Assignment 3

Probabilistic Reasoning

Abid Azad (netID: aa2177, RUID: 202005452)

CS440, April 14th, 2024

Contents

1 Problem 1

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

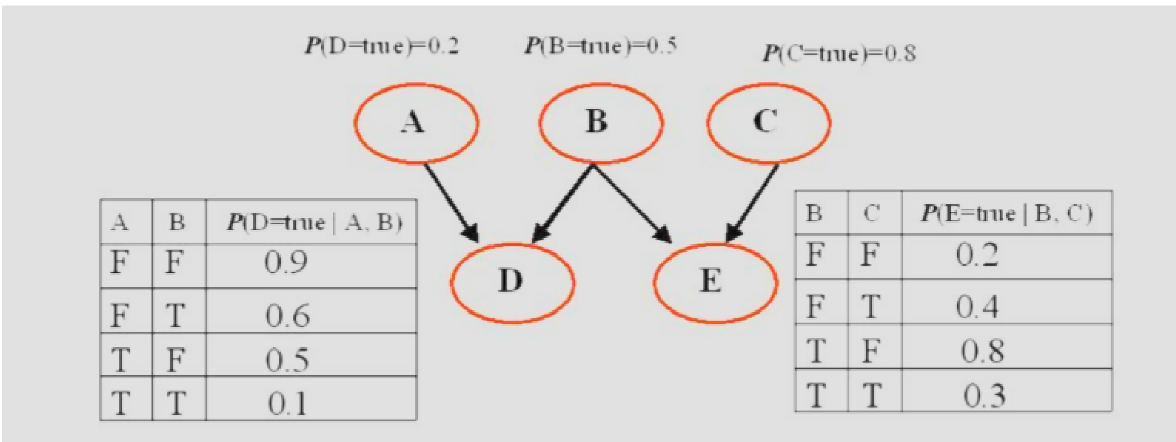

Figure 1: Problem 1's Bayesian Network

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.]

Given the expression:

$$P(A, B, C, D, E) = P(A) \cdot P(B) \cdot P(C) \cdot P(D|A, B) \cdot P(E|B, C).$$

The probability of $A = \mathrm{T}, B = \mathrm{T}, C = \mathrm{T}, D = \mathrm{T}, E = \mathrm{T}$ is calculated as:

$$P(A = \mathrm{T}, B = \mathrm{T}, C = \mathrm{T}, D = \mathrm{T}, E = \mathrm{T}) = P(A = \mathrm{T}) \cdot P(B = \mathrm{T}) \cdot P(C = \mathrm{T}) \cdot P(D = \mathrm{T}) \cdot P(E = \mathrm{T})$$

$$= 0.2 \cdot 0.5 \cdot 0.8 \cdot 0.1 \cdot 0.3$$

$$= 0.0024.$$

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

Given the expression:

$$P(A = F, B = F, C = F, D = F, E = F) = P(A = F) \cdot P(B = F) \cdot P(C = F) \cdot P(D = F) \cdot P(E = F)$$

The probability of $A = \mathrm{F}, B = \mathrm{F}, C = \mathrm{F}, D = \mathrm{F}, E = \mathrm{F}$ is calculated as:

$$= 0.8 \cdot 0.5 \cdot 0.2 \cdot 0.1 \cdot 0.8$$

$$= 0.0064.$$

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

We can find $P(\neg A|B, C, D, E)$ using the formula:

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

Where α is calculated as:

$$\alpha = \frac{1}{P(A, B, C, D, E) + P(\neg A, B, C, D, E)}$$

$$= \frac{1}{(0.2 \times 0.5 \times 0.8 \times 0.1 \times 0.3) + (0.8 \times 0.5 \times 0.8 \times 0.6 \times 0.3)}$$

$$= \frac{1}{0.0024 + 0.0576}$$

$$= \frac{50}{3}$$

Therefore:

$$P(\neg A | B, C, D, E) = \frac{50}{3} \times 0.0576$$

$$= 0.96$$

2 Problem 2

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

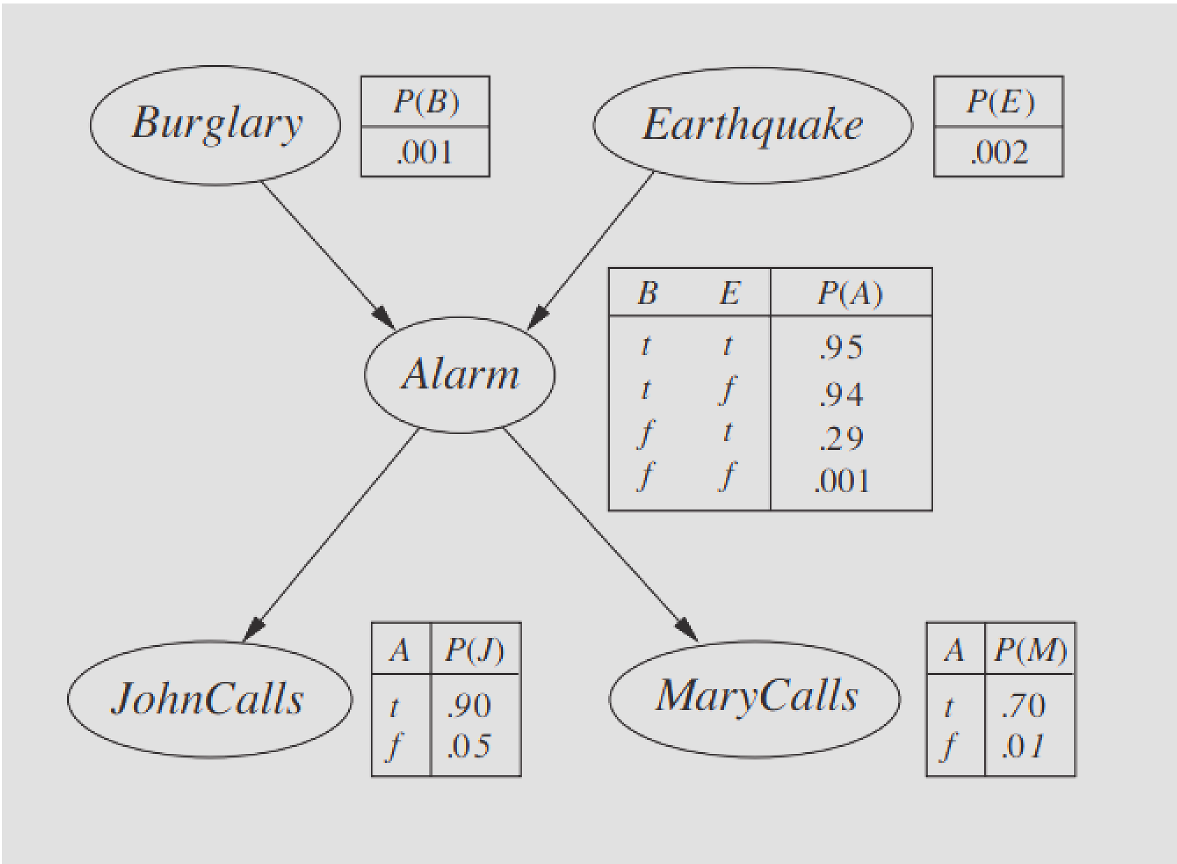

Figure 2: Problem 2's Bayesian Network

(By definition of conditional probability)

P(B|J,M) $\equiv \frac{P(B,J,M)}{P(J,M)}$

(Using the definition of joint probability and the chain rule)

$\equiv \alpha * P(B) \sum_E P(E) \sum_A P(A|B,E) * P(J|A) * P(M|A)$

(Expanding the expression using Bayes' theorem)

$$\equiv \alpha {*} P(B) \sum_A P(E) {*} \left((0.9 \times 0.7 \times \begin{bmatrix} 0.95 & 0.29 \\ 0.94 & 0.001 \end{bmatrix} + 0.5 \times 0.01 \times \begin{bmatrix} 1-0.95 & 1-0.29 \\ 1-0.94 & 1-0.001 \end{bmatrix}) \right)$$

(Calculating the conditional probabilities given evidence)

$$\equiv \alpha * P(B) \sum_A P(E) * \left((0.9 \times 0.7 \times \begin{bmatrix} 0.95 & 0.29 \\ 0.94 & 0.001 \end{bmatrix} + 0.5 \times 0.01 \times \begin{bmatrix} 0.05 & 0.71 \\ 0.06 & 0.999 \end{bmatrix}) \right)$$

(Combining probabilities)

$$\equiv \alpha * P(B) \sum_A P(E) * \begin{bmatrix} 0.598525 & 0.183055 \\ 0.59223 & 0.011295 \end{bmatrix}$$

(Calculating the weighted sum)

$$\equiv \alpha * P(B) * \left(0.002 \times \begin{bmatrix} 0.598525 \\ 0.183055 \end{bmatrix} + 0.998 \times \begin{bmatrix} 0.59223 \\ 0.0011295 \end{bmatrix} \right)$$

(Calculating the product)

$$\equiv \alpha * \begin{bmatrix} 0.001 \\ 0.999 \end{bmatrix} * \begin{bmatrix} 0.59224259 \\ 0.001493351 \end{bmatrix}$$

(Element-wise multiplication)

$$\equiv \alpha * \begin{bmatrix} 0.00059224259 \\ 0.0014918576 \end{bmatrix}$$

(We can calculate α using the same method. $\alpha = \frac{1}{0.0020853609}$)

$$\equiv \frac{1}{0.0020853609} * \begin{bmatrix} 0.00059224259 \\ 0.0014918576 \end{bmatrix}$$

(Final conditional probability)

$$\equiv \begin{bmatrix} 0.284 \\ 0.716 \end{bmatrix}$$

Here, 0.284 is the probability that a burglary will happen if John and Mary both call, whereas 0.716 is the probability that burglary will not happen if they both call.

b) Suppose a Bayesian network has the form 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 $P(X_1|X_n = \text{true})$ using enumeration? What is the complexity with variable elimination?

Consider the computation of $P(A|E_1 = \text{true})$ in a Bayesian network resembling a chain, where the variables $E_1, E_2, \ldots, E_n$ are Boolean and $Parents(E_i) = \{E_{i-1}\}$ for $i = 2, \ldots, n$.

When employing enumeration to calculate $P(A|E_1 = \text{true})$, each possible value of A necessitates evaluating two binary trees, each with depth $n-2$. Consequently, the overall complexity of enumeration is $O(2^n)$.

Now, transitioning to variable elimination, the factors never expand beyond two variables. For instance,

$$\begin{aligned} P(A|E_1 = \text{true}) &= \alpha * P(A) \ldots \sum_{e_{n-2}} P(e_{n-2}|e_{n-3}) \sum_{e_{n-1}} P(e_{n-1}|e_{n-2}) P(E_n = \text{true}|e_{n-1}) \\ &= \alpha * P(A) \ldots \sum_{e_{n-2}} P(e_{n-2}|e_{n-3}) \sum_{e_{n-1}} f_{E_{n-1}}(e_{n-1}, e_{n-2}) f_{E_n}(e_{n-1}) \\ &= \alpha * P(A) \ldots \sum_{e_{n-2}} P(e_{n-2}|e_{n-3}) f\left(\frac{e_{n-2}}{E_{n-1} * E_n}\right) \end{aligned}$$

This representation reveals that the problem becomes isomorphic to one involving $n-1$ variables rather than n. Consequently, the complexity becomes a constant independent of n, resulting in a total complexity of $O(n)$.

3 Problem 3

Problem 3: 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 cardholder is traveling abroad, fraudulent transactions are more likely since tourists are prime targets for thieves. More precisely, 1% of transactions are fraudulent when the cardholder is traveling, whereas only 0.4% of the transactions are fraudulent when she is not traveling. On average, 5% of all transactions happen while the cardholder is traveling. If a transaction is fraudulent, then the likelihood of a foreign purchase increases, unless the cardholder happens to be traveling. More precisely, when the cardholder is not traveling, 10% of the fraudulent transactions are foreign purchases, whereas only 1% of the legitimate transactions are foreign purchases. On the other hand, when the cardholder is traveling, 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 cardholders who don't own any computer. Currently, 75% of the population owns a computer or smartphone and for those cardholders, 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 smartphone, a mere 0.1% of their legitimate transactions are done over the internet, but that number increases to 1.1% for fraudulent transactions. Unfortunately, the credit card company doesn't know whether a cardholder owns a computer or smartphone; 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 smartphone 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 smartphone.

a) Construct a Bayes Network to identify fraudulent transactions.

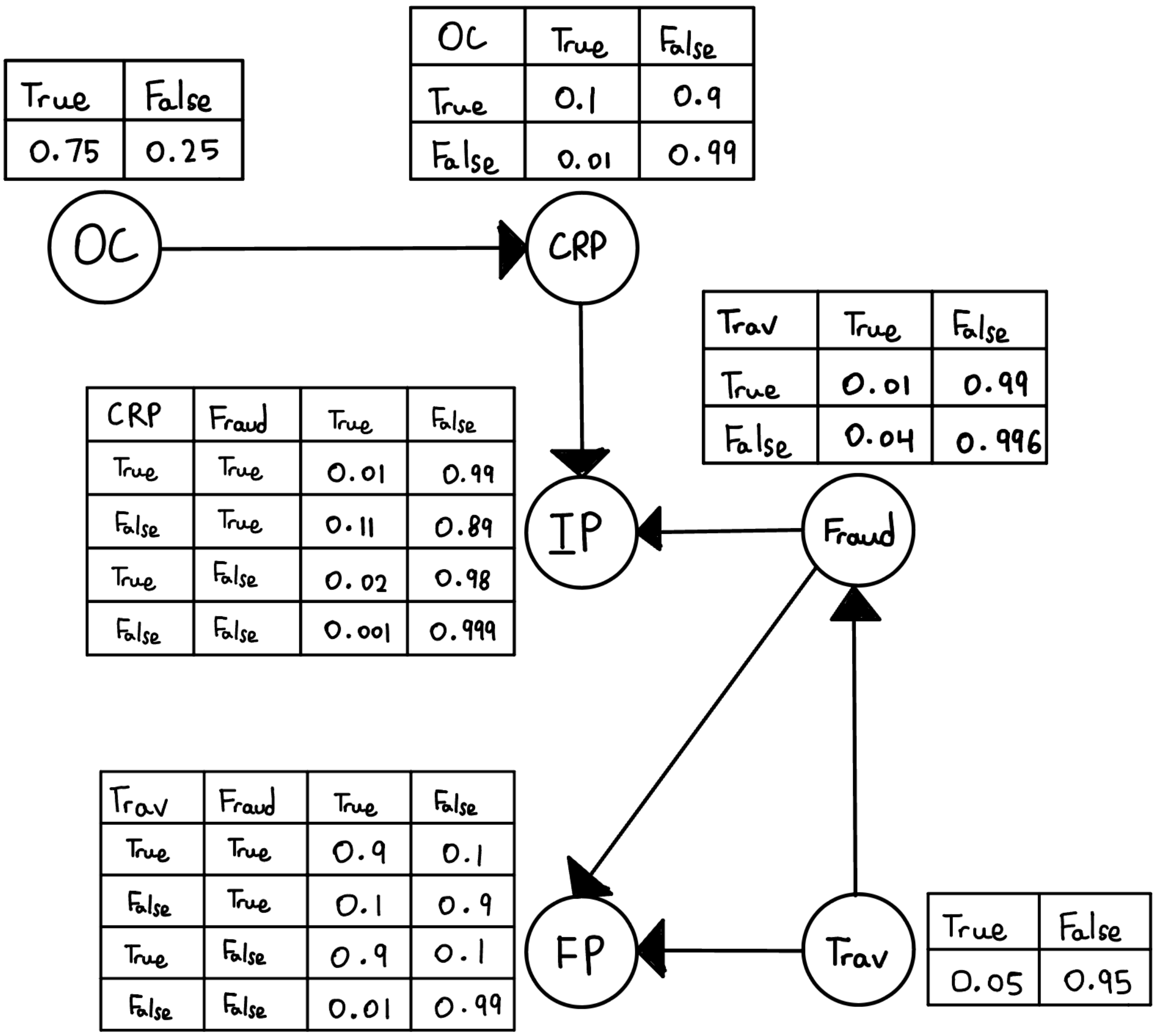

Figure 3: Problem 3's Bayesian Network

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?

First, the probability of fraud ($P(Fraud)$) is calculated using the given probabilities as follows:

$$\begin{aligned} P(Fraud) &= P(Fraud = T|Trav = T) \times P(Trav = T) + P(Fraud = T|Trav = F) \times P(Trav = F) \\ &= 0.01 \times 0.05 + 0.004 \times 0.95 \\ &= 0.004275 \end{aligned}$$

Now, suppose we want to find the probability that a transaction is fraudulent given that it is a foreign transaction, but not an internet purchase, and the cardholder has purchased computer-related accessories in the past week ($P(Fraud|FP)$):

$$\begin{aligned} P(Fraud|FP) &= P(Fraud|Trav) \times P(FP|Trav, Fraud) \times P(Trav) \\ &\quad + P(Fraud|\neg Trav) \times P(FP|\neg Trav, Fraud) \times P(\neg Trav) \\ &= 0.01 \times 0.90 \times 0.05 + 0.004 \times 0.10 \times 0.95 \\ &= 0.00045 + 0.00038 \\ &= 0.00083 \end{aligned}$$

Next, let's find the probability of fraud given that it's not an internet purchase, the cardholder purchased computer-related accessories, and other conditions ($P(Fraud|\neg IP, CRP, OC)$):

$$
\begin{aligned}
&P(Fraud|\neg IP, CRP, OC)\\
&= P(\neg IP|CRP, Fraud) \times P(CRP)\\
&\quad + P(\neg IP|\neg CRP, Fraud) \times P(\neg CRP)\\
&= \Big(P(\neg IP|CRP, Fraud)\times\\
&\quad (P(CRP|OC) \times P(OC) + P(CRP|\neg OC) \times P(\neg OC))\\
&\quad + P(\neg IP|\neg CRP, Fraud)\times\\
&\quad (P(\neg CRP|OC) \times P(OC) + P(\neg CRP|\neg OC) \times P(\neg OC))\Big)\\
&= 0.99 \times (0.1 \times 0.75 + 0.01 \times 0.25)\\
&\quad + 0.89 \times (0.9 \times 0.75 + 0.99 \times 0.25)\\
&= 0.99 \times 0.02125 + 0.89 \times 0.41625\\
&= 0.379
\end{aligned}
$$

Finally, the overall probability (P) is calculated as the product of the probabilities obtained from the previous calculations:

$$P = 0.379 \times 0.00083 = 0.00031457$$

Thus, the prior probability that the current transaction is a fraud, before considering previous computer-related purchases and verifying whether it's a foreign and/or an internet purchase, is 0.004275. Furthermore, once it is verified that the transaction is a foreign transaction, but not an internet purchase, and the cardholder has purchased computer-related accessories in the past week, the probability that the transaction is fraudulent becomes 0.00031457.