COMP-424: Artificial intelligence

Homework 4

Due on *myCourses* Tuesday, Apr 13, 9:00pm.

Due to the overlap with the final project submission deadline, you may submit A4 without penalty until Friday, Apr 16, 9:00pm.

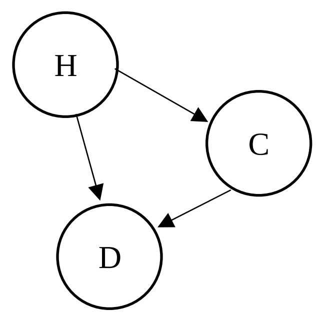
General instructions.

- This is an individual assignment. You can discuss solutions with your classmates, but should only exchange information orally, or else if in writing through the discussion board on *myCourses*. All other forms of written exchange are prohibited.
- Unless otherwise mentioned, the only sources you should need to answer these questions are your course notes, the textbook, and the links provided. Any other source used should be acknowledged with proper referencing style in your submitted solution.
- Submit a single pdf document containing all your pages of your written solution on your McGill's *myCourses* account. You can scan-in hand-written pages. If necessary, learn how to combine many pdf files into one.

Question 1: Health Behaviours

Consider the following causal graphical model involving three Bernoulli random variables, which is a simple model of health status and behaviours: H (health status), C (cautious behaviour), D (disease).

People's health status influences whether they adopt cautious behaviour, and their health status together with their behaviour influence their probability of disease.



We collect an observational dataset of a population of 1000 people:

H	**C**	**D**	**#instances**
0	0	0	10
0	0	1	13
0	1	0	44
0	1	1	28
1	0	0	689
1	0	1	124
1	1	0	84
1	1	1	8

a) Using maximum likelihood estimation, estimate the observed conditional probabilities of disease given cautiousness: P(D=1|C=1) and P(D=1|C=0)

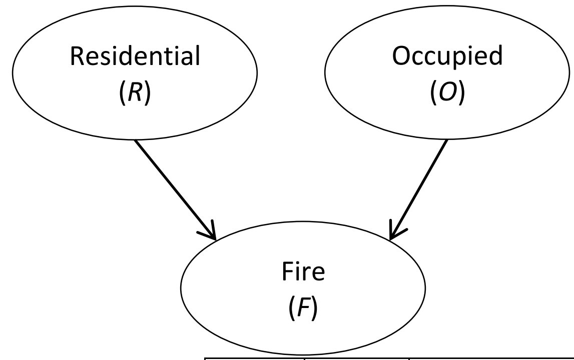
b) Suppose that we can intervene and perfectly persuade people to be cautious or not. Estimate P(D=1|do(C=1)) and P(D=1|do(C=0)) using maximum likelihood estimation. What is the relative risk reduction of adopting cautious behaviour? (RRR = 1 – (P(D=1|do(C=1)) / P(D=1|do(C=0)))

c) A talk show host on TV points to your results in part a) to say that there is no point in being cautious. How do you rebut this argument?

R (F)	**R (T)**
0.25	0.75

O (F)	**O (T)**
0.15	0.85

Question 2: Fire Hazard



R	**O**	**F (T)**
F	F	0.005
F	T	0.02
T	F	0.005
T	T	0.01

Consider the Bayes Net shown here, with all Bernoulli variables, which involve building types and their risk of fire. Having a fire has a utility of -1000 if the building was insured, but has a utility of -50000 if the building was not insured (or the insurance claim is denied). Having insurance when there is no fire has a utility of -50, and not having insurance and no fire has a utility of 0.

Use the principle of Maximum Expected Utility and Value of Information to answer the following questions. For parts a)-c) assume the insurance company pays for 100% of the cases.

a) Given no information about whether a building is residential or whether it is occupied, how much should the insurance company charge to insure the building to break even, in utility points?
b) How much should the company charge for insurance if you know for certain that the building is commercial (R=F) and occupied (O=T) to break even?
c) How much should they charge if the building is residential (R=T)?
d) A company is offering cheaper insurance but has a reputation of rejecting 25% of insurance claims. How much should they charge for this insurance, to make it competitive with the insurance offered by the more reliable company? (*Hint*: Set the cost of the new insurance to have the same MEU as the other insurance.)

Question 3: Bandits

Consider the following 6-armed bandit problem. The initial value estimates of the arms are given by Q = {1, 2, 2, 1, 0, 3}, and the actions are represented by A = {1, 2, 3, 4, 5, 6}. Suppose we observe that each lever is played in turn: (from lever 1 to lever 6, and then start from lever 1 again):

$$A_t = ((t-1) \, mod \; 6) \; + 1 \qquad (1)$$

We also observe that the rewards R_t seem to fit the following function:

$$R_t = \; 2\cos\left[\frac{\pi}{6}(t-1)\right] \qquad (2)$$

So, the first two action-reward pairs are $A_1 = 1, \; R_1 = 2$, and $A_2 = 2, \; R_2 = \sqrt{3}$.

a) Show the estimated Q values from t=1 to t=12 of the trajectory using the average of the observed rewards, where available. Do not consider the initial estimates as samples.

b) It turns out the player was following an ϵ-greedy strategy, which just happened to coincide with the scheme described above in (1) for the first 12 time steps. For each time step t from 1 to 12, report whether it can be concluded with certainty that a random action was selected.

c) Suppose now we continue to visit the levers iteratively as in (1), and that the observed rewards continue to fit the pattern established by (2). Is there a limiting expected reward $Q^*(a)$ for each action $a \in A$ as t approaches infinity? Justify your answer.