



Computer  
Science

# CSC380: Principles of Data Science

**Probability 1**

**Xinchen Yu**

- What is probability?
- Events
- Calculating probabilities

# What is probability?

# What is probability?

- Suppose I flip a coin, What is the probability it will come up heads? Most people say 50%, but why?
- “Nolan’s new movie is coming out next weekend! There’s a 100% chance you’re going to love it.”



# Interpreting probabilities

Basically two different ways to interpret:

- Objective probability
  - based on logical analysis or long-run frequency of an event. It's derived from known facts, symmetry, or repeated experiments.
- Subjective probability
  - based on personal belief, opinion, or information about how likely an event is, especially when there's uncertainty or limited data.

# Subjective probability

- Probabilities aren't in the world itself; they're in our knowledge/beliefs about the world.
- Can assign a probability to the truth of any statement that I have a degree of belief about.

We will focus on **objective probability** in this class.

# Objective probability

- The probability of an event represents the long run proportion of the time the event occurs under repeated, controlled experimentation.
  - e.g. 00011101001111101000110
- Famous experiments in history on coin tosses

Experimenter	# Tosses	# Heads	Half # Tosses
De Morgan	4092	2048	2046
Buffon	4040	2048	2020
Feller	10000	4979	5000
Pearson	24000	12012	12000

# Outcome, Event and Probability



# Outcome

- Outcome is a **single** result/observation of a **random experiment**.
- Experiment 1: You flip a coin once.
  - The outcomes are: "Heads" or "Tails"
- Experiment 2: You roll a 6-sided die.
  - The outcomes are: 1, or 2, or 3, or 4, or 5, or 6
- Experiment 3: You tap "shuffle play" on your favorite Spotify playlist with 100 songs and the first song played.
  - The outcomes are: the 1st song in the list, or the 2nd song, ....

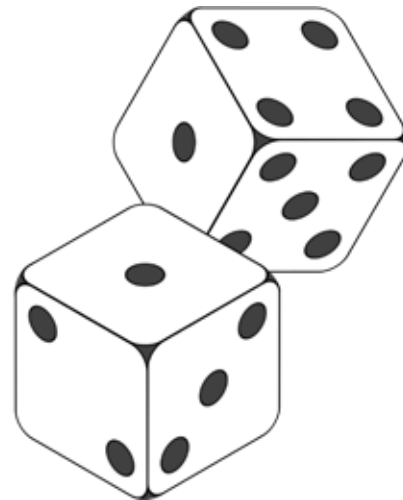
# Random Events and Probability

***Suppose we roll two fair dice...***



***Suppose we roll two fair dice...***

- ◆ What are the possible outcomes?
- ◆ What is the *probability* of the following:
  - ◆ rolling **even** numbers?
  - ◆ having two numbers sum to 6?
  - ◆ If one die rolls 1, the second die also rolling 1?



***...this is a random process.***

How to formalize all these quantitatively?

# The Sample Space

- The set of **all** possible outcomes of a random experiment is called the sample space, written as  $S$ .
- In math, the standard notation for a set is to write the individual members in curly braces:
  - $S = \{\text{Outcome1}, \text{Outcome2}, \dots, \}$
- Useful to visualize the sample space with an actual space.

# The Sample Space

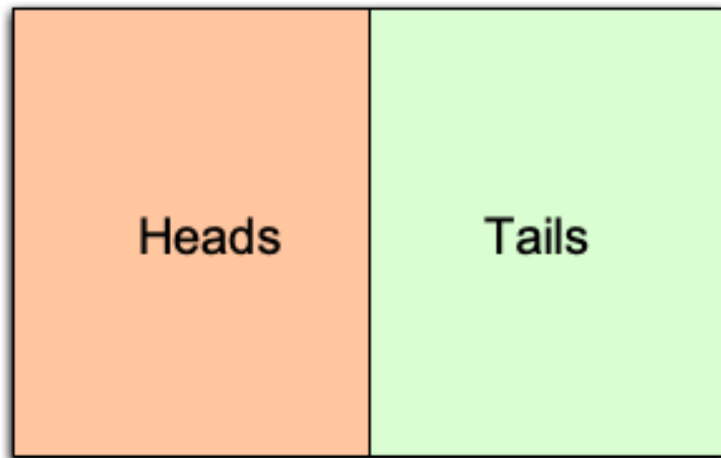


Figure: Visualization of a Sample Space

# Examples of Sample Spaces

What's the sample space for a single coin flip?

- $S = \{\text{Heads}, \text{Tails}\}$



# Examples of Sample Spaces

What is the sample space of rolling a die?

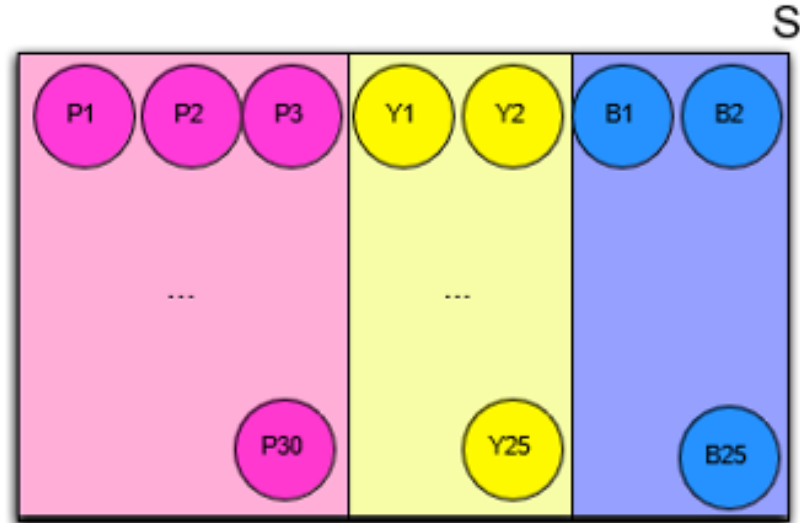
- $S = \{1, 2, 3, 4, 5, 6\}$

1	2	3
4	5	6

# Examples of Sample Spaces



What is the sample space of drawing a ball out of a box containing 30 pink, 25 yellow, and 25 blue balls?

- $S = \{P1, P2, \dots, P30, Y1, \dots, Y25, B1, \dots, B25\}$





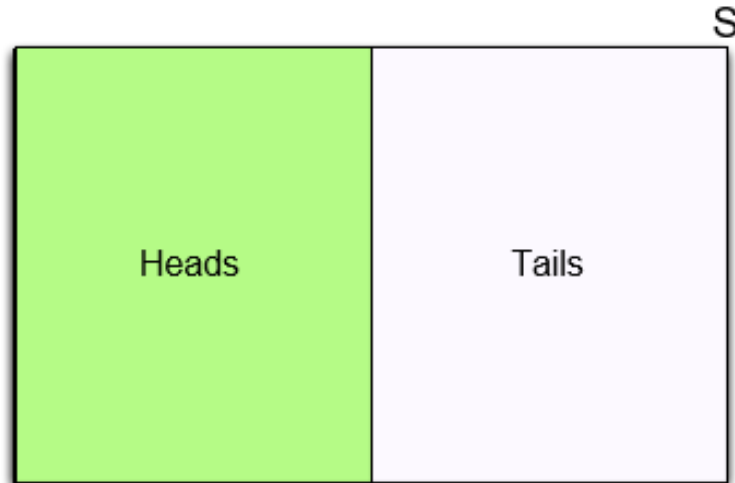
# Events

- An event  $E$  is a **subset** of the sample space.
- An event  $E$  is a **set** of outcomes
  - When we make a particular observation, it is either “in”  $E$  or not.
  - Helpful to think about events as propositions (TRUE/FALSE): TRUE when the outcome is among the elements of the event set, and FALSE otherwise.
  - Is 4 in event  $E = \{2, 4, 6\}$ ?  YES  $\rightarrow$  the **proposition is TRUE**
  - Is 4 in event  $F = \{1, 3, 5\}$ ?  NO  $\rightarrow$  the **proposition is FALSE**

# Examples of Events

What's the event set corresponding to the following propositions?

- “The coin comes up heads”
- $E = \{\text{Heads}\}$



# Examples of Events

What's the event set corresponding to the following propositions?

- “The die comes up an even number”
- $E = \{2, 4, 6\}$

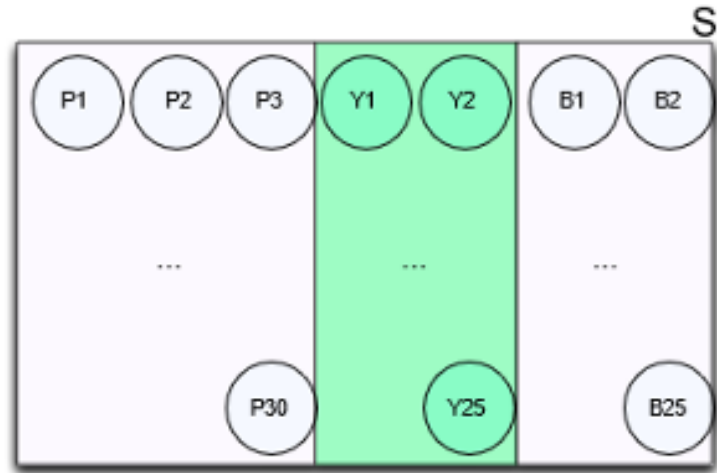
1	2	3
4	5	6

S

# Examples of Events

What's the event set corresponding to the following propositions?

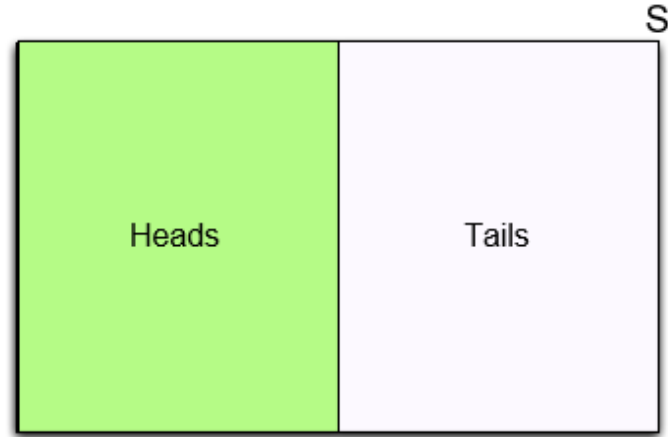
- “A yellow ball is chosen”
- $E = \{Y1, Y2, \dots, Y25\}$



# Calculating Probabilities

# Calculating probability

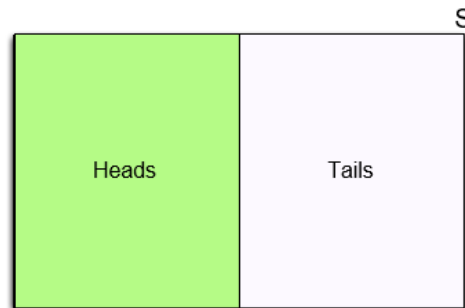
- We can think of the probability of an event  $E$  as its area, where  $S$  (sample space) always has a total area of 1.0
- So, the probability of  $E$  is the fraction of  $S$  that it takes up.



# Calculating probability using symmetry

- If we have a sample space for which every outcome is equally likely (has the same area), then we can find event probabilities easily.
- Classic probability model to get the probability of an event  $E$ :

$$P(E) = \frac{\text{\#outcomes in } E}{\text{\#outcomes in } S}$$



# Probability as Area

What is the probability of

$$P(E) = \frac{\text{\#outcomes in } E}{\text{\#outcomes in } S}$$

- Rolling a fair die and see an even number?

- $E = \{2, 4, 6\}$

- $$P(E) = \frac{\text{\#}\{2, 4, 6\}}{\text{\#}\{1, 2, 3, 4, 5, 6\}} = \frac{3}{6} = \frac{1}{2}$$

1	2 $P(2) = 1/6$	3
4 $P(4) = 1/6$	5	6 $P(6) = 1/6$

$$P(S) = 1$$

$$P(\text{Even}) = P(2) + P(4) + P(6) = 3/6$$

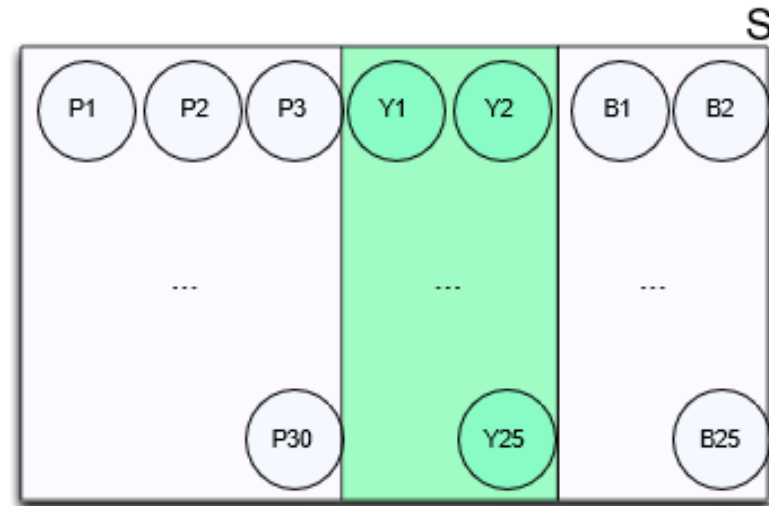


# Probability as Area

What is the probability of

- Selecting a yellow ball?
- $E = \{Y1, Y2, \dots, Y25\}$

- $$P(E) = \frac{\# E}{\# S} = \frac{25}{30+25+25} = \frac{5}{16}$$

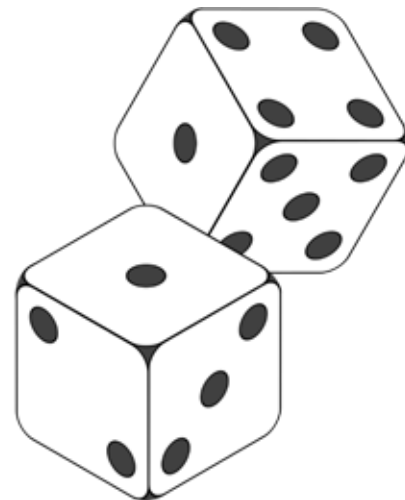


$$P(\text{Yellow}) = P(Y1) + \dots + P(Y25) = 25 * (1/80)$$

- Suppose we throw two fair dice
  - What is the sample space  $S$  (space of all possible outcomes)?

Event  $E$ : the two dice's outcomes sum to 6

- What is the size of  $E$ ?
- What is the probability of  $E$ ?

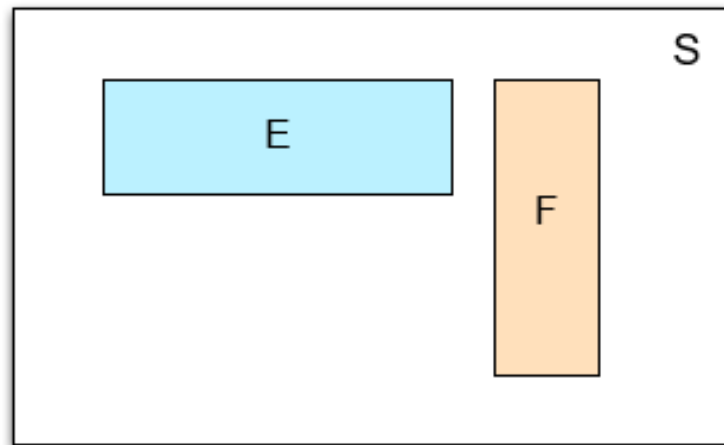


# Disjoint events

- In general, breaking an event into *disjoint events* preserves the total probability
- **Disjoint:**  $E$  and  $F$  are *disjoint* if they cannot happen simultaneously,
- e.g.  $E = \{2, 4\}$ ,  $F = \{1, 3, 5\}$

- In such cases,

$$P(E \text{ or } F) = P(E) + P(F)$$



# Elementary events

- Notice that we can find the total probability of an event by breaking it into pieces and adding up the probabilities of the pieces:

$$P(\text{Even}) = P(2) + P(4) + P(6) = 3/6$$

1	2 $P(2) = 1/6$	3
4 $P(4) = 1/6$	5	6 $P(6) = 1/6$

- These pieces are called ‘**elementary events**’:  
Events that correspond to exactly one outcome
- These elementary events are disjoint events

# Partition

- We say that disjoint events  $E_1, \dots, E_n$  form a *partition* of  $E$  if any outcome in  $E$  lies in exactly one  $E_i$
- e.g.
  - {Fr.}  $E_1$ , {Soph.}  $E_2$  form a partition of {Lower division}  $E$
  - {Fr.}, {Soph.} {J.} {Sen.} form a partition of  $S$  (sample space)

Freshmen	Sophomores	Juniors	Seniors
----------	------------	---------	---------

# Probability as Area

For disjoint events  $E, F$ :

$$P(E \text{ or } F) = P(E) + P(F)$$

If disjoint events  $E_1, \dots, E_n$  forms a partition of  $E$  :

$$P(E) = P(E_1) + P(E_2) + \dots + P(E_n)$$

If disjoint events  $E_1, \dots, E_n$  forms a partition of  $S$  , for event  $F$  (law of total probability):

$$P(F) = P(E_1, F) + P(E_2, F) + \dots + P(E_n, F)$$

Notation:  $P(A, B)$  is a shorthand for  $P(A \text{ and } B)$

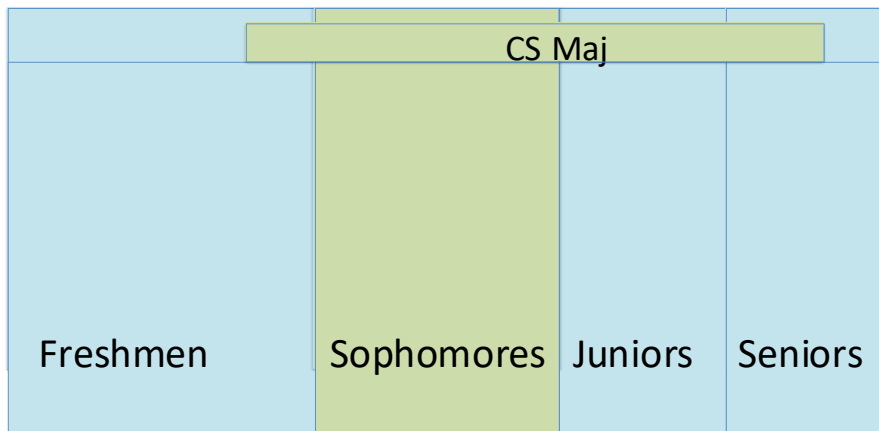
# Probability as Area

## Examples

- $P(\text{CS}) = P(\text{Fr.}, \text{CS}) + P(\text{Soph.}, \text{CS}) + P(\text{J.}, \text{CS}) + P(\text{Sen.}, \text{CS})$

Notation:  $P(A, B)$  is a shorthand for  $P(A \text{ and } B)$

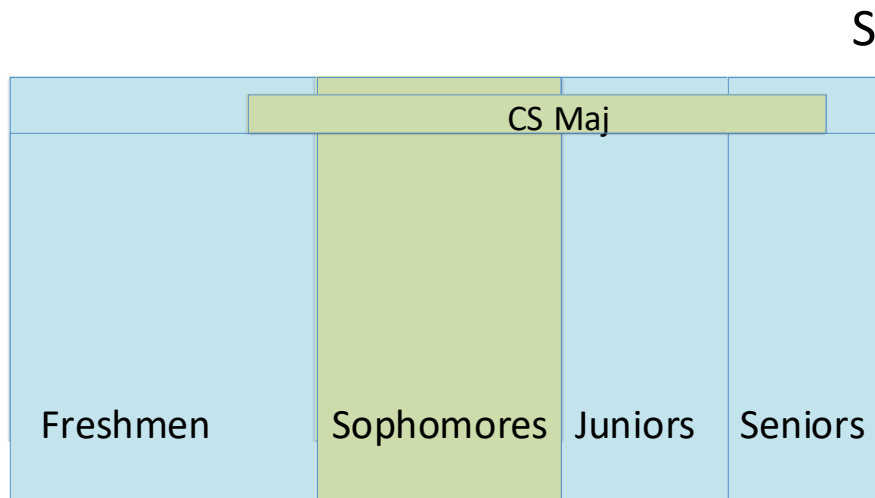
- $P(\text{Soph.}) = P(\text{CS}, \text{Soph.}) + P(\text{nonCS}, \text{Soph.})$   
S



# Probability as Area

If events  $E, F$  are non-disjoint, what is  $P(E \text{ or } F)$ ?

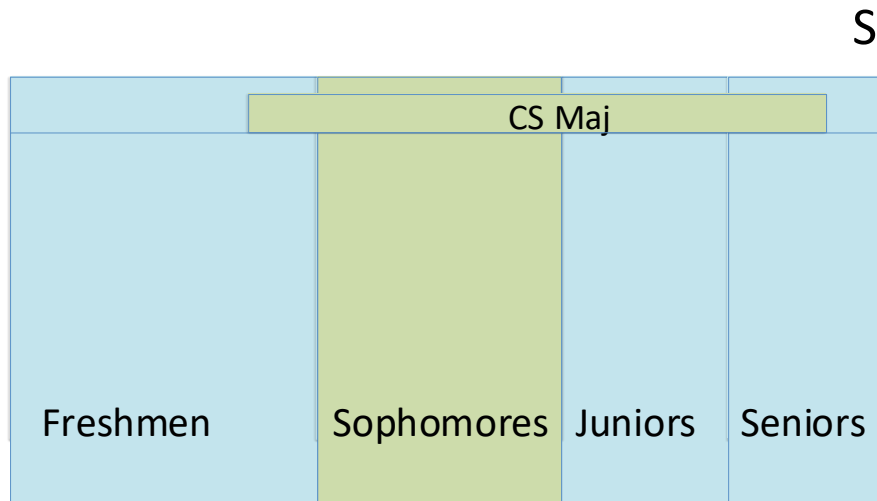
- $P(\text{Soph. or CS})$ ?
- Note: events “Sophomore” and “CS Major” may overlap





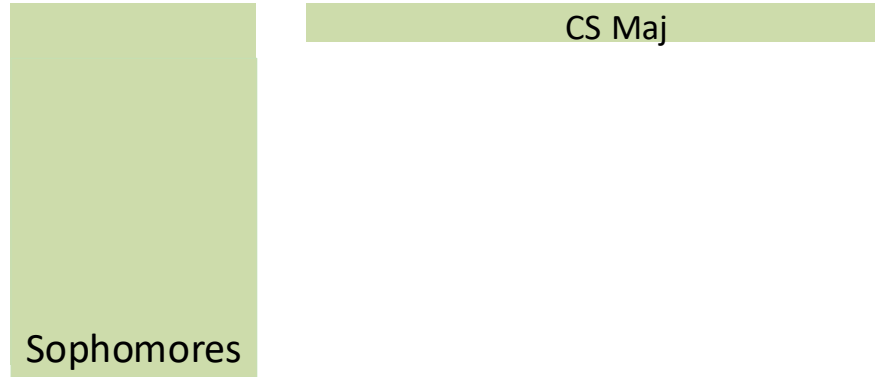
# Probability as Area

- $E = \{ \text{Soph OR CS} \}$
- Is  $P(E) = P(\text{Soph}) + P(\text{CS})$ ?
  - No
- Which one is larger?
  - Let's see..

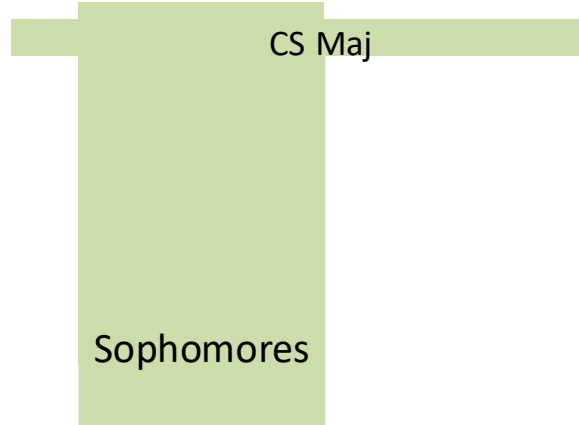


# Probability as Area

$$P(\text{Soph}) + P(\text{CS}) =$$



$$P(\text{Soph or CS}) =$$



# Inclusion-Exclusion Principle

**Inclusion-Exclusion Principle** For any events  $E$  and  $F$ ,

$$P(E \text{ or } F) = P(E) + P(F) - P(E \text{ and } F)$$

Accounting for overlap  
between  $E$  and  $F$

S

