, etc.) What would happen if we perform rejection sampling or likelihood weighting with these events as observed?

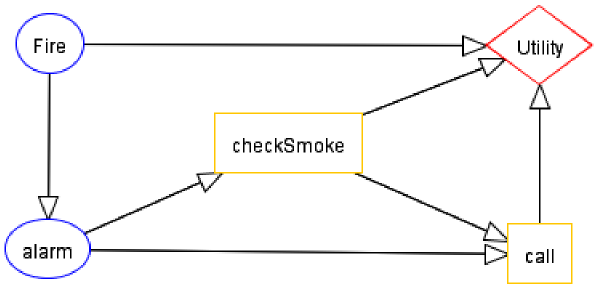
3. We perform Gibbs sampling and want to resample the Fire variable conditioned on the current sample $(+t, +a, +s, +l, +r)$. Find a minimal analytical expression for $\Pr(\text{Fire} \mid \textit{sample})$ (or its unnormalized version) in terms of the Bayes net CPTs, and then numerically compute it using the applet parameters. You can use the applet to verify that this comes out to $(0.35, 0.65)$, but you should carry out the computations yourself and show your work.

4. Repeat the above, but now suppose that we have lost the value of Alarm in the sample. In other words, compute the distribution $\Pr(\text{Fire} \mid +t, +s, +l, +r)$. Why does conditioning on less information lead to potentially *greater* computation?

Problem 4: Fire Alarm Decision Network (16 points)

The following Bayes net is a simplified version of the "Fire Alarm Decision Problem" from the Sample Problems of the Belief and Decision Networks tool on AIspace. All variables are binary, and the chance node CPTs as well as utility values are shown below.

While you are not required to do so, we recommend that you use `pandas` functionality to carry out the operations in this and the next problem. It is also acceptable if you prefer to perform all calculations manually. Please include your work in either case.



Fire	CheckSmoke	Call	Utility
True	True	True	-220
True	True	False	-5020
True	False	True	-200
True	False	False	-5000
False	True	True	-220
False	True	False	-20
False	False	True	-200
False	False	False	0

Fire	Pr(Fire)
T	0.01
F	0.99

Fire	Alarm	Pr(Alarm | Fire)
True	True	0.95
True	False	0.05
False	True	0.01
False	False	0.99

1. Compute the joint factor required to sum out the chance nodes that are not parents of any decision nodes. Show the resultant factor after performing the elimination.

2. Determine the optimal decision function and expected utilities for Call. Show the table containing this information.

3. Briefly explain whether the optimal decision function for Call actually depends on both of its parent nodes. Would the result change if we had determined the optimal decision function for CheckSmoke first, followed by Call?

4. Determine the optimal decision function and expected utilities for CheckSmoke. Show the table containing this information. Then compute the maximum expected utility.

Problem 5: Fire Alarm Decision Network Redux (20 points)

Now suppose that the Call decision also depends on Fire; in other words, we get to observe Fire prior to making the decision.

1. Explain why we should not eliminate any chance nodes prior to computing optimal decision functions in this modified network.

2. Determine the optimal decision function and expected utilities for Call. Show the table containing this information.

3. Eliminate any chance nodes that are not parents of the remaining decision node, CheckSmoke. Show the resultant factor.

4. Determine the optimal decision function and expected utilities for CheckSmoke. Show the table containing this information. Then compute the maximum expected utility, as well as the VPI of Fire for the Call decision.

5. Give an interpretation for the VPI of Fire. Why might it be beneficial to be able to observe Fire directly, in addition to (or instead of) relying on the "after-effects" of a Fire event?