Welcome to ECE 3077

Intro to Probability and Statistics for ECEs

Instructor: Justin Romberg

May 13, 2014

Course goal:

To develop a mathematical framework for modeling and understanding uncertainty

Probability

What is probability?

Probability

What is probability?

- Notoriously hard to define

Probability

What is probability?

- Notoriously hard to define
- Definitions often circular
  - likelihood
  - chance
  - odds

Probability

What is probability?

- Notoriously hard to define
- Definitions often circular
  - likelihood
  - chance
  - odds

Frequentist interpretation

Probability

What is probability?

- Notoriously hard to define
- Definitions often circular
  - likelihood
  - chance
  - odds

Frequentist interpretation

- **Aristotle:** "the probable is that for which the most part happens"

Probability

What is probability?

- Notoriously hard to define
- Definitions often circular
  - likelihood
  - chance
  - odds

Frequentist interpretation

- **Aristotle:** "the probable is that for which the most part happens"
- **Modern frequentists:** Let n_t denote the total number of "trials" and let n_x denote the number of trials where "event x" occurs. Then the probability of x occurring is given by

$$P(x) = \lim_{n_t \to \infty} \frac{n_x}{n_t}$$

.

Probability

Is the frequentist interpretation adequate?

Probability

Is the frequentist interpretation adequate?

What is the probability that

- the next coin toss will land heads?

Probability

Is the frequentist interpretation adequate?

What is the probability that

- the next coin toss will land heads?
- the next circuit off the assembly line will be faulty?

Probability

Is the frequentist interpretation adequate?

What is the probability that

- the next coin toss will land heads?
- the next circuit off the assembly line will be faulty?
- it will rain tomorrow?

Probability

Is the frequentist interpretation adequate?

What is the probability that

- the next coin toss will land heads?
- the next circuit off the assembly line will be faulty?
- it will rain tomorrow?
- I will die of cancer?

Probability

Is the frequentist interpretation adequate?

What is the probability that

- the next coin toss will land heads?
- the next circuit off the assembly line will be faulty?
- it will rain tomorrow?
- I will die of cancer?
- I am telling a lie?

Probability

Is the frequentist interpretation adequate?

What is the probability that

- the next coin toss will land heads?
- the next circuit off the assembly line will be faulty?
- it will rain tomorrow?
- I will die of cancer?
- I am telling a lie?
- an electron will be “spin up”

Probability

Is the frequentist interpretation adequate?

What is the probability that

- the next coin toss will land heads?
- the next circuit off the assembly line will be faulty?
- it will rain tomorrow?
- I will die of cancer?
- I am telling a lie?
- an electron will be "spin up"
- the "big bang theory" of cosmology is correct?

Probability

Is the frequentist interpretation adequate?

What is the probability that

- the next coin toss will land heads?
- the next circuit off the assembly line will be faulty?
- it will rain tomorrow?
- I will die of cancer?
- I am telling a lie?
- an electron will be "spin up"
- the "big bang theory" of cosmology is correct?
- a flying saucer crashed at Roswell?

“Probability is common sense reduced to computation”

— *Laplace*

“Probability is common sense reduced to computation”

— *Laplace*

Human intuition is very bad when it comes to probability ...

Simpson's "Paradox"

Batting averages for the DJs:

	1995	1996
Derek Jeter	.250	.314
David Justice	**.253**	**.321**

(Ross, *A Mathematician at the Ballpark: Odds and Probabilities for Baseball Fans*, 2004.)

Simpson's "Paradox"

Batting averages for the DJs:

	1995	1996	Combined
Derek Jeter	.250	.314	**.310**
David Justice	**.253**	**.321**	.270

(Ross, *A Mathematician at the Ballpark: Odds and Probabilities for Baseball Fans*, 2004.)

Simpson's "Paradox"

Batting averages for the DJs:

	1995		1996		Combined	
Derek Jeter	12/48	.250	183/582	.314	195/360	**.310**
David Justice	104/411	**.253**	45/140	**.321**	149/551	.270

(Ross, *A Mathematician at the Ballpark: Odds and Probabilities for Baseball Fans*, 2004.)

Simpson's "Paradox": A less whimsical example

Graduate admissions data for UC Berkeley, 1973:

	Men		Women	
	Applicants	Admitted	Applicants	Admitted
University	8442	**44%**	4321	35%

Was gender bias rampant at UC Berkeley in the early 70s?

(Bickel et al, "Sex Bias in Graduate Admissions: Data From Berkeley", *Science*, 1975.)

Simpson's "Paradox": A less whimsical example

Graduate admissions data for UC Berkeley, 1973:

	Men		Women	
	Applicants	Admitted	Applicants	Admitted
Dept. A	825	62%	108	**82%**
Dept. B	560	63%	25	**68%**
Dept. C	325	**37%**	593	34%
Dept. D	417	33%	375	**35%**
Dept. E	191	**28%**	393	24%
Dept. F	272	6%	341	**7%**

(Bickel et al, "Sex Bias in Graduate Admissions: Data From Berkeley", *Science*, 1975.)

Beware of narratives

A study of the incidence of kidney cancer in the 3141 counties in the US shows that the counties in which the incidence of kidney cancer is the *lowest* are mostly

- rural,
- sparsely populated,
- located in traditionally Republican states in the Midwest, South, and West.

Why?

(Kahneman, *Thinking, Fast and Slow*, 2011)

Beware of narratives

The same study of the incidence of kidney cancer in the 3141 counties in the US shows that the counties in which the incidence of kidney cancer is the *highest* are mostly

- rural,
- sparsely populated,
- located in traditionally Republican states in the Midwest, South, and West.

How can this be?

(Kahneman, *Thinking, Fast and Slow*, 2011)

"Law of Small Numbers"

The overall rate of kidney cancer in the US is 15 per 100,000 ($p = 0.00015$)

Armstrong County, Texas has a population of $n = 2148$

Probability of 0 cases: 0.72

Probability of 1 case: 0.24
(3.1x national incidence rate)

Probability of 2 cases: 0.038
(6.2x national incidence rate)

$\Rightarrow$ 28% chance of
incidence rate $\geq 3\times$ national average

Law of Large Numbers

The overall rate of kidney cancer in the US is 15 per 100,000
($p = 0.00015$)

Los Angeles County, California has a population of $n = 9.8$ million

mean cases $= 1479$
std dev $= 38.3$

$\Rightarrow$ odds are less than 1 in 10^{1277} that
incidence rate $\geq 3\times$ national average

Confirmation Bias

the tendency to seek out evidence that confirms our existing beliefs or hypotheses

Confirmation Bias

Is there such a thing as a "hot hand" in basketball?

Every fan and player alive swears there is ...

Confirmation Bias

Is there such a thing as a "hot hand" in basketball?

Tversky says: "No"

Kobe is skeptical

Gilovich, Vallone, and Tversky charted shots for the Philadelphia 76ers for the entire 1980-81 season, and found absolutely no evidence of the influence of the success/failure of previous shots on a current shot

(Gilovich, Vallone, and Tversky, "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology*, 1985.)

Confirmation Bias

Is there such a thing as a "hot hand" in basketball?

Tversky says: "No"

Kobe is skeptical

They repeated the experiment several seasons later with the Boston Celtics with the same results.

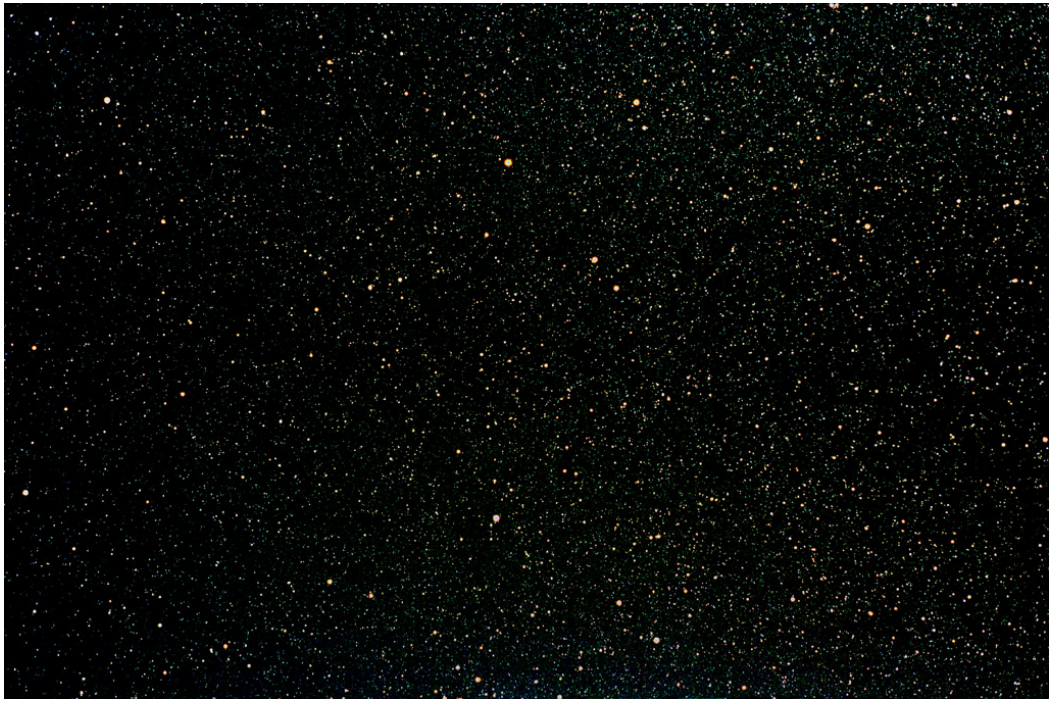
(Gilovich, Vallone, and Tversky, "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology*, 1985.)

Recognizing randomness when we see it

We seem to be hard-wired to look for patterns, and are pretty good at finding them even when none exist...

Recognizing randomness when we see it

We seem to be hard-wired to look for patterns, and are pretty good at finding them even when none exist...



The constellation Pegasus

Apple iPod in 2001

Probabilistic inference

- The incidence rate for disease X is 15 in 100,000
- There is a test for disease X which is 95% accurate — if you have X, there is a 95% chance the test comes back positive, and if you don't have X, there is a 95% chance it comes back negative.

Probabilistic inference

- The incidence rate for disease X is 15 in 100,000
- There is a test for disease X which is 95% accurate — if you have X, there is a 95% chance the test comes back positive, and if you don't have X, there is a 95% chance it comes back negative.

You test positive for X. What is the probability that you actually have the disease?

Probabilistic inference

- The incidence rate for disease X is 15 in 100,000
- There is a test for disease X which is 95% accurate — if you have X, there is a 95% chance the test comes back positive, and if you don't have X, there is a 95% chance it comes back negative.

You test positive for X. What is the probability that you actually have the disease?

Answer: 0.28%

Probabilistic inference

- The incidence rate for disease X is 15 in 100,000
- There is a test for disease X which is 95% accurate — if you have X, there is a 95% chance the test comes back positive, and if you don't have X, there is a 95% chance it comes back negative.

You test positive for X. What is the probability that you actually have the disease?

Answer: 0.28%

The base rate of 15/100000 carries a lot of information

Probability can be counter-intuitive

Entire industries are based on the fact that humans are bad at probability

Probability can be counter-intuitive

Entire industries are based on the fact that humans are bad at probability

Probability can be counter-intuitive

Entire industries are based on the fact that humans are bad at probability

All this makes it seem like humans are complete idiots!

So are we?

Sir Francis Galton (1822–1911)

So are we?

Sir Francis Galton (1822–1911)

- cousin of Charles Darwin

So are we?

Sir Francis Galton (1822–1911)

- cousin of Charles Darwin
- statistician (introduced notions of correlation and standard deviation)

So are we?

Sir Francis Galton (1822–1911)

- cousin of Charles Darwin
- statistician (introduced notions of correlation and standard deviation)
- father of eugenics

So are we?

Sir Francis Galton (1822–1911)

- cousin of Charles Darwin
- statistician (introduced notions of correlation and standard deviation)
- father of eugenics

Galton was wary of democracy and distrusted "the mob"

So are we?

Sir Francis Galton (1822–1911)

- cousin of Charles Darwin
- statistician (introduced notions of correlation and standard deviation)
- father of eugenics

Galton was wary of democracy and distrusted "the mob"

How much does this ox weigh?

Vox populi

If we collect hundreds of uneducated farmers from the English countryside (with no particular expertise in weighing oxen), how well will they do ?

Vox populi

If we collect hundreds of uneducated farmers from the English countryside (with no particular expertise in weighing oxen), how well will they do ?

Average (mean) of the guesses: 1,197 pounds

Vox populi

If we collect hundreds of uneducated farmers from the English countryside (with no particular expertise in weighing oxen), how well will they do ?

Average (mean) of the guesses: 1,197 pounds
Actual weight: 1,198 pounds

Vox populi

If we collect hundreds of uneducated farmers from the English countryside (with no particular expertise in weighing oxen), how well will they do ?

Average (mean) of the guesses: 1,197 pounds
Actual weight: 1,198 pounds

The average guess was significantly more accurate than any single guess

Vox populi

If we collect hundreds of uneducated farmers from the English countryside (with no particular expertise in weighing oxen), how well will they do ?

Average (mean) of the guesses: 1,197 pounds
Actual weight: 1,198 pounds

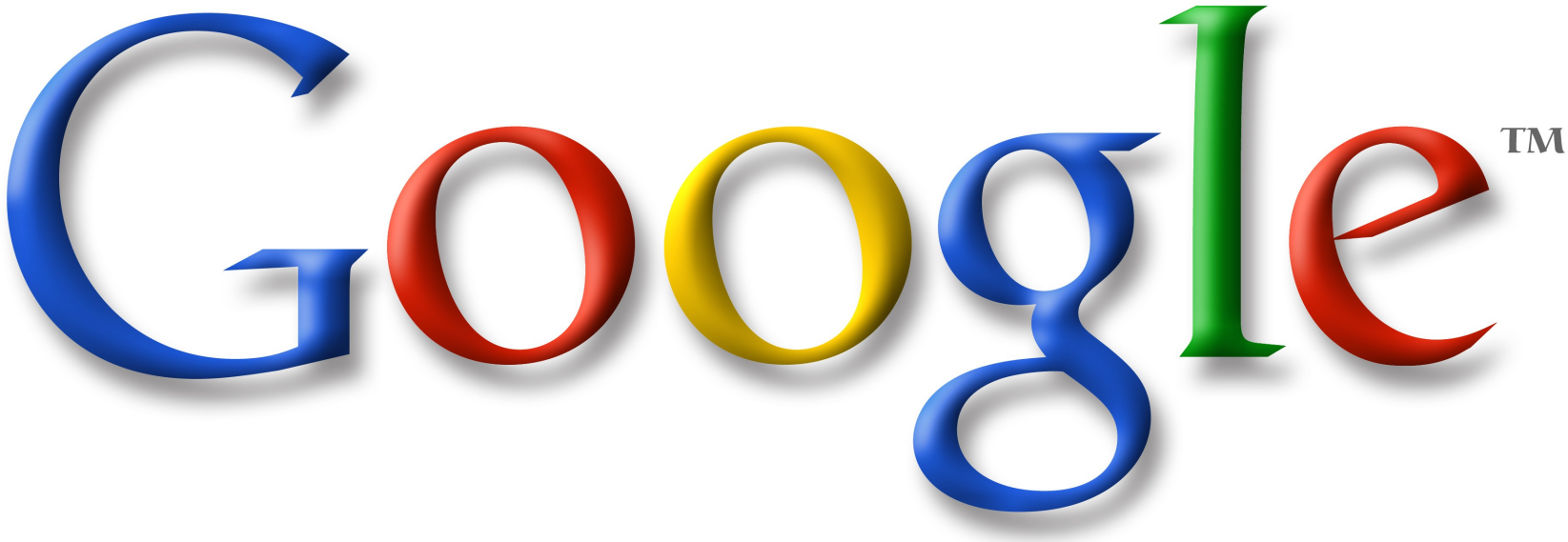
The average guess was significantly more accurate than any single guess

Example of the **wisdom of the crowds**.

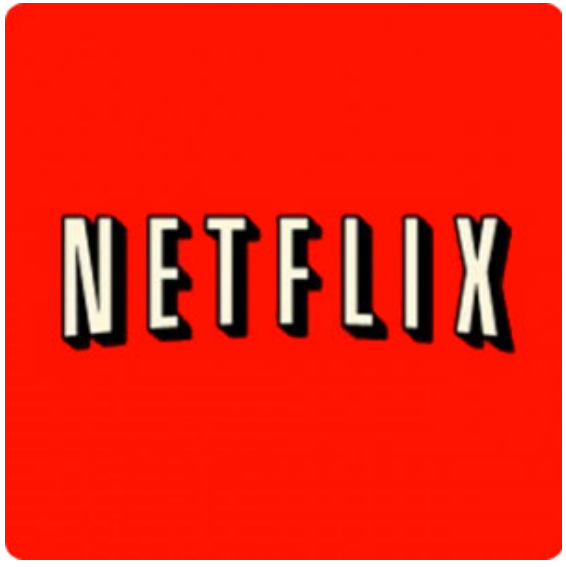
Examples of technology that would be impossible without a clear understanding of probability and statistics are endless ...

Insurance

Web search



Collaborative filtering / recommender systems



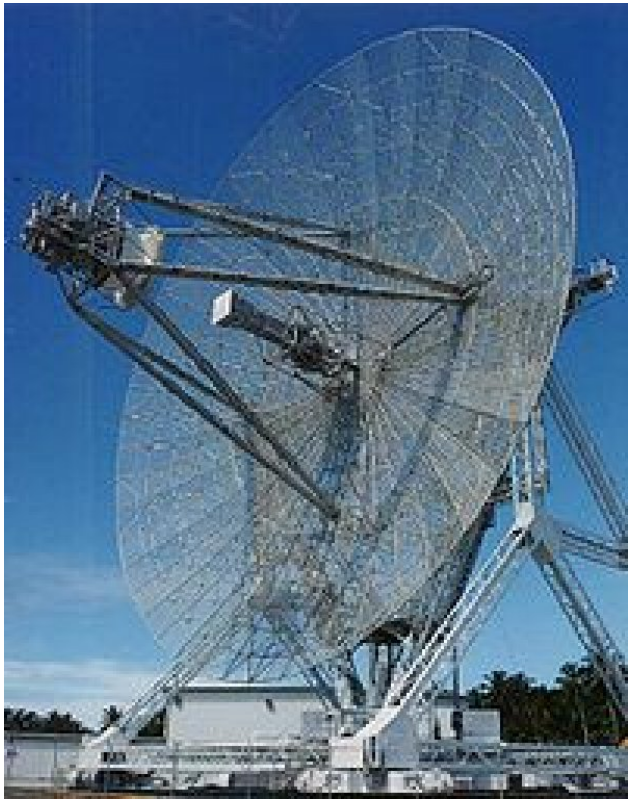
Artificial intelligence

Google's self-driving car

Data compression and encoding

Digital communications

Radar



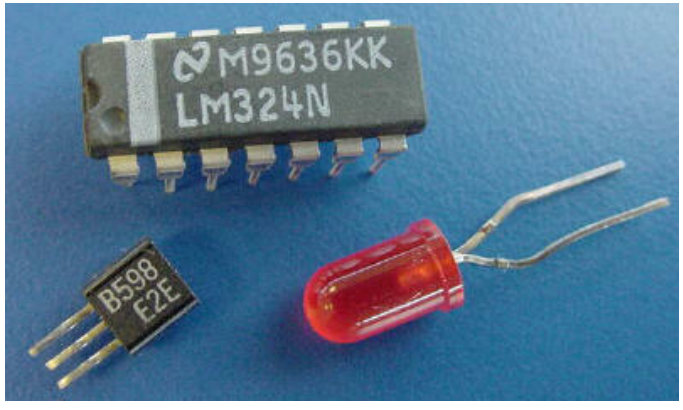
Quantum mechanics

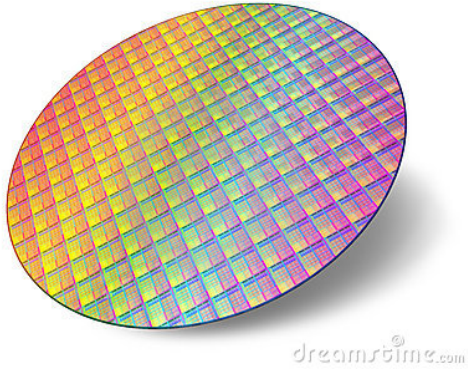
Quantum mechanics describes a world that is inherently probabilistic

Quantum mechanics

Quantum mechanics describes a world that is inherently probabilistic

Without quantum mechanics, there are no semiconductors...

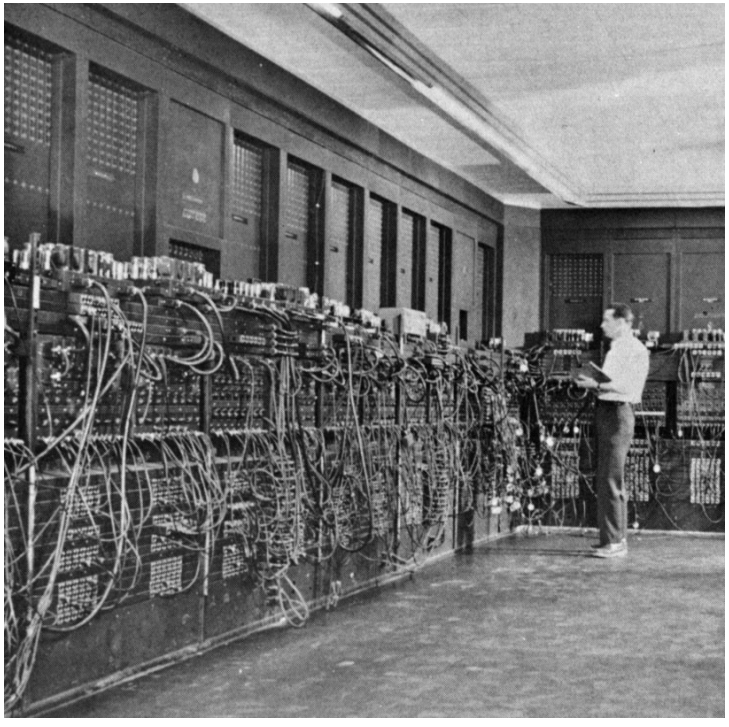




Quantum mechanics

Without semiconductors, we are stuck ...

... with this ...



ENIAC, 1946
$\approx$ 18,000 cubic feet
$\approx$ 500 flops

... instead of this:

Samsung Galaxy S4, 2013
$\approx$ 0.0026 cubic feet
$\approx 50 \cdot 10^9$ flops