Soc500: Applied Social Statistics
Week 1: Introduction and Probability

Brandon Stewart[1]

Princeton

September 14, 2016

[1] These slides are heavily influenced by Matt Blackwell, Adam Glynn and Matt Salganik. The spam filter segment is adapted from Justin Grimmer and Dan Jurafsky. Illustrations by Shay O'Brien.

Where We've Been and Where We're Going...

- Last Week
  - methods camp
  - pre-grad school life
- This Week
  - Wednesday
    - welcome
    - basics of probability
- Next Week
  - random variables
  - joint distributions
- Long Run
  - probability $\rightarrow$ inference $\rightarrow$ regression

Questions?

Welcome and Introductions

Welcome and Introductions

- Soc500: Applied Social Statistics

Welcome and Introductions

- Soc500: Applied Social Statistics
- I

Welcome and Introductions

- Soc500: Applied Social Statistics
- I
  - ...am an Assistant Professor in Sociology.

Welcome and Introductions

- Soc500: Applied Social Statistics
- I
  - ...am an Assistant Professor in Sociology.
  - ...am trained in political science and statistics

Welcome and Introductions

- Soc500: Applied Social Statistics
- I
  - ...am an Assistant Professor in Sociology.
  - ...am trained in political science and statistics
  - ...do research in methods and statistical text analysis

Welcome and Introductions

- Soc500: Applied Social Statistics
- I
  - . . . am an Assistant Professor in Sociology.
  - . . . am trained in political science and statistics
  - . . . do research in methods and statistical text analysis
  - . . . love doing collaborative research

Welcome and Introductions

- Soc500: Applied Social Statistics
- I
  - ...am an Assistant Professor in Sociology.
  - ...am trained in political science and statistics
  - ...do research in methods and statistical text analysis
  - ...love doing collaborative research
  - ...talk very quickly

Welcome and Introductions

- Soc500: Applied Social Statistics
- I
  - . . . am an Assistant Professor in Sociology.
  - . . . am trained in political science and statistics
  - . . . do research in methods and statistical text analysis
  - . . . love doing collaborative research
  - . . . talk very quickly
- Your Preceptors

Welcome and Introductions

- Soc500: Applied Social Statistics
- I
  - ...am an Assistant Professor in Sociology.
  - ...am trained in political science and statistics
  - ...do research in methods and statistical text analysis
  - ...love doing collaborative research
  - ...talk very quickly
- Your Preceptors
  - sage guides of all things

Welcome and Introductions

- Soc500: Applied Social Statistics
- I
  - ...am an Assistant Professor in Sociology.
  - ...am trained in political science and statistics
  - ...do research in methods and statistical text analysis
  - ...love doing collaborative research
  - ...talk very quickly
- Your Preceptors
  - sage guides of all things
  - Ian Lundberg

Welcome and Introductions

- Soc500: Applied Social Statistics
- I
  - ...am an Assistant Professor in Sociology.
  - ...am trained in political science and statistics
  - ...do research in methods and statistical text analysis
  - ...love doing collaborative research
  - ...talk very quickly
- Your Preceptors
  - sage guides of all things
  - Ian Lundberg
  - Simone Zhang

The Core Strategy

The Core Strategy

- Goal: get you ready to quantitative work

The Core Strategy

- Goal: get you ready to quantitative work
- First in a two course sequence $\rightsquigarrow$ replication project

The Core Strategy

- Goal: get you ready to quantitative work
- First in a two course sequence $\rightsquigarrow$ replication project (for graduate students, part of a longer arc)

The Core Strategy

- Goal: get you ready to quantitative work
- First in a two course sequence $\rightsquigarrow$ replication project (for graduate students, part of a longer arc)
- Difficult course but with many resources to support you.

The Core Strategy

- Goal: get you ready to quantitative work
- First in a two course sequence $\rightsquigarrow$ replication project (for graduate students, part of a longer arc)
- Difficult course but with many resources to support you.
- When we are done you will be able to teach yourself many things

Specific Goals

Specific Goals

- For the semester

Specific Goals

- For the semester
  - critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences

Specific Goals

- For the semester
  - critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
  - conduct, interpret, and communicate results from analysis using multiple regression

Specific Goals

- For the semester
  - critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
  - conduct, interpret, and communicate results from analysis using multiple regression
  - explain the limitations of observational data for making causal claims

Specific Goals

- For the semester
  - critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
  - conduct, interpret, and communicate results from analysis using multiple regression
  - explain the limitations of observational data for making causal claims
  - write clean, reusable, and reliable `R` code.

Specific Goals

- For the semester
  - critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
  - conduct, interpret, and communicate results from analysis using multiple regression
  - explain the limitations of observational data for making causal claims
  - write clean, reusable, and reliable R code.
  - feel empowered working with data

Specific Goals

Specific Goals

- For the year

Specific Goals

- For the year
    - conduct, interpret, and communicate results from analysis using generalized linear models

Specific Goals

- For the year
  - conduct, interpret, and communicate results from analysis using generalized linear models
  - understand the fundamental ideas of missing data, modern causal inference, and hierarchical models

Specific Goals

- For the year
  - conduct, interpret, and communicate results from analysis using generalized linear models
  - understand the fundamental ideas of missing data, modern causal inference, and hierarchical models
  - build a solid, reproducible research pipeline to go from raw data to final paper

Specific Goals

- For the year
  - conduct, interpret, and communicate results from analysis using generalized linear models
  - understand the fundamental ideas of missing data, modern causal inference, and hierarchical models
  - build a solid, reproducible research pipeline to go from raw data to final paper
  - provide you with the tools to produce your own research (e.g. second year empirical paper).

Why R?

- It's free

Why R?

- It's free
- It's extremely powerful, but relatively simple to do basic stats

Why R?

- It's free
- It's extremely powerful, but relatively simple to do basic stats
- It's the *de facto* standard in many applied statistical fields

Why R?

- It's free
- It's extremely powerful, but relatively simple to do basic stats
- It's the *de facto* standard in many applied statistical fields
    - great community support

Why R?

- It's free
- It's extremely powerful, but relatively simple to do basic stats
- It's the *de facto* standard in many applied statistical fields
    - great community support
    - continuing development

Why R?

- It's free
- It's extremely powerful, but relatively simple to do basic stats
- It's the *de facto* standard in many applied statistical fields
    - great community support
    - continuing development
    - massive number of supporting packages

Why R?

- It's free
- It's extremely powerful, but relatively simple to do basic stats
- It's the *de facto* standard in many applied statistical fields
    - great community support
    - continuing development
    - massive number of supporting packages
- It will help you do research

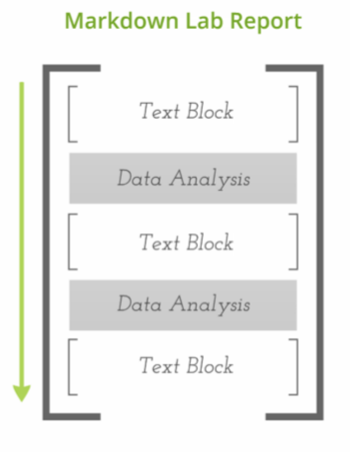
Why RMarkdown?

What you've done before

Baumer et al (2014)

Why RMarkdown?

RMarkdown



Baumer et al (2014)

Mathematical Prerequisites

Mathematical Prerequisites

- No formal pre-requisites

Mathematical Prerequisites

- No formal pre-requisites
- Balancing rigor and intuition

Mathematical Prerequisites

- No formal pre-requisites
- Balancing rigor and intuition
  - no rigor for rigor's sake

Mathematical Prerequisites

- No formal pre-requisites
- Balancing rigor and intuition
  - no rigor for rigor's sake
  - we will tell you *why* you need the math, but also feel free to ask

Mathematical Prerequisites

- No formal pre-requisites
- Balancing rigor and intuition
  - no rigor for rigor's sake
  - we will tell you *why* you need the math, but also feel free to ask
- We will teach you any math you need as we go along

Mathematical Prerequisites

- No formal pre-requisites
- Balancing rigor and intuition
  - no rigor for rigor's sake
  - we will tell you *why* you need the math, but also feel free to ask
- We will teach you any math you need as we go along
- Crucially though- this class is not about statistical aptitude, it is about effort

Ways to Learn

Ways to Learn

- Lecture
  learn broad topics

Ways to Learn

- Lecture
  learn broad topics
- Precept
  learn data analysis skills, get targeted help on assignments

Ways to Learn

- Lecture
  learn broad topics
- Precept
  learn data analysis skills, get targeted help on assignments
- Readings
  support materials for lecture and precept

Reading

- Required reading:

Reading

- Required reading:
  - Fox (2016) *Applied Regression Analysis and Generalized Linear Models*

Reading

- Required reading:
  - Fox (2016) *Applied Regression Analysis and Generalized Linear Models*
  - Angrist and Pischke (2008) *Mostly Harmless Econometrics*

Reading

- Required reading:
  - Fox (2016) *Applied Regression Analysis and Generalized Linear Models*
  - Angrist and Pischke (2008) *Mostly Harmless Econometrics*
  - Imai (2017) *A First Course in Quantitative Social Science**

Reading

- Required reading:
  - Fox (2016) *Applied Regression Analysis and Generalized Linear Models*
  - Angrist and Pischke (2008) *Mostly Harmless Econometrics*
  - Imai (2017) *A First Course in Quantitative Social Science**
  - Aronow and Miller (2017) *Theory of Agnostic Statistics**

Reading

- Required reading:
  - Fox (2016) *Applied Regression Analysis and Generalized Linear Models*
  - Angrist and Pischke (2008) *Mostly Harmless Econometrics*
  - Imai (2017) *A First Course in Quantitative Social Science**
  - Aronow and Miller (2017) *Theory of Agnostic Statistics**
- Suggested reading

Reading

- Required reading:
  - Fox (2016) *Applied Regression Analysis and Generalized Linear Models*
  - Angrist and Pischke (2008) *Mostly Harmless Econometrics*
  - Imai (2017) *A First Course in Quantitative Social Science**
  - Aronow and Miller (2017) *Theory of Agnostic Statistics**
- Suggested reading
- When and how to do the reading

Ways to Learn

- Lecture
  learn broad topics
- Precept
  learn data analysis skills, get targeted help on assignments
- Readings
  support materials for lecture and precept

Ways to Learn

- Lecture
  learn broad topics
- Precept
  learn data analysis skills, get targeted help on assignments
- Readings
  support materials for lecture and precept
- Problem Sets
  reinforce understanding of material, practice

Problem Sets

- Schedule (available Wednesday, due 8 days later at precept)

Problem Sets

- Schedule (available Wednesday, due 8 days later at precept)
- Grading and solutions

Problem Sets

- Schedule (available Wednesday, due 8 days later at precept)
- Grading and solutions
- Code conventions

Problem Sets

- Schedule (available Wednesday, due 8 days later at precept)
- Grading and solutions
- Code conventions
- Collaboration policy

Problem Sets

- Schedule (available Wednesday, due 8 days later at precept)
- Grading and solutions
- Code conventions
- Collaboration policy

Note: You may find these difficult. Start early and seek help!

Ways to Learn

- Lecture
  learn broad topics
- Precept
  learn data analysis skills, get targeted help on assignments
- Readings
  support materials for lecture and precept
- Problem Sets
  reinforce understanding of material, practice

Ways to Learn

- Lecture
  learn broad topics
- Precept
  learn data analysis skills, get targeted help on assignments
- Readings
  support materials for lecture and precept
- Problem Sets
  reinforce understanding of material, practice
- Piazza
  ask questions of us and your classmates

Ways to Learn

- Lecture
  learn broad topics
- Precept
  learn data analysis skills, get targeted help on assignments
- Readings
  support materials for lecture and precept
- Problem Sets
  reinforce understanding of material, practice
- Piazza
  ask questions of us and your classmates
- Office Hours
  ask even more questions.

Ways to Learn

- Lecture
  learn broad topics
- Precept
  learn data analysis skills, get targeted help on assignments
- Readings
  support materials for lecture and precept
- Problem Sets
  reinforce understanding of material, practice
- Piazza
  ask questions of us and your classmates
- Office Hours
  ask even more questions.

Your Job: get help when you need it!

Attribution and Thanks

Attribution and Thanks

- My philosophy on teaching: don't reinvent the wheel- customize, refine, improve.

Attribution and Thanks

- My philosophy on teaching: don't reinvent the wheel- customize, refine, improve.
- Huge thanks to those who have provided slides particularly: Matt Blackwell, Adam Glynn, Justin Grimmer, Jens Hainmueller, Kevin Quinn

Attribution and Thanks

- My philosophy on teaching: don't reinvent the wheel- customize, refine, improve.
- Huge thanks to those who have provided slides particularly: Matt Blackwell, Adam Glynn, Justin Grimmer, Jens Hainmueller, Kevin Quinn
- Also thanks to those who have discussed with me at length including Dalton Conley, Chad Hazlett, Gary King, Kosuke Imai, Matt Salganik and Teppei Yamamoto.
- Shay O'Brien produced the hand-drawn illustrations used throughout.

Outline of Topics

Outline in reverse order:

Outline of Topics

Outline in reverse order:

- Regression:
  how to determine the relationship between variables.

Outline of Topics

Outline in reverse order:

- Regression:
  how to determine the relationship between variables.
- Inference:
  how to learn about things we don't know from the things we do know

Outline of Topics

Outline in reverse order:

- Regression:
  how to determine the relationship between variables.
- Inference:
  how to learn about things we don't know from the things we do know
- Probability:
  learn what data we would expect if we did know the truth.

Outline of Topics

Outline in reverse order:

- Regression:
  how to determine the relationship between variables.
- Inference:
  how to learn about things we don't know from the things we do know
- Probability:
  learn what data we would expect if we did know the truth.
- Probability $\rightarrow$ Inference $\rightarrow$ Regression

What is Statistics?

- branch of mathematics studying collection and analysis of data

What is Statistics?

- branch of mathematics studying collection and analysis of data
- the name statistic comes from the word state

What is Statistics?

- branch of mathematics studying collection and analysis of data
- the name statistic comes from the word state
- the arc of developments in statistics

What is Statistics?

- branch of mathematics studying collection and analysis of data
- the name statistic comes from the word state
- the arc of developments in statistics
  1. an applied scholar has a problem

What is Statistics?

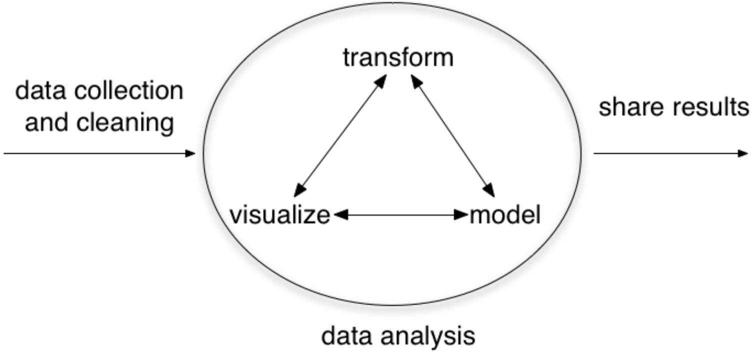
- branch of mathematics studying collection and analysis of data
- the name statistic comes from the word state
- the arc of developments in statistics
  1. an applied scholar has a problem
  2. they solve the problem by inventing a specific method

What is Statistics?

- branch of mathematics studying collection and analysis of data
- the name statistic comes from the word state
- the arc of developments in statistics
  1. an applied scholar has a problem
  2. they solve the problem by inventing a specific method
  3. statisticians generalize and export the best of these methods

Quantitative Research in Theory

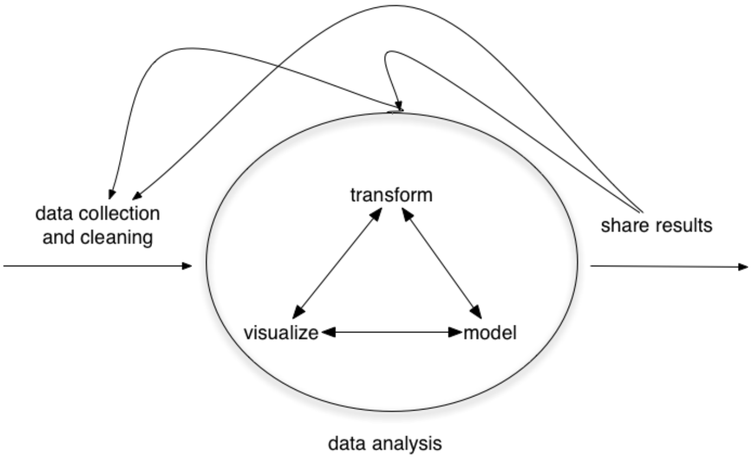
Quantitative Research in Theory



Inspiration: Hadley Wickham, Image: Matt Salganik

Quantitative Research in Practice

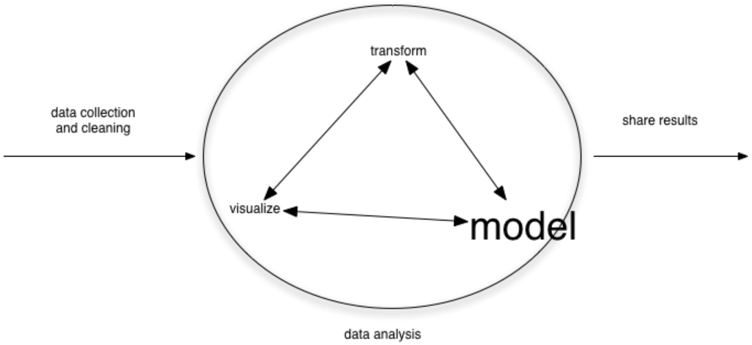
Quantitative Research in Practice



Inspiration: Hadley Wickham, Image: Matt Salganik

Traditional Statistics Class

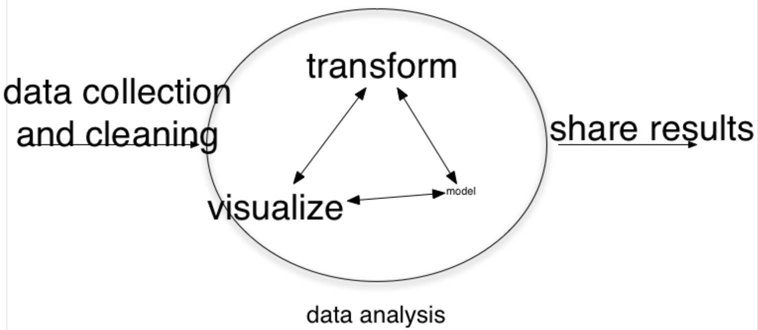
Traditional Statistics Class



Inspiration: Hadley Wickham, Image: Matt Salganik

Time Actually Spent

Time Actually Spent



Inspiration: Hadley Wickham, Image: Matt Salganik

This Class

This Class

- Strike a balance between practice and theory

This Class

- Strike a balance between practice and theory
- Heavy emphasis on applied data analysis

This Class

- Strike a balance between practice and theory
- Heavy emphasis on applied data analysis
  - problem sets with real data

This Class

- Strike a balance between practice and theory
- Heavy emphasis on applied data analysis
  - problem sets with real data
  - replication project next semester

This Class

- Strike a balance between practice and theory
- Heavy emphasis on applied data analysis
  - problem sets with real data
  - replication project next semester
- Teaching select key principles from statistics

Deterministic vs. Stochastic

- “what is the relationship between hours spent studying and performance in Soc500?”

Deterministic vs. Stochastic

- "what is the relationship between hours spent studying and performance in Soc500?"
- One way to approach this:

Deterministic vs. Stochastic

- "what is the relationship between hours spent studying and performance in Soc500?"
- One way to approach this:
  - generate a deterministic account of performance
    $\text{performance}_i = f(\text{hours}_i)$

Deterministic vs. Stochastic

- "what is the relationship between hours spent studying and performance in Soc500?"
- One way to approach this:
  - generate a deterministic account of performance
    $\text{performance}_i = f(\text{hours}_i)$
  - but studying isn't the only indicator of performance!

Deterministic vs. Stochastic

- "what is the relationship between hours spent studying and performance in Soc500?"
- One way to approach this:
  - generate a deterministic account of performance
    $\text{performance}_i = f(\text{hours}_i)$
  - but studying isn't the only indicator of performance!
  - we could try to account for everything
    $\text{performance}_i = f(\text{hours}_i) + g(\text{other}_i)$.

Deterministic vs. Stochastic

- "what is the relationship between hours spent studying and performance in Soc500?"
- One way to approach this:
  - generate a deterministic account of performance
    $\text{performance}_i = f(\text{hours}_i)$
  - but studying isn't the only indicator of performance!
  - we could try to account for everything
    $\text{performance}_i = f(\text{hours}_i) + g(\text{other}_i)$.
  - but that's impossible

Deterministic vs. Stochastic

- "what is the relationship between hours spent studying and performance in Soc500?"
- One way to approach this:
  - generate a deterministic account of performance
    $\text{performance}_i = f(\text{hours}_i)$
  - but studying isn't the only indicator of performance!
  - we could try to account for everything
    $\text{performance}_i = f(\text{hours}_i) + g(\text{other}_i)$.
  - but that's impossible
- A better approach

Deterministic vs. Stochastic

- "what is the relationship between hours spent studying and performance in Soc500?"
- One way to approach this:
  - generate a deterministic account of performance $\text{performance}_i = f(\text{hours}_i)$
  - but studying isn't the only indicator of performance!
  - we could try to account for everything $\text{performance}_i = f(\text{hours}_i) + g(\text{other}_i)$.
  - but that's impossible
- A better approach
  - Instead treat other factors as stochastic

Deterministic vs. Stochastic

- "what is the relationship between hours spent studying and performance in Soc500?"
- One way to approach this:
  - generate a deterministic account of performance
    $\text{performance}_i = f(\text{hours}_i)$
  - but studying isn't the only indicator of performance!
  - we could try to account for everything
    $\text{performance}_i = f(\text{hours}_i) + g(\text{other}_i)$.
  - but that's impossible
- A better approach
  - Instead treat other factors as stochastic
  - Thus we often write it as
    $\text{performance}_i = f(\text{hours}_i) + \epsilon_i$

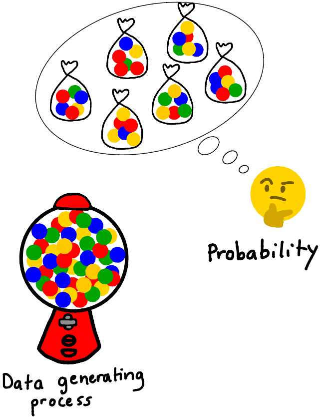
Deterministic vs. Stochastic

- "what is the relationship between hours spent studying and performance in Soc500?"
- One way to approach this:
  - generate a deterministic account of performance
    $\text{performance}_i = f(\text{hours}_i)$
  - but studying isn't the only indicator of performance!
  - we could try to account for everything
    $\text{performance}_i = f(\text{hours}_i) + g(\text{other}_i)$.
  - but that's impossible
- A better approach
  - Instead treat other factors as stochastic
  - Thus we often write it as
    $\text{performance}_i = f(\text{hours}_i) + \epsilon_i$
  - This allows us to have uncertainty over outcomes given our inputs

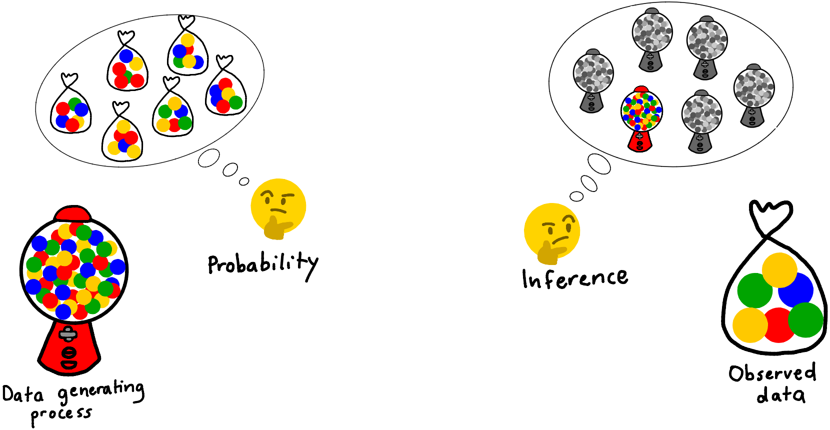
Deterministic vs. Stochastic

- "what is the relationship between hours spent studying and performance in Soc500?"
- One way to approach this:
  - generate a deterministic account of performance
    $\text{performance}_i = f(\text{hours}_i)$
  - but studying isn't the only indicator of performance!
  - we could try to account for everything
    $\text{performance}_i = f(\text{hours}_i) + g(\text{other}_i)$.
  - but that's impossible
- A better approach
  - Instead treat other factors as stochastic
  - Thus we often write it as
    $\text{performance}_i = f(\text{hours}_i) + \epsilon_i$
  - This allows us to have uncertainty over outcomes given our inputs
- Our way of talking about stochastic outcomes is probability.

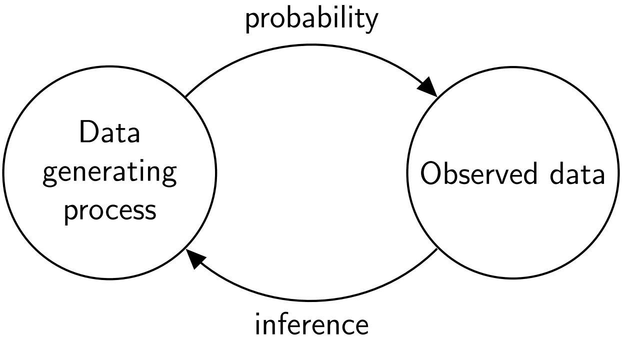
In Picture Form



In Picture Form



In Picture Form



Statistical Thought Experiments

Statistical Thought Experiments

- Start with probability

Statistical Thought Experiments

- Start with probability
- Allows us to contemplate world under hypothetical scenarios

Statistical Thought Experiments

- Start with probability
- Allows us to contemplate world under hypothetical scenarios
  - hypotheticals let us ask- is the observed relationship happening by chance or is it systematic?

Statistical Thought Experiments

- Start with probability
- Allows us to contemplate world under hypothetical scenarios
  - hypotheticals let us ask- is the observed relationship happening by chance or is it systematic?
  - it tells us what the world would look like under a certain assumption

Statistical Thought Experiments

- Start with probability
- Allows us to contemplate world under hypothetical scenarios
  - hypotheticals let us ask- is the observed relationship happening by chance or is it systematic?
  - it tells us what the world would look like under a certain assumption
- We will review probability today, but feel free to ask questions as needed.

Example: Fisher's Lady Tasting Tea

Example: Fisher's Lady Tasting Tea

- The Story Setup

Example: Fisher's Lady Tasting Tea

- The Story Setup
  (lady discerning about tea)

Example: Fisher's Lady Tasting Tea

- The Story Setup
  (lady discerning about tea)
- The Experiment

Example: Fisher's Lady Tasting Tea

- The Story Setup
  (lady discerning about tea)
- The Experiment
  (perform a taste test)

Example: Fisher's Lady Tasting Tea

- The Story Setup
  (lady discerning about tea)
- The Experiment
  (perform a taste test)
- The Hypothetical

Tea-Tasting Distribution

Success count	Permutations of selection	Number of permutations
0	oooo	1 × 1 = 1
1	ooox, ooxo, oxoo, xooo	4 × 4 = 16
2	ooxx, oxox, oxxo, xoxo, xxoo, xoox	6 × 6 = 36
3	oxxx, xoxx, xxox, xxxo	4 × 4 = 16
4	xxxx	1 × 1 = 1
Total		70

Example: Fisher's Lady Tasting Tea

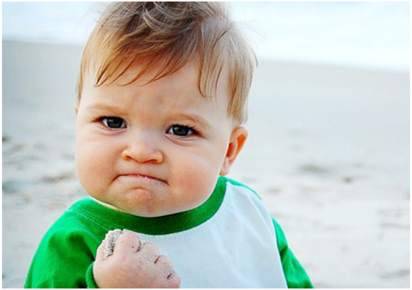
- The Story Setup
  (lady discerning about tea)
- The Experiment
  (perform a taste test)
- The Hypothetical
  (count possibilities)

Tea-Tasting Distribution

Success count	Permutations of selection	Number of permutations
0	oooo	1 × 1 = 1
1	ooox, ooxo, oxoo, xooo	4 × 4 = 16
2	ooxx, oxox, oxxo, xoxo, xxoo, xoox	6 × 6 = 36
3	oxxx, xoxx, xxox, xxxo	4 × 4 = 16
4	xxxx	1 × 1 = 1
Total		70

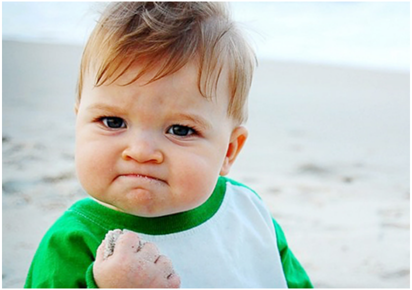
Example: Fisher's Lady Tasting Tea

- The Story Setup
  (lady discerning about tea)
- The Experiment
  (perform a taste test)
- The Hypothetical
  (count possibilities)
- The Result



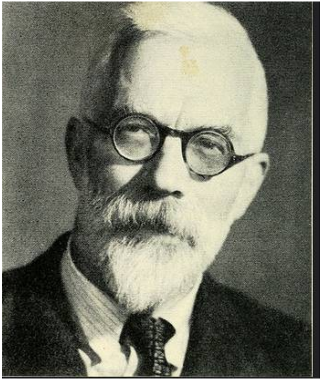
Example: Fisher's Lady Tasting Tea

- The Story Setup
  (lady discerning about tea)
- The Experiment
  (perform a taste test)
- The Hypothetical
  (count possibilities)
- The Result
  (boom she was right)



Example: Fisher's Lady Tasting Tea

- The Story Setup
  (lady discerning about tea)
- The Experiment
  (perform a taste test)
- The Hypothetical
  (count possibilities)
- The Result
  (boom she was right)



This became the Fisher Exact Test.

Why Probability?

Why Probability?

- Helps us envision hypotheticals

Why Probability?

- Helps us envision hypotheticals
- Describes uncertainty in how the data is generated

Why Probability?

- Helps us envision hypotheticals
- Describes uncertainty in how the data is generated
- Data Analysis: estimate probability that something will happen

Why Probability?

- Helps us envision hypotheticals
- Describes uncertainty in how the data is generated
- Data Analysis: estimate probability that something will happen
- Thus: we need to know how probability gives rise to data

Intuitive Definition of Probability

Intuitive Definition of Probability

While there are several interpretations of what probability is, most modern (post 1935 or so) researchers agree on an axiomatic definition of probability.

Intuitive Definition of Probability

While there are several interpretations of what probability is, most modern (post 1935 or so) researchers agree on an axiomatic definition of probability.

3 Axioms (Intuitive Version):

Intuitive Definition of Probability

While there are several interpretations of what probability is, most modern (post 1935 or so) researchers agree on an axiomatic definition of probability.

3 Axioms (Intuitive Version):

1. The probability of any particular event must be non-negative.

Intuitive Definition of Probability

While there are several interpretations of what probability is, most modern (post 1935 or so) researchers agree on an axiomatic definition of probability.

3 Axioms (Intuitive Version):

1. The probability of any particular event must be non-negative.
2. The probability of anything occurring among all possible events must be 1.

Intuitive Definition of Probability

While there are several interpretations of what probability is, most modern (post 1935 or so) researchers agree on an axiomatic definition of probability.

3 Axioms (Intuitive Version):

1. The probability of any particular event must be non-negative.
2. The probability of anything occurring among all possible events must be 1.
3. The probability of one of many mutually exclusive events happening is the sum of the individual probabilities.

Intuitive Definition of Probability

While there are several interpretations of what probability is, most modern (post 1935 or so) researchers agree on an axiomatic definition of probability.

3 Axioms (Intuitive Version):

1. The probability of any particular event must be non-negative.
2. The probability of anything occurring among all possible events must be 1.
3. The probability of one of many mutually exclusive events happening is the sum of the individual probabilities.

All the rules of probability can be derived from these axioms.

Sample Spaces

Sample Spaces

To define probability we need to define the set of possible outcomes.

Sample Spaces

To define probability we need to define the set of possible outcomes.

The sample space is the set of all possible outcomes, and is often written as **S** or Ω.

Sample Spaces

To define probability we need to define the set of possible outcomes.

The sample space is the set of all possible outcomes, and is often written as **S** or Ω.

For example, if we flip a coin twice, there are four possible outcomes,

Sample Spaces

To define probability we need to define the set of possible outcomes.

The sample space is the set of all possible outcomes, and is often written as **S** or Ω.

For example, if we flip a coin twice, there are four possible outcomes,

$$\mathbf{S} =$$

Sample Spaces

To define probability we need to define the set of possible outcomes.

The sample space is the set of all possible outcomes, and is often written as **S** or Ω.

For example, if we flip a coin twice, there are four possible outcomes,

$$\mathbf{S} = \{\{heads, heads\},$$

Sample Spaces

To define probability we need to define the set of possible outcomes.

The sample space is the set of all possible outcomes, and is often written as **S** or Ω.

For example, if we flip a coin twice, there are four possible outcomes,

$$\mathbf{S} = \Big\{\{heads, heads\}, \{heads, tails\},$$

Sample Spaces

To define probability we need to define the set of possible outcomes.

The sample space is the set of all possible outcomes, and is often written as **S** or Ω.

For example, if we flip a coin twice, there are four possible outcomes,

$$\mathbf{S} = \big\{\{heads, heads\}, \{heads, tails\}, \{tails, heads\},$$

Sample Spaces

To define probability we need to define the set of possible outcomes.

The sample space is the set of all possible outcomes, and is often written as **S** or Ω.

For example, if we flip a coin twice, there are four possible outcomes,

$$\mathbf{S} = \Big\{\{heads, heads\}, \{heads, tails\}, \{tails, heads\}, \{tails, tails\}\Big\}$$

Sample Spaces

To define probability we need to define the set of possible outcomes.

The sample space is the set of all possible outcomes, and is often written as **S** or Ω.

For example, if we flip a coin twice, there are four possible outcomes,

$$\mathbf{S} = \Big\{\{heads, heads\}, \{heads, tails\}, \{tails, heads\}, \{tails, tails\}\Big\}$$

Thus the table in Lady Tasting Tea was defining the sample space.

Sample Spaces

To define probability we need to define the set of possible outcomes.

The sample space is the set of all possible outcomes, and is often written as **S** or Ω.

For example, if we flip a coin twice, there are four possible outcomes,

$$\mathbf{S} = \big\{\{heads, heads\}, \{heads, tails\}, \{tails, heads\}, \{tails, tails\}\big\}$$

Thus the table in Lady Tasting Tea was defining the sample space.
(Note we defined illogical guesses to be prob$= 0$)

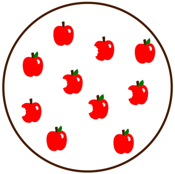
A Running Visual Metaphor

A Running Visual Metaphor

Imagine that we sample an apple from a bag.

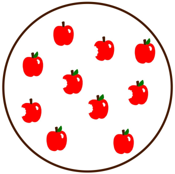
A Running Visual Metaphor

Imagine that we sample an apple from a bag.
Looking in the bag we see:

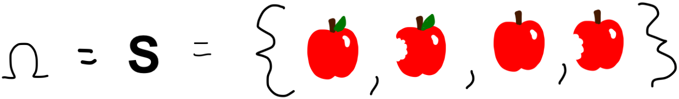


A Running Visual Metaphor

Imagine that we sample an apple from a bag.
Looking in the bag we see:



The sample space is:



Events

Events are subsets of the sample space.

Events

Events are subsets of the sample space.

For Example, if

Events

Events are subsets of the sample space.

For Example, if

then

and

are both **events**.

Events Are a Kind of Set

Events Are a Kind of Set

Sets are collections of things, in this case collections of outcomes

Events Are a Kind of Set

Sets are collections of things, in this case collections of outcomes

One way to define an event is to describe the common property that all of the outcomes share. We write this as

$$\{\omega | \omega \text{ satisfies } P\},$$

where P is the property that they all share.

Events Are a Kind of Set

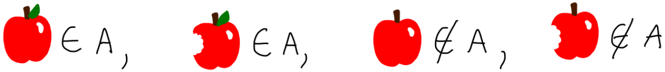
Sets are collections of things, in this case collections of outcomes

One way to define an event is to describe the common property that all of the outcomes share. We write this as

$$\{\omega | \omega \text{ satisfies } P\},$$

where P is the property that they all share.

If $A = \{\omega | \omega \text{ has a leaf}\}$:

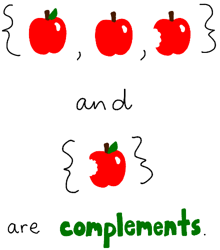


Complement

A complement of event A is a set: A^c, is collection of all of the outcomes not in A. That is, it is "everything else" in the sample space.

Complement

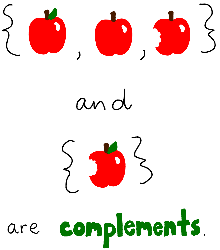
A complement of event A is a set: A^c, is collection of all of the outcomes not in A. That is, it is "everything else" in the sample space.



$$A^c = \{\omega \in \Omega | \omega \notin A\}.$$

Complement

A complement of event A is a set: A^c, is collection of all of the outcomes not in A. That is, it is "everything else" in the sample space.



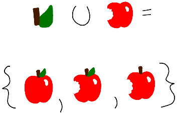
$$A^c = \{\omega \in \Omega | \omega \notin A\}.$$

Important complement: $\Omega^c = \emptyset$, where $\emptyset$ is the empty set—it's just the event that nothing happens.

Operations on Events

Operations on Events

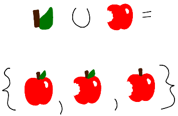
The union of two events, A and B is the event that A or B occurs:



$$A \cup B = \{\omega | \omega \in A \text{ or } \omega \in B\}.$$

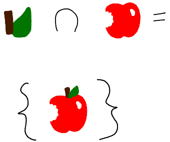
Operations on Events

The union of two events, A and B is the event that A or B occurs:



$$A \cup B = \{\omega | \omega \in A \text{ or } \omega \in B\}.$$

The intersection of two events, A and B is the event that both A and B occur:



$$A \cap B = \{\omega | \omega \in A \text{ and } \omega \in B\}.$$

Operations on Events

We say that two events A and B are disjoint or mutually exclusive if they don't share any elements or that $A \cap B = \emptyset$.

Operations on Events

We say that two events A and B are disjoint or mutually exclusive if they don't share any elements or that $A \cap B = \emptyset$.

An event and its complement A and A^c are disjoint.

Operations on Events

We say that two events A and B are disjoint or mutually exclusive if they don't share any elements or that $A \cap B = \emptyset$.

An event and its complement A and A^c are disjoint.

Operations on Events

We say that two events A and B are disjoint or mutually exclusive if they don't share any elements or that $A \cap B = \emptyset$.

An event and its complement A and A^c are disjoint.

Sample spaces can have infinite outcomes $A_1, A_2, \ldots$.

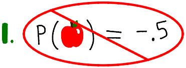
Probability Function

A probability function $P(\cdot)$ is a function defined over all subsets of a sample space $\mathbf{S}$ that satisfies the following three axioms:

Probability Function

A probability function $P(\cdot)$ is a function defined over all subsets of a sample space **S** that satisfies the following three axioms:

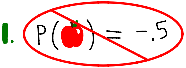
1. $P(A) \geq 0$ for all A in the set of all events.



Probability Function

A probability function $P(\cdot)$ is a function defined over all subsets of a sample space **S** that satisfies the following three axioms:

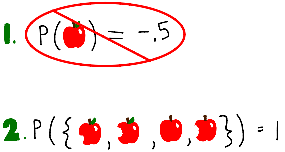
1. $P(A) \geq 0$ for all A in the set of all events. nonnegativity



Probability Function

A probability function $P(\cdot)$ is a function defined over all subsets of a sample space $\mathbf{S}$ that satisfies the following three axioms:

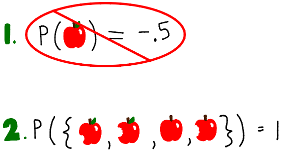
1. $P(A) \geq 0$ for all A in the set of all events. nonnegativity
2. $P(\mathbf{S}) = 1$



Probability Function

A probability function $P(\cdot)$ is a function defined over all subsets of a sample space $\mathbf{S}$ that satisfies the following three axioms:

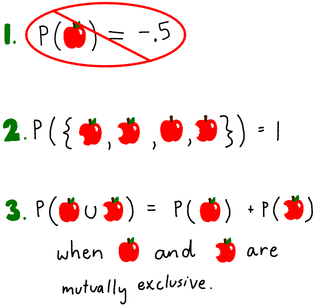
1. $P(A) \geq 0$ for all A in the set of all events. nonnegativity
2. $P(\mathbf{S}) = 1$ normalization



Probability Function

A probability function $P(\cdot)$ is a function defined over all subsets of a sample space $\mathbf{S}$ that satisfies the following three axioms:

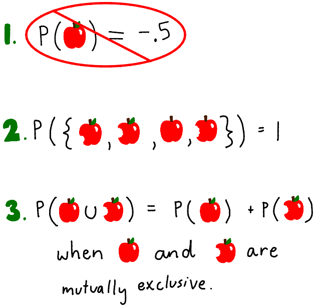
1. $P(A) \geq 0$ for all A in the set of all events. nonnegativity
2. $P(\mathbf{S}) = 1$ normalization
3. if events $A_1, A_2, \ldots$ are mutually exclusive then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.



Probability Function

A probability function $P(\cdot)$ is a function defined over all subsets of a sample space $\mathbf{S}$ that satisfies the following three axioms:

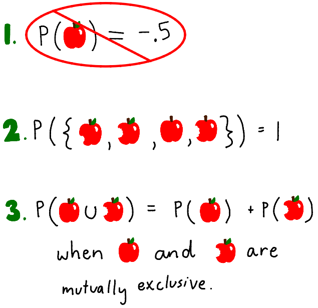
1. $P(A) \geq 0$ for all A in the set of all events. nonnegativity
2. $P(\mathbf{S}) = 1$ normalization
3. if events $A_1, A_2, \ldots$ are mutually exclusive then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. additivity



Probability Function

A probability function $P(\cdot)$ is a function defined over all subsets of a sample space $\mathbf{S}$ that satisfies the following three axioms:

1. $P(A) \geq 0$ for all A in the set of all events. nonnegativity
2. $P(\mathbf{S}) = 1$ normalization
3. if events $A_1, A_2, \ldots$ are mutually exclusive then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. additivity

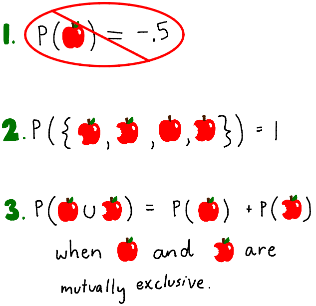


All the rules of probability can be derived from these axioms.

Probability Function

A probability function $P(\cdot)$ is a function defined over all subsets of a sample space $\mathbf{S}$ that satisfies the following three axioms:

1. $P(A) \geq 0$ for all A in the set of all events. nonnegativity
2. $P(\mathbf{S}) = 1$ normalization
3. if events $A_1, A_2, \ldots$ are mutually exclusive then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. additivity



All the rules of probability can be derived from these axioms.

Intuition: probability as allocating chunks of a unit-long stick.

A Brief Word on Interpretation

A Brief Word on Interpretation

Massive debate on interpretation:

A Brief Word on Interpretation

Massive debate on interpretation:

- Subjective Interpretation

A Brief Word on Interpretation

Massive debate on interpretation:

- Subjective Interpretation
  - Example: The probability of drawing 5 red cards out of 10 drawn from a deck of cards is whatever you want it to be.

A Brief Word on Interpretation

Massive debate on interpretation:

- Subjective Interpretation
  - Example: The probability of drawing 5 red cards out of 10 drawn from a deck of cards is whatever you want it to be. But...

A Brief Word on Interpretation

Massive debate on interpretation:

- Subjective Interpretation
  - Example: The probability of drawing 5 red cards out of 10 drawn from a deck of cards is whatever you want it to be. But...
  - If you don't follow the axioms, a bookie can beat you

A Brief Word on Interpretation

Massive debate on interpretation:

- Subjective Interpretation
  - Example: The probability of drawing 5 red cards out of 10 drawn from a deck of cards is whatever you want it to be. But...
  - If you don't follow the axioms, a bookie can beat you
  - There is a correct way to update your beliefs with data.

A Brief Word on Interpretation

Massive debate on interpretation:

- Subjective Interpretation
  - Example: The probability of drawing 5 red cards out of 10 drawn from a deck of cards is whatever you want it to be. But...
  - If you don't follow the axioms, a bookie can beat you
  - There is a correct way to update your beliefs with data.
- Frequency Interpretation

A Brief Word on Interpretation

Massive debate on interpretation:

- Subjective Interpretation
  - Example: The probability of drawing 5 red cards out of 10 drawn from a deck of cards is whatever you want it to be. But...
  - If you don't follow the axioms, a bookie can beat you
  - There is a correct way to update your beliefs with data.
- Frequency Interpretation
  - Probability of is the relative frequency with which an event would occur if the process were repeated a large number of times under similar conditions.

A Brief Word on Interpretation

Massive debate on interpretation:

- Subjective Interpretation
  - Example: The probability of drawing 5 red cards out of 10 drawn from a deck of cards is whatever you want it to be. But...
  - If you don't follow the axioms, a bookie can beat you
  - There is a correct way to update your beliefs with data.
- Frequency Interpretation
  - Probability of is the relative frequency with which an event would occur if the process were repeated a large number of times under similar conditions.
  - Example: The probability of drawing 5 red cards out of 10 drawn from a deck of cards is the frequency with which this event occurs in repeated samples of 10 cards.

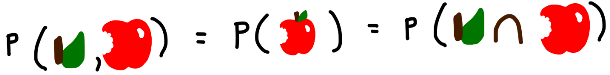
Marginal and Joint Probability

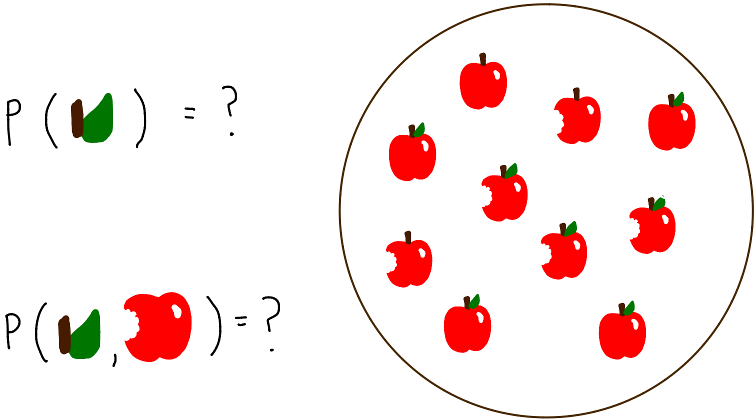
So far we have only considered situations where we are interested in the probability of a single event A occurring. We've denoted this $P(A)$. $P(A)$ is sometimes called a marginal probability.

Marginal and Joint Probability

So far we have only considered situations where we are interested in the probability of a single event A occurring. We've denoted this $P(A)$. $P(A)$ is sometimes called a marginal probability.

Suppose we are now in a situation where we would like to express the probability that an event A and an event B occur. This quantity is written as $P(A \cap B)$, $P(B \cap A)$, $P(A, B)$, or $P(B, A)$ and is the joint probability of A and B.




P( ) = ?
P(,) = ?

Conditional Probability

Conditional Probability

The "soul of statistics"

Conditional Probability

The "soul of statistics"
If $P(A) > 0$ then the probability of B conditional on A can be written as

Conditional Probability

The "soul of statistics"
If $P(A) > 0$ then the probability of B conditional on A can be written as

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

Conditional Probability

The "soul of statistics"
If $P(A) > 0$ then the probability of B conditional on A can be written as

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

This implies that

Conditional Probability

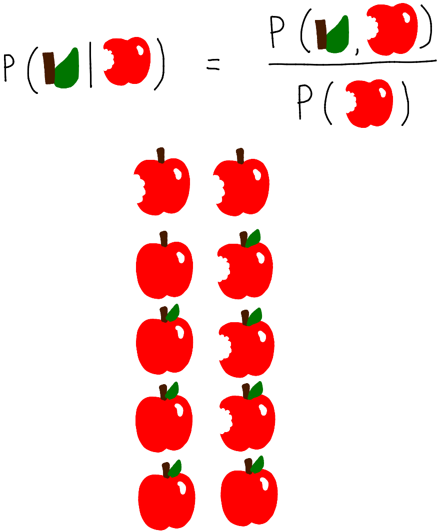
The "soul of statistics"
If $P(A) > 0$ then the probability of B conditional on A can be written as

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

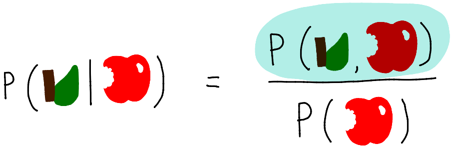
This implies that

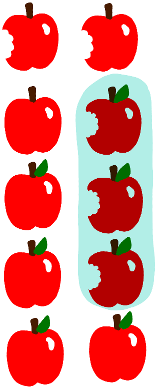
$$P(A,B) = P(A) \times P(B|A)$$

Conditional Probability: A Visual Example

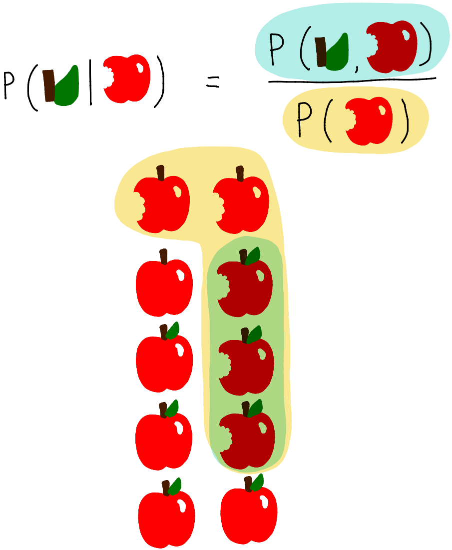


Conditional Probability: A Visual Example





Conditional Probability: A Visual Example



A Card Player's Example

If we randomly draw two cards from a standard 52 card deck and define the events
$A = \{\text{King on Draw 1}\}$ and $B = \{\text{King on Draw 2}\}$, then

A Card Player's Example

If we randomly draw two cards from a standard 52 card deck and define the events
$A = \{\text{King on Draw 1}\}$ and $B = \{\text{King on Draw 2}\}$, then

- $P(A) =$

A Card Player's Example

If we randomly draw two cards from a standard 52 card deck and define the events
$A = \{\text{King on Draw 1}\}$ and $B = \{\text{King on Draw 2}\}$, then

- $P(A) = 4/52$

A Card Player's Example

If we randomly draw two cards from a standard 52 card deck and define the events
$A = \{\text{King on Draw 1}\}$ and $B = \{\text{King on Draw 2}\}$, then

- $P(A) = 4/52$
- $P(B|A) =$

A Card Player's Example

If we randomly draw two cards from a standard 52 card deck and define the events
$A = \{\text{King on Draw 1}\}$ and $B = \{\text{King on Draw 2}\}$, then

- $P(A) = 4/52$
- $P(B|A) = 3/51$

A Card Player's Example

If we randomly draw two cards from a standard 52 card deck and define the events
$A = \{\text{King on Draw 1}\}$ and $B = \{\text{King on Draw 2}\}$, then

- $P(A) = 4/52$
- $P(B|A) = 3/51$
- $P(A, B) = P(A) \times P(B|A) =$

A Card Player's Example

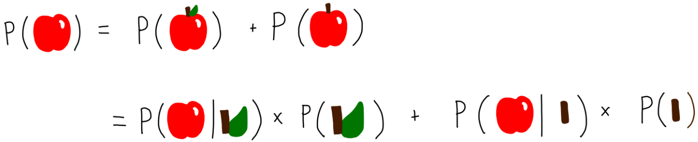
If we randomly draw two cards from a standard 52 card deck and define the events
$A = \{\text{King on Draw 1}\}$ and $B = \{\text{King on Draw 2}\}$, then

- $P(A) = 4/52$
- $P(B|A) = 3/51$
- $P(A, B) = P(A) \times P(B|A) = 4/52 \times 3/51 \approx .0045$

Law of Total Probability (LTP)

With 2 Events:

$$\begin{aligned} P(B) &= P(B, A) + P(B, A^c) \\ &= P(B|A) \times P(A) + P(B|A^c) \times P(A^c) \end{aligned}$$



Recall, if we randomly draw two cards from a standard 52 card deck and define the events $A = \{\text{King on Draw 1}\}$ and $B = \{\text{King on Draw 2}\}$, then

- $P(A) = 4/52$
- $P(B|A) = 3/51$
- $P(A, B) = P(A) \times P(B|A) = 4/52 \times 3/51$

Question: $P(B) = ?$

Confirming Intuition with the LTP

Confirming Intuition with the LTP

$$
\begin{aligned}
P(B) &= P(B, A) + P(B, A^c) \\
&= P(B|A) \times P(A) + P(B|A^c) \times P(A^c)
\end{aligned}
$$

Confirming Intuition with the LTP

$$
\begin{aligned}
P(B) &= P(B, A) + P(B, A^c) \\
&= P(B|A) \times P(A) + P(B|A^c) \times P(A^c)
\end{aligned}
$$

$$
\begin{aligned}
P(B) &= 3/51 \times 1/13 + 4/51 \times 12/13 \\
&= \frac{3 + 48}{51 \times 13} = \frac{1}{13} = \frac{4}{52}
\end{aligned}
$$

Example: Voter Mobilization

Suppose that we have put together a voter mobilization campaign and we want to know what the probability of voting is after the campaign: Pr[vote].

Example: Voter Mobilization

Suppose that we have put together a voter mobilization campaign and we want to know what the probability of voting is after the campaign: Pr[vote]. We know the following:

Example: Voter Mobilization

Suppose that we have put together a voter mobilization campaign and we want to know what the probability of voting is after the campaign: Pr[vote]. We know the following:

- $\Pr(\text{vote}|\text{mobilized}) = 0.75$

Example: Voter Mobilization

Suppose that we have put together a voter mobilization campaign and we want to know what the probability of voting is after the campaign: Pr[vote]. We know the following:

- $\Pr(\text{vote}|\text{mobilized}) = 0.75$
- $\Pr(\text{vote}|\text{not mobilized}) = 0.15$

Example: Voter Mobilization

Suppose that we have put together a voter mobilization campaign and we want to know what the probability of voting is after the campaign: Pr[vote]. We know the following:

- $\Pr(\text{vote}|\text{mobilized}) = 0.75$
- $\Pr(\text{vote}|\text{not mobilized}) = 0.15$
- $\Pr(\text{mobilized}) = 0.6$ and so $\Pr(\text{ not mobilized}) = 0.4$

Example: Voter Mobilization

Suppose that we have put together a voter mobilization campaign and we want to know what the probability of voting is after the campaign: Pr[vote]. We know the following:

- $\Pr(\text{vote}|\text{mobilized}) = 0.75$
- $\Pr(\text{vote}|\text{not mobilized}) = 0.15$
- $\Pr(\text{mobilized}) = 0.6$ and so $\Pr(\text{ not mobilized}) = 0.4$

Note that mobilization partitions the data. Everyone is either mobilized or not.

Example: Voter Mobilization

Suppose that we have put together a voter mobilization campaign and we want to know what the probability of voting is after the campaign: Pr[vote]. We know the following:

- $\Pr(\text{vote}|\text{mobilized}) = 0.75$
- $\Pr(\text{vote}|\text{not mobilized}) = 0.15$
- $\Pr(\text{mobilized}) = 0.6$ and so $\Pr(\text{ not mobilized}) = 0.4$

Note that mobilization partitions the data. Everyone is either mobilized or not. Thus, we can apply the LTP:

Example: Voter Mobilization

Suppose that we have put together a voter mobilization campaign and we want to know what the probability of voting is after the campaign: Pr[vote]. We know the following:

- $\Pr(\text{vote}|\text{mobilized}) = 0.75$
- $\Pr(\text{vote}|\text{not mobilized}) = 0.15$
- $\Pr(\text{mobilized}) = 0.6$ and so $\Pr(\text{ not mobilized}) = 0.4$

Note that mobilization partitions the data. Everyone is either mobilized or not. Thus, we can apply the LTP:

$$\begin{aligned}\Pr(\text{vote}) = & \Pr(\text{vote}|\text{mobilized}) \Pr(\text{mobilized}) + \\ & \Pr(\text{vote}|\text{not mobilized}) \Pr(\text{not mobilized})\end{aligned}$$

Example: Voter Mobilization

Suppose that we have put together a voter mobilization campaign and we want to know what the probability of voting is after the campaign: Pr[vote]. We know the following:

- $\Pr(\text{vote}|\text{mobilized}) = 0.75$
- $\Pr(\text{vote}|\text{not mobilized}) = 0.15$
- $\Pr(\text{mobilized}) = 0.6$ and so $\Pr(\text{ not mobilized}) = 0.4$

Note that mobilization partitions the data. Everyone is either mobilized or not. Thus, we can apply the LTP:

$$\begin{aligned}\Pr(\text{vote}) =& \Pr(\text{vote}|\text{mobilized}) \Pr(\text{mobilized}) + \\ & \Pr(\text{vote}|\text{not mobilized}) \Pr(\text{not mobilized}) \\ =& 0.75 \times 0.6 + 0.15 \times 0.4\end{aligned}$$

Example: Voter Mobilization

Suppose that we have put together a voter mobilization campaign and we want to know what the probability of voting is after the campaign: Pr[vote]. We know the following:

- $\Pr(\text{vote}|\text{mobilized}) = 0.75$
- $\Pr(\text{vote}|\text{not mobilized}) = 0.15$
- $\Pr(\text{mobilized}) = 0.6$ and so $\Pr(\text{ not mobilized}) = 0.4$

Note that mobilization partitions the data. Everyone is either mobilized or not. Thus, we can apply the LTP:

$$\begin{aligned}
\Pr(\text{vote}) =& \Pr(\text{vote}|\text{mobilized}) \Pr(\text{mobilized}) + \\
& \Pr(\text{vote}|\text{not mobilized}) \Pr(\text{not mobilized}) \\
=& 0.75 \times 0.6 + 0.15 \times 0.4 \\
=& .51
\end{aligned}$$

Bayes' Rule

Bayes' Rule

- Often we have information about $\Pr(B|A)$, but require $\Pr(A|B)$ instead.

Bayes' Rule

- Often we have information about $\Pr(B|A)$, but require $\Pr(A|B)$ instead.
- When this happens, always think: Bayes' rule

Bayes' Rule

- Often we have information about $\Pr(B|A)$, but require $\Pr(A|B)$ instead.
- When this happens, always think: Bayes' rule
- Bayes' rule: if $\Pr(B) > 0$, then:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

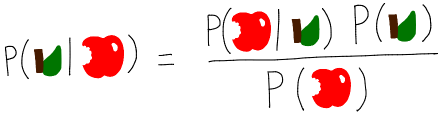
Bayes' Rule

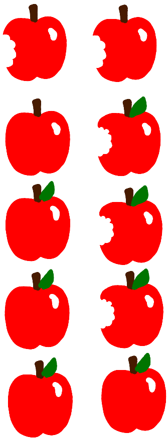
- Often we have information about $\Pr(B|A)$, but require $\Pr(A|B)$ instead.
- When this happens, always think: Bayes' rule
- Bayes' rule: if $\Pr(B) > 0$, then:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

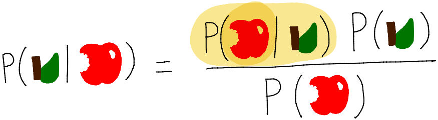
- Proof: combine the multiplication rule $\Pr(B|A)\Pr(A) = P(A \cap B)$, and the definition of conditional probability

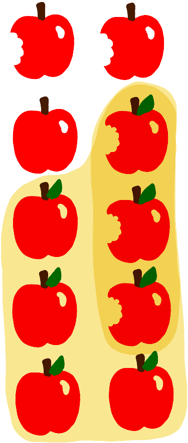
Bayes' Rule Mechanics



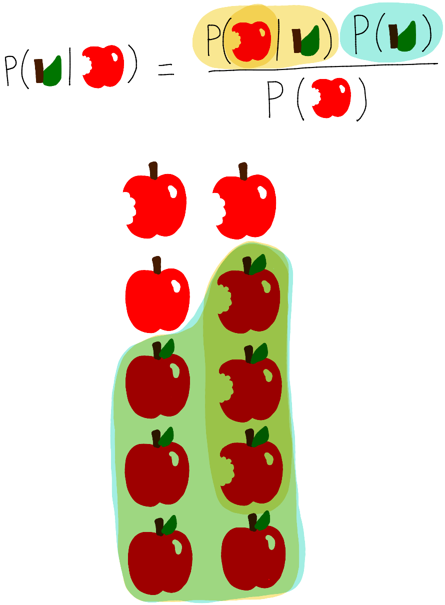


Bayes' Rule Mechanics

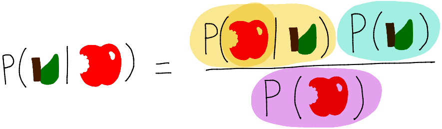


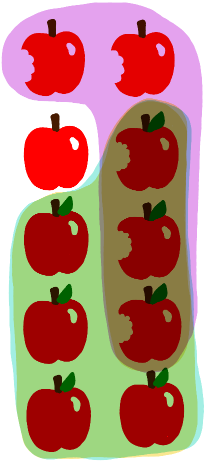


Bayes' Rule Mechanics



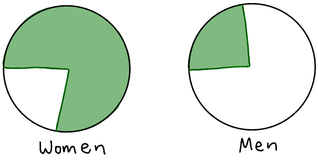
Bayes' Rule Mechanics





Bayes' Rule Example

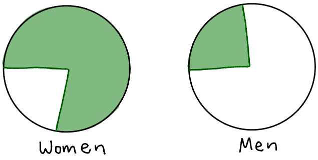
U.S. Billionaires, 2014



- 76.5% of female billionaires inherited their fortunes, compared to 24.5% of male billionaires
- So is P(woman | inherited billions) greater than P(man | inherited billions)?

Bayes' Rule Example

U.S. Billionaires, 2014

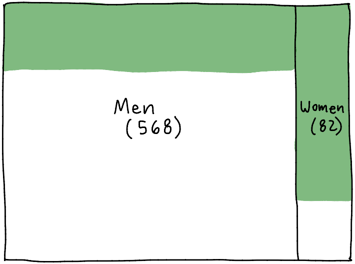


- 76.5% of female billionaires inherited their fortunes, compared to 24.5% of male billionaires
- So is P(woman | inherited billions) greater than P(man | inherited billions)?

$$P(W \mid I) = \frac{P(I \mid W)\, P(W)}{P(I)}$$

$$= \frac{.765\left(\frac{82}{568+82}\right)}{\frac{.765(82) + .245(568)}{568+82}}$$

$$= .31$$



* Data source = Billionaires characteristics database

Example: Race and Names

Example: Race and Names

- Enos (2015): how do we identify a person's race from their name?

Example: Race and Names

- Enos (2015): how do we identify a person's race from their name?
- First, note that the Census collects information on the distribution of names by race.

Example: Race and Names

- Enos (2015): how do we identify a person's race from their name?
- First, note that the Census collects information on the distribution of names by race.
- For example, Washington is the most common last name among African-Americans in America:

Example: Race and Names

- Enos (2015): how do we identify a person's race from their name?
- First, note that the Census collects information on the distribution of names by race.
- For example, Washington is the most common last name among African-Americans in America:
  - $\Pr(\text{AfAm}) = 0.132$

Example: Race and Names

- Enos (2015): how do we identify a person's race from their name?
- First, note that the Census collects information on the distribution of names by race.
- For example, Washington is the most common last name among African-Americans in America:
  - $\Pr(\text{AfAm}) = 0.132$
  - $\Pr(\text{not AfAm}) = 1 - \Pr(\text{AfAm}) = .868$

Example: Race and Names

- Enos (2015): how do we identify a person's race from their name?
- First, note that the Census collects information on the distribution of names by race.
- For example, Washington is the most common last name among African-Americans in America:
  - $\Pr(\text{AfAm}) = 0.132$
  - $\Pr(\text{not AfAm}) = 1 - \Pr(\text{AfAm}) = .868$
  - $\Pr(\text{Washington}|\text{AfAm}) = 0.00378$

Example: Race and Names

- Enos (2015): how do we identify a person's race from their name?
- First, note that the Census collects information on the distribution of names by race.
- For example, Washington is the most common last name among African-Americans in America:
  - $\Pr(\text{AfAm}) = 0.132$
  - $\Pr(\text{not AfAm}) = 1 - \Pr(\text{AfAm}) = .868$
  - $\Pr(\text{Washington}|\text{AfAm}) = 0.00378$
  - $\Pr(\text{Washington}|\text{not AfAm}) = 0.000061$

Example: Race and Names

- Enos (2015): how do we identify a person's race from their name?
- First, note that the Census collects information on the distribution of names by race.
- For example, Washington is the most common last name among African-Americans in America:
  - $\Pr(\text{AfAm}) = 0.132$
  - $\Pr(\text{not AfAm}) = 1 - \Pr(\text{AfAm}) = .868$
  - $\Pr(\text{Washington}|\text{AfAm}) = 0.00378$
  - $\Pr(\text{Washington}|\text{not AfAm}) = 0.000061$
- We can now use Bayes' Rule

Example: Race and Names

- Enos (2015): how do we identify a person's race from their name?
- First, note that the Census collects information on the distribution of names by race.
- For example, Washington is the most common last name among African-Americans in America:
  - $\Pr(\text{AfAm}) = 0.132$
  - $\Pr(\text{not AfAm}) = 1 - \Pr(\text{AfAm}) = .868$
  - $\Pr(\text{Washington}|\text{AfAm}) = 0.00378$
  - $\Pr(\text{Washington}|\text{not AfAm}) = 0.000061$
- We can now use Bayes' Rule

$$\Pr(\text{AfAm}|\text{Wash}) = \frac{\Pr(\text{Wash}|\text{AfAm})\Pr(\text{AfAm})}{\Pr(\text{Wash})}$$

Example: Race and Names

Example: Race and Names

Note we don't have the probability of the name Washington.

Example: Race and Names

Note we don't have the probability of the name Washington.

Remember that we can calculate it from the LTP since the sets African-American and not African-American partition the sample space:

$$\frac{\Pr(\text{Wash}|\text{AfAm})\Pr(\text{AfAm})}{\Pr(\text{Wash})} =$$

Example: Race and Names

Note we don't have the probability of the name Washington.

Remember that we can calculate it from the LTP since the sets African-American and not African-American partition the sample space:

$$\frac{\Pr(\text{Wash}|\text{AfAm})\Pr(\text{AfAm})}{\Pr(\text{Wash})} = \frac{\Pr(\text{Wash}|\text{AfAm})\Pr(\text{AfAm})}{\Pr(\text{Wash}|\text{AfAm})\Pr(\text{AfAm}) + \Pr(\text{Wash}|\text{not AfAm})\Pr(\text{not AfAm})}$$

Example: Race and Names

Note we don't have the probability of the name Washington.

Remember that we can calculate it from the LTP since the sets African-American and not African-American partition the sample space:

$$\frac{\Pr(\text{Wash}|\text{AfAm})\Pr(\text{AfAm})}{\Pr(\text{Wash})} = \frac{\Pr(\text{Wash}|\text{AfAm})\Pr(\text{AfAm})}{\Pr(\text{Wash}|\text{AfAm})\Pr(\text{AfAm}) + \Pr(\text{Wash}|\text{not AfAm})\Pr(\text{not AfAm})}$$

$$= \frac{0.132 \times 0.00378}{0.132 \times 0.00378 + .868 \times 0.000061}$$

Example: Race and Names

Note we don't have the probability of the name Washington.

Remember that we can calculate it from the LTP since the sets African-American and not African-American partition the sample space:

$$\begin{aligned}\frac{\Pr(\text{Wash}|\text{AfAm})\Pr(\text{AfAm})}{\Pr(\text{Wash})} &= \frac{\Pr(\text{Wash}|\text{AfAm})\Pr(\text{AfAm})}{\Pr(\text{Wash}|\text{AfAm})\Pr(\text{AfAm}) + \Pr(\text{Wash}|\text{not AfAm})\Pr(\text{not AfAm})} \\ &= \frac{0.132 \times 0.00378}{0.132 \times 0.00378 + .868 \times 0.000061} \\ &\approx 0.9\end{aligned}$$

Independence

Independence

Intuitive Definition

Independence

Intuitive Definition

Events A and B are independent if knowing whether A occurred provides no information about whether B occurred.

Independence

Intuitive Definition

Events A and B are independent if knowing whether A occurred provides no information about whether B occurred.

Formal Definition

Independence

Intuitive Definition

Events A and B are independent if knowing whether A occurred provides no information about whether B occurred.

Formal Definition

$$P(A, B) = P(A)P(B) \implies A \perp\!\!\!\perp B$$

Independence

Intuitive Definition

Events A and B are independent if knowing whether A occurred provides no information about whether B occurred.

Formal Definition

$$P(A, B) = P(A)P(B) \implies A \perp\!\!\!\perp B$$

With all the usual > 0 restrictions, this implies

Independence

Intuitive Definition

Events A and B are independent if knowing whether A occurred provides no information about whether B occurred.

Formal Definition

$$P(A, B) = P(A)P(B) \implies A \perp\!\!\!\perp B$$

With all the usual > 0 restrictions, this implies

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

Independence

Intuitive Definition

Events A and B are independent if knowing whether A occurred provides no information about whether B occurred.

Formal Definition

$$P(A, B) = P(A)P(B) \implies A \perp\!\!\!\perp B$$

With all the usual > 0 restrictions, this implies

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

Independence is a massively important concept in statistics.

Next Week

- Random Variables

Next Week

- Random Variables
- Reading

Next Week

- Random Variables
- Reading
  - Aronow and Miller (1.1) on Random Events
  - Aronow and Miller (1.2-2.3) on Probability Theory, Summarizing Distributions
  - Optional: Blitzstein and Hwang Chapters 1-1.3 (probability), 2-2.5 (conditional probability), 3-3.2 (random variables), 4-4.2 (expectation), 4.4-4.6 (indicator rv, LOTUS, variance), 7.0-7.3 (joint distributions)
  - Optional: Imai Chapter 6 (probability)

Next Week

- Random Variables
- Reading
  - Aronow and Miller (1.1) on Random Events
  - Aronow and Miller (1.2-2.3) on Probability Theory, Summarizing Distributions
  - Optional: Blitzstein and Hwang Chapters 1-1.3 (probability), 2-2.5 (conditional probability), 3-3.2 (random variables), 4-4.2 (expectation), 4.4-4.6 (indicator rv, LOTUS, variance), 7.0-7.3 (joint distributions)
  - Optional: Imai Chapter 6 (probability)
- A word from your preceptors

Fun With

Fun with

Fun with History

Fun with History

Legendre

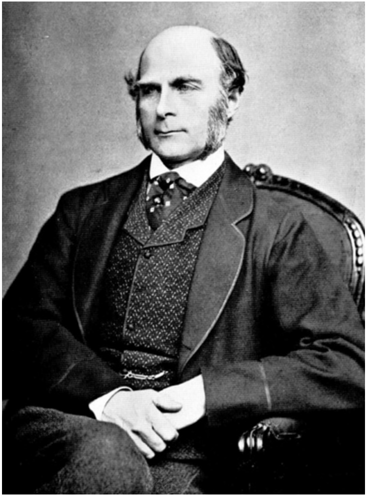
Fun with History

Gauss

Fun with History

Quetelet

Fun with History



Galton

References

- Enos, Ryan D. "What the demolition of public housing teaches us about the impact of racial threat on political behavior." *American Journal of Political Science* (2015).
- J. Andrew Harris "Whats in a Name? A Method for Extracting Information about Ethnicity from Names" *Political Analysis* (2015).
- Salsburg, David. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century* (2002).