Chapter 4: Probability and Combinatorics

DSCC 462

Computational Introduction to Statistics

Anson Kahng

Fall 2022

Probability

Probability

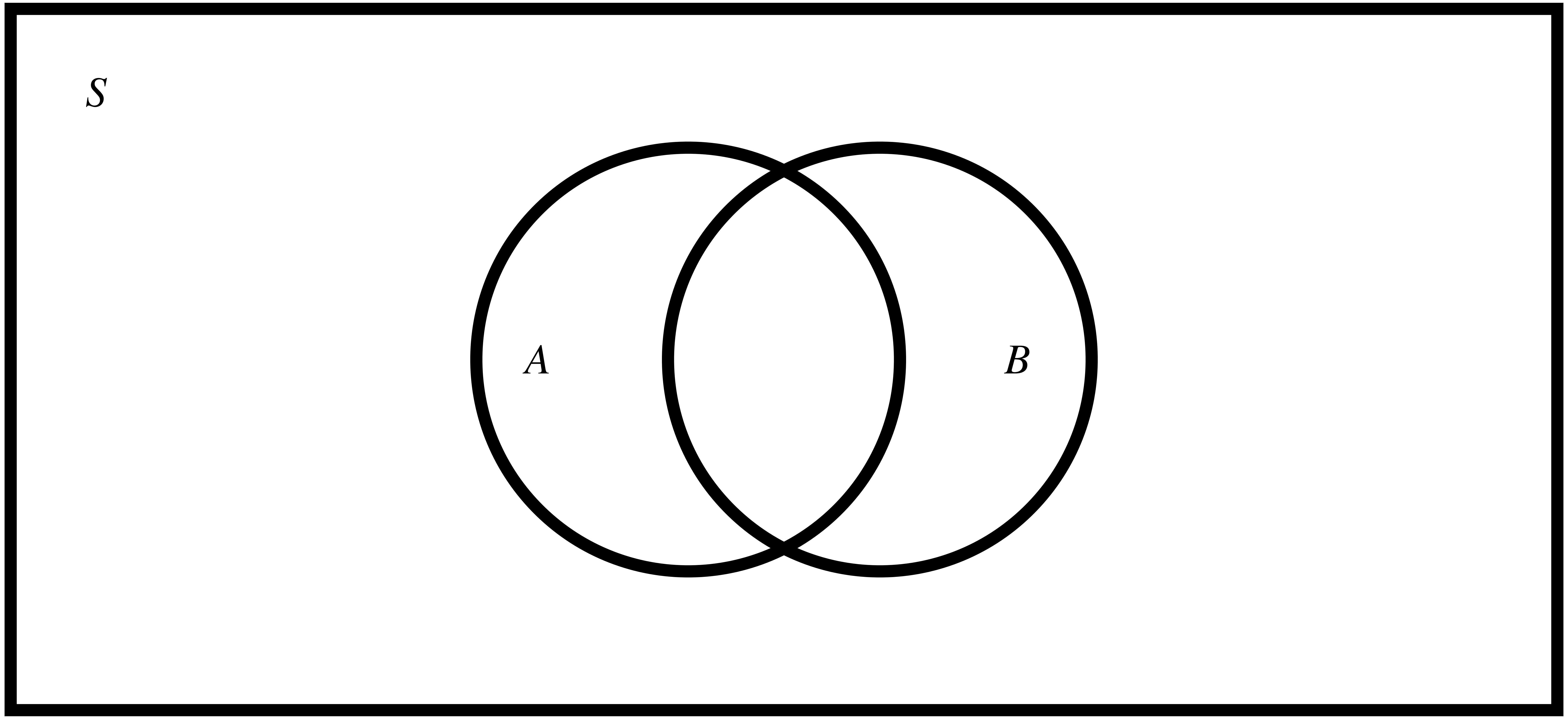
- The outcome that we will observe is often uncertain
    - Flip a coin
    - Draw a card
    - Roll a die
    - Income of a selected individual
- We want to find the *probability* of each event happening
- Probability is the mathematics of random occurrences

Events

- **Sample space**: All possible outcomes that can be observed in a given situation, denoted S
  - Example: Flip of a coin, $S = \{\text{Heads, Tails}\}$
- A *random experiment* occurs when an element of S is randomly selected
- **Event**: The basic element to which probability can be applied
  - “Probability of an event happening”
  - Events can be possible outcomes or observed values
  - Either happens or it does not
- Events are represented by uppercase letters: $A, B, C, \ldots$
- List the event in $\{\ \}$ brackets
- Example: $A = \{\text{roll an even number on a six-sided die}\} = \{2,4,6\}$

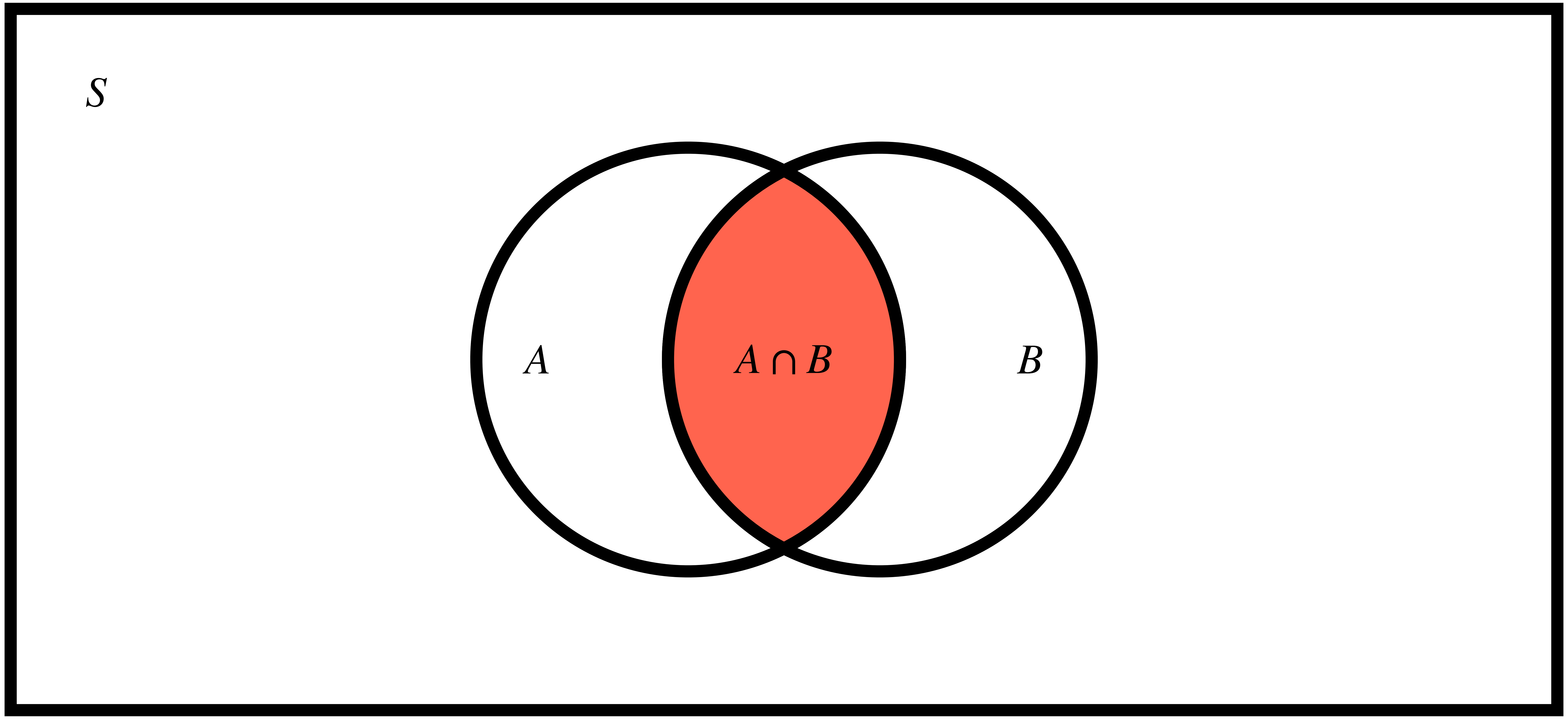
Operations on Events

- Let A and B be events, or subsets of S, where $A \subset S$ and $B \subset S$



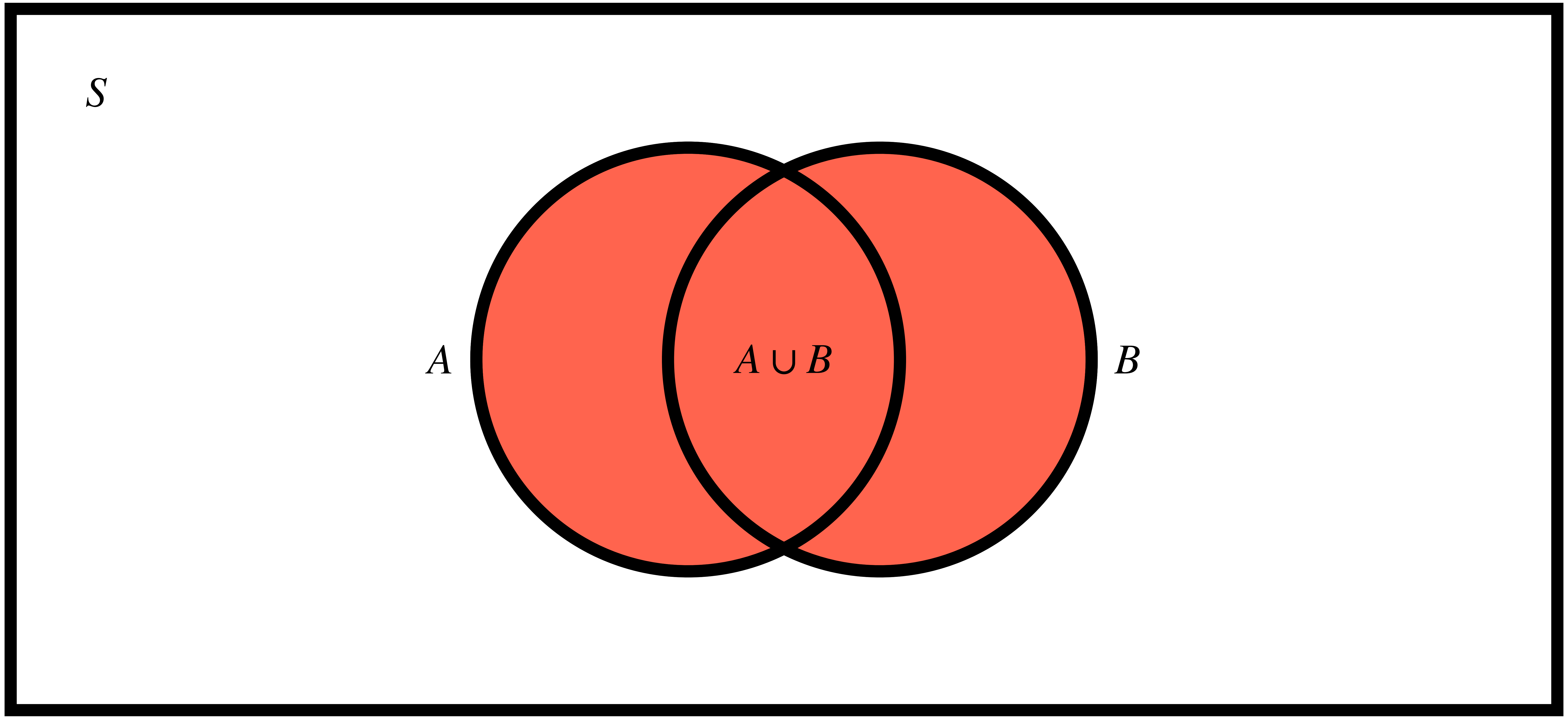
Intersection

- Intersection ($A \cap B$): The event "both A and B", or all elements in S in both A and B



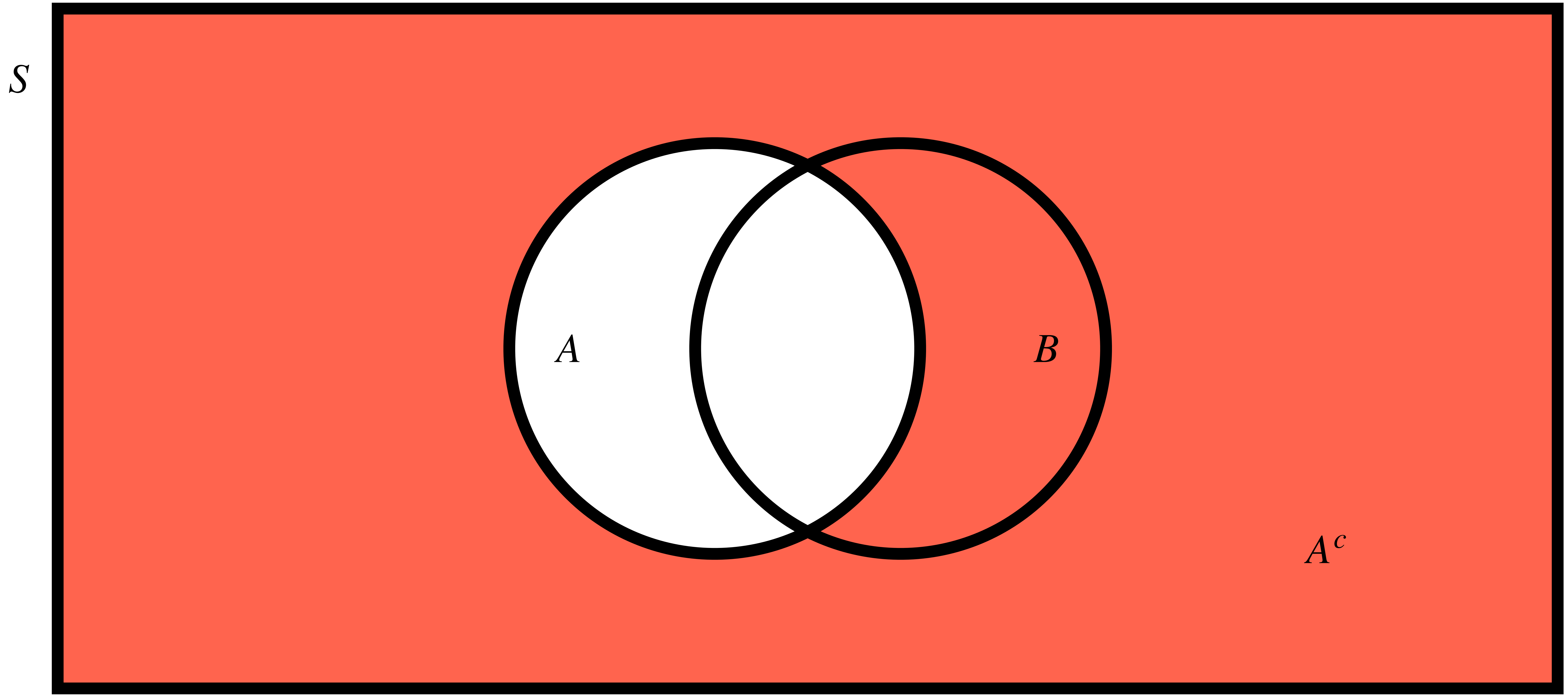
Union

- Union ($A \cup B$): The event "either A or B", or all elements in S in either A or B



Complement

- Complement (A^c, $\overline{A}$, or A'): The event "not A", or all elements in S not in A



Operations Example

- Suppose we have the following, where $A \subset S, B \subset S,$ and $C \subset S$:

$$S = \{1,2,3,4,5,6,7,8\}$$

$$A = \{1,2,3,4\}$$

$$B = \{2,4,6,8\}$$

$$C = \{7,8\}$$

- Evaluate the following expressions:

$$A \cap B =$$

$$(A \cup C) \cap B =$$

$$A^c \cap C =$$

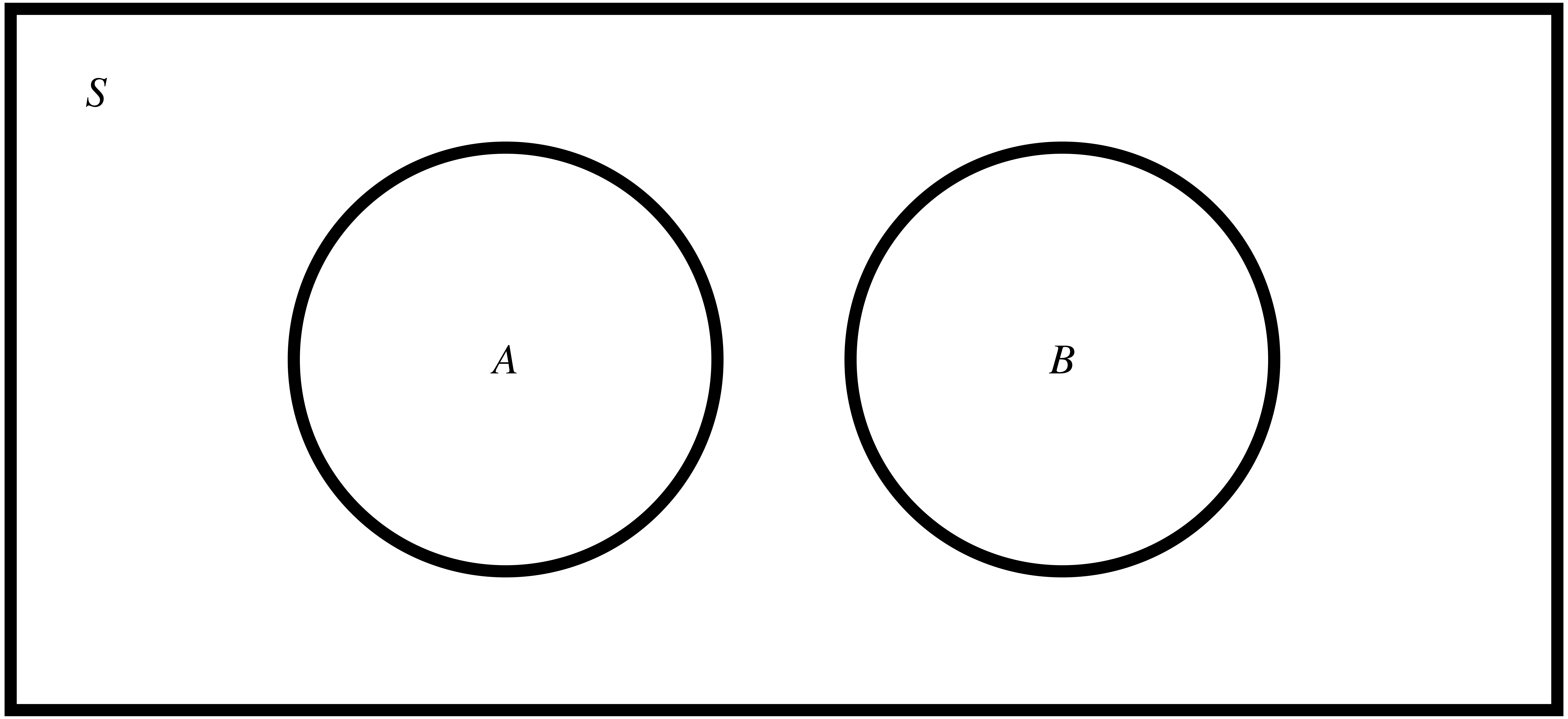
$$(A \cap B^c) \cup C =$$

Operations on Events: De Morgan's Laws

- De Morgan's Laws:
  - $(A \cup B)^c = A^c \cap B^c$
  - $(A \cap B)^c = A^c \cup B^c$

Events

- Null events are events that can never occur, represented as $\varnothing$
- Disjoint or mutually exclusive events are events that cannot occur simultaneously; A and B are disjoint if and only if $A \cap B = \varnothing$



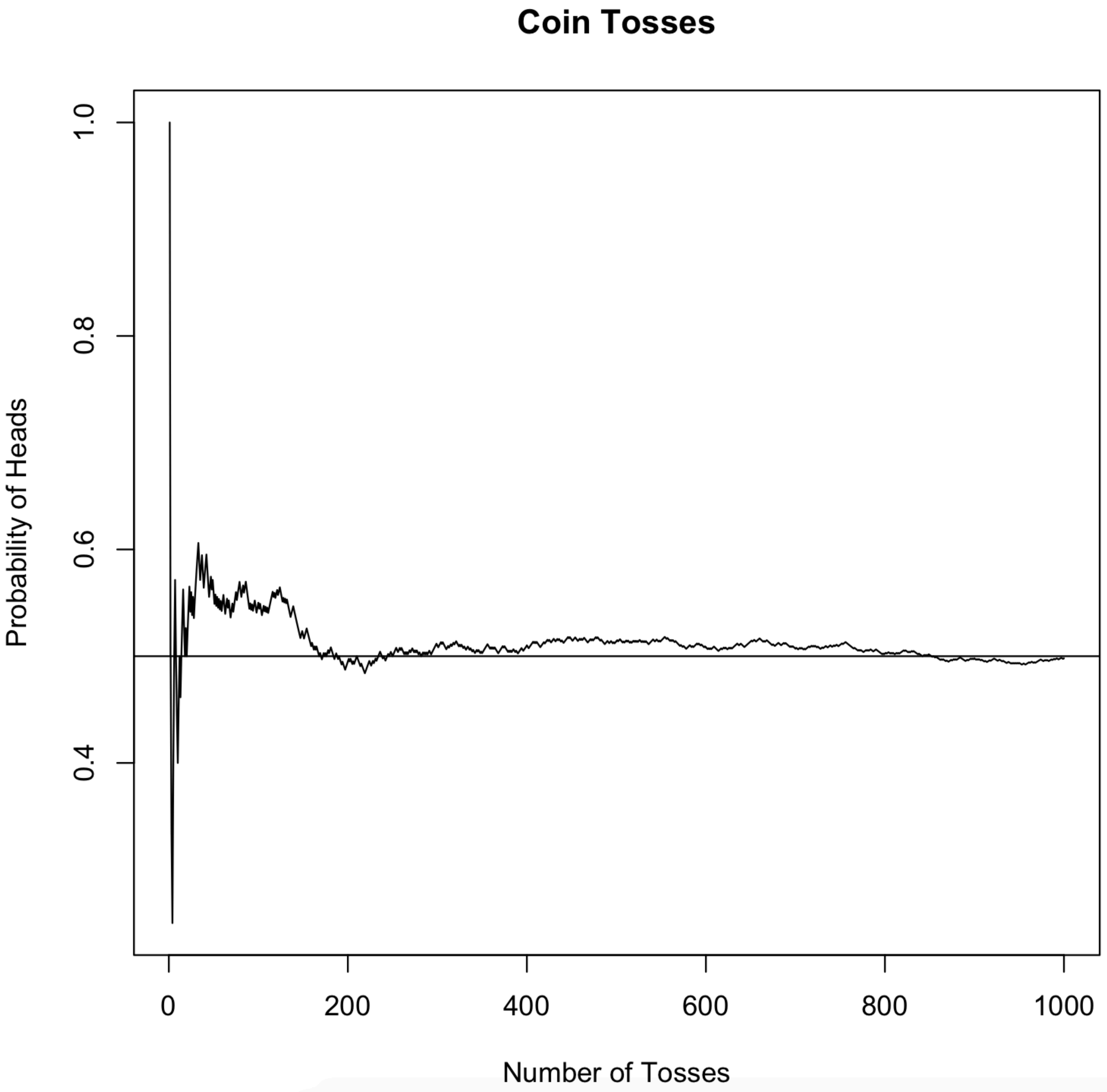
Cardinality

- The *cardinality* of A is the number of elements in the set, denoted $|A|$
- Three types of cardinality:
  - Finite: $|A| < \infty$
  - Countable: $|A| = \infty$ but elements can be listed as $x_1, x_2, \ldots$
  - Uncountable: $|A| = \infty$ and elements cannot be listed as $x_1, x_2, \ldots$

Probability

- **Probability**: If an experiment is repeated n times under identical conditions, and if event A occurs m times, then as n grows large, the ratio m/n approaches a fixed limit that is the probability of event A: $\Pr(A) = \frac{m}{n}$
- Relative frequency of occurrence of an event when repeated many times
- $\Pr(A) = \frac{\text{\# of times } A \text{ occurs}}{\text{total \# of trials}}$

Probability

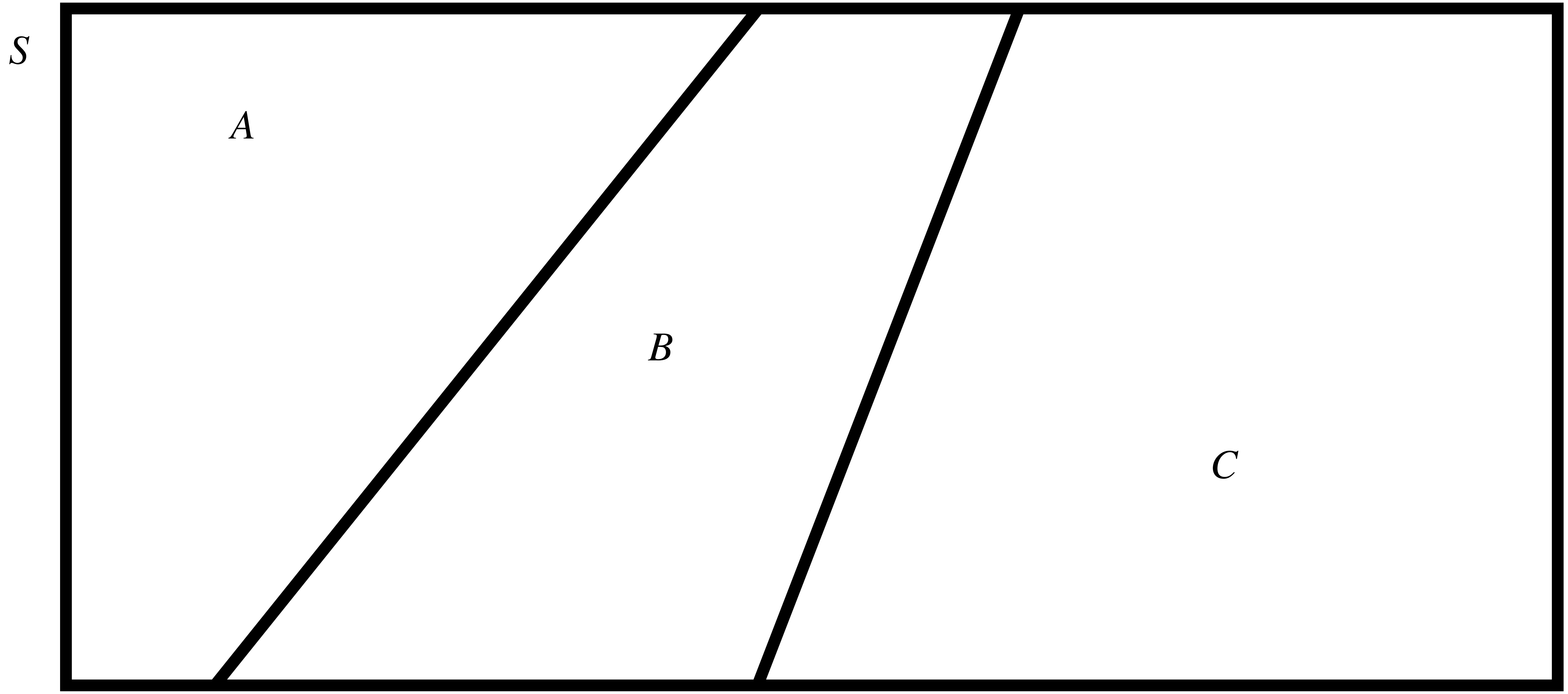


Probability Rules

- $0 \leq \Pr(A) \leq 1$
- $\Pr(S) = 1$
- $\Pr(\varnothing) = 0$
- $\Pr(A^c) = 1 - \Pr(A)$
- If $A \subset B$, then $\Pr(A) \leq \Pr(B)$

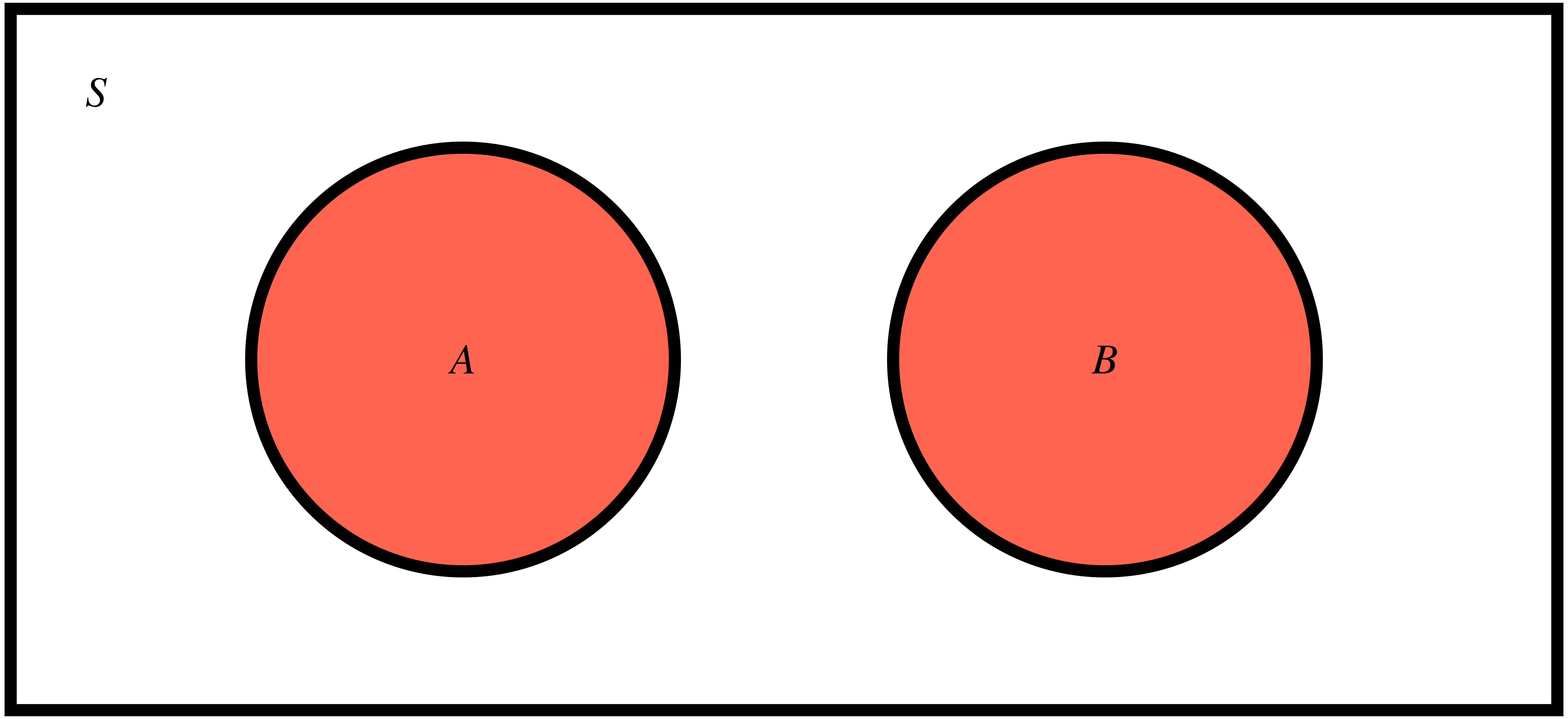
Mutual Exclusivity and Exhaustiveness

- When the probabilities of mutually exclusive events sum to 1, the events are *exhaustive* (i.e., no other possible outcomes)



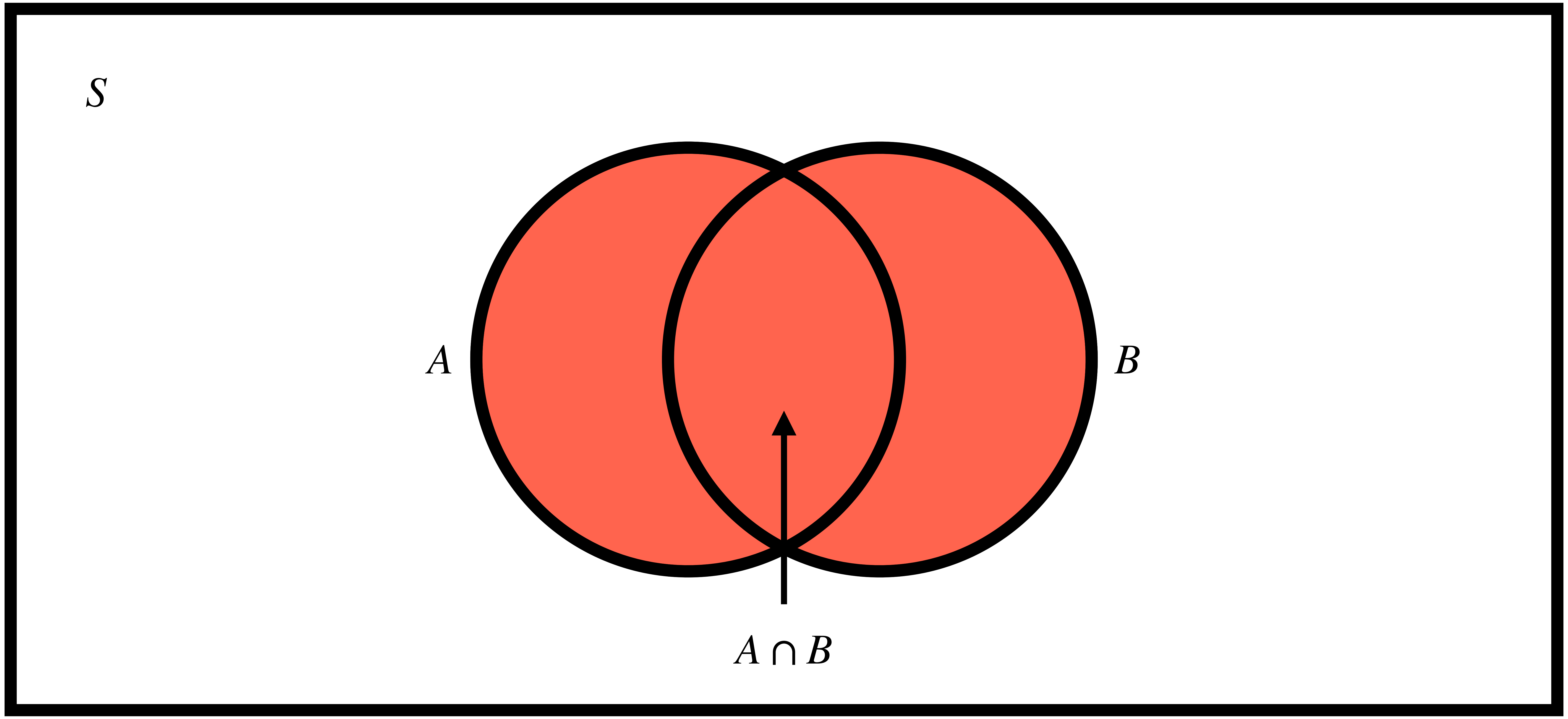
Addition Rule: Mutually Exclusive Events

- If A and B are mutually exclusive, we have $\Pr(A \cup B) = \Pr(A) + \Pr(B)$



Addition Rule: General

- In general, we have $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$



Probability Example

- Suppose that 55% of cancer patients are female, 20% of cancer patients have previously undergone chemotherapy, and 15% of cancer patients are both female and have undergone chemotherapy
- What is the probability that a patient is female or has undergone chemotherapy?

Conditional Probability

- Often, we are interested in determining the probability that an event will occur given that we already know the outcome of another event
  - Example: What is the probability that it rains tomorrow given that it rained today?
- **Conditional Probability**: The probability that event A will occur given that we already know the outcome of event B
- $\Pr(A \mid B) =$ probability of A given B

Multiplicative Rule

- The *multiplicative rule of probability* tells us the following:

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B \mid A)$$

$$\Pr(A \cap B) = \Pr(B) \cdot \Pr(A \mid B)$$

- Rearranging yields *conditional probability expressions*:

$$\Pr(B \mid A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Conditional Probability Example

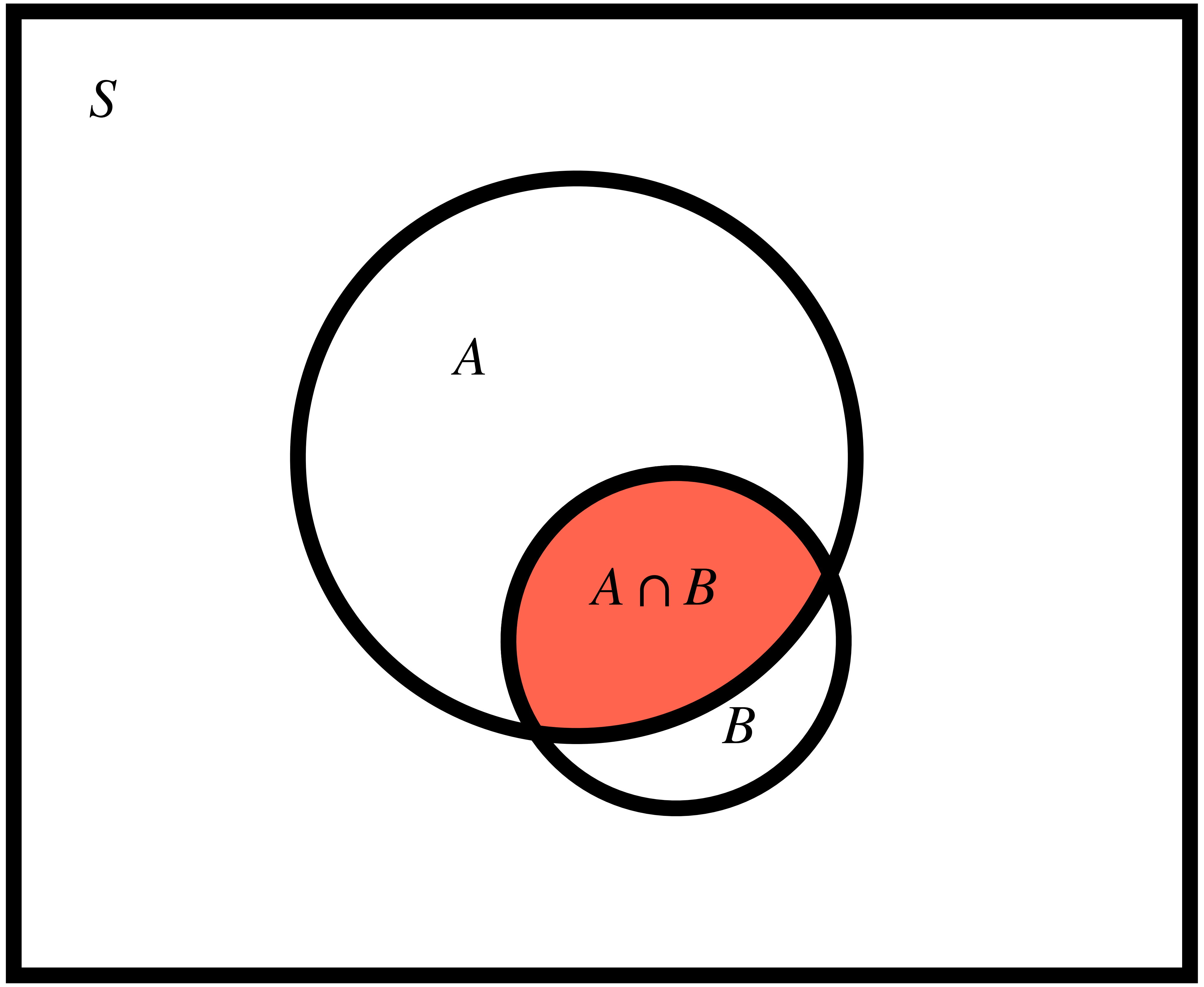
- Setup:
  - Suppose 10,000 students enter college
  - 450 students changed majors
  - 300 students who changed majors were males
  - 3000 students were males
- Q1: What is the probability of changing majors given that you are a male?
- Q2: What is the probability of changing majors given that you are not a male?

Multiplicative Rule Example

- Setup:
  - The probability that you will be sick tomorrow is 0.6
  - If you are sick tomorrow, the probability that you will be sick the next day is 0.7
  - If you are not sick tomorrow, the probability that you will be sick the next day is 0.2
- Q1: What is the probability that you are sick tomorrow and the next day?
- Q2: What is the probability that you are not sick tomorrow but sick the following day?

Conditional Probability

- Note, $\Pr(B \mid A) \neq 1 - \Pr(A \mid B)$
- Similarly, $\Pr(B \mid A) \neq 1 - \Pr(B \mid A^c)$
- But, $\Pr(B \mid A) = 1 - \Pr(B^c \mid A)$



Conditional Probability Example

- Setup:
  - Consider a random experiment where 3 balls are randomly selected (without replacement) from 5 balls labeled 1, 2, 3, 4, 5. Sample space:

    123, 124, 125, 134, 135, 145
    234, 235, 245
    345

- Let $A = \{1 \text{ is selected}\}$ and $B = \{5 \text{ is selected}\}$. What is $\Pr(A \mid B)$?

Independence

- **Independence**: The outcome of one event has no effect on the outcome of another event
  - If A and B are independent, then $\Pr(A \mid B) = \Pr(A)$ (and $\Pr(B \mid A) = \Pr(B)$)
- This is because intersection is decomposable:
  - If A and B are independent, then $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$
  - From this, we see that $\Pr(A \mid B) = \dfrac{\Pr(A \cap B)}{\Pr(B)} = \dfrac{\Pr(A) \cdot \Pr(B)}{\Pr(B)} = P(A)$

Independence Example

- Setup:
  - Suppose we flip a coin twice; tosses are independent
  - Let $A = \{\text{first flip is heads}\}$ and $B = \{\text{second flip is heads}\}$
  - $\Pr(A) = \Pr(B) = 1/2$
- What is $\Pr(A \cap B)$ (probability that both flips are heads)?

Mutual Independence

- Suppose we have n events, N. These n events are **mutually independent** iff, for every subset of events $M \subseteq N$, we have
$$\Pr\left(\bigcup_{i \in M} A_i\right) = \prod_{i \in M} \Pr(A_i)$$
- Consider the case of $n = 3$. Events A_1, A_2, A_3 are independent iff the following hold:
$$\begin{aligned} \Pr(A_1 \cap A_2) &= \Pr(A_1) \cdot \Pr(A_2) \\ \Pr(A_1 \cap A_3) &= \Pr(A_1) \cdot \Pr(A_3) \\ \Pr(A_2 \cap A_3) &= \Pr(A_2) \cdot \Pr(A_3) \\ \Pr(A_1 \cap A_2 \cap A_3) &= \Pr(A_1) \cdot \Pr(A_2) \cdot \Pr(A_3) \end{aligned}$$
- If all but the last equality hold, A_1, A_2, A_3 are *pairwise independent*, but not mutually independent

Pairwise Independence: Example

- Setup: Consider rolling a fair six-sided die. Consider the events $A = \{1,2\}$, $B = \{1,3\}$, and $C = \{2,3\}$
  - $\Pr(A) = \Pr(B) = \Pr(C) =$
  - $\Pr(A \cap B) =$
  - $\Pr(A \cap C) =$
  - $\Pr(B \cap C) =$
  - $\Pr(A \cap B \cap C) =$
- These events are pairwise independent but not mutually independent

Independence vs. Mutual Exclusivity

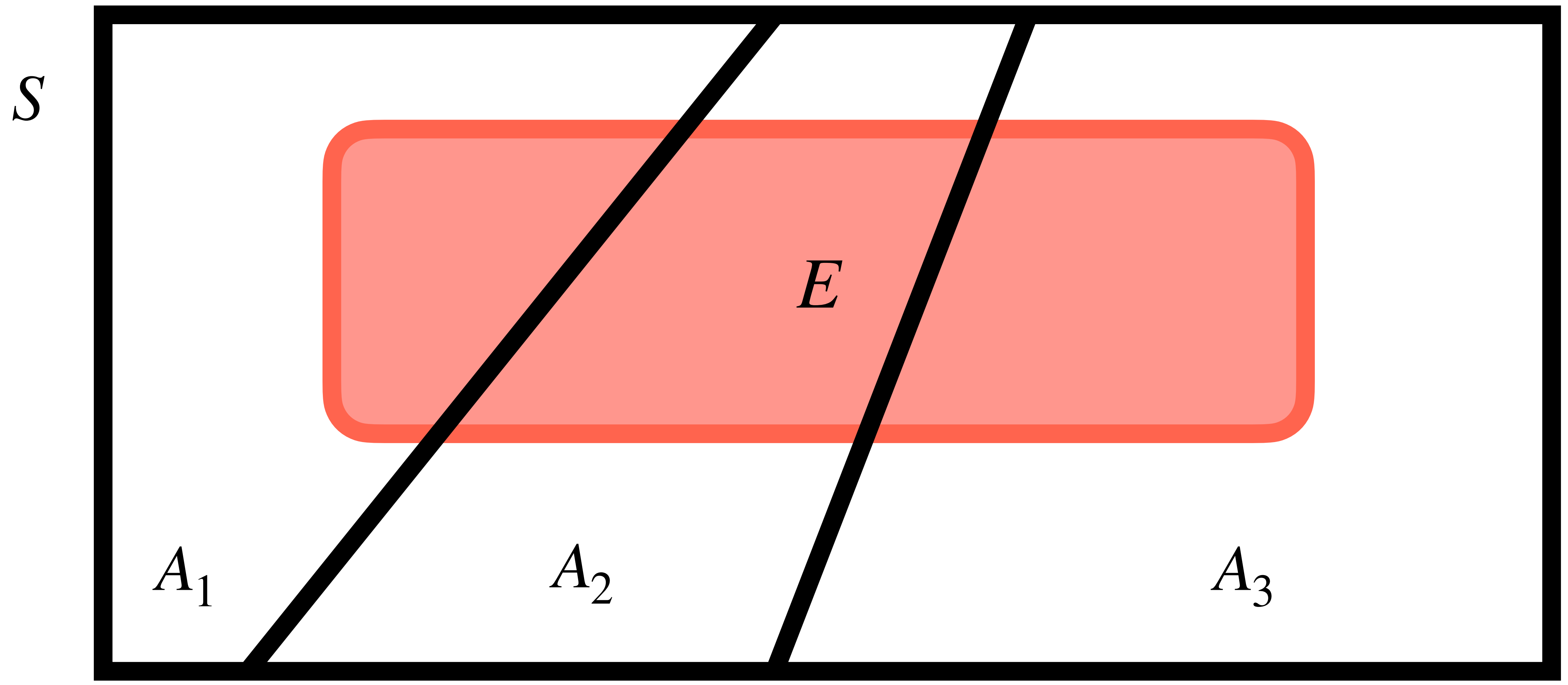
- Independence and mutual exclusivity are not the same thing
- If A and B are mutually exclusive, then $\Pr(A \mid B) = 0$ and $\Pr(B \mid A) = 0$
- This is not the same thing as independence, where $\Pr(A \mid B) = \Pr(A)$ and $\Pr(B \mid A) = \Pr(B)$
- Independence: the other event still may occur; its probability is unaffected

Law of Total Probability

- Consider a collection of mutually exclusive and exhaustive events $A_1, A_2, \ldots, A_n$ that *partitions* the sample space S

- Then, for any event E, the law of total probability states the following:

$$\begin{aligned}\Pr(E) &= \Pr(E \cap A_1) + \Pr(E \cap A_2) + \ldots + \Pr(E \cap A_n) \\ &= \Pr(E \mid A_1) \cdot \Pr(A_1) + \Pr(E \mid A_2) \cdot \Pr(A_2) + \ldots + \Pr(E \mid A_n) \cdot \Pr(A_n)\end{aligned}$$



Bayes' Theorem

- Let's say you have an idea of $\Pr(B \mid A)$ but want to know about $\Pr(A \mid B)$
- Recall that $\Pr(A \mid B) \cdot \Pr(B) = \Pr(B \mid A) \cdot \Pr(A) = \Pr(A \cap B)$
- Rearranging yields Bayes' Theorem:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)} = \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B \mid A) \cdot \Pr(A) + \Pr(B \mid A^c) \cdot \Pr(A^c)}$$

Posterior Likelihood Prior

Bayes' Theorem: Example

- Setup:
  - Given that you have diabetes, there is a 70% chance you are also overweight
  - Given that you do not have diabetes, there is a 35% chance you are overweight
  - 10% of people have diabetes
- Q: Given that a randomly selected person is overweight, what is the probability that he has diabetes?

Diagnostic Tests

- Apply Bayes' theorem to diagnostic testing and screening
- Assume there are two mutually exclusive and exhaustive states of health:
  - D_1: the event that a subject has the disease
  - D_2: the event that a subject does not have the disease
- Assume that we run a screening test on a patient to determine if they have the disease, with two mutually exclusive and exhaustive outcomes:
  - T^+: the test is positive
  - T^-: the test is negative
- Typically, we are interested in $\Pr(D_1 \mid T^+)$ (true positive rate of a test)

Diagnostic Tests

- **Sensitivity**: Probability of a positive test result given that the individual tested actually has the disease (true positive):
  - $\Pr(T^+ \mid D_1)$
- **False negative probability**: Probability of a negative test result given that the individual tested actually has the disease (false negative):
  - $\Pr(T^- \mid D_1) = 1 -$ Sensitivity
- **Specificity**: Probability of a negative test result given that the individual tested does not have the disease (true negative):
  - $\Pr(T^- \mid D_2)$
- **False positive probability**: Probability of a positive test result given that the individual tested does not have the disease (false positive):
  - $\Pr(T^+ \mid D_2) = 1 -$ Specificity

Positive Predictive Value (PPV)

- **Positive predictive value (PPV)**: The probability that a person with a positive test result actually has the disease
  - $\Pr(D_1 \mid T^+)$
- Using Bayes' Rule, sensitivity, and specificity:

$$\Pr(D_1 \mid T^+) = \frac{\Pr(D_1 \cap T^+)}{\Pr(T^+)}$$
$$= \frac{\Pr(T^+ \mid D_1) \cdot \Pr(D_1)}{\Pr(T^+ \mid D_1) \cdot \Pr(D_1) + \Pr(T^+ \mid D_2) \cdot \Pr(D_2)}$$

- What are $\Pr(D_1)$ and $\Pr(D_2)$?
  - $\Pr(D_1)$: probability of having the disease, or prevalence of the disease
  - $\Pr(D_2) = 1 - \Pr(D_1)$

Negative Predictive Value (NPV)

- **Negative predictive value (NPV)**: The probability that a person with a negative test result actually does not have the disease
  - $\Pr(D_2 \mid T^-)$
- Using Bayes' Rule, sensitivity, and specificity:

$$\Pr(D_2 \mid T^-) = \frac{\Pr(D_2 \cap T^-)}{\Pr(T^-)}$$

$$= \frac{\Pr(T^- \mid D_2) \cdot \Pr(D_2)}{\Pr(T^- \mid D_2) \cdot \Pr(D_2) + \Pr(T^- \mid D_1) \cdot \Pr(D_1)}$$

Diagnostic Tests: Example

- Cancer test has the following properties:
  - The test gives a positive result 95% of the time when the patient has cancer
  - The test gives a negative result 90% of the time when the patient does not have cancer
  - About 12% of patients have cancer
- Q: A patient tested positive for cancer. What is the probability that they have cancer?

Diagnostic Tests: Example

- Cancer test has the following properties:
  - The test gives a positive result 95% of the time when the patient has cancer
  - The test gives a negative result 90% of the time when the patient does not have cancer
  - About 12% of patients have cancer
- Q: A patient tested positive for cancer. What is the probability that they have cancer?

$$
\begin{aligned}
\Pr(C \mid pos) &= \frac{\Pr(C \cap pos)}{\Pr(pos)} \\
&= \frac{\Pr(pos \mid C) \cdot \Pr(C)}{\Pr(pos \mid C) \cdot \Pr(C) + \Pr(pos \mid C^c) \cdot \Pr(C^c)} \\
&= \frac{0.95 \cdot 0.12}{0.95 \cdot 0.12 + (1 - 0.90) \cdot (1 - 0.12)} \\
&= 0.5644
\end{aligned}
$$

Combinatorics

Counting Outcomes

- If each outcome in the sample space is equally likely, then computing probabilities is an exercise in counting
- For a sample space S and an event $E \subseteq S$, the probability of E (under an equiprobable model) is $\Pr(E) = \frac{N}{D}$
  - Where N is the total number of outcomes in E and D is the total number of outcomes in S
- We're going to learn how to count the number of outcomes

Ordered vs. Unordered Selection

- **Ordered selection** of size n from sample space S: select n distinct objects from S where order of selection matters
  - Care about the names and order of choices
- **Unordered selection** of size n from sample space S: select n distinct objects from S where order of selection does not matter
  - Care about the names of choices (think of it as a set)

Rule of Product

- Suppose a procedure can be broken down into m tasks
- There are n_i distinct ways to perform the i^{th} task, for $i = 1, \dots, m$
- Then, there are $n_1 \cdot n_2 \cdot \ldots \cdot n_m$ distinct ways to perform the entire procedure

Rule of Product: Example

- How many valid three-digit numbers (i.e., between 100 and 999, inclusive) have three different digits and only a single odd number in the middle?

Rule of Product: Example

- How many valid three-digit numbers (i.e., between 100 and 999, inclusive) have three different digits and only a single odd digit in the center?

Break this down into $m = 3$ tasks

Task 1: Select an odd (center) digit, $n_1 = 5$

Task 2: Select a first (even) digit that is not 0, $n_2 = 4$

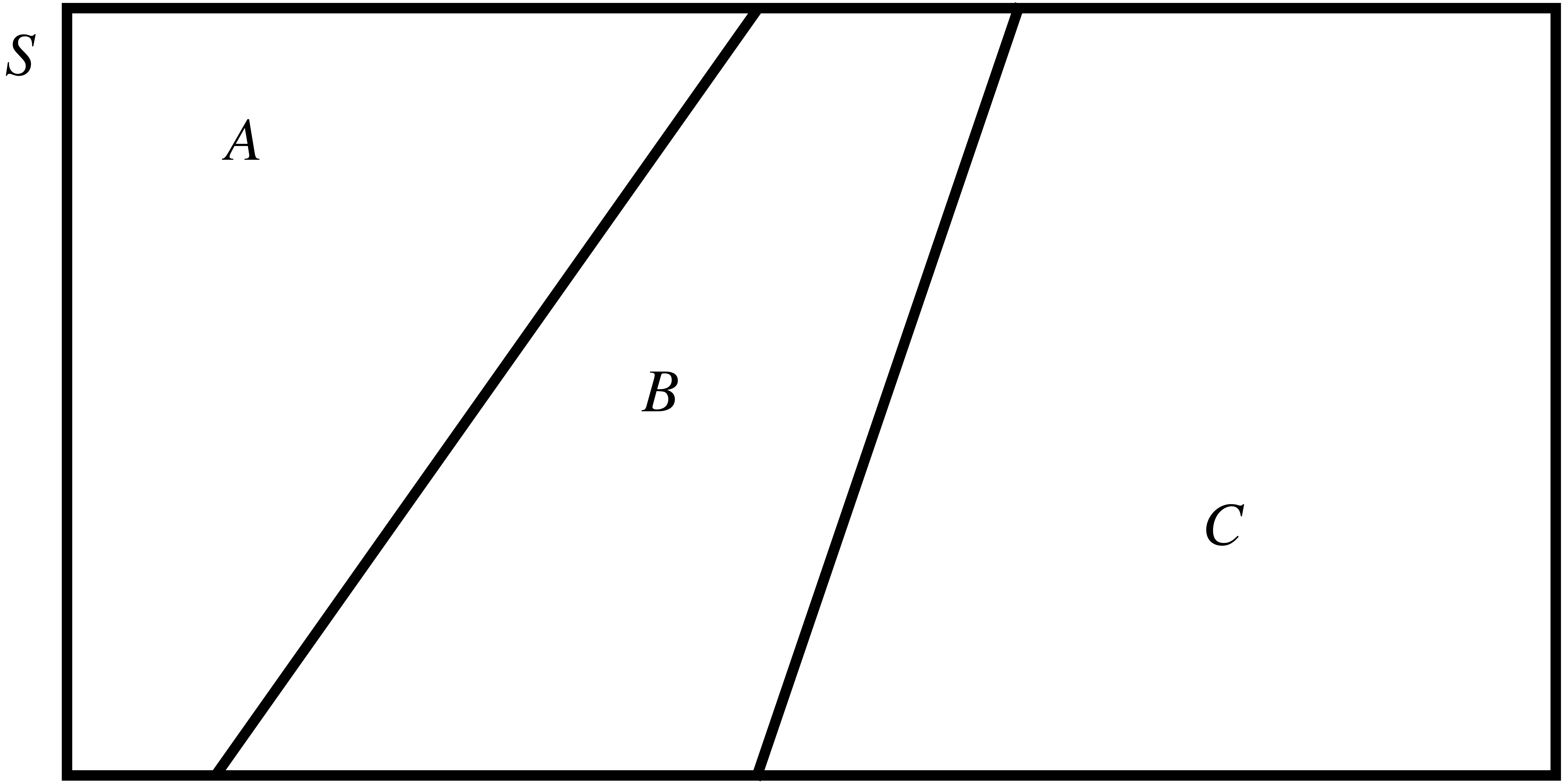
Task 3: Select a last (even) digit, $n_3 = 4$

Total: $n_1 \cdot n_2 \cdot n_3 = 5 \cdot 4 \cdot 4 = 80$

Tree Method (Rule of Sum)

- Suppose a procedure can be broken down into m disjoint and exhaustive cases
- There are n_i distinct ways to get the i^{th} case, for $i = 1,\ldots,m$
- Then, there are $n_1 + n_2 + \ldots + n_m$ distinct ways to perform the entire procedure
- Often, use the rule of sum (tree method) and the rule of product together

Rule of Sum (OR) and Rule of Product (AND)



Factorials

- Factorial: $n!$ is the product of all positive integers less than or equal to n
  - $n! = n \cdot (n-1) \cdot \ldots \cdot 1$
- Allows us to calculate the number of ways in which n objects can be ordered
- By convention, $0! = 1$ (there is one way of ordering zero things)
- In R: use `factorial(x)`

Permutation

- Suppose we want to select and order k objects from a total of n objects
  - Ordered selection
- There are n ways to select the first object, $n-1$ ways to select the second object, and so on until we have $n-k+1$ ways to select the final object

$$
\begin{aligned}
P(n,k) &= n \cdot (n-1) \cdot \ldots \cdot (n-k+1) \\
&= \frac{n!}{(n-k)!}
\end{aligned}
$$

Permutation: Example

- Q1: How many four-letter “words” are there where each letter is distinct?

- Q2: How many ways are there of assigning three students to seven orientation groups, where each student must go to a different group?

Combination

- Suppose we want to select k objects from n objects (unordered selection)
- There are $P(n, k)$ ways to select and order k out of n objects
- There are $k!$ ways to order k distinct objects
- Therefore, we have $C(n, k) = \dfrac{P(n, k)}{k!} = \dfrac{n!}{k!(n-k)!} = \dbinom{n}{k}$
- Interpretation: $C(n, k)$ is the number of ways in which k objects can be selected from a total of n objects (without replacement) without regard to order
- In R, use `choose(n, k)`
- *Binomial coefficient*

Combination: Example (Poker Hands)

- Setting: A poker hand consists of five cards dealt from a standard deck of 52 cards (4 suits of 13 values)

- Q1: How many different five-card hands are there?

choose(52, 5)

how to get the full house? (3 of one value and 2 of another value)

number

- Q2: What is the probability of getting four of the same kind?

(13 * choose (4,4) * 12) / choose (52, 5)

Combination: Example (Urn)

“story proof”

- Setting: An urn contains 35 yellow balls (numbered 1-35) and 35 pink balls (numbered 1-35). Four balls are chosen at random

- Q1: What is the probability that there are two pairs of balls which have the same number?

  choose(35, 2) * choose(2,2) * choose(2,2)/ choose(70, 4)

- Q2: What is the probability that there is exactly one pair of balls with matching numbers?

Combination: Example (Urn)

- Setting: An urn contains 35 yellow balls (numbered 1-35) and 35 pink balls (numbered 1-35). Four balls are chosen at random
- Q3: What is the probability that the balls are all the same color and consecutively numbered?

32 * 2/ choose(70, 4)

1, 2, 3, 4 … 32, 33, 34, 35 * 2 colors

Stars and Bars: Intuition

pick up two places of 5 between stars

- How many ways are there of choosing three *positive* numbers, x_1, x_2, x_3, such that $x_1 + x_2 + x_3 = 6$?
  - $\binom{6-1}{3-1} = \binom{5}{2}$: $\quad \star \;\Big|\; \star \quad \star \;\Big|\; \star \quad \star \quad \star$
- How many ways are there of choosing three *nonnegative* numbers, x_1, x_2, x_3, such that $x_1 + x_2 + x_3 = 6$?
  - $\binom{6+3-1}{3-1} = \binom{8}{2}$: $\quad \Big|\star \quad \star \;\Big|\; \star \quad \star \quad \star \quad \star$

Stars and Bars: More Formally

- Suppose there are n objects and k bins. Bins are distinguishable, but objects are not. The only thing we care about is the number of objects in each bin.
- If each bin has to have at least one object in it:
  - Total number of ways = $\binom{n-1}{k-1}$ (think of filling in gaps between objects)
- For nonnegative (not positive) constraints:
  - Total number of ways = $\binom{n+k-1}{k-1}$ (think of arranging n objects and $k-1$ dividers)

Stars and Bars: Example

- Setup: Six children are choosing ice cream flavors from {vanilla, strawberry, chocolate, caramel}. Each child picks exactly one flavor. Requests are placed in the form: {# vanilla, # strawberry, # chocolate, # caramel}.
- Q1: How many different requests are possible if at least one child must choose each flavor?
- Q2: How many different requests are possible without this restriction?