Spring 2019

Lecture: Probability and Statistics - Basic Probability I - Week Three

Evan Agovino
CUNY City College

NYC Tech-in-Residence Corps

Week Three: Basic Probability I

CS 217

Statistics vs. Probability

- **Probability** deals with the likelihood of predicting future events, while **statistics** involves the analysis of the frequency of past events
- Say we have a jar of red and green jelly beans
- A **probablist** will know the proportion of red and green jelly beans in the jar and ask the probability of drawing a red jelly bean
- A **statistician** will infer the proportion of red jelly beans in the jar by sampling from the jar
- That statistician will use probabilistic concepts to make their inference
- **Probability** is to **Statistics** as **Calculus** is to **Physics**
- To be an effective statistician (or data scientist), you must have an understanding of the essential probability concepts, models and distributions

A Basic Example

- We have a ('fair') coin. What is the probability that we flip it and it lands on heads?

A Basic Example

- We have a ('fair') coin. What is the probability that we flip it and it lands on heads?
  - There are two possibilities, heads or tails.
  - The desired outcome is heads.
  - The probability that we flip it and it lands on heads is **1 out of 2**, or **½**, as the probability of getting heads or tails is equally likely.

A Basic Example

- We have a ('fair') coin. What is the probability that we flip it and it lands on heads?
  - There are two possibilities, heads or tails.
  - The desired outcome is heads.
  - The probability that we flip it and it lands on heads is **1 out of 2**, or **½**, as the probability of getting heads or tails is equally likely.
- **Sample Space**: Heads, Tails
- **Event**: Heads
- **Probability Function:** P(H) = .5, P(T) = .5
- **Probability of Our Event**: 0.5

Definitions

- **Sample Space**: the set of all possible outcomes
- **Event**: a subset of the sample space
- **Probability Function**: a function giving the probability for each outcome

Rules

- All probabilities are between 0 and 1 (or between 0% and 100%)
- The sum of all probabilities of all possible outcomes is 1 (or 100%)

A Basic Example

- We have a (‘fair’) coin. What is the probability that we flip it three times and it lands on heads once?

A Basic Example

- We have a (‘fair’) coin. What is the probability that we flip it three times and it lands on heads once?
- What is the sample space for this problem?

A Basic Example

- We have a (‘fair’) coin. What is the probability that we flip it three times and it lands on heads once?
- What is the sample space for this problem?
  - Sample Space: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT

A Basic Example

- We have a ('fair') coin. What is the probability that we flip it three times and it lands on heads once?
- What is the sample space for this problem?
  - Sample Space: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
- What is the event for this problem?

A Basic Example

- We have a (‘fair’) coin. What is the probability that we flip it three times and it lands on heads once?
- What is the sample space for this problem?
  - Sample Space: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
- What is the event for this problem?
  - Event: HTT, THT, TTH

A Basic Example

- We have a (‘fair’) coin. What is the probability that we flip it three times and it lands on heads once?
- What is the sample space for this problem?
  - Sample Space: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
- What is the event for this problem?
  - Event: HTT, THT, TTH
- What is the probability function for this problem?

A Basic Example

- We have a ('fair') coin. What is the probability that we flip it three times and it lands on heads once?
- What is the sample space for this problem?
  - Sample Space: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
- What is the event for this problem?
  - Event: HTT, THT, TTH
- What is the probability function for this problem?
  - Each outcome is equally likely with a probability of ⅛

A Basic Example

- We have a ('fair') coin. What is the probability that we flip it three times and it lands on heads once?
- What is the sample space for this problem?
  - Sample Space: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
- What is the event for this problem?
  - Event: HTT, THT, TTH
- What is the probability function for this problem?
  - Each outcome is equally likely with a probability of ⅛
- What is the probability of our event occurring?

A Basic Example

- We have a ('fair') coin. What is the probability that we flip it three times and it lands on heads once?
- What is the sample space for this problem?
    - Sample Space: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
- What is the event for this problem?
    - Event: HTT, THT, TTH
- What is the probability function for this problem?
    - Each outcome is equally likely with a probability of $\frac{1}{8}$
- What is the probability of our event occurring?
    - The probability of our event occurring is $\frac{3}{8}$.

A Basic Example

- Let’s say we have a set of 10 animals:
  - Antelope, Bee, Cat, Dog, Elephant, Frog, Gnat, Hyena, Iguana, Jaguar
- From this set of 10 animals, we have a **subset** of six mammals:
  - Antelope, Cat, Dog, Elephant, Hyena, Jaguar
- From this set of 10 animals, we have a **subset** of eight animals who live in the wild:
  - Antelope, Bee, Elephant, Frog, Gnat, Hyena, Iguana, Jaguar
- From this set of 10 animals, how many wild mammals are there?
- From this set of 10 animals, how many animals are **either** wild **or** a mammal?

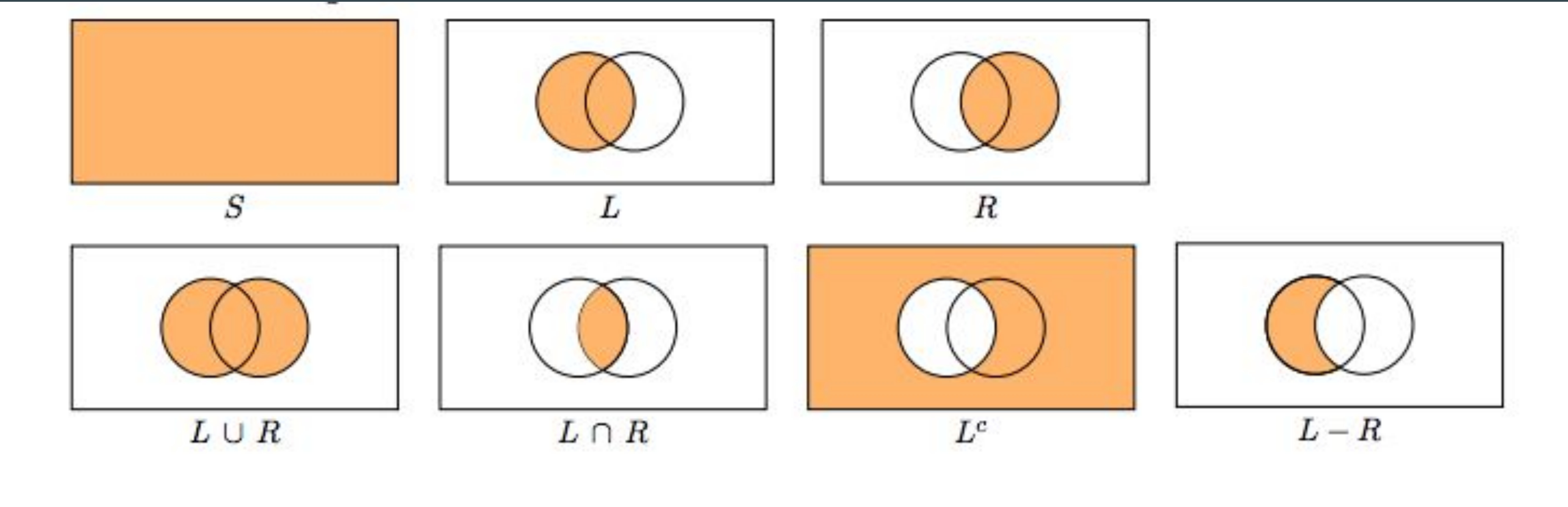
A Basic Example

- Let's say we have a set of 10 animals:
  - Antelope, Bee, Cat, Dog, Elephant, Frog, Gnat, Hyena, Iguana, Jaguar
- From this set of 10 animals, we have a **subset** of six mammals:
  - Antelope, Cat, Dog, Elephant, Hyena, Jaguar
- From this set of 10 animals, we have a **subset** of eight animals who live in the wild:
  - Antelope, Bee, Elephant, Frog, Gnat, Hyena, Iguana, Jaguar
- From this set of 10 animals, how many wild mammals are there?
  - There are **four** wild mammals: Antelope, Elephant, Hyena, Jaguar
- From this set of 10 animals, how many animals are **either** wild **or** a mammal?
  - All **ten** animals are either wild or a mammal. There are eight animals who live in the wild, six mammals, and four wild mammals, or 8 + 6 - 4 = 10

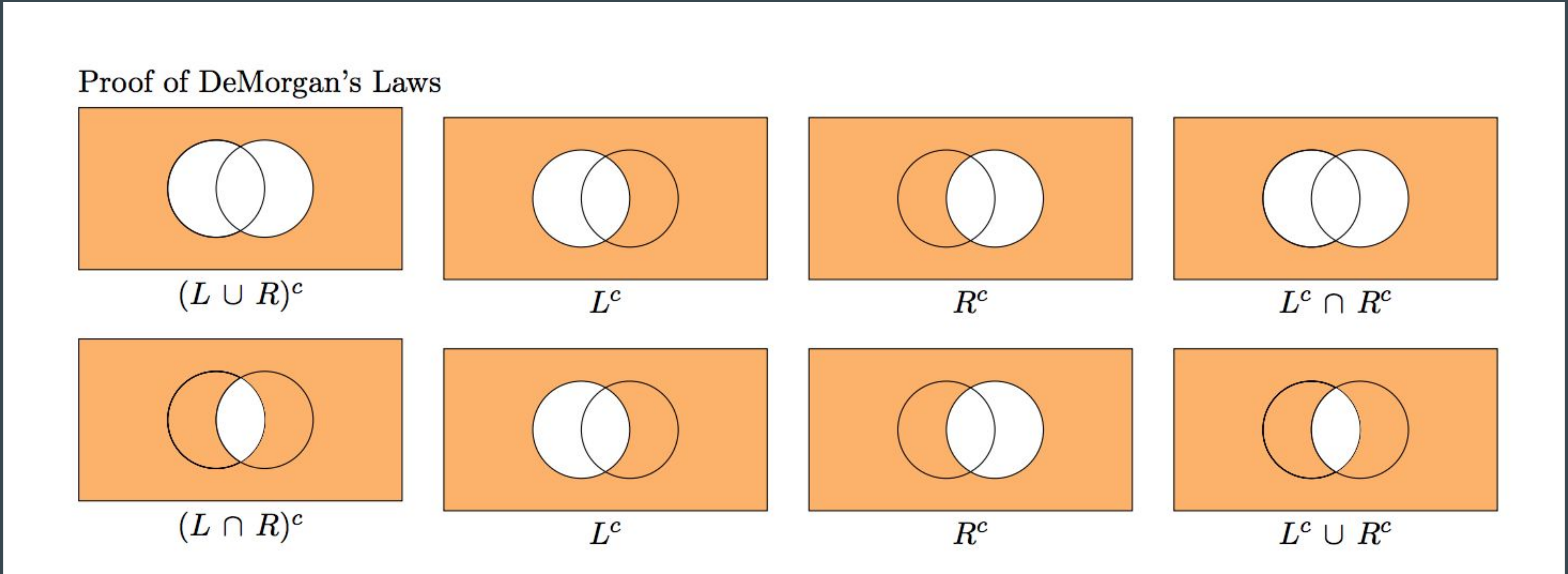
Definitions

- Intersection: All animals who are both **wild** and **mammals**:
  - Antelope, Elephant, Hyena, Jaguar
- Union: All animals who are either **wild** or **mammals**:
  - Antelope, Bee, Cat, Dog, Elephant, Frog, Gnat, Hyena, Iguana, Jaguar
- Complement: All animals who are **not wild**:
  - Dog, Cat
- Difference: All animals who are **wild**, but not **mammals**
  - Bee, Frog, Gnat, Iguana

Venn Diagrams



Venn Diagrams



Definitions

- **Inclusion-Exclusion Principle**: The probability of the union of two sets is equal to the probability of the first set + the probability of the second set - the probability of the **intersection** of the two sets

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- To reiterate our previous example, the probability of an animal either being from the wild or being a mammal is equal to the probability of an animal being from the wild plus the probability of an animal being a mammal minus the probability of an animal being both a mammal and from the wild.

Your Turn

- Say I flip three coins. What is the probability that the first coin is a head or that two of the three coins are heads?

Your Turn

- Say I flip three coins. What is the probability that the first coin is a head or that two of the three coins are heads?
- $P(H_1) = ½$
    - Sample Space: [HHH, HHT, HTH, THH, HTT, THT, TTH, TTT]
    - Event: [HHH, HHT, HTH, HTT]

Your Turn

- Say I flip three coins. What is the probability that the first coin is a head or that two of the three coins are heads?
- $P(H_1) = ½$
  - Sample Space: [HHH, HHT, HTH, THH, HTT, THT, TTH, TTT]
  - Event: [HHH, HHT, HTH, HTT]
- $P(2H) = ⅜$
  - Sample Space: [HHH, HHT, HTH, THH, HTT, THT, TTH, TTT]
  - Event: [HHT, HTH, THH]

Your Turn

- Say I flip three coins. What is the probability that the first coin is a head or that two of the three coins are heads?
- $P(H_1) = ½$
  - Sample Space: [HHH, HHT, HTH, THH, HTT, THT, TTH, TTT]
  - Event: [HHH, HHT, HTH, HTT]
- $P(2H) = ⅜$
  - Sample Space: [HHH, HHT, HTH, THH, HTT, THT, TTH, TTT]
  - Event: [HHT, HTH, THH]
- $P(H \& 2H) = ¼$
  - Sample Space: [HHH, HHT, HTH, THH, HTT, THT, TTH, TTT]
  - Event: [HHT, HTH]

Your Turn

- Say I flip three coins. What is the probability that the first coin is a head or that two of the three coins are heads?
- $P(H_1) = ½$
  - Sample Space: [HHH, HHT, HTH, THH, HTT, THT, TTH, TTT]
  - Event: [HHH, HHT, HTH, HTT]
- P(2H) = ⅜
  - Sample Space: [HHH, HHT, HTH, THH, HTT, THT, TTH, TTT]
  - Event: [HHT, HTH, THH]
- P(H & 2H) = ¼
  - Sample Space: [HHH, HHT, HTH, THH, HTT, THT, TTH, TTT]
  - Event: [HHT, HTH]
- P(H or 2H) = ⅝
  - Sample Space: [HHH, HHT, HTH, THH, HTT, THT, TTH, TTT]
  - Event: [HHH, HHT, HTH, HTT, THH]

Your Turn

- Say I flip three coins. What is the probability that the first coin is a head or that two of the three coins are heads?

$$\frac{4}{8} + \frac{3}{8} - \frac{2}{8} = \frac{5}{8}$$

Definitions

- Two sets are **disjoint** is they don't have any common elements
- Say we have a set of numbers between 3 and 100
- The probability that a number in that set is **even** and the probability that a number in that set is **prime** is disjoint, as no numbers in the set are both **prime** and **even** (2 is the only prime number that is even)
- What is the probability that a number in that set is prime or even?

Definitions

- Two sets are **disjoint** is they don't have any common elements
- Say we have a set of numbers between 3 and 100
- The probability that a number in that set is **even** and the probability that a number in that set is **prime** is disjoint, as no numbers in the set are both **prime** and **even** (2 is the only prime number that is even)
- What is the probability that a number in that set is prime or even?

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Definitions

- Two sets are **disjoint** is they don't have any common elements
- Say we have a set of numbers between 3 and 100
- The probability that a number in that set is **even** and the probability that a number in that set is **prime** is disjoint, as no numbers in the set are both **prime** and **even** (2 is the only prime number that is even)
- What is the probability that a number in that set is prime or even?

$$P(A \cup B) = P(A) + P(B)$$

Definitions

- **Rule of Product**: if there are A ways of doing something, and B ways of doing another thing, then there are A * B ways of doing both things
  - If I have 4 shirts, and 3 pants, I have 4*3=12 different potential outfits
  - Note that in this example all shirts can be worn with all pants
  - If I have a four-digit pin number, I have 10*10*10*10 = 10,000 different potential pin numbers
  - The rule of product is **with replacement**, since in this example we can use pin numbers repeatedly (your pin can be 0000)

Basic Probability

- **Permutation**: a selection of items **without replacement** where the order our items are selected in matters
- A permutation is a **rule of product** without replacement
  - Say we need to pick a 4-digit ATM pin number but we cannot use the same digit twice.
  - Our first digit can be any of the 10 available digits
  - Our second digit can be any of the 9 remaining digits
  - Our third digit can be any of the 8 remaining digits
  - Our fourth digit can be any of the 7 remaining digits
  - Thus there are 10 x 9 x 8 x 7 = 5,040 possible pin numbers

$$ {}_nP_k = \frac{n!}{(n-k)!} = n(n-1)\ldots(n-k+1) $$

Basic Probability

- **Combination**: a selection of items **without replacement** where the order our items are selected in doesn't matter
  - Say we want to get all four-digit pin combinations without replacement, but we do not care about the order in a given four-digit combination, i.e. we will treat 1234 and 1324 the same (for a permutation these will be considered two separate outcomes)
  - There are only 5040/24 = 210 possible outcomes

$$ {}_{n}C_{k} = \frac{n!}{k!(n-k)!} = \frac{{}_{n}P_{k}}{k!} $$

Basic Probability

- **Combination With Replacement**: a selection of items **with replacement** where the order our items are selected in doesn't matter
- Rarer than the rule of product, permutation, and combination
  - Say we want to get all four-digit pin combinations with replacement, but we do not care about the order in a given four-digit combination, i.e. we will treat 1234 and 1324 the same (for a permutation these will be considered two separate outcomes)
  - There are only 715 possible outcomes

$${}_{n+k-1}C_k = \frac{(n+k-1)!}{k!(n-1!)} = \frac{{}_{n+k-1}P_k}{k!}$$

Your Turn

- You’re on a strict diet and will only eat two of three meals in a day (breakfast, lunch, dinner).
- Say you can repeat the same meal twice per day. What are all the combinations of meals you can have?
- Say you can repeat the same meal twice per day. What are all the permutations of meals you can have?
- Say you cannot repeat the same meal twice per day. What are all the combinations of meals you can have?
- Say you cannot repeat the same meal twice per day. What are all the permutations of meals you can have?

- You’re on a strict diet and will only eat two of three meals in a day (breakfast, lunch, dinner).
- Say you can repeat the same meal twice per day. What are all the combinations of meals you can have?
    - Six combinations: BB, BL, BD, LL, LD, DD
- Say you can repeat the same meal twice per day. What are all the permutations of meals you can have?
    - Nine permutations: BB, BL, BD, LB, LL, LD, DB, DL, DD
- Say you cannot repeat the same meal twice per day. What are all the combinations of meals you can have?
    - Three combinations: BL, BD, LD
- Say you cannot repeat the same meal twice per day. What are all the permutations of meals you can have?
    - Six permutations: BL, BD, LB, LD, DB, DL

Full Circle

- What are all the combinations of ways you can get two heads in four flips?

Full Circle

- What are all the combinations of ways you can get two heads in four flips?
- This is simply a **combination** problem with a slightly different twist. It is a given that the actual combination of outcomes will be some permutation of [H,H,T], but the actual combinations are in the trials in which you will land on heads: i.e. First and Second Trial, First and Third Trial, Second and Third Trial
- It is **not** a permutation problem because we don't care about the order i.e. whether a **given** coin landed heads on the first or third trial, just that **a** coin landed then
- This is confusing and a bit counter-intuitive, try and sit with it to understand exactly what is going on
- This is the foundation of many of the discrete distributions we'll be working with later in the courses