Intro to Probability

Statistics and Applications

Quiz 2 on Gradescope

Some comments regarding variance and standard deviation.

- Why square difference from mean? This is still debated in the field!
  - Gorard, Stephen. "Revisiting a 90-year-old debate: the advantages of the mean deviation." *British Journal of Educational Studies* 53.4 (2005): 417-430.
    - Recounts back-and-forth publications of statisticians Fisher and Eddington from the early 1900’s on standard vs. mean deviation (SD vs. MD). Fisher contends SD is best.
  - Gerstenberger, Carina, and Daniel Vogel. "On the efficiency of Gini’s mean difference." *Statistical Methods & Applications* 24 (2015): 569-596.
    - Tukey found MD to be better in certain cases in the 1960’s.
      - Tukey, John Wilder. "A survey of sampling from contaminated distributions." *Contributions to probability and statistics* (1960): 448-485.
  - Pinsky, Eugene, and Sidney Klawansky. "MAD (about median) vs. quantile-based alternatives for classical standard deviation, skewness, and kurtosis." *Frontiers in Applied Mathematics and Statistics* 9 (2023): 1206537.
    - Alternatives to SD can be better for certain distributions and problems.

Science and technology are continually developing.

- There are very good reasons for why things are the way they are beyond just memorizing an equation.
- At some level, it can come down to someone just making a choice that has long-term impacts.
    - Examples from computer history:
        - The web's most important decision
        - Folklore: swedish campground

Degrees of Freedom

- Number of independent values that can vary without breaking any constraints.

	Original data
	5
	15
	10
	9
	6
Mean	**9**
Standard deviation	**3.52**

Changed
7
14
9
?
?
9
3.52

Constrained
7
14
9
3.72
11.28
9
3.52

Degrees of Freedom

- William Sealy Gosset, who often wrote under the pen name of “Student”:
  - Developed a number of statistical tests for data analysis.
  - Found that the number of values calculated from your data removes that number of values that can vary in the data.
    - Called calculated values “constraints” and values that can vary “degrees of freedom”

Probability

Motivation for Probability.

- We collect data to help answer a question.
- Visualizing and describing data are helpful, especially for **populations**.
- In almost everything in real life, we work with **samples** only.
  - With a sample, you must ask how your data may relate to the actual population.
- Thus, we study probability.
  - Using our sample data and math, we want to make predictions about future data.

Probability.

Probability is a measure of how likely an event is to occur.

Some Terminology

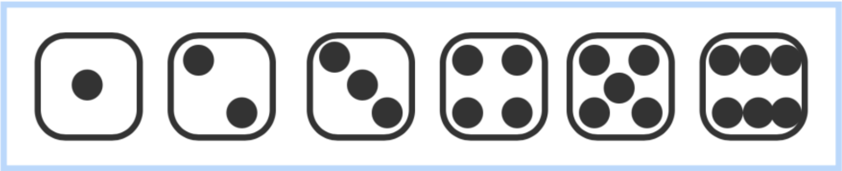
- Probability: how likely an **event** is to occur.
- Event: a subset of a **sample space**.
- Sample Space: the set of all possible **outcomes**.
- Outcome: the result of an **experiment**.
- Experiment: a procedure resulting in an **outcome**.

Experiment and Outcome Example.

- Consider that we have a set of 6-sided dice.
- Rolling one of the dice would be an experiment.
- The side that ends facing up will be the outcome.

Sample Space Example.

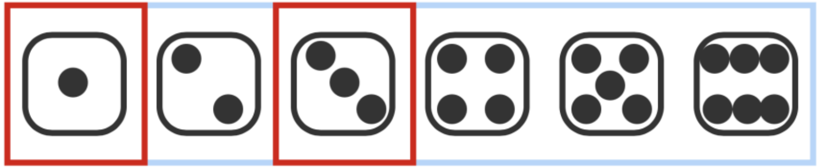
- For a single 6-sided die, the sample space is the set of numbers on the die:
  - $S = \{1,2,3,4,5,6\}$
- This is the set of all possible outcomes; it will land on one of these sides.



Sample space

Event Example.

- An event is a subset of this sample space.
- For instance, the event containing the possible outcomes that the number is odd and less than 4.
  - $A = \{1,3\}$



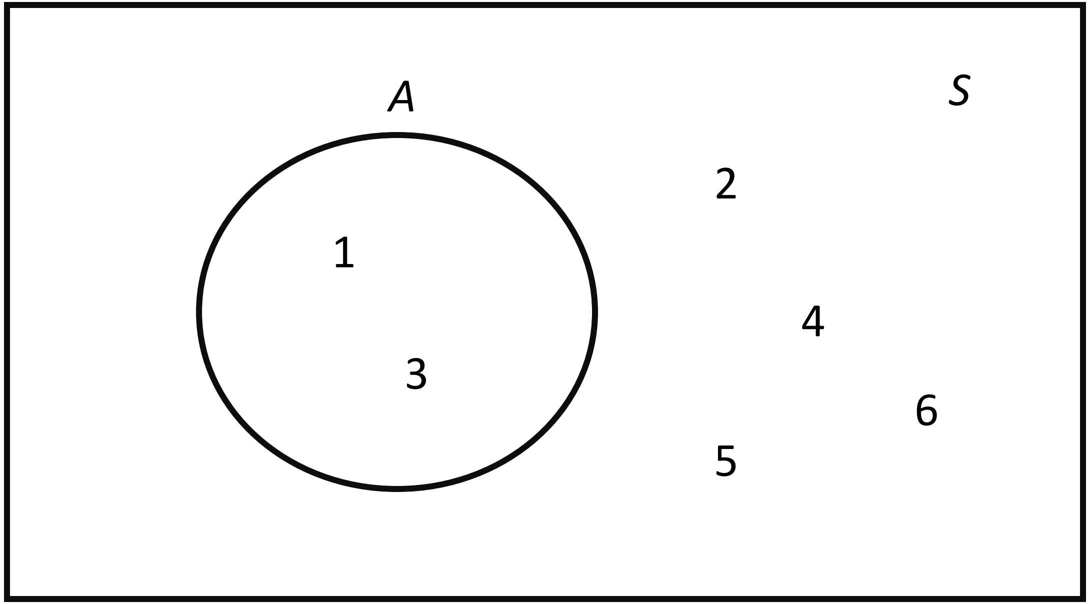
- In this case, A is a **compound event** because it contains multiple outcomes.
- A **simple event** is an event containing a single outcome.
  - E.g., the event you roll a 1 could be written as $C = \{1\}$

Set Notation.

- We often express sample spaces and events as **sets**.
- A set is a data structure that stores unique elements of the same type.
  - They may be sorted or unsorted, depending on the implementation. Python sets are unsorted by default.
- Examples: $S = \{1,2,3,4,5,6\}$ and $A = \{1,3\}$
- An empty set has no outcomes and would simply be written as {} or $\emptyset$

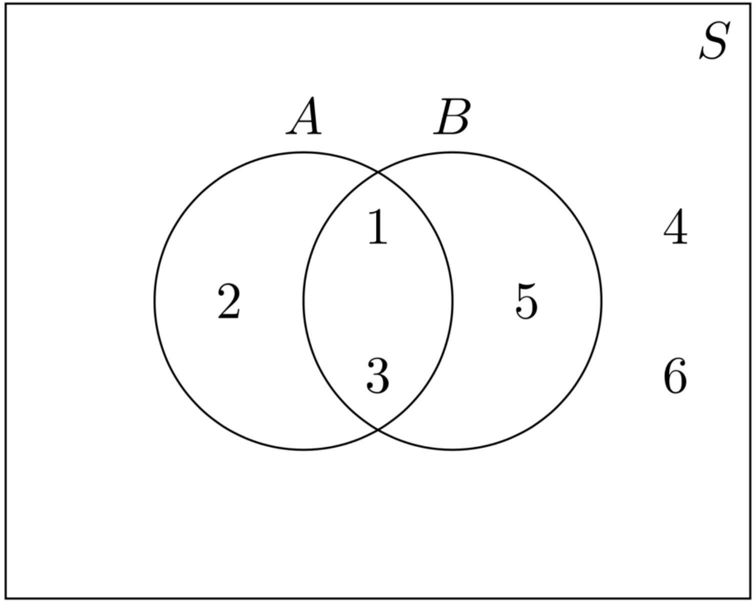
Venn Diagrams.

- One common way to visualize sets is to use Venn Diagrams.
- For sample space $S = \{1,2,3,4,5,6\}$ and event $A = \{1,3\}$:



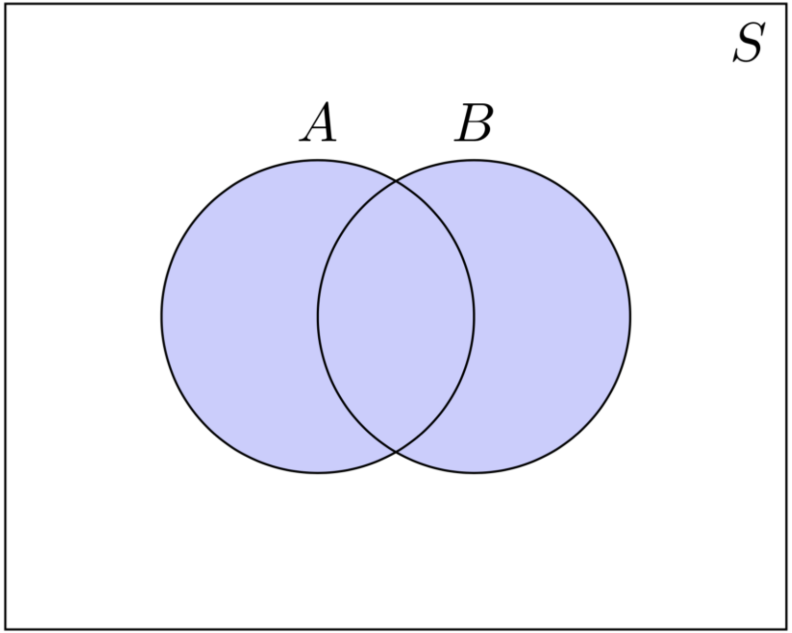
Venn Diagrams with Multiple Events.

- Most commonly, we see Venn Diagrams with multiple events to help visualize how they relate to each other.
- For $S = \{1,2,3,4,5,6\}$ and events $A = \{1,2,3\}$ and $B = \{1,3,5\}$:



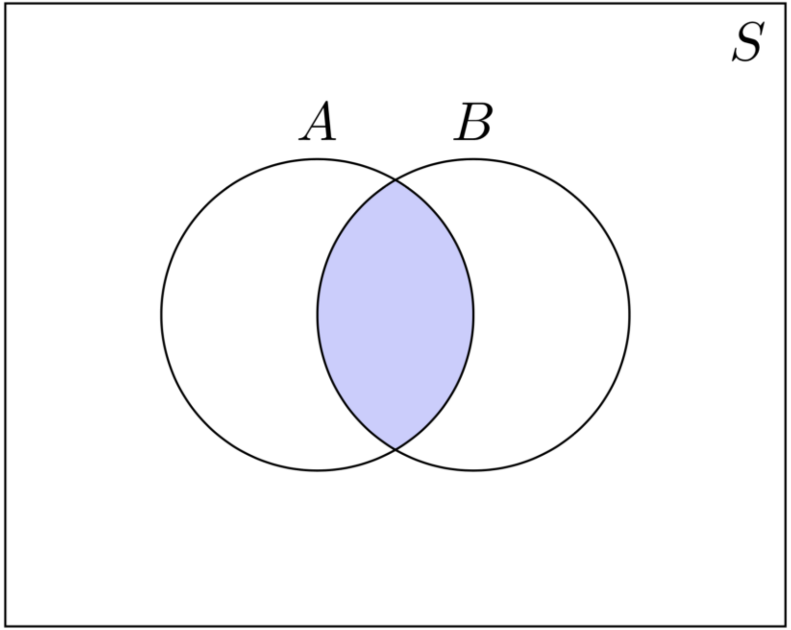
Set Operations: Union

- The **union** of two sets is written with $\cup$.
- $A \cup B$ is the event including outcomes of A or B or both.



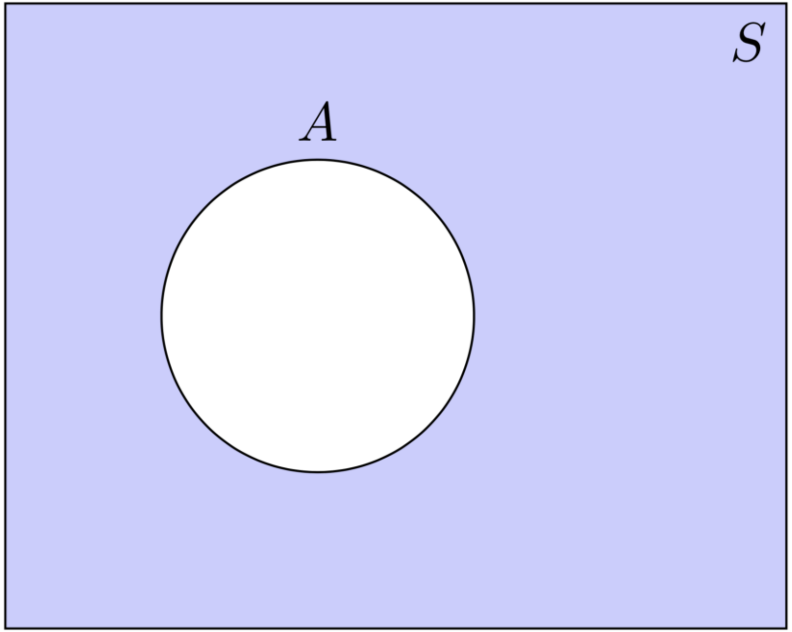
Set Operations: Intersection

- The **intersection** of two sets is written with $\cap$.
- $A \cap B$ is the event including outcomes that are in both A and B.



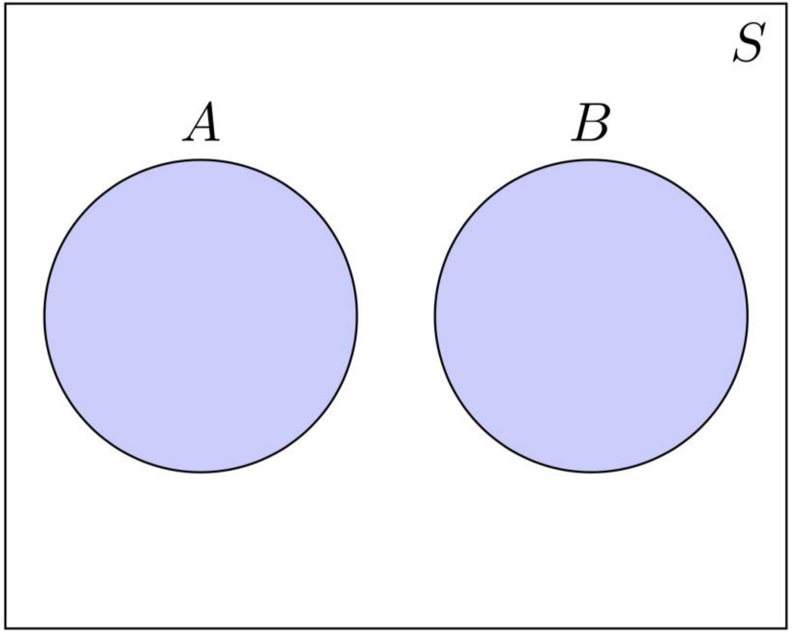
Set Operations: Complement

- The **complement** of A is written as $\bar{A}$ or $\neg A$ or A'.
- It denotes all the outcomes that are *not* in A.

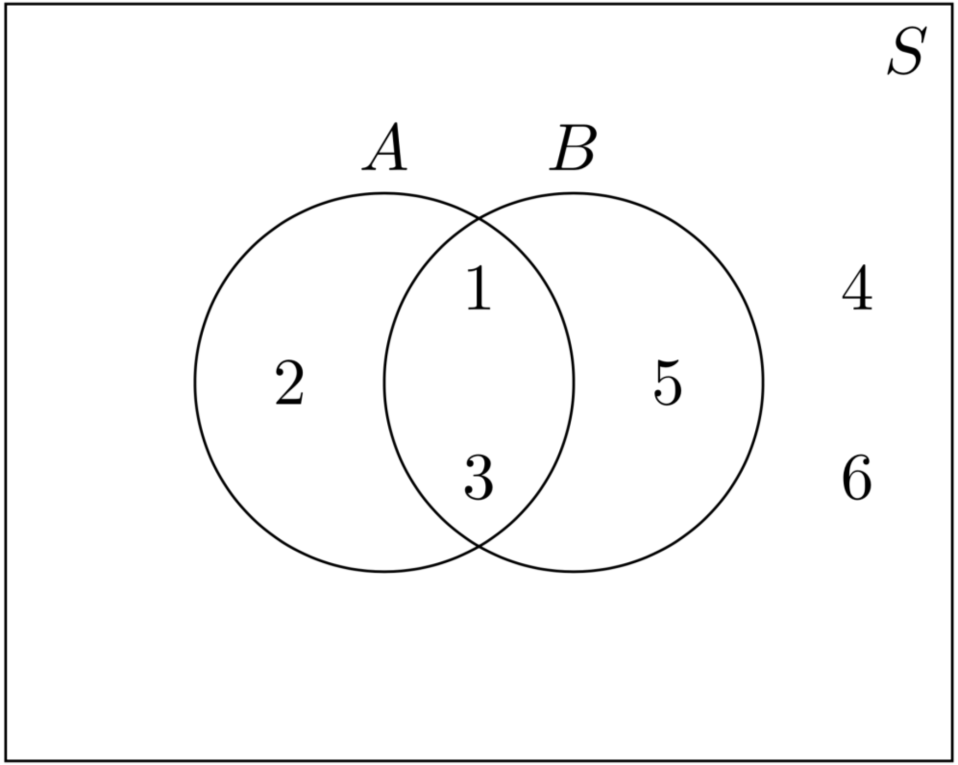


Mutual Exclusion

- Two events are considered **mutually exclusive** if they contain no shared outcomes.
  - i.e., the Venn diagram for the subsets has no overlap.



These events are not mutually exclusive:



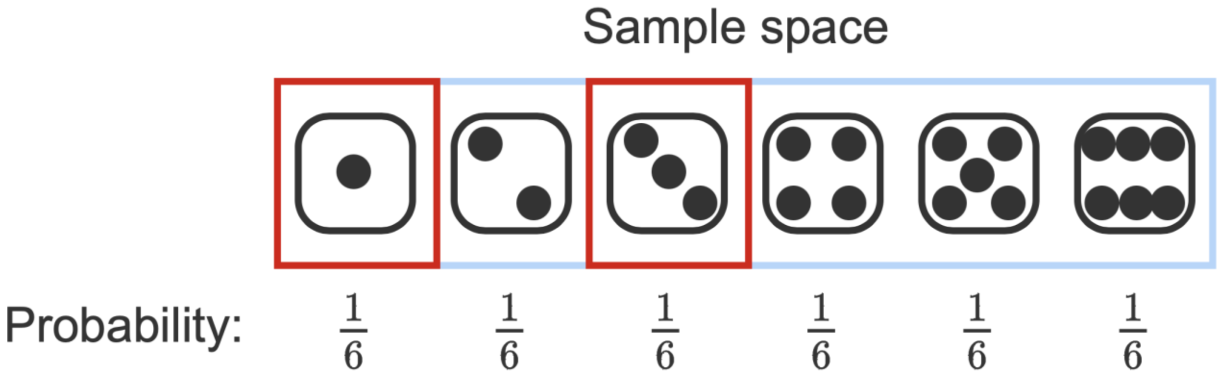
Let’s bring all this knowledge and terminology back to probability!

The Axioms of Probability

- Three fundamental properties form the basis of probability theory. They are called the probability axioms or the Kolmogorov axioms.
- For *A* and *B* mutually exclusive events in sample space *S*:

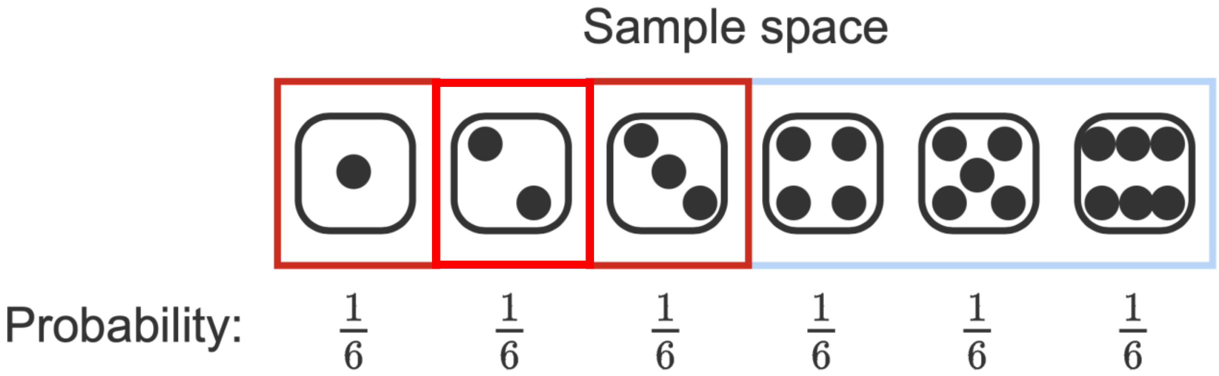
1. **$0 \leq P(A) \leq 1$**
   The probability of an outcome from an event is between 0% and 100%.
2. **$P(S) = 1$**
   The probability of *an* outcome at all is 100%. (The dice will show a number)
3. **$P(A \cup B) = P(A) + P(B)$**
   For mutually exclusive events, the probability of an outcome from either event is the sum of probabilities of each event.

- For a 6-sided fair die with sample space $S = \{1,2,3,4,5,6\}$, all sides have equal probability.
- Given the events $A = \{1\}$ and $B = \{3\}$, what is the probability of the event $C = A \cup B$?
  - That is, what is the likelihood that the die will show either 1 or 3?
- These are mutually exclusive—the outcome will be in one event or the other—so we can use axiom 3.
- $P(A \cup B) = P(A) + P(B) = 1/6 + 1/6 = 2/6 = 1/3$



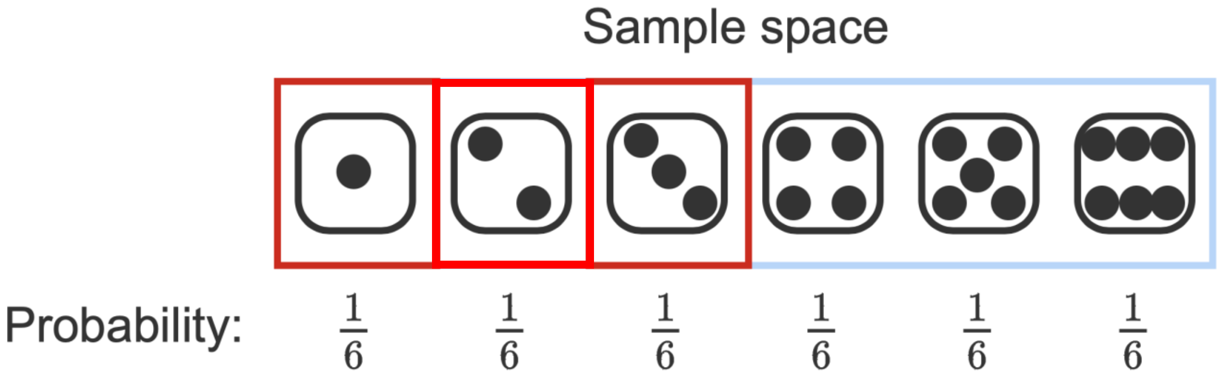
What if the events are A = {1,3} and B = {2}?

- These are still mutually exclusive:
  - P(A ∪ B) = P(A) + P(B)
- We just established that P({1,3}) = 2/6, so P(A ∪ B) = 2/6 + 1/6 = 3/6 = 1/2
- This should make sense when we look at the image since it's asking P({1,2,3}).



What if the events are **not** mutually exclusive?

- Let's say A = {1,2,3} and B = {3}.
- Can we directly use P(A ∪ B) = P(A) + P(B)?
- P(A) + P(B) = 1/2 + 1/6 = 4/6 = 2/3
- This is wrong because we know that it is not a 2/3 chance to land on {1,2,3} on a 6-sided die!



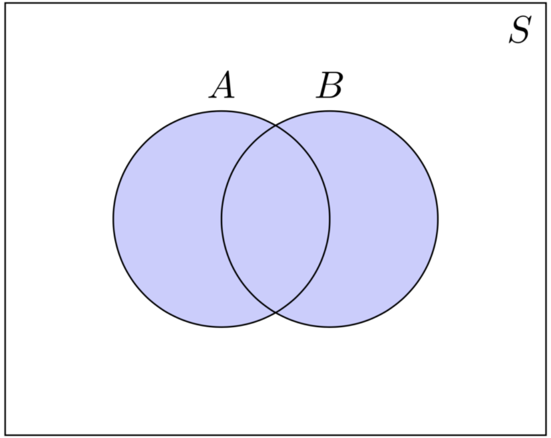
The Addition Rule

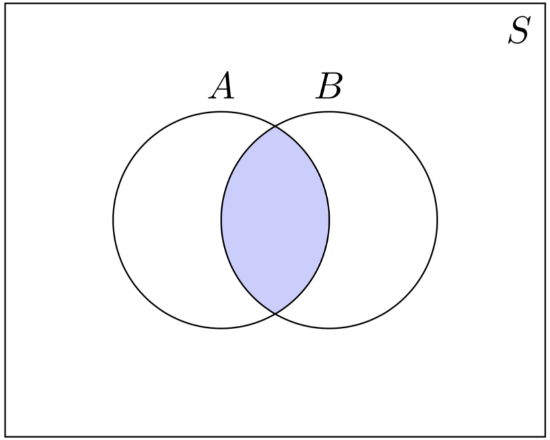
- A generalization of axiom 3 is the addition rule for any events *A* and *B*:
- $P(A \cup B) = P(A) + P(B) – P(A \cap B)$

- For our example, we must subtract the probability of $P(A \cap B)$.
    - The intersection of A = {1,2,3} and B = {3} is {3}.
    - The probability of the outcome being 3 is 1/6.
- For our non-mutually exclusive sets
- $P(A \cup B) = P(A) + P(B) – P(A \cap B) = 1/2 + 1/6 – 1/6 = 1/2$ as expected!

Summary

- Probability terminology:
  - Sample space is the set of all possible outcomes of an experiment.
  - Event is a subset of the sample space.
- Venn Diagrams are a useful way to visualize sample spaces, events, and their relationships to one another:



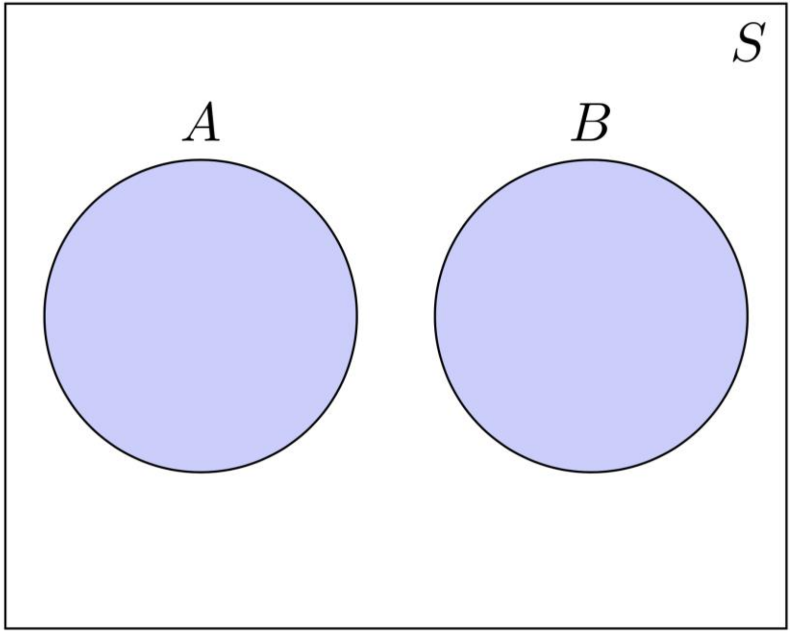


Summary

- Axioms of probability:
  - $0 \leq P(A) \leq 1$
  - $P(S) = 1$
  - $P(A \cup B) = P(A) + P(B)$ for mutually exclusive events A and B
- In cases of random selection with equal chances:
  - Probability = (# desired outcomes)/(# possible outcomes)
- Addition Rule for any events A and B:
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Mutual Exclusion

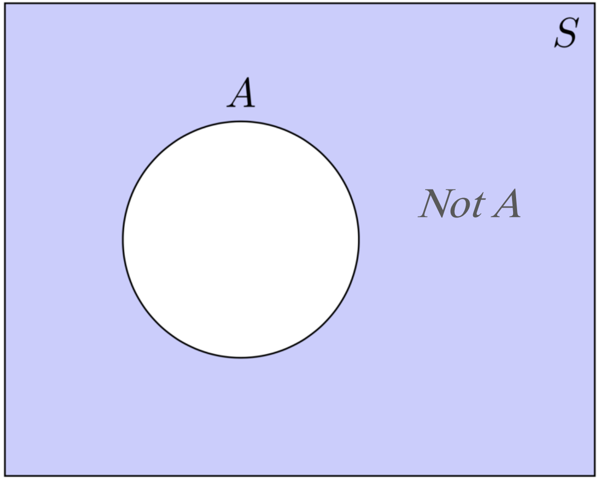
- Two events are considered **mutually exclusive** if they contain no shared outcomes.
  - We call this disjoint because $A \cap B = \emptyset$



Jupyter Notebook Example

Subtraction Rule

- Recall the axiom P(S) = 1.
  - The sample space covers all possible outcomes, so the probability of an outcome is 100%.
- For event A, its complement $\bar{A}$ is the set of all outcomes **not** in A.
  - An event and its complement will contain all possible outcomes between them. E.g., the union of A and $\bar{A}$ is the set of all possible outcomes, S.
- Thus, P(S) = P($A \cup \bar{A}$) = 1



Subtraction Rule

- A and $\bar{A}$ are mutually exclusive by definition, so we may use the Addition Rule to simplify:
  $P(A \cup \bar{A}) = P(A) + P(\bar{A}) = 1$
  $P(\bar{A}) = 1 - P(A)$
- The Subtraction Rule states that $P(\bar{A}) = 1 - P(A)$.
  - E.g., if there's a 1/6 chance you'll roll a 1 on a fair 6-sided die, there's a 5/6 chance you won't!