

# Assignment 3 - Part 1

Xinyu Meng netid:xm73 RUID:193002567

Qichao Lin netid:ql180 RUID:196005340

April 2023

## 1 Problem 1

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  $P(A = \text{true}) = 0.2$  instead of  $P(D = \text{true}) = 0.2$ .

### 1.1 a)

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

$$P(A, B, C, D, E) = P(A)P(B)P(C)P(D | A, B)P(E | B, C) = 0.2 * 0.5 * 0.8 * 0.1 * 0.3 = 0.0024$$

### 1.2 b)

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

$$P(\neg A, \neg B, \neg C, \neg D, \neg E) = P(\neg A)P(\neg B)P(\neg C)P(\neg D | \neg A, \neg B)P(\neg E | \neg B, \neg C) = 0.8 * 0.5 * 0.2 * 0.1 * 0.8 = 0.0064$$

### 1.3 c)

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

$$P(\neg A | B, C, D, E) = P(B, C, D, E | \neg A)P(\neg A) \div P(B, C, D, E)$$

based on product rule, we know that  $P(B, C, D, E | \neg A)P(\neg A) = P(B, C, D, E, \neg A) = P(\neg A)P(B)P(C)P(D | \neg A, B)P(E | B, C) = 0.8 * 0.5 * 0.8 * 0.6 * 0.3 = 0.0576$

and  $P(B, C, D, E) = P(B, C, D, E, A) + P(B, C, D, E, \neg A) = 0.0024 + P(\neg A)P(B)P(C)P(D | \neg A, B)P(E | B, C) = 0.0024 + 0.8 * 0.5 * 0.8 * 0.6 * 0.3 = 0.06$

Therefore,  $P(\neg A | B, C, D, E) = P(B, C, D, E | \neg A)P(\neg A) \div P(B, C, D, E) =$

$$P(B, C, D, E, \neg A) \div (P(B, C, D, E, A) + P(B, C, D, E, \neg A)) = 0.0576 \div 0.06 = 0.96$$

## 2 Problem 2

### 2.1 a)

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

$$\begin{aligned} P(\text{Burglary} \mid \text{JohnsCalls} = \text{true}, \text{MaryCalls} = \text{true}) &= \alpha P(\text{Burglary}, \text{JohnsCalls} = \text{true}, \text{MaryCalls} = \text{true}) \\ &= \alpha \sum_{\text{Earthquake}} \sum_{\text{Alarm}} P(\text{Burglary}, \text{JohnsCalls}, \text{MaryCalls}, \text{Earthquake}, \text{Alarm}) \\ &= \alpha \sum_{\text{Earthquake}} \sum_{\text{Alarm}} P(\text{Burglary}) P(\text{JohnsCalls} \mid \text{Alarm}) P(\text{MaryCalls} \mid \text{Alarm}) P(\text{Earthquake}) P(\text{Alarm} \mid \text{Burglary}, \text{Earthquake}) \\ &= \alpha P(\text{Burglary}) \sum_{\text{Earthquake}} P(\text{Earthquake}) \sum_{\text{Alarm}} P(\text{JohnsCalls} \mid \text{Alarm}) P(\text{MaryCalls} \mid \text{Alarm}) P(\text{Alarm} \mid \text{Burglary}, \text{Earthquake}) \\ &= \alpha < P(\text{Burglary}) * ((0.02 * (0.9 * 0.7 * 0.95 + 0.05 * 0.01 * 0.05)) + (0.98 * (0.9 * 0.7 * 0.94 + 0.05 * 0.01 * 0.06))), P(\neg \text{Burglary}) * ((0.02 * (0.9 * 0.7 * 0.95 + 0.05 * 0.01 * 0.05)) + (0.98 * (0.9 * 0.7 * 0.94 + 0.05 * 0.01 * 0.06))) >= \alpha < 0.001 * ((0.002 * (0.9 * 0.7 * 0.95 + 0.05 * 0.01 * 0.05)) + (0.998 * (0.9 * 0.7 * 0.94 + 0.05 * 0.01 * 0.06))), 0.999 * ((0.002 * (0.9 * 0.7 * 0.29 + 0.05 * 0.01 * 0.71)) + (0.998 * (0.9 * 0.7 * 0.001 + 0.05 * 0.01 * 0.999))) > \\ &= \alpha < 0.00059224, 0.0014919 > \approx < 0.28416517124, 0.71583482875 > \end{aligned}$$

### 2.2 b)

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

For enumeration inference,  $P(X_1 \mid X_n = \text{true}) = \alpha \sum_{x_2} \sum_{x_3} \dots \sum_{x_{n-1}} P(X_1, X_2, \dots, X_n)$ , which will be  $P(X_1 \mid X_n = \text{true}) = \alpha \sum_{x_2} \sum_{x_3} \dots \sum_{x_{n-1}} P(X_1) P(X_2 \mid X_1) \dots P(X_n \mid X_{n-1})$  there are  $X_1 \dots X_n$  variables, Each of the sums over the values of the variables has 2 terms, so the total number of terms to be multiplied is  $2^{n-2}$ . Therefore, the time complexity of computing  $P(X_1 \mid X_n = \text{true})$  using enumeration is  $O(2^{n-2})$ .

For variable elimination and a singly connected bayesian network with a chain, the complexity for variable elimination is linear therefore,  $O(n)$  as the chain structure allows for the computation happen once for each variable.

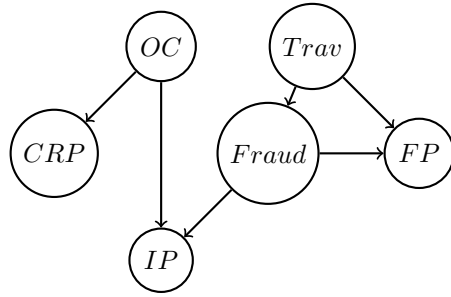
### 3 Problem 3

#### 3.1 a)

Problem3(20points): 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.



OC	0.75
$\neg OC$	0.25

Transaction during Trav
0.05

OC	P(CRP)
True	0.1
False	0.001

Trav	P(Fraud)
True	0.01
False	0.004

OC	Fraud	P(IP)
True	True	0.02
True	False	0.01
False	True	0.011
False	False	0.001

Fraud	Trav	P(FP)
True	True	0.9
False	True	0.9
True	False	0.1
False	False	0.01

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.,  $P(\text{variables} \mid \text{evidence})$ ) you used to compute those two probabilities. Show each step of the calculation

For the probability of a transaction being fraud,  $P(\text{Fraud}) = P(\text{Fraud} \mid \text{Trav}) * P(\text{Trav}) + P(\text{Fraud} \mid \neg \text{Trav}) * P(\neg \text{Trav}) = 0.01 * 0.05 + 0.004 * 0.95 = 0.0043$

$$\begin{aligned}
P(Fraud \mid FP, \neg IP, CRP) &= \alpha P(Fraud, FP, \neg IP, CRP) \\
&= \alpha \sum_{Trav} \sum_{OC} P(Fraud, FP, \neg IP, CRP, Trav, OC) \\
&= \alpha \sum_{Trav} \sum_{OC} P(Fraud \mid Trav) P(FP \mid Trav, Fraud) P(\neg IP \mid OC, Fraud) P(CRP \mid \\
&\quad OC) P(Trav) P(OC) \\
&= \alpha \sum_{Trav} P(Trav) \sum_{OC} P(OC) P(Fraud \mid Trav) P(FP \mid Trav, Fraud) P(\neg IP \mid \\
&\quad OC, Fraud) P(CRP \mid OC) \\
&= \alpha < 0.00006121, 0.0040239 > \\
&\approx < 0.0149837, 0.9850163 > \\
\text{Therefore, } P(Fraud \mid FP, \neg IP, CRP) &= 0.0149837
\end{aligned}$$