Lectures on Introduction to Discrete Structures

Contents

Lecture 1
Setting Things up: Set Theory

Today's lecture will be on set theory, a foundational topic in discrete mathematics. Set theory provides the framework for understanding collections of objects and their relationships, essential for various areas of mathematics and computer science. In this lecture, we will explore the basic concepts of sets, operations on sets, and their applications in problem-solving.

What is a Set?

A **set** refers to a collection of objects identified by particular properties. It is represented by $\mathcal{S} = \{\text{objects} \,|\, \text{properties}\}$. The **elements** of a set are the objects it contains. We use the notation $x \in \mathcal{S}$ to denote that x is an element of set $\mathcal{S}$. For instance, $2 \in \{1, 2, 3, 4, 5\}$.

Example. 1. The set of natural numbers: $\mathbb{N} = \{0, 1, 2, 3, 4, \ldots\}$.

2. The set of even natural numbers: $\mathbb{E} = \{x \in \mathbb{N} \,|\, x \text{ is an even integer}\}$.

3. The set of days in a week: $\mathcal{D} = \{\text{Sunday}, \text{Monday}, \text{Tuesday}, \text{Wednesday}, \text{Thursday}, \text{Friday}, \text{Saturday}\}$.

4. The set of days in a week when we have a class: $\mathcal{DC} = \{x \in \mathcal{D} \,|\, \text{we have a class}\} = \{\text{Monday}, \text{Wednesday}\}$.

Subsets and Supersets

A **subset** $A \subseteq B$ means that every element in A is also in B. If $A \subset B$, then A is a subset of B but not equal to B. Similarly, $A \supseteq B$ indicates that every element in B is also in A, and $A \supset B$ means A is a **superset** of B but not equal to B.

Example. Let's consider a few real-life examples:

Fruits and Apples. Set A can represent the set of all fruits in a grocery store. Set B can represent the set of all apples in that store. Since every apple is a fruit, we have $B \subseteq A$.

Math Courses and Calculus. Set A could be the set of all math courses offered in a university, and set B could be the set of all calculus courses. In this case, $B \subseteq A$ because all calculus courses are math courses, but not all math courses are necessarily calculus courses.

Animals and Dogs. Set A represents all animals on a farm, and set B represents all the dogs on that farm. Here, $B \subseteq A$ since every dog is an animal.

Operations on Sets

Sets can be combined and manipulated in various ways.

Definition. 1. The **union** of sets A and B, denoted $A \cup B$, is the set containing all elements that belong to A, or to B, or to both.

2. The **intersection** of sets A and B, denoted $A \cap B$, is the set containing all elements that belong to both A and B.

3. The **difference** of sets A and B, denoted $A \setminus B$, is the set containing all elements that belong to A but not to B.

4. The **symmetric difference** of sets A and B, denoted $A \oplus B$, is the set containing all elements that belong to either A or B, but not to both.

5. A set X is a **subset** of set Y, denoted $X \subseteq Y$, if every element of X is also an element of Y.

6. The **universal set**, denoted by $\mathcal{U}$, is the set that contains all elements under consideration for a particular discussion or problem.

7. The **complement** of set A with respect to the universal set $\mathcal{U}$, denoted A^C, contains all elements that are in $\mathcal{U}$ but not in A.

Definition. Two sets A and B are called **disjoint** if their intersection is the empty set, i.e., $A \cap B = \varnothing$.

Set Operations in Action

Let's try some examples!

Example. 1. Consider the sets $A = \{1, 2, 3, 4\}$ and $B = \{3, 4, 5, 6\}$ with the universal set $\mathcal{U} = \{1, 2, 3, \ldots, 10\}$.

Union: $A \cup B = \{1, 2, 3, 4, 5, 6\}$.

Intersection: $A \cap B = \{3, 4\}$.

Difference: $A \setminus B = \{1, 2\}$ and $B \setminus A = \{5, 6\}$.

Symmetric Difference: $A \oplus B = \{1, 2, 5, 6\}$.

Complement: $A^C = \{5, 6, 7, 8, 9, 10\}$.

2. Let's consider sets $C = \{2, 4, 6, 8\}$ and $D = \{3, 6, 9\}$ with the universal set $\mathcal{U} = \{1, 2, 3, \ldots, 10\}$.

Union: $C \cup D = \{2, 3, 4, 6, 8, 9\}$.

Intersection: $C \cap D = \{6\}$.

Difference: $C \setminus D = \{2, 4, 8\}$ and $D \setminus C = \{3, 9\}$.

Symmetric Difference: $C \oplus D = \{2, 3, 4, 8, 9\}$.

Complement: $C^C = \{1, 3, 5, 7, 9, 10\}$.

3. Lastly, consider sets $X = \{1, 2, 3, 4\}$ and $Y = \{3, 4, 5\}$ with the universal set $\mathcal{U} = \{1, 2, 3, 4, 5\}$

Union: $X \cup Y = \{1, 2, 3, 4, 5\}$.

Intersection: $X \cap Y = \{3, 4\}$.

Difference: $X \setminus Y = \{1, 2\}$ and $Y \setminus X = \{5\}$.

Symmetric Difference: $X \oplus Y = \{1, 2, 5\}$.

Complement: $X^C = \{5\}$.

Unions and Intersections of Many Sets

Unions and intersections can involve not just a couple, but many sets. The union of multiple sets consists of elements that belong to at least one set from the collection under consideration, while the intersection consists of elements that are common to all sets in the collection. Let's explore some intriguing examples that dive into the realm of countably many sets.

Example. Consider the sets $A_1, A_2, A_3, \ldots$ where $A_n = \{n, n+1, n+2, \ldots\}$ for each positive integer n.

1. Notice that $A_1 \supset A_2 \supset A_3 \ldots$, therefore, the union of all sets A_n is the set A_1 containing all positive integers. In other words, $\bigcup\limits_{n=1}^{\infty} A_n = \{1, 2, 3, \ldots\}$.
2. The intersection of all sets A_n is the empty set $\varnothing$. This is because there is no number that belongs to every set A_n since each set starts from a different positive integer.

Example. Consider the sets $C_n = \left(-\frac{1}{n}, \frac{1}{n}\right)$ for each positive integer n.

1. Again, $C_1 \supset C_2 \supset C_3 \ldots$, so the union of all sets C_n is the open interval $(-1, 1)$.

2. The intersection of all sets C_n is the singleton set $\{0\}$. This is because 0 is the only number that belongs to every set C_n.

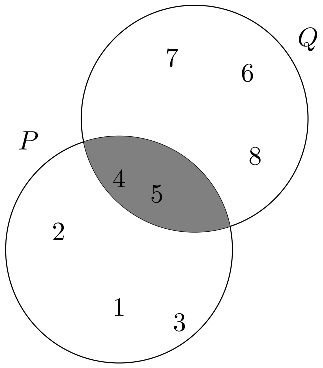
Example. Consider the sets $D_n = (-n, n)$ for each positive integer n.

1. This time we have the opposite sequence of inclusions of sets: $D_1 \subset D_2 \subset D_3 \subset \ldots$, so the intersection of all sets D_n is equal to D_1, the open interval $(-1, 1)$.
2. The union of all sets D_n is the real line $\mathbb{R} = (-\infty, \infty)$.

Venn Diagrams

Venn diagrams provide a visual representation of sets and their relationships. For example, if $A \subseteq B$, we can draw a Venn diagram showing A inside B.

Example. Let's use Venn diagrams to illustrate set relationships.
Consider sets $P = \{1, 2, 3, 4, 5\}$ and $Q = \{4, 5, 6, 7, 8\}$.



In this Venn diagram, P is represented by the left circle and Q by the right circle. The overlapping region contains the elements that are in both P and Q, which are $\{4, 5\}$.

Cardinality and Power Set

Definition. The **cardinality** of a set A, denoted by $|A|$, represents the number of elements it contains.

Definition. The power set of a set A, denoted $\mathcal{P}(A)$, is the set of all subsets of A, including the empty set, $\varnothing$, and A itself.

Example. Consider a set $B = \{a, b\}$.

Cardinality: the set B contains 2 elements, so the cardinality of B is $|B| = 2$.

Power Set: the power set of B is

$$\mathcal{P}(B) = \{\varnothing, \{a\}, \{b\}, \{a, b\}\}.$$

Now, let's explore the cardinality of the power set for a set with 3 elements.
Consider the set $C = \{x, y, z\}$.

Cardinality: the set C contains 3 elements, so the cardinality of C is $|C| = 3$.

Power Set: the power set of C is

$$\mathcal{P}(C) = \{\varnothing, \{x\}, \{y\}, \{z\}, \{x, y\}, \{x, z\}, \{y, z\}, \{x, y, z\}\}.$$

Cardinality of Power Set: the cardinality of the power set of C is $|\mathcal{P}(C)| = 2^{|C|}$, which equals $2^3 = 8$. This can be understood by considering that for each element in C, there are 2 options: include it in a subset or exclude it. Since there are 3 elements in C, there are $2 \cdot 2 \cdot 2 = 2^3 = 8$ possible subsets in the power set.

De Morgan's Laws

De Morgan's Laws describe the relationship between set operations, specifically unions and intersections, along with complements.

Definition. For any sets A and B, the De Morgan's Laws state:

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

Example. Let us understand De Morgan's law with the help of a simple example. Let the universal set $U = \{7, 8, 9, 10, 11, 12, 13\}$. The two subsets are given by $A = \{11, 12, 13\}$ and $B = \{7, 8\}$.

De Morgan's Law of Union

$$A \cup B = \{7, 8, 11, 12, 13\},$$
$$A^C = \{7, 8, 9, 10\},$$
$$B^C = \{9, 10, 11, 12, 13\},$$
$$A^C \cap B^C = \{9, 10\}.$$

Thus, $(A \cup B)^C = A^C \cap B^C$.

De Morgan's Law of Intersection

$$A \cap B = \varnothing,$$
$$(A \cap B)^C = U = \{7, 8, 9, 10, 11, 12, 13\}.$$
$$A^C \cup B^C = \{7, 8, 9, 10, 11, 12, 13\}.$$

Hence, $(A \cap B)^C = A^C \cup B^C$.

Paradoxes (Optional Fun Facts)

Liar's Paradox

In the novel 'Don Quixote' written by Miguel de Cervantes, there's an intriguing scenario that includes a paradox from set theory known as the 'liar's paradox'. The book is highly recommended for its captivating tales and thought-provoking moments.

On the second day of his governing, Sancho eats a meager breakfast and goes into the courtroom. A man comes in and begins to describe a dilemma. A river cuts a lord's estate in two parts, and a bridge crosses over it. The owner of the river decreed that every person that wants to cross the bridge must state his purpose to several judges; if he tells the truth, he can cross, but if he lies, he must be hung. One man told the judges that his purpose is to be hung. If the judges allow him to cross, then his statement will have been a lie, and he should have been hung; if they hang him, then he told the truth, and he should have been allowed to cross. Sancho responds that this man deserves to live as much as he deserves to die, so it's better to be merciful and let him live. Everyone is satisfied with the decision.

The scenario presents a classic example of a self-referential paradox, often referred to as the "liar's paradox." The contradiction arises from the man's statement itself:

The man states, 'My purpose is to be hung'.

If the judges allow him to cross, then his statement is false, because he said he intended to be hung but was not. This would mean he should have been hung according to the rules.

On the other hand, if the judges hang him, then his statement is true, because he said he intended to be hung and he was. However, this also leads to a contradiction, because if he was telling the truth, then he should have been allowed to cross according to the rules.

So, no matter what the judges decide, they run into a logical contradiction. This creates a perplexing situation where it seems impossible to make a decision based on the rules established.

Sancho's response, that the man deserves to live as much as he deserves to die, is a recognition of this inherent contradiction. By showing mercy and allowing the man to live, Sancho essentially acknowledges that the situation is beyond a straightforward application of the rules.

This scenario illustrates the complexities and philosophical conundrums that can arise from self-referential statements and logical paradoxes, making it a thought-provoking moment in the novel.

Anticipated Movie Premiere Paradox

Imagine there's a highly anticipated movie premiere scheduled for next Monday. Fans have been eagerly awaiting this film for months, and the excitement is palpable. The organizers of the premiere have decided to add an element of surprise. Eight days before the premiere, on Sunday, they announced that tickets would go on sale on one of the weekdays at noon. However, it will impossible to determine which day in advance (the previous day).

Question. Given the information provided, which day can the ticket sale start?

The tickets sale cannot start on Sunday. If it did, then it would be known after noon on Saturday (if the sale did not start) that it has to start on Sunday, which contradicts the assumption that it is impossible to determine which day the sale starts in advance. Hence, Sunday (being the last day of the week) can be eliminated as a possibility. Now, Saturday becomes the last day, and the same logic applies to eliminate it as a possibility as well. Continuing this way, we can eliminate Friday, Thursday, Wednesday, Tuesday, and Monday, sequentially. This situation presents a paradox, as the assumptions are impossible to combine: the sale cannot start on any day without violating one of the conditions. Therefore, the conclusion is that the tickets will not go on sale as originally planned.

Lecture 2
More about Sets

In the previous lecture, we discussed that the cardinality of a finite set is the number of elements in it. However, there exist sets with an infinite number of elements, such as the natural numbers, integers, and points on the interval $[0, 1]$. A set is said to be *countable* if its elements can be enumerated, i.e. put into a one-to-one correspondence with the natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$.

Example. Some examples of countable sets include:

- The set of natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$.
- The set of integers $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$. We can assign numbers to integers by starting from 0 and alternating between positive and negative numbers. For example, we can assign 0 to 1, 1 to 2, -1 to 3, 2 to 4, -2 to 5, and so on.
- The set of pairs of integers $\mathbb{Z}^2 = (m, n) \mid m, n \in \mathbb{Z}$ forms a two-dimensional integer grid. One way to number the elements of this set is to draw a spiral path and assign numbers according to the order of appearance on the path of a ladybug 🐞 that starts its journey at $(0, 0)$ and traverses along the path:

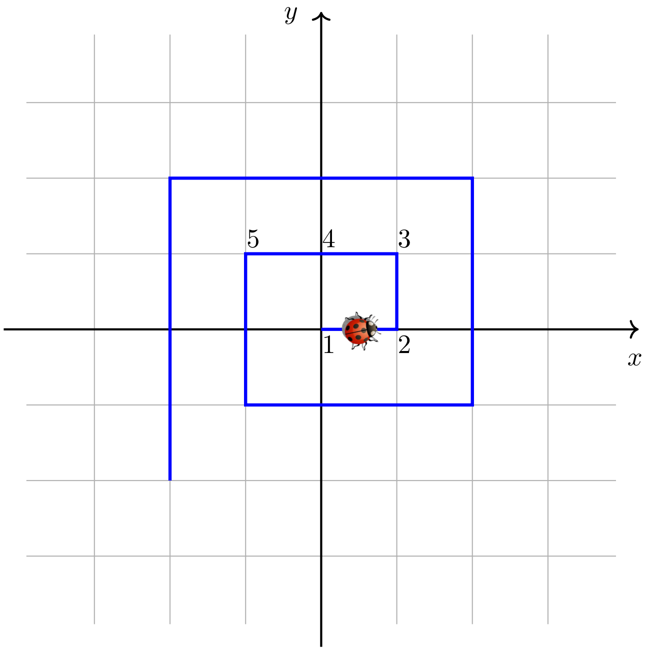
A set is said to be *uncountable* if its elements cannot be put into a one-to-one correspondence with the natural numbers. For example, the interval $[0, 1]$ is an uncountable set.

Disjoint Sets and Partitioning

Disjoint sets are sets whose intersection is the empty set. By partitioning a finite set, we understand an expression of it as a disjoint union of nonempty subsets.

Example. Consider the set $X = \{1, 2, 3, 4\}$. We can partition X in the following ways.

- Partition into 2 sets:



- $\{1,2,3\} \sqcup \{4\}$, $\{1,2,4\} \sqcup \{3\}$, $\{1,3,4\} \sqcup \{2\}$ and $\{2,3,4\} \sqcup \{1\}$
- $\{1,2\} \sqcup \{3,4\}$, $\{1,3\} \sqcup \{2,4\}$ and $\{1,4\} \sqcup \{2,3\}$

- Partition into 3 sets:
  - $\{1,2\} \sqcup \{3\} \sqcup \{4\}$, $\{1,3\} \sqcup \{2\} \sqcup \{4\}$, $\{1,4\} \sqcup \{2\} \sqcup \{3\}$, $\{2,3\} \sqcup \{1\} \sqcup \{4\}$, $\{2,4\} \sqcup \{1\} \sqcup \{3\}$ and $\{3,4\} \sqcup \{1\} \sqcup \{2\}$.
- Unique partitioning into 4 one-element sets:
  - $\{1\} \sqcup \{2\} \sqcup \{3\} \sqcup \{4\}$.

There are a total of 14 possible ways to partition the set $X = \{1,2,3,4\}$.

Lecture 3

Exclusive Offer: All-Inclusive Mathematical Induction

Today we will explore two fundamental techniques in discrete mathematics: the inclusion-exclusion principle, which provides a powerful method for counting, and mathematical induction, a versatile proof technique used to establish statements for all positive integers.

Inclusion-Exclusion Principle for Two Sets

Imagine a scenario in which a company has two open positions. Specifically, 30 individuals have applied for the first position, while 25 have applied for the second. Interestingly, 10 candidates have shown interest in both positions. The director of the company would like to determine the overall number of applicants.

To address this, let's define A as the set of applicants for the first position and B as the set for the second position. At first glance, it might seem logical to simply add the number of applicants for A and B together. However, this approach would lead to an inaccurate count, as those 10 applicants who applied for both positions would be counted twice. Therefore, to arrive at the correct answer, we need to account for this overcounting. This can be achieved by subtracting the number of applicants who applied for both positions from the total count of applicants for both positions. This gives us the accurate total number of unique applicants, which equals 45.

In mathematical terms:

$$|A \cup B| = |A| + |B| - |A \cap B| = 30 + 25 - 10 = 45.$$

Now that we've successfully determined the total number of unique applicants in this scenario, let's take a moment to understand the broader principle behind this.

This process of correcting for overcounting by subtracting the number of shared elements is a fundamental idea in mathematics. It's known as the Inclusion-Exclusion Principle. This principle provides a powerful tool for dealing with overlapping sets.

In the case of two sets, like our applicants for positions A and B, the Inclusion-Exclusion Principle can be expressed as:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

This formula allows us to find the total number of unique elements in two sets, even when there is overlap.

The Inclusion-Exclusion Principle proves to be invaluable in various areas of mathematics, from combinatorics to probability theory and beyond.

Example. Consider a group of friends who have the following common interests in birthday parties:

- A: Friends who enjoy outdoor activities.
- B: Friends who enjoy indoor activities.

Out of the group of 30 friends, we assume that every friend enjoys at least one activity.

- 15 friends enjoy outdoor activities ($|A| = 15$).
- 20 friends enjoy indoor activities ($|B| = 20$).
- 10 friends enjoy both indoor and outdoor activities ($|A \cap B| = 10$).

We want to calculate the number of friend who enjoy either outdoor or indoor activities.

Using the inclusion-exclusion principle:

$$|A \cup B| = |A| + |B| - |A \cap B| = 15 + 20 - 10 = 25.$$

So, 25 friends enjoy either outdoor or indoor activities.

Proof Idea for Two Sets

To prove the inclusion-exclusion principle for two sets, we start by considering the sizes of A, B, and their intersection $A \cap B$.

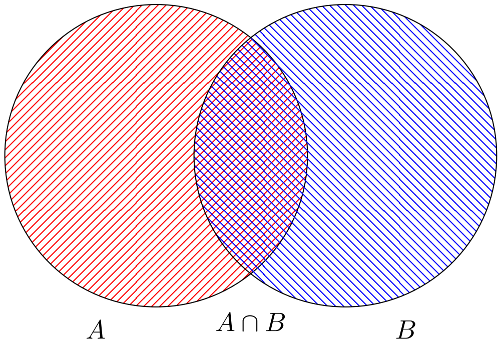
$$\begin{aligned} |A| &= \text{number of elements in } A \\ |B| &= \text{number of elements in } B \\ |A \cap B| &= \text{number of elements in both } A \text{ and } B \end{aligned}$$

When we sum the sizes of A and B, we count the elements in $A \cap B$ twice, so we need to subtract $|A \cap B|$ to correct for this double-counting.

$$|A| + |B| - |A \cap B|$$

This accounts for all elements in either A or B without double-counting.

In the Venn diagram below, we illustrate sets A and B, along with their intersection $A \cap B$. Notice that $A \cap B$ is shaded with a line segment pattern in both A and B to emphasize that it is counted twice:



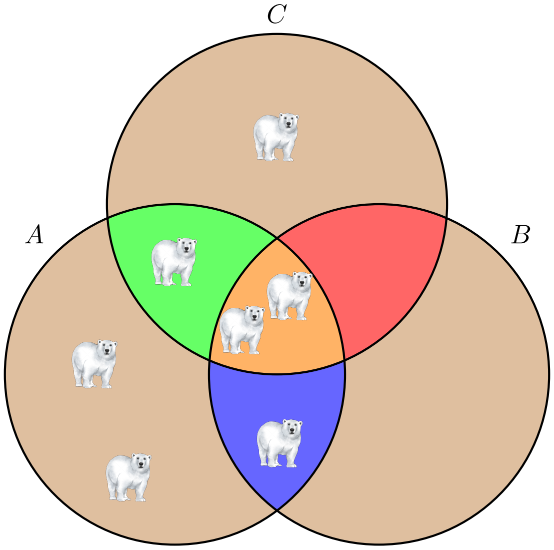
Generalization to Three or More Sets

Note: this section contains additional material for those interested in further exploring the topic.

The inclusion-exclusion principle can be extended to more than two sets. For three sets A, B, and C, it is given by the formula

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

Example. Imagine we are studying the habitats of polar bears. In this context, A, B and C represent distinct environments where these bears reside. For instance, A may be a region with 6 polar bears, B another area with 3 bears, and C a third area with 4 bears. Now, it's not uncommon for a polar bear to traverse different habitats. We could humorously attribute this to the bears' keen instinct for locating the richest salmon feeding grounds. When faced with a choice between two thriving areas, our bear may gravitate towards the one with the most abundant salmon population. Suppose we are interested in computing the total number of bears, taking into account that there are 3 bears migrating between habitats A and B, 3 bears between A and C, 2 bears alternating between B and C, and finally, 2 intrepid bears constantly moving around all three habitats:



Using the inclusion-exclusion principle for three sets we get

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C| = 6 + 3 + 4 - 3 - 3 - 2 + 2 = 7.$$

This means that there are 7 polar bears in total across all three habitats. This example demonstrates how the inclusion-exclusion principle helps us count elements in the union of multiple sets, taking into account their intersections.

When applying the inclusion-exclusion principle, think of it as an 'iteration' process. At each successive intersection, we either add, subtract, or ignore an element, ensuring that it is taken into account exactly once by the formula overall. This systematic approach allows us to handle complex scenarios involving multiple sets.

In general, for n sets $A_1, A_2, \ldots, A_n$, the inclusion-exclusion principle can be expressed as:

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| - \sum_{i<j} |A_i \cap A_j| \quad + \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \ldots + (-1)^{n-1} |A_1 \cap A_2 \cap \ldots \cap A_n|$$

This formula expresses the total number of elements in the union of n sets, in terms of *cardinalities* (numbers of elements) of all possible intersections.

Introduction to Mathematical Induction

Mathematical induction is a powerful technique used to prove statements for all natural numbers. It involves two steps: establishing a base case and proving that if the statement holds for $n = k$, then it also holds for $n = k + 1$.

Let's illustrate mathematical induction with two examples.

Arithmetic Progression

Consider the statement:

$$1 + 2 + \ldots + n = \frac{n(n+1)}{2}$$

We will prove this statement using mathematical induction.

Base Case ($n = 1$):

$$1 = \frac{1(1+1)}{2} = 1\checkmark$$

Induction Step ($n = k \Rightarrow n = k + 1$): Assume the statement holds for $n = k$, i.e.,

$$1 + 2 + \ldots + k = \frac{k(k+1)}{2}$$

Now, we need to prove it holds for $n = k + 1$:

$$\begin{aligned} 1 + 2 + \ldots + k + (k+1) &= \frac{k(k+1)}{2} + (k+1) \quad \text{(by induction assumption)} \\ &= \frac{k(k+1) + 2(k+1)}{2} = \frac{(k+1)(k+2)}{2}. \end{aligned}$$

Thus, the statement holds for $n = k + 1$ as well.

Sum of Odd Integers

Consider the statement:

$$1 + 3 + 5 + \ldots + (2n-1) = n^2$$

Again, we will use mathematical induction to prove this.

Base Case ($n = 1$):

$$1 = 1^2 = 1\checkmark$$

Induction Step ($n = k \Rightarrow n = k + 1$): Assume the statement holds for $n = k$, i.e.,

$$1 + 3 + 5 + \ldots + (2k-1) = k^2$$

Now, we need to prove it holds for $n = k + 1$:

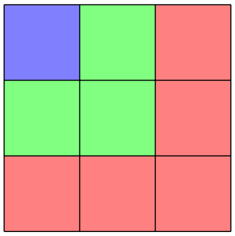
$$\begin{aligned} 1 + 3 + 5 + \ldots + (2k-1) + (2k+1) &= k^2 + 2k + 1 \quad \text{(by induction assumption)} \\ &= (k+1)^2 \end{aligned}$$

Thus, the statement holds for $n = k + 1$ as well.

Gauss's Famous Proof

An extraordinary anecdote in the history of mathematics involves the young Carl Friedrich Gauss, who, as the legend goes, astounded his teacher by deriving a formula for the sum of consecutive integers. The method elegantly illustrates the power of geometric visualization in mathematical reasoning.

Consider a sequence of consecutive odd numbers: $1, 3, 5, \ldots, (2n-1)$. Gauss envisioned these numbers arranged as layers of squares, with each layer representing the terms in the sequence. Consider a $1 \cdot 1$ square, representing 1^2. Gauss realizes that to form the next square, he needs to add 3 more unit squares to the right. Each subsequent square requires adding an additional $2n-1$ unit squares, where n represents the length of the side of the square:



Thus, as Gauss observes, the total area of the $n \cdot n$ square, which is equal to n^2, can be expressed as the sum $1 + 3 + 5 + \ldots + (2n-1)$.

Thus far we focused primarily on proving equalities for all positive integers n. However, mathematical induction is a powerful tool that can also be used to prove inequalities.

Example. Prove by mathematical induction that $n! > 2^n$ for all integers $n \geq 4$.

Base Case: first, we verify the inequality for the base case $n = 4$. We have:

$$4! = 24 > 2^4 = 16 \qquad \checkmark$$

It's worth noting that for $n = 3$, we have $3! = 6 < 8 = 2^3$, and $n = 4$ is the smallest value for which the inequality holds.

Inductive Hypothesis: assume that the inequality $k! > 2^k$ holds true for some positive integer $k \geq 4$.

Inductive Step: we want to show that $(k+1)! > 2^{k+1}$.

Using the inductive hypothesis, we get a chain of inequalities

$$(k+1)! = (k+1)k! > (k+1)2^k > 2^{k+1}.$$

Therefore, we conclude that the inequality $n! > 2^n$ holds true for all integers $n \geq 4$.

This example demonstrates how mathematical induction can be used to prove inequalities. By establishing the base case, assuming the inductive hypothesis, and then proving the inductive step, we can confidently conclude that the inequality holds true for all relevant values of n.

Lecture 4
Products of Sets and Relations

The (Cartesian) product of two sets $A \cdot B$ is defined as the set of all ordered pairs (a, b) where $a \in A$ and $b \in B$. Similarly, the product of finitely many sets is the set of all tuples. If all sets involved are finite, then the cardinality of the product set is the product of the cardinalities of the individual sets.

Example.

1. Let $A = B = \mathbb{R}$. The product $\mathbb{R} \cdot \mathbb{R}$ represents the Cartesian product of the set of real numbers with itself. Geometrically, this represents the Cartesian plane, where each point has an x-coordinate and a y-coordinate, both of which are real numbers.

2. The product $\mathbb{R} \cdot \mathbb{Z}$ is the Cartesian product of the set of real numbers with the set of integers. This results in pairs where the first element is a real number and the second element is an integer.

3. Let $A = \{1, 2\}$ and $B = \{a, b, c\}$. Then, $A \cdot B = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c)\}$.

Remark. If the sets A and B are finite, then the number of elements in the Cartesian product $A \cdot B$ is equal to the product of the number of elements in A and the number of elements in B. In other words, if $|A|$ denotes the number of elements in A and $|B|$ denotes the number of elements in B, then $|A \cdot B| = |A| \cdot |B|$.

One can extend the concept of the Cartesian product to involve more than two sets. For instance, the Cartesian product of three sets A, B, and C, denoted as $A \cdot B \cdot C$, consists of all possible ordered triples where the first element is from set A, the second element is from set B, and the third element is from set C.

Relations

Definition. Let X and Y be two sets. A **relation** from X to Y is a subset $\mathcal{R} \subseteq X \cdot Y$. The **domain** of $\mathcal{R}$ comprises all the first elements of the ordered pairs that belong to $\mathcal{R}$, while the **range** consists of the second elements.

Example.

1. Consider the sets $A = \{1, 2, 3\}$ and $B = \{x, y, z\}$. Let $\mathcal{R} = \{(1, y), (1, z), (3, y)\}$. We observe that $\mathcal{R}$ is a relation from A to B because $\mathcal{R}$ is a subset of $A \cdot B$. The domain of $\mathcal{R}$ is $\{1, 3\}$ and the range is $\{y, z\}$.
2. Let $A = B = \mathbb{Z}$, the set of integers, and consider the relation $\mathcal{R} = \{(x, y) \in \mathbb{Z}^2 \mid y = |x| + 3\}$. Here, the domain of $\mathcal{R}$ is the entire set of integers, $\mathbb{Z}$, while the range consists of integers greater than or equal to three, denoted as $\mathbb{Z}_{\geq 3}$.

More generally, a function $f : X \to Y$ gives rise to a relation defined by its graph, the subset $\{(x, f(x)) \mid x \in X\}$. However, a relation may have multiple values associated with the same x, making it a more general notion than a function.

Example. Consider the unit circle, a subset of points on the plane given by $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$. This is a relation but not a function, as it assigns multiple y values for some x values.

Definition. The **inverse of a relation** $\mathcal{R}$ from a set A to a set B, denoted by $\mathcal{R}^{-1}$, is a relation from B to A. It comprises those ordered pairs which, when reversed, belong to $\mathcal{R}$; formally,

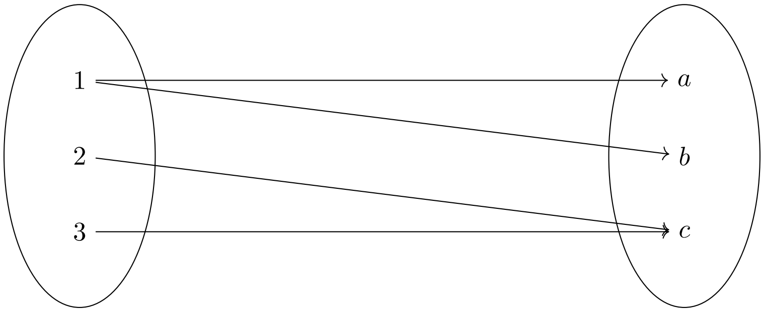
$$\mathcal{R}^{-1} = \{(b, a) \mid (a, b) \in \mathcal{R}\}.$$

Ways of Representing Relations

When the relation $\mathcal{R} \subseteq A \cdot B$ is a finite set of small cardinality, it can be conveniently represented using a diagram of dots and arrows. Such diagrams represent directed graphs, where the dots represent elements of set A (the domain) and set B (the range), and the arrows indicate the direction of the relation $\mathcal{R}$ between them.

A directed graph, is a graph in which edges (arrows) have a direction associated with them. In the context of representing relations, this directionality allows us to distinguish between the domain and range of the relation.

Example. Consider a relation $\mathcal{R}$ where $A = \{1, 2, 3\}$ and $B = \{a, b, c\}$. Let $\mathcal{R} = \{(1, a), (1, b), (2, c), (3, c)\}$. The diagram representing $\mathcal{R}$ is illustrated below:



This diagram illustrates the mapping of elements from set A to set B in the relation $\mathcal{R}$. Arrows from elements in set A point to the corresponding elements in set B, indicating the direction of the relation.

The adjacency matrix is a square matrix used to represent a directedgraph. In this case, it represents the directed graph formed by the relation $\mathcal{R}$. The rows and columns of the matrix correspond to the elements of set A and B, respectively. A non-zero entry in the matrix indicates the presence of an arrow from the corresponding element in set A to the corresponding element in set B.

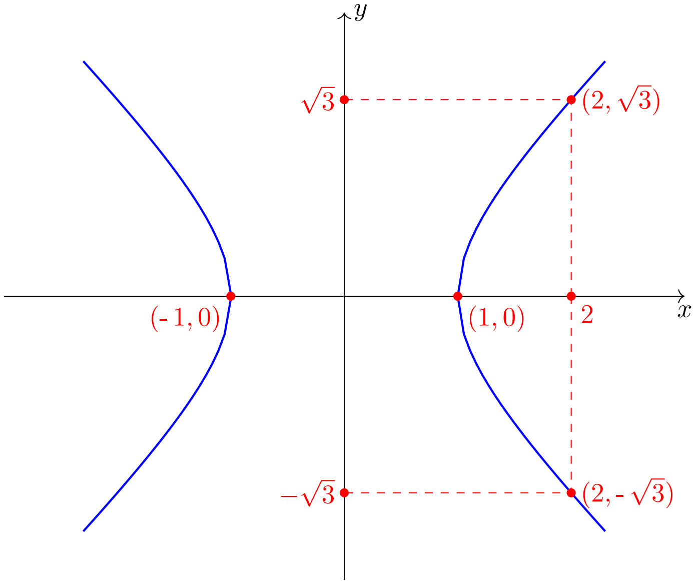
For the given example, the adjacency matrix is

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

In this matrix, a non-zero entry in row i and column j means that there is an arrow from element i in set A to element j in set B. For instance, the entry highlighted in red corresponds to the element $(1, b) \in \mathcal{R}$ (and the edge $1 \to b$).

When the relation $\mathcal{R}$ is infinite or has a large cardinality, and is defined by a formula, it may be convenient to represent it with a graph. Here, "graph" refers to a subset of the plane, rather than a collection of nodes and edges.

Example. When we consider the relation $\mathcal{R} = \{(x, y) \in \mathbb{R}^2 \mid x^2 - y^2 = 1\}$, and plot it in the Cartesian plane, we obtain a hyperbola.



To find the corresponding elements in the range for an element a in the domain of $\mathcal{R}$, we can look at the y-coordinates of the points of intersection where a vertical line $x = a$ intersects the graph of $\mathcal{R}$. Conversely, for an element b in the range of $\mathcal{R}$, we can determine the corresponding elements in the domain by examining the x-coordinates of the points of intersection where a horizontal line $y = b$ intersects the graph of $\mathcal{R}$.

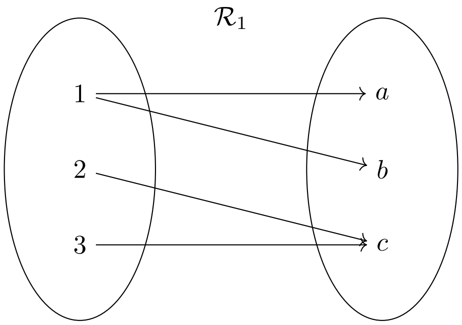
Lecture 5

Types of Relations, Composition, Closure and Equivalence

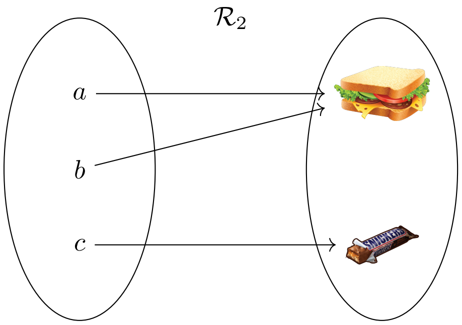
Composition of Relations

Similar to the composition of functions, more general relations can be composed. This allows for the chaining of relations to form new relations.

Example. Suppose we would like to compose two relations

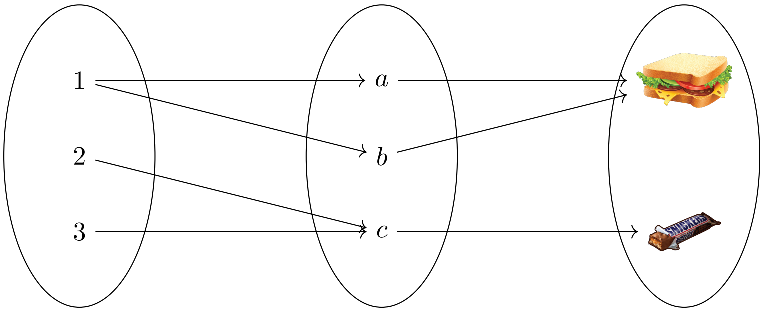


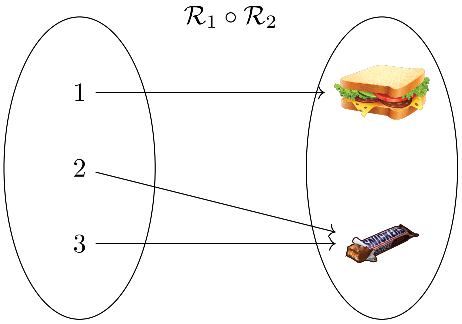
and



,

where $\mathcal{R}_1$ is a relation from the set $\{1, 2, 3\}$ to the set $\{a, b, c\}$ and $\mathcal{R}_2$ is a relation from the set $\{a, b, c\}$ to the set $\{$🥪, 🍫$\}$. The composition $\mathcal{R}_1 \circ \mathcal{R}_2$ is obtained by merging the two diagrams along the common set $\{a, b, c\}$:





The matrix representing the composition of two relations is obtained by multiplying the corresponding matrices representing the respective relations. Subsequently, each nonzero entry in the resulting matrix is replaced by 1:

$$M_{\mathcal{R}_1 \circ \mathcal{R}_2} = M_{\mathcal{R}_1} M_{\mathcal{R}_2} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Remark. Since the target of the first relation is the source of the second, the column dimension of $M_{\mathcal{R}_1}$ coincides with the row dimension of $M_{\mathcal{R}_2}$. Therefore, the product $M_{\mathcal{R}_1} M_{\mathcal{R}_2}$ is well-defined.

Types of Relations

Relations between sets are classified into different types based on certain properties they exhibit. Let's explore three fundamental types of relations: reflexive, symmetric, and transitive.

Reflexive Relations

A relation $\mathcal{R}$ on a set A is said to be reflexive if every element in A is related to itself.

Definition. A relation $\mathcal{R}$ is **reflexive** if $(a, a) \in \mathcal{R}$ for every $a \in A$.

Example. 1. The relation "is equal to" on the set of real numbers is reflexive, as $a = a$ for every real number a.

2. The relation "is a subset of" on the power set of a set A is reflexive, as every set is a subset of itself.

Non-example. 1. The relation "is less than" on the set of natural numbers is not reflexive, as $a < a$ is never true for any natural number a.

2. The relation "is perpendicular to" on the set of all lines in a plane is not reflexive, as no line is perpendicular to itself.

Symmetric Relations

A relation $\mathcal{R}$ on a set A is said to be symmetric if whenever (a, b) is in $\mathcal{R}$, then (b, a) is also in $\mathcal{R}$.

Definition. A relation $\mathcal{R}$ is symmetric if $(a, b) \in \mathcal{R}$ implies $(b, a) \in \mathcal{R}$ for all $a, b \in A$.

Example. 1. The relation "is congruent to" on the set of all triangles is symmetric, as if triangle ABC is congruent to triangle DEF, then triangle DEF is also congruent to triangle ABC.

2. The relation "is perpendicular to" on the set of all lines in a plane is symmetric, as if line AB is perpendicular to line CD, then the line CD is perpendicular to line AB.

Non-example. 1. The relation "is a parent of" on the set of all people is not symmetric, as if A is a parent of B, it is not true that B is a parent of A.

2. The relation "is divisible by" on the set of positive integers is not symmetric. If a is divisible by b, it does not imply that b is divisible by a.

Transitive Relations

A relation $\mathcal{R}$ on a set A is said to be transitive if whenever (a, b) and (b, c) are in $\mathcal{R}$, then (a, c) is also in $\mathcal{R}$.

Definition. A relation $\mathcal{R}$ is called **transitive** if $(a, b) \in \mathcal{R}$ and $(b, c) \in \mathcal{R}$ implies $(a, c) \in \mathcal{R}$ for all $a, b, c \in A$.

Example. 1. The relation "is less than" on the set of real numbers is transitive, as if $a < b$ and $b < c$, then $a < c$.

2. The relation "is taller than" on the set of all people is transitive. If person A is taller than person B and person B is taller than person C, then person A is taller than person C.

Non-example. 1. The relation "is adjacent to" on the set of all vertices in a graph is not transitive. If vertex A is adjacent to vertex B and vertex B is adjacent to vertex C, it does not necessarily mean that vertex A is adjacent to vertex C.

2. The relation "is a friend of" on the set of all people is not transitive. If person A is a friend of person B and person B is a friend of person C, it does not necessarily mean that person A is a friend of person C.

Closures

A relation with a specific property, denoted by P, is referred to as a P-relation. To capture these properties in a relation, we introduce the concept of the P-closure.

Definition. The ***P*-closure** of a relation $\mathcal{R} \subseteq A \cdot A$ on a set A, denoted by $P(\mathcal{R})$, is a P-relation that satisfies two conditions:

1. $\mathcal{R}$ is a subset of $P(\mathcal{R})$, ensuring that the original relation's elements are retained.
2. $P(\mathcal{R})$ is a subset of any other P-relation S containing $\mathcal{R}$, guaranteeing that $P(\mathcal{R})$ is the smallest P-relation containing $\mathcal{R}$.

It is important to note that not all relations have a P-closure. However, under certain conditions, the existence of $P(\mathcal{R})$ can be guaranteed. Suppose P is a property such that there exists at least one P-relation containing $\mathcal{R}$, and the intersection of any collection of P-relations is again a P-relation. In this case, it can be proven that:

$$P(\mathcal{R}) = \bigcap \{S \mid S \text{ is a } P\text{-relation and } \mathcal{R} \subseteq S\}$$

This expression ensures the existence of $P(\mathcal{R})$ under specific conditions.

Example. Consider the set $A = \{a, b, c, d\}$.

1. Let $\mathcal{R}$ be the relation $\mathcal{R} = \{(a, a), (a, c), (c, a), (d, d)\}$. This relation is symmetric but not reflexive. Its reflexive closure, is the relation

   $$P(\mathcal{R}) = \{(a, a), (a, c), (c, a), (d, d), (b, b), (c, c)\}.$$

   More generally, for an arbitrary set X and relation $\mathcal{R}$, the reflexive closure of $\mathcal{R}$ is the relation $P(\mathcal{R}) = \mathcal{R} \cup \triangle_X$, where $\triangle_X = \{(x, x) \mid x \in X\}$ represents the *diagonal* of X.

2. Let $\mathcal{R}$ be the relation $\mathcal{R} = \{(a, a), (a, b), (d, a), (a, c), (c, a), (d, d)\}$. This relation is not symmetric. Its symmetric closure, is the relation

   $$P(\mathcal{R}) = \{(a, a), (a, b), (b, a), (d, a), (a, d), (a, c), (c, a), (d, d)\}.$$

   More generally, for an arbitrary set X and relation $\mathcal{R}$, the symmetric closure of $\mathcal{R}$ is the relation $P(\mathcal{R}) = \mathcal{R} \cup \mathcal{R}^{-1}$.

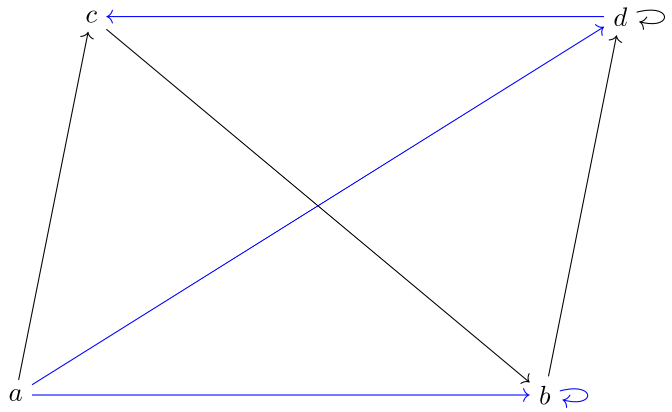
3. Let $\mathcal{R}$ be the relation $\mathcal{R} = \{(a, c), (c, b), (b, d), (d, d)\}$. This relation is not transitive and has the transitive closure

   $$P(\mathcal{R}) = \{(a, c), (c, b), (b, d), (d, d), (a, b), (d, c), (b, b), (a, d)\}.$$

   The transitive closure of a relation on a finite set can be naturally constructed by adding edges to the directed graph that represents it.

   To ensure transitivity, we examine all possible paths of length greater than one between pairs of vertices. If there exists a path from vertex a to vertex c via intermediate vertices, we add a direct edge from a to c. This process is repeated until no new edges can be added, resulting in the transitive closure of the relation.

   In essence, this construction ensures that for any two elements in the set related by the original relation, there exists a path between them in the graph representing the transitive closure, reflecting the transitive property of the relation. The graph corresponding to the transitive closure of the relation from our example is depicted below. The added edges are highlighted in blue:



In general, the transitive closure of a relation $\mathcal{R}$ on a finite set X with n elements is the union $\mathcal{R} \cup \mathcal{R}^{\circ 2} \cup \ldots \cup \mathcal{R}^{\circ n}$. Here $\mathcal{R}^{\circ i} = \underbrace{\mathcal{R} \circ \mathcal{R} \circ \ldots \circ \mathcal{R}}_{i}$ represents i compositions of relation $\mathcal{R}$ with itself.

Lecture 6

Equivalence Relations and Partitions; Functions

Equivalence Relations and Partitions

An equivalence relation on a nonempty set A is a relation that satisfies three fundamental properties: reflexivity, symmetry, and transitivity. Formally, a relation R on A is an equivalence relation if it meets the following criteria:

1. $(a, a) \in \mathcal{R}$ for every $a \in A$ (reflexivity).
2. If $(a, b) \in \mathcal{R}$, then $(b, a) \in \mathcal{R}$ (symmetry).
3. If $(a, b) \in \mathcal{R}$ and $(b, c) \in \mathcal{R}$, then $(a, c) \in \mathcal{R}$ (transitivity).

The concept behind an equivalence relation is that it classifies objects that are 'alike' in some manner.

Example. 1. The relation 'equals' (denoted by $=$) on any set A is an equivalence relation.

2. Define $a \equiv b \pmod{m}$ if a and b have the same residue modulo m. This is an equivalence relation.

3. On the set of all rational numbers, define $\frac{a}{b} \sim \frac{c}{d}$ if $ad = bc$. This is an equivalence relation.

Equivalence relations give rise to partitions by grouping together elements that are related to each other. Specifically, if $\mathcal{R}$ is an equivalence relation on a set A, the equivalence classes of $\mathcal{R}$ partition A, meaning that every element of A belongs to one and only one equivalence class.

Example. 1. Consider the set $\mathbb{Z}$ of all integers, and define $a \equiv b \pmod{7}$ if a and b have the same residue modulo 7. This defines an equivalence relation on $\mathbb{Z}$.

There are 7 equivalence classes determined by the residues modulo 7. The equivalence classes can be listed as follows:

$$\begin{aligned}
[0] &= \{\ldots, -14, -7, 0, 7, 14, \ldots\} \\
[1] &= \{\ldots, -13, -6, 1, 8, 15, \ldots\} \\
[2] &= \{\ldots, -12, -5, 2, 9, 16, \ldots\} \\
[3] &= \{\ldots, -11, -4, 3, 10, 17, \ldots\} \\
[4] &= \{\ldots, -10, -3, 4, 11, 18, \ldots\} \\
[5] &= \{\ldots, -9, -2, 5, 12, 19, \ldots\} \\
[6] &= \{\ldots, -8, -1, 6, 13, 20, \ldots\}
\end{aligned}$$

This set of equivalence classes forms a partition of $\mathbb{Z}$, where every integer is contained in exactly one equivalence class.

2. On the set of all rational numbers, we define $\frac{a}{b} \sim \frac{c}{d}$ if $ad = bc$. This definition yields an equivalence relation on the set of rational numbers.

   For instance, to find the equivalence classes of $\frac{1}{3}$, we consider all rational numbers that are equivalent to $\frac{1}{3}$ under this relation:

   $$\left[\frac{1}{3}\right] = \left\{\frac{a}{b} \mid b = 3a\right\} = \left\{\frac{1}{3}, \frac{2}{6}, \frac{3}{9}, \ldots\right\}$$

   These are all rational numbers where the denominator is three times the numerator.

Partial Ordering Relations

Definition. A relation $\mathcal{R}$ on a set A is **antisymmetric** if for all $(a, b) \in A$, if $(a, b) \in \mathcal{R}$ and $(b, a) \in \mathcal{R}$, then $a = b$.

A relation $\mathcal{R}$ on a set A is called a partial ordering or a partial order of A if $\mathcal{R}$ is reflexive, antisymmetric, and transitive. A set A together with a partial ordering $\mathcal{R}$ is called a *partially ordered set* or *poset*.

Example. 1. **Subset Relation**. On the set of all sets, define $A \leq B$ if A is a subset of B. This is a partial ordering relation.

2. **Divisibility Relation**. On the set of all positive integers, define $a \leq b$ if a divides b without leaving a remainder. This is a partial ordering relation.

3. **Lexicographic Order**. On the set of all strings of characters, define $s \leq t$ if s comes before t in the dictionary order. This is a partial ordering relation.

Functions

Functions play a central role in mathematics, serving as fundamental tools for modeling relationships and processes. At their core, a **function** can be conceptualized as a rule that uniquely assigns an output to each input. Consider functions as mathematical machines, akin to a blender seamlessly transforming various fruits into a delicious smoothie.

Example. • **Weight of Animals in a Zoo.** Consider an imaginary zoo with the following animals and weights:

  - Tiger: 500 lbs
  - Lion: 450 lbs
  - Giraffe: 1800 lbs
  - Penguin: 30 lbs
  - Polar Bear: 800 lbs

- **Movie Genres and Student Preferences.** Consider a class of about 130 students and genres such as Comedy, Romance, Horror, Science-Fiction, Documentaries, Thriller, and Action. For each genre, note the number of students who like to watch it:

  - Comedy: 45 students;
  - Romance: 40 students;
  - Horror: 50 students (most popular);
  - Science-Fiction: 30 students;
  - Documentaries: 15 students;
  - Thriller: 25 students;

  - Action: 20 students.

- Let x be a real number, and consider the function $f(x) = 2x - 5$
- Let x be a real number, and consider the function $g(x) = 16 - x^2$.

Some Descriptive Characteristics of Functions

Similar to machines having technical specifications, functions have descriptive characteristics as well.

1. **Domain:** the set of all possible inputs.
2. **Range:** the set of all possible outputs.
3. **Intercepts:** points where the graph intersects the axes (x and y intercepts).
4. **Max/Min:** maximum and minimum values of the function.

Function	Domain	Range	x- and y-intercepts	max/min
Movies	{Comedy,Romance,Horror,Sci-Fi, Documentaries,Thriller,Action}	$\{45, 40, 50, 30, 15, 25, 20\}$	-	50 and 15
Zoo Animals	{Tiger,Lion,Giraffe, Penguin,Polar Bear}	$\{500, 450, 1800, 30, 800\}$	-	800 and 30
$f(x)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-2.5, 0)$, $(0, -5)$	None
$g(x)$	$(-\infty, \infty)$	$(-\infty, 16]$	$(-4, 0), (4, 0); (0, 16)$	max at $(0, 16)$

Operations on Functions

Functions can undergo various operations similar to arithmetic operations on numbers. These include addition, multiplication, and composition.

When we add two functions $f(x)$ and $g(x)$, denoted as $(f + g)(x)$, the resulting function is obtained by adding the values of $f(x)$ and $g(x)$ for each input x:

$$(f + g)(x) = f(x) + g(x).$$

Similarly, functions can be multiplied pointwise. This means that for each input x, we multiply the values of $f(x)$ and $g(x)$ to get the value of the resulting function. It is denoted as $(f \cdot g)(x)$ and defined as

$$(f \cdot g)(x) = f(x) \cdot g(x).$$

Composition of functions, denoted as $(f \circ g)(x)$, involves applying one function after another. In other words, the output of the inner function $g(x)$ becomes the input of the outer function $f(x)$:

$$(f \circ g)(x) = f(g(x)).$$

Inverse Functions

A function g is called the **inverse** of a function f if $(f \circ g)(x) = x$ (and automatically $(g \circ f)(x) = x$). The inverse of f is denoted by f^{-1}.

Remark. The inverse of a function is **NOT** the multiplicative inverse, in other words,

$$f^{-1} \neq \frac{1}{f}.$$

Example. The formula for converting the temperature from Celsius to Fahrenheit is $F(x) = \frac{9}{5}x + 32$. We would like to find the formula for converting 'backwards': from Fahrenheit to Celsius.

Said differently, we would like to find the function $C(x)$ with $F(C(x)) = x$ (since conversion from Fahrenheit to Celsius and then back to Fahrenheit should give the initial value).

Let's denote $C(x)$ by a dummy variable y. Then we need to find y from the equation $F(y) = \frac{9}{5}y + 32 = x$. This can be done via the following sequence of algebraic transformations:

$$\frac{9}{5}y + 32 = x \Leftrightarrow \frac{9}{5}y = x - 32 \Leftrightarrow y = \frac{5}{9}(x - 32).$$

Therefore, $C(x) = \frac{5}{9}(x - 32)$.

We can (and should) check that the functions F and C are indeed inverse:

$$C(F(x)) = \frac{5}{9}\left(\frac{9}{5}x + 32 - 32\right) = \frac{5}{9} \cdot \frac{9}{5}x = x.$$

Question. There are two natural questions at this point.

1. For which functions f does the inverse function f^{-1} exist?
2. How can we find a formula for f^{-1} (provided it exists)?

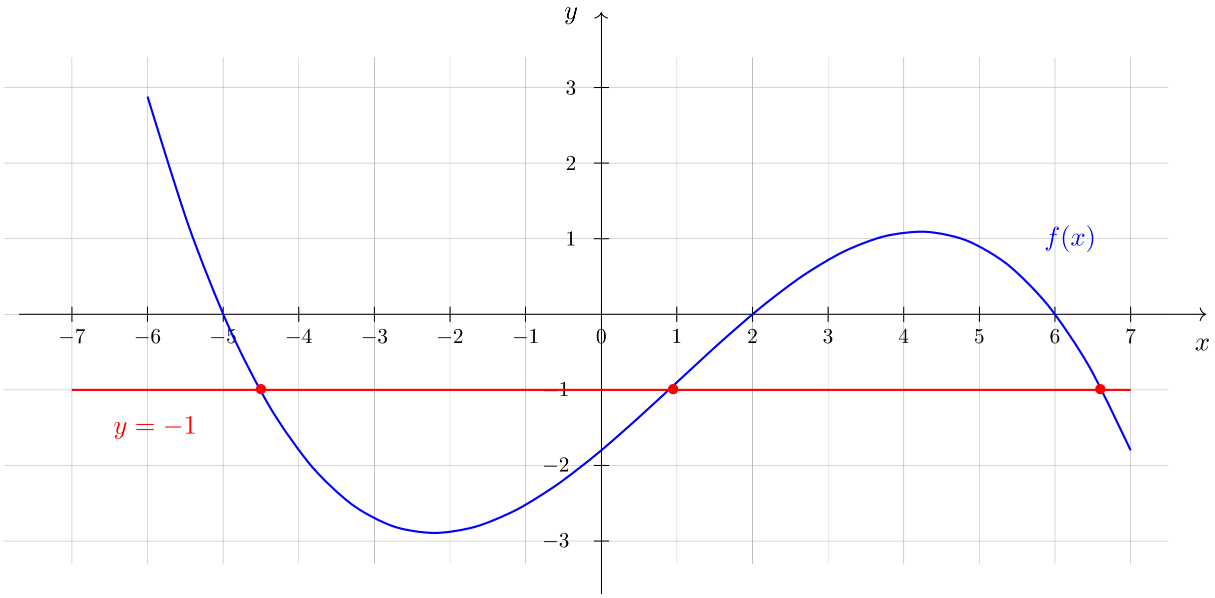
Example. Consider the function $f(x) = x^2$, then $f(1) = 1^2 = 1$, but $f(-1) = (-1)^2 = 1$ as well. If f^{-1} existed then what would $f^{-1}(1)$ be equal to? Recall that f^{-1} is a function, so we can not have $f^{-1}(1) = \{1, -1\}$. Therefore, the function $f(x) = x^2$ does not have an inverse.

One-to-one functions

A function f is called **one-to-one** (injective) if for any numbers $a \neq b$ in the domain of f, one has $f(a) \neq f(b)$.

Remark. A function f is one-to-one if and only if any horizontal line $y = c$ intersects the graph in at most one point. This statement is known as the **horizontal line test**. A function f is invertible if and only if it is one-to-one.

Example. The function $f(x)$ below is not one-to-one.



Definition. A function $f : A \to B$ is said to be **onto** (or **surjective**) if each element of B is the image of some element of A. In other words, $f : A \to B$ is onto if the image of f covers all of B, i.e., if $f(A) = B$.

Remark. Let A be a finite set and $f : A \to A$ a function. Then f is one-to-one if and only if it is onto.

Explanation. If f is one-to-one, it must map every element of A to a distinct element in A, implying the equality of cardinalities $|f(A)| \geq |A|$. But, in addition, $f(A) \subseteq A$, meaning $|f(A)| \leq |A|$. Therefore, $|f(A)| = |A|$, so f is onto. It is similar to show the reverse implication.

Lecture 7

Examples of Functions; Recursively Defined Functions

In this lecture, we will explore the concept of functions through practical examples, including recursively defined functions and algorithms.

Floor and Ceiling Functions

For any real number x, we define two important functions: the floor function and the ceiling function.

The *floor* of x, denoted by $\lfloor x \rfloor$, represents the greatest integer that does not exceed x. Similarly, the *ceiling* of x, denoted by $\lceil x \rceil$, represents the least integer that is not less than x.

When x is an integer itself, the floor and ceiling functions coincide; otherwise, the floor is one less than the ceiling.

Example. 1. $\lfloor 5.14 \rfloor = 5, \quad \lfloor \sqrt{5} \rfloor = 2, \quad \lfloor -7.3 \rfloor = -8.$

2. $\lceil 5.14 \rceil = 6, \quad \lceil \sqrt{5} \rceil = 3, \quad \lceil -7.3 \rceil = -7.$

Exponential and Logarithmic Functions

Exponential and logarithmic functions are fundamental in mathematics and have wide-ranging applications in various fields. Let's discuss their definitions and meanings.

Exponential Function

An exponential function is of the form $f(x) = a^x$, where a is a positive constant called the base, and x is the exponent. These functions represent rapid growth or decay phenomena. For example, if $a > 1$, the function grows exponentially as x increases, while if $0 < a < 1$, it decays exponentially.

Logarithmic Function

The logarithmic function is the inverse of the exponential function. It is of the form $f(x) = \ell og_a(x)$, where a is the base. This function gives the exponent to which the base must be raised to obtain x. Logarithmic functions often represent the magnitude or intensity of quantities, such as sound levels or earthquake magnitudes.

Exponential and logarithmic functions with coinciding bases are mutually inverse. An exponential function, $f(x) = a^x$, is the inverse of a logarithmic function, $g(x) = \ell og_a(x)$, and vice versa. They "undo" each other's operations. For instance, taking the logarithm of an exponential function with base a yields the exponent, and raising a to the power of the logarithm gives back the original value:

$$\ell og_a(a^x) = a^{\ell og_a(x)} = x.$$

Sequences

In mathematics, a sequence is a concept closely related to functions. It's essentially a function from the set of positive integers (or sometimes nonnegative integers) into another set. Each integer corresponds to an element in the sequence, and the sequence is typically denoted by listing its elements in order, separated by commas.

Example. 1. Consider the sequence of even numbers: $\{a_1, a_2, a_3, \ldots\} = \{2, 4, 6, \ldots\}$. In this sequence, each positive integer n corresponds to the even number $2n$, represented by the formula $a_n = 2n$.

2. Another example is the sequence of odd numbers: $\{b_1, b_2, b_3, \ldots\} = \{1, 3, 5, 7, \ldots\}$. Here, each positive integer n corresponds to the odd number $2n - 1$, given by the formula $b_n = 2n - 1$.

Sequences can also be finite, where there are only a limited number of terms. In such cases, we typically specify the range of integers over which the sequence is defined, such as 1 to m for a sequence of length m.

Summation Symbol

In mathematics, particularly in algebra and calculus, we employ the summation symbol $\sum$ (often represented by the capital Greek letter sigma) to succinctly represent the sum of a sequence of terms. Consider a sequence $a_1, a_2, a_3, \ldots$. The notation

$$\sum_{i=1}^{m} a_i = a_1 + a_2 + \ldots + a_m$$

denotes the sum of the first m terms in the sequence. We can adjust the bounds of the summation to capture specific portions of the sequence. For example,

$$\sum_{i=3}^{7} a_i = a_3 + a_4 + a_5 + a_6 + a_7$$

represents the sum of terms from the third to the seventh position. If we wish to encompass all terms from a certain point to infinity, we utilize notation such as $\sum\limits_{i=3}^{\infty} a_i$.

Here are some illustrative examples:

Example. 1. $\sum\limits_{j=1}^{5} j = 1 + 2 + 3 + 4 + 5 = 15$.

2. $\sum\limits_{k=1}^{10} (2k - 1) = 2(1) - 1 + 2(2) - 1 + \ldots + 2(10) - 1 = 1 + 3 + 5 + \ldots + 19 = 10^2 = 100$.

3. $\sum\limits_{i=0}^{3} 2^i = 2^0 + 2^1 + 2^2 + 2^3 = 1 + 2 + 4 + 8 = 15$.

Recursively Defined Functions

A function is considered to be recursively defined if its definition refers to itself. However, for this definition to be valid and non-circular, it must satisfy two essential properties:

1. There exist specific arguments, known as base values, for which the function does not reference itself.
2. Each time the function refers to itself, the argument of the function must be moving closer to a base value.

A recursively defined function that meets these criteria is said to be well-defined.

Example. Here are a few examples of recursively defined functions.

1. The factorial function, denoted as $n!$, is defined recursively as follows:

$$n! = \begin{cases} 1, & \text{if } n = 0 \\ n \cdot (n-1)!, & \text{if } n > 0 \end{cases}$$

2. The Fibonacci sequence, denoted as $F(n)$, is defined recursively by the following equations:

$$F(n) = \begin{cases} 0, & \text{if } n = 0 \\ 1, & \text{if } n = 1 \\ F(n-1) + F(n-2), & \text{if } n > 1 \end{cases}$$

3. The Ackermann function is a two-argument function where each argument takes a nonnegative integer value. It is defined as follows:

   (a) If $m = 0$, then $A(m, n) = n + 1$.

   (b) If $m \neq 0$ but $n = 0$, then $A(m, n) = A(m - 1, 1)$.

   (c) If $m \neq 0$ and $n \neq 0$, then $A(m, n) = A(m - 1, A(m, n - 1))$.
   Let's break down the computation of $A(1, 1)$ step by step:

$$A(1,1) \overset{(c)}{=} A(1-1, A(1, 1-1)) = A(0, A(1,0)) \overset{(b)}{=} A(0, A(1-1,1)) = A(0, A(0,1)) \overset{(a)}{=} A(0,2) \overset{(a)}{=} 3.$$

   Then, we can also find $A(2, 0) \overset{(c)}{=} A(2 - 1, 1) = A(1, 1) = 3$.

Lecture 8
Algorithms, Big $\mathcal{O}$ Notation and Complexity

An algorithm M is a finite step-by-step list of well-defined instructions for solving a particular problem, such as finding the output $f(X)$ for a given function f with input X. The input X may be a list or set of values. There may be more than one way to obtain $f(X)$, and the choice of algorithm M may depend on factors such as efficiency or complexity.

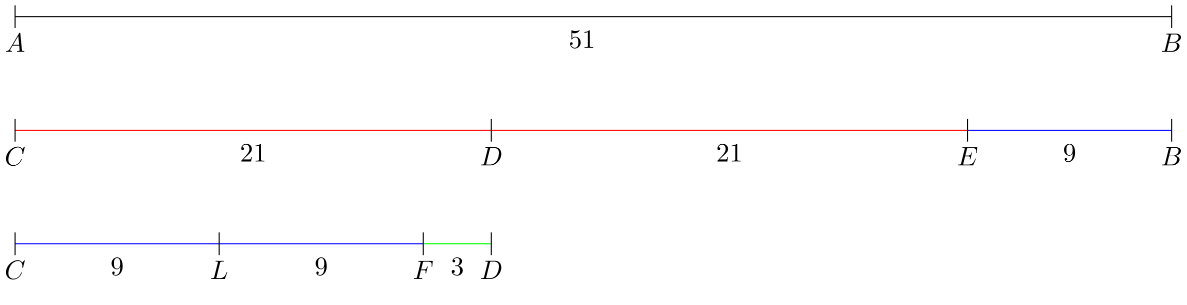
Example. The Euclidean algorithm, attributed to the ancient Greek mathematician Euclid around 300 BC, stands as a testament to the enduring power of elegant problem-solving methods. This algorithm presents a remarkably efficient approach to determining the greatest common divisor (GCD) of two integers. The GCD represents the largest integer that evenly divides both given numbers without leaving a remainder.

Rooted in Euclid's seminal work, ”Elements,” the algorithm remains a cornerstone of number theory and computational mathematics. As one of the oldest algorithms still widely utilized today, it exemplifies the timeless nature of mathematical principles.

Conceptually, Euclid's method resonates with geometric intuition. Consider two line segments, AB and CD, defined as multiples of a common ”unit” length. Visualize AB as 51 units long and CD as 21 units long. Euclid's approach involves iteratively ”measuring” the shorter segment (CD) against the longer one (AB), seeking the remainder each time.

Starting with CD measuring AB, we find a remainder, EB, of length 9. Next, EB becomes the new measuring stick, measuring CD twice, yielding a remainder, FD, of length 3. Continuing this process, FD measures EB three times with no remainder, signaling the end of the calculation.

Thus, the smallest length attainable from AB and CD corresponds to the length of FD, which equals 3 units. This result precisely mirrors the GCD of the original numbers, 51 and 21:



Euclid's algorithm underlies countless mathematical and cryptographic computations.

Let's elucidate this method with an example. Consider $a = 316$ and $b = 56$:

$$\begin{aligned} 316 &= 56 \cdot 5 + 36 \\ 56 &= 36 \cdot 1 + 20 \\ 36 &= 20 \cdot 1 + 16 \\ 20 &= 16 \cdot 1 + 4 \\ 16 &= 4 \cdot 4 + 0 \end{aligned}$$

Following this series of divisions, the process concludes with a remainder of 0, indicating that we have found the greatest common divisor. In this case, $\text{GCD}(316, 56) = 4$.

Complexity of Algorithms

One of the important characteristics of an algorithm is its running time. Let's consider an example to understand this concept better.

Suppose we have a collection (array) Arr of 1000 ordered numbers and a separate number a. We would like an algorithm that determines if a belongs to Arr.

The naive approach would be to check all the numbers one by one: 1000 checks in the worst case.

Now, let's discuss a more efficient approach using binary search. In binary search, we repeatedly divide the search interval in half until we narrow down the search to just one element.

For the worst case, let's say a is not in Arr. In each step of binary search, we reduce the size of the search interval by half. Initially, we have 1000 elements, then 500, then 250, and so on.

After k steps, where k is the number of times we divide by 2 until we reach one element, the size of the search interval is 2^{-k} times the size of the original interval.

Therefore, to find a single element, binary search takes at most $\ell og_2 1000 \approx 10$ steps.

In conclusion, binary search is significantly faster than the naive approach for searching in an ordered collection.

Example. Let's illustrate how binary search works step by step using a concrete example: an array $\text{Arr} = \{1, 2, 3, 3, 5, 7, 8, 11, 14, 17, 23, 23, 24, 29, 31\}$ and a number $a = 29$.

Step 1. We compare a (which is 29) with the middle element of the array, which is 11. Since 29 is greater than 11, we discard the left half of the array. The updated array becomes $\{14, 17, 23, 23, 24, 29, 31\}$.

Step 2. We repeat the process. Now, we compare 29 with the middle element of the new array, which is 23. Since 29 is greater than 23, we discard the left half again. The new array becomes $\{24, 29, 31\}$.

Step 3. Now, we compare 29 with the middle element of the new array, which is 29. Since 29 equals 29, we have found the element we were searching for.

Thus, using binary search, we have successfully found that 29 belongs to the array Arr in 3 checks. If we checked each element starting from the leftmost one, it would require examining 14 elements of the array, which is noticeably less efficient.

One of the main mathematical tools for analyzing complexity is **asymptotic analysis**, which aims to understand the behavior of algorithms as the size of the input grows. This analysis often involves quantifying the runtime and memory usage of algorithms in terms of the input size. A fundamental concept in asymptotic analysis is the 'Big $\mathcal{O}$', which provides a formal way to express the upper bound on the growth rate of an algorithm's runtime or space complexity in terms of the input size.

Big $\mathcal{O}$ Notation

Big $\mathcal{O}$ notation provides insights into the efficiency of algorithms by indicating how their performance scales with input size. Consider a scenario where you have a list of size n. With a simple search algorithm, each element needs to be checked, resulting in n operations. In terms of Big $\mathcal{O}$ notation, this runtime is expressed as $\mathcal{O}(n)$. But where is the time measurement? There isn't one—Big $\mathcal{O}$ notation abstracts away from time units like seconds. Instead, it enables comparison based on the number of operations, elucidating how the algorithm's efficiency grows with increasing input size. Binary search needs $\ell og(n)$ operations to check a list of size n. What's the running time in Big $\mathcal{O}$ notation? It is $\mathcal{O}(\ell og(n))$.

We need a more precise definition to understand this concept better and be able to operate with it.

Definition. Let $f(x)$ and $g(x)$ be two real-valued functions, with $g(x)$ being strictly positive for large x-values. We say that $f(x)$ is **big $\mathcal{O}$ of** $g(x)$, denoted as $f(x) = \mathcal{O}(g(x))$ as $x \to \infty$, if the absolute value of $f(x)$ is at most a positive constant multiple of $g(x)$ for all sufficiently large values of x. In other words, $f(x) = \mathcal{O}(g(x))$ if there exists a positive real number M and a real number x_0 such that $|f(x)| \leq Mg(x)$ for all $x \geq x_0$. Alternatively, for those familiar with limit notation, $f(x) = \mathcal{O}(g(x))$ if $\lim\limits_{x\to\infty} \frac{|f(x)|}{g(x)} = M$.

In this definition, $g(x)$ serves as the comparison function. Common choices for $g(x)$ include $\ell og_2(x)$, x, $x\ell og_2(x)$, x^k, and 2^k.

Remark. It is important to remember that although the constant M in the definition of big $\mathcal{O}$ may not appear significant from a theoretical standpoint, it holds considerable practical importance. For example, consider the functions $0.0001x$ and 10^6x, both of which are $\mathcal{O}(x)$. However, an algorithm with the former complexity will execute noticeably faster, unless the input has an exceptionally large order of magnitude.

Example. Consider an array $Arr = [5, 4, 3, 2, 1]$. We want to arrange the elements of Arr in increasing (non-decreasing) order.

Bubble sort is a simple sorting algorithm that repeatedly steps through the list, compares adjacent elements, and swaps them if they are in the wrong order. This process continues until the entire array is sorted.

For our example array, the sorting process would proceed as follows:

- Pass 1: Compare 5 and 4, they are in the wrong order ($5 > 4$), swap them. Compare 5 and 3, wrong order, swap them. Compare 5 and 2, wrong order, swap them. Compare 5 and 1, wrong order, swap them. Result: $[4, 3, 2, 1, 5]$. Number of swaps: 4.
- Pass 2: Compare 4 and 3, wrong order, swap them. Compare 4 and 2, wrong order, swap them. Compare 4 and 1, wrong order, swap them. Result: $[3, 2, 1, 4, 5]$. Number of swaps: 3.
- Pass 3: Compare 3 and 2, wrong order, swap them. Compare 3 and 1, swap them. Result: $[2, 1, 3, 4, 5]$. Number of swaps: 2.
- Pass 4: Compare 2 and 1, wrong order, swap them. Result: $[1, 2, 3, 4, 5]$. Number of swaps: 1.

The array is now sorted in ascending order. In this example, the total number of swaps performed is $4 + 3 + 2 + 1 = 10$.

Bubble sort gets its name from the way smaller elements "bubble" to the top of the list in each pass, analogous to bubbles rising in liquid.

In bubble sort, the worst-case time complexity for an array of size n is $\mathcal{O}(n^2)$. This occurs when, for each element in the array, we must compare and potentially swap it with every other element. Specifically, we may need to perform $\frac{(n-1)n}{2}$ operations, as discussed in a previous lecture.

To demonstrate that $f(x) = \frac{(x-1)x}{2}$ belongs to $\mathcal{O}(x^2)$ as $x \to \infty$, we can evaluate the limit:

$$\lim_{x\to\infty} \frac{f(x)}{x^2} = \lim_{x\to\infty} \frac{\frac{(x-1)x}{2}}{x^2} = 0.5 \lim_{x\to\infty} \frac{x^2 - x}{x^2} = 0.5 \lim_{x\to\infty} \left(1 - \frac{1}{x}\right) = 0.5.$$

Thus, as x approaches infinity, the ratio $\frac{f(x)}{x^2}$ approaches 0.5, indicating that $f(x)$ grows asymptotically no faster than x^2. Therefore, $f(x)$ is indeed $\mathcal{O}(x^2)$.

Example. Let $f(x) = a_kx^k + a_{k-1}x^{k-1} + \ldots + a_0$ be a polynomial of degree k. We want to show that $\lim\limits_{x\to\infty} \frac{f(x)}{x^k} = a_k$, which implies that $f(x) = \mathcal{O}(x^k)$.

To prove this, let's divide every term of $f(x)$ by x^k and take the limit as x approaches infinity:

$$\frac{f(x)}{x^k} = a_k + \frac{a_{k-1}}{x} + \ldots + \frac{a_0}{x^k}.$$

Now, as x approaches infinity, all terms with x in the denominator go to zero, leaving only a_k:

$$\lim_{x\to\infty} \frac{f(x)}{x^k} = \lim_{x\to\infty} \left(a_k + \frac{a_{k-1}}{x} + \ldots + \frac{a_0}{x^k}\right) = a_k$$

This proves that $\lim\limits_{x\to\infty} \frac{f(x)}{x^k} = a_k$, which implies that $f(x) = \mathcal{O}(x^k)$.

Algorithms with such complexity are called **polynomial-time algorithms**.

Lecture 9

Logic and Propositional Calculus, I

Logic serves as the foundation of reasoning and argumentation, providing a systematic framework for analyzing and evaluating the validity of statements and arguments. We will explore the fundamental concepts and principles of propositional calculus, which forms the basis of formal logic, allowing us to express and manipulate propositions using logical operators and truth values.

Propositions and Compound Statements

A proposition (or statement) is a declarative sentence that can be either true or false, but not both simultaneously. Let's examine the following sentences:

1. "Water boils at 100 degrees Celsius." This statement is true.
2. "$6 \cdot 9 = 42$." This statement is false.
3. "How was your day?" This is not a proposition, as it is a question, not a declarative statement.
4. "Australia is a country." This statement is true.
5. "$7 - 7 = 0$." This statement is true.
6. "Clean your room." This is not a proposition, as it is a directive or command.

Compound Propositions

Compound propositions are those comprised of multiple subpropositions connected by various logical operators, as discussed later in this section. These operators allow us to combine simple propositions into more complex ones. Conversely, a proposition that cannot be broken down into simpler components is called *primitive.*

Example. 1. "The Earth orbits around the Sun." (Primitive)

2. "$2 + 2 = 4$" (Primitive)
3. "Water is essential for life, and oxygen is necessary for breathing." (Compound - made of 2 primitives)
4. "The moon is made of cheese or the Earth is flat." (Compound - made of 2 primitives)
5. "The cat is sleeping and the dog is barking or the mailman is delivering letters." (Compound, made of 3 primitives)

The essential characteristic of a compound proposition lies in its truth value, which is entirely dictated by the truth values of its constituent subpropositions and the manner in which they are interconnected to form the compound statement.

Basic Logical Operations

We will begin by examining the three fundamental logical operations: conjunction, disjunction, and negation. These operations align with the English words "and," "or," and "not," respectively.

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

Truth table for $p \wedge q$

Conjunction, $p \wedge q$

Any two propositions can be combined by the word "and" to form a compound proposition called the *conjunction* of the original propositions. Symbolically,

$$p \wedge q$$

is read as "p and q," indicating the conjunction of p and q. Since $p \wedge q$ is a proposition, it has a truth value, and this truth value depends only on the truth values of p and q.

Definition. If p and q are true, then $p \wedge q$ is true; otherwise, $p \wedge q$ is false:

In the truth table above:

- the first column represents the truth value of proposition p;
- the second column represents the truth value of proposition q;
- the third column represents the truth value of the conjunction $p \wedge q$;
- the letters "T" and "F" stand for "True" and "False," respectively.

Example.

1. Consider the statement "Birds have feathers and $6-2=4$". Let p represent the proposition "Birds have feathers" (p is true) and q represent the proposition "$6-2=4$" (q is true).

   Since both p and q are true, according to the first row of truth Table 1, $p \wedge q$ is true. Therefore, the statement "Birds have feathers and $6-2=4$" is true.

2. Similarly, for the statement "Birds have feathers and $6-2=5$", p is true and q is false ($6-2=4 \neq 5$). According to the second row of truth Table 1, $p \wedge q$ is false. Therefore, the statement "Birds have feathers and $6-2=5$" is false.

Disjunction, $p \vee q$

Any two propositions can be combined by the word "or" to form a compound proposition called the **disjunction** of the original propositions, denoted by $p \vee q$. The truth value of $p \vee q$ depends only on the truth values of p and q as follows:

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

Truth table for $p \vee q$

Example.

1. Consider the statement "Birds have feathers or $6-2=20$".

   Let p represent the statement 'Birds have feathers'. Let q represent the statement '$6-2=20$'. Given that p is true (since birds indeed have feathers) and q is false (as $6-2=4 \neq 20$), according to the second line of the truth Table 2, the compound statement $p \vee q$ is true.

Hence, the statement ”Birds have feathers or $6 - 2 = 20$” evaluates to true. Despite the falsehood of the second component, the truth of the first component is sufficient to render the entire statement true.

2. Let’s take a look at one more statement: ”Humans have wings or elephants can climb trees”.

   Let p represent the statement ”Humans have wings” and q represent the statement ”Elephants can climb trees.” Since both statements p and q are false, according to the truth table for the logical OR operator, the statement ”Humans have wings or elephants can climb trees” is false. This corresponds to the last row of Table 2.

Remark. The English word “or” is commonly employed in two distinct contexts. In one usage, it denotes “p or q or both,” indicating that at least one of the alternatives occurs. Conversely, it can signify “p or q but not both,” specifying that exactly one of the alternatives occurs. For instance, consider the sentence: “The team will either triumph in the championship or face defeat in the playoffs,” which employs “or” in the latter sense. This exclusive disjunction is denoted by $\oplus$ and its truth table is depicted below

p	q	$p \oplus q$
T	T	F
T	F	T
F	T	T
F	F	F

Truth table for $p \oplus q$

Unless explicitly stated otherwise, the term “or” is assumed to signify the inclusive sense. This discussion underscores the precision afforded by our symbolic language: $p \vee q$ is defined by its truth table and consistently represents “p and/or q.”

Negation, $\neg p$

Given any proposition p, another proposition, called the negation of p, can be formed by writing “It is not true that . . .” or “It is false that . . .” before p or, if possible, by inserting in p the word “not.” Symbolically, the negation of p, read “not p,” is denoted by $\neg p$.

The truth value of $\neg p$ depends on the truth value of p as follows:

Definition. If p is true, then $\neg p$ is false; and if p is false, then $\neg p$ is true:

p	$\neg p$
T	F
F	T

Truth table for $\neg p$

The truth value of the negation of p is always the opposite of the truth value of p.

Example. Let p be the statement ”The sun rises in the west.” Then p is false because the sun rises in the east. Therefore, $\neg p$ is true (second line of Table 4).

Truth Tables for Compound Propositions

The main property of a compound proposition $P(p, q, \ldots)$ is that its truth value depends exclusively upon the truth values of its constituents. In other words, the truth value of a proposition can be uniquely determined once the truth value of each of its variables is known.

Example. Suppose we would like to construct the truth table for the proposition $P = (\neg p \wedge \neg q) \vee r$. This means that for each triplet of truth values for statements p, q, and r, we can determine the truth value of P directly.

For instance, suppose all statements are true. Then $\neg p$ and $\neg q$ are both false, so $\neg p \wedge \neg q$ is false as well. However, r is true, so $(\neg p \wedge \neg q) \vee r$ is true as well. Alternatively, we could start by noticing that r is true, so regardless of the value of the statement that "or" is taken with, the value of P will be true. However, we need to perform similar analysis for the 7 other combinations of values of p, q, and r.

In case $P(p_1, \ldots, p_n)$ consists of n statements, there will be 2^n n-tuples to analyze. Instead, we can find the truth table for P faster if we consistently use Tables 1, 2 and 4 to substitute P with an equivalent proposition that has the same truth table but includes one less logical operation. The process terminates with no logical operations remaining.

Let's illustrate this process with our proposition $P = (\neg p \wedge \neg q) \vee r$. We'll demonstrate step-by-step how to construct its truth table efficiently.

First, we begin by establishing the truth values for $\neg p$ and $\neg q$ (Table 5). This involves negating the truth values of p and q respectively:

p	q	r	$\neg p$	$\neg q$
T	T	T	F	F
T	T	F	F	F
T	F	T	F	T
T	F	F	F	T
F	T	T	T	F
F	T	F	T	F
F	F	T	T	T
F	F	F	T	T

Truth table for $\neg p, \neg q$

Next, utilizing the truth values obtained for $\neg p$ and $\neg q$, we construct the truth table for the statement $\neg p \wedge \neg q$ (Table 6). This involves applying the logical operation AND to the truth values of $\neg p$ and $\neg q$:

p	q	r	$\neg p$	$\neg q$	$\neg p \wedge \neg q$
T	T	T	F	F	F
T	T	F	F	F	F
T	F	T	F	T	F
T	F	F	F	T	F
F	T	T	T	F	F
F	T	F	T	F	F
F	F	T	T	T	T
F	F	F	T	T	F

Truth table for $\neg p \wedge \neg q$

Subsequently, we use the truth table for $\neg p \wedge \neg q$ to derive the truth table for P (Table 7). This entails applying the logical operation OR between the result of $\neg p \wedge \neg q$ and the truth values of r.

Finally, to present a simplified view of the truth table for P, we exclude the auxiliary columns used in computation, resulting in Table 8.

This systematic approach allows us to efficiently construct the truth table for complex propositions by leveraging simpler components and logical operations.

p	q	r	$\neg p$	$\neg q$	$\neg p \wedge \neg q$	$P = (\neg p \wedge \neg q) \vee r$
T	T	T	F	F	F	T
T	T	F	F	F	F	F
T	F	T	F	T	F	T
T	F	F	F	T	F	F
F	T	T	T	F	F	T
F	T	F	T	F	F	F
F	F	T	T	T	T	T
F	F	F	T	T	F	F

Truth table for P

p	q	r	P
T	T	T	T
T	T	F	F
T	F	T	T
T	F	F	F
F	T	T	T
F	T	F	F
F	F	T	T
F	F	F	F

Simplified truth table for P

Lecture 10
Logic and Propositional Calculus, II

We continue with our exploration of logic and propositional calculus, introducing more essential elements of formal logic to deepen our understanding of these fundamental concepts.

Tautologies and Contradictions

Some propositions $P(p, q, \ldots)$ exhibit a particular pattern in their truth tables, where the last column contains only T or F. Propositions with this characteristic are termed tautologies and contradictions, respectively.

A proposition $P(p, q, \ldots)$ is classified as a *tautology* if its truth table reveals that it evaluates to true for all possible combinations of truth values of its variables. Conversely, a proposition is labeled a *contradiction* if its truth table indicates that it evaluates to false for all possible combinations of truth values of its variables.

For instance, consider the proposition "p or not p", represented as $p \vee \neg p$. This proposition is identified as a tautology because its truth table demonstrates that it evaluates to true regardless of the truth value of p.

Conversely, the proposition "p and not p", denoted as $p \wedge \neg p$, is recognized as a contradiction. This designation is confirmed by observing its truth table, which illustrates that it evaluates to false for all possible truth values of p.

This distinction between tautologies and contradictions can be seen from their respective truth tables:

p	$p \vee \neg p$
T	T
F	T

p	$p \wedge \neg p$
T	F
F	F

Truth tables for tautology and contradiction.

It is worth noting that the negation of a tautology results in a contradiction, as a tautology is always true. Conversely, the negation of a contradiction yields a tautology, as a contradiction is always false.

Now, consider a proposition $P(p, q, \ldots)$ that is a tautology, meaning it evaluates to true for all possible truth values of its variables $p, q, \ldots$. Let $P_1(p, q, \ldots)$, $P_2(p, q, \ldots)$, and so forth, be any propositions. Since the truth value of $P(p, q, \ldots)$ is independent of the specific truth values of its variables, we can substitute P_1 for p, P_2 for q, and so on, in the tautology $P(p, q, \ldots)$ and still retain a tautology.

In other words, if $P(p, q, \ldots)$ is a tautology, then $P(P_1, P_2, \ldots)$ is also a tautology for any propositions $P_1, P_2, \ldots$.

Logical Equivalence

Logical equivalence is a fundamental concept in propositional logic. Two propositions $P(p, q, \ldots)$ and $Q(p, q, \ldots)$ are considered logically equivalent, denoted by $P(p, q, \ldots) \equiv Q(p, q, \ldots)$, if they share identical truth tables.

In other words, if for every possible combination of truth values for the variables p, q, and so forth, the truth values of P and Q match, then they are pronounced logically equivalent.

This notion of equivalence, often referred to simply as equivalence or equality, signifies that despite potential differences in their symbolic expressions, P and Q convey the same logical meaning. It underscores the idea that these propositions are indistinguishable in terms of their truth conditions.

Example. De Morgan's laws provide a classic example of logical equivalence. They state:

$$\neg(p \wedge q) \equiv \neg p \vee \neg q$$
$$\neg(p \vee q) \equiv \neg p \wedge \neg q$$

These laws assert that the negation of a conjunction is logically equivalent to the disjunction of the negations of its components, and vice versa.

We can verify the validity of De Morgan's laws by employing truth tables:

p	q	$\neg(p \wedge q) \equiv \neg p \vee \neg q$
T	T	F
T	F	T
F	T	T
F	F	T

p	q	$\neg(p \vee q) \equiv \neg p \wedge \neg q$
T	T	F
T	F	F
F	T	F
F	F	T

Conditionals and Biconditionals

In many scientific contexts, statements often follow the structure "If p, then q." These are termed conditional statements and are represented by $p \to q$, interpreted as "if p, then q" or "p only if q".

Another common type of statement is the biconditional, which expresses "p if and only if q". This is denoted by $p \leftrightarrow q$ and essentially represents the conjunction of two conditionals: $p \to q$ and $q \to p$.

The truth values of $p \to q$ and $p \leftrightarrow q$ are determined by the following truth tables:

p	q	$p \to q$
T	T	T
T	F	F
F	T	T
F	F	T

p	q	$p \leftrightarrow q$
T	T	T
T	F	F
F	T	F
F	F	T

We make the following observations.

1. The conditional $p \to q$ is false only when p is true and q is false. However, if p is false, the conditional $p \to q$ is true regardless of the truth value of q.

2. The biconditional $p \leftrightarrow q$ is true if p and q have the same truth values, and false otherwise.

Remark. It's worth noting that the truth table for $\neg p \vee q$ is identical to that of $p \to q$, indicating that these propositions are logically equivalent.

Arguments

An argument is a claim that a set of propositions $P_1, P_2, \ldots, P_n$, known as *premises*, leads to another proposition Q, known as the *conclusion*. This relationship is symbolized as $P_1, P_2, \ldots, P_n \vdash Q$.

The concept of a ”logical argument” or ”valid argument” is defined as follows.

Definition. An argument $P_1, P_2, \ldots, P_n \vdash Q$ is valid if Q is true whenever all the premises $P_1, P_2, \ldots, P_n$ are true. An argument that fails this condition is termed a *fallacy*.

Consider the following examples:

Example.

1. Consider the argument $p, p \to q \vdash q$. We can confirm its validity using the truth table. Specifically, when both p and $p \to q$ are true, q is also true. This argument is known as the Law of Detachment.

2. The argument $p \to q, q \vdash p$ is a fallacy: if $p \to q$ and q are both true, it does not necessarily follow that p is true.

We note that the propositions $P_1, P_2, \ldots, P_n$ are true simultaneously if and only if the proposition $P_1 \wedge P_2 \wedge \ldots \wedge P_n$ is true. Therefore, the validity of the argument $P_1, P_2, \ldots, P_n \vdash Q$ hinges on Q's truth whenever $P_1 \wedge P_2 \wedge \ldots \wedge P_n$ is true, or equivalently, if the proposition $P_1 \wedge P_2 \wedge \ldots \wedge P_n \to Q$ is a tautology.

Law of Syllogism

The Law of Syllogism is a fundamental concept in logic. It states that if $p \to q$ and $q \to r$ are true statements, then $p \to r$ is also true. In simpler terms, if p implies q and q implies r, then p implies r. This law bears resemblance to the concept of transitivity, where the relationship between p and r is inferred through the intermediate step of q.

Example. Let's illustrate this law with a real-life example. Suppose p represents the statement ”It is raining,” q represents ”I will take an umbrella,” and r represents ”I will stay dry.”

If it's true that if it's raining (p), I will take an umbrella (q), and if I take an umbrella (q), I will stay dry (r), then it logically follows that if it's raining (p), I will stay dry (r).

Now, let's verify this law using a truth table:

p	q	r	$p \to q$	$q \to r$	$(p \to q) \wedge (q \to r)$	$p \to r$	$((p \to q) \wedge (q \to r)) \to (p \to r)$
T	T	T	T	T	T	T	T
T	T	F	T	F	F	F	T
T	F	T	F	T	F	T	T
T	F	F	F	T	F	F	T
F	T	T	T	T	T	T	T
F	T	F	T	F	F	T	T
F	F	T	T	T	T	T	T
F	F	F	T	T	T	T	T

This truth table confirms that $p \to r$ holds true whenever both $p \to q$ and $q \to r$ are true, validating the Law of Syllogism.

Propositional Functions and Quantifiers

Propositional functions and quantifiers are foundational concepts in logic, providing a framework for expressing statements about sets of objects and their properties.

Definition. Let A be a set. A **propositional function** (also referred to as an open sentence or condition) defined on A is a function $p : A \to \{\text{True}, \text{False}\}$ that assigns either true or false to each element $a \in A$.

In simpler terms, substituting any element $a \in A$ for the variable x transforms $p(x)$ into a statement with a definite truth value. We refer to the set A as the *domain* of $p(x)$. Furthermore, the set T_p, which consists of all elements of A for which $p(a)$ is true, is termed the *truth set* of $p(x)$. In other words, T_p can be represented as $\{x \mid x \in A, p(x) \text{ is true}\}$ or simply $\{x \mid p(x)\}$.

Often, when A represents a set of numbers, the condition $p(x)$ typically manifests as an equation or inequality involving the variable x.

Example. Let's explore three examples akin to determining the truth set for the propositional function $p(x)$ defined on the set $\mathbb{Z}$ of integer integers.

1. Consider $p(x)$ as $x > -1$. Its truth set consists of all integers greater than -1, i.e. $T_p = \{-1, 0, 1, \ldots\}$.
2. Suppose $p(x)$ represents $x^2 - 3x + 2 = 0$. The truth set is $\{1, 2\}$, as these are the integers satisfying the equation.
3. Let's examine $p(x)$ as x being an even number. Here, the truth set is $\{\ldots, -4, -2, 0, 2, 4, 6, \ldots\}$, encompassing all positive even integers.

These examples showcase how propositional functions help us identify subsets of a given set based on specific conditions.

Furthermore, we can employ quantifiers to express the extent to which a property holds true within a set. The *universal quantifier* $\forall$ denotes that a property holds true for all elements in a set, while the *existential quantifier* $\exists$ signifies that there exists at least one element in a set for which the property holds true.

Example.

1. Consider the proposition $p(x)$ defined as $x \equiv 0 \pmod 2$, meaning "x is divisible by 2."

   The statement $\forall x \in \mathbb{Z}$, $p(x)$ asserts that every integer is divisible by 2. This statement is false because there exist integers that are not divisible by 2 (e.g., 1 or -3).

2. Consider the proposition $q(x)$ defined as $x \equiv 0 \pmod 5$, meaning " x leaves a remainder of 0 when divided by 5."

   The statement $\exists x \in \mathbb{Z}$, $q(x)$ implies that there exists at least one integer such that when divided by 5, it leaves a remainder of 0. This statement is true because, for example, $x = 10$ satisfies the condition.

Negation of Quantified Statements

Consider the statement: "All planets in the solar system have rings." Its negation reads: "It is not the case that all planets in the solar system have rings" or, equivalently, "There exists at least one planet in the solar system that does not have rings."

Symbolically, let P denote the set of planets in the solar system. The negation can be expressed as:

$$\neg(\forall x \in P)(x \text{ has rings}) \equiv (\exists x \in P)(x \text{ does not have rings})$$

or, using $p(x)$ to denote "x has rings,"

$$\neg(\forall x \in P)p(x) \equiv (\exists x \in P)\neg p(x) \text{ or } \neg\forall x\ p(x) \equiv \exists x\ \neg p(x)$$

This concept holds true for any proposition $p(x)$. The following results are attributed to De Morgan.

1. For any proposition $p(x)$ defined over a set A, De Morgan's first law states:

$$\neg(\forall x \in A)p(x) \equiv (\exists x \in A)\neg p(x)$$

   In simpler terms, this law asserts that the negation of a universal quantification is equivalent to the existence of a counterexample. More explicitly:

   - The negation of "For all x in A, $p(x)$ holds true" is "There exists an x in A such that $p(x)$ does not hold true."

2. De Morgan’s second law complements the first, offering a parallel expression for negating existentially quantified statements:

$$\neg(\exists x \in A)p(x) \equiv (\forall x \in A)\neg p(x)$$

   Here, the negation of an existential quantification corresponds to the assertion that no counterexample exists within the set A. To put it succinctly:

   - The negation of ”There exists an x in A such that $p(x)$ holds true” is ”For all x in A, $p(x)$ does not hold true.”

De Morgan’s laws are foundational in logic, providing powerful tools for reasoning about the negation of quantified statements.

Example. Let’s examine the statement $\forall x \in \mathbb{R}$, $x^2 - x > 0$. However, this statement is not universally true, as illustrated by the counterexample $x = 0.1$. Specifically, when $x = 0.1$, $0.1^2 - 0.1 = -0.09$, which is not greater than zero; instead, it is less than or equal to zero.

Lecture 11
Basic Counting Principles

Our focus today centers around the multiplication rule and counting principles, crucial elements in the world of probability. These tools serve as the cornerstone of probability theory with finite sample spaces, providing us with the means to address a diverse array of probabilistic situations.

The Multiplication Rule

When we have multiple experiments, each with its own set of outcomes, we can calculate the total number of possible outcomes by multiplying the number of outcomes for each experiment.

Building a Sandwich

Imagine you have 4 types of bread, 5 types of fillings, and 3 types of condiments to choose from. You’d like to create a sandwich with 1 piece of bread, one filling, and one condiment. To find the total number of possible sandwich combinations, you need to consider all the choices you have for each component. You have:

- 4 options for the type of bread,
- 5 options for the filling, and
- 3 options for the condiment.

To find the total number of combinations, you multiply these options together:

$$4 \cdot 5 \cdot 3 = 60.$$

This means you have a total of 60 unique sandwich combinations to choose from.

Drawing Cards Without Replacement

Imagine we have a standard deck of 52 cards. If we draw three cards without replacement, the number of possible outcomes can be calculated as:

$$52 \cdot 51 \cdot 50$$

Each draw affects the available choices for the next draw, giving us a total of 132,600 possible combinations.

General Formula

Let's say we have k experiments, where the first experiment has m_1 outcomes, the second experiment has m_2 outcomes, and so on, up to the k-th experiment with m_k outcomes. The total number of possible outcomes is given by:

$$m_1 \cdot m_2 \cdot \ldots \cdot m_k$$

This is known as the General Multiplication Rule.

Number of Arrangements & Factorial

Suppose you have 5 different books and you want to arrange them on a shelf.

The first spot can be occupied by any of the 5 books. So, there are 5 choices for the first spot.

Now, for the second spot, you have 4 remaining books to choose from, since you've already placed one book in the first spot.

Similarly, for the third spot, there are 3 remaining choices, and so on.

To find the total number of arrangements, you multiply the number of choices for each spot:

$$5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120.$$

This can be expressed as 5!, which is the factorial of 5.

Definition. The **factorial** of a non-negative integer n, denoted as $n!$, is the product of all positive integers up to n, with the convention that $0! = 1$. Mathematically, it is defined as:

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1$$

.

Example. Consider the word "WORK". If we want to rearrange its letters, there are 4! ways to do so:

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

This means there are 24 different permutations of the letters in the word "WORK".

Arrangements: Order Matters

The number of ways to arrange k items from a set of n distinct items, considering the order, is denoted as $P(n, k)$ and can be calculated using the formula:

$$P(n, k) = \frac{n!}{(n-k)!}$$

This formula represents permutations, where the order of arrangement matters.

Explanation

To understand why this formula computes the number of arrangements, let's break it down:

- $n!$ represents the number of ways to arrange all n items without any restrictions. This is because there are n choices for the first item, $n-1$ choices for the second item, and so on, down to 1 choice for the last item.

- $(n-k)!$ represents the number of ways to arrange the remaining $(n-k)$ items after k items have been selected and arranged. This accounts for the fact that we are considering k items and their order, so we don't want to count the arrangements of the unselected items.

Dividing $n!$ by $(n-k)!$ eliminates the arrangements of the unselected items, leaving us with the number of arrangements of the k selected items.

Therefore, $P(n, k)$ gives us the total number of permutations of k items chosen from n distinct items, considering the order of arrangement.

Example. In a basketball tournament, there are 10 players competing for 5 spots on the starting lineup, and each spot has a specific position. How many different starting lineups can be formed?

$$P(10,5) = \frac{10!}{5!} = 30240$$

There are 30240 different possible starting lineups with positions considered.

Arrangements: Order Doesn't Matter

The number of ways to choose k items from a set of n distinct items, without considering the order, is denoted as $C(n,k)$ or $\binom{n}{k}$, and can be calculated using the formula:

$$C(n,k) = \frac{n!}{k! \cdot (n-k)!}$$

This formula represents combinations, where the order of selection doesn't matter.

Explanation

To understand why this formula counts the right number, let's refer back to the concept of permutations discussed earlier.

In permutations, we calculated $P(n,k)$ to find the number of arrangements of k items from n distinct items where order matters. This means that if we had the same k items, there would be $k!$ different ways to arrange them.

However, in combinations, we want to find the number of ways to choose k items without considering the order. This means that for each combination of k items, there are $k!$ different arrangements that are equivalent. Therefore, to get the correct count of combinations, we divide $P(n,k)$ by $k!$ to eliminate the $k!$ duplicate arrangements.

This leads us to the formula for combinations, $C(n,k) = \frac{n!}{k! \cdot (n-k)!}$, which accurately counts the number of ways to choose k items from n distinct items, without considering the order.

Example. If you have 8 people and you want to form a committee of 3 members, the number of ways to do this is:

$$C(8,3) = \frac{8!}{3! \cdot 5!} = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} = 56$$

So, there are 56 different ways to choose a committee of 3 from a group of 8 people.

Example. In a basketball tournament, there are 10 players competing for 5 spots on the starting lineup. How many different starting lineups can be formed?

$$C(10,5) = \frac{10!}{5! \cdot 5!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 252$$

There are 252 different possible starting lineups.

Permutations with Replacement

When the order of selection matters and items can be chosen more than once, we use the formula for permutations with replacement. If we have n choices and we make k selections with replacement, the number of permutations is n^k.

Example. Consider a 4-digit PIN where each digit can be any number from 0 to 9. The number of possible PINs is $10^4 = 10,000$.

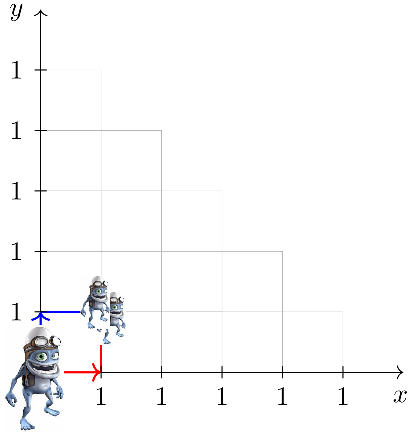
Binomial Coefficients and Pascal's Triangle

In this section, we'll explore the concept of binomial coefficients and their relationship with Pascal's Triangle. But before we dive into the theory, let's consider an engaging problem involving Crazy Frog!

Crazy Frog's Adventures on the Grid

Imagine Crazy Frog, the whimsical character known for his jumps and antics, navigating through a grid. In this grid, he starts at the origin $(0, 0)$ and can hop either one step right or one step up at each move.

The number of paths (ways) the frog can hop to each point on the grid below. For example, there are two paths to $(1, 1)$: either first hop 'right' and then 'up' or first hop 'up' followed by 'right' (the 'red' and 'blue' paths below).



Now, let's express each path as a monomial, where 'right' hops are represented by a and 'up' hops are represented by b. For instance, a 'two-hop' path can be represented as:

$$\begin{aligned} \rightarrow\rightarrow &\longleftrightarrow aa \longleftrightarrow a^2 \\ \rightarrow\uparrow &\longleftrightarrow ab \longleftrightarrow ab \\ \uparrow\rightarrow &\longleftrightarrow ba \longleftrightarrow ab \\ \uparrow\uparrow &\longleftrightarrow bb \longleftrightarrow b^2. \end{aligned}$$

This implies $(a + b)^2 = a^2 + 2ab + b^2$. The choice of direction in the hop corresponds to choosing either a or b in the corresponding $(a + b)$ factor.

Exercise. Using our findings, expand the following.

(a) $(a + b)^3$ (using '3-hop' paths).

(b) $(a + b)^4$ (using '4-hop' paths).

(c) $(a + b)^5$ (using '5-hop' paths).

Now, let's examine our findings. Notice that the number of paths to grid point $(k, n - k)$ coincides with the coefficient of $a^k b^{n-k}$ in the expansion of $(a + b)^n$. This observation leads us to an intriguing pattern known as Pascal's Triangle.

Pascal's Triangle is a triangular array of numbers in which each number is the sum of the two directly above it. The rows of the triangle correspond to the coefficients in the expansion of $(a + b)^n$.

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array}$$

The Pigeonhole Principle

The Pigeonhole Principle, also known as Dirichlet's Box Principle, is a fundamental concept in combinatorics. It is named after the idea that if there are more pigeons than pigeonholes, at least one pigeonhole must contain more than one pigeon.

Example. 1. Suppose we have a drawer containing 10 socks: 5 black socks and 5 white socks. If we randomly select socks from the drawer, how many socks must we pick to guarantee that we have a matching pair? In this scenario, there are two ”pigeonholes” representing the two colors of socks: black and white. We want to ensure that we have at least two socks of the same color, which corresponds to having a ”pigeon” in one of the pigeonholes.

So, to guarantee that we have a matching pair of socks, we need to pick socks until we have two socks of the same color. This means we need to pick at least 3 socks.

2. Show that it is impossible to place 9 rooks on an 8×8 grid without any two rooks threatening each other.

   According to the Pigeonhole Principle, if we have more pigeons (rooks) than pigeonholes (rows in the grid), then at least one pigeonhole must contain more than one pigeon. In this case, if we have 9 rooks and only 8 rows in the grid, there must be at least one row that accommodates at least two rooks.

   If we assume the contrary, that each row contains at most one rook, then we would have a total of 8 rooks distributed across 8 rows. However, this contradicts the fact that we have 9 rooks in total.

   Since each rook placed in the same row threatens each other due to their shared row, having at least two rooks in the same row violates the condition of placing the rooks without any two threatening each other.

   Therefore, it is impossible to place 9 rooks on an 8×8 grid without any two rooks threatening each other.

The generalized Pigeonhole Principle states: If n items are placed into m containers, and $n > km$ for some positive integer k, then at least one container must contain at least $k + 1$ items.

Example. 1. Find the minimum number of cards to guarantee three of them are of the same suit in a standard deck of 52 playing cards.

Here, the $n = 4$ suits are the pigeonholes, and $k + 1 = 3$, so $k = 2$. Among any $kn + 1 = 9$ cards (pigeons), three of them are guaranteed to be of the same suit.

2. We can employ the generalized Pigeonhole Principle to demonstrate that in a class of 33 students in MATH11, which includes freshmen, sophomores, juniors, and seniors, there are at least 9 students of the same academic year.

By setting $n = 4$ (the number of academic years) and $k + 1 = 9$, we can apply the principle to ensure that among $kn + 1 = 8 \cdot 4 + 1 = 33$ students, nine of them are guaranteed to belong to the same academic year.

Tree Diagrams

Tree diagrams are hierarchical diagrams that represent relationships between different entities or data points. They consist of nodes, which represent the entities, and edges, which represent the connections or relationships between them.

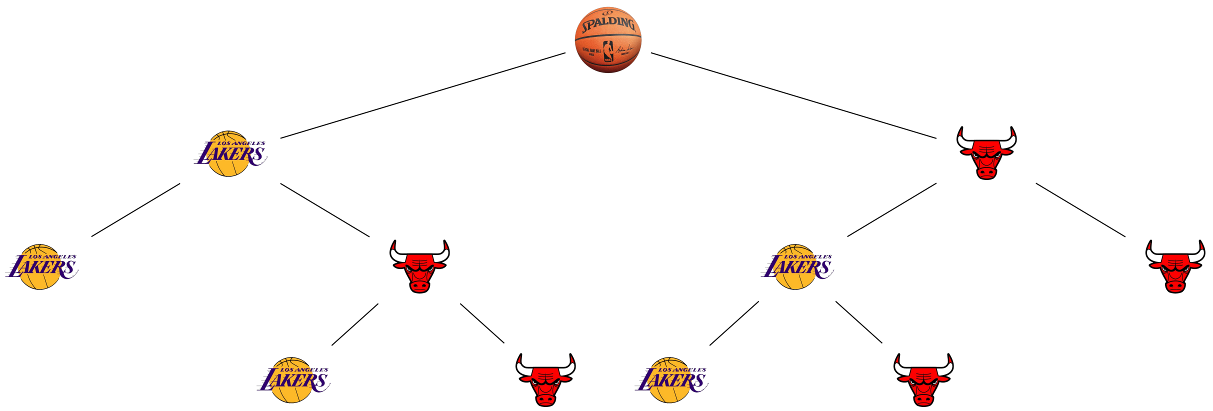
These diagrams are useful for encoding data because they provide a visual representation of complex relationships in a structured and organized manner. They help in understanding the hierarchical structure of data, identifying patterns, and visualizing dependencies between different elements.

Example.

1. A genealogical tree represents the familial relationships between individuals, such as parents, children, and siblings. It helps in tracing ancestry and understanding familial connections.
2. A file directory tree diagram illustrates the hierarchical structure of files and folders on a computer. It shows how files are organized into directories and subdirectories, making it easier to navigate and manage files.

Tree diagrams are often used in game theory for a pure analysis of possible outcomes and decision-making strategies. They provide a structured visual representation of the various potential sequences of events, allowing for a comprehensive evaluation of outcomes (and their associated payoffs).

Example. Two friends, Kobe and Michael, are playing one-on-one basketball games, with the first player to win two games declared the overall winner. The tree diagram below illustrates the possible outcomes, where a vertex labeled 🐂 indicates a game won by Michael, and a vertex labeled LOS ANGELES LAKERS signifies a game won by Kobe.



Lecture 12
Advanced Counting Principles

Mathematics often finds its way into popular culture, offering unique perspectives and inspiring stories. While exploring various movie genres, consider venturing into the intriguing realm of mathematics through movies.

Mathematics Meets the Silver Screen

Here are some recommendations with brief descriptions.

- **A Beautiful Mind.** A captivating journey into the life of the brilliant mathematician John Nash, exploring his contributions to game theory.

- **Good Will Hunting.** Follow the story of a self-taught mathematical prodigy working as a janitor at MIT, showcasing the power of raw talent.
- **Proof.** Dive into the complexities of mathematics and mental health as a daughter grapples with her father's legacy in the field.
- **21.** Experience the thrill of mathematics applied to the art of card counting in this exciting tale of blackjack and probability.
- **The Man Who Knew Infinity.** Explore the extraordinary life of the self-taught Indian mathematician Srinivasa Ramanujan, who made significant contributions to number theory.

The Stars and Bars Formula

The Stars and Bars Formula is a combinatorial technique used to solve problems involving the distribution of indistinguishable objects into distinguishable containers. It visualizes a configuration of n stars (representing the objects to be distributed) and $k-1$ bars (representing the separators), with the number of stars between two consecutive bars indicating the number of objects allocated to the corresponding container:

$$\star \quad \star \quad \star \mid \star \quad \star \quad \star \quad \star \quad \star \mid \star \quad \star$$

$$\boxed{\star\star\star} \qquad \boxed{\star\star\star\star\star} \qquad \boxed{\star\star}$$

The formula states that if we have n identical objects (stars) to distribute into k distinct containers, then the number of ways to do this, considering that each container can hold any number of objects (including zero), is given by $\binom{n+k-1}{k-1}$.

Example.

- We can use the stars and bars formula to find the number of non-negative integer solutions to the equation $x_1 + x_2 + x_3 = 10$.

  To solve this equation using the stars and bars formula: 1. Arrange 10 stars to represent the number 10. 2. Insert 2 bars to divide the stars into 3 groups, representing x_1, x_2, and x_3. The number of stars to the left of the first bar represents x_1, the stars between the first and second bars represent x_2, and the stars to the right of the second bar represent x_3.

  For instance, if we place the bars after the 3rd and 6th stars, we get the arrangement:

  $$\star\star\star|\star\star|\star\star\star\star\star,$$

  corresponding to the solution $10 = 3 + 2 + 5$.

  Any such arrangement of stars and bars corresponds to a unique solution to the equation $x_1+x_2+x_3 = 10$, where x_1, x_2, and x_3 are non-negative integers.

  Therefore, the number of non-negative integer solutions is equal to the number of ways we can arrange 10 stars and 2 bars, which is calculated to be

  $$\binom{10+3-1}{3-1} = \binom{12}{2} = \frac{12 \cdot 11}{2} = 66.$$

  So, there are 66 non-negative integer solutions to the equation $x_1 + x_2 + x_3 = 10$.

- In a project team of 4 members, there are 10 indistinguishable tasks to be assigned. Each member needs to take on at least one task. Using the stars and bars formula, we can determine the number of ways to distribute the tasks among the team members. Let's start by assigning one task to each team member. After that, we have 6 tasks remaining to be distributed among the team members.

  With 6 remaining tasks and 4 team members, the number of ways to distribute the tasks is:

$$\binom{6+4-1}{4-1} = \binom{9}{3} = \frac{9 \cdot 8 \cdot 7}{6} = 84.$$

So, there are 84 ways to distribute the tasks among the team members.

Multinomial Coefficients

As we discussed in the previous lecture, the coefficient of $a^k b^{n-k}$ in the expansion of $(a+b)^n$ is equal to $\binom{n}{k}$, the number of arrangements of k letters 'a' and $n-k$ letters $'b'$. Similarly, the coefficient of $a^{k_1} b^{k_2} c^{k_3}$ with $k_1 + k_2 + k_3 = n$ in the expansion of $(a+b+c)^n$ is equal to $\frac{n!}{k_1! k_2! k_3!}$, the *multinomial coefficient*. In general, the multinomial coefficient $\frac{n!}{k_1! k_2! \ldots k_\ell}$ is equal to the coefficient of the monomial $a_1^{k_1} a_2^{k_2} \ldots a_\ell^{k^\ell}$ in the expansion of $(a_1 + a_2 + \ldots a_\ell)^n$. It counts the number of ways to distribute n distinct objects into ℓ distinct categories, where each category receives k_i objects and $\sum\limits_{i=1}^{\ell} k_i = n$.

Example. Imagine a school election with 3 candidates, Alice, Bob, and Charlie, and 10 voters. Each voter indicates his most preferable candidate among the three. Such a voting procedure is called the *plurality rule*, and a collection of voting preferences is called a *voting profile.*

How many profiles result in Alice receiving 5 points, Bob 3, and Charlie 2?

We can build a bijection (one-to-one correspondence) between the voting profiles and the rearrangements of the string $'aaaaabbbcc'$. Each profile corresponds to a unique arrangement of the votes (letters $'a','b'$ and $'c'$), and vice versa. Therefore, the multinomial coefficient $\frac{10!}{5!3!2!}$ represents the number of different voting profiles that result in Alice receiving 5 points, Bob 3, and Charlie 2 in the school election.

If the order of the categories doesn't matter, we have the extra freedom of rearrangement of the categories with the same number of objects.

Example. Suppose we have 9 distinct books that we want to distribute into 3 distinct bookshelves, where each bookshelf has a certain capacity: 5 books, 2 books, and 2 books respectively. If the order of the bookshelves does not matter, how many ways can we distribute the books?

1. **Total Arrangements.**
   First, let's calculate the total number of arrangements if the order of the bookshelves matters. We can use the multinomial coefficient for this: $\dfrac{9!}{5!2!2!}$.

2. **Adjusting for Order.**
   Since the order of the bookshelves doesn't matter, we need to adjust our count. Therefore, we must account for this by dividing by the number of ways we can rearrange the shelves with coinciding capacity (of two books), which is $2! = 2$.

   Our final answer is:
   $$\text{Total Arrangements} = \frac{9!}{5!2!2! \cdot 2!}.$$

Applications of the Inclusion-Exclusion Principle

Recall that in the third lecture, we discussed the inclusion-exclusion principle. It is a formula that allows us to express the total number of elements in the union of n sets, in terms of cardinalities (numbers of elements) of all possible intersections. For n sets $A_1, A_2, \ldots, A_n$, the formula is given by:

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \ldots + (-1)^{n-1} |A_1 \cap A_2 \cap \ldots \cap A_n|.$$

In this section, we will apply this formula in order to count two important numbers: the number of onto maps from a finite set A to a finite set B and the number of derangements.

Number of Onto Functions

Let A and B be two finite sets, such that $|A| = n$ and $|B| = 3$. We want to find the number of onto (surjective) functions from A to B.

Let b_1, b_2, b_3 be the elements in B. Let $\mathcal{F}$ be the set of all functions from A into B. Furthermore, let $\mathcal{F}_1 \subset \mathcal{F}$ be the subset of functions which do not map any element of A into b_1, i.e., b_1 is not in the range of any function in $\mathcal{F}_1$. Similarly, let $\mathcal{F}_2$ and $\mathcal{F}_3$ be the subsets of functions which do not send any element of A into b_2 and b_3, respectively. Then the set $\mathcal{F}_i^C$, complement of the set $\mathcal{F}_i$, consists of functions whose range contains b_i. Therefore, the onto functions from A to B comprise the set $\mathcal{F}_1^C \cap \mathcal{F}_2^C \cap \mathcal{F}_3^C$, and the number of such functions is equal to the cardinality of this set.

The inclusion-exclusion principle for three sets A, B, and C reads

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

Plugging in $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$ for A, B, and C, we get

$$|\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3| = |\mathcal{F}_1| + |\mathcal{F}_2| + |\mathcal{F}_3| - |\mathcal{F}_1 \cap \mathcal{F}_2| - |\mathcal{F}_1 \cap \mathcal{F}_3| - |\mathcal{F}_2 \cap \mathcal{F}_3| + |\mathcal{F}_1 \cap \mathcal{F}_2 \cap \mathcal{F}_3|.$$

Next, we compute the required cardinalities. As $\mathcal{F}_1$ consists of maps $A \to \{b_2, b_3\}$, we have $|\mathcal{F}_1| = 2^n$ (every element of A can be sent to either b_2 or b_3). Similarly, we establish $|\mathcal{F}_2| = |\mathcal{F}_3| = 2^n$.

Now the set $\mathcal{F}_1 \cap \mathcal{F}_2$ consists of functions which do not send any element of A into b_1 or b_2. Said differently, the image of every element $a \in A$ is equal to b_3 (as $B \setminus \{b_1, b_2\} = \{b_3\}$), and the set $\mathcal{F}_1 \cap \mathcal{F}_2$ consists of just one (constant) function: $f(a) = b_3$ for all $a \in A$. Similarly, $\mathcal{F}_1 \cap \mathcal{F}_3$ has only the function $g(a) = b_2$ for all $a \in A$, and $\mathcal{F}_2 \cap \mathcal{F}_3$ has only the function $h(a) = b_1$ for all $a \in A$. We conclude that $|\mathcal{F}_1 \cap \mathcal{F}_2| = |\mathcal{F}_1 \cap \mathcal{F}_3| = |\mathcal{F}_2 \cap \mathcal{F}_3| = 1$. Finally, the set $\mathcal{F}_1 \cap \mathcal{F}_2 \cap \mathcal{F}_3$ is empty and has cardinality zero.

We get

$$|\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3| = |\mathcal{F}_1| + |\mathcal{F}_2| + |\mathcal{F}_3| - |\mathcal{F}_1 \cap \mathcal{F}_2| - |\mathcal{F}_1 \cap \mathcal{F}_3| - |\mathcal{F}_2 \cap \mathcal{F}_3| + |\mathcal{F}_1 \cap \mathcal{F}_2 \cap \mathcal{F}_3| = 3 \cdot 2^n - 3 \cdot 1 + 0 = 3 \cdot 2^n - 3.$$

Applying De Morgan's law, we obtain $\mathcal{F}_1^C \cap \mathcal{F}_2^C \cap \mathcal{F}_3^C = (\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)^C$. Hence,

$$|\mathcal{F}_1^C \cap \mathcal{F}_2^C \cap \mathcal{F}_3^C| = |\mathcal{F}| - |\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3| = 3^n - (3 \cdot 2^n - 3) = 3^n - 3 \cdot 2^n + 3$$

is the number of onto functions from A to B.

Number of Derangements

Imagine a beleaguered postman, tasked with delivering four letters to four different addresses. Each letter is meticulously addressed to a specific recipient and placed inside an envelope labeled E_1, E_2, E_3, and E_4. However, fate takes a mischievous turn when the postman, burdened with his parcel, stumbles and drops all four envelopes. Remarkably, none of the letters end up in their intended envelopes. How many ways could this chaotic mishap have unfolded?

Let's unravel the mystery step by step.

1. Initially, there are $4! = 24$ possible ways for the postman to distribute the letters, each representing a permutation of the envelopes.

2. However, we are interested in the scenarios where none of the letters ended up in their original envelopes. We can use the principle of inclusion-exclusion to count the number of such situations. Let D_i denote the event that the letter intended for envelope E_i ends up in E_i. We aim to find $|D_1^C \cap D_2^C \cap D_3^C \cap D_4^C|$, where D_i^C represents the complement of event D_i.

3. To compute the cardinality of each individual event and their intersections, we use permutations:

   - $|D_1|$ represents the number of permutations where the letter intended for E_1 ends up back in E_1. This is 3!, as once the letter for E_1 is fixed, there are 3! ways to distribute the remaining letters. Similarly, $|D_2| = |D_3| = |D_4| = 3!$.

   - $|D_1 \cap D_2|$ represents the number of permutations where the letters intended for both E_1 and E_2 end up back in their respective envelopes. This is 2!, as once the letters for E_1 and E_2 are fixed, there are 2! ways to distribute the remaining letters. Similarly, $|D_1 \cap D_3| = |D_1 \cap D_4| = |D_2 \cap D_3| = |D_2 \cap D_4| = |D_3 \cap D_4| = 2!$.

   - $|D_1 \cap D_2 \cap D_3|$ represents the number of permutations where the letters intended for E_1, E_2, and E_3 all end up back in their respective envelopes. This is $1! = 1$. Similarly, $|D_1 \cap D_2 \cap D_4| = |D_1 \cap D_3 \cap D_4| = |D_2 \cap D_3 \cap D_4| = 1$.

   - Since there is exactly one way for all letters to end up back in their original envelopes simultaneously, $|D_1 \cap D_2 \cap D_3 \cap D_4| = 1$.

   Using the principle of inclusion-exclusion, we find:

   $$|D_1^C \cap D_2^C \cap D_3^C \cap D_4^C| = 4! - |D_1 \cup D_2 \cup D_3 \cup D_4| = 4! - \binom{4}{1} \cdot 3! + \binom{4}{2} \cdot 2! - \binom{4}{3} \cdot 1! + \binom{4}{4} \cdot 0! =$$

   $$24 - 4 \cdot 6 + 6 \cdot 2 - 4 \cdot 1 + 1 = 24 - 24 + 12 - 4 + 1 = 9.$$

   So, there are 9 possible derangements.

A *derangement* of a set of n elements is a permutation in which no element appears in its original position. The number of derangements of n elements, denoted by $!n$, can be calculated using the formula:

$$n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! - \cdots + (-1)^n = n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right).$$

Remark. The expression $\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right)$, as $n \to \infty$, converges to $\frac{1}{e}$. In the context of derangements, this implies that the probability that no element returns to its original position approaches $\frac{1}{e}$ as $n \to \infty$.

Lecture 13

Generating Functions and Linear Recurrence Relations

In this lecture, we will explore the powerful tools of generating functions and their applications in solving linear recurrence relations, enabling us to find closed-form expressions for sequences defined recursively.

Generating Functions

Example. Consider the geometric progression $S(q) = 1 + q + q^2 + \ldots$. To find the sum of this series, we can exploit generating functions.

We write $qS(q)$, shifting the series by 1 unit, and notice that $S(q) - qS(q) = 1$. Solving this equation gives us $S(q) = \frac{1}{1-q}$.

In general, it is often helpful to encode sequence values into power series using a dummy variable. Let's see how we can encode a sequence into a generating function with some examples.

Encoding Sequence into Generating Function

To encode a sequence into a generating function, we represent each term of the sequence as a coefficient of the corresponding power of the dummy variable.

For example, consider the sequence $a_n = 2^n$. Its generating function is given by $A(q) = \sum_{n=0}^{\infty} 2^n q^n = \frac{1}{1-2q}$.

Another example is the sequence of Fibonacci numbers, F_n. Its generating function is $F(q) = \sum_{n=0}^{\infty} F_n q^n$, where $F_0 = 0$, $F_1 = 1$, and $F_{n+2} = F_{n+1} + F_n$.

Linear Homogeneous Recurrence Relation

Definition. A **linear recurrence relation** is an equation that expresses each term of a sequence as a linear combination of previous terms, with constant coefficients:

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k} + f(n)$$

where $c_1, c_2, \ldots, c_k \neq 0$ are constants and $f : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ is a function. The number k is called the **order** of the recurrence relation. If the function f is identically zero, the relation is called **homogeneous**.

This recurrence relation gives rise to an equation involving the generating function $A(t) = \sum\limits_{n\geq 0} a_n t^n$ encoding the sequence $\{a_n\}$. By multiplying both sides of the recurrence relation by t, we "align" the terms involving a_{n-1} and a_n as coefficients of t^n. Similarly, by multiplying both sides by t^2, we align the terms involving a_{n-2} and a_n as coefficients of t^n, etc.:

$$\begin{array}{rlll} A(t) = & a_0 + a_1 t & +a_2 t^2 + \ldots & +a_n t^n + \ldots \\ tA(t) = & 0 + a_0 t & +a_1 t^2 + \ldots & +a_{n-1} t^n + \ldots \\ t^2 A(t) = & 0 + 0 & +a_0 t^2 + \ldots & +a_{n-2} t^n + \ldots \end{array}$$

This way, we obtain the equation:

$$A(t) = a_0 + (a_1 - c_1 a_0)t + (a_2 - c_1 a_1 - c_2 a_0)t^2 + \ldots + (a_{k-1} - c_1 a_{k-2} - c_2 a_{k-3} - \ldots - c_{k-1} a_0)t^{k-1} +$$
$$c_1 t A(t) + c_2 t^2 A(t) + \cdots + c_k t^k A(t) \Leftrightarrow$$
$$A(t) = \frac{a_0 + (a_1 - c_1 a_0)t + (a_2 - a_1 c_1 - a_0 c_0)t^2 + \ldots + (a_k - c_1 a_{k-1} - c_2 a_{k-2} - \ldots - c_k a_0)t^{k-1}}{1 - c_1 t - c_2 t^2 - \ldots - c_k t^k}$$

from the recurrence relation $a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k}$.

Example. For the Fibonacci sequence $F_n = F_{n-1} + F_{n-2}$, where $F_0 = 1$ and $F_1 = 1$, we have

$$F(t) = F_0 + (F_1 - F_0)t + tF(t) + t^2 F(t) = 1 + tF(t) + t^2 F(t)$$

Simplifying this equation, we get:

$$F(t) = \frac{1}{1 - t - t^2}$$

Homogeneous Linear Recurrence Relations of Orders 1 and 2

We will discuss linear recurrence relations of small orders. A general linear recurrence relation of order 1 takes the form:

$$a_n = c a_{n-1} + f(n)$$

A homogeneous recurrence relation of order 1 simplifies to:

$$a_n = c a_{n-1}$$

This represents a geometric progression with ratio c. Every term can be expressed in terms of the initial term as $a_n = c^n a_0$.

Let $A(t) = \sum\limits_{n\geq 0} a_n t^n = \sum\limits_{n\geq 0} c^n a_0 t^n$ be the corresponding generating function. It satisfies the equation:

$$A(t) - ctA(t) = a_0$$

Therefore, we obtain:

$$A(t) = \frac{a_0}{1 - ct}.$$

Remark. Let's denote the polynomial in the denominator by $f(t) := 1 - ct$. Then the polynomial $\chi(t) := tf(1/t) = t(1 - c \cdot \frac{1}{t}) = t - c$ has the root $t = c$ and $a_n = c^n a_0$.

A homogeneous linear recurrence relation of order 2 takes the form:

$$a_n = \beta a_{n-1} + \theta a_{n-2}$$

The corresponding generating function $A(t) = \sum\limits_{n\geq 0} a_n t^n$ satisfies the equation:

$$A(t) - \beta t A(t) - \theta t^2 A(t) = a_0 + (a_1 - \beta a_0)t$$

From this, we get the expression:

$$A(t) = \frac{a_0 + (a_1 - \beta a_0)t}{1 - \beta t - \theta t^2}.$$

Definition. Similarly to the polynomial in the remark above, we define the *characteristic polynomial of the recurrence relation* as

$$\chi(t) := t^2 f(1/t) = t^2 f(1/t) = t^2(1 - \beta \cdot \frac{1}{t} - \theta \cdot \frac{1}{t^2}) = t^2 - \beta t - \theta.$$

From the discussion above with a little bit of algebra, one can prove the following result.

Proposition.

1. Suppose the characteristic polynomial $\chi(t) = t^2 - \beta t - \theta$ of the recurrence relation $a_n = \beta a_{n-1} + \theta a_{n-2}$ has distinct roots t_1 and t_2. Then the general solution of the recurrence relation is given by

$$a_n = c_1 t_1^n + c_2 t_2^n \quad \text{for} \quad n \geq 2.$$

Here, the constants c_1 and c_2 are uniquely determined from the *initial conditions*, which are the values of a_0 and a_1.

2. If the characteristic polynomial $\chi(t)$ has a repeated root t_1, then the general solution of the recurrence relation is given by

$$a_n = c_1 t_1^n + c_2 n t_1^n \quad \text{for} \quad n \geq 2.$$

Example.

1. Consider the sequence defined recursively by $a_0 = 1$, $a_1 = 2$, and $a_n = 5a_{n-1} - 4a_{n-2}$ for $n \geq 2$. Here, $\beta = 5$ and $\theta = -4$, so the characteristic polynomial for this sequence is given by

$$\chi(t) = t^2 - \beta t - \theta = t^2 - 5t + 4.$$

The roots of this characteristic polynomial are $t_1 = 4$ and $t_2 = 1$. According to the proposition for solving linear homogeneous recurrence relations, the general term of the sequence can be expressed as:

$$a_n = c_1 \cdot 4^n + c_2 \cdot 1^n = c_1 \cdot 4^n + c_2$$

To determine the constants c_1 and c_2, we use the initial conditions $a_0 = 1$ and $a_1 = 2$:

$$\begin{cases} a_0 = c_1 \cdot 4^0 + c_2 = c_1 + c_2 = 1 \\ a_1 = c_1 \cdot 4^1 + c_2 = 4c_1 + c_2 = 2. \end{cases}$$

Solving this system of equations, we get:

$$\begin{cases} c_1 + c_2 = 1 \\ 4c_1 + c_2 = 2. \end{cases}$$

Subtracting the first equation from the second:

$$(4c_1 + c_2) - (c_1 + c_2) = 2 - 1 = 1 \Leftrightarrow 3c_1 = 1 \Leftrightarrow c_1 = \frac{1}{3}.$$

Substituting $c_1 = \frac{1}{3}$ into $c_1 + c_2 = 1$ gives

$$\frac{1}{3} + c_2 = 1 \Leftrightarrow c_2 = 1 - \frac{1}{3} = \frac{2}{3}.$$

Therefore, the general term of the sequence is

$$a_n = \frac{4^n + 2}{3}.$$

For instance, we can verify this formula for a_2 and a_3:

$$a_2 = \frac{4^2 + 2}{3} = \frac{16 + 2}{3} = 6,$$
$$a_3 = \frac{4^3 + 2}{3} = \frac{64 + 2}{3} = 22.$$

We can also confirm these values directly from the definition of the sequence:

$$a_2 = 5a_1 - 4a_0 = 5 \cdot 2 - 4 \cdot 1 = 10 - 4 = 6,$$
$$a_3 = 5a_2 - 4a_1 = 5 \cdot 6 - 4 \cdot 2 = 30 - 8 = 22.$$

2. For the Fibonacci sequence $F_n = F_{n-1} + F_{n-2}$ with initial conditions $F_0 = 0$ and $F_1 = 1$, the characteristic polynomial is $\chi(t) = t^2 - t - 1$. This polynomial has distinct roots $t_1 = \frac{1-\sqrt{5}}{2}$ and $t_2 = \frac{1+\sqrt{5}}{2}$. Therefore, the general solution to the recurrence relation can be written as

$$F_n = c_1 \left(\frac{1 - \sqrt{5}}{2} \right)^n + c_2 \left(\frac{1 + \sqrt{5}}{2} \right)^n$$

To determine the constants c_1 and c_2, we use the initial conditions:

$$\begin{cases} F_0 = 0 & \Rightarrow c_1 + c_2 = 0 \\ F_1 = 1 & \Rightarrow c_1 \left(\frac{1-\sqrt{5}}{2} \right) + c_2 \left(\frac{1+\sqrt{5}}{2} \right) = 1 \end{cases}$$

From the first equation, $c_2 = -c_1$. Substituting this into the second equation gives:

$$c_1 \left(\frac{1 - \sqrt{5}}{2} \right) - c_1 \left(\frac{1 + \sqrt{5}}{2} \right) = 1 \Leftrightarrow c_1 \left(\frac{-2\sqrt{5}}{2} \right) = 1 \Leftrightarrow c_1 = -\frac{1}{\sqrt{5}}$$

Thus, $c_2 = \frac{1}{\sqrt{5}}$. Therefore, the general term of the Fibonacci sequence is

$$F_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{2} \right)^n - \left(\frac{1 - \sqrt{5}}{2} \right)^n \right)$$

For instance, we can verify this formula for F_2:

$$F_2 = \frac{1}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{2} \right)^2 - \left(\frac{1 - \sqrt{5}}{2} \right)^2 \right) = \frac{1}{\sqrt{5}} \left(\frac{1 + 2\sqrt{5} + 5}{4} - \frac{1 - 2\sqrt{5} + 5}{4} \right) =$$
$$\frac{1}{\sqrt{5}} \left(\frac{6 + 2\sqrt{5}}{4} - \frac{6 - 2\sqrt{5}}{4} \right) = \frac{1}{\sqrt{5}} \left(\frac{4\sqrt{5}}{4} \right) = \frac{1}{\sqrt{5}} \cdot \sqrt{5} = 1.$$

Notice that $\frac{1-\sqrt{5}}{2} \approx -0.618$, which means the term $\left(\frac{1-\sqrt{5}}{2}\right)^n$ becomes very small for large n. Therefore, since F_n is an integer by definition, we have:

$$F_n = \left\lceil \frac{1}{\sqrt{5}} \cdot \left(\frac{1+\sqrt{5}}{2}\right)^n \right\rceil$$

This shows that the Fibonacci numbers grow exponentially with the golden ratio $\frac{1+\sqrt{5}}{2}$.

3. Consider the sequence defined recursively by $a_0 = 1$, $a_1 = 4$, and $a_{n+2} = 6a_{n+1} - 9a_n$. The characteristic polynomial for this sequence is given by

$$\chi(t) = t^2 - 6t + 9$$

and has a repeated root $t_1 = 3$. Therefore, the second part of the proposition implies that the general term of the sequence can be expressed as:

$$a_n = c_1 \cdot 3^n + c_2 \cdot n \cdot 3^n$$

To determine the constants c_1 and c_2, we use the initial values:

$$\begin{cases} a_0 = 1 & \Rightarrow c_1 = 1 \\ a_1 = 4 & \Rightarrow 3c_1 + 3c_2 = 4. \end{cases}$$

Substituting $c_1 = 1$ into the second equation:

$$3 \cdot 1 + 3c_2 = 4 \Leftrightarrow 3 + 3c_2 = 4 \Leftrightarrow 3c_2 = 1 \Leftrightarrow c_2 = \frac{1}{3}.$$

Therefore, the general term of the sequence is

$$a_n = 3^n + \frac{1}{3} \cdot n \cdot 3^n = 3^n \left(1 + \frac{n}{3}\right) = 3^{n-1}(3+n).$$

For instance, we can verify this formula for a_2 and a_3:

$$\begin{aligned} a_2 &= 3^{2-1}(3+2) = 3 \cdot 5 = 15, \\ a_3 &= 3^{3-1}(3+3) = 3^2 \cdot 6 = 54. \end{aligned}$$

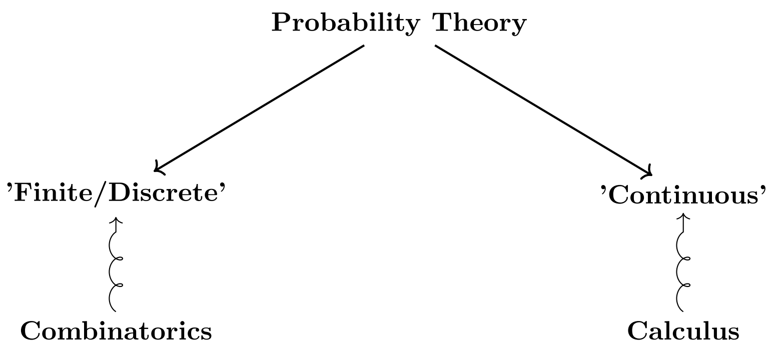
We can also confirm these values directly from the definition of the sequence:

$$\begin{aligned} a_2 &= 6a_1 - 9a_0 = 6 \cdot 4 - 9 \cdot 1 = 24 - 9 = 15, \\ a_3 &= 6a_2 - 9a_1 = 6 \cdot 15 - 9 \cdot 4 = 90 - 36 = 54. \end{aligned}$$

Lecture 14
Introduction to Probability Theory

Today, we are setting sail into the intriguing realm of probability, a branch of mathematics that helps us navigate uncertainty and randomness.

Probability theory can be broadly categorized into two main branches: 'Finite/Discrete' and 'Continuous'. These branches are distinguished by the mathematical tools they employ, with combinatorics playing a key role in the former and calculus in the latter:



Finite/Discrete Probability

In finite or discrete probability, we deal with a finite number of distinct outcomes. The main tool here is combinatorics, which involves counting principles.

Example: Flipping a Coin

Consider the simple act of flipping a coin. If we were to ask, 'What is the probability of getting heads? without any further information, we would find ourselves in a bit of a conundrum. Why? Because without knowing how likely it is for the coin to land on heads versus tails, we can't assign a meaningful probability.

To make sense of this, think about the scenario in which the coin lands near a wall, wobbles, but never fully flips onto a side. In this case, it wouldn't be accurate to say that the outcome is either heads or tails; rather, it's undetermined.

Now, let's make some additional assumptions. Suppose we have a fair coin, one that's perfectly balanced, and we flip it once. In this case, we have two possible outcomes: heads or tails. Since the coin is fair, we assume that the likelihood of getting heads is the same as getting tails, each with a probability of 0.5 or 50%.

What if we decide to flip the coin twice? Now, the possible outcomes expand. We could get heads on both flips, tails on both flips, or a combination of heads and tails. There are four possible outcomes: {(H, H), (H, T), (T, H), (T, T)}. Assuming the coin flips are independent (the outcome of one flip doesn't affect the outcome of the next), each of these outcomes has a probability of 0.25 or 25%.

This illustrates how probabilities come into play in simple situations like coin flips. By making appropriate assumptions and considering the possible outcomes, we can assign meaningful probabilities to events.

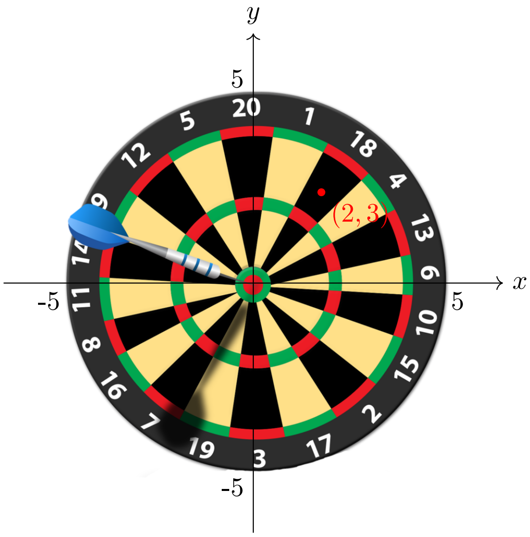
Continuous Probability

In continuous probability, we deal with outcomes from a continuous set, such as real numbers within a certain range. The main tool here is calculus.

Example: Throwing a Dart

When we throw a dart at a circular target, the set of possible outcomes encompasses all points within the circle (assuming we never miss the target!).

Question. What is the probability of hitting the point $(2, 3)$?



This question, as stated, doesn't quite make sense yet. To assign probabilities, we need a distribution, which essentially tells us the likelihood of hitting different points. For simplicity, we can assume a 'uniform' distribution. In pedestrian terms, this means that each point within the circle has an equal chance of being hit.

Here's a fascinating paradox to consider: when we throw a dart, we do actually hit some point on the board. However, under this uniform distribution assumption, the probability of hitting any specific point, like $(2, 3)$, is zero. This is because there are infinitely many points in the circle, and the 'target' at each specific point is effectively a single point with zero area.

This concept aligns with the intuition that hitting an exact point on the target, out of countless possible points, is an extremely rare event. It's a paradox that challenges our everyday understanding of probabilities!

Interestingly, the probability of hitting a region, like the bullseye, can be calculated using calculus. It's simply the ratio of the area of the bullseye to the total area of the target.

Sample Spaces and Events

To begin, let's provide a more precise definition of the concept of a sample space, building upon our familiarity with set theory.

Definition. A **sample space** is the set of all possible outcomes of an experiment.

Consider the following examples:

Rolling a six-sided die
Sample Space: $\{1, 2, 3, 4, 5, 6\}$

Drawing Cards from a Deck
Sample Space: {Ace of Spades, Two of Hearts, Jack of Diamonds, ...}

Weather Forecasting
Sample Space: {Sunny, Cloudy, Rainy, Snowy}

Now, onto events.

Definition. An **event** is a subset of the sample space.

Let's take a look at some events related to our examples.

- For the Die Roll.
  - Event A: Rolling an even number = {2, 4, 6}

  - Event B: Rolling a number greater than 4 = {5, 6}

- For Drawing Cards.
  - Event C: Drawing a red card = $\{\heartsuit, \diamondsuit\}$
  - Event D: Drawing a face card = {Jack, Queen, King}

  Here's a bit more detail:

  - **Hearts:** This suit is symbolized by red heart shapes, $\heartsuit$. It consists of thirteen cards, ranging from Ace to King. These include numbered cards (2 through 10), face cards (Jack, Queen, King), and the Ace.
  - **Diamonds:** Similarly, this suit is represented by red diamond shapes, $\diamondsuit$. It also contains thirteen cards, comprising the same range as Hearts (Ace to King).

  So, when we talk about the event 'Drawing a red card', we're essentially referring to the combined set of Hearts and Diamonds. In other words, if you draw a card and it's either a Heart or a Diamond, it falls under the event 'Drawing a red card'.

- For Weather Forecasting:
  - Event E: Predicting a rainy day = {Rainy}
  - Event F: Expecting a non-sunny day = {Cloudy, Rainy, Snowy}

Probabilities as Functions

Now, probabilities come into play. These are functions that assign a likelihood to each event. They provide us with a numerical representation of how probable an outcome is. This is a crucial tool that aids us in making informed decisions, whether it's in games of chance, weather forecasting, or many other real-world applications.

So, as we venture further into probability theory, keep in mind that it's about understanding all potential outcomes and assigning values that signify their likelihood. This is the key to unlocking the power of probability.

We start with an exploration of the key characteristics that define probability functions.

Definition. A **probability function** on a sample space Ω is a function $P : \mathcal{P}(\Omega) \to [0, 1]$ that satisfies the following conditions:

1. **Non-Negativity**: $P(A) \geq 0$ for any event $A \subseteq \Omega$. This is because probabilities cannot be negative. We wouldn't want any *negative vibes* in our probability space!

2. **Normalization**: $P(\Omega) = 1$. This ensures that the total probability of all possible outcomes is 1, reflecting certainty. We're closing all doors to uncertainty!

3. **Additivity**: For any pair of disjoint sets A and B, $P(A \cup B) = P(A) + P(B)$. This is an extension of the idea that the probability of either of two mutually exclusive events occurring is the sum of their individual probabilities. We're making sure our probabilities *add up* nicely!

Let's explore some examples of probability functions.

Fair Coin Toss

Consider a fair coin toss. In this simple experiment, the sample space Ω consists of two possible outcomes: heads (H) or tails (T). Mathematically, we represent this sample space as $\Omega = \{H, T\}$, where H represents heads and T represents tails.

We define $P(A)$ for any subset $A \subseteq \Omega$ as the ratio of the number of outcomes in A to the total number of outcomes in Ω. Mathematically, it is given by:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{2}$$

In particular, $P(\{H\}) = P(\{T\}) = \frac{1}{2}$.
This probability function satisfies all three conditions we discussed earlier.

1. **Non-Negativity**: Since both $|A|$ is a non-negative integer, $P(A)$ is always non-negative.

2. **Normalization**: If we consider the entire sample space, i.e., $A = \Omega$, we have:

   $$P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$$

   This means that the probability of getting either heads or tails in a fair coin toss is 1.

3. **Additivity**: Since there are only two possible outcomes (heads or tails) in the sample space, any two events are necessarily disjoint. Therefore, the additivity condition holds trivially.

In summary, the probability function defined for a fair coin toss is indeed fair, as it meets all the required conditions. It accurately reflects the inherent fairness of the experiment, where each outcome is equally likely.

So, we can confidently say that this probability function is as *fair* as a coin toss can get!

Biased Die

Consider a biased six-sided die with probabilities $P(\{i\}) = \frac{i}{21}$ for $i = 1, 2, \ldots, 6$. This means that the probability of rolling a 1 is $\frac{1}{21}$, of rolling a 2 is $\frac{2}{21}$, and so on, up to the probability of rolling a 6 which is $\frac{6}{21}$.

1. **Non-Negativity**: The probabilities assigned are all positive fractions, ensuring non-negativity.

2. **Normalization**: Let's calculate the probability for the entire sample space:

   $$P(\Omega) = P(\{1\})+P(\{2\})+P(\{3\})+P(\{4\})+P(\{5\})+P(\{6\}) = \frac{1}{21}+\frac{2}{21}+\frac{3}{21}+\frac{4}{21}+\frac{5}{21}+\frac{6}{21} = \frac{21}{21} = 1$$

   This demonstrates that the probabilities sum up to 1, confirming the normalization condition.

3. **Additivity**: This property holds because the function is defined on one-element subsets (single outcomes), and extended by the additivity property to unions of events. For instance, $P(\{1\} \cup \{2\} \cup \{3\}) = P(\{1\}) + P(\{2\}) + P(\{3\}) = \frac{1}{21} + \frac{2}{21} + \frac{3}{21} = \frac{6}{21}$.

While this probability function results in a die that is biased towards higher values, this does not invalidate it as a probability function. It simply means that the die is not fair, and the outcomes are more likely to be on the higher end.

In summary, the function $P(\{i\}) = \frac{i}{21}$ for $i = 1, 2, \ldots, 6$ is indeed a valid probability function, even though it represents a biased die.

The extensive explanation here was for clarity You won't need to provide such detailed answers on upcoming assignments.

As we briefly touched upon earlier, understanding counting principles is paramount when it comes to tackling probability on finite sample spaces. These principles provide the fundamental tools needed to navigate through various probability scenarios. Let's apply these concepts to a range of examples.

Example. 1. In an urn, there are 4 red balls, 3 green balls, and 2 blue balls. If I draw 3 balls at random without replacement, what is the probability of getting exactly 2 red balls?

The total number of ways to draw 3 balls out of 9 is $C(9,3) = \dfrac{9!}{3! \cdot 6!} = 84$. The number of ways to get exactly 2 red balls and 1 non-red ball is $C(4,2) \cdot C(5,1) = 30$. Therefore, the probability is $\dfrac{30}{84} = \dfrac{5}{14}$.

2. If you draw 3 cards from a standard deck of 52 cards, what is the probability that all 3 are hearts?

There are $C(13,3)$ ways to choose 3 hearts out of the 13 hearts in the deck. The total number of ways to draw 3 cards from 52 is $C(52,3)$. Therefore, the probability is $\dfrac{C(13,3)}{C(52,3)}$.

3. If you roll two fair six-sided dice, what is the probability of getting a sum of 7?

In the table below, n_1 represents the number on the first die, and n_2 represents the number on the second die.

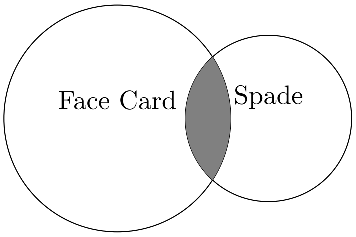
n_1 / n_2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The highlighted red cells represent combinations where the sum is 7.

There are 6 ways to get a sum of 7 $(1+6,\ 2+5,\ 3+4,\ 4+3,\ 5+2,\ 6+1)$ out of a total of 36 possible outcomes (since each die has 6 sides). Therefore, the probability is $\dfrac{6}{36} = \dfrac{1}{6}$.

Conditional Probability

Consider a standard deck of 52 playing cards. You draw one card at random, but before revealing it, you're informed that the card is a face card (jack, queen, or king). With this knowledge in hand, what is the probability that the card is a spade?



Face Card ∩ Spade

Intuitively, knowing that you have a face card should increase the likelihood of having a spade. This is where conditional probability comes into play.

Definition. The **conditional probability** of event A given that event B has already occurred, denoted as $P(A|B)$, is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ where } P(B) \neq 0.$$

Example. The 'Four of a Kind' is a rare and formidable hand in poker, featuring four cards of the same rank.

1. **Four of a Kind (Four Aces).**

   To find the probability of getting four Aces in a five-card hand, we first calculate the number of ways to choose four Aces from the deck (which is $\binom{4}{4} = 1$) and the number of ways to choose the fifth card (which is $\binom{48}{1} = 48$). The total number of five-card hands is $\binom{52}{5} = 2,598,960$. Therefore, the probability is

$$P(\text{Four Aces}) = \frac{\binom{4}{4} \cdot \binom{48}{1}}{\binom{52}{5}} \approx 0.000018 \approx 0.0018\%.$$

2. **Conditional Probability (Given One Ace).**

   If we already know that one of the cards is an Ace, there are $\binom{3}{3} \cdot \binom{48}{1}$ ways to choose the remaining three cards from the three remaining Aces and the other 48 cards. The total number of four-card hands with one Ace is $\binom{51}{4} = 249,900$. Therefore, the conditional probability is

$$P(\text{Four Aces} \,|\, \text{One Ace}) = \frac{\binom{3}{3} \cdot \binom{48}{1}}{\binom{51}{4}} \approx 0.00019 \approx 0.019\%.$$

3. **Conditional Probability (Given Two Aces).**

   If we know that both cards in hand are Aces, the conditional probability is

$$P(\text{Four Aces} \,|\, \text{Two Aces}) = \frac{\binom{2}{2} \cdot \binom{48}{1}}{\binom{50}{3}} \approx 0.0024 \approx 0.24\%.$$

Now, let's address a crucial question that often arises when dealing with conditional probabilities.

Question. Let A and B be two events. Why the probabilities $P(A|B)$ (probability of A given B) and $P(B|A)$ (probability of B given A) differ even though they both involve conditional probabilities??

A very important observation. The formulas for computing conditional probabilities are similar, but **NOT** the same:

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

while

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

The numerators in both formulas coincide and are equal to the probability of occurrence of both events A and B. The difference between the two formulas is that the former formula 'measures the proportion of the intersection relative to B', while the latter 'measures the proportion of the intersection relative to A'.

Here is an example, where the difference between the two probabilities is strikingly evident.

Example. Consider two events:

- A: a person is a natural born citizen of the United States;
- B: a person is the president of the United States.

Remark. Recall that 'No Person except a natural born Citizen, or a Citizen of the United States, at the time of the Adoption of the Constitution, shall be eligible to the Office of President', so $B \subset A$ is a subset.

Then $P(B|A)$ stands for the probability that a person who is a natural born citizen of the United States, is the president of the country, while $P(A|B)$ is the probability that someone who is a president of the United States is a natural born citizen of the United States. Notice that $P(B|A) < \dfrac{1}{10^7}$ is incredibly small, while $P(A|B) = 1$ (see the remark above).

Notice that the formulas above give rise to two expressions for the probability of intersection of two events:

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

Let's take a look at some more examples.

Example. 1. There is an amazing basketball team ABT and the best player on that team is the player JM.[1] Consider two events:

- A: team ABT wins the game;
- B: player JM participates in the game.

Then $P(B|A)$ stands for the probability that JM takes part in the game that ABT team wins, while $P(A|B)$ is the probability that ABT team wins the game, when JM plays.

2. Consider two events:

- A: A child loves snickers bars ;
- B: A child's parent buys the child a snickers bar .

Now $P(A|B)$ stands for the probability that a child loves snickers bars given his (her) parent bought him (her) one, while $P(B|A)$ is the probability that a child's parent bought the child a snickers given that he (she) loves those.

3. Consider two events:

- A: A person has disease $\mathfrak{h}$;
- B: The result of diagnostic test for disease $\mathfrak{h}$ is correct.

This time $P(B|A)$ stands for the probability that a person, who has the disease $\mathfrak{h}$, tested positive (the test detected the disease correctly), while $P(A|B)$ stands for the probability that the outcome of the test was correct, and the person, who took the test, has the disease.

Interpretation: In this context, it's crucial to understand the implications of false positives and false negatives. A false positive occurs when the test indicates a disease is present, but it is not. This can lead to unnecessary worry and further testing. On the other hand, a false negative occurs when the test indicates no disease, but it is actually present. This can delay necessary treatment. Therefore, interpreting test results should be done cautiously, considering both types of errors.

Lecture 15

Independence Day Comes Early: Expectations, Random Variables and More

Independence of Two Events

We are interested in situations where the occurrence of one event doesn't influence the occurrence of another: $P(A|B) = P(A)$. This concept is known as the independence of events.

Definition. Two events A and B are considered **independent** if the occurrence of one does not affect the occurrence of the other. In mathematical terms, this can be expressed as:

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

Example. 1. Consider flipping a fair coin twice. Let A be the event of getting heads on the first flip and B be the event of getting tails on the second flip. These events are independent because the outcome of the first flip does not impact the outcome of the second flip.

[1] Feel free to read from right to left

2. Suppose you roll a fair six-sided die twice. Let A be the event of getting an even number on the first roll and B be the event of getting a 6 on the second roll. These events are independent because the outcome of one roll does not affect the outcome of the other roll.

Independence of Many Events

When the occurrence of multiple events is not influenced by the occurrence of any other event, we have independence of many events.

Definition. Events $A_1, A_2, \ldots, A_n$ are considered **mutually independent** if the occurrence of any one event does not affect the occurrence of any combination of the other events. In mathematical terms, this can be expressed as:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n).$$

It's important to note that pairwise independence doesn't necessarily imply mutual independence.

Example. Consider three events in a game of throwing two fair dice.

- A: Getting a sum of 7.
- B: The first die rolls a 4.
- C: The second die rolls a 5.

We want to determine if these events are independent.

We know that $P(A) = \frac{1}{6}$, as there are 6 possible ways to get a sum of 7 out of 36 equally likely outcomes. The probability of event B is also $\frac{1}{6}$ since there is one way to get a 4 on the first die. Similarly, the probability of event C is $\frac{1}{6}$.

Now, let's compute the probabilities of the pairwise intersections:

$$\begin{aligned}
P(A \cap B) &= P(\text{Sum is 7 and first die is 4}) = P(\{(4,3)\}) = \frac{1}{36}, \\
P(A \cap C) &= P(\text{Sum is 7 and second die is 5}) = P(\{(2,5)\}) = \frac{1}{36}, \\
P(B \cap C) &= P(\text{First die is 4 and second die is 5}) = P(\{(4,5)\}) = \frac{1}{36}.
\end{aligned}$$

Since $P(A \cap B) = P(A) \cdot P(B)$, $P(A \cap C) = P(A) \cdot P(C)$, and $P(B \cap C) = P(B) \cdot P(C)$, we can conclude that these events are pairwise independent.

$$P(A \cap B \cap C) = P(\text{Sum is 7, first die is 4, and second die is 5}) = 0.$$

As $P(A \cap B \cap C) = 0 \neq \dfrac{1}{216} = P(A) \cdot P(B) \cdot P(C)$, we see that these events are not mutually independent.

Random Variables

In probability theory, we often encounter situations where weare more interested in some measure or quantification of the outcomes, rather than the outcomes themselves. This is where the concept of random variables comes into play.

Examples

Game: Coin Flip

Sample Space: $\{H, T\}$ (Heads or Tails)

Rule: If the coin lands **Heads (H)**, you win 2 dollars. If it lands **Tails (T)**, you lose 1 dollar. Thus, we assign the values $X(H) = 2$ and $X(T) = -1$ to the respective outcomes.

Dice

Let's say we roll two six-sided dice.

Sample Space: $\{(1,1), (1,2), \ldots, (6,6)\}$

We can define a random variable $\mathcal{S}$ as the sum of the numbers rolled. The possible values of $\mathcal{S}$ are $\{2, 3, \ldots, 12\}$.

These examples provide different contexts for *random variables* and highlight their usefulness in various scenarios.

Definition. A **random variable** is a function that assigns a real number to each outcome in the sample space of an experiment.

In other words, a random variable is a rule that takes outcomes and maps them to real numbers. Let's explore some examples.

Example.

1. In a card game, we can define a random variable Y as the number of aces drawn from a shuffled deck of cards in a single draw. The possible values of Y are $\{0, 1, 2, 3, 4\}$.
2. Consider a manufacturing process that produces light bulbs. Let Z be the brightness (measured in lumens) of a randomly selected bulb. The possible values of Z form a continuous range of positive real numbers.
3. Consider a scenario where you're waiting for a bus. Let's say the sample space of possible waiting times is $\{1, 2, 3\}$ minutes. We want to define a random variable $W(k)$ that represents the number of buses that arrive within k minutes.

In the previous lecture, we introduced the concepts of 'discrete' and 'continuous' probability theories. These terms relate to the cardinalities of the range of the corresponding random variable.

Definition. A random variable is said to be **discrete** if its range is finite or *countable*. In mathematics, a set is considered countable if its elements can be put into one-to-one correspondence with the natural numbers $(1, 2, 3, \ldots)$, meaning there is a systematic way to list all the elements in the set.

The probability mass function, often abbreviated as PMF, is a fundamental concept in the study of discrete random variables. It gives the probability of each possible outcome.

Definition. For a discrete random variable X with possible values $x_1, x_2, \ldots$, the **probability mass function (PMF)** gives the probability that the random variable X takes on the value x_i.

Motivation: the PMF provides a precise way to describe the likelihood of each outcome in a discrete random experiment.

Example. Consider a pair of six-sided dice.

1. Let X be the random variable representing the outcome. There are 36 equally likely outcomes, so the PMF of X is:

$$P(X = (i, j)) = \frac{1}{36}, \quad \text{for } i, j \in \{1, 2, \ldots, 6\}$$

This reflects the equal probability of each outcome.

2. Let Y be the random variable representing the sum of outcomes. The range of Y is $\{2, 3, \dots, 12\}$. The PMF of Y is given by:

$$P(Y = k) = \begin{cases} \frac{k-1}{36} & \text{for } k = 2, 3, \dots, 7 \\ \frac{13-k}{36} & \text{for } k = 8, 9, \dots, 12. \end{cases}$$

Moreover, the probabilities $P(Y = k)$ sum up to 1:

$$\sum_{k=2}^{12} P(Y = k) = \sum_{k=2}^{7} \frac{k-1}{36} + \sum_{k=8}^{12} \frac{13-k}{36} = 1.$$

3. Let Z be the random variable representing the parity of the sum of outcomes, where 1 represents odd and 2 represents even. The PMF of Z is given by:

$$P(Z = i) = \begin{cases} \frac{1}{2} & \text{for } i = 1 \\ \frac{1}{2} & \text{for } i = 2. \end{cases}$$

This reflects the equal probability of getting an odd or even sum.

Expected Value

We are all familiar with the arithmetic mean - it's the usual way to compute an average. However, there are scenarios where we need to compute a 'weighted' average to account for varying probabilities.

Example. Consider a biased six-sided die, where the probabilities of rolling each face are as follows:

$$P(X = 1) = 0.1, \quad P(X = 2) = 0.2, \quad P(X = 3) = 0.3,$$
$$P(X = 4) = 0.15, \quad P(X = 5) = 0.1, \quad P(X = 6) = 0.15.$$

The expected value, which is a weighted average is computed as

$$\mathbb{E}(X) = (1 \cdot 0.1) + (2 \cdot 0.2) + (3 \cdot 0.3) + (4 \cdot 0.15) + (5 \cdot 0.1) + (6 \cdot 0.15) = 0.1 + 0.4 + 0.9 + 0.6 + 0.5 + 0.9 = 3.4$$

So, for this biased die, the expected value is 3.4. This expanded example includes the given probabilities for each face of the biased die and computes the weighted average to find the expected value. For a fair six-sided die, each face has an equal probability of $\frac{1}{6}$. The expected value is the arithmetic mean:

$$\mathbb{E}(X) = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3.5.$$

Comparing the expected values, we see that the biased die has an expected value of 3.4, while the fair die has an expected value of approximately 3.5. This illustrates how weighting the outcomes affects the expected value.

Definition. The **expected value** of a random variable X, denoted as $\mathbb{E}(X)$ or μ_X, is defined as

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x)$$

if X is discrete.

Example. 1. **Unfair Coin Toss** Suppose we have a biased coin with probabilities: $P(H) = 0.4$ and $P(T) = 0.6$. We define a random variable X as follows:

$$X(H) = 1, \quad X(T) = 2.$$

The expected value is calculated as:

$$\mathbb{E}(X) = P(H) \cdot 1 + P(T) \cdot 2 = 1 \cdot 0.4 + 2 \cdot 0.6 = 1.6.$$

2. **Biased Coin Toss** Suppose we have a biased coin with probabilities: $P(H) = 0.4$ and $P(T) = 0.6$. We want to find the expected number of steps needed to get a Heads. Define N as the number of steps. The expected value is calculated as

$$\mathbb{E}(N) = P(H) \cdot 0 + P(T)P(H) \cdot 1 + P(T)^2 P(H) \cdot 2 + \dots$$

Let $t = P(T)$ and consider the generating function with a dummy variable t:

$$G(t) = P(H) \sum_{n=0}^{\infty} t^n.$$

Notice that $\mathbb{E}(N) = P(H) \sum_{n=0}^{\infty} nt^n = tG'(t)$ and compute

$$\mathbb{E}(N) = tG'(t) = P(H) \cdot \frac{t}{(1-t)^2} = P(H) \cdot \frac{P(T)}{(1-P(T))^2} = 0.4 \cdot \frac{0.6}{(1-0.6)^2} = 1.5.$$

Question. Notice that $\mathbb{E}(N) = \dfrac{P(T)}{P(H)}$. Is there a quicker way to deduce this?

Alternatively, we can notice that the expected value $\mathbb{E}(N)$ satisfies the linear equation:

$$\mathbb{E}(N) = 0.4 \cdot 0 + 0.6(\mathbb{E}(N) + 1).$$

Solving for $\mathbb{E}(N)$ recovers

$$\mathbb{E}(N) = 1.5.$$

This equation arises from the probabilities of getting Heads or Tails on the first flip. With a probability of 0.4, we get Heads on the first flip, so the number of tosses before the first Heads is zero. With a probability of 0.6, we get Tails, and in that case, the expected number of tosses increases by 1 (from this extra flip).

Properties of Expectation

1. $\mathbb{E}(c) = c$: the expectation of a constant random variable c is simply the constant itself.

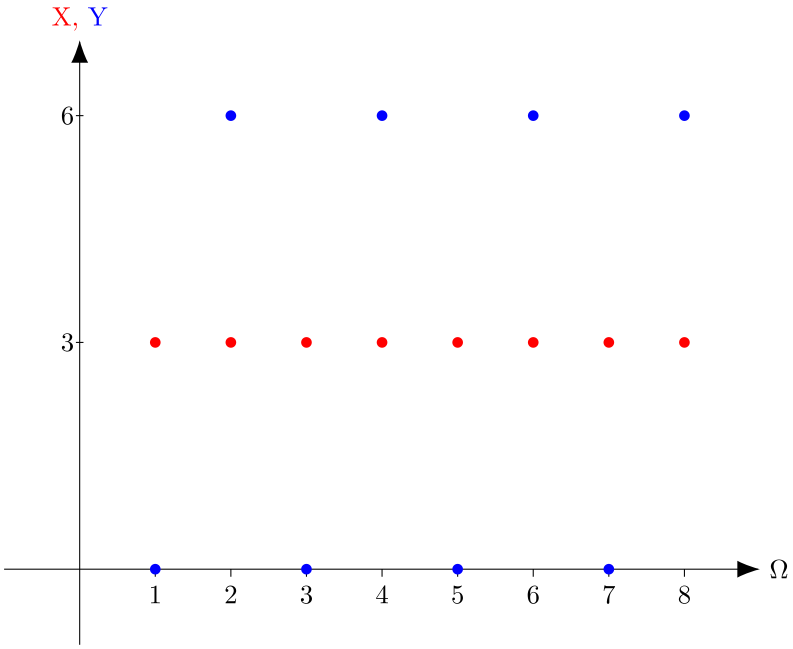
2. Linearity: for any constants a and b, and random variables X and Y,

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

   This property states that the expectation of a linear combination of random variables is equal to the linear combination of their individual expectations.

Variance and Standard Deviation

In statistics, variance and standard deviation are crucial measures that help us understand the spread or dispersion of a random variable's outcomes. While expectation gives us a measure of central tendency, variance and standard deviation provide us with information about how spread out the values are. Consider two random variables X and Y uniformly distributed on sample space $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$ with the same mean $\mathbb{E}(X) = \mathbb{E}(Y) = 3$. The first one, X, has all its values equal to that mean, while Y has a much more spread out distribution.



Definition. The **variance** of a random variable X, denoted as $\mathrm{Var}(X)$, is defined as

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2.\right]$$

Another useful formula for variance is

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X).$$

This formula is often more convenient for practical calculations.

Definition. The **standard deviation** of a random variable X, denoted as σ_X, is the square root of the variance

$$\sigma_X = \sqrt{\mathrm{Var}(X).}$$

Example. 1. **Coin Toss.** Consider the biased coin toss with probabilities: $P(H) = 0.4$ and $P(T) = 0.6$. Let X be the random variable defined as $X(H) = 1$ and $X(T) = 2$. The expected value is $\mathbb{E}(X) = 1.6$. The variance is calculated as:

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = (1 - 1.6)^2 \cdot 0.4 + (2 - 1.6)^2 \cdot 0.6 = 0.24.$$

The standard deviation is then $\sigma_X = \sqrt{\mathrm{Var}(X)} \approx 0.49$.

2. **Biased Die.** Consider the biased six-sided die with probabilities:

$$P(X = 1) = 0.1,\ P(X = 2) = 0.2,\ P(X = 3) = 0.3,$$

$$P(X = 4) = 0.15,\ P(X = 5) = 0.1,\ P(X = 6) = 0.15.$$

The expected value is $\mathbb{E}(X) = 3.4$. The variance is computed as:

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = (1 - 3.4)^2 \cdot 0.1 + (2 - 3.4)^2 \cdot 0.2 + \ldots + (6 - 3.4)^2 \cdot 0.15 = 2.34$$

The standard deviation is then $\sigma_X = \sqrt{\mathrm{Var}(X)} \approx 1.53$. Let's use the alternative formula for variance to verify that it indeed yields the same result:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = (1^2 \cdot 0.1) + (2^2 \cdot 0.2) + \ldots + (6^2 \cdot 0.15) - (3.4)^2 = 2.34.$$

As expected, the result coincides with the previous calculation using the original formula. This demonstrates that both formulas for variance are equivalent and yield the same result.

Lecture 16
A Trial and a Law That Most Lawyers Are Clueless About: Understanding Bernoulli Trials and the Law of Large Numbers

Bernoulli and Binomial Distributions

The Bernoulli distribution is a discrete probability distribution for a random variable which takes the value 1 (representing successful outcome of an experiment) with probability p and the value 0 (representing failure) with probability $1 - p$.

Probability Mass Function (PMF)

$$P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

where $0 \leq p \leq 1$.

Mean and Variance

- **Mean**: $E(X) = p$
- **Variance**: $\text{Var}(X) = p(1 - p)$

Example. 1. Consider a coin toss where getting a head is a success (1) and getting a tail is a failure (0). If the coin is fair, then $p = 0.5$.

2. Consider a medical test for a rare disease. If the probability of testing positive (success) is $p = 0.01$, then the probability of testing negative (failure) is $1 - p = 0.99$.

Binomial Distribution

The binomial distribution describes the number of successes in a fixed number of independent Bernoulli trials, each with the same probability of success p.

Probability Mass Function (PMF)

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

where:

- n is the number of trials,
- k is the number of successes,
- $0 \leq k \leq n$.

Mean and Variance

- **Mean**: $E(X) = np$
- **Variance**: $\text{Var}(X) = np(1 - p)$

Example. 1. Consider a basketball player who has a 70% chance of making a free throw. If she takes 10 free throws, the number of successful free throws follows a binomial distribution with $n = 10$ and $p = 0.7$.

2. Another example is a quality control inspector who checks 20 items from a production line, where each item has a 95% chance of meeting quality standards. The number of items meeting the quality standards follows a binomial distribution with $n = 20$ and $p = 0.95$.

Math meets Bio: Hardy-Weinberg Principle (optional)

The Hardy-Weinberg principle is a fundamental concept in population genetics that describes the genetic equilibrium within a population. It states that allele and genotype frequencies in a population remain constant from generation to generation in the absence of evolutionary influences.

Alleles, Homozygotes, and Heterozygotes

- *Alleles* are different forms of a gene that exist at a specific locus on a chromosome. For a given gene, individuals can have two copies, one inherited from each parent.
- An individual is *homozygous* for a gene if they have two identical alleles (e.g., AA or aa).
- An individual is *heterozygous* for a gene if they have two different alleles (e.g., Aa).

Formulation for Two Alleles

For a gene with two alleles, A (dominant) and a (recessive), with frequencies p and q respectively (where $p + q = 1$), the genotype frequencies in the first generation are given by:

- $P(AA) = p^2$
- $P(Aa) = 2pq$
- $P(aa) = q^2$

These frequencies can be derived from the binomial expansion of $(p + q)^2$.

In the second generation, these frequencies remain the same due to the principle of genetic equilibrium:

$$(p^2 + 2pq + q^2) \times (p + q) = p^2 + 2pq + q^2.$$

In general, for any subsequent generation, the genotype frequencies remain the same:

- $P(AA) = p^2$
- $P(Aa) = 2pq$
- $P(aa) = q^2$

Generalization to Three and More Alleles

For a gene with three alleles, A_1, A_2, and A_3, with frequencies p_1, p_2, and p_3 respectively (where $p_1+p_2+p_3 = 1$), the genotype frequencies in the first generation are given by:

- $P(A_1A_1) = p_1^2$
- $P(A_1A_2) = 2p_1p_2$
- $P(A_1A_3) = 2p_1p_3$
- $P(A_2A_2) = p_2^2$
- $P(A_2A_3) = 2p_2p_3$
- $P(A_3A_3) = p_3^2$

These frequencies can be derived from the multinomial expansion of $(p_1+p_2+p_3)^2$ or from the multinomial expansion of $(p_1 + p_2 + \ldots + p_k)^2$ in case of k alleles.

Example. Consider a population where the allele frequency of A is 0.6 and a is 0.4. Using the Hardy-Weinberg principle:

- Frequency of AA: $0.6^2 = 0.36$
- Frequency of Aa: $2 \times 0.6 \times 0.4 = 0.48$
- Frequency of aa: $0.4^2 = 0.16$

Polyploidy

Polyploidy refers to the condition in which an organism has more than two complete sets of chromosomes. In the context of the Hardy-Weinberg principle, polyploidy affects the calculation of genotype frequencies.

For an organism with ploidy level c and two alleles, the genotype frequencies can be derived from the binomial expansion of $(p+q)^c$.

Example. Consider a polyploid organism with three sets of chromosomes (triploid, $c = 3$). For two alleles A and a with frequencies p and q, the genotype frequencies are given by the expansion of $(p+q)^3$:

- $P(AAA) = p^3$
- $P(AAa) = 3p^2q$
- $P(Aaa) = 3pq^2$
- $P(aaa) = q^3$

The Hardy-Weinberg equilibrium is named after G. H. Hardy, an English mathematician, and Wilhelm Weinberg, a German physician. Weinberg had already formulated the principle in Germany, but it was Hardy who provided the mathematical rigor needed to formalize it.

Markov's Inequality

In probability theory, Markov's inequality provides a powerful tool for bounding probabilities involving random variables. This inequality allows us to make statements about the likelihood of a random variable deviating from its expected value.

Consider a non-negative random variable X and let $a > 0$ be any positive number. Markov's inequality states

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

This means that the probability that a non-negative random variable exceeds a certain positive value is bounded by the expected value divided by that value.

Example. 1. Suppose we have a biased coin with $P(H) = 0.2$ and we flip it ten times. We want to estimate the probability of obtaining at least 8 Heads.

In our case, X represents the number of Heads obtained in ten flips. The expected value of X is $\mathbb{E}(X) = n \cdot p = 10 \cdot 0.2 = 2$.

Using Markov's inequality with $a = 8$, we get

$$P(X \geq 8) \leq \frac{2}{8} = 0.25.$$

Now, let's compute the actual probability:

$$P(X \geq 8) = P(X = 8) + P(X = 9) + P(X = 10) = \binom{10}{8} \cdot 0.2^8 \cdot 0.8^2 + \binom{10}{9} \cdot 0.2^9 \cdot 0.8 + 0.2^{10} \approx 0.00007793.$$

Conclusion: the inequality $P(X \geq a) \leq \frac{E[X]}{a}$ gives an upper bound for the probability. In this case, the upper bound was 0.25, which is much larger than the actual probability of approximately 0.00007793. While this bound is often not very tight, it is a quick and easy way to obtain an estimate for the probability of rare events using the expected value.

2. In a company with 200 employees, it is found that the average number of years each employee has worked at the company is 6. Using Markov's inequality, we can find

$$P(\text{Years of Service} \geq 10) \leq \frac{\mathbb{E}(\text{Years of Service})}{10} = \frac{6}{10} = 0.6.$$

This means that the probability of an employee having worked at the company for 10 or more years is less than or equal to 0.6.

3. In a city with a population of $100,000$, you know that the average number of cars per household is 1.5. Using Markov's inequality, we can find

$$P(\text{Number of Cars in a Household} \geq 3) \leq \frac{\mathbb{E}(\text{Number of Cars in a Household})}{3} = \frac{1.5}{3} = 0.5$$

This means that the probability of a household having 3 or more cars is less than or equal to 0.5.

Optional: Proof of Markov's Inequality

Note: This section is optional.

We will now prove Markov's inequality for continuous random variables.

Consider a non-negative random variable X with probability mass function $p(x)$. The expected value of X is given by:

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} np(x=n) = \sum_{n=0}^{a-1} np(x=n) + \sum_{n=a}^{\infty} np(x=n) \geq$$
$$\geq \sum_{n=a}^{\infty} np(x=n) \geq \sum_{n=a}^{\infty} ap(x=n) = a\sum_{n=a}^{\infty} p(x=n) = aP(X \geq a)$$

It follows that

$$\mathbb{E}(X) \geq aP(X > a) \Leftrightarrow \frac{\mathbb{E}(X)}{a} \geq P(X \geq a).$$

In the above inequalities, we used the fact that the sum over the entire range is greater than or equal to the sum over any subset, since $n \geq 0$. This establishes Markov's inequality for discrete random variables.

Chebyshev's Inequality and Properties of Variance

Next we turn our attention to Chebyshev's inequality, a versatile concept with wide-ranging applications. It plays a pivotal role in various fields. In finance, it assists in managing investment risk by estimating the likelihood of significant deviations from expected returns. In manufacturing, it ensures that a substantial majority of products meet quality standards, even when the distribution of characteristics is uncertain. For betting strategies, especially in sports betting, it provides insights into the probability of experiencing specific losses over a given period. In epidemiology and public health, Chebyshev's inequality aids in estimating the proportion of a population at risk of contracting a disease within a defined timeframe, based on the mean and variance of infection rates.

For any random variable X (not necessarily non-negative) with finite variance σ^2, and any $k > 0$, Chebyshev's inequality states

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

where $\mu = \mathbb{E}(X)$ is the mean of X.

This inequality provides a bound on the probability of a random variable deviating from its mean by a certain number of standard deviations.

Example. 1. Let W be a random variable representing the weight of a certain type of fruit. The mean weight of a fruit is 150 grams and the variance is 25 grams. Suppose we would like to estimate the probability of the weight of this type of fruit not falling within the range of $130 - 170$ grams. Chebyshev's inequality can be applied to find

$$P(|W - 150| \geq 20) \leq \frac{1}{4^2} = 0.0625,$$

where $\sigma = \sqrt{25} = 5$, giving $k\sigma = 5k = 20$ and $k = 4$. This means that the probability of the weight of this type of fruit not falling within the range of $130 - 170$ grams is less than or equal to 6.25%.

2. Suppose X is a random variable such that $\mathbb{E}(X) = 3$ and $\mathbb{E}(X^2) = 13$, we can use Chebyshev's inequality to determine a lower bound for the probability $P(-2 < X < 8)$.

   In this case, we have $\mathbb{E}(X) = 3$ and $\mathbb{E}(X^2) = 13$. The variance of X can be calculated using the formula

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 13 - (3)^2 = 13 - 9 = 4.$$

   So, $\sigma^2 = 4$ and $\sigma = 2$.

   We want to find $P(-2 < X < 8)$, which can be rephrased as $P(|X - 3| < 5)$. Using Chebyshev's inequality

$$P(|X - 3| < 5) \geq 1 - \frac{1}{k^2} = 1 - \frac{4}{5^2} = 1 - \frac{4}{25} = \frac{21}{25} \approx 0.84,$$

   where we have used that $k = \frac{5}{\sigma} = \frac{5}{2}$. So, Chebyshev's inequality gives a lower bound of approximately 84% for the probability $P(-2 < X < 8)$.

While in principle, Chebyshev's inequality pertains to distances from the mean in either direction, it can still provide an estimate of how frequently a random variable may assume large values. It typically offers more precise bounds compared to Markov's inequality. To illustrate, let's revisit the example from our previous lecture, where we tossed a weighted coin with a 20% chance of landing heads, ten times. Recall that Markov's inequality provided an upper bound of $\frac{1}{4}$ on the probability of getting at least 8 heads. First we let $\widetilde{X}$ be the random variable given by $\widetilde{X}(H) = 1$ and $\widetilde{X}(T) = 0$ and compute $\mathbb{E}(\widetilde{X}) = 0.2 \cdot 1 + 0.8 \cdot 0 = 0.2$, $Var(\widetilde{X}) = 0.2 \cdot 1^2 + 0.8 \cdot 0^2 - 0.2^2 = 0.16$. The random variable, giving the number of heads in 10 coin flips, is $X = \widetilde{X}_1 + \widetilde{X}_2 + \ldots + \widetilde{X}_{10}$, where $\widetilde{X}_i$'s are independent and have the same probability distribution as $\widetilde{X}$. It follows that $\mathbb{E}(X) = \mathbb{E}(\widetilde{X}_1 + \ldots + \widetilde{X}_{10}) = \mathbb{E}(\widetilde{X}_1) + \ldots + \mathbb{E}(\widetilde{X}_{10}) = 10\mathbb{E}(\widetilde{X}) = 10 \cdot 0.2 = 2$, while $Var(X) = Var(\widetilde{X}_1 + \ldots + \widetilde{X}_{10}) = Var(\widetilde{X}_1) + \ldots + Var(\widetilde{X}_{10}) = 10 \cdot Var(\widetilde{X}) = 10 \cdot 0.16 = 1.6$, giving $\sigma = \sqrt{1.6}$. As $P(|X - 2| \geq 6) = P(X \leq -4 \cup X \geq 8) = P(X \geq 8)$, since X is nonnegative, Chebyshev's inequality allows to obtain $P(X \geq 8) = P\left(|X - 2| \geq 6 = \frac{6}{\sqrt{1.6}} \cdot \sqrt{1.6}\right) \leq \frac{1}{\left(\frac{6}{\sqrt{1.6}}\right)^2} \approx 0.044.$

The Law of Large Numbers

The Law of Large Numbers states that:

Weak Law of Large Numbers (WLLN): Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with finite mean μ and finite variance σ^2. Then, for any $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) = 0.$$

Example. Consider flipping a fair coin 6 times. Each flip can result in either heads (H) or tails (T), representing a success or failure, respectively.

Let's consider two concrete outcomes:

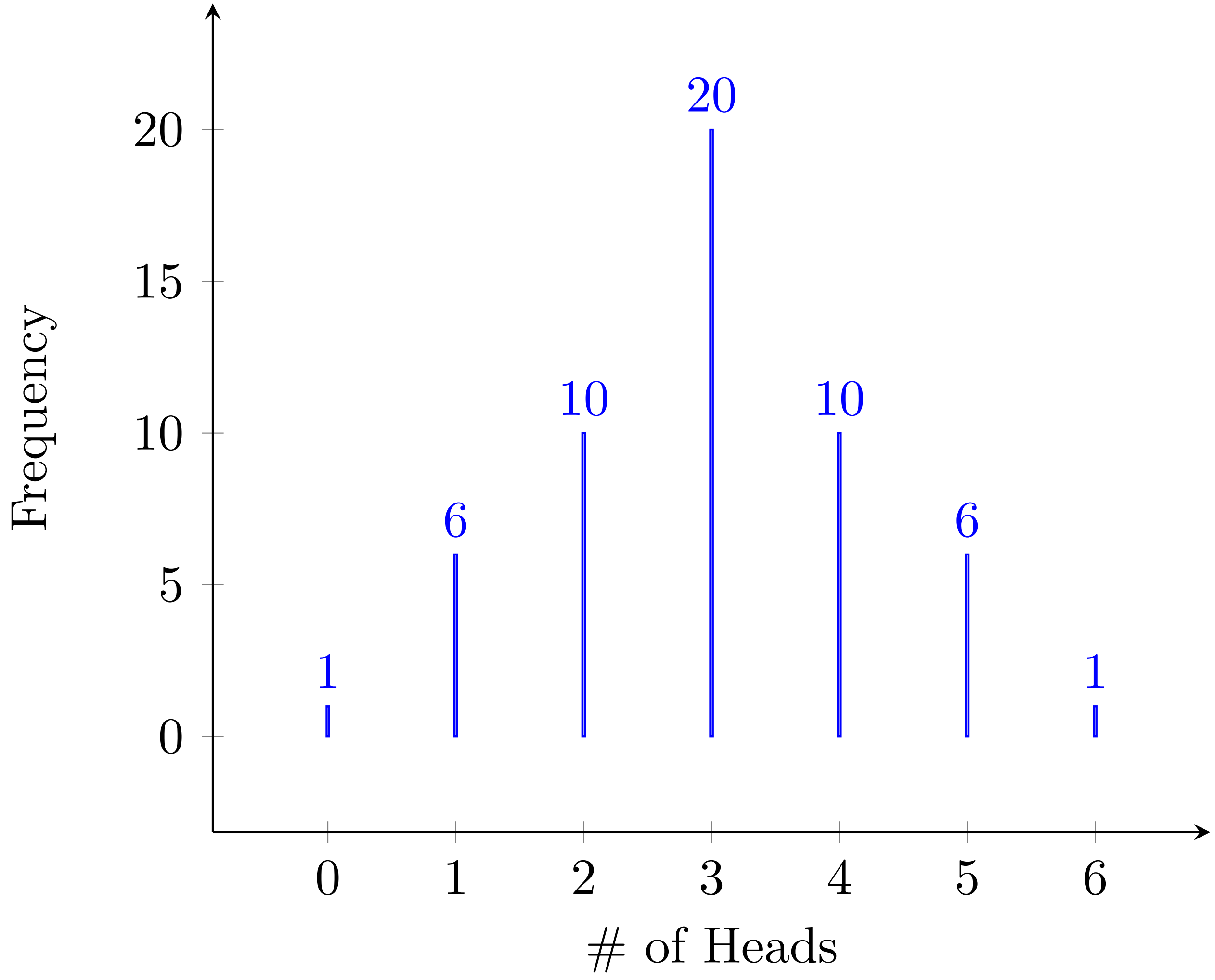
1. Outcome 1: HHTHTH

2. Outcome 2: TTHHHT

For Outcome 1 (HHTHTH), the number of heads is $\frac{4}{6}$, while for Outcome 2 (TTHHHT), the average number of heads is $\frac{3}{6}$.

The number of heads obtained in the coin flipping experiment follows the binomial distribution since we conduct independent Bernoulli experiments multiple times. From the binomial distribution, we know that outcomes with values close to the average have a higher probability. This is because the corresponding binomial coefficients are larger.

For example, in our experiment, the binomial coefficient for getting exactly 3 heads out of 6 flips is $\binom{6}{3} = 20$. This means that getting 3 heads is more probable than getting $0, 1, 2, 4, 5$, or 6 heads:


Frequency
# of Heads