Conditional Probability

Axioms of Probability

Recall: S = all possible outcomes. E = the event.

- Axiom 1: $0 \leq P(E) \leq 1$
- Axiom 2: $P(S) = 1$
- Axiom 3: $P(E^c) = 1 - P(E)$

Aside: axiom 3 is often stated as the probability of mutually exclusive events. We'll come back to that later in the lecture…

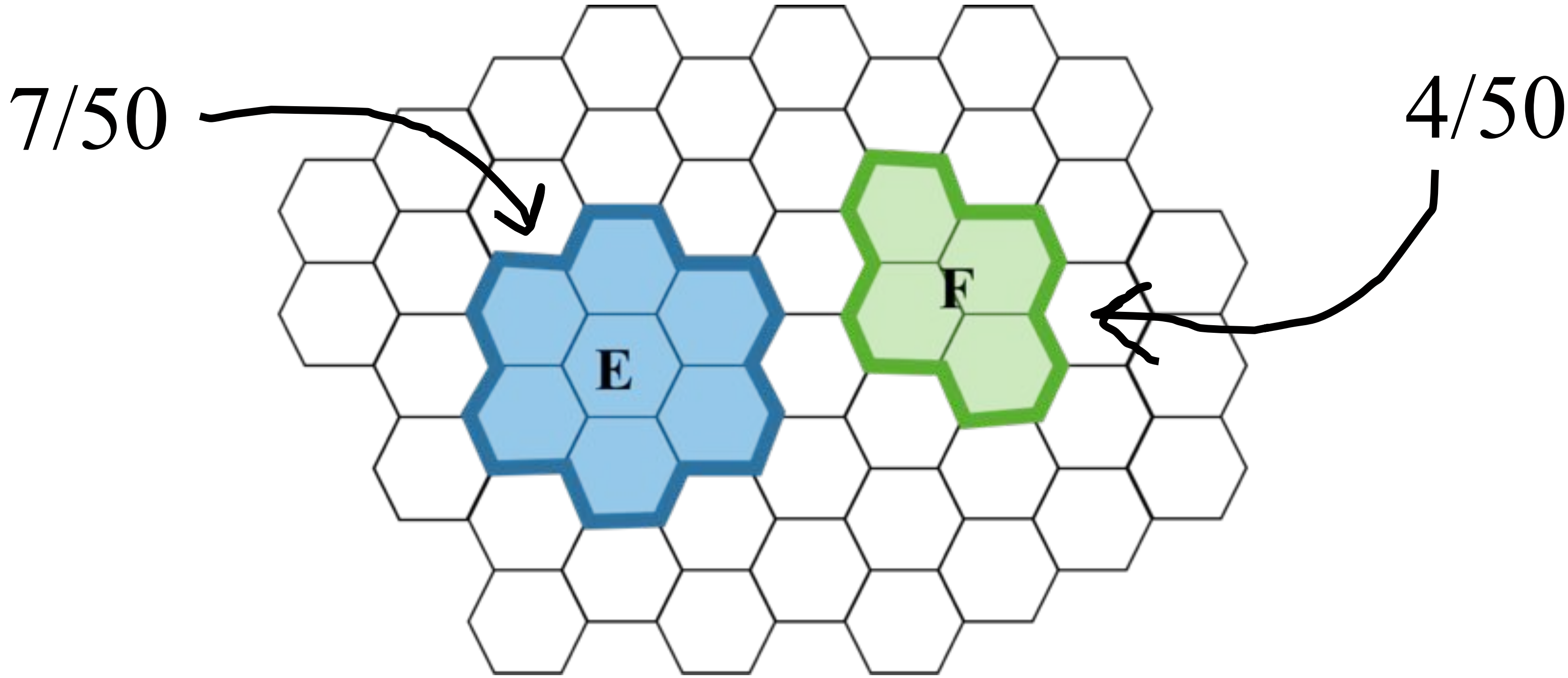
Axioms of Probability

Recall: S = all possible outcomes. E = the event.

- Axiom 1: $0 \leq P(E) \leq 1$
- Axiom 2: $P(S) = 1$
- Axiom 3: If events E and F are mutually exclusive:

$$P(E \cup F) = P(E) + P(F)$$

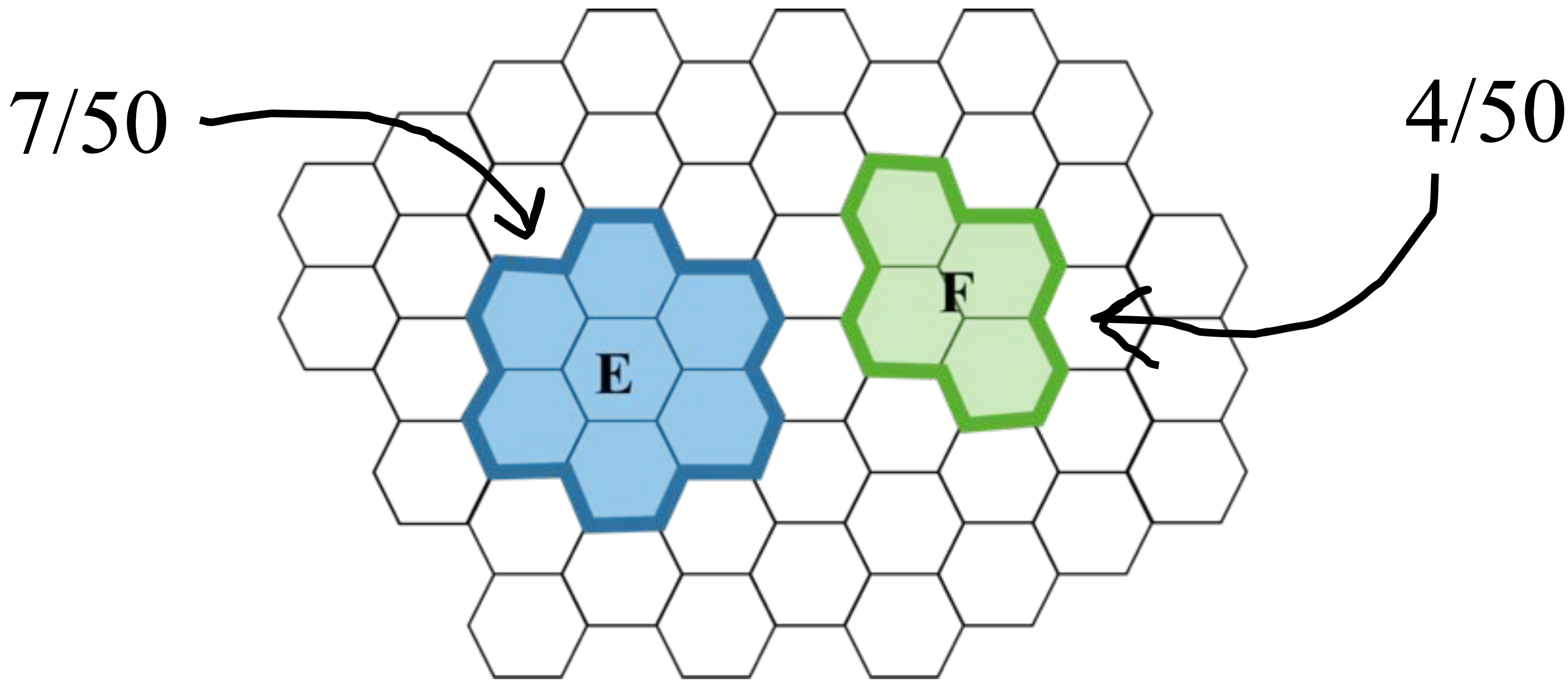
Mutually Exclusive Events



If events are mutually exclusive, probability of OR is simple:

$$P(E \cup F) = P(E) + P(F)$$

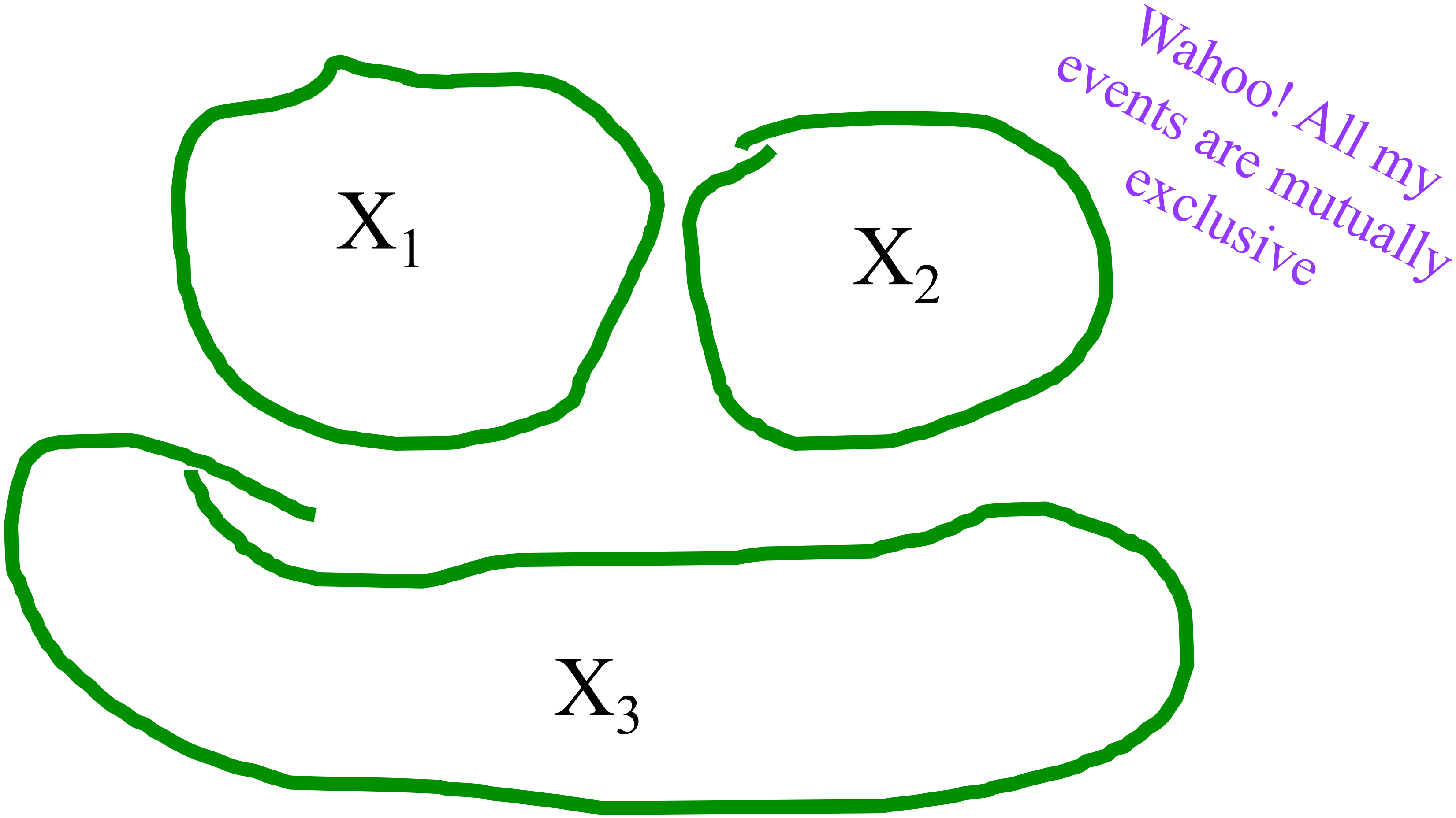
Mutually Exclusive Events



If events are mutually exclusive, probability of OR is simple:

$$P(E \cup F) = \frac{7}{50} + \frac{4}{5} = \frac{11}{50}$$

OR with Many Mutually Exclusive Events



$$P(X_1 \cup X_2 \cup \cdots \cup X_n) = \sum_{i=1}^{n} P(X_i)$$

If events are *mutually exclusive* probability of OR is easy!

$P(E^c) = 1 - P(E)$?

$$P(E \cup E^c) = P(E) + P(E^c)$$

Since E and E^c are mutually exclusive

$$P(S) = P(E) + P(E^c)$$

Since everything must either be in E or E^c

$$1 = P(E) + P(E^c)$$

Axiom 2

$$P(E^c) = 1 - P(E)$$

Rearrange

Why study probability?

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 $D_1 + D_2 = 4$

- Q: What do you think is the best outcome for D_1 ?

- Your Choices:
  - 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"

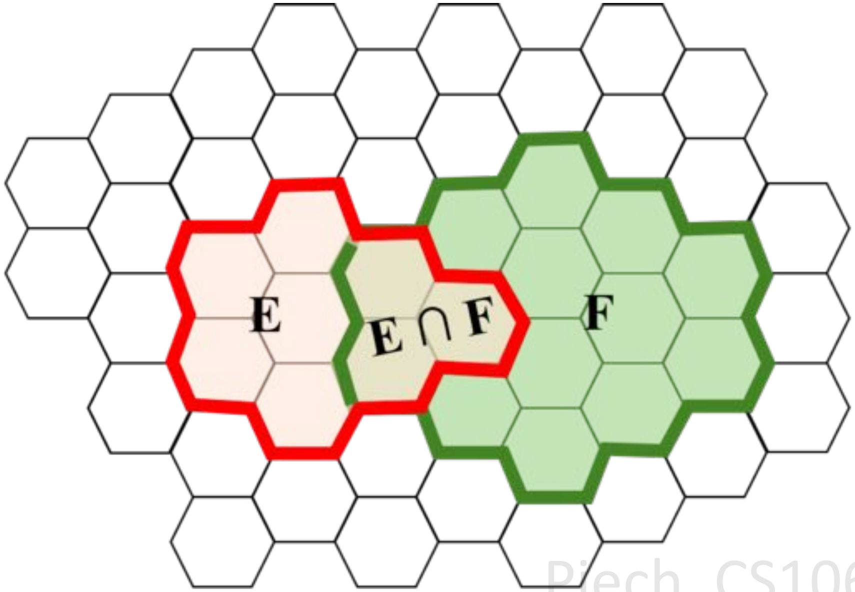
- Written as $P(E|F)$
  - Means "P(E, given F already observed)"
  - Sample space, S, reduced to those elements consistent with F (i.e. $S \cap F$)
  - Event space, E, reduced to those elements consistent with F (i.e. $E \cap F$)

Conditional Probability

With equally likely outcomes:

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

$$= \frac{|EF|}{|SF|} = \frac{|EF|}{|F|}$$



$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) = \frac{3}{14} \approx 0.21$$

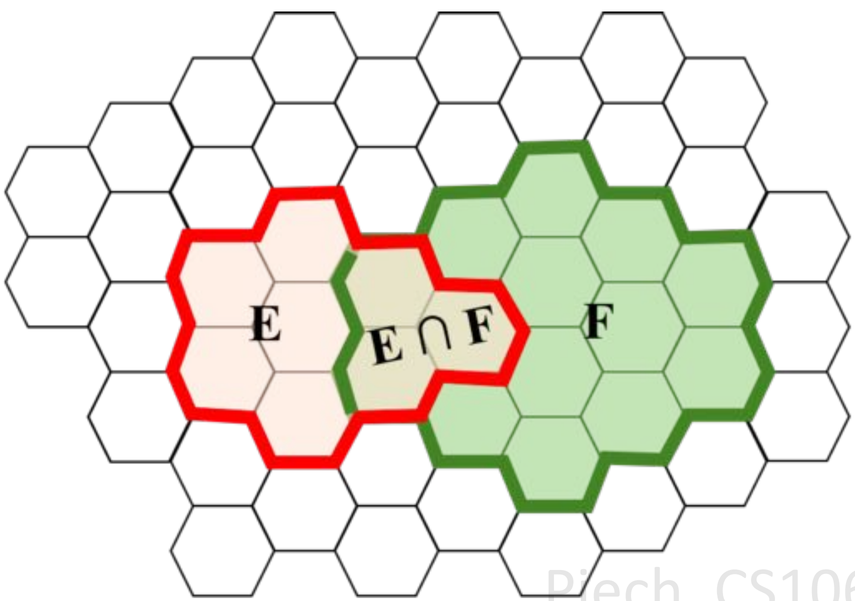
Conditional Probability

With equally likely outcomes:

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

$$= \frac{|EF|}{|SF|} = \frac{|EF|}{|F|}$$

Shorthand notation for set intersection (aka set "and")



$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) = \frac{3}{14} \approx 0.21$$

Conditional Probability

- General definition:

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

- 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:

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

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

Conditional Paradigm

Name of Rule	Original Rule	Conditional Rule
First axiom of probability	$0 \leq P(E) \leq 1$	$0 \leq P(E \mid G) \leq 1$
Complement Rule	$P(E) = 1 - P(E^C)$	$P(E \mid G) = 1 - P(E^C \mid G)$
Chain Rule	$P(EF) = P(E \mid F)P(F)$	$P(EF \mid G) = P(E \mid FG)P(F \mid G)$

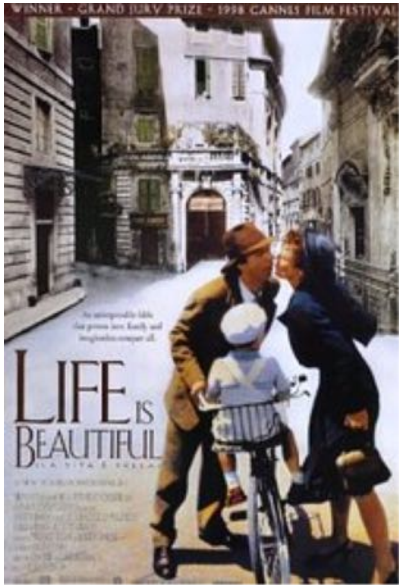
NETFLIX

+ Learn

Netflix and Learn

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

$P(E)$



S = {Watch, Not Watch}

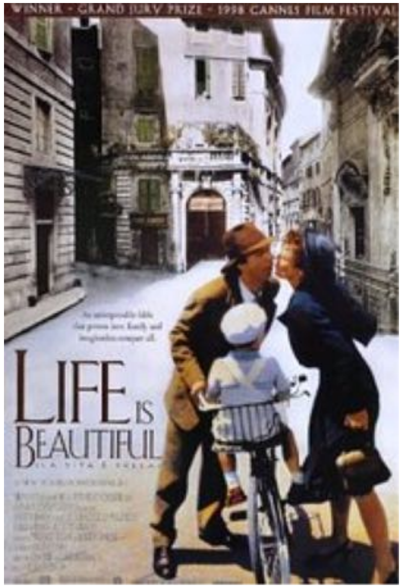
E = {Watch}

P(E) = ½ ?

Netflix and Learn

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

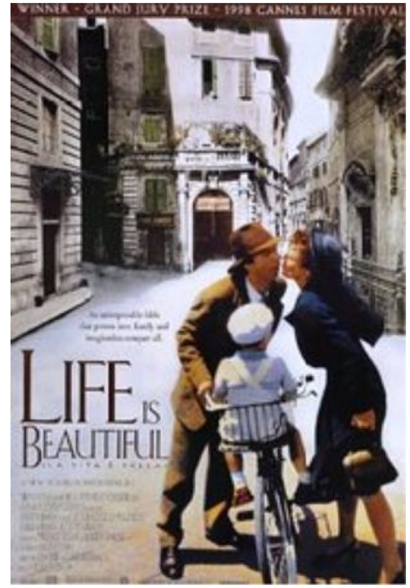
$$P(E)$$



Netflix and Learn

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

$P(E)$

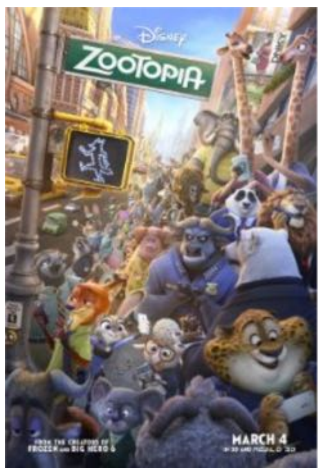


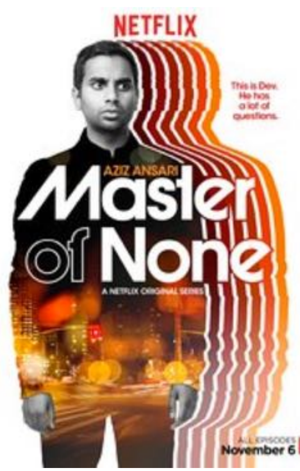
$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \approx \frac{\#\text{people who watched movie}}{\#\text{people on Netflix}}$$

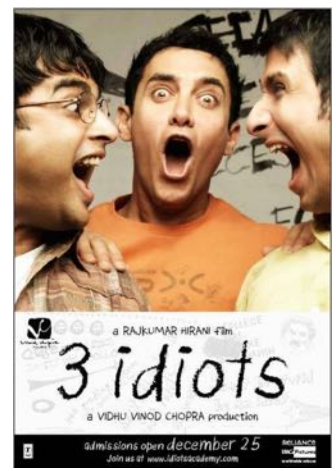
$$P(E) = 10{,}234{,}231 \;/\; 50{,}923{,}123 = 0.20$$

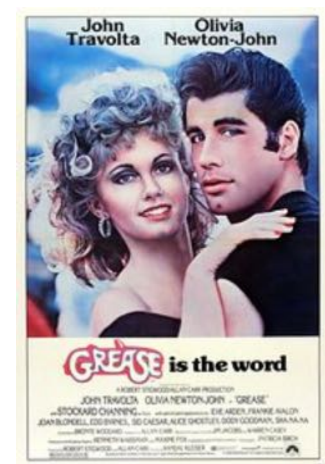
Netflix and Learn

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

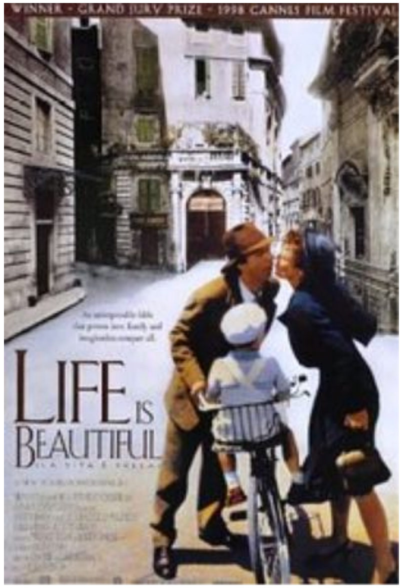
P(E) = 0.23

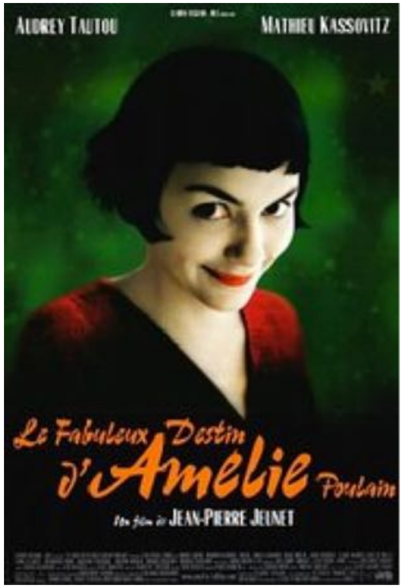
* *These are the actual estimates*

Netflix and Learn

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

$P(E|F)$



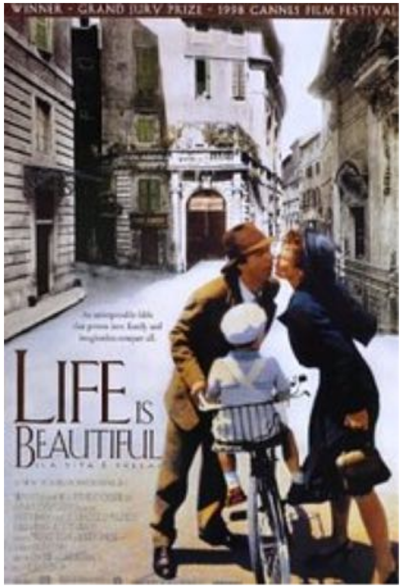


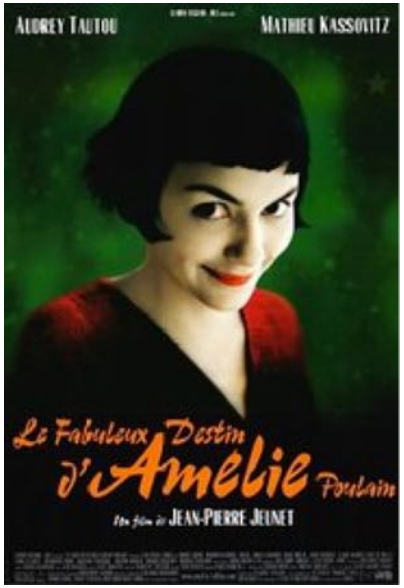
$$P(E|F) = \frac{P(EF)}{P(F)}$$

Netflix and Learn

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

$$P(E|F)$$



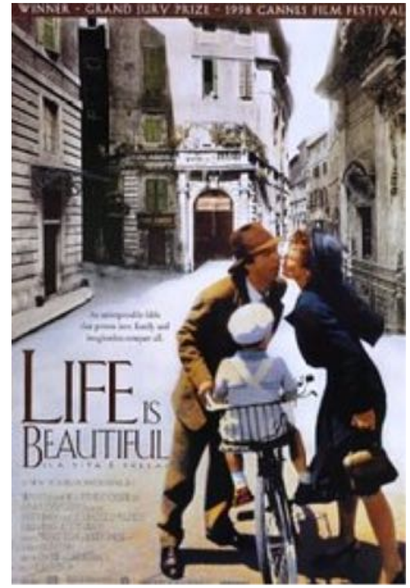


$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\frac{\#\text{people who watched both}}{\#\text{people on Netflix}}}{\frac{\#\text{people who watched } F}{\#\text{people on Netflix}}}$$

Netflix and Learn

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

$P(E|F)$



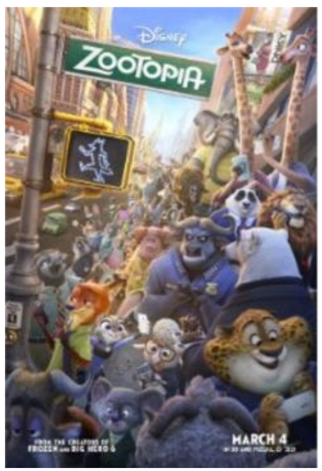
$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\#\text{people who watched both}}{\#\text{people who watched } F}$$

$P(E|F) = 0.42$

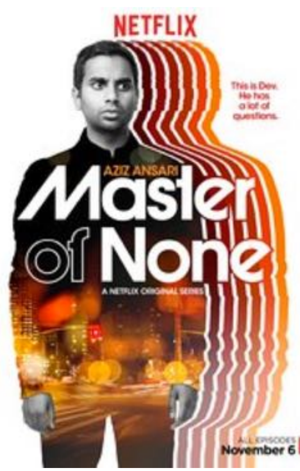
Netflix and Learn

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

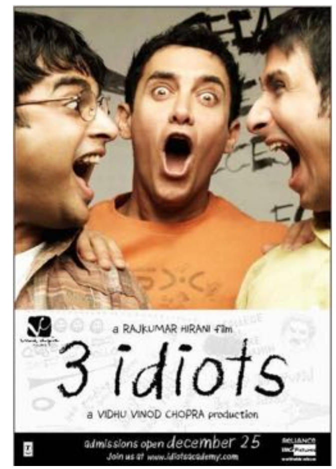
$P(E|F) =$ 0.14



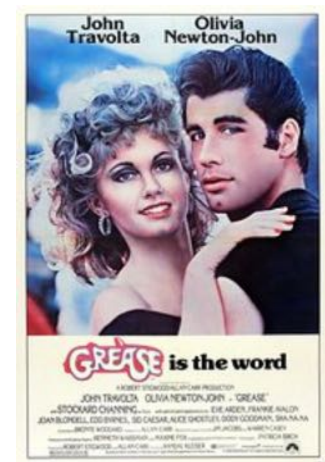
$P(E|F) =$ 0.35



$P(E|F) =$ 0.20



$P(E|F) =$ 0.72



$P(E|F) =$ 0.49

Machine Learning

Machine Learning is:
Probability + Data + Computers

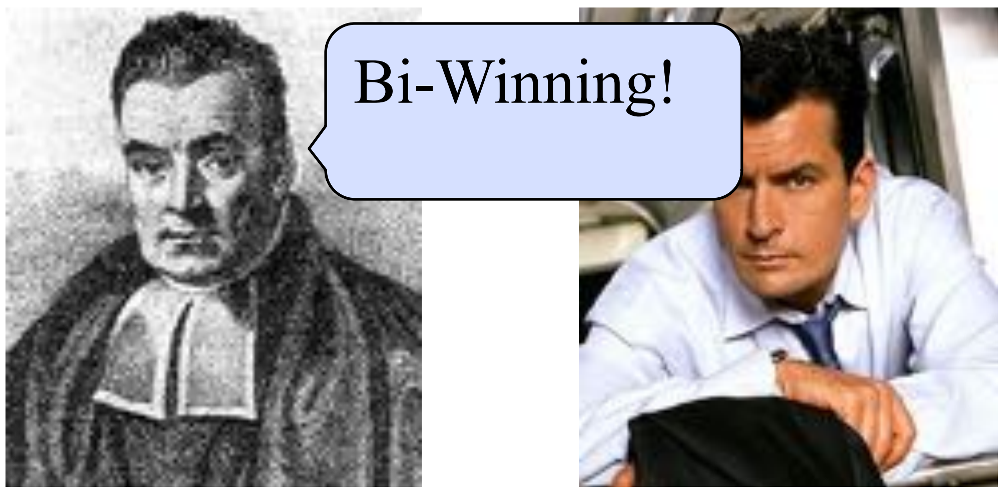
Sophomores

- There are 400 students in CS109:
  - Probability that a random student in CS109 is a Sophomore is 0.43
  - 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

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

Thomas Bayes

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



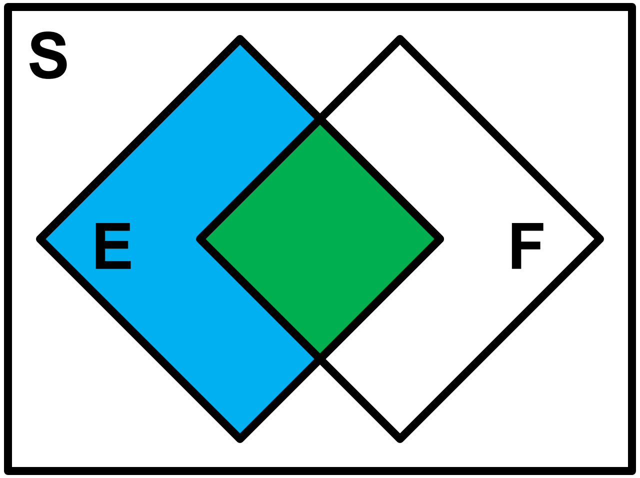
- He looked remarkably similar to Charlie Sheen
  - But that’s not important right now...

But First!

Background Observation

- Say E and F are events in S

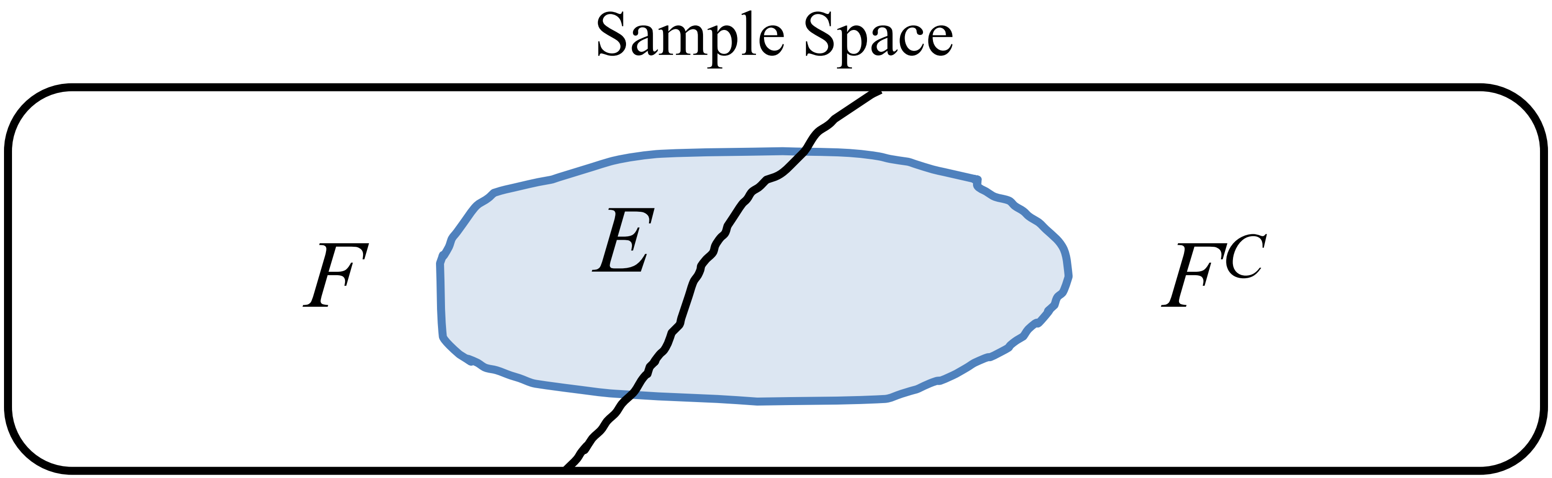
$$E = EF \cup EF^c$$



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

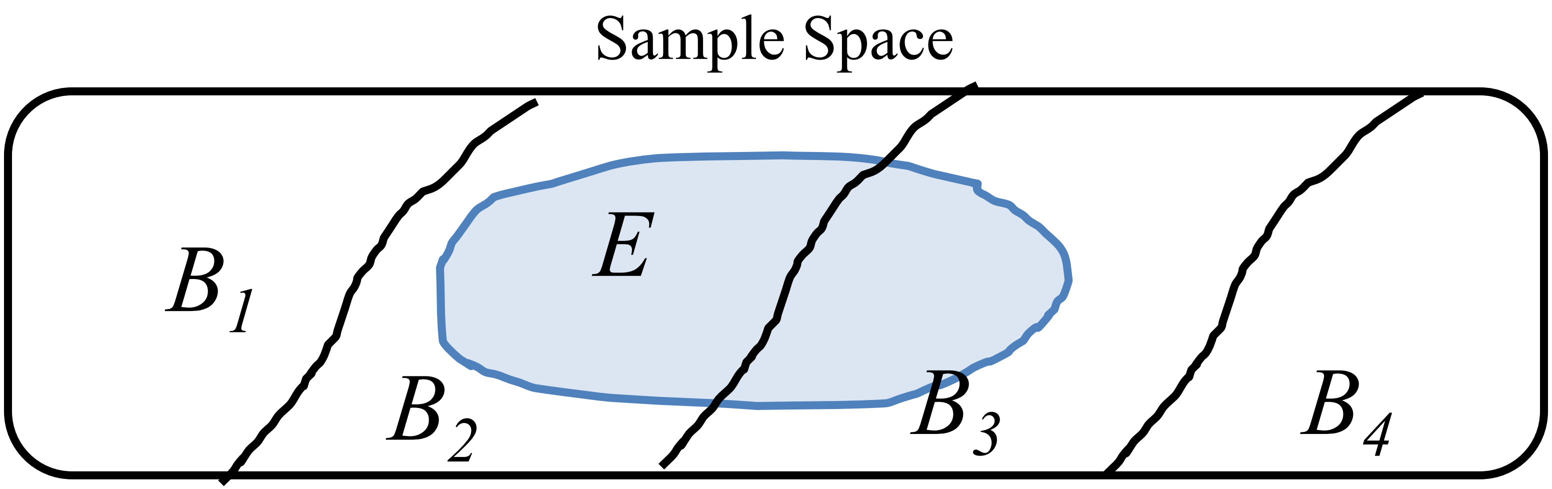
So, $P(E) = P(EF) + P(EF^c)$

Law of Total Probability



$$P(E) = P(EF) + P(EF^C)$$
$$= P(E|F)P(F) + P(E|F^C)P(F^C)$$

Law of Total Probability



$$P(E) = \sum_i P(B_i \cap E)$$

$$= \sum_i P(E|B_i)P(B_i)$$

Moment of Silence…

Bayes Theorem

$P(F \mid E)$

I want to calculate
P(State of the world F | Observation E)
It seems so tricky!…

The other way around is easy
P(Observation E | State of the world F)
What options to I have, chief?

$P(E \mid F)$

Bayes Theorem

Want $P(\ F \mid E\ )$. Know $P(\ E \mid F\ )$

$$P(F|E) = \frac{P(EF)}{P(E)} \quad \text{Def. of Conditional Prob.}$$

A little while later…

$$= \frac{P(E|F)P(F)}{P(E)} \quad \text{Chain Rule}$$

Bayes Theorem

- Most common form:

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

- Expanded form:

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

H1N1 Testing

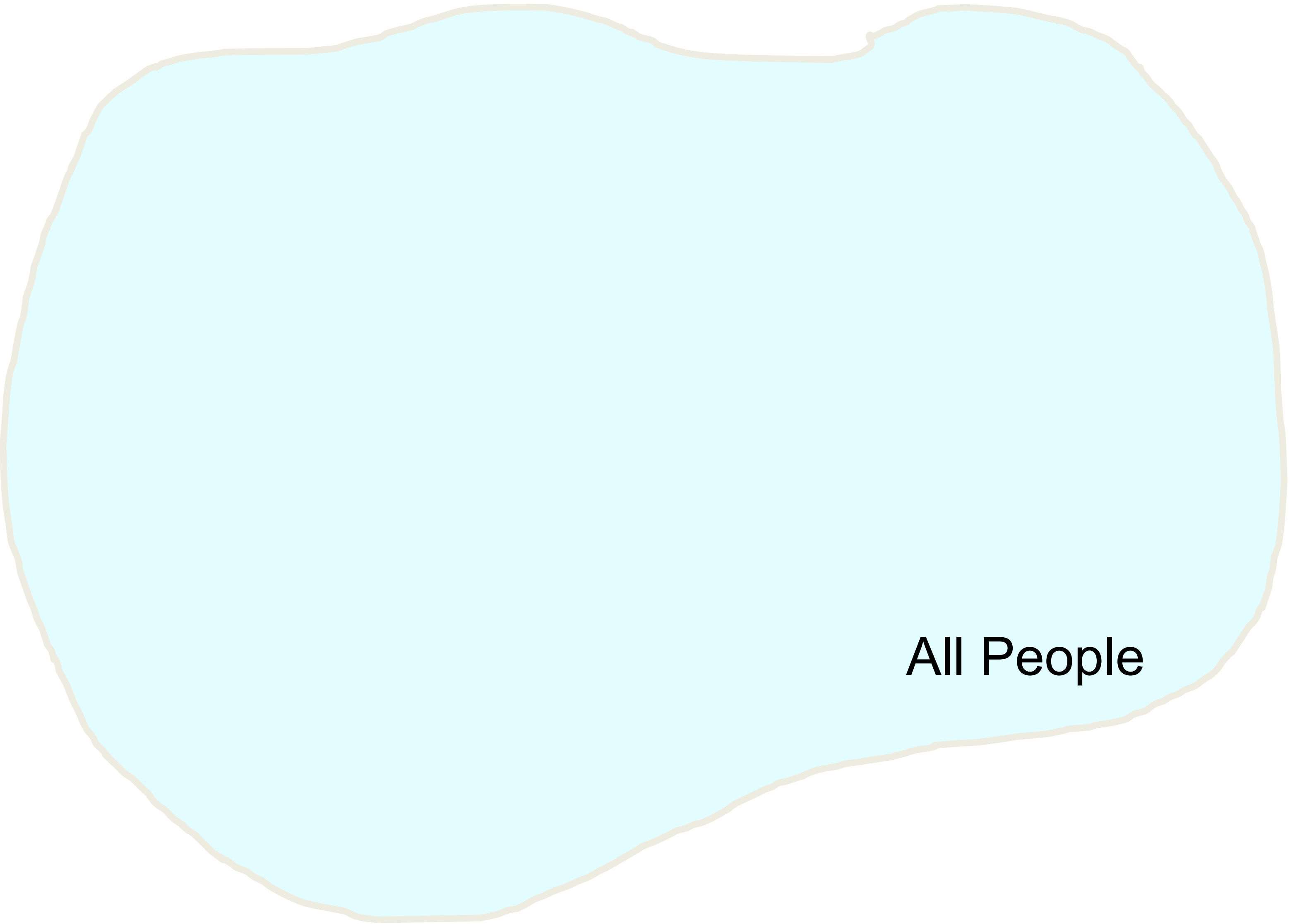
- A test is 98% effective at detecting H1N1
  - However, test has a "false positive" rate of 1%
  - 0.5% of US population has H1N1
  - Let E = you test positive for H1N1 with this test
  - Let F = you actually have H1N1
  - What is P(F | E)?
- Solution:

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

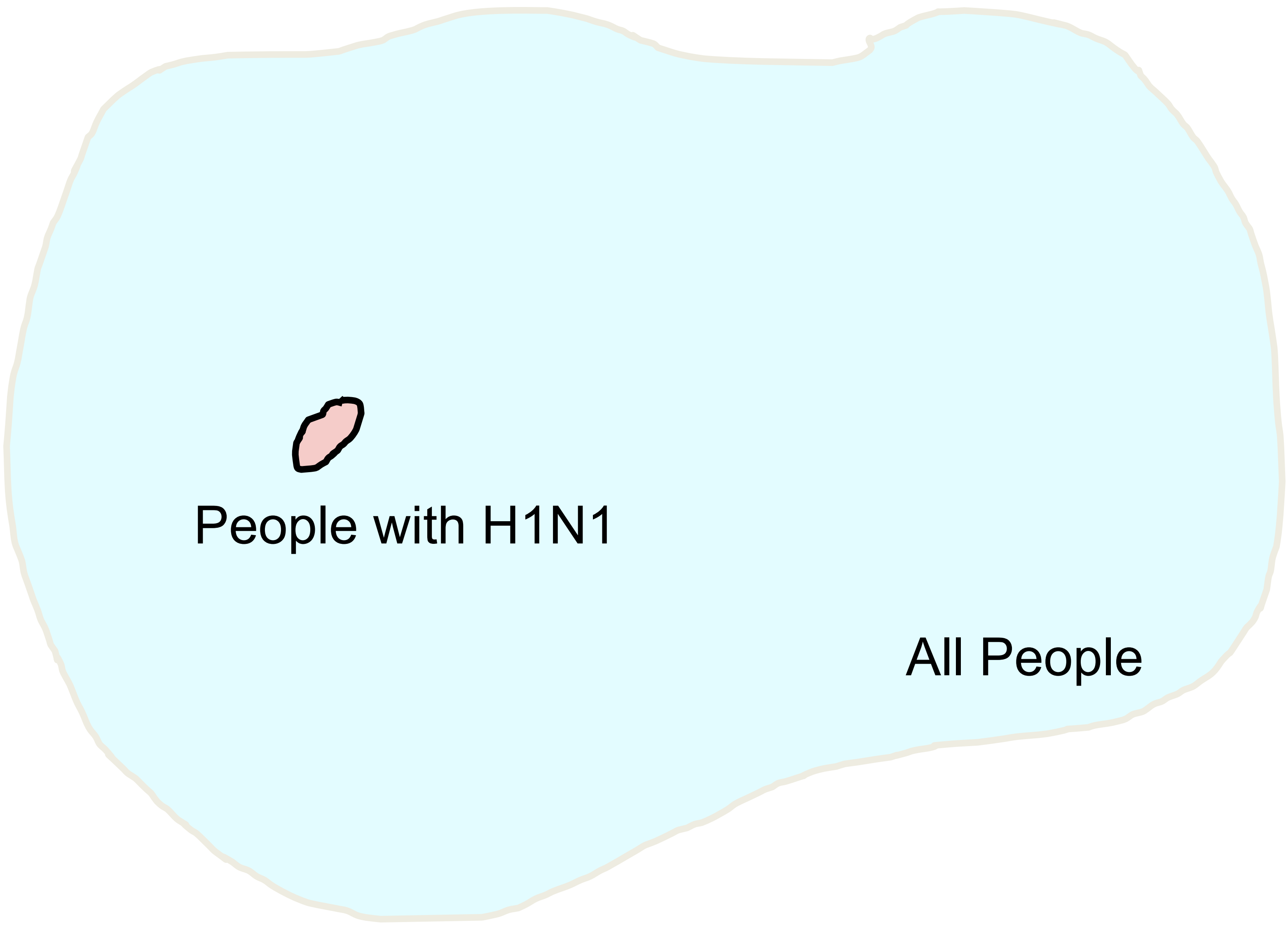
$$P(F \mid E) = \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.01)(1 - 0.005)} \approx 0.330$$

Intuition Time

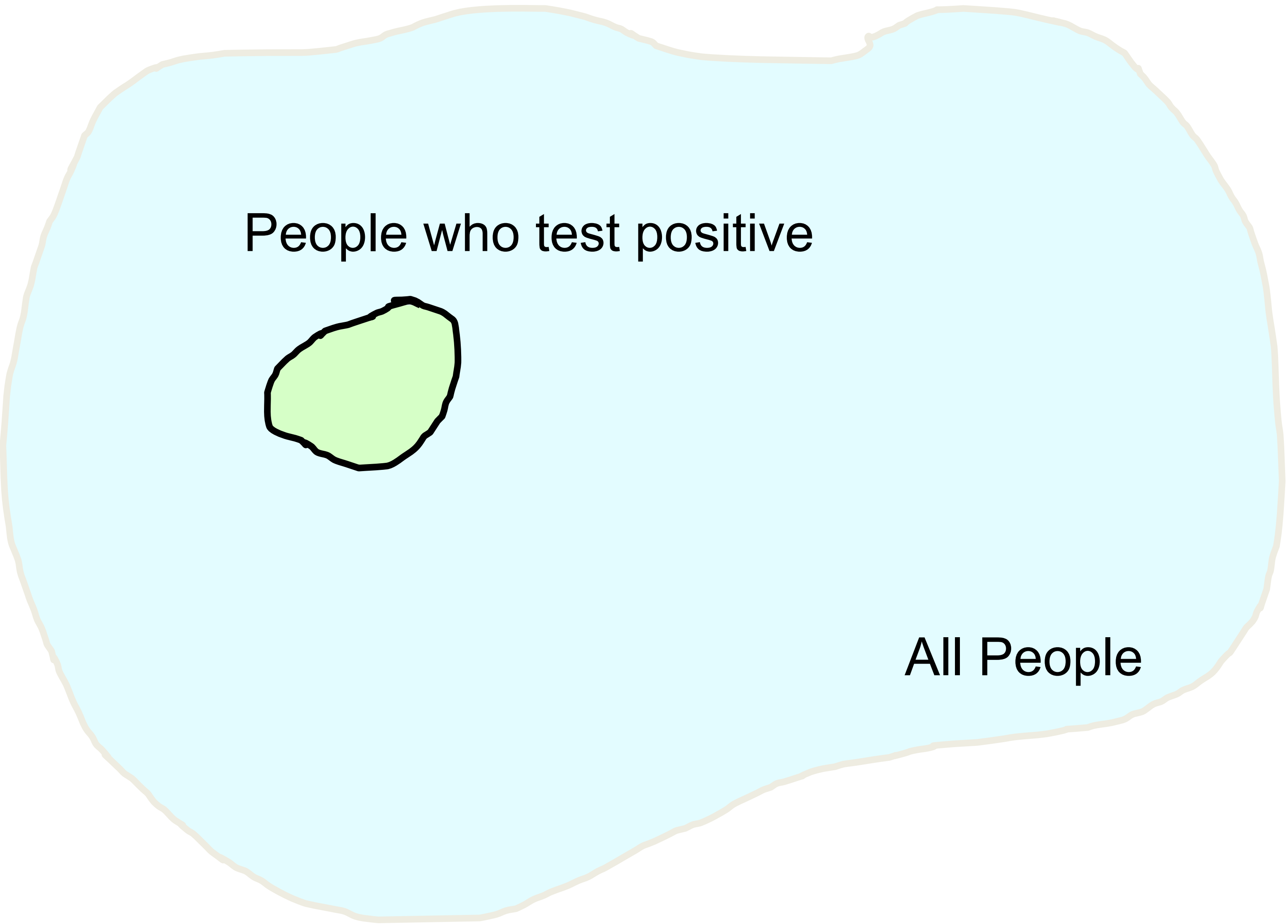
Bayes Theorem Intuition



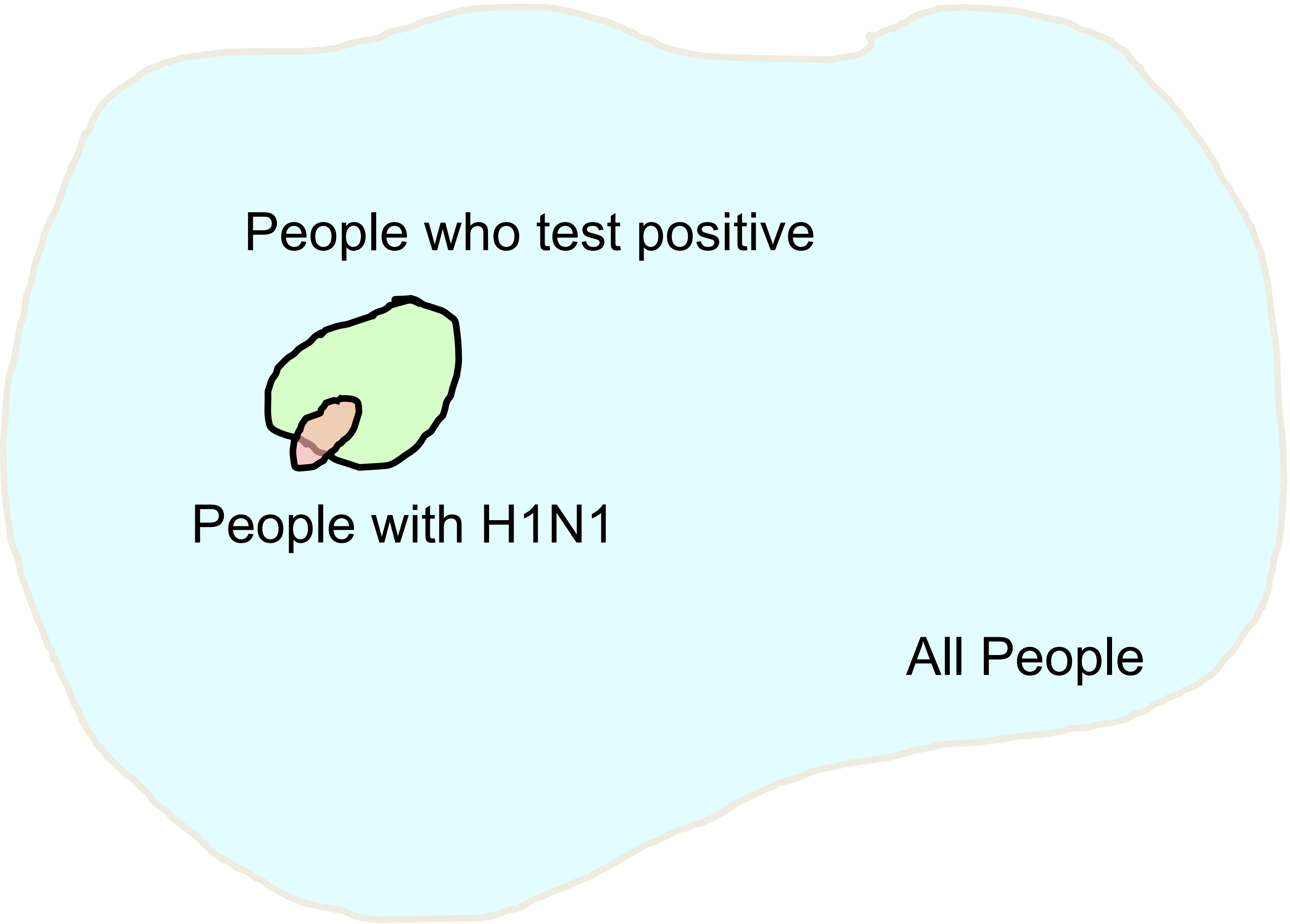
Bayes Theorem Intuition



Bayes Theorem Intuition

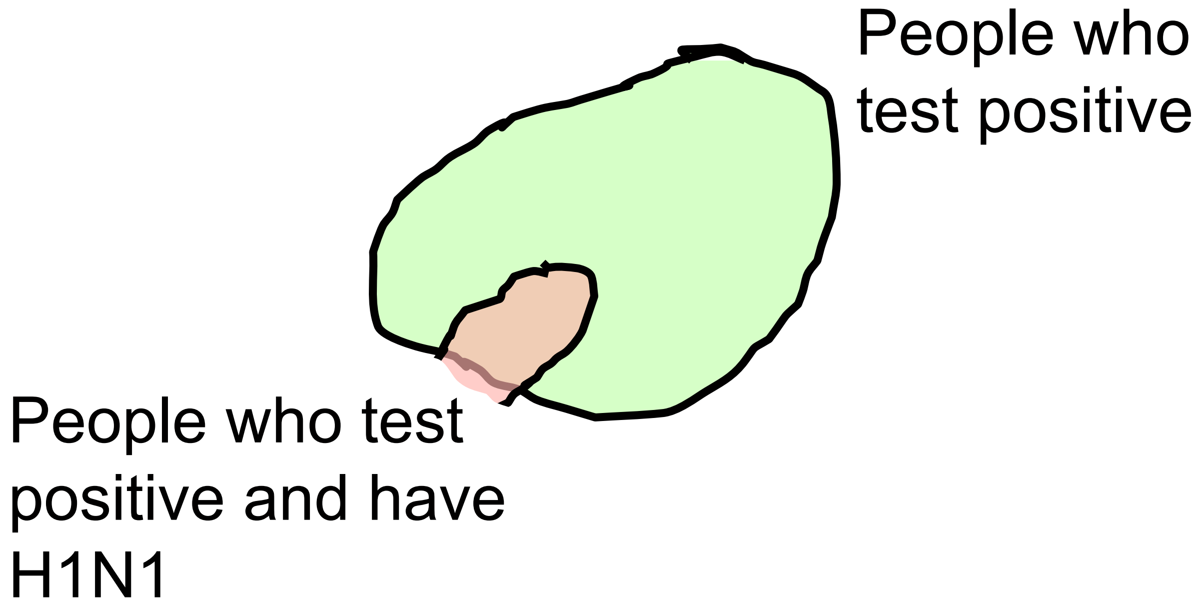


Bayes Theorem Intuition



Bayes Theorem Intuition

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



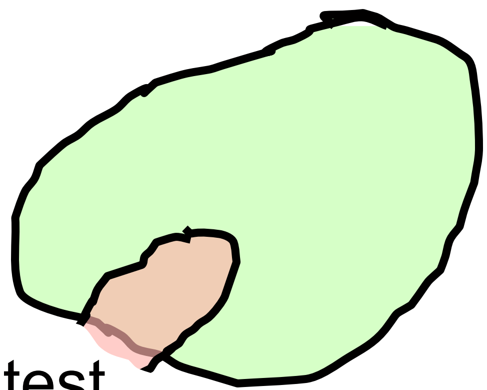
≈ 0.330

Bayes Theorem Intuition

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

People who test positive

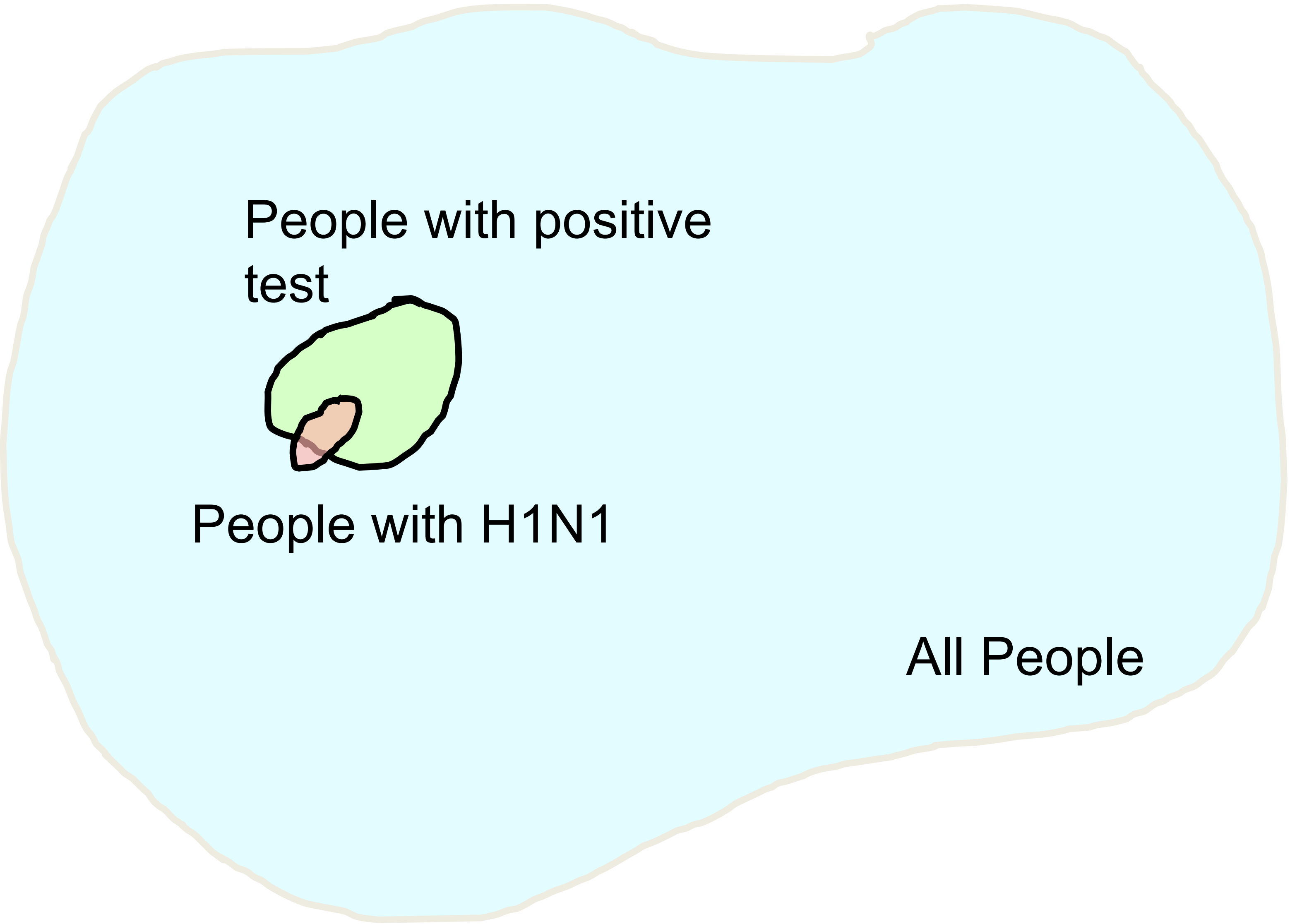
$$P(F)P(E|F) + P(F^c)P(E|F^c)$$



People who test positive and have H1N1 $P(F)P(E|F)$

≈ 0.330

Bayes Theorem Intuition



Bayes Theorem Intuition

Say we have 1000 people:

5 have H1N1 and test positive, 985 **do not** have H1N1 and test negative. 10 **do not** have H1N1 and test positive.

≈ 0.333

Why It's Still Good to get Tested

	H1N1 +	H1N1 –
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 H1N1 with this test
- Let F = you actually have H1N1
- What is $P(F \mid E^c)$?

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

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

Slicing Up Spam

In 2010 88% of email was spam

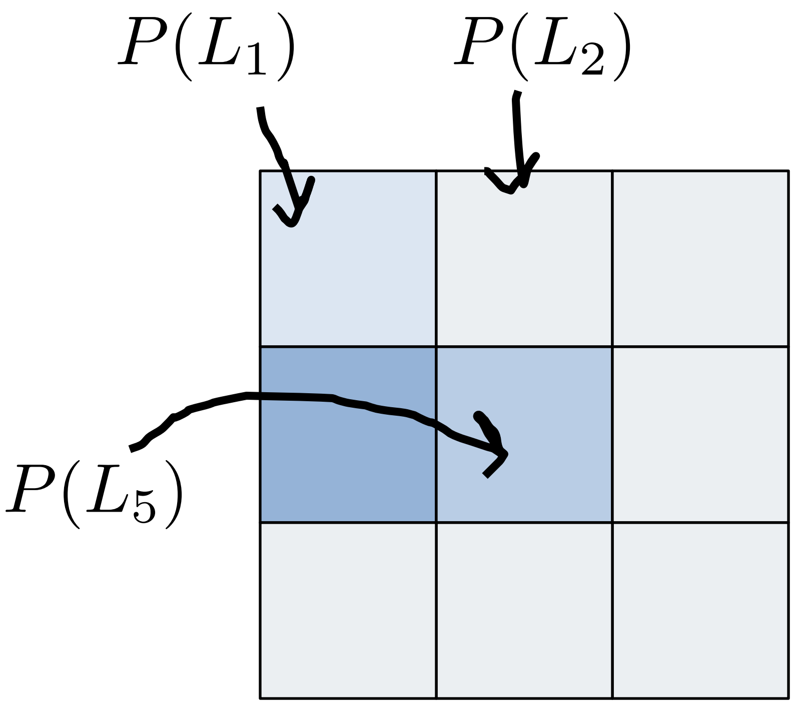
Simple Spam Detection

- 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 $P(F \mid E)$?
- Solution:

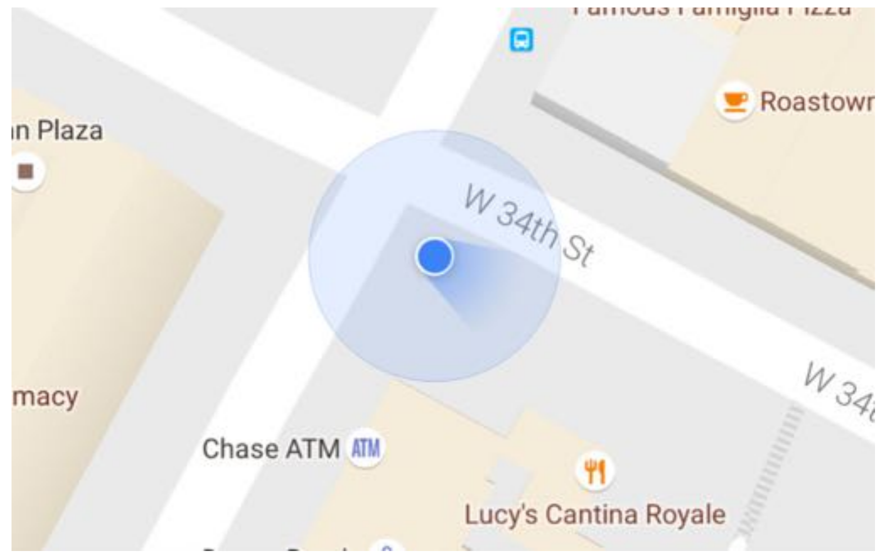
$$P(F \mid E) = \frac{P(E \mid F)\, \mathrm{P}(F)}{P(E \mid F)\, \mathrm{P}(F) + \mathrm{P}(E \mid F^c)\, P(F^c)}$$

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

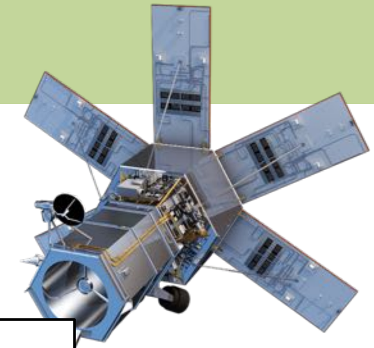
Update Belief



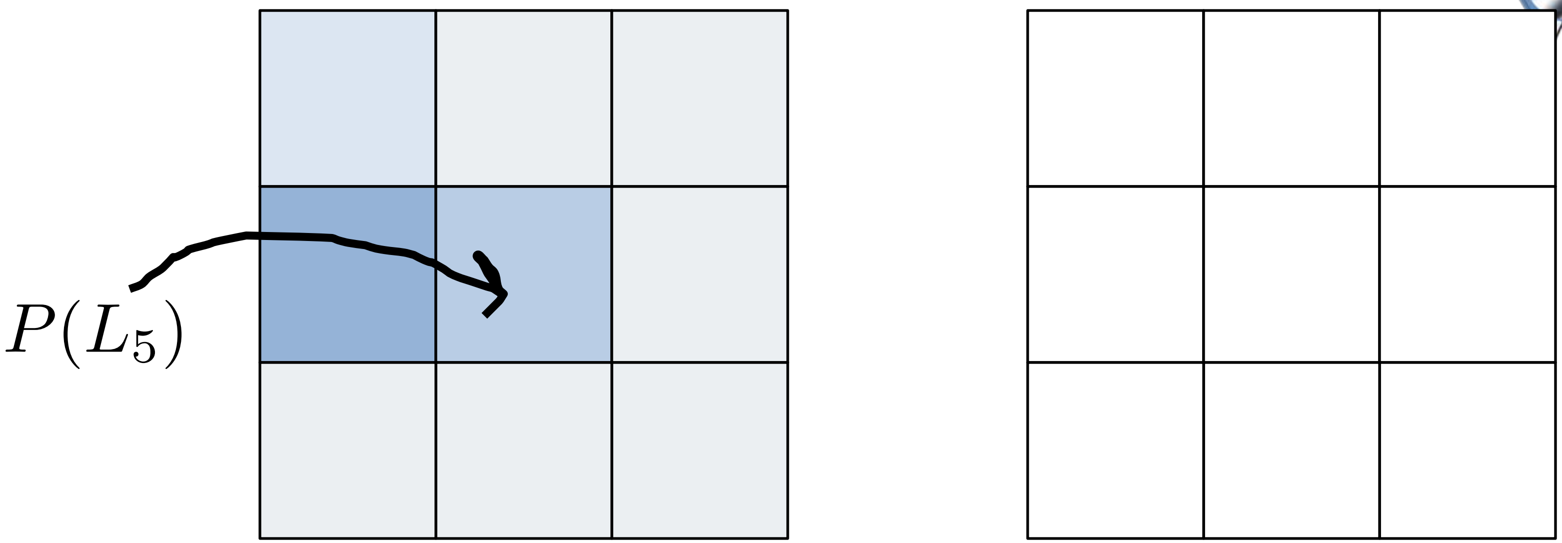
Before Observation



Update Belief



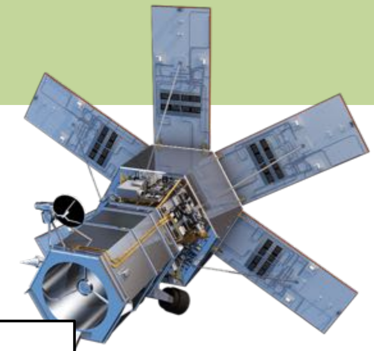
Know: $P(O|L_i)$



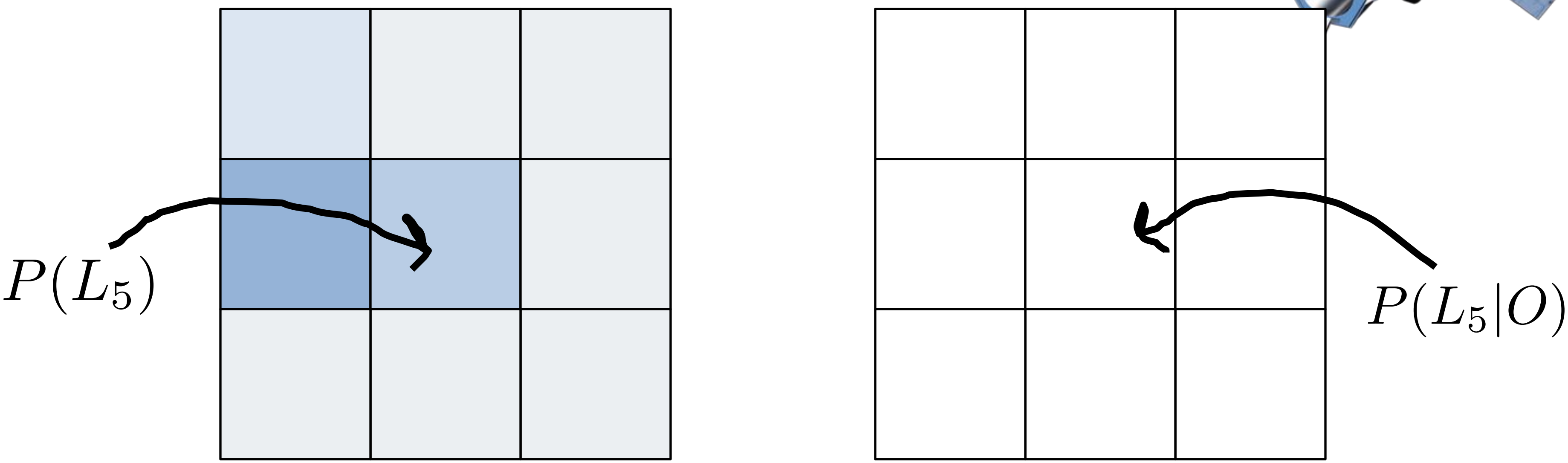
Before Observation

After Observation

Update Belief



Know: $P(O|L_i)$

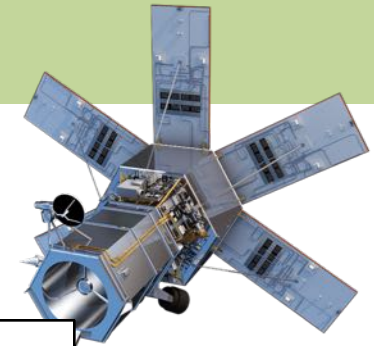


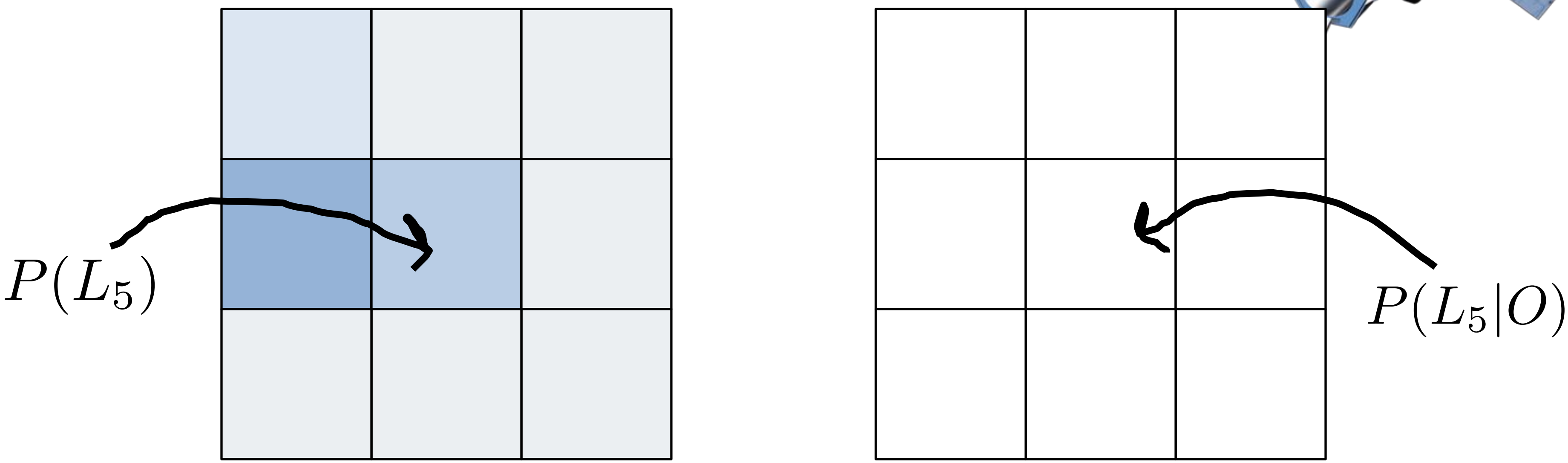
Before Observation

After Observation

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

Update Belief



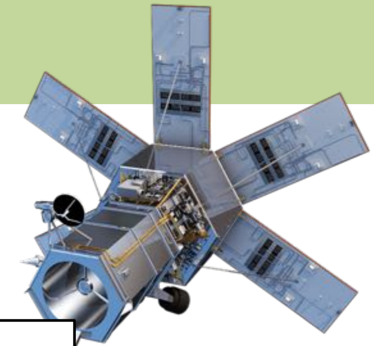


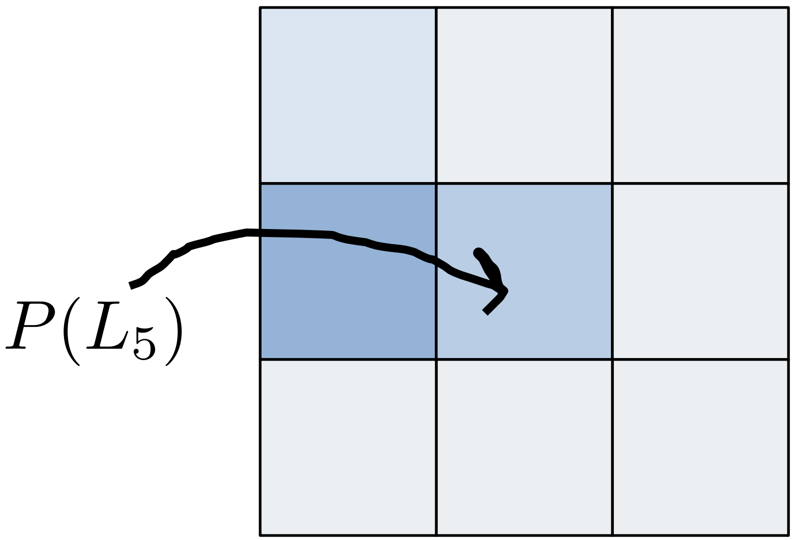
Before Observation

After Observation

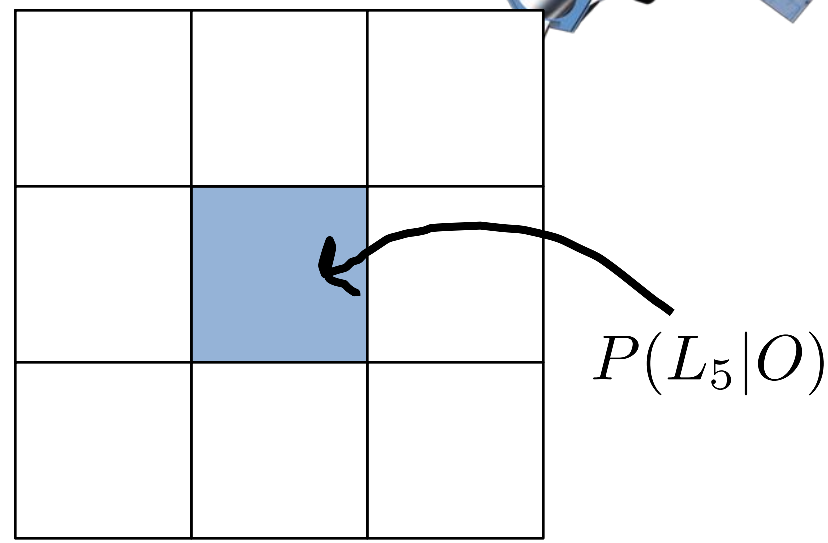
$$P(L_5|O) = \frac{P(O|L_5)P(L_5)}{\sum_i P(O|L_i)P(L_i)}$$

Update Belief





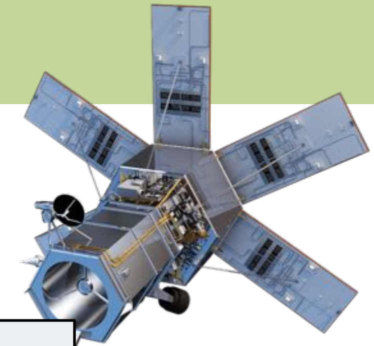
Before Observation

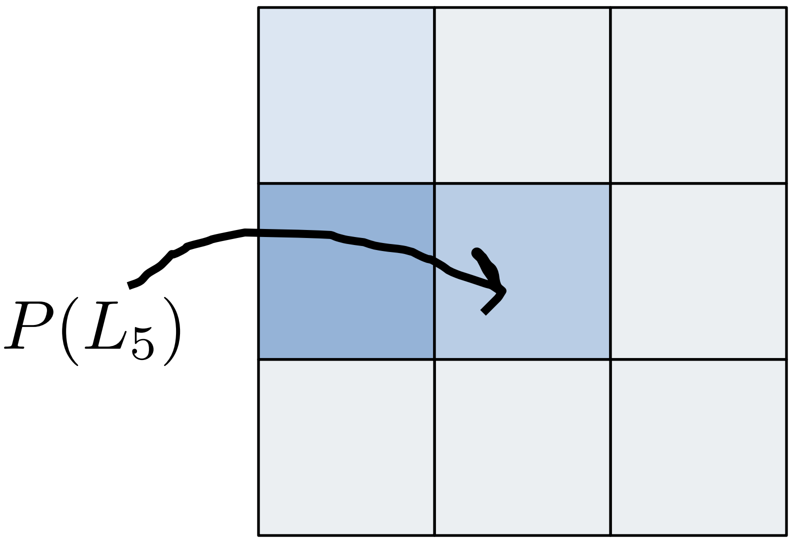


After Observation

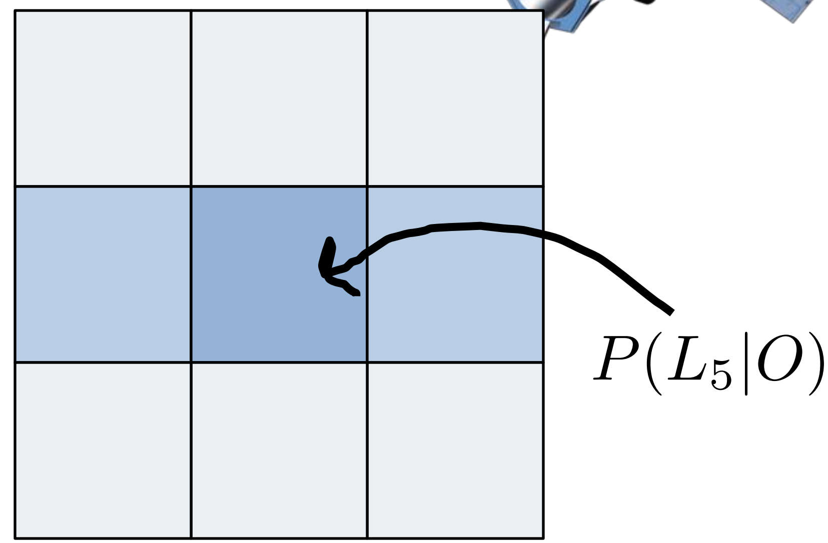
$$P(L_5|O) = \frac{P(O|L_5)P(L_5)}{\sum_i P(O|L_i)P(L_i)}$$

Update Belief





Before Observation



After Observation

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

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 always lose by switching
    - P(win | A is winner, picked A, switched) = 0
  - P(B 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(win | B is winner, picked A, switched) = 1
  - P(C 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(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 = $\frac{\text{original \# envelopes - 1}}{\text{original \# envelopes}}$