Zain Chalisa, Akash Shah, Rishi Ajjarapu
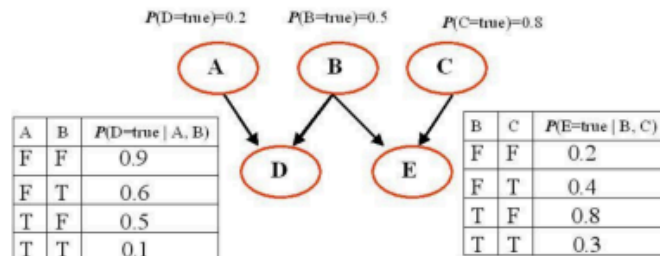Intro to AI - Assignment 3

Problem 1 (*15 points*): Consider the following Bayesian network, where variables A through E are all Boolean valued.
Note: there is a typo in the image, it should be $P(A = true) = 0.2$ instead of $P(D = true) = 0.2$.

P(D=true)=0.2   P(B=true)=0.5   P(C=true)=0.8

A   B   C

A	B	P(D=true | A, B)
F	F	0.9
F	T	0.6
T	F	0.5
T	T	0.1

D   E

B	C	P(E=true | B, C)
F	F	0.2
F	T	0.4
T	F	0.8
T	T	0.3

a) What is the probability that all five of these Boolean variables are simultaneously true?
[Hint: You have to compute the joint probability distribution. The structure of the Bayesian network suggests how the joint probability distribution is decomposed to the conditional probabilities available.]

b) What is the probability that all five of these Boolean variables are simultaneously false?
[Hint: Answer similarly to above.]

c) What is the probability that A is false given that the four other variables are all known to be true?

a) What is the probability that all five of these Boolean variables are simultaneously true?

We can decompose the equation P(A = true, B = true, C = true, D= true, E = true) into
P(A = true) * P(B = true) * P(C = true) * P(D = true | A, B) * P(E = true | B, C)

P(A = true) = 0.2
P(B = true) = 0.5
P(C = true) = 0.8
P(D = true | A, B) = 0.1
P(E = true | B, C) = 0.3

If we substitute these values back into the equation above, we get 0.2 * 0.5 * 0.8 * 0.1 * 0.3 = 0.0024, representing the probability of all the 5 variables being true simultaneously.

b) What is the probability that all five of these Boolean variables are simultaneously false?

We can decompose the equation P(A = False, B = False, C = False, D= False, E = False) into
P(A = False) * P(B = False) * P(C = False) * P(D = False | A, B) * P(E = False | B, C)

P(A = false) = 0.8
P(B = false) = 0.5
P(C = false) = 0.2

P(D = false | A =false, B=false) = 0.1
P(E = false | B=false, C=false) = 0.8

If we substitute these values back into the equation above, we get 0.8 * 0.5 * 0.2 * 0.1 * 0.8 = 0.0064, which represents the probability of all the 5 variables being false simultaneously.

c) What is the probability that A is false given that the four other variables are all known to be true?

P(A = false | B, C, D, E) = P(A = false, B, C, D, E = True) / P(B, C, D, E)

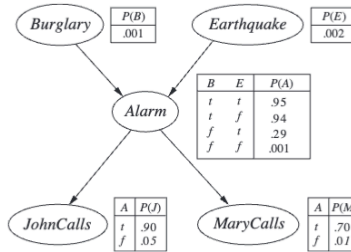P(A = false, B, C, D, E = True) = P(D = True | A = False, B= True) * P(E = True | B = True, C= True) * P (A = False) * P(B=True) * P(C = True)
= 0.0576

P(B, C, D, E) = P(A = false, B, C, D, E = True) + P(A = True, B, C, D, E = True)
P(A = True, B, C, D, E = True) = 0.0024

So P(A = false | B, C, D, E) = P(A = false, B, C, D, E = True) / P(B, C, D, E)
 = (0.0576) / (0.0576 + 0.0024)
= 0.96

Problem 2 (*15 points*):

a) Calculate $P(Burglary|JohnsCalls = true, MaryCalls = true)$ and show in detail the calculations that take place. Use your book to confirm that your answer is correct.

b) Suppose a Bayesian network has the from of a *chain*: a sequence of Boolean variables $X_1, \ldots X_n$ where $Parents(X_i) = \{X_{i-1}\}$ for $i = 2, \ldots, n$. What is the complexity of computing $P(X_1|X_n = true)$ using enumeration? What is the complexity with variable elimination?

a)

a) $P(B \mid John = True, Mary = True)$

$= \dfrac{P(B, j, m)}{P(j, m)}$ ← all variables / ← evidence

$= a\, P(B, j, m)$

$= a\, \hat{\Sigma}_e \hat{\Sigma}_a\, P(B, e, a, j, m)$

$= a\, \Sigma_e \Sigma_a\, P(B)\, P(e)\, P(a|B, e)\, P(j|a)\, P(m|a)$

$= a\, P(B)\, \hat{\Sigma}_e\, P(e)\, \hat{\Sigma}_a\, P(a|B, e)\, P(j|a)\, P(m|a)$

$= a\, P(B)\, \hat{\Sigma}_e\, P(e) \left[.9 \times .7 \times \begin{bmatrix} .95 & .29 \\ .94 & .001 \end{bmatrix} + .5 \times .01 \times \begin{pmatrix} .05 & .71 \\ .06 & .999 \end{pmatrix} \right]$

$= a\, P(B)\, \hat{\Sigma}_e\, P(e) \begin{bmatrix} .599 & .183 \\ .592 & .001 \end{bmatrix}$

$= a\, \begin{pmatrix} .001 \\ .999 \end{pmatrix} \times \begin{bmatrix} .592 \\ .001 \end{bmatrix}$

$= \langle 0.284, 0.716 \rangle$

The probability that the burglary will happen is going to be 0.284.

b)

b)

$$P(x_1 \mid x_n = true) = a \sum_{x_2 \ldots x_{n-1}} P(x_n = true \mid x_{n-1}) P(x_{n-1} \mid x_{n-2}) \ldots P(x_3 \mid x_2) P(x_2 \mid x_1) P(x_1)$$

Being that are $n-2$ variables to sum over we can see that each term reads $n-1$ multiplication. For enumeration it will be $O(2^n)$. runtime.

$$P(x_1 \mid x_n = true) = a \sum_{x_{n-1}} P(x_n = true \mid x_{n-1}) \sum_{x_{n-2}} P(x_{n-1} \mid x_{n-2}) \ldots \sum_{x_2} P(x_3 \mid x_2)$$

$$= a \sum_{x_{n-1}} P(x_n = true \mid x_{n-1}) \sum_{x_{n-2}} P(x_{n-1} \mid x_{n-2}) \ldots \sum_{x_3} P(x_4 \mid x_3) f_{x_2}(x_3)$$

Looking at the formula we can deduce they're all summations of each resulting in an $O(n)$ runtime.

Problem 3 (*20 points*): Suppose you are working for a financial institution and you are asked to implement a fraud detection system. You plan to use the following information:

- When the card holder is travelling abroad, fraudulent transactions are more likely since tourists are prime targets for thieves. More precisely, 1% of transactions are fraudulent when the card holder is travelling, where as only 0.4% of the transactions are fraudulent when she is not travelling. On average, 5% of all transactions happen while the card holder is travelling. If a transaction is fraudulent, then the likelihood of a foreign purchase increases, unless the card holder happens to be travelling. More precisely, when the card holder is not travelling, 10% of the fraudulent transactions are foreign purchases where as only 1% of the legitimate transactions are foreign purchases. On the other hand, when the card holder is travelling, then 90% of the transactions are foreign purchases regardless of the legitimacy of the transactions.

- Purchases made over the internet are more likely to be fraudulent. This is especially true for card holders who don't own any computer. Currently, 75% of the population owns a computer or smart phone and for those card holders, 1% of their legitimate transactions are done over the internet, however this percentage increases to 2% for fraudulent

  transactions. For those who don't own any computer or smart phone, a mere 0.1% of their legitimate transactions is done over the internet, but that number increases to 1.1% for fraudulent transactions. Unfortunately, the credit card company doesn't know whether a card holder owns a computer or smart phone, however it can usually guess by verifying whether any of the recent transactions involve the purchase of computer related accessories. In any given week, 10% of those who own a computer or smart phone purchase (with their credit card) at least one computer related item as opposed to just 0.1% of those who don't own any computer or smart phone.

a) Construct a Bayes Network to identify fraudulent transactions.
   What to hand in: Show the graph defining the network and the Conditional Probability Tables associated with each node in the graph. This network should encode the information stated above. Your network should contain exactly six nodes, corresponding to the following binary random variables:
   OC : card holder owns a computer or smart phone.
   Fraud : current transaction is fraudulent.
   Trav : card holder is currently travelling.
   FP : current transaction is a foreign purchase.
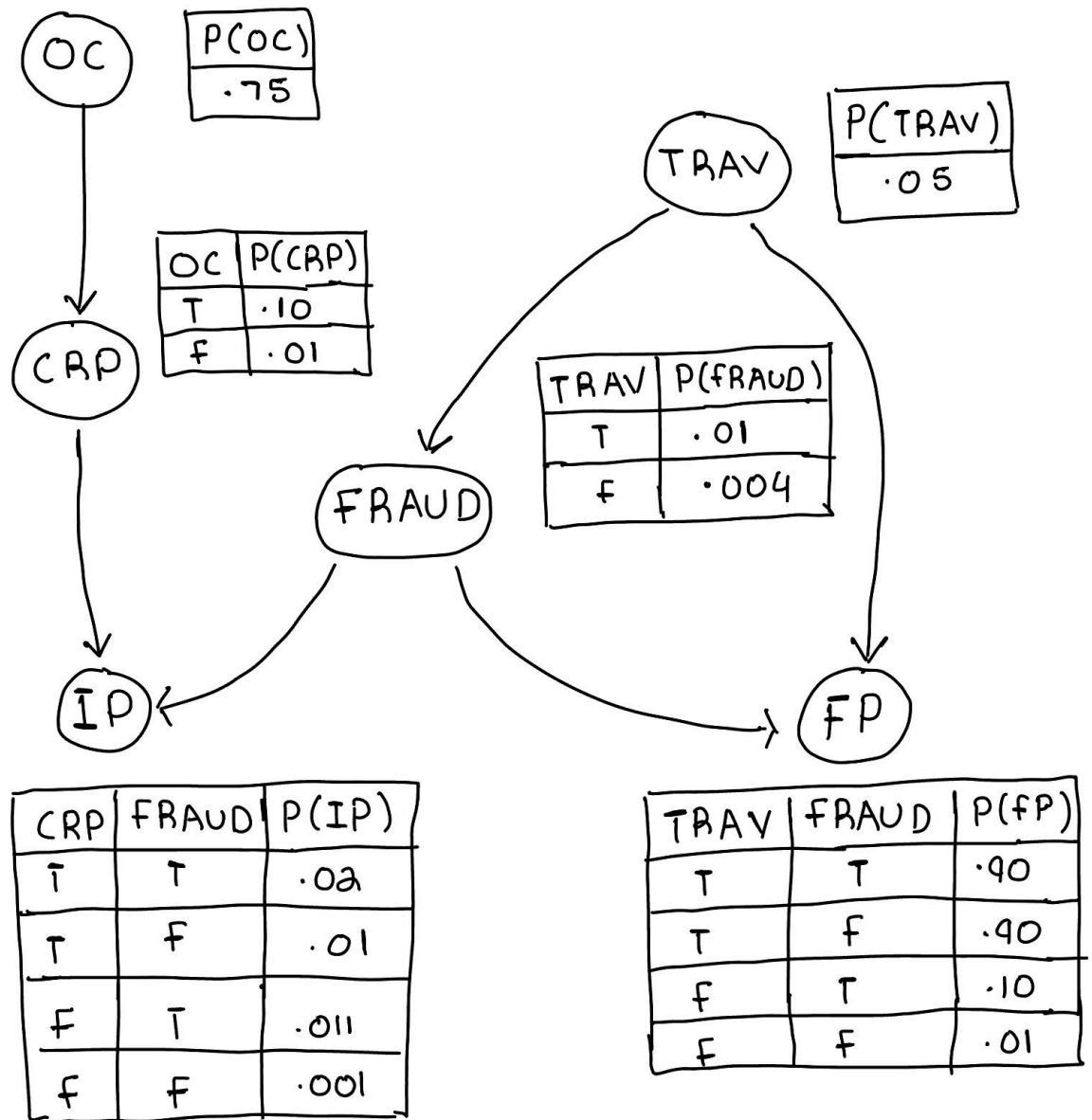   IP : current purchase is an internet purchase.
   CRP : a computer related purchase was made in the past week.
   The arcs defining your Network should accurately capture the probabilistic dependencies between these variables.

b) What is the prior probability (i.e., before we search for previous computer related purchases and before we verify whether it is a foreign and/or an internet purchase) that the current transaction is a fraud? What is the probability that the current transaction is a fraud once we have verified that it is a foreign transaction, but not an internet purchase and that the card holder purchased computer related accessories in the past week?
   What to hand in: Indicate the two queries (i.e., $Pr(variables|evidence)$) you used to compute those two probabilities. Show each step of the calculation

a)

OC

P(OC)
.75

TRAV

P(TRAV)
.05

CRP

OC	P(CRP)
T	.10
f	.01

FRAUD

TRAV	P(FRAUD)
T	.01
f	.004

IP

FP

CRP	FRAUD	P(IP)
T	T	.02
T	f	.01
f	T	.011
f	f	.001

TRAV	FRAUD	P(FP)
T	T	.90
T	f	.90
f	T	.10
f	f	.01

b)

$P(Fraud) = P(Fraud \mid Trav) * P(Trav) + P(Fraud \mid \neg Trav) * P(\neg Trav)$

$= 0.01 * 0.05 + 0.004 * 0.95$

$= 0.004275$

The probability that the current transaction is Fraud prior to other verification checks is 0.004275

$P(Fraud \mid FP) = P(Fraud \mid Trav) * P(FP \mid Trav, Fraud) * P(Trav) + P(Fraud \mid \neg Trav) * P(FP \mid \neg Trav, Fraud) * P(\neg Trav)$

$= 0.01 * 0.90 * 0.05 + 0.004 * 0.10 * 0.95$

$= 0.00045 + 0.00038$

$= 0.00083$

$P(Fraud \mid \neg IP, CRP) = P(\neg IP \mid CRP, Fraud) * P(CRP) + P(\neg IP \mid \neg CRP, Fraud) * P(\neg CRP)$

$= P(\neg IP \mid CRP, Fraud) * [P(CRP \mid OC) * P(OC) + P(CRP \mid \neg OC) * P(\neg OC)] +$

$P(\neg IP \mid \neg CRP, Fraud) * [P(\neg CRP \mid OC) * P(OC) + P(\neg CRP \mid \neg OC) * P(\neg OC)]$

$= 0.98 * (0.10 * 0.75 + 0.01 * 0.25) + 0.989 * (0.01 * 0.75 + .99 * .25)$

$= 0.98 * 0.0775 + 0.989 * 0.255$

$= 0.328145$

The probability that a given transaction is fraudulent and a foreign purchase while also not being an internet purchase from a cardholder who purchased computer related accessories in the last week can be expressed as

$= 0.328145 * 0.00083$

$= 0.00027236035$