



WEEK #3: PROBABILITY

Problem 1. Probability is a field of mathematics that allows us to think about random events and quantify how often particular outcomes will occur. There are two main ways we can think about probability. The first is known as the frequentist approach. This approach looks at probability as a long term proportion: if we repeated this trial/experiment an infinite number of times, how often would a particular event occur. Let's begin by thinking of an unfair coin. This coin is weighted such that heads will show up 3 times more often than tails.

- a. If heads will show up 3 times more often than tails, what proportion of the time do you expect to get heads? What proportion of the time do you expect to get tails.
- b. What will the proportion of tails be after one flip? What are the possible values?
- c. What happens to the proportion of tails as we flip the coin more and more times?

Problem 2. Estimating probabilities by looking at long term proportions is a wonderful idea in many cases, but it is not always feasible. Sometimes, we have to use what is called a subjective or personal probability. For instance, if you wanted to know what is the probability of you getting an A in a particular class, you wouldn't be able to look at the long term proportion to estimate this. Can you think of another example where you would need to use a subjective probability?

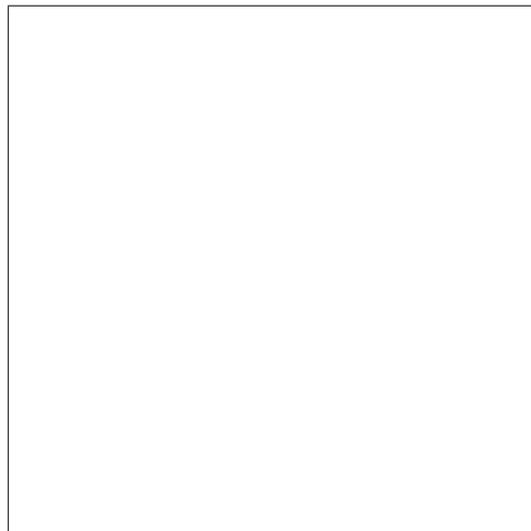
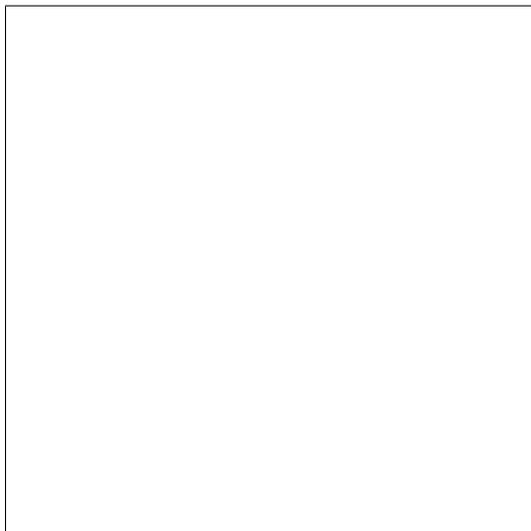
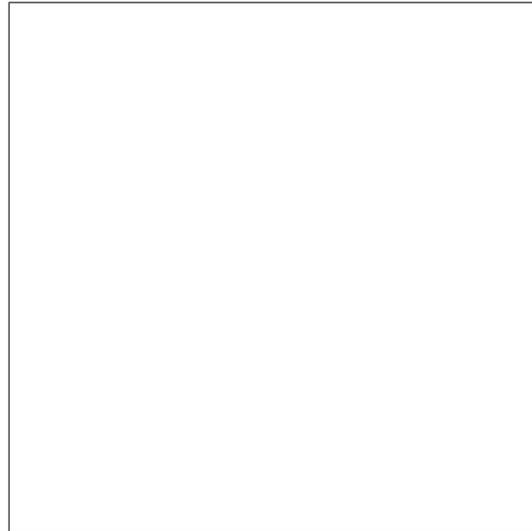
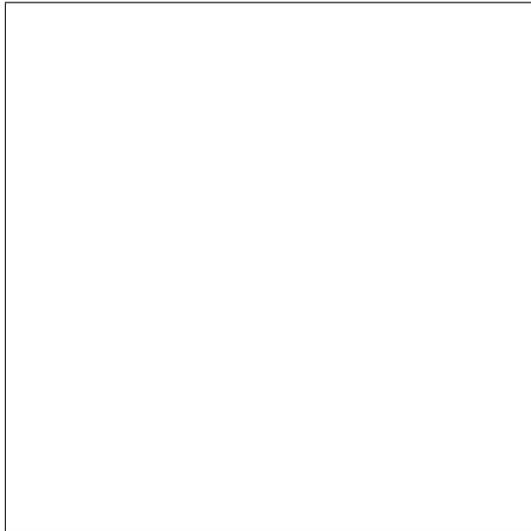


Problem 3. You decide to flip a fair coin many times and calculate the proportion of heads.

- a. On the first ten coin flips, 9 of them are heads and 1 is tails. What is the proportion of heads on the first 10 flips
- b. You flip the coin another 10 times-half are heads and half are tails. What is the proportion of heads on the first 20 flips?
- c. You flip the coin another 100 times-half are heads and half are tails. What is the proportion of heads on the first 120 flips?
- d. You flip the coin another 1000 times-half are heads and half are tails. What is the proportion of heads on the first 1120 flips?
- e. You flip the coin another 10,000 times-half are heads and half are tails. What is the proportion of heads on the first 11,120 flips?
- f. What does this tell us about long term probabilities?



Problem 4. Use the space below to illustrate the following four concepts with Venn Diagrams: Disjoint Events, Complementary Events, Union of Two Events, Intersection of Two Events.





Problem 5. Many times when we are looking at probability problems, we are not asking about just one event, but rather a combination of events. Let's look at an example where we are looking at two different events: A student is taking a history class that has one midterm and one final exam. They believe that the chance of them getting an A on the midterm is 50%. They also believe the chance of them getting an A on the final exam is 50%. The student states that their chance of getting an A on at least one of the two tests is 100%. Is their reasoning correct? Why or why not? If it is incorrect, how would you solve for the probability of them getting an A on at least one of the two tests?

Problem 6. Let's look at another example of a probability question involving two events. A software company provides an email filtering service to protect email users from spam. The company has advertised that their software is 95% accurate. This could mean one of four things. Listed below are the four things this could mean. Write out each as a conditional probability statement.

- a. 95% of the blocked emails are spam
- b. 95% of spam emails are blocked
- c. 95% of the valid emails are allowed through
- d. 95% of the emails allowed through are valid



Problem 7. While the most common die has six sides, there are a number of other varieties of die that are used in various games. One such die is the four-sided die. Assume we have two fair four-sided die, each labeled with the numbers from one to four.

- a. What do we mean by a fair die?
- b. What is the sample space for rolling one die?
- c. What is the sample space for rolling two die?
- d. Assume you roll both die and add them together. What is the sample space?
- e. Assume you roll both die and add them together. What is the probability the sum is greater than 6? It may be a good idea to use the sample space from part c to solve this.



Problem 8. Many times, we organize our information about a random event by creating a table or diagram that describes the probability distribution. The table below shows the probability distribution of Educational Attainment for Americans 25 years and older in 2019, as reported by the Census Bureau.

x	None	No HS Diploma	HS Diploma	Some College	Assoc.	Bach	Bach+
$P(x)$	0.003	0.096	0.281	0.157	0.103	0.225	0.135

- What type of variable is educational attainment?
- What two properties must be true for a probability distribution to be valid? Is this a valid probability distribution?
- What is the probability that a randomly selected person does not have a high school diploma?
- What is the probability that a randomly selected person has some college degree?



Problem 9. Let's continue looking at the 2019 Educational Attainment data, however, now let's look at it split up by gender.

	None	No HS Diploma	HS Diploma	Some College	Assoc.	Bach	Bach+	Total
Males	0.002	0.048	0.141	0.075	0.045	0.107	0.063	0.482
Females	0.002	0.047	0.140	0.082	0.058	0.118	0.072	0.518
Total	0.004	0.095	0.281	0.157	0.103	0.225	0.135	1

- Are having an associates degree and being female mutually exclusive?
- What is the probability that a randomly selected person has an associates degree or is a female?
- What is the probability that a randomly selected person has an associates degree and is a female?
- What is the probability that a randomly selected female has an associates degree?
- What is the probability that a randomly selected person has an associates degree if they are a male?
- Does it appear as though gender and educational attainment are dependent or independent? Why?



Problem 10. The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey designed to identify risk factors in the adult population and report emerging health trends. The following tables displays the distribution of health status of respondents to this survey (excellent, very good, good, fair, poor) as well as their health coverage (whether or not they have health insurance).

	Excellent	Very Good	Good	Fair	Poor	Total
No	0.0230	0.0364	0.0427	0.0192	0.0050	0.1262
Yes	0.2099	0.3123	0.2410	0.0817	0.0289	0.8738
Total	0.2329	0.3486	0.2838	0.1009	0.0338	1.0000

- Are being in excellent health and having health coverage mutually exclusive?
- What is the probability that a randomly chosen individual has excellent health?
- What is the probability that a randomly chosen individual has excellent health and health coverage?
- What is the probability that a randomly chosen individual has excellent health or health coverage?
- What is the probability that a randomly chosen individual has excellent health given that they have health coverage?
- If an individual doesn't have health coverage, what is the probability that they have excellent health?
- Does it seem like health status and health coverage are independent?



Problem 11. According to the Anxiety and Depression Association of America, Anxiety Disorders and Depression are very common in the US population. The most common mental illness in the US are anxiety disorders, which affect 18.1% of the population. Of those who have an anxiety disorder, 36.9% receive treatment. One anxiety disorder, Generalized Anxiety Disorder affects 3.1% of the US population, with women twice as likely to be affected than men. Major Depressive Disorder affects approximately 6.7% of the US population. About 50% of individuals with Major Depressive Disorder are also diagnosed with an anxiety disorder.

- a. Are anxiety disorders and Major Depressive Disorders mutually exclusive? Are they independent?
- b. Is Generalized Anxiety Disorder independent of gender?
- c. We know that 36.9% of those who have an anxiety disorder are receiving treatment. How would we write this using mathematical notation?
- d. We know that 50% of individuals with Major Depressive Disorder are also diagnosed with an anxiety disorder. How would we write this using mathematical notation. What is the probability that someone has both Major Depressive Disorder and an anxiety disorder?



Problem 12. Lupus is a medical phenomenon where antibodies that are supposed to attack foreign cells to prevent infections instead see plasma proteins as foreign bodies, leading to a high risk of blood clotting. It is believed that 2% of the population suffers from this disease. The test is 98% accurate if a person actually has lupus. The test is 74% accurate if a person does not have lupus. There is a line from the Fox television show *House* that is often used after a patient tests positive for lupus: "It's never lupus." Do you think there is some truth to this statement? Use appropriate probabilities to support your answer.



Problem 13. About 30% of human twins are identical, and the rest are fraternal. Identical twins are necessarily the same sex. Half of all identical twins are males and the other have are females. One-quarter of fraternal twins are both male, one-quarter are both female, and one-half are mixes: one male, one female. You have just become a parent of twins and are told they are both girls. Given this information, what is the probability that they are identical?



Problem 14. Swaziland has the highest HIV prevalence in the world: 25.9% of this country's population is infected with HIV. The ELISA test is one of the first and most accurate tests for HIV. For those who carry HIV, the ELISA test is 99.7% accurate. For those who do not carry HIV, the test is 92.6% accurate. If an individual from Swaziland has tested positive, what is the probability that they have HIV?



Problem 15. Many times, we see that the probability of a certain event depends on what events happened before. Let's say Mrs. Fleming is the advisor of choir at a 7th-8th grade middle school. She wants to randomly select two students to attend a conference.

- a. What is the probability that at least one student is in seventh grade?
- b. What is the probability that both students are in seventh grade?
- c. What is the probability that both students are in the same grade?



Problem 16. A professor surveyed his students and asked them, on a typical weekend, how many days do you spend studying. The probability distribution is shown below.

x	0	1	2	3
$P(X=x)$	0.10	0.35	0.25	0.30

- Verify that this is a valid probability distribution.
- What is the probability that a randomly selected individual studied at least 2 days?
- What is the probability that a randomly selected individual studied no more than 2 days?
- What is the expected value?



Problem 17. An airline charges the following baggage fees: \$25 for the first bag and \$35 for the second bag. They do not allow individuals to bring more than two bags. Suppose 54% of passengers have no checked luggage, 34% have one piece of checked luggage, and 12% have two pieces.

- a. Draw the probability model for the amount of money spent by passengers on luggage.
- b. What is the expected value?