



Random Variables

Chris Piech

CS109, Stanford University

Let's Play a Game

- Game set-up
 - We have a fair coin (come up “heads” with $p = 0.5$)
 - Let n = number of coin flips (“heads”) before first “tails”
 - You win $\$2^n$
- How much would you pay to play?

Learning Goals

1. Be able to use conditional independence
2. Be able to define a random variable (R.V.)
3. Be able to use + produce a PMF of a R.V.
4. Be able to calculate the expectation of the R.V.



Conditional Paradigm

- Recall:

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

$$P(A \cap B) = P(A | B) P(B)$$



Conditional Paradigm

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

$$P(A \cap B | E) = P(B | A \cap E)$$

$$P(A \cap B | E) = P(A | B \cap E) P(B | E)$$

- Can think of E as “everything you already know”
- Formally, $P(\bullet | E)$ satisfies 3 axioms of probability



BAE's Theorem?

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





In the conditional paradigm, the formulas of probability are preserved.



Independence
relationships can change
with conditioning.

If E and F are independent, that does not mean they will still be independent given another event G.

There is additional reading about this in the course reader. You will explore this more in depth in CS228



Two Great Tastes

Conditional Probability

Independence



Conditional Independence

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

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

- Or, equivalently if:

$$P(E|FG) = P(E|G)$$



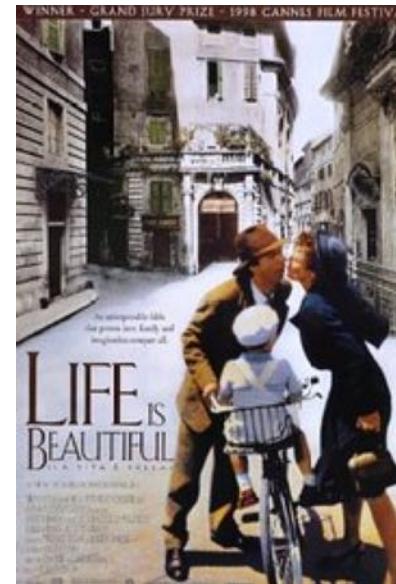
NETFLIX

And Learn

Netflix and Learn

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

$$P(E)$$



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

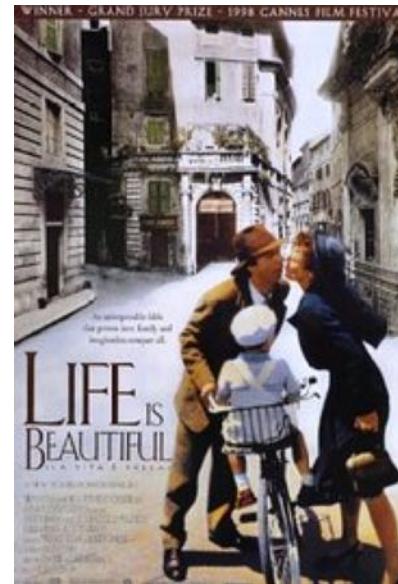
$$P(E) = 10,234,231 / 50,923,123 = 0.20$$



Netflix and Learn

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

$$P(E|F)$$



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

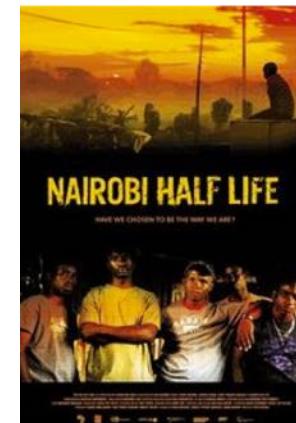
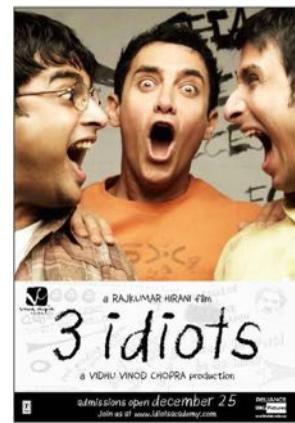
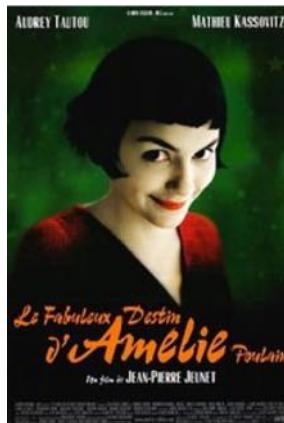
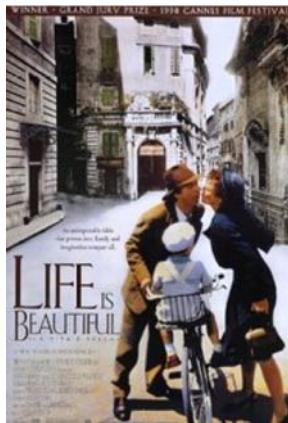
$$P(E|F) = 0.42$$



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

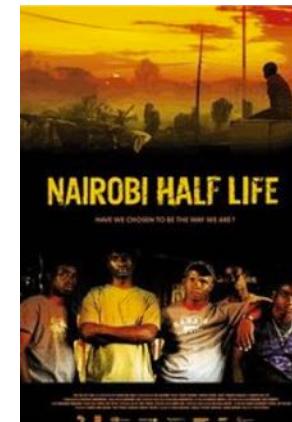
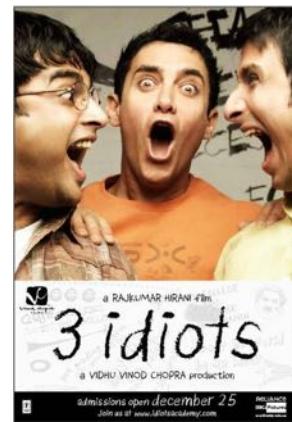
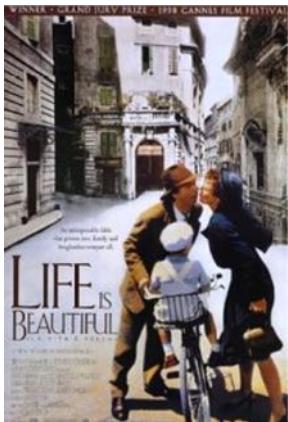
$$P(E_4 | E_1, E_2, E_3) ?$$



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

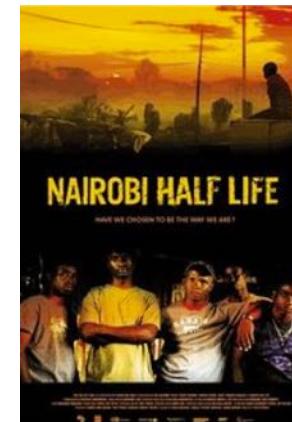
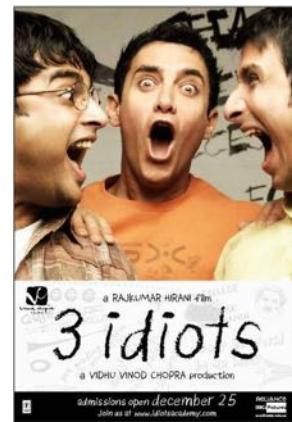
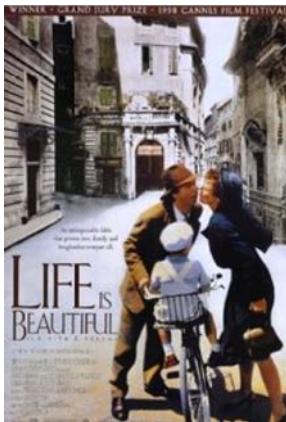
E_4

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



Netflix and Learn

Is E_4 independent of E_1, E_2, E_3 ?



E_1

E_2

E_3

E_4

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



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:

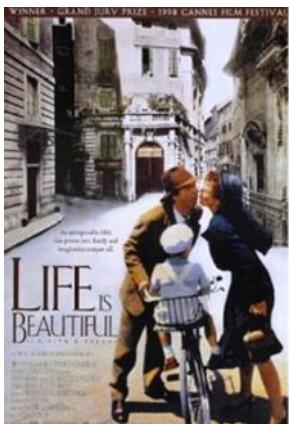
$$P(E) = \frac{\binom{4}{4} \binom{12996}{24}}{\binom{13000}{30}} = 10^{-11}$$

Watch those four *Choose 24 movies
not in the set*
*Choose 30 movies
from netflix*

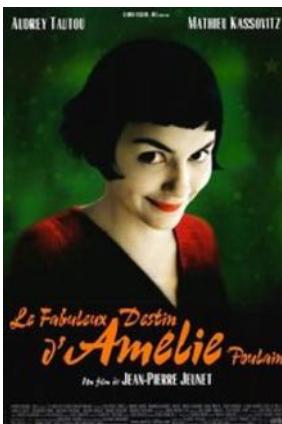


What can Netflix Do?

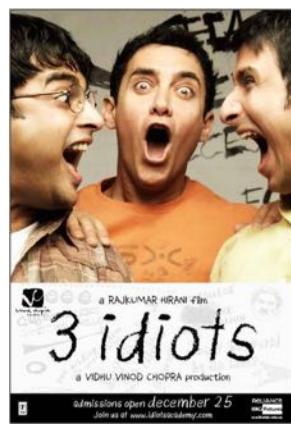
$$P(E_4|E_1, E_2, E_3) ?$$



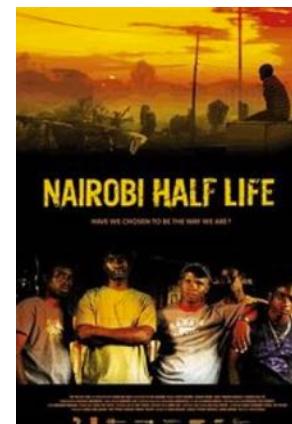
E_1



E_2



E_3



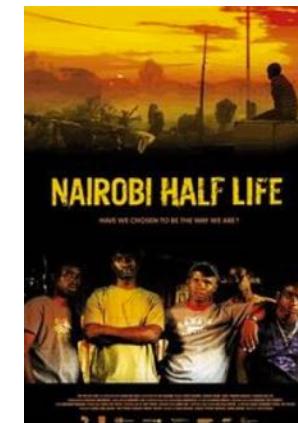
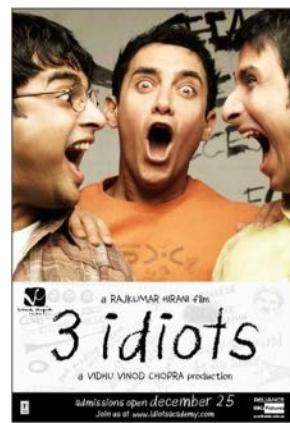
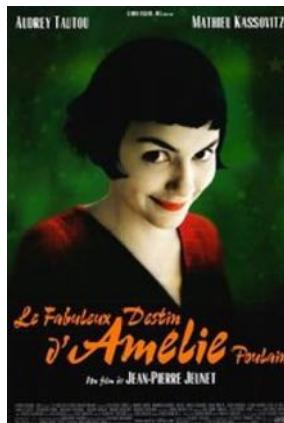
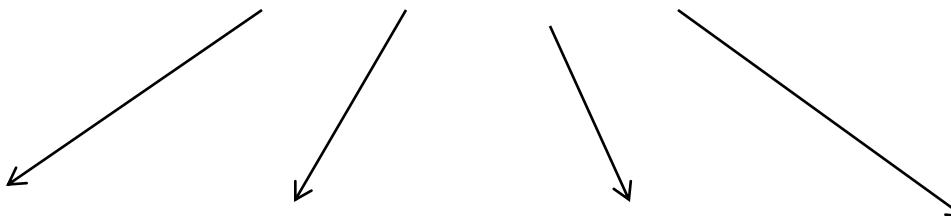
E_4



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



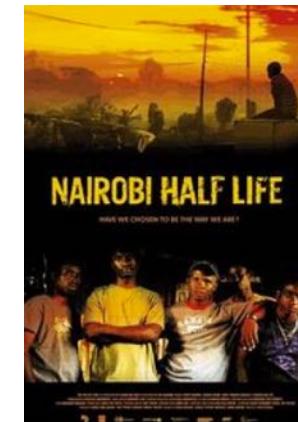
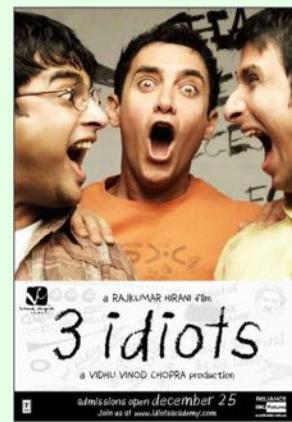
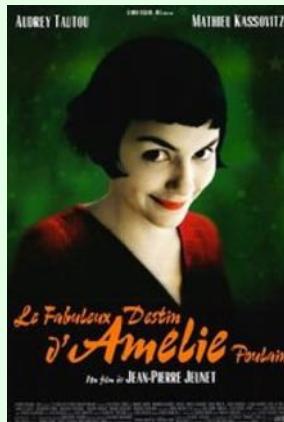
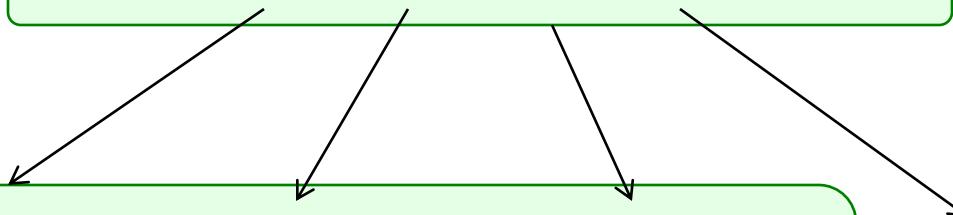
Assume E_1 , E_2 , E_3 and E_4 are conditionally independent given K_1



Netflix and Learn

K_1

Like foreign emotional comedies



E_1

E_2

E_3

E_4

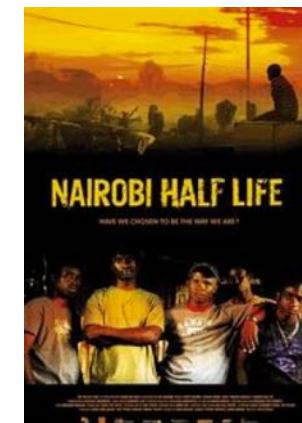
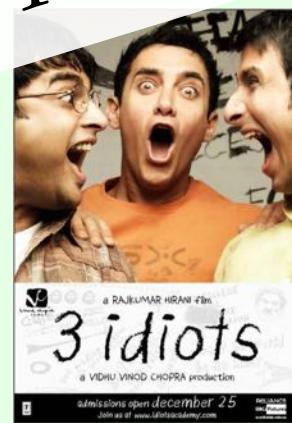
Assume E_1 , E_2 , E_3 and E_4 are conditionally independent given K_1



Netflix and Learn

K_1

Like foreign emotional comedies



E_1

E_2

E_3

E_4

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

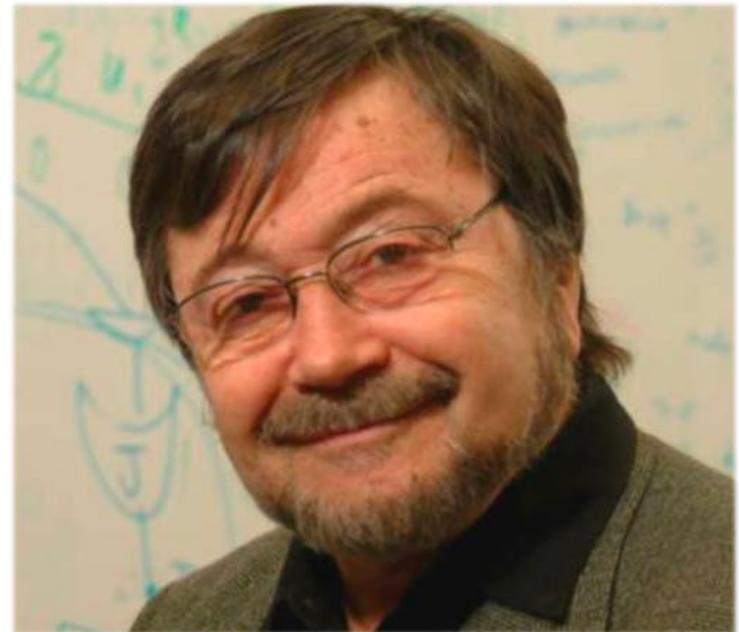


Conditional independence is a practical, real world way of decomposing hard probability questions.

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*”





G_1

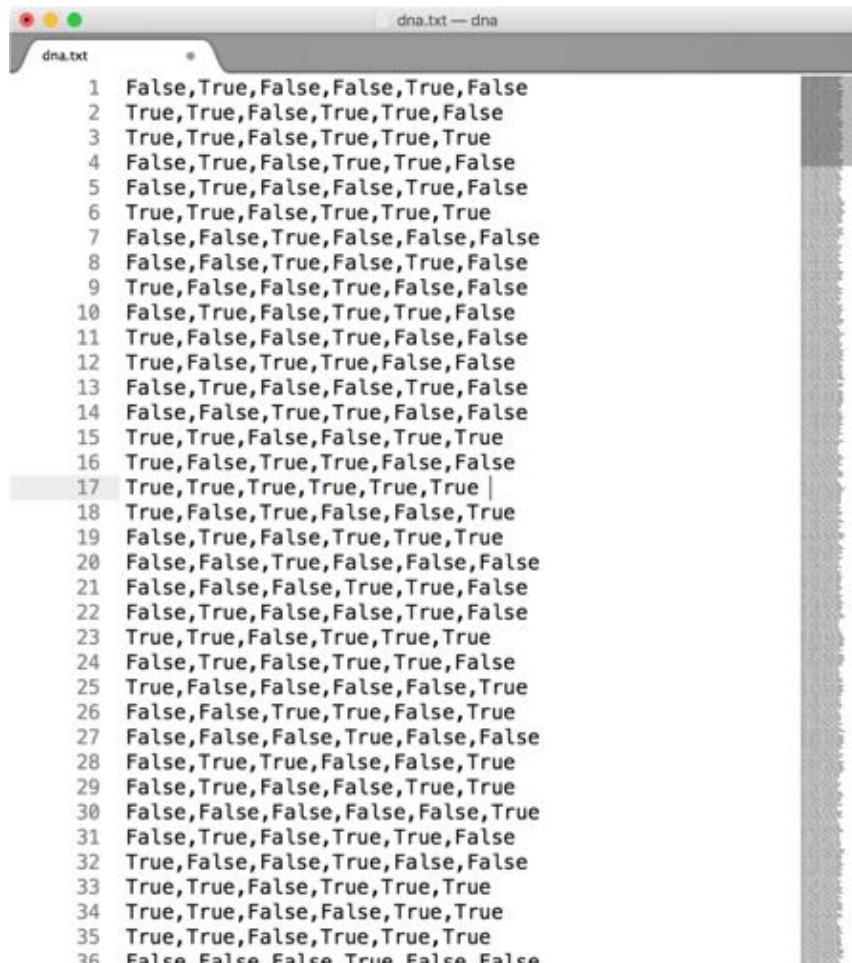
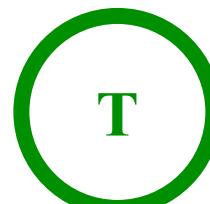
G_2

G_3

G_4

G_5

T



A screenshot of a Mac OS X window titled "dna.txt" in a "TextEdit" application. The window displays a list of 100,000 samples, each consisting of six binary observations (True or False). The first few lines of the text are:

```
1 False,True,False,False,True,False  
2 True,True,False,True,True,False  
3 True,True,False,True,True,True  
4 False,True,False,True,True,False  
5 False,True,False,False,True,False  
6 True,True,False,True,True,True  
7 False,False,True,False,False,False  
8 False,False,True,False,True,False  
9 True,False,False,True,False,False  
10 False,True,False,True,True,False  
11 True,False,False,True,False,False  
12 True,False,True,True,False,False  
13 False,True,False,False,True,False  
14 False,False,True,True,False,False  
15 True,True,False,False,True,True  
16 True,False,True,True,False,False  
17 True,True,True,True,True,True  
18 True,False,True,False,False,True  
19 False,True,False,True,True,True  
20 False,False,True,False,False,False  
21 False,False,False,True,True,False  
22 False,True,False,False,True,False  
23 True,True,False,True,True,True  
24 False,True,False,True,True,False  
25 True,False,False,False,False,True  
26 False,False,True,True,False,True  
27 False,False,False,True,False,False  
28 False,True,True,False,False,True  
29 False,True,False,False,True,True  
30 False,False,False,False,False,True  
31 False,True,False,True,True,False  
32 True,False,False,True,False,False  
33 True,True,False,True,True,True  
34 True,True,False,False,True,True  
35 True,True,False,True,True,True  
36 False,False,False,True,False,False
```

6 observations per sample

100,000
samples



Discovered Pattern

```
[Piech-2:dna piech$ python findStructure.py
size data = 100000
p(G1) = 0.500
p(G2) = 0.545
p(G3) = 0.299
p(G4) = 0.701
p(G5) = 0.600
p(T) = 0.390
p(T and G1) = 0.291 , P(T)p(G1) = 0.195
p(T and G2) = 0.300 , P(T)p(G2) = 0.213
p(T and G3) = 0.116 , P(T)p(G3) = 0.117
p(T and G4) = 0.273 , P(T)p(G4) = 0.273
p(T and G5) = 0.309 , P(T)p(G5) = 0.234
```

• • •

$$p(T \text{ and } G5 \mid G2) = 0.450$$
$$p(T \mid G2)p(G5 \mid G2) = 0.450$$



Discovered Pattern

```
[Piech-2:dna piech$ python findStructure.py
size data = 100000
p(G1) = 0.500
p(G2) = 0.545
p(G3) = 0.299
p(G4) = 0.701
p(G5) = 0.600
p(T) = 0.390
p(T and G1) = 0.291 , P(T)p(G1) = 0.195
p(T and G2) = 0.300 , P(T)p(G2) = 0.213
p(T and G3) = 0.116 , P(T)p(G3) = 0.117
p(T and G4) = 0.273 , P(T)p(G4) = 0.273
p(T and G5) = 0.309 , P(T)p(G5) = 0.234
```

• • •

$$p(T \text{ and } G5 | G2) = 0.450$$
$$p(T | G2)p(G5 | G2) = 0.450$$



Discovered Pattern

```
[Piech-2:dna piech$ python findStructure.py
size data = 100000
p(G1) = 0.500
p(G2) = 0.545
p(G3) = 0.299
p(G4) = 0.701
p(G5) = 0.600
p(T) = 0.390
p(T and G1) = 0.291 , P(T)p(G1) = 0.195
p(T and G2) = 0.300 , P(T)p(G2) = 0.213
p(T and G3) = 0.116 , P(T)p(G3) = 0.117
p(T and G4) = 0.273 , P(T)p(G4) = 0.273
p(T and G5) = 0.309 , P(T)p(G5) = 0.234
```

• • •

$$p(T \text{ and } G5 | G2) = 0.450$$
$$p(T | G2)p(G5 | G2) = 0.450$$



Discovered Pattern

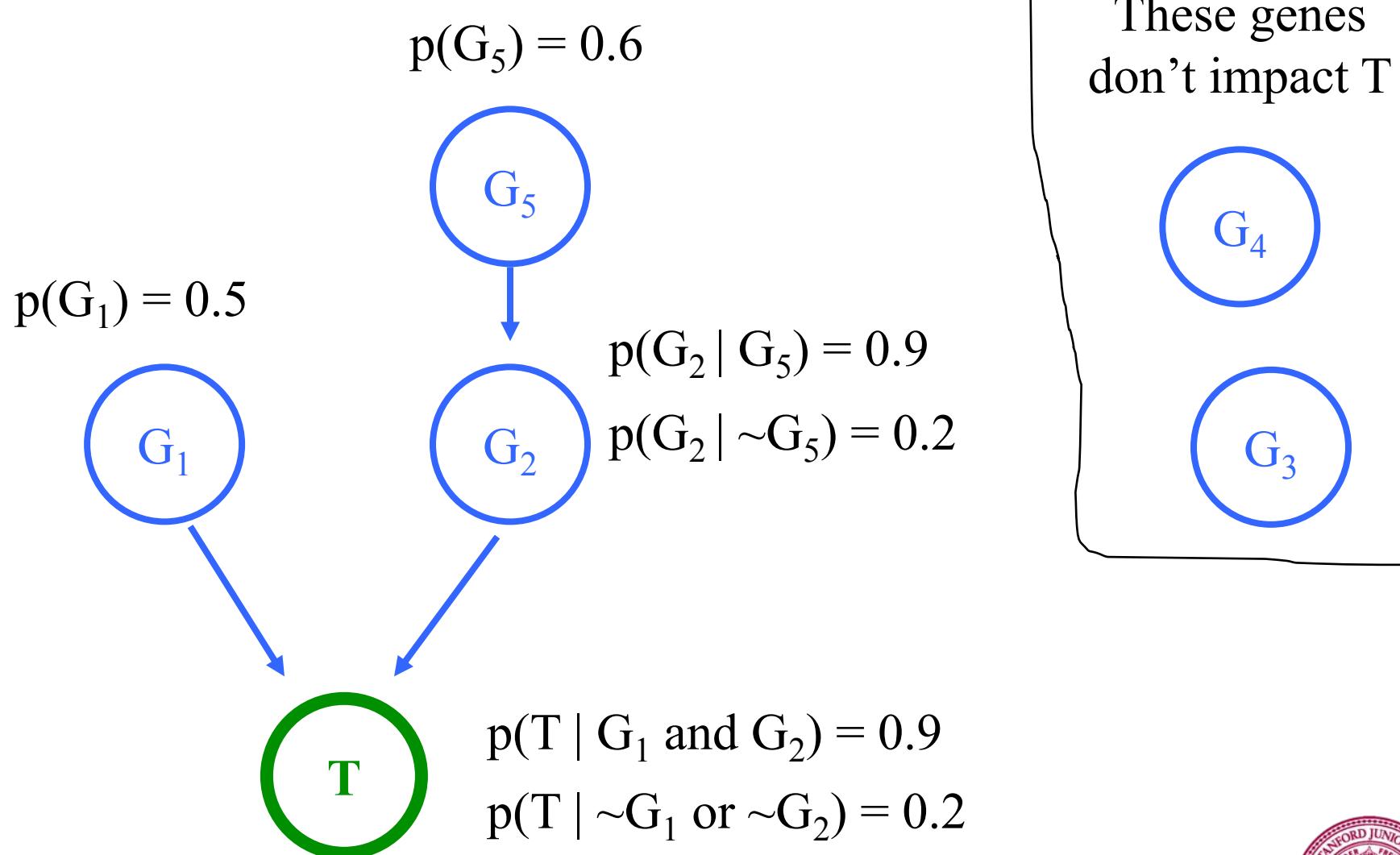
```
[Piech-2:dna piech$ python findStructure.py  
size data = 100000  
p(G1) = 0.500  
p(G2) = 0.545  
p(G3) = 0.299  
p(G4) = 0.701  
p(G5) = 0.600  
p(T) = 0.390  
p(T and G1) = 0.291 , P(T)p(G1) = 0.195  
p(T and G2) = 0.300 , P(T)p(G2) = 0.213  
p(T and G3) = 0.116 , P(T)p(G3) = 0.117  
p(T and G4) = 0.273 , P(T)p(G4) = 0.273  
p(T and G5) = 0.309 , P(T)p(G5) = 0.234
```

• • •

$p(T \text{ and } G5 | G2) = 0.450$
 $p(T | G2)p(G5 | G2) = 0.450$



Only Causal Structure that Fits



Ready for the next (cs109) episode

Random Variables

Remember Learning to Code?

name
value

type

```
int a = 5;  
double b = 4.2;  
bit c = 1;  
choice d = medium;
```

$$z \in \{\text{high}, \text{medium}, \text{low}\}$$

Random variables are like programming variables, with uncertainty

Pirates of the Random Variables

int a = 5;

A is the number of pirate ships in our *future* armada.

$$A \in \{1, 2, \dots, 10\}$$



double b = 4.2;

B is the amount of money we get *after* we defeat Blackbeard.

$$B \in \mathbb{R}^+$$



bit c = 1;

C is 1 *if* we successfully raid Isla de Muerta. 0 otherwise.

$$C \in \{0, 1\}$$



Random Variable

- A **Random Variable** is a real-valued function defined on a sample space
- Example:
 - 3 fair coins are flipped.
 - Y = number of “heads” on 3 coins
 - Y is a random variable
 - $P(Y = 0) = 1/8 \quad (T, T, T)$
 - $P(Y = 1) = 3/8 \quad (H, T, T), (T, H, T), (T, T, H)$
 - $P(Y = 2) = 3/8 \quad (H, H, T), (H, T, H), (T, H, H)$
 - $P(Y = 3) = 1/8 \quad (H, H, H)$
 - $P(Y \geq 4) = 0$

It is confusing that both random variables
and events use the same notation



Random variables and
events are two *different*
things





We can define an event to
be a particular assignment
to a random variables

Example Random Variable

- Consider 5 coin flips, each which independently come up heads with probability p

- Recall:

$$P(2 \text{ heads}) = \binom{5}{2} p^2 (1-p)^3$$

$$P(3 \text{ heads}) = \binom{5}{3} p^3 (1-p)^2$$

- Y = number of “heads” on 5 flips

$$Y \in \{1, 2, \dots, 5\}$$

$$P(Y = k) = \binom{5}{k} p^k (1-p)^{5-k}$$

* Pro tip: no coin works like this... but many real world binary events do

Fun with Random Variables

- Probability Mass Function:

$$P(X = a)$$

- Expectation:

$$E[X]$$

- Variance:

$$\text{Var}(X)$$



Learning
goals for
today

1. Probability Mass Function

All the different assignments to a random variable make a function

If this is a number

$$P(Y = 2)$$

Then this is a number

For example Y is the number of heads in 5 coin flips

If this is a variable

$$P(Y = k)$$

Then this is a function

For example Y is the number of heads in 5 coin flips

Random Variables -> Functions

$$P(Y = k) \xrightarrow{k = 5} 0.03125$$

For example Y is the number of heads in 5 coin flips

Random Variables -> Functions

$$P(Y = k)$$

```
private double eventProbability(int k) {  
    int ways = choose(N, k);  
    double a = Math.pow(P, k);  
    double b = Math.pow(P, N-k);  
    return ways * a * b;  
}  
  
private static final int N = 5;  
private static final double P = 0.5;
```

For example Y is the number of heads in 5 coin flips



If a random variable is discrete we call this function
the **Probability Mass**
Function



Probability Mass Function

Let X be a random variable that represents the result of a **single dice roll**. X can take on the values $\{1, 2, 3, 4, 5, 6\}$

$$P(X = x)$$

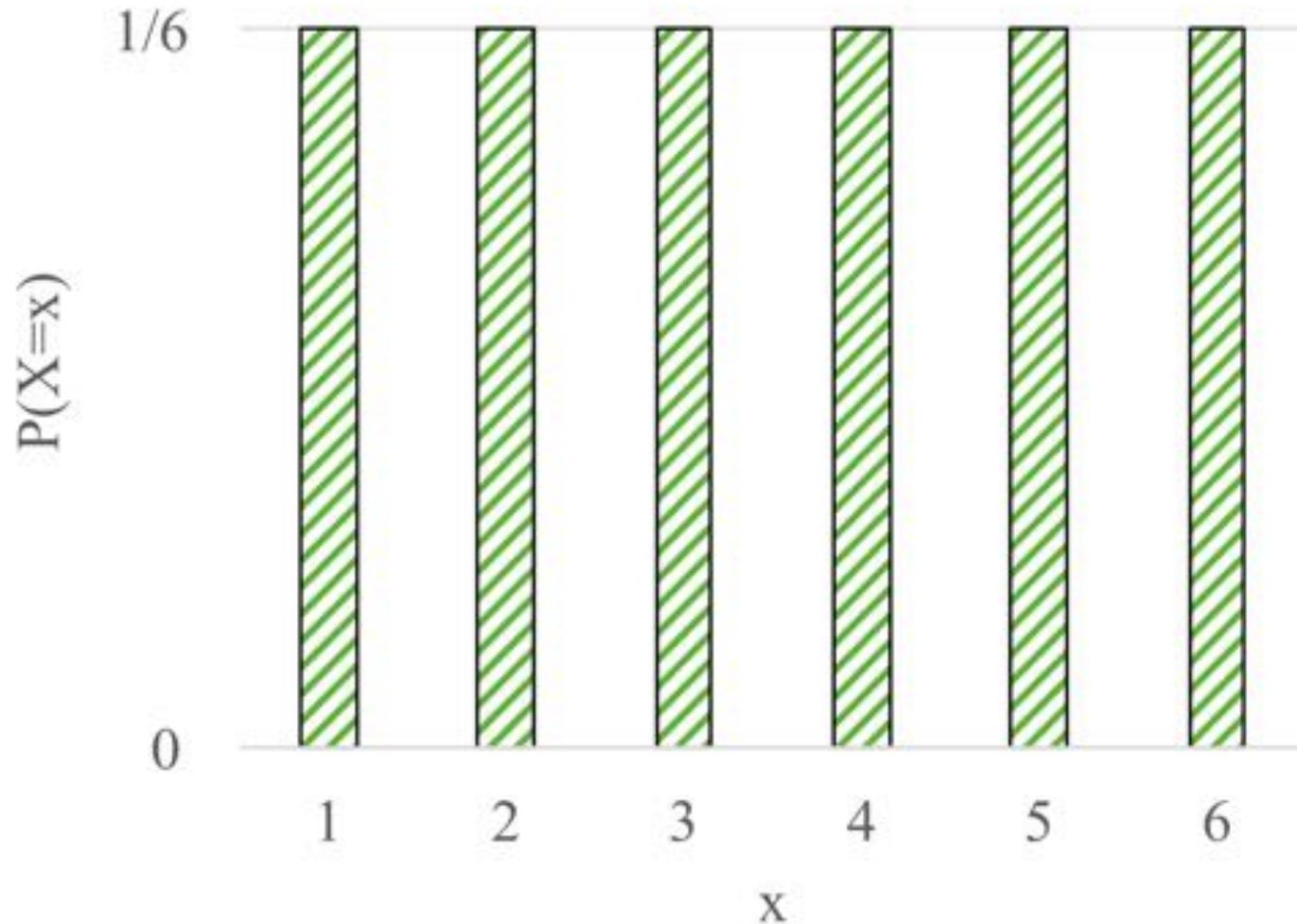
$$p(x)$$

This is shorthand notation for the PMF

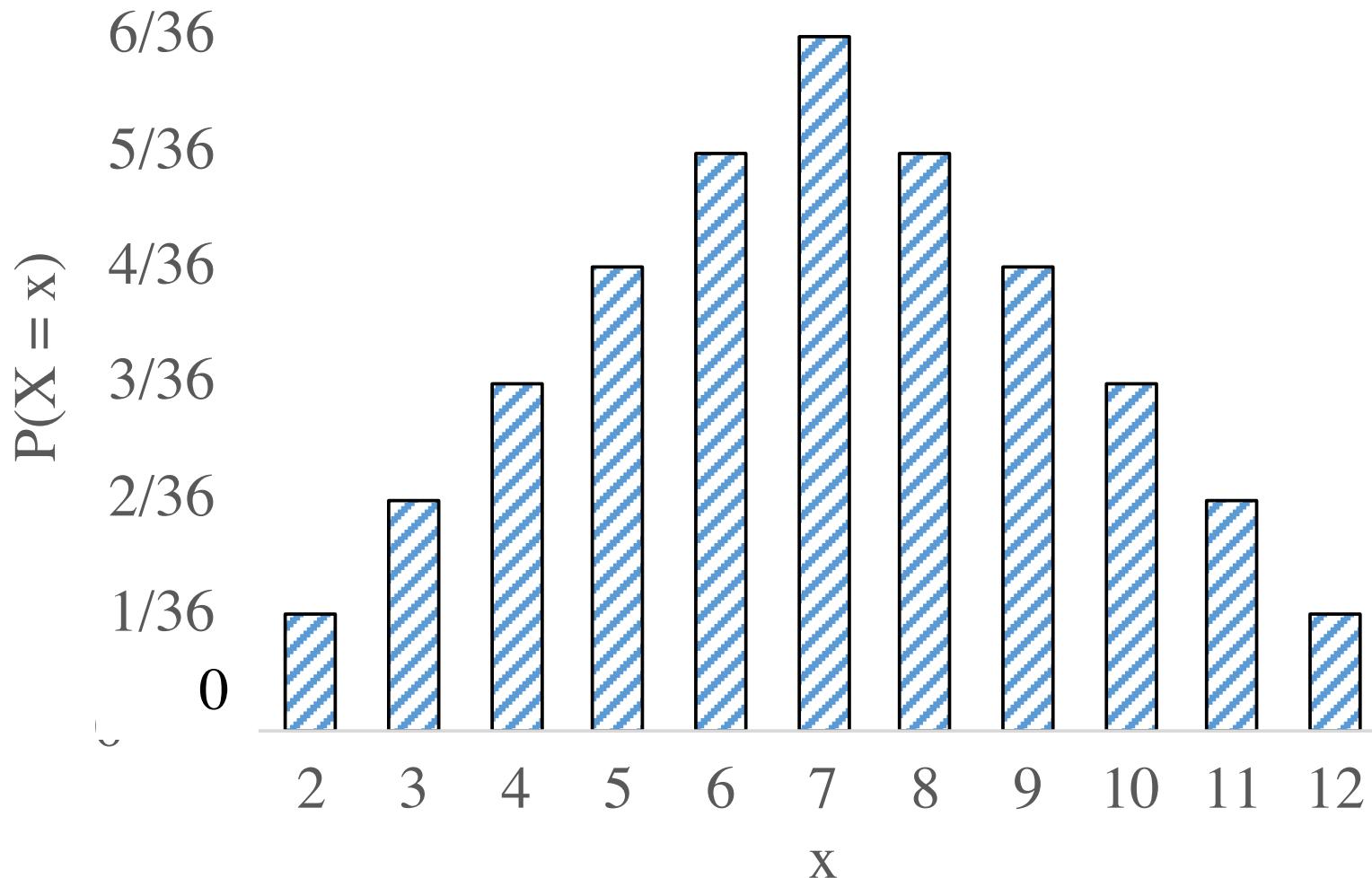
$$p_X(x)$$

This is also shorthand notation for the PMF

PMF For a Single 6 Sided Dice



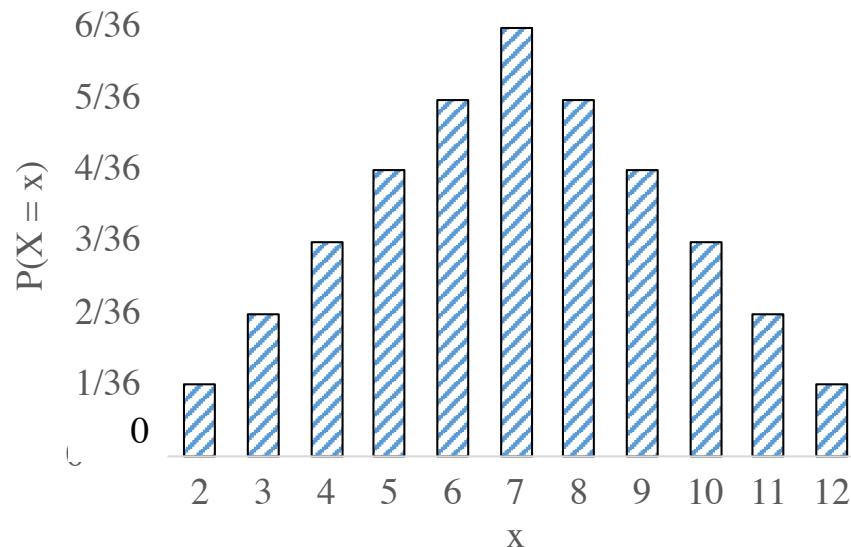
PMF for the sum of two dice



PMF as an Equation

$$P(X = x) = \begin{cases} \frac{x}{36} & \text{if } x \in \mathbb{R}, 0 \leq x \leq 6 \\ \frac{12-x}{36} & \text{if } x \in \mathbb{R}, x \leq 7 \\ 0 & \text{else} \end{cases}$$

Again, this is the probability for the sum of two dice



2. Expectation

Expected Value

- The Expected Values for a discrete random variable X is defined as:

$$E[X] = \sum_{x:p(x)>0} x \cdot p(x)$$

- Note: sum over all values of x that have $p(x) > 0$.
- Expected value also called: *Mean, Expectation, Weighted Average, Center of Mass, 1st Moment*

Expected Value

- Roll a 6-Sided Die. X is outcome of roll
 - $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6$
- $E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$
- Y is random variable
 - $P(Y = 1) = 1/3, P(Y = 2) = 1/6, P(Y = 3) = 1/2$
- $E[Y] = 1 (1/3) + 2 (1/6) + 3 (1/2) = 13/6$

Lying with Statistics

“There are three kinds of lies:
lies, damned lies, and statistics”

– *Mark Twain*

- School has 3 classes with 5, 10 and 150 students
- Randomly choose a class with equal probability
- X = size of chosen class
- What is $E[X]$?
 - $$\begin{aligned} E[X] &= 5(1/3) + 10(1/3) + 150(1/3) \\ &= 165/3 = 55 \end{aligned}$$

Lying with Statistics

“There are three kinds of lies:
lies, damned lies, and statistics”

– *Mark Twain*

- School has 3 classes with 5, 10 and 150 students
- Randomly choose a student with equal probability
- Y = size of class that student is in
- What is $E[Y]$?
 - $$\begin{aligned} E[Y] &= 5(5/165) + 10(10/165) + 150(150/165) \\ &= 22635/165 \approx 137 \end{aligned}$$
- Note: $E[Y]$ is students' perception of class size
 - But $E[X]$ is what is usually reported by schools!

Properties of Expectation

- **Linearity:**

$$E[aX + b] = aE[X] + b$$

- Consider X = 6-sided die roll, $Y = 2X - 1$.
- $E[X] = 3.5$ $E[Y] = 6$

- **Expectation of a sum** is the sum of expectations

$$E[X + Y] = E[X] + E[Y]$$

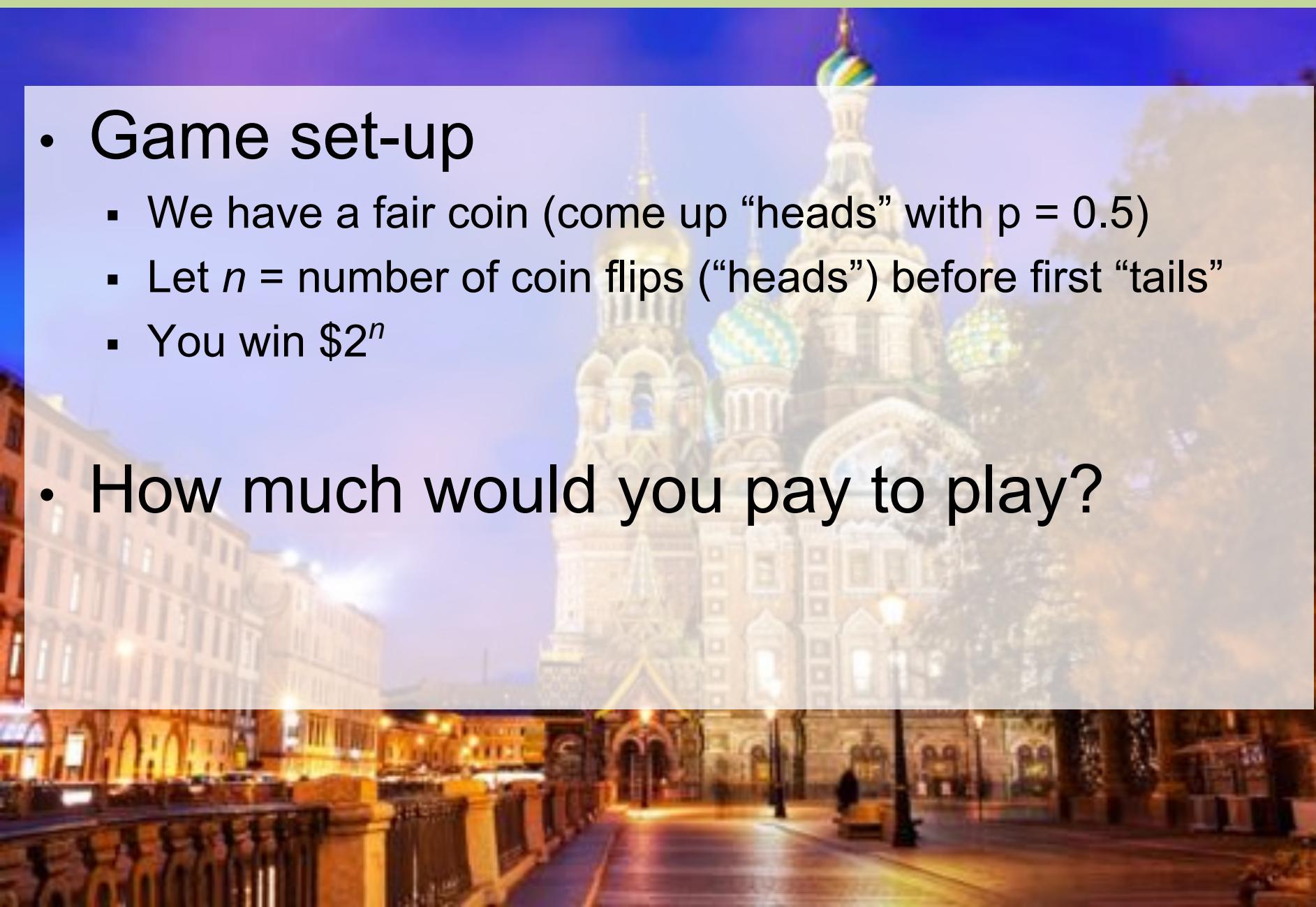
- **Unconscious statistician:**

$$E[g(x)] = \sum_x g(x)p(x)$$

Wonderful

St Petersburg

- Game set-up
 - We have a fair coin (come up “heads” with $p = 0.5$)
 - Let n = number of coin flips (“heads”) before first “tails”
 - You win $\$2^n$
- How much would you pay to play?



St Petersburg

- Game set-up
 - We have a fair coin (come up “heads” with $p = 0.5$)
 - Let n = number of coin flips (“heads”) before first “tails”
 - You win $\$2^n$
- How much would you pay to play?
- Solution
 - Let X = your winnings
 - $E[X] = \left(\frac{1}{2}\right)^1 2^0 + \left(\frac{1}{2}\right)^2 2^1 + \left(\frac{1}{2}\right)^3 2^2 + \left(\frac{1}{2}\right)^4 2^3 + \dots = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{i+1} 2^i$
 $= \sum_{i=0}^{\infty} \frac{1}{2} = \infty$
 - I'll let you play for \$1 thousand... but just once! Takers?

St Petersburg + Reality

- What if Chris has only \$65,536?
 - Same game
 - If you win over \$65,536 I leave the country.
- Solution
 - Let X = your winnings
 - $$\begin{aligned} E[X] &= \left(\frac{1}{2}\right)^1 2^0 + \left(\frac{1}{2}\right)^2 2^1 + \left(\frac{1}{2}\right)^3 2^2 + \left(\frac{1}{2}\right)^4 2^3 + \dots \\ &= \sum_{i=0}^k \left(\frac{1}{2}\right)^{i+1} 2^i \text{ s.t. } k = \log_2(65,536) \\ &= \sum_{i=0}^{16} \frac{1}{2} = 8.5 \end{aligned}$$

Learning Goals

1. Be able to use conditional independence
2. Be able to define a random variable (R.V.)
3. Be able to use + produce a PMF of a R.V.
4. Be able to calculate the expectation of the R.V.

