# Probability Homework 3:

Conditional Probability
Law of Total Probability
Independence
Bayes Rule

Tong Jin
NYU Steinhardt A3SR
October 7, 2019

**Question 1**: Based on records of automobile accidents in a recent year, the Department of Highway Safety and Motor Vehicles in Florida reported the counts who survived (S) and died (D) according to whether they wore a seat belt or not. The data are presented in the following table.

| Wore Seat Belt | Survived | Died | Total |
|---|---|---|---|
| Yes | 412,368 | 510 | 412,878 |
| No | 162,527 | 1,601 | 164,128 |
| Total | 574,895 | 2,111 | 577,006 |

(a) What is the sample space of possible outcomes for a randomly selected individual involved in an auto accident?

**Solution**: There are two events under the argument "Wore seat belt or not":

Yes (Wore seat belt)
No (Did not wear seat belt)

There are also two events under the argument "Survived or not":

Survived (Survived)
Died (Died)

Therefore, the sample space is:

**(Yes, Survived), (Yes, Died), (No, Survived), (No, Died)**

(b) Using these data, estimate Pr(Died), Pr(Survived), Pr(Yes) and Pr(No).

**Solution**:
The probability of Died, in this case, is:

$$Pr(Died) = \frac{TotalDied}{Total} = \frac{2,111}{577,006} \approx \mathbf{0.00365854} \tag{1.1}$$

The probability of cases with survived victims, in this case, is:

$$Pr(Survived) = \frac{TotalSurvived}{Total} = \frac{574,895}{577,006} \approx \mathbf{0.99634146} \tag{1.2}$$

The probability of cases with victims wearing seat belts, in this case, is:

$$Pr(Yes) = \frac{TotalWoreSeatBelt}{Total} = \frac{412,878}{577,006} \approx \mathbf{0.71555235} \tag{1.3}$$

The probability of cases with victims not wearing seat belts, in this case, is:

$$Pr(No) = \frac{TotalDidNotWearSeatBelt}{Total} = \frac{164,128}{577,006} \approx \mathbf{0.28444765} \tag{1.4}$$

(c) Find the probability that an individual did not wear a seat belt and died.

**Solution**:
The probability, in this case, is:

$$Pr(No \cap Died) = \frac{1,601}{577,006} \approx \mathbf{0.002774668} \tag{1.5}$$

(d) Given that an individual did not wear a seat belt, find the probability that they died.

**Solution**:
The probability, in this case, is:

$$Pr(Died|No) = \frac{Pr(Died \cap No}{Pr(No)} \frac{1,601}{164,128} \approx \mathbf{0.009754582} \tag{1.6}$$

(e) Given that an individual did wear a seat belt, find the probability that they died. Interpret the results of (d) and (e) in the context of the problem.

**Solution**:
The probability, in this case, is:

$$Pr(Died|Yes) = \frac{510}{412,878} \approx \mathbf{0.001235232} \tag{1.7}$$

The probability of an individual who died in an accident, given that the person wore seat belt, is approximately 0.00124. The probability of an individual who died in an accident, given that the person did not wear seat belt, is approximately 0.00975.

The ratio between Pr(Died—No) and Pr(Died—Yes) is:

$$\frac{Pr(Died|No)}{Pr(Died|Yes)} \approx 7.896963 \approx \mathbf{8} \tag{1.8}$$

We can interpret that the chance of fatality of an individual who does not wear seat belt in a car accident is approximately 8 times higher than the change of fatality of an individual who does wear seat belt.

(f) Are the events Died and No independent? Why or why not?

**Solution**: The probability that an individual did not wear a seat belt and died is:

$$Pr(No\&Died) \approx \mathbf{0.001040663} \tag{1.9}$$

The probability that an individual died given that the individual did not wear a seat belt is:

$$Pr(Died|No) \approx \mathbf{0.009754582} \tag{1.10}$$

Since Pr(No & Died) and Pr(Died—No) are not equal, we can say the the event Died and No are not independent. They are dependent events.

**Question 2**: In a certain country, the probability that a woman reaches the age of 70 is 0.7 and the probability that a woman lives to be 80 is 0.5.

(a) If a woman from that country is 70 years old, what is the conditional probability that she will survive to 80 years?

**Solution**: The probability of reaching to 70 years old is:

$$Pr(70) = 0.7 \tag{2.1}$$

The probability of reaching to 80 years old is:

$$Pr(80) = 0.5 \tag{2.2}$$

The probability, in this case, is:

$$Pr(80|70) = \frac{Pr(80) \cap Pr(70)}{Pr(70)} = \frac{0.5}{0.7} \approx \mathbf{0.7142857} \tag{2.3}$$

(b) Suppose that three sisters born in that country live completely independent lives. What is the probability that they will be alive at the age of 70?

**Solution**: Suppose there are sister A, B and C. The probability of A who lives to 70 is:

$$Pr_A(70) = 0.7 \tag{2.4}$$

Similarly, the probability of B and C who live to 70 is:

$$Pr_B(70) = 0.7 Pr_C(70) = 0.7 \tag{2.5}$$

Now, because three sisters live completely independent lives, we can say that these are independent events. Therefore, the probability of three sisters all live to 70 is:

$$Pr(All70) = Pr_A(70) \cdot Pr_B(70) \cdot Pr_C(70) = 0.7^3 = \mathbf{0.343} \tag{2.6}$$

(c) If the three sisters are all alive at the age of 70, what is the probability that they will all be alive at the age of 80?

**Solution**: The probability, in this case, is:

$$Pr(\text{All 80}|\text{All 70}) = \frac{Pr(\text{All 80} \cap \text{All 70})}{Pr(\text{All 70}} = \frac{0.5^3}{0.343} \approx \mathbf{0.3644315} \tag{2.7}$$

**Question 3**: Consider the following experiment: First, a fair six-sided die is rolled and the top number is recorded. If the number recorded is even, a fair coin is flipped and the outcome is recorded. If the number recorded from the die is odd, a card is drawn from a standard (52 card) deck and the suit of the card (spade ♠, heart ♡, diamond ♢ or club ♣) is recorded.

(a) List the sample points in this experiment. Note that each sample point should be a pair of outcomes consisting of a die roll and either a suit or coin flip. What is the probability of each sample point?

**Solution**:
If we note (1, 2, 3, 4, 5, 6) for die results and (Head, Tail) for coin results, then:
Each with probability of $\frac{1}{12}$:

(2, Head), (2, Tail)
(4, Head), (4, Tail)
(6, Head), (6, Tail)

Each with probability of $\frac{1}{24}$:

(1, Spade), (1, Heart), (1, Diamond), (1, Club)
(3, Spade), (3, Heart), (3, Diamond), (3, Club)
(5, Spade), (5, Heart), (5, Diamond), (5, Club)

(b) Find the probability that a heart is drawn.

**Solution**:
The probability that a heart is drawn is:

$$Pr(Heart) = 3 \cdot \frac{1}{24} = \mathbf{0.125} \tag{3.1}$$

(c) Find the probability that either a heart is drawn or a tail is flipped.

**Solution**:
The probability, in this case, is:

$$Pr(Heart \cup Tail) = 3 \cdot \frac{1}{24} + 3 \cdot \frac{1}{12} = \mathbf{0.375} \tag{3.2}$$

(d) Find the probability that both a heart is drawn and a tail is flipped. What can be said about these two events?

**Solution**:
The probability, in this case, is:

$$Pr(Heart \cap Tail) = \mathbf{0} \tag{3.3}$$

They are mutually exclusive events.

(e) Given that a heart was drawn, what is the probability that the die rolled a 3?

**Solution**:
The probability, in this case, is:

$$Pr(3|Heart) = \frac{1}{3} \approx \mathbf{0.3333333} \tag{3.4}$$

**Question 4**: Consider the following experiment: A ball is drawn from an urn containing an equal number of red, white and blue balls. If the ball drawn is white, a fair coin is flipped and the outcome is recorded. If the ball is blue, a card is drawn from a standard (52 card) deck and the suit (club ♣, diamond ♢, heart ♡ and spade ♠, present in equal proportions in the deck) is recorded. If the ball is red, a fair six-sided die is rolled and the top number is recorded.

(a) List the sample points in this experiment. Note that sample points should be pairs of occurrences consisting of a color and a suit or roll. What is the probability of each sample point?

**Solution**:
If we note (1, 2, 3, 4, 5, 6) for die results, (Heart, Diamond, Club, Spade) for card results, and (Head, Tail) for coin results, then:
Each with probability of $\frac{1}{18}$:

**(1, Red), (2, Red), (3, Red), (4, Red), (5, Red), (6, Red)**

Each with probability of $\frac{1}{6}$:

**(Head, White), (Tail, White)**

Each with probability of $\frac{1}{12}$:

**(Club, Blue), (Diamond, Blue), (Heart, Blue), (Spade, Blue)**

(b) What is the probability that the die rolls an odd number?

**Solution**:
The probability, in this case, is:

$$Pr(odd) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6} \approx \mathbf{0.1666667} \tag{4.1}$$

(c) What is the probability that a blue ball is drawn and a heart is drawn?

**Solution**:
The probability, in this case, is:

$$Pr(Blue \cap Heart) = Pr(Blue) \cdot Pr(Heart) = \frac{1}{3} \cdot \frac{1}{4} \approx \mathbf{0.0833333} \tag{4.2}$$

(d) Are the events {Blue ball is drawn} and {Heart is drawn} independent? Justify your answer.

**Solution**: The probability that Blue ball is drawn is $\frac{1}{3}$.
The probability that Heart is drawn is $\frac{1}{12}$.
The probability that Blue ball is drawn is:

$$Pr(Blue) = \frac{1}{3} \tag{4.3}$$

They are not independent events because $Pr(Heart) \cdot Pr(Blue) \neq Pr(Blue \cap Heart)$

**Question 5**: In *Molecular Systems Biology* (Vol. 3, 2007), geneticists at the Università di Napoli (Italy) used reverse engineering to identify genes. They calculated $\Pr(G|D)$, where $D$ is a gene expression dataset of interest and $G$ is a graphical identifier. Several graphical identifiers were investigated. Suppose that two different graphical identifiers, $G_1$ and $G_2$, are investigated. Suppose also that $\Pr(D|G_1) = 0.5$, $\Pr(D|G_2) = 0.3$, $\Pr(G_1) = 0.2$ and $\Pr(G_2) = 0.8$.

(a) Find the probability of the first graphical identifier, $G_1$, given the gene expression dataset of interest, $D$.

**Solution**:
First, since two graphical identifiers are investigated in this case, then:

$$Pr(G_1^c) = 1 - Pr(G_1) = Pr(G_2) = 0.8$$
$$Pr(G_2^c) = 1 - Pr(G_2) = Pr(G_1) = 0.2$$

The probability, in this case, is:

$$Pr(G_1|D) = \frac{Pr(G_1 \cap D)}{Pr(D)} = \frac{Pr(G_1) \cdot Pr(D|G_1)}{Pr(D|G_1) \cdot Pr(G_1) + Pr(D|G_2) \cdot Pr(G_2)} = \frac{0.1}{0.34} \approx \mathbf{0.294} \tag{5.1}$$

(b) Find the probability of the second graphical identifier, $G_2$, given the gene expression dataset of interest, $D$.

**Solution**:
The probability, in this case, is:

$$Pr(G_2|D) = \frac{Pr(G_2 \cap D)}{Pr(D)} = \frac{Pr(G_2) \cdot Pr(D|G_2)}{Pr(D|G_2) \cdot Pr(G_2) + Pr(D|G_1) \cdot Pr(G_1)} = \frac{0.24}{0.34} \approx \mathbf{0.706} \quad (5.2)$$

**Question 6**: John is a very forgetful person, especially when it comes to his umbrella. Suppose that in the town where John lives, the probability that it is raining on any given day is 0.6. Furthermore, suppose that on 40% of all days, it is both raining and John has forgotten his umbrella at home.

(a) Calculate the probability that on a rainy day, John will forget his umbrella at home.

**Solution**:
First, suppose we use "Leave" and "Take" to notate the events that John forgets or doesn't forget his umbrella. We know that:
$Pr(Rain) = 0.6$
$Pr(NotRain) = 1 - Pr(Rain) = 0.4$
$Pr(Rain) \cap Pr(Leave) = 0.4$
The probability of $Pr(Leave|Rain)$ is:

$$Pr(Leave|Rain) = \frac{Pr(Rain) \cap Pr(Leave)}{Pr(Rain)} = \frac{0.4}{0.6} \approx \mathbf{0.667} \quad (6.1)$$

(b) Suppose John realizes that given it is not raining, he has not forgotten his umbrella 80% of the time. Find the probability that, regardless of the weather, John forgets his umbrella.

**Solution**:
We know that:
$Pr(Take|NotRain) = 0.8$
$Pr(Leave|NotRain) = 1 - Pr(Take|NotRain) = 0.2$
The probability of $Pr(Leave)$ is:

$$Pr(Leave) = Pr(Leave|Rain) \cdot Pr(Rain) + Pr(Leave|NotRain) \cdot Pr(NotRain) = \mathbf{0.48} \quad (6.2)$$

(c) On his way to work, John realizes that he has forgotten his umbrella. Knowing this, what is the probability that it is raining outside?

**Solution**:
The probability of $Pr(Rain|Leave)$ is:

$$Pr(Rain|Leave) = \frac{Pr(Rain) \cap Pr(Leave)}{Pr(Leave)} = \frac{0.4}{0.48} \approx \mathbf{0.833} \tag{6.3}$$

**Question 7**: We know the following about a calorimetric method used to test lake water for nitrates. If a water specimen contains nitrates, a solution dropped into the water will cause the specimen to turn red 95% of the time. When used on water specimens without nitrates, the solution causes the water to turn red 10% of the time (because the chemicals other than nitrates are sometimes present and they also react to the agent). Past experience in a lab indicates that nitrates are contained in 30% of the water specimens that are sent to the lab for testing.

(a) If a water specimen is randomly selected from among those sent to the lab, what is the probability that it will turn red when tested?
**Solution**:
Suppose we use "Yes" and "No" to notate whether a water specimen contains nitrates. We know that:
$Pr(Red|Yes) = 0.95$
$Pr(Red|No) = 0.1$
$Pr(Yes) = 0.3$
$Pr(No) = 1 - Pr(Yes) = 0.7$
The probability of $Pr(Red)$ is:

$$Pr(Red) = Pr(Red|Yes) \cdot Pr(Yes)|Pr(Red|No) \cdot Pr(No) = 0.95 \cdot 0.3 + 0.1 \cdot 0.7 = \mathbf{0.355} \tag{7.1}$$

(b) If a water specimen is randomly selected and turns red when tested, what is the probability that it actually contains nitrates?
**Solution**:
The probability of $Pr(Yes|Red)$ is:

$$Pr(Yes|Red) = \frac{Pr(Yes) \cap Pr(Red)}{Pr(Red)} = \frac{0.95 \cdot 0.3}{0.355} \approx \mathbf{0.803} \tag{7.2}$$

**Question 8**: A restaurant is interested in understanding the behaviours of its customers. Suppose that they realize that 25% of their customers are under the age of 30, and another 40% of their customers are between 30 and 50, with the rest over the age of 50. Furthermore, suppose that the amount of money spent at the restaurant depends on the age of the customer. If bills are categorized into two groups, over \$100 and under \$100, then for customers under the age of 30, the probability that the bill is over \$100 at this restaurant is 0.4. Similarly, for customers between the ages of 30 and 50, the probability that the bill is over \$100 is 0.7, and for customers over 50 years of age, the probability that the bill is over \$100 is 0.6.

(a) Find the probability that a randomly selected bill at this restaurant will be over \$100.

**Solution**:
First, we know that:
$Pr(30-) = 0.25$
$Pr(30-50) = 0.4$
$Pr(50+) = 0.35$
$Pr(\$100+|30-) = 0.4$
$Pr(\$100+|30-50) = 0.7$
$Pr(\$100+|50+) = 0.6$
In this case, the probability of $Pr(\$100+)$ is:

$$Pr(\$100+) =$$
$$Pr(\$100+|30-) \cdot Pr(30-) + Pr(\$100+|30-50) \cdot Pr(30-50) + Pr(\$100+|50+) \cdot Pr(50+) =$$
$$0.25 \cdot 0.4 + 0.4 \cdot 0.7 + 0.35 \cdot 0.6 =$$
$$\mathbf{0.59}$$

(b) If a bill is over \$100, which age group is it most likely for the customer to be in? Justify your answer by finding the appropriate probabilities.

**Solution**:
Let A, B and C represent three different age groups. The possibility for a bill over \$100 to be landed in each age group is:

$$Pr_A(\$100) = Pr(30-) \cdot Pr(\$100+|30-) = 0.1$$
$$Pr_B(\$100) = Pr(30-50) \cdot Pr(\$100+|30-50) = \mathbf{0.28}$$
$$Pr_C(\$100) = Pr(50+) \cdot Pr(\$100+|50+) = 0.21$$

**Therefore, it is more likely that a bill over \$100 will be landed in the 30-50 years old age group.**

**Question 9**: The worldwide population of Muslims was estimated in 2010 by Pew Research to be 1.6 billion[1]. A report from 2014 from the Bipartisan Policy institute estimated membership of terrorist groups at between 85,000 and 106,000.[2] The world population is currently 7.5 billion people. A politician is claiming that 95% of terrorists are Muslims.

[1] http://www.pewresearch.org/fact-tank/2016/07/22/muslims-and-islam-key-findings-in-the-u-s-and-around-the-world/
[2] http://bipartisanpolicy.org/library/2014-jihadist-terrorism-and-other-unconventional-threats/

(a) Using this percentage and the data above to calculate the probability that a person is a terrorist given that they are Muslims, using the upper estimate for terrorist group membership.

**Solution**:
From that politician's claim, we know that the probability that a Muslim is a terrorist is:

$$Pr(Muslim|Terrorist) = 0.95$$

The probability of an individual being a Muslim is:

$$Pr(Muslim) = \frac{1.6}{7.5} \approx 0.2133333$$

The probability of an individual being a terrorist is:

$$Pr(Terrorist) = \frac{106{,}000}{7{,}500{,}000{,}000} \approx 0.00001413333$$

Then, the probability of $Pr(Terrorist|Muslim)$ is:

$$Pr(Terrorist|Muslim) = \frac{Pr(Muslim|Terrorist) \cdot Pr(Terrorist}{Pr(Muslim)} \approx \mathbf{0.00006293746} \tag{9.1}$$

b) The politician's claim was actually made up on the spot. According to an FBI report, of terrorist attacks between 1980 and 2005, only 6% of them were committed by muslims.[3] Using this to mean that 6% of terrorists are Muslims, revise your calculation to include this more accurate estimate.

**Solution**:
Now we know that:
$Pr(Muslim|Terrorist) = 0.06$
Therefore,

$$Pr(Terrorist|Muslim) = \frac{Pr(Muslim|Terrorist) \cdot Pr(Terrorist}{Pr(Muslim)} \approx \mathbf{0.000003975} \tag{9.2}$$

**Question 10**: Jay and Maurice are playing a tennis match. In one particular game, they have reached a deuce, which means that each player has won three points. To finish the game, one of the two players must get two points ahead of the other. For example, Jay will win if he wins the next two points (JJ), or if Maurice wins the next point and Jay the three points after that (MJJJ), or if the result of the next six points is JMMJJJ, and so on.

a) Suppose that the probability of Jay winning a point is 0.6 and outcomes of successive points are independent of one another. What is the probability that Jay wins the game? [Hint: In the law of total probability, let $A_1$ be the event that Jay wins each of the next two points, $A_2$ be the event

[3] https://www.fbi.gov/stats-services/publications/terrorism-2002-2005

that Maurice wins each of the next two points, and $A_3$ be the event that each player wins one of the next two points. Also, let $p = P(\text{Jay wins the game}|A_3)$]

**Solution**:
First, let $p = P(\text{Jay wins the game}|A_3)$].
Let $A_1$ be the event that Jay wins each of the next two points. The probability of $A_1$ is:

$$Pr(A_1) = 0.6 \cdot 0.6 = 0.36 \tag{10.1}$$

Let $A_3$ be the event that each player wins one of the next two points. The probability of $A_3$ is:

$$Pr(A_3) = 0.6 \cdot 0.4 + 0.6 \cdot 0.4 = 0.48 \tag{10.2}$$

Then,

$$Pr(\text{Jay Wins}) = Pr(A_1) + Pr(A_3) \cdot p \tag{10.3}$$

Since $Pr(\text{Jay Wins})$ is based on $Pr(A_3)$, which is the deuce situation,
we can say that $p = Pr(\text{Jay wins the game}|A_3) = Pr(\text{Jay Wins})$. Then,

$$p = Pr(A_1) + Pr(A_3) \cdot p \tag{10.4}$$

The probability, in this case, is:

$$p = \frac{0.36}{0.52} \approx \mathbf{0.6923077} \tag{10.5}$$

b) If Jay wins the game, what is the probability that he needed only two points to do so?

**Solution**:
The probability, in this case, is:

$$Pr(\text{Jay Wins in first two sets}|\text{Jay Wins}) = \frac{Pr(A_1 \cap \text{Jay Wins})}{Pr(\text{Jay Wins})} = \frac{0.36}{0.6923077} = \mathbf{0.52} \tag{10.6}$$