The Big Picture

30 April 2020
End of semester updates
End of semester updates

DON'T PANIC
Assignment 8 is due today.

For any assignments (or Exam 2) that you haven’t submitted, the final deadline is the end of study week.

On Monday we will have an in-class review for Exam 3.

Thank you for your patience as I’ve fallen behind on grading this semester. I’ll do my best to have everything back to you in enough time to study.
A last word on probability: *Simpson’s paradox*
The Restaurant at the End of the Semester
This final section of the course, on probability, has been a lot. Let’s break for a nice dinner.

Where do we go?

(Remind the *Before Times*, when you went out for dinner?)
The restaurant guide tells you:

Tom’s Bistro is recommended by a higher percent of both locals and out-of-towners than JJ’s Diner is!

However, JJ’s is recommended by a higher percentage of all users!
The restaurant guide tells you:

Tom’s Bistro is recommended by a higher percent of both locals and out-of-towners than JJ’s Diner is!

However, JJ’s is recommended by a higher percentage of all users!

This doesn’t seem right.
We have entered the strange-but-true world of Simpson’s Paradox, where

a restaurant can be both better and worse than its competitor,
exercise can lower and increase the risk of disease,
and the same data set can be used to prove two opposing arguments.
**Simpson’s Paradox** occurs when trends that appear when a dataset is separated into groups reverse when the data are aggregated.
<table>
<thead>
<tr>
<th></th>
<th>Recommend JJ’s Diner</th>
<th>Recommend Tom’s Bistro</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Locals</strong></td>
<td>50 / 150 = 30%</td>
<td>180 / 360 = 50%</td>
</tr>
<tr>
<td><strong>Out-of-towners</strong></td>
<td>200 / 250 = 80%</td>
<td>36 / 40 = 90%</td>
</tr>
<tr>
<td><strong>Combined</strong></td>
<td>250 / 400 = 62.5%</td>
<td>216 / 400 = 54%</td>
</tr>
</tbody>
</table>
Instead of just looking at percentages / probabilities, look at the sample size.

<table>
<thead>
<tr>
<th></th>
<th>JJ’s Diner</th>
<th>Tom’s Bistro</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Locals</strong></td>
<td>50 / 150 = 30%</td>
<td>180 / 360 = 50%</td>
</tr>
<tr>
<td><strong>Out-of-towners</strong></td>
<td>200 / 250 = 80%</td>
<td>36 / 40 = 90%</td>
</tr>
<tr>
<td><strong>Combined</strong></td>
<td>250 / 400 = 62.5%</td>
<td>216 / 400 = 54%</td>
</tr>
</tbody>
</table>

Tom’s has way more reviews from locals while JJ’s has way more from out-of-towners.
Kidney stones
How can this happen?

Consider the data generation process ("causal model"), informed by domain knowledge.

Small stones are considered less serious cases, and treatment A is more invasive than treatment B.

Doctors are more likely to recommend the inferior treatment, B, for small kidney stones, where the patient is more likely to recover successfully in the first place because the case is less severe.

For large, serious stones, doctors more often go with the better – but more invasive – treatment A.

Even though treatment A performs better on these cases, because it is applied to more serious cases, the overall recovery rate for treatment A is lower than treatment B.
That is, the size – seriousness – of the kidney stone is a **confounding variable** because it affects both the **independent variable** (treatment method) and the **dependent variable** (recovery).

We don’t see this in the data table!
To determine which treatment actually works better, we need to control for the confounding variable by segmenting the two groups and comparing recovery rates within groups rather than aggregated over groups:

If you have a small stone, you prefer treatment A.
If you have a large stone, you also prefer treatment A.
Since you must have a small or a large stone, you always prefer treatment A.
“Paradox” resolved!
A Ford, not a Lincoln
Gerald Ford lowered taxes for every income group.
Gerald Ford raised taxes nation-wide.

Both true!
<table>
<thead>
<tr>
<th>Adjusted Gross Income</th>
<th>Income</th>
<th>Tax</th>
<th>Rate</th>
<th>Income</th>
<th>Tax</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>under $ 5,000</td>
<td>41,651,643</td>
<td>2,244,467</td>
<td>.054</td>
<td>19,879,622</td>
<td>689,318</td>
<td>.035</td>
</tr>
<tr>
<td>$ 5,000 to $ 9,999</td>
<td>146,400,740</td>
<td>13,646,348</td>
<td>.093</td>
<td>122,853,315</td>
<td>8,819,461</td>
<td>.072</td>
</tr>
<tr>
<td>$ 10,000 to $14,999</td>
<td>192,688,922</td>
<td>21,449,597</td>
<td>.111</td>
<td>171,858,024</td>
<td>17,155,758</td>
<td>.100</td>
</tr>
<tr>
<td>$ 15,000 to $99,999</td>
<td>470,010,790</td>
<td>75,038,230</td>
<td>.160</td>
<td>865,037,814</td>
<td>137,860,951</td>
<td>.159</td>
</tr>
<tr>
<td>$ 100,000 or more</td>
<td>29,427,152</td>
<td>11,311,672</td>
<td>.384</td>
<td>62,806,159</td>
<td>24,051,698</td>
<td>.383</td>
</tr>
<tr>
<td>Total</td>
<td>880,179,247</td>
<td>123,690,314</td>
<td></td>
<td>1,242,434,934</td>
<td>188,577,186</td>
<td></td>
</tr>
<tr>
<td>Overall Tax Rate</td>
<td></td>
<td>.141</td>
<td></td>
<td></td>
<td>.152</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adjusted Gross Income</th>
<th>1974</th>
<th></th>
<th>1978</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Income</td>
<td>Tax</td>
<td>Income</td>
<td>Tax</td>
</tr>
<tr>
<td>under $ 5,000</td>
<td>41,651,643</td>
<td>2,244,467</td>
<td>19,879,622</td>
<td>689,318</td>
</tr>
<tr>
<td>$ 5,000 to $ 9,999</td>
<td>146,400,740</td>
<td>13,646,348</td>
<td>122,853,315</td>
<td>8,819,461</td>
</tr>
<tr>
<td>$ 10,000 to $14,999</td>
<td>192,688,922</td>
<td>21,449,597</td>
<td>171,858,024</td>
<td>17,155,758</td>
</tr>
<tr>
<td>$ 15,000 to $99,999</td>
<td>470,010,790</td>
<td>75,038,230</td>
<td>865,037,814</td>
<td>137,860,951</td>
</tr>
<tr>
<td>$ 100,000 or more</td>
<td>29,427,152</td>
<td>11,311,672</td>
<td>62,806,159</td>
<td>24,051,698</td>
</tr>
<tr>
<td>Total</td>
<td>880,179,247</td>
<td>123,690,314</td>
<td>1,242,434,934</td>
<td>188,577,186</td>
</tr>
<tr>
<td>Overall Tax Rate</td>
<td>.141</td>
<td></td>
<td>.152</td>
<td></td>
</tr>
</tbody>
</table>

The overall tax rate is a function both of the *individual bracket tax rates*, and also the *amount of taxable income in each bracket*.

Due to inflation (or wage increases), there was more income in the upper tax brackets with higher rates in 1978 and less income in lower brackets with lower rates.

Therefore, the overall tax rate increased!
Why does Simpson’s Paradox matter?
The data we are shown is not all the data there is.

We can’t be satisfied only with the numbers or a figure; we need to consider the data generation process – the causal model – responsible for the data.
So, where do we eat?

Whether to combine the data or look at them independently depends on the problem you’re looking at.

In this case, are we bringing just locals to eat? Just out-of-towners? A mix?

If it’s a mix – or you’re not sure – it makes sense to aggregate the data.

And, besides, JJ’s has great waffles.
The Big Picture
Imagine what it must have been like to discover all of the ideas in this class.
**STEP 1: INSIGHT**

*My god. I wonder if this is true.*

**STEP 2: RESISTANCE**

Impossible! Insane! It’s not just incorrect; it’s an entirely new category of stupid!

**STEP 3: DEBATE**

It looks right, but it can’t be right. Perhaps we could restructure all of mathematics in a way that makes it wrong.

**STEP 4: ADDITIONAL DECADES OF DEBATE.**

**STEP 5: CHANGING OF THE GUARD**

I say your mother’s a whore. With Leibniz, the faculty of Cambridge.

I will never understand it. I will never believe it. As I go into death, with my final breath I spit on your theorem.

**STEP 6: TRANSMISSION TO STUDENTS.**

How do you not get this concept? We spent an hour on it yesterday.
We started with the idea of a set — our all-purpose mathematical collection.

Sets are used throughout computer science and mathematics. You’ll see them in computer algorithms, in machine learning papers, in theoretical proofs.
We saw how we can manipulate sets, looking at their intersections and unions, and drawing diagrams to represent the results.
the venn diagram of people protesting for salons and restaurants to be reopened and people who don't tip servers, hairdressers, or nail technicians because "it's not my responsibility to pay their wage" is a circle
Well, actually that would be an Euler diagram. A Venn diagram always shows all the possibilities, even if some are empty.
We also saw how to use sets to think about all possible subsets, the *power set*. 
Cantor’s theorem: $|S| < |\mathcal{P}(S)|$

This isn’t just a claim about sets; this turns out to be the basis for proving that there exist problems that can never be solved by a computer!
How could we ever know that there are problems that can’t be solved?

You’ll see that proof in CMPU 240, but to get there you need more background than just set theory!
First, we need to learn how to prove results with certainty.

Otherwise, how can we be confident that we’re right about anything?
We also should be sure we have some rules about reasoning itself.

Let’s add some logic into the mix.
Let’s study how we can model relationships, and particularize them to mathematical functions.

We can define these a lot like we define functions in Racket!

In fact, we’ll see that programming languages like Racket are built on the mathematical ideas of this course.
Now we need to learn how to prove things about processes that proceed step-by-step.

So, let’s learn induction.
Combinatorics lets us look at ways of counting possibilities.

E.g., in how many different orders could we have learned these topics
Which we then formalize into a notion of how likely different events are – probability.
There’s much more to each of these topics – and no end to the topics they’ll let you learn about.

Theory is about exploring and experimenting.
Discovery isn’t a straight road

The Transfăgărășan, a road in the Carpathian Mountains of Romania
c. 350 BCE: Aristotle discusses several modes of logical syllogism in *Prior Analytics*, considered by many to be the first treatise on formal logic. The four Aristotelian forms were first described in this work.

1202: Leonardo of Pisa, better known as Fibonacci, publishes Liber Abaci, introducing the Hindu-Arabic numeral system to the Western world. The book also included a puzzle about population growth rates, whose solution is the Fibonacci sequence.
1654: Although the general technique had been in use for some time, mathematical induction is used in its first modern form in Blaise Pascal’s “Traité du triangle arithmétique”.

1834: Peter Dirichlet gives the first instance of the use of the pigeonhole principle. Dirichlet referred to this principle as the “box principle” or “drawer principle,” and so some mathematicians still refer to it as “Dirichlet's box principle.”
1854: **George Boole** publishes *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*, outlining a mathematical system for formal logic. His approach was based on representing logical notation as arithmetical statements involving addition and multiplication and ultimately gave rise to propositional logic.

1860: In his work “On the Syllogism, No. IV, and on the logic of relations”, **Augustus de Morgan** introduces binary relations and their properties as an attempt to move beyond the four Aristotelian forms described over two millenia earlier.
1874: *Georg Cantor* proves that $|\mathbb{N}| \neq |\mathbb{R}|$ in his paper “Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen.” The proof Cantor outlined in this paper is not his famous diagonal argument, but rather a more technical argument involving infinite sequences.
1879: *Gottlob Frege* publishes *Begriffsschrift* (“Concept Script”), a formal notation for logical reasoning that is considered the precursor of first-order logic and other modern logical systems. Frege's system was based on implications and negations only and contained one of the earliest references to quantifiers.

1880: *John Venn* publishes “On the Diagrammatic and Mechanical Representation of Propositions and Reasonings,” containing a detailed catalog of ways to use overlapping regions to describe how different logical statements overlapped with one another.
1888: *Guisepppe Peano* publishes “*Calcolo geometrico secondo l’Ausdehnungslehre di H. Grassmann,*” introducing the symbols $\cup$ and $\cap$ for union and intersection.

1891: *Georg Cantor* introduces the diagonal argument in his paper “Über eine elementare Frage der Mannigfaltigkeitslehre.”

1908: *Ernst Zermelo* publishes “*Untersuchungen über die Grundlagen der Mengenlehre,*” introducing the idea that sets could be defined via a small number of axioms describing their behavior. This paper introduced the now familiar notation of describing sets using curly braces.
1935: **Gehard Gentzen** publishes “Untersuchungen ueber das logische Schliessen,” which, among other things, introduces the symbol ∀ as the universal quantifier. Previously, authors had used a bunch of other symbols, such as Π or (x) to indicate universal quantification. The existential quantifier ∃ had been introduced at least thirty years before this.

1954: A group of French mathematicians working under the pseudonym “**Nicolas Bourbaki**” introduce the terms “injection,” “surjection,” and “bijection” in their highly influential textbook Théorie des ensembles. The Bourbaki group has since had a profound influence on the standard mathematics curriculum.
These developments lead us to modern computability theory, complexity theory, Gödel’s incompleteness theory, and more!
Where to go from here?
Congratulations on making it this far!
Racket
Set theory
Power sets
Cantor’s theorem
Venn diagrams
Euler diagrams
Ordered tuples
Direct proofs
Proof by contradiction
Proof by contrapositive
Propositional logic
First-order logic
Logic translations
Logical negations
Quantifiers and scope
Binary relations
Equivalence relations
Relation composition
Functions
Injections
Surjections
Bijections
Inverse functions
Recursive definitions
Weak induction
Strong induction
Structural induction
Peano axioms
Combinations
Permutations
Simulations
Probability
Expected value
Simpson’s paradox
You’ve done more than tick off a bunch of boxes.

You’ve given yourself the foundation to tackle problems from all over computer science.
CMPU 240: Language Theory and Computation

Asks the fundamental theoretical question “what problems can we solve with computers?” This course builds directly on 145, using set theory to represent problems as languages of strings, giving definitions in logic, and writing proofs.
CMPU 241: Algorithms

A mix of theory and practice, where you’ll learn about computational complexity: Which decidable problems can we compute efficiently and which are so inefficient they might as well be undecidable?
CMPU 365: Artificial Intelligence

The work of many mathematicians and theoretical computer scientists, including Alan Turing, has been motivated by the question of how we might create a thinking machine.

Decades of work in this field has been devoted to knowledge representation and reasoning, using logic.

Modern machine learning is built on probability, especially Bayes’ rule.

The language of AI programming is Lisp, the basis for Racket.
CMPU 224: Computer Organization

You don’t need to be a theoretician to love computer science! See how we use the ideas from this course to make real computers work. Expect to see more Boolean algebra!
Final thoughts
“A computer is not dependent so much on technology as on ideas.”

W. Daniel Hillis, *The Pattern on the Stone*
CS theory is all about asking what’s possible in computer science.
There’s so much more to explore and so many big questions to ask – *many of which haven’t been asked yet!*
A whole world of theory and practice awaits.
That’s it!
Next time, we’ll review for the final exam by working through practice problems and answering your questions.
Acknowledgments

This lecture incorporates material from:

David Makinson
Will Koehrsen