Math 224 **Spring 2022**
Homework 2 **Drew Armstrong**

1. Tricks With the Binomial Theorem. For any real number x and integer $n \geq 0$, the Binomial Theorem tells us that

$$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n.$$

Use this to prove the following identities:

(a) $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$

(b) $1\binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \cdots + n\binom{n}{n} = n2^{n-1}$

(c) $\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots$

[Hint: Differentiate the polynomial and/or substitute various values of x.]

(a): Substitute $x = 1$ to get

$$\begin{aligned}
(1+1)^n &= \binom{n}{0} + \binom{n}{1}1 + \binom{n}{2}1^2 + \cdots + \binom{n}{n}1^n \\
2^n &= \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n}.
\end{aligned}$$

(b): Differentiate with respect to x and then substitute $x = 1$ to get

$$\begin{aligned}
n(1+x)^{n-1} &= 0 + \binom{n}{1} + \binom{n}{2}(2x) + \cdots + \binom{n}{n}(nx^{n-1}) \\
n(1+1)^{n-1} &= 0 + \binom{n}{1} + \binom{n}{2}(2 \cdot 1) + \cdots + \binom{n}{n}(n \cdot 1^{n-1}) \\
n2^{n-1} &= 0 + \binom{n}{1} + 2\binom{n}{2} + \cdots + n\binom{n}{n}.
\end{aligned}$$

(c): Substitute $x = -1$ and then rearrange to get

$$\begin{aligned}
(1+(-1))^n &= \binom{n}{0} + \binom{n}{1}(-1) + \binom{n}{2}(-1)^2 + \cdots + \binom{n}{n}(-1)^n \\
0 &= \binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots \\
\binom{n}{1} + \binom{n}{3} + \cdots &= \binom{n}{0} + \binom{n}{2} + \cdots.
\end{aligned}$$

Remark: There are other ways to prove the identities in (a) and (b).

- To prove part (a) we recall that $\binom{n}{k}$ is the number of such words with k copies of H and $n-k$ copies of T. But the total number of words made from the letters $\{H, T\}$ is 2^n, hence 2^n is the sum of $\binom{n}{k}$ over all possible values of k. Another interpretation is that 2^n is the total number of subsets of a set of size n, while $\binom{n}{k}$ is the number of subsets of size k from a set of size n.

- It is slightly trickier to give a counting proof of (b). One way to do it is to count the number of committees with presidents that can be formed from a set of n people. On the one hand, we could first choose the president in n ways, then choose the remaining committee members from the remaining $n-1$ people in in 2^{n-1} ways, for a total of $n \times 2^{n-1}$ choices. On the other hand, we could first choose a commiittee of k people in $\binom{n}{k}$ ways, then choose one of these k people to be the president, for a total of $k \times \binom{n}{k}$ choices.
- It is **quite tricky** to give a counting proof of (c), so I won't. But here is an interesting application: Let X be the number of heads that occur in n flips of a **fair** coin. Then

$$\begin{aligned}
P(X \text{ is even}) &= P(X=0) + P(X=2) + \cdots \\
&= \binom{n}{0}/2^n + \binom{n}{2}/2^n + \cdots \\
&= \frac{1}{2^n}\left[\binom{n}{0} + \binom{n}{2} + \cdots\right] \\
&= \frac{1}{2^n}\left[\binom{n}{1} + \binom{n}{3} + \cdots\right] \\
&= \binom{n}{1}/2^n + \binom{n}{3}/2^n + \cdots \\
&= P(X=1) + P(X=3) + \cdots \\
&= P(X \text{ is odd}).
\end{aligned}$$

  This is not true for biased coins.

2. Counting License Plates. Find the total number of possible license plates with the following properties:

(a) Four letters followed by two digits.
(b) The same as (a), but letters and digits cannot be repeated.
(c) Four letters and two digits in any order. Repetition is okay. [Hint: First choose the positions for the letters and digits. Then count how many ways to fill them in.]

Remark: We use an alphabet with 26 letters.

(a): Since the characters are ordered we use the multiplication principle. Since characters are allowed to be repeated, the number of possibilities is

$$\underbrace{26}_{\text{1st letter}} \times \underbrace{26}_{\text{2nd letter}} \times \underbrace{26}_{\text{3rd letter}} \times \underbrace{26}_{\text{4th letter}} \times \underbrace{10}_{\text{1st digit}} \times \underbrace{10}_{\text{2nd digit}} = 45,697,600.$$

(b): If characters cannot be repeated then the number of possibilities is

$$\underbrace{26}_{\text{1st letter}} \times \underbrace{25}_{\text{2nd letter}} \times \underbrace{24}_{\text{3rd letter}} \times \underbrace{23}_{\text{4th letter}} \times \underbrace{10}_{\text{1st digit}} \times \underbrace{9}_{\text{2nd digit}} = 32,292,000.$$

(c): First we choose where to puts the letters and digits. Label the positions from left to right as $\{1,2,3,4,5,6\}$. Then choose two of these positions to place the digits. There are $\binom{6}{2} = 15$ ways to do this. For example, suppose we place the digits in positions 2 and 5. Then the number of ways to fill in the letters and digits is the same as in part (a):

$$\underbrace{26}_{\text{1st letter}} \times \underbrace{10}_{\text{1st digit}} \times \underbrace{26}_{\text{2nd letter}} \times \underbrace{26}_{\text{34d letter}} \times \underbrace{10}_{\text{2nd digit}} \times \underbrace{26}_{\text{4th letter}} = 45,697,600.$$

We get the same total for every position of the digits, hence the total is 15 times this:

$$\binom{6}{2} \cdot 26^4 \cdot 10^2 = 15 \times 45{,}697{,}600 = 685{,}464{,}000.$$

Remark: This problem is based on Florida license plates, which have 4 letters and 2 digits. The position of the digits changes over time. Since 2009 the 2 digits appear on the right as in part (a).

3. Multinomial Probability. A fair eight-sided die has 2 sides painted red, 3 sides painted green and 3 sides painted blue. Suppose you roll the die 5 times and let R, G, B be the number of times you get red, blue and green, respectively. Compute the following probabilities:

(a) $P(R = 2, G = 1, B = 2)$
(b) $P(R \geq 1)$ [Hint: Think of the die as a coin with "heads"="you get red".]
(c) $P(R = G)$ [Hint: If $R = G$ then what are the possible values of R, G, B?]

(a): We can interpret this experiment as a "strange 3-sided die" with sides painted {red, green, blue} and with probabilities $P(\text{red}) = 2/8)$, $P(\text{green}) = 3/8$, $P(\text{blue}) = 3/8$. Then using the formula for multinomial probability gives

$$\begin{aligned} P(R = 2, G = 1, B = 2) &= \frac{5!}{2!1!2!} P(\text{red})^2 P(\text{green})^1 P(\text{blue})^2 \\ &= 30 \left(\frac{2}{8}\right)^2 \left(\frac{3}{8}\right)^1 \left(\frac{3}{8}\right)^2 \\ &= \frac{30 \cdot 2^2 \cdot 3^1 \cdot 3^2}{8^5} \quad \text{or } 9.89\% \end{aligned}$$

Remark: The formula $5!/(2!1!2!) = 30$ counts the words of length 5 made from the symbols r, r, g, b, b. We can think of each such word as an outcome of the experiment. The probability of each of these 30 outcomes is the same: $(2/3)^2(3/8)^1(3/8)^2 \approx 0.30\%$.

(b): There is a short way to do this and a long way. The short way views each die roll as a "strange coin" where "heads"="we get red" and "tails"="we don't get red", so that $P(\text{heads}) = 2/8 = 1/4$ and $P(\text{tails}) = 6/8 = 3/4$. Then the formula for binomial probability gives

$$\begin{aligned} P(R \geq 1) &= 1 - P(R = 0) \\ &= 1 - P(\text{all tails}) \\ &= 1 - P(\text{tails})^5 \\ &= 1 - (3/4)^5 \quad \text{or } 76.27\% \end{aligned}$$

The long way sums over all the possible values of R, G and B corresponding to $R = 0$:[1]

R	G	B
0	0	5
0	1	4
0	2	3
0	3	2
0	4	1
0	5	0

[1] It would take even **longer** to sum over all possible values of R, G and B corresponding to $R \geq 1$.

Thus we have

$$\begin{aligned} P(R=0) = \quad & P(R=0,G=0,B=5) \\ &+P(R=0,G=1,B=4) \\ &+P(R=0,G=2,B=3) \\ &+P(R=0,G=3,B=2) \\ &+P(R=0,G=4,B=1) \\ &+P(R=0,G=5,B=0). \end{aligned}$$

We can use the formula for multivariable probability to compute each of these six numbers and then we add them up.

(c): This time there is no shortcut. Our only option is to sum the probabilities over all possible values of R, G and B corresponding to $R = G$:

R	G	B
0	0	5
1	1	3
2	2	1

Thus we have

$$\begin{aligned} &P(R=G) \\ &= P(R=0,G=0,B=5)+P(R=1,G=1,B=3)+P(R=2,G=2,B=1) \\ &= \frac{5!}{0!0!5!}\left(\frac{2}{8}\right)^0\left(\frac{3}{8}\right)^0\left(\frac{3}{8}\right)^5+\frac{5!}{1!1!3!}\left(\frac{2}{8}\right)^1\left(\frac{3}{8}\right)^1\left(\frac{3}{8}\right)^3+\frac{5!}{2!2!1!}\left(\frac{2}{8}\right)^2\left(\frac{3}{8}\right)^2\left(\frac{3}{8}\right)^1 \\ &= 1\cdot\frac{3^5}{8^5}+20\cdot\frac{2^1 3^1 3^3}{8^5}+30\cdot\frac{2^2 3^2 3^1}{8^5} \quad \text{or } 20.52\% \end{aligned}$$

Remark: All of the probabilities add to 1 because of the *trinomial theorem*:

$$1=\left(\frac{2}{8}+\frac{3}{8}+\frac{3}{8}\right)^5=\sum_{i+j+k=5}\frac{5!}{i!j!k!}\left(\frac{2}{8}\right)^i\left(\frac{3}{8}\right)^j\left(\frac{3}{8}\right)^k.$$

4. Hypergeometric Probability. A deck of cards contains 26 red cards and 26 black cards. Choose 5 cards at random and let R be the number of red cards that you get.

(a) What is the size of the sample space?
(b) How many outcomes are there with $R = k$? [Hint: Your answer will involve k.]
(c) Use parts (a) and (b) to compute the probabilities $P(R = k)$ for $k = 0, 1, 2, 3, 4, 5$. [Hint: Use a calculator.]

(a): This problem is easier to analyze if we think of the cards as unordered. In fact, we can think of them as balls in an urn. Consider an urn containing 26 red balls and 26 black balls. Reach in and grab 5 balls at random and let R be the number of red balls you get. The sample space is

$$S = \text{all possible ways to choose 5 balls from the 52 balls in the urn.}$$

Since the choices are unordered[2] we have

$$\#S = \binom{52}{5} = \frac{52!}{5!47!} = \frac{52\cdot 51\cdot 50\cdot 49\cdot 48}{5\cdot 4\cdot 3\cdot 2\cdot 1} = 2,598,960.$$

(b): Let $E \subseteq S$ be the event defined as follows:

$$E = \text{all possible ways to choose } k \text{ red balls and } 5-k \text{ black balls.}$$

By the multiplication principle we can choose the red and black balls separately and then multiply the possibilities:[3]

$$\begin{aligned}\#E &= \#(\text{ways to choose the red balls}) \cdot \#(\text{ways to choose the black balls}) \\ &= \binom{26}{k}\binom{26}{5-k}.\end{aligned}$$

(c): Since all of the outcomes are equally likely we have

$$P(R=k) = \frac{\#E}{\#S} = \frac{\binom{26}{k}\binom{26}{5-k}}{\binom{52}{5}}.$$

I used my computer to find the probabilities:

k	0	1	2	3	4	5
$P(R=k)$	2.53%	14.96%	32.51%	32.51%	14.96%	2.53%

5. The Birthday Problem. We choose n people at random and record their birthdays as a sequence of n numbers from the set $\{1, 2, \ldots, 365\}$. (We ignore leap years.) Assume that each birthday is equally likely.

(a) What is the size of the sample space?
(b) Compute the probability that no two people have the same birthday. [Hint: Count the sequences with no repeated numbers. Your answer will involve n.]
(c) Let $f(n)$ be the probability that in a random group of n people there exist two people with the same birthday. Note that $f(1) = 0\%$ and $f(366) = 100\%$. Find the smallest n such that $f(n) > 50\%$. [Hint: Use part (b) to find a formula for $f(n)$. Then use a calculator to plug in various values of n. The answer is surprisingly small.]

Remark: Equivalently, suppose that we roll a fair 365-sided die n times.

(a): Each outcome is an ordered list of numbers from the set $\{1, 2, \ldots, n\}$, so that

$$\#S = \underbrace{365}_{\substack{\text{1st student's}\\ \text{birthday}}} \times \underbrace{365}_{\substack{\text{2nd student's}\\ \text{birthday}}} \times \cdots \times \underbrace{365}_{\substack{n\text{th student's}\\ \text{birthday}}} = 365^n.$$

[2] We are free to order the balls if we want so. In this case we would have $\#S = 52\cdot 51\cdot 50\cdot 49\cdot 48$. But then part (b) is much more difficult.

[3] This is for unordered choices. The number of ordered choices is $\binom{26}{k}\binom{26}{5-k}\binom{5}{k}k!(5-k)!$. (First choose the unordered red and black balls. Then choose where the red balls will come in the ordering. Then place the balls in order.) I don't recommend doing it this way.

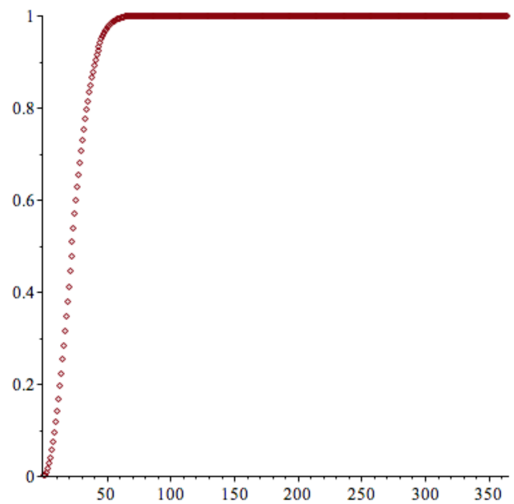
(b): Let E be the set of outcomes where all the birthdays are different. Each of these corresponds to a list of length n using numbers of the set $\{1, 2, \ldots, 365\}$ where no number is repeated. Hence

$$\#E = \underbrace{365}_{\substack{\text{1st student's} \\ \text{birthday}}} \times \underbrace{364}_{\substack{\text{2nd student's} \\ \text{birthday}}} \times \cdots \times \underbrace{(365 - n + 1)}_{\substack{n\text{th student's} \\ \text{birthday}}} = \frac{365!}{(365 - n)!}.$$

(c): Let $f(n)$ be the probability that some pair of people have the same birthday. Then

$$f(n) = 1 - P(\text{all different birthdays}) = 1 - \frac{\#E}{\#S} = 1 - \frac{365!/(365 - n)!}{365^n}.$$

This is a hard function to analyze because it is only defined for whole numbers $1 \leq n \leq 365$. A typical graphing calculator can't graph it but a more powerful computer algebra system can.[4] My computer made the following plot of the points $(n, f(n))$ for $n \in \{1, \ldots, 365\}$:



Note that it goes up really fast. For $n \geq 50$ the chance is almost 100% that a group of n people contains two with the same birthday. From the picture it seems $f(n)$ crosses the 50% threshold around $n = 25$. To be precise, I used my computer to show that $n = 23$ is the smallest value with $f(n) > 50\%$:

$$f(22) = 47.57\%, f(23) = 50.73\%.$$

Do you find the number 23 surprisingly small? That's why this problem is sometimes also called the **birthday paradox**. In our class of $n = 33$ students there is a $f(33) = 77.5\%$ chance that two of you have the same birthday.

6. Bayes' Theorem. A random person is tested for a certain disease. Consider the events

$$D = \text{the person has the disease,}$$
$$T = \text{the test returns positive.}$$

Assume that the disease prevalence is $P(D) = 1\%$. Assume that the test has false positive rate $P(T|D') = 0.1\%$ and false negative rate $P(T'|D) = 5\%$.

(a) Find the true negative rate $P(T'|D')$ and the true positive rate $P(T|D)$. [Hint: Easy.]

[4] I use Maple because I grew up in Canada. University of Miami students have free access to Mathematica which does all the same things. WolframAlpha.com is a cheap and slow version of Mathematica.

(b) Find the probability $P(T)$ that the test returns positive. [Hint: Use the definition of conditional probability, $P(A \cap B) = P(A)P(B|A)$, and the Law of Total Probability, which tells us that $P(T) = P(D \cap T) + P(D' \cap T)$.]

(c) Compute the backwards probabilities $P(D|T)$ and $P(D'|T')$. [Hint: The definition of conditional probability says that $P(D|T) = P(D \cap T)/P(T)$.]

(a): We use the formula $P(A|B) + P(A'|B) = 1$ to obtain

$$\begin{aligned} P(T'|D') &= 1 - P(T|D') = 100\% - 0.1\% = 99.9\%, \\ P(T|D) &= 1 - P(T'|D) = 100\% - 5\% = 95\%. \end{aligned}$$

(b): We can express $P(T)$ in terms of probabilities that we know:

$$\begin{aligned} P(T) &= P(D \cap T) + P(D' \cap T) \\ &= P(D)P(T|D) + P(D')P(T|D') \\ &= (1\%)(95\%) + (99\%)(0.1\%) = 1.049\%. \end{aligned}$$

(c): We compute $P(D|T)$ as follows:

$$\begin{aligned} P(D|T) &= P(D \cap T)/P(T) \\ &= P(D)P(T|D)/P(T) = (1\%)(95\%)/(1.049\%) = 90.56\%. \end{aligned}$$

And we compute $P(D'|T')$ as follows:

$$\begin{aligned} P(D'|T') &= P(D' \cap T')/P(T') \\ &= P(D')P(T'|D')/P(T') = (99\%)(99.9\%)/(100\% - 1.049\%) = 99.95\%. \end{aligned}$$

Remark: If the test returns negative then the patient almost certainly does not have the disease. However, if the test returns positive then there is only a 90.56% chance that the person has the disease. Why so low? The general formula is

$$P(D|T) = \frac{P(D)P(T|D)}{P(D)P(T|D) + P(D')P(T|D')}.$$

The only way to make $P(D|T)$ larger is to increase the rate of true positives $P(T|D)$ from 95% or decrease the rate of false positives $P(T|D')$ from 0.1%.[5] But note that increasing $P(T|D)$ from 95% to 100% only increases $P(D|T)$ from 90.56% to 90.99%. Thus our only option is to decrease $P(T|D')$ from 0.1% to an even smaller number. This illustrates the difficulty of testing for a rare disease.

[5]Increasing $P(D)$ would also increase $P(D|T)$ but we don't to do that!