



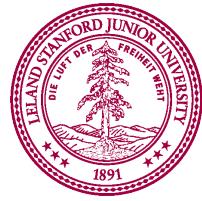
Conditional Probability

What is a probability?

Number between 0 and 1

What is a Probability

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



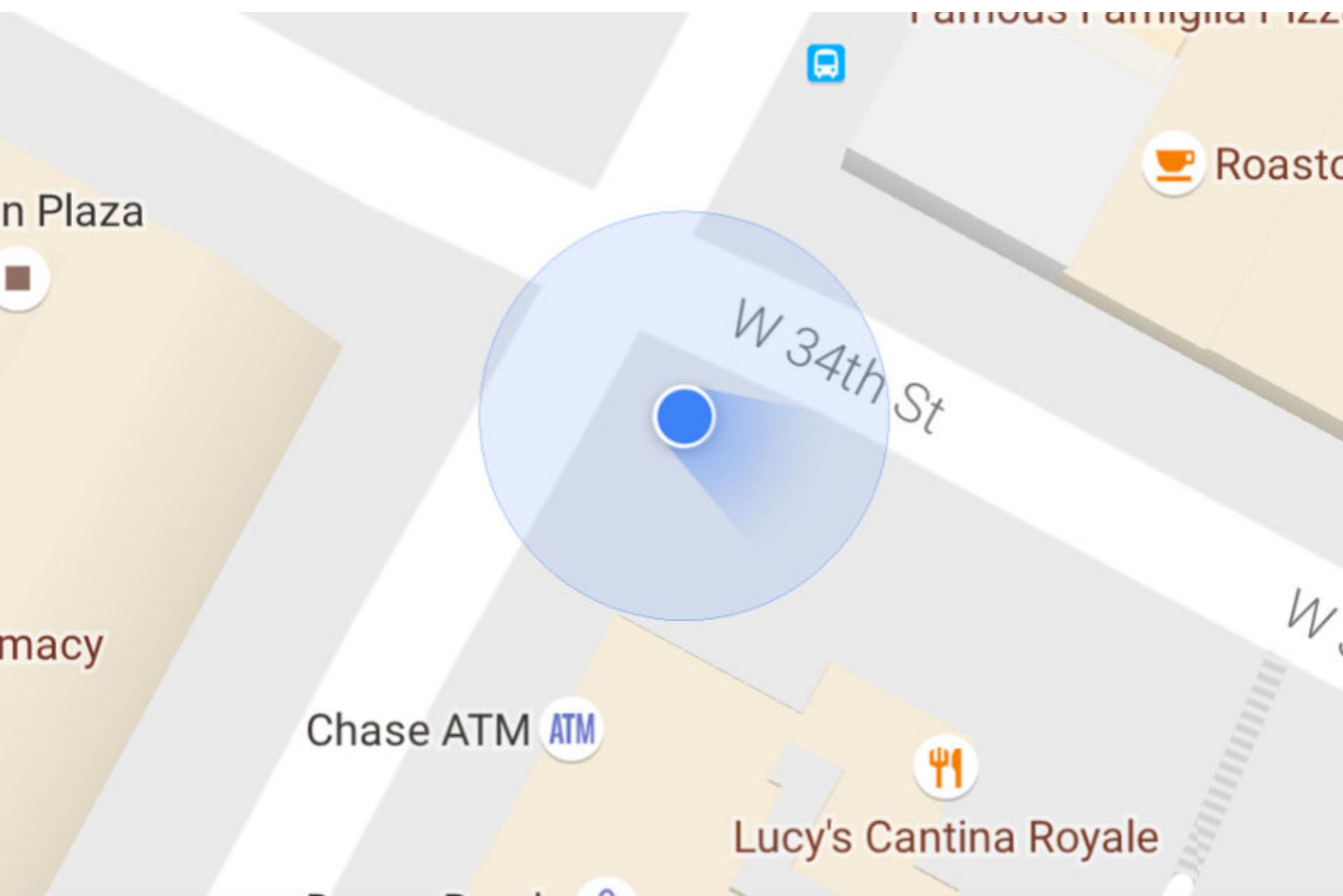
Two Sources of Probabilities

1. Randomness in the world

2. Our own uncertainties

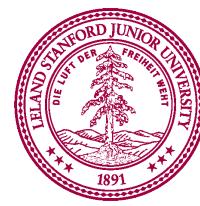


Probabilities in The Real World



Equally Likely Outcomes

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



Birthdays

- What is the probability that of n people, none share the same birthday (regardless of year)?
 - $|S| = (365)^n$
 - $|E| = (365)(364)\dots(365 - n + 1)$
 - $P(\text{no matching birthdays})$
 $= (365)(364)\dots(365 - n + 1)/(365)^n$
- Interesting values of n
 - $n = 23$: $P(\text{no matching birthdays}) < \frac{1}{2}$ (least such n)
 - $n = 75$: $P(\text{no matching birthdays}) < 1/3,000$
 - $n = 100$: $P(\text{no matching birthdays}) < 1/3,000,000$
 - $n = 150$:

$P(\text{no matching birthdays}) < 1/3,000,000,000,000,000$



Birthdays

- What is the probability that of n other people, none of them share the same birthday as you?
 - $|S| = (365)^n$
 - $|E| = (364)^n$
 - $P(\text{no birthdays matching yours}) = (364)^n / (365)^n$
- Interesting values of n
 - $n = 23$: $P(\text{no matching birthdays}) \approx 0.9388$
 - $n = 150$: $P(\text{no matching birthdays}) \approx 0.6626$
 - $n = 253$: $P(\text{no matching birthdays}) \approx 0.4995$
 - Least such n for which $P(\text{no matching birthdays}) < \frac{1}{2}$
 - Why are these probabilities much higher than before?
 - Anyone born on April 25th?
 - Is today anyone's birthday?



Not Everything is Equally Likely

- Say n balls are placed in m urns
 - Each ball is equally likely to be placed in any urn
- Counts 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: Counts:

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

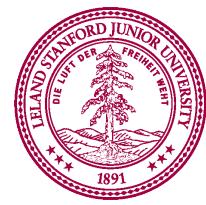
Urn 1	Urn 2	Prob
2	0	1/4
1	1	2/4
0	2	1/4



End Review

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: $D_1 = 2$
- $P(E, \text{ given } F \text{ already observed})?$
 - $S = \{(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6)\}$
 - $E = \{(2, 2)\}$
 - $P(E, \text{ given } F \text{ already observed}) = 1/6$



Dice – Our Misunderstood Friends

- Two people each roll a die, yielding D_1 and D_2 .
You win if $D_1 + D_2 = 4$
- Q: What do you think is the best outcome for D_1 ?
- 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

- **Conditional probability** 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. $E \cap F$)



Conditional Probability

With equally likely outcomes:

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



Conditional Probability

- General definition:

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

- 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:

$$P(E_1 E_2 E_3 \dots E_n)$$

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

- Ross calls this the “multiplication rule”
- You can call it either (just be consistent)



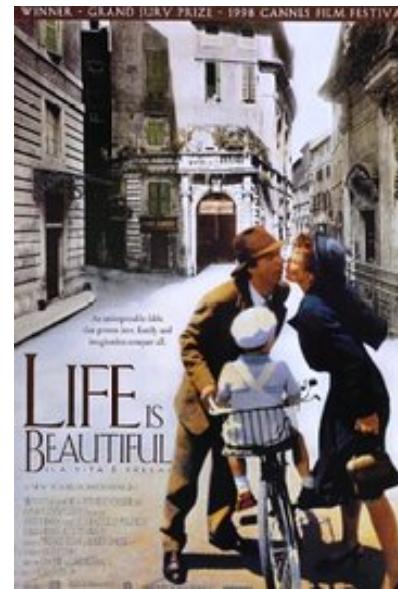
NETFLIX

And Learn

Netflix and Learn

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

$$P(E)$$



S = {Watch, Not Watch}

E = {Watch}

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



Netflix and Learn

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

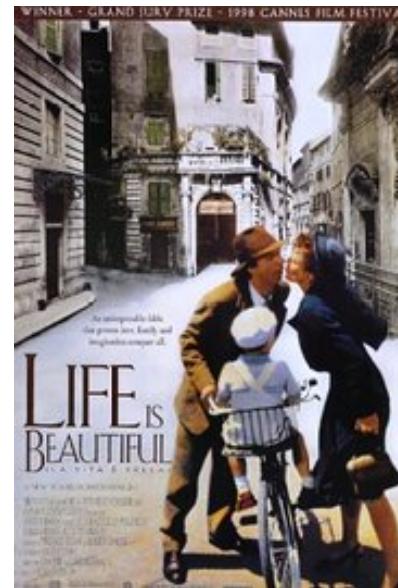
$$P(E)$$



Netflix and Learn

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

$$P(E)$$

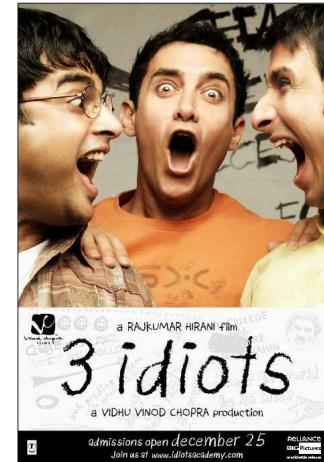
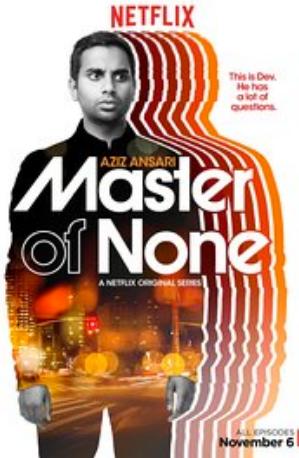


$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n} \approx \frac{\text{\#people who watched movie}}{\text{\#people on Netflix}}$$

$$P(E) = 10,234,231 / 50,923,123 = 0.20$$

Netflix and Learn

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



$$P(E) = \\ 0.19$$

$$P(E) = \\ 0.32$$

$$P(E) = \\ 0.20$$

$$P(E) = \\ 0.09$$

$$P(E) = \\ 0.23$$

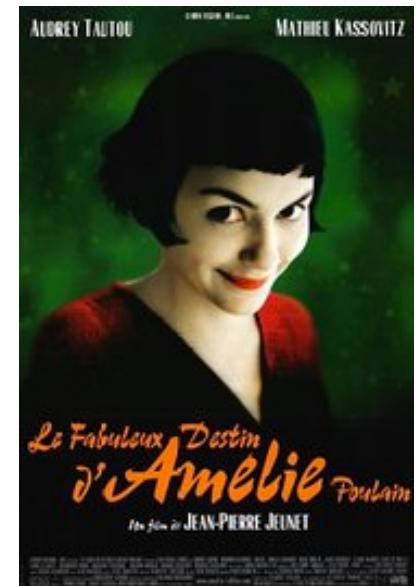
* These are the actual estimates
Piech, CS106A, Stanford University



Netflix and Learn

What is the probability
that a user will watch
Life is Beautiful, given
they watched Amelie?

$$P(E|F)$$



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

Netflix and Learn

What is the probability
that a user will watch
Life is Beautiful, given
they watched Amelie?

$$P(E|F)$$

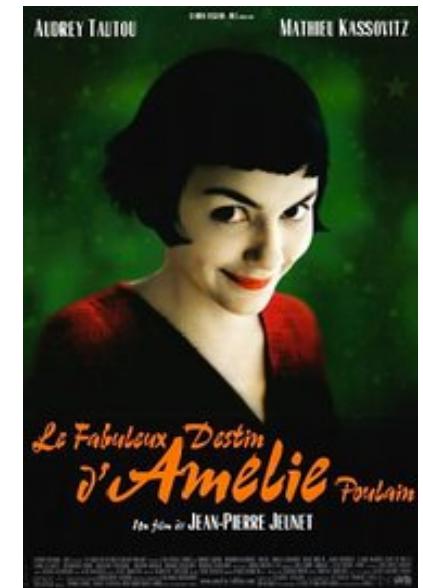
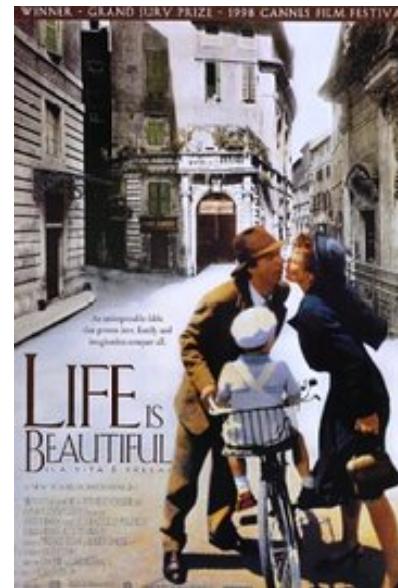


$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\frac{\# \text{people who watched both}}{\# \text{people on Netflix}}}{\frac{\# \text{people who watched } F}{\# \text{people on Netflix}}}$$

Netflix and Learn

What is the probability
that a user will watch
Life is Beautiful, given
they watched Amelie?

$$P(E|F)$$



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

$$P(E|F) = 0.42$$

Piech, CS106A, Stanford University



Netflix and Learn

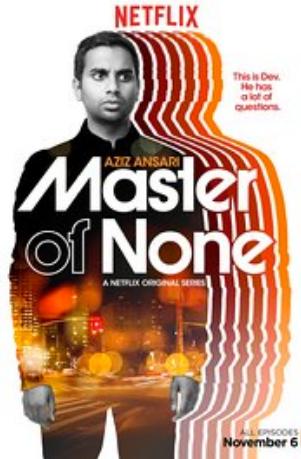
Let E be the event that a user watched the given movie,
Let F be the event that the same user watched Amelie:



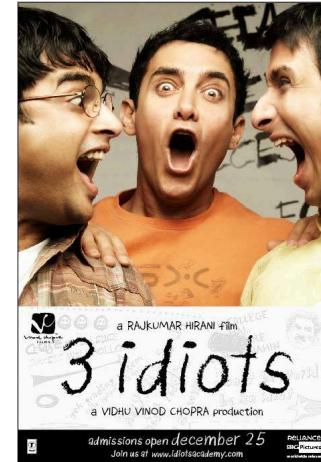
$$P(E|F) = \\ 0.14$$



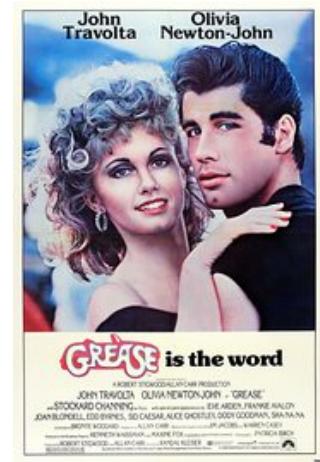
$$P(E|F) = 0.35$$



$$P(E|F) = \\ 0.20$$



$$P(E|F) = \\ 0.72$$



$$P(E|F) = \\ 0.49$$



Machine Learning

Machine Learning is:
Probability + Data + Computers



Conditioned on liking a genre?

Coming soon ...

Sophomores

- There are 230 students in CS109:
 - Probability that a random student in CS109 is a Sophomore is 0.28
 - We can observe the probability that a student is both a Sophomore and is in class
 - What is the conditional probability of a student coming to class given that they are a Sophomore?
- Solution:
 - S is the event that a student is a sophomore
 - A is the event that a student is in class

$$P(A|S) = \frac{P(SA)}{P(S)}$$



Card Piles

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



Card Piles

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

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

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

$E_4 = \{\text{All 4 aces in different piles}\}$

$$P(E_1) = 1$$

$$P(E_2 | E_1) = 39/51 \quad (\text{39 cards not in AS pile})$$

$$P(E_3 | E_1 E_2) = 26/50 \quad (\text{26 cards not in AS or AH piles})$$

$$P(E_4 | E_1 E_2 E_3) = 13/49 \quad (\text{13 cards not in AS, AH, AD piles})$$

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



Sending Bit Strings

- Bit string with m 0's and n 1's sent on network
 - All distinct arrangements of bits equally likely
 - $E = \text{first bit received is a } 1$
 - $F = k \text{ of first } r \text{ bits received are } 1\text{'s}$ $P(E|F)?$



Sending Bit Strings

- Bit string with m 0's and n 1's sent on network
 - All distinct arrangements of bits equally likely
 - $E = \text{first bit received is a } 1$
 - $F = k \text{ of first } r \text{ bits received are } 1\text{'s}$
- $P(E|F)?$



*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
- $P(E|F)?$

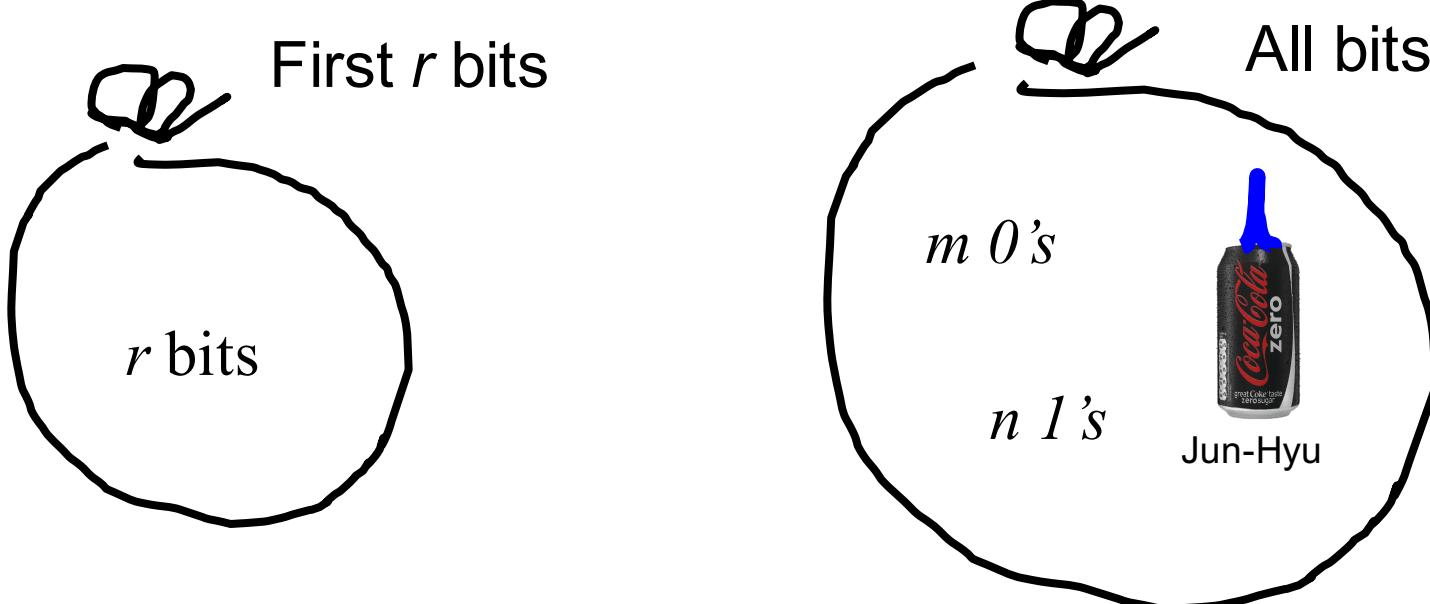


*Think of the bits as distinct so that all outcomes are equally likely

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:

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

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

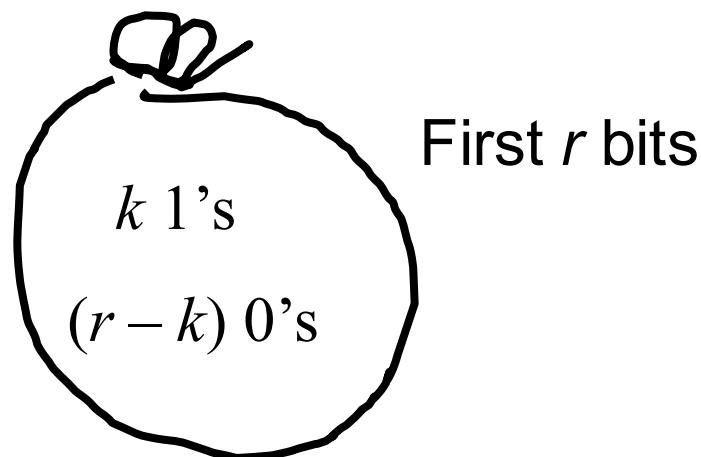
$$P(E) = \frac{n}{m+n}$$

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



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



Thomas Bayes

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



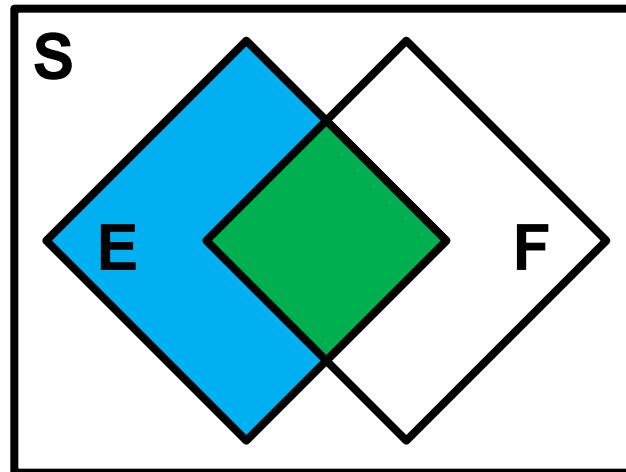
- He looked remarkably similar to Charlie Sheen
 - But that's not important right now...



Background for Bayes Theorem

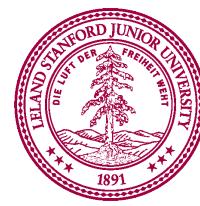
- Say E and F are events in S

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



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

$$\text{So, } P(E) = P(EF) + P(EF^c)$$



Bayes Theorem

- Ross's form:

$$\begin{aligned} P(E) &= P(EF) + P(EF^C) \\ &= P(E|F)P(F) + P(E|F^C)P(F^C) \end{aligned}$$

- Most common form:

$$\begin{aligned} P(F|E) &= \frac{P(EF)}{P(E)} \\ &= \frac{P(E|F)P(F)}{P(E)} \end{aligned}$$

- Expanded form:

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$



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 $P(F | E)$?
- Solution:

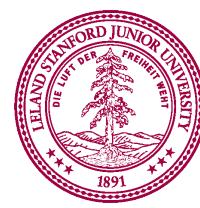


HIV Testing

- 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:

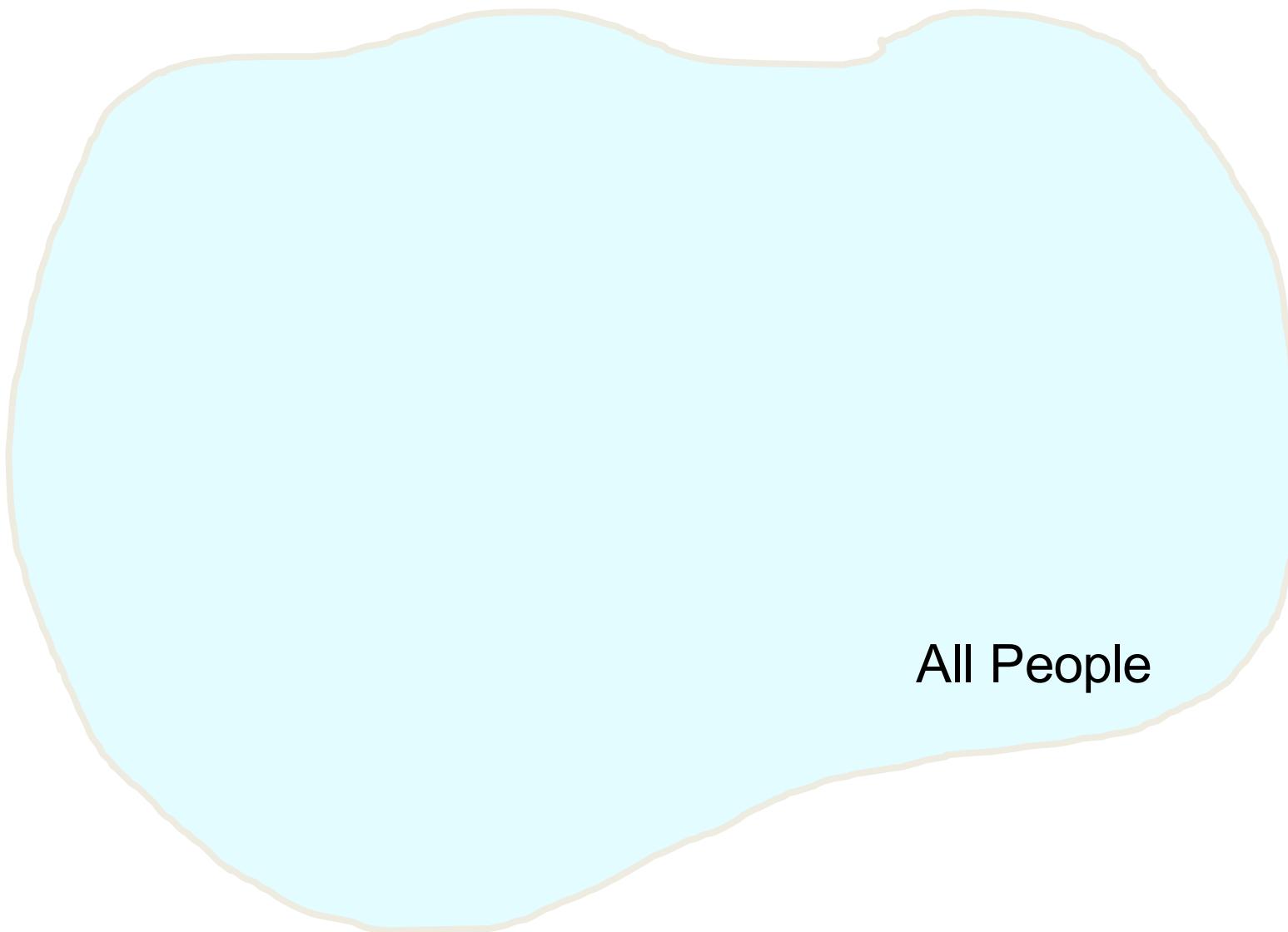
$$P(F | E) = \frac{P(E | F) P(F)}{P(E | F) P(F) + P(E | F^c) P(F^c)}$$

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



Intuition Time

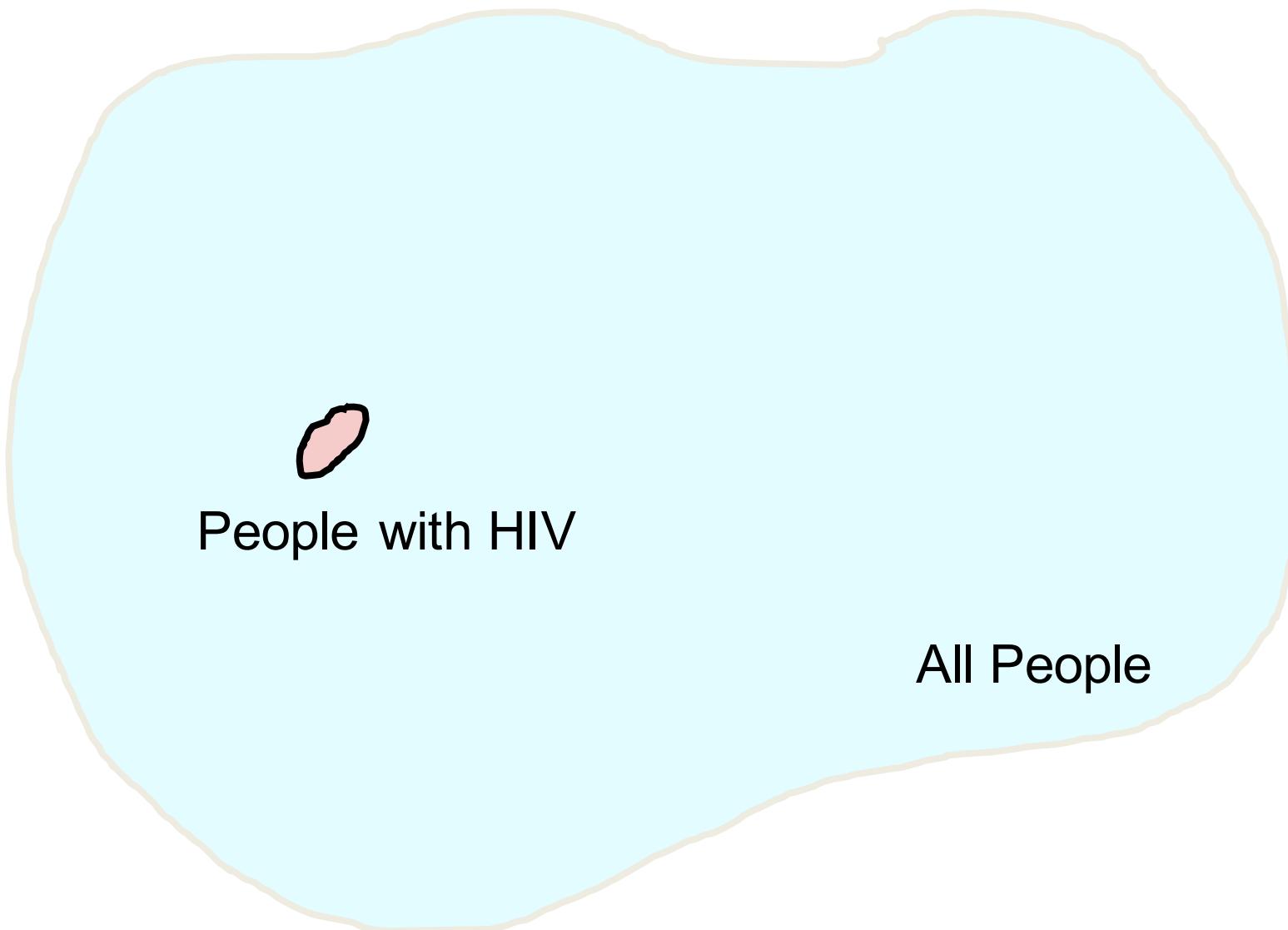
Bayes Theorem Intuition



All People

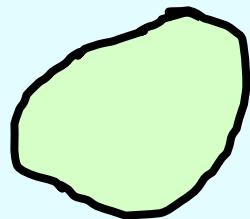


Bayes Theorem Intuition



Bayes Theorem Intuition

People who test positive

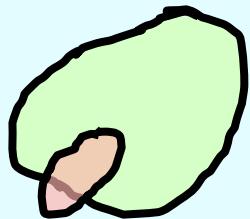


All People



Bayes Theorem Intuition

People who test positive



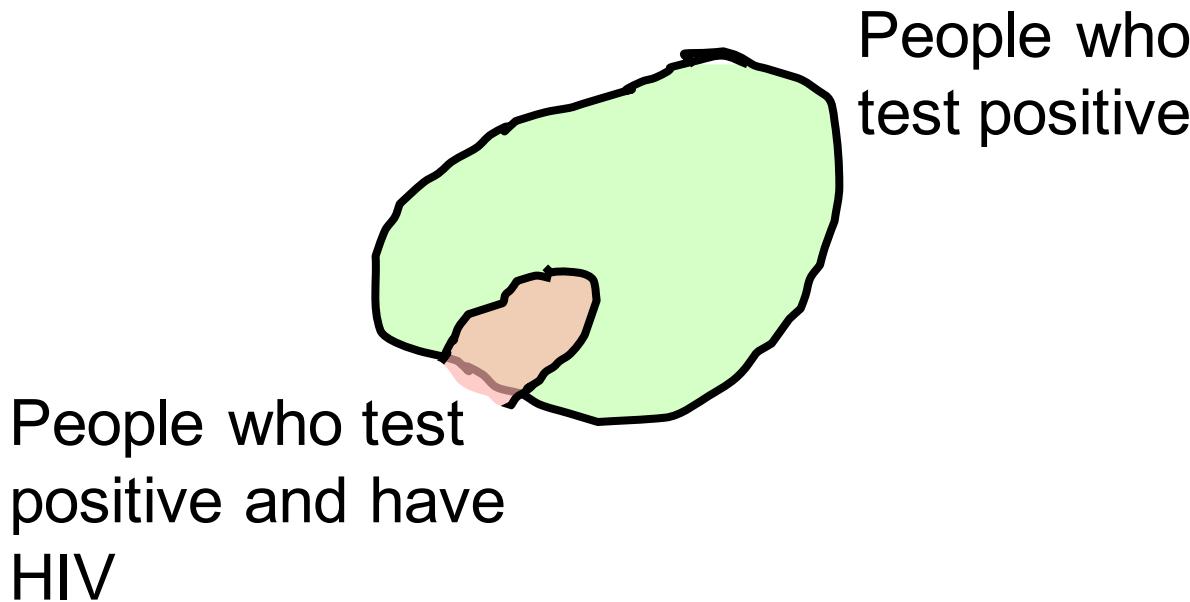
People with HIV

All People



Bayes Theorem Intuition

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

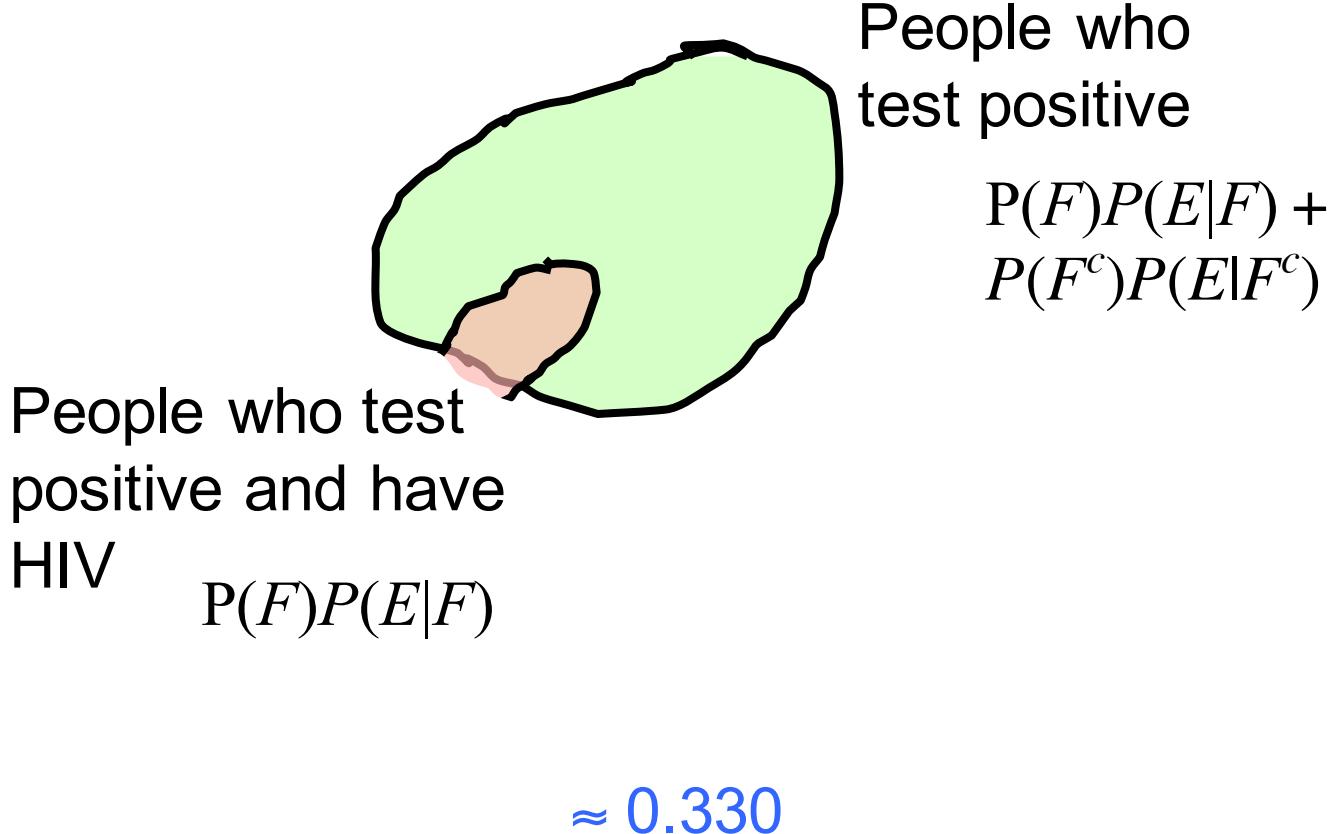


≈ 0.330

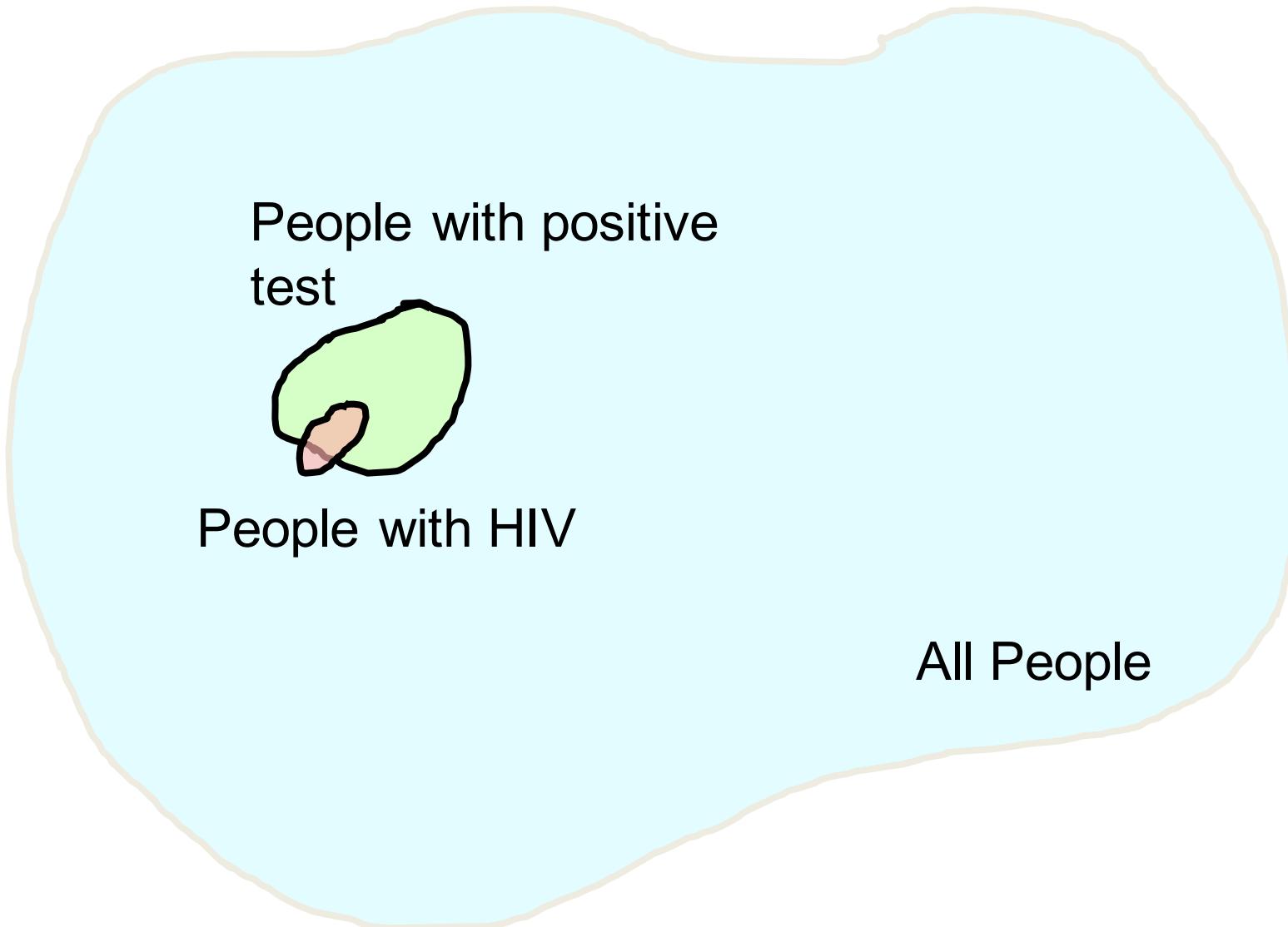


Bayes Theorem Intuition

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

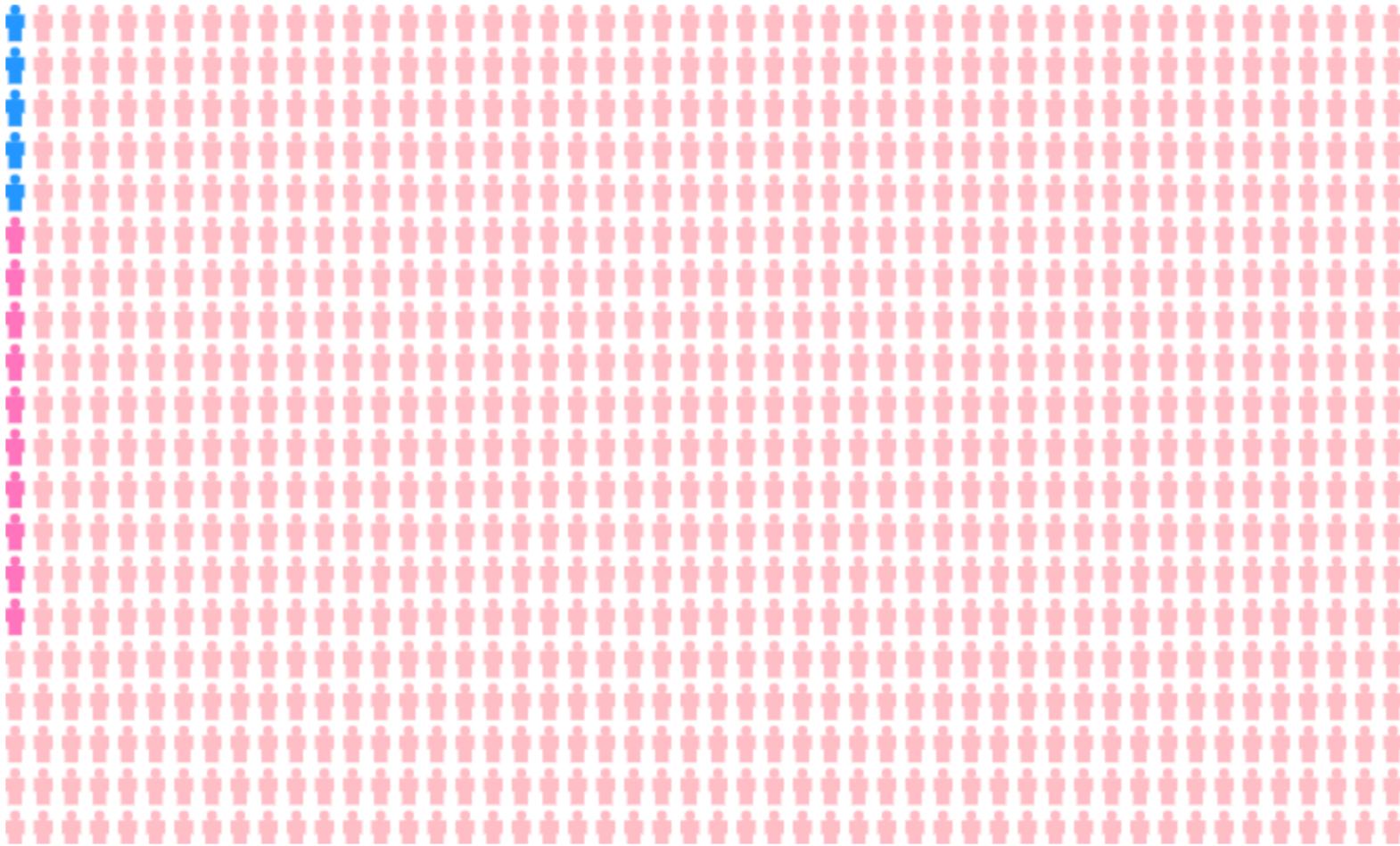


Bayes Theorem Intuition



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

Piech, CS106A, Stanford University



$$\approx 0.333$$

Why It's Still Good to get Tested

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

- Let E^c = you test negative for HIV with this test
- Let F = you actually have HIV
- What is $P(F | E^c)$?

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

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



Slicing Up Spam



In 2010 88% of email was spam

Piech, CS106A, Stanford University



Simple Spam Detection

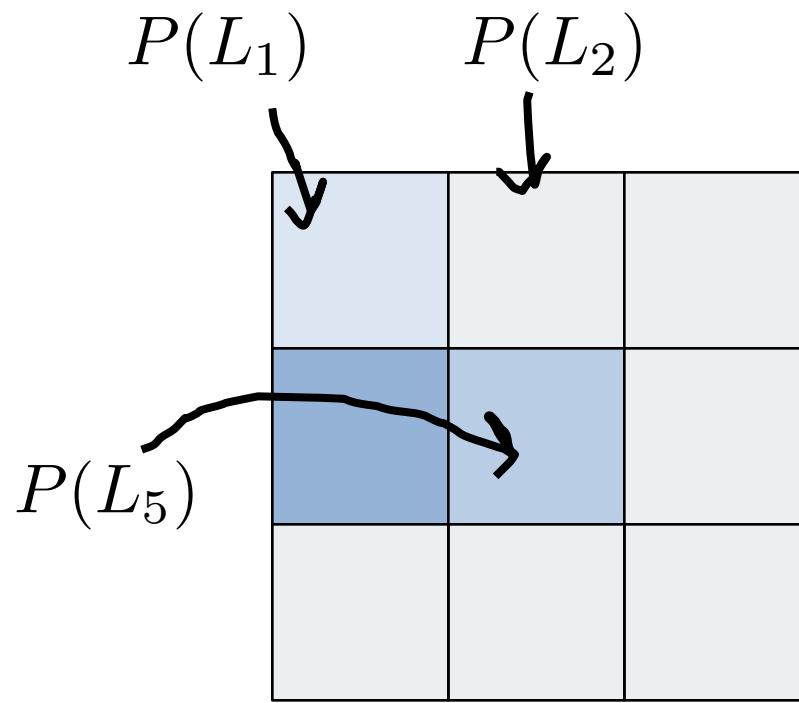
- Say 60% of all email is spam
 - 90% of spam has a forged header
 - 20% of non-spam has a forged header
 - Let E = message contains a forged header
 - Let F = message is spam
 - What is $P(F | E)$?

- Solution:
$$P(F | E) = \frac{P(E | F) P(F)}{P(E | F) P(F) + P(E | F^c) P(F^c)}$$

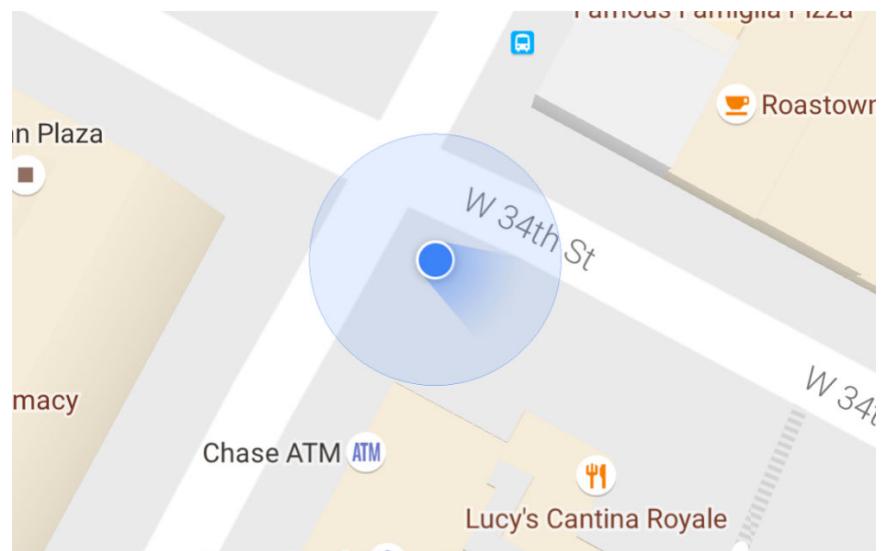
$$P(F | E) = \frac{(0.9)(0.6)}{(0.9)(0.6) + (0.2)(0.4)} \approx 0.871$$



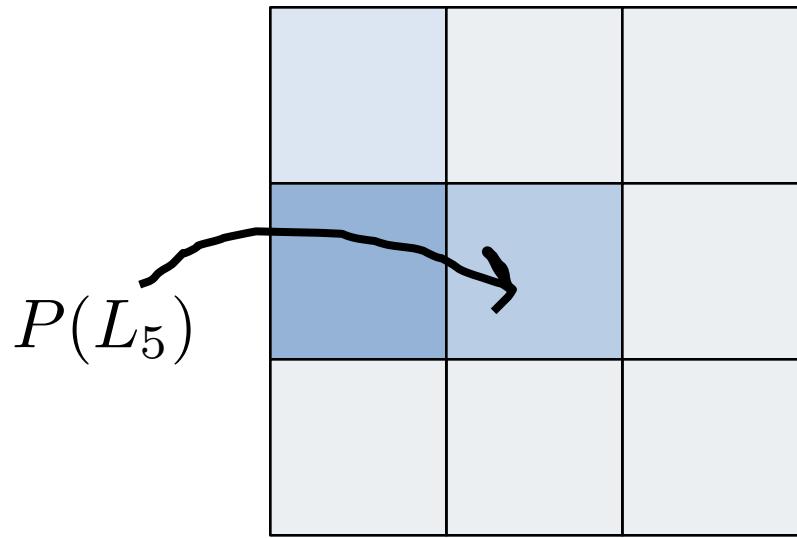
Update Belief



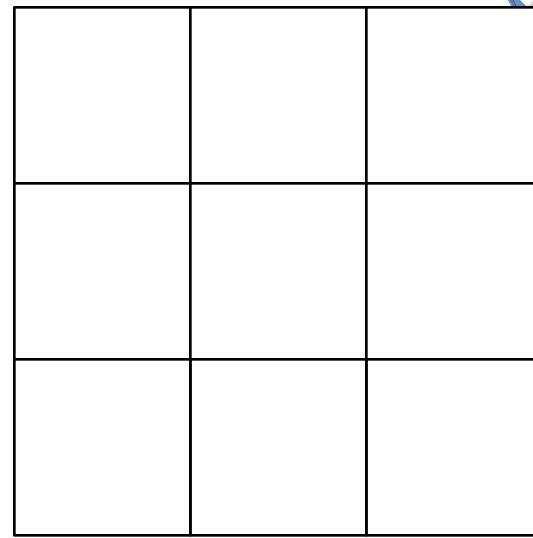
Before Observation



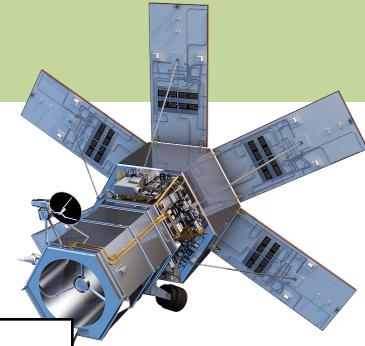
Update Belief



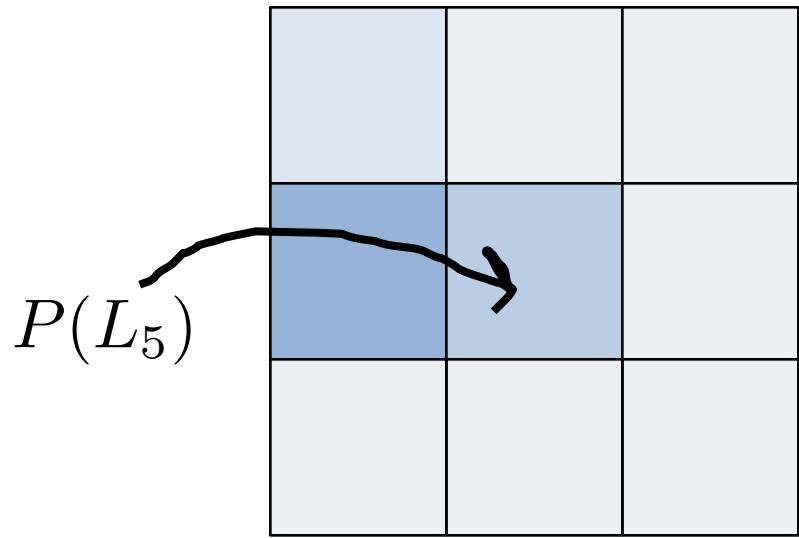
Before Observation



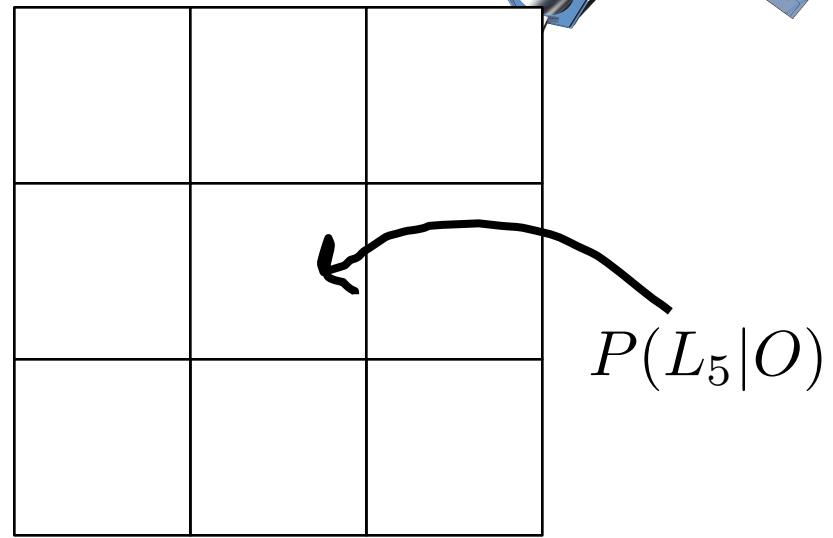
After Observation



Update Belief

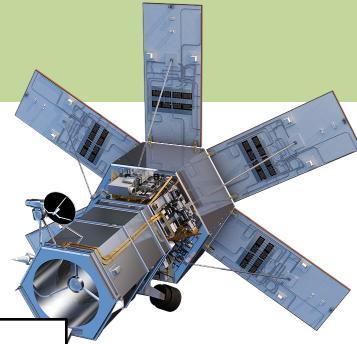


Before Observation

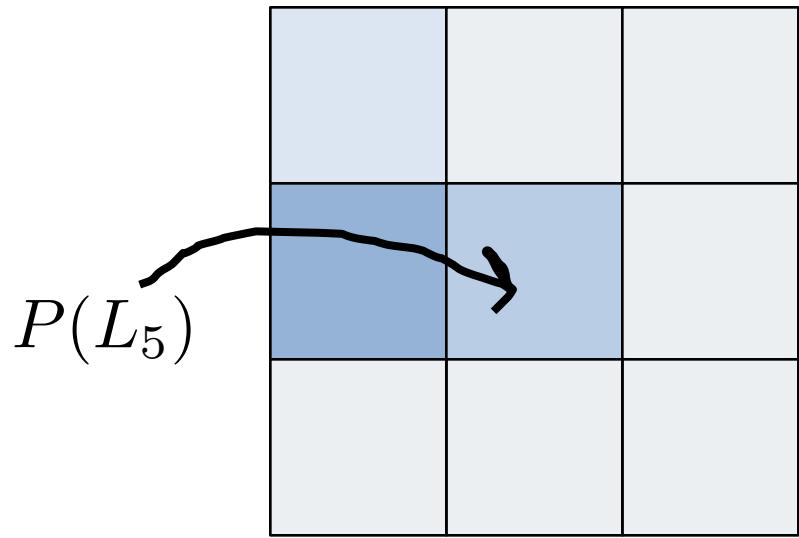


After Observation

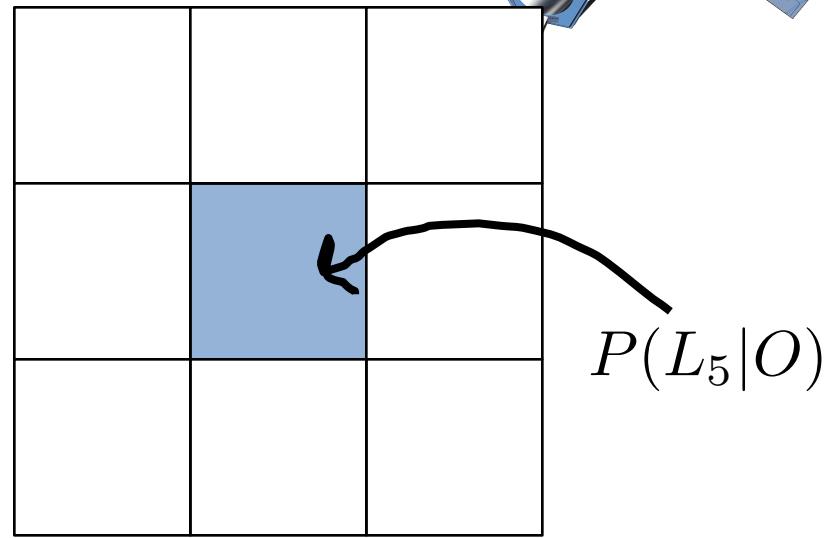
$$P(L_5|O) = \frac{P(O|L_5)P(L_5)}{P(O)}$$



Update Belief



Before Observation

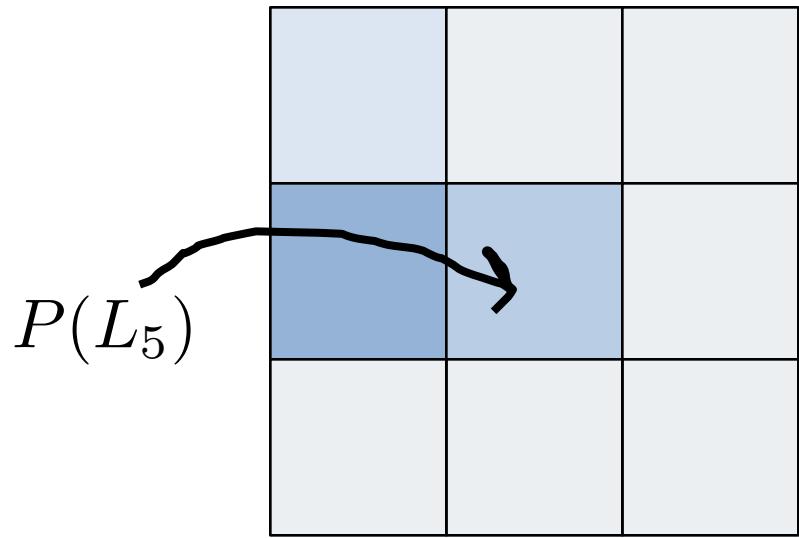


After Observation

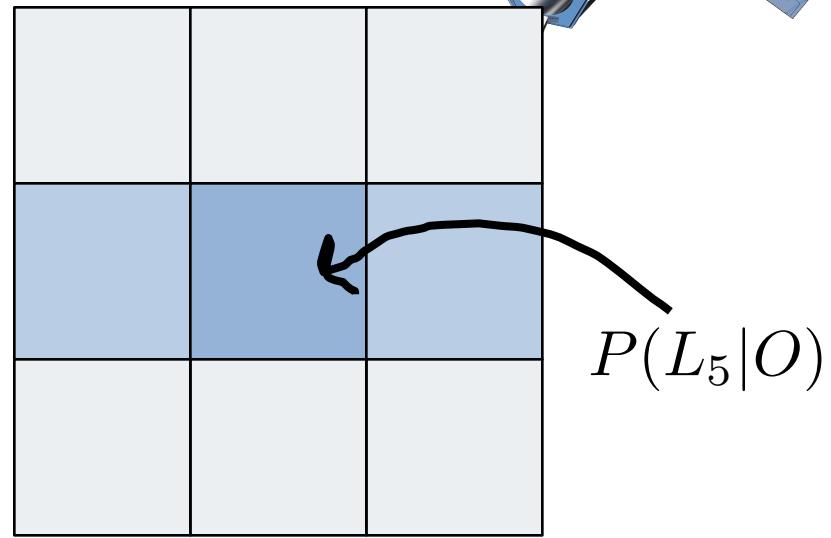
$$P(L_5|O) = \frac{P(O|L_5)P(L_5)}{P(O)}$$



Update Belief



Before Observation



After Observation

$$P(L_5|O) = \frac{P(O|L_5)P(L_5)}{P(O)}$$



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, $P(\text{win}) = 1/3$ (random)



Let's Make a Deal

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



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 = $\frac{\text{original \# envelopes} - 1}{\text{original \# envelopes}}$

