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…

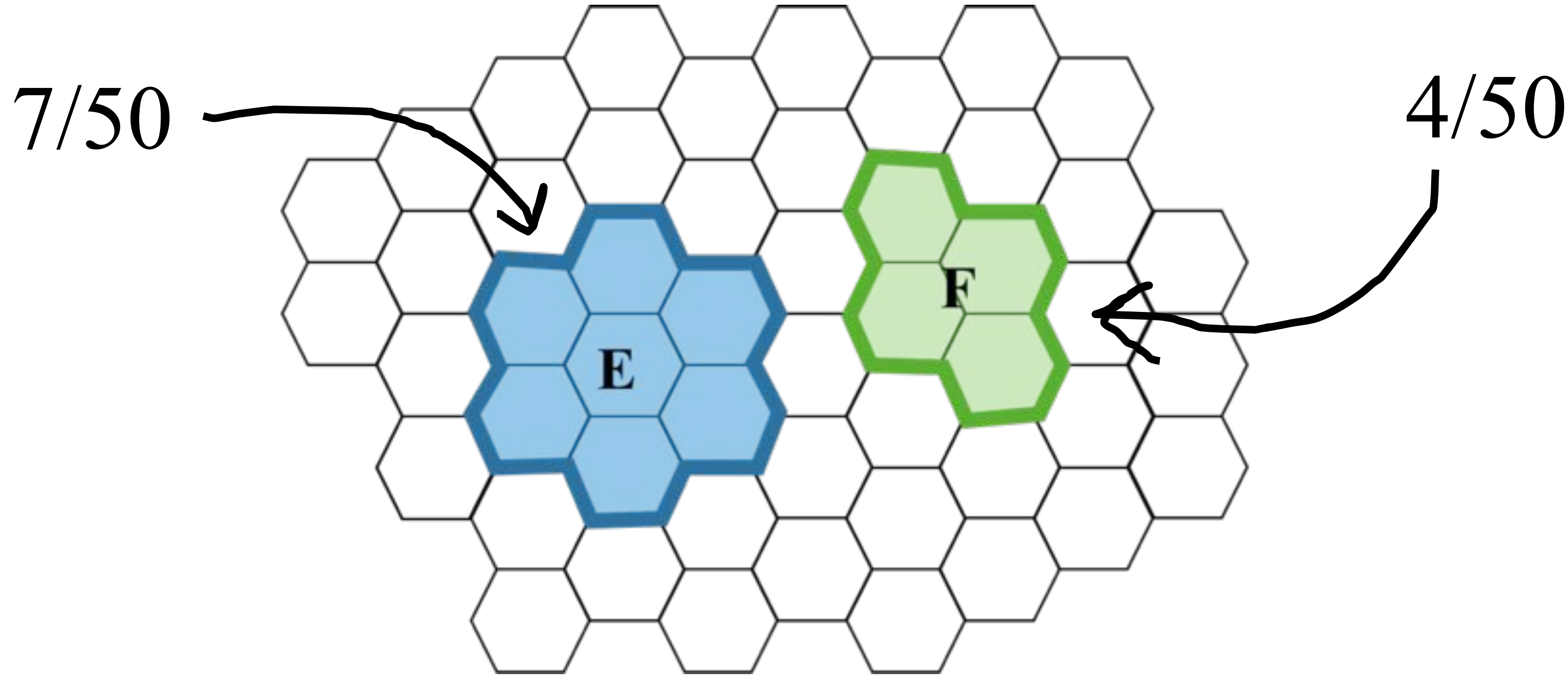
Axioms of Probability

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



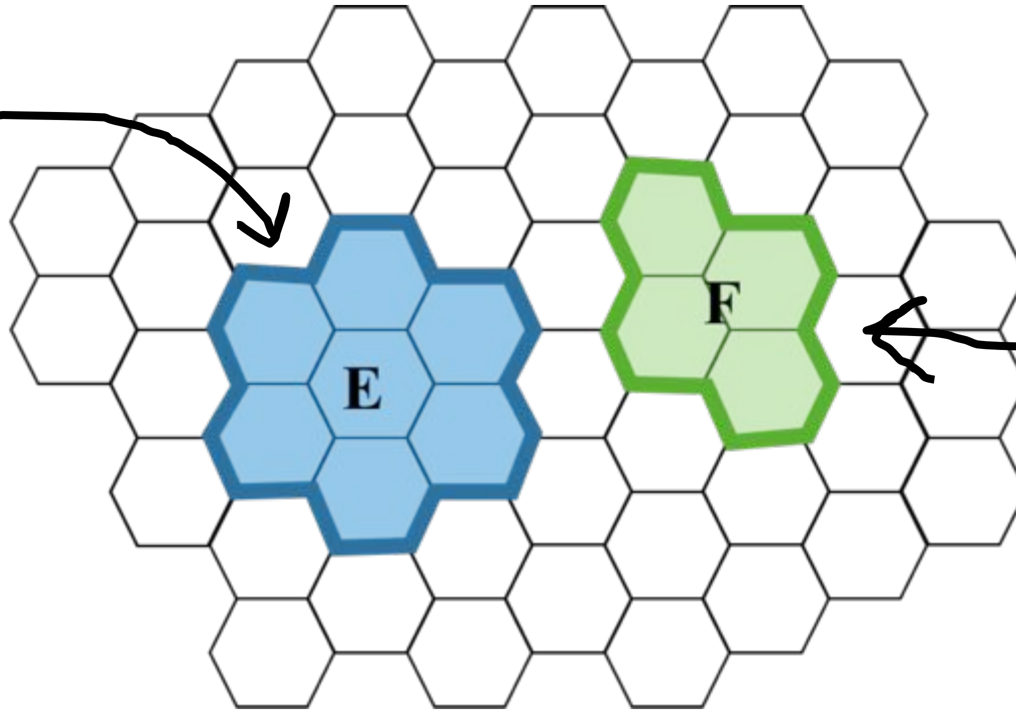
If events are mutually exclusive, probability of OR is simple:

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

Mutually Exclusive Events

7/50

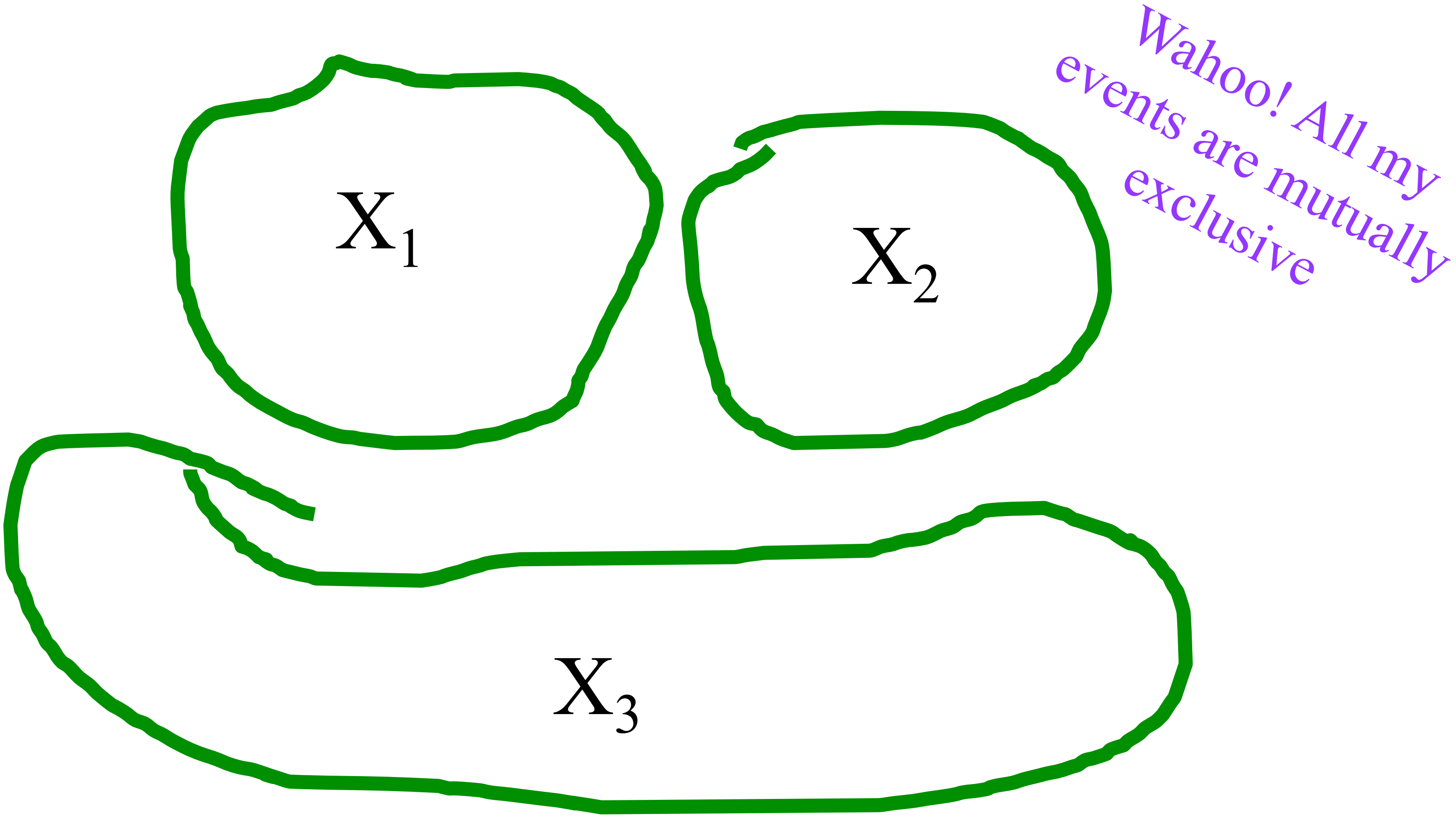
4/50



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"

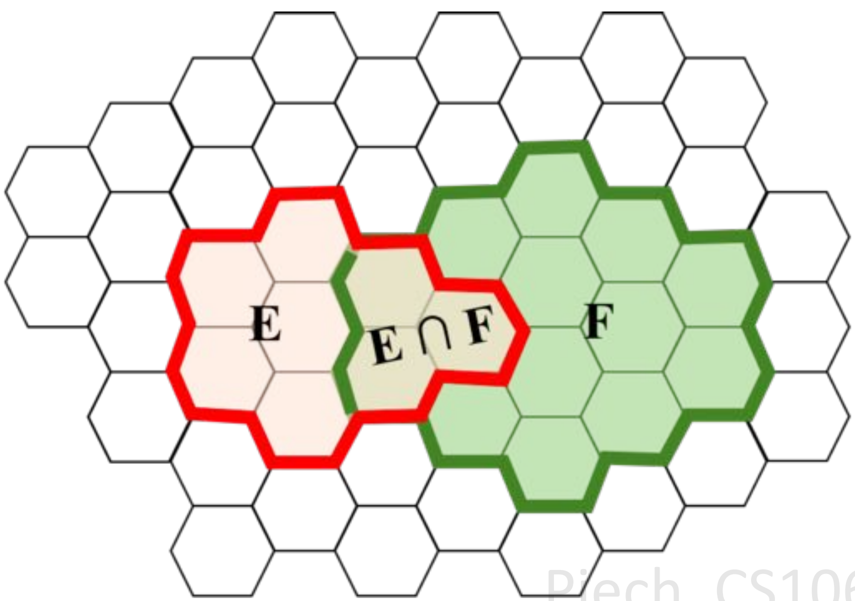
- Written as $P(E|F)$
  - Means "P(E, given F already observed)"
  - Sample space, S, reduced to those elements consistent with F (i.e. $S \cap F$)
  - Event space, E, reduced to those elements consistent with F (i.e. $E \cap F$)

Conditional Probability

With equally likely outcomes:

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

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



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

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

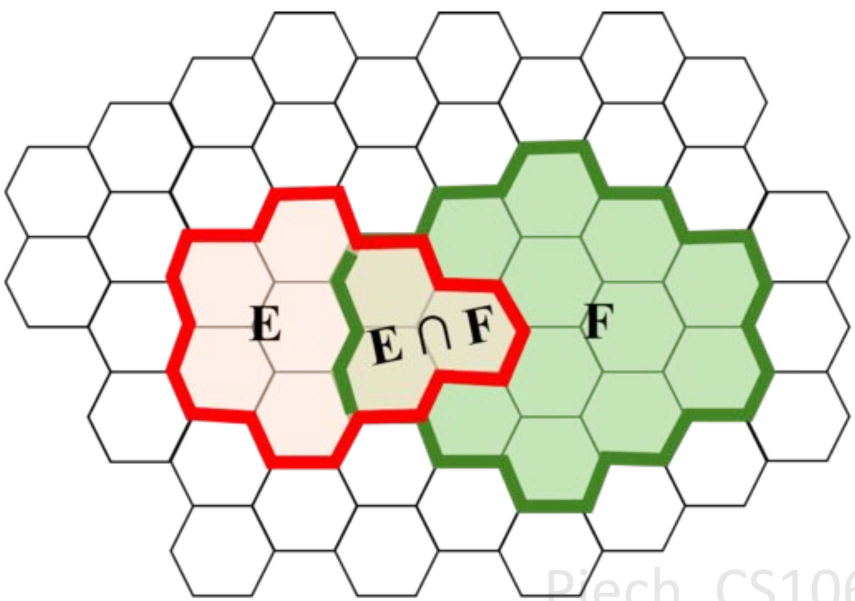
Conditional Probability

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:

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

Conditional Paradigm

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

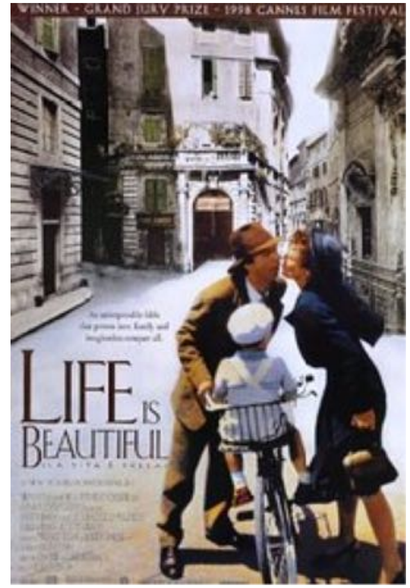
NETFLIX

+ Learn

Netflix and Learn

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

$P(E)$



S = {Watch, Not Watch}

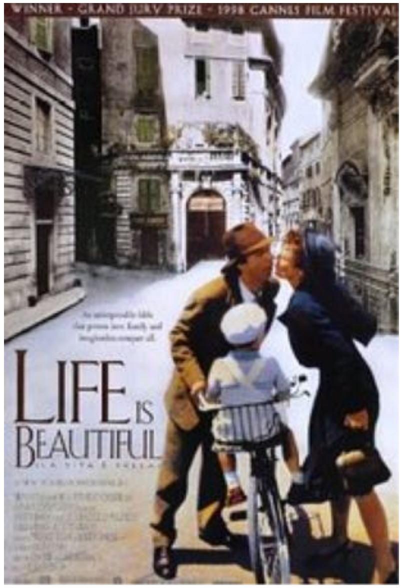
E = {Watch}

P(E) = ½ ?

Netflix and Learn

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

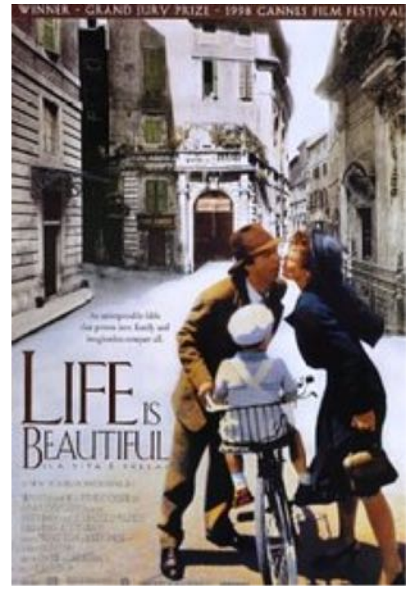
$$P(E)$$



Netflix and Learn

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

$P(E)$

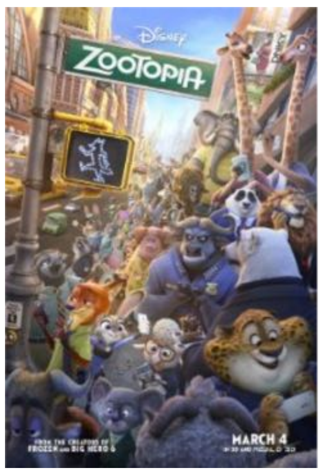


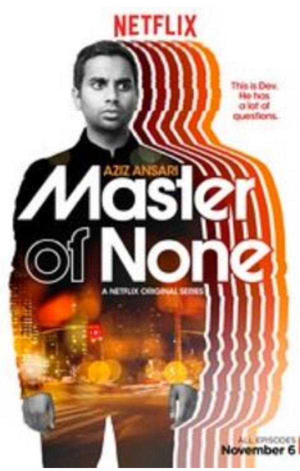
$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \approx \frac{\#\text{people who watched movie}}{\#\text{people on Netflix}}$$

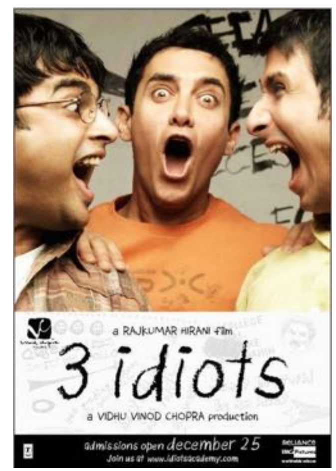
$$P(E) = 10{,}234{,}231 \,/\, 50{,}923{,}123 = 0.20$$

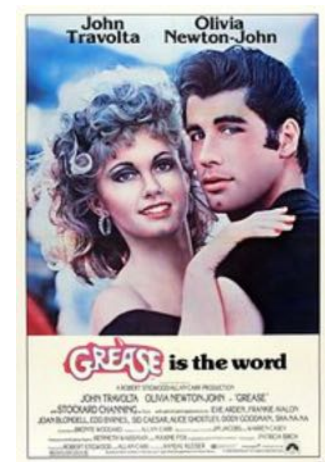
Netflix and Learn

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

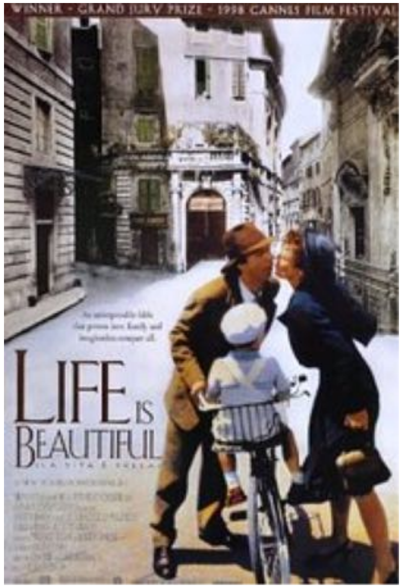
P(E) = 0.23

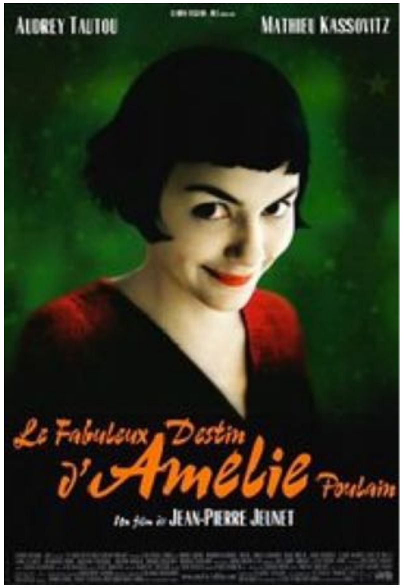
* These are the actual estimates

Netflix and Learn

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

$$P(E|F)$$



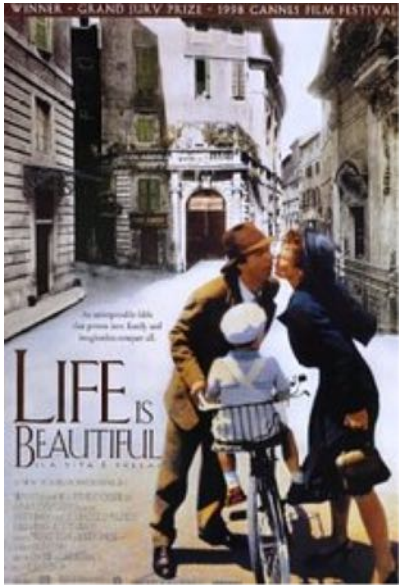


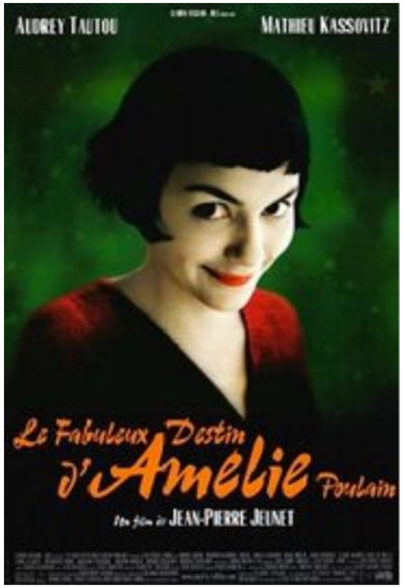
$$P(E|F) = \frac{P(EF)}{P(F)}$$

Netflix and Learn

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

$$P(E|F)$$



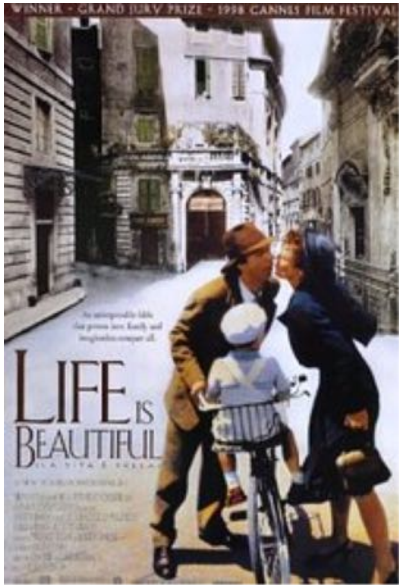


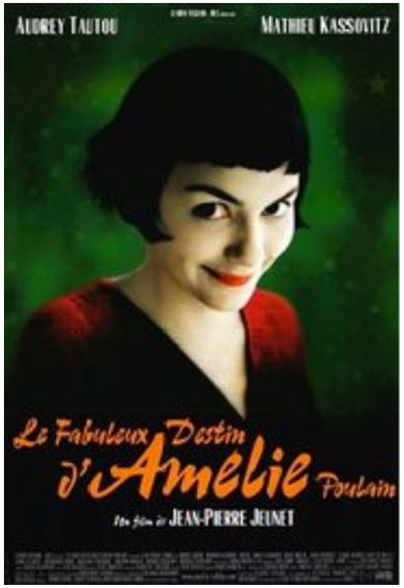
$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\frac{\#\text{people who watched both}}{\#\text{people on Netflix}}}{\frac{\#\text{people who watched } F}{\#\text{people on Netflix}}}$$

Netflix and Learn

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

$P(E|F)$





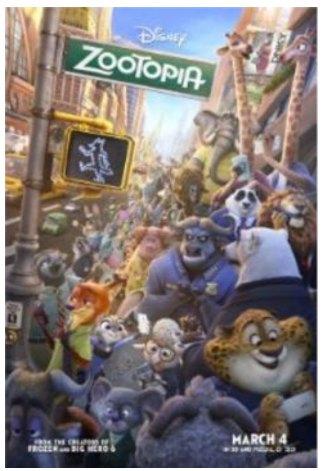
$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\#\text{people who watched both}}{\#\text{people who watched } F}$$

$P(E|F) = 0.42$

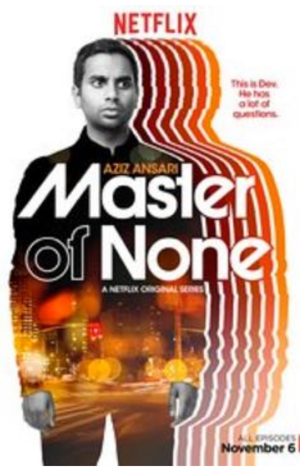
Netflix and Learn

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

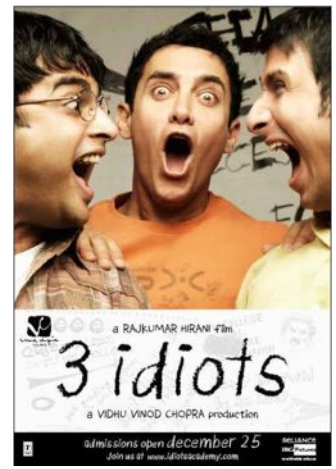
$P(E|F) =$ 0.14



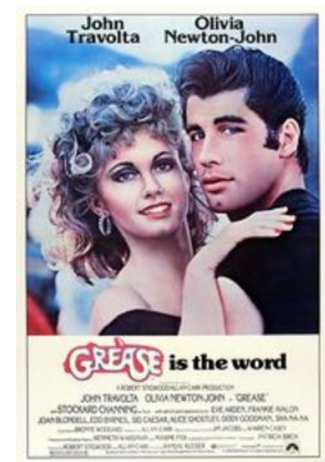
$P(E|F) =$ 0.35



$P(E|F) =$ 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

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