Homework 3

By signing my name, I agree to abide by the Stern Code of Conduct ______________________

Question 1

a. Suppose A and B are disjoint. What is $P(A|B)$? Are A and B independent?

$P(A|B) = P(A \cap B)/P(B) = 0/P(B) = 0$.

They are *not independent*, because knowing one event happens tells us the other event can't happen.

The law of total probability says that...

If E_1 and E_2 are disjoint and $E_1 \cup E_2 = S$ is the whole sample space, then for any event A, $P(A) = P(A \cap E_1) + P(A \cap E_2) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2)$. This is useful when it's hard to calculate $P(A)$ by the counting rule but easy to calculate $P(A|E_1)P(E_1)$ and so on. Use this to solve the following problem:

b. Suppose that half of the people on campus get the flu shot (vaccine). Write $V = 1$ if someone has the vaccine and $V = 0$ otherwise. Let F be the event that a person gets the flu, and suppose $P(F|V = 0) = 3/10$ and $P(F|V = 1) = 1/10$. Use the law of total probability to find the probability that a person on campus gets the flu. (Hint: this person either got the flu shot or didn't get the flu shot...)

$P(F) = P(F|V = 1)P(V = 1) + P(F|V = 0)P(V = 0) = (1/10)(1/2) + (3/10)(1/2) = 1/5$

c. Continuing the previous problem, what percent of people who got the flu were vaccinated?

Use Bayes' rule!

$P(V = 1|F) = P(F|V = 1)P(V = 1)/P(F) = (1/10)(1/2)/(1/5) = 1/4$ or 25%

d. Suppose I claim I got the vaccine but ended up getting the flu anyway. Do you find this surprising?

It's not very surprising. It happens to about 25% of the people who get the vaccine.

e. Which of the quantities above would be different if, due to an improved vaccine, $P(F|V = 1)$ became much smaller? And for each of these, would they become larger or smaller?

Consider the extreme case when $P(F|V = 1) = 0$.

Looking at the law of total probability in part (b), we see that $P(F|V = 1)$ is one of the terms, so $P(F)$ would get smaller, but would still be bigger than 0 (due to people who didn't get the vaccine).

In Bayes' rule, $P(F|V = 1)$ is in the numerator and becomes zero, but the denominator stays above zero, so the whole fraction becomes zero. Hence, $P(V = 1|F)$ would also get smaller.

This makes sense: if the vaccine becomes more effective, (1) fewer people get the flu, and (2) if you know someone got the flu, you can be more certain they didn't have the vaccine.

Question 2

- **Factorials**: $n!$ means "multiply all the numbers between 1 and n," for example $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. Note that by default $0! = 1$.
- **A cool trick with factorials**: $n! = n \cdot (n-1)! = n \cdot (n-1) \cdot (n-2)!$, etc.
- **Permutations**: the number of ways to rank (or arrange) n objects in order, where the order matters, is $n!$. (There are n possible items to go 1st, then there are $n-1$ remaining to go in second place, then there are $n-2$ for third, and so on.)
- **Scientific notation in `R`**: A number like 4.015e+11 means, roughly, 4 followed by 11 zeros.

a. A deck of cards has 52 cards. Use the command `factorial(52)` in `R` to find the number of ways a deck can be ordered (permutations). The answer is roughly an 8 followed by how many zeros?

```
factorial(52)
```

```
## [1] 8.065818e+67
```

Roughly equals 8 followed **67 zeros**.

b. To pick k items out of n, and arrange them in a specific order, there are $n!/(n-k)!$ wasy to do this. For example, there are $n!/(n-1)! = n(n-1)!/(n-1)! = n$ ways to pick 1 item. How many ways are there to pick 3 items? (Simplify as much as you can)

We use the trick $n! = n(n-1)(n-2)(n-3)!$ to simplify the expression:

$$n!/(n-3)! = n(n-1)(n-2)[(n-3)!/(n-3)!] = n(n-1)(n-2)$$

c. To pick k items out of n, *ignoring the order* (e.g., picking item1 and then item2 is the same as picking item2 and then item1), you first count the number of ways to pick k out of n where the order does matter as in part (b), and then you divide by the number of ways to permute k items. The final answer is $n!/((n-k)!k!)$. We have another name and notation for this: it's called the *Binomial coefficient* and denoted as $\binom{n}{k}$. In `R`, you can compute this number using the function `choose(n, k)` or the formula `factorial(n)/(factorial(n-k)*factorial(k))`. How many ways are there to pick a subset of 5 items out of 10?

```
choose(10,5)
```

```
## [1] 252
```

b. Suppose I toss a coin 10 times and am interested in the number X of times it comes up heads. Each individual sequence of 10 coin tosses has probability $(1/2)^{10}$. How many of these sequences have 5 heads? (Hint: this is the same as picking which 5 out of the 10 tosses are heads).

By the hint, this is just the same number as the previous problem: 252

c. To find $P(X=5)$ we can add up the probability of each individual outcome $(1/2)^{10}$ for each of the outcomes that have 5 out of 10 coming up heads. What is this probability?

$$\binom{10}{5}(1/2)^{10}$$

which equals:

```r
choose(10,5)*(1/2)^(10)
```

```
## [1] 0.2460938
```

d. What's the probability that the number of heads is either 4 or 6? (Hint: the event that it equals 4 is *disjoint*, or mutually exclusive, from the event that it equals 6).

$$\binom{10}{4}(1/2)^{10} + \binom{10}{6}(1/2)^{10}$$

which equals:

```r
choose(10,4)*(1/2)^(10) + choose(10,6)*(1/2)^(10)
```

```
## [1] 0.4101562
```

e. Suppose that instead of fair coin tosses, we have coin tosses with probability p of being heads. Each coin toss has a Ber(p) distribution, and the number of heads is distributed as Bin(n, p). Now, to compute probabilities, we still need to count which of the tosses are heads, but each individual outcome no longer has probability $(1/2)^n$. Instead, to get k heads and $n-k$ tails, independently, each outcome with k heads now has probability $p^k(1-p)^{n-k}$ (there are k copies of p, one for each success, and $n-k$ copies of (1-p), one for each failure). What's the probability of 5 heads out of 10 when the coin has probability $p = 0.2$ of heads?

$$\binom{10}{5}(0.2)^5(1-0.2)^{10-5}$$

which equals:

```r
choose(10,5)*(0.2)^5*(0.8)^5
```

```
## [1] 0.02642412
```

f. During a game of Dungeons and Dragons, I roll a 20-sided dice 5 times. What is the probability that it lands on 20 exactly one of those 5 times?

This is a $\mathrm{Bin}(5, 1/20)$ random variable, so the probability is

$$\binom{5}{1}(1/20)^1(1-1/20)^{5-1}$$

which equals:

```r
choose(5,1)*(1/20)*(19/20)^4
```

```
## [1] 0.2036266
```

g. Can you think of any real world examples, aside from coin tosses, where the numbers might have a Binomial distribution? Some hints are in the terms used for Binomial: there are n independent "trials" (this just means event) and each one has probability p of being a "success" (note that we could switch "success" with "failure" and it would be a Bin($n, 1-p$), so success is just an arbitrary label), and we are interested in the distribution of the number of successes. Examples could come from sports, business, science, or whatever you can imagine. Try to come up with two examples and briefly describe them below.

Any kind of example where there are a pre-determined number of success/failure or yes/no type events, and the probability of success is the same for each one, and they are independent (the outcome from one doesn't affect the probability of success for any of the others).

- **Sports**: We can think of a team as having a probability p of winning each game, so the number of wins in a season with a fixed number of games could be Binomial. (For sports where there are playoffs and some teams get to play more games if they keep winning then the assumption of a pre-determined number of trials would be wrong and Binomial might be a bad model)
- **Medicine**: Suppose there is an illness which people sometimes recover from without any treatment, perhaps because of lifestyle changes or other factors. Say p is the proportion of people with this illness who spontaneously recover during a period of a year. If 100 people with the illness are recruited into a year long drug trial, and 50 of them are assigned to the *control group*, then the number of people in the control group who recover during the study might be modeled as Bin(50,p). (This is not taking into account any possible placebo effect)
- **Quality control**: Imagine a manufacturing process that produces a certain number of items every day. Each item has some probability (hopefully small) of being defective. Then the number of defective items produced in a given day might have a Binomial distribution.