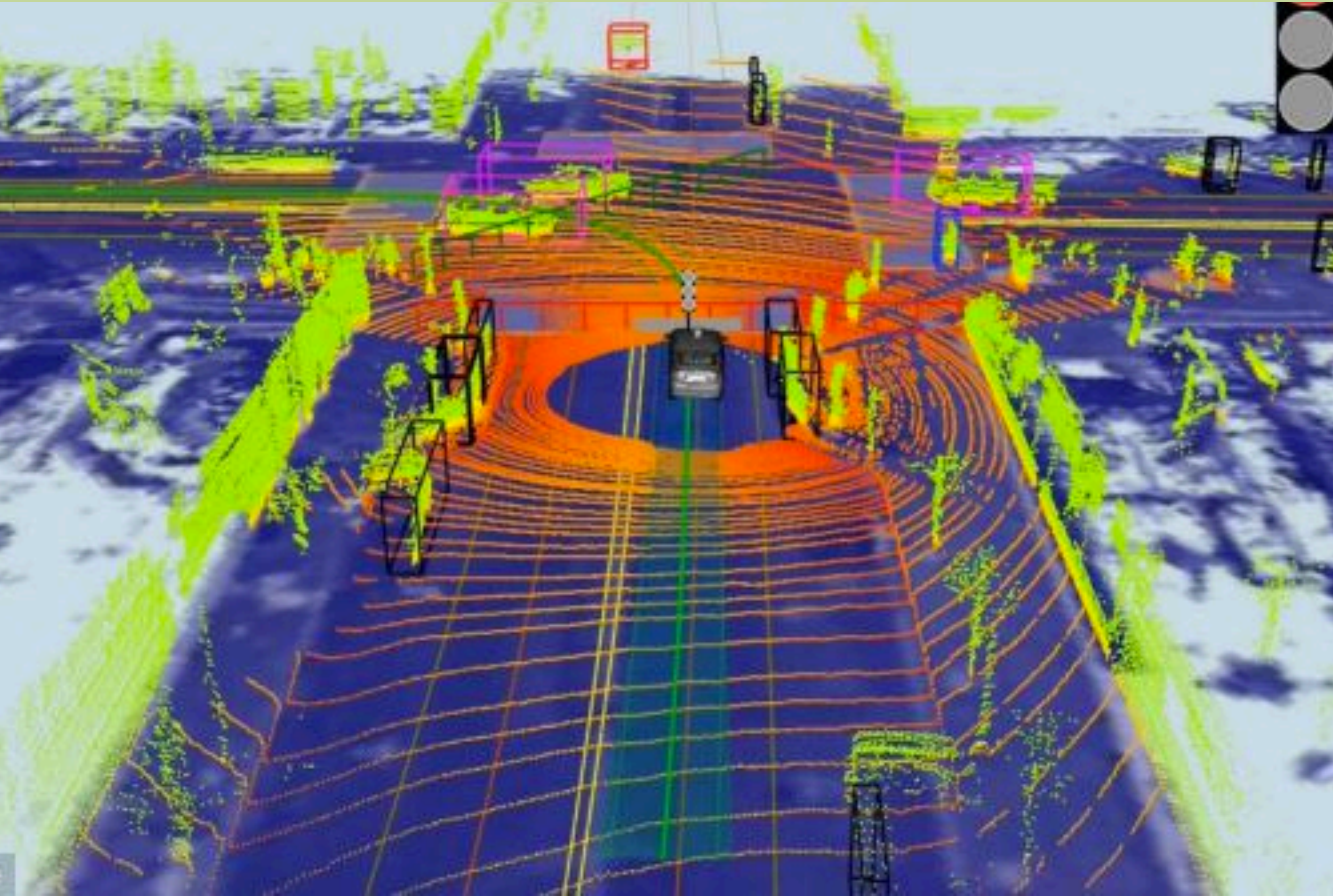
Conditional Probability

CS 106B
Lecture 3
March 28th, 2016

Probability Philosophy

Probabilities in Self Driving Car



Two Sources of Probabilities

1. Randomness in the world

2. Our own uncertainties

End Probability Philosophy

Today's Topics

Last time:
Probability Definition
Probability Identities
Equally Likely Events

Today:
Conditional Probability
Bayes Theorem

Next time:
Independence

Review:

Counting

Counting Elements

Distinct

Indistinct

Distinct + Ordered	Distinct + Unordered	Indistinct + Ordered	Indistinct + Unordered
Permutations	Combinations	Perms with Rep	No formula
$n!$	$\binom{n}{r}$	$\dfrac{n!}{n_1!n_2!\ldots n_m!}$	

What is a Probability

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

Equally Likely Outcomes

- Some sample spaces have equally likely outcomes
  - Coin flip: S = {Head, Tails}
  - Flipping two coins: S = {(H, H), (H, T), (T, H), (T, T)}
  - Roll of 6-sided die: S = {1, 2, 3, 4, 5, 6}
- P(Each outcome) = $\frac{1}{|S|}$
- In that case, P(E) = $\frac{\text{number of outcomes in E}}{\text{number of outcomes in S}} = \frac{|E|}{|S|}$

End Review

Not Everything is Equally Likely

- Say *n* balls are placed in *m* urns
  - Each ball is equally likely to be placed in any urn
- <u>Counts</u> of balls in urns are not equally likely!
  - Example: two balls (A and B) placed with equal likelihood in two urns (Urn 1 and Urn 2)
  - Possibilities:

Urn 1	Urn 2
A, B	-
A	B
B	A
-	A, B

Counts:

Urn 1	Urn 2	Prob
2	0	1/4
1	1	2/4
0	2	1/4

Dice – Our Misunderstood Friends

- Roll two 6-sided dice, yielding values D_1 and D_2
- Let E be event: $D_1 + D_2 = 4$
- What is P(E)?
  - $|S| = 36$, $E = \{(1, 3), (2, 2), (3, 1)\}$
  - $P(E) = 3/36 = 1/12$
- Let F be event: $D_1 = 2$
- P(E, given F already observed)?
  - $S = \{(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6)\}$
  - $E = \{(2, 2)\}$
  - P(E, given F already observed) = 1/6

Dice – Our Misunderstood Friends

- Two people each roll a die, yielding D_1 and D_2. You win if $D_1 + D_2 = 4$

- Socrative:
  - A. 1 and 3 tie for best
  - B. 1, 2 and 3 tie for best
  - C. 2 is the best
  - D. Other/none/more than one

Conditional Probability

- <u>**Conditional probability**</u> is probability that E occurs *given* that F has already occurred
  - "Conditioning on F"
- Written as P(E | F)
  - Means "P(E, given F already observed)"
  - Sample space, S, reduced to those elements consistent with F (i.e. $S \cap F$)
  - Event space, E, reduced to those elements consistent with F (i.e. $E \cap F$)
- With equally likely outcomes:

$$P(E \mid F) = \frac{\text{\# of outcomes in E consistent with F}}{\text{\# of outcomes in S consistent with F}} = \frac{|EF|}{|SF|} = \frac{|EF|}{|F|}$$

Conditional Probability

- General definition:

$$P(E \mid F) = \frac{P(EF)}{P(F)}$$

  where P(F) > 0

- Holds even when outcomes are not equally likely
- Implies: P(EF) = P(E | F) P(F) (chain rule)

- What if P(F) = 0?
  - P(E | F) undefined
  - *Congratulations! You observed the impossible!*

Generalized Chain Rule

- General definition of Chain Rule:

$$P(E_1 E_2 E_3 \ldots E_n)$$
$$= P(E_1)P(E_2 \mid E_1)P(E_3 \mid E_1 E_2)\ldots P(E_n \mid E_1 E_2 \ldots E_{n-1})$$

- Ross calls this the "multiplication rule"

- You can call it either (just be consistent)

Slicing Up Spam

In 2010 88% of email was spam

Slicing Up Spam

- 24 emails are sent 6 each to 4 users.
- 10 of the 24 emails are spam.
  - All possible outcomes equally likely
  - E = user 1 receives 3 spam emails
  - What is P(E)?

Choose 3 spam of the 10 spam

Choose 3 non-spam of the 14 non-spam

$$\frac{\binom{10}{3}\binom{14}{3}}{\binom{24}{6}} \approx 0.3245$$

Choose 6 emails of the 24 emails

Slicing Up Spam

- 24 emails are sent 6 each to 4 users.
- 10 of the 24 emails are spam.
  - All possible outcomes equally likely
  - E = user 1 receives 3 spam emails
  - F = user 2 receives 6 spam emails
  - What is P(E | F)?

Choose 3 spam of the remaining 4

Choose 3 non-spam of the 14

$$\frac{\binom{4}{3}\binom{14}{3}}{\binom{18}{6}} \approx 0.0784$$

Choose 6 email of the remaining 18

Slicing Up Spam

- 24 emails are sent 6 each to 4 users.
- 10 of the 24 emails are spam.
  - All possible outcomes equally likely
  - E = user 1 receives 3 spam emails
  - F = user 2 receives 6 spam emails
  - G = user 3 receives 5 spam emails
  - What is P(G | F)?

$$\frac{\boxed{\binom{4}{5}}\binom{14}{1}}{\binom{18}{6}} = 0$$

No way to choose 5 spam from 4 remaining spam emails!

Sending Bit Strings

- Bit string with *m* 0's and *n* 1's sent on network
  - All distinct arrangements of bits equally likely
  - E = first bit received is a 1
  - F = *k* of first *r* bits received are 1's

P(E|F)?

Sending Bit Strings

- Bit string with m 0's and n 1's sent on network
  - All distinct arrangements of bits equally likely
  - E = first bit received is a 1
  - F = k of first r bits received are 1's

P(E|F)?

Sending Bit Strings

- Bit string with *m* 0's and *n* 1's sent on network
  - All distinct arrangements of bits equally likely
  - E = first bit received is a 1
  - F = *k* of first *r* bits received are 1's

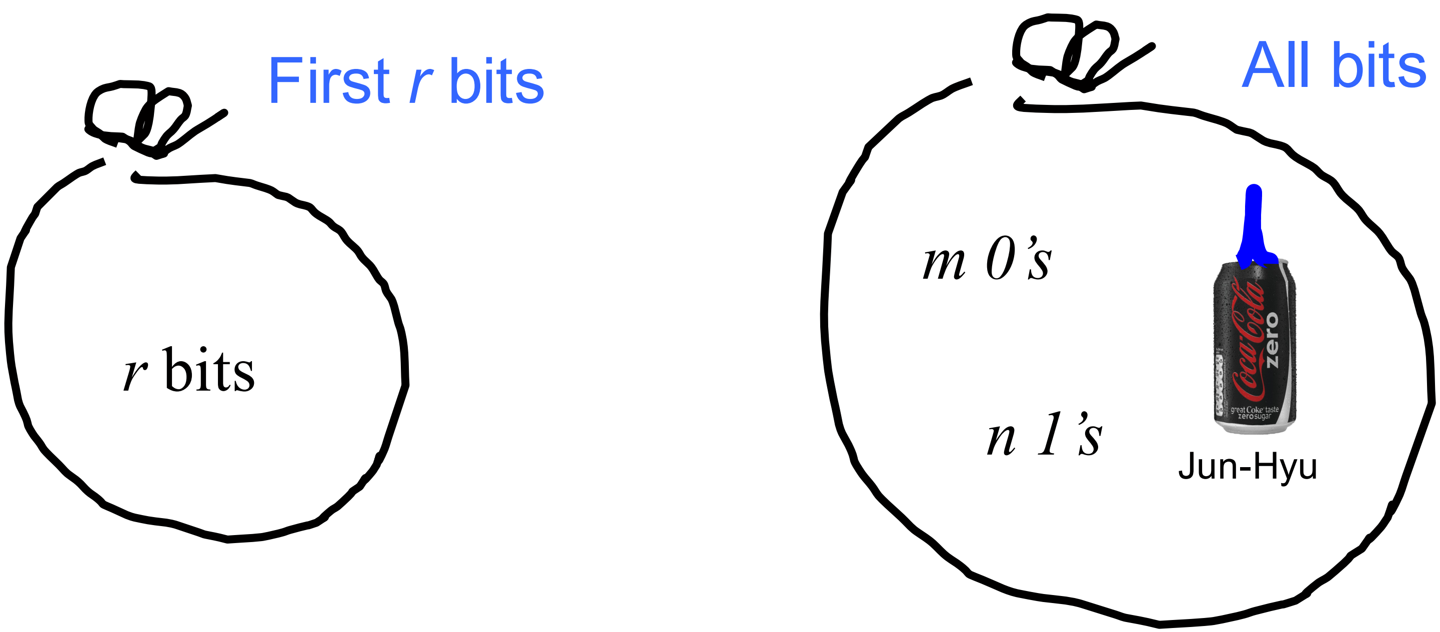
P(E|F)?

*Think of the bits as distinct so that all outcomes are equally likely

Sending Bit Strings

- Bit string with m 0's and n 1's sent on network
  - All distinct arrangements of bits equally likely
  - E = first bit received is a 1
  - F = k of first r bits received are 1's

P(E|F)?



Sending Bit Strings

- Bit string with *m* 0's and *n* 1's sent on network
  - All distinct arrangements of bits equally likely
  - E = first bit received is a 1
  - F = *k* of first *r* bits received are 1's

P(E|F)?

- Solution 1:

$$P(E \mid F) = \frac{P(EF)}{P(F)} = \frac{P(F \mid E)P(E)}{P(F)} = \boxed{\frac{k}{r}}$$

$$P(E) = \frac{n}{m+n}$$

$$P(F \mid E) = \frac{\binom{n-1}{k-1}\binom{m}{r-k}}{\binom{m+n-1}{r-1}}$$

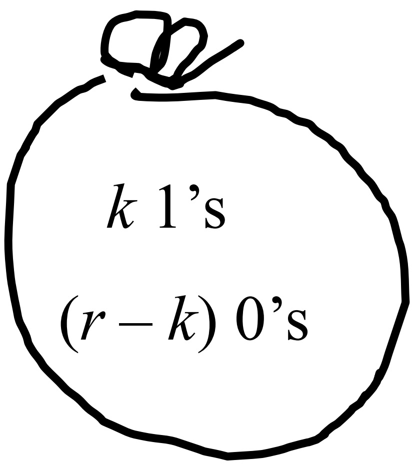
$$P(F) = \frac{\binom{n}{k}\binom{m}{r-k}}{\binom{m+n}{r}}$$

Sending Bit Strings

- Bit string with m 0's and n 1's sent on network
  - All distinct arrangements of bits equally likely
  - E = first bit received is a 1
  - F = k of first r bits received are 1's

P(E|F)?

- Solution 2:
  - Realize P(E | F) = P(picking one of k 1's out of r bits)
  - $P(E \mid F) = \frac{k}{r}$
  - Rock on!



First r bits

Card Piles

- Deck of 52 cards randomly divided into 4 piles
  - 13 cards per pile
  - Compute P(each pile contains exactly one ace)
- Solution:
  - E_1 = {Ace Spades (AS) in any one pile}
  - E_2 = {AS and Ace Hearts (AH) in different piles}
  - E_3 = {AS, AH, Ace Diamonds (AD) in different piles}
  - E_4 = {All 4 aces in different piles}
  - Compute $P(E_1 E_2 E_3 E_4)$

$$= P(E_1)\, P(E_2 \mid E_1)\, P(E_3 \mid E_1 E_2)\, P(E_4 \mid E_1 E_2 E_3)$$

Card Piles

E_1 = {Ace Spades (AS) in any one pile}

E_2 = {AS and Ace Hearts (AH) in different piles}

E_3 = {AS, AH, Ace Diamonds (AD) in different piles}

E_4 = {All 4 aces in different piles}

$P(E_1) = 1$

$P(E_2 \mid E_1) = 39/51$ (39 cards not in AS pile)

$P(E_3 \mid E_1 E_2) = 26/50$ (26 cards not in AS or AH piles)

$P(E_4 \mid E_1 E_2 E_3) = 13/49$ (13 cards not in AS, AH, AD piles)

$$P(E_1 E_2 E_3 E_4) = \frac{39 \cdot 26 \cdot 13}{51 \cdot 50 \cdot 49} \approx 0.105$$

NETFLIX

And Learn

Netflix and Learn

What is the probability that a user will like Life is Beautiful?

$P(E)$

$S = \{\text{Like}, \text{Not Like}\}$

$E = \{\text{Like}\}$

$P(E) = \frac{1}{2}$?

Netflix and Learn

What is the probability that a user will like Life is Beautiful?

$P(E)$

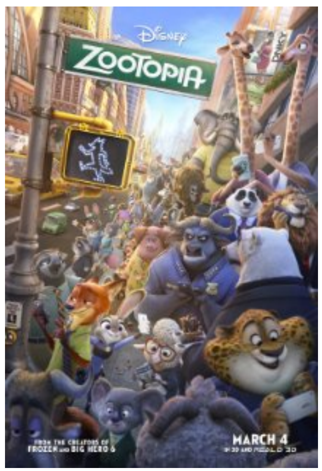
$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

$P(E) = 50{,}234{,}231 \,/\, 50{,}923{,}123 = 0.97$

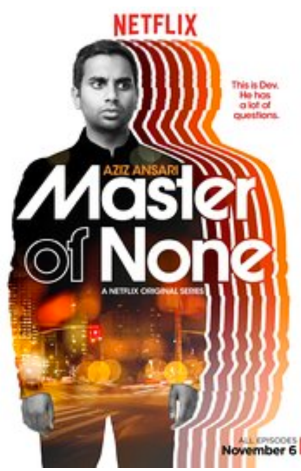
Netflix and Learn

Let E be the event that a user liked the given movie:

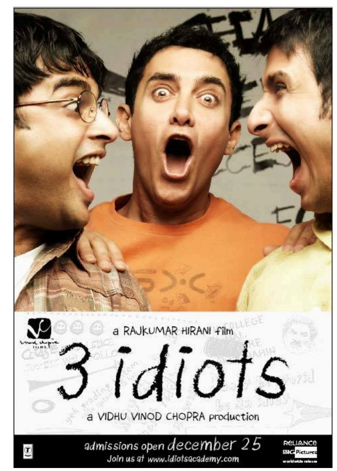
P(E) = 0.89



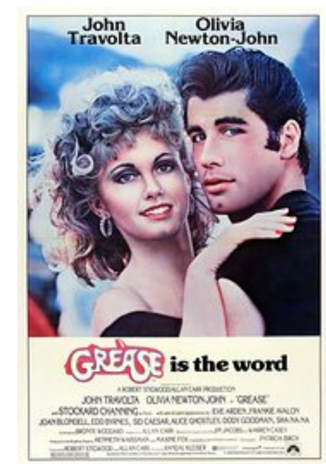
P(E) = 0.95



P(E) = 0.89



P(E) = 0.92



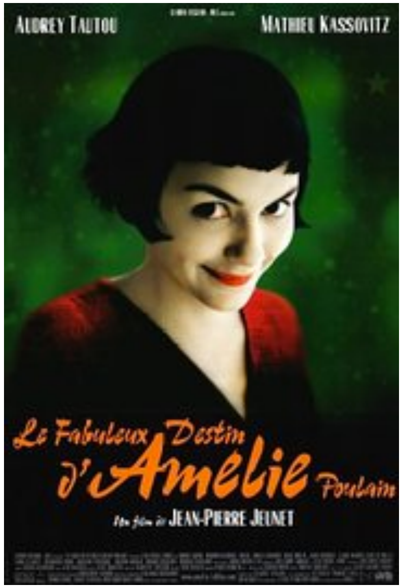
P(E) = 0.88

** These are the actual estimates*

Netflix and Learn

What is the probability that a user will like Life is Beautiful, given they liked Amelie?

$$P(E|F)$$



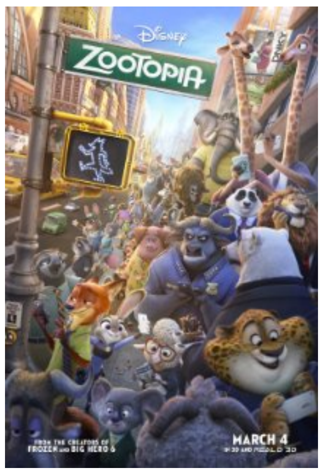
$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\frac{\text{people who liked both}}{\text{people who watched both}}}{\frac{\text{people who liked amelie}}{\text{people who watched amelie}}}$$

$$P(E|F) = 0.99$$

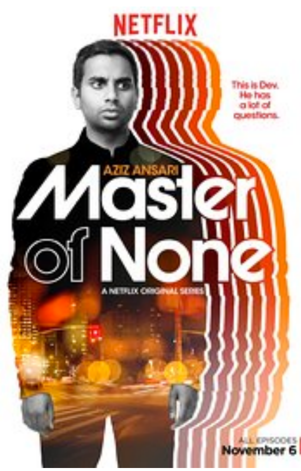
Netflix and Learn

Let E be the event that a user liked the given movie,
Let F be the event that the same user liked Amelie:

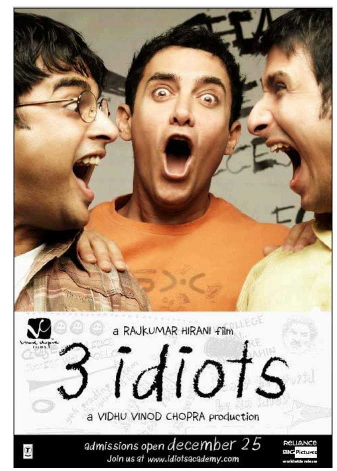
$P(E|F) =$ 0.82



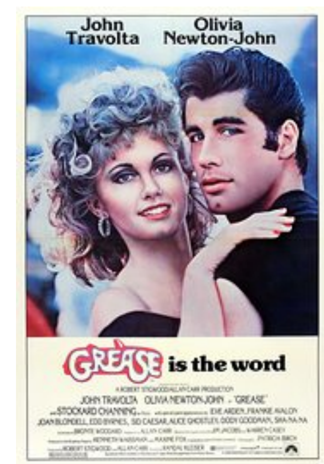
$P(E|F) =$ 0.96



$P(E|F) =$ 0.93



$P(E|F) =$ 0.99



$P(E|F) =$ 0.90

Machine Learning

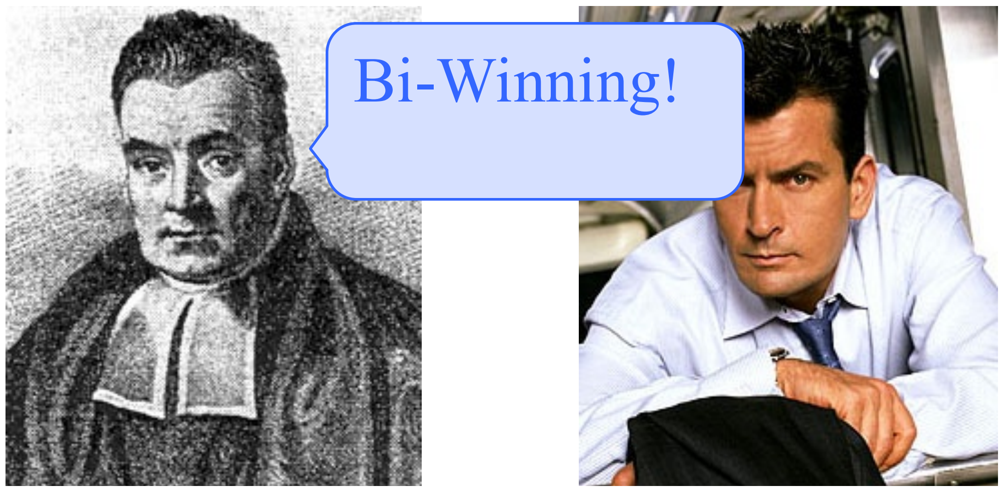
Machine Learning is:
Probability + Data + Computers

Conditioned on liking a set of movies?

Coming soon to a lecture hall near you…

Thomas Bayes

- Rev. Thomas Bayes (1702 –1761) was a British mathematician and Presbyterian minister

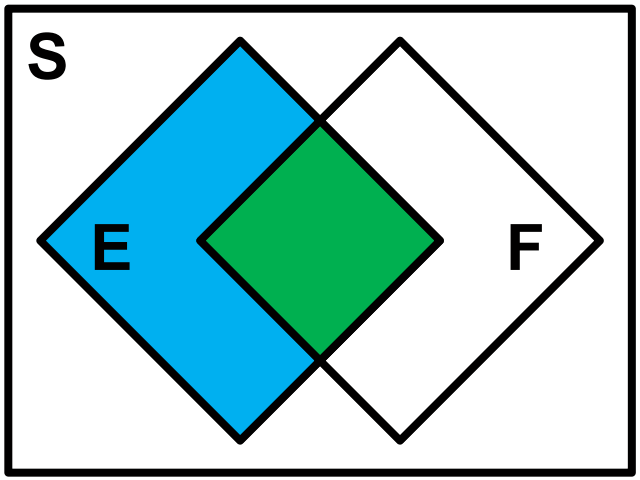


- He looked remarkably similar to Charlie Sheen
  - But that’s not important right now...

Background for Bayes Theorem

- Say E and F are events in S

$$E = EF \cup EF^c$$



Note: $EF \cap EF^c = \emptyset$

So, $P(E) = P(EF) + P(EF^c)$

Bayes Theorem

- Ross's form:

$$
\begin{aligned}
P(E) &= P(EF) + P(EF^c) \\
&= P(E \mid F)\, P(F) + P(E \mid F^c)\, P(F^c)
\end{aligned}
$$

- Most common form:

$$
\begin{aligned}
P(F \mid E) &= P(EF) / P(E) \\
&= [P(E \mid F)\, P(F)] / P(E)
\end{aligned}
$$

- Expanded form:

$$
P(F \mid E) = \frac{P(E \mid F)\, P(F)}{P(E \mid F)\, P(F) + P(E \mid F^c)\, P(F^c)}
$$

HIV Testing

- A test is 98% effective at detecting HIV
  - However, test has a “false positive” rate of 1%
  - 0.5% of US population has HIV
  - Let E = you test positive for HIV with this test
  - Let F = you actually have HIV
  - What is P(F | E)?

HIV Testing

- A test is 98% effective at detecting HIV
  - However, test has a “false positive” rate of 1%
  - 0.5% of US population has HIV
  - Let E = you test positive for HIV with this test
  - Let F = you actually have HIV
  - What is P(F | E)?

First, just do a “gut reaction” prediction of how this math is going to work out:

A. $P(F|E)$ **≤ 1/2**
B. **1/2 <** $P(F|E)$ **≤ 3/4**
C. **3/4 <** $P(F|E)$

STUDENT

HIV Testing

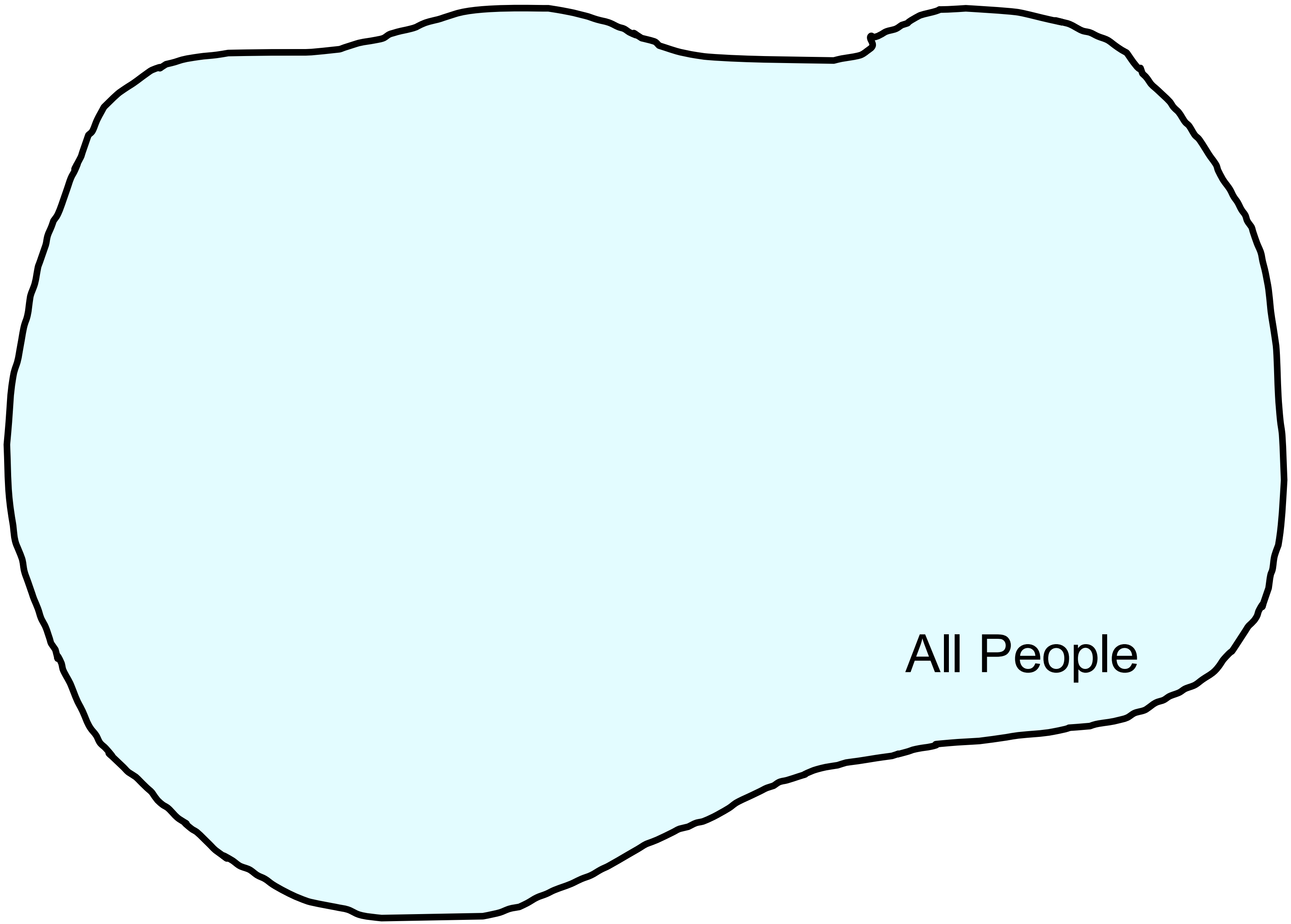
- A test is 98% effective at detecting HIV
  - However, test has a “false positive” rate of 1%
  - 0.5% of US population has HIV
  - Let E = you test positive for HIV with this test
  - Let F = you actually have HIV
  - What is P(F | E)?
- Solution:

$$P(F \mid E) = \frac{P(E \mid F)\, P(F)}{P(E \mid F)\, P(F) + P(E \mid F^c)\, P(F^c)}$$

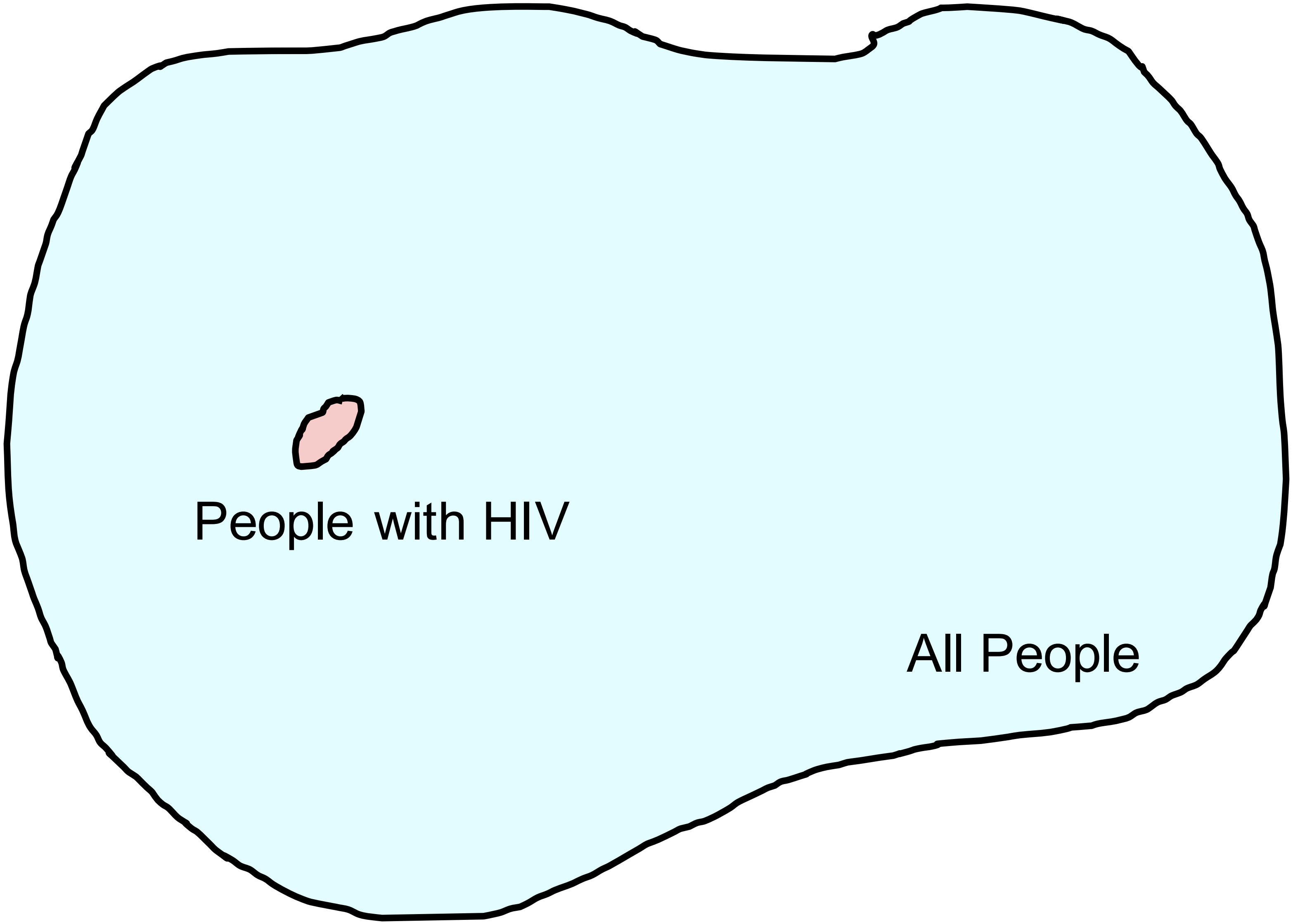
$$P(F \mid E) = \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.01)(1 - 0.005)} \approx 0.330$$

Intuition Time

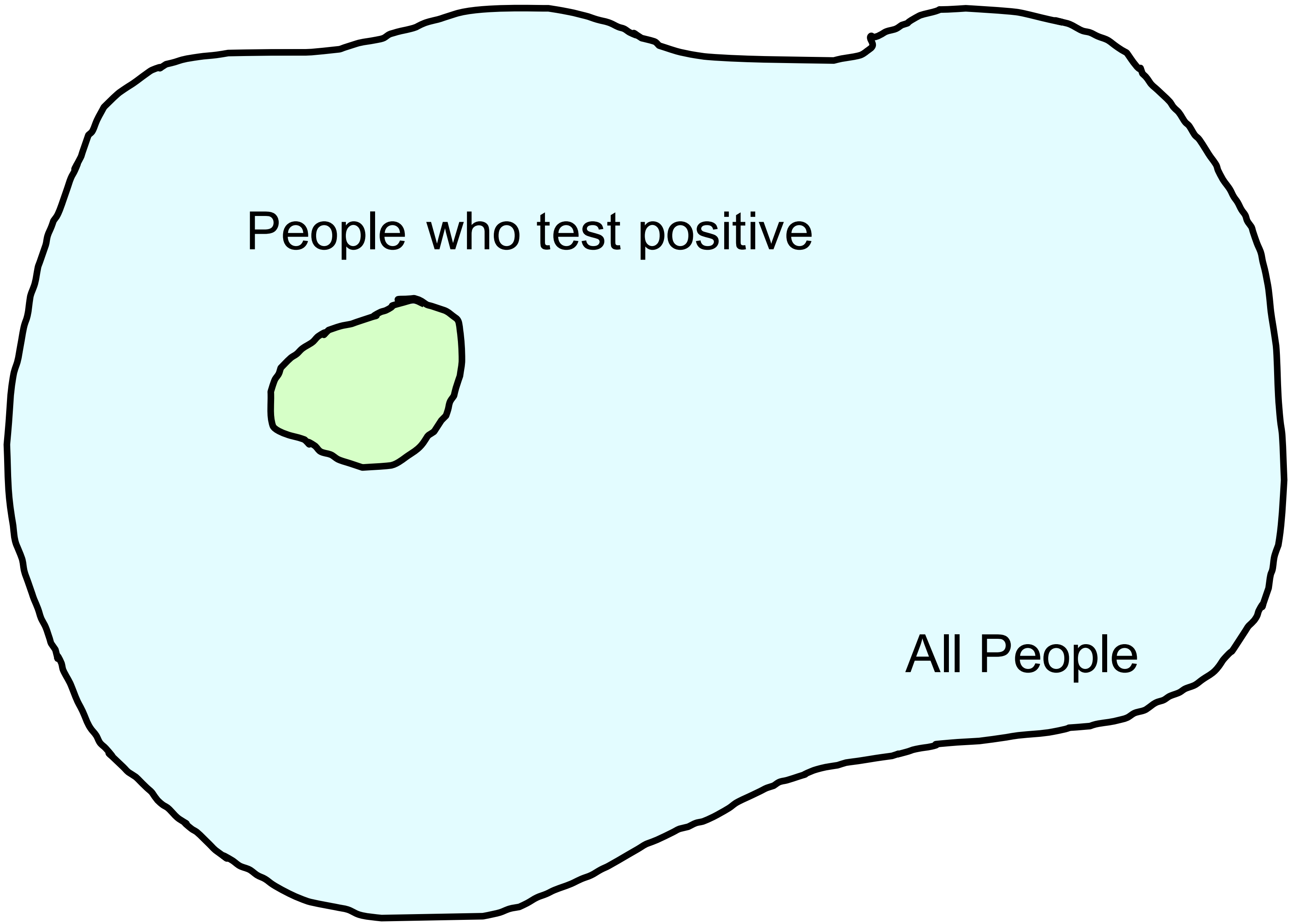
Bayes Theorem Intuition



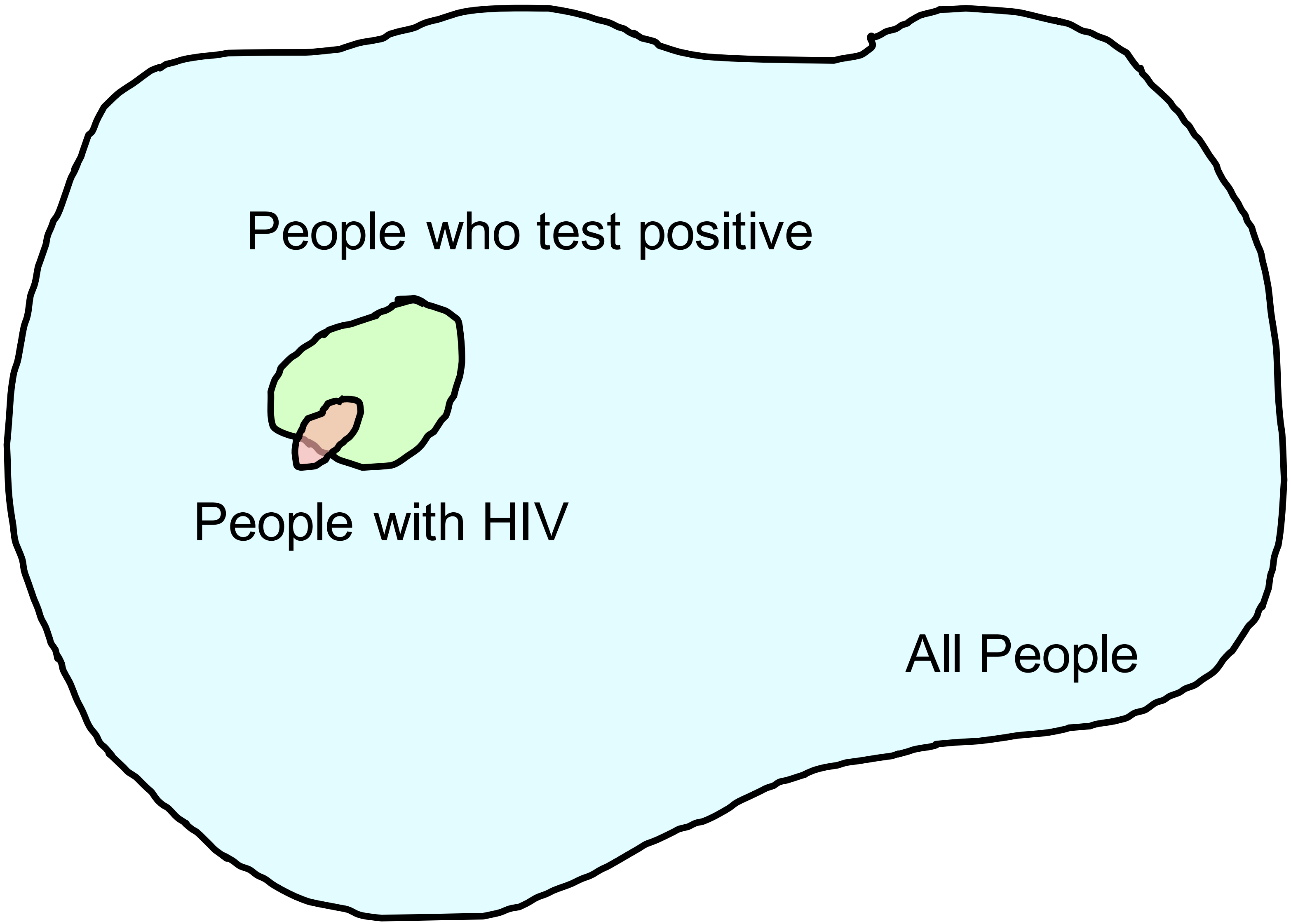
Bayes Theorem Intuition



Bayes Theorem Intuition

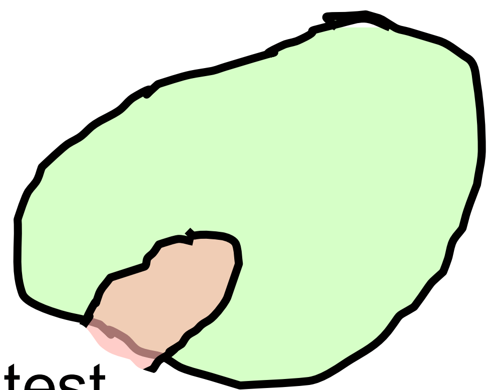


Bayes Theorem Intuition



Bayes Theorem Intuition

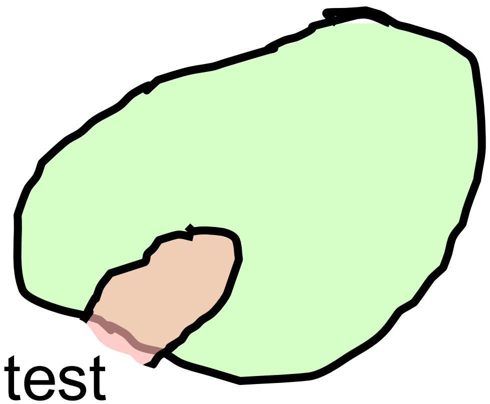
Conditioning on a positive result changes the sample space to this:



≈ 0.330

Bayes Theorem Intuition

Conditioning on a positive result changes the sample space to this:



People who test positive

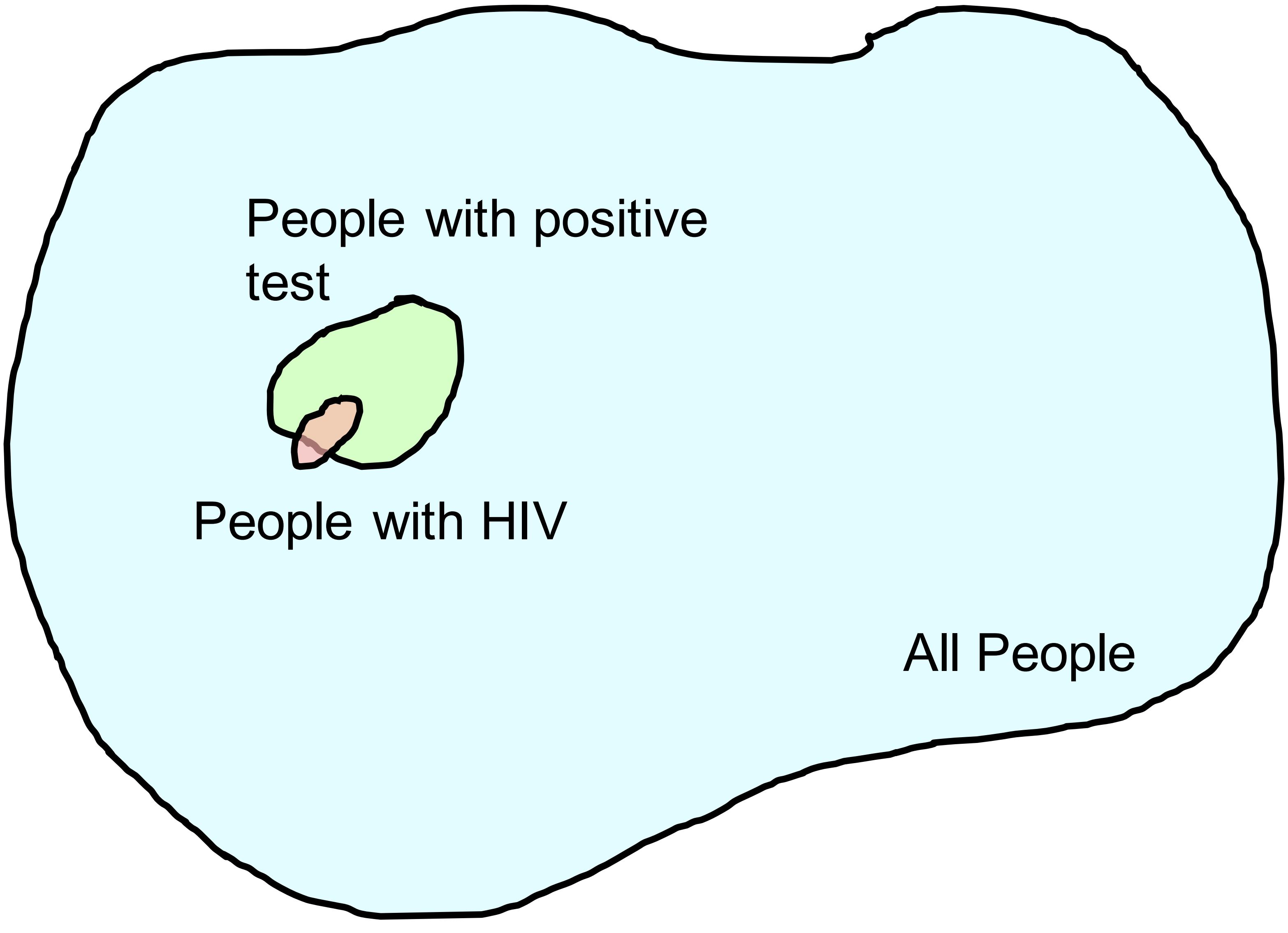
$$P(F)P(E|F) + P(F^c)P(E|F^c)$$

People who test positive and have HIV

$$P(F)P(E|F)$$

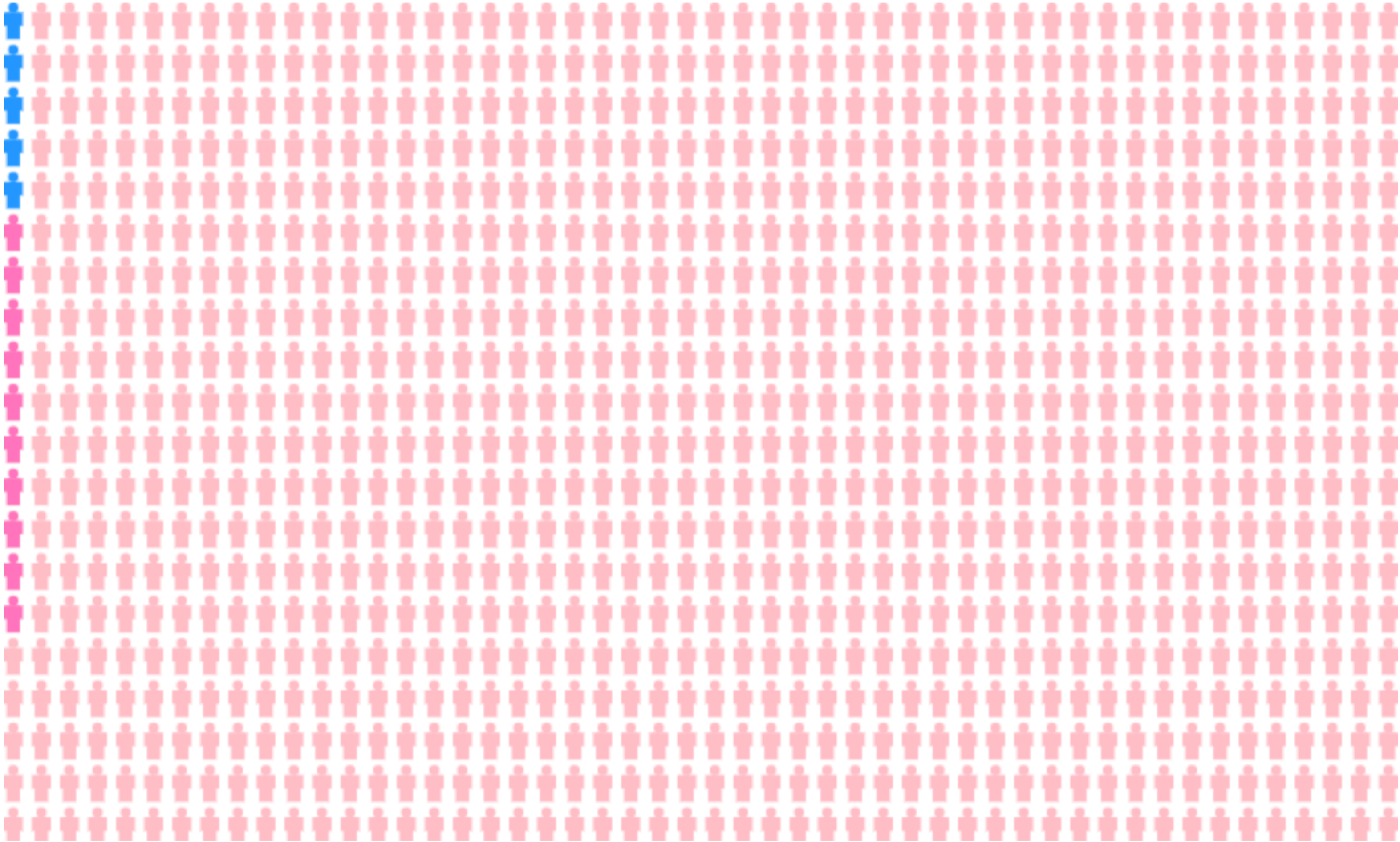
≈ 0.330

Bayes Theorem Intuition



Bayes Theorem Intuition

Say we have 1000 people:



5 have HIV and test positive, 985 **do not** have HIV and test negative.
10 **do not** have HIV and test positive.

≈ 0.333

Why It's Still Good to get Tested

	HIV +	HIV –
Test +	$0.98 = P(E \mid F)$	$0.01 = P(E \mid F^c)$
Test –	$0.02 = P(E^c \mid F)$	$0.99 = P(E^c \mid F^c)$

- Let E^c = you test <u>negative</u> for HIV with this test
- Let F = you actually have HIV
- What is $P(F \mid E^c)$?

$$P(F \mid E^c) = \frac{P(E^c \mid F)\, P(F)}{P(E^c \mid F)\, P(F) + P(E^c \mid F^c)\, P(F^c)}$$

$$P(F \mid E^c) = \frac{(0.02)(0.005)}{(0.02)(0.005) + (0.99)(1 - 0.005)} \approx 0.0001$$

Simple Spam Detection

- Say 60% of all email is spam
  - 90% of spam has a forged header
  - 20% of non-spam has a forged header
  - Let E = message contains a forged header
  - Let F = message is spam
  - What is P(F | E)?
- Solution:

$$P(F \mid E) = \frac{P(E \mid F)\,P(F)}{P(E \mid F)\,P(F) + P(E \mid F^c)\,P(F^c)}$$

$$P(F \mid E) = \frac{(0.9)(0.6)}{(0.9)(0.6) + (0.2)(0.4)} \approx 0.871$$

Monty Hall

Let's Make a Deal

- Game show with 3 doors: A, B, and C

  - Behind one door is prize (equally likely to be any door)
  - Behind other two doors is nothing
  - We choose a door
  - Then host opens 1 of other 2 doors, revealing nothing
  - We are given option to change to other door
- Should we?
  - Note: If we don't switch, P(win) = 1/3 (random)

Let's Make a Deal

- Without loss of generality, say we pick A
  - P(A is winner) = 1/3
    - Host opens either B or C, we <u>always lose</u> by switching
    - P(win | A is winner, picked A, switched) = 0
  - P(B is winner) = 1/3
    - Host <u>must</u> open C (can't open A and can't reveal prize in B)
    - So, by switching, we switch to B and <u>always win</u>
    - P(win | B is winner, picked A, switched) = 1
  - P(C is winner) = 1/3
    - Host <u>must</u> open B (can't open A and can't reveal prize in C)
    - So, by switching, we switch to C and <u>always win</u>
    - P(win | C is winner, picked A, switched) = 1
  - Should always switch!
    - P(win | picked A, switched) = (1/3*0) + (1/3*1) + (1/3*1) = 2/3

Slight Variant to Clarify

- Start with 1,000 envelopes, of which 1 is winner
  - You get to choose 1 envelope
    - Probability of choosing winner = 1/1000
  - Consider remaining 999 envelopes
    - Probability one of them is the winner = 999/1000
  - I open 998 of remaining 999 (showing they are empty)
    - Probability the last remaining envelope being winner = 999/1000
  - Should you switch?
    - Probability winning without switch = $\frac{1}{\text{original \# envelopes}}$
    - Probability winning with switch = $\frac{\text{original \# envelopes - 1}}{\text{original \# envelopes}}$