COGS 300 — Fall 2013
Assignment 1

Due: 11:00am, Tuesday 29 October 2013 (hardcopy in class and email to cogs300@gmail.com)

This can be done in groups of size 1, 2 or 3. Working alone is not recommended. A group of size n can choose any $n + 1$ questions from questions 1-5. All members of the group need to be able to explain the group's answer. Please look at all of the questions, as the final exam will assume that you have thought about all of the questions. Everyone should do question 6. The group should hand in one copy of the assignment.

You are encouraged to discuss this assignment and collaborate with other classmates, as long as (a) you list the people with whom you discussed the assignment and (b) you give your own answers and explanations. Please post questions to the Connect web site.

Question 1

In the early 1970's there was a controversy about sex biases for graduate admissions at UC Berkeley. This example is based on that case, but the numbers are fictional.

There are two departments, which we will call $dept\#1$ and $dept\#2$ (so $Dept$ is a random variable with values $dept\#1$ and $dept\#2$) which students can apply to. Assume students apply to one, but not both. Students have a sex (male or female), and are either admitted or not. Suppose we have the following table of number of students in each category:

Dept	Sex	Admitted	Number
dept#1	male	true	90
dept#1	male	false	50
dept#1	female	true	20
dept#1	female	false	10
dept#2	male	true	15
dept#2	male	false	40
dept#2	female	true	20
dept#2	female	false	40

In the semantics of possible worlds, we will treat the students as possible worlds, each with the same measure.

(a) What is $P(Admitted = true \mid Sex = male)$? What is $P(Admitted = true \mid Sex = female)$? Which sex is more likely to be admitted?

(b) What is $P(Admitted = true \mid Sex = male, Dept = dept\#1)$? What is $P(Admitted = true \mid Sex = female, Dept = dept\#1)$? Which sex is more likely to be admitted to $dept\#1$?

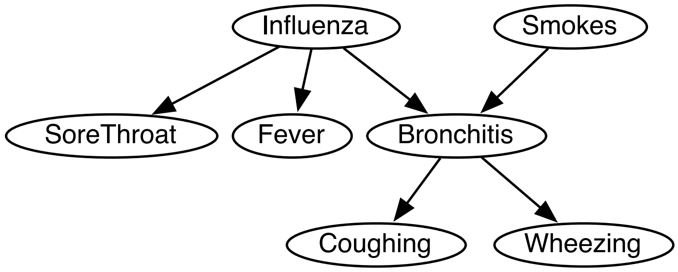
(c) What is $P(Admitted = true \mid Sex = male, Dept = dept\#2)$? What is $P(Admitted = true \mid Sex = female, Dept = dept\#2)$? Which sex is more likely to be admitted to $dept\#2$?

(d) This is an instance of Simpson's paradox. Why is it a paradox? Explain why it happened in this case.

(e) Give another scenario where Simpson's paradox likely occurs. [Don't just use one from the Internet!]

Question 2

Consider the following Bayesian network:



This can be loaded into the AISpace Belief network tool at `http://www.aispace.org/bayes/`, as follows: From "File" choose "Load Sample Problem", then "Simple diagnostic example". To see the posterior distribution of variables: in "solve mode", toggle monitoring and look at the distribution of each variable.

(a) The posterior probabilities of which variables change when *Smokes* is observed to be true? [I.e., give the variables X such that $P(X \mid Smoke = true) \neq P(X)$?]

(b) Starting from the original network, the posterior probabilities of which variables change when *Fever* is observed to be true? [I.e., specify the X where $P(X \mid Fever = true) \neq P(X)$?]

(c) Does the probability of *Fever* change when *Wheezing* is observed to be true? [I.e., is $P(Fever \mid Wheezing = true) \neq P(Fever)$?] Explain why (in terms of the domain; in language that could be understood by someone who didn't know about Bayesian networks).

(d) Suppose *Wheezing* is observed to be true. Does the observing *Fever* change the probability of *Smokes*? [I.e., is $P(Smokes \mid Wheezing) \neq P(Smokes \mid Wheezing, Fever)$?] Explain why (in terms that could be understood by someone who didn't know about Bayesian networks).

(e) What could be observed so that subsequently observing *Wheezing* does not change the probability of *SoreThroat*. [That is, specify a variable or variables X such that $P(SoreThroat \mid X) = P(SoreThroat \mid X, Wheezing)$, or state that there are none.] Explain why.

(f) Suppose *Allergies* could be another explanation of *Sore Throat*. Change the diagram so that *Allergies* also affects *Sore Throat* but is independent of the other variables in the network. Give reasonable probabilities. [You need to move to Create mode to do this.] You need to hand in the resulting network and show what new probabilities are specified.

(g) What could be observed so that observing *Wheezing* changes the probability of *Allergies*? Explain why.

(h) What could be observed so that observing *Smokes* changes the probability of *Allergies*? Explain why.

Note that parts (a), (b), (c) only involve observing a single variable.

Question 3

D. Kahneman [Thinking Fast and Slow, 2011, p. 166] gives the following example:

"A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

- 85% of the cabs in the city are Green and 15% are Blue.

- A witness identified the cab as Blue. The court tested the reliability of the witness in the circumstances that existed on the night of the accident and concluded that the witness correctly identifies each one of the two colours 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue?"

(a) Represent this story as a belief network. Explain all variables and conditional probabilities. What is observed, what is the answer?

(b) Suppose there were three independent witnesses, two of which claimed the cab was Blue and one of whom claimed the cab was Green. Show the corresponding belief network. What is the probability the cab was Blue? What if all three claimed the cab was Blue?

(c) Suppose it was found that the two witnesses who claimed the cab was Blue were not independent, but there was a 60% chance they colluded. (What might this mean?) Show the corresponding belief network, and the relevant probabilities. What is the probability that the cab is Blue, (both for the case where all three witnesses claim that cab was Blue and the case where the other witness claimed the cab was Green)?

(d) In a variant of this scenario, Kahneman [p 167] replaced the first condition with: "The two companies operate the same number of cabs, but Green cabs are involved in 85% of the accidents." How can this new scenario be represented as a belief network? Your belief network should allow observations about whether there is an accident as well as the colour of the taxi. Show examples of inferences in your network. Make reasonable choices for anything that is not fully specified. Be explicit about any assumptions you make.

Question 4

Give a possible exam question (perhaps with sub-parts) to test students either about control, hierarchical control, probability, belief networks and/or causality. It should be worth 15 marks, and take students approximately 15 minutes to complete in an exam setting. It must be clear what the question is asking for and must be self-contained. Give a solution and a marking scheme (how much each part is worth).

Question 5

Play with something and report what you found.

On the wiki `http://wiki.ubc.ca/Course:COGS300`, create some online learning resources. You should create pedagogical or real-world examples that use useful for COGS 300 students to learn about agents and control, hierarchical control or probabilistic or causal reasoning. Please add references for any external resources used. You will need to login with your CWL to edit. This is intended to be an open-ended creative question. This is a cooperative question, as anyone can edit other people's questions. It is possible to gain credit by improving other's contributions. Please help to build a useful resource.

Explain clearly what your contribution was. It can be worth multiple questions; justify any claim of how many questions your contribution is worth. It is even possible to do the whole assignment just by creating useful resources.

Question 6

How long did this assignment take? What did you learn? Was it reasonable?