Tutorial Sections 15-16

1. Consider a small town where people prefer either tea or coffee in the morning. Additionally, people in this town either read the newspaper or watch the news on TV. We are interested in understanding the relationship between these preferences.
   From a survey in the town, it was estimated that 60% of the people prefer tea and 50% revealed a preference for reading the newspaper. The preference for having tea and reading the newspaper simultaneously was estimated at 30%.
   (a) Calculate the probability of preferring coffee.
   (b) Calculate the probability of preferring coffee and watching TV news.
   (c) Calculate the probability that a person prefers tea given that they read the newspaper.
   (d) Calculate the probability that a person watches TV news given that they prefer coffee.

Solution: This is a simple problem to practice understanding marginal, joint, and conditional probabilities.
The problem considers two random variables (RVs), one that considers the preference for the drink and the other that considers the preference for being informed. Let's denote the preference for tea or coffee by a random variable T, where $T = 1$ represents tea, and $T = 0$ represents coffee. Similarly, the preference for reading the newspaper or watching TV news can be denoted by another random variable, N, where $N = 1$ represents reading the newspaper, and $N = 0$ represents watching TV news.
From the problem statement, $P(T = 1) = 0.6$, $P(N = 1) = 0.5$, and $P(T = 1, N = 1) = P(T = 1 \cap N = 1) = 0.3$.

(a) Since we are considering only two possibilities for drink preference (i.e., our sample space is $\Omega_T = \{0, 1\}$), and the probability of preferring tea is 0.6, the probability of preferring coffee is the complement,

$$P(T = 0) = 1 - P(T = 1) = 1 - 0.6 = 0.4$$

(b) We want to find $P(T = 0, N = 0)$. From one of the axioms of probability, we know that

$$P(A \cup B) = P(A) + P(B).$$

However, in the event that there can be an overlap between A and B, we need to account for the fact that we would be summing these elements twice by summing $P(A)$ and $P(B)$. So, whenever $A \cap B \neq \emptyset$, we would have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Applying this reasoning to our specific problem here, we have that:

$$P(T \cup N) = P(T) + P(N) - P(T \cap N).$$

In part (a), we found that $P(T = 0) = 0.4$. Applying the same reasoning for the preference of news, we have

$$P(N = 0) = 1 - P(N = 1) = 1 - 0.5 = 0.5.$$

Similarly, the probability of either preferring coffee or watching TV news or both can be calculated as:

$$P(T = 0 \cup N = 0) = 1 - P(T = 1 \cap N = 1) = 1 - 0.3 = 0.7$$

So,

$$\begin{aligned} P(T = 0 \cup N = 0) &= P(T = 0) + P(N = 0) - P(T = 0 \cap N = 0) \\ P(T = 0 \cap N = 0) &= 0.4 + 0.5 - 0.7 = 0.2. \end{aligned}$$

(c) We want to find $P(T=1|N=1)$, which can be done by using the definition of conditional probability:

$$P(T=1|N=1) = \frac{P(T=1, N=1)}{P(N=1)} = \frac{0.3}{0.5} = 0.6$$

(d) Similarly, we can find $P(N=0|T=0)$ using the definition of conditional probability:

$$P(N=0)|T=0) = \frac{P(N=0, T=0)}{P(T=0)} = \frac{0.2}{0.4} = 0.5$$

2. Hepatitis C is a liver disease caused by the Hepatitis C virus (HCV), which can range from a mild illness lasting a few weeks to a serious, lifelong illness. Despite its seriousness, it is fortunately relatively rare, present in 1% of the general population. Due to its significant impact on public health, Hepatitis C is often diagnosed through serological testing. A common serological test for Hepatitis C has a sensitivity (true positive rate) of 95% and a specificity (true negative rate) of 90%. Sensitivity refers to the test's ability to correctly identify individuals with Hepatitis C, while specificity refers to correctly identifying individuals without the disease.

   (a) A patient undergoes this test and receives a positive result. What is the probability that the patient actually has Hepatitis C?

   (b) If another patient from the same population undergoes the test and receives a negative result, what is the probability that this patient is actually free from Hepatitis C?

   (c) How would the probability of having Hepatitis C change if the prevalence of the disease in the general population was higher, say 5% instead of 1%?

Solution: This is a problem to practice Bayes' theorem, which is typically confusing for most people because we naturally tend to ignore the prior probability when reasoning about likelihoods.
When applied to this problem, Bayes' theorem tells us that

$$P(\text{Disease}|\text{Test}) = \frac{P(\text{Test}|\text{Disease}) \cdot P(\text{Disease})}{P(\text{Test})}.$$

From the text, we know

- the prior probability: $P(\text{Disease}=1) = 0.01$
- the sensitivity: $P(\text{Test}=P|\text{Disease}=1) = 0.95$
- the specificity: $P(\text{Test}=N|\text{Disease}=0) = 0.90$

(a) According to Bayes' theorem,

$$P(\text{Disease}=1|\text{Test}=P) = \frac{P(\text{Test}=P|\text{Disease}=1) \cdot P(\text{Disease}=1)}{P(\text{Test}=P)}.$$

According to the law of total probability, the probability of getting a positive test result is

$$\begin{aligned} P(\text{Test}=P) &= \sum_{D \in \Omega_D} P(\text{Test}=P|D) \cdot P(D) \\ &= P(\text{Test}=P|\text{Disease}=1) \cdot P(\text{Disease}=1) + P(\text{Test}=P|\text{Disease}=0) \cdot P(\text{Disease}=0) \\ &= 0.95 \times 0.01 + (1-0.90) \times (1-0.01) = 0.1085 \end{aligned}$$

So,

$$P(\text{Disease} = 1|\text{Test} = P) = \frac{0.95 \times 0.01}{0.1085} = 0.0876 \text{ or } 8.76\%$$

(i.e., unlike "common sense" would expect, receiving a positive result for a rare disease drastically increases your chances of being a disease, but it is still way less than 50%!)

(b) For this part, we want to know

$$P(\text{Disease} = 0|\text{Test} = N) = \frac{P(\text{Test} = N|\text{Disease} = 0) \cdot P(\text{Disease} = 0)}{P(\text{Test} = N)}.$$

Similarly,

$$\begin{aligned} P(\text{Test} = N) &= \sum_{D \in \Omega_D} P(\text{Test} = N|D) \cdot P(D) \\ &= P(\text{Test} = N|\text{Disease} = 0) \cdot P(\text{Disease} = 0) + P(\text{Test} = N|\text{Disease} = 1) \cdot P(\text{Disease} = 1) \\ &= 0.90 \times (1 - 0.01) + (1 - 0.95) \times 0.01 = 0.8915, \end{aligned}$$

so that

$$P(\text{Disease} = 0|\text{Test} = N) = \frac{0.9 \times (1 - 0.01)}{0.8915} = 0.9994 \text{ or } 99.94\%.$$

(c) If the prior probability changes to $P(\text{Disease} = 1) = 0.05$, then

$$P(\text{Test} = P) = 0.95 \times 0.05 + (1 - 0.90) \times (1 - 0.05) = 0.1425$$

and, therefore,

$$P(\text{Disease} = 1|\text{Test} = P) = \frac{0.95 \times 0.05}{0.1425} = 0.3333 \text{ or } 33.3\%,$$

which demonstrates how the probability of actually having the disease, given a positive test result, significantly increases with the prevalence of the disease.

3. Consider a given weather prediction model used to determine the likelihood of snowfall in a particular region based on two observable factors: Temperature (T) and Wind Speed (W). The event of Snow (S) and these two factors can be represented as binary random variables (e.g., $S = 1$ for snowfall, $S = 0$ for no snowfall; $T = 1$ for low temperature, $T = 0$ for high temperature; $W = 1$ for high wind speed, $W = 0$ for low wind speed).

   Based on historical weather data, analysts have estimated the following probabilities:

   - prior probability of snow = 25%
   - probability of low temperature given snow = 90%
   - probability of high wind speed given snow = 50%
   - probability of low temperature without Snow = 20%
   - probability of high wind speed without Snow = 30%

   (a) Create a Bayesian network that represents the relationship between Temperature, Wind Speed, and Snowfall.

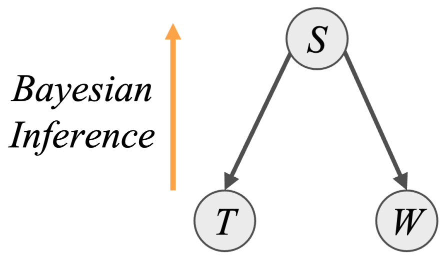
   (b) Calculate the probability of Snow given that both low temperature and high wind are observed.

   (c) Suppose you were given a dataset with the following test observations:
   For each observation, determine which class has the highest likelihood probability for Snow.

Observation	Temperature	Wind Speed
1	low	high
2	high	high
3	low	low

Solution: This problem provides an example of probabilistic classification using Bayes' theorem. We want to determine the probability $P(S|T,W)$ of the class $S = 1$, considering $\Omega_S = \{0,1\}$, to which T, W belong, taking into account the prior $P(S)$, i.e., the probability of Snow before having seen the data $P(T,W|S)$.

(a) The PGM is designed to capture the dependencies of T and W on S. The root node is S; it has no parent nodes, as $P(S)$ gives the prior probability. Since the probability of temperature being low or high depends on whether it's snowing, T is a child of S. Similarly, the probability of wind speed is conditional on the occurrence of Snow. This scenario has no direct dependence between T and W. Given all that, the PGM of this naive Bayes classifier can be represented by the figure below:



(b) Given the data $P(S=1) = 0.25$, $P(T=1|S=1) = 0.9$, $P(W=1|S=1) = 0.5$, $P(T=1|S=0) = 0.20$, and $P(W=1|S=0) = 0.3$, we want to calculate $P(S=1|T=1,W=1)$. Using Bayes' theorem,

$$P(S=1|T=1,W=1) = \frac{P(T=1,W=1|S=1)\cdot P(S=1)}{P(T=1,W=1)}.$$

Since T and W are conditionally independent,

$$\begin{aligned} P(T=1,W=1|S=1) &= P(T=1|S=1) \times P(W=1|S=1) \\ &= 0.9 \times 0.5 = 0.45 \end{aligned}$$

The total probability of both low temperature and high wind speed is given by:

$$P(T=1,W=1) = P(T=1,W=1|S=1)\cdot P(S=1) + P(T=1,W=1|S=0)\cdot P(S=0),$$

where

$$P(T=1,W=1|S=0) = P(T=1|S=0) \times P(W=1|S=0) = 0.2 \times 0.3 = 0.06.$$

Therefore,

$$P(T=1,W=1) = 0.45 \times 0.25 + 0.06 \times (1-0.25) = 0.1575.$$

Plugging the numbers into Bayes' theorem, we have

$$P(S=1|T=1,W=1) = \frac{0.45 \times 0.25}{0.1575} = 0.71$$

(c) The goal is to select the value of S that maximises the posterior $P(S|T,W)$ (maximum *a-posteriori* decision rule):

$$s^* = \arg\max_{s \in \Omega_S} P(T = t_i, W = w_i | S = s) P(S = s),$$

which can be rewritten as

$$s^* = \arg\max_{s \in \Omega_S} \left(P(T = t_i | S = s) \times P(W = w_i | S = s) \times P(S = s) \right)$$

since T and W are conditionally independent and $P(W,T)$ is independent of S. For Observation 1, we have $T = 1$ and $W = 1$, so:

$$\begin{aligned}
\text{Probability of snow } (s = 1 \in \Omega_S) &= P(T = 1|S = 1) \times P(W = 1|S = 1) \times P(S = 1) \\
&= 0.9 \times 0.5 \times 0.25 = 0.1125 \\
\text{Probability of no snow } (s = 0 \in \Omega_S) &= P(T = 1|S = 0) \times P(W = 1|S = 0) \times P(S = 0) \\
&= 0.2 \times 0.3 \times 0.75 = 0.045
\end{aligned}$$

So, the classifier decision can be represented as

$$s^* = \arg\max\,(0.1125, 0.045) = 0.1125,$$

which means the highest likelihood is for Snow ($S = 1$), given Observation 1.
For Observation 2, we have $T = 0, W = 1$, so:

$$\begin{aligned}
\text{Probability of snow } (s = 1 \in \Omega_S) &= P(T = 0|S = 1) \times P(W = 1|S = 1) \times P(S = 1) \\
&= (1 - 0.9) \times 0.5 \times 0.25 = 0.0125 \\
\text{Probability of no snow } (s = 0 \in \Omega_S) &= P(T = 0|S = 0) \times P(W = 1|S = 0) \times P(S = 0) \\
&= (1 - 0.2) \times 0.3 \times 0.75 = 0.18,
\end{aligned}$$

So, the classifier decision is

$$s^* = \arg\max\,(0.0125, 0.18) = 0.18,$$

which means the highest likelihood is for no Snow ($S = 0$).
Last, for Observation 3, we have $T = 1, W = 0$:

$$\begin{aligned}
\text{Probability of snow } (s = 1 \in \Omega_S) &= P(T = 1|S = 1) \times P(W = 0|S = 1) \times P(S = 1) \\
&= 0.9 \times (1 - 0.5) \times 0.25 = 0.1125 \\
\text{Probability of no snow } (s = 0 \in \Omega_S) &= P(T = 1|S = 0) \times P(W = 0|S = 0) \times P(S = 0) \\
&= 0.2 \times (1 - 0.3) \times 0.75 = 0.105,
\end{aligned}$$

So, the classifier decision is

$$s^* = \arg\max\,(0.1125, 0.105) = 0.1125,$$

and, therefore, the highest likelihood given Observation 3 is Snow ($S = 1$).