CSCI 3202, Fall 2025

Homework 5

Total: 100 pts

Due March 5, 2025 by 11:59 pm

Your name: Gil Eskayo

In [1]:
```
from probabilityAe import *
from utils4e import print_table

```

This homework uses the aima-python github repository. You will need a python function from the aima-python folder. Since we will be using the function in this repository quite a bit, it is probably easiest to clone it:

git clone https://github.com/aimacode/aima-python.git

Once you have the repository, you will need to copy the .py file(s) you need into the directory containing your homework, or copy the the homework notebook (not the folder containing the .ipynb file) into the aima-data directory.

For this assignment, you will need the file:

probabilityAe.py
utils4e.py

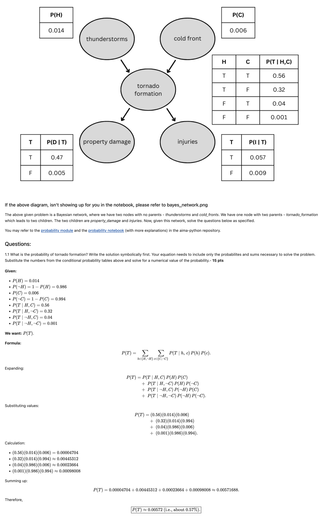
Note: You should copy the file(s) from the repository, not move them. If you move the files, they won't be in the repository for your next assignment.

After you do this, if the import statement below fails, you will need to restart either the Jupyter server or your Python kernel. We are restarting to clear the cache for the import statement.

Use enumeration_ask() from the AIMA repository to solve the problems below.

Problem 1: Solve on paper

For question 1, we will solve the Bayes Network initially on paper and then using probabilityAe module from aima-python repository.



If the above diagram, isn't showing up for you in the notebook, please refer to bayes_network.png

The above given problem is a Bayesian network, where we have two nodes with no parents - thunderstorms and cold_fronts. We have one node with two parents - tornado_formation which leads to two children. The two children are property_damage and injuries. Now, given this network, solve the questions below as specified.

You may refer to the probability module and the probability notebook (with more explanations) in the aima-python repository.

Questions:

1.1 What is the probability of tornado formation? Write the solution symbolically first. Your equation needs to include only the probabilities and sums necessary to solve the problem. Substitute the numbers from the conditional probability tables above and solve for a numerical value of the probability. - **15 pts**.

Given:

- $P(H) = 0.014$
- $P(\neg H) = 1 - P(H) = 0.986$
- $P(C) = 0.006$
- $P(\neg C) = 1 - P(C) = 0.994$
- $P(T \mid H, C) = 0.56$
- $P(T \mid H, \neg C) = 0.32$
- $P(T \mid \neg H, C) = 0.04$
- $P(T \mid \neg H, \neg C) = 0.001$

We want: $P(T)$,

Formula:

$$P(T) = \sum_{h \in \{H, \neg H\}} \sum_{c \in \{C, \neg C\}} P(T \mid h, c) P(h) P(c).$$

Expanding:

$$P(T) = P(T \mid H, C) P(H) P(C) + P(T \mid H, \neg C) P(H) P(\neg C) + P(T \mid \neg H, C) P(\neg H) P(C) + P(T \mid \neg H, \neg C) P(\neg H) P(\neg C).$$

Substituting values:

$$P(T) = (0.56)(0.014)(0.006) + (0.32)(0.014)(0.994) + (0.04)(0.986)(0.006) + (0.001)(0.986)(0.994).$$

Calculation:

- $(0.56)(0.014)(0.006) = 0.00047544$
- $(0.32)(0.014)(0.994) = 0.00445222$
- $(0.04)(0.986)(0.006) = 0.00023664$
- $(0.001)(0.986)(0.994) = 0.00097958$

Summing up:

$$P(T) = 0.00047544 + 0.00445222 + 0.00023664 + 0.00097958 = 0.00073668.$$

Therefore,

$$\boxed{P(T) = 0.00073\ \text{(i.e., about 0.07\%)}}$$

1.2 What is the probability of property damage? Write the solution symbolically first. Your equation needs to include only the probabilities and sums necessary to solve the problem. Substitute the numbers from the conditional probability tables above and solve for a numerical value of the probability. You may use results from earlier questions if they are helpful - **10 pts**

Given:

- From Question 1.1, the probability of tornado formation is
  P(T) (approx 0.00073)
  Thus, P(Not T) = 1 - P(T) (approx 0.99926.
- From the conditional probability table for property damage:
  - P(D |not T) = 0.05
  - P(D |not Tnot T) = 0.01

We want: P(D), the probability of property damage.

Formula:

$$P(D) = P(D \mid T) P(T) + P(D \mid \neg T) P(\neg T)$$

Substituting the values:

$$P(D) = (0.95)(0.00072) + (0.01)(0.99426)$$
$$= 0.000684 + 0.0099426$$
$$= 0.0103768.$$

Thus,

$$\boxed{P(D) = 0.01038\ \text{(i.e., about 1.04\%)}}$$

1.3 What is the probability of thunderstorms given there is tornado formation? Write the solution symbolically first. Your equation needs to include only the probabilities and sums necessary to solve the problem. Substitute the numbers from the conditional probability tables above and solve for a numerical value of the probability. - **10 pts**

Given:

- $P(H) = 0.014$
- $P(\neg H) = 0.986$and$P(\neg C) = 0.994$
- $P(T \mid H, C) = 0.56$
- $P(T \mid H, \neg C) = 0.32$
- From Question 1.1, $P(T) = 0.00072$

We want:

Find P(H |not T), the probability of thunderstorms given tornado formation.

Step 1: Write the expression symbolically

Using the definition of conditional probability:

$$P(H \mid T) = \frac{P(H, T)}{P(T)}$$

Since T depends on both H and C, we sum over the possible states of C:

$$P(H, T) = P(H) \left[P(T \mid H, C) P(C) + P(T \mid H, \neg C) P(\neg C) \right].$$

Step 2: Substitute the given values

Substitute into the expression:

$$P(H, T) = 0.014 \left((0.56)(0.006) + (0.32)(0.994) \right]$$
$$= 0.014 \left[0.00336 + 0.31808 \right]$$
$$= 0.014 \times 0.32144$$
$$= 0.00450016.$$

Now, substitute into the conditional probability formula:

$$P(H \mid T) = \frac{0.00450016}{0.00073668} \approx 0.767$$

Conclusion:

$$\boxed{P(H \mid T) = 0.767\ \text{(i.e., about 76.7\%)}}$$

1.4 Using enumeration, write the equation for the P(injuries). Your equation must include the joint distribution for all 5 of the variables and the sums necessary to solve for P(injuries). You only need to write the equation. You do not need to solve for a numerical answer. - **10 pts**

Equation for P (I) (Probability of injuries) using enumeration:

First, note that the full joint distribution for the five variables is given by

$$P(H, C, T, D, I) = P(H) P(C) P(T \mid H, C) P(D \mid T) P(I \mid T, D),$$

where:

- (H) = Thunderstorms,
- (C) = Cold Front,
- (T) = Tornado,
- (D) = Property Damage,
- (I) = Injuries.

Then, to obtain $P(I)$, we marginalize over the other variables:

$$P(I) = \sum_{h \in \{H, \neg H\}} \sum_{c \in \{C, \neg C\}} \sum_{t \in \{T, \neg T\}} \sum_{d \in \{D, \neg D\}} P(H, C, T, D, I),$$

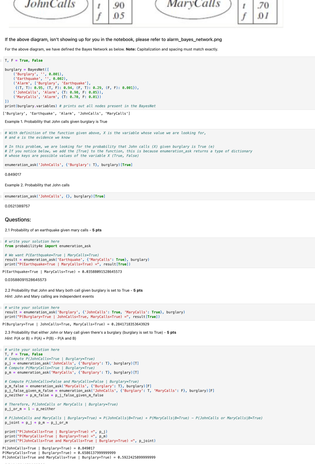
Substituting the factorized form, we have:

$$\boxed{P(I) = \sum_{h \in \{H, \neg H\}} \sum_{c \in \{C, \neg C\}} \sum_{t \in \{T, \neg T\}} \sum_{d \in \{D, \neg D\}} P(I \mid t, d) P(D \mid t) P(T \mid h, c) P(H) P(C).}$$

Solve through code -

Note: In some of the questions, the answer is below the cell to allow you to check your answer.

Problem 2



If the above diagram, isn't showing up for you in the notebook, please refer to alarm_bayes_network.png

For the above diagram, we have defined the Bayes Network as below. **Note:** Capitalization and spacing must match exactly.

In [2]:
```
T, F = True, False

burglary = BayesNet([
    ('Burglary', '', 0.001),
    ('Earthquake', '', 0.002),
    ('Alarm', 'Burglary Earthquake',
     {(T, T): 0.95, (T, F): 0.94, (F, T): 0.29, (F, F): 0.001}),
    ('JohnCalls', 'Alarm', {T: 0.90, F: 0.05}),
    ('MaryCalls', 'Alarm', {T: 0.70, F: 0.01})
])
print(burglary.variables) # prints out all nodes present in the BayesNet
```

```
['Burglary', 'Earthquake', 'Alarm', 'JohnCalls', 'MaryCalls']
```

Example 1 Probability that John calls given burglary is True

In [3]:
```
# With definition of the function given above, X is the variable whose value we are looking for,
# and e is the evidence we know.
# In this problem, we are looking for the probability that John calls (X) given burglary is True (e)
# If you notice above, we use the {True} in the function, this is because enumeration_ask returns a type of dictionary
# whose keys are possible values of the variable X (True, False)

enumeration_ask('JohnCalls', {'Burglary': T}, burglary).show()
```

0.849317

Example 2: Probability that John calls

In [4]:
```
enumeration_ask('JohnCalls', {}, burglary).show()
```

0.052158097

Questions:

2.1 Probability of an earthquake given mary calls - **5 pts**

In [5]:
```
# write your solution here
from probabilityAe import enumeration_ask

## We want P(Earthquake=True | MaryCalls=True)
result = enumeration_ask('Burglary', {'JohnCalls': True, 'MaryCalls': True}, burglary)
print("P(Burglary=True | JohnCalls=True, MaryCalls=True) =", result[True])
```

```
P(Burglary=True | JohnCalls=True, MaryCalls=True) = 0.2841718353643929
```

2.3 Probability that either John or Mary call given there's a burglary (burglary is set to True) - **5 pts**

Hint: P(A or B) = P(A) + P(B) - P(A and B)

In [6]:
```
# write your solution here
T, F = True, False
# P_JoinCallsGivenTrue = burglaryTrue)
p_j = enumeration_ask('JohnCalls', {'Burglary': T}, burglary)[T]
# Compute P(MaryCallsTrue | BurglaryTrue)
p_m = enumeration_ask('MaryCalls', {'Burglary': T}, burglary)[T]

# Compute P(JohnCallsTrue and MaryCallsTrue | BurglaryTrue)
p_j_and_m = enumeration_ask('JohnCalls', {'Burglary': T, 'MaryCalls': T}, burglary)[T]
p_neither = p_A_false_given_B_false

# Therefore, P(JohnCalls or MaryCalls | BurglaryTrue)
p_j_or_m = 1 - p_neither

# P(JohnCalls and MaryCalls | BurglaryTrue) = P(JohnCallsTrue) + P(MaryCallsTrue) - P(JohnCalls or MaryCallsTrue)
p_JAnM = p_j + p_m - p_j_or_m

print("P(JohnCallsTrue | BurglaryTrue) =", p_j)
print("P(MaryCallsTrue | BurglaryTrue) =", p_m)
print("P(JohnCallsTrue and MaryCallsTrue | BurglaryTrue) =", p_JAnM)

P(JohnCallsTrue | BurglaryTrue) = 0.84983
P(MaryCallsTrue | BurglaryTrue) = 0.6599999999999999
P(JohnCallsTrue and MaryCallsTrue | BurglaryTrue) = 0.5624169999999999
```

Problem 3

If the above diagram, isn't showing up for you in the notebook, please refer to bayes_network_2.png

Questions

3.1 For this network, what is P(Heart Disease)? - **10 pts**

In [7]:
```
# For convenience
T, F = True, False
# Define the Heart Disease Bayes Network
heart_disease_net = BayesNet([
    # Smoking & alcohol (SA) is an unconditional node:
    ('SA', '', 0.25),  # P(SmokeOrAlc_is_TRUE)=0.25

    # Moderate exercise (ME) is unconditional:
    ('ME', '', 0.60),  # P(ExerciseOK_is_TRUE)=0.60

    # High BP depends on (SA, ME)
    # Order: (SA, ME) in that table:
    ('HBP', 'SA ME', {
        (T, T): 0.52,  # SmkAlc=T ExOK=T
        (T, F): 0.84,  # SmkAlc=T ExOK=F
        (F, T): 0.30,  # SmkAlc=F ExOK=T
        (F, F): 0.75   # SmkAlc=F ExOK=F
    }),

    # Atherosclerosis (ATH) is unconditional:
    ('ATH', '', 0.33),  # P(AtherosMal_is_TRUE)=0.33

    # Family history (FH) is unconditional:
    ('FH', '', 0.15),  # P(FamilyHist_is_TRUE)=0.15

    # Heart disease (HD) depends on (HBP, ATH, FH)
    # Order: (HBP, ATH, FH) in that table:
    ('HD', 'HBP ATH FH', {
        (T, T, T): 0.89,
        (T, T, F): 0.82,
        (T, F, T): 0.79,
        (T, F, F): 0.61,
        (F, T, T): 0.72,
        (F, T, F): 0.77,
        (F, F, T): 0.55,
        (F, F, F): 0.22
    }),

    # Angina Pectoris (AngP) depends on:
    # P(AngP | HDcTrue)=0.2, P(AngP | HDcFalse)=0.42
    ('AngP', 'HD', {T: 0.2, F: 0.42}),

    # Rapid heartbeats (RapidH) depends on HD:
    # P(RapidH | HDcTrue)=0.84, P(RapidH | HDcFalse)=0.38
    ('RapidH', 'HD', {T: 0.84, F: 0.38})
])

# Write your solution here
p_hd = enumeration_ask('HD', {}, heart_disease_net)[T]
print("P(Heart Disease) =", p_hd)
```

```
P(Heart Disease) = 0.7351541000000001
```

3.2 You have a patient who Smokes and Consumes alcohol, doesn't exercise at all, and has a family history of heart disease. What is the probability that patient will have Angina Pectoris? - **5 pts**

In [8]:
```
# Write your solution here
p_ang = enumeration_ask('AngP', {'SA': T, 'ME': F, 'FH': T}, heart_disease_net)[T]
print("P(Angina Pectoris | SmkTrue, MEcNo, FHcYes) =", p_ang)
```

```
P(Angina Pectoris | SmkTrue, MEcNo, FHcYes) = 0.7036653800000001
```

3.3 You have a patient who has Angina Pectoris and Rapid Heartbeats. The patient also has a Family History of Heart Disease. What is the probability the patient has Atherosclerosis? - **5 pts**

In [9]:
```
# Write your answer here
p_ath = enumeration_ask('ATH', {'AngP': T, 'RapidH': T, 'FH': T}, heart_disease_net)[T]
print("P(Atherosclerosis | AnginaTrue, RapidHeartTrue, FHcYes) =", p_ath)
```

```
P(Atherosclerosis | AnginaTrue, RapidHeartTrue, FHcYes) = 0.5306818181818182
```

3.4 In your own words, describe the difference between diagnostic and causal reasoning in Bayesian networks. - **5 points**

Diagnostic reasoning goes from effect to cause; for example, we observe symptoms like prolonged coughing and infer the probability of underlying causes like lung cancer. Causal reasoning goes from cause to effect; for example, we start with risk factors like Smoking, and deduce the probability of outcomes like lung cancer.

3.5 How could a physician use this Bayes Network to help with counseling patients who have a family history of heart disease and they smoke and drink alcohol and don't exercise? Hint: You may want to compare your results from 4.1 to those from 4.2 to determine how the probabilities change depending on how the patient behaves. - **10 points**

A physician can use the network to show how lifestyle factors combine with a family history to increase the probability of Heart Disease and its complications. By comparing the probabilities with and without these risk factors, the physician can demonstrate the benefits of quitting smoking, reducing alcohol, and increasing exercise. This helps patients understand how changing their behavior can lower their overall risk.