Conditional Probability

Review

Make indistinct items **distinct** to get equally likely sample space outcomes

*You will need to use this "trick" with high probability

Many times it is easier to calculate $P(E^C)$.

Making History

- What is the probability that in the *n* shuffles seen since the start of time, yours is unique?
  - $|S| = (52!)^n$
  - $|E| = (52! - 1)^n$
  - P(no deck matching yours) = $(52!-1)^n/(52!)^n$

- For $n = 10^{20}$,
  - P(deck matching yours) < 0.000000001

* Assume 7 billion people have been shuffling cards once a second since cards were invented

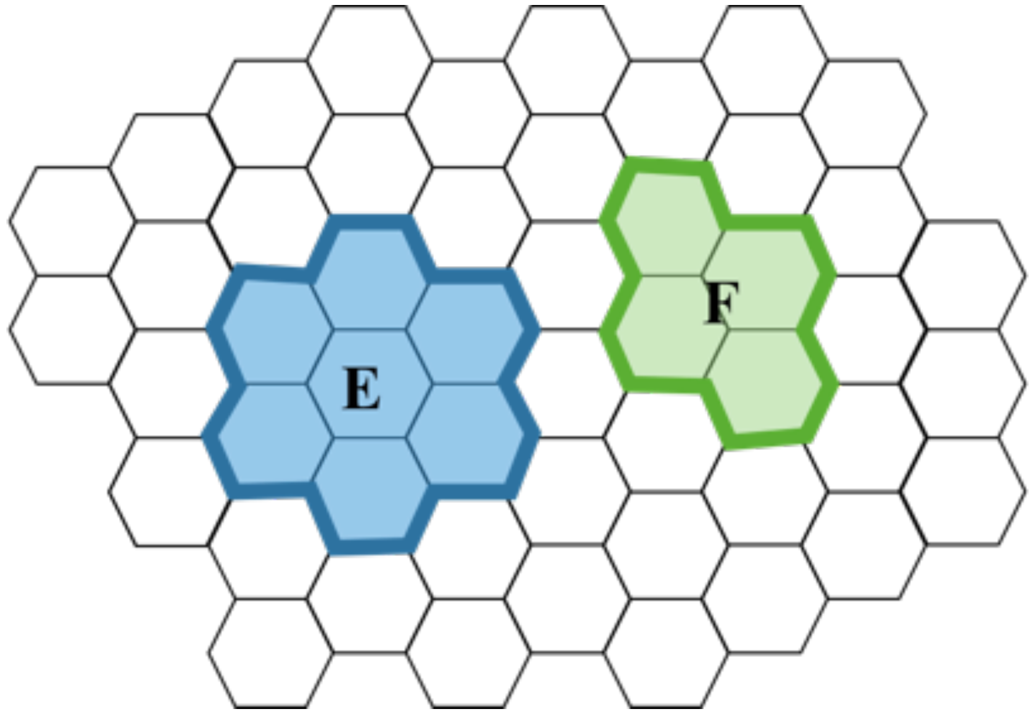
Axioms of Probability

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:

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

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

Conditional Probability

- **Conditional probability** is probability that E occurs *given* that F has already occurred "Conditioning on F"

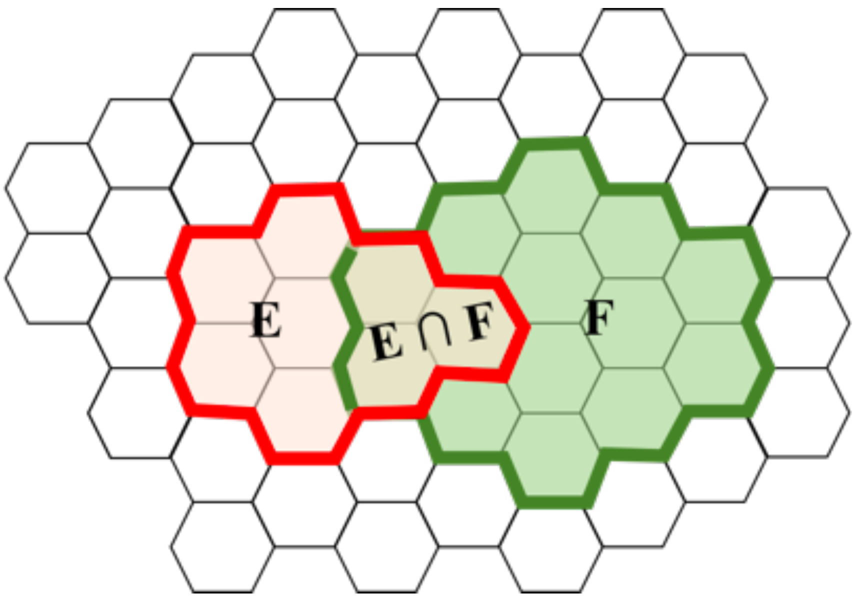
- Written as $P(E|F)$
  - Means "P(E, given F already observed)"
  - Sample space, S, reduced to those elements consistent with F (i.e. $S \cap F$)
  - Event space, E, reduced to those elements consistent with F (i.e. $E \cap F$)

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

- 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)

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	$P(EF) = P(E \mid F)P(F)$	$P(EF \mid G) = P(E \mid FG)P(F \mid G)$

End Review

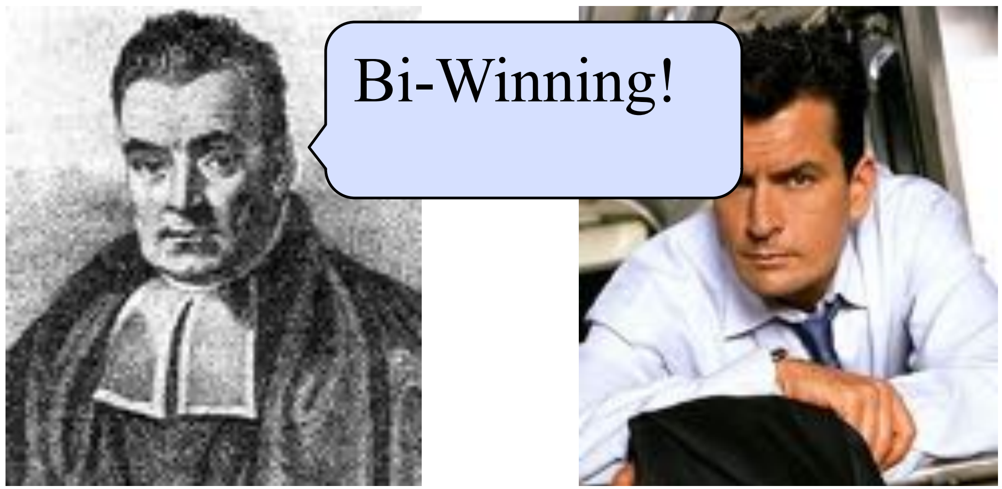
Undergraduates

- There are 46 students in CS109:
  - Probability that a random student in CS109 is an undergraduate is 0.48
  - We can observe the probability that a student is both an undergraduate and is in class
  - What is the conditional probability of a student coming to class given that they are an undergraduate?
- Solution:
  - S is the event that a student is an undergraduate
  - 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



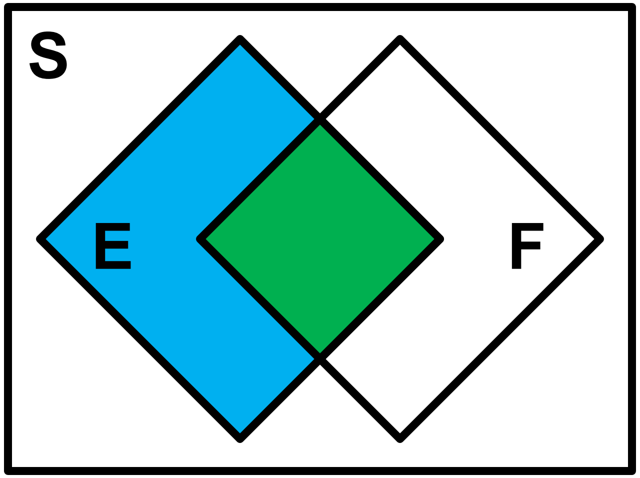
- He looked remarkably similar to Charlie Sheen
  - But that's not important right now...

But First!

Background Observation

- Say E and F are events in S

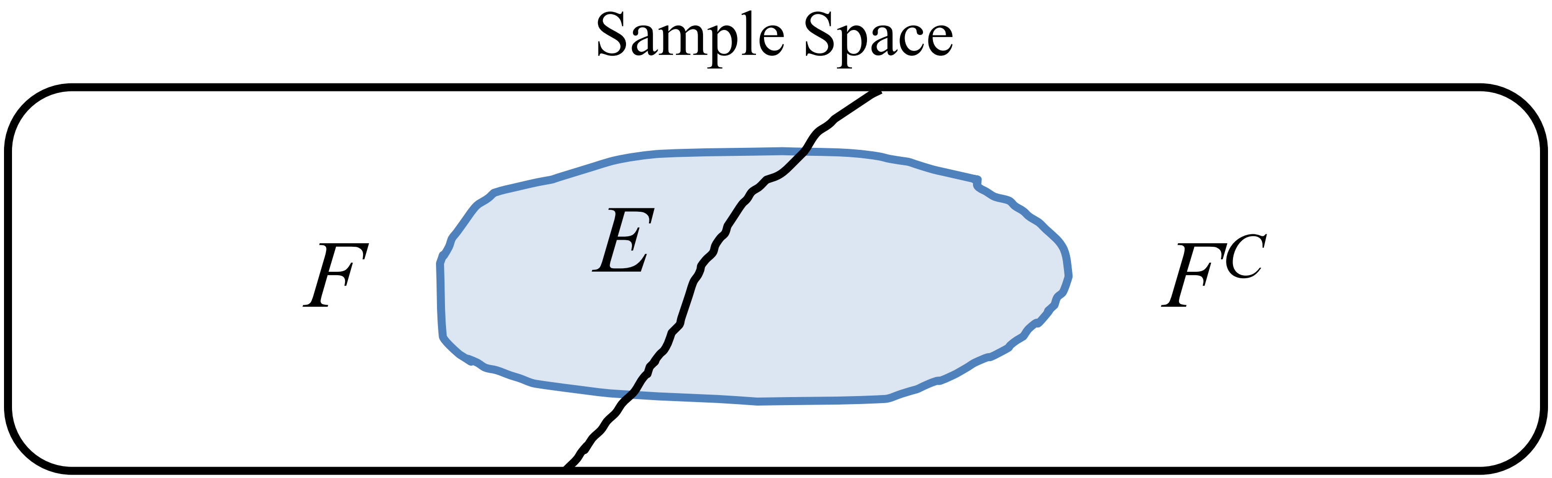
$$E = EF \cup EF^c$$



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

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

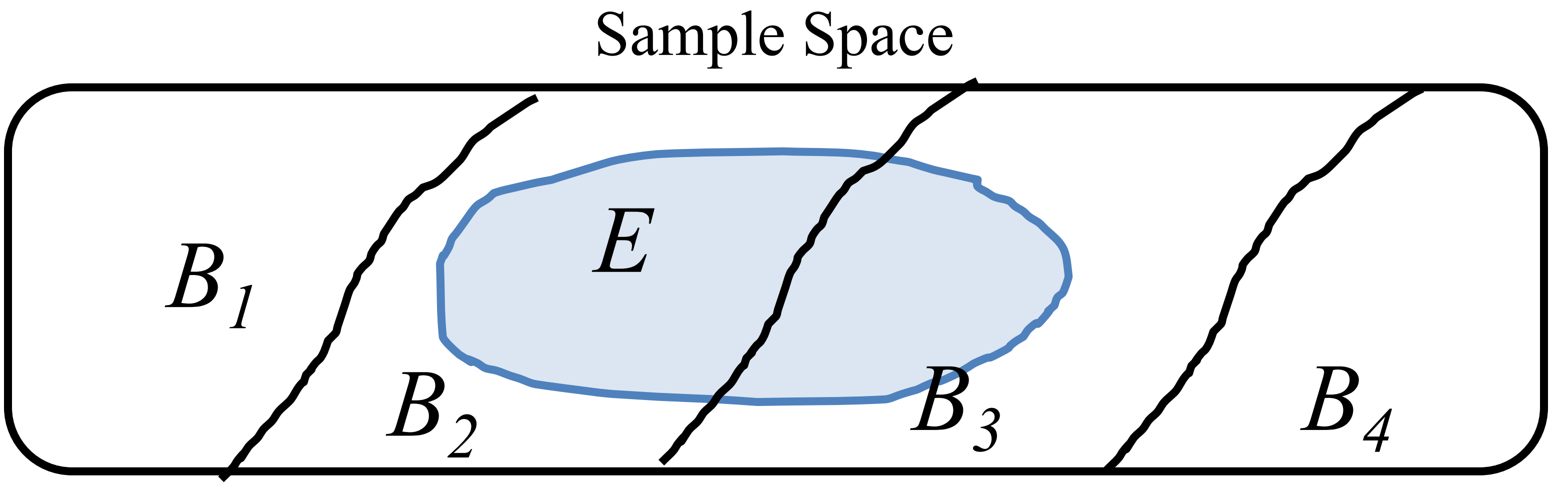
Law of Total Probability



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

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

Law of Total Probability



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

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

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

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 $P(E \mid F)$

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

A little while later…

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

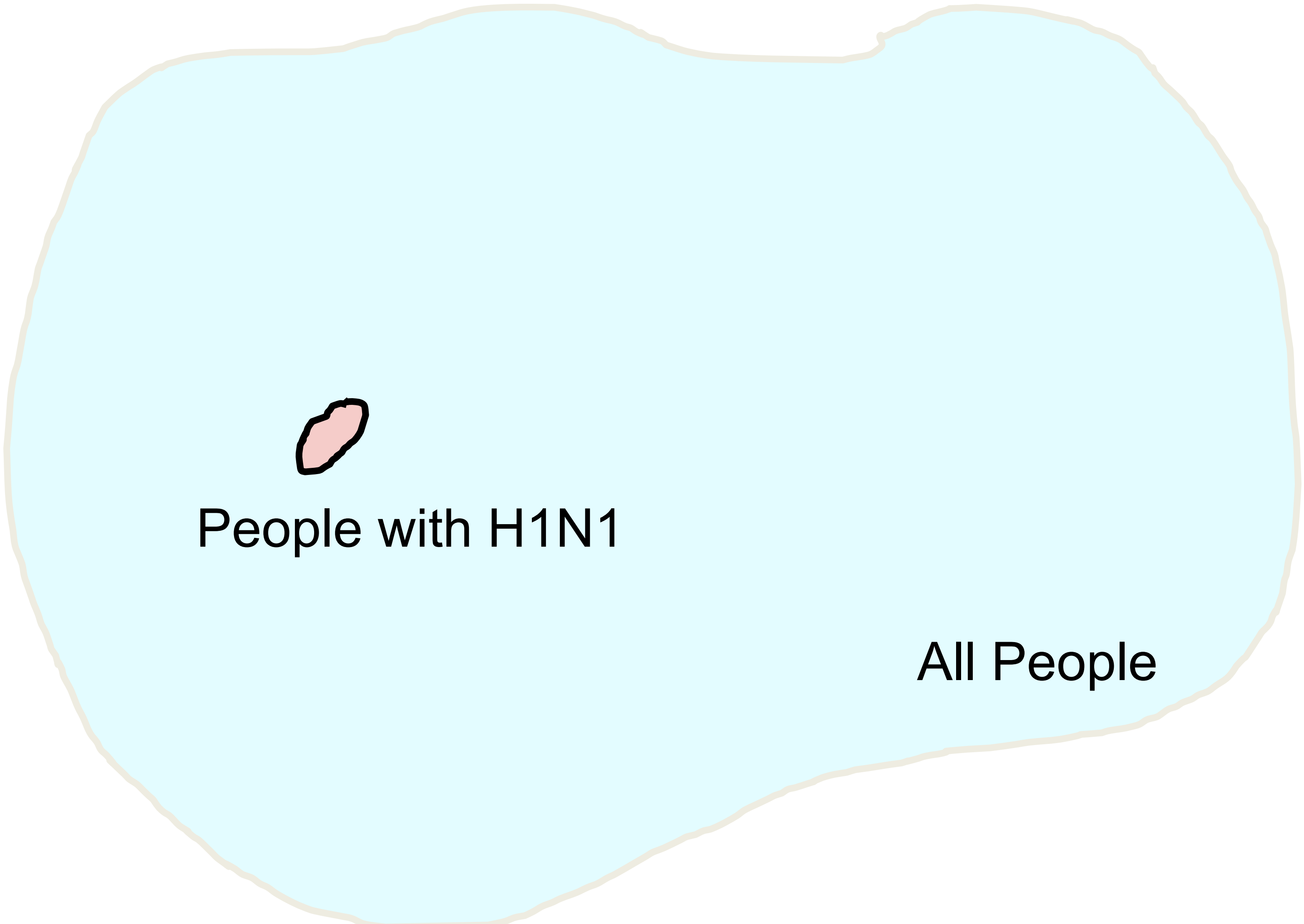
$$P(F \mid E) = \frac{P(E \mid F)\, P(F)}{P(E \mid F)\, P(F) + P(E \mid F^c)\, P(F^c)}$$

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

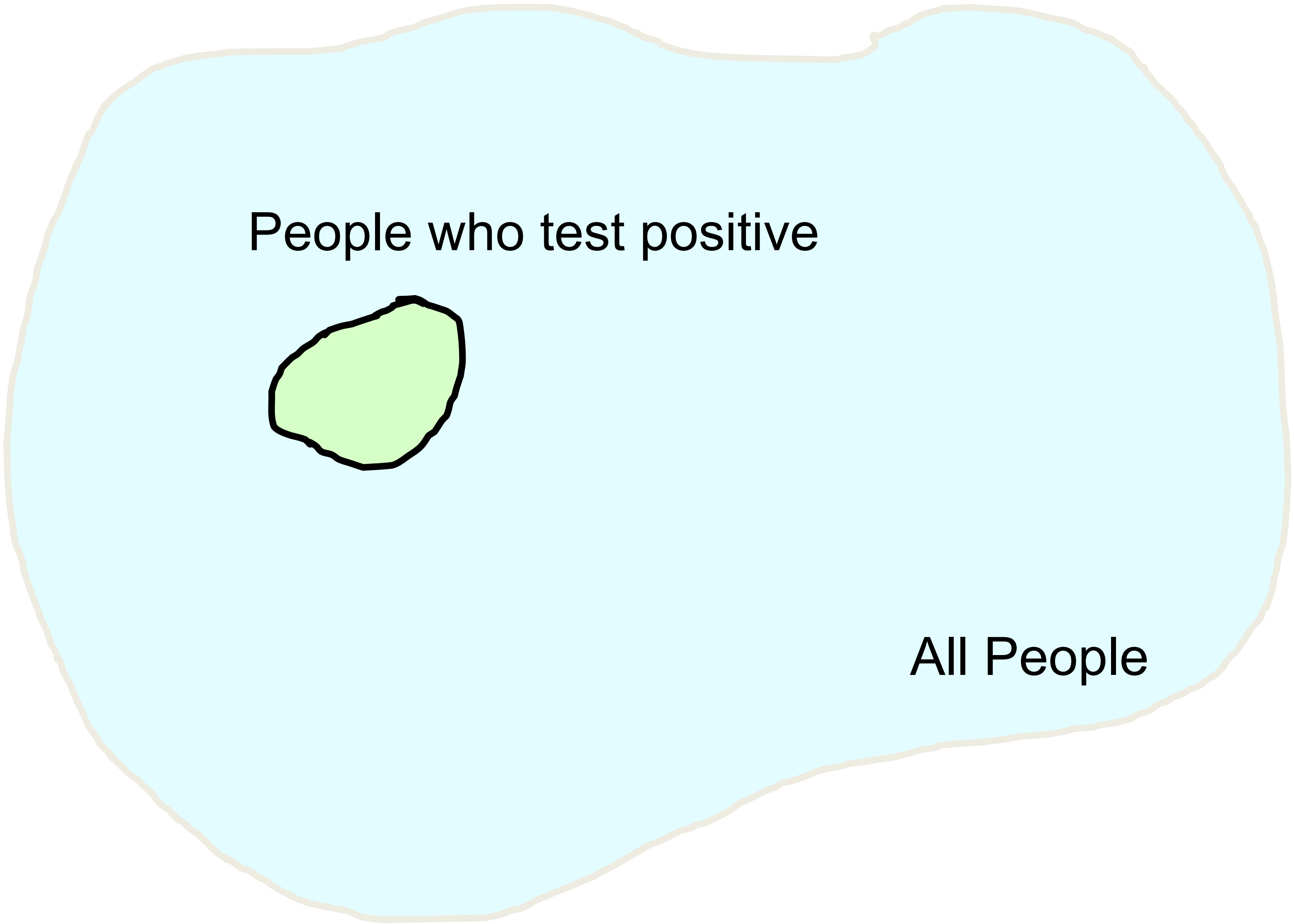
Intuition Time

Bayes Theorem Intuition

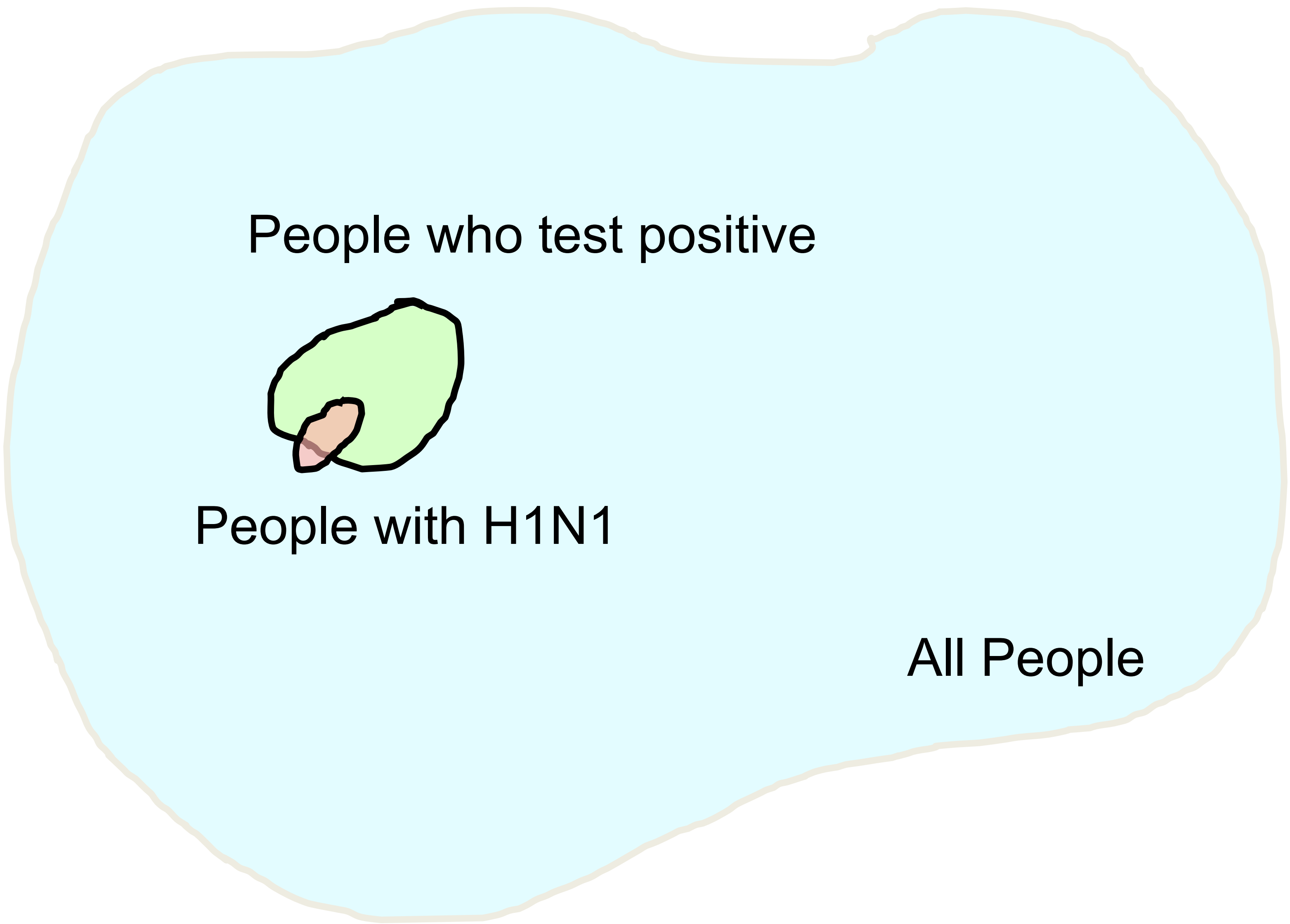
Bayes Theorem Intuition



Bayes Theorem Intuition

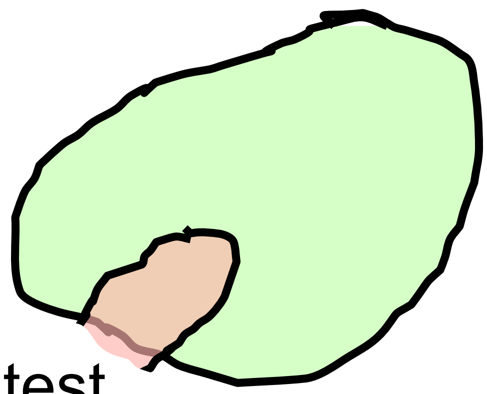


Bayes Theorem Intuition



Bayes Theorem Intuition

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



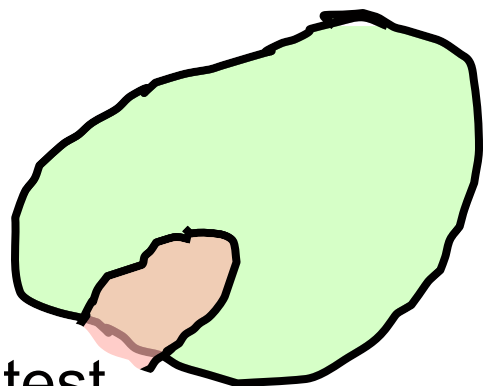
People who test positive

People who test positive and have H1N1

≈ 0.330

Bayes Theorem Intuition

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



People who test positive

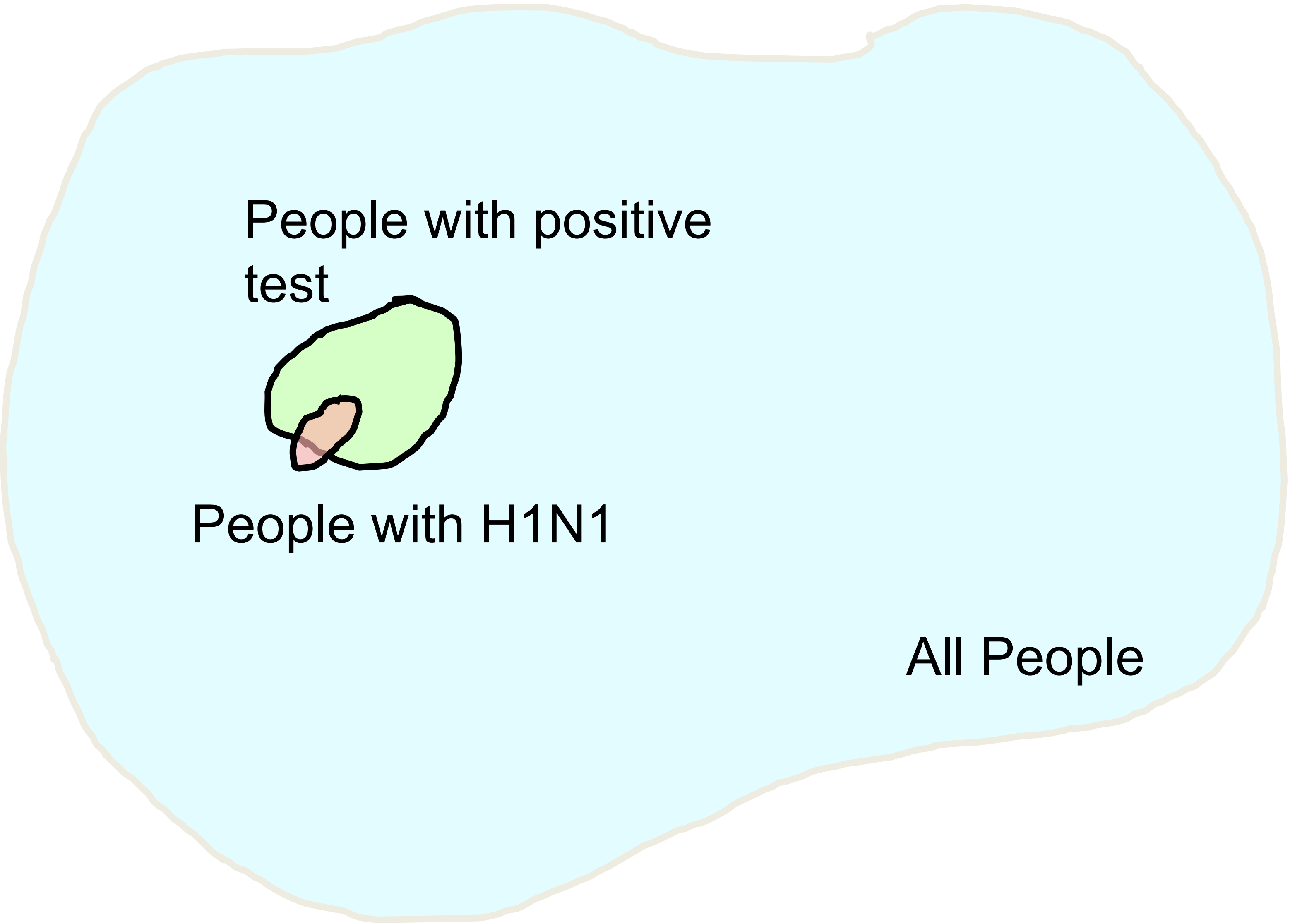
$$P(F)P(E|F) + P(F^c)P(E|F^c)$$

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

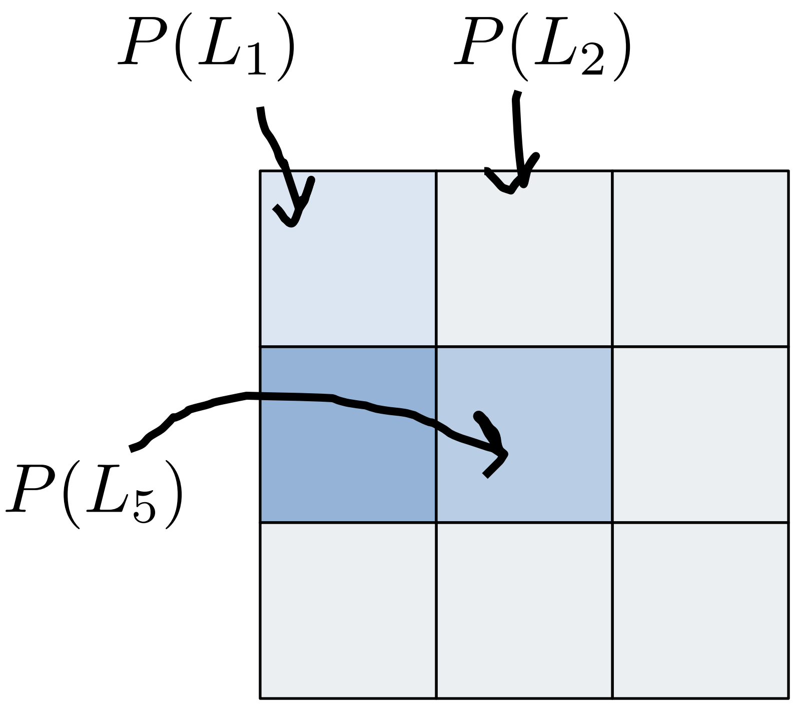
Simple Spam Detection

- 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:

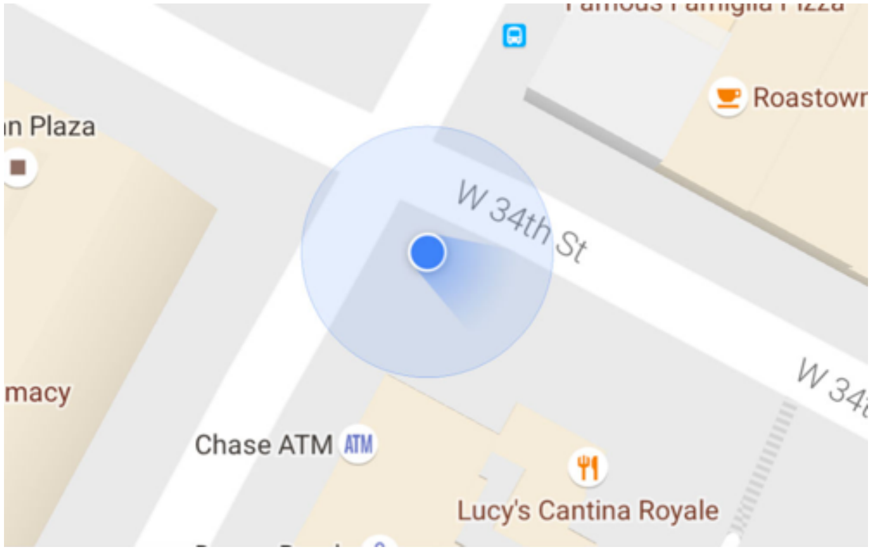
$$P(F \mid E) = \frac{P(E \mid F)\, \mathrm{P}(F)}{P(E \mid F)\, \mathrm{P}(F) + \mathrm{P}(E \mid F^c)\, P(F^c)}$$

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

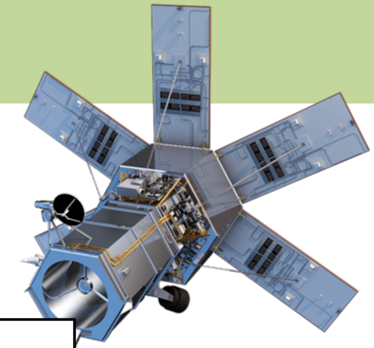
Update Belief



Before Observation

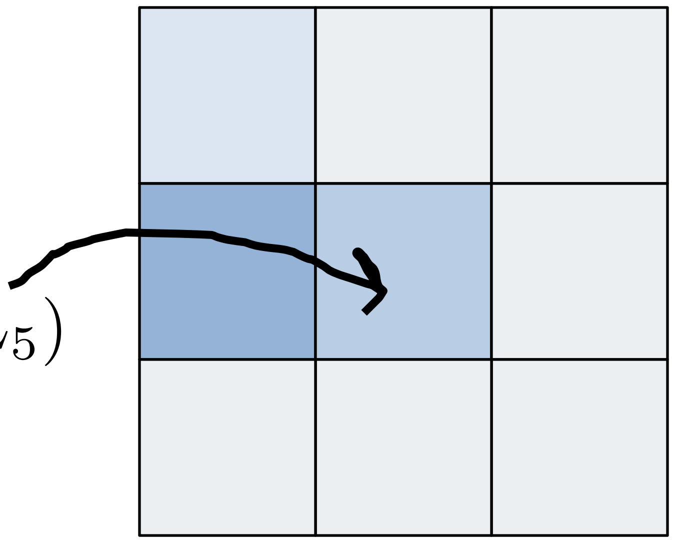


Update Belief

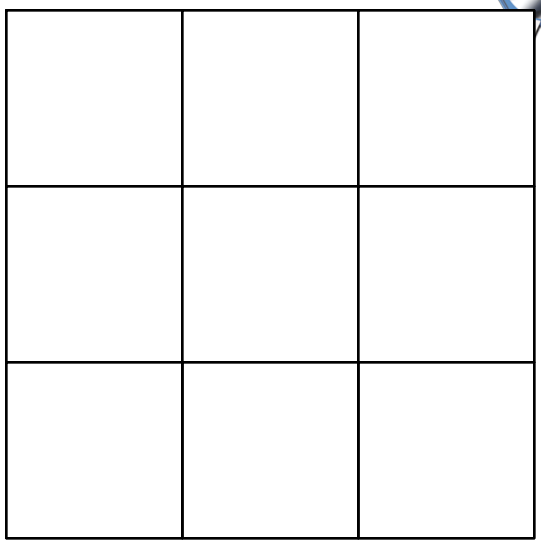


Know: $P(O|L_i)$

$P(L_5)$

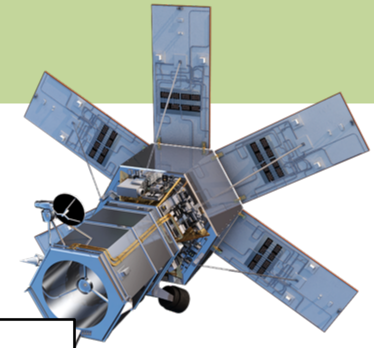


Before Observation

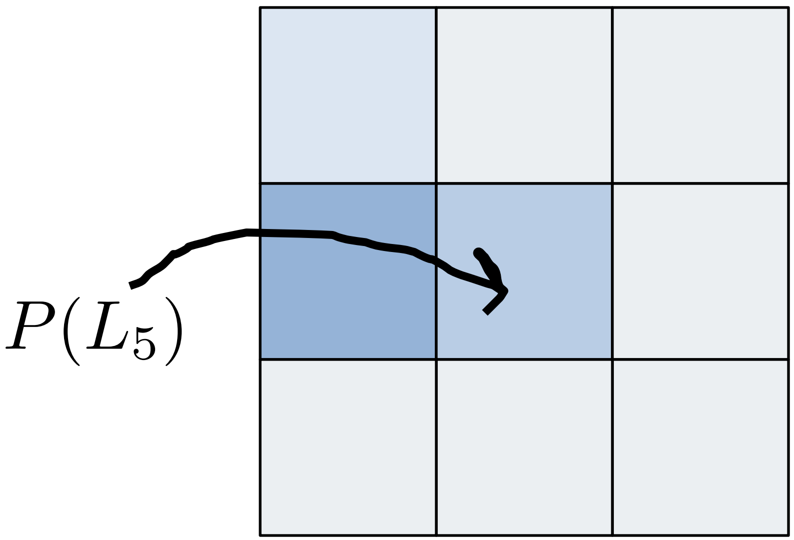


After Observation

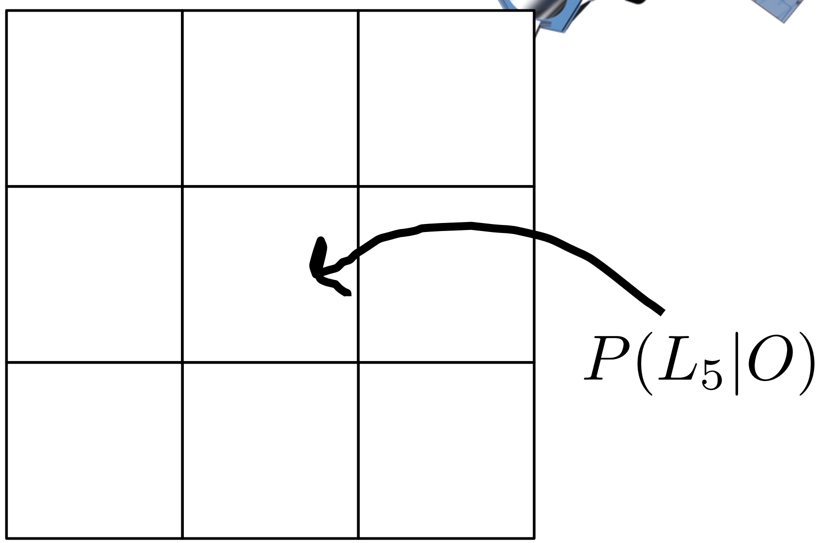
Update Belief



Know: $P(O|L_i)$



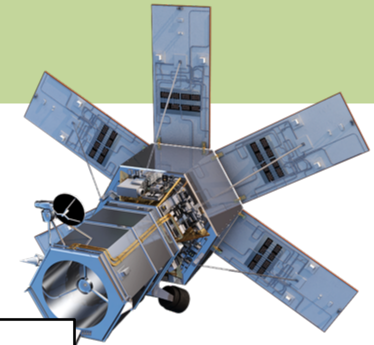
Before Observation

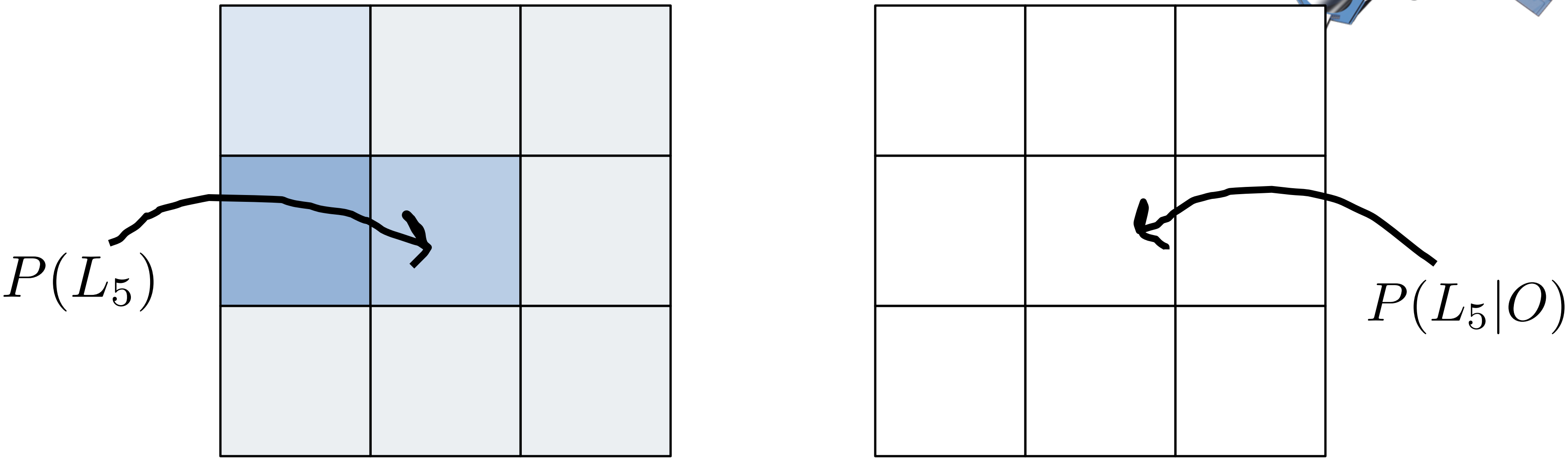


After Observation

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

Update Belief



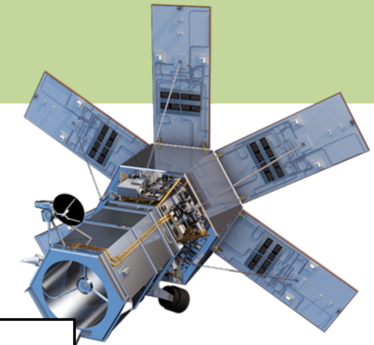


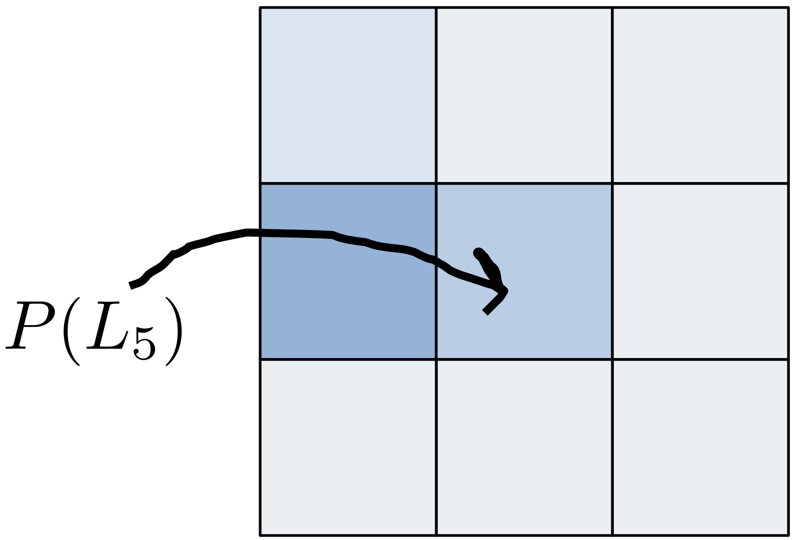
Before Observation

After Observation

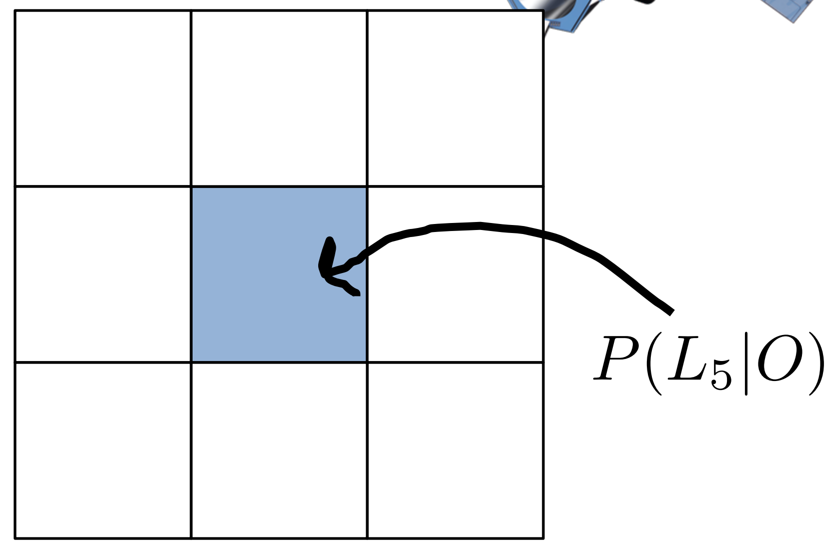
$$P(L_5|O) = \frac{P(O|L_5)P(L_5)}{\sum_i P(O|L_i)P(L_i)}$$

Update Belief





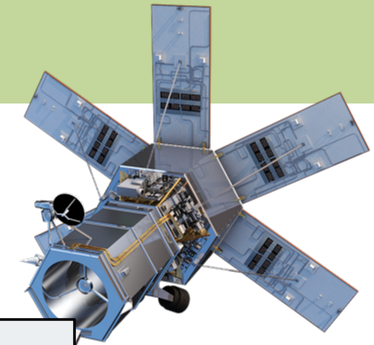
Before Observation

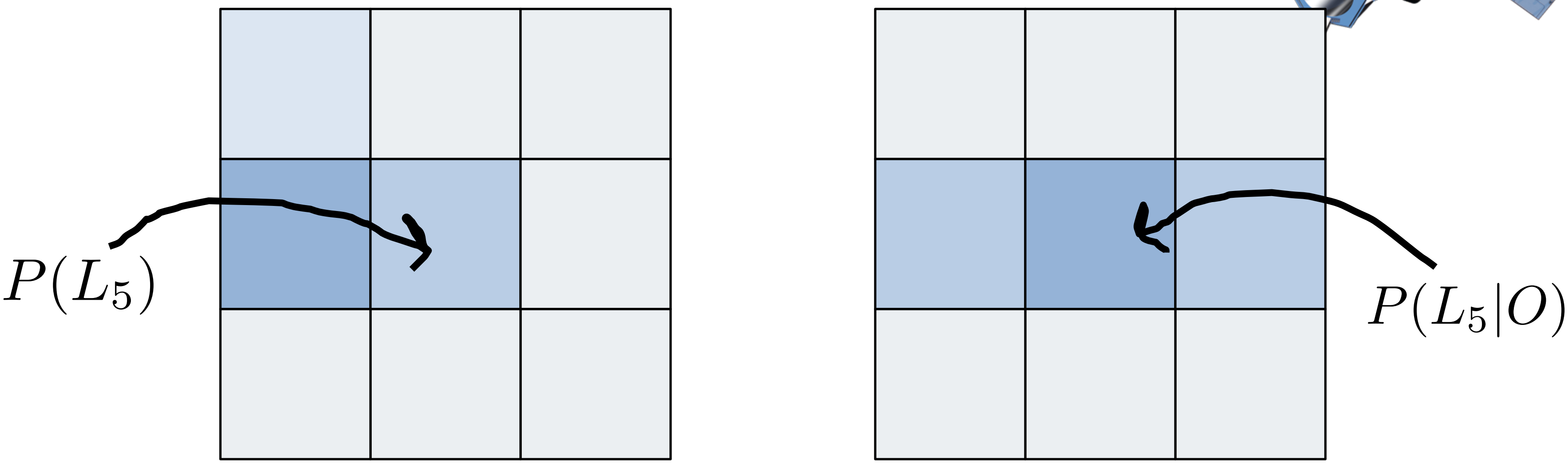


After Observation

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

Update Belief





Before Observation

After Observation

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

Stretch!

Bayes + Condition

Say we want $P(E|FG)$ ...

1. Combining F, G: $P(E|FG) = \dfrac{P(FG|E)P(E)}{P(FG)}$

2. World of G: $P(E|FG) = \dfrac{P(F|EG)P(E|G)}{P(F|G)}$

3. World of F: $P(E|FG) = \dfrac{P(G|EF)P(E|F)}{P(G|F)}$

Which Rule When?

Chain Rule: $P(EF) = P(E|F)P(F)$

- Goes from an "and" to a conditional and vice versa
- Think about which event you want to condition on

LotP: $P(E) = P(E|F)P(F) + \mathrm{P}(\mathrm{E}|F^C)\mathrm{P}(F^C)$

- We don't know about E but we do know about E|F
- Don't forget about the "and" version and "summation" version

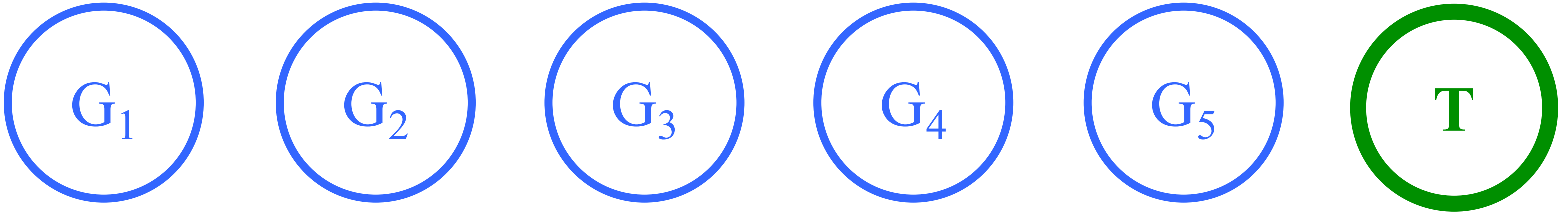
Bayes: $P(E|F) = \dfrac{P(F|E)P(E)}{P(F)}$

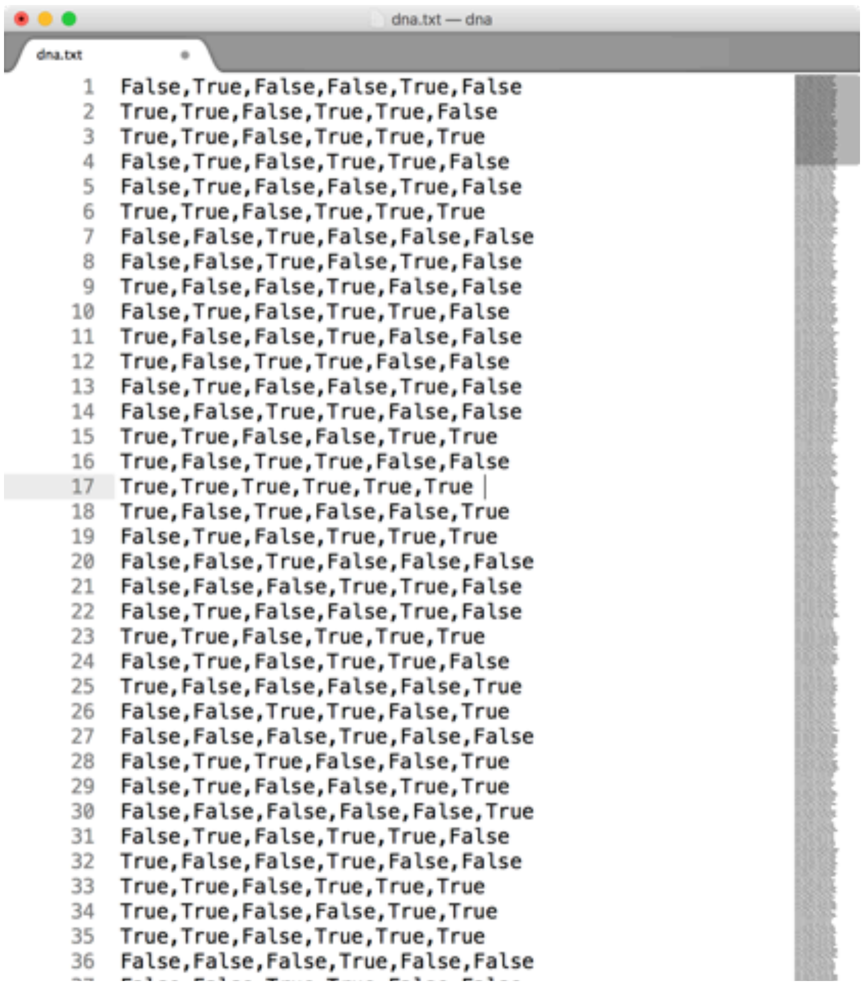
- Good for when E|F is hard but F|E is not so hard
- Common mistake: not trying chain rule first

Independence

Start with a cool program


G_1
G_2
G_3
G_4
G_5
T



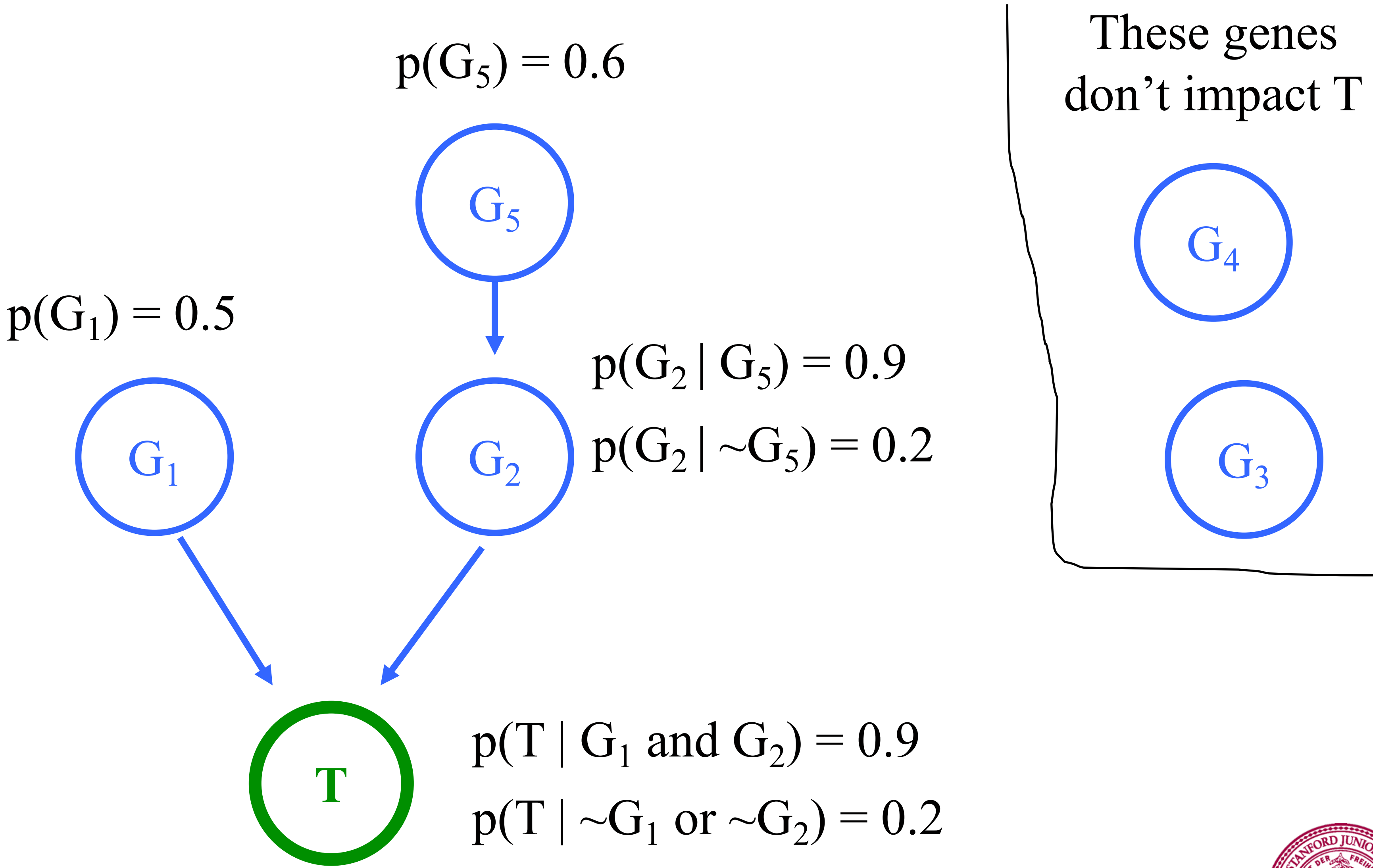



```
False,True,False,False,True,False
True,True,False,True,True,False
True,True,False,True,True,True
False,True,False,True,True,False
False,True,False,False,True,False
True,True,False,True,True,True
False,False,True,False,False,False
False,False,True,False,True,False
True,False,False,True,False,False
False,True,False,True,True,False
True,False,False,True,False,False
True,False,True,True,False,False
False,True,False,False,True,False
False,False,True,True,False,False
True,True,False,False,True,True
True,False,True,True,False,False
True,True,True,True,True,True
True,False,True,False,False,True
False,True,False,True,True,True
False,False,True,False,False,False
False,False,False,True,True,False
False,True,False,False,True,False
True,True,False,True,True,True
False,True,False,True,True,False
True,False,False,False,False,True
False,False,True,True,False,True
False,False,False,True,False,False
False,True,True,False,False,True
False,True,False,False,True,True
False,False,False,False,False,True
False,True,False,True,True,False
True,False,False,True,False,False
True,True,False,True,True,True
True,True,False,False,True,True
True,True,False,True,True,True
False,False,False,True,False,False
```

100,000 samples

6 observations per sample

Discovered Pattern



We’ve gotten ahead of ourselves

Source: The Ho

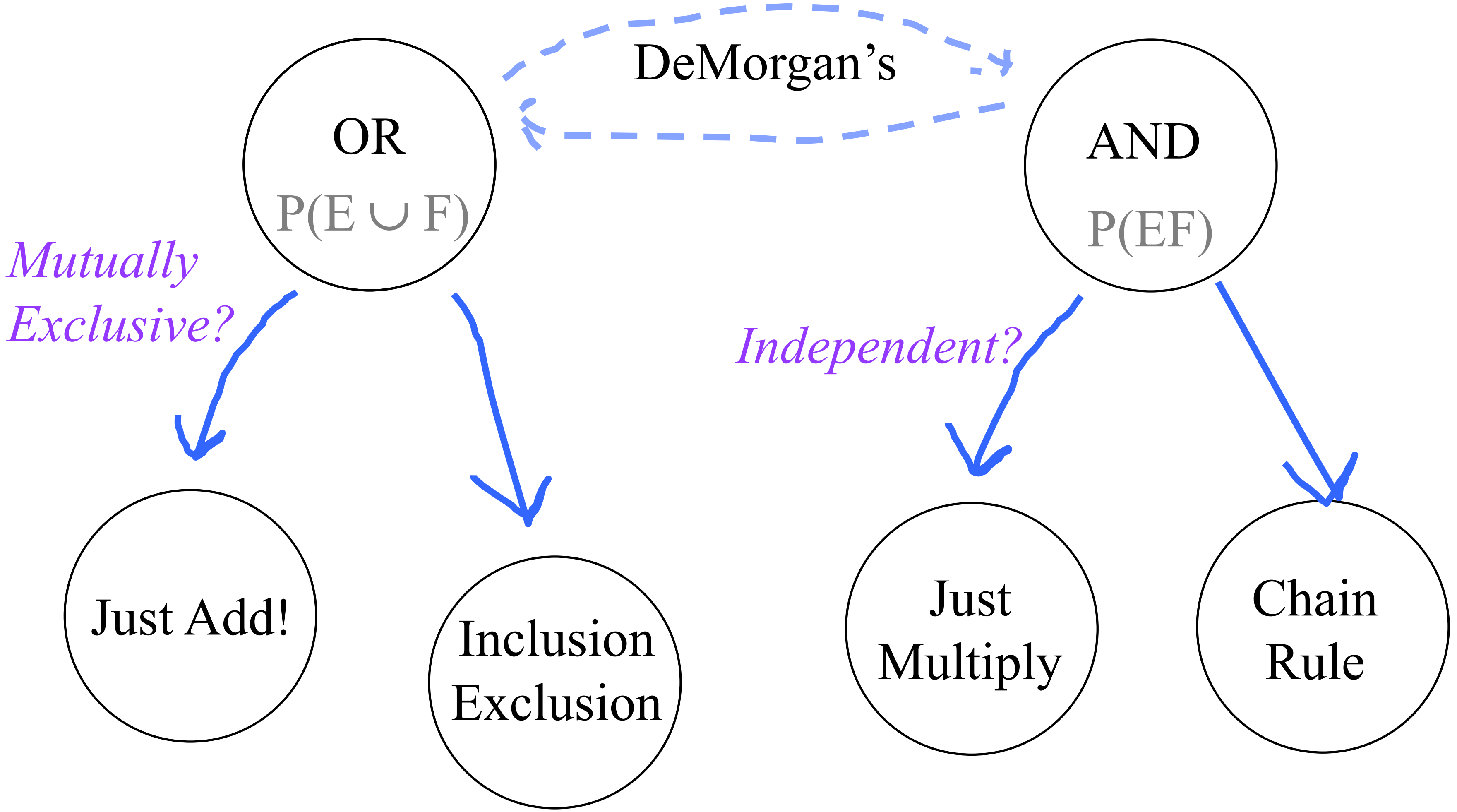
Start at the beginning

And vs Condition

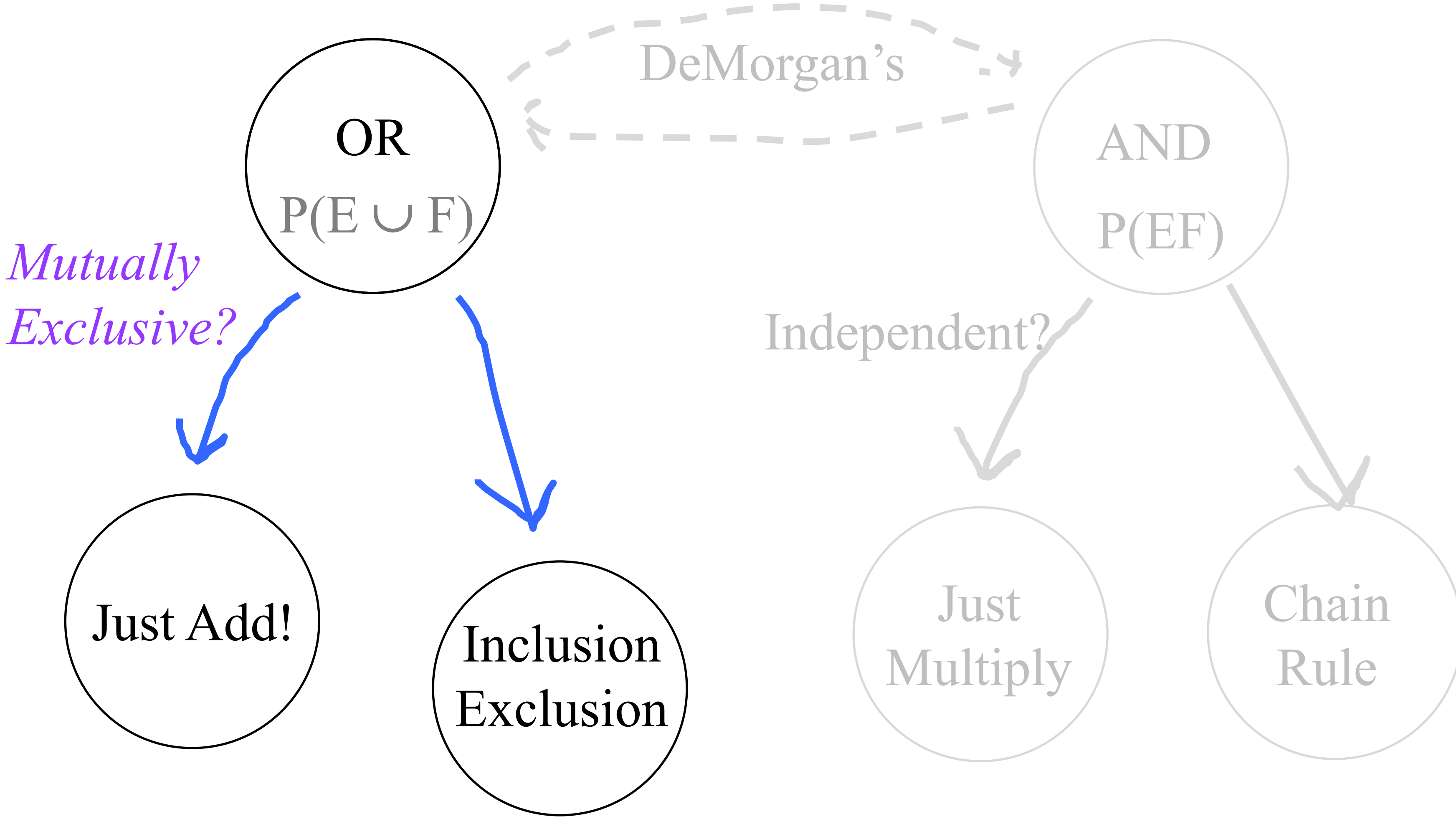
$$P(AB) \text{ vs } P(A|B)$$

$$P(AB) = P(A|B)P(B)$$

Overview


DeMorgan's
OR
P(E ∪ F)
AND
P(EF)
Mutually Exclusive?
Independent?
Just Add!
Inclusion Exclusion
Just Multiply
Chain Rule

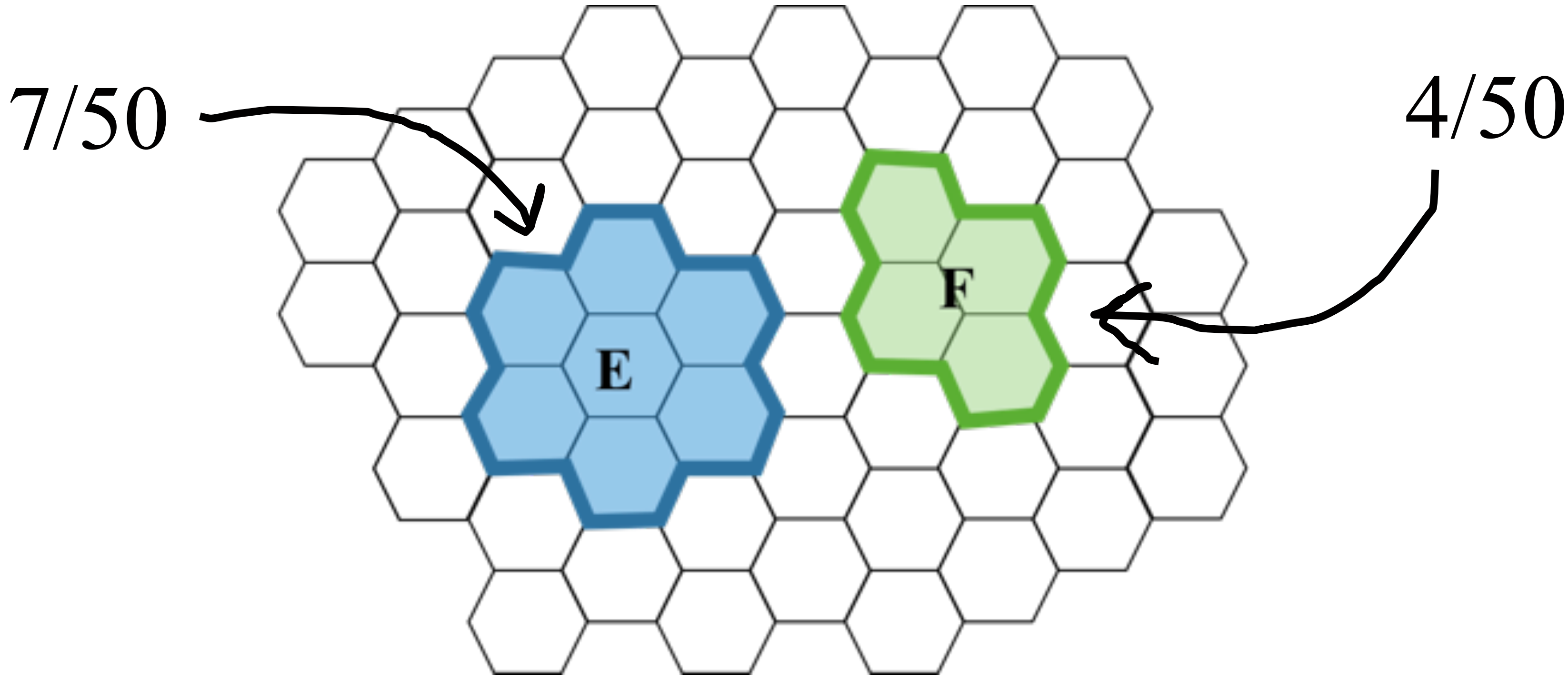
Today



OR
P(E ∪ F)
Mutually
Exclusive?
Just Add!
Inclusion
Exclusion
DeMorgan's
AND
P(EF)
Independent?
Just
Multiply
Chain
Rule

Probability of “OR”

OR with Mutually Exclusive Events

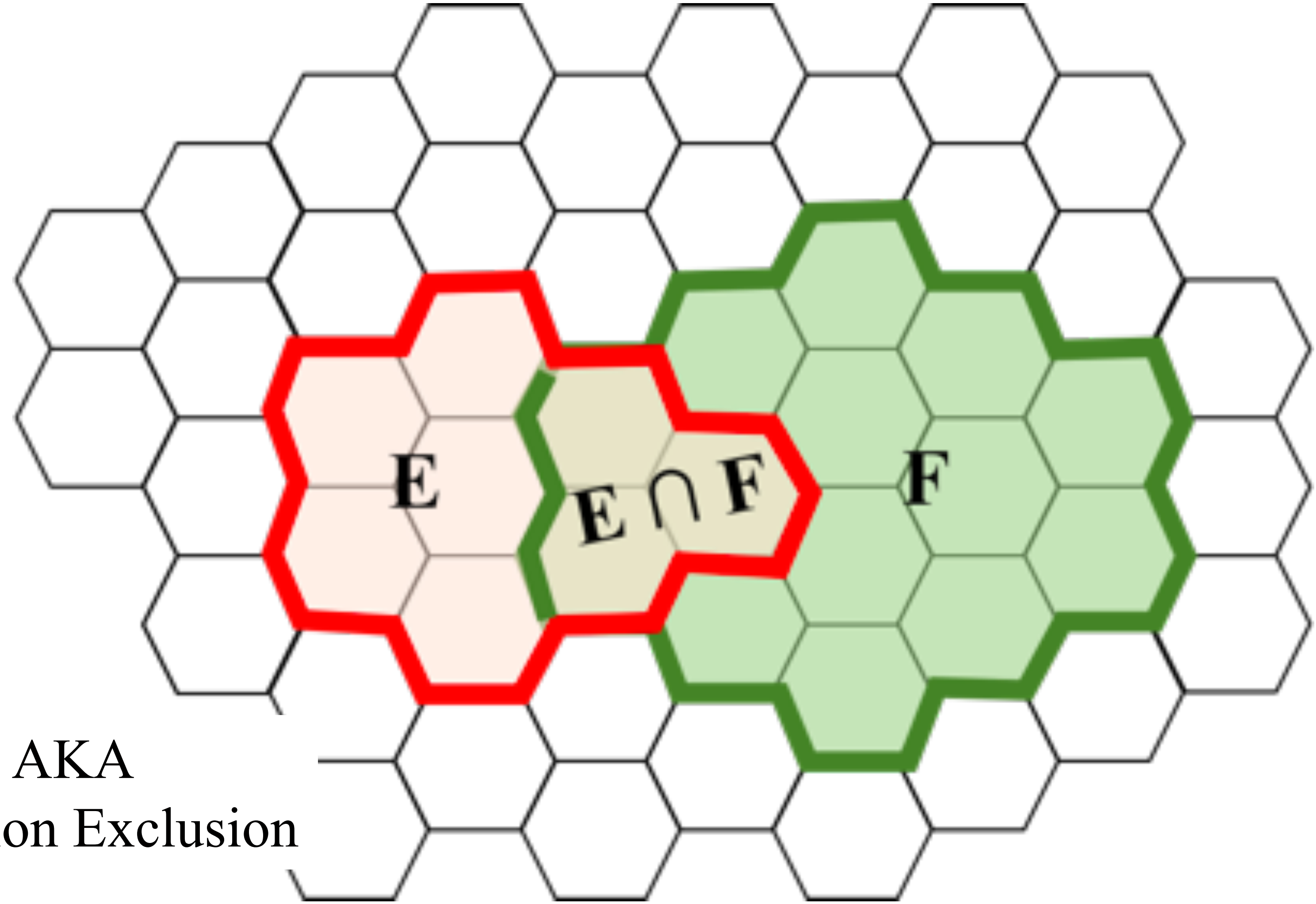


If events are mutually exclusive, probability of OR is simple:

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

What about when they are not *Mutually exclusive*?

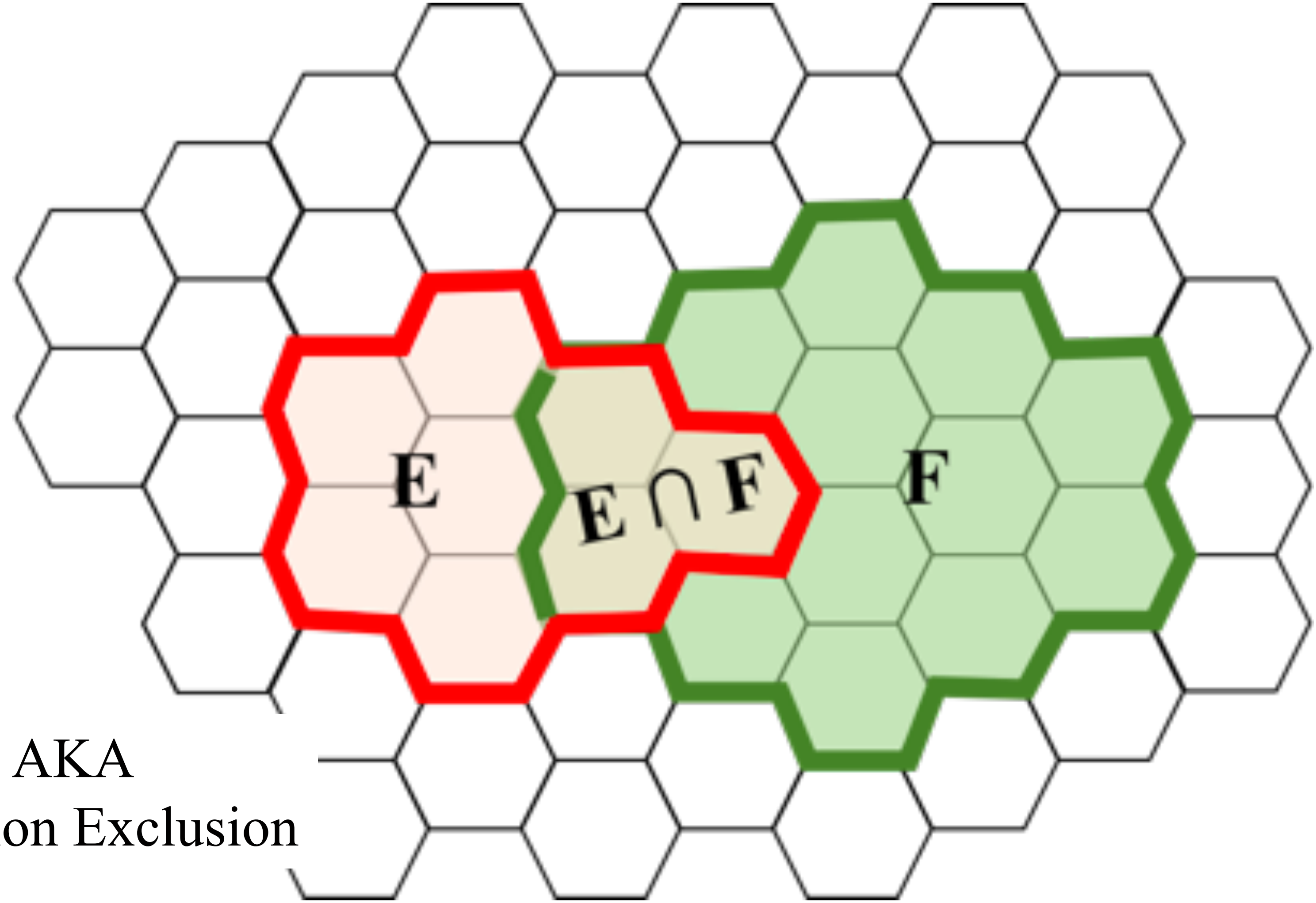
OR without Mutually Exclusivity



AKA
Inclusion Exclusion

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

OR without Mutually Exclusivity



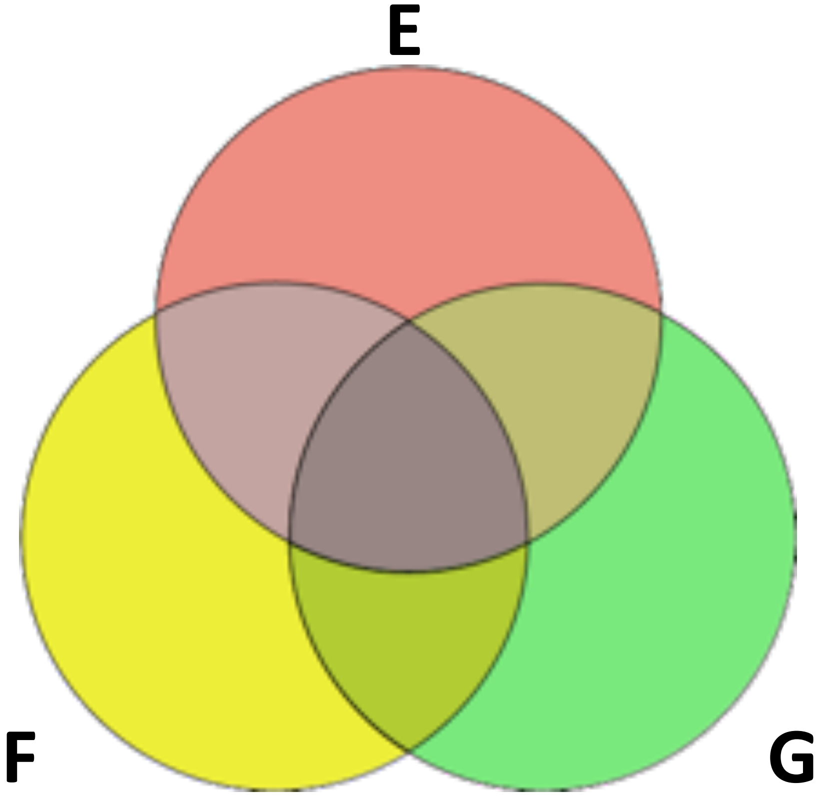
AKA
Inclusion Exclusion

$$P(E \cup F) = \frac{8}{50} + \frac{14}{50} - \frac{3}{50} = \frac{19}{50}$$

More than two sets?

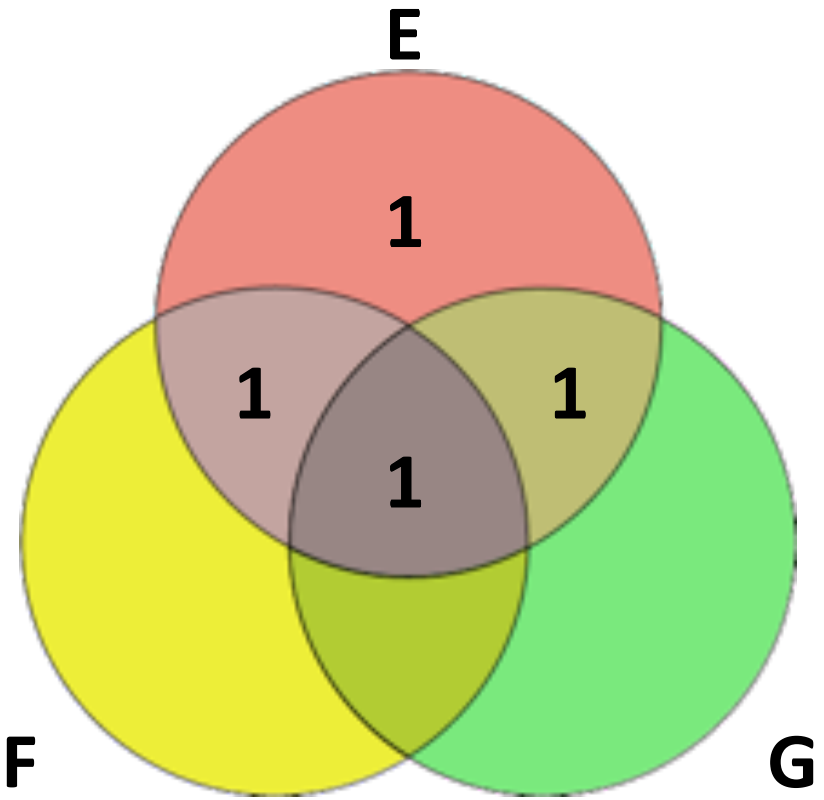
Inclusion Exclusion with Three Sets

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



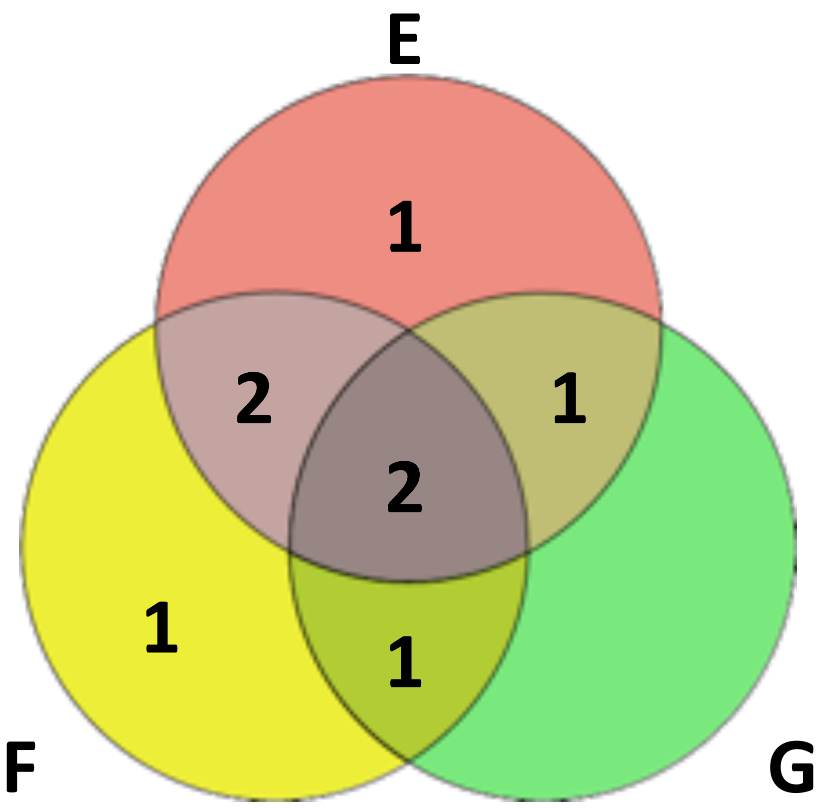
Inclusion Exclusion with Three Sets

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



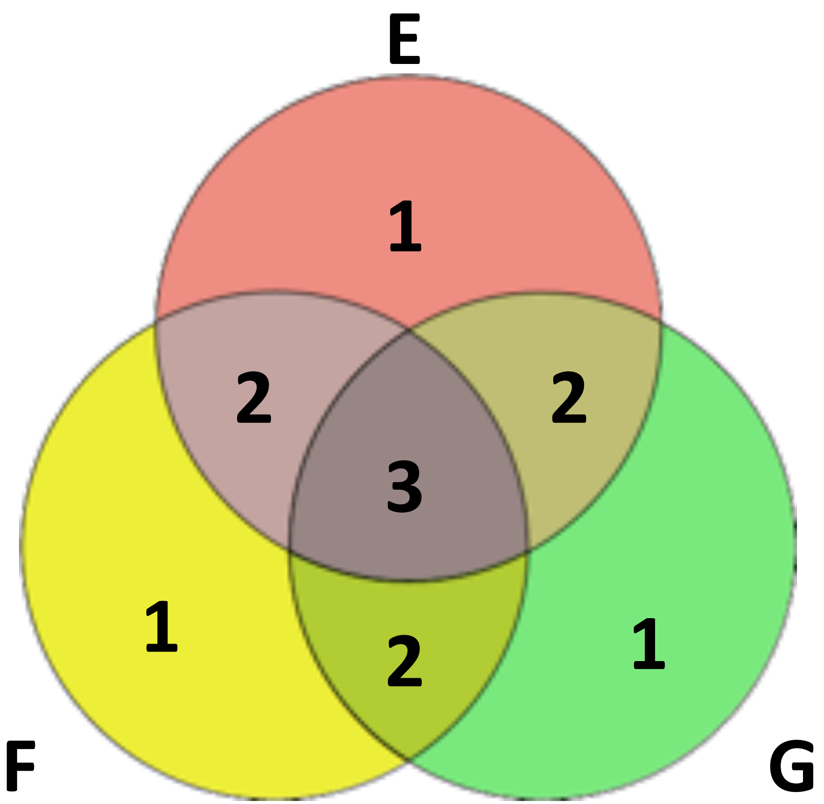
Inclusion Exclusion with Three Sets

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



Inclusion Exclusion with Three Sets

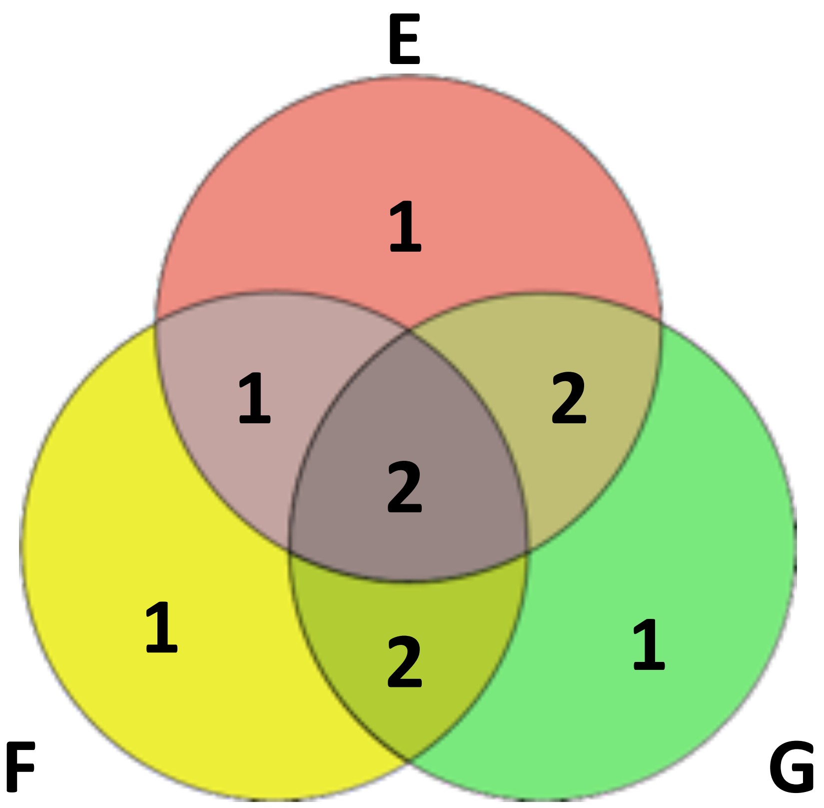
$$P(E \cup F \cup G) = P(E) + P(F) + P(G)$$



Inclusion Exclusion with Three Sets

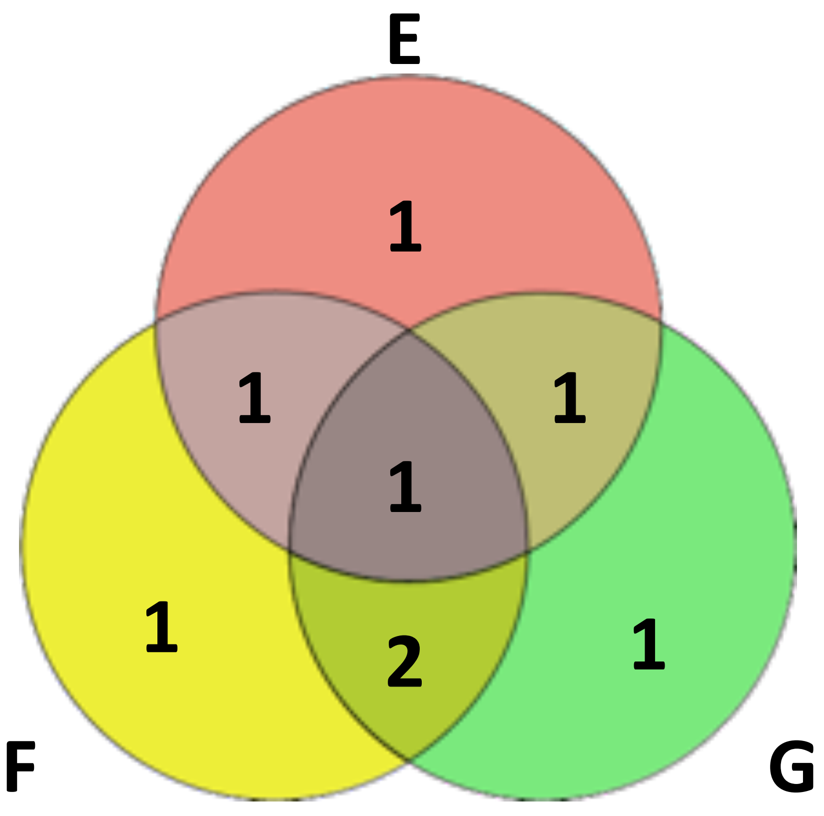
$$P(E \cup F \cup G) = P(E) + P(F) + P(G)$$

$$-P(EF)$$



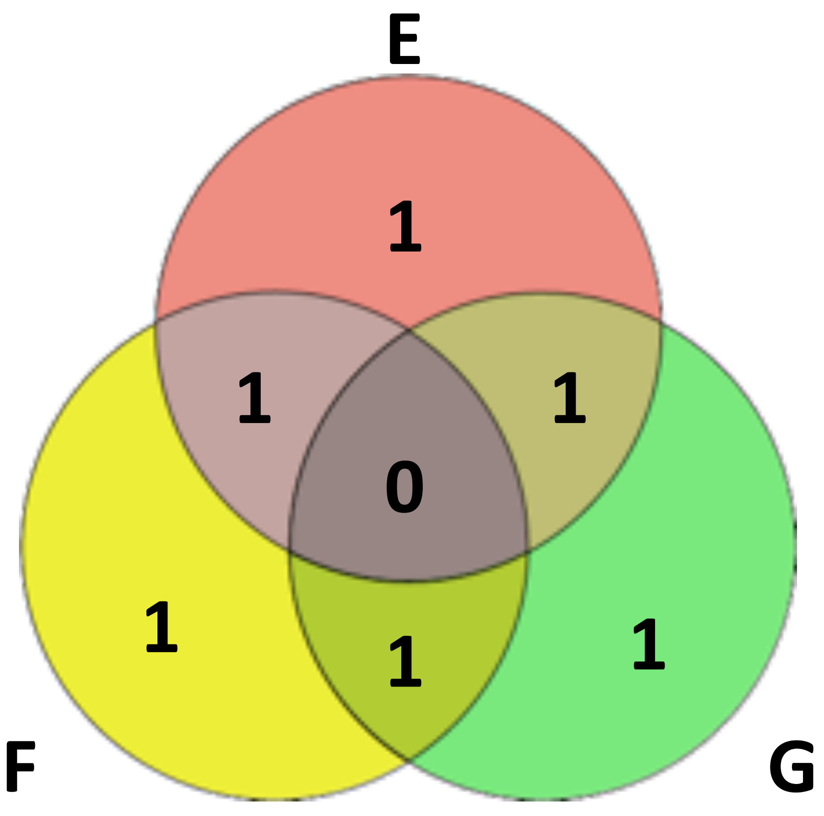
Inclusion Exclusion with Three Sets

$$P(E \cup F \cup G) = P(E) + P(F) + P(G)$$
$$-P(EF) - P(EG)$$



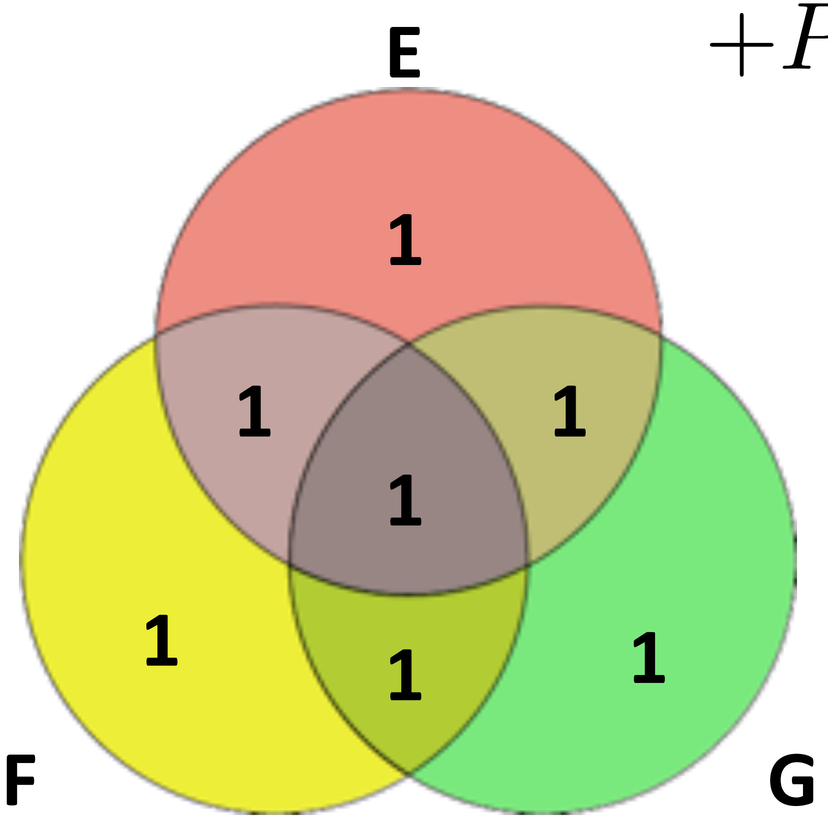
Inclusion Exclusion with Three Sets

$$P(E \cup F \cup G) = P(E) + P(F) + P(G)$$
$$-P(EF) - P(EG) - P(FG)$$



Inclusion Exclusion with Three Sets

$$P(E \cup F \cup G) = P(E) + P(F) + P(G)$$
$$-P(EF) - P(EG) - P(FG)$$
$$+P(EFG)$$



General Inclusion Exclusion

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{r=1}^{n} (-1)^{r+1} Y_r$$

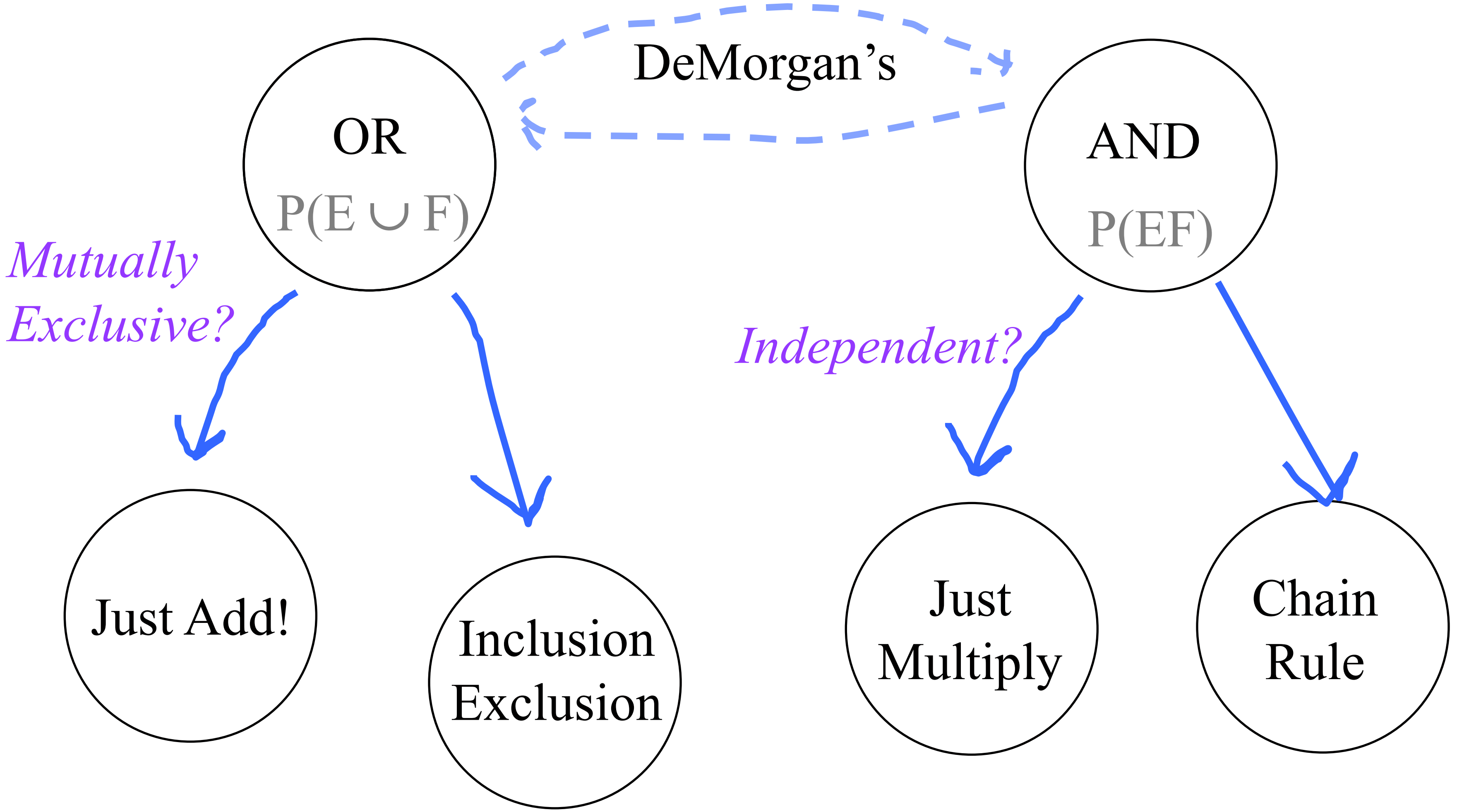
Y_1 = Sum of all events on their own $\quad \sum_i P(E_i)$

Y_2 = Sum of all pairs of events $\quad \sum_{i,j} P(E_i \cap E_j) \quad \text{s.t.} i \neq j$

Y_3 = Sum of all triples of events $\quad \sum_{i,j,k} P(E_i \cap E_j \cap E_k) \quad \text{s.t.} i \neq j, j \neq k, i \neq k$

* Where Y_r is the sum, for all combinations of r events, of the probability of the union those events.

Overview


DeMorgan's
OR
P(E ∪ F)
AND
P(EF)
Mutually Exclusive?
Independent?
Just Add!
Inclusion Exclusion
Just Multiply
Chain Rule