Will Monroe
July 3, 2017

with materials by
Mehran Sahami
and Chris Piech

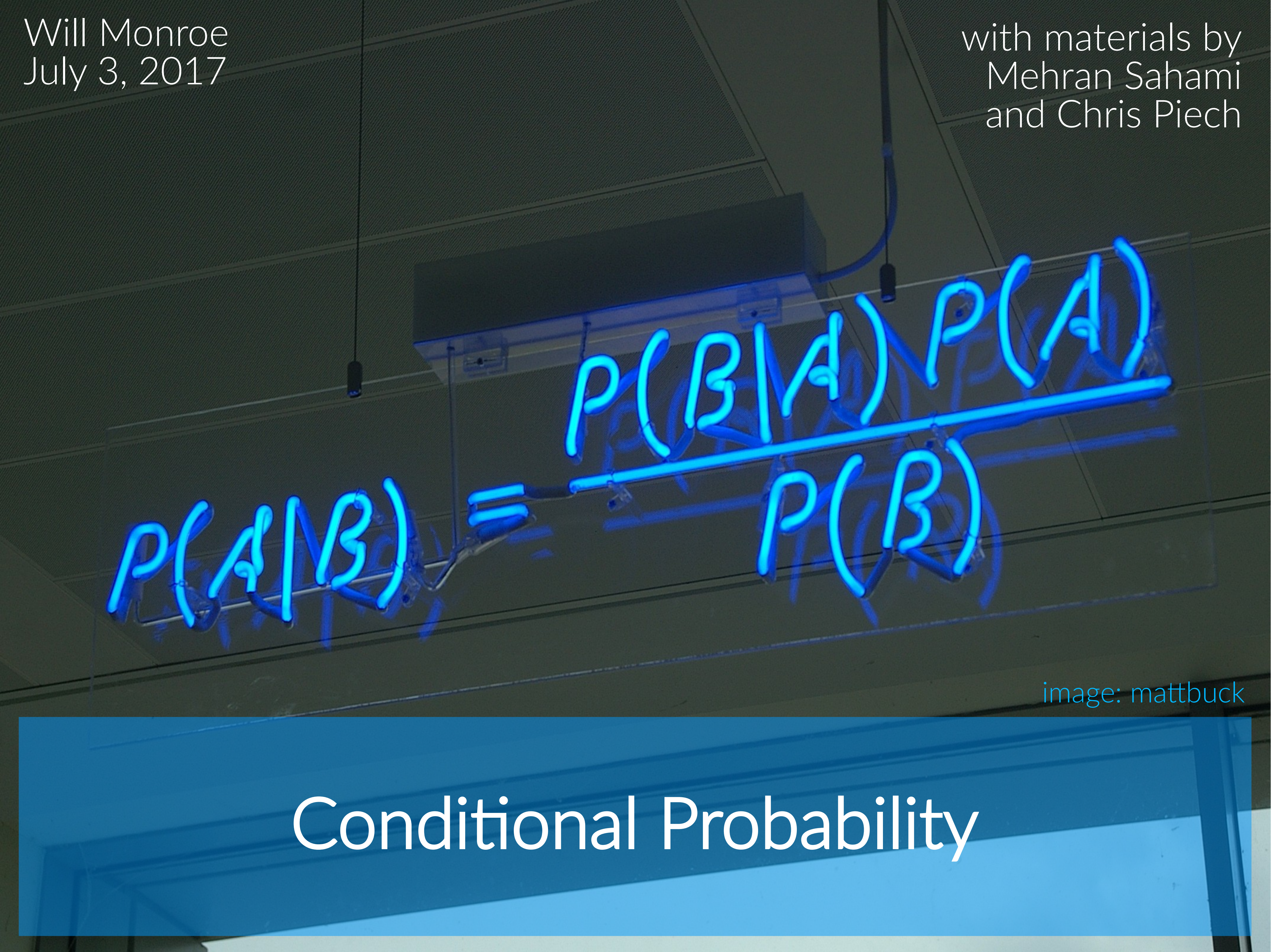
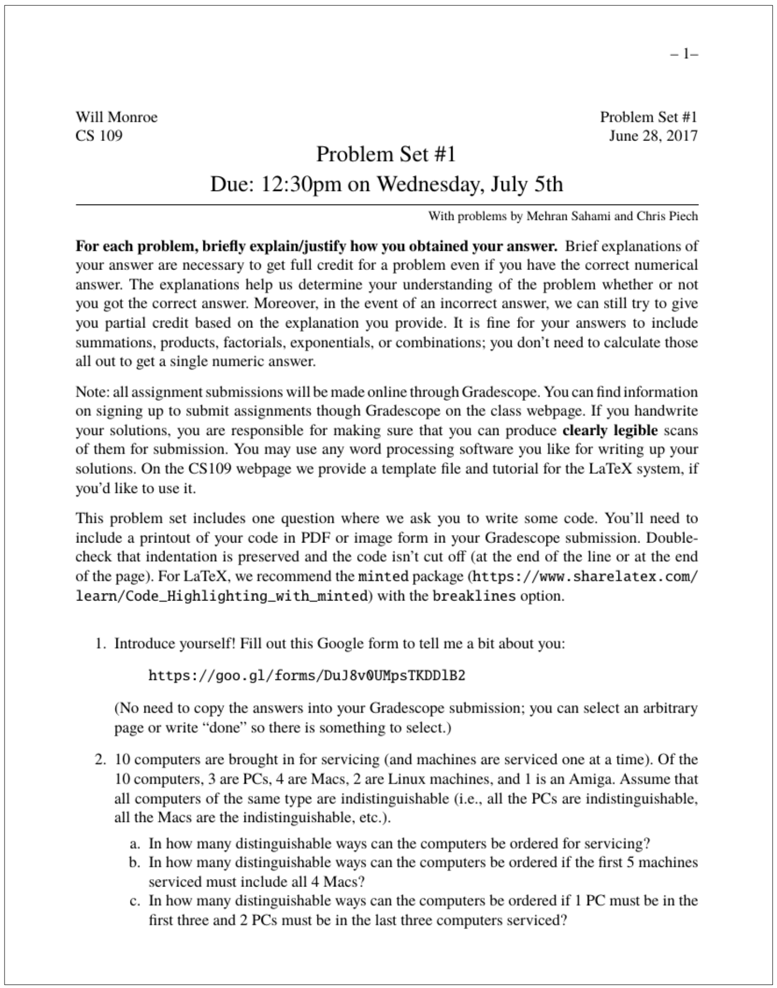


image: mattbuck

Conditional Probability

Announcements: Problem Set #1



– 1–

Will Monroe
CS 109

Problem Set #1
June 28, 2017

Problem Set #1
Due: 12:30pm on Wednesday, July 5th

With problems by Mehran Sahami and Chris Piech

For each problem, briefly explain/justify how you obtained your answer. Brief explanations of your answer are necessary to get full credit for a problem even if you have the correct numerical answer. The explanations help us determine your understanding of the problem whether or not you got the correct answer. Moreover, in the event of an incorrect answer, we can still try to give you partial credit based on the explanation you provide. It is fine for your answers to include summations, products, factorials, exponentials, or combinations; you don't need to calculate those all out to get a single numeric answer.

Note: all assignment submissions will be made online through Gradescope. You can find information on signing up to submit assignments though Gradescope on the class webpage. If you handwrite your solutions, you are responsible for making sure that you can produce **clearly legible** scans of them for submission. You may use any word processing software you like for writing up your solutions. On the CS109 webpage we provide a template file and tutorial for the LaTeX system, if you'd like to use it.

This problem set includes one question where we ask you to write some code. You'll need to include a printout of your code in PDF or image form in your Gradescope submission. Double-check that indentation is preserved and the code isn't cut off (at the end of the line or at the end of the page). For LaTeX, we recommend the `minted` package (`https://www.sharelatex.com/learn/Code_Highlighting_with_minted`) with the `breaklines` option.

1. Introduce yourself! Fill out this Google form to tell me a bit about you:

   `https://goo.gl/forms/DuJ8v0UMpsTKDDlB2`

   (No need to copy the answers into your Gradescope submission; you can select an arbitrary page or write "done" so there is something to select.)

2. 10 computers are brought in for servicing (and machines are serviced one at a time). Of the 10 computers, 3 are PCs, 4 are Macs, 2 are Linux machines, and 1 is an Amiga. Assume that all computers of the same type are indistinguishable (i.e., all the PCs are indistinguishable, all the Macs are the indistinguishable, etc.).
   a. In how many distinguishable ways can the computers be ordered for servicing?
   b. In how many distinguishable ways can the computers be ordered if the first 5 machines serviced must include all 4 Macs?
   c. In how many distinguishable ways can the computers be ordered if 1 PC must be in the first three and 2 PCs must be in the last three computers serviced?

Due **this Wednesday**!

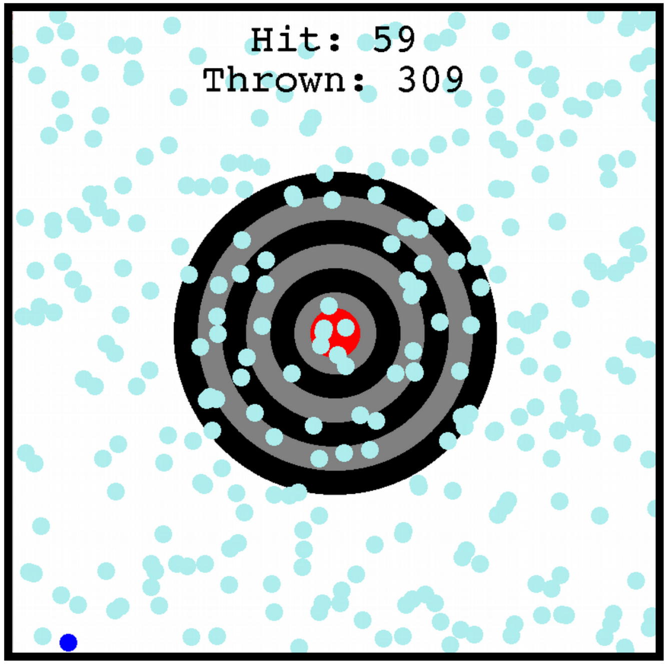
4.c is particularly challenging.

Announcements: Python!

Handout on website

Tutorial: Wed. 7/5, 2:30pm

Review: What is a probability?



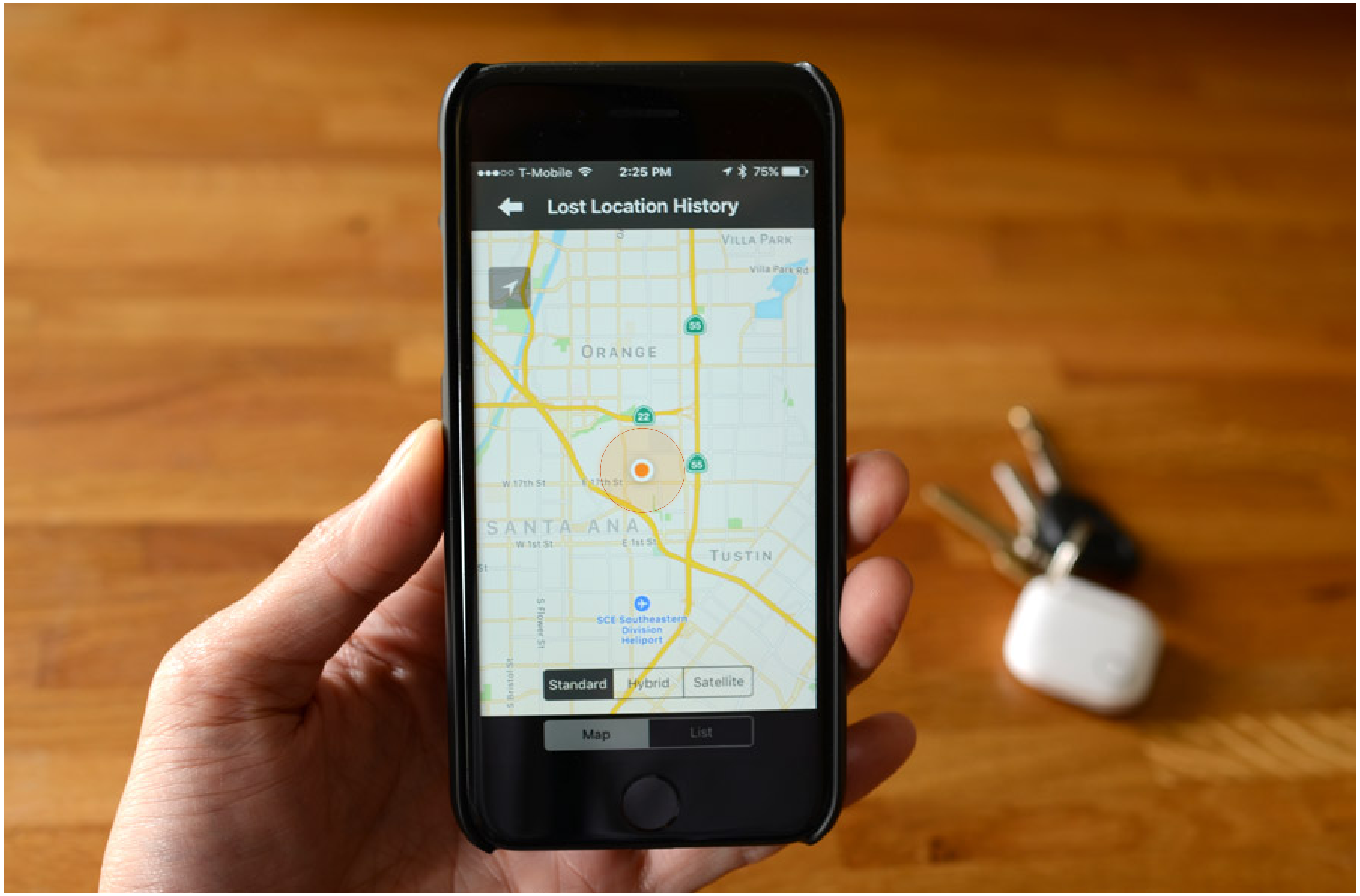
$$P(E) = \lim_{n \to \infty} \frac{\#(E)}{n}$$

Review: Meaning of probability

A quantification of ignorance

image: Frank Derks

Review: Meaning of probability



A quantification of ignorance

image: www.yourbestdigs.com

Review: Axioms of probability

(1) $$0 \leq P(E) \leq 1$$

(2) $$P(S) = 1$$

(3) $$\text{If } E \cap F = \emptyset, \quad \text{then}$$
$$P(E \cup F) = P(E) + P(F)$$

(Sum rule, but with probabilities!)

Review: Corollaries

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

$$\text{If } E \subseteq F, \quad \text{then} \quad P(E) \leq P(F)$$

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

(Principle of inclusion/exclusion, but with probabilities!)

Review: Inclusion/exclusion with more than two sets

prob. of OR

add or subtract (based on size)

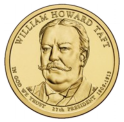
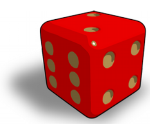
prob. of AND

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n} (-1)^{(r+1)} \sum_{i_1 < \cdots < i_r} P\left(\bigcap_{j=1}^{r} E_{i_j}\right)$$

sum over subset sizes

sum over all **subsets** of that size

Equally likely outcomes

	Coin flip	S = {Heads, Tails}
	Two coin flips	S = {(H, H), (H, T), (T, H), (T, T)}
	Roll of 6-sided die	S = {1, 2, 3, 4, 5, 6}

$$P(\text{Each outcome}) = \frac{1}{|S|}$$

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

(counting!)

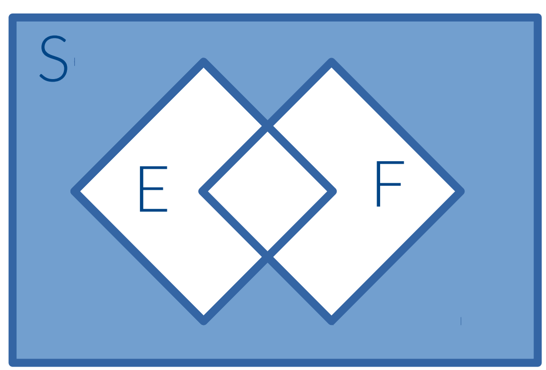
Review: How do I get started?

For word problems involving probability, start by **defining events**!

Review: Getting rid of ORs

Finding the probability of an OR of events can be nasty. Try **using De Morgan's laws** to turn it into an AND!

$$P\left(A \cup B \cup \cdots \cup Z\right) = 1 - P\left(A^c B^c \cdots Z^c\right)$$



Birthdays

WolframAlpha computational knowledge engine.

birthday problem calculator

Web Apps Examples Random

- number of people: 35
- number of possible birthdays: 365

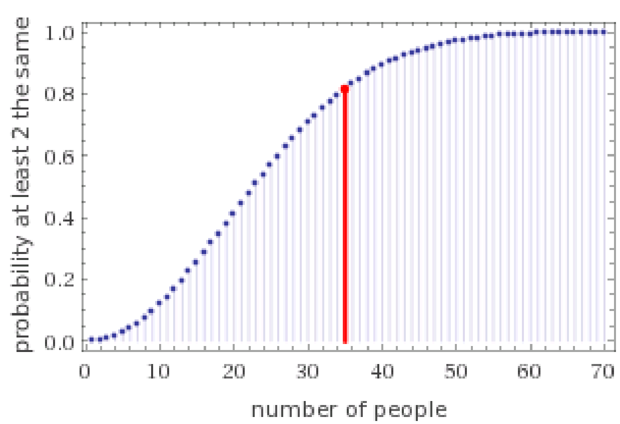
Probabilities that people have the same birthday:

More

	probability	chance
at least 2 the same	0.8144	≈ 1 in 1.2
at least 3 the same	0.0452	≈ 1 in 22

(assuming people chosen independently and all 365 possible birthdays are equally likely)

Probability vs. number of people:



Wolfram Alpha

Review: Flipping cards

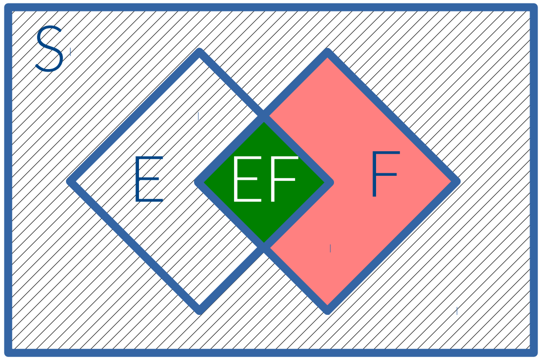
- Shuffle deck.
- Reveal cards from the top until we get an Ace. Put Ace aside.
- What is P(*next* card is the Ace of Spades)?
- P(next card is the 2 of Clubs)?

P(Ace of Spades) **=** P(2 of Clubs)

Definition of conditional probability

The conditional probability P(E | F) is the probability that E happens, **given** that F has happened. F is the new sample space.

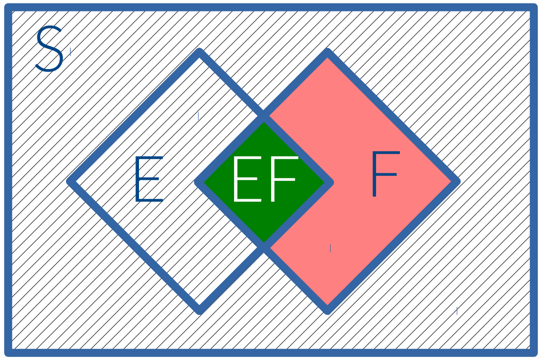
$$P(E|F) = \frac{P(EF)}{P(F)}$$



Equally likely outcomes

If all outcomes are equally likely:

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



Rolling two dice

D_1 D_2

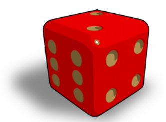
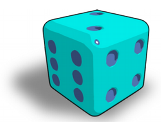
E: {all outcomes such that the sum of the two dice is 4}

What should you hope for D_1 to be?

A) 2
B) 1 and 3 are equally good
C) 1, 2, 3 are equally good
D) other

https://bit.ly/1a2ki4G → https://b.socrative.com/login/student/
Room: CS109SUMMER17

Rolling two dice

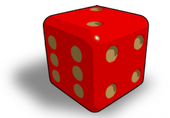
 

D_1 D_2

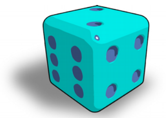
$$P(D_1 + D_2 = 4) = ?$$

E: {all outcomes such that the sum of the two dice is 4}

Rolling two dice



D_1



D_2

$$P(D_1 + D_2 = 4) = \frac{3}{36} = \frac{1}{12}$$

S = {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),

(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),

(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),

(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),

(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),

(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

$$|E| = 3$$

$$|S| = 36$$

Rolling two dice

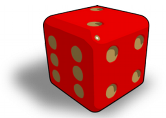
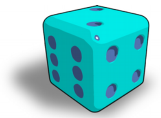
D_1 D_2

$$P(D_1 + D_2 = 4 | D_1 = 2) = ?$$

E: {all outcomes such that the sum of the two dice is 4}

F: {all outcomes such that the first die is 2}

Rolling two dice

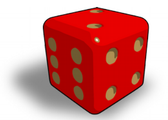
D_1 D_2

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

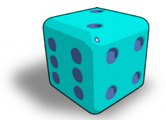
E: {all outcomes such that the sum of the two dice is 4}

F: {all outcomes such that the first die is 2}

Rolling two dice



D_1



D_2

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

S = {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),

(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),

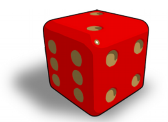
(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),

(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),

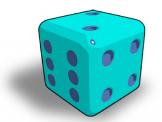
(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),

(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

Rolling two dice



D_1



D_2

$$P(E|F)=\frac{1}{6}$$

$$|EF|=1$$

$$|F|=6$$

S = {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),

(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),

(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),

(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),

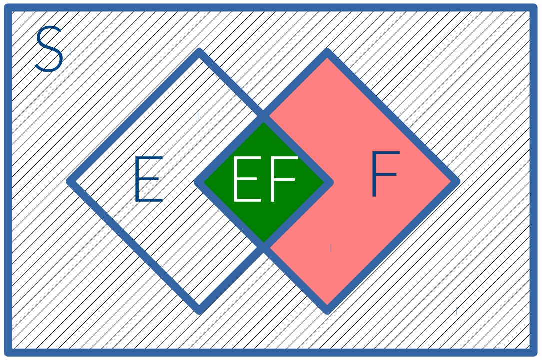
(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),

(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

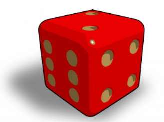
Definition of conditional probability

The conditional probability P(E | F) is the probability that E happens, **given** that F has happened. F is the new sample space.

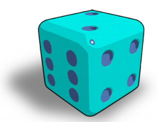
$$P(E|F) = \frac{P(EF)}{P(F)}$$



Rolling two dice



D_1



D_2

$$P(E|F) = \frac{1}{6} = \frac{1/36}{6/36}$$

$$|EF| = 1$$

$$|F| = 6$$

S = {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),

(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),

(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),

(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),

(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),

(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

Juniors

What is the probability that a randomly chosen (rising) junior in CS 109 comes to class?

C: event that randomly chosen student comes to class
J: event that randomly chosen student is a junior

$$P(C|J) = \frac{P(CJ)}{P(J)} = \frac{?/65}{16/65}$$

What if P(F) = 0?

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

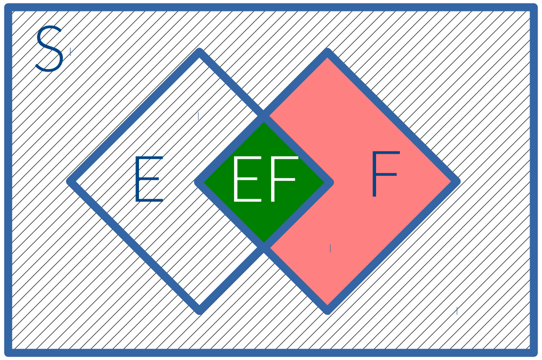
```
ZeroDivisionError:
float division by zero
```

Congratulations! You've observed the impossible!

Definition of conditional probability

The conditional probability P(E | F) is the probability that E happens, **given** that F has happened. F is the new sample space.

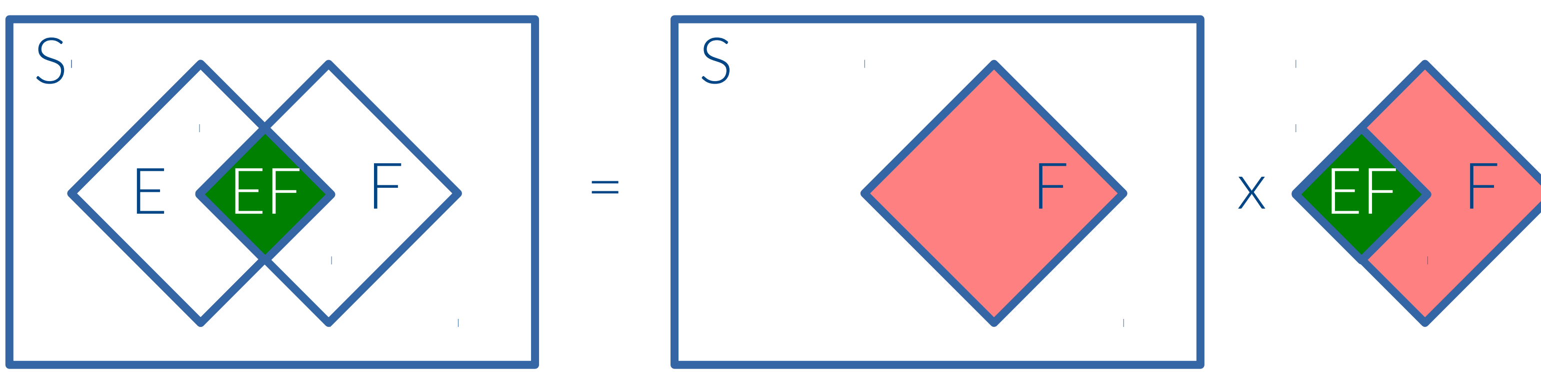
$$P(E|F) = \frac{P(EF)}{P(F)}$$



Chain rule of probability

The probability of **both** events happening is the probability of **one happening** times the probability of **the other happening given the first one**.

$$P(EF) = P(F)\,P(E|F)$$



General chain rule of probability

The probability of **all** events happening
is the probability of **the first** happening
times the prob. of **the second given the first**
times the prob. of **the third given the first two**
...etc.

$$P(EFG\ldots) = P(E)\,P(F|E)\,P(G|EF)\ldots$$

Four piles of cards

- Divide deck randomly into 4 piles of 13 cards each.
- What is P(one Ace in each pile)?

S: ways of labeling 52 cards with 4 types of labels
E: ways resulting in all Aces getting different labels

$$|E| = 4! \cdot \binom{48}{12,12,12,12}$$

$$|S| = \binom{52}{13,13,13,13}$$

$$P(E) = \frac{|E|}{|S|} \approx 0.105$$

Four piles of cards

- Divide deck randomly into 4 piles of 13 cards each.
- What is P(one Ace in each pile)?

E_1: Ace of Spades goes in any one pile

E_2: Ace of Clubs goes in different pile from Spades

E_3: Ace of Hearts goes in different pile from first two

E_4: Ace of Diamonds goes in different pile from first three

$$P(E) = P(E_1 E_2 E_3 E_4) = P(E_1) P(E_2 | E_1) P(E_3 | E_1 E_2) P(E_4 | E_1 E_2 E_3)$$

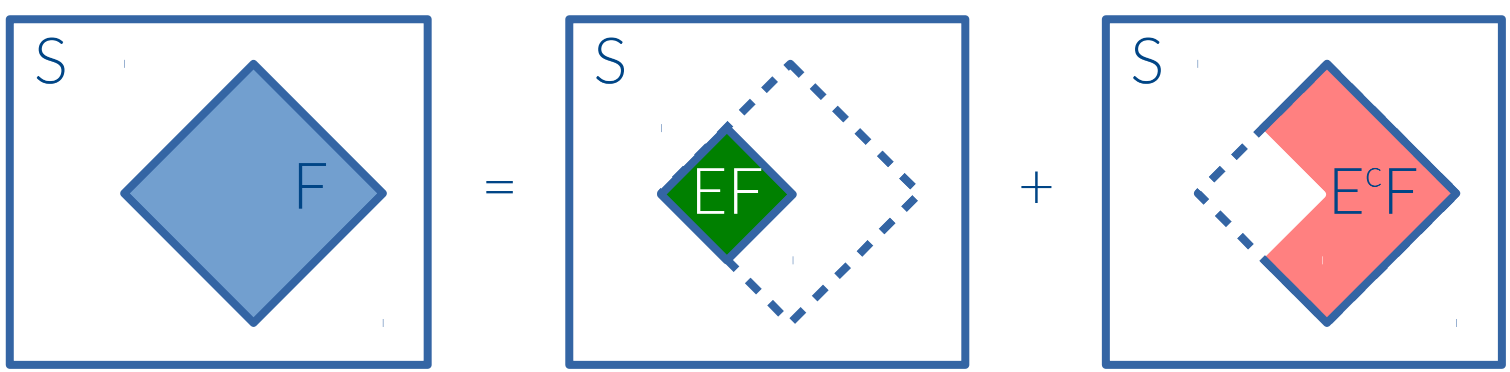
$$= \frac{52}{52} \cdot \frac{39}{51} \cdot \frac{26}{50} \cdot \frac{13}{49} \approx 0.105$$

Law of total probability

You can compute an overall probability by adding up the case when an event **happens** and when it **doesn't happen**.

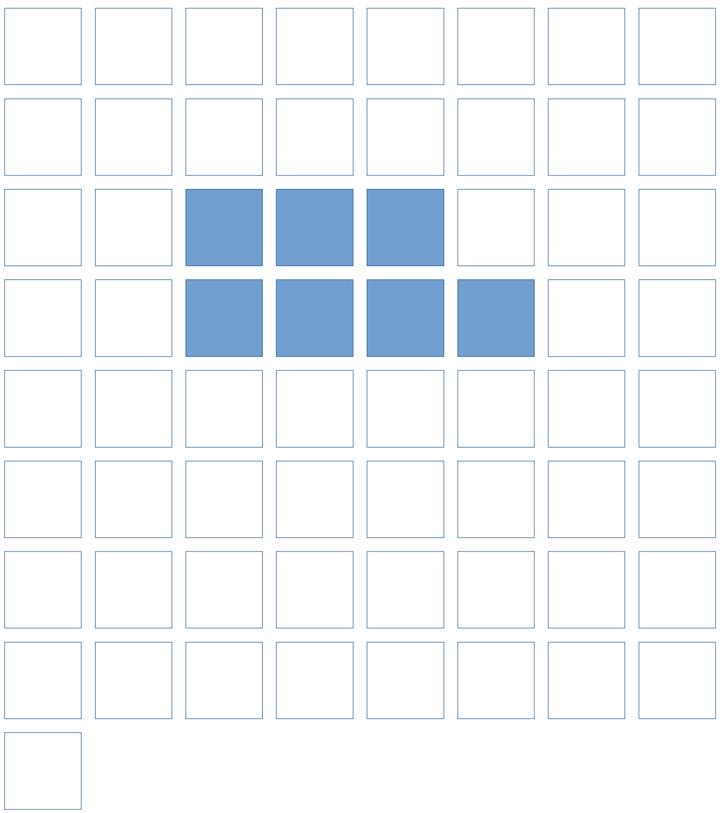
$$P(F) = P(EF) + P(E^C F)$$

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



Majoring in CS

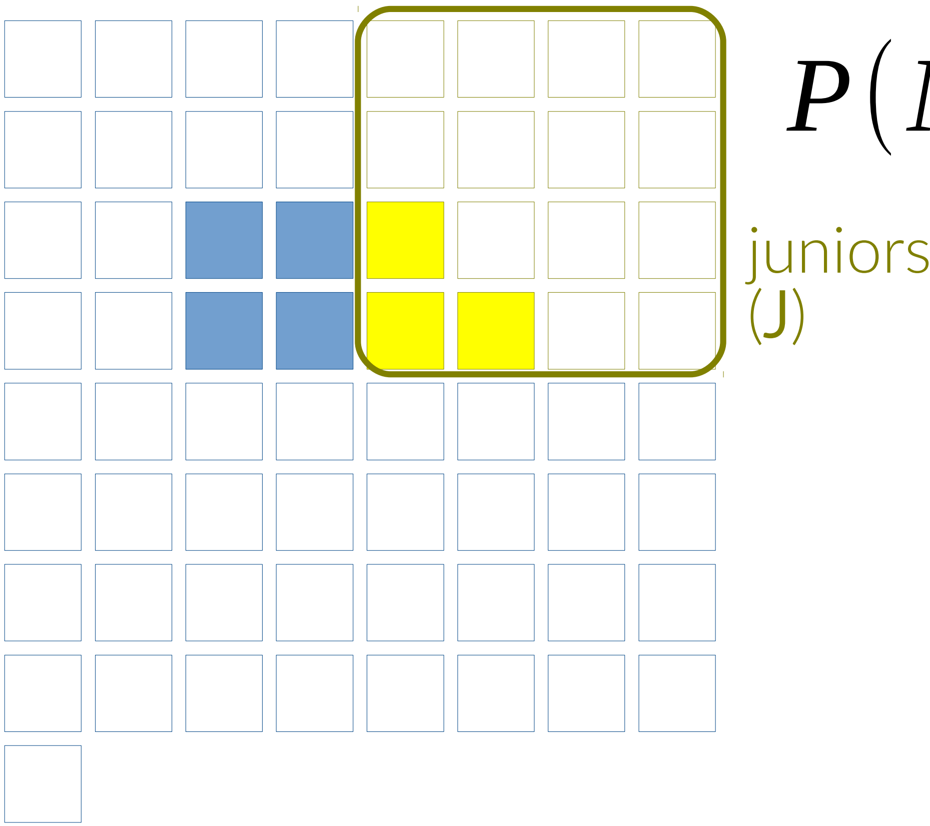
What is the probability that a randomly chosen student in CS 109 is a (declared) CS major (M)?



$$P(M)=\frac{|M|}{|S|}=\frac{7}{65}$$

Majoring in CS

What is the probability that a randomly chosen student in CS 109 is a (declared) CS major (M)?



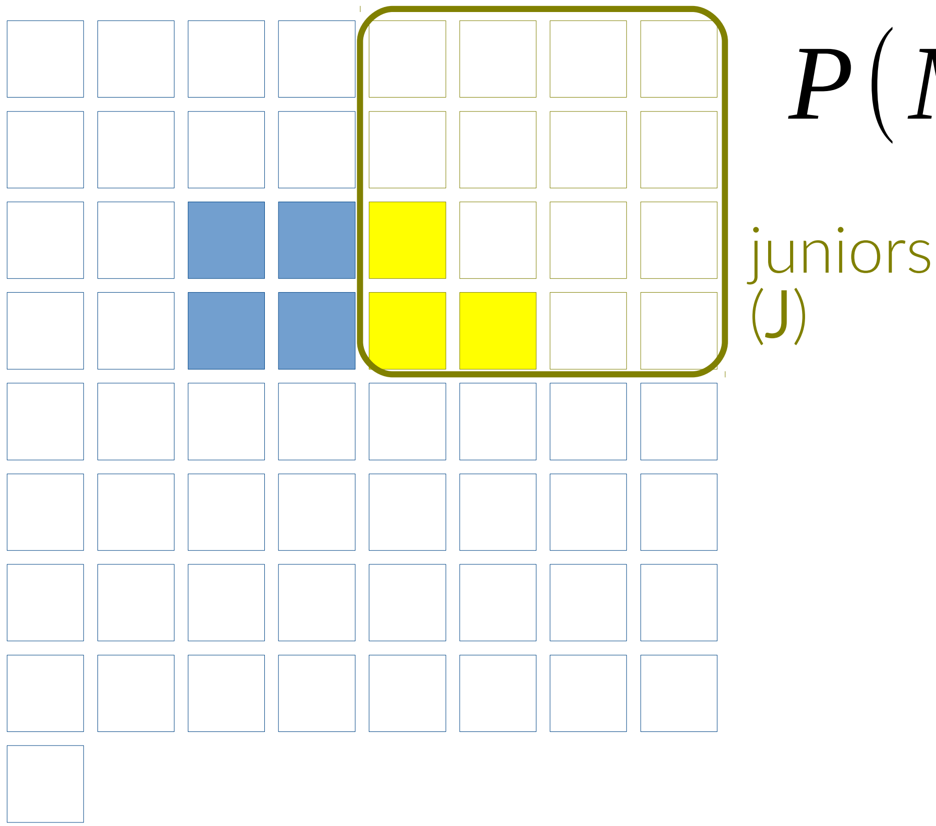
$$P(M) = P(JM) + P(J^C M)$$

$$= \frac{|JM|}{|S|} + \frac{|J^C M|}{|S|}$$

$$= \frac{3}{65} + \frac{4}{65} = \frac{7}{65}$$

Majoring in CS

What is the probability that a randomly chosen student in CS 109 is a (declared) CS major (M)?



$$P(M) = P(J)P(M|J) + P(J^C)P(M|J^C)$$

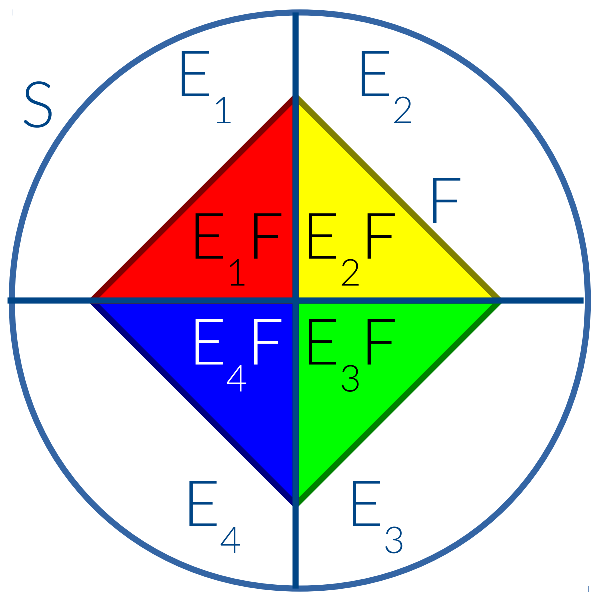
$$= \frac{16}{65} \cdot \frac{3}{16} + \frac{49}{65} \cdot \frac{4}{49}$$

$$= \frac{7}{65}$$

General law of total probability

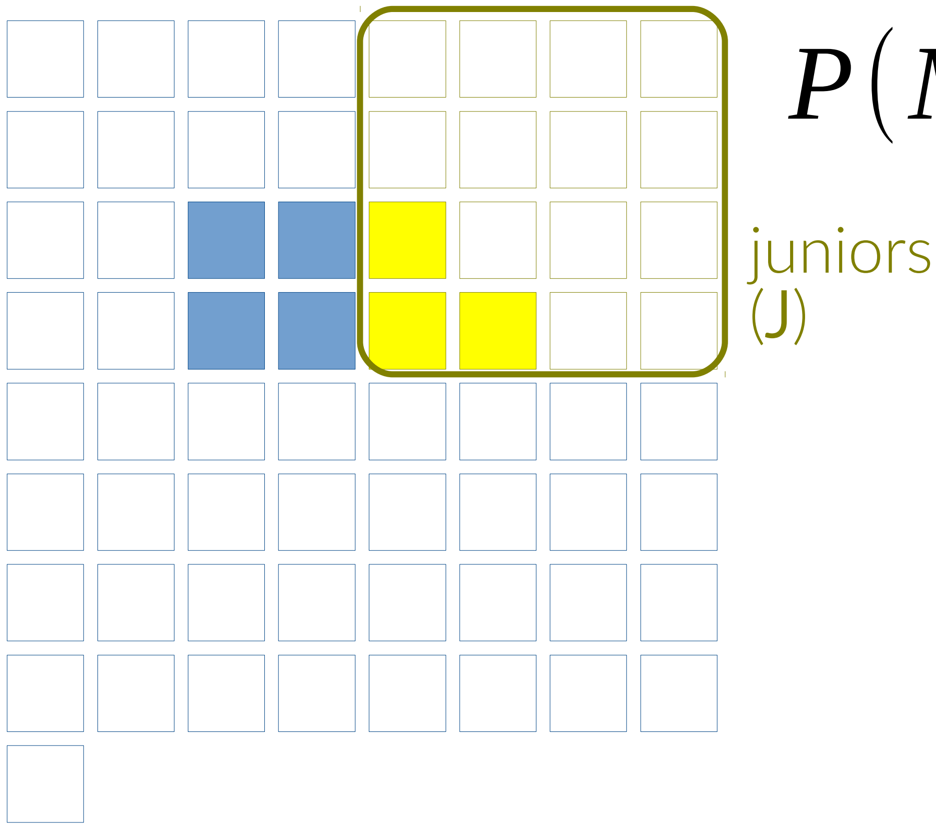
You can compute an overall probability by summing over **mutually exclusive** and **exhaustive** sub-cases.

$$P(F) = \sum_i P(E_i F)$$
$$= \sum_i P(E_i) P(F|E_i)$$



Majoring in CS

What is the probability that a randomly chosen student in CS 109 is a (declared) CS major (M)?



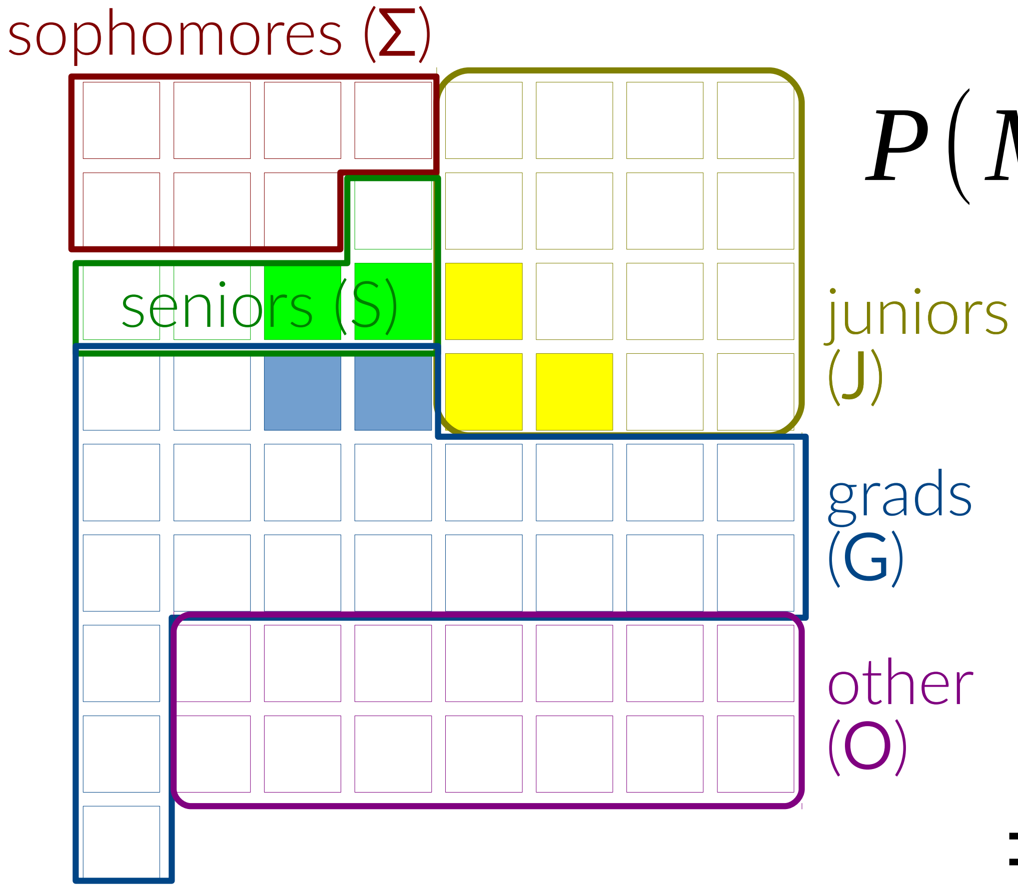
$$P(M) = P(J)P(M|J) + P(J^C)P(M|J^C)$$

$$= \frac{16}{65} \cdot \frac{3}{16} + \frac{49}{65} \cdot \frac{4}{49}$$

$$= \frac{7}{65}$$

Majoring in CS

What is the probability that a randomly chosen student in CS 109 is a (declared) CS major (M)?



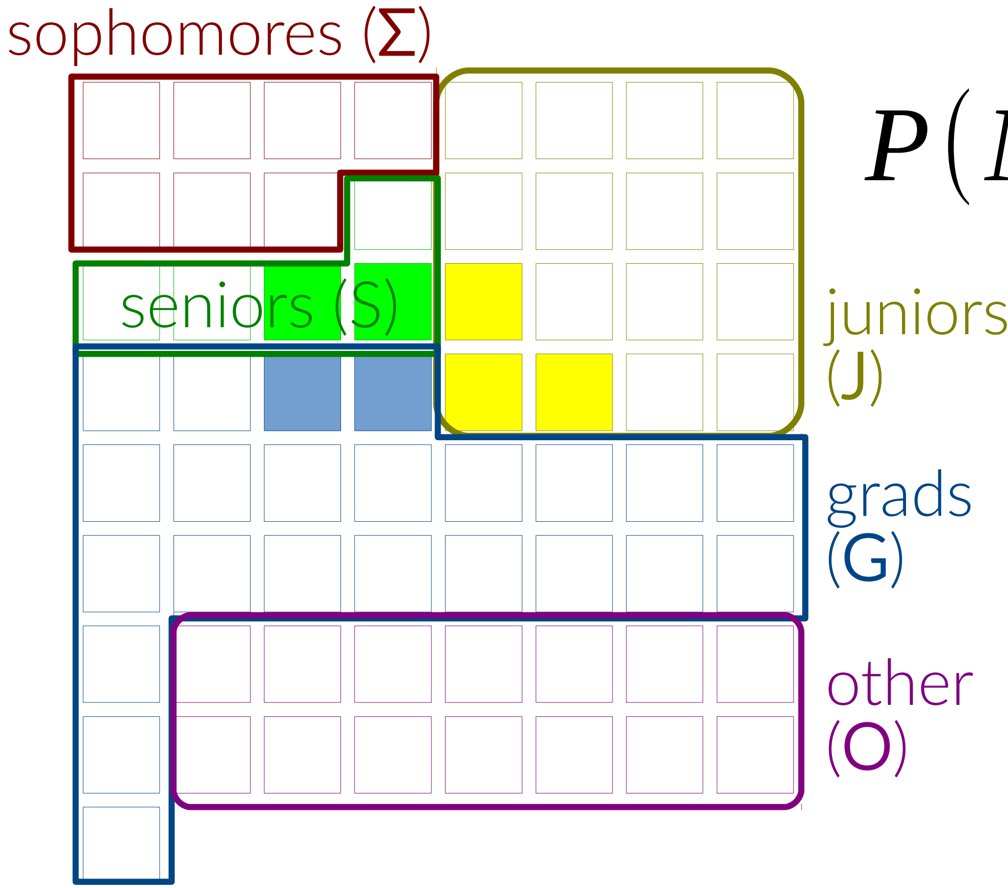
$$P(M) = P(M\Sigma) + P(MJ) + P(MS) + P(MG) + P(MO)$$

$$= \frac{0}{65} + \frac{3}{65} + \frac{2}{65} + \frac{2}{65} + \frac{0}{65}$$

$$= \frac{7}{65}$$

Majoring in CS

What is the probability that a randomly chosen student in CS 109 is a (declared) CS major (M)?



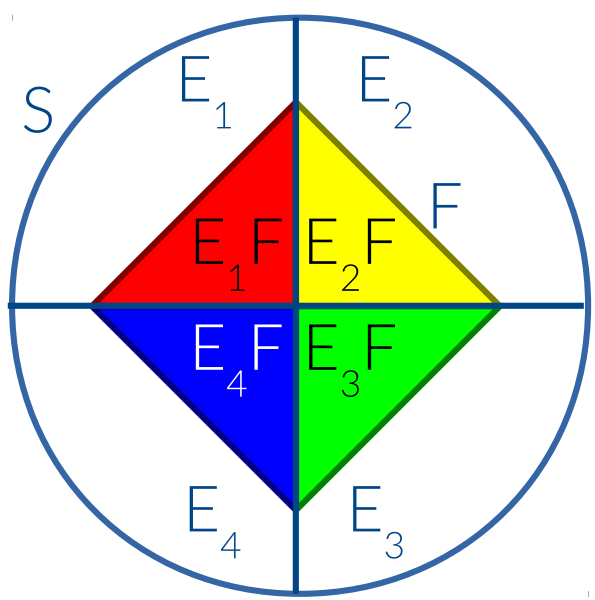
$$P(M) = P(\Sigma)P(M|\Sigma) + P(J)P(M|J) + P(S)P(M|S) + P(G)P(M|G) + P(O)P(M|O)$$

$$= \frac{0}{7}\cdot\frac{7}{65} + \frac{3}{16}\cdot\frac{16}{65} + \frac{2}{5}\cdot\frac{5}{65} + \frac{2}{23}\cdot\frac{23}{65} + \frac{0}{14}\cdot\frac{14}{65} = \frac{7}{65}$$

General law of total probability

You can compute an overall probability by summing over **mutually exclusive** and **exhaustive** sub-cases.

$$P(F) = \sum_i P(E_i F)$$
$$= \sum_i P(E_i) P(F|E_i)$$



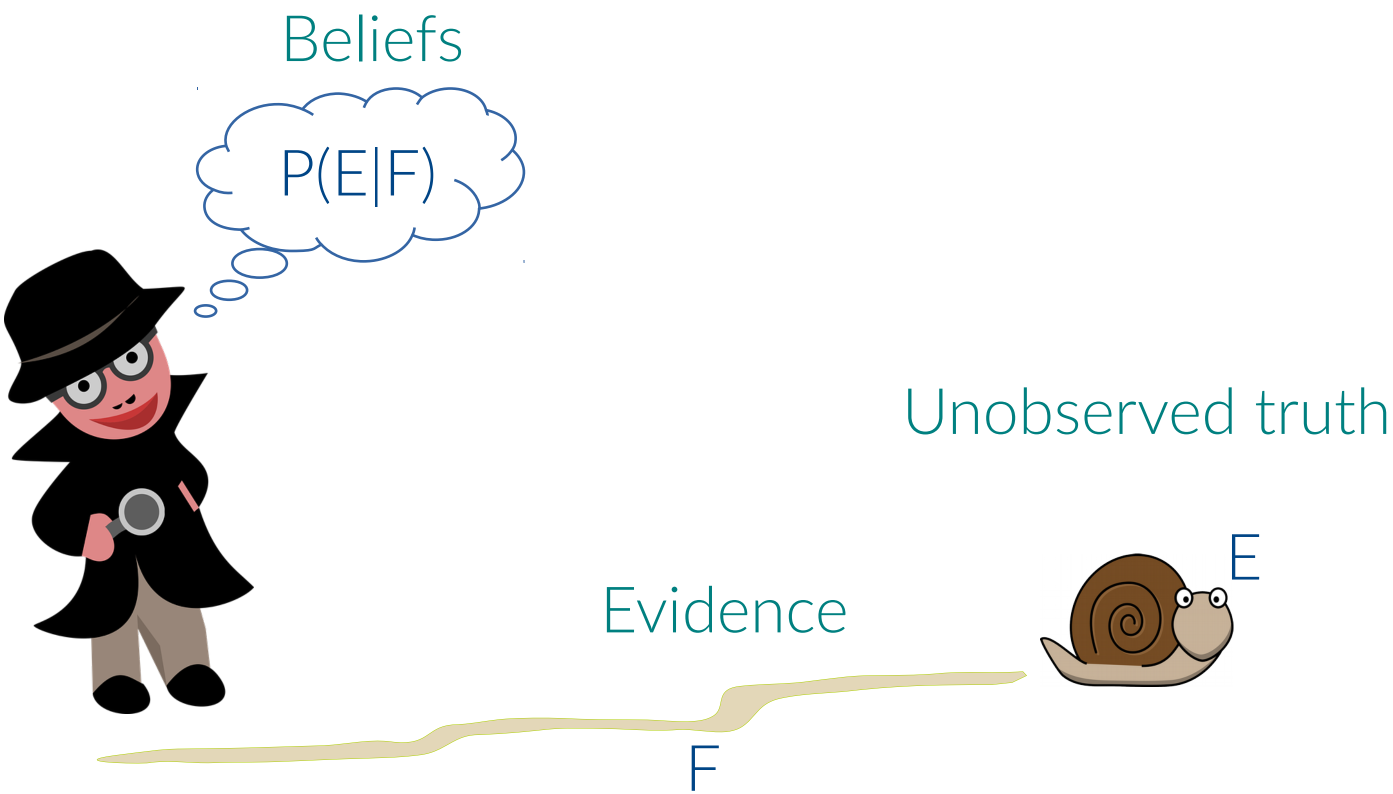
Break time!

Bayes' theorem

You can “flip” a conditional probability if you multiply by the probability of the **hypothesis** and divide by the probability of the **observation**.

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

Probabilistic inference



Probabilistic inference

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

definition of conditional probability

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

chain rule (aka same def'n, plus algebra)

Finding the denominator

If you don't know P(F) on the bottom,
try using the **law of total probability**.

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

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

Zika testing

0.08% of people have Zika

$$P(Z) = 0.0008$$

90% of people with Zika test positive

$$P(T|Z) = 0.90$$

7% of people without Zika test positive

$$P(T|Z^C) = 0.07$$

Someone tests positive. What's the probability they have Zika?

Z: event that person has Zika
T: event that person tests positive

$$P(Z|T) = \frac{P(T|Z)P(Z)}{P(T)}$$

?

Zika testing

0.08% of people have Zika

$$P(Z) = 0.0008$$

90% of people with Zika test positive

$$P(T|Z) = 0.90$$

7% of people without Zika test positive

$$P(T|Z^C) = 0.07$$

Someone tests positive. What's the probability they have Zika?

Z: event that person has Zika
T: event that person tests positive

$$P(Z|T) = \frac{P(T|Z)P(Z)}{P(T|Z)P(Z) + P(T|Z^C)P(Z^C)}$$

Zika testing

0.08% of people have Zika

$$P(Z) = 0.0008$$

90% of people with Zika test positive

$$P(T|Z) = 0.90$$

7% of people without Zika test positive

$$P(T|Z^C) = 0.07$$

Someone tests positive. What's the probability they have Zika?

Z: event that person has Zika
T: event that person tests positive

$$P(Z|T) = \frac{P(T|Z)P(Z)}{P(T|Z)P(Z) + P(T|Z^C)P(Z^C)} = \frac{0.90 \cdot 0.0008}{0.90 \cdot 0.0008 + 0.07 \cdot 0.9992} \approx 0.01$$

Bayes: Terminology

hypothesis

$$P(Z|T) = \frac{P(T|Z)P(Z)}{P(T)}$$

observation

Bayes: Terminology

likelihood → $P(T|Z)$

prior → $P(Z)$

posterior → $P(Z|T)$

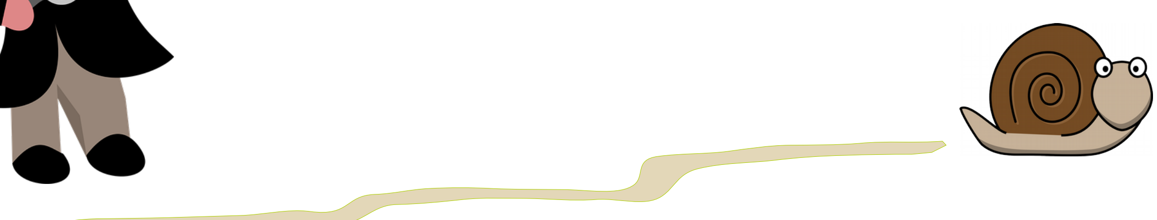
$$P(Z|T) = \frac{P(T|Z)\,P(Z)}{P(T)}$$

normalizing constant → $P(T)$

Bayes' theorem

You can "flip" a conditional probability if you <u>multiply</u> by the probability of the **hypothesis** and <u>divide</u> by the probability of the **observation**.

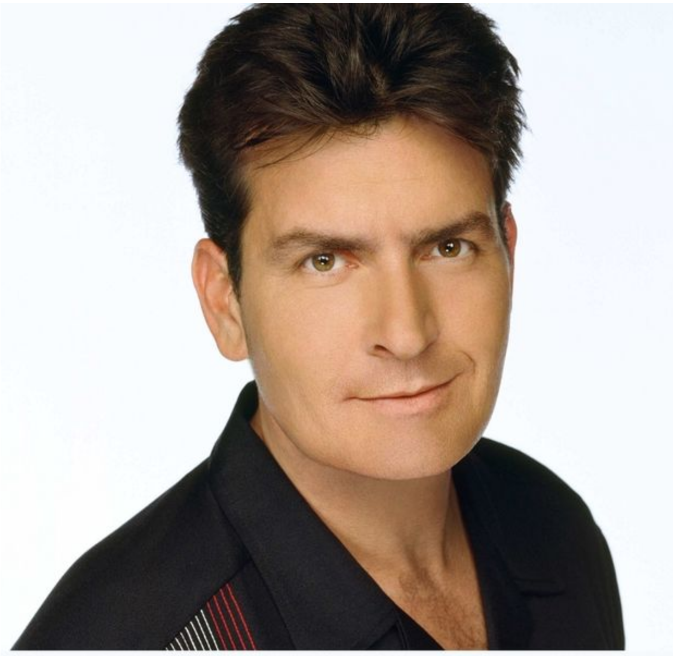
$$P(E|F) = \frac{P(F|E)\,P(E)}{P(F)}$$



Thomas Bayes

Rev. Thomas Bayes (~1701-1761):
British mathematician and Presbyterian minister

[*citation needed*]



images: (left) unknown, (right) Martin ormazabal

Implicatures

Work is work.

Examples adapted from Grice (1970)

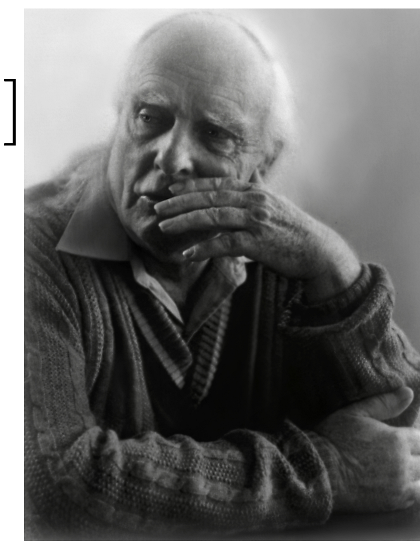
Implicatures

Will produced a series of sounds that corresponded closely to the tune of "Hey Jude."

Examples adapted from Grice (1970)

Implicatures: Grice's maxims

- “Make your contribution as informative as required [...]
- Do not make your contribution more informative than is required. [...]
- Do not say what you believe to be false. [...]
- Avoid obscurity of expression.
- Avoid ambiguity.
- Be brief (avoid unnecessary prolixity).”



(Grice, 1970)

Image credit: Siobhan Chapman

Implicatures: Grice's maxims

Work is work.

“Make your contribution as informative as required”

Implicatures: Grice's maxims

Will produced a series of sounds that corresponded closely to the tune of "Hey Jude."

"Be brief (avoid unnecessary prolixity)."

Implicatures: Grice's maxims

How do you like
my new haircut?

...It's shorter in
the back!

“Be relevant.”

Image credit: Gage Skidmore

Implicatures

1 2 3

Image credit: Chris Potts

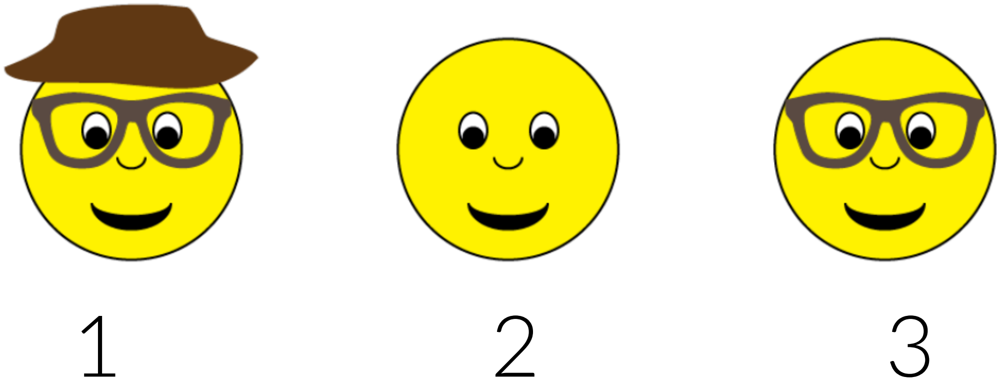
Implicatures

1 2 3

“glasses”

Image credit: Chris Potts

Implicatures



"person"

Image credit: Chris Potts

RSA: Bayesian pragmatic reasoning

"hat"	✓	✗	✗
"glasses"	✓	✗	✓
"person"	✓	✓	✓

RSA: Bayesian pragmatic reasoning

"hat"	1	0	0	literal (naive) listener
"glasses"	0.5	0	0.5	
"person"	0.33	0.33	0.33	

RSA: Bayesian pragmatic reasoning

"hat"	0.33	0	0
"glasses"	0.33	0	0.5
"person"	0.33	1	0.5

literal (naive) speaker

RSA: Bayesian pragmatic reasoning

t

m

"hat"	1	0	0
"glasses"	0.4	0	0.6
"person"	0.18	0.55	0.27

pragmatic listener

$$L(t|m) = \frac{S(m|t)\,P(t)}{P(m)}$$

(Frank and Goodman, 2012)

Conditional probabilities are still probabilities

Everything you know about some set of events is still true if you condition **consistently** on some other event!

$$0 \leq P(E) \leq 1$$

Conditional probabilities are still probabilities

Everything you know about some set of events is still true
if you condition **consistently** on some other event!

$$0 \leq P(E|G) \leq 1$$

Conditional probabilities are still probabilities

Everything you know about some set of events is still true if you condition **consistently** on some other event!

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

Conditional probabilities are still probabilities

Everything you know about some set of events is still true if you condition **consistently** on some other event!

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

Conditional probabilities are still probabilities

Everything you know about some set of events is still true if you condition **consistently** on some other event!

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

must be the same!

Conditional probabilities are still probabilities

Everything you know about some set of events is still true if you condition **consistently** on some other event!

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

Conditional probabilities are still probabilities

Everything you know about some set of events is still true if you condition **consistently** on some other event!

$$P(EF|G) = P(E|FG)\,P(F|G)$$

Conditional probabilities are still probabilities

Everything you know about some set of events is still true if you condition **consistently** on some other event!

$$P(EF|G) = P(E|FG)\,P(F|G)$$

same as P((E|F) | G)

"conditioning twice" = conditioning on AND

Conditional probabilities are still probabilities

Everything you know about some set of events is still true if you condition **consistently** on some other event!

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

Conditional probabilities are still probabilities

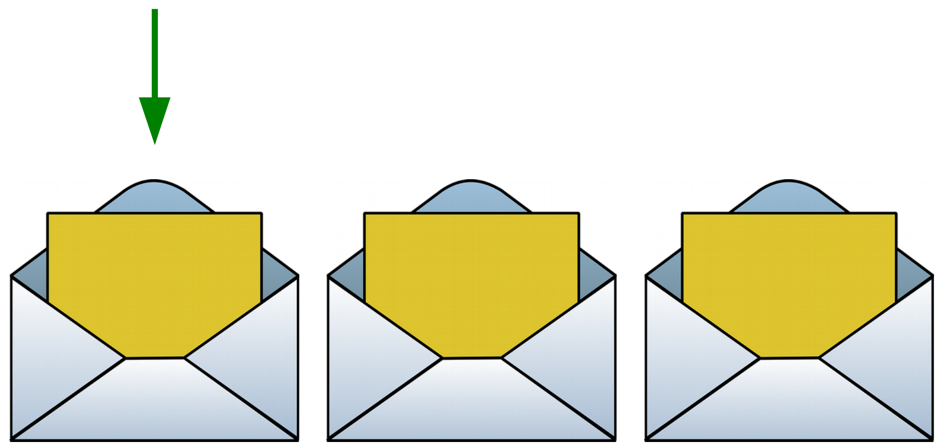
Everything you know about some set of events is still true if you condition **consistently** on some other event!

$$P(E|FG) = \frac{P(F|EG)\,P(E|G)}{P(F|G)}$$

Let's play a game

$1 **$20** $1

Let's Make a Deal



A has $20 (P = 1/3): staying wins

B has $20 (P = 1/3): switching wins

C has $20 (P = 1/3): switching wins

Important assumption: host must open a $1!