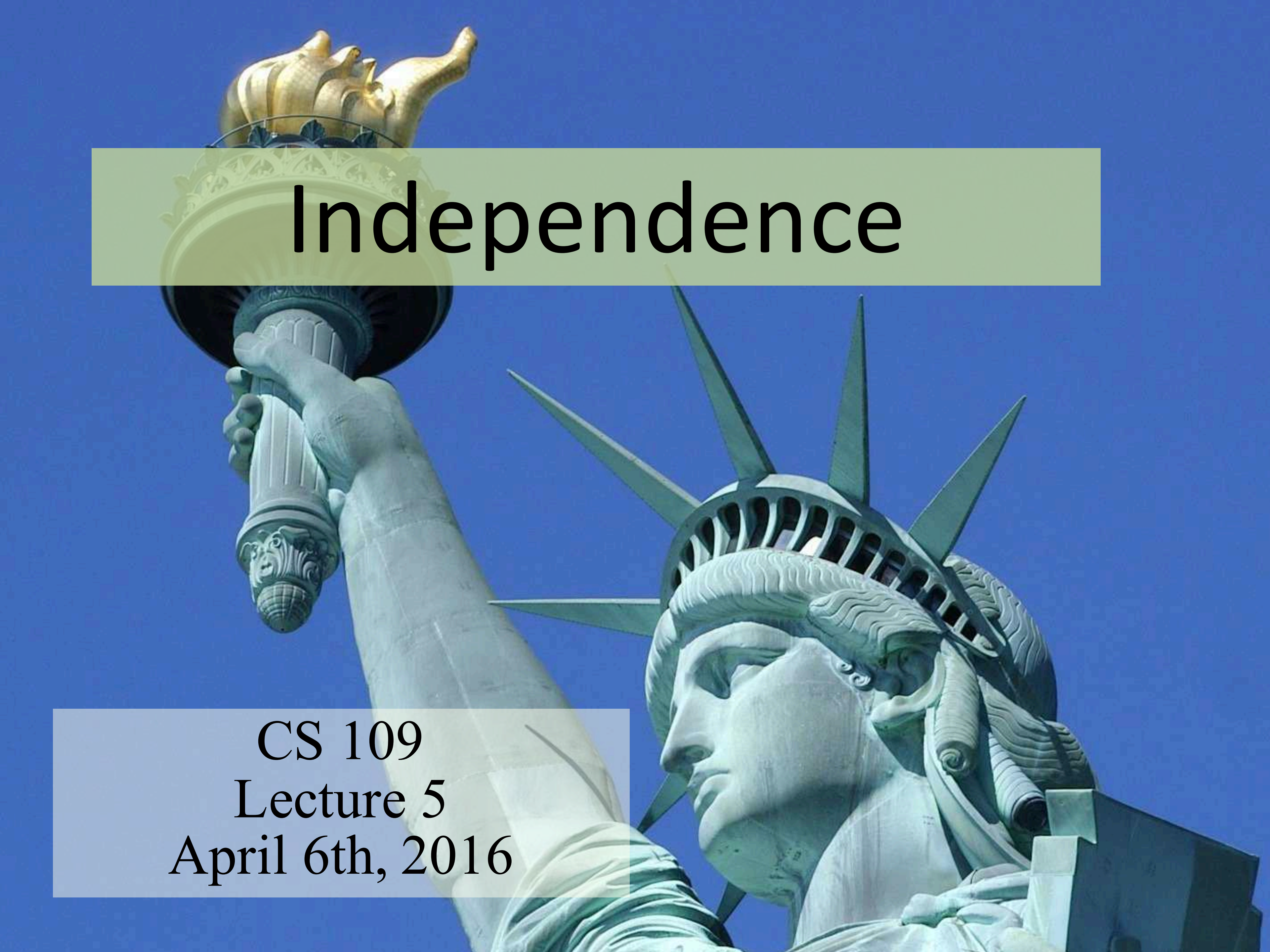
Independence



CS 109
Lecture 5
April 6th, 2016

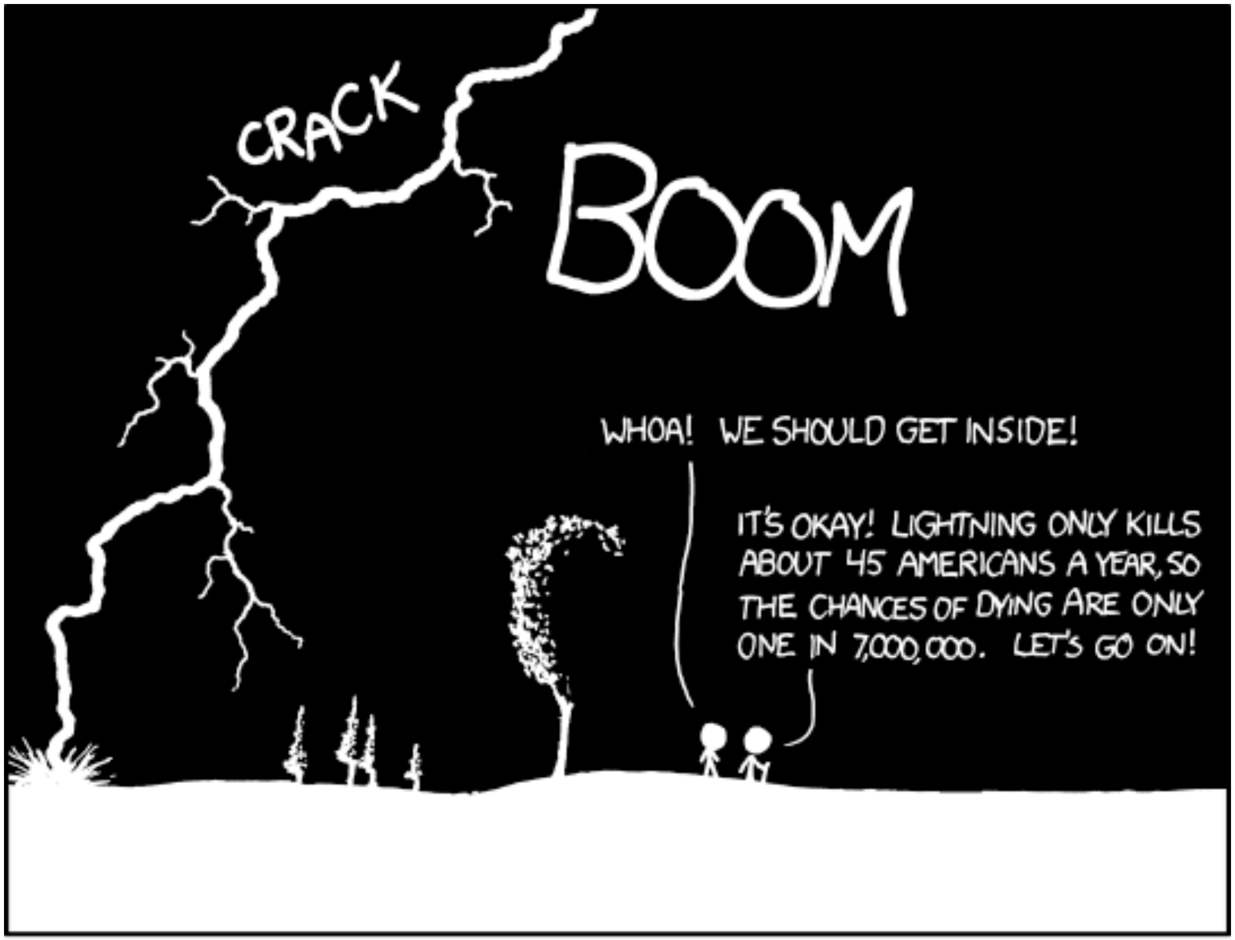
Today's Topics

Last time:
Conditional Probability
Bayes Theorem

Today:
Independence
Conditional Independence

Next time:
Random Variables

The Tragedy of Conditional Prob



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

Thanks xkcd! `http://xkcd.com/795/`

A Few Useful Formulas

- For any events A and B:

$P(A\ B) = P(B\ A)$ (Commutativity)

$P(A\ B) = P(A \mid B)\,P(B)$ (Chain rule)

$= P(B \mid A)\,P(A)$

$P(A\ B^c) = P(A) – P(AB)$ (Intersection)

$P(A\ B) \geq P(A) + P(B) – 1$ (Bonferroni)

Generality of Conditional Probability

- For any events A, B, and E, you can condition consistently on E, and these formulas still hold:

$$P(A\ B \mid E) = P(B\ A \mid E)$$

$$P(A\ B \mid E) = P(A \mid B\ E)\ P(B \mid E)$$

$$P(A \mid B\ E) = \frac{P(B \mid A\ E)\ P(A \mid E)}{P(B \mid E)} \qquad \text{(Bayes' Thm.)}$$

BAE's Theorem?

$$P(A \mid B\ E) = \frac{P(B \mid A\ E)\ P(A \mid E)}{P(B \mid E)}$$

Generality of Conditional Probability

- For any events A, B, and E, you can condition consistently on E, and these formulas still hold:

$$P(A\,B \mid E) = P(B\,A \mid E)$$

$$P(A\,B \mid E) = P(A \mid B\,E)\,P(B \mid E)$$

$$P(A \mid B\,E) = \frac{P(B \mid A\,E)\,P(A \mid E)}{P(B \mid E)} \qquad \text{(Bayes' Thm.)}$$

- Can think of E as "everything you already know"
- Formally, P(• | E) satisfies 3 axioms of probability

Our Still Misunderstood Friend

- Roll two 6-sided dice, yielding values D_1 and D_2
  - Let E be event: $D_1 = 1$
  - Let F be event: $D_2 = 1$
- What is P(E), P(F), and P(EF)?
  - $P(E) = 1/6$, $P(F) = 1/6$, $P(EF) = 1/36$
  - $P(EF) = P(E)\,P(F)$ → E and F *independent*
- Let G be event: $D_1 + D_2 = 5$ {(1, 4), (2, 3), (3, 2), (4, 1)}
- What is P(E), P(G), and P(EG)?
  - $P(E) = 1/6$, $P(G) = 4/36 = 1/9$, $P(EG) = 1/36$
  - $P(EG) \neq P(E)\,P(G)$ → E and G *dependent*

Independence

- Two events E and F are called **independent** if:

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

  Or, equivalently: $P(E \mid F) = P(E)$
- Otherwise, they are called **dependent** events

- Three events E, F, and G independent if:

  $P(EFG) = P(E)\,P(F)\,P(G)$, and

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

  $P(EG) = P(E)\,P(G)$, and

  $P(FG) = P(F)\,P(G)$

Independence

- Given independent events E and F, prove that E and F^C are independent

Independence

- Given independent events E and F, prove that E and F^C are independent
- Proof:

  $P(E F^c)$

We want to show that this is equal to $P(E)P(F^C)$

Independence

- Given independent events E and F, prove that E and F^C are independent
- Proof:

$$P(E F^c) = P(E) - P(EF)$$

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

Independence

- Given independent events E and F, prove that E and F^C are independent
- Proof:

$$P(E F^c) = P(E) - P(EF)$$
$$= P(E) - P(E)\,P(F)$$

Since we are told E and F are independent

Independence

- Given independent events E and F, prove that E and F^C are independent
- Proof:

$$P(E F^c) = P(E) - P(EF)$$
$$= P(E) - P(E)\,P(F)$$
$$= P(E)\,[1 - P(F)]$$

Factoring!

Independence

- Given independent events E and F, prove that E and F^C are independent
- Proof:

$$
\begin{aligned}
P(E F^c) &= P(E) - P(EF) \\
&= P(E) - P(E)\,P(F) \\
&= P(E)\,[1 - P(F)] \\
&= P(E)\,P(F^c)
\end{aligned}
$$

Yep, that's the complement

Independence

- Given independent events E and F, prove that E and F^C are independent
- Proof:

$$P(E F^c) = P(E) – P(EF)$$
$$= P(E) – P(E) P(F)$$
$$= P(E) [1 – P(F)]$$
$$= P(E) P(F^c)$$

So, E and F^c independent, implying that:

$$P(E \mid F^c) = P(E) = P(E \mid F)$$

Independence

- Given independent events E and F, prove that E and F^C are independent
- Proof:

$$
\begin{aligned}
P(E F^c) &= P(E) - P(EF) \\
&= P(E) - P(E)\,P(F) \\
&= P(E)\,[1 - P(F)] \\
&= P(E)\,P(F^c)
\end{aligned}
$$

So, E and F^c independent, implying that:

$P(E \mid F^c) = P(E) = P(E \mid F)$

- Intuitively, if E and F are independent, knowing whether F holds gives us no information about E

Generalized Independence

- General definition of Independence:

  Events E_1 , E_2 , ..., E_n are independent if for every subset with r elements (where $r \leq n$) it holds that:

$$P(E_{1'}E_{2'}E_{3'}\ldots E_{r'}) = P(E_{1'})P(E_{2'})P(E_{3'})\ldots P(E_{r'})$$

- Example: outcomes of *n* separate flips of a coin are all independent of one another
  - Each flip in this case is called a "trial" of the experiment

Two Dice

- Roll two 6-sided dice, yielding values D_1 and D_2
  - Let E be event: $D_1 = 1$
  - Let F be event: $D_2 = 6$
  - Are E and F independent? Yes!
- Let G be event: $D_1 + D_2 = 7$
  - Are E and G independent? Yes!
  - P(E) = 1/6, P(G) = 1/6, P(E G) = 1/36 [roll (1, 6)]
  - Are F and G independent? Yes!
  - P(F) = 1/6, P(G) = 1/6, P(F G) = 1/36 [roll (1, 6)]
  - Are E, F and G independent? No!
  - $P(EFG) = 1/36 \neq 1/216 = (1/6)(1/6)(1/6)$

Generating Random Bits

- A computer produces a series of random bits, with probability *p* of producing a 1.
  - Each bit generated is an independent trial
  - E = first *n* bits are 1's, followed by a single 0
  - What is P(E)?
- Solution
  - P(first *n* 1's) $= P(1^{st}\text{ bit}=1)\ P(2^{nd}\text{ bit}=1)\ \ldots\ P(n^{th}\text{ bit}=1)$
    $= p^n$
  - P(*n+1* bit=0) = (1 – p)
  - P(E) = P(first *n* 1's) P(*n+1* bit=0) = $p^n\,(1 - p)$

Coin Flips

- Say a coin comes up heads with probability p
  - Each coin flip is an independent trial

- P(*n* heads on *n* coin flips) = p^n
- P(*n* tails on *n* coin flips) = $(1 - p)^n$

- P(first *k* heads, then *n* – *k* tails) = $p^k(1-p)^{n-k}$

- P(exactly *k* heads on *n* coin flips) =?

Explain…

P(exactly *k* heads on *n* coin flips)? $\binom{n}{k} p^k (1-p)^{n-k}$

Think of the flips as ordered:

Ordering 1: T, H, H, T, T, T….

Ordering 2: H, T, H, T, T, T….

And so on…

The coin flips are independent!

$$P(F_i) = p^k (1-p)^{n-k}$$

Let's make each ordering with *k* heads an event… F_i

P(exactly *k* heads on *n* coin flips) = P(any one of the events)

P(exactly *k* heads on *n* coin flips) = P(F_1 or F_2 or F_3…)

Those events are mutually exclusive!

Moment of Crystallization

Add vs Multiply?

Batman vs Superman

Add vs Multiply

+ vs x

Add vs Multiply

Probabilities

multiply

$P(AB)$	Generally:	$P(A)P(B\|A)$
	Independent:	$P(A)P(B)$
$P(A \cup B)$	Generally:	$P(A) + P(B) - P(AB)$
	Mutually Exclusive:	$P(A) + P(B)$

add

Add vs Multiply

Counting

A then B	Generally:	*Not handled*
	Independent:	$|A||B|$
$A \cup B$	Generally:	$|A| + |B| - |AB|$
	Mutually Exclusive:	$|A| + |B|$

multiply

add

Next up…

And vs Condition

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

$$P(A\,B) = P(A \mid B)\,P(B)$$

Hash Tables

- *m* strings are hashed (unequally) into a hash table with *n* buckets
  - Each string hashed is an independent trial, with probability p_i of getting hashed to bucket *i*
  - E = at least one string hashed to first bucket
  - What is P(E)?
- Solution

To the chalk board!

Yet More Hash Tables

- m strings are hashed (unequally) into a hash table with n buckets
  - Each string hashed is an independent trial, with probability p_i of getting hashed to bucket i
  - E = At least 1 of buckets 1 to k has ≥ 1 string hashed to it
- Solution
  - F_i = at least one string hashed into i-th bucket
  - $P(E) = P(F_1 \cup F_2 \cup \ldots \cup F_k) = 1 - P((F_1 \cup F_2 \cup \ldots \cup F_k)^c)$

    $= 1 - P(F_1^c F_2^c \ldots F_k^c)$ (DeMorgan's Law)
  - $P(F_1^c F_2^c \ldots F_k^c)$ = P(no strings hashed to buckets 1 to k)

    $= (1 - p_1 - p_2 - \ldots - p_k)^m$
  - $P(E) = 1 - (1 - p_1 - p_2 - \ldots - p_k)^m$

No, Really, More Hash Tables

- *m* strings are hashed (unequally) into a hash table with *n* buckets
  - Each string hashed is an independent trial, with probability p_i of getting hashed to bucket *i*
  - E = Each of buckets 1 to *k* has ≥ 1 string hashed to it
- Solution
  - F_i = at least one string hashed into *i*-th bucket
  - $P(E) = P(F_1 F_2 \ldots F_k) = 1 - P((F_1 F_2 \ldots F_k)^c)$

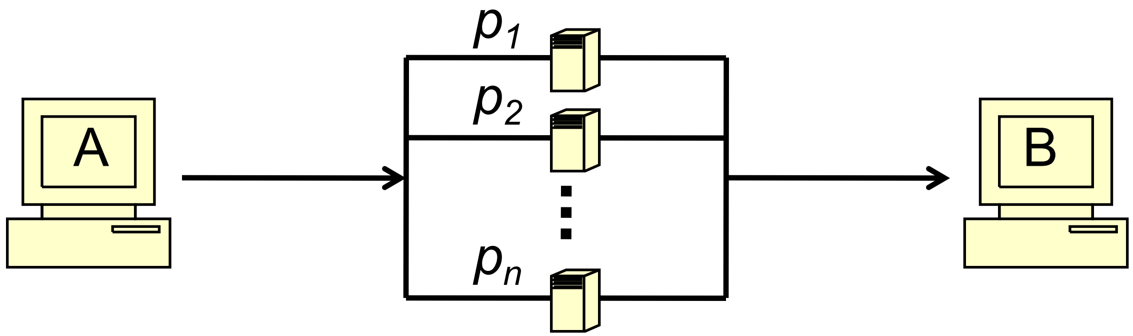
$$= 1 - P(F_1^c \cup F_2^c \cup \ldots \cup F_k^c) \qquad \text{(DeMorgan's Law)}$$

$$= 1 - P\left(\bigcup_{i=1}^{k} F_i^c\right) = 1 - \sum_{r=1}^{k} (-1)^{(r+1)} \sum_{i_1 < \ldots < i_r} P(F_{i_1}^{\ c} F_{i_2}^{\ c} \ldots F_{i_r}^{\ c})$$

where $P(F_{i_1}^{\ c} F_{i_2}^{\ c} \ldots F_{i_r}^{\ c}) = (1 - p_{i_1} - p_{i_2} - \ldots - p_{i_r})^m$

Sending a Message Through Network

- Consider the following parallel network:



  - n independent routers, each with probability p_i of functioning (where $1 \le i \le n$)
  - E = functional path from A to B exists. What is P(E)?
- Solution:
  - P(E) = 1 – P(all routers fail)

$$= 1 - (1 - p_1)(1 - p_2)\ldots(1 - p_n)$$

$$= 1 - \prod_{i=1}^{n}(1 - p_i)$$

Phew…

2 min pedagogical pause

Digging Deeper on Independence

- Recall, two events E and F are called independent if

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

- If E and F are independent, does that tell us whether the following is true or not:

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

  where G is an arbitrary event?

- In general, No!

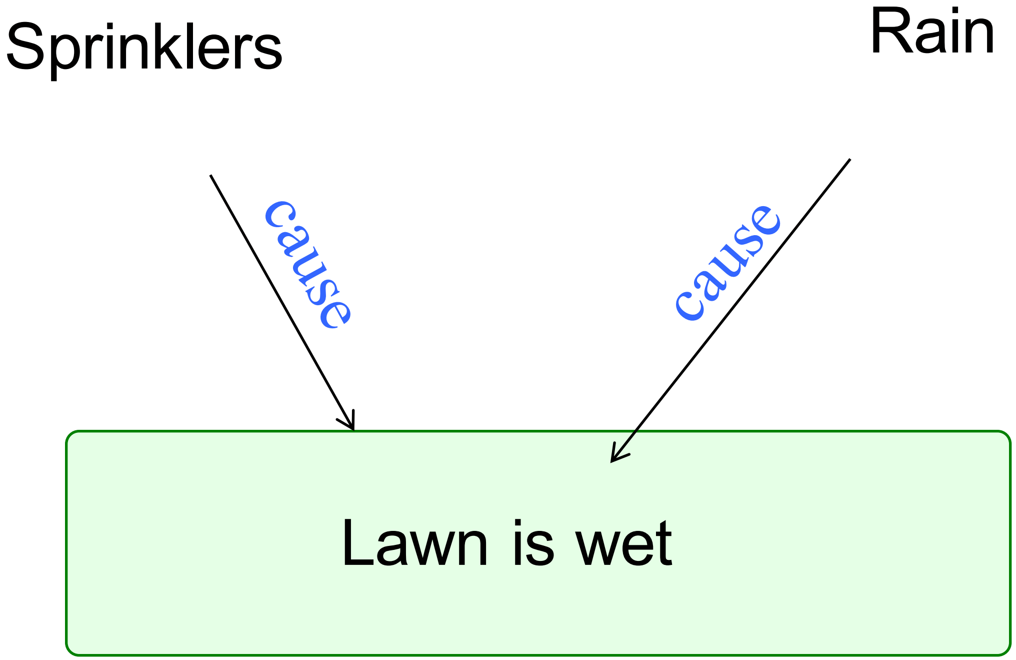
Not So Independent Dice

- Roll two 6-sided dice, yielding values D_1 and D_2
  - Let E be event: $D_1 = 1$
  - Let F be event: $D_2 = 6$
  - Let G be event: $D_1 + D_2 = 7$
- E and F are independent
  - $P(E) = 1/6$, $P(F) = 1/6$, $P(EF) = 1/36$
- Now condition both E and F on G:
  - $P(E|G) = 1/6$, $P(F|G) = 1/6$, $P(EF|G) = 1/6$
  - $P(EF|G) \neq P(E|G)\, P(F|G)$ → E|G and F|G *dependent*
- Independent events can become dependent by conditioning on additional information

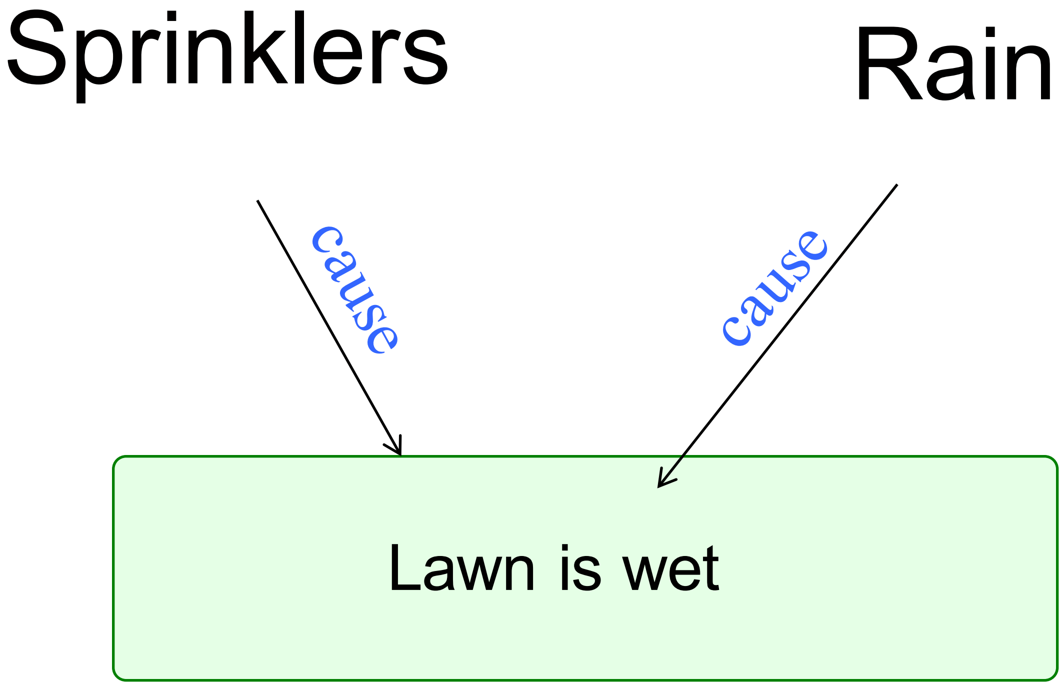
Explaining Away

- Say you have a lawn
  - It gets watered by rain or sprinklers
  - P(rain) and P(sprinklers were on) are independent
  - Now, you come outside and see the grass is wet
    - You know that the sprinklers were on
    - Does that lower probability that rain was cause of wet grass?
  - This phenomena is called "explaining away"
    - One cause of an observation makes other causes less likely

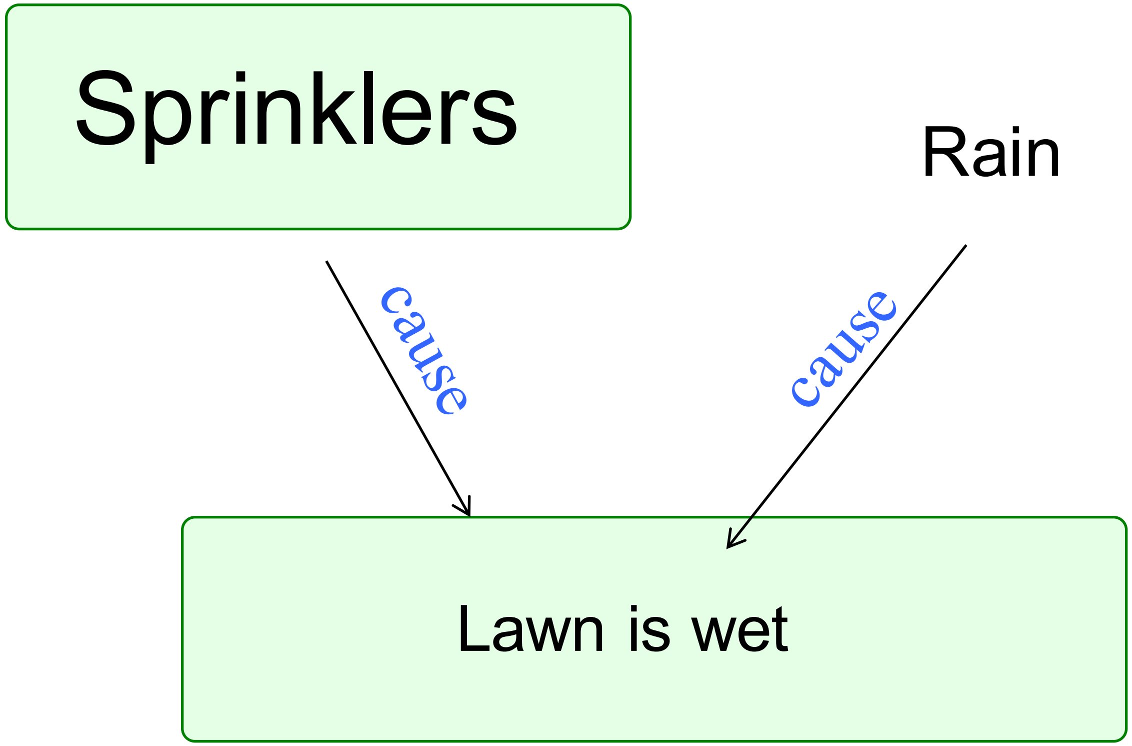
Explaining Away



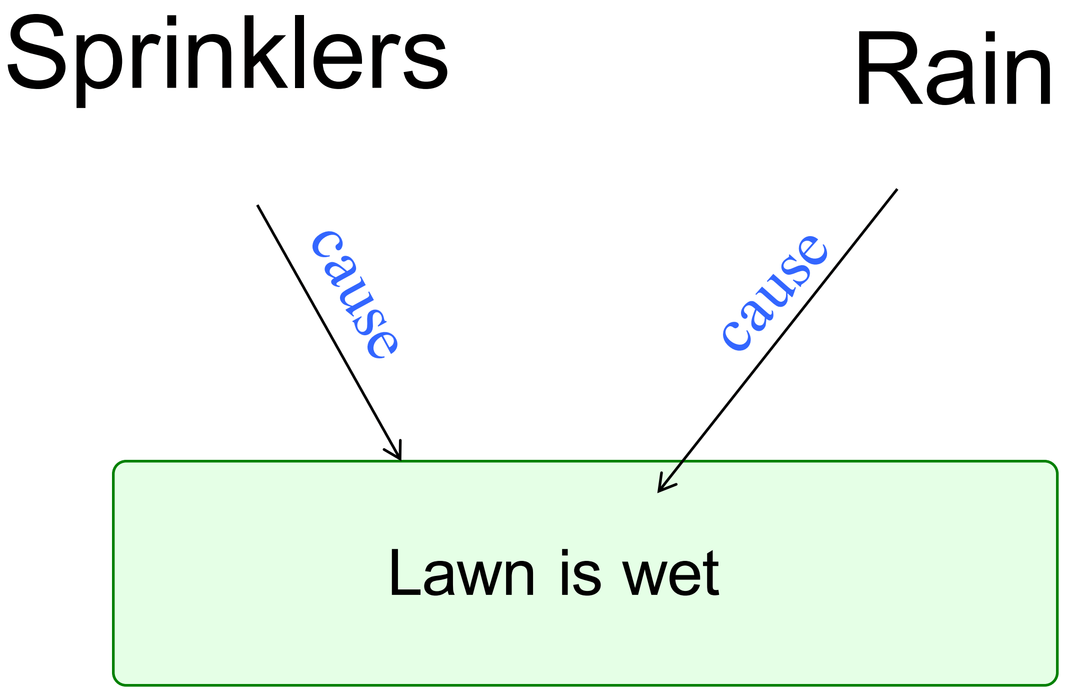
Explaining Away



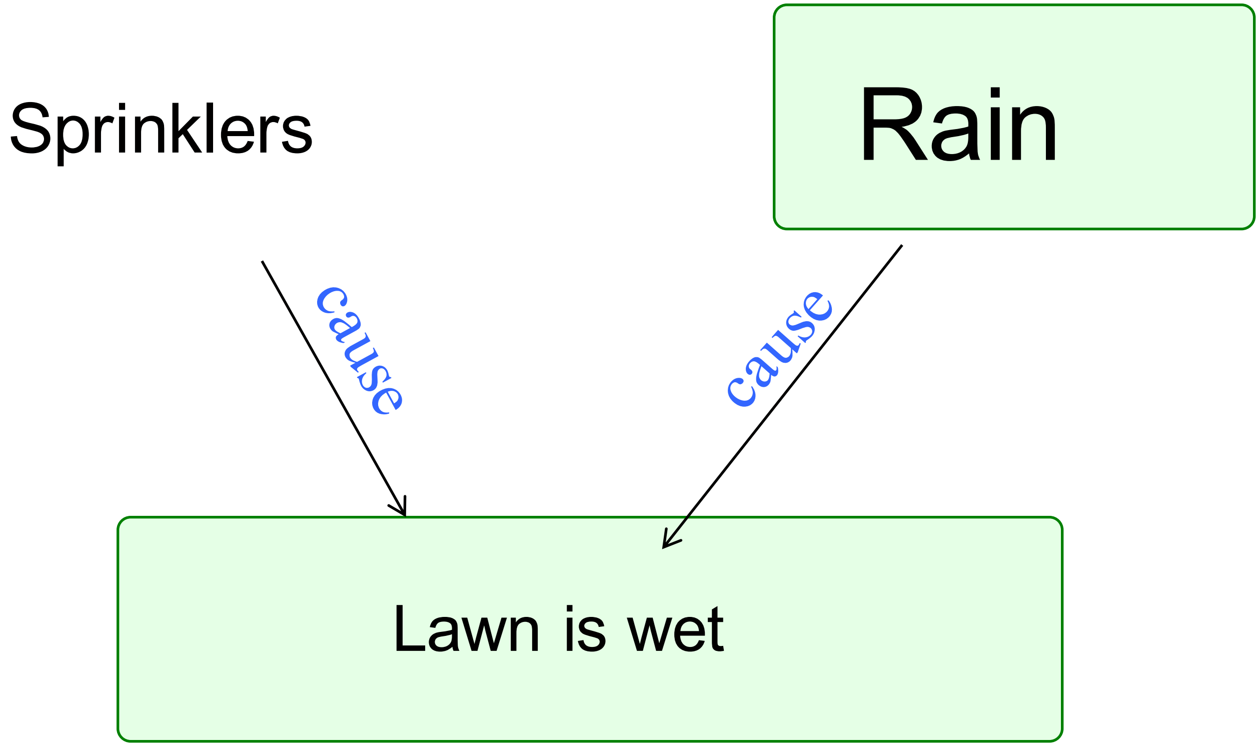
Explaining Away



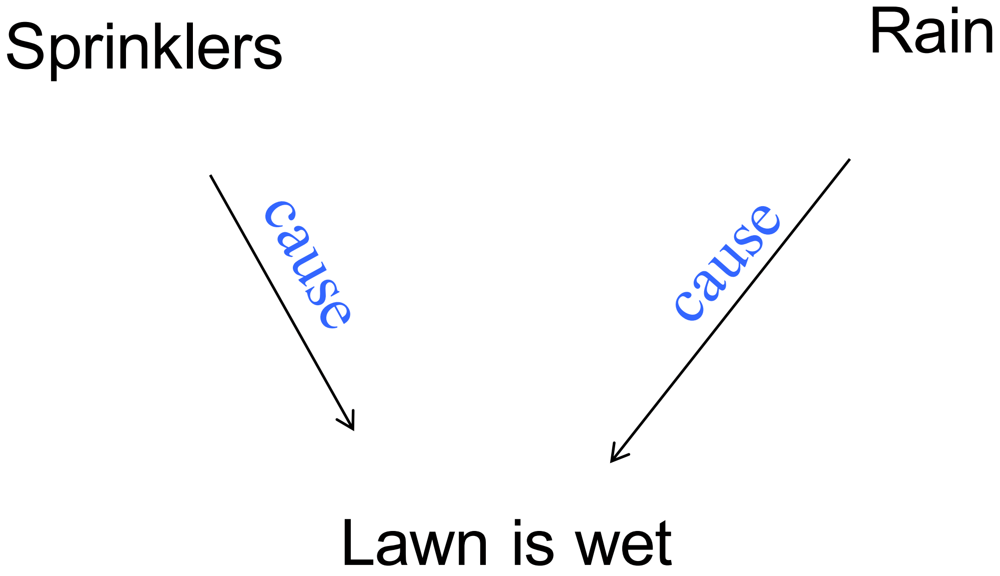
Explaining Away



Explaining Away



Explaining Away



Conditioning Can Make Independence

- Consider a randomly chosen day of the week
  - Let A be event: It is not Monday
  - Let B be event: It is Saturday
  - Let C be event: It is the weekend
- A and B are dependent
  - $P(A) = 6/7$, $P(B) = 1/7$, $P(AB) = 1/7 \neq (6/7)(1/7)$
- Now condition both A and B on C:
  - $P(A|C) = 1$, $P(B|C) = 1/2$, $P(AB|C) = 1/2$
  - $P(AB|C) = P(A|C)\, P(B|C) \rightarrow$ A|C and B|C *independent*
- Dependent events can become independent by conditioning on additional information

Conditional Independence

- Two events E and F are called **conditionally independent given G**, if

  $$P(E\,F \mid G) = P(E \mid G)\,P(F \mid G)$$

  Or, equivalently: $P(E \mid F\,G) = P(E \mid G)$

NETFLIX

And Learn

Netflix and Learn

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

$P(E)$

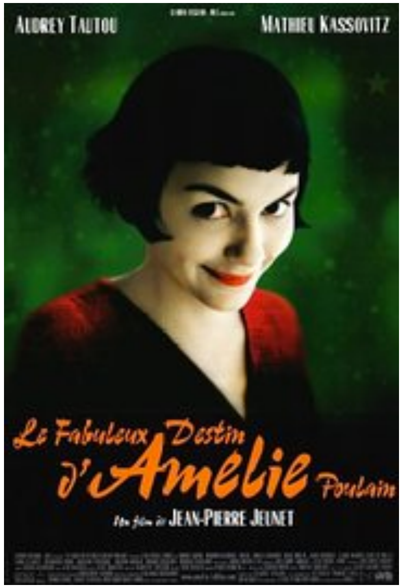
$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \approx \frac{\#\text{ people who liked movie}}{\#\text{ people who watched movie}}$$

$P(E) = 50{,}234{,}231 / 50{,}923{,}123 = 0.97$

Netflix and Learn

What is the probability that a user will like Life is Beautiful, given they liked Amelie?

$P(E|F)$



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

$$P(E|F) = 0.99$$

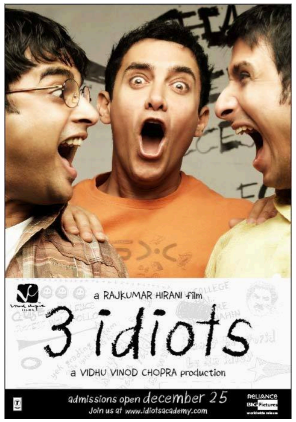
Conditioned on liking a set of movies?

Netflix and Learn

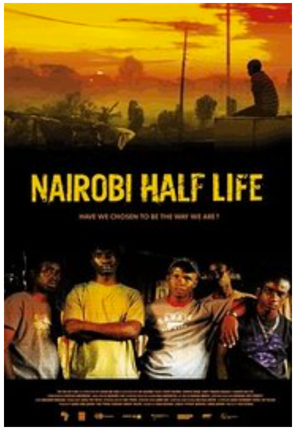
Each event corresponds to liking a particular movie

E_1

E_2



E_3



E_4

$$P(E_4 | E_1, E_2, E_3)?$$

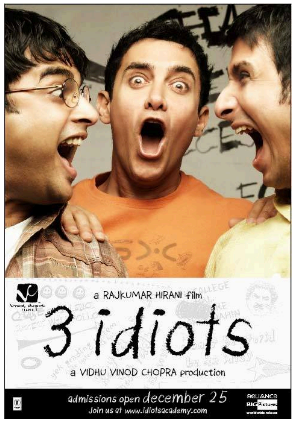
Is E_4 independent of E_1, E_2, E_3?

Netflix and Learn

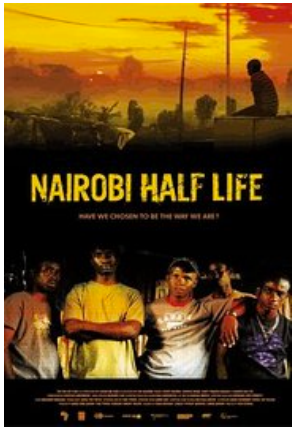
Is E_4 independent of E_1,E_2,E_3?

E_1

E_2



E_3



E_4

$$P(E_4 | E_1, E_2, E_3) \stackrel{?}{=} P(E_4)$$

Netflix and Learn

- What is the probability that a user watched four particular movies?
  - There are 13,000 titles on Netflix
  - The user watches 30 random titles
  - E = movies watched include the given four.
- Solution:

Watch those four

Choose 24 movies not in the set

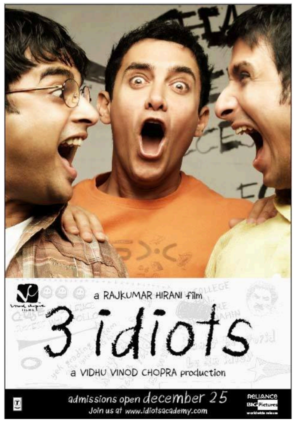
$$P(E) = \frac{\binom{4}{4}\binom{12996}{24}}{\binom{13000}{30}} = 10^{-11}$$

Choose 30 movies from netflix

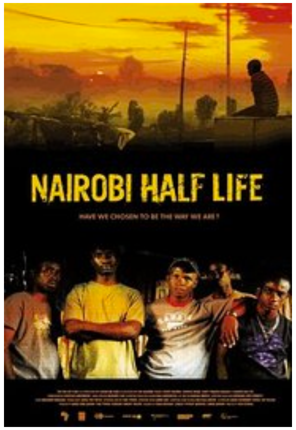
Netflix and Learn

E_1

E_2



E_3



E_4

Netflix and Learn

K_1

Like foreign emotional comedies

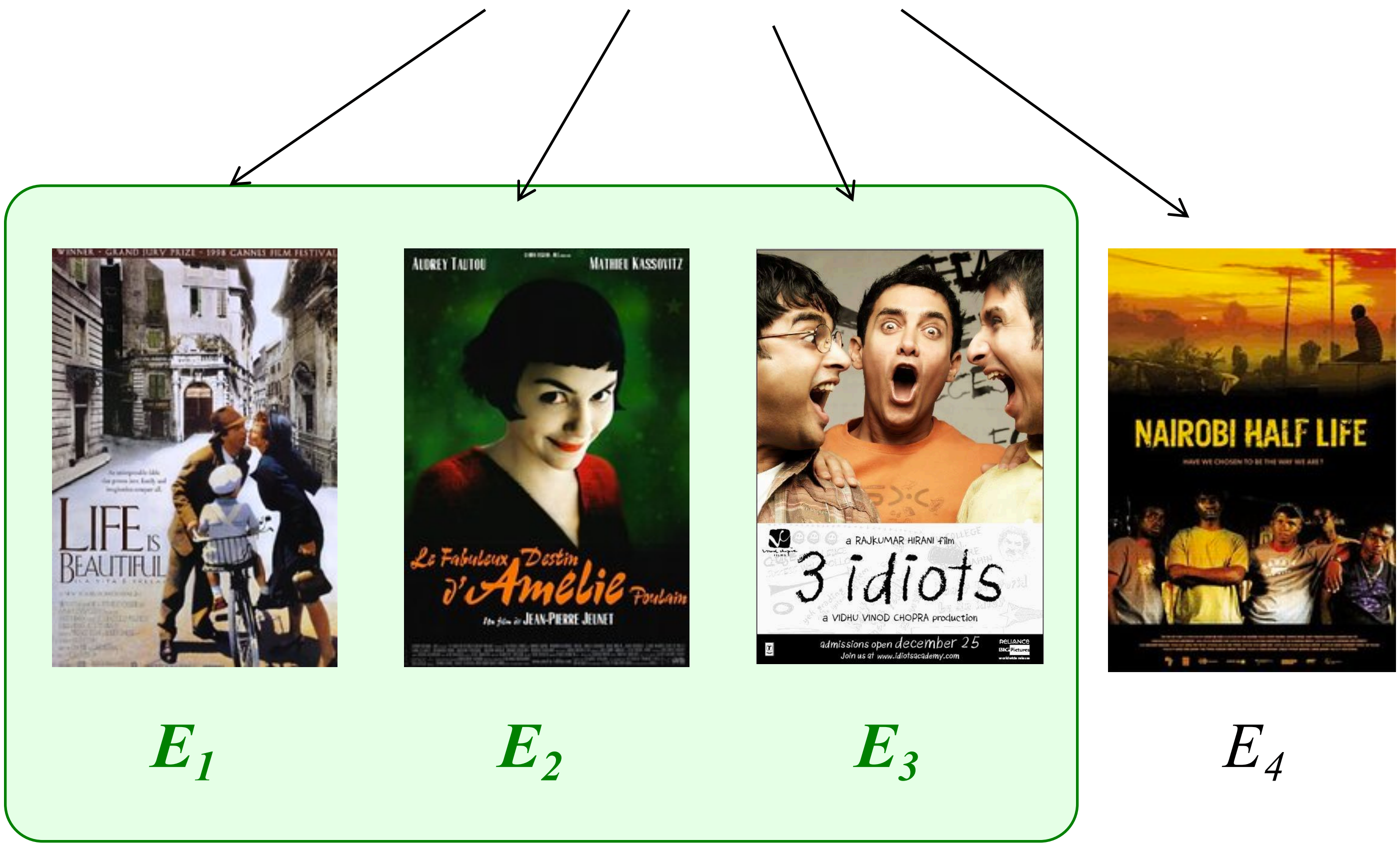
E_1 E_2 E_3 E_4

Assume E_1, E_2, E_3 and E_4 are conditionally independent given K_1

Netflix and Learn

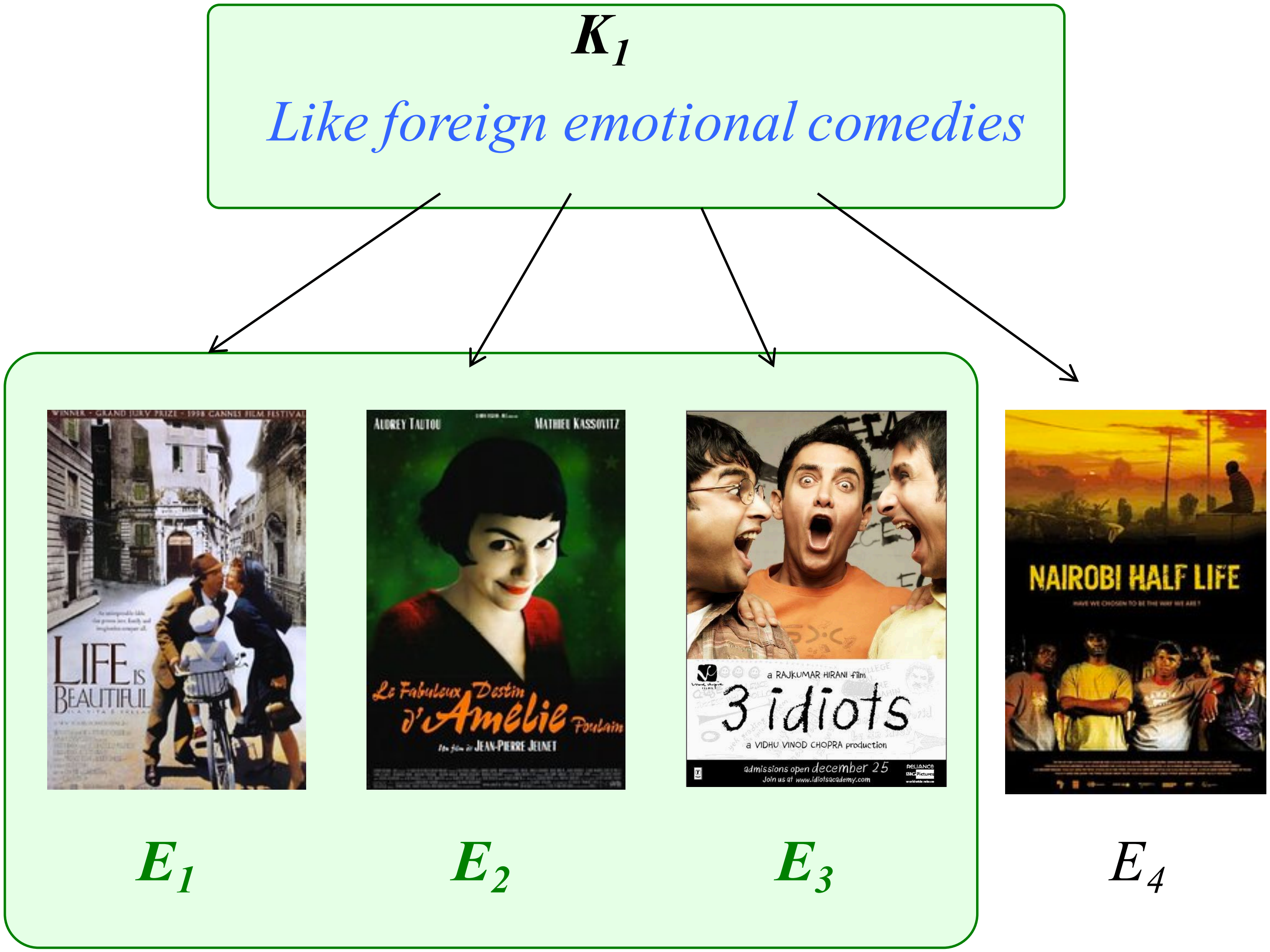
K_1

Like foreign emotional comedies



Assume E_1, E_2, E_3 and E_4 are conditionally independent given K_1

Netflix and Learn



Assume E_1, E_2, E_3 and E_4 are conditionally independent given K_1

Netflix and Learn

K_1

Like foreign emotional comedies

$$P(E_4 | E_1, E_2, E_3, K_1) = P(E_4 | K_1)$$

E_1 ***E_2*** ***E_3*** *E_4*

Assume E_1, E_2, E_3 and E_4 are conditionally independent given K_1

Conditional independence is a practical, real world way of decomposing hard probability questions.

Big Deal

"Exploiting conditional independence to generate fast probabilistic computations is one of the main contributions CS has made to probability theory"

-Judea Pearl wins 2011 Turing Award, *"For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning"*

Extra problem given time

Reminder: Geometric Series

- Geometric series: $x^0 + x^1 + x^2 + x^3 + \ldots + x^n = \sum_{i=0}^{n} x^i$
- If x is greater than 0 and less than 1:

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$$

- From your "Calculation Reference" handout:

Simplified Craps

- Two 6-sided dice repeatedly rolled (roll = ind. trial)
  - E = 5 is rolled before a 7 is rolled
  - What is P(E)?
- Solution
  - F_n = no 5 or 7 rolled in first n – 1 trials, 5 rolled on nth trial
  - P(E) = $P\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} P(F_n)$
  - P(5 on any trial) = 4/36 P(7 on any trial) = 6/36
  - $P(F_n) = (1 - (10/36))^{n-1}\,(4/36) = (26/36)^{n-1}\,(4/36)$
  - P(E) = $\frac{4}{36}\sum_{n=1}^{\infty}\left(\frac{26}{36}\right)^{n-1} = \frac{4}{36}\sum_{n=0}^{\infty}\left(\frac{26}{36}\right)^{n} = \frac{4}{36}\frac{1}{\left(1-\frac{26}{36}\right)} = \frac{2}{5}$