Introduction to Probability Lessons 1 - 4

Erin Desmond, Sean Costello, Alicia Key

Learning Objectives

- ❏ Define Probability and articulate a probability question
- ❏ Describe the three axioms of probability theory
- ❏ Illustrate the difference between:
    - ❏ Union, Intersection, Difference & Complement
- ❏ Define De Morgan's Laws
- ❏ Define Mutual Exclusivity

Probability

The likelihood that an event will occur

Some Basic Probability Questions

- What is the probability that it will rain tomorrow?
- What is the probability that I will draw a queen of hearts from a standard deck of 52 cards?
- Based upon a positive test result, what is the probability that I have a disease?

Three Axioms of Probability

First Axiom

- The probability of an event occurring is a non-negative real number.

- $P(A) \in \mathbb{R}, P(A) \geq 0$

Second Axiom

- The probability of an entire sample space is equal to 1.

- $P(S) = 1$

Third Axiom

- If two events are mutually exclusive, the probability of the union of the events is equal to the sum of their individual probabilities.

- If $A \cap B = 0$, then $P(A \cup B) = P(A) + P(B)$

More on this later

Following from the Axioms

...

Empty set, monotonicity, numeric bound and non-event

Following from the Axioms

Probability of the empty set:

- The probability of an empty set (an impossible event) is equal to 0.

- $P(\emptyset) = 0$

Following from the Axioms

Monotonicity

- If A is a subset of B, or is equal to B, then the probability of A happening is less than or equal to the probability of B happening.

- If $A \subseteq B$, then $P(A) \leq P(B)$

Following from the Axioms

Numeric Bound

- The probability of an event must be between 0 and 1.

- $0 \leq P(A) \leq 1$

Following from the Axioms

Non-Event

- The probability of an event not occuring is equal to 1 minus the probability of the event occuring.

- $P(E^C) = 1 - P(E)$

Probability Definitions

Foundational terms for further discussion

Definitions

Set

- A collection of distinct objects.
- A set: {1, 4, 15, apple, John}
- Not a set: [1, 2, 2, 6]

Definitions

Experiment

- Observing and recording the result of a random phenomenon.

- Often denoted with a capital ‘X’.

Definitions

Random

- Something whose outcome can only be described using probability.

Definitions Continued

Outcome

- The possible results of a random experiment.
- Ex: “All the possible individual numbers you would get if you rolled a dice: 1, 2, 3, 4, 5, or 6”
- Often denoted with a lowercase ‘x’.

Definitions Continued

Event

- Sets of outcomes from an experiment.

- Ex: “You actually roll the dice (perform an experiment). The outcome is the event.”

Class Question

Which is the sample space, the outcome, and the event?

A. A deck of 52 cards.

B. Drawing a king of spades.

C. Each individual card in the deck.

Set Relationships

Of the people in this room, how many use Python or R?

The above question describes a set relationship.

Class Question:

What are the possible outcomes of our sample space?

Set Relationships

Of the people in this room, how many use Python or R?

All possible outcomes:

- Those who use both Python and R
- Those who use only Python
- Those who use only R
- Those who use neither

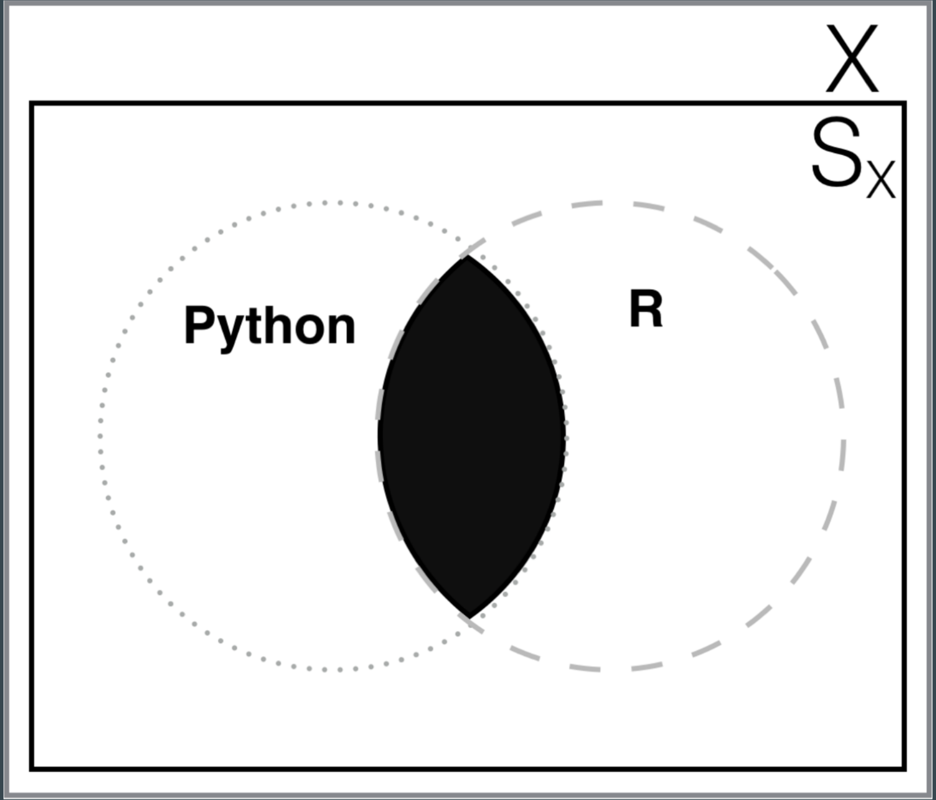
Set Operators: Intersection and Union

- Intersection:
  - The set of all people who use both Python AND R
    - $E_{Python} \cap E_R$
- Union:
  - The set of all people who use either Python OR R (or both)
    - $E_{Python} \cup E_R$

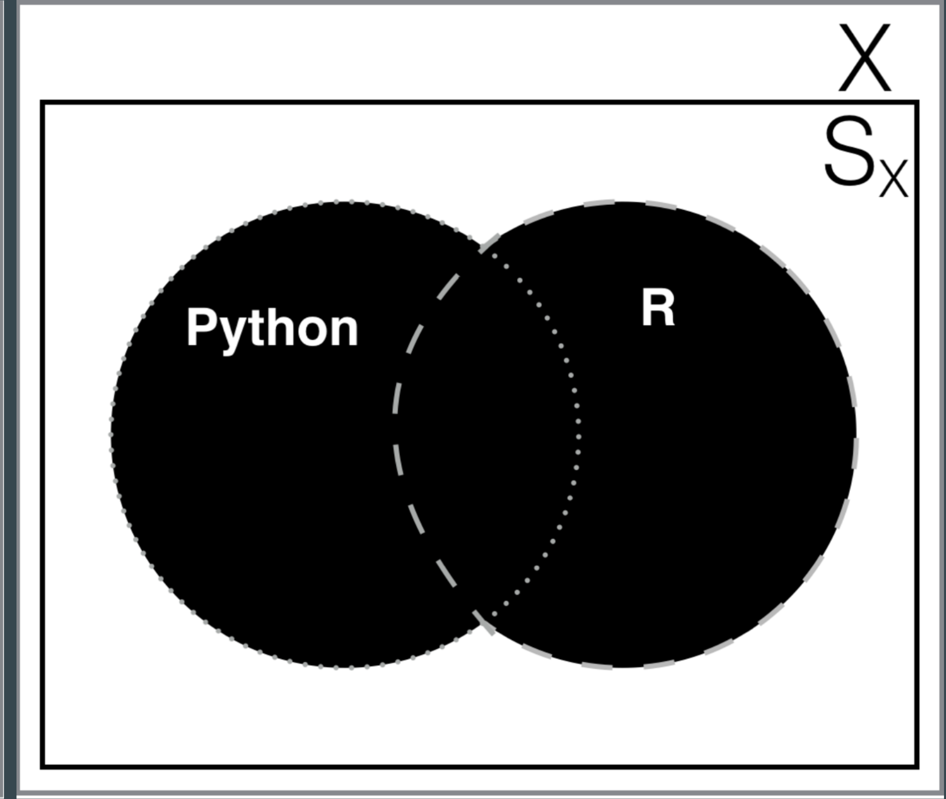
Set Operators: Difference and Complement

- Difference:
  - People who use Python but do not use R and vice versa.
    - $E_{Python} \setminus E_R$ OR $E_R \setminus E_{Python}$
- Complement:
  - Every member of the universal set that does NOT belong to the set of interest.
    - Not R: E^C_R
    - Not Python: E^C_{Python}

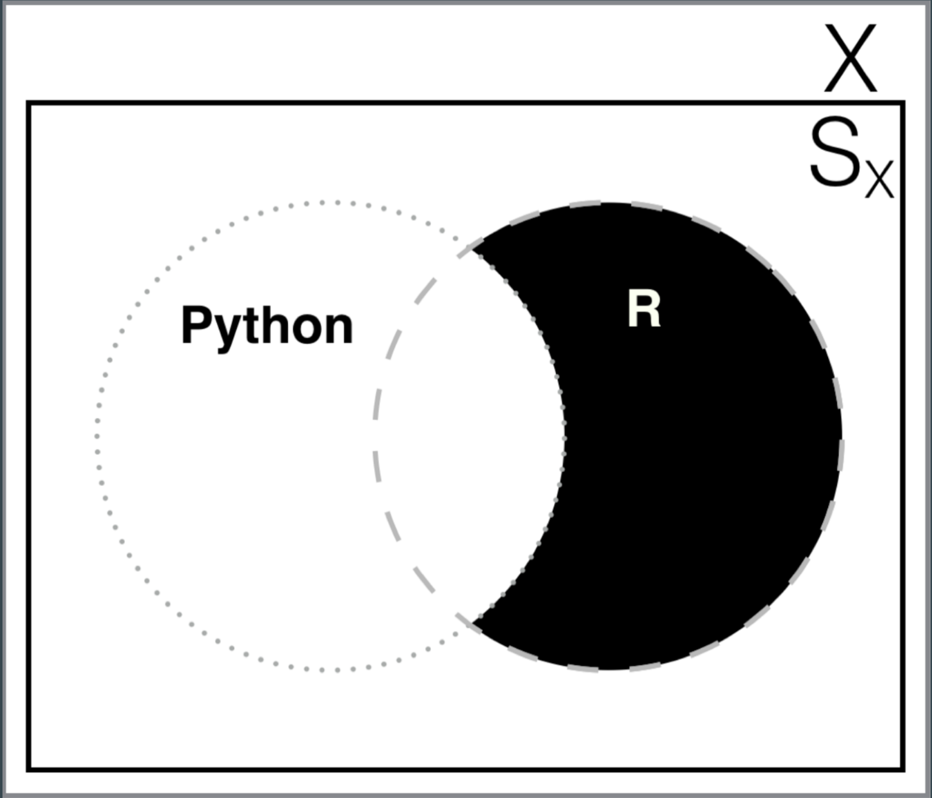
$E_{Python} \cap E_R$



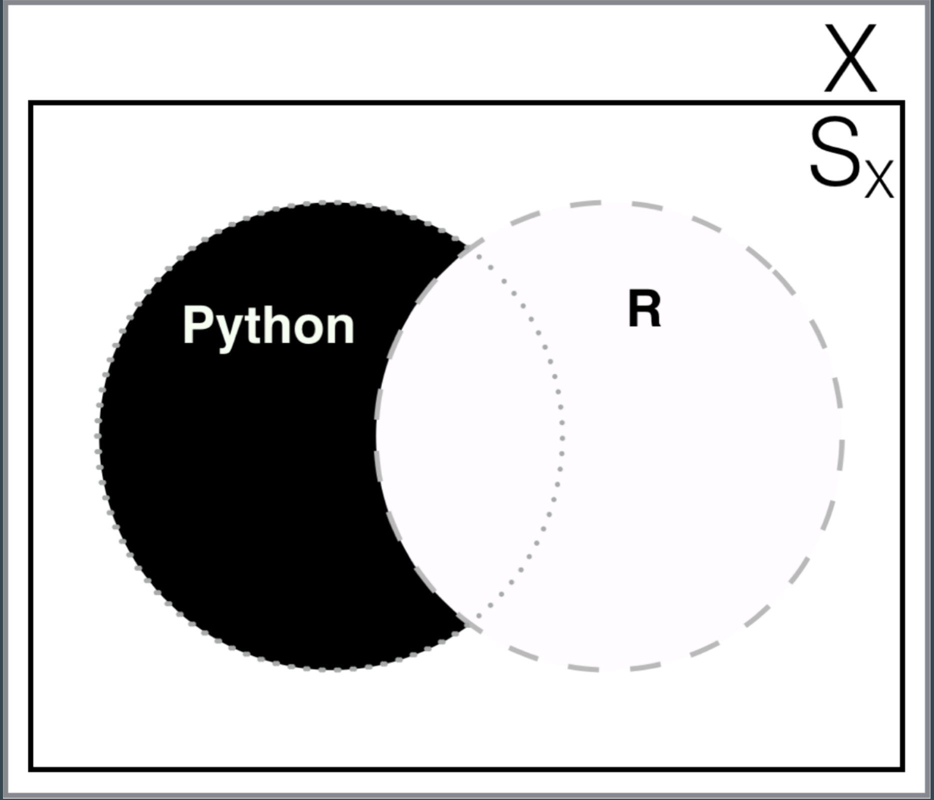
$E_{Python} \cup E_R$



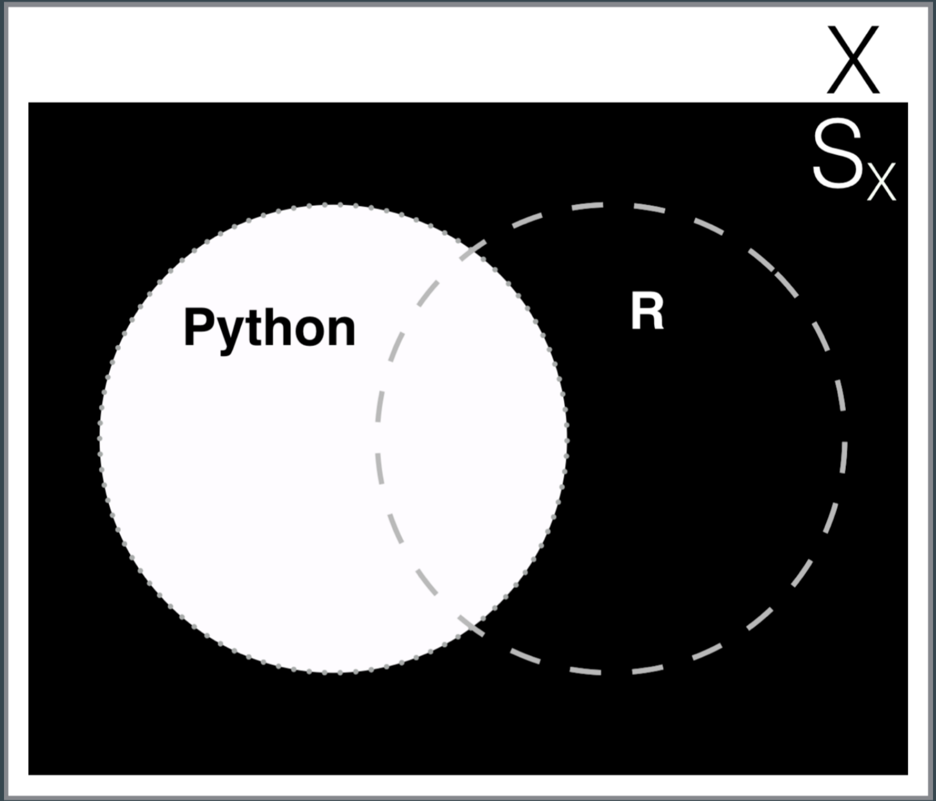
$E_R \setminus E_{Python}$



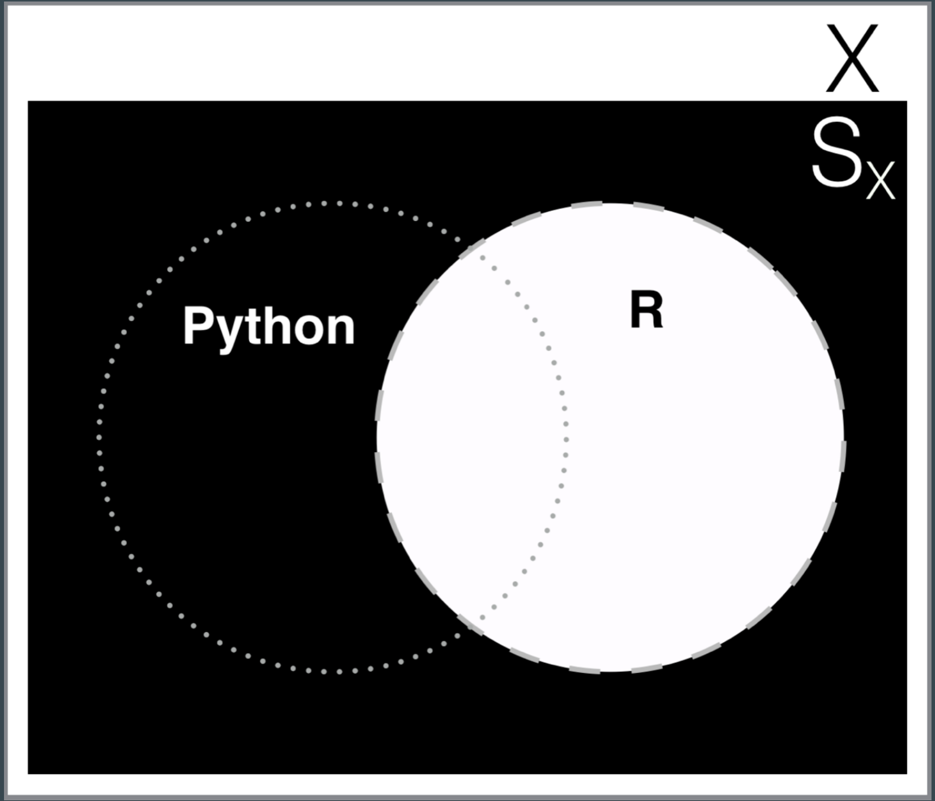
$E_{Python} \setminus E_R$



E^C_{Python}



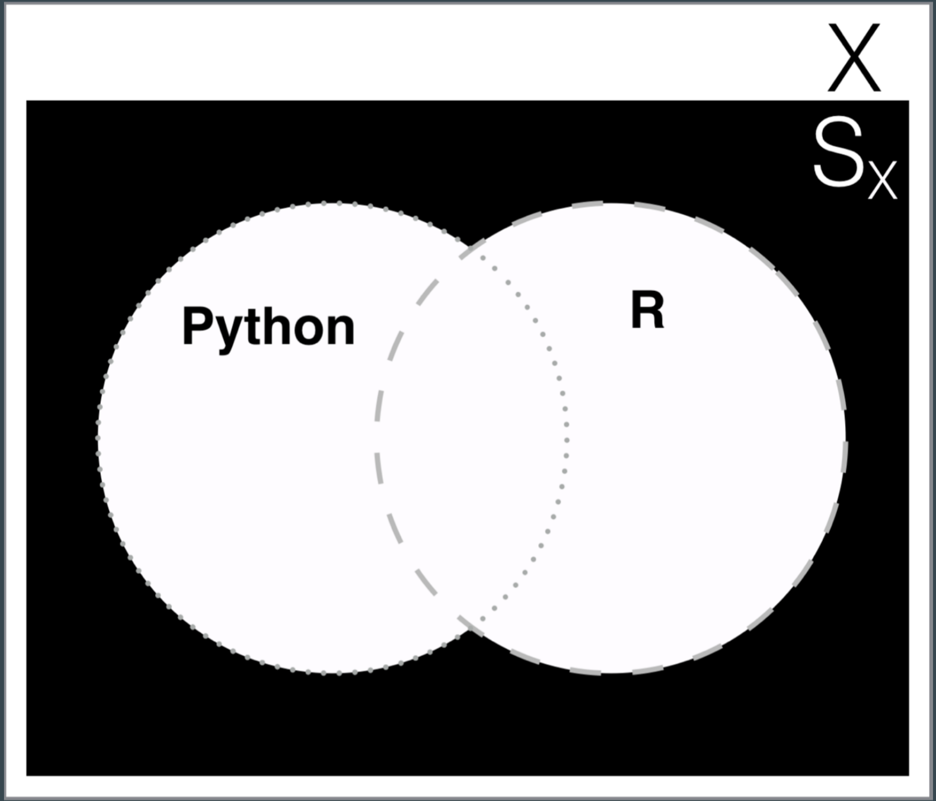
E^C_R

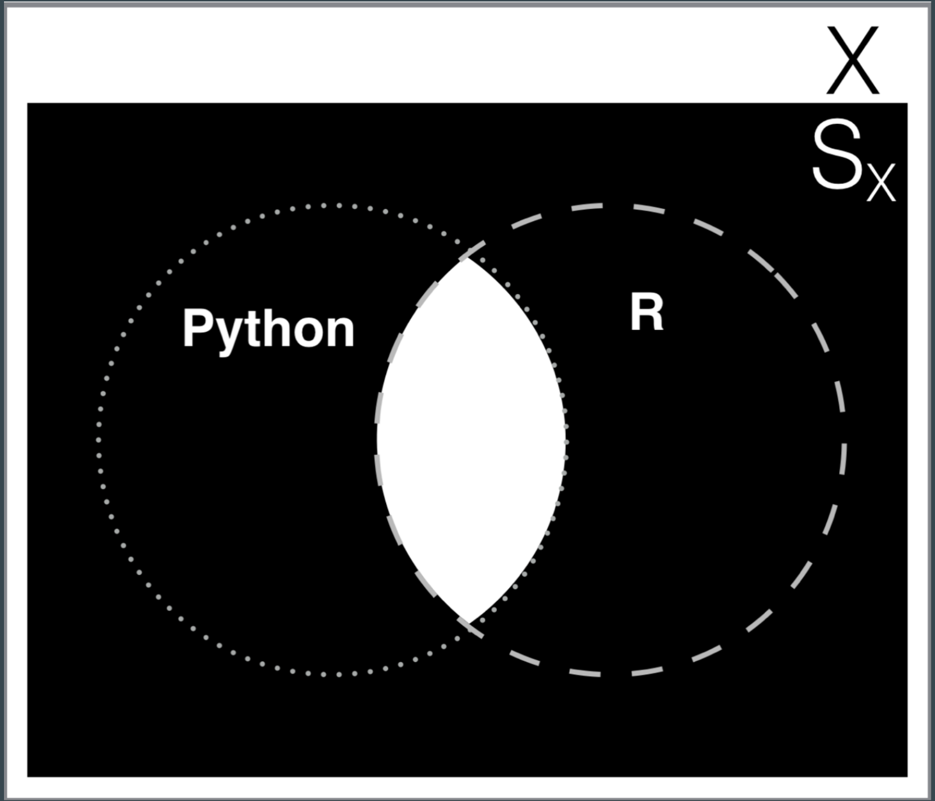


Class Challenge

A. How can you represent the set of people who don’t use either Python or R?
B. What about the people who might know one, but don’t know both languages?
   - Visually with a Venn Diagram
   - Mathematically

$(E_{Python} \cup E_R)^C$ *OR* $E^C_{Python} \cap E^C_R$ $(E_{Python} \cap E_R)^C$ *OR* $E^C_{Python} \cup E^C_R$





De Morgan's Laws

- The complement of the union of two sets is equal to the intersection of their complements.
    - $(E_{Python} \cup E_R)^C \quad = \quad E^C{}_{Python} \cap E^C{}_R$
- The complement of the intersection of two sets is equal to the union of their complements.
    - $(E_{Python} \cap E_R)^C \quad = \quad E^C{}_{Python} \cup E^C{}_R$

Partitions

- Partition of a set:
  - A grouping of the set's elements into non-empty subsets, in such a way that every element is included into one, and only one, subset. Also called 'disjoint'.
    - They use both Python and R
    - They only use Python
    - They only use R
    - They don't use either
- Are these partitions?
    - They use Python
    - They use R

Mutual Exclusivity

- Two or more events that cannot occur simultaneously.
  - The occurrence of one event is not influenced or caused by the occurrence of another event.
    - Two events, E_i and E_j are said to be mutually exclusive when:
      - $E_i \cap E_j = \emptyset$, where $\emptyset$ is the event containing no outcomes.

- Class challenge:
  - Name some mutually exclusive events.

Summary

- ❑ Defined Probability and articulate a probability question
- ❑ Described the three axioms of probability theory
- ❑ Illustrated the difference between:
    - ❑ Union, Intersection, Difference & Complement
- ❑ Defined De Morgan's Laws
- ❑ Defined Mutual Exclusivity

Introduction to Probability
Lessons 5 - 7

...

Erin Desmond, Sean Costello, Alicia Key

Learning Objectives

- Use probability concepts to determine whether two events are dependent or independent
- Articulate the difference between correlation and causation
- Define dependent and independent events and provide an example of each
- Define and write the Conditional Probability Rule
- Define and write the Chain Rule
- Define, write, and use the Law of Total Probability

Class Discussion Aside: Correlation vs. Causation

- Describe the difference between correlation and causation.
- List an example of correlation and an example of causation.
- Discuss the ethical implications of confusing the two terms and give an example of why data scientists need to pay attention to the difference.

Probability Review

The probability of an event is the number of outcomes that meet some condition divided by the total number of equally likely outcomes in the sample space of the experiment.

P(E) = outcomes that meet a condition / total number of equally likely outcomes

Dependent Events

- Two events are dependent if the result of the first event affects the outcome of the second event such that the probability of that event is changed.
- Two events are dependent if:
  - $P(E_2 \mid E_1) \neq P(E_2)$
  - $P(E_2)$ “given that E_1 has happened” is not equal to the $P(E_2)$.

Independent Events

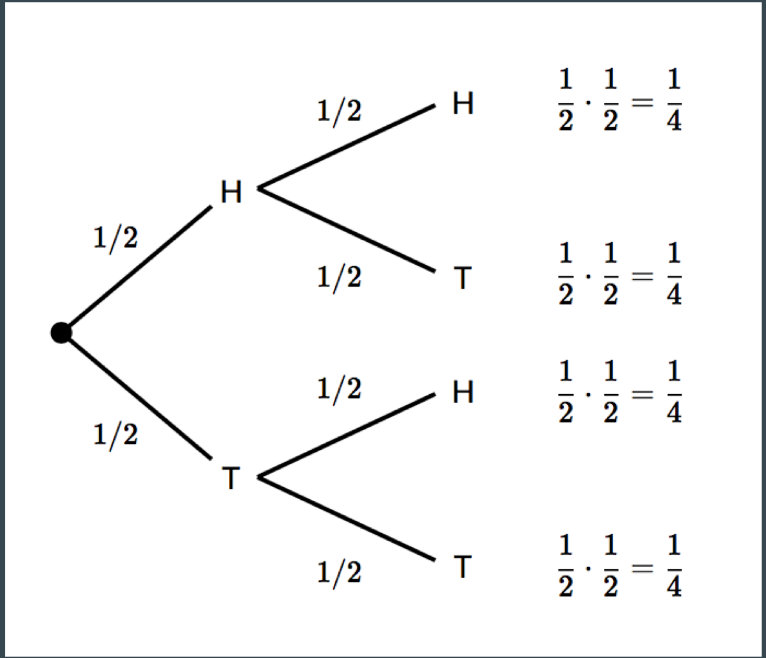
- Two events are independent if the result of the second event is not affected by the result of the first event.
- Two events are independent if:
    - $P(E_2 \mid E_1) = P(E_2)$ OR $P(E_1 \mid E_2) = P(E_1)$
    - $P(E_2)$ “given that E_1 has happened” is equal to the $P(E_2)$.

Multiplication Rule

- Independent Events:
  - P(A and B) = P(A) ・ P(B)
  - $P(A \cap B) = P(A) \cdot P(B)$
- Dependent Events:
  - The multiplication rule still applies, but with an important difference - your sample space has changed. This is called the conditional probability rule and we'll get to that in a moment.
  - Ex: Drawing any king from a standard deck of 52 cards without replacement followed by a 2 of hearts.
    - P(King) = 4/52
    - P(2♡ after drawing a king without replacement) = 13/51
    - I'll let you figure out the rest...

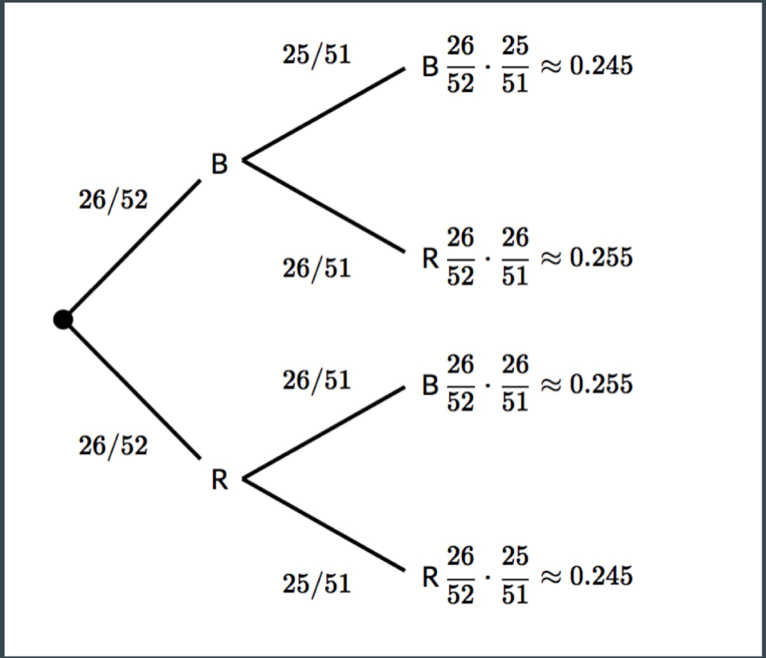
Independent Events Trees

What's the probability of flipping a fair coin and getting "heads" twice in a row? Simply multiply along the branches to find the answer.



Dependent Events Trees

Draw two cards from a standard 52 card deck without replacement. What's the probability that both cards selected are black?



Class Challenge

- There is a bag of marbles with 5 red marbles, 3 yellow marbles and 2 blue marbles. You draw one marble from the bag, record the color, return the marble to the bag and draw again.
    - Are these events dependent or independent?
    - What is the probability of drawing a red marble followed by a blue marble?
- From the same bag of marbles, you draw one marble, record the color, then draw another marble without returning the first one to the bag.
    - Are these events dependent or independent?
    - What is the probability of drawing a yellow marble followed by a red marble?

Conditional Probability Rule

- Independent events:
  - $P(B \mid A) = P(B)$
- Dependent events:
  - $P(A \text{ and } B) = P(B \mid A) \cdot P(A)$
- Rearranging these, we get the Conditional Probability Rule:
  - $P(B \mid A) = P(A \text{ and } B) / P(A)$
  - $P(B \mid A) = P(A \cap B) / P(A)$

The Chain Rule

Useful for calculating the probability of multiple events occurring.

- Recall the Conditional Probability Rule:
  - $P(E_2 \mid E_1) = P(E_1 \cap E_2) / P(E_1)$
  - P(B | A) = P(A and B) / P(A)
- Multiply both sides of the equation by $P(E_1)$:
  - $P(E_1 \cap E_2) = P(E_2 \mid E_1) \cdot P(E_1)$ OR $P(E_2 \cap E_1) = P(E_1 \mid E_2) \cdot P(E_2)$
  - P(A and B) = P(B | A) ・ P(A) OR P(A and B) = P(A | B) ・ P(B)

Summary

- Independent events:
  - P(B | A) = P(B)
  - P(A | B) = P(A)
- Dependent events:
  - P(A and B) = P(B | A) ・ P(A)
  - P(A and B) = P(A | B) ・ P(B)
- Conditional Probability Rule (divide equation for dependent events by P(A) or P(B)):
  - P(B | A) = P(A and B) / P(A)
  - P(A | B) = P(A and B) / P(B)
- Chain Rule (multiply both sides of conditional probability rule by P(A) or P(B)):
  - P(A and B) = P(B | A) ・ P(A)
  - P(A and B) = P(A | B) ・ P(B)

Definitions

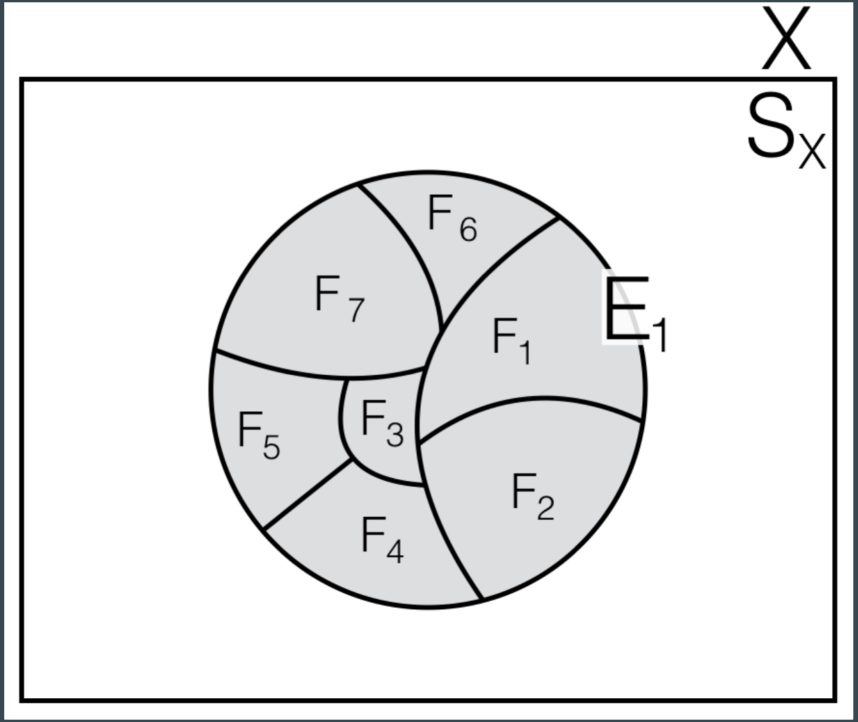
- Conditional Probability:
  - Ex: $P(A \mid B)$
- Joint Probability:
  - Ex: $P(A \text{ and } B)$ OR $P(A \cap B)$
- Marginal Probability:
  - Ex: $P(A)$

Law of Total Probability

Used to find the probability of an event, A, when you don't know enough about A's probabilities to calculate it directly. Instead, you can take a related event, B, and use that to calculate the probability for A.

Basically, it's a method of reducing joint probabilities to marginal probabilities.

Visual Law of Total Probability



Assuming mutual exclusivity for all set of events, F_1 - F_7

Law of Total Probability Formula

$$\Pr(E_2) = \Pr(E_2 \cap G_1) + \Pr(E_2 \cap G_2) + \Pr(E_2 \cap G_3) + \cdots + \Pr(E_2 \cap G_p)$$

Using the Chain Rule

$$\Pr(E_1) = \Pr(E_1|F_1) \cdot P(F_1) + \Pr(E_1|F_2) \cdot P(F_2) + \Pr(E_1|F_3) \cdot P(F_3) + \cdots + \Pr(E_1|F_p) \cdot P(F_p)$$

Class Challenge

Suppose that a school will be closed 90% of the time when it snows, and 5% of the time when it rains, and 1% of the time when it neither snows nor rains. And suppose you're interested in calculating the probability that the school will be closed on any given day.

Using the Law of Total Probability:

- Write the Law of Total Probability in the context of this equation.
- Given the following, calculate the probability that the school will be closed:
  - P(snow) = 0.02
  - P(rain) = 0.15
  - P(good weather) = 0.83

Class Challenge Answers

- $P(\text{closed}) = P(\text{closed} \mid \text{snow}) \cdot P(\text{snow})$

  $+ P(\text{closed} \mid \text{rain}) \cdot P(\text{rain})$

  $+ P(\text{closed} \mid \text{good weather}) \cdot P(\text{good weather})$

- $P(\text{closed}) = (0.90 \cdot 0.02) + (0.05 \cdot 0.15) + (0.01 \cdot 0.83) = 0.0338$

Introduction to Probability Lessons 8 - 10

...

Erin Desmond, Sean Costello, Alicia Key

Learning Objectives

- ❏ Understand the difference between Bayesian and Frequentist approaches to statistics.
- ❏ Define Bayes' Rule.
- ❏ Derive Bayes' Rule from the Chain Rule.
- ❏ Use Bayes' Rule in a real life situation.
- ❏ Articulate the difference between Precision and Recall
- ❏ Use a probability tree to solve Bayesian problems

Bayesian Statistics

- Bayesian Statistics:
    - Probability expresses a degree of belief in an event. The degree of belief may be based upon prior knowledge, such as results from a previous experiment, or on personal beliefs about the event.
- Frequentist Statistics:
    - Contrasted with this interpretation is the Frequentist interpretation which views probability as the limit of the relative frequency of an event after a large number of trials.

Lost Phone for Bayesian vs. Frequentist

- Frequentist Reasoning:
    - I can hear the phone beeping. I also have a mental model which helps me identify the area from which the sound is coming. Therefore, upon hearing the beep, I infer the area of my home I must search to locate the phone.
- Bayesian Reasoning:
    - I can hear the phone beeping. Now, apart from a mental model which helps me identify the area from which the sound is coming from, I also know the locations where I have misplaced the phone in the past. So, I combine my inferences using the beeps and my prior information about the locations I have misplaced the phone in the past to identify an area I must search to locate the phone

Bayesian Pros

- Incorporates relevant prior probabilities.
- Represents a degree of belief rather than a relative frequency (Frequentist cannot assign a probability to a hypothesis).
- Can calculate the probability that a hypothesis is true.
- Allows for more nuanced analysis.
- Bayes' Rule derived from logical foundation of probability theory.

Bayesian Cons

- Introduces subjectivity.
- Priors are often difficult to justify and can be a major source of inaccuracy.
- There can be too many hypotheses.
- More computationally expensive.
- Analytical solutions can be difficult to derive.
- Can be more difficult to explain.

Review of the Chain Rule

Recall the Chain Rule:

- $P(E_1 \cap E_2) = P(E_2 \mid E_1) \cdot P(E_1)$ OR $P(E_2 \cap E_1) = P(E_1 \mid E_2) \cdot P(E_2)$
- $P(A \text{ and } B) = P(B \mid A) \cdot P(A)$ OR $P(A \text{ and } B) = P(A \mid B) \cdot P(B)$

Rearranging this rule, we can derive Bayes' Rule…

Bayes' Rule

- Recall the Chain Rule:
  - $P(A \text{ and } B) = P(B|A) \cdot P(A)$
  - $P(A \text{ and } B) = P(A|B) \cdot P(B)$
- Rearrange to isolate a conditional probability:
  - $P(B|A) = P(A \text{ and } B) / P(A)$
  - $P(A|B) = P(A \text{ and } B) / P(B)$
- Bayes' Rule:
  - $P(B|A) = P(A|B) \cdot P(B) / P(A)$
  - $P(A|B) = P(B|A) \cdot P(A) / P(B)$

Some Terminology

- Posterior:
  - How probable was our hypothesis before observing the evidence?
- Likelihood:
  - How probable is the evidence given that our hypothesis is true?
- Prior:
  - How probable is our hypothesis given the observed evidence?
- Marginal:
  - How probable is the new evidence under all possible hypotheses?

$$P(A|B) = P(B|A) \cdot P(A) \,/\, P(B)$$

Posterior = Likelihood · Prior / Marginal

The Sneaky Marginal Probability

- Recall the Law of Total Probability:
  - $P(A) = P(A|B) \cdot P(A) + P(A|C) \cdot P(C) + \cdots + P(A|x) \cdot P(x)$

  - Recall that P(A) is also called the "marginal probability".

- Application of the Law of Total Probability in Bayes Rule':
  - $P(A|B) = P(B|A) \cdot P(A) / P(B|x_1) \cdot P(x_1) + P(B|x_2) \cdot P(x_2) + \cdots + P(B|x_i) \cdot P(x_i)$
  - Sometimes, the marginal probability will be given to you. Often, you will have to derive it from implicit information.

Class Challenge

Imagine that there's a rare disease that affects about 0.02% of the population. The test for this disease correctly detects the presence of the disease 99.5% of the time - that is, if someone is infected, there is a 99.5% chance the test will come back positive - the True Positive Rate. There is also a False Positive Rate - if someone does not have the disease, there is a 10% chance the test will return a positive result anyway.

Your patient's test comes back positive. What is the probability that your patient has the disease?

Class Challenge Answer

- $P(\text{disease}) = 0.0002$
- $P(\text{no disease})$:
  - $1 - P(\text{disease})$
  - $1 - 0.0002$
  - $\approxeq 0.9998$
- $P(+ \mid \text{no disease}) = 0.10$
- $P(+ \mid \text{disease}) = 0.995$
- $P(+)$:
  - $P(+ \mid \text{disease}) \cdot P(\text{disease})$

    $+ P(+ \mid \text{no disease}) \cdot P(\text{no disease})$
  - $(0.995 \cdot 0.0002) + (0.10 \cdot 0.9998)$
  - $\approxeq 0.100179$

$$P(\text{disease}|+) = P(+|\text{disease}) \cdot P(\text{disease}) / P(+)$$

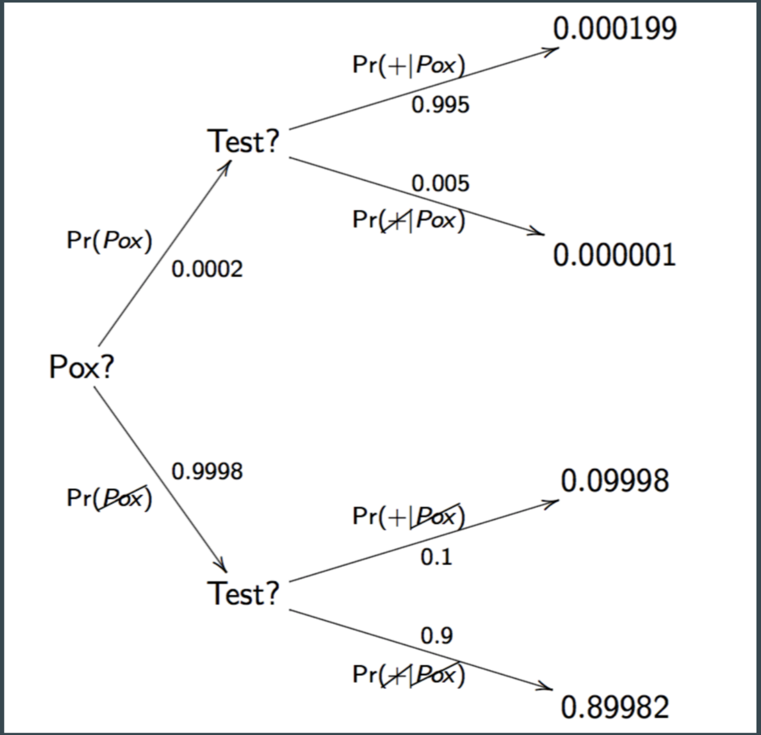
$$P(\text{disease}|+) = (0.995 \cdot 0.0002) / 0.101179$$

$$P(\text{disease}|+) \approxeq 0.002\%$$

Probability Trees (Bayes for your Eyes)

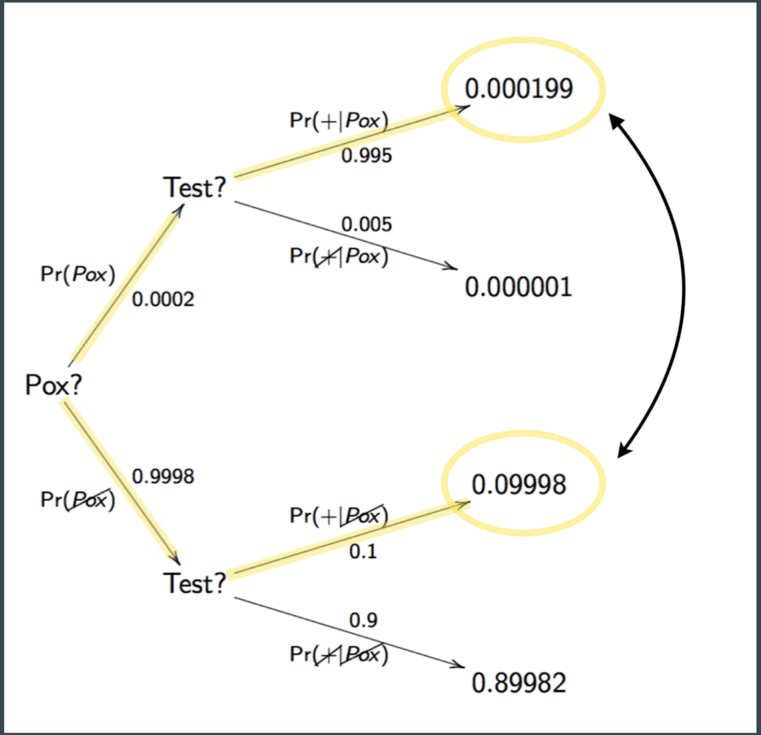
- P(disease) = 0.0002
- P(no disease):
  - 1 - P(disease)
  - 1 - 0.0002
  - ≈ 0.9998
- P(+ | no disease) = 0.10
- P(+ | disease) = 0.995
- P(+):
  - P(+ | disease) ・ P(disease)

    + P(+ | no disease) ・ P(no disease)
  - (0.995 ・ 0.0002) + (0.10 ・ 0.9998)
  - ≈ 0.100179



Bayes for your Eyes Continued

- Follow the paths that are relevant for the problem at hand.
    - In this case, all the paths that produce a positive test result.
- Multiply down the tree. Sum the terminal nodes.
- Compute P(disease|+):
    - Add the nodes that meet the positive test result condition. This is the marginal probability.
    - The likelihood and prior were given, so just complete the calculation.



Some Terminology

- Precision, False Positive Rate:
  - What proportion of positive identifications was actually correct?
  - TP / (TP + FP)
  - TP / Total number *predicted* positive
- Recall, True Positive Rate, Sensitivity, Probability of Detection:
  - What proportion of actual positives was identified correctly?
  - TP / (TP + FN)
  - TP / Total number *actually* positive
- Accuracy:
  - The ratio of correctly predicted observations.
  - TP + TN / TP + FP + FN + TN

Class Discussion

When would you optimize for Recall? What about Precision?

Introduction to Probability
Lessons 11 - 12

Erin Desmond, Sean Costello, Alicia Key

Class Probability Review

- Concrete
  - What is the probability of rolling a die and getting a 5?
  - What is the probability of rolling a die and getting a 5, **and then getting a 6**?
- Abstract
  - What is the probability of drawing one specific observation out of data set of size *N?*
  - What is the probability of **not** drawing that one specific observation out of a data set of size *N?*
- Little Harder
  - What is the probability of drawing a king out of a deck of cards **after I have drawn 2 kings and didn't replace them?**
  - What is the probability of drawing a king out of a deck of cards **after I have drawn 3 queens and didn't replace them?**
- Here we go...
  - What is the probability of drawing a 2 ace's back-to-back out of the same deck and then rolling "snake eyes?" (a one followed by a one)?

Probability Review Answers

- Concrete
  - What is the probability of rolling a die and getting a 5? 1/6
  - What is the probability of rolling a die and getting a 5, **and then getting a 6**? (⅙) • (⅙) = 1/36
- Abstract
  - What is the probability of drawing one specific observation out of data set of size *N*? 1/N
  - What is the probability of **not** drawing that one specific observation out of a data set of size *N*? (N - 1) / N
- Little Harder
  - What is the probability of drawing a king out of a deck of cards **after I have drawn 2 kings and didn't replace them?** 2/50 = 1/25
  - What is the probability of drawing a king out of a deck of cards **after I have drawn 3 queens and didn't replace them?** 4/49
- Here we go...
  - What is the probability of drawing 2 ace's back-to-back out of the same deck and then rolling "snake eyes?" (a one followed by a one)? (4/52) • (3/51) • (⅙) • (⅙) ≈ 0.00012563

Bayes' Rule Review

- The prevalence of a genetic condition in the population is 0.1%. The probability that someone with the condition tests positive is 93%. I take the test and get a positive result (bad news). What is the probability that I have the genetic condition?

- I have a bag with two coins in it; one coin is normal (i.e. has a 50% chance of heads and a 50% chance of tails) and one that is biased towards heads (there is a 75% chance that it lands on heads). I pull one coin out, flip it and get heads. What is the probability that the coin I flipped was the fair coin?

Bayes' Rule Review Answers

$$P(D) = 0.001,\ therefore\ P(\bar{D}) = 0.999$$

$$P(+|D) = 0.93,\ therefore\ P(+|\bar{D}) = 0.07$$

$$
\begin{aligned}
P(D|+) &= \frac{P(+|D)P(D)}{P(+)} \\
&= \frac{P(+|D)P(D)}{P(+|given\ any\ condition)} \\
&= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|\bar{D})P(\bar{D})} \\
&= \frac{(0.93 \cdot 0.001)}{(0.93 \cdot 0.001) + (0.07 \cdot 0.999)}
\end{aligned}
$$

$$P(D|+) = 0.0131$$

$$P(F) = 0.5,\ therefore\ P(\bar{F}) = 0.5$$

$$P(H|F) = 0.5,\ and\ P(H|\bar{F}) = 0.75$$

$$
\begin{aligned}
P(F|H) &= \frac{P(H|F)P(F)}{P(H)} \\
&= \frac{P(H|F)P(F)}{P(H|given\ any\ condition)} \\
&= \frac{P(H|F)P(F)}{P(H|F)P(F) + P(H|\bar{F})P(\bar{F})} \\
&= \frac{(0.5 \cdot 0.5)}{(0.5 \cdot 0.5) + (0.75 \cdot 0.5)}
\end{aligned}
$$

$$P(F|H) = 0.4$$

Learning Objectives

- Define, write the formula, and calculate:
  - Factorials
  - Permutations
  - Combinations
- Attempt factorials, permutations, and combinations in Python
- Review the Multiplication Rule
- Define Independent and Identically Distributed
- Define Cardinality

Combinatorics

A field of mathematics concerned with counting the possible orderings and combinations of elements within a set.

The following are incorporated into equations to solve ordering problems:

- Factorials
- Permutations
- Combinations

Factorial

- Represents the number of possible orderings of n items.
- Represented by an exclamation mark, !.
- Factorial of a positive integer n is the product of n and all positive integers less than n.
- Ex: Imagine you are drawing 4 lottery balls, how many possible orderings of these balls can you choose?
  - $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$
- Zero is a special case:
  - $0! = 1$

Permutations

- The number of ways to order a set of things, when order is important.
- Ex: Randomly select 3 different (unique) letters from the alphabet. X-B-T would be a different permutation than T-B-X.
- Formula:
  - nPk = n! / (n - k)!
  - Where ***n*** is the number of things you'd like to order and ***k*** is the total number of items.
- Class Challenge:
  - From 10 different lottery balls, how many orderings are possible if you draw 4 of them without replacement? The winning ticket must match the 4 numbers in the order in which you drew them.

Permutations

- The number of ways to order a set of things, when order is important.
- Formula:
  - $nPk = n! / (n - k)!$
  - Where n is the total number of items and k is the number of things you'd like to order.
- Class Challenge:
  - From 10 different lottery balls, how many orderings are possible if you draw 4 of them without replacement? The winning ticket must match the 4 numbers in the order in which you drew them.
  - ${}_{10}P_4 = 10! / (10 - 4)!$
  - ${}_{10}P_4 = (10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1) / (6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)$

    $= 3{,}628{,}800 / 720 = 5{,}040$

Combinations

- The number of ways to order a set of things, when order is ***not*** important.
- Formula:
  - nCk = n! / ((n - k)! • k!)
  - Where ***n*** is the number of things you'd like to order and ***k*** is the total number of items.
- Class Challenge:
  - From 10 different lottery balls, how many orderings are possible if you draw 4 of them without replacement? The winning ticket allows for any order of lottery balls.

Combinations

- The number of ways to order a set of things, when order is ***not*** important.
- Formula:
  - nCk = n! / ((n - k)! • k!)
  - Where ***n*** is the number of things you'd like to order and ***k*** is the total number of items.
- Class Challenge:
  - From 10 different lottery balls, how many orderings are possible if you draw 4 of them without replacement? The winning ticket allows for any order of lottery balls.
  - $_{10}C_4$ = 10! / ((10 - 4)! • 4!)
  - $_{10}C_4$ = (10 • 9 • 8 • 7 • 6 • 5 • 4 • 3 • 2 • 1) / ((6 • 5 • 4 • 3 • 2 • 1) • (4 • 3 • 2 • 1))

    = 3,628,800 / 17,280 = 210

Factorials, Permutations, Combinations in Python

```python
def factorial(n):

    if n == 0:
        return 1
    else:
        return n * factorial(n-1)

def permutation(n,k):

    numer = factorial(n)
    denom = factorial(n-k)

    return numer / denom
```

```python
def combination(n,k):

    numer = factorial(n)
    denom_1 = factorial(n-k)
    denom_2 = factorial(k)

    return numer / (denom_1 * denom_2)

if __name__ == '__main__':

    n = 10
    k = 4

    print(factorial(n))
    print(permutation(n,k))
    print(combination(n,k))
```

Multiplication Rule Review

- Recall: When we calculate probabilities involving one event AND another event, we multiply their probabilities.
  - Dependent Events: First event impacts the probability of the second event.
  - Independent Events: First event has no impact on the probability of the second event.
- How many four-letter strings can you make out of the 26 letters of the English alphabet?
  - 26 options for the 1st character, 26 options for the 2nd character, 26 options for the 3rd character...and so on. Therefore:
    - 26^4
    - Notice the number of options didn't change as the process progressed...this is because...

Independent and Identically Distributed

- Independent and Identically Distributed:
  - Randomly choosing a letter from the alphabet is unaffected by the previous or subsequent letters drawn. Independent.
  - The options available from one step to the next do not change. Identical.
  - This means that the randomness in question is independent and identical.
- Each random variable has the same probability distribution as the others and are all mutually independent. (Don’t worry, more on this in stats.)
- In many situations, i.i.d. is often assumed, but is not necessarily a reality. Checks exist to test i.i.d of a set of events.

Independent and Identically Distributed Examples

- Some examples of when i.i.d is assumed:
  - Each roll of a dice
  - Each time a deck of cards is dealt
  - Flipping coins
- Thought experiment:
  - Imagine giving a test to children to assess abilities. The test would only work if they took it independently. If they interacted with each other, you'd probably measure the abilities of the most clever child, or the loudest child. The children also need to be identically distributed: all the children would need to come from the same country, speak the same language, be the same age, come from the same socio-economic background...
  - You see how complicated it can be to find data that is i.i.d. Data Scientists use various tools to account for these variables.

Cardinality

- All probability boils down to the notion of set size, which is a count.
- You make it a probability by dividing your set size (number of possible outcomes) by the total number of outcomes in the whole sample space S_X.
- Cardinality:
  - Number of elements in a set.
  - Denoted as:
    - $\text{card}(E_1) = |E_1|$

Summary

- Independent and Identically Distributed:
    - All events provide the same kind of information independently of each other.
- i.i.d is often assumed to be true.
- Events that influence each other give us some of the same information, therefore when attempting to use a set of events to make a prediction, we should check to make sure that our events are i.i.d. At the very least, we should be aware that we are making that assumption.
- Cardinality is a measure of the number of elements in a set.
    - $card(E_1) = |E_1|$

Class Question

- In your own words, describe Independent and Identically Distributed Events.
- Name some examples of when i.i.d is assumed.
- Why is i.i.d data important in the context of making predictions?