Homework 2 solution

Question 1

a.

Consider a game involving tossing a coin three times, and use set notation (e.g. { item1, item2, ... }) to answer the following: What is the sample space S? What is the event E_1 = "the first toss is H"? What is the intersection of E_1 and E_2 = "the second toss is H"?

(Remember that the order of items in a set doesn't matter).

$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

$E_1 = \{HHH, HHT, HTH, HTT\}$,

$E_2 = \{HHH, HHT, THH, THT\}$,

and $E_1 \cap E_2 = \{HHH, HHT\}$

b.

A 6-sided die is rolled 5 times. What is the probability it comes up six every time? What is the probability of no sixes? What is the probability every roll is less or equal to 5?

Using independence and the multiplication rule:

$P(\text{all } 6) = (1/6)^5 \approx 0.00013$,

$P(\text{no } 6) = (5/6)^5 \approx 0.402$,

$\{\text{all } \leq 5\} = \{\text{no } 6\}$ so same answer as previous.

c.

Every time I roll a 20-sided die there is a 1/20 chance it lands on 20. So if I roll it 10 times, the probability that at least one of these will be a 20 is $10/20 = 1/2$. True or false? Explain your answer.

False. By the complement rule,

$P(\text{at least one } 20) = 1 - P(\text{no } 20) = 1 - (19/20)^{10} \approx 0.401$

d.

What is the equation for the conditional probability of E_2 given E_1? Suppose $P(E_1) > P(E_2)$ and $P(E_1 \cap E_2) > 0$. Which is larger, $P(E_2|E_1)$ or $P(E_1|E_2)$, or is there not enough information to tell? Explain, and also draw a Venn diagram as an example.

$P(E_2|E_1) = P(E_2 \cap E_1)/P(E_1)$

The two conditional probabilities both have $P(E_1 \cap E_2)$ in the numerator, since $E_2 \cap E_1$ is the same as $E_1 \cap E_2$. So the one with a larger denominator will be smaller overall. Since $P(E_1) > P(E_2)$, the one with the larger denominator is $P(E_2|E_1)$, so we conclude

$P(E_2|E_1) < P(E_1|E_2)$

The Venn diagram should show the events have an overlap because $P(E_1 \cap E_2) > 0$, and the overlap should be a *smaller portion* of E_1 than it is compared to E_2.

e.

The mortality rates from a hypothetical drug trial are in the table below. Is survival independent from treatment/placebo, adherence/non-adherence, both of these, or neither? Explain.

	Treatment	Placebo
Adherers	15%	15%
Non-adherers	25%	25%
Total	20%	20%

$P(\text{survival}|\text{treatment}) = P(\text{survival}|\text{placebo}) = 0.20$ so survival is independent of treatment/placebo.

$P(\text{survival}|\text{adherence}) = 0.15 \neq 0.20 = P(\text{survival})$ so survival is **not** independent of adherence/non-adherence.

Question 2

a.

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

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

They are not independent: if we know that A occurs then we know that B does not occur.

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). Let V be the event that a person has the vaccine, and F be the event that a person gets the flu, and suppose $P(F|V^c) = 3/10$ and $P(F|V) = 1/10$. 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...)

$P(F) = P(F|V)P(V) + P(F|V^c)P(V^c) = (1/10)(1/2) + (3/10)(1/2) = 2/10$

c.

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

Using Bayes' rule:

$P(V|F) = P(F|V)P(V)/P(F) = (1/10)(1/2)/(2/10) = 0.25$

d.

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

From part (a) we see that $P(F)$ would decrease.

In part (c), the numerator would decrease. The denominator $P(F)$ would also decrease, but not as much, so the overall $P(V|F)$ would also decrease.

Question 3

- **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 permute (i.e. 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?

67

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 2 items out of n? What about 3? Simplify each until the answer is not a fraction.

For 2 items: $n!/(n-2)! = n(n-1)(n-2)!/(n-2)! = n(n-1)$

For 3 items: $n!/(n-3)! = n(n-1)(n-2)(n-3)!/(n-3)! = n(n-1)(n-2)$

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?

$\binom{10}{5} = \frac{10!}{5!(10-5)!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5 \cdot 4 \cdot 3 \cdot 2} = 252$

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

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

d.

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?

Same answer as part (c) above.

e.

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

$\binom{10}{5}(1/2)^{10} \approx 0.246$

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

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

f.

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).

Using the addition rule for disjoint events, this is

$\binom{10}{4}(1/2)^{10} + \binom{10}{6}(1/2)^{10} \approx 0.41$

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

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

g.

Suppose that instead of fair coin tosses, we have coin tosses with probability p of being heads. Each coin toss has a $\text{Ber}(p)$ distribution, and the number of heads is distributed as $\text{Bin}(n, p)$. 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?

There are still $\binom{10}{5}$ different outcomes with 5 heads, but now each them of them has probability $(0.2)^5(0.8)^{10-5}$, so the answer is

$\binom{10}{5}(0.2)^5(0.8)^5 \approx 0.0264$

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

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

h.

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 2 out of those 5 times?

The success probability is $p = 1/20$, and there are $\binom{5}{2}$ different possible outcomes with 2 successes, so the probability is

$\binom{5}{2}(1/20)^2(19/20)^3 \approx 0.0214$

```
choose(5,2) * (1/20)^2 * (19/20)^3
```

```
## [1] 0.02143438
```

i.

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 $\text{Bin}(n, 1-p)$, 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.

- If a basketball player gets to take a certain number of free throws, the number of baskets they make could be considered a binomial random variable. (Note: different players might have different success probabilities)
- If a student takes a multiple choice exam without studing, and picks answers randomly, the number of questions they get correct would be Binomial.
- 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.