CSC 384 A3 Solutions

Module 1

3. Screening Test(short answer) [2 points]

There is a screening test for prostate cancer that looks at the level of PSA (prostate specific antigen) in the blood. There are a number of reasons besides prostate cancer that a man can have elevated PSA levels. In addition, many types of prostate cancer develop so slowly that that they are never a problem. Unfortunately, there is currently no test to distinguish the different types and using the test is controversial because it is hard to quantify the accuracy rates and the harm done by false positives.

For this problem we'll call a positive test a true positive if it catches a dangerous type of prostate cancer. We'll assume the following numbers:

Rate of prostate cancer among men over 50 = 0.0005 (.05%)
True positive rate for the test = 0.9 (90%)
False positive rate for the test = 0.01 (1%)

Let T be the event a man has a positive test and let D be the event a man has a dangerous type of the disease.

a. What is P(T)?
P(T) = 0.9*0.0005 + 0.01 * 0.9995 = 0.010445

b. What is P(D|T)?
P(D|T) = P(T|D)P(D)/P(T) = 0.9 * 0.0005/P(T) = 0.043

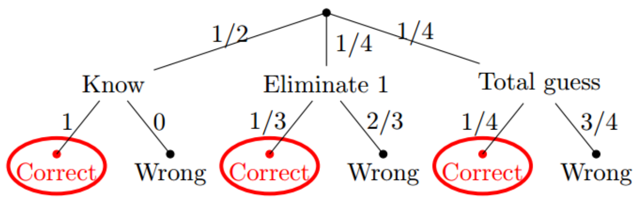
4 Multiple Choice Exam (short answer) [4 points]

A multiple choice exam has 4 choices for each question. A student has studied enough so that the probability they will know the answer to a question is 0.5. The probability that they will be able to eliminate one choice is 0.25; otherwise all 4 choices seem equally plausible. If the student knows the answer they will get the question right. If not, they have to guess.

As the teacher you want the test to measure what the student knows.

1. If the student answers a question correctly, what's the probability they knew the answer?
2. If the student answers a question correctly, what's the probability they didn't know the answer?

We have the following tree:



For a given problem let C be the event the student gets the problem correct and K the event the student knows the answer.

$$P(K|C) = \frac{P(C|K)\,P(K)}{P(C)} = \frac{1 \cdot 1/2}{1/2 + 1/12 + 1/16} = \frac{24}{31} \approx 0.774 = 77.4\%$$

P(~K|C) = P(C|~K)P(~K)/P(C) = [(¼*⅓)+(¼*¼)]/(½+1/12+1/16) = 7/31 = 0.2258

5. Dice [9 points]

Part A

There are four dice in a drawer: one tetrahedron (4 sides), one hexahedron (i.e., cube, 6-sides), and two octahedra (8 sides). Your friend secretly grabs one of the four dice at random. Let S be the number of sides on the chosen die. Now, your friend rolls the chosen die without showing it to you. Let R be the result of the roll.

a. What is P(S = 4 | R = 2) ?

3/8

b. What is P(S = 6 | R = 1) ?

1/4

c. What is P(S = 8 | R = 3) ?

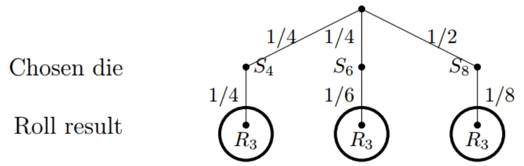
3/8

Solutions:

(b) Bayes' rule says

$$P(S=k|R=3) = \frac{P(R=3|S=k)P(S=k)}{P(R=3)}.$$

We summarize what we know in a tree. In the tree the notation S_4 means the 4-sided die ($S=4$), likewise R_3 means a 3 was rolled ($R=3$). Because we only care about the case $R=3$ the tree does not include other possible rolls.



From the tree and Bayes' Rule, we can get the correct answers.

Module 2

2. Gas Station [4 points]

At a certain gas station 45% of the customers request regular gas, 30% request unleaded gas, and 25% request premium gas. Of those customers requesting regular gas, only 25% fill their tanks. Of those customers requesting unleaded gas, 50% fill their tanks, while of those requesting premium, 60% fill their tanks.

1. What's the probability that the next customer fills the tank? [1 point]

P(F) = P(R)*P(F|R)+P(U)*P(F|U)+P(P)*P(F|P) =
0.45*0.25+0.3*0.5+0.25*0.6 = 0.4125

2. If the next customer fills the tank, what is the probability that regular gas is requested?[1 point]

P(R|F) = P(R^F)/P(F)
=P(R)*P(F|R)/{P(R)*P(F|R)+P(U)*P(F|U)+P(P)*P(F|P)}
=0.45*0.25/(0.45*0.25+0.3*0.5+0.25*0.6) = 0.273

3. If the next customer doesn't fill the tank, what is the probability that unleaded gas is requested?[1 point]

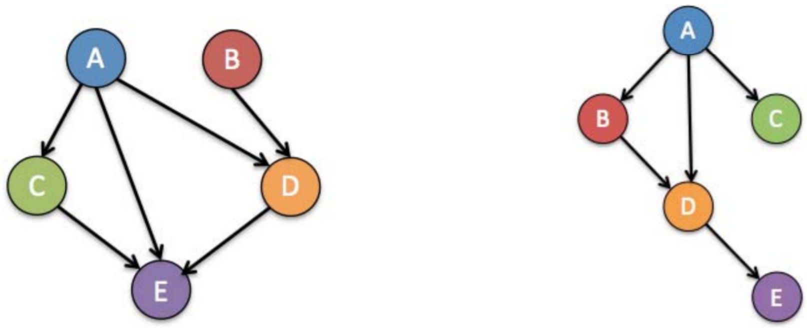
P(U|~F) = P(U^~F)/P(~F)
=P(U)P(~F|U)/(1-P(F)) = 0.255

4. If the next customer fills the tank, what is the probability that premium gas is requested?[1 point]

P(P|F) = P(P)P(F|P)/P(F) = 0.364

4. Gene Expression

You are given two different Bayesian network structures 1 and 2, each consisting of 5 binary random variables A, B, C, D, E. Each variable corresponds to a gene, whose expression can be either "ON" or "OFF".



Network 1 **Network 2**

1. How many parameters will represent the Network 2? [1 point]

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

1+2+2+4+2 = 11 (SAME comment: Look at module 3 Q3)

2. How many parameters will represent the Network 1? [1 point]

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

1+1+2+4+8 = 16

3. Using Network 2 and the probabilities given below, calculate the probability of the following: [3 points]

$$P(A=ON)=0.6$$

$$P(D=ON\mid A,B)=\begin{cases} 0.1 & A=OFF, B=OFF \\ 0.9 & A=ON, B=OFF \\ 0.3 & A=OFF, B=ON \\ 0.95 & A=ON, B=ON \end{cases}$$

$$P(B=ON\mid A)=\begin{cases} 0.1, & A=OFF \\ 0.95, & A=ON \end{cases}$$

$$P(C=ON\mid A)=\begin{cases} 0.8, & A=OFF \\ 0.5, & A=ON \end{cases}$$

$$P(E=ON\mid D)=\begin{cases} 0.8, & D=OFF \\ 0.1, & D=ON \end{cases}$$

a. P(A=ON, B=ON, C=ON, D=ON, E=OFF)

P(A=ON, B=ON, C=ON, D=ON, E=OFF) = P(A=ON)P(B=ON|A=ON)P(C=ON|A=ON)P(D=ON|A=ON,B=ON)P(E=OFF|D=ON)

= 0.6*0.95*0.5*0.95*(1-0.1) = 0.244

b. P(E = OFF | A = ON)

B, D, and E are conditionally independent of C given A, so C drops out. Therefore, we sum over the 4 {B, D} possibilities:
P(E = OFF| A =ON) = Sum _{B,D} {P(E = OFF| D)P(D|A= ON, B)P(B|A=ON)}

B	D	P(B|A=ON)	P(D|A=ON, B)	P(E=OFF|D)	P(E=O,B,D|A=ON)
ON	ON	0.95	0.95	0.9	0.81225
ON	OFF	0.95	0.05	0.2	0.0095
OFF	ON	0.05	0.9	0.9	0.0405
OFF	OFF	0.05	0.1	0.2	0.001

P(E = OFF | A = ON) = 0.86325

c. P(A = ON | E = OFF)

<u>Bayes Rule:</u>
P(A = ON | E = OFF) = P(E = OFF|A=ON)P(A=ON)/P(E=OFF)
=P(E = OFF|A=ON)P(A=ON)/{P(E = OFF|A=ON)P(A=ON)+P(E=OFF|A=OFF)P(A=OFF)

B	D	P(B|A=OFF)	P(D|A=OFF,B)	P(E=OFF|D)	P(E=O,B,D|A=OFF)
ON	ON	0.1	0.3	0.9	0.027
ON	OFF	0.1	0.7	0.2	0.014

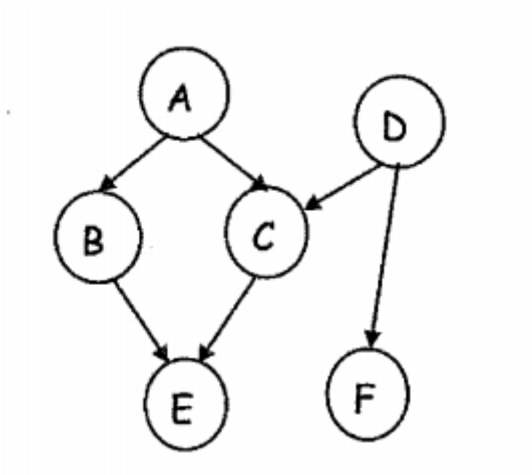
OFF	ON	0.9	0.1	0.9	0.081
OFF	OFF	0.9	0.9	0.2	0.162

P(E=OFF|A=OFF) = 0.284

P(A = ON | E = OFF) = 0.86325 * 0.6/(0.86325 * 0.6 + 0.284 * 0.4) = 0.8201

Module 3

3. <u>Bayes Net</u> [9 points]



a. Suppose A has 3 possible values, B has 2 possible values, C has 4 possible values, D has 4 possible values, F has 2 possible values, and E has 5 possible values, how many values are there in the all of the conditional probability tables combined (and assuming you include entries for all combinations of values)? [1 point]

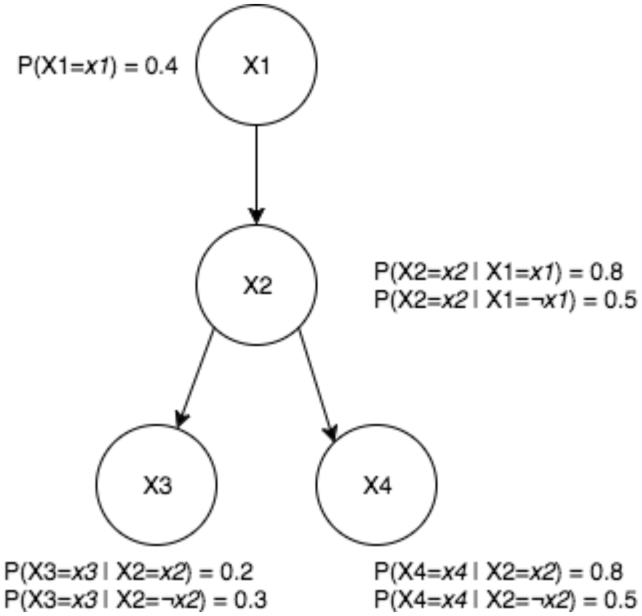
93 (**3 + 2*3 + 4*3*4 + 4 + 2*4 + 5*4*2)**

b. If you had to include entries for every possible combination of values to variables in the joint, what would be the size of the resulting conditional probability table for A,B,C,D and E? [1 point]

960 (2*2*3*4*4*5”)

Module 4

1. Variable elimination Q1 [9 points]



c. Solve P(X2 = x2 | X3 = ¬x3). [4 points] * Hint: Using variable elimination with elimination ordering X4, X1, X2, X3 will greatly simplify your calculations! The two previous questions should also guide your calculations.

Final answer = 0.6509

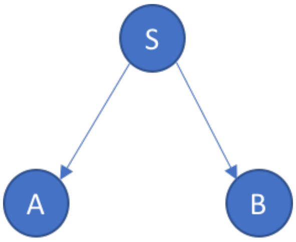
Solution:

- Eliminate X4:
  - f4(X2) = SUM_{X4} P(X4|X2)
  - f4(X2 = x2) = 1
  - f4(X2 = -x2) = 1
- Eliminate X1:
  - f1(X2) = SUM_{X1} P(X1) P(X2|X1)
  - f1(X2 = x2) = 0.4(0.8) + 0.6(0.5) = 0.62
  - f1(X2 = -x2) = 0.4(0.2) + 0.6(0.5) = 0.38
- Eliminate X2:
  - f2(X3) = SUM_{X2) P(X3|X2) f1(X2)

    - $f2(X3 = x3) = 0.2(0.62) + 0.3(0.38) = 0.238$
    - $f2(X3 = -x3) = 0.8(0.62) + 0.7(0.38) = 0.762$
- $P(X3 = -x3) = f2(X3 = -x3) = 0.762$
- $P(X2 = x2, X3 = -x3) = P(X3 = -x3 \mid X2 = x2) f1(X2 = x2) f4(X2 = x2) = 0.496$
- $P(X2 = -x2, X3 = -x3) = 0.266$
- $0.496/0.762 = 0.496/(0.496 + 0.266) = 0.6509$

2. The Spacefleet collegiate has decided to create a class of cyborg students. 90% of these cyborgs study hard for their exams. Out of the cyborgs who study hard for an exam, 80% get an A. Out of the cyborgs who do not study, only half get an A. Cyborgs who study hard have a 75% probability of depleting their battery in less that a day. Cyborgs who do not study hard have a longer battery life: only 10% of them deplete their batteries within the next day.

a. Which bayes net correctly captures this problem? [1 point]



b. Which factors have S as one of their variables in the original Bayes Net? **3** [1 point]

c. What is the probability of a cyborg getting an A if it does not study hard? **0.5** [1 point]

d. Assume you want to find the probability of the battery dying (or not dying) given that the cyborg gets an A. Before you eliminate S, you should have a single factor remaining, f(S, B). What are the four values of this factor (i.e. f(s, b), f(s, -b), f(-s, b), f(-s, -b))? [4 points]

f(s, b) = 0.9 * 0.8 * 0.75 = 0.54
f(s, -b) = 0.9 * 0.8 * 0.25 = 0.18
f(-s, b) = 0.1 * 0.5 * 0.1 = 0.005
f(-s, -b) = 0.1 * 0.5 * 0.9 = 0.045

e. What is the probability of the battery being depleted given that a cyborg got an A?

P(b | a) = 0.54 + 0.005 / (0.54 + 0.005 + 0.045 + 0.18) = 0.708 [2 points]

f. Your friend does not believe that cyborgs are much good at studying. He says he is willing to pay you $10 if your cyborg gets an A in the class. Recharging the battery costs $5 Suppose that you could program your cyborg to study or not to study at will (this is not very ethical, but it is technically feasible). What is the best course of action for you? [2 points]

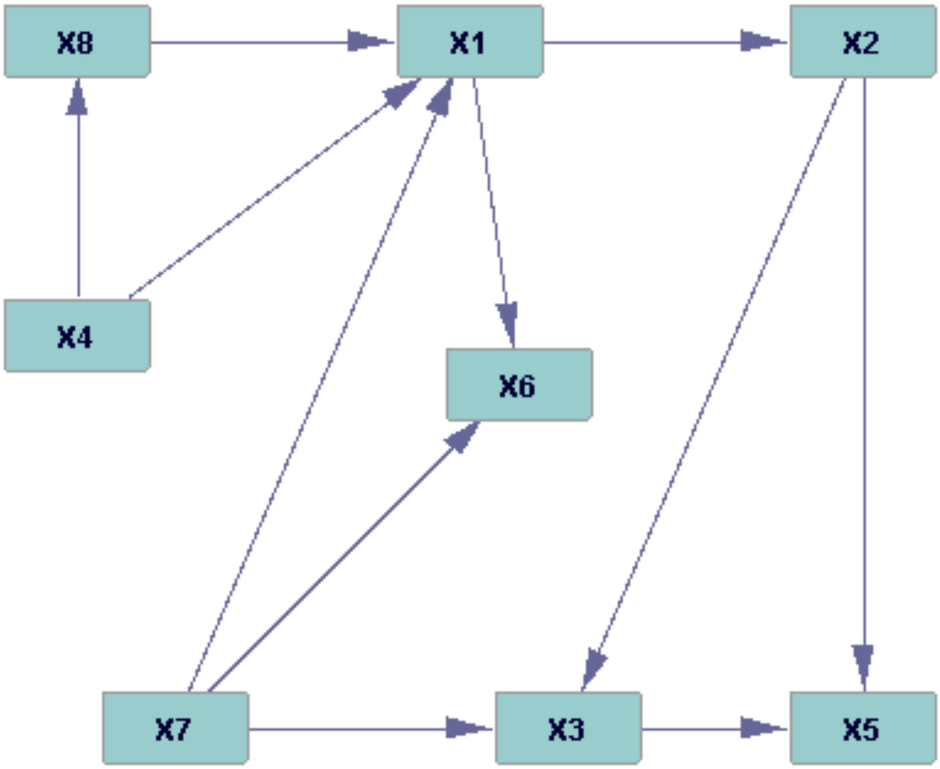
R(s) = 0.8 * (10) - 0.75 * 5 = 4.25

$R(-s) = 0.5 * (10) - 0.1 * 5 = 4.5$

Cyborg should not study

Module 5

1. <u>D-separation</u> [4 points]



Based on the above graph, use D-separation to answer.

a. Which of the following is True?[1 point]

i. X4 and X2 are d-separated (F)

ii. X4 and X6 are d-separated (F)

iii. X8 and X6 are d-separated conditional on X1 (F)

iv. X7 and X2 are d-separated conditional on X1 (T)

b. Which variables are conditionally independent of X2 given X1 and X6? (Check all that apply)

. X3

i. X4 (correct)

ii. X5

iii. X7(correct)

iv. X8(correct)

c. Which variables are conditionally Independent of X2 given X1, X3 and X6? (Check all that apply)

. None

i. X4

ii. X5

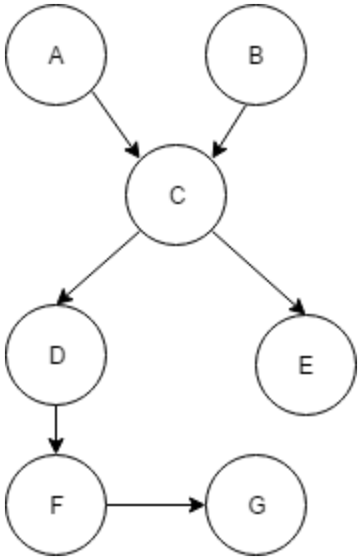
iii. X7

iv. X8 (correct)

d. Which variables do we have to condition on to d-separate X7 and X2?

Answer: X1

2. D-seperation Part 2 [5 points]



Based on the above graph, use D-separation to answer.

a. Are A and B conditionally independent, given D and F?

i. Yes

ii. No

b. Are A and B conditionally independent, given C?

. *Yes*

i. No

c. Are D and E conditionally independent, given C?

. *Yes*

i. *No*

d. Are D and E conditionally independent, given A and B?

. Yes

i. No

e. Does P(D|CEG) equal to P(D|C)?

. Yes

i. No