Random Variables

CS109, Stanford University

Announcements

- Pset #2 will be out today. Pset #1 due next Monday.
- Will will begin his lectures next Monday

Problem Set #2

PS2

1 2a 2b 3 4a 4b 5 6 7 8 9 10 11

Medical Test

Write a function:

```
def predict_positive_given_test_result(
    prior_disease,
    p_true_given_disease,
    p_true_given_no_disease,
    test_result):
```

That can be used for any noisy (binary) medical test, such as a Covid-19 test, or an Ebola test. Your function takes in a prior belief that a patient has a disease, statistics on a noisy test, and the test result from the noisy test. Based off this information, you should compute the probability that the patient is "positive" for the disease (in other words, they have the disease). Your return value must be a number between 0 and 1, not a boolean prediction. This problem requires you to code up a general implementation of Bayes' Theorem for a binary prediction!

Hint: you might find it helpful to read the medical example from the [Bayes Theorem] chapter.

Answer Editor | Solution

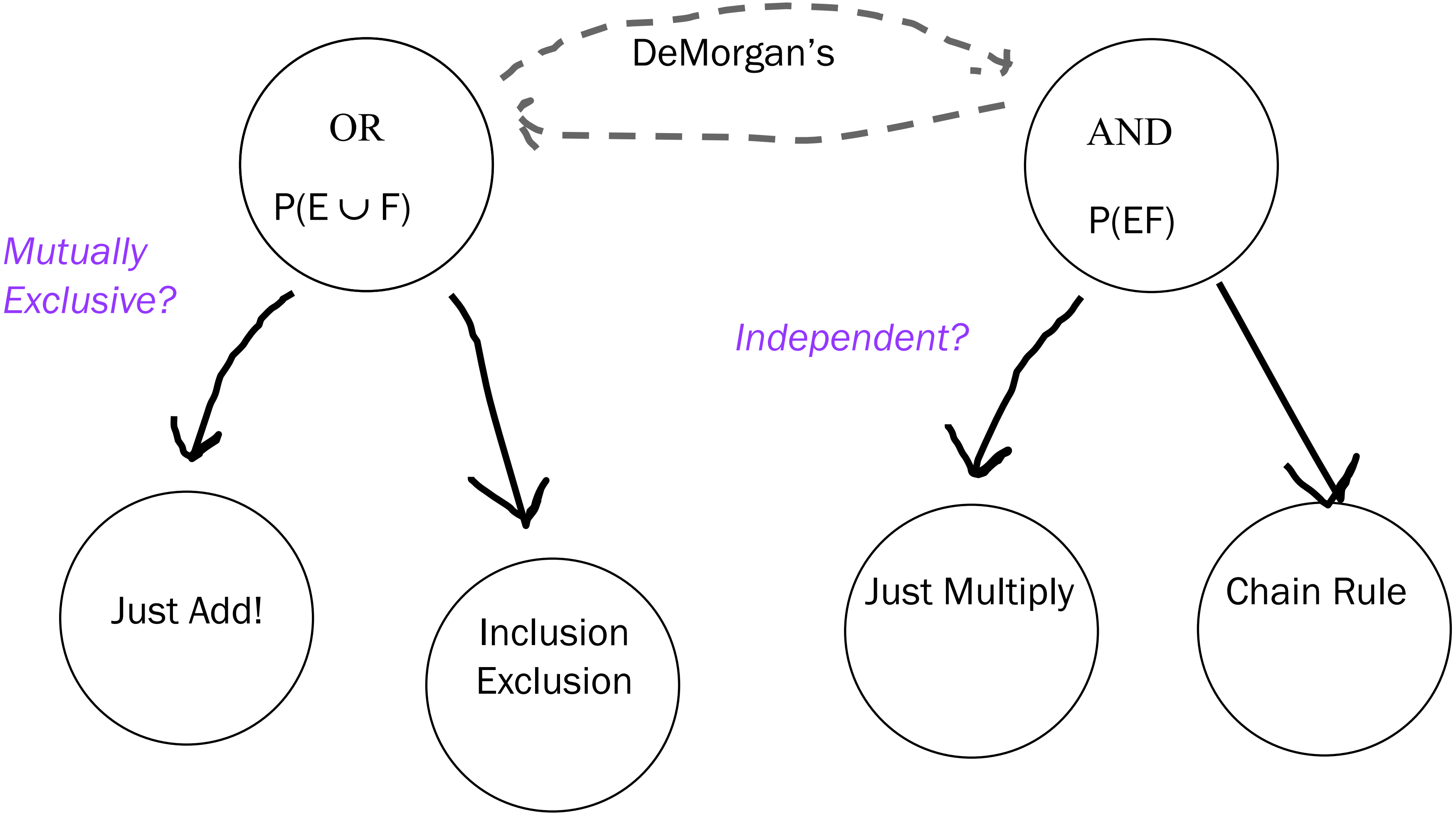
Agent:

```
def predict_positive_given_test_result(
        prior_disease,           # prior prob that the patient has the disease
        p_true_given_disease,    # the "true positive" probability
        p_true_given_no_disease, # the "false positive" probability
        test_result):            # True/False test result
    # TODO: your code here
    return 0.5
```

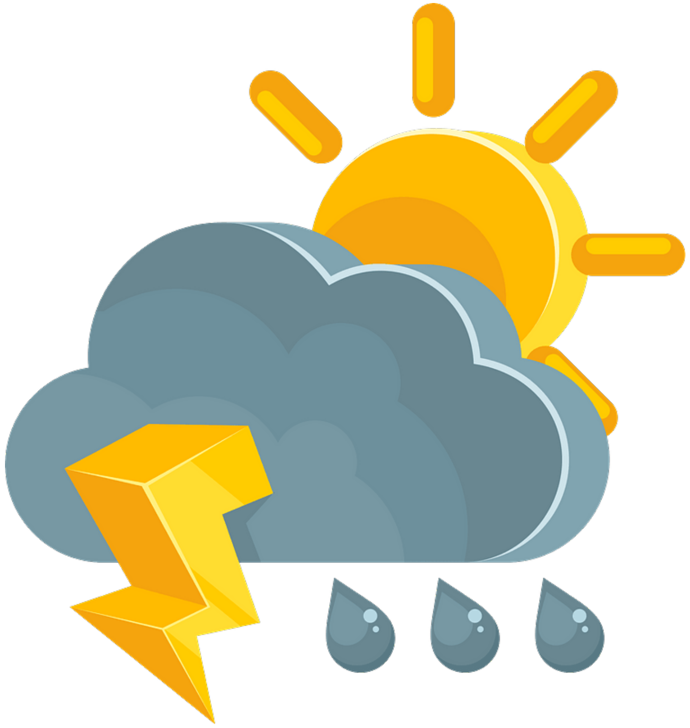
Run One Game | Test Agent

Review!

3. Review: Probability of Or and And



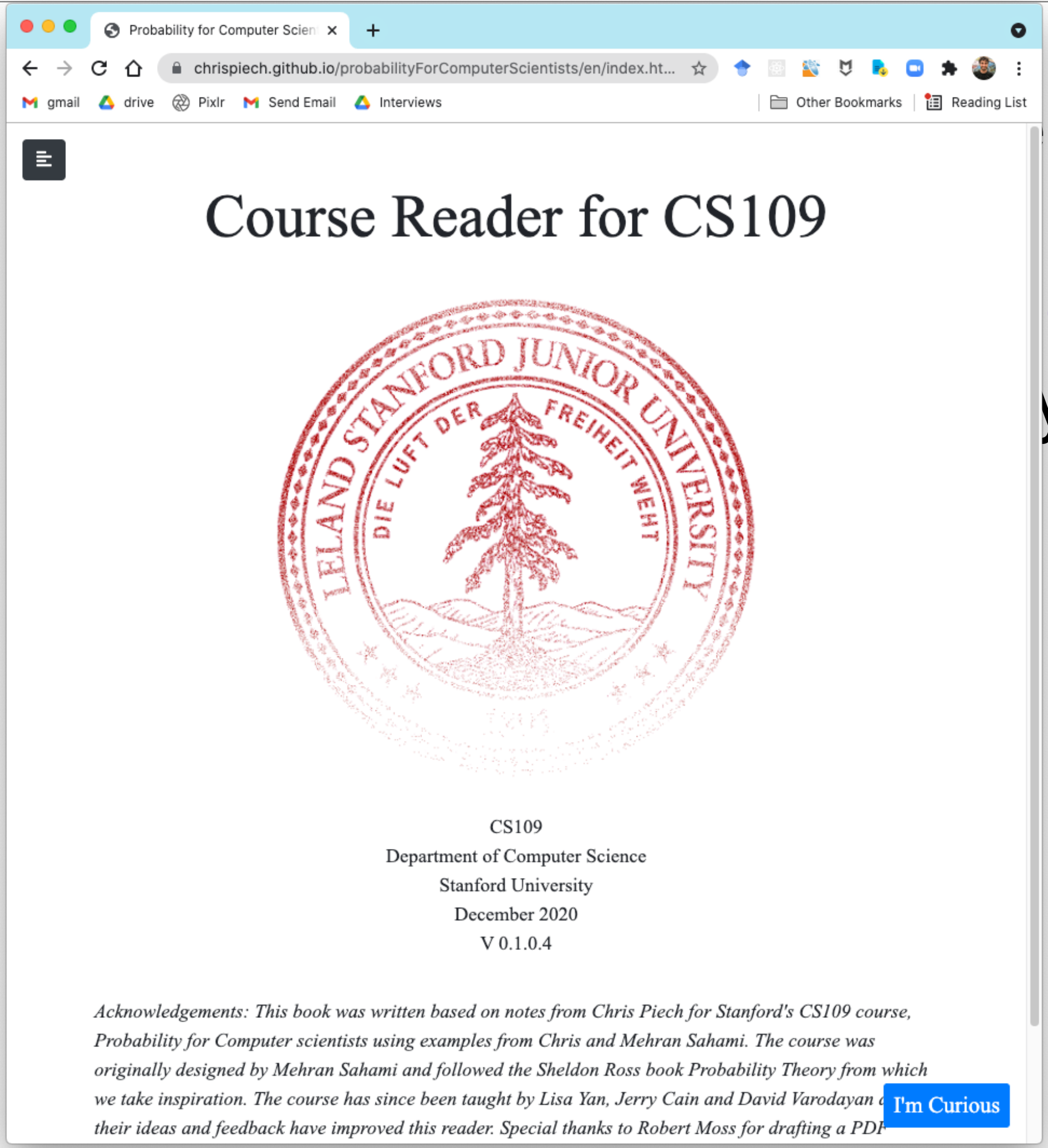
Independence Review: Weather Forecast



There is a 80% chance that the weather is sunny on any given day in Palo Alto. If the weather on each day is independent of the weather on other days, what is the probability that there are exactly 7 sunny days during the next 10-day period?

$$P(E) = \overbrace{\binom{10}{7}}^{\text{Number of ways to have exactly 7 sunny days in the next 10 day-period}} \underbrace{0.8^7(1-0.8)^3}_{\text{Probability of each outcome}} = 0.201\ldots$$

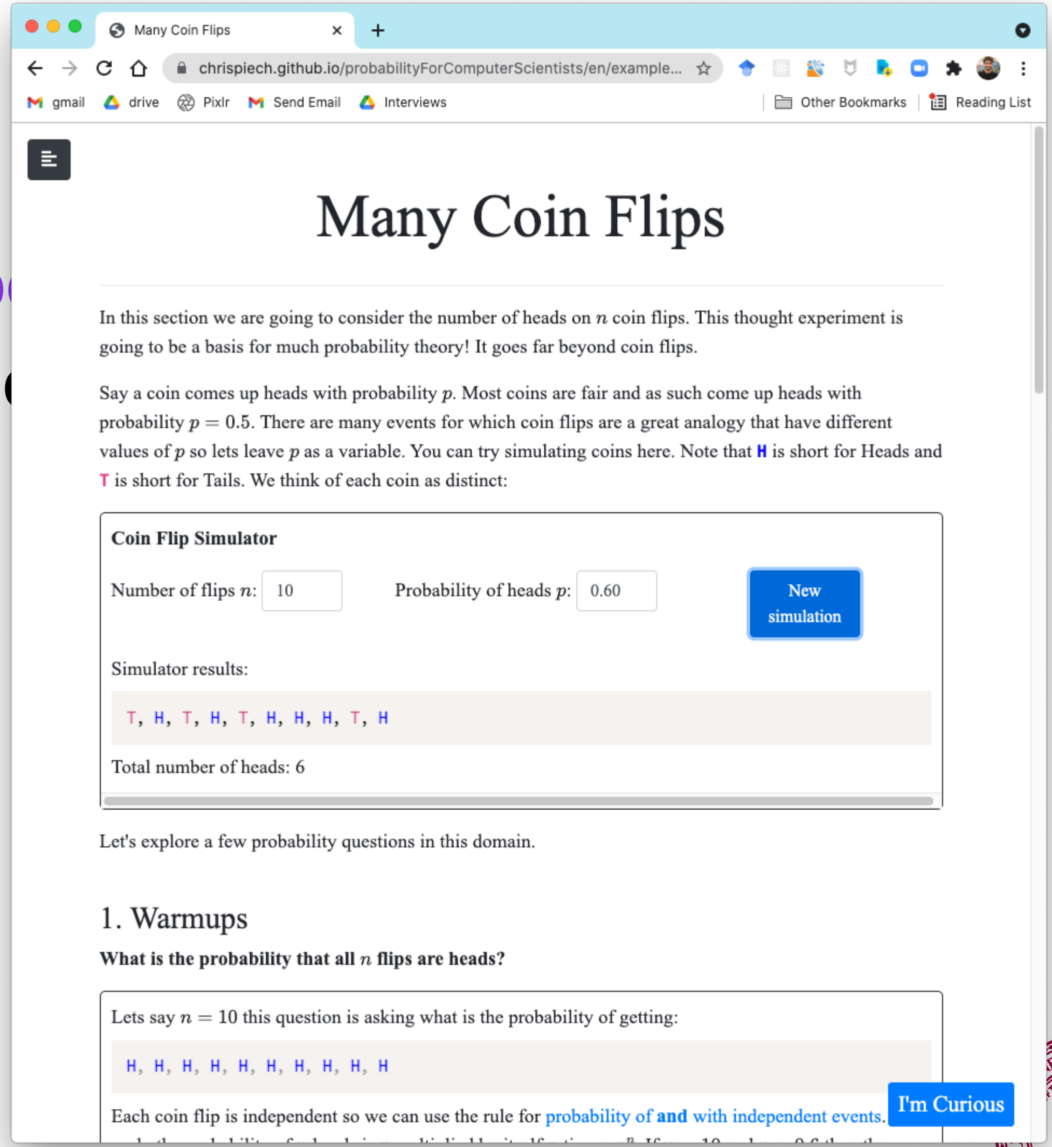
The Most Important Core Probability Question



ads

dep

y of



End of Review

Conditional Paradigm

Conditional Paradigm

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

Axiom 1	$0 \leq P(A\|E) \leq 1$
Corollary 1 (Complement)	$P(A\|E) = 1 - P(A^C\|E)$
Transitivity	$P(AB\|E) = P(BA\|E)$
Chain Rule	$P(AB\|E) = P(B\|E)P(A\|BE)$
Bayes' Theorem	$P(A\|BE) = \dfrac{P(B\|AE)P(A\|E)}{P(B\|E)}$

😍 BAE's Theorem

Conditional Paradigm:
When consistently conditioned on the same event, the ***formulas*** of probability are preserved.

Conditional Independence

Independent events E and F $\Leftrightarrow$ $P(EF) = P(E)P(F)$ $P(E|F) = P(E)$

To events A and B are defined as **conditionally independent given E** if:

$$P(AB|E) = P(A|E)P(B|E)$$

An equivalent definition:

A. $P(A|B) = P(A)$

B. $P(A|BE) = P(A)$

C. $P(A|BE) = P(A|E)$

E is the new sample space, so left and right side must both be conditioned on E

Conditional Paradigm: 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.

Conditional Dependence

If E and F are independent,

that does not mean E and F will be independent when another event is observed.

(See section 2 handout for an example)

Conditional Independence

If E and F are dependent,

that does not mean E and F will be dependent when another event is observed.

(See section 2 handout for an example)

NETFLIX

And Learn

Netflix Learning

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 Learning

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

$P(E|F)$

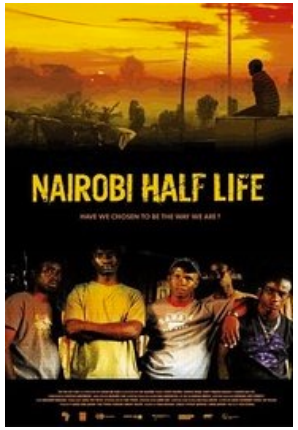
$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\#\text{people who watched both}}{\#\text{people who watched } F}$$

$P(E|F) = 0.42$

Conditioned on liking a set of movies?

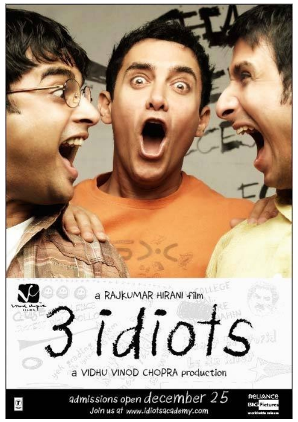
Netflix Learning

Each event corresponds to liking a particular movie



E_1

E_2



E_3

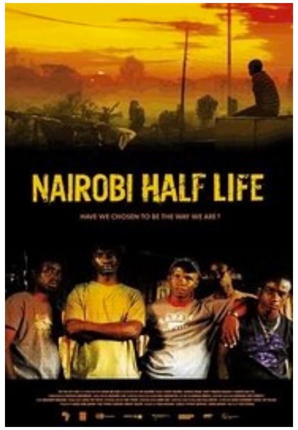
E_4

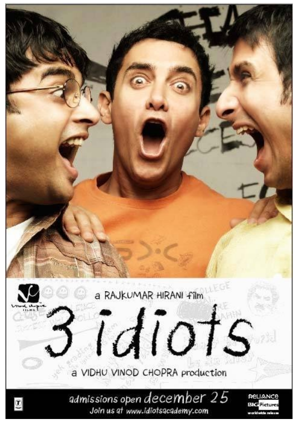
$$P(E_4 | E_1, E_2, E_3)?$$

Is E_4 independent of E_1,E_2,E_3?

Netflix Learning

Is E_4 independent of E_1,E_2,E_3?



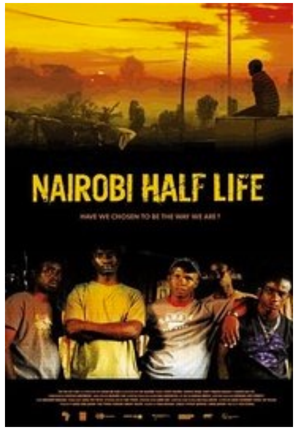


E_1 E_2 E_3 E_4

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

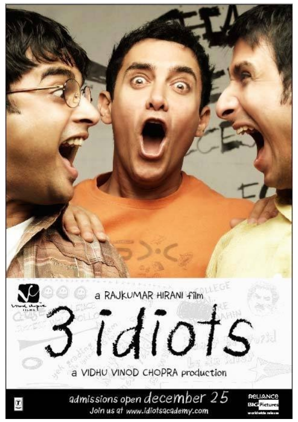
Netflix Learning

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 Learning

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

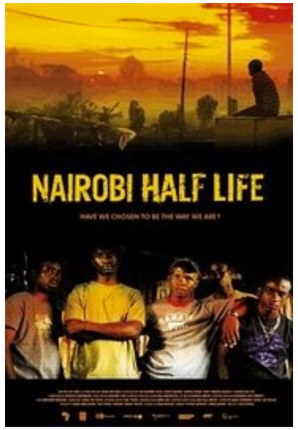
Choose 24 movies not in the set

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

Choose 30 movies from netflix

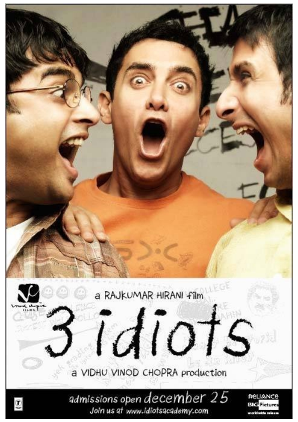
CHRIS!! 30 – 4 is not 24

Netflix Learning



E_1

E_2



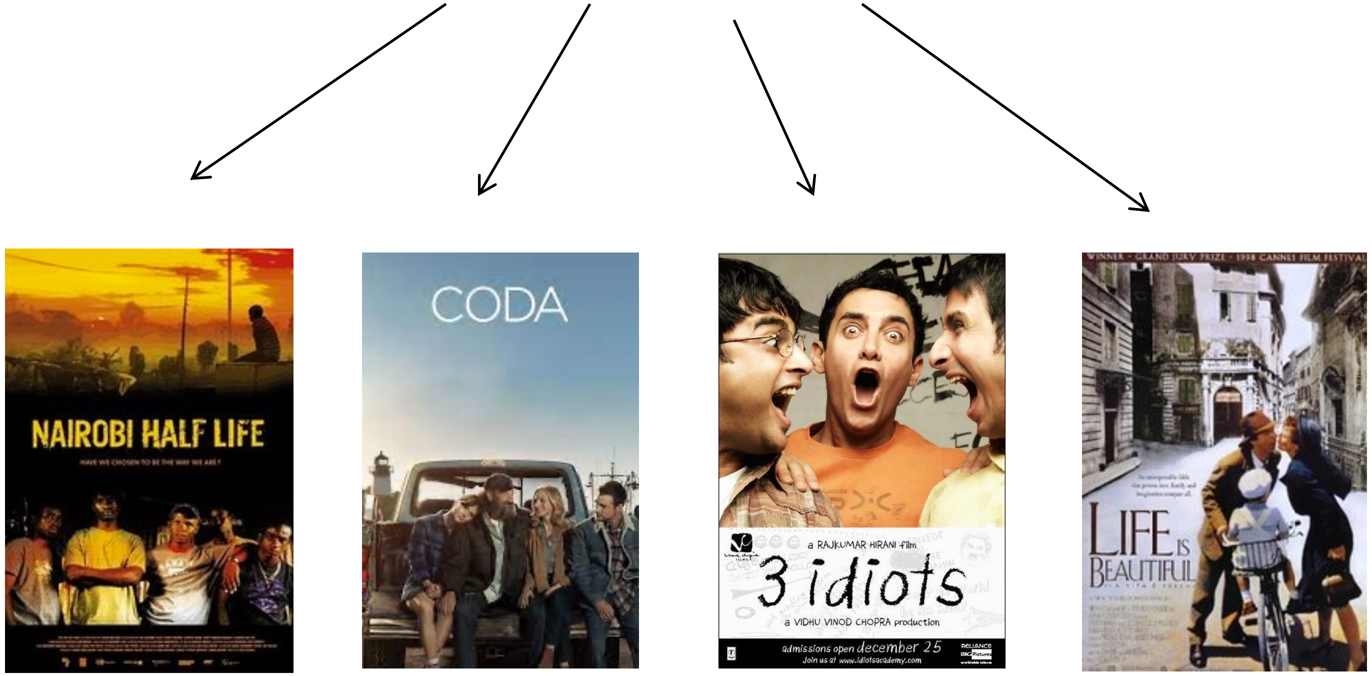
E_3

E_4

Netflix Learning: Advanced, Conditional Independence

K_1

Like foreign emotional comedies



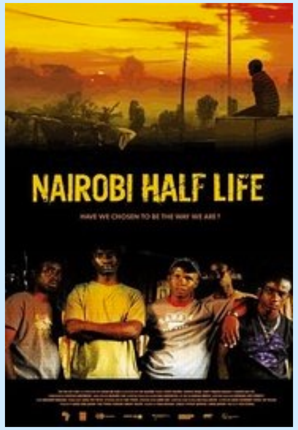
E_1 E_2 E_3 E_4

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

Netflix Learning: Advanced, Conditional Independence

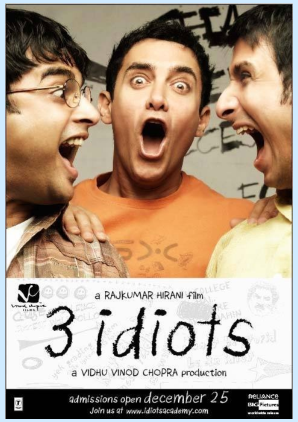
K_1

Like foreign emotional comedies



E_1

E_2

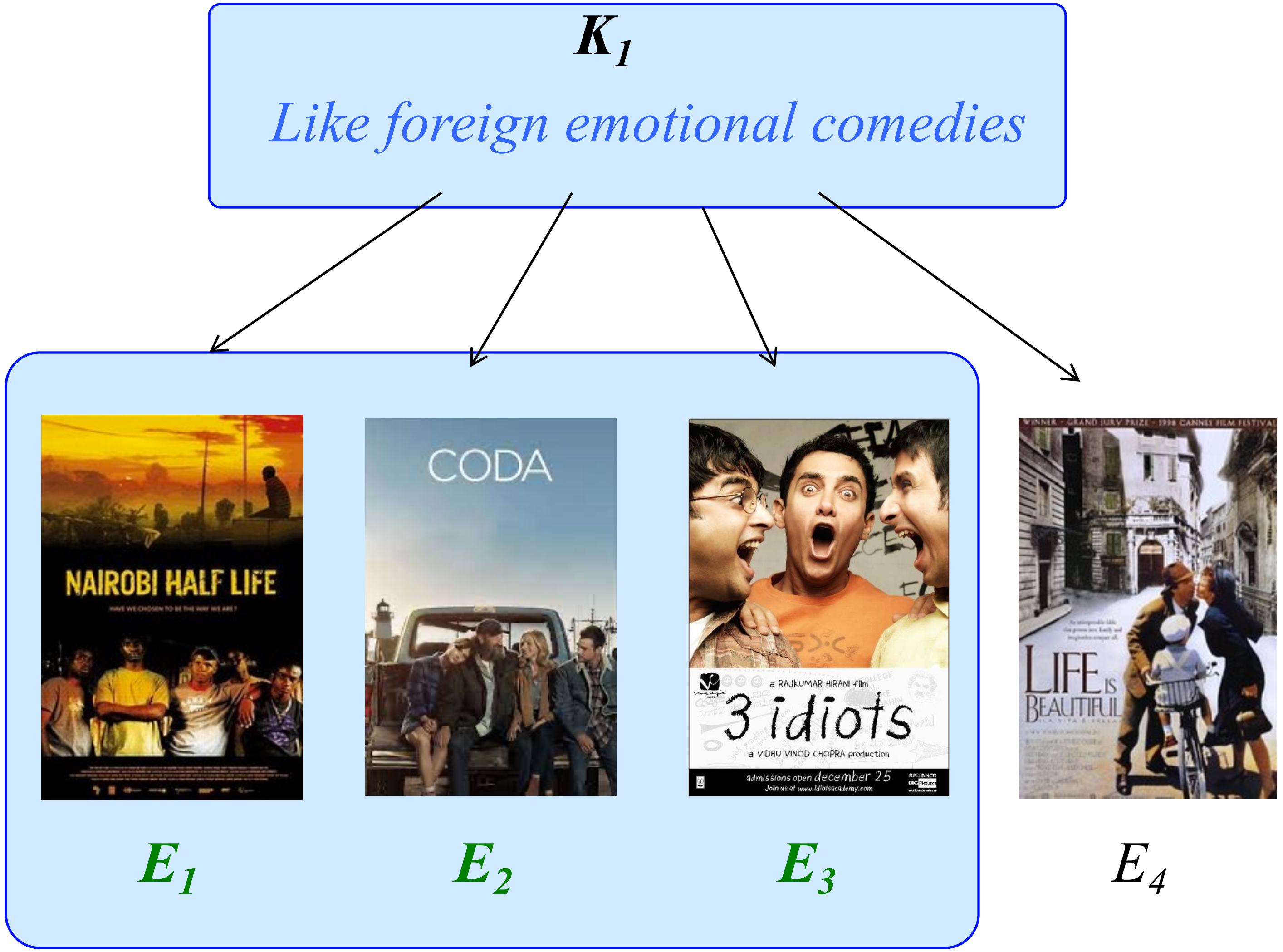


E_3

E_4

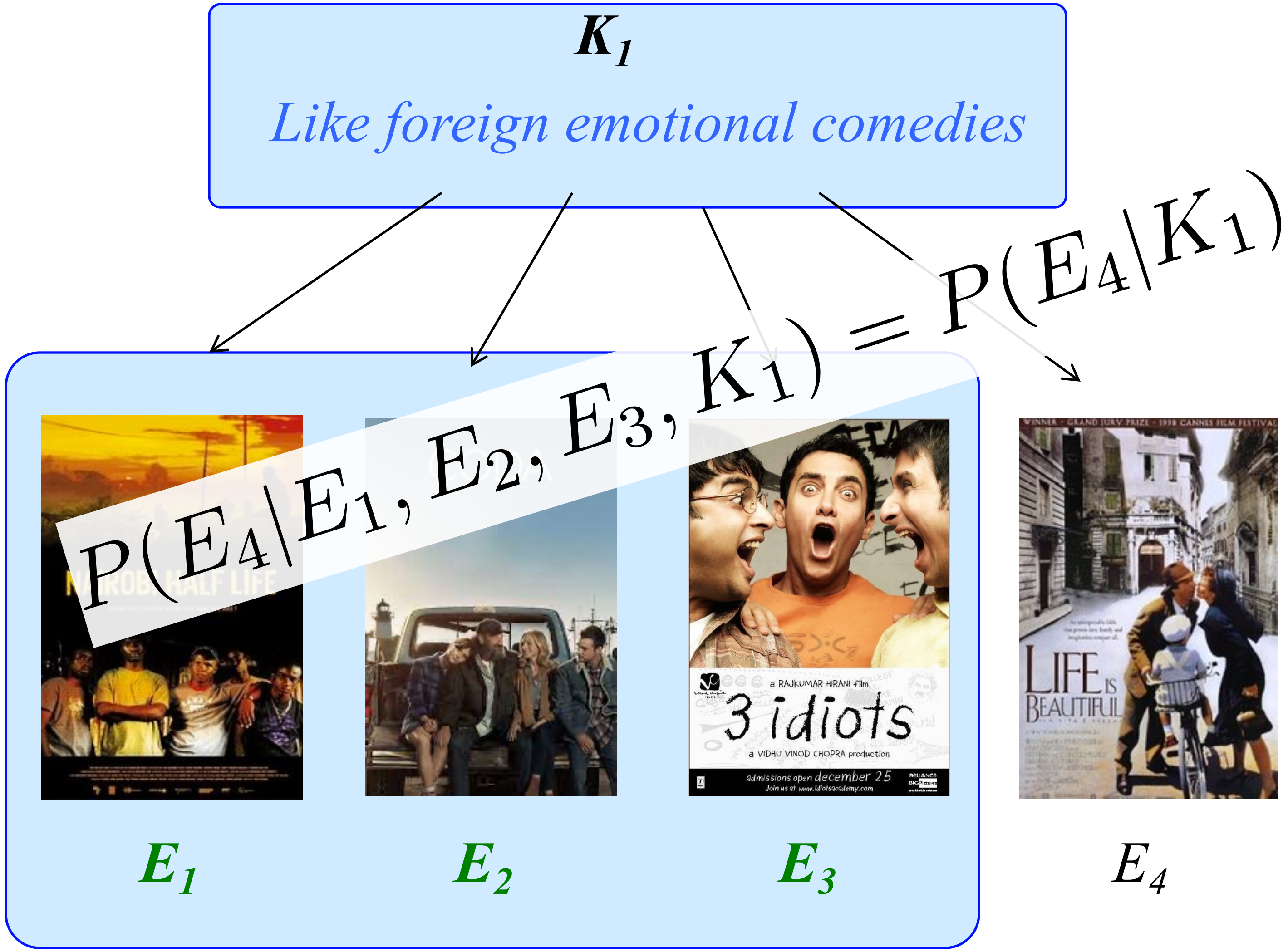
Assume E_1, E_2, E_3 and E_4 are conditionally independent given K_1

Netflix Learning: Advanced, Conditional Independence



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

Netflix Learning: Advanced, Conditional Independence



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

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"*

Random Variables

You Are Here

Overview of Topics

Counting Theory

Core Probability

Random Variables

Probabilistic Models

Uncertainty Theory

Machine Learning

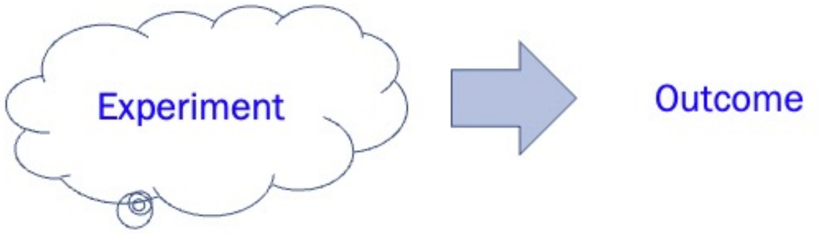
Learning Goals

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

Flashback to Week 1…

What is Counting?

An experiment in probability:



$$P(E)$$

- **Sample space**, S, is set of all possible outcomes of an experiment
  - Coin flip: S = {Head, Tails}
  - Flipping two coins: S = {(H, H), (H, T), (T, H), (T, T)}
  - Roll of 6-sided die: S = {1, 2, 3, 4, 5, 6}
  - # emails in a day: $S = \{x \mid x \in \mathbf{Z}, x \geq 0\}$ {non-neg. ints}
  - YouTube hrs. in day: $S = \{x \mid x \in \mathbf{R}, 0 \leq x \leq 24\}$

- **Event**, E, is some subset of S $\{E \subseteq S\}$
  - Coin flip is heads: E = {Head}
  - ≥ 1 head on 2 coin flips: E = {(H, H), (H, T), (T, H)}
  - Roll of die is 3 or less: E = {1, 2, 3}
  - # emails in a day ≤ 20: $E = \{x \mid x \in \mathbf{Z}, 0 \leq x \leq 20\}$
  - Wasted day {≥ 5 YT hrs.}: $E = \{x \mid x \in \mathbf{R}, 5 \leq x \leq 24\}$

More than Just Outcomes…

A dog is learning to do a "give paw" trick. She has ½ chance of succeeding each attempt, and all attempts are independent. We're interested in trying "give paw" 10 times and thinking about <u>how many times she'll succeed</u>.

We could ask…

P(Succeeds 3 or 4 times)

P(Succeeds 8 times)

P(More than 5 success)

P(Succeeds 0 times)

"What's a 'typical' number of successes?"

Outcomes of this experiment:

[S,F,S,S,F,S,F,S,F,F]

[S,S,S,S,S,F,S,S,F,F]

[S,S,F,F,S,F,F,S,F,F]

[S,F,F,F,F,F,S,S,F,F]

S: Success
F: Failure

More than Just Outcomes…

A dog is learning to do a "give paw" trick. She has ½ chance of succeeding each attempt, and all attempts are independent. We're interested in trying "give paw" 10 times and think about how many times she'll succeed.

We could ask…

P(Succeeds 3 or 4 times)

P(Succeeds 8 times)

P(More than 5 success)

P(Succeeds 0 times)

"What's a 'typical' number of successes?"

What really matters to us:

5 Successes

7 Successes

4 Successes

3 Successes

More than Just Outcomes…

A dog is learning to do a "give paw" trick. She has ½ chance of succeeding each attempt, and all attempts are independent. We're interested in trying "give paw" 10 times and think about how many times she'll succeed.

We could ask…

P(Succeeds 3 or 4 times)

P(Succeeds 8 times)

P(More than 5 success)

P(Succeeds 0 times)

"What's a 'typical' number of successes?"

We are often interested more in the numbers associated with the outcomes than the outcomes themselves

Definition of Random Variables

A Random Variable is an **assignment of numbers** to **outcomes**.

More formally, it is a **function** that maps outcomes to numbers

$$X(\cdot): S \to V$$

where V is a set of numbers.

$V = \{x \mid -2 \leq x \leq 2\}$

$V = \mathbb{R}$

$V = \{0, 1, \ldots\}$

$V = \{0, 1\}$

Examples of Random Variables

A Random Variable is an **assignment of numbers** to **outcomes**.

X: Number of successful attempts of “give paw!” in 4 attempts

X maps the series of attempts (outcome) to 0,1,…,4

$$X([\mathrm{S,S,S,S}]) = 4$$

$$X([\mathrm{S,F,S,F}]) = 2$$

$$X([\mathrm{F,F,S,S}]) = 2$$

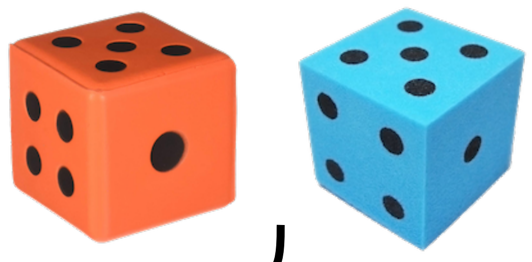
$$X([\mathrm{F,F,F,F}]) = 0$$

S: Success
F: Failure

Examples of Random Variables

A Random Variable is an **assignment of numbers** to **outcomes**.

Each outcome



Value
dice 1

Value
dice 2

X: Sum of the rolls of two 6-sided dice

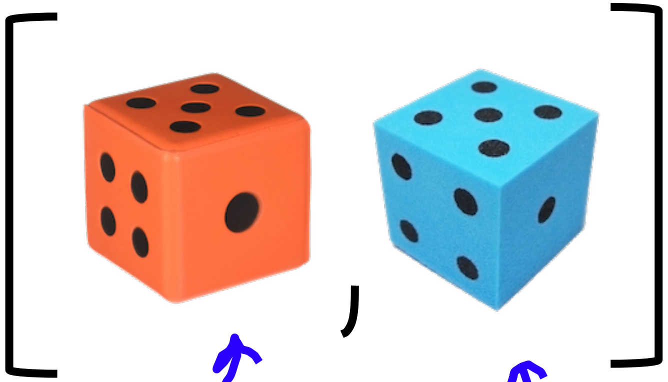
X maps each pair of dice rolls to 2,3,…,12

$$
S = \{\begin{matrix}
[1,1] & [1,2] & [1,3] & [1,4] & [1,5] & [1,6] \\
[2,1] & [2,2] & [2,3] & [2,4] & [2,5] & [2,6] \\
[3,1] & [3,2] & [3,3] & [3,4] & [3,5] & [3,6] \\
[4,1] & [4,2] & [4,3] & [4,4] & [4,5] & [4,6] \\
[5,1] & [5,2] & [5,3] & [5,4] & [5,5] & [5,6] \\
[6,1] & [6,2] & [6,3] & [6,4] & [6,5] & [6,6]
\end{matrix}\}
$$

Examples of Random Variables

A Random Variable is an **assignment of numbers** to **outcomes**.

Each outcome



Value dice 1

Value dice 2

X: Sum of the rolls of two 6-sided dice

X maps each pair of dice rolls to 2,3,…,12

2 3 4 5 6 7

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] 8

[3,1] [3,2] [3,3] [3,4] [3,5] [3,6] 9

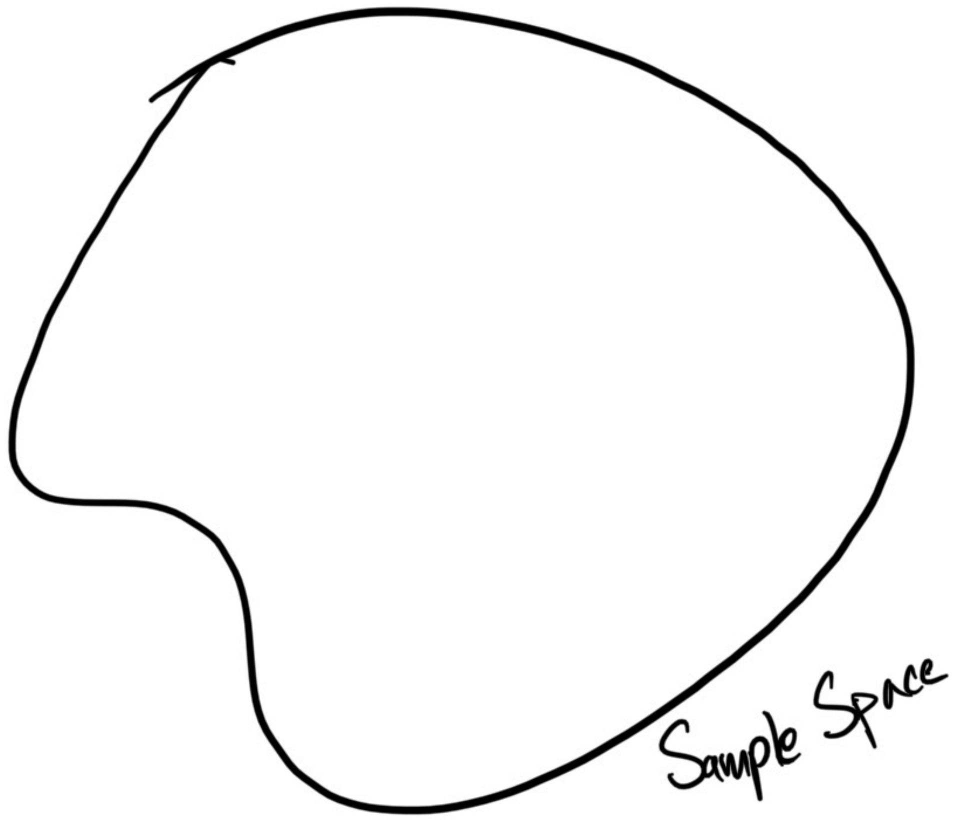
[4,1] [4,2] [4,3] [4,4] [4,5] [4,6] 10

[5,1] [5,2] [5,3] [5,4] [5,5] [5,6] 11

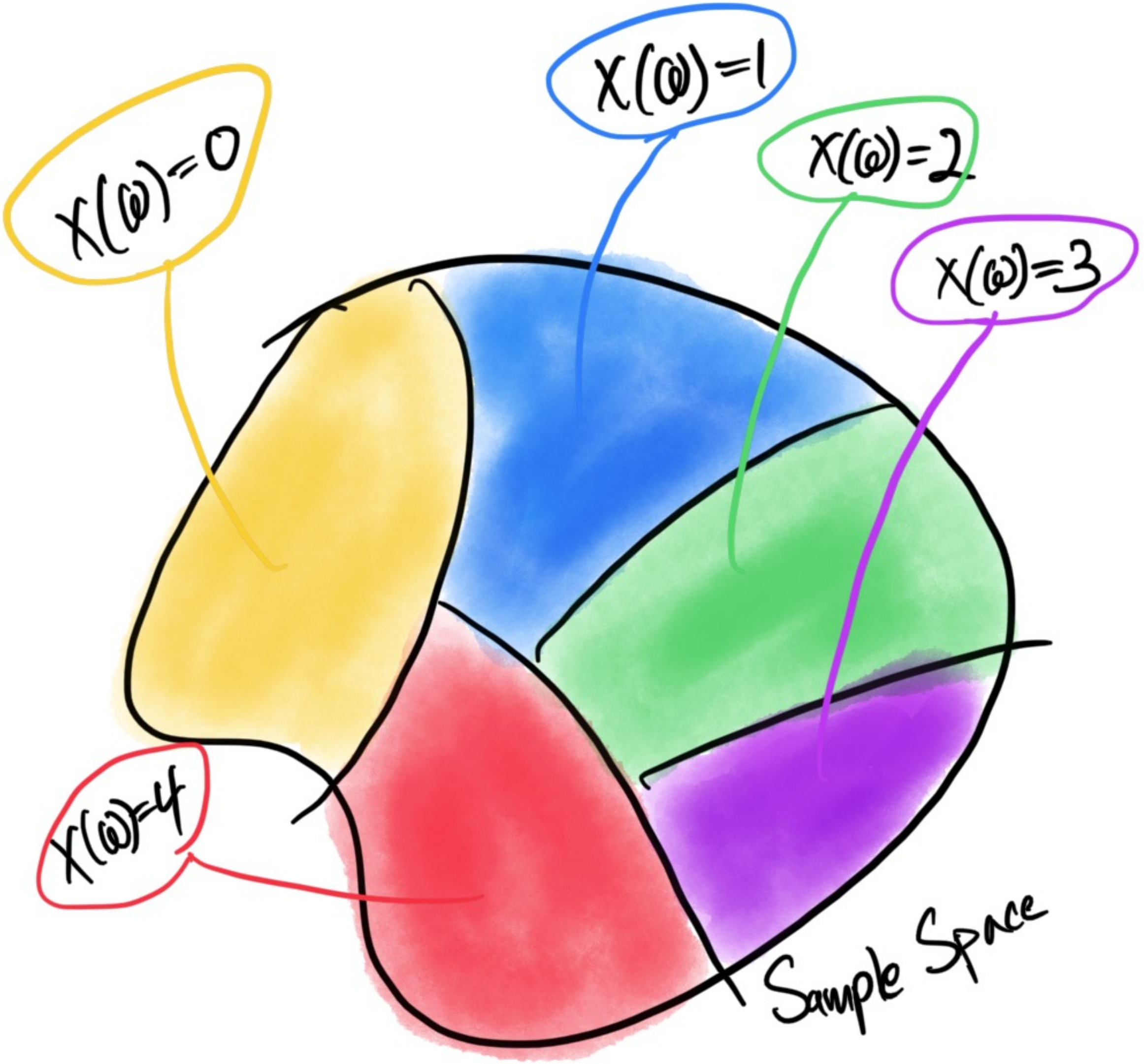
[6,1] [6,2] [6,3] [6,4] [6,5] [6,6]} 12

Visualizing Random Variables

Random Variables Group Outcomes



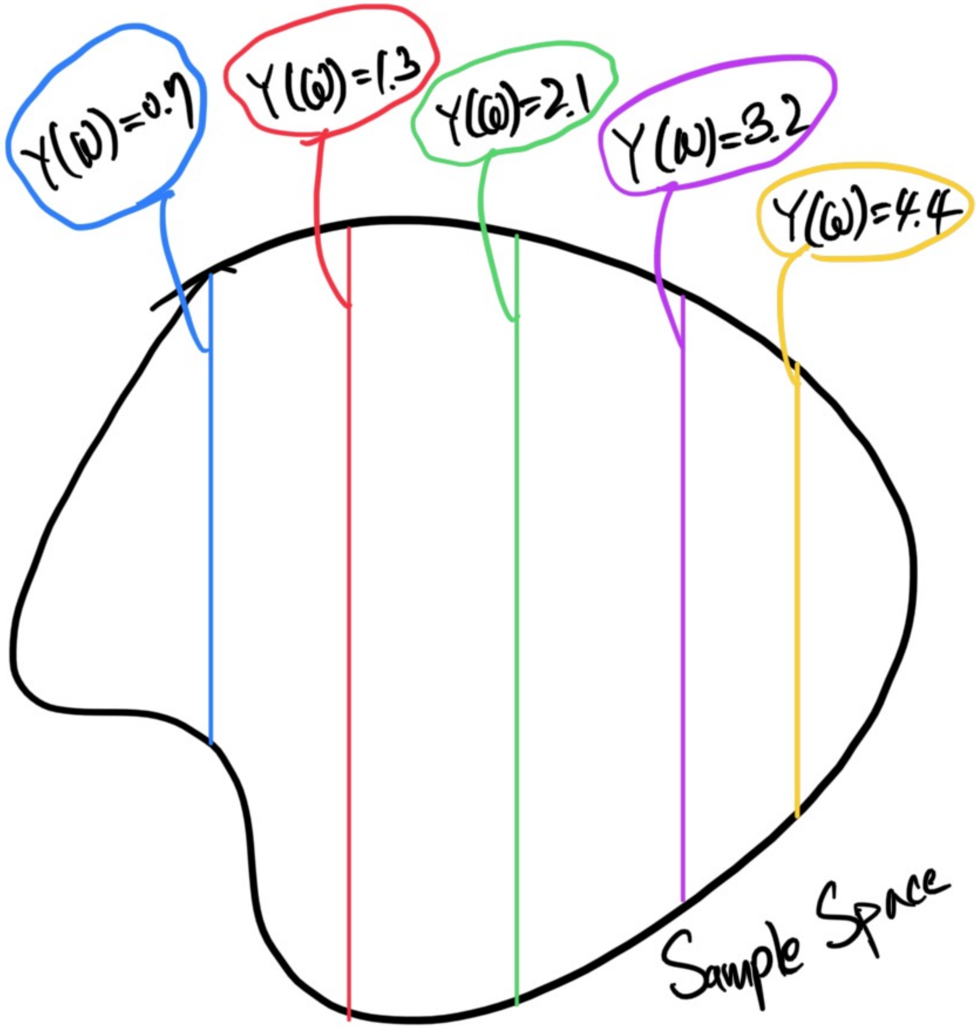
Random Variables Group Outcomes



X: Random Variable

ω: Outcome

Random Variables Group Outcomes

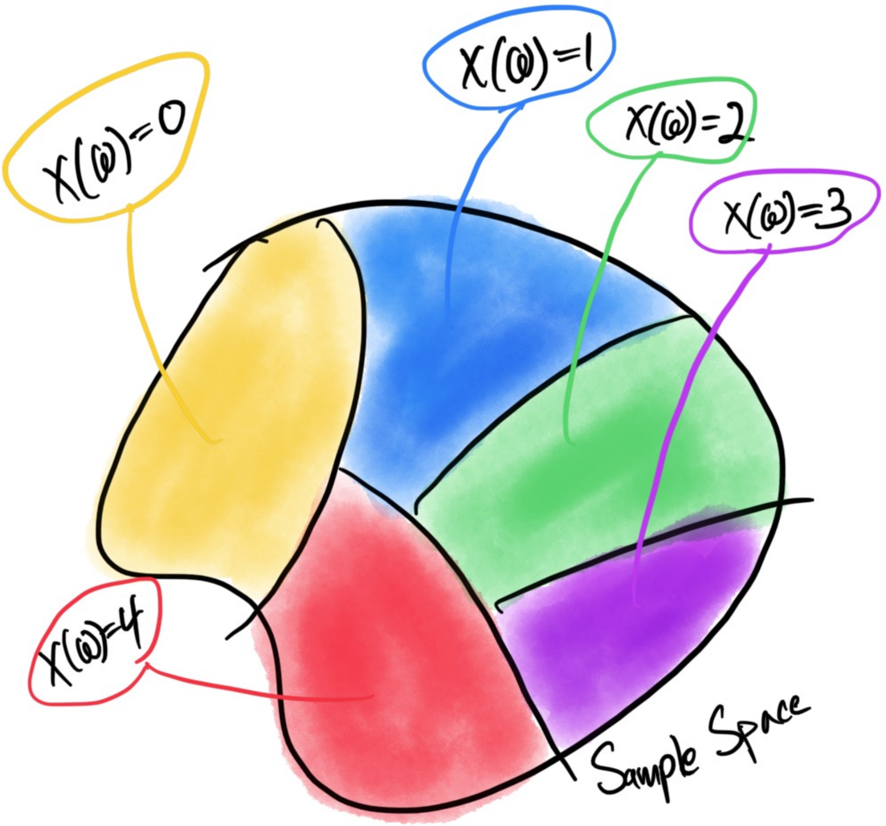


Y: Random Variable

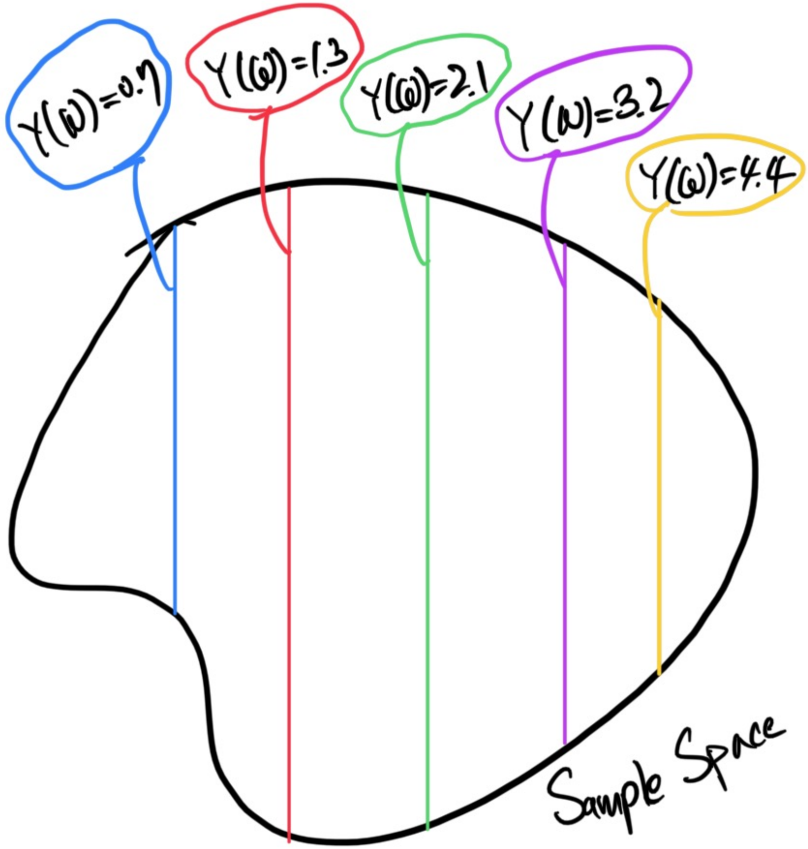
ω: Outcome

Random Variables Group Outcomes

X, Y: Random Variables ω: Outcome



Discrete Random Variables

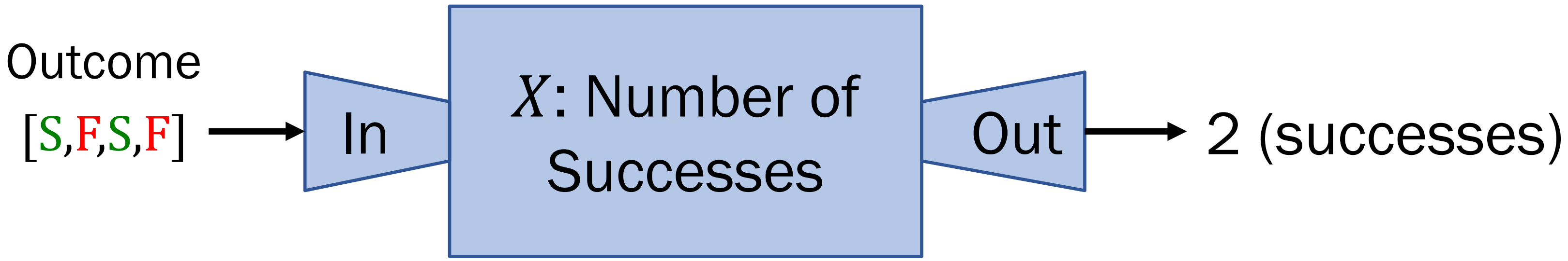


Continuous Random Variables

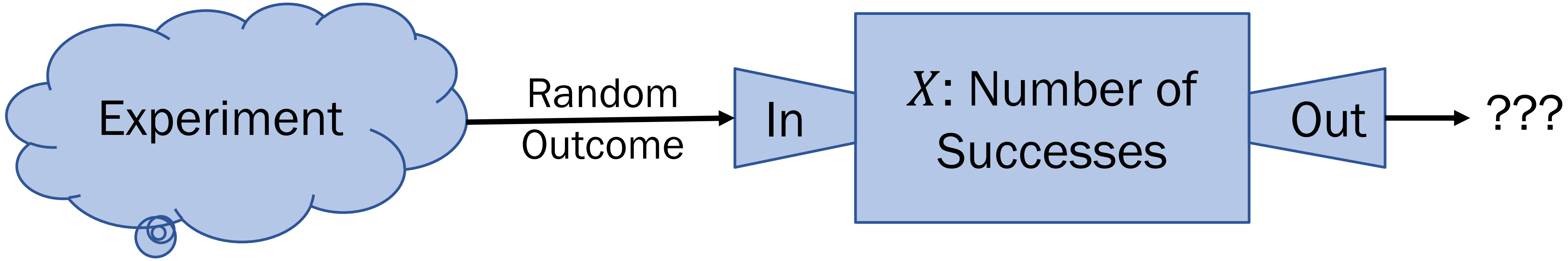
But Why Are They Called “Random” “Variables”?

Random Variables as "Variables with Uncertainty"

A random variable is a function that maps outcomes in the sample space to numbers of interest. On its own, it is just a function.



When combined with the random experiment and its outcome is fed through this function, we say that X "**<u>takes on</u>**" the value it outputs. Since the outcome is random, so is this value.



Random Variables as “Variables with Uncertainty”

$X =$ 

When combined with the random experiment and its outcome is fed through this function, we say that X “**<u>takes on</u>**” the value it outputs. Since the outcome is random, so is this value.

Random Variables as "Variables with Uncertainty"

$X =$ 

When combined with the random experiment and its outcome is fed through this function, we say that X "**<u>takes on</u>**" the value it outputs. Since the outcome is random, so is this value.

Random Variables as "Variables with Uncertainty"

$X =$ 

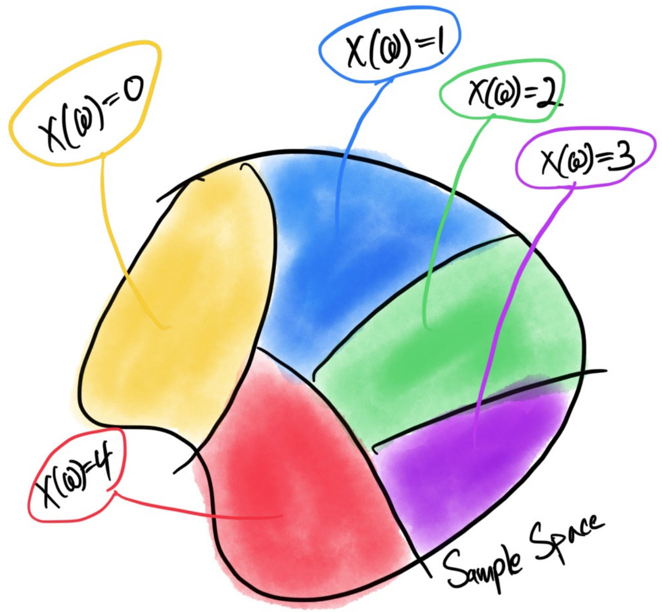
When combined with the random experiment and its outcome is fed through this function, we say that X "**takes on**" the value it outputs. Since the outcome is random, so is this value.

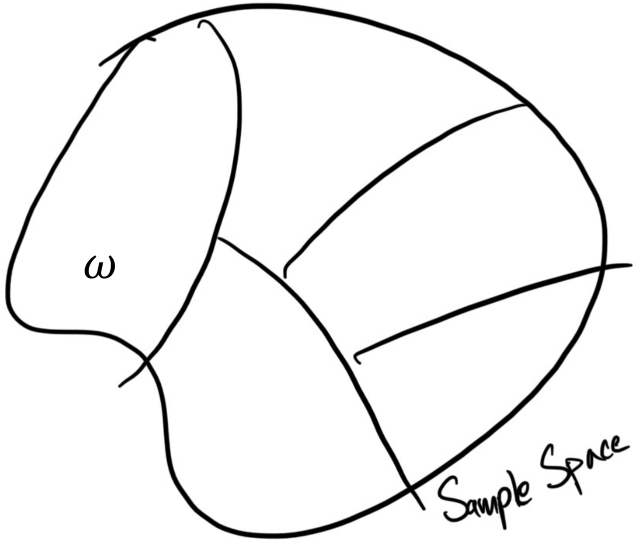
Random Variables as "Variables with Uncertainty"

X "takes on" a random value when combined with random experiment

$X =$

Under the hood…



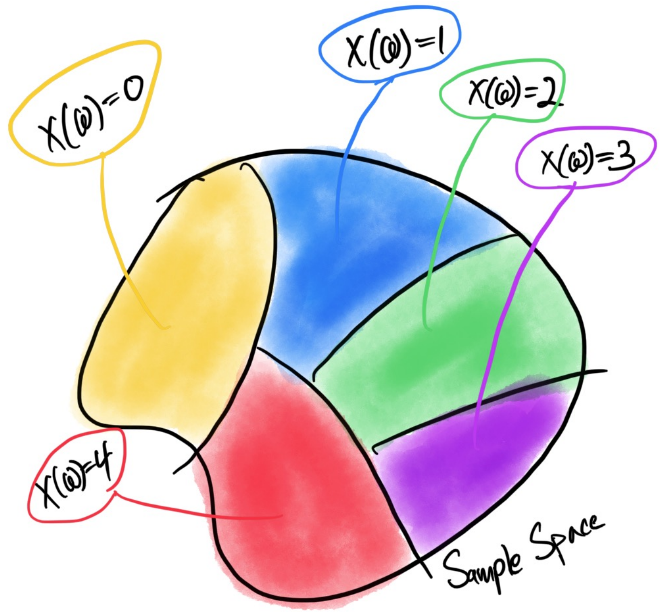


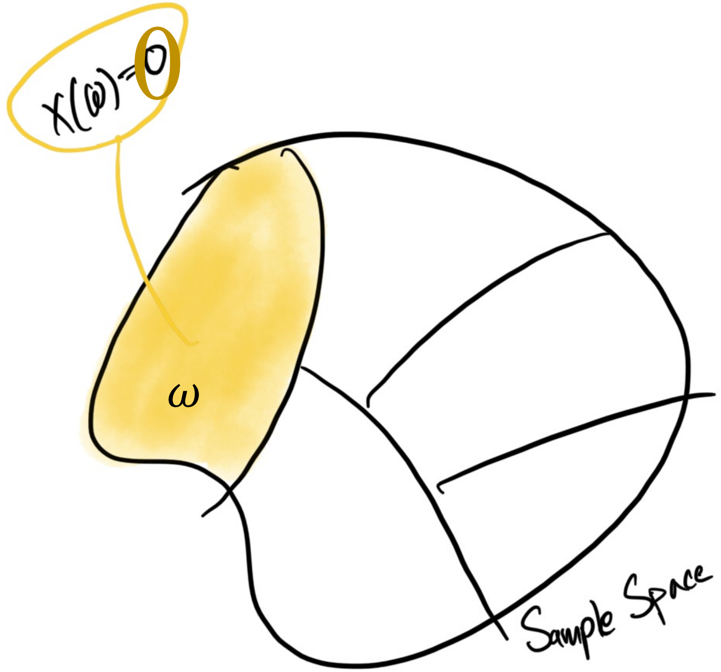
Random Variables as "Variables with Uncertainty"

X "takes on" a random value when combined with random experiment

$X =$ 

Under the hood…



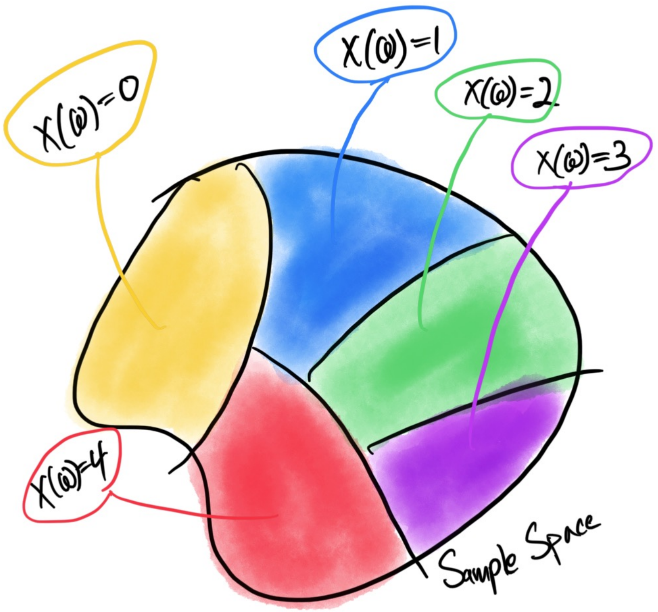


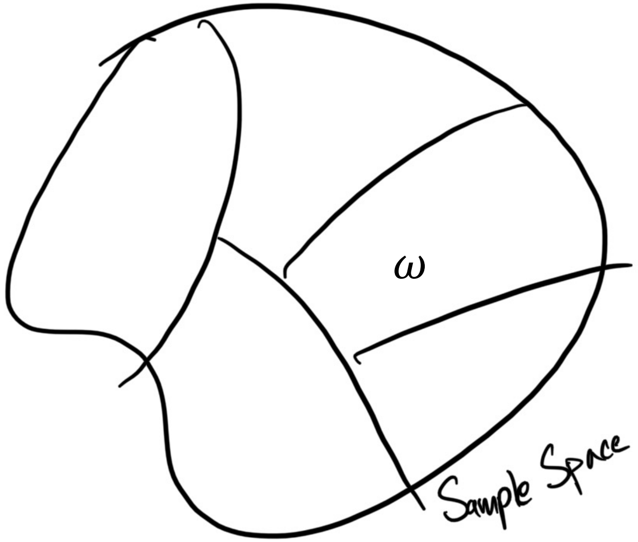
Random Variables as "Variables with Uncertainty"

X "takes on" a random value when combined with random experiment

$X =$ 

Under the hood…

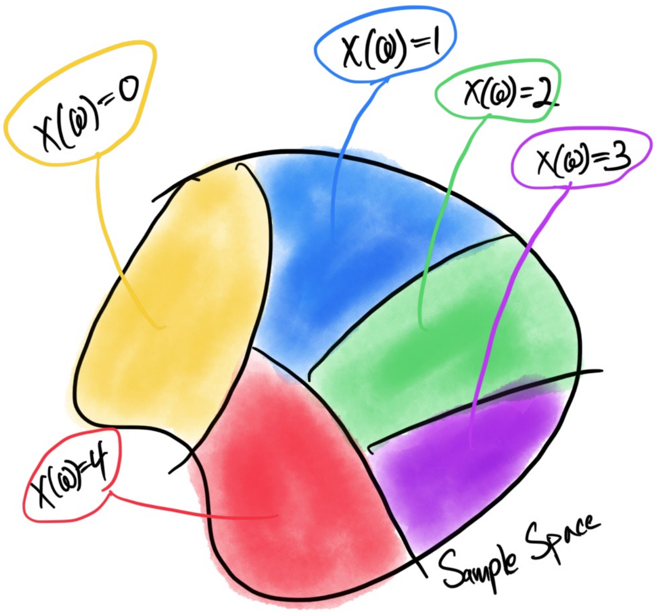


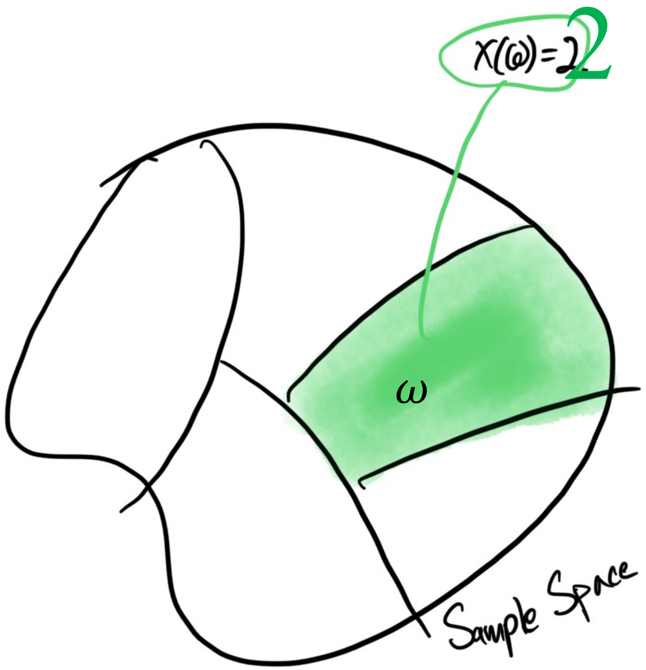


Random Variables as "Variables with Uncertainty"

X "takes on" a random value when combined with random experiment

$X =$ 





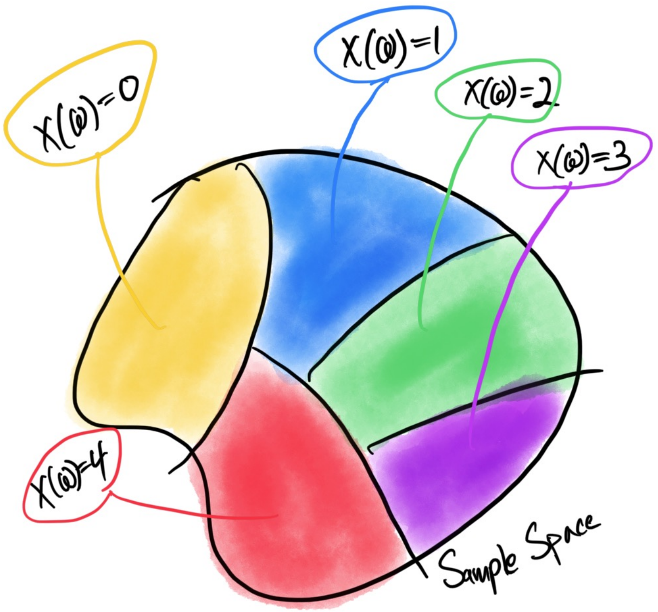
Under the hood…

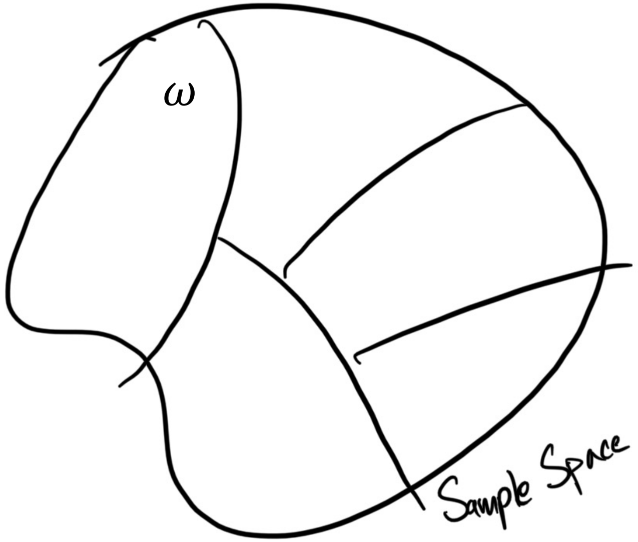
Random Variables as "Variables with Uncertainty"

X "takes on" a random value when combined with random experiment

$$X =$$

Under the hood…

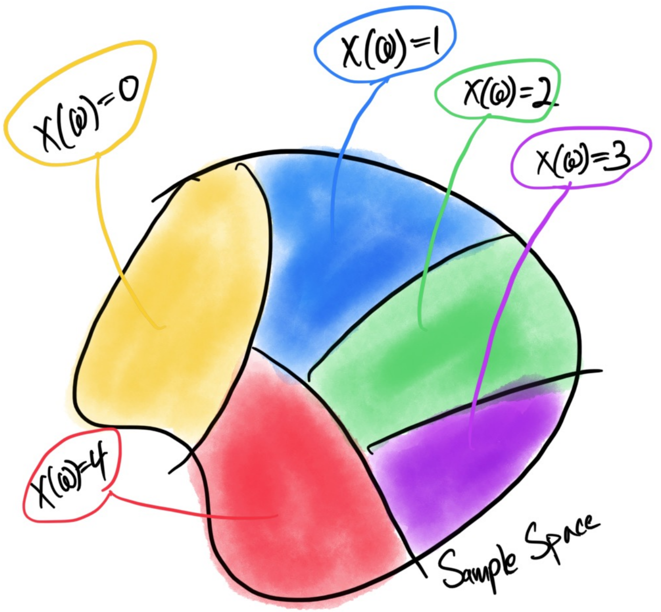


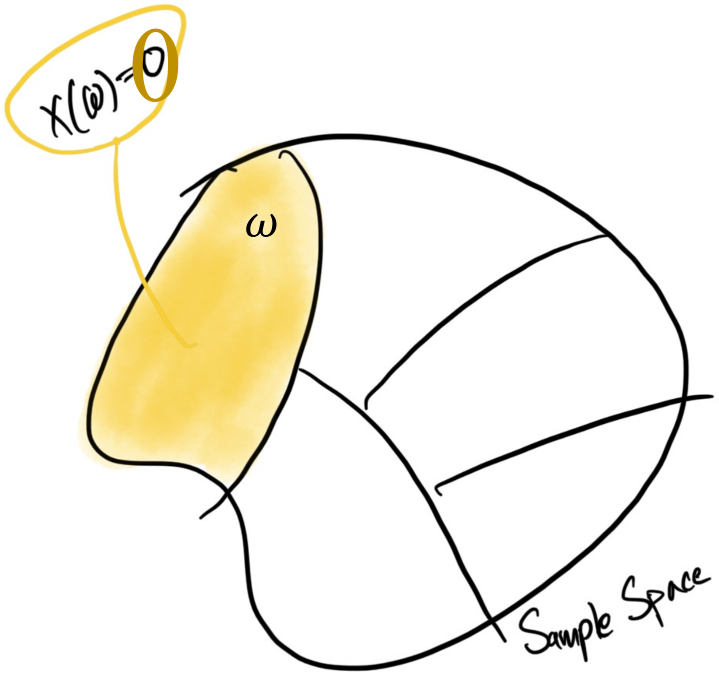


Random Variables as “Variables with Uncertainty”

X “takes on” a random value when combined with random experiment

$X =$ 





Under the hood…

Random Variables as "Variables with Uncertainty"

X "takes on" a random value when combined with random experiment

name

value

type

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

double X =

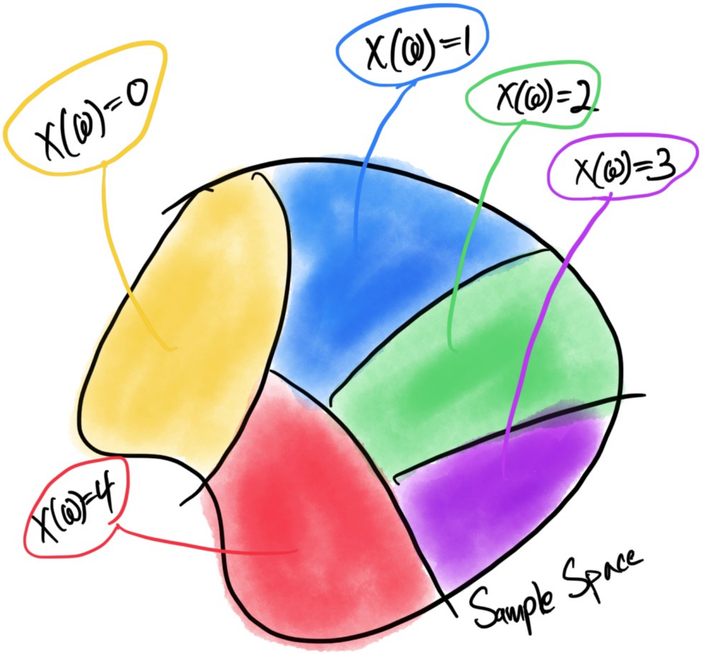
Random variables are like programming variables, with uncertainty

End of Visualization

Talking Probability using Random Variables

We can calculate the probability that a random variable X "takes on" a value.

It is common to write: $P(X = 0)$



and it is equivalent to

$$P(\{\omega \in S | X(\omega) = 0\})$$

Any True/False statement about a random variable "X satisfies …" denotes the event $\{\omega \in S | X(\omega) \text{ satisfies } \ldots\}$

Talking Probability using Random Variables

4 independent tries of "give paw".
Each try has ½ chance of success.

Y: Number of successful tries

Y is a random variable.

$P(Y=0) = 1/16$	{[F,F,F,F]}
$P(Y=1) = 4/16$	{[S,F,F,F],[F,S,F,F],[F,F,S,F],[F,F,F,S]}
$P(Y=2) = 6/16$	{[S,S,F,F],[S,F,S,F],[S,F,F,S],[F,S,S,F],[F,S,F,S],[F,F,S,S]}
$P(Y=3) = 4/16$	{[S,S,S,F],[S,S,F,S],[S,F,S,S],[F,S,S,S]}
$P(Y=4) = 1/16$	{[S,S,S,S]}

Properties of Random Variables

Probability Mass Function:

$$P(X = a)$$

Expectation:

$$E[X]$$

Variance:

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

Learning goals for today

1. Probability Mass Function

The relationship between values a random variable can take on, and the corresponding probability, is a ***function***!

Let Y be a random variable

$$Y$$

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

$$Y = 2$$

It is an event when Y takes on a value

$$\{\omega | Y(\omega) = 2\}$$

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

If this is a number

$$P(Y = 2)$$

Then this is a number (between 0 and 1)

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

If this is a variable

$$\underbrace{P(Y = k)}$$

Then this is a function

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

Random Variables -> Functions

$$P(Y = k)$$

$$k = 5$$

$$0.03125$$

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

Random Variables -> Functions

$$P(Y = k)$$

```
def event_probability(k):
    # probability mass function of Y in python
    N = 5    # number of coin flips
    P = 0.5  # probability of heads

    ways = math.comb(N, k);
    prob_heads = math.pow(P, k)
    prob_tails = math.pow(P, N-k)
    return ways * a * b
```

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 (PMF)

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

$$P(X = x)$$

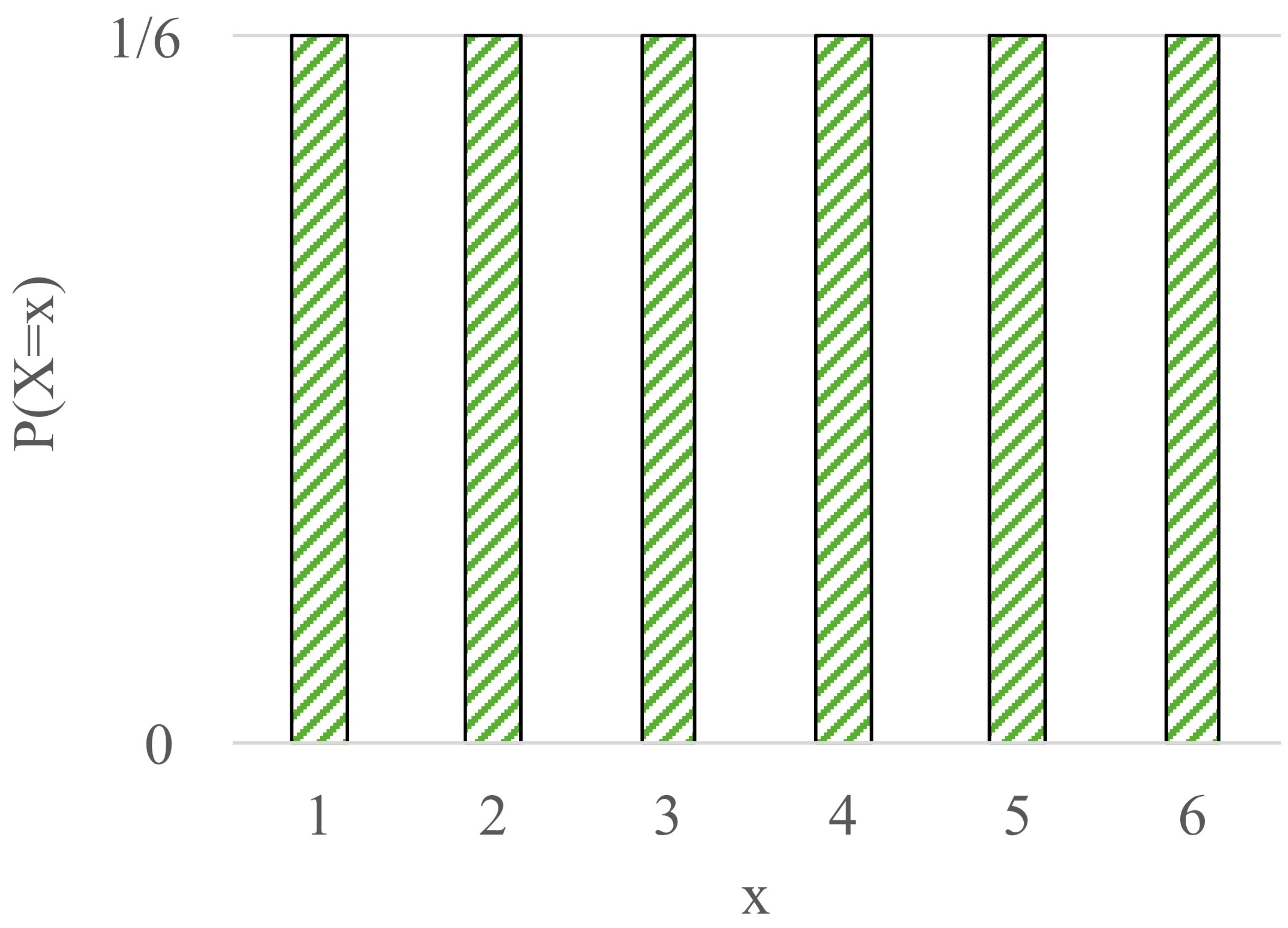
$$p(x)$$

This is shorthand notation for the PMF

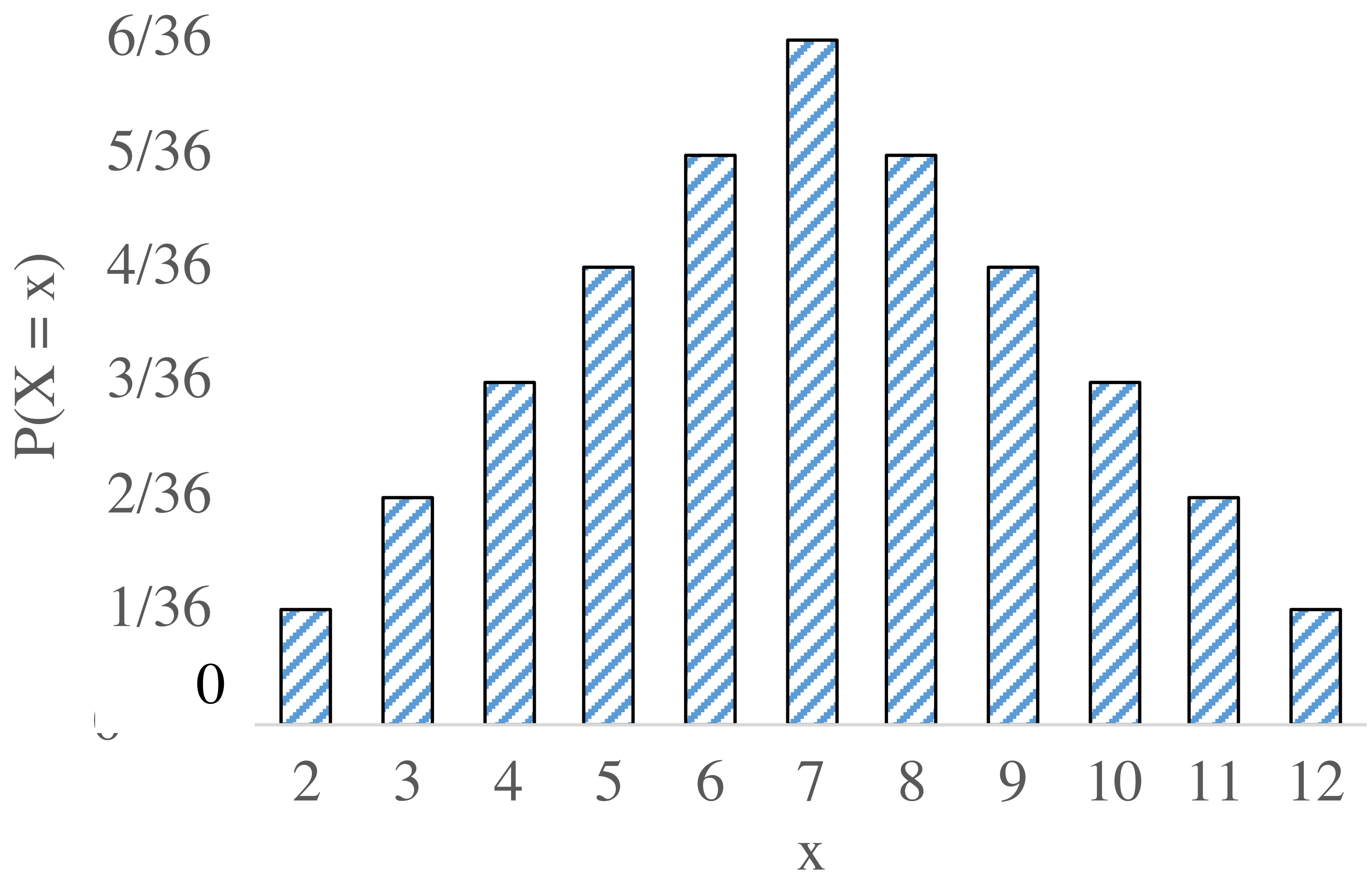
$$p_X(x)$$

This is also shorthand notation for the PMF

PMF for *X the outcome of a die roll*



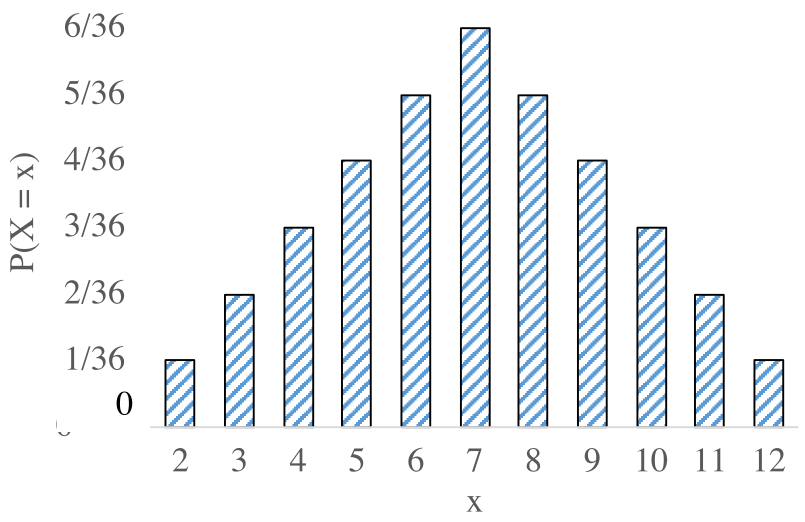
PMF for *X the sum of two dice rolls*



PMF as an equation

$$p(X = x) = \begin{cases} \frac{x-1}{36} & \text{if } x \in \mathbb{Z}\ ,\ 1 \leq x \leq 6 \\ \frac{13-x}{36} & \text{if } x \in \mathbb{Z}\ ,\ 7 \leq x \leq 12 \\ 0 & \text{else} \end{cases}$$

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

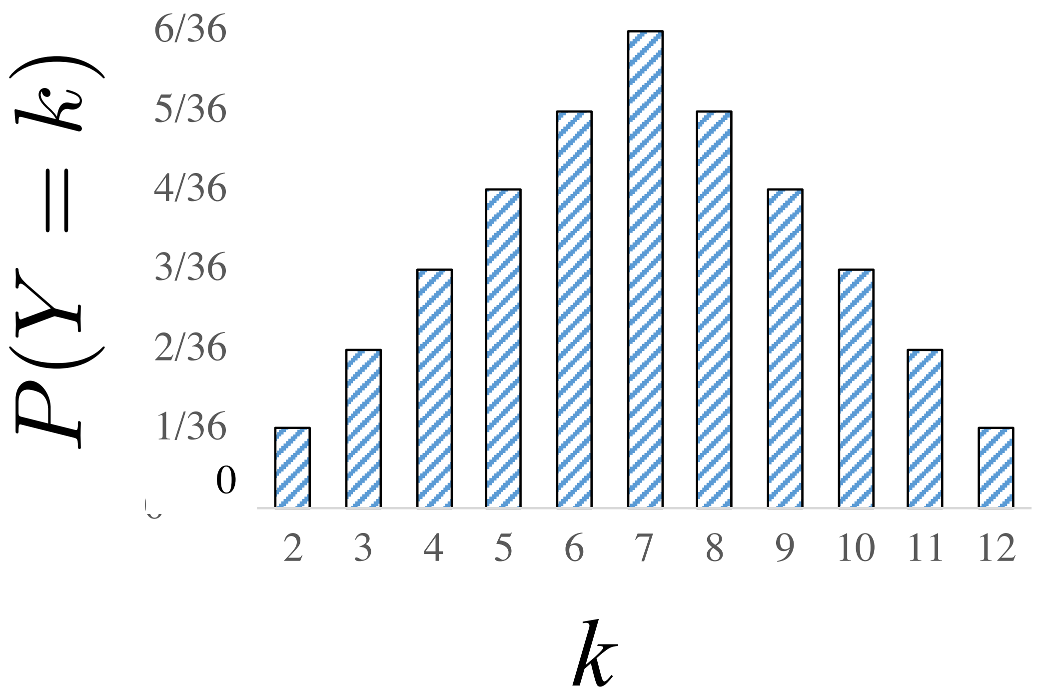


This is Fine Check

$$\sum_{\text{all } k} P(Y = k) \stackrel{?}{=}$$

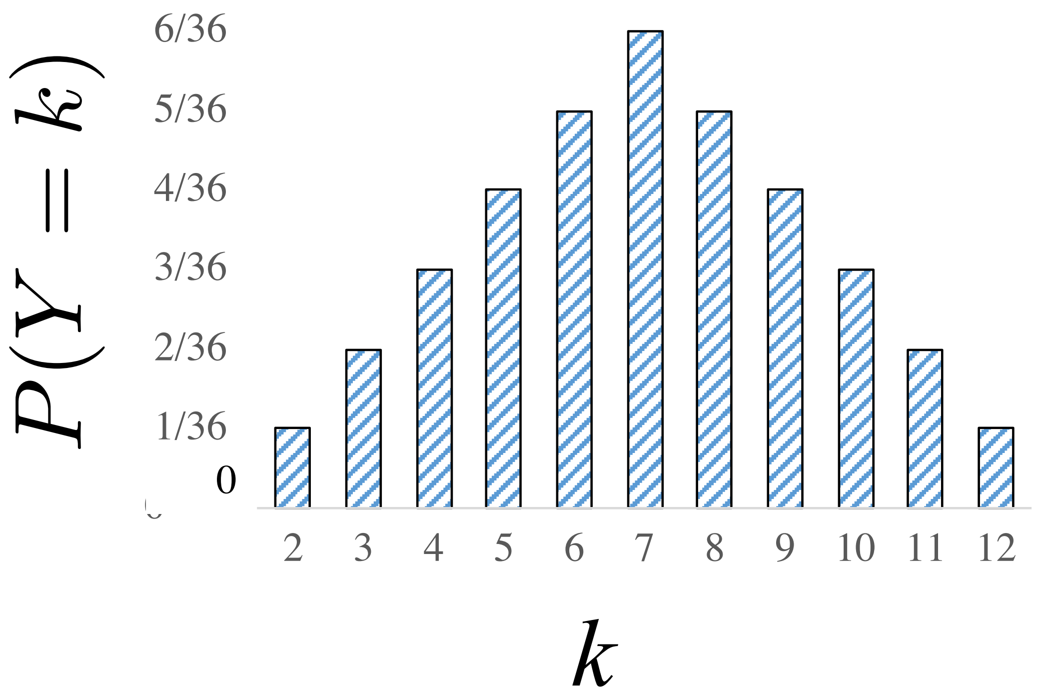
This is Fine Check

$$\sum_{\text{all } k} P(Y = k) \stackrel{?}{=}$$



This is Fine Check

$$\sum_k P(Y = k) = 1$$



2. Expectation

Properties of Random Variables

Probability Mass Function:

$$P(X = a)$$

Expectation:

$$E[X]$$

Variance:

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

Learning goals for today

Expected Value

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

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

$P(X = x)$

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

- $p(X=1) = p(X=2) = p(X=3) = p(X=4) = p(X=5) = p(X=6) = 1/6$

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

Y is random variable

- $P(Y = 1) = 1/3, \quad P(Y = 2) = 1/6, \quad P(Y = 3) = 1/2$

$E[Y] = 1\ (1/3) + 2\ (1/6) + 3\ (1/2) = 13/6$

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 <u>class</u> with equal probability

X = size of chosen class

What is E[X]?

- $E[X] = 5\,(1/3) + 10\,(1/3) + 150\,(1/3)$
  $= 165/3 = 55$

Same expectation if we had 55, 55, 55 student classes

Lying With Statistics Part 2

"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 <u>student</u> with equal probability

Y = size of class that student is in

What is E[Y]?

- $E[Y] = 5\,(5/165) + 10\,(10/165) + 150\,(150/165)$
  $= 22635/165 \approx 137$

Note: E[Y] is students' perception of class size

- But E[X] is what is usually reported by schools!

Different expectation if we had 55, 55, 55 student classes

→ this expectatiaon would have been 55

Properties of Expectation (more on this later)

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

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

Law of the Unconscious statistician (LOTUS):

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

Properties of Random Variables

Probability Mass Function:

$$P(X = a)$$

Expectation:

$$E[X]$$

Variance:

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

Learning goals for today

Learning Goals

1. Know what is meant by 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.

Have a Great Weekend!