

CS109 - lecture 4

Let's derive the 3rd Axioms from last class
corollary

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

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

$$\text{Is due to } \boxed{\text{axiom 3}} \quad P(E \cup F) = P(E) + P(F)$$

BTW:

$$\boxed{\text{Axiom 1:}} \quad 0 \leq P(E) \leq 1$$

$$\boxed{\text{Axiom 2:}} \quad P(S) = 1$$

$$P(E \cup E^c) = P(S) = 1$$

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

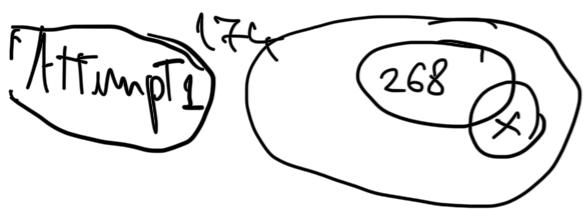
$$\text{so } P(E) = 1 - P(E^c)$$

Serendipity

- Population of Stanford is 17000 people
- You are friends with X people
- You walk into a room and see 268 people
- What is the probability of seeing someone you know?
 $\text{know: } 1, 2, 3, \dots$

$$P(\text{seeing someone you know}) = P(E) = \frac{\text{Prob(Know1)} + \dots + \text{Prob(Know}X\text{)}}{X}$$

know in a room) - $\overline{P(S)} = \frac{\text{_____}}{(17000)} \quad (268)$



$$\binom{x}{1} \binom{17000-1}{268-1} \rightarrow \text{But just for one!!!}$$

Pro Tip: Sometimes it is easy to calculate the probability of the Event not happen

$$\text{Sample Space} = \binom{17000}{268}$$

$$\text{Event Space for not happen} = \binom{17000-X}{268} \quad \begin{matrix} \text{Remove the} \\ \text{People I know} \end{matrix}$$

(do not know anyone)

$$X = 100$$

$$\text{Prob not seeing someone} = \frac{\binom{17000-100}{268}}{\binom{17000}{268}} = 0,2$$

$$\text{Prob of seeing someone} = 1 - 0,2$$

↓
Can be see 1 or 2 or 3 or
000

Much easier to calculate

Goal for the Lesson: Conditional Probability

$$P(E \text{ and } F)$$

↑ ↓ Definition of Conditional
Probability

$$P(E | F)$$

Law of total
Probability

$$P(E)$$

Bayes'
theorem

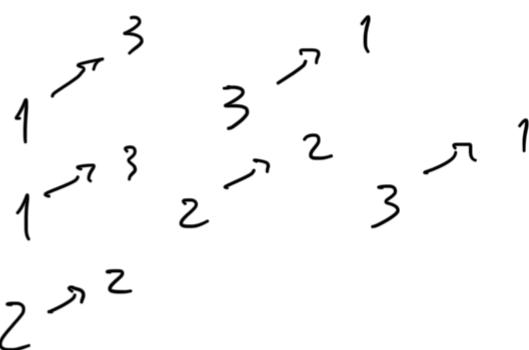
$$P(F | E)$$

Conditional Probability

Roll 2 dice (6 sided) so that the sum is 4
• What is the best outcome for $P(D_1)$?

Choices:

A: 1 and 3 tie



B: 1, 2, 3 tie

C: 2 is best

D : other

$$|S| = \left[(1,1) \ (1,2) \ (\cancel{1,3}) \ \dots \ (1,6) \atop (2,1) \ (2,2) \atop (3,1) \right] \atop (6,1) \ \dots \ \dots \ (6,6) \Big]$$
$$= 36$$

$$E = \{(1,3), (2,2), (3,1)\} \quad |E| = 3$$

$$P(E) = \frac{3}{36} = \frac{1}{12}$$

What is $P(E)$, given F already observed ? $F=2$

$$\text{Now } S = \{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)\}$$

$$E' = \{(2,2)\} \quad \text{So } P(E') = \frac{1}{6}$$

$$\text{So } P(E) \xrightarrow{\text{from: } \boxed{0}} \xrightarrow{\text{to: }} P(E')$$

the probability increased!

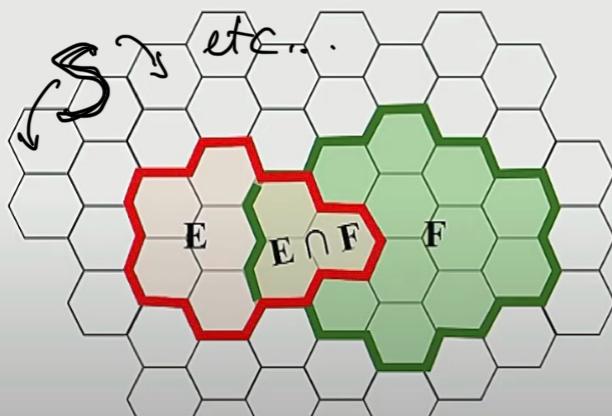
$$\frac{1}{12} \rightarrow \underline{\frac{1}{6}}$$

Increased because the Sample

Space decreased more than the Event Space

Conditional Probability, visual intuition

The conditional probability of E given F is the probability that E occurs given that F has already occurred. This is known as conditioning on F .



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

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

So with equally likely outcomes:

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

Implicit AND

$$= \frac{|E \cap F|}{|S \cap F|} = \frac{|E \cap F|}{|F|}$$

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

General Definition
of Conditional Probability

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

Chain Rule of
P.I.

If $P(F) = 0$ then (and now)

$P(E|F) = \text{undefined}$ (observed the impossible)

Netflix \rightarrow "Life is beautiful"

What is the probability that a user will watch
"Life is beautiful" $P(E)$

$S = \{\text{watch, not watch}\}$

$E = \{\text{watch}\}$ $P(E) \neq \frac{1}{2}$

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

$$\text{So: } P(E) = \frac{10234231}{50923123} = 0.20$$

Now the conditional prob:

"What is the probability that a user watches
"Life is beautiful" given they watched "CDA"?"
 $\underbrace{E \text{ AND } F}_{P(E|F)}$

$D \cap E = 1$

$$\begin{aligned}
 P(E|F) &= \frac{\dots}{P(F)} \\
 &= \frac{\frac{\# \text{ users that watched both}}{\# \text{ total users}}}{\frac{\# \text{ user that watched Coda}}{\# \text{ total users}}} \\
 &= \frac{\# \text{ Users that watched both} \text{ (outcomes)}}{\# \text{ Users that watched Coda} \text{ (outcomes)}}
 \end{aligned}$$

≈ 0.42

Why all this?

Because Machine Learning is:

Probability + Data + Computers

Notation

And	Or	Given
$P(E \text{ and } F)$	$P(E \text{ or } F)$	$P(E F)$
$P(E, F)$	$P(E \cup F)$	$P(E F, G)$
$P(E F)$		

$P(E \cap F)$

Different notations are used across the world.

Chain Rule via baby poop

In the morning when she wakes up, a baby has 50% chance of having pooped. The chance that a baby cries if she has pooped is 50%.

- What is the probability that a baby has pooped and cries?

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

$$P(E|F) = \frac{P(EF)}{P(F)} \rightarrow \text{General definition of Conditional Probability}$$

$$P(EF) = P(E|F) \cdot P(F) \rightarrow \text{Chain Rule} \quad !!$$

$$P(\text{cries and Pooped}) = P(\text{cries given it pooped}) \cdot P(\text{if pooped})$$

$$= 0,5 \cdot 0,5 = 0,25$$

In conditional probabilities the order matters

- - - - - / or - - - 1

$$\text{So } P(E|F) \neq P(F|E)$$

Generalized chain Rule

$$P_n(E_1 \text{ and } E_2 \text{ and } \dots \text{ and } E_m)$$

$$= P_n(E_1) \cdot P_n(E_2 | E_1) \cdot P_n(E_3 | E_1, E_2) \cdots$$

$$P_n(E_m | E_1, E_2, \dots, E_{m-1})$$

LAW OF TOTAL PROBABILITY

Say E and F are events in S



$$P(E) = P(E \cap F) + P(E \cap F^c)$$

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

We were missing the probability of crying given "no poop"

If we say it is e.g. $\frac{1}{8}$

$$P(E) = 0,25 + \frac{1}{8} \cdot 0,5 = 0,375$$

Evolution of Bacteria

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

Law of Total Probability



- You have bacteria in your gut which is causing a disease.
 - 10% have a mutation which makes them resistant to anti-biotics
 - You take half a course of anti-biotics...
- Probability a bacteria survives given it has the mutation: 20%
- Probability a bacteria survives given it doesn't have the mutation: 1%
- What is the probability that a randomly chosen bacteria survives?

$$P(F) = 10\%$$

F = mutation

$$P(E|F) = 20\%$$

E = Survives

$$P(E|F^C) = 1\%$$

BTW
 $P(E^C|F) = 0.8$
 $P(E^C|F^C) = 0.99$
 are also known

$$\begin{aligned}
 P(E) &= P(E|F) \cdot P(F) + P(E|F^C) \cdot P(F^C) \\
 &= 0.2 \cdot 0.1 + 0.01 \cdot 0.9 \\
 &= 0.02 + 0.009 \\
 &= 0.029 = 2.9\%
 \end{aligned}$$

"The Idea that wouldn't die" - Bayes' Theorem

Allows us to get from $P(E|F)$ to $P(F|E)$!

v v