STA 247 — Solutions to Assignment #1, Part I

1. Suppose that A, B, and C are events for which $P(B)$, $P(B \cap C)$, and $P(B \cap C^c)$ are not zero. Using only the basic axioms of probability (on page 8), the definition of conditional probability (on page 27), and properties of sets, prove that

$$P(A \,|\, B) \;=\; P(A \,|\, B \cap C)P(C \,|\, B) \;+\; P(A \,|\, B \cap C^c)P(C^c \,|\, B)$$

Solution: Using the definition of conditional probability, we can rewrite the right side as

$$\begin{aligned} & P(A \,|\, B \cap C)P(C \,|\, B) \;+\; P(A \,|\, B \cap C^c)P(C^c \,|\, B) \\ &\quad = \; \frac{P(A \cap B \cap C)}{P(B \cap C)} \frac{P(C \cap B)}{P(B)} \;+\; \frac{P(A \cap B \cap C^c)}{P(B \cap C^c)} \frac{P(C^c \cap B)}{P(B)} \end{aligned}$$

Since $B \cap C = C \cap B$ and $B \cap C^c = C^c \cap B$, we can cancel the factors of $P(B \cap C)$ and $P(B \cap C^2)$ in the numerators and denominators, giving

$$\begin{aligned} P(A \,|\, B \cap C)P(C \,|\, B) \;+\; P(A \,|\, B \cap C^c)P(C^c \,|\, B) \;&=\; \frac{P(A \cap B \cap C)}{P(B)} \;+\; \frac{P(A \cap B \cap C^c)}{P(B)} \\ &=\; \frac{P(A \cap B \cap C) \;+\; P(A \cap B \cap C^c)}{P(B)} \end{aligned}$$

$A \cap B \cap C$ and $A \cap B \cap C^c$ are mutually exclusive, since $(A \cap B \cap C) \cap (A \cap B \cap C^c) \;=\; A \cap B \cap (C \cap C^c) \;=\; A \cap B \cap \emptyset \;=\; \emptyset$ (informally, C and C^c can't both happen). Also, $(A \cap B \cap C) \cup (A \cap B \cap C^c) \;=\; A \cap B \cap (C \cup C^c) \;=\; A \cap B$ (informally, $A \cap B \cap C$ and $A \cap B \cap C^c$ split $A \cap B$ into the two parts where C does or doesn't happen). The third basic axiom of probability therefore allows us to conclude that $P(A \cap B \cap C) \;+\; P(A \cap B \cap C^c) = P(A \cap B)$. Substituting this in the equation above, and using the definition of conditional probability, we obtain the desired result:

$$P(A \,|\, B \cap C)P(C \,|\, B) \;+\; P(A \,|\, B \cap C^c)P(C^c \,|\, B) \;=\; \frac{P(A \cap B)}{P(B)} \;=\; P(A \,|\, B)$$

2. A programmer writes a program to construct poems automatically, as follows. First, the program randomly selects ten distinct words from a dictionary. Then, starting with a poem with no words, the program repeatedly picks a word randomly from among these ten words, with equal probabilities, and adds it to the end of the poem. Words can be selected and added to the poem more than once, but when the word selected is the same as the word selected immediately before, the program stops, without adding this word to the end of the poem a second time.

   What is the probability that a poem produced by this program will be exactly three words long? If the poem is exactly three words long, how likely it is that the first and last words will be the same?

   Solution: Let W be the event that the poem is exactly three words long, and S be the event that the poem is three words long and the first and third words of the poem are the same.

This problem can be solved in at least two ways.

One way is to identify a suitable sample space, and then identify the subsets of the sample space corresponding to W and S. First, note that we needn't worry about which 10 words were picked from the dictionary, since all that matters is that they are distinct. Since the poems can be indefinitely long, so it might seem that we need an infinite sample space. However, the questions we're interested in can be answered using a sample space that describes only the first four words (from the set of 10) that the program picks. (The program will pick fewer than four words if the poem is only one or two words long, but we can imagine that it picks at least four words even if it doesn't need them.) These first four words are enough to determine whether the poem is exactly three words long or not.

Using as the sample space the sequence of four words, we see that there are $10^4 = 10000$ points in the sample space, all of which are equally likely. The number of these points in the event W is $10 \times 9 \times 9 \times 1 = 810$. This is obtained from 10 possibilities for the first word picked, 9 possibilities for the second that don't lead to the poem being only one word long, 9 possibilities for the third that don't lead to the poem being only two words long, and 1 possibility for the fourth if the poem is to end after three words. This gives

$$P(W) \;=\; 810/10000 \;=\; 0.081$$

The number of points in this sample space in the event S is $10 \times 9 \times 1 \times 1 = 90$. There are 10 possibilities for the first word, 9 possibilities for the second that don't lead to the poem being only one word long, 1 possibility for the third word if it is to be the same as the first, and 1 possibility for the fourth if the poem is to end after three words. Note that if the third word is the same as the first word, it will also be different from the second word. From this, we have that

$$P(S) \;=\; 90/10000 \;=\; 0.009$$

This isn't the answer, however. The probability that the first and third words are the same if the poem is exactly three words long is

$$\begin{aligned}
&P(\text{first and third words the same} \mid \text{three words long}) \\
&\quad= \frac{P(\text{three words long and first and third words the same})}{P(\text{three words long})} \\
&\quad= P(S)/P(W) \;=\; 0.009/0.081 \;=\; 1/9
\end{aligned}$$

The second way to solve the problem is to think directly in terms of conditional probabilities for the program to do things that are compatible with the event W or the event S. The choice for the first word doesn't matter. The second word must not be the same as the first if W is to occur, which has probability 9/10. The third must be different from the second for W to occur, which also has probability 9/10. The fourth must be the same as the third for W to occur, which has probability 1/10. Multiplying these together gives

$$P(W) \;=\; (9/10)(9/10)(1/10) \;=\; 0.081$$

Similary, we find that

$$P(S) \;=\; (9/10)(1/10)(1/10) \;=\; 0.009$$

And hence $P(S)/P(W) = 1/9$, as before.

3. Suppose that computers A and C can communicate only via computer B. So, for a network packet to get from computer A to computer C, it must first be sent from computer A to computer B, and then from computer B to computer C. Packets sent from computer A to computer B usually take 5 milliseconds, but with probability 0.1, the packet will take 15 milliseconds. Packets sent from computer B to computer C usually take 10 milliseconds, but with probability 0.2, the packet will take 20 milliseconds. The time taken from A to B is independent of the time taken from B to C. The time required for other operations is negligible.

   Let X be the total time for a packet to be sent from computer A to computer C. Find the probability mass function for X, and compute $E(X)$ and $STD(X)$.

   Solution: Let U be the time to send a packet from computer A to computer B, and let V be the time to send a packet from computer B to computer C. We know that $X = U + V$. We also know the probability mass functions for U and V, which are as follows:

u	5	15
$p(u)$	0.9	0.1

v	10	20
$p(v)$	0.8	0.2

   To find the probability mass function for X, we need to consider all the ways that a particular value for X could come about from a combination of values for U and for V. There are four possible combinations of values for U and V, but these result in only three possible values for X, with probabilities as follows:

x	15	25	35
$p(x)$	0.72	0.26	0.02

   These probabilities are found as follows, using the fact that U and V are independent:

$$\begin{aligned}
P(X = 15) &= P(U = 5, V = 10) \;=\; 0.9 \times 0.8 \\
P(X = 25) &= P(U = 5, V = 20) + P(U = 15, V = 10) \;=\; 0.9 \times 0.2 + 0.1 \times 0.8 \\
P(X = 35) &= P(U = 15, V = 20) \;=\; 0.1 \times 0.2
\end{aligned}$$

   We can now easily find $E(X)$:

$$E(X) \;=\; 15 \times 0.72 + 25 \times 0.26 + 35 \times 0.02 \;=\; 18$$

   Alternatively, we could have found this by

$$E(X) \;=\; E(U + V) \;=\; E(U) + E(V) \;=\; (5 \times 0.9 + 15 \times 0.1) + (10 \times 0.8 + 20 \times 0.2) \;=\; 18$$

   We can find the variance of X as follows:

$$\begin{aligned}
\mathrm{Var}(X) &= E((X - E(X))^2) \\
&= (15 - 18)^2 \times 0.72 + (25 - 18)^2 \times 0.26 + (35 - 18)^2 \times 0.02 \;=\; 25
\end{aligned}$$

   The standard deviation of X is the square root of this, which is 5.

   Since U and V are independent, we could also have found the variance of X from

$$\mathrm{Var}(X) \;=\; \mathrm{Var}(U + V) \;=\; \mathrm{Var}(U) + \mathrm{Var}(V)$$