Unit 1: Probability

Math 361

Welcome to our MATH 361 Class!

- This class is all about probability and how probability is used in statistics
  - How are probability and statistics connected?
    - The answer is RANDOMNESS
- We will make heavy use of computers and oscillate between theory and computation
  - You are ALWAYS encouraged to complete problems using technology (this will make more sense as we go along)
- Let's get started!

Probability

- ► Branch of mathematics concerned with chance

- ► Randomness is unpredictable but probability can help us see patterns

- ► A **random process** is a process for which the individual outcome is unpredictable.

- ► We can think of probability as a mathematical model for chance phenomena/random process

- ► Probability is concerned with situations in which the outcomes of the random process occur randomly

- ► Important in the sciences, engineering, and CS because scientific processes can involve chance and be modeled with chance

► In this unit, we are going to carry out several random processes to develop terminology, understand probabilistic concepts, and develop formulas to compute probabilities.

Investigation 1: Blueberry Pancake

► The Blueberry Pancake House (BPH) prides itself on having at least one blueberry in all of the pancakes they serve. In fact, their slogan is "No Pancake Should Be Without a Blueberry!"

► Our class won a breakfast from BPH as a prize for our outstanding students! BPH will deliver pancakes to the class. Each student in the class will get 6 silver dollar size pancakes.

► We hope that every pancake will have at least one blueberry but we know that no one should get upset if some pancakes have none. Can we figure out the probability of getting a pancake with no blueberries?

► BPH tells us that for an order of 6 pancakes, they use a total of 20 blueberries.

► Answer the following investigative question:

   ► **What is the probability of obtaining a pancake with no blueberries in an order of 6 pancakes?**

Rules

- To help you model the random process of assigning 20 blueberries on 6 pancakes, you have at your disposal the following manipulatives
  - dice
  - pennies
  - paper
  - cards
- You can use some or all of these manipulatives to model the pancake making
- Using these materials, set up a simulation to model the process
- Carry out the simulation at least 20 times using your model
- Compute the empirical probability of getting a pancake without any blueberries in an order of 6

Empirical Probability

- The empirical probability offers a glimpse at the long term behavior of the relative frequency of getting zero blueberries on a pancake

- In the long-run, the empirical probability will converge to the theoretical probability

Main Concepts

- ► Probability shows which events are likely and unlikely.

- ► A simulation is an action that mimics the random process using physical tools (e.g., dice) or software. A simulation is a model of the random process.

- ► When a probability is not known, a simulation can be used to help see the long-run patterns.

- ► The empirical probability of an outcome of a random process is the relative frequency of that outcome for a fixed number of trials.

- ► The theoretical probability of an outcome of a random process is the relative frequency of that outcome as the number of trials tends to infinity.

DAY 2

Investigation 2: The Last Banana

► https://ed.ted.com/lessons/the-last-banana-a-thought-experiment-in-probability-leonardo-barichello#review

► Two people are on a deserted island and they decide to play a game in order to see who will get the last banana to eat. They have two dice with them on the island. The game is the following:

► Both players roll their dice. If a 1, 2, 3, or 4 is the highest number rolled, then player A wins. If instead a 5 or a 6 is the highest number rolled, then player B wins.

 ► **Which player would you rather be?**

Rules

- ► Simulate the game using two dice at minimum 20 times
- ► Use class Google Spreadsheet to record your results
  - ► https://docs.google.com/spreadsheets/d/1l8ztKPZmzcE-6CIj-nfPAxprpOcrK5d-RoK5eYuqA JM/edit?gid=501435346#gid=501435346

- ► Compute the empirical probability of winning for each player using ONLY YOUR simulated results
- ► Compute the empirical probability of winning for each player using the class collected simulated results

- ► Is there a difference between the two empirical probabilities you computes?
- ► Which one do you think is more accurate? Why?

Theoretical Probability Computation

- Find the sample space = the set of all possible outcomes one could get when the two dice are rolled

- Find the probability of winning if you are player A

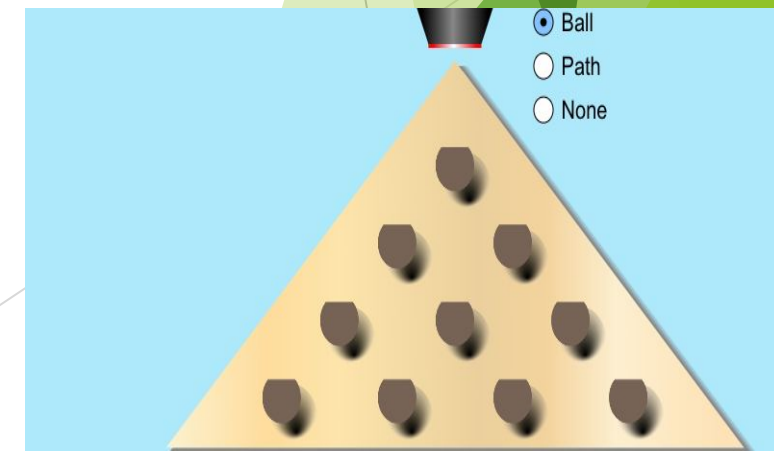- Find the probability of winning if you are player B

The Law of Large Numbers

- Law of Large Numbers states that the relative frequency of a specific outcome of a random process tends towards the theoretical probability of that outcome as the number of repetitions tends to infinity.

Main Concepts

- The empirical probability computed through simulations will converge to the theoretical probability as the number of simulations performed increases.

- Simulations offer a way to estimate the long-run relative frequency of an event.

- If X and Y are independent, then P(X and Y) = P(X)*P(Y)

- P(Y) = 1 - P(X) if P(Y) + P(X) = 1

Start: Investigation 3: Game Board

► The committee for a school fundraiser is organizing a game to raise money for the school. To play the game, parents need to purchase tickets. For each ticket they purchase, they get one turn at the game. The game the committee designs is the following:

► A wooden peg board is constructed and a player releases one ball from the top of the board. The ball follows some pathway and ends in a cell at the bottom of the board.

► The cell either has a prize or no prize associated with it. The students of the school want to create a game so that the chances of winning a prize are low and those of not winning a prize are high. The students pose the following investigative question:

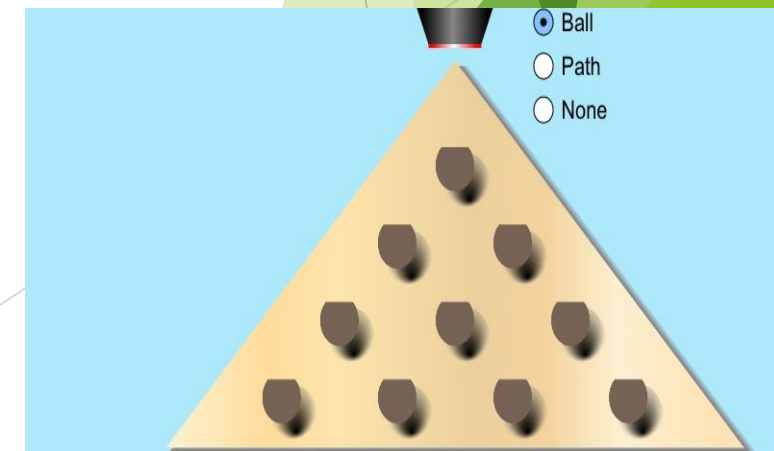 ► **Where should the prizes be placed in order to have the lowest probability of winning?**

Day 3

Investigation 3: Game Board

► The committee for a school fundraiser is organizing a game to raise money for the school. To play the game, parents need to purchase tickets. For each ticket they purchase, they get one turn at the game. The game the committee designs is the following:

► A wooden peg board is constructed and a player releases one ball from the top of the board. The ball follows some pathway and ends in a cell at the bottom of the board.

► The cell either has a prize or no prize associated with it. The students of the school want to create a game so that the chances of winning a prize are low and those of not winning a prize are high. The students pose the following investigative question:

  ► **Where should the prizes be placed in order to have the lowest probability of winning?**

Rules

- Each group of 2 students have a game board showing all possible pathways the ball could take, a coin, and two pegs.

- Both students start their pegs on the start square. In order to advance down the board, the students have to flip the coin. If the student flips to a heads, then they advance their peg to the right. If the student flips to a tails, then they advance to the left. The players keep going until they end in a cell at the bottom.

- Each player plays the game 10 times. For each game, the players mark in what cell they ended.

Main Concepts from Game Board

- ► Probability computations can help make decisions

- ► Implementing hand simulations are useful to build conceptual understanding before moving to software to carry out large number of simulations

- ► Using software to simulate does not necessarily mean needing to program...although that helps because you are more free 😊

- ► Now, we are ready to move to software!

Lab: Downloading R and RStudio

- ► Open up your computers
- ► Do a search for R and Rstudio
- ► Download it and open up Rstudio

Navigating RStudio and R

- Open up R Studio

Investigation 4: Soccer Games

► *Goal: Introduce the Addition Rule through simulations*

► A new high school soccer team has been formed. To motivate the players, the coach plans to play a fun game at the end of each practice. The game consists of trying to score a goal from a corner kick and dribbling the ball in the air. To win the game each practice, a player needs to do one of the following:

► Make a shot directly on goal from a corner kick

► Dribble the ball in the air at least 6 times

► At each practice the team plays the game. There are a total of 100 practices throughout the season.

► Before playing the game at a practice, Bella practices at a field by her house. She makes 2 shots directly on goal from a corner kick out of 10 attempted shots. She also dribbles the ball in the air 6 times without dropping the ball 7 times out of 10 attempts. She then goes home and sets up a simulation of the 100 upcoming practices based on this preliminary practice to compute her probability of winning the game.

Set up the Simulation

- Use a paperclip, a piece of paper, a cup, and colored pencils

Day 4

Investigation 4: Soccer Games

► *Goal: Introduce the Addition Rule through simulations*

► A new high school soccer team has been formed. To motivate the players, the coach plans to play a fun game at the end of each practice. The game consists of trying to score a goal from a corner kick and dribbling the ball in the air. To win the game each practice, a player needs to do one of the following:

► Make a shot directly on goal from a corner kick

► Dribble the ball in the air at least 6 times

► At each practice the team plays the game. There are a total of 100 practices throughout the season.

► Before playing the game at a practice, Bella practices at a field by her house. She makes 2 shots directly on goal from a corner kick out of 10 attempted shots. She also dribbles the ball in the air 6 times without dropping the ball 7 times out of 10 attempts. She then goes home and sets up a simulation of the 100 upcoming practices based on this preliminary practice to compute her probability of winning the game.

Set up the Simulation

- Use a paperclip, a piece of paper, a cup, and colored pencils

Summary

- **The main concepts developed in the Soccer Games Investigation are:**

- Addition rule for events X and Y in probability states:

  - P(X or Y) = P(X) + P(Y) – P(X intersection Y).

- When X and Y are independent, P(X and Y) = P(X)*P(Y)

- Simulations are useful to model and understand complex probabilistic scenarios.

Investigation 5: Dominant Traits

► Sarah is about to become an older sister. She is interested in finding out the chances that her new sibling will look like her. Specifically, she is interested in understanding the chances of her sibling being right-handed like her. Right handedness is a dominant gene. Sarah wants to find the probability that her sibling is born with this dominant trait. **In order for a child to be born with a dominant trait, either the mother's or the father's dominant gene must be passed on to the child.**

► Based on family history, Sarah's family doctor tells her that her mother has the dominant trait with a probability of .6 and that her father has the dominant trait with probability .8. The probability of both of them having it is .5.

► Sarah wants to know the following:

  ► **What is the probability that the mom and dad dominant trait prevails and the child also gets the dominant trait?**

Sample Space

- ► Sarah recognizes that there are four possible outcomes for her mom and dad passing along the gene

- ► She knows that her sibling will receive the trait for outcomes 1, 2 and 3 and not in case 4.

	Dad Dominant	Dad Recessive
Mom Dominant	Outcome 1: both dominant	Outcome 2: Dominant, Recessive
Mom Recessive	Outcome 3: Recessive, Dominant	Outcome 4: both recessive

Sketch out how to carry out a simulation by hand

- Ideas?

Theoretical Probability

► Now that we have modeled the Dominant Trait through simulation, lets compute the probability of the child getting the dominant trait theoretically

► Our scenario amounts to computing the probability of when **either** the mom **or** the dad contribute the dominant gene.

► Lets model this in a Venn Diagram

► Therefore, we have:

  ► P(getting dominant gene) = P(mom giving dominant gene OR dad giving dominant gene) = P(mom giving dominant gene) + P(dad giving dominant gene) - P(mom and dad both giving the dominant gene)

EITHER/OR Statements

► One must subtract off the "both" because that probability is already counted in the general probabilities of mom giving dominant gene and dad giving dominant gene.

► If we do not subtract off, she would be counting the intersection twice.

► Therefore,

► P(getting dominant gene) = P(mom giving dominant gene OR dad giving dominant gene) = P(mom giving dominant gene) + P(dad giving dominant gene) - P(mom and dad both giving the dominant gene) = .6 + .8 - .5 = .9

► How close is our empirical probability?

Main Concepts from Dominant Traits

- The addition rule for events X and Y states:
  - $P(X \text{ or } Y) = P(X) + P(Y) - P(X \text{ intersection } Y)$
- Simulations are useful to model and understand complex probabilistic scenarios.

HW: Finish Lab: Modeling Dominant Traits in R

- Carry out the Dominant Traits simulation in R

- Open up RStudio and we can write a Markdown file together for this simulation

Day 5

Main Concepts from Dominant Traits

- The addition rule for events X and Y states:
  - P(X or Y) = P(X) + P(Y) – P(X intersection Y)
- Simulations are useful to model and understand complex probabilistic scenarios.

Lab: Monty Hall Problem in R

- https://video.search.yahoo.com/yhs/search;_ylt=AwrSw7XN_WhaYDkAn3o0nIlQ;_ylu=X3oDMT
BncGdyMzQ0BHNlYwNzZWFyY2gEdnRpZAM-;_ylc=X1MDMTM1MTE5NTcwMARfcgMyBGFjdG4DY2
xrBGJjawM3azYwYTA1Y2k2azlrJTI2YiUzRDMlMjZzJTNENmEEY3NyY3B2aWQDTlNrRkpERXdMako
2R0FvQldTTlJOQTFhTVRVNUxnQUFBQUNLdUdiNARmcgN5aHMtaWJhLTEZnlyA3NhLWdwBGdwc
mlkA1hEQ3JrLjJSUkZ5empiclpkYlc0UkEEbXRlc3RpZANudWxsBG5fcnNsAM2MARuX3N1Z2cDNgR
vcmlnaW4W4DdmlkZW8uc2VhcmNoLnlhaG9vLmNvbQRwb3MDMARwcXN0cgMEcHFzdHJsAwRxc3Ry
bAMyNARxdWVyeQNsZXRzIG1ha2UgYSBkZWFsIHlvdXR1YmUEdF9zdG1wAzE1MTY4MzA0MTkEdnR
lc3RpZANudWxs?gprid=XDCrk.2RRFyzjbrZdbW4RA&pvid=NSkFJDEwLjJ6GAoBWSNRNA1aMTU5Lg
AAAACKuGb4&p=lets+make+a+deal+youtube&ei=UTF-8&fr2=p%3As%2Cv%3Av%2Cm%3Asa&fr=y
hs-iba-1&hsimp=yhs-1&hspart=iba#id=8&vid=6b80aae65d124a211eb59ea7dedc2ef7&action=vi
ew

Monty Hall Problem

► Monty Hall was the host of the famous game show *Let's Make a Deal* which was popular in the 70s and 80s. After a series of deals, each competitor on the show would be confronted with three doors. Only one of the doors led to a good prize, usually a car or a trip. Behind the other two doors, there were things that nobody would want (e.g., a goat).

► The competitors would initially select a door and then the host would open up one of the remaining doors. (always the one that did not have the prize of course).

► After that, the competitor would have the choice to switch doors or not to switch doors.

► **What would you do?**

Simulation

- How could you simulate this situation in order to see whether you have a larger chance at winning if you switch or do not switch doors?

- Sketch out your simulation.

Simulation in R

- Carry out the simulation in R
  - See Lab handout

Theoretical Probability Computation

- Using Tree Diagram
  - Resource: https://www.youtube.com/watch?v=cphYs1bCeDs

- Using Bayes' Theorem
  - Resource: https://blogs.cornell.edu/info2040/2022/11/10/the-monty-hall-problem-using-bayes-theorem/

Main Concepts

- Simulations in complex, sometimes counterintuitive situations, can greatly help clarify and convince us of the probabilities.

- Bayes' Theorem states:

  - P(X|Y) = P(Y|X) * P(X) / P(Y)

- The conditional probability of X given Y is given by:

  - P(X | Y) = P(X intersection Y) / P(Y)

- Multiplication Rule: P(A intersection B) = P(A|B)P(B)

  - Note that P(A|B) = P(A) when A and B are independent

Homework Lab: Detecting Disease Lab

► Strep throat is a common virus. To detect whether someone has this virus, doctors administer a swab test. Even though the swab test helps doctors make a diagnoses, it does, at times, give the wrong results. Specifically, the test could come back negative when in fact a patient has strep (false negative) or the test could come back positive when in fact a patient does not have strep (false positive). As a doctor, it is important to understand the chances of these false positives and false negatives occurring so that risks could be communicated accurately with patients.

► A doctor asks:

   ► **What is the probability of not having the virus given that you get a positive test result?**

   ► ***Must do the lab through simulation in R (or other software) and theoretically. Check that answers "match." Minimum of 10,000 runs for the simulation.**

Day 6

Main Concepts

- Simulations in complex, sometimes counterintuitive situations, can greatly help clarify and convince us of the probabilities.

- Bayes' Theorem states:
  - $P(X|Y) = P(Y|X) * P(X) / P(Y)$

- The conditional probability of X given Y is given by:
  - $P(X | Y) = P(X \text{ intersection } Y) / P(Y)$

- Multiplication Rule: $P(A \text{ intersection } B) = P(A|B)P(B)$
  - Note that $P(A|B) = P(A)$ when A and B are independent

Bayes' Theorem Practice

- https://betterexplained.com/articles/an-intuitive-and-short-explanation-of-bayes-theorem/

Bayes Theorem Example

► An airport screens bags for forbidden items, and an alarm is supposed to be triggered when a forbidden item is detected.

  ► Suppose that 5% percent of bags contain forbidden items.

  ► If a bag contains a forbidden item, there is a 98% percent chance that it triggers the alarm.

  ► If a bag doesn't contain a forbidden item, there is an 8% percent chance that it triggers the alarm.

► **Given a randomly chosen bag triggers the alarm, what is the probability that it contains a forbidden item?**

Main Concepts

► Simulations in complex, sometimes counter-intuitive situations, can greatly help clarify and convince us of the probabilities.

► Bayes' Theorem states:

  ► $P(X|Y) = P(Y|X) * P(X) / P(Y)$

► The conditional probability of X given Y is given by:

  ► $P(X | Y) = P(X \text{ intersection } Y) / P(Y)$

► Multiplication Rule: $P(X \text{ intersection } Y) = P(X|Y)P(Y)$

  ► Note that $P(X|Y) = P(X)$ when X and Y are independent

Review: Basics from Probability Unit

► Sample space = the set of all possible outcomes of a random process

► Probability of an event must be between 0 and 1

► P(A U B) = P(A) + P(B) – P(A intersection B)

► P(A complement) = 1 – P(A)

► P(A) = number of ways A can occur / total number of outcomes

Review: Using Counting in Probability

- There are n! different ways to order n objects (permutations of n objects)

- How many ways are there to choose r items from a list of n items?

  - If sampling without replacement: (n!) / ((n-r)! * r!)= n choose r

Review

- Conditional probability: P(A | B) = P(A intersection B) / P(B)
  - Rewrite this as the multiplication law
  - If A and B are independent, then P(A | B) = P(A)
  - A and B are independent if: P(A intersection B) = P(A)*P(B)
  - Regardless of whether two events are independent, you ALWAYS can use the following:
    - P(A U B) = P(A) + P(B) – P(A intersection B)

In-Class Review Problems

1. Company A supplies 40% of the computers sold and is late 5% of the time. Company B supplies 30% of the computers sold and is late 3% of the time. Company C supplies another 30% and is late 2.5% of the time. A computer arrives late - what is the probability that it came from Company A?

2. In a study of pleas and prison sentences, it is found that 45% of the subjects studied were sent to prison. Among those sent to prison, 40% chose to plead guilty. Among those not sent to prison, 55% chose to plead guilty.

A. If one of the study subjects is randomly selected, find the probability of getting someone who was not sent to prison.

B. If a study subject is randomly selected and it is then found that the subject entered a guilty plea, find the probability that this person was not sent to prison.

3. On a game show, a contestant can select one of four boxes. The red box contains one $100 bill and nine $1 bills. A green box contains two $100 bills and eight $1 bills. A blue box contains three $100 bills and seven $1 bills. A yellow box contains five $100 bills and five $1 bills. The contestant selects a box at random and selects a bill from the box at random. If a $100 bill is selected, find the probability that it came from the yellow box.

In-Class Review Problems

1. If you take out a single card from a regular pack of cards, what is probability that the card is either an ace or spade?

2. Suppose an investor considers investing in two stocks, A and B. The probability of stock A increasing in value over the next year is 0.4, and the probability of stock B increasing in value over the next year is 0.6. The investor wants to know the probability that at least one of the two stocks will increase in value. Find it.

3. Let's say a bank is considering giving loans to two borrowers, X and Y. The probability of borrower X defaulting on the loan is 0.3, and the probability of borrower Y defaulting on the loan is 0.4. The bank wants to know the probability that at least one borrower will default. Find it.

More Problems

1. Suppose you take out two cards from a standard pack of cards one after another, without replacing the first card. What is probability that the first card is the ace of spades, and the second card is a heart?

2. Helen plays basketball. For free throws, she makes the shot 75% of the time. Helen must now attempt two free throws. C= the event that Helen makes the first shot. P(C)=0.75. D= the event Helen makes the second shot. P(D)=0.75
   a. Find the probability that Helen makes the second free throw given that she made the first is 0.85.
   b. What is the probability that Helen makes both free throws?

3. A community swim team has **150** members. **Seventy-five** of the members are advanced swimmers. **Forty-seven** of the members are intermediate swimmers. The remainder are novice swimmers. **Forty** of the advanced swimmers practice four times a week. **Thirty** of the intermediate swimmers practice four times a week. **Ten** of the novice swimmers practice four times a week. Suppose one member of the swim team is chosen randomly.

   a. What is the probability that the member is a novice swimmer?

   b. What is the probability that the member practices four times a week?

   c. What is the probability that the member is an advanced swimmer and practices four times a week?

   d. What is the probability that a member is an advanced swimmer and an intermediate swimmer? Are being an advanced swimmer and an intermediate swimmer mutually exclusive? Why or why not?

   e. Are being a novice swimmer and practicing four times a week independent events? Why or why not?

Due on day of Chapter 1 Test

- Lab: Detecting Disease
- Practice Test

- TEST 1: some multiple choice, some short answer, straight off the practice test, and labs (both in class and assigned)