Week 1

Introduction to Probability

Paul Laskowski and Alex Hughes

January 11, 2024

UC Berkeley, School of Information

The Nature of Statistical Models

What is a Model?

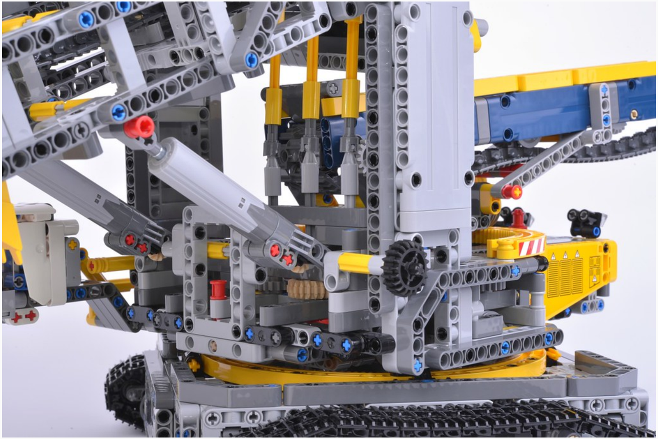


Image: www.flickr.com/photos/brickset/ (CC BY 2.0)

What is a Statistical Model?

Course Plan

1. Putting together statistical models
2. Fitting models to data

First Models

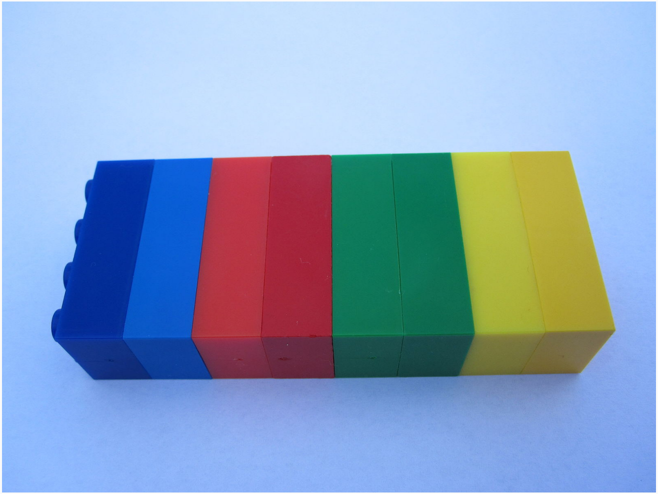


Image: Hans Schou (CC BY-SA 3.0)

Useful Models

Image: www.flickr.com/photos/brickset/ (CC BY 2.0)

Probability is Rules for Putting Pieces Together

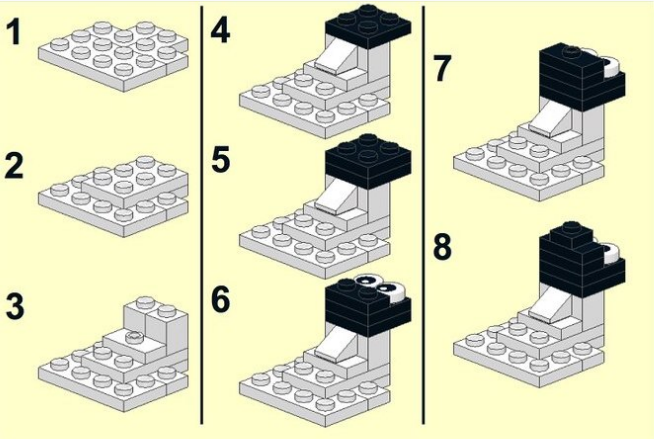


Image: www.flickr.com/photos/billward/ (CC BY 2.0)

Unit Plan

Necessary Foundation

- A precise definition of probability
- How mathematicians build from a set of axioms to useful properties
- How to connect these properties to problems that we want to solve

Reading Assignment

Reading Assignment

Throughout this course, we intersperse reading, lectures, and work that you complete as a bundled *async*. This way, you can very quickly move from

- An introduction to a concept
- An explanation of its use
- A proof of that concept
- An application using data or problem solving

We have designed the course so that you can complete the async, reading, and lecture activity components over a few self-contained study sessions.

Reading Assignment (cont.)

Read the following sections:

- Introduction, page xv - xviii

Probability is Reasoning Under Uncertainty

Probablity is A System of Reasoning

"Essentially, the theory of probability is nothing but good common sense reduced to mathematics. It provides an exact appreciation of what sound minds feel with a kind of instinct, frequently without being able to account for it."
- Pierre-Simon Laplace

Probability is A Model of the World

Apocryphal quotations

"*All models are wrong; some are useful.*"

—George Box

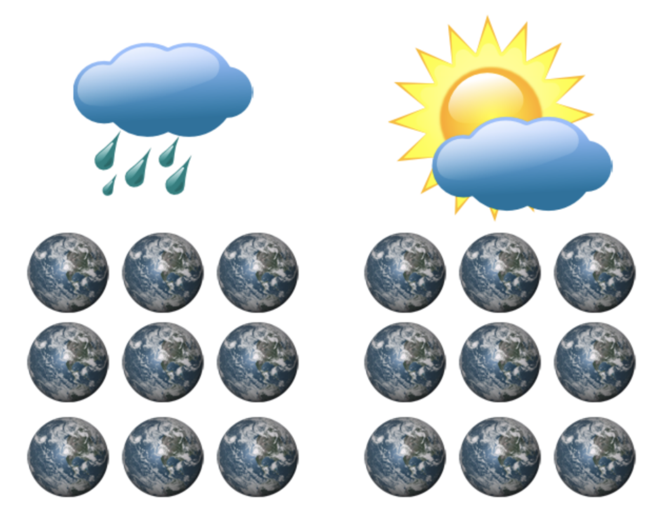
Probability is:

- A mathematical construct
- A model or abstraction of the world

Probability is a Model - Example 1

Probability is a Model - Example 2

Probability is a Model - Example 2 (cont.)



Probability is a Model - Example 3

An Axiomatic Approach

Why take an axiomatic approach?

- Define a minimal set of statements that are consistent with the world we observe
- Deduce statements that must then be logically entailed

Thomas Kuhn's Philosophy of Science

- Axiomatic statements
- Intermediate theories
- Testable hypotheses

Learnosity: Set Theory Diagnostic Quiz

Set Theory

If you already understand the following notation and can solve the problem, we recommend that you skip this sequence and go straight to the next segment about partitions.

Notation:

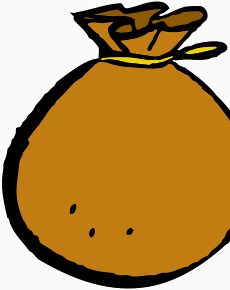
A and B are sets.

union	$A \cup B$
intersection	$A \cap B$
set difference	$A \setminus B$
complement	A^C
empty set	$\emptyset$

Simplify the expression: $(A \cap B) \cup (B \setminus A)$

The Set Operations, Part 1

Visualizing a Set



- A set contains objects.
- Items don't have an order.
- Items can't repeat.

Elements of Sets

Defining a new set:

- Elements go in curly braces.
- A colon means "such that".

Specifying elements:

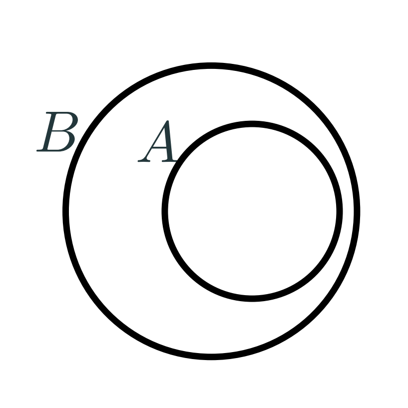
- '$\in$' means ""is an element of."
- '$\notin$' means "is not an element of."
- '$\emptyset$' means the empty set.

Subsets and Supersets

$A = \{1, 2\},\ B = \{1, 2, 3, 4\}$

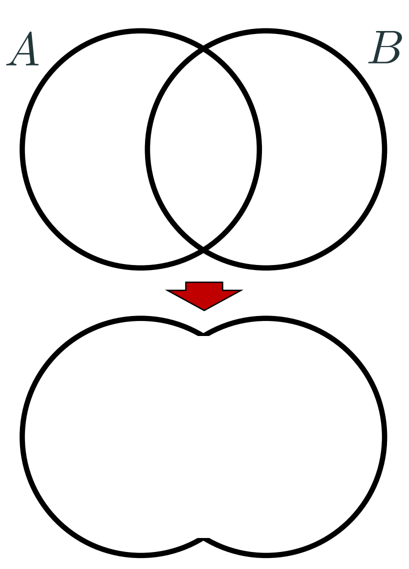
Subset: $A \subseteq B$

Strict Subset: $A \subsetneq B$



Set Union

$A = \{1, 2, 3\},\ B = \{3, 4\}$
Union is $A \cup B$



Properties of Unions

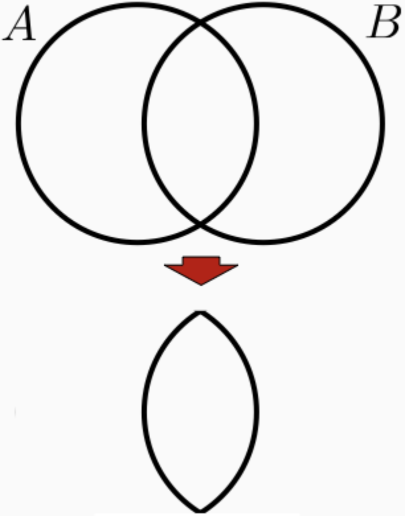
- Symmetry: $A \cup B = B \cup A$
- Association: $(A \cup B) \cup C = A \cup (B \cup C)$
- $A \cup A = A$
- $A \cup \emptyset = A$

The Set Operations, Part 2

Set Intersection

$A = \{1, 2, 3\}, \;\; B = \{3, 4\}$

Intersection is $A \cap B$



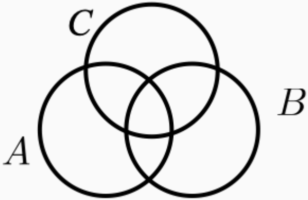
Two sets are *disjoint* if their intersection is empty.

Properties of Intersections

- Symmetry: $A \cap B = B \cap A$
- Association: $(A \cap B) \cap C = A \cap (B \cap C)$
- $A \cap A = A$
- $A \cap \emptyset = \emptyset$

Distributing Union and Intersection

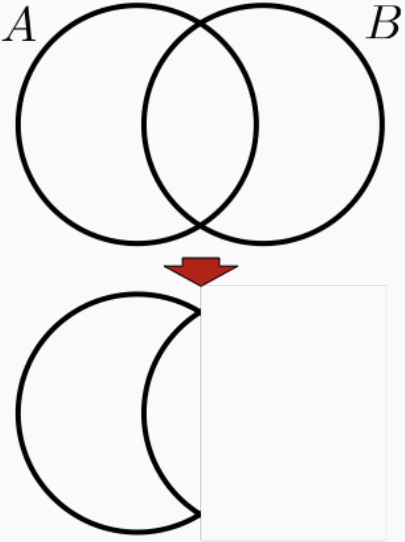
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$



$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

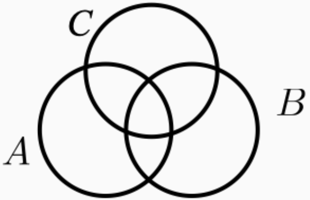
Set Difference

$A = \{1, 2, 3\},\ B = \{3, 4\}$
Difference is $A \setminus B$



Distributing Set Difference

$$A \cap (B \setminus C) = (A \cap B) \setminus C$$

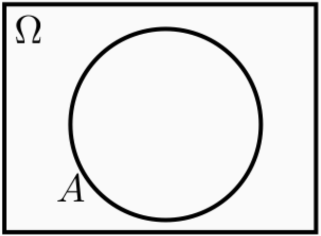


$$C \setminus (A \cap B) = (C \setminus A) \cup (C \setminus B)$$

$$C \setminus (A \cup B) = (C \setminus A) \cap (C \setminus B)$$

Set Complement

Ω - ALL objects of interest.
$A = \{1, 2, 3\},\ \Omega = \{1, 2, 3, 4, 5\}$
Complement: $A^C = \Omega \setminus A$



Properties:

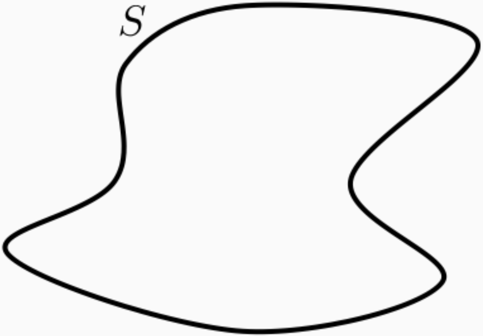
- $A^C \cup A = \Omega$
- $A \setminus B = A \cap B^C$

Set Partitions

Set Partitions

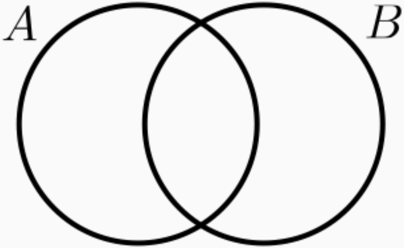
Definition 1.1.12: partition

A set of sets $A_1, A_2, \ldots, A_n$ is a *partition* of set S, if $A_1, A_2, \ldots, A_n$ are nonempty and pairwise disjoint, and if $S = A_1 \cup A_2 \cup \cdots \cup A_n$.



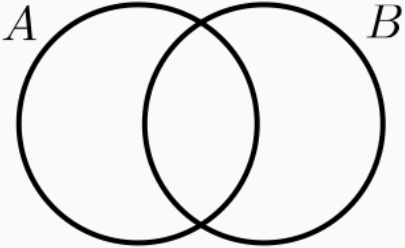
Partition Rule 1

Given a set A and another set B, a partition of A is...



Partition Rule 2

Given a set A and another set B, a partition of $A \cup B$ is...



Call to Reading

- Chapter 1, from page 3 to the top of page 8, stopping before *Theorem 1.1.1*

Probability Spaces

Fundamental Components of Probability Space

A Probability Space

- A **sample space**, denoted as Ω, is the collection of *all* possible outcomes.
- An **event space**, denoted as S, is made up of sets of outcomes.
- A **probability measure** is a function that maps events to numbers in $[0, 1]$.

Sample Spaces

Choosing sample space Ω to represent...

- A die
- A dartboard
- A survey of 5 U.S. senators
- A spoken conversation

Events

An event is a set of outcomes.

- Rolling an even number.
- Hitting a bullseye.
- Starting a sentence with 'Um'.

Forming New Events

$$A = \{1, 2\}, B = \{2, 4, 6\}$$

- A OR B:
- A AND B:
- NOT A:

Event Spaces

Intuition for an event space

- S represents all possible events
- However, there are technical reasons that some sets of outcomes must be excluded (non-measurable).

Requirements for a σ-algebra

- Nonempty: $S \neq \emptyset$
- Closed under complements: If $A \in S$ then $A^C \in S$
- Closed under countable unions: if $A_1, A_2, ... \in S$ then $A_1 \cup A_2 \cup ... \in S$

Probability Measure

Definition of a probability measure

A function $P : S \to \mathbb{R}$ is a probability measure if:

- **Axiom 1.** There are no negative probabilities: $\forall A \in S : P(A) \geq 0$
- **Axiom 2.** The probability that *some* outcome occurs is one: $P(\Omega) = 1$
- **Axiom 3.** Probabilities "add up": If $A_1, A_2, ...$ is a countable set of disjoint events, then $P(A_1 \cup A_2 \cup ...) = \sum_i P(A_i)$

Probability Measure

Definition of a probability measure

A function $P : S \to \mathbb{R}$ is a probability measure if:

- **Axiom 1.** There are no negative probabilities: $\forall A \in S : P(A) \geq 0$
- **Axiom 2.** The probability that *some* outcome occurs is one: $P(\Omega) = 1$
- **Axiom 3.** Probabilities "add up": If $A_1, A_2, ...$ is a countable set of disjoint events, then $P(A_1 \cup A_2 \cup ...) = \sum_i P(A_i)$

Learnosity Concept Check

You run an experiment in which you flip 3 coins, and each one lands heads or tails. Your sample space can be written: $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. Assume all outcomes have equal probability.

- (Assuming that you include all possible sets) what is the size of the event space?
- How many outcomes are in the event that the first coin lands heads?
- If A is the event that the first coin lands heads, and B is the event that the second and third coin land the same way, what is the probability of $A \cup B$

A Single Coin Example

A fair coin

Suppose that a coin is fair. Then, let H be the event that the coin comes up "heads" and T be the event that it comes up "tails."

- What is the sample space, Ω?
- What is the probability measure for any single outcome, ω?
- How many outcomes are in event H?
- If you apply the probability measure to each outcome in H, what do you compute for $P(H)$?

Learnosity: A Double Coin Example

Two fair coins

Suppose that you have two coins, and both are fair. Then, let H be the event that when you toss both coins, at least one comes up heads.

- What is the sample space, Ω?
- What is the probability measure for any particular outcome, ω? Are any of these more likely than any other?
- How many outcomes are in event H?
- If you apply the probability measure to each outcome in H, what do you compute for $P(H)$?

Reading: Properties of Probability

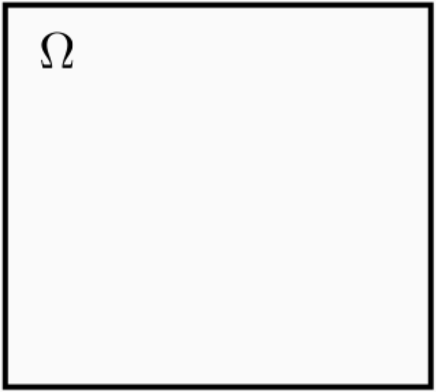
Reading: Properties of Probability

Read page 8 through the end of page 10.

Deriving Properties of Probability

A Helpful Analogy

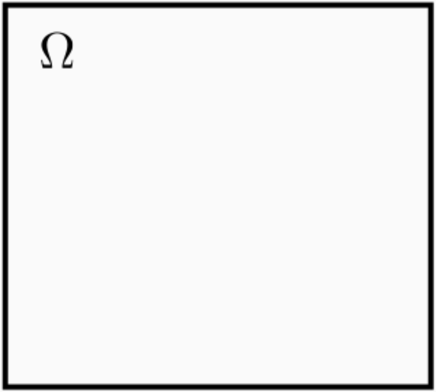
Probability $\Longleftrightarrow$ **Area**



The Subset Subtraction Rule

Given $A, B \in S$, $A \subseteq B$. Compute $P(B \setminus A)$.

Ex: A = being over 7', B = being over 6'.



The Complement Rule

Given $A \in S$. Compute $P(A^C)$.

Basic Properties of Probability

Given a probability space (Ω, S, P), For all events $A, B \in S$:

$$\text{If } A \subseteq B\text{, then } P(A) \leq P(B) \qquad (\textit{Monotonicity})$$

$$\text{If } A \subseteq B\text{, then the probability assigned to } B \qquad (\textit{Subtraction})$$
$$\text{but not } A\text{, written } P(BnA) = P(B) - P(A)$$

$$0 \leq P(A) \leq 1 \qquad (\textit{Probability Bounds})$$

$$P(A^C) = 1 - P(A) \qquad (\textit{Compliments})$$

Learnosity Concept Check - Addition Rule

Here is a proof of the addition rule, select the best justification for each line.

$$
\begin{aligned}
A \cup B &= (A \setminus B) \cup (A \cap B) \cup (B \setminus A) \\
P(A \cup B) &= P(A \setminus B) + P(A \cap B) + P(B \setminus A) \\
&= P(A \setminus B) + P(A \cap B) + P(B \setminus A) + P(A \cap B) - P(A \cap \\
&= P\big((A \setminus B) \cup (A \cap B)\big) + P\big((B \setminus A) \cup (A \cap B)\big) - P(A \cap \\
&= P(A) + P(B) - P(A \cap B)
\end{aligned}
$$

Options (to be scrambled):

- Partition Rule 2
- Axiom 3, countable additivity.
- Because adding something and also subtracting it doesn't change anything.
- Axiom 3, countable additivity.
- Partition Rule 1.

Learnosity Concept Check - Applying the Addition Rule

Applying the Addition Rule

A station along Route 66 sells gas and postcards. The probability that a customer buys postcards is .4. The probability that a customer leaves without buying anything is .3. The probability that the customer buys both gas and postcards is .2. What is the probability that the customer buys gas?

Answer: .5

Conditional Probability

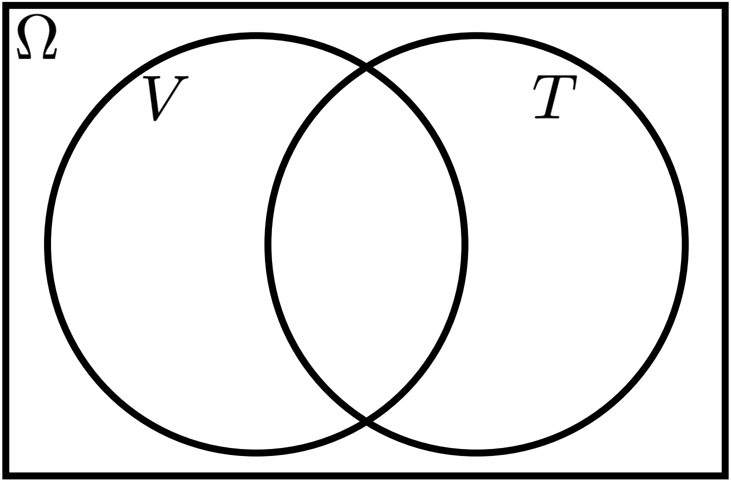
Conditional Probability, Part I

Conditional probability is a rescoping of the sample space from Ω—every possible outcome—to a smaller set called the *conditioning set*

- What is the probability that a student on campus is a volleyball player (V)?
- What is the probability that a student on campus is a volleyball player (V), given that they are 6'3' (T)'?

Uses information to produce a *better* statement of the likelihood of an event

Think of a Dartboard...

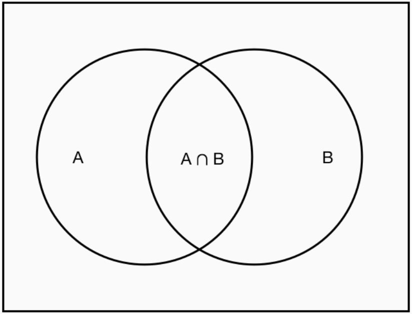


Conditional Probability, Part II

Conditional probability

For events $A, B \in \Omega$ with $P(B) > 0$, define the *conditional probability* of A given B to be

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

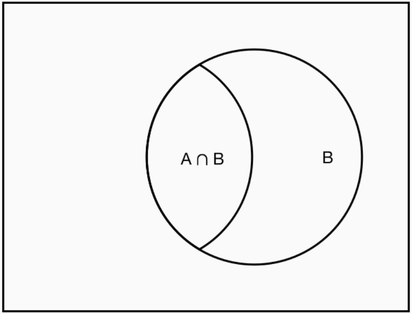


Conditional Probability, Part III

Conditional probability

For events $A, B \in \Omega$ with $P(B) > 0$, define the *conditional probability* of A given B to be

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



Conditional Probability, Part IV

Multiplicative Law of Probability

Rearranging the conditional probability statement minimally produces the *Multiplicative Law of Probability*.

For $A, B \in \Omega$ with $P(B) > 0$,

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

Conditional Probability, Part IV

Multiplicative Law of Probability

Rearranging the conditional probability statement minimally produces the *Multiplicative Law of Probability*.

For $A, B \in \Omega$ with $P(B) > 0$,

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

Learnosity Check

Suppose that the probability that a student plays volleyball is .4, the probability that the student is tall is .5, and the probability that the student is both tall and plays volleyball is .3.

What is the conditional probability that the student plays volleyball, given that they are tall?

Answer: .6

Reading: Bayes' Rule

Reading: Bayes' Rule

Read from the top of page 11 to the end of page 13.

Total Probability

Total Probability, Part I

The Law of Total Probability

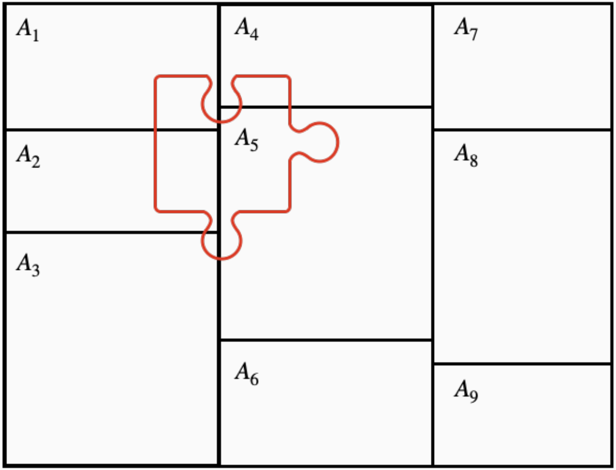
If $A_1, A_2, \ldots, A_n$ is a partition of Ω, and if $B \in \Omega$, then

$$P(B) = \sum_i P(B \cap A_i)$$

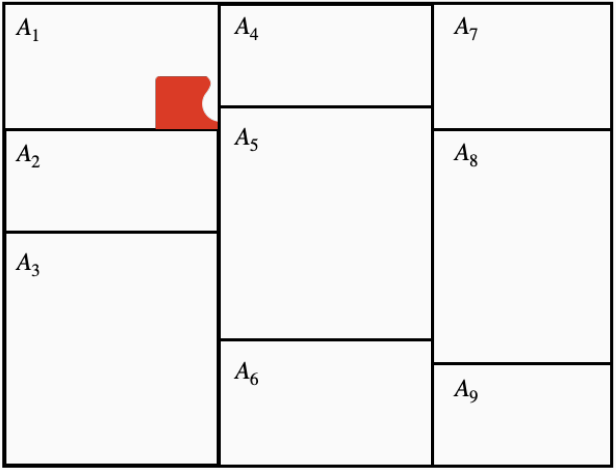
If there is a positive probability for each partition A_i, then by the *Multiplicative Law of Probability*

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

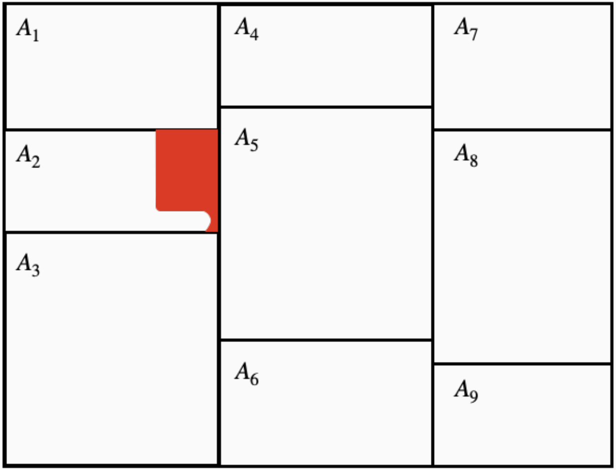
Total Probability, Part II



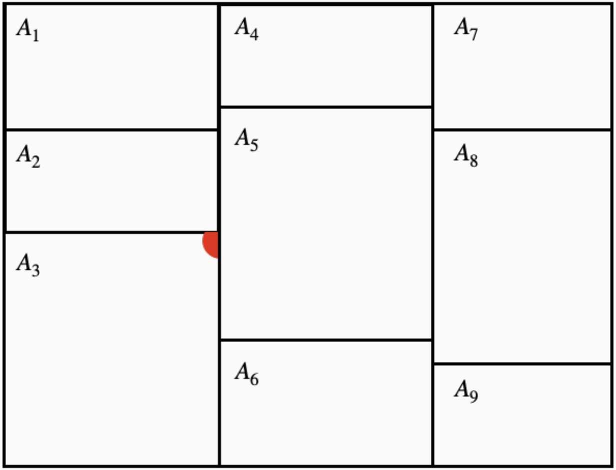
Total Probability, Part III



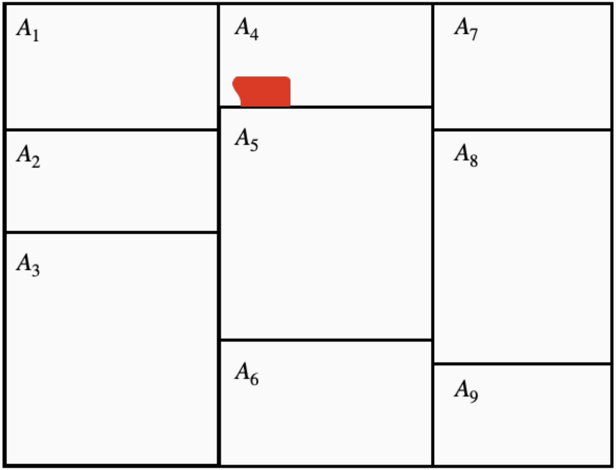
Total Probability, Part IV



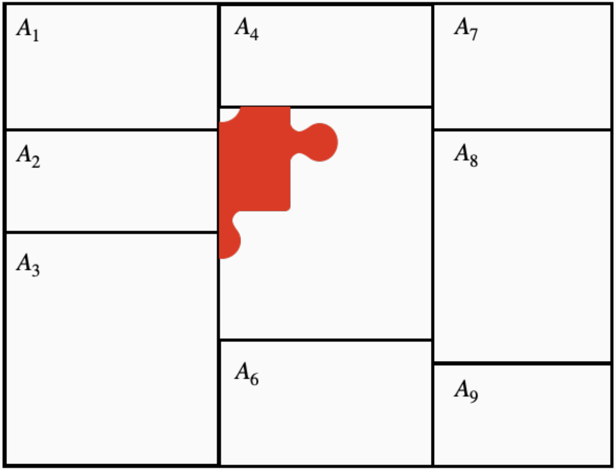
Total Probability, Part V



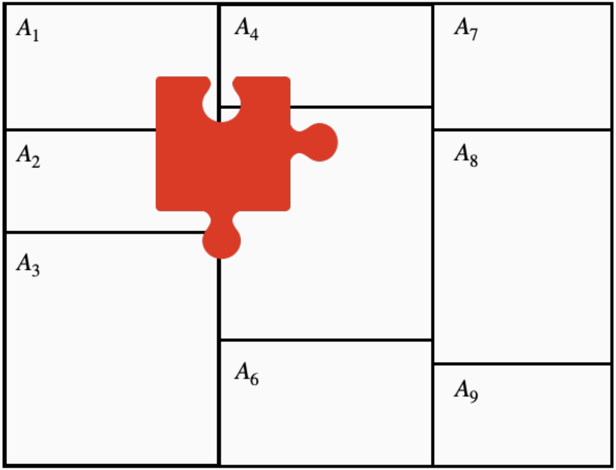
Total Probability, Part VI



Total Probability, Part VII



Total Probability, Part VIII



Total Probability, Part I

The Law of Total Probability

If $A_1, A_2, \ldots, A_n$ is a partition of Ω, and if $B \in \Omega$, then

$$P(B) = \sum_i P(B \cap A_i)$$

If there is a positive probability for each partition A_i, then by the *Multiplicative Law of Probability*

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

Bayes' Rule

Conditional Inverses

Does $P(X|Y) = P(Y|X)$?

Conditional Inverses

Does $P(X|Y) = P(Y|X)$?

Example:

- $X =$ Visited Berkeley
- $Y =$ Attend MIDS

Deriving Bayes' Rule

$$P(X \cap Y)$$

Bayes' Rule Example

- A rare disease affects 2 out of 10,000 people.
- The test is right 99% of the time for healthy people.
- The test is right 100% of the time for sick people.

Given that you get a positive test, what is the probability that you have the disease?

- Let D be the event you have the disease
- Let T be the event the test is positive.

Bayes' Rule Example Cont.

We know: $P(D) = .0002, P(D^C) = .9998$

$P(T|D) = 1, P(T|D^C) = .01$

We want: $P(D|T)$

Bayesian Statistics

Let H be the event a hypothesis is true.

- We begin with prior (subjective) belief $P(H)$

Let D be the data we collect

- Use Bayes' Rule to compute posterior belief, $P(H|D)$.

Reading: Independence

Reading: Independence

Read section 1.1.4, Independence of Events

Independence

Independence

Independence of events

Events $A, B \in S$ are *independent* if $P(A \cap B) = P(A)P(B)$.

Independence

Independence of events

Events $A, B \in S$ are *independent* if $P(A \cap B) = P(A)P(B)$.

Theorem: conditional probability and independence

For $A, B \in S$ and $P(B) > 0$, A and B are independent if and only if $P(A|B) = P(A)$.

Conditional Probability and Information

- Let C be event that our subscriber cancels.
- Let W be the event our subscriber waited 2 hours on the phone.

$$P(C) < P(C|W)$$

W gives us *information* about C.

Understanding Independence

- Let C be event that our subscriber cancels.
- Let H be the event our subscriber hates cilantro.

Is $P(C)$ different from $P(C|H)$?

Importance of Independence

Later in the course: In a *random sample* every unit is independent of every other unit.

Survey Example:

- Let D_1 be event that the first respondent has a dog.
- Let D_2 be event that the second respondent has a dog.
- $\ldots$

Framing the Following Segment

Fatal Officer Involved Shootings

We're going to take on an important, but challenging topic here – the use of police force.

- In particular, we're going to read a paper that investigates whether office race plays any role in disparate use of force against people of color
- This is an important, durable question and one that at the time that we're putting the course together is the predominant political issue.

Fatal Officer Involved Shootings

As faculty, we stand is solidarity with people of color, without exception and without qualification.

- We're going to read an academic, data-based conversation where the authors come to the conclusion that there is no evidence of differential treatment of people of color by white officers.
- And, we're going to critically engage with this article, to assess whether the data supports this claim

Probability and Police Shootings

Are Black citizens subjected to unfairly violent treatment by law enforcement?

Johnson et al. (2019)

Johnson et al. (2019)

Are minority groups subjected to unfairly violent treatment by law enforcement?

1. Are Black civilians overrepresented in FOIS?
2. If so, is this because of racial discrimination by White officers?

- Develop Database that is a near-total recording of all *fatal officer involved shootings* (FOIS) in the US in 2015.
- "Model" whether race of civilian is a factor that predicts being involved in a FOIS.

Johnson et al. (2019) Conclusions

Study conclusions

- *"We find no overall evidence of anti-Black or anti-Hispanic disparities in fatal shootings."*
- *"White officers are not more likely to shoot minority civilians than non-White officers."*

Probability Translation

Crystal clear probability statement about officer race:

$$P(\text{shot}|\text{minority civilian, white officer}, X) \leq P(\text{shot}|\text{minority civilian, minority officer}, X)$$

Question: Is it possible to justify this statement using only data of officer shootings?

Measuring Fairness

- Twenty-six percent of civilians killed by police shootings in 2015 were Black.
- The U.S. population is 12% Black.

- White and Black civilians have different rates of exposure to situations that might result in FOIS.