Math 10B Probability Worksheet 2

1. (a) Suppose your friend offers to play the following game with you: she will roll three six-sided dice. If all of the dice show different numbers, she will pay you \$10. How much would you be willing to pay her to be allowed to play this game?

   The sample space here is all sequences of three numbers, each between 1 and 6. So there are 6^3 possible outcomes and (assuming the dice are not loaded) each is equally likely. The number of outcomes in which all three dice show different numbers is $P(6,3) = 6 \cdot 5 \cdot 4$. Thus the probability that you win the \$10 is
   $$\frac{6 \cdot 5 \cdot 4}{6^3} = \frac{5}{9}.$$
   So if you play this game many times, you should expect that $\frac{5}{9}$ of the time you will win \$10. Thus the average amount of money you win if you play many times is $\$\frac{50}{9}$. So it is reasonable to be willing to pay any amount less than that value to be allowed to play the game.

   (b) What if she rolls seven dice instead of three?

   You should not pay any money to play the game. If there are seven 6-sided dice then they cannot all show different numbers—this is an example of the pigeonhole principle.

2. Show that your belief in something should never increase both when some other event occurs and when it doesn't occur. Formally, show that if $P(A \mid B) > P(A)$ then $P(A \mid B^c) < P(A)$.

   Let's assume for a moment that $P(A \mid B^c) \geq P(A)$ and see what that implies. As we saw in class, we can always write
   $$P(A) = P(A \cap B) + P(A \cap B^c).$$
   By definition of conditional probability, we also have that
   $$P(A \cap B) = P(A \mid B)P(B)$$
   $$P(A \cap B^c) = P(A \mid B^c)P(B^c).$$
   Therefore
   $$P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c).$$
   But since we have assumed that $P(A \mid B) > P(A)$ and $P(A \mid B^c) \geq P(A)$, this means that
   $$P(A) > P(A)P(B) + P(A)P(B^c) = P(A)(P(B) + P(B^c)).$$

By the laws of probability, $P(B) + P(B^c) = 1$. Therefore the above equation implies that

$$P(A) > P(A).$$

Since this conlcusion is absurd, our original assumption must have been false. So we can conclude that $P(A \mid B^c) < P(A)$.
By the way, there are many situations in which humans fail to act in a way that is consistent with this fact of probability theory.

3. Suppose there is a test for checking the presence of skin cancer. When cancer is present, the test is positive 90% of the time and negative the other 10%. When cancer is not present, the test is positive 10% of the time, and negative the other 90%. Furthermore, the probability of having cancer is 1%. If someone receives the test and the result is positive, what is the probability that they have cancer? *Hint:* Use Bayes' theorem.

As the hint suggests, we can calculate this probability using Bayes' theorem. Let C be the event that cancer is present and let P be the event that the test is positive. We want to find $P(C \mid P)$ and we are given $P(C)$, $P(P \mid C)$, and $P(P \mid C^c)$. So by Bayes' theorem,

$$\begin{aligned} P(C \mid P) &= \frac{P(P \mid C)P(C)}{P(P)} \\ &= \frac{P(P \mid C)P(C)}{P(P \cap C) + P(P \cap C^c)} \\ &= \frac{P(P \mid C)P(C)}{P(P \mid C)P(C) + P(P \mid C^c)P(C^c)} \\ &= \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.1 \cdot (1 - .01)} \\ &= .0833. \end{aligned}$$

In other words, if the test is positive then the probability that you actually have cancer is about 8.3%.
This result may seem surprising since the test appears to be pretty accurate based on the given information. However, the probability of having cancer starts out very low. This is common in uses of Bayes' theorem and demonstrates a common flaw in human ability to intuitively judge probabilities. Namely, humans have a tendency to overweight new evidence and underweight the importance of the original probability (often referred to as the "prior probability.") By the way, this example helps explain why it is now recommended that people delay getting certain types of cancer screenings until later in life. When you are young, your chances of getting cancer are much lower and so a positive test result is likely to be a false positive and will lead to unnecessary stress and medical fees.

4. Kidney stones is an affliction that comes in two varieties: small stones and large stones.

Suppose that there are two treatments for kidney stones: treatment A and treatment B. Suppose that the success probabilities of these two types of treatment are as shown in the following table.

	Treatment A	Treatment B
Small Stones	93%	87%
Large Stones	73%	68%

Also suppose that a patient with kidney stones is equally likely to have small stones or large stones and that patients with small stones receive treatment A with probability 20% and patients with large stones receive treatment A with probability 80%. All patients who don't receive treatment A receive treatment B.

Given that a patient receives treatment A, what is the chance that it is successful? Given that a patient receives treatment B, what is the chance that it is successful? Which treatment do you think is better?

By the way, this is a real example. The general phenomenon is known as "Simpson's paradox."

There is a lot of information here, so it is helpful to be careful and organized. First let's give names to the relevant events. Let A be the event that the patient receives treatment A, L be the event that the patient has large stones, and S the event that the treatment is successful. So A^c is the event that the patient receives treatment B, L^c is the event that the patient has small stones, and S^c is the event that the treatment is not successful.
We want to find $P(S \mid A)$ and $P(S \mid B)$. By definition of conditional probability,

$$P(S \mid A) = \frac{P(S \cap A)}{P(A)}.$$

To calculate these two probabilities, we will use that facts that

$$\begin{aligned} P(A) &= P(A \cap L) + P(A \cap L^c) \\ &= P(A \mid L)P(L) + P(A \mid L^c)P(L^c) \\ P(S \cap A) &= P(S \cap A \cap L) + P(S \cap A \cap L^c) \\ &= P(S \mid A \cap L)P(A \cap L) + P(S \mid A \cap L^c)P(A \cap L^c) \\ &= P(S \mid A \cap L)P(A \mid L)P(L) + P(S \mid A \cap L^c)P(A \mid L^c)P(L^c). \end{aligned}$$

Using the numbers given above, we have that

$$\begin{aligned} P(A) &= 0.8 \cdot 0.5 + 0.2 \cdot 0.5 = 0.5 \\ P(S \cap A) &= 0.73 \cdot 0.8 \cdot 0.5 + 0.93 \cdot 0.2 \cdot 0.5 = 0.385. \end{aligned}$$

Therefore

$$P(S \mid A) = \frac{P(S \cap A)}{P(A)} = \frac{0.385}{0.5} = 0.77.$$

By similar calculations, we have that

$$P(S \mid A^c) = \frac{P(S \cap A^c)}{P(A^c)} = \frac{0.416}{0.5} = 0.832.$$

If you only see these final answers, it is tempting to believe that treatment B is better. However, when you look at the table of success probabilities above, it seems that treatment A is better for both small stones and large stones! What is going on? The answer is that despite that fact that treatment A is apparently more successful, treatment A is used more frequently for patients with large stones. But both treatments A and B are less likely to succeed on patients with large stones. Here's one way to think about this: treatment A is most often used on patients for which both methods have trouble, whereas treatment B is most often used on patients for which both methods work well, creating the illusion that treatment B is better overall. The lesson here is not that treatment A is better—although based on the evidence given above, that seems likely—but rather that you should be careful when reasoning about probabilities: they don't always indicate what you intuitively think they should.