Day Three: Probability

SOC Methods Camp

September 5th, 2019

Outline

- Introduction: Why Do We Need Probability?

Outline

- Introduction: Why Do We Need Probability?
- Counting and Sets

Outline

- Introduction: Why Do We Need Probability?
- Counting and Sets
  - Methods of Counting/Sampling

Outline

- Introduction: Why Do We Need Probability?
- Counting and Sets
  - Methods of Counting/Sampling
  - Sets and Operations

Outline

- Introduction: Why Do We Need Probability?
- Counting and Sets
  - Methods of Counting/Sampling
  - Sets and Operations
- Basic Probability

Outline

- Introduction: Why Do We Need Probability?
- Counting and Sets
  - Methods of Counting/Sampling
  - Sets and Operations
- Basic Probability
  - Probability Axioms

Outline

- Introduction: Why Do We Need Probability?
- Counting and Sets
  - Methods of Counting/Sampling
  - Sets and Operations
- Basic Probability
  - Probability Axioms
  - Calculations with Probabilities

Outline

- Introduction: Why Do We Need Probability?
- Counting and Sets
  - Methods of Counting/Sampling
  - Sets and Operations
- Basic Probability
  - Probability Axioms
  - Calculations with Probabilities
  - Law of Total Probability

Outline

Outline

- Introduction: Why Do We Need Probability?
- Counting and Sets
  - Methods of Counting/Sampling
  - Sets and Operations
- Basic Probability
  - Probability Axioms
  - Calculations with Probabilities
  - Law of Total Probability
- Conditional Probability
  - Conditional Probability

Outline

- Introduction: Why Do We Need Probability?
- Counting and Sets
  - Methods of Counting/Sampling
  - Sets and Operations
- Basic Probability
  - Probability Axioms
  - Calculations with Probabilities
  - Law of Total Probability
- Conditional Probability
  - Conditional Probability
  - Bayes' Rule

Outline

- Introduction: Why Do We Need Probability?
- Counting and Sets
  - Methods of Counting/Sampling
  - Sets and Operations
- Basic Probability
  - Probability Axioms
  - Calculations with Probabilities
  - Law of Total Probability
- Conditional Probability
  - Conditional Probability
  - Bayes' Rule
  - Independence

Why do we need probability?

- Social processes are not deterministic. The "effects" of social causes are difficult to isolate and estimate

Why do we need probability?

- Social processes are not deterministic. The "effects" of social causes are difficult to isolate and estimate
- We need a framework for communicating our uncertainty about the inferences we draw in our empirical work

Why do we need probability?

- Social processes are not deterministic. The "effects" of social causes are difficult to isolate and estimate
- We need a framework for communicating our uncertainty about the inferences we draw in our empirical work
- **Probability theory is the root of social statistics**

Why do we need probability?

- Social processes are not deterministic. The "effects" of social causes are difficult to isolate and estimate
- We need a framework for communicating our uncertainty about the inferences we draw in our empirical work
- **Probability theory is the root of social statistics**
- Being able to think rigorously and intuitively about uncertainty is also helpful for case selection and small-*n* inference in qualitative research

Counting and Sets

What is counting?

- In this section we aren't just thinking about counting how many apples are in a tree, or how many dogs are at a park

What is counting?

- In this section we aren't just thinking about counting how many apples are in a tree, or how many dogs are at a park
- Instead, we're interested in thinking about counting *possibilities*

What is counting?

- In this section we aren't just thinking about counting how many apples are in a tree, or how many dogs are at a park
- Instead, we're interested in thinking about counting *possibilities*
  - Homework example: how many ways to assemble a cohort from a pool of applicants?

What is counting?

- In this section we aren't just thinking about counting how many apples are in a tree, or how many dogs are at a park
- Instead, we're interested in thinking about counting *possibilities*
  - Homework example: how many ways to assemble a cohort from a pool of applicants?
  - Stats example: sampling! How many ways can we construct a sample of size *n* from a pool of data?

What is counting?

- In this section we aren't just thinking about counting how many apples are in a tree, or how many dogs are at a park
- Instead, we're interested in thinking about counting *possibilities*
  - Homework example: how many ways to assemble a cohort from a pool of applicants?
  - Stats example: sampling! How many ways can we construct a sample of size *n* from a pool of data?
- This will be useful for probability when we need to think about all the possibilities for something that could exist...

Methods of counting

Two things to consider when counting events/objects:

Methods of counting

Two things to consider when counting events/objects:

1. Does the order of occurrence matter?

Methods of counting

Two things to consider when counting events/objects:

1. Does the order of occurrence matter?
2. Are events/objects counted more than once?

Methods of counting

	Order Matters	**Order Doesn't Matter**
Counted more than once	Ordered with replacement	Unordered with replacement
Counted once	Ordered without replacement	Unordered without replacement

Methods of counting: ordered with replacement

- We have n objects and we want to pick $k < n$ from them, and replace the choice back into the available set of options each time.

Methods of counting: ordered with replacement

- We have n objects and we want to pick $k < n$ from them, and replace the choice back into the available set of options each time.
- How many possible ways to draw? We have n choices for k iterations: $n * n * ..n = n^k$

Methods of counting: ordered with replacement

- How might we do this in R?

Methods of counting: ordered with replacement

- How might we do this in R?
- Generate a vector of numbers from 1 to 1000 (*hint:* use the *seq* command). This is our **population**. Draw a **sample** of 100 from the population with the *sample* command ($n = 1000$ and $k = 100$).

```
population <- seq(1, 1000, by = 1)
ordered.replace <- sample(population, 100, replace = TRUE)
```

Methods of counting: ordered without replacement

- We have n objects and we want to pick $k < n$ from them, but we are not replacing objects back into the original set, meaning we have fewer choices to pick from in each iteration.

Methods of counting: ordered without replacement

- We have n objects and we want to pick $k < n$ from them, but we are not replacing objects back into the original set, meaning we have fewer choices to pick from in each iteration.
- How many possible ways to draw? We have n choices for the first object, *n-1* choices for the second object... and so on until we have k choices left:

$$n * (n-1) * (n-2) * ... * (k+1) * k = \frac{n!}{(n-k)!}$$

Methods of counting: ordered without replacement

- How might we do this in R?

Methods of counting: ordered without replacement

- How might we do this in R?
- Generate a vector of numbers from 1 to 1000 (*hint:* use the *seq* command). This is our **population**. Draw a **sample** of 100 from the population with the *sample* command ($n = 1000$ and $k = 100$).

```
ordered.no.replace <- sample(population, 100, replace = FALSE)
```

Methods of counting: unordered without replacement

- Just like ordered without replacement, except we can't see or don't care about the order of events (e.g. heads, tails = tails, heads)

Methods of counting: unordered without replacement

- Just like ordered without replacement, except we can't see or don't care about the order of events (e.g. heads, tails = tails, heads)
- How many possible ways to draw? We have $k!$ fewer choices than we did with ordered without replacement:

$$\frac{n!}{(n-k)!k!} = \binom{n}{k}$$

Methods of counting: unordered with replacement

- rare and pretty unintuitive – think about unordered without replacement (combinations) but then adjusted upwards to reflect increased number of choices in each iteration.

Methods of counting: unordered with replacement

- rare and pretty unintuitive – think about unordered without replacement (combinations) but then adjusted upwards to reflect increased number of choices in each iteration.
- How many possible ways to draw?

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

Sets and operations

Characteristics of a set

A **set** is a collection of objects.

Characteristics of a set

A **set** is a collection of objects.

Characteristics of a Set

Characteristics of a set

A **set** is a collection of objects.

Characteristics of a Set

- **Countability**: a countable set has elements that can be placed in one-to-one correspondence with positive integers

Characteristics of a set

A **set** is a collection of objects.

Characteristics of a Set

- **Countability**: a countable set has elements that can be placed in one-to-one correspondence with positive integers
- **Finiteness**: a finite set has finite number of contained events.

Characteristics of a set

A **set** is a collection of objects.

Characteristics of a Set

- **Countability**: a countable set has elements that can be placed in one-to-one correspondence with positive integers
- **Finiteness**: a finite set has finite number of contained events.
- A set can be countably finite, countably infinite, or uncountably infinite

Characteristics of a set

Characteristics of a Set

Characteristics of a set

Characteristics of a Set

- **Cardinality**: the number of elements in the set

Characteristics of a set

Characteristics of a Set

- **Cardinality**: the number of elements in the set
- **Empty Set**: a set with no elements ($\emptyset = \{ \quad \}$)

Operations

Operations compare, contrast, construct, or identify sets.

Operations

Operations compare, contrast, construct, or identify sets. **Operations**

Operations

Operations compare, contrast, construct, or identify sets. **Operations**

- **Subset**: A set that is composed entirely of elements of another set (e.g. Set A is subset of Set B if every element of A was also an element of B – Set B *contains* Set A)

Operations

Operations compare, contrast, construct, or identify sets. **Operations**

- **Subset**: A set that is composed entirely of elements of another set (e.g. Set A is subset of Set B if every element of A was also an element of B – Set B *contains* Set A)
- **Union**: The union of two sets contains all the elements that belong to either sets ($A \cup B = \{X|X \in A \textbf{ or } X \in B\}$)

Operations

Operations compare, contrast, construct, or identify sets. **Operations**

- **Subset**: A set that is composed entirely of elements of another set (e.g. Set A is subset of Set B if every element of A was also an element of B – Set B *contains* Set A)
- **Union**: The union of two sets contains all the elements that belong to either sets ($A \cup B = \{X | X \in A \textbf{ or } X \in B\}$)
- **Intersection**: The intersection of two sets contains only those elements found in both sets ($\{A \cap B = \{X | X \in A \textbf{ and } X \in B\}$)

Operations

Operations compare, contrast, construct, or identify sets. **Operations**

- **Subset**: A set that is composed entirely of elements of another set (e.g. Set A is subset of Set B if every element of A was also an element of B – Set B *contains* Set A)
- **Union**: The union of two sets contains all the elements that belong to either sets ($A \cup B = \{X|X \in A \textbf{ or } X \in B\}$)
- **Intersection**: The intersection of two sets contains only those elements found in both sets ($\{A \cap B = \{X|X \in A \textbf{ and } X \in B\}$)
- **Complement**: The complement of a given set is the set that contains all elements not in the original set ($A^C = \{X \in \Omega | X \notin A\}$)

Characteristics of multiple sets

Characteristics of multiple sets

Characteristics of multiple sets

Characteristics of multiple sets

- **Disjoint**: Two sets are disjoint when their intersection is empty (*note*: complements are by definition disjoint)

Characteristics of multiple sets

Characteristics of multiple sets

- **Disjoint**: Two sets are disjoint when their intersection is empty (*note*: complements are by definition disjoint)
- **Mutually Exclusive**: When k sets are all pairwise disjoint with each other

Basic probability

Axioms for a probability function

A probability function maps defined event(s) onto a metric in the interval of [0:1]. It enables us to discuss various *degrees of likelihood of occurrence* in a systematic way. It satisfies three axioms:

Axioms for a probability function

A probability function maps defined event(s) onto a metric in the interval of [0:1]. It enables us to discuss various *degrees of likelihood of occurrence* in a systematic way. It satisfies three axioms:

1. **Nonnegativity**: for any event A, $P(A) \geq 0$

Axioms for a probability function

A probability function maps defined event(s) onto a metric in the interval of [0:1]. It enables us to discuss various *degrees of likelihood of occurrence* in a systematic way. It satisfies three axioms:

1. **Nonnegativity**: for any event A, $P(A) \geq 0$
2. **Normalization**: $P(\mathbf{S}) = 1$

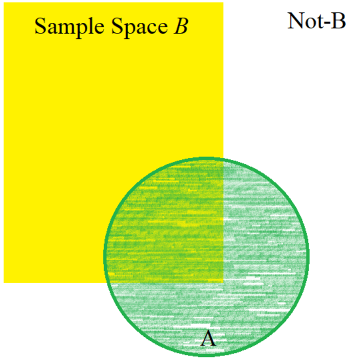
Axioms for a probability function

A probability function maps defined event(s) onto a metric in the interval of [0:1]. It enables us to discuss various *degrees of likelihood of occurrence* in a systematic way. It satisfies three axioms:

1. **Nonnegativity**: for any event A, $P(A) \geq 0$
2. **Normalization**: $P(\mathbf{S}) = 1$
3. **Additivity**: the probability of unions of n mutually exclusive events is the sum of their individual probabilities: $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$

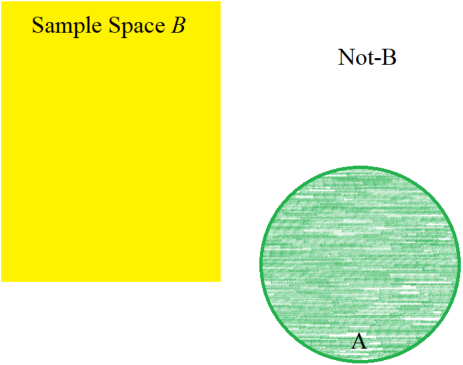
Law of total probability

The probability of event A can be always be decomposed into two parts: one that intersects with sample space B and one that does not.



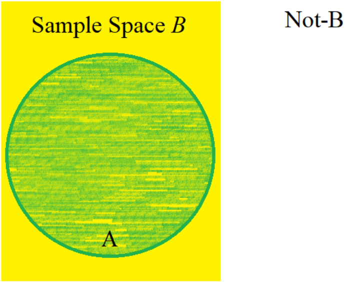
Law of total probability

The probability of event A can be always be decomposed into two parts: one that intersects with sample space B and one that does not.



Law of total probability

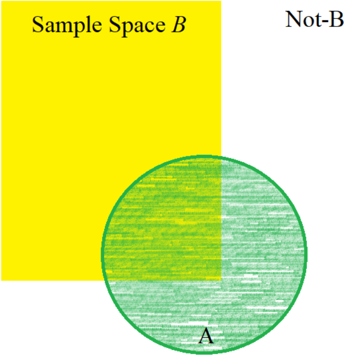
The probability of event A can be always be decomposed into two parts: one that intersects with sample space B and one that does not.



Law of total probability

Formally:

$$P(A) = P(A \cap B) + P(A \cap B^C)$$



Conditional probability, Bayes' Rule, and Independence

Conditional probability

- We think about outcomes of experiments/processes differently if we have partial information about them versus no information at all. Conditional probability statements recognize that some prior information bears on the determination of subsequent probabilities.

Conditional probability

- We think about outcomes of experiments/processes differently if we have partial information about them versus no information at all. Conditional probability statements recognize that some prior information bears on the determination of subsequent probabilities.
- For example: Incumbent senator Debbie Stabenow (D) is up for re-election next year in Michigan. Her probability of reelection can be expressed as $P(A)$. That probability depends on the winner of the GOP primary. One of the GOP candidates in the race so far is musician Kid Rock. Let's express the event of Kid Rock winning the GOP primary as B. So $P(A|B)$ would express the probability of Debbie Stabenow being reelected **conditional** on Kid Rock winning the GOP primary.

Conditional probability

The probability of a Debbie Stabenow reelection given a Kid Rock victory can be calculated as the probability of Debbie Stabenow winning the reelection AND Kid Rock winning the primary divided by the probability that Kid Rock wins the primary:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

which can be rearranged as:

$$P(A|B) * P(B) = P(A \cap B)$$

Conditional probability

We can also use conditional probability statements to express the Law of Total Probability:

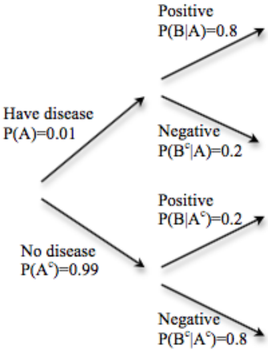
$$\text{LoTP: } P(A) = P(A \cap B) + P(A \cap B^C)$$

$$\text{Conditional Prob: } P(A|B) * P(B) = P(A \cap B)$$

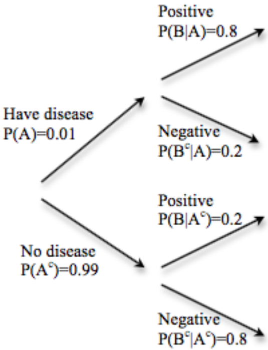
$$\text{Combined: } P(A) = P(A|B) * P(B) + P(A|B^C) * P(B^C)$$

Bayes' Rule

New example: An individual who recently traveled to South America either catches Zika or does not. They are tested for the disease and receive either a positive or negative test result. The test is imperfect though, so people who have the disease will sometimes test negative and vice versa.

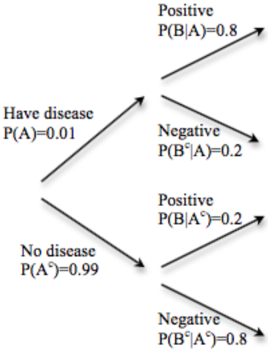


Bayes' Rule



So say you test positive for Zika. What is the probability that you are *actually* sick?

Bayes' Rule



We know the probability of having the disease ($P(A)$). We also know the probability of getting a positive test result given that you have the disease ($P(B|A)$), and of getting a false positive ($P(B|A^C)$). But what we *want* to know is the probability of having the disease given a positive result ($P(A|B)$).

Bayes' Rule

We don't know $P(A|B)$, but we have $P(A)$ and $P(B|A)$, as well as all varieties of their complements.

AND we know conditional probability expressions showed us that

$$P(A \cap B) = P(A) * P(B|A)$$

and

$$P(A \cap B) = P(B) * P(A|B)$$

because

$$P(A \cap B) = P(B \cap A)$$

Bayes' Rule

This means we can rearrange to:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) * P(A)}{P(B)}$$

This is Bayes' Rule.

Too bad we still don't know $P(B)$...

Bayes' Rule

But wait! The Law of Total Probability tells us that

$$P(B) = (P(B|A) * P(A)) + (P(B|A^C) * P(A^C))$$

So plugging that back into Bayes' Rule gives us:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B|A) * P(A) + (P(B|A^C) * P(A^C)}$$

Now solve on worksheet – what's the probability of actually having Zika given a positive test? What's the probability of actually *not* having Zika given a negative test?

Bayes' Rule: Prosecutor's Fallacy

One surprising place where Bayes' Rule is highly important is the courtroom. Let's imagine that there's been a murder in LA, a city that has just under 4 million people (we'll say it's 4 million on the nose). One person out of the 4 million in the city is guilty of that murder. 3,999,999 people out of 4 million are innocent. But guess what? There's DNA evidence! A person is brought to court, where the prosecutor explains that their DNA matches the DNA found at the crime scene. According to him, there's only a $1/1{,}000{,}000$ chance of an innocent person matching the DNA. *Obviously* that means this person is guilty, because given that this person matched there's only a $1/1{,}000{,}000$ chance they're innocent!

Right?

Bayes' Rule: Prosecutor's Fallacy

One surprising place where Bayes' Rule is highly important is the courtroom. Let's imagine that there's been a murder in LA, a city that has just under 4 million people (we'll say it's 4 million on the nose). One person out of the 4 million in the city is guilty of that murder. 3,999,999 people out of 4 million are innocent. But guess what? There's DNA evidence! A person is brought to court, where the prosecutor explains that their DNA matches the DNA found at the crime scene. According to him, there's only a $1/1{,}000{,}000$ chance of an innocent person matching the DNA. *Obviously* that means this person is guilty, because given that this person matched there's only a $1/1{,}000{,}000$ chance they're innocent!

Right?

Wrong. This is known as the prosecutor's fallacy.

Bayes' Rule: Prosecutor's Fallacy

The prosecutor knows that the probability of a DNA match (P(Match)) given a person being innocent (P(Match | Innocent)) is 1/1,000,000.

But by saying that this means there is only a 1/100,000,000 chance the defendant is innocent, he is claiming that P(Match|Innocent) = P(Innocent|Match).

Bayes' Rule: Prosecutor's Fallacy

So what we want to know is P(Innocent | Match). What we already know is:

$$P(\text{Match} \mid \text{Innocent}) = 1/1,000,000$$

$$P(\text{Innocent}) = 399,999/4,000,000$$

$$P(\text{Guilty}) = 1/4,000,000$$

$$P(\text{Match} \mid \text{Guilty}) = 1 \text{ (assuming that's the murderer's DNA)}$$

Given the above, what is the true probability that a defendant is innocent given that their DNA was a match?

Independence

Sometimes we have information about the outcome of event A but it doesn't change the probability of event B happening (e.g. – knowing that today is Thursday does not help predict whether I will get heads or tails when I flip a coin). This is the intuitive description of **independence.**

Independence

Formally, A and B are independent if $P(A \cap B) = P(A) * P(B)$.

From that, we can also deduce that when A and B are independent:

$$P(A) = P(A|B)$$

and

$$P(B) = P(B|A)$$