CIS 160 Lecture Notes
January 18, 2020

Counting

Counting is a part of combinatorics, an area of mathematics which is concerned with the arrangement of objects of a set into patterns that satisfy certain constraints. We will mainly be interested in the number of ways of obtaining an arrangement, if it exists.

Before we delve into the subject, let's take a small detour and understand what a *set* is. Below are some relevant definitions.

- A *set* is an *unordered* collection of distinct objects. The objects of a set are sometimes referred to as its elements or members. If a set is finite and not too large it can be described by listing out all its elements, e.g., $\{a, e, i, o, u\}$ is the set of vowels in the English alphabet. Note that the order in which the elements are listed is not important. Hence, $\{a, e, i, o, u\}$ is the same set as $\{i, a, o, u, e\}$. If V denotes the set of vowels then we say that e belongs to the set V, denoted by $e \in V$ or $e \in \{a, e, i, o, u\}$.
- Two sets are *equal* if and only if they have the same elements.
- The *cardinality* of S, denoted by $|S|$, is the number of distinct elements in S.
- A set A is said to be a *subset* of B if and only if every element of A is also an element of B. We use the notation $A \subseteq B$ to denote that A is a subset of the set B, e.g., $\{a, u\} \subseteq \{a, e, i, o, u\}$. Note that for any set S, the empty set $\{\} = \emptyset \subseteq S$ and $S \subseteq S$. If $A \subseteq B$ and $A \neq B$ the we say that A is a *proper subset* of B; we denote this by $A \subset B$. In other words, A is a proper subset of B if $A \subseteq B$ and there is an element in B that does not belong to A.
- A *power set* of a set S, denoted by $P(S)$, is a set of all possible subsets of S. For example, if $S = \{1, 2, 3\}$ then $P(S) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\}$. In this example $|P(S)| = 8$.
- Some of the commonly used sets in discrete mathematics are: $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, $\mathbb{Q} = \{p/q \,|\, p \in \mathbb{Z} \text{ and } q \in \mathbb{Z} \text{ ,and } q \neq 0\}$, and $\mathbb{R}$ is the set of real numbers.
- Another way to describe a set is by explicitly stating the properties that all members of the set must have. For instance, the set of all positive even integers less than 100 can be written as $\{x \,|\, x \text{ is a positive even integer less than } 100\}$ or $\{x \in \mathbb{Z}^+ \,|\, x < 100 \text{ and } x = 2k, \text{ for some integer } k\}$. Similarly, the set $\{2, 4, \ldots, 12\}$ can be written as $\{2n \,|\, 1 \leq n \leq 6 \text{ and } n \in \mathbb{N}\}$ or $\{n + 1 \,|\, n \in \{1, 3, 5, 7, 11\}\}$.

Understanding the above terminology related to sets is enough to get us started on counting.

Theorem. If m and n are integers and $m \leq n$, then there are $n - m + 1$ integers from m to n inclusive.

Example. How many three-digit integers (integers from 100 to 999 inclusive) are divisible by 5?

Solution. The first number in the range that divisible by 5 is 100 (5×20) and the last one that is divisible by 5 is 995 (5×199). Using the above theorem, there are $199 - 20 + 1 = 180$ numbers from 100 to 999 that are divisible by 5.

Tree Diagram. A tree diagram is a very useful tool for systematically keeping track of all possible outcomes of a combinatorial process. We will also use this tool when we study probability.

Example. Teams A and B are to play each other in a best-of-three match, i.e., they play each other until one team wins two games in a row or a total of three games are played. What is the number of possible outcomes of the match? What does the possibility tree look like if they play three games regardless of who wins the first two?

Solution. The possibility trees for the two cases are shown in Figure 1. From the tree diagram it is clear that there are 6 outcomes in the first case and 8 in the second case.

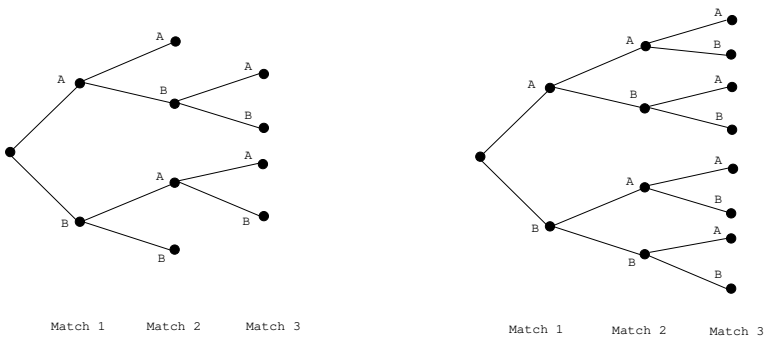


Figure 1: Tree diagrams.

Multiplication Rule. If a procedure can be broken down into k steps and

> the first step can be performed in n_1 ways,
> the second step can be performed in n_2 ways, regardless of how the first step was performed,
> $\vdots$
> the k^{th} step can be performed in n_k ways, regardless of how the preceding steps were performed, then

the entire procedure can be performed in $n_1 \cdot n_2 \cdots n_k$ ways.

To apply the multiplication rule think of objects that you are trying to count as the output of a multi-step operation. The possible ways to perform a step may depend on how the

preceding steps were performed, but the number of ways to perform each step must be constant regardless of the action taken in prior steps.

Example. An ordered pair (a, b) consists of two things, a and b. We say that a is the first member of the pair and b is the second member of the pair. If M is an m-element set and N is an n-element set, how many ordered pairs are there whose first member belongs to M and whose second member belongs to N?

Solution. An ordered pair can be formed using the following two steps.

Step 1. Choose the first member of the pair from the set M.

Step 2. Choose the second member of the pair from the set N.

Step 1 can be done in m ways and Step 2 can be done in n ways. From the multiplication rule it follows that the number of ordered pairs is mn.

Example. A local deli that serves sandwiches offers a choice of three kinds of bread and five kinds of filling. How many different kinds of sandwiches are available?

Solution. A sandwich can be made using the following two steps.

Step 1. Choose the bread.

Step 2. Choose the filling.

Step 1 can be done in 3 ways and Step 2 can be done in 5 ways. From the multiplication rule it follows that the number of available sandwich offerings is 15.

Example. The chairs of an auditorium are to be labeled with a upper-case letter and a positive integer not exceeding 100. What is the largest number of chairs that can be labeled differently?

Solution. A chair can be labeled using the following two steps.

Step 1. Choose the upper-case letter.

Step 2. Choose the number.

Step 1 can be done in 26 ways and Step 2 can be done in 100 ways. From the multiplication rule it follows that the number of possible labelings is 2600.

Example. A typical PIN is a sequence of any four symbols chosen from 26 letters in the alphabet and the 10 digits, with repetition allowed. How many different PINS are possible? What happens if repetition is not allowed?

Solution. A PIN can be formed using the following steps.

Step 1. Choose the alphanumeric for the first position.

Step 2. Choose the alphanumeric for the second position.

Step 3. Choose the alphanumeric for the third position.

Step 4. Choose the alphanumeric for the fourth position.

When repetition is allowed, each step can be done in 36 ways and hence the number of possible PINS is 36^4. When repetition is not allowed, the number of ways of doing Step 1 is 36, the number of ways of doing Step 2 is 35, the number of ways of doing Step 3 is 34, and the number of ways of doing Step 4 is 33. By multiplication rule, the number of PINs in this case is $36 \times 35 \times 34 \times 33$.

Example. Three officers - a president, a treasurer, and a secretary - are to be chosen from among four people: Ann, Bob, Clyde, and Dan. Suppose that for various reasons, Ann cannot be the president and either Clyde or Dan must be the secretary. In how many ways can the officers be chosen?

Solution. Attempt 1. A set of three officers can be formed as follows.

Step 1. Choose the president.

Step 2. Choose the treasurer.

Step 3. Choose the secretary.

There are 3 ways to do Step 1. There are 3 ways of doing Step 2 (all except the person chosen in Step 1), and 2 ways of doing Step 3 (Clyde or Dan). By multiplication rule, the number of different ways of choosing the officers is $3 \times 3 \times 2 = 18$.

The above solution is incorrect because the number of ways of doing Step 3 depends upon the outcome of Steps 1 and 2 and hence the multiplication rule cannot be applied. It is easy to see this from the tree diagram in Figure 2.

Attempt 2. A set of three officers can be formed as follows.

Step 1. Choose the secretary.

Step 2. Choose the president.

Step 3. Choose the treasurer.

Step 1 can be done in 2 ways (Clyde or Dan). Step 2 can be done in 2 ways (Ann cannot be the president and the person chosen in Step 1 cannot be the president). Step 3 can be done in 2 ways (either of the two remaining people can be the treasurer). By multiplication rule, the numberof ways in which the officers can be chosen is $2 \times 2 \times 2 = 8$.

From the previous example we learn that there may not be a fixed order in which the operations have to be performed, and by changing the order a problem may be more readily solved by the multiplication rule. A rule of thumb to keep in mind is to *make the most restrictive choice first.*

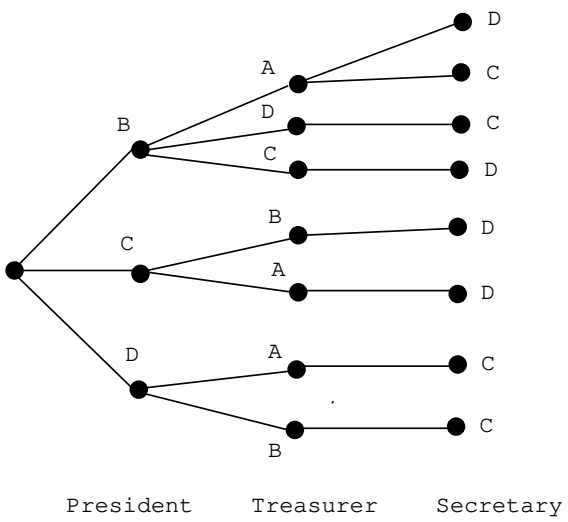


Figure 2: Tree diagram. In the tree A, B, C, stand for Ann, Bob, and Clyde respectively.

Example. Recall that the power set $P(S)$ of a set S is the set of all possible subsets of S. If $S = \{x_1, x_2, \ldots, x_n\}$, what is $P(S)$?

Solution. A subset of S can be constructed in n steps such that in step $i, 1 \leq i \leq n$, we decide whether to choose x_i or not. Each step can be performed in 2 ways regardless of the decisions made in the previous steps. By using the multiplication rule, the total number of subsets of S equals 2^n.

Example. How many odd numbers between 1000 and 9999 have distinct digits?

Solution.
Attempt 1: An odd number from 1000 through 9999 can be constructed in four steps as follows.

Step 1. Choose the first digit.

Step 2. Choose the second digit.

Step 3. Choose the third digit.

Step 4. Choose the fourth digit.

Observe that the number of ways of performing Step 4 depends upon the choices made in the earlier steps. For example, if the choices made in the first three steps are 1, 3, and 5, then Step 4 can be performed in two ways. However, if the choices made in the first three steps are 2, 4, and 6 then Step 4 can be performed in five ways. Hence, we cannot apply multiplication rule to solve the problem in the above manner.

Attempt 2: An odd number from 1000 through 9999 can be constructed in four steps as follows.

Step 1. Choose the fourth digit.

Step 2. Choose the third digit.

Step 3. Choose the second digit.

Step 4. Choose the first digit.

Note that the number of ways of performing Step 4 depends upon whether a zero was chosen in the earlier steps. If a zero was chosen in either Step 2 or Step 3 then the number of ways of performing Step 4 is 7, otherwise it is 6. Hence, multiplication rule cannot be applied to solve the problem in the above manner.

Attempt 3. An odd number from 1000 through 9999 can be constructed in four steps as follows.

Step 1. Choose the fourth digit.

Step 2. Choose the first digit.

Step 3. Choose the second digit.

Step 4. Choose the third digit.

There are 5 ways to perform Step 1, 8 ways to perform Step 2, 8 ways to perform Step 3, and 7 ways to perform Step 4. Note that the number of ways of doing each step is independent of the choices made in the earlier steps. By the multiplication rule, the number of odd numbers from 1000 through 9999 equals $5 \times 8 \times 8 \times 7 = 2240$.

Q. How many even numbers between 1000 and 9999 have distinct digits? Note that the solution to the above problem does not work for this one.

Permutations.

A permutation of a set of distinct objects is an ordering of the objects in a row. For example, the set of elements x, y, and z has six permutations: $xyz, xzy, yxz, yzx, zxy, zyx$.

In general, how many permutations are possible if we have a set of n distinct objects?

A permutation can be obtained in a sequence of n steps such that in step $i, 1 \leq i \leq n$, we choose the ith element in the ordering. Note that step i can be performed in i ways regardless of the choices made in the first $i-1$ steps. By multiplication rule, the number of permutations is

$$n \times n-1 \times n-2 \times \cdots \times 2 \times 1 = n!.$$

Example. Consider the set of letters $\{a, b, c, d, e, f, g, h\}$. (a) How many possible permutations are there of these letters? (b) How many permutations of these letters contain the substring abc?

Solution.
(a) There are 8 distinct elements and hence 8! permutations.
(b) We consider the string *abc* as one unit and that along with the remaining elements amounts to 6 distinct elements. Hence there are 6! possible permutations.

The following question was raised in class. Can we solve part (b) from first principles? We can do it as follows.
A permutation of letters consisting of substring *abc* can be constructed in eight steps as follows. In Steps 1, 2, and 3, choose the positions for a, b, and c, respectively. In Step i, $4 \leq i \leq 8$, choose position for the ith element in the set. Step 1 can be performed in 6 ways as a can be placed only in the first six positions. Choosing a position for a also decides positions for b and c. Hence, Steps 2 and 3 can be performed in exactly 1 way. Step i, $4 \leq i \leq 8$ can be performed in $8 - (i - 1)$ ways regardless of the choices made in the earlier steps. By the multiplication rule, the number of required permutations is given by

$$6 \times 1 \times 1 \times 5 \times 4 \times 3 \times 2 \times 1 = 6!$$

Permutations of Selected Elements.

We looked at permutations of n elements out of the available n elements. Now we will consider permutations of r elements out of the available n elements. Such an arrangement is called an r-permutation. For example, ab, ba, ac, ca, bc, cb are all 2-permutations of the set $\{a, b, c\}$.

Let $P(n, r)$ denote the number of r-permutations of a set of n elements. What is the value of $P(n, r)$?

Forming an r-permutation of a set of n elements can be thought of as an r-step process such that in step $i, 1 \leq i \leq r$, we choose the ith element of the ordering. There are $n - (i - 1) = n - i + 1$ ways of performing step i. By the multiplication rule, the number of r-permutations equals

$$\begin{aligned} P(n, r) &= n \times n - 1 \times n - 2 \times \cdots \times n - (r - 1) \\ &= n \times n - 1 \times n - 2 \times \cdots \times n - r + 1 \\ &= \frac{n \times (n - 1) \times \cdots \times (n - r + 1) \times (n - r) \times \cdots \times 1}{n - r \times (n - r - 1) \times (n - r - 2) \times \cdots \times 1} \\ &= \frac{n!}{(n - r)!} \end{aligned}$$

Example. How many ways are there to select a first-prize winner, a second-prize winner, and a third-prize winner from 100 different contestants?

Solution. Selecting the winners can be done in 3 steps with each step $i, 1 \leq i \leq 3$ choosing the winner in the ith place. Step i can be performed in $100 - (i - 1)$ ways. By

multiplication rule, the total number of possible ways in which the prizes can be given is $100 \times 99 \times 98 = 970200$. Note that this is same as $P(100, 3)$.

Example. In how many ways can we order 26 letters of the alphabet so that no two of the vowels a, e, i, o, u occur consecutively?

Solution. The task of ordering the letters so that no two vowels appear consecutively can be performed in two steps.

Step 1. Order the 21 consonants.

Step 2. Choose locations for the 5 vowels. The vowels can be placed before the consonants, between the consonants and after the consonants.

Step 1 can be performed in 21! ways. To count the number of ways of performing Step 2, observe that there is only one location for placing a vowel before and after the consonants, and 20 locations for placing the vowels between the consonants. This gives a total of 22 valid locations for placing 5 vowels. Thus the number of ways of placing the 5 vowels in 5 of the 22 locations is $P(22, 5)$. This is because there are 22 locations for a, 21 for e, 20 for i, 19 for o, and 18 for u. By multiplication rule, the total number of orderings in which no two vowels occur consecutively equals

$$21! \times P(22, 5) = \frac{21! \times 22!}{17!}$$

The Inclusion-Exclusion Formula.

If A, B, and C are any finite sets, then

$$\begin{aligned} |A \cup B| &= |A| + |B| - |A \cap B| \\ |A \cup B \cup C| &= |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| \\ &\quad + |A \cap B \cap C| \end{aligned}$$

If we have finite sets $A_1, A_2, \ldots, A_n$ then

$$\left| \bigcup_{i=1}^{n} A_i \right| = \sum_{i=1}^{n} |A_i| - \sum_{\substack{i,j \\ i<j}} |A_i \cap A_j| + \sum_{\substack{i,j,k \\ i<j<k}} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n-1} | \cap_{i=1}^{n} A_i|$$

Observe that if the sets A, B, and C are mutually disjoint, i.e., $A \cap B = A \cap C = B \cap C = \emptyset$ then we get

$$\begin{aligned} |A \cup B| &= |A| + |B| \\ |A \cup B \cup C| &= |A| + |B| + |C| \end{aligned}$$

This is often called the *addition rule* or the *sum rule*.

Example. In how many ways can we select two books from different subjects among five distinct computer science books, three distinct math books, and two distinct art books?

Solution. The set of all possible two books from different subjects can be partitioned into three subsets, S_1, S_2, and S_3. The subset S_1 contains two books belonging to computer science and math, the subset S_2 contains two books belonging to computer science and art, and the subset S_3 contains two books belonging to math and art. We have

$$\begin{aligned} |S_1| &= 5 \times 3 = 15 \\ |S_2| &= 5 \times 2 = 10 \\ |S_3| &= 3 \times 2 = 6 \end{aligned}$$

By the addition rule, total number of ways of selecting 2 books from different subjects equals $|S_1| + |S_2| + |S_3| = 31$.

Example. A PIN is typically made of four symbols chosen from 26 letters of the alphabet and the 10 digits, with repetitions allowed. How many PINS contain repeated symbols?

Solution. Let S denote the set of all possible PINs of four alpha-numeric characters. Let S_1 denote the set of all possible PINs of four alpha-numeric characters with no repeated symbols. Let S_2 denote the set of all possible PINs of four alpha-numeric characters with some symbols repeated. By the addition rule,

$$|S| = |S_1| + |S_2|$$

By simple application of multiplication rule, we see that $|S| = 36^4 = 1679616$ and $|S_1| = 36 \times 35 \times 34 \times 33 = 1413720$. Plugging these values in the above equation, we get $|S_2| = 265896$.

Example. (a) How many integers from 1 through 1000 are multiples of 3 or multiples of 5?
(b) How many integers from 1 through 1000 are neither multiples of 3 nor multiples of 5?

Solution. (a) Let $S = \{1, 2, 3, \ldots, 1000\}$. Let $M \subseteq S$ be the set of integers that are multiples of 3 or multiples of 5. Let $M_1 \subseteq S$ be the set of integers that are multiples of 3. Let $M_2 \subseteq S$ be the set of integers that are multiples of 5. Note that the first integer in S that is divisible by 3 is $3 = 3 \times 1$. The last integer in S that is divisible by 3 is $999 = 3 \times 333$. Thus, $|M_1| = 333$. Similarly, $|M_2| = 200$. Note that M_1 and M_2 are not disjoint, i.e., there are integers like 15 that are divisible by 3 and by 5 and hence exist in M_1 as well as M_2. We have double-counted them. So now, let's find the size of the set $M_1 \cap M_2$. Observe that each element in $M_1 \cap M_2$ must be a multiple of $3 \times 5 = 15$. The first number in S that is a multiple of 15 is $15 = 15 \times 1$ and the last number in S that is a multiple of 15 is $990 = 15 \times 66$. Thus, $|M_1 \cap M_2| = 66$. By the inclusion-exclusion formula, we get

$$|M| = |M_1| + |M_2| - |M_1 \cap M_2| = 333 + 200 - 66 = 467$$

(b) Let $N \subseteq S$ be the set of integers that are neither multiples of 3 nor multiples of 5. Note that the sets M and N form a partition of the set S. Applying the addition rule we get

$$\begin{aligned} |S| &= |M| + |N| \\ \therefore |N| &= |S| - |M| \\ &= 1000 - 467 \\ &= 533 \end{aligned}$$

Example. How many strings are there of four lower-case letters that have the letter x in them?

Solution. Let S be the set of all possible four-letter strings that can be constructed using lower-case letters. The set S can be partitioned into two sets S_1 and S_2 where S_1 is the set of all strings that contain at least one x and S_2 is the set of strings that do not contain x. Hence we have

$$|S| = |S_1| + |S_2| \tag{1}$$

Each string in S and S_2 can be constructed using the following four steps. In Step $i, 1 \leq i \leq 4$, we choose the letter in the ith location of the string.

While constructing a string in S each of the four steps can be performed in 26 ways. While constructing a string in S_2 each of the four steps can be performed in 25 ways. Thus $|S| = 26^4$ and $|S_2| = 25^4$. Substituting these values in equation (1) we get

$$|S_1| = 26^4 - 25^4 = 66351$$

Incorrect Solution. Here is an incorrect solution. Can you figure out what is wrong?

A four letter string that contains x can be constructed in two steps as follows. In Step 1 we choose one of the four positions for x (4 ways of doing this). In Step 2 we choose three letters for the remaining three places (26^3 ways of doing this). By the multiplcation rule, there are $4 \cdot 26^3 = 70304$ four letter strings that contain x.

Example. How many even 4-digit numbers have no repeated digits?

Solution. Let S be the set of all 4-digit numbers with distinct digits. Let S_0 be a set that contains all 4-digit numbers with distinct digits that end in a zero. Let S_1 be the set of all 4-digit numbers with distinct digits that end in $2, 4, 6, 8$. Note that the sets S_0 and S_1 partition the set S and hence we have

$$|S| = |S_0| + |S_1|$$

The procedure for constructing a number in S_0 is as follows: in step 1, we choose the digit in position 4, in steps 2,3,4, we choose the digits in positions 1,2,3, respectively. There is only 1 way to do step 1, 9 ways to do step 2, 8 ways to do step 3, and 7 ways to do step 4. By the Multiplication Rule, $|S_0| = 1 \times 9 \times 8 \times 7 = 504$.

A number in S_1 can be constructed similarly and hence $|S_1| = 4 \times 8 \times 8 \times 7 = 1792$.

Hence, $|S| = 504 + 1792 = 2296$.

Combinations.

Let n and r be non-negative integers. An r-combination of a set of n elements means an unordered selection of r of the n elements of S. The symbol $\binom{n}{r}$ (read as "n choose r") denotes the number of r-combinations of a set of n elements. This is same as the number of subsets of size r that can be chosen from a set of n elements.

The following numbers can be verified easily.

$$\binom{n}{r} = \begin{cases} 0 & \text{if } r > n \\ 1 & \text{if } r = 0 \text{ or } r = n \\ n & \text{if } r = 1 \end{cases}$$

Do you see the distinction between a r-permutation and a r-combination? A r-permutation is an *ordered* selection of r elements, i.e., both, which r elements, as well as the order in which they are chosen are important. Two r-permutations are the same if the r elements chosen are the same and they are chosen in the same order. In contrast, in a r-combination, only the choice of r elements is important. The order in which the r elements are chosen is irrelevant. Two r-combinations are the same if they have the same r elements regardless of the orders of selection of these elements.

In general, what is the value of $\binom{n}{r}$, i.e., how many r-combinations are possible if we have a set of n distinct objects?
We will answer this question by giving an expression that relates $\binom{n}{r}$ and $P(n, r)$. A r-permutation can be obtained in two steps as follows.

Step 1. Choose r elements from the available n elements.

Step 2. Arrange the chosen r elements.

Step 1 can be performed in $\binom{n}{r}$ ways. Step 2 can be performed in $r!$ ways. By the multiplication rule, the total number of r-permutations is given by

$$P(n, r) = \binom{n}{r} \times r!$$

Rearranging the terms of the above equation we get

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

Example. We have a pool of 14 players from which 11 players must be chosen to play a cricket match? How many 11-member teams are possible?

Solution. The number of distinct 11-member teams is the same as the number of subsets of size 11 from the set of 14 players. This is given by

$$\binom{14}{11} = \frac{14!}{11!3!} = \frac{12 \times 13 \times 14}{1 \times 2 \times 3} = 364.$$

Example. Consider a set of twenty-five points, no three of which are collinear. How many straight lines do they determine? How many triangles do they determine?

Solution. Since no three points lie on a straight line, every two points determine a straight line. The number of straight lines equals the number of 2-combinations of a 25-element set. This is given by

$$\binom{25}{2} = \frac{25!}{2!23!} = \frac{24 \times 25}{1 \times 2} = 300.$$

Similarly every three points determine a triangle. Thus the number of triangles is given by

$$\binom{25}{3} = \frac{25!}{3!22!} = \frac{23 \times 24 \times 25}{1 \times 2 \times 3} = 2300.$$

Example. From a group of 8 women and 6 men, how many different committes consisting of 3 women and 2 men can be formed? What if 2 of the men are feuding and refuse to serve on the committee together?

Solution. The procedure of forming a committee of 3 women and 2 men is as follows.

Step 1. Choose the 3 women.
Step 2. Choose the 2 men

Step 1 can be done in $\binom{8}{3}$ ways. Step 2 can be done in $\binom{6}{2}$ ways. Using the multiplication rule, the total number of possible committees is $\binom{8}{3} \times \binom{6}{2} = 840$.

The second part of the question can be solved as follows. Let S_1 be the set of all possible committees that do not contain the two feuding men. Let S_2 be the the set of all possible committees that contain exactly one of the two feuding men. Clearly, the no. of possible committes that do not contain the two feuding men together equals $|S_1| + |S_2|$. Using the reasoning used in the first part of the question we get $|S_1| = \binom{8}{3} \times \binom{4}{2} = 336$ and $|S_2| = 2\binom{8}{3} \times \binom{4}{1} = 448$. Hence the total number of committees without the two feuding men together is $336 + 448 = 784$.

The answer to the second part could also be derived by finding the number of all possible committees and then subtracting the number of committees in which the two feuding men are together. There are $\binom{8}{3}\binom{6}{2} = 840$ committees in all out of which $\binom{8}{3}\binom{2}{2} = 56$ committees contain the two feuding men. Thus there are $840 - 56 = 784$ committees in all that have non-feuding men.

Example. There are 15 students enrolled in a course, but exactly 12 students attend on any given day. The classroom for the course has 25 distinct seats. How many different classroom seatings are possible?

Solution. A classroom seating can be constructed in two steps as follows.

Step 1. Choose 12 students out of 15 that are enrolled.

Step 2. Arrange 12 students in 25 distinct seats available.

Step 1 can be performed in $\binom{15}{12}$ ways. Step 2 can be performed in $P(25, 12)$ ways. By the multiplication rule, the number of different classroom seatings possible is given by

$$\binom{15}{12} \times P(25, 12) = \frac{15!}{12!3!} \times \frac{25!}{13!}$$

Example. How many 8-letter strings can be constructed by using the 26 letters of the alphabet if each string contains 3,4, or 5 vowels? There is no restriction on the number of occurrences of a letter in the string.

Solution. Let E be the set of 8-letter strings that contain at least 3 vowels. Let E_i be the set of 8-letter strings containing exactly i vowels.

An element of E_i, i.e., a 8-letter string with exactly i vowels, can be constructed using the following steps.

Step 1. Choose i locations out of the available 8 locations for vowels.

Step 2. Choose the vowels for each of the i locations.

Step 3. Choose the consonants for each of the remaining $8 - i$ locations.

Step 1 can be performed in $\binom{8}{i}$ ways. Step 2 can be performed in 5^i ways. Step 3 can be performed in 21^{8-i} ways. By the multiplication rule, the number of 8-letter strings with exactly i vowels is given by

$$|E_i| = \binom{8}{i} 5^i 21^{8-i}$$

Since the sets E_3, E_4, and E_5 partition the set E, by the addition rule we get

$$|E| = \sum_{i=3}^{5} |E_i| = \sum_{i=3}^{5} \binom{8}{i} 5^i 21^{8-i}$$

The following question was raised in class. What if we want to count all 8-letter strings with distinct letters that have 3, 4, or 5 vowels? In this case, the above procedure still applies. However, the number of ways of doing each step changes. Step 1 can be performed in $\binom{8}{i}$ ways. Step 2 can be performed in $P(5, i)$ ways. Step 3 can be performed in $P(21, 8 - i)$ ways. By the multiplication rule, the number of 8-letter strings with distinct letters that have exactly i vowels is given by

$$\binom{8}{i} P(5, i) P(21, 8 - i)$$

The total number of 8-letter strings with distinct letters that have 3, 4, or 5 vowels is

$$\sum_{i=3}^{5} \binom{8}{i} P(5, i) P(21, 8 - i)$$

Permutations of Multisets.

Let S be a multiset that consists of n objects of which

n_1 are of type 1 and indistinguishable from each other.
n_2 are of type 2 and indistinguishable from each other.
$\vdots$
n_k are of type k and indistinguishable from each other.

and suppose $n_1 + n_2 + \ldots + n_k = n$. What is the number of distinct permutations of the n objects in S?

A permutation of S can be constructed by the following k-step process:

Step 1. Choose n_1 places out of n places for type 1 objects.

Step 2. Choose n_2 places out of the remaining $n - n_1$ places for type 2 objects.

. . . . . .

Step k. Choose n_k places of the remaining unused places for type k objects.

By the multiplication rule, the total number of permutations of n objects in S is

$$\begin{aligned} & \binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-n_2-\cdots-n_{k-1}}{n_k} \\ = \; & \frac{n!}{n_1!(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdots\frac{n-n_1-n_2-\cdots-n_{k-1}}{n_k!(n-n_1-\cdots-n_k)!} \\ = \; & \frac{n!}{n_1!n_2!\cdots n_k!} \end{aligned}$$

Example. How many permutations are there of the word MISSISSIPPI?

Solution. We want to find the number of permutations of the multiset $\{1\cdot M, 4\cdot I, 4\cdot S, 2\cdot P\}$. Thus, $n = 11, n_1 = 1, n_2 = 4, n_3 = 4, n_4 = 2$. Then number of permutations is given by

$$\frac{n!}{n_1!n_2!n_3!n_4!} = \frac{11!}{1!4!4!2!}$$

Example. Consider n distinct objects and k bins labeled $B_1, B_2, \ldots, B_k$. How many ways are there to distribute the objects in the bins so that bin B_i receives n_i objects and $\sum_{i=1}^{k} n_i = n$?

Solution. A partition of n objects into k labeled bins, $B_1, B_2, \ldots, B_k$ such that bin B_i gets n_i objects can be constructed in k steps. Step $i, 1 \le i \le k$ chooses n_i objects that go in box B_i from the remaining objects. Step $i, 1 \le i \le k$ can be performed in $\binom{n-n_1-n_2-\cdots-n_{i-1}}{n_i}$

ways. By the multiplication rule, the total number of ways to achieve the required partition equals

$$\begin{aligned} & \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-n_2-\cdots-n_{k-1}}{n_k} \\ = \quad & \frac{n!}{n_1!n_2!\cdots n_k!} \end{aligned}$$

Another way of arriving at the solution is as follows. Let the distinct objects be numbered $1, 2, \ldots, n$. Consider the multiset $A = \{n_1 \cdot B_1, n_2 \cdot B_2, \ldots, n_k \cdot B_k\}$. The procedure of obtaining the required partition can be done in $k+1$ steps as follows. In Step 0, we obtain a permutation P of the multiset A. In step $i, 1 \leq i \leq k$, bin B_i gets the objects corresponding to the positions of 'B_i' in P.

Step 1 can be done in $\frac{n!}{n_1!n_2!\cdots n_k!}$ ways. There is exactly one way to do each of the remaining steps. Hence, by the multiplication rule, the required answer is

$$\frac{n!}{n_1!n_2!\cdots n_k!}$$

Example. In how many ways can eight distinct books be divided among three students if Bill gets four books and Sharon and Marian each get two books?

Solution. Such partition can be obtained in three steps.

Step 1. Choose 4 books for Bill out of the available 8 books.

Step 2. Choose 2 books for Sharon out of the remaining 4 books.

Step 3. Choose 2 books for Marian out of the remaining 2 books.

Step 1 can be performed in $\binom{8}{4}$ ways. Step 2 can be performed in $\binom{4}{2}$ ways. Step 3 can be performed in $\binom{2}{2} = 1$ way. By the multiplication rule, the total number of possible divisions is given by

$$\binom{8}{4}\binom{4}{2} = \frac{8!}{4!4!} \times \frac{4!}{2!2!} = 420.$$

---

r-Combinations with Repetition Allowed.

We have seen that there are $\binom{n}{r}$ ways of choosing r *distinct* elements from a set of n distinct elements. What if we allow elements to be repeated? In other words, we want to find the number of ways there are to choose a multiset of r elements from a multiset of n distinct elements with infinite copies of each of the n elements available?

The following method was suggested in class.

A multiset of r elements can be constructed in r steps as follows. In Step i, choose one of the n elements. Since each step can be done in n ways, there are n^r multisets of r elements. Is this correct? No, this is not correct. For example, let $S = \{a, b\}$. Suppose we want to find the number of 2-combinations of S with repetition allowed. Note that the above procedure would consider the sets $\{a, b\}$ and $\{b, a\}$ as different whereas they are the same multiset and should not be counted twice. Using the above solution we get the answer as 4, but the correct answer is 3. In other words, the above procedure gives incorrect answer as it pays attention to the order of the r elements. We give the correct solution below.

Think of the n elements of the set as categories formed using $n-1$ vertical bars (sticks). Then each multiset of size r can be represented as a string of $n-1$ vertical bars (to separate the n categories) and r crosses (to represent the r elements to be chosen). The number of crosses in each category represents the number of times the object represented by that category is chosen. Note that each multiset of size r (chosen from a multiset of n objects, with infinite copies of each object), corresponds to exactly one way to arrange the $n-1$ sticks and r crosses and for each arrangement of $n-1$ sticks and r crosses, there is exactly one multiset of size r. Thus the number of multisets of size r is the same as the number of permutations of the multiset $\{(n-1) \cdot |, r \cdot \times\}$. The number of strings of $n-1$ vertical bars and r crosses is the number of ways to choose r positions from the available $r+n-1$ positions. The r positions chosen will contain the crosses and the remaining positions will have the vertical bars. Thus the total number of possible ways to choose multisets of size r from a multiset of n objects with infinite copies of each object available is given by

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{(n-1)!r!}$$

Example. Consider 3 books: a computer science book, a math book, and a history book. Suppose the library has at least 6 copies of each of these books. How many ways are there to select 6 books?

Solution. The no of ways is $\binom{6+3-1}{6} = \frac{8!}{6!2!} = 28$.

Example. How many solutions are there to the equation $x_1+x_2+x_3+x_4 = 10$ if x_1, x_2, x_3, and x_4 are non-negative integers? What if each $x_i \geq 1$?

Solution. Think of x_1, x_2, x_3, and x_4 as categories in which we must place 10 $\times$'s. The number of $\times$'s in each category represents the value of the corresponding variable in the equation. The number of solutions is the number of 10 multisets of a 4-element set. This is given by

$$\binom{4+10-1}{10} = \binom{13}{10} = 286$$

If each $x_i \geq 1$, we put one $\times$ in each category to start with. Then we distribute the remaining 6 $\times$'s among the categories. Such a distribution can be represented by a string

of 3 vertical bars and 6 crosses. The number of such distributions are

$$\binom{6+3}{6} = \binom{9}{6} = 84$$

Example. What is the number of non-decreasing sequences of length 10 whose terms are taken from 1 through 25?

Solution. The procedure of constructing a non-decreasing sequence of length 10 using integers from 1 through 25 is as follows: in step 1, choose 10 numbers with repetition allowed, from $\{\infty \cdot 1, \infty \cdot 2, \ldots, \infty \cdot 25\}$, and in step 2, order the chosen numbers in non-decreasing order. Note that the number of ways to do Step 1 is the same as the number of permutations of the multiset $\{24 \cdot |, 10 \cdot \times\}$, since we can think of the 25 digits as 25 categories (created using 24'|') in which 10 $\times$'s are to be placed. There is exactly one way to do step 2. Thus, the total number of ways that this can be done is given by $\binom{34}{10}$.

Example. How many ways are there to choose a 5-letter strings from the 26-letter English alphabet with replacement, where strings that are anagrams are considered the same?

Solution. Let S be the set of all 5-letter strings such that if a string is in S then its anagrams are not in S. We are interested in finding $|S|$. Note that two words are anagrams of each other iff the number of occurrences of each letter in the alphabet is the same same in both words. Thus the $|S|$ is the same as the number of 5-combinations with repetitions allowed from a multiset $\{\infty \cdot a, \infty \cdot b, \infty \cdot c, \ldots, \infty \cdot z\}$. Thus $|S| = \binom{26+5-1}{5} = \binom{30}{5}$.

Example. There are 15 identical customers and 4 distinct cashiers. How many ways can the customers line up to the cashiers?

Solution. This problem can be solved using the sticks and crosses method in which $n = 4$ and $r = 15$. Thus the answer is $\binom{4+15-1}{3} = \binom{18}{3}$

Combinatorial Proofs

Example. Prove that

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \ldots + (-1)^n \binom{n}{n} = 0$$

Solution. We want to prove that

$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \ldots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \ldots$$

Consider a set $X = \{x_1, x_2, x_3, \ldots, x_n\}$. We want to show that the total number of subsets of X that have even size equals the total number of subsets of X that have odd size. We

will now show that both these quantities equal 2^{n-1} from which the claim follows. An even-sized subset of X can be constructed as follows.

Step 1. Decide whether x_1 belongs to the subset of not.
Step 2. Decide whether x_2 belongs to the subset of not.
$\vdots$
Step n. Decide whether x_n belongs to the subset of not.

Note that there are 2 choices for each of the first $n-1$ steps but exactly one choice for performing step n. This is because if we have choose an even number of elements from $X \setminus \{x_n\}$ then we must leave out x_n otherwise we must include x_n in the subset. Hence using the multiplication rule the total number of even-sized subsets of X equals 2^{n-1}. Since we know that the total number of subsets of X is 2^n, the total number of odd-sized subsets of X is $2^n - 2^{n-1} = 2^{n-1}$.

We will prove a few identities using counting techniques. Specifically, we will use the following technique. To prove an identity we will pose a counting question. We will then answer the question in two ways, one answer will correspond to LHS and the other would correspond to the RHS. Since both answers are to the same question, the two answers must be the same.

Example. Show that $\binom{n}{r} = \binom{n}{n-r}$.

Solution. We can of course prove it algebrically. However, here is a combinatorial argument which provides more intuition. Observe that for every set of r elements that is chosen there is exactly one set of $n-r$ elements that is not chosen. Thus if a set A has k subsets of size r: $B_1, B_2, \ldots, B_k$ then each B_i can be paired up with exactly one set of size $n-r$, namely its complement $A \setminus B_i$. Hence the number of subsets of sizaue r is same as the number of subsets of size $n-r$.

We can also prove it by answering the following counting question in two different ways.

Given a set S of n distinct elements how many r-subsets are there of the set S?

Clearly, one answer is $\binom{n}{r}$, which gives us the left hand side. Another way to solve the problem is as follows. The procedure of forming a r-subset is as follows.

Step 1: Choose the $n-r$ elements that we want to leave out.
Step 2: Include the remaining r elements in the set.

There are $\binom{n}{n-r}$ ways to do step1 and exactly one way to do step 2. Hence, by the multiplication rule, the total number of ways of chooseing r-subsets of S is $\binom{n}{n-r}$, which gives us the right hand side.

Pascal's Formula. If n and k are positive integers with $n \geq k$ then

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

Proof. We will prove the claim by answering the following counting question in two different ways.

> Given a set $X = \{x_1, x_2, \ldots, x_n\}$ of n distinct elements how many k-subsets are there of the set X?

Let S be the set of all possible k-subsets of X. Clearly, $|S| = \binom{n}{k}$, which gives us the left hand side of the claim. Another way to find $|S|$ is as follows. The set S can be partitioned into sets S_1 and S_2, where S_1 is the set of all possible k-subsets of X that contain the element x_n and S_2 is the set of all possible k-subsets of X that do not contain the element x_n. In any k-subset of X that is in S_1, the other $k-1$ elements (since x_n is already in the subset) come from $X \setminus \{x_n\}$. Since there are $\binom{n-1}{k-1}$ ways of choosing these subsets, $|S_1| = \binom{n-1}{k-1}$. The k elements of any set in S_2 must be chosen from $X \setminus \{x_n\}$. There are $\binom{n-1}{k}$ ways of doing this. Since S_1 and S_2 partition the set S, we have

$$|S| = |S_1| + |S_2| = \binom{n-1}{k-1} + \binom{n-1}{k}$$

This gives us the right hand side of the claim.

Example. Prove that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

Solution. We pose the following counting question.

> Given a set S of n distinct elements how many subsets are there of the set S?

From earlier lectures, we know that the answer is 2^n. This gives us the RHS.

Another way to compute the answer to the question is as follows. The power set $\mathcal{P}(S)$ containing all possible subsets can be partitioned into $S_0, S_1, \ldots, S_n$, where S_i, $0 \leq i \leq n$, is the set of all subsets of S that have cardinality i. Thus

$$\begin{aligned} |\mathcal{P}(S)| &= |S_0| + |S_1| + \ldots + |S_n| \\ &= \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n} \\ &= \sum_{k=0}^{n} \binom{n}{k} = \text{LHS} \end{aligned}$$

This proves the claim.

Example. Prove that $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.

Solution. We pose the following counting question.

> How many ways are there to choose two numbers from $S = \{0, 1, 2, \ldots, n\}$?

By definition, there are $\binom{n+1}{2} = \frac{n(n+1)}{2}$ distinct pairs of S. This gives us the RHS.

We can also compute the answer as follows. Let P be the set of all pairs of S. P can be partitioned into $S_1, S_2, \ldots, S_n$, where S_i, $1 \leq i \leq n$, is the set of pairs in which i is the bigger element in the pair. Clearly,

$$\begin{aligned} |P| &= |S_1| + |S_2| + \ldots + |S_n| \\ &= 1 + 2 + \ldots + n \\ &= \sum_{k=1}^{n} k = \text{LHS} \end{aligned}$$

This proves the claim.

Example. Give a combinatorial proof to show that

$$\sum_{k=0}^{r} \binom{n}{k}\binom{m}{r-k} = \binom{n+m}{r}$$

Solution. We pose the following counting question.

> There are n men and m women, where $n \geq r$ and $m \geq r$. How many ways are there to form a committee of r people from this group of people?

By definition, there are $\binom{n+m}{r}$ distinct committees of r people. This gives us the RHS.

The set S of all possible committees of r people can be partitioned into subsets $S_0, S_1, S_2, \ldots, S_r$, where S_k is the set of committees in which there are exactly k men and the rest $r-k$ are women. Note that $|S_k| = \binom{n}{k}\binom{m}{r-k}$. Thus we have

$$\begin{aligned} |S| &= \sum_{k=0}^{r} |S_k| \\ &= \sum_{k=0}^{r} \binom{n}{k}\binom{m}{r-k} \end{aligned}$$

which gives us the left hand side of the expression.

The Binomial Theorem

A binomial is a sum of two terms, such as $a+b$. The *binomial theorem* gives an expression for $(a+b)^n$ where a and b are real numbers and n is a positive integer.

Theorem. For any real numbers a and b and non-negative integer n

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

Proof. Observe that each term in the expansion of $(a+b)^n$ is of the form $a^{n-k}b^k, k = 0, 1, 2, \ldots, n$. How many terms are there of the form $a^{n-k}b^k$? This is the same number of times as there are orderings of $n-k$ a's and k b's. This is equal to $\binom{n}{k}$. Thus the coefficient of like terms of the form $a^{n-k}b^k$ is $\binom{n}{k}$. This proves the theorem.

Example. Prove that $2^n = \sum_{k=0}^{n} \binom{n}{k}$

Solution. Last week we proved this claim using a counting argument in which we showed that L.H.S. and R.H.S. count the number of subsets of a set of n elements. Now we will prove this using the binomial theorem as follows.

$$\begin{aligned} 2^n &= (1+1)^n \\ &= \sum_{k=0}^{n} \binom{n}{k} (1)^{n-k}(1)^k \\ &= \sum_{k=0}^{n} \binom{n}{k} \\ &= \text{R.H.S.} \end{aligned}$$

Example. Let n be a positive integer. Then, for all x prove that $(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$.

Solution. Using the binomial theorem we get

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} 1^{n-k} x^k = \sum_{k=0}^{n} \binom{n}{k} x^k$$

Example. Prove that

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \ldots + (-1)^n \binom{n}{n} = 0$$

Solution. One way to solve this problem is by substituting $x = -1$ in the previous example. When $x = -1$ the above equation becomes

$$0^n = 0 = \sum_{k=0}^{n} \binom{n}{k} (-1)^k.$$

A combinatorial proof of the claim was presented earlier.

The Pigeonhole Principle

If $k+1$ or more objects are distributed among k bins then there is at least one bin that has two or more objects. For example, the pigeon hole principle can be used to conclude that in any group of thirteen people there are at least two who are born in the same month.

Example. There are n pairs of socks. How many socks must you pick without looking to ensure that you have at least one matching pair?

Solution. The pigeonhole principle can be applied by letting n bins correspond to the n pairs of socks. If we select $n+1$ socks and put each one in the box corresponding to the pair it belongs to then there must be at least one box containing a matched pair.

The Generalized Pigeonhole Principle

If n objects are placed into k boxes, then there is at least one box containing at least $\lceil n/k \rceil$ objects.

Proof: We will prove the contrapositive. That is, we will show that if each box contains at most $\lceil n/k \rceil - 1$ objects then the total number of objects is not equal to n. Assume that each box contains at most $\lceil n/k \rceil - 1$ objects. Then, the total number of objects is at most

$$k\left(\left\lceil \frac{n}{k} \right\rceil - 1\right) < k\left(\frac{n}{k} + 1 - 1\right) = n$$

Thus we have shown that the total number of objects is less than n. This completes the proof.

Using the generalized pigeonhole principle we can conclude that among 100 people, there are at least $\lceil 100/12 \rceil = 9$ who are born in the same month.

Example. Suppose each point in the plane is colored either red or blue. Show that there always exist two points of the same color that are exactly one feet apart.

Solution. Consider an equilateral triangle with the length of each side being one feet. The three corners of the triangle are colored red or blue. By pigeonhole principle, two of these three points must have the same color.

Example. Given a sequence of n integers, show that there exists a subsequence of consecutive integers whose sum is a multiple of n.

Solution. Let $x_1, x_2, \ldots, x_n$ be the sequence of n integers. Consider the following n sums.

$$x_1, x_1 + x_2, x_1 + x_2 + x_3, \ldots, x_1 + x_2 + \cdots + x_n$$

If any of these n sums is divisible by n, then we are done. Otherwise, each of the n sums have a non-zero remainder when divided by n. There are at most $n-1$ different possible remainders: $1, 2, \ldots n-1$. Since there are n sums, by the pigeonhole principle, at least two

of the n sums have the same remainder when divided by n. Let p and q, $p < q$, be integers such that for some integers c_1 and c_2,

$$x_1 + x_2 + \cdots + x_p = c_1 n + r \text{ and } x_1 + x_2 + \cdots + x_q = c_2 n + r$$

Subtracting the two sums, we get

$$x_{p+1} + \cdots + x_q = (c_2 - c_1)n$$

Hence, $x_{p+1} + \cdots + x_q$ is divisible by n.

Example. Show that in any group of six people there are either three mutual friends or three mutual strangers.

Solution. Consider one of the six people, say A. The remaining five people are either friends of A or they do not know A. By the pigeonhole principle, at least $\lceil 5/2 \rceil = 3$ of the five people are either friends of A or are unacquainted with A. In the former case, if any two of the three people are friends then these two along with A would be mutual friends, otherwise the three people would be strangers to each other. The proof for the latter case, when three or more people are unacquainted with A, proceeds in the same manner.

Example. A chess master who has 11 weeks to prepare for a tournament decides to play at least one game every day but, in order not to tire himself, he decides not to play more than 12 games during any calendar week. Show that there exists consecutive days during which the chess master will have played exactly 21 games.

Solution. Let a_i, $1 \leq i \leq 77$, be the total number of games that the chess master has played during the first i days. Note that the sequence of numbers $a_1, a_2, \ldots, a_{77}$ is a strictly increasing sequence. We have

$$1 \leq a_1 < a_2 < \ldots < a_{77} \leq 11 \times 12 = 132$$

Now consider the sequence $a_1 + 21, a_2 + 21, \ldots, a_{77} + 21$. We have

$$22 \leq a_1 + 21 < a_2 + 21 < \ldots < a_{77} + 21 \leq 153$$

Clearly, this sequence is also a strictly increasing sequence. The numbers $a_1, a_2, \ldots, a_{77}, a_1 + 21, a_2 + 21, \ldots, a_{77} + 21$ (154 in all) belong to the set $\{1, 2, \ldots, 153\}$. By the pigeonhole principle there must be two numbers out of the 154 numbers that must be the same. Since no two numbers in $a_1, a_2, \ldots, a_{77}$ are equal and no two numbers in $a_1 + 21, a_2 + 21, \ldots, a_{77} + 21$ are equal there must exist i and j such that $a_i = a_j + 21$. Hence during the days $j+1, j+2, \ldots, i$, exactly 21 games must have been played.

Benjamin Judd suggested the following nice proof in class. For $1 \leq i \leq 77$, let g_i denote the number of games played by the chessmaster on day i. Consider the number of games

played by the chessmaster during each day of the first three weeks: $g_1, g_2, \ldots, g_{21}$. By the constraints described in the question, we have

$$g_i \geq 1, i = 1, 2, \ldots, 21 \text{ and } \sum_{i=1}^{21} g_i \leq 36 \tag{2}$$

We know that in the sequence of positive integers $g_1, g_2, \ldots, g_{21}$, there must be a subsequence $S : g_l, g_{l+1}, g_{l+2}, \ldots, g_k, 1 \leq l < k \leq 21$ of consecutive integers whose sum is divisible by 21 (we proved this earlier in the lecture). Combining this with (2), we conclude that the sum of the numbers in S must be exactly 21. This means that during the days $l, l+1, l+2, \ldots, k$, the chessmaster played exactly 21 games.

Example. Prove that every sequence of $n^2 + 1$ distinct real numbers, $x_1, x_2, \ldots, x_{n^2+1}$, contains a subsequence of length $n+1$ that is either strictly increasing or strictly decreasing.

Solution. We will prove this as follows. We suppose that there is no strictly increasing subsequence of length $n+1$ and show that there must be a strictly decreasing subsequence of length $n+1$. Let $m_k, k = 1, 2, \ldots, n^2+1$, be the length of the longest increasing subsequence that begins with x_k. Since there is no increasing subsequence of length $n+1$, for $k = 1, 2, \ldots, n^2+1$, we have $1 \leq m_k \leq n$.

Solution. We will prove this as follows. We suppose that there is no strictly increasing subsequence of length $n+1$ and show that there must be a strictly decreasing subsequence of length $n+1$. Let $m_k, k = 1, 2, \ldots, n^2+1$, be the length of the longest increasing subsequence that begins with x_k. Since there is no increasing subsequence of length $n+1$, for $k = 1, 2, \ldots, n^2+1$, we have $1 \leq m_k \leq n$. Using the generalized pigeonhole principle, we conclude that $n+1$ of the numbers $m_1, m_2, \ldots, m_{n^2+1}$ are equal. Let

$$m_{k_1} = m_{k_2} = \cdots = m_{k_{n+1}}$$

where $1 \leq k_1 < k_2 < \cdots < k_{n+1} \leq n^2+1$. We will now argue that $x_{k_1} > x_{k_2} > \cdots > x_{k_{n+1}}$, which will complete the proof as we will have a decreasing subsequence of length $n+1$. Assume for contradiction that this is not the case, which means that there is a i, $1 \leq i \leq n+1$, such that $x_{k_i} < x_{k_{i+1}}$. Then, since $k_i < k_{i+1}$, we could take a longest increasing subsequence starting with $x_{k_{i+1}}$ and put x_{k_i} in front to obtain an increasing subsequence that begins with x_{k_i}. This implies that $m_{k_i} > m_{k_{i+1}}$, which is a contradiction. Hence, for all $i = 1, 2, \ldots, n, x_{k_i} > x_{k_{i+1}}$. Thus, we have a decreasing subsequence of length $n+1$. Similarly, we can show that if there is no decreasing subsequence of length $n+1$ then there must be an increasing sequence of length $n+1$.

Introduction to Probability

Probability theory has many applications in engineering, medicine, etc. It has also found many useful applications in computer science, such as cryptography, networking, game theory etc. Many algorithms are randomized and we need probability theory to analyze them. In this course, our goal is to understand how to describe uncertainty using probabilistic arguments. To do this we first have to define a probabilistic model.

A probabilistic model is a mathematical description of a random process or an experiment. In a random process exactly one outcome from a set of outcomes is sure to occur but no outcome can be predicted with certainty. For example, tossing a coin is an experiment. Below are definitions of entities associated with the probabilistic model.

- The *sample space* of a random process or experiment is the set of all possible outcomes. The sample space is often denoted by Ω. Since we are going to study discrete probability Ω will be finite or countably infinite (such as integers and not real numbers).
- The *probability space* is a sample space together with a *probability distribution* in which a probability is assigned to each outcome $\omega \in \Omega$, such that
  - $0 \leq \Pr[\omega] \leq 1$
  - $\sum_{\omega \in \Omega} \Pr[\omega] = 1$

In an experiment we are usually interested in the probability with an event occurs. For example, when tossing a coin we may be interested in knowing the probability that the result is heads. Below we define formally what an event is and what does it mean to calculate the probability of an event.

- A subset of the sample space is called an *event*.
- For any event, $A \subseteq \Omega$, the probability of A is defined as

$$\Pr[A] = \sum_{\omega \in A} \Pr[\omega]$$

We are now ready to work through some problems. Before we proceed, keep in mind that probability is a slippery topic; it is very easy to make mistakes. Solving the problem systematically is the key to avoid mistakes. The following four-step process that is described in the notes by Lehman and Leighton is a way to systematically solve the problem at hand.

(a) Define the sample space, Ω, of the experiment, i.e., find the set of all possible outcomes of the experiment.

(b) Define the probability distribution.

(c) Find the event of interest, A, i.e., find the subset of outcomes, $A \subseteq \Omega$ that are of interest.

(d) Compute the probability of A by adding up the probabilities of the outcomes in A.

Example. On flipping a fair coin what is the probability that the result is heads?

Solution. $\Omega = \{H, T\}, \Pr[H] = \Pr[T] = 1/2, A = \{H\}, \Pr[A] = 1/2.$

Example. Consider a biased coin in which the probability of heads is 1/3. Suppose we flip the coin twice. What is the probability that we obtain one tails and one heads?

Solution. $\Omega = \{HH, HT, TH, TT\}$. The probability distribution is given by

$$\begin{aligned} \Pr[HH] &= \frac{1}{3} \times \frac{1}{3} = \frac{1}{9} \\ \Pr[HT] &= \frac{1}{3} \times \frac{2}{3} = \frac{2}{9} \\ \Pr[TH] &= \frac{2}{3} \times \frac{1}{3} = \frac{2}{9} \\ \Pr[TT] &= \frac{2}{3} \times \frac{2}{3} = \frac{4}{9} \end{aligned}$$

Note that the assigned probabilities form a valid probability distribution. Event $A = \{HT, TH\}$. The probability of the event A is given by

$$\Pr[A] = \Pr[HT] + \Pr[TH] = \frac{4}{9}$$

Example. We roll two dice. Compute the probability that the two numbers are equal when (i) two dice are distinct, (ii) the dice are indistinguishable.

Solution. (a) Each outcome of the experiment can be denoted by an ordered pair $(\omega_1, \omega_2), 1 \leq \omega_1, \omega_2 \leq 6$, where ω_1 and ω_2 are the numbers on dice 1 and dice 2 respectively. Note that $|\Omega| = 36$ and each outcome is equally likely. The event that the two numbers are equal is given by $A = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$. The probability that A occurs is given by

$$\Pr[A] = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$$

(b) When the die are indistinguishable, the order of the two numbers is not important, hence each outcome of the experiment can be denoted by a 2-set $\{\omega_1, \omega_2\}, 1 \leq \omega_1, \omega_2 \leq 6$, where ω_1 and ω_2 are the numbers on the two die. Note that $|\Omega| = 21$. Each outcome of the form $\{\omega_1, \omega_2\}, \omega_1 \neq \omega_2$ occurs with a probability of $\frac{2}{36} = \frac{1}{18}$ and outcomes of the form $\{\omega, \omega\}$ occur with the probability of $\frac{1}{36}$. The event that the two numbers are equal is given by $A = \{\{1,1\}, \{2,2\}, \{3,3\}, \{4,4\}, \{5,5\}, \{6,6\}\}$. The probability that A occurs is given by

$$\Pr[A] = 6 \times \frac{1}{36} = \frac{1}{6}$$

Example. Suppose we throw m distinct balls into n distinct bins. Assume that there is no bound on the number of balls that a bin contains. What is the probability that a particular bin, say bin 1, contains all the m balls?

Solution. Each outcome can be represented by a m-tuple $(\omega_1, \omega_2, \ldots, \omega_m)$, where ω_i denotes the bin that contains the ith ball. Note that $|\Omega| = n^m$ and each outcome is equally likely. Since there is only one way in which all balls can be in bin 1, the probability of this event is $\frac{1}{n^m}$.

Example. What is the probability of rolling a six-sided die six times and having all the numbers 1 through 6 result (in any order)?

Solution. Each element in Ω can be represented by $(\omega_1, \omega_2, \ldots, \omega_6)$, where ω_i is the number that results on the ith roll of the die. Using the multiplication rule we get $|\Omega| = 6^6$. Let $A \subseteq \Omega$ be the set of outcomes in which the numbers of the rolls are different. By multiplication rule $|A| = 6!$. Since each outcome is equally likely, the desired probability is given by

$$\frac{|A|}{|\Omega|} = \frac{6!}{6^6} = \frac{5}{324}$$

Example. On "Let's Make a Deal" show, there are three doors. There is a prize behind one of the doors and goats behind the other two. The contestant chooses a door. Then the host opens a different door behind which there is a goat. The contestant is then given a choice to either switch doors or to stay put. The contestant wins the prize if and only if the contestant chooses the door with the prize behind it. Is it to the contestant's benefit to switch doors?

Solution. Each outcome of the sample space can be denoted by a 3-tuple $(\omega_1, \omega_2, \omega_3)$, where ω_1 denotes the door hiding the prize, ω_2 denotes the door chosen by the contestant initially, and ω_3 is the door chosen by the host. Now, lets assign probabilities to each of the outcomes[1]. There are two types of outcomes, those in which $\omega_1 = \omega_2$ and those in which $\omega_1 \neq \omega_2$. It is easy to verify that there are 6 outcomes of each type. Each outcome of the first type occurs with a probability of $\frac{1}{3} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{18}$. If the outcome is of the second type then there is only one choice for ω_3, i.e., there is only one choice of door for the host. Each outcome of the second type occurs with a probability $\frac{1}{3} \times \frac{1}{3} \times 1 = \frac{1}{9}$. The event in which the contestant switches doors and wins is the set of all outcomes in which $\omega_1 \neq \omega_2$. Since the size of this set is 6 and each outcome occurs with a probability of $\frac{1}{9}$ the probability of the contestant winning the prize by switching is $\frac{6}{9} = \frac{2}{3}$. Thus, it is to contestant's benefit to switch.

Inclusion-Exclusion Formula

For two events A and B we have

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B].$$

[1]We are making the following assumptions:(i) the prize is equally likely to be behind any of the doors, (ii) the contestant is equally likely to choose any of the three doors, (iii) the host opens any of the possible doors with equal probability

For three events A, B, and C, we have

$$\Pr[A \cup B \cup C] = \Pr[A] + \Pr[B] + \Pr[C] - \Pr[A \cap B] - \Pr[B \cap C] - \Pr[A \cap C] + \Pr[A \cap B \cap C].$$

For events $A_1, A_2, \ldots, A_n$ in some probability space, let $S_1 = \{(i_1) | 1 \leq i_1 \leq n\}$, $S_2 = \{(i_1, i_2) | 1 \leq i_1 < i_2 \leq n\}$, and more generally let $S_p = \{(i_1, i_2, \ldots, i_p) | 1 \leq i_1 < i_2 < \ldots \leq i_p \leq n\}$. Then we have

$$\Pr[\cup_{i=1}^{n} A_i] = \sum_{i \in S_1} \Pr[A_i] - \sum_{(i_1,i_2) \in S_2} \Pr[A_{i_1} \cap A_{i_2}] + \sum_{(i_1,i_2,i_3) \in S_3} \Pr[A_{i_1} \cap A_{i_2} \cap A_{i_3}] - \cdots + (-1)^{n-1} \Pr[\cap_{x=1}^{n} A_x]$$

Note that there are $2^n - 1$ non-empty subsets of a set of n events. To compute the probability of the intersection of every such subset is not possible when n is large. In such cases we have to approximate the probability of a union of n events. The successive terms of the above formula actually give an overestimate and underestimate respectively of the actual probability. In many situations the upper-bound given by the first term itself is quite useful. It is called the *union-bound* and is given by

$$\Pr[\cup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} \Pr[A_i]$$

Note that when the events are pairwise disjoint, the inequality in the above expression becomes an equality.

Example. Consider three flips of a fair coin. What is the probability that result is heads on the first flip or the third flip?

Solution. Let H_1 and H_2 denote the events that the first flip results in heads and the third flip results in heads respectively. By the inclusion-exclusion formula, we have

$$\begin{aligned} \Pr[H_1 \cup H_2] &= \Pr[H_1] + \Pr[H_2] - \Pr[H_1 \cap H_2] \\ &= \frac{1}{2} + \frac{1}{2} - \frac{1}{4} \\ &= \frac{3}{4} \end{aligned}$$

Example. When three dice are rolled what is the probability that one of the dice results in 4?

Solution. Let $F_i, i \in \{1, 2, 3\}$ be the event that the ith dice results in a 4. We are interested in $\Pr[F_1 \cup F_2 \cup F_3]$. By inclusion-exclusion formula we have

$$\Pr[F_1 \cup F_2 \cup F_3] = \Pr[F_1] + \Pr[F_2] + \Pr[F_3] - \Pr[F_1 \cap F_2] - \Pr[F1 \cap F_3] - \Pr[F_2 \cap F_3] + \Pr[F_1 \cap F_2 \cap F_3]$$

Since the events F_1, F_2, F_3 are mutually independent we can rewrite the above expression as

$$\begin{aligned}\Pr[F_{\cup}F_2 \cup F_3] &= \Pr[F_1] + \Pr[F_2] + \Pr[F_3] - \Pr[F_1]\Pr[F_2] - \Pr[F1]\Pr[F_3] - \Pr[F_2]\Pr[F_3] \\ &\quad + \Pr[F_1]\Pr[F_2]\Pr[F_3] \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} - \left(\frac{1}{6} \times \frac{1}{6}\right) - \left(\frac{1}{6} \times \frac{1}{6}\right) - \left(\frac{1}{6} \times \frac{1}{6}\right) + \left(\frac{1}{6} \times \frac{1}{6} \times \frac{1}{6}\right) \\ &= \frac{91}{216}\end{aligned}$$

An easier way to solve this is as follows. Let $\overline{F_i}$ be the complement of event F_i, $i = 1, 2, 3$.

$$\Pr[F_1 \cup F_2 \cup F_3] = 1 - \Pr[\overline{F_1} \cap \overline{F_2} \cap \overline{F_3}] = 1 - (5/6)^3 = \frac{91}{216}$$

Example. A coin is tossed 10 times. What is the probability that eight or more heads turn up?

Solution. Let E_i denote the event that exactly i heads turn up. We are interested in $\Pr[E_8 \cup E_9 \cup E_{10}]$. Since the events E_i are disjoint, we have

$$\Pr[E_8 \cup E_9 \cup E_{10}] = \Pr[E_8] + \Pr[E_9] + \Pr[E_{10}]$$

Note that $\Pr[E_i] = \binom{10}{i}/2^{10}$. Hence, we have

$$\Pr[E_8 \cup E_9 \cup E_{10}] = \frac{1}{2^{10}}\left(\binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right) = \frac{56}{2^{10}}$$

Example. (Birthday Paradox) Suppose there are k people in a room and n days in a year. We are interested in the probability that there are at least two people in the room with the same birthday. What is the smallest value of k for which this probability is at least 1/2? Assume that it is equally likely for a person to be born on any of the n days of the year.

Solution. Let B be the event that at least two people in the room have the same birthday. We are interested in $\Pr[B]$.

$$\begin{aligned}\Pr[B] &= 1 - \Pr[\overline{B}] \\ &= 1 - \frac{P(n,k)}{n^k}\end{aligned}$$

For $n = 365$, the smallest value of k for which the RHS is at least 1/2 is 23. If $k = 40$ then $\Pr[B] = 0.89$, and if $k = 60$ then $\Pr[B] = 0.994$. This means that if there are 60 people then it is almost certain that there exists two among them sharing the same birthday. To illustrate how good our model is, consider the set of presidents of the United States of America. Through Bill Clinton, 41 people belong to this set. The chances of two of them sharing the same birthday is at least 89%. Indeed, James Polk (11th president) and Warren Harding (29th president) are both born on Nov. 2.

Conditional Probability

We now introduce a very important concept of conditional probability. Conditional probability allows us to calculate the probability of an event when some partial information about the result of an experiment is known. As we shall see conditional probability is often a convenient way to calculate probabilities even when no information about the result of an experiment is available.

Suppose we want to calculate the probability of event A given that event B has already occured. We denote this by $\Pr[A|B]$ (read as "the probability of A given B"). Since we know that event B has occured our sample space reduces to the outcomes in B. Is this a valid probability space? No, because the sum of probabilities of the outcomes in B is less than 1. How do we change the probabilities so that this is a valid probability distribution while making sure that the relative probabilities of outcomes in B do not change? We do this by scaling the probability of all sample points in B by $\frac{1}{\Pr[B]}$. Thus for each sample point $\omega \in B$,

$$\Pr[\omega|B] = \frac{\Pr[\omega]}{\Pr[B]}$$

To calculate $\Pr[A|B]$ we just sum up the probabilities of sample points in $A \cap B$. Thus we get

$$\Pr[A|B] = \sum_{\omega \in A \cap B} \Pr[\omega|B] = \sum_{\omega \in A \cap B} \frac{\Pr[\omega]}{\Pr[B]} = \frac{\Pr[A \cap B]}{\Pr[B]}$$

In order to avoid division by 0, we only define $\Pr[A|B]$ when $\Pr[B] > 0$. Conditional probabilities can sometimes get tricky. To avoid pitfalls, it is best to use the above mathematical definition of conditional probability. Note that the R.H.S. of the above equation are unconditional probabilities.

Example. Suppose we flip two fair coins. What is the probability that both tosses give heads given that one of the flips results in heads? What is the probability that both tosses give heads given that the first coin results in heads?

Example. Suppose we flip two fair coins. What is the probability that both tosses give heads given that one of the flips results in heads? What is the probability that both tosses give heads given that the first coin results in heads?

Solution. We consider the following events to answer the question.

A: event that both flips give heads.
B: event that one of the flips gives heads.
C: event that the first coin flip gives heads.

Let's first calculate $\Pr[A|B]$.

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{\Pr[A]}{\Pr[B]} = \frac{1/4}{3/4} = \frac{1}{3}.$$

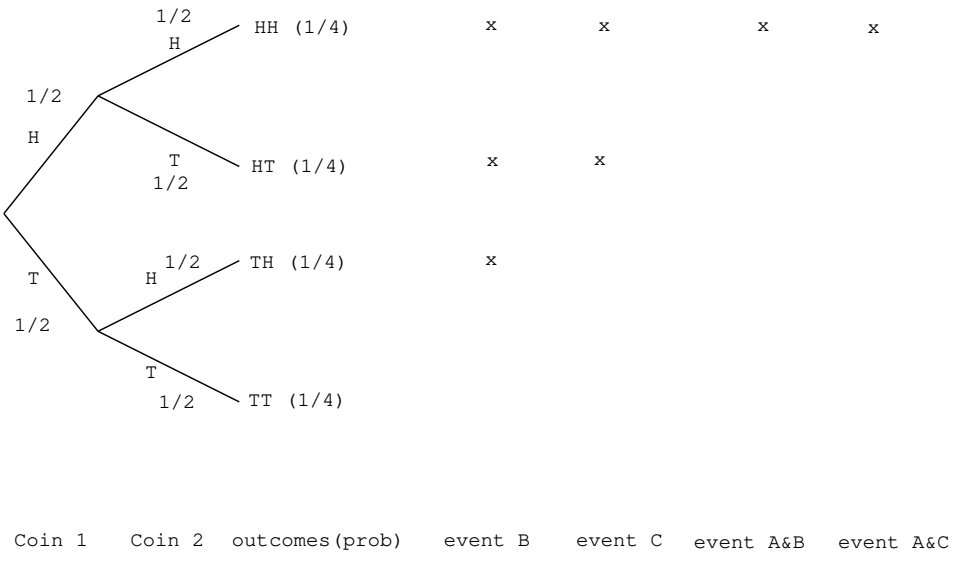


Figure 3: Tree diagram for the experiment in Example 1.

Similarly we can calculate $\Pr[A|C]$ as follows.

$$\Pr[A|C] = \frac{\Pr[A \cap C]}{\Pr[C]} = \frac{\Pr[A]}{\Pr[C]} = \frac{1/4}{1/2} = \frac{1}{2}.$$

The above analysis also follows from the tree diagram in Figure 3.

The Multiplication Rule. For any two events A_1 and A_2 we have

$$\Pr[A_1 \cap A_2] = \Pr[A_1] \cdot \Pr[A_2|A_1]$$

The above formula follows from the definition of $\Pr[A_2|A_1]$. This formula can be generalized to n events. We state the generalization without proof.

$$\Pr[A_1 \cap A_2 \cap \cdots \cap A_n] = \Pr[A_1] \cdot \Pr[A_2|A_1] \cdot \Pr[A_3|A_1 \cap A_2] \cdots \Pr[A_n|A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_{n-1}]$$

Example. The probability that a new car battery functions for over 10,000 miles is 0.8, the probability that it functions for over 20,000 miles is 0.4, and the probability that it functions for over 30,000 miles is 0.1. If a new car battery is still working after 10,000 miles, what is the probability that (i) its total life will exceed 20,000 miles, (ii) its additional life will exceed 20,000 miles?

Solution. We will consider the following events to answer the question.

L_{10}: event that the battery lasts for more than 10K miles.
L_{20}: event that the battery lasts for more than 20K miles.
L_{30}: event that the battery lasts for more than 30K miles.

We know that $\Pr[L_{10}] = 0.8$, $\Pr[L_{20}] = 0.4$ and $\Pr[L_{30}] = 0.1$. We are interested in calculating $\Pr[L_{20}|L_{10}]$ and $\Pr[L_{30}|L_{10}]$.

$$\begin{aligned}\Pr[L_{20}|L_{10}] &= \frac{\Pr[L_{20} \cap L_{10}]}{\Pr[L_{10}]} \\ &= \frac{\Pr[L_{20}] \cdot \Pr[L_{10}|L_{20}]}{0.8} \\ &= \frac{0.4 \times 1}{0.8} \\ &= \frac{1}{2}\end{aligned}$$

By doing similar calculations it is easy to verify that $\Pr[L_{30}|L_{10}] = \frac{1}{8}$.

Example. An urn initially contains 5 white balls and 7 black balls. Each time a ball is selected, its color is noted and it is replaced in the urn along with two other balls of the same color. Compute the probability that the first two balls selected are black and the next two white.

Solution. We will consider the following events to answer the question.

B_1: event that the first ball chosen is black.
B_2: event that the second ball chosen is black.
W_3: event that the third ball chosen is white.
W_4: event that the fourth ball chosen is white.

We are interested in calculating $\Pr[B_1 \cap B_2 \cap W_3 \cap W_4]$. Using the Multiplication rule we get,

$$\begin{aligned}\Pr[B_1 \cap B_2 \cap W_3 \cap W_4] &= \Pr[B_1] \cdot \Pr[B_2|B_1] \cdot \Pr[W_3|B_1 \cap B_2] \cdot \Pr[W_4|B_1 \cap B_2 \cap W_3] \\ &= \frac{7}{12} \times \frac{9}{14} \times \frac{5}{16} \times \frac{7}{18} \\ &= \frac{35}{768}\end{aligned}$$

The Total Probability Theorem. Consider events E and F. Consider a sample point $\omega \in E$. Observe that ω belongs to either F or $\overline{F}$. Thus, the set E is a disjoint union of two sets: $E \cap F$ and $E \cap \overline{F}$. Hence we get

$$\begin{aligned}\Pr[E] &= \Pr[E \cap F] + \Pr[E \cap \overline{F}] \\ &= \Pr[F] \times \Pr[E|F] + \Pr[\overline{F}] \times \Pr[E|\overline{F}]\end{aligned}$$

In general, if $A_1, A_2, \ldots, A_n$ form a partition of the sample space and if $\forall i, \Pr[A_i] > 0$, then for any event B in the same probability space, we have

$$\Pr[B] = \sum_{i=1}^{n} \Pr[A_i \cap B] = \sum_{i=1}^{n} \Pr[A_i] \times \Pr[B|A_i]$$

Example. A medical test for a certain condition has arrived in the market. According to the case studies, when the test is performed on an affected person, the test comes up positive 95% of the times and yields a "false negative" 5% of the times. When the test is performed on a person not suffering from the medical condition the test comes up negative in 99% of the cases and yields a "false positive" in 1% of the cases. If 0.5% of the population actually have the condition, what is the probability that the person has the condition given that the test is positive?

Solution. We will consider the following events to answer the question.

C: event that the person tested has the medical condition.
$\overline{C}$: event that the person tested does not have the condition.
P: event that the person tested positive.

We are interested in $\Pr[C|P]$. From the definition of conditional probability and the total probability theorem we get

$$\begin{aligned}
\Pr[C|P] &= \frac{\Pr[C \cap P]}{\Pr[P]} \\
&= \frac{\Pr[C]\Pr[P|C]}{\Pr[P \cap C] + \Pr[P \cap \overline{C}]} \\
&= \frac{\Pr[C]\Pr[P|C]}{\Pr[C]\Pr[P|C] + \Pr[\overline{C}]\Pr[P|\overline{C}]} \\
&= \frac{0.005 \times 0.95}{0.005 \times 0.95 + 0.995 \times 0.01} \\
&= 0.323
\end{aligned}$$

This result means that 32.3% of the people who are tested positive actually suffer from the condition!

Example. A transmitter sends binary bits, 80% 0's and 20% 1's. When a 0 is sent, the receiver will detect it correctly 80% of the time. When a 1 is sent, the receiver will detect it correctly 90% of the time.
(a) What is the probability that a 1 is sent and a 1 is received?
(b) If a 1 is received, what is the probability that a 1 was sent?

Solution. We will consider the following events.

S_0: event that the transmitter sent a 0.
S_1: event that the transmitter sent a 1.
R_1: event that 1 was received.

(a) We are interested in finding $\Pr[S_1 \cap R_1]$.

$$\begin{aligned}
\Pr[S_1 \cap R_1] &= \Pr[S_1] \times \Pr[R_1|S_1] \\
&= 0.2 \times 0.9 \\
&= 0.18
\end{aligned}$$

(b) We are interested in finding $\Pr[S_1|R_1]$.

$$\begin{aligned}
\Pr[S_1|R_1] &= \frac{\Pr[S_1 \cap R_1]}{\Pr[R_1]} \\
&= \frac{0.18}{\Pr[R_1 \cap S_1] + \Pr[R_1 \cap S_0]} \\
&= \frac{0.18}{0.18 + \Pr[S_0] \times \Pr[R_1|S_0]} \\
&= \frac{0.18}{0.18 + 0.8 \times 0.2} \\
&= 0.5294
\end{aligned}$$

Example. An urn contains 5 white and 10 black balls. A fair die is rolled and that number of balls are chosen from the urn.
(a) What is the probability that all of the balls selected are white?
(b) What is the conditional probability that the die landed on 3 if all the balls selected are white?

Solution. We will consider the following events.

W: event that all of the balls chosen are white.
D_i: event that the die landed on i, $1 \leq i \leq 6$.

(a) We want to find $\Pr[W]$. We can do this as follows.

$$\begin{aligned}
\Pr[W] &= \sum_{i=1}^{6} \Pr[W \cap D_i] \\
&= \sum_{i=1}^{6} \Pr[D_i] \Pr[W|D_i] \\
&= \sum_{i=1}^{6} \frac{1}{6} \frac{\binom{5}{i}}{\binom{15}{i}} \\
&= \frac{1}{6}\left(\frac{5}{15} + \frac{10}{105} + \frac{10}{455} + \frac{5}{1365} + \frac{1}{3003}\right) \\
&= 0.075
\end{aligned}$$

(b) We want to find $\Pr[D_3|W]$. This can be done as follows.

$$\begin{aligned}
\Pr[D_3|W] &= \frac{\Pr[D_3 \cap W]}{\Pr[W]} \\
&= \frac{\Pr[D_3] \times \Pr[W|D_3]}{\Pr[W]} \\
&= \frac{1/6 \times \binom{5}{3}/\binom{15}{3}}{0.075} \\
&= \frac{1/6 \times 10/455}{0.075} \\
&= \frac{0.00366}{0.075} \\
&= 0.048
\end{aligned}$$

Independent Events. Two events A and B are *independent* if and only if $\Pr[A \cap B] = \Pr[A] \times \Pr[B]$. This definition also implies that if the conditional probability $\Pr[A|B]$ exists, then A and B are independent events if and only if $\Pr[A|B] = \Pr[A]$.

Events $A_1, A_2, \ldots, A_n$ are *mutually independent* if $\forall i, 1 \leq i \leq n$ A_i does not "depend" on any combination of the other events. More formally, for every subset $I \subseteq \{1, 2, \ldots, n\}$,

$$\Pr[\cap_{i \in I} A_i] = \prod_{i \in I} \Pr[A_i]$$

In other words, to show that $A_1, A_2, \ldots, A_n$ are mutually independent we must show that all of the following hold.

$$\begin{aligned}
\Pr[A_i \cap A_j] &= \Pr[A_i] \cdot \Pr[A_j] \quad \forall \text{ distinct } i, j \\
\Pr[A_i \cap A_j \cap A_k] &= \Pr[A_i] \cdot \Pr[A_j] \cdot \Pr[A_k] \quad \forall \text{ distinct } i, j, k \\
\Pr[A_i \cap A_j \cap A_k \cap A_l] &= \Pr[A_i] \cdot \Pr[A_j] \cdot \Pr[A_k] \cdot \Pr[A_l] \quad \forall \text{ distinct } i, j, k, l \\
&\ldots \\
\Pr[A_1 \cap A_2 \cap \cdots \cap A_n] &= \Pr[A_1] \Pr[A_2] \cdots \Pr[A_n]
\end{aligned}$$

The above definition implies that if $A_1, A_2, \ldots, A_n$ are mutually independent events then

$$\Pr[A_1 \cap A_2 \cap \ldots \cap A_n] = \Pr[A_1] \times \Pr[A_2] \times \cdots \times \Pr[A_n]$$

However, note that $\Pr[A_1 \cap A_2 \cap \ldots \cap A_n] = \Pr[A_1] \times \Pr[A_2] \times \cdots \times \Pr[A_n]$ is not a sufficient condition for $A_1, A_2, \ldots, A_n$ to be mutually independent.

Do not confuse the concept of disjoint events and independent events. If two events A and B are disjoint and have a non-zero probability of happening then given that one event happens reduces the chances of the other event happening to zero, i.e., $\Pr[A|B] = 0 \neq \Pr[A]$. Thus by definition of independence, events A and B are not independent.

Example. Two cards are sequentially drawn (without replacement) from a well-shuffled deck of 52 cards. Let A be the event that the two cards drawn have the same value (e.g. both 4s) and let B be the event that the first card drawn is an ace. Are these events independent?

Solution. To decide whether the two events are independent we need to check whether $\Pr[A \cap B] = \Pr[A]\Pr[B]$.

$$\begin{aligned}
\Pr[A] &= \frac{3}{51} = \frac{1}{17} \\
\Pr[B] &= \frac{4}{52} = \frac{1}{13} \\
\Pr[A \cap B] &= \frac{1}{13} \times \frac{3}{51} \\
&= \frac{1}{221} \\
&= \frac{1}{17} \times \frac{1}{13} \\
&= \Pr[A]\Pr[B]
\end{aligned}$$

Example. Suppose that a fair coin is tossed twice. Let A be the event that a head is obtained on the first toss, B be the event that a head is obtained on the second toss, and C be the event that either two heads or two tails are obtained. (a) Are events A, B, C pairwise independent? (b) Are they mutually independent?

Solution. Note that $\Omega = \{HH, HT, TH, TT\}$. $A = \{HH, HT\}$, $B = \{HH, TH\}$, $C = \{HH, TT\}$, $A \cap B = \{HH\}$, $A \cap C = \{HH\}$, $B \cap C = \{HH\}$, $A \cap B \cap C = \{HH\}$. The probabilities of the relevant events are as follows.

$$\begin{aligned}
\Pr[A] &= 1/2 \\
\Pr[B] &= 1/2 \\
\Pr[C] &= 1/2 \\
\Pr[A \cap B] &= 1/4 = \Pr[A]\Pr[B] \\
\Pr[A \cap C] &= 1/4 = \Pr[A]\Pr[C] \\
\Pr[B \cap C] &= 1/4 = \Pr[B]\Pr[C] \\
\Pr[A \cap B \cap C] &= 1/4 \neq \Pr[A]\Pr[B]\Pr[C]
\end{aligned}$$

Thus we see that A, B, C are pairwise independent but not mutually independent.

Example. Consider the experiment in which we roll a dice twice. Consider the following events.

A: event that the first roll results in a 1, 2, or a 3.
B: event that the first roll results in a 3, 4, or a 5.
C: event that the sum of the two rolls is a 9

Are events A, B, and C mutually independent?

Solution. We show below that the events are not mutually independent as they are not pairwise independent.

$$\begin{aligned}
A &= \{(i,j) \,|\, 1 \le i \le 3 \text{ and } 1 \le j \le 6\} \\
B &= \{(i,j) \,|\, 3 \le i \le 5 \text{ and } 1 \le j \le 6\} \\
C &= \{(3,6),(6,3),(4,5),(5,4)\} \\
A \cap B &= \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\} \\
A \cap C &= \{(3,6)\} \\
B \cap C &= \{(3,6),(4,5),(5,4)\} \\
A \cap B \cap C &= \{(3,6)\} \\
\Pr[A] &= 1/2 \\
\Pr[B] &= 1/2 \\
\Pr[C] &= 1/9 \\
\Pr[A \cap B \cap C] &= 1/36 = \Pr[A] \cdot \Pr[B] \cdot \Pr[C] \\
\Pr[A \cap B] &= 1/6 \neq \Pr[A] \cdot \Pr[B] \\
\Pr[A \cap C] &= 1/36 \neq \Pr[A] \cdot \Pr[C] \\
\Pr[B \cap C] &= 3/36 \neq \Pr[B] \cdot \Pr[C]
\end{aligned}$$

Random Variables

In an experiment we are often interested in some value associated with an outcome as opposed to the actual outcome itself. For example, consider an experiment that involves tossing a coin three times. We may not be interested in the actual head-tail sequence that results but be more interested in the number of heads that occur. These quantities of interest are called *random variables*.

Definition. A *random variable* X on a sample space Ω is a real-valued function that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.

In this course we will study discrete random variables which are random variables that take on only a finite or countably infinite number of values.

For a discrete random variable X and a real value a, the event "X=a" is the set of outcomes in Ω for which the random variable assumes the value a, i.e., $X = a \equiv \{\omega \in \Omega | X(\omega) = a\}$. The probability of this event is denoted by

$$\Pr[X = a] = \sum_{\omega \in \Omega : X(\omega) = a} \Pr[\omega]$$

Definition. The *distribution* or the *probability mass function*(PMF) of a random variable X gives the probabilitites for the different possible values of X. Thus, if x is a value that

X can assume then $p_X(x)$ is the probability mass of X and is given by

$$p_X(x) = \Pr[X = x]$$

Observe that $\sum_x p_X(x) = \sum_x \Pr[X = x] = 1$. This is because the events $X = x$ are disjoint and hence partition the sample space Ω.

Consider the experiment of tossing three fair coins. Let X be the random variable that denotes the number of heads that result. The PMF or the distribution of X is given below.

$$p_X(x) = \begin{cases} 1/8 & \text{if } x = 0 \text{ or } x = 3 \\ 3/8 & \text{otherwise} \end{cases}$$

The definition of independence that we developed for events extends to random variables.

Definition. Two random variables X and Y are independent if and only if

$$\Pr[(X = x) \cap (Y = y)] = \Pr[X = x] \times \Pr[Y = y]$$

for all values x and y. In other words, two random variables X and Y are independent if every event determined by X is independent of every event determined by Y.

Similarly, random variables $X_1, X_2, \ldots, X_k$ are mutually independent if and only if, for any subset $I \subseteq [1, k]$ and any values $x_i, i \in I$,

$$\Pr[\cap_{i \in I} X_i = x_i] = \prod_{i \in I} \Pr[X_i = x_i]$$

Expectation

The PMF of a random variable, X, provides us with many numbers, the probabilities of all possible values of X. It would be desirable to summarize this distribution into a representative number that is also easy to compute. This is accomplished by the *expectation* of a random variable which is the weighted average (proportional to the probabilities) of the possible values of X.

Definition. The *expectation* of a discrete random variable X, denoted by $\mathbf{E}[X]$, is given by

$$\mathbf{E}[X] = \sum_i i p_X(i) = \sum_i i \Pr[X = i]$$

Intuitively, $\mathbf{E}[X]$ is the value we would expect to obtain if we repeated a random experiment several times and took the average of the outcomes of X.

In our running example, in expectation the number of heads is given by

$$\mathbf{E}[X] = 0 \times \frac{1}{8} + 3 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} = \frac{3}{2}$$

As seen from the example, the expectation of a random variable may not be a valid value of the random variable.

Example. When we roll a die what is the result in expectation?

Solution. Let X be the random variable that denotes the result of a single roll of dice. The PMF for X is given by

$$p_X(x) = \frac{1}{6}, x = 1, 2, 3, 4, 5, 6.$$

The expectation of X is given by

$$\mathbf{E}[X] = \sum_{x=1}^{6} p_x(x) \cdot x = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

Example. When we roll two dice what is the expected value of the sum?

Solution. Let S be the random variable denoting the sum. The PMF for S is given by

$$p_S(x) = \begin{cases} \frac{1}{36}, x = 2, 12 \\ \frac{2}{36}, x = 3, 11 \\ \frac{3}{36}, x = 4, 10 \\ \frac{4}{36}, x = 5, 9 \\ \frac{5}{36}, x = 6, 8 \\ \frac{6}{36}, x = 7 \end{cases}$$

The expectation of S is given by

$$\begin{aligned} \mathbf{E}[S] &= \sum_{x=2}^{12} p_S(x) \cdot x \\ &= \frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \frac{4}{36} \times 4 + \frac{5}{36} \times 6 + \frac{6}{36} \times 7 + \\ &\quad \frac{5}{36} \times 8 + \frac{4}{36} \times 9 + \frac{3}{36} \times 10 + \frac{2}{36} \times 11 + \frac{1}{36} \times 12 \\ &= \frac{252}{36} = 7 \end{aligned}$$

Linearity of Expectation

One of the most important properties of expectation that simplifies its computation is the *linearity of expectation*. By this property, the expectation of the sum of random variables equals the sum of their expectations. This is given formally in the following theorem. I didn't cover the proof in the class but I am including it here for anyone who is interested.

Theorem. For any finite collection of random variables $X_1, X_2, \ldots, X_n$,

$$\mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i]$$

Proof. We will prove the statement for two random variables X and Y. The general claim can be proven using induction.

$$\begin{aligned}
\mathbf{E}[X+Y] &= \sum_i \sum_j (i+j)\Pr[X=i \cap Y=j] \\
&= \sum_i \sum_j (i\Pr[X=i \cap Y=j] + j\Pr[X=i \cap Y=j]) \\
&= \sum_i \sum_j i\Pr[X=i \cap Y=j] + \sum_i \sum_j j\Pr[X=i \cap Y=j] \\
&= \sum_i i \sum_j \Pr[X=i \cap Y=j] + \sum_j j \sum_i \Pr[X=i \cap Y=j] \\
&= \sum_i i\Pr[X=i] + \sum_j j\Pr[Y=j] \\
&= \mathbf{E}[X] + \mathbf{E}[Y]
\end{aligned}$$

It is important to note that no assumptions have been made about the random variables while proving the above theorem. For example, the random variables do not have to be independent for linearity of expectation to be true.

Lemma. For any constant c and discrete random variable X,

$$\mathbf{E}[cX] = c\mathbf{E}[X]$$

Proof. The lemma clearly holds for $c = 0$. For $c \neq 0$

$$\begin{aligned}
\mathbf{E}[cX] &= \sum_j j\Pr[cX=j] \\
&= c\sum_j (j/c)\Pr[X=j/c] \\
&= c\sum_k k\Pr[X=k] \\
&= c\mathbf{E}[X]
\end{aligned}$$

Example. Using linearity of expectation calculate the expected value of the sum of the numbers obtained when two dice are rolled.

Solution. Let X_1 and X_2 denote the random variables that denote the result when die 1 and die 2 are rolled respectively. We want to calculate $\mathbf{E}[X_1 + X_2]$. By linearity of expectation

$$\begin{aligned}
\mathbf{E}[X_1+X_2] &= \mathbf{E}[X_1] + \mathbf{E}[X_2] \\
&= \frac{1}{6}(1+2+3+4+5+6) + \frac{1}{6}(1+2+3+4+5+6) \\
&= 3.5 + 3.5 \\
&= 7
\end{aligned}$$

Example. Suppose that n people leave their hats at the hat check. If the hats are randomly returned what is the expected number of people that get their own hat back?

Solution. Let X be the random variable that denotes the number of people who get their own hat back. Let $X_i, 1 \leq i \leq n$, be the random variable that is 1 if the ith person gets his/her own hat back and 0 otherwise. Clearly,

$$X = X_1 + X_2 + X_3 + \ldots + X_n$$

By linearity of expectation we get

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} \frac{(n-1)!}{n!} = n \times \frac{1}{n} = 1$$

Example. Suppose we throw n balls into n bins with the probability of a ball landing in each of the n bins being equal. What is the expected number of empty bins?

Solution. First Approach: The following approach was discussed in class. Let X be the random variable denoting the number of empty bins. For $0 \leq i \leq n$, let X_i be a random variable that is i if exactly i bins are empty and 0, otherwise. We have

$$X = \sum_{i=1}^{n} X_i$$

By the linearity of expectation, we have

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} i \Pr[X_i = i] = \sum_{i=1}^{n} i \Pr[X = i]$$

The last equality follows because exactly one of the X_is will be non-zero and if $X_i \neq 0$ then $X = X_i$. Note that we have not made any progress as we are back to using the original definition of expectation to solve the problem.

Second Approach: Let X be the random variable denoting the number of empty bins. Let X_i be a random variable that is 1 if the ith bin is empty and is 0 otherwise. Clearly

$$X = \sum_{i=1}^{n} X_i$$

By linearity of expectation, we have

$$\begin{aligned}\mathbf{E}[X] &= \sum_{i=1}^{n} \mathbf{E}[X_i] \\ &= \sum_{i=1}^{n} \Pr[X_i = 1] \\ &= \sum_{i=1}^{n} \left(\frac{n-1}{n}\right)^n \\ &= \sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^n\end{aligned}$$

As $n \to \infty$, $(1 - \frac{1}{n})^n \to \frac{1}{e}$. Hence, for large enough values of n we have

$$\mathbf{E}[X] = \frac{n}{e}$$

Example. The following pseudo-code computes the minimum of n distinct numbers that are stored in an array A. What is the expected number of times that the variable *min* is assigned a value if the array A is a random permutation of the n elements.

```
FindMin(A, n):
min ← A[1]
for i ← 2 to n do
    if (A[i] < min) then
        min ← A[i]
return min
```

Solution. Let X be the random variable denoting the number of times that *min* is assigned a value. We want to calculate $\mathbf{E}[X]$. Let X_i be the random variable that is 1 if *min* is assigned $A[i]$ and 0 otherwise. Clearly,

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

Using the linearity of expectation we get

$$\begin{aligned}\mathbf{E}[X] &= \sum_{i=1}^{n} \mathbf{E}[X_i] \\ &= \sum_{i=1}^{n} \Pr[X_i = 1] \qquad (3)\end{aligned}$$

Note that $\Pr[X_i = 1]$ is the probability that $A[i]$ contains the smallest element among the elements $A[1], A[2], \ldots, A[i]$. Since the smallest of these elements is equally likely to be in any of the first i locations, we have $\Pr[X_i = 1] = \frac{1}{i}$. Thus equation (3) becomes

$$\mathbf{E}[X] = \sum_{i=1}^{n} \frac{1}{i} = H(n) \approx \ln n + c$$

where c is a constant less than 1.

Example. Suppose there are k people in a room and n days in a year. On average how many pairs of people share the same birthday?

Solution. Let X be the random variable denoting the number of pairs of people sharing the same birthday. For any two people i and j, let X_{ij} be an indicator random variable that is 1 if i and j have the same birthday and is 0 otherwise. Clearly $X = \sum_{i,j} X_{ij}$. Using the linearity of expectation we get

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{i,j} \mathbf{E}[X_{ij}] \\
&= \sum_{i,j} \Pr[X_{ij} = 1] \\
&= \sum_{i,j} \frac{1}{n} \\
&= \frac{\binom{k}{2}}{n} \\
&= \frac{k(k-1)}{2n}
\end{aligned}$$

Assuming $n = 365$, the smallest value of k for which the RHS is at least 1 is 28.

Example (Markov's Inequality). Let X be a non-negative random variable. Then for all $a > 0$, prove that

$$\Pr[X \geq a] \leq \frac{\mathbf{E}[X]}{a}$$

Solution. Intuitively, the claim means that if there is too much of probability mass associated with values above $\mathbf{E}[X]$ then the total contribution of such values to $\mathbf{E}[X]$ would be very large. Formally, the proof is as follows.

$$\begin{aligned}
\mathbf{E}[X] &= \sum_x x \Pr[X = x] \\
&\geq \sum_{x \geq a} x \Pr[X = x] \\
&\geq a \sum_{x \geq a} \Pr[X = x] \\
&= a \Pr[X \geq a] \\
\therefore \Pr[X \geq a] &\leq \frac{\mathbf{E}[X]}{a}
\end{aligned}$$

Example. Suppose we flip a fair coin n times. Using Markov's inequality bound the the probability of obtaining at least $3n/4$ heads.

Solution. Let X be the random variable denoting the total number of heads in n flips of a fair coin. We know that $\mathbf{E}[X] = n/2$. Applying the above inequality we get

$$\Pr[X \geq 3n/4] \leq \frac{\mathbf{E}[X]}{3n/4} = \frac{n/2}{3n/4} = \frac{2}{3}$$

Example. Suppose we roll a die. Using Markov's inequality bound the probability of obtaining a number greater than or equal to 7.

Solution. Let X be the random variable denoting the result of the roll of a die. We know that $\mathbf{E}[X] = 3.5$. Using the Markov's inequality we get

$$\Pr[X \geq 7] \leq \frac{\mathbf{E}[X]}{7} \leq \frac{1}{2}$$

As this result shows, Markov's inequality gives a loose bound in some cases.

Variance

We are interested in calculating how much a random variable deviates from its mean. This measure is called *variance*. Formally, for a random variable X we are interested in $\mathbf{E}[X - \mathbf{E}[X]]$. By the linearity of expectation we have

$$\mathbf{E}[X - \mathbf{E}[X]] = \mathbf{E}[X] - \mathbf{E}[\mathbf{E}[X]] = \mathbf{E}[X] - \mathbf{E}[X] = 0$$

Note that we have used the fact that $\mathbf{E}[X]$ is a constant and hence $\mathbf{E}[\mathbf{E}[X]] = \mathbf{E}[X]$. This is not very informative. While calculating the deviations from the mean we do not want the positive and the negative deviations to cancel out each other. This suggests that we should take the absolute value of $X - \mathbf{E}[X]$. But working with absolute values is messy. It turns out that squaring of $X - \mathbf{E}[X]$ is more useful. This leads to the following definition.

Definition. The *variance* of a random variable X is defined as

$$\mathrm{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$$

The *standard deviation* of a random variable X is

$$\sigma[X] = \sqrt{\mathrm{Var}[X]}$$

The standard deviation undoes the squaring in the variance. In doing the calculations it does not matter whether we use variance or the standard deviation as we can easily compute one from the other.

We show as follows that the two forms of variance in the definition are equivalent.

$$\begin{aligned}
\mathbf{E}[(X - \mathbf{E}[X])^2] &= \mathbf{E}[X^2 - 2X\mathbf{E}[X] + \mathbf{E}[X]^2] \\
&= \mathbf{E}[X^2] - 2\mathbf{E}[X\mathbf{E}[X]] + \mathbf{E}[X]^2 \\
&= \mathbf{E}[X^2] - 2\mathbf{E}[X]^2 + \mathbf{E}[X]^2 \\
&= \mathbf{E}[X^2] - \mathbf{E}[X]^2
\end{aligned}$$

In step 2 we used the linearity of expectation and the fact that $\mathbf{E}[X]$ is a constant.

Example. Consider three random variables X, Y, Z. Their probability mass distribution is as follows.

$$\Pr[X = x] = \begin{cases} \frac{1}{2}, x = -2 \\ \frac{1}{2}, x = 2 \end{cases}$$

$$\Pr[Y = y] = \begin{cases} 0.001, y = -10 \\ 0.998, y = 0 \\ 0.001, y = 10 \end{cases}$$

$$\Pr[Z = z] = \begin{cases} \frac{1}{3}, z = -5 \\ \frac{1}{3}, z = 0 \\ \frac{1}{3}, z = 5 \end{cases}$$

Which of the above random variables is more "spread out"?

Solution. It is easy to see that $\mathbf{E}[X] = \mathbf{E}[Y] = \mathbf{E}[Z] = 0$.

$$\begin{aligned} \mathrm{Var}[X] &= \mathbf{E}[X^2] \\ &= 0.5 \cdot (-2)^2 + 0.5 \cdot (2)^2 \\ &= 4 \\ \mathrm{Var}[Y] &= \mathbf{E}[Y^2] \\ &= 0.001 \cdot (-10)^2 + 0.998 \cdot 0^2 + 0.001 \cdot (10)^2 \\ &= 0.2 \\ \mathrm{Var}[Z] &= \mathbf{E}[Z^2] \\ &= (1/3) \cdot (-5)^2 + (1/3) \cdot 0^2 + (1/3) \cdot (5)^2 \\ &= 16.67 \end{aligned}$$

Thus Z is the most spread out and Y is the most concentrated.

Example. In the experiment where we roll one die let X be the random variable denoting the number that appears on the top face. What is $\mathrm{Var}[X]$?

Solution. From the definition of variance, we have

$$\begin{aligned} \mathrm{Var}[X] &= \mathbf{E}[X^2] - \mathbf{E}[X]^2 \\ &= \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) + \left(\frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6)\right)^2 \\ &= \frac{91}{6} - \frac{49}{4} \\ &= \frac{35}{12} \end{aligned}$$

Example. In the hat-check problem that we did in one of the earlier lectures, what is the variance of the random variable X that denotes the number of people who get their own hat back?

Solution. We can express X as

$$X = X_1 + X_2 + \cdots + X_n$$

where X_i is the random variable that denotes that is 1 if the ith person receives his/her own hat back and 0 otherwise. We already know from an earlier lecture that $\mathbf{E}[X] = 1$. If $n = 1$ then $\mathbf{E}[X^2] = \mathbf{E}[X_1^2] = \Pr[X_1 = 1] = 1$. In this case, $\mathrm{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = 1 - 1 = 0$, as expected. If $n \geq 2$, $\mathbf{E}[X^2]$ can be calculated as follows.

$$\begin{aligned}
\mathbf{E}[X^2] &= \sum_{i=1}^{n} \mathbf{E}[X_i^2] + 2\sum_{i<j} \mathbf{E}[X_i \cdot X_j] \\
&= \sum_{i=1}^{n} \mathbf{E}[X_i^2] + 2\sum_{i<j} 1 \cdot \Pr[X_i = 1 \cap X_j = 1] \\
&= \sum_{i=1}^{n} \frac{1}{n} + 2\left(\frac{n(n-1)}{2}\right)\left(\frac{1}{n(n-1)}\right) \\
&= n \cdot \frac{1}{n} + 1 \\
&= 2
\end{aligned}$$

$\mathrm{Var}[X]$ is given by

$$\mathrm{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = 2 - 1 = 1$$

Note that like the expectation, the variance is independent of n. This means that it is not likely for many people to get their own hat back even if n is large.

Example (Chebyshev's Inequality). Let X be a random variable. Show that for any $a > 0$,

$$\Pr[|X - \mathbf{E}[X]| \geq a] \leq \frac{\mathrm{Var}[X]}{a^2}$$

Solution. The inequality that we proved in the earlier homework is called Markov's inequality. We will use it to prove the above tail bound called Chebyshev's inequality.

$$\begin{aligned}
\Pr[|X - \mathbf{E}[X]| \geq a] &= \Pr[(X - \mathbf{E}[X])^2 \geq a^2] \\
&\leq \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{a^2} \qquad \text{(using Markov's Inequality)} \\
&= \frac{\mathrm{Var}[X]}{a^2}
\end{aligned}$$

Theorem. If X and Y are independent real-valued random variables then

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] \text{ and } \mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$$

The result can be extended to a finite number of random variables.

Note that the converse of the above statement is not true as illustrated by the following example. Let $\Omega = \{a, b, c\}$, with all three outcomes equally likely. Let X and Y be random variables defined as follows: $X(a) = 1, X(b) = 0, X(c) = -1$ and $Y(a) = 0, Y(b) = 1, Y(c) = 0$. Note that X and Y are not independent since

$$\Pr[X = 0 \wedge Y = 0] = 0, \text{ but } \Pr[X = 0] \cdot \Pr[Y = 0] = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9} \neq 0.$$

Note that for all $\omega \in \Omega$, $X(\omega)Y(\omega) = 0$. Also, $\mathbf{E}[X] = 0$ and $\mathbf{E}[Y] = 1/3$. Thus we have

$$\mathbf{E}[XY] = 0 = \mathbf{E}[X]\mathbf{E}[Y]$$

It is also easy to verify that $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$.

Example (Chebyshev's Inequality). Let X be a random variable. Show that for any $a > 0$,

$$\Pr[|X - \mathbf{E}[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}$$

Solution. The inequality that we proved in the earlier homework is called Markov's inequality. We will use it to prove the above tail bound called Chebyshev's inequality.

$$\begin{aligned}
\Pr[|X - \mathbf{E}[X]| \geq a] &= \Pr[(X - \mathbf{E}[X])^2 \geq a^2] \\
&\leq \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{a^2} \qquad \text{(using Markov's Inequality)} \\
&= \frac{\text{Var}[X]}{a^2}
\end{aligned}$$

Example. Use Chebyshev's inequality to bound the probability of obtaining at least $3n/4$ heads in a sequence of n fair coin flips.

Solution. Let X denote the random variable denoting the total number of heads that result in n flips of a fair coin. For $1 \leq i \leq n$, let X_i be a random variable that is 1, if the ith flip results in Heads, 0, otherwise. Thus,

$$X = X_1 + X_2 + \cdots + X_n$$

By the linearity of expectation, $\mathbf{E}[X] = n/2$. Since the random variables X_is are independent, we have

$$\text{Var}[X] = \sum_{i=1}^{n} \text{Var}[X_i] = \sum_{i=1}^{n} (1/2 - 1/4) = \frac{n}{4}$$

Using Chebyshev's inequality, we get

$$\begin{aligned}\Pr[X \geq 3n/4] &= \Pr[X - n/2 \geq n/4] \\ &= \Pr[X - \mathbf{E}[X] \geq n/4] \\ &= \frac{1}{2} \cdot \Pr[|X - \mathbf{E}[X]| \geq n/4] \\ &\leq \frac{1}{2} \cdot \frac{\mathrm{Var}[X]}{n^2/16} \\ &= \frac{2}{n}\end{aligned}$$

Probability Distributions

Tossing a coin is an experiment with exactly two outcomes: heads ("success") with a probability of, say p, and tails ("failure") with a probability of $1 - p$. Such an experiment is called a *Bernoulli trial*. Let Y be a random variable that is 1 if the experiment succeeds and is 0 otherwise. Y is called a *Bernoulli* or an *indicator* random variable. For such a variable we have

$$\mathbf{E}[Y] = p \cdot 1 + (1 - p) \cdot 0 = p = \Pr[Y = 1]$$

Thus for a fair coin if we consider heads as "success" then the expected value of the corresponding indicator random variable is 1/2.

A sequence of Bernoulli trials means that the trials are independent and each has a probability p of success. We will study two important distributions that arise from Bernoulli trials: the *geometric distribution* and the *binomial distribution*.

The Geometric Distribution

Consider the following question. Suppose we have a biased coin with heads probability p that we flip repeatedly until it lands on heads. What is the distribution of the number of flips? This is an example of a *geometric distribution*. It arises in situations where we perform a sequence of independent trials until the first success where each trial succeeds with a probability p.

Note that the sample space Ω consists of all sequences that end in H and have exactly one H. That is

$$\Omega = \{H, TH, TTH, TTTH, TTTTH, \ldots\}$$

For any $\omega \in \Omega$ of length i, $\Pr[\omega] = (1 - p)^{i-1}p$.

Definition. A *geometric random variable* X with parameter p is given by the following distribution for $i = 1, 2, \ldots$:

$$\Pr[X = i] = (1 - p)^{i-1}p$$

We can verify that the geometric random variable admits a valid probability distribution as follows:

$$\sum_{i=1}^{\infty}(1-p)^{i-1}p = p\sum_{i=1}^{\infty}(1-p)^{i-1} = \frac{p}{1-p}\sum_{i=1}^{\infty}(1-p)^{i} = \frac{p}{1-p}\cdot\frac{1-p}{1-(1-p)} = 1$$

Note that to obtain the second-last term we have used the fact that $\sum_{i=1}^{\infty} c^i = \frac{c}{1-c}$, $|c| < 1$.

Let's now calculate the expectation of a geometric random variable, X. We can do this in several ways. One way is to use the definition of expectation.

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{i=0}^{\infty} i\Pr[X=i] \\
&= \sum_{i=0}^{\infty} i(1-p)^{i-1}p \\
&= \frac{p}{1-p}\sum_{i=0}^{\infty} i(1-p)^{i} \\
&= \left(\frac{p}{1-p}\right)\left(\frac{1-p}{(1-(1-p))^2}\right) \quad \left(\because \sum_{i=0}^{\infty} kx^k = \frac{x}{(1-x)^2}, \text{ for } |x|<1.\right) \\
&= \left(\frac{p}{1-p}\right)\left(\frac{1-p}{p^2}\right) \\
&= \frac{1}{p}
\end{aligned}$$

Another way to compute the expectation is to note that X is a random variable that takes on non-negative values. From a theorem proved in last class we know that if X takes on only non-negative values then

$$\mathbf{E}[X] = \sum_{i=1}^{\infty}\Pr[X \geq i]$$

Using this result we can calculate the expectation of the geometric random variable X. For the geometric random variable X with parameter p,

$$\Pr[X \geq i] = \sum_{j=i}^{\infty}(1-p)^{j-1}p = (1-p)^{i-1}p\sum_{j=0}^{\infty}(1-p)^{j} = (1-p)^{i-1}p \times \frac{1}{1-(1-p)} = (1-p)^{i-1}$$

Therefore

$$\mathbf{E}[X] = \sum_{i=1}^{\infty}\Pr[X \geq i] = \sum_{i=1}^{\infty}(1-p)^{i-1} = \frac{1}{1-p}\sum_{i=1}^{\infty}(1-p)^{i} = \frac{1}{1-p}\cdot\frac{1-p}{1-(1-p)} = \frac{1}{p}$$

Memoryless Property. For a geometric random variable X with parameter p and for $n > 0$,

$$\Pr[X = n + k \,|\, X > k] = \Pr[X = n]$$

Conditional Expectation. The following is the definition of conditional expectation.

$$\mathbf{E}[Y \mid Z = z] = \sum_y y \Pr[Y = y \mid Z = z],$$

where the summation is over all possible values y that the random variable Y can assume.

Example. For any random variables X and Y,

$$\mathbf{E}[X] = \sum_y \Pr[Y = y]\mathbf{E}[X \mid Y = y]$$

We can also calculate the expectation of a geometric random variable X using the memoryless property of the geometric random variable. Let Y be a random variable that is 0, if the first flip results in tails and that is 1, if the first flip is a heads. Using conditional expectation we have

$$\begin{aligned}
\mathbf{E}[X] &= \Pr[Y = 0]\mathbf{E}[X|Y = 0] + \Pr[Y = 1]\mathbf{E}[X|Y = 1] \\
&= (1 - p)(\mathbf{E}[X] + 1) + p \cdot 1 \qquad \text{(using the memoryless property)} \\
\therefore p\mathbf{E}[X] &= 1 \\
\mathbf{E}[X] &= \frac{1}{p}
\end{aligned}$$

Binomial Distributions

Consider an experiment in which we perform a sequence of n coin flips in which the probability of obtaining heads is p. How many flips result in heads?

If X denotes the number of heads that appear then

$$\Pr[X = j] = \binom{n}{j} p^j (1 - p)^{n-j}$$

Definition. A *binomial* random variable X with parameters n and p is defined by the following probability distribution on $j = 0, 1, 2, \ldots, n$:

$$\Pr[X = j] = \binom{n}{j} p^j (1 - p)^{n-j}$$

We can verify that the above is a valid probability distribution using the binomial theorem as follows

$$\sum_{j=1}^{n} \binom{n}{j} p^j (1 - p)^{n-j} = (p + (1 - p))^n = 1$$

What is the expectation of a binomial random variable X? We can calculate $\mathbf{E}[X]$ is two ways. We first calculate it directly from the definition.

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{j=0}^{n} j\binom{n}{j}p^j(1-p)^{n-j} \\
&= \sum_{j=0}^{n} j\frac{n!}{j!(n-j)!}p^j(1-p)^{n-j} \\
&= \sum_{j=1}^{n} j\frac{n!}{j!(n-j)!}p^j(1-p)^{n-j} \\
&= \sum_{j=1}^{n} \frac{n!}{(j-1)!(n-j)!}p^j(1-p)^{n-j} \\
&= np\sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!((n-1)-(j-1))!}p^{j-1}(1-p)^{(n-1)-(j-1)} \\
&= np\sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!}p^k(1-p)^{(n-1)-k} \\
&= np\sum_{k=0}^{n-1} \binom{n-1}{k}p^k(1-p)^{(n-1)-k} \\
&= np
\end{aligned}$$

The last equation follows from the binomial expansion of $(p+(1-p))^{n-1}$.

We can obtain the result in a much simpler way by using the linearity of expectation. Let $X_i, 1 \leq i \leq n$ be the indicator random variable that is 1 if the ith flip results in heads and is 0 otherwise. We have $X = \sum_{i=1}^{n} X_i$. By the lineartity of expectation we have

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} p = np$$

What is the variance of the binomial random variable X? Since $X = \sum_{i=1}^{n} X_i$, and $X_1, X_2, \ldots, X_n$ are independent we have

$$\begin{aligned}
\mathrm{Var}[X] &= \sum_{i=1}^{n} \mathrm{Var}[X_i] \\
&= \sum_{i=1}^{n} \mathbf{E}[X_i^2] - \mathbf{E}[X_i]^2 \\
&= \sum_{i=1}^{n} (p - p^2) \\
&= np(1-p)
\end{aligned}$$

Coupon Collector's Problem.

We are trying to collect n different coupons that can be obtained by buying cereal boxes. The objective is to collect at least one coupon of each of the n types. Assume that each cereal box contains exactly one coupon and any of the n coupons is equally likely to occur. How many cereal boxes do we expect to buy to collect at least one coupon of each type?

Solution. Let the random variable X denote the number of cereal boxes bought until we have at least one coupon of each type. We want to compute $\mathbf{E}[X]$. Let X_i be the random variable denoting the number of boxes bought to get the ith new coupon. Clearly,

$$X = X_1 + X_2 + X_3 + \ldots + X_n$$

Using the linearity of expectation we have

$$\mathbf{E}[X] = \mathbf{E}[X_1] + \mathbf{E}[X_2] + \mathbf{E}[X_3] + \ldots + \mathbf{E}[X_n] \tag{4}$$

What is the distribution of random variable X_i? Observe that the probability of obtaining the ith new coupon is given by

$$p_i = \frac{n-(i-1)}{n} = \frac{n-i+1}{n}$$

Thus the random variable $X_i, 1 \leq i \leq n$ is a geometric random variable with parameter p_i.

$$\mathbf{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$$

Combining this with equation (4) we get

$$\mathbf{E}[X] = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{2} + \frac{n}{1} = n\sum_{i=1}^{n}\frac{1}{i}$$

The summation $\sum_{i-1}^{n}\frac{1}{i}$ is known as the *harmonic number* $H(n)$ and $H(n) = \ln n + c$, for some constant $c < 1$.

Hence the expected number of boxes needed to collect n coupons is about $nH(n) < n(\ln n + 1)$.

The Probabilistic Method

A *tournament graph* is a directed graph with exactly one directed edge between any pair of vertices. Every tournament graph has at least one Hamiltonian path, a path that visits each vertex exactly once (can be proved using induction). In 1943, Szele used the Probabilistic Method to show the existence of a tournament graph with a large number of Hamiltonian paths. Note that there are tournaments in which there is exactly one Hamiltonian path. For example, the tournament on vertices $\{1, 2, \ldots, n\}$ in which there is a directed edge (i, j) iff $i < j$ has exactly one Hamiltonian path.

Example. Prove that there is a n-vertex tournament with at least $\frac{n!}{2^{n-1}}$ distinct Hamiltonian paths.

Solution. Let $G = (n, 1/2)$ be a n-vertex tournament graph, in which an edge between any two vertices u and v is directed towards u with probability $\frac{1}{2}$ and towards v with probability $\frac{1}{2}$. Let X denote the total number of Hamiltonian paths in G and let X_σ be an indicator random variable that is 1, iff a permutation σ of the vertices in G yields a Hamiltonian path. Clearly, $X = \sum_\sigma X_\sigma$. Applying the Linearity of Expectation, we get

$$\begin{aligned}
\mathbf{E}[X] &= \sum_\sigma \mathbf{E}[X_\sigma] \\
&= \sum_\sigma \Pr[X_\sigma = 1] \\
&= \sum_\sigma \left(\frac{1}{2}\right)^{n-1} \\
&= \frac{n!}{2^{n-1}}
\end{aligned}$$

Since a random orientation of the edges, i.e., a random tournament, yields us the above number in expectation, there must be an orientation of the edges, i.e., a tournament, in which the number of Hamiltonian paths is at least $n!/2^{n-1}$.

An *independent set* S in G is a subset of vertices such that no two vertices in S share an edge. The *independence number* of a graph G, denoted by $\alpha(G)$ is the size of the largest independent set in G.

Example. Let n be the number of vertices in G and m be the number of edges, and let $d = \frac{2m}{n} \geq 1$ be the average degree. Then

$$\alpha(G) \geq \frac{n}{2d}$$

This is a weaker version of the celebrated Turán's theorem.

Solution. Construct a random subset S of vertices by placing each vertex in S independently with probability p (to be determined later). Let X be the random variable denoting the number of vertices in S and let Y be the random variable denoting the number of edges whose both endpoints are in S. Let Y_e be an indicator random variable that is 1 iff both endpoints of e are in S. By the Linearity of Expectation we have

$$\mathbf{E}[X] = np \qquad \text{and} \qquad \mathbf{E}[Y] = \sum_e \mathbf{E}[Y_e] = \sum_e \Pr[Y_e = 1] = mp^2 = \frac{nd}{2}p^2$$

Note that the quantity $X - Y$ denotes the number of vertices in S minus the number of edges with both endpoints in S. By the Linearity of Expectation we get

$$\mathbf{E}[X - Y] = np - \frac{nd}{2}p^2 = np\left(1 - \frac{dp}{2}\right)$$

This means that there exists a set S such that the number of vertices in S exceeds the number of edges in S by the above quantity. We now modify set S by deleting an arbitrary endpoint of each edge. The resulting set S' has at least $np\left(1-\frac{dp}{2}\right)$ vertices left and has no edges between any of its vertices. We want to maximize $|S'|$, so we set $p = 1/d$ (using $d \geq 1$), giving us $|S'| = \frac{n}{2d}$.

For any graph $G = (V, E)$, a set of vertices $D \subseteq V$ is called a *dominating set* if every vertex in $V \setminus D$ is adjacent to a vertex in D.

Example. Prove that any connected graph $G = (V, E)$ with $n \geq 2$ vertices and minimum degree $\delta(G) = \delta$, contains a dominating set of size at most $\frac{n(1+\log(1+\delta))}{1+\delta}$.

Solution. For each vertex $v \in V$, add it to the set X independently with probability p. Let $Y \subseteq V \setminus X$ be the vertices that are not dominated by X, i.e., they are vertices in $V \setminus X$ that are not dominated by X. Then $X \cup Y$ is a dominating set for G. We will now show that $\mathbf{E}[X \cup Y]$ is not too large. Since X and Y are disjoint sets, we have

$$\mathbf{E}[X \cup Y] = \mathbf{E}[X] + \mathbf{E}[Y] \tag{5}$$

We consider the following random variables.
X_v: random variable that is 1 if vertex v is in X, 0, otherwise.
Y_v: random variable that is 1 if vertex v and all of its neighbors are not in X, 0, otherwise.

$$\begin{aligned} X &= \sum_v X_v \\ \therefore \mathbf{E}[X] &= \sum_v \Pr[X_v = 1] \\ &= np \end{aligned}$$

$$\begin{aligned} Y &= \sum_v Y_v \\ \therefore \mathbf{E}[Y] &= \sum_v \Pr[Y_v = 1] \\ &= \sum_v (1-p)^{\deg(v)+1} \\ &\leq \sum_v (1-p)^{\delta+1} \\ &= n(1-p)^{\delta+1} \end{aligned}$$

Plugging the values of $\mathbf{E}[X]$ and $\mathbf{E}[Y]$ in (5) we get

$$\mathbf{E}[X \cup Y] \leq np + n(1-p)^{\delta+1} \leq np + ne^{-p(\delta+1)},$$

The last expression is minimized when

$$p = \frac{\ln(1+\delta)}{1+\delta}$$

Thus, we can find a dominating set of size at most $\frac{n(1+\ln(1+\delta))}{1+\delta}$.

Recall that a tournament is a directed graph with exactly one directed edge between any pair of vertices. A tournament $G = (V, E)$ is called k-dominated if for every set of k vertices $v_1, v_2, \ldots, v_k$, there exists another vertex $u \in V$ such that $(u, v_i) \in E$, for $i = 1, 2, \ldots, k$.

Example. Prove that for any positive integer k, if n is large enough then there is a k-dominated tournament on n vertices.

Solution. Construct a random tournament G in which an edge between any two vertices u and v is directed towards u with probability $\frac{1}{2}$ and towards v with probability $\frac{1}{2}$. The bad event for our random process is that G is not k-dominated. We will calculate the probability of this bad event as follows. Let S be a fixed set of k vertices in G. The probability that a vertex u outside of S does not dominate set S is given by $1 - (1/2)^k$. Thus the probability that S is not dominated by any of the $n - k$ vertices outside of S is given by $(1 - 1/2^k)^{n-k}$. Since there are $\binom{n}{k}$ possibilities for set S, the probability of some set of k vertices in G not being dominated is at most

$$\binom{n}{k}\left(1 - \frac{1}{2^k}\right)^{n-k} \tag{6}$$

If the above expression is less than 1, it means that the probability of the random tournament G being k-dominated is strictly larger than 0, which means that such a tournament exists. We will now show that if $n/\ln n > k2^k$ then the expression (6) is less than 1.

$$\begin{aligned}
\binom{n}{k}\left(1 - \frac{1}{2^k}\right)^{n-k} &\leq \frac{n^k}{k!} \cdot e^{\frac{k-n}{2^k}} && (\text{using } 1 + x \leq e^x, \forall x \in \mathbb{R}) \\
&= \frac{e^{k \ln n}}{k!} \cdot e^{\frac{k}{2^k} - \frac{n}{2^k}} && (\text{since } n = e^{\ln n}) \\
&= \frac{e^{\frac{k}{2^k}}}{k!} \cdot e^{k \ln n - \frac{n}{2^k}} \\
&\leq \frac{e}{k!} \cdot \frac{1}{e} && (\text{since } n/\ln n > k2^k) \\
&= \frac{1}{k!} < 1
\end{aligned}$$

Note that for large values of k, $n > k^2 2^k$ satisfies the inequality $n/\ln n > k2^k$. This is because when $n = k^2 2^k$, we have

$$\frac{n}{\ln n} = \frac{k^2 2^k}{\ln(k^2 2^k)}$$

Note that for the last term to be larger than $k2^k$, it must be that

$$\ln(k^2 2^k) < k \Rightarrow k^2 2^k < e^k \Rightarrow k^2 < \left(\frac{e}{2}\right)^k$$

which is true for sufficiently large values of k.

The *Ramsey number* $R(k,l)$ is the smallest number n such that any graph with n vertices has clique of size k or an independent set of size l. Another way to formulate this is: in any two-coloring on edges of the complete graph on n vertices, there is a monochromatic clique of size k or a monochromatic clique of size l. Diagonal Ramsey Number asks for the value of $R(k,k)$ for any integer k. Finding a diagonal Ramsey number even for $k = 6$ is very difficult. We want to find a lower bound on $R(k,k)$.

Example. If $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, then $R(k,k) > n$. In particular, $R(k,k) > \lfloor 2^{\frac{k}{2}} \rfloor$, for $k \geq 3$.

Solution. Consider a complete graph G in which each edge is colored red or blue with a probability of 1/2. Let S be a any subset of k vertices and $E(S)$ be the set of edges with both endpoints in S.

$$\Pr[\text{edges in } E(S) \text{ are monochromatic}] = 2 \cdot 2^{-\binom{k}{2}}$$

Since there are $\binom{n}{k}$ subsets of size k, the probability that some subset of size k is monochromatic is at most

$$2\binom{n}{k}2^{-\binom{k}{2}} = \binom{n}{k}2^{1-\binom{k}{2}} \tag{7}$$

Since the last expression is less than 1 (given as a condition in the problem statement), there is a 2-coloring of edges of a complete graph on n vertices in which there is no monochromatic clique of size k. Thus $R(k,k) > n$.

If $n = \lfloor 2^{k/2} \rfloor$ then

$$\binom{n}{k}2^{1-\binom{k}{2}} \leq \frac{n^k}{k!} \cdot 2^{1-\frac{k(k-1)}{2}} \leq \left(\frac{2^{k^2/2}}{k!}\right) 2^{1-\frac{k^2}{2}+\frac{k}{2}} = \frac{2^{1+\frac{k}{2}}}{k!}$$

Note that the last expression is less than 1, if $k \geq 3$.

It can be shown that $R(k,k) < 4^k$. These are the best known bounds on the size of $R(k,k)$, so a lot of progress is yet to be made. What is known is that $R(2,2) = 2, R(3,3) = 6$, and $R(4,4) = 18$. The values of $R(k,k)$ are not known for $k \geq 5$.

A fundamental question in graph coloring is: what is the relation between $\chi(G)$ and the size of the largest clique? We state without proof that simply bounding the size of the largest clique does not allow us to bound $\chi(G)$.

Example. For any $k \geq 1$, there exist triangle-free graphs (size of the largest clique is at most 2) with chromatic number greater than k.

Solution. Let $G = (n, p)$ be a n-vertex graph, in which an edge between any two vertices is included with a probability of p.

Note that if $\chi(G) = k$ then there must be an independent set in G of size $\lceil n/k \rceil$. Thus, to show that $\chi(G) \geq k$, it suffices to show that the largest independent set in G is at most $\lceil n/k \rceil$. We will show that with a high probability, for a suitable value of p, G does not have an independent set of size $\lceil n/2k \rceil$.

Let I be the random variable denoting the number of independent sets of size $\lceil n/2k \rceil$ in G. For any set S consisting of $\lceil n/2k \rceil$ vertices, let I_S be an indicator random variable that is 1, iff S is an independent set. Thus we have

$$\begin{aligned}
\mathbf{E}[I_S] &= \Pr[I_S = 1] \\
&= (1-p)^{\binom{\lceil n/2k \rceil}{2}} \\
&\leq (1-p)^{\binom{n/2k}{2}} \\
&= (1-p)^{\frac{(n/2k)(n/2k-1)}{2}} \\
&= e^{-p(\frac{n^2}{8k^2} - \frac{n}{4k})} (\text{using } 1 + x \leq e^x, \text{ for all } x) \\
&\leq e^{-p(\frac{n^2}{16k^2})} \quad (\text{for } n \geq 4k) \\
&< 2^{-\frac{n^{1+\epsilon}}{16k^2}} \qquad (8)
\end{aligned}$$

The expected value of I can now be calculated as follows.

$$\begin{aligned}
I &= \sum_S I_S \\
\mathbf{E}[I] &= \sum_S \mathbf{E}[I_S] \\
&< \sum_S 2^{-\frac{n^{1+\epsilon}}{16k^2}} \qquad (\text{using } (8)) \\
&= \binom{n}{\lceil n/2k \rceil} 2^{-\frac{n^{1+\epsilon}}{16k^2}} \\
&< 2^n \times 2^{-\frac{n^{1+\epsilon}}{16k^2}} \\
&= 2^{n(1-\frac{n^\epsilon}{16k^2})}
\end{aligned}$$

We want $\mathbf{E}[I] \leq 1/2$. For this to happen, it suffices that

$$\begin{aligned} n(1 - \frac{n^\epsilon}{16k^2}) &\leq -1, \text{which holds if} \\ n - \frac{n^{1+\epsilon}}{16k^2} &\leq -1, \text{which holds if} \\ n + 1 &\leq \tfrac{n^{1+\epsilon}}{16k^2}, \text{which holds if} \\ 2n &\leq \tfrac{n^{1+\epsilon}}{16k^2}, \text{which holds if} \\ n &\leq \tfrac{n^{1+\epsilon}}{32k^2}, \text{which holds if} \\ n^\epsilon &\geq 32k^2, \text{which holds if} \\ n &\geq (32k^2)^{1/\epsilon} \end{aligned} \tag{9}$$

Thus, we have that for all $n \geq (32k^2)^{\frac{1}{\epsilon}}$, $\mathbf{E}[I] < 1/2$. By Markov's inequality, we have

$$\Pr[I \geq 1] \leq \mathbf{E}[I] < \frac{1}{2}$$

Let T be the random variable denoting the number of triangles. Fix a set of 3 vertices; the probability that they form a triangle is p^3. Summing this over all 3-subsets, we get

$$\begin{aligned} \mathbf{E}[T] &= \binom{n}{3} p^3 \\ &< \frac{n^3}{3!}(n^{\epsilon-1})^3 \\ &= \frac{n^{3\epsilon}}{6} \end{aligned}$$

Using Markov's inequality, we have

$$\Pr[T \geq n/2] \leq \frac{\mathbf{E}[T]}{n/2} < \frac{n^{3\epsilon}/6}{n/2} = \frac{1}{3n^{1-3\epsilon}}$$

Setting the last expression to be at most $1/3$, we have

$$\begin{aligned} \frac{1}{3n^{1-3\epsilon}} &\leq 1/3 \\ \epsilon &\leq 1/3 \end{aligned}$$

By plugging $\epsilon = 1/3$ in (9), we get $n \geq 2^{15}k^6$. Thus, we have that for all $n \geq 2^{15}k^6$, we have $\Pr[I \geq 1] + \Pr[T \geq n/2] < 1$. This means that there exists a graph G for which $I = 0$ and $T < n/2$. We now alter this graph G by deleting one vertex from each triangle in G. Let G' be the resulting triangle-free graph. We remove less than $n/2$ vertices from G, thus G' has at least $n/2$ vertices. Since G does not have an independent set of size $\lceil n/2k \rceil$, G' does not have an independent set of size $\lceil n/2k \rceil \leq \lceil |G'|/k \rceil$. Thus $\chi(G') > k$.

The *girth* of a graph G, $g(G)$, is the length of the smallest cycle in G. In triangle-free graphs, $g(G) > 3$. In 1954 B. Descartes constructively showed that triangle-free graphs can have high chromatic number, but this construction was complicated and contained many short cycles. In 1959, Paul Erdős used the probabilistic method to prove the existence of graphs with arbitrarily high girth and chromatic number.

Example (Erdős 1959) For every $g, k > 0$, there exists a graph G with $\chi(G) \geq k$ and $g(G) \geq g$.

Proofs

Introduction to Logic

A *proposition* is a statement that is either true or false. For example, "$2 + 2 = 4$" and "Donald Knuth is a faculty at Rutgers-Camden" are propositions, whereas "What time is it?", $x^2 < x + 40$ are not propositions.

We can construct compound propositions from simpler propositions by using some of the following connectives. Let p and q be arbitrary propositions.

Negation: $\tilde{p}$ (read as "not p") is the proposition that is true when p is false and vice-versa.

Conjunction: $p \wedge q$ (read as "p and q") is the proposition that is true when both p and q are true.

Disjunction: $p \vee q$ (read as "p or q") is the proposition that is true when at least one of p or q is true.

Exclusive Or: $p \oplus q$ (read as "p exclusive-or q") is the proposition that is true when exactly one of p and q is true is false otherwise.

Implication: $p \to q$ (read as "p implies q") is the proposition that is false when p is true and q is false and is true otherwise.

The implication $q \to p$ is called the *converse* of the implication $p \to q$. The implication $\neg p \to \neg q$ is called the *inverse* of $p \to q$. The implication $\neg q \to \neg p$ is the *contrapositive* of $p \to q$. *p only if q* means "if not q then not p", or equivalently if p then q.

Biconditional: $p \leftrightarrow q$ (read as "p if, and only if, q") is the proposition that is true if p and q have the same truth values and is false otherwise. "If and only if" is often abbreviated as iff.

The following truth table makes the above definitions precise.

p	q	$\neg p$	$p \wedge q$	$p \vee q$	$p \oplus q$	$p \to q$	$q \to p$	$p \leftrightarrow q$
T	T	F	T	T	F	T	T	T
T	F	F	F	T	T	F	T	F
F	T	T	F	T	T	T	F	F
F	F	T	F	F	F	T	T	T

Necessary and Sufficient Conditions: For propositions p and q,

p is a *sufficient* condition for q means that $p \to q$.
p is a *necessary* condition for q means that $\neg p \to \neg q$, or equivalently $q \to p$.
Why is $p \wedge q$ not the correct answer?

Thus p is a necessary and sufficient condition for q means "p iff q".

Logical Equivalence

Two compound propositions are logically equivalent if they always have the same truth value. Two statement p and q can be proved to be logically equivalent either with the aid of truth tables or using a sequence of previously derived logically equivalent statements.

Example. Show that $p \to q \equiv \neg p \vee q \equiv \neg q \to \neg p$.

Solution. The truth table below proves the above equivalence.

p	q	$\neg p$	$\neg q$	$p \to q$	$\neg p \vee q$	$\neg q \to \neg p$
T	T	F	F	T	T	T
T	F	F	T	F	F	F
F	T	T	F	T	T	T
F	F	T	T	T	T	T

Example. Show that $p \equiv \neg p \to C$ and $p \to q \equiv (p \wedge \neg q) \to C$.

p	q	$\neg p$	$\neg q$	$p \to q$	$p \wedge \neg q$	C	$\neg p \to C$	$(p \wedge \neg q) \to C$
T	T	F	F	T	F	F	T	T
T	F	F	T	F	T	F	T	F
F	T	T	F	T	F	F	F	T
F	F	T	T	T	F	F	F	T

The above equivalence forms the basis of proofs by contradiction.

The logic of Quantified Statements

Consider the statement $x < 15$. We can denote such a statement by $P(x)$, where P denotes the predicate "is less than 15" and x is the variable. This statement $P(x)$ becomes a proposition when x is assigned a value. In the above example, $P(8)$ is true while $P(18)$ is false.

Another way to convert the statement $P(x)$ into a proposition is through *quantification*. The two types of quantification that we will study are *universal quantification* and *existential quantification*. Using universal quantifier $\forall$ ("for all") alongside $P(x)$ means that the statement $P(x)$ is true for all elements in the domain of x. Thus the proposition $\forall x \in D, P(x)$ is true when $P(x)$ is true for all $x \in D$ and is false if there is an element $x' \in D$ for which $P(x')$ is false. Using existential quantifier $\exists$ ("there exists") alongside $P(x)$ means that there exists an element in the domain of x for which $P(x)$ is true. Thus the proposition $\exists x \in D, P(x)$ is true if there is an $x' \in D$ for which $P(x')$ is true and is false if $P(x)$ is false for all $x \in D$.

Examples of propositions using quatifiers are as follows.

1. $\forall x \in \mathbb{Z}, x^3 + 1$ is composite.
2. $\forall x \in \mathbb{Z}, x$ is even $\rightarrow x + 1$ is odd.
3. $\exists x \in \mathbb{N}, x^2 \not> x$.
4. $\exists x \in \mathbb{Z}, 2|x$ and $2|x + 1$.
5. $\forall x \in \mathbb{Z}\, \exists y \in \mathbb{Z}, x + y = 0$.
6. $\exists x \in \mathbb{Z}\, \forall y \in \mathbb{Z}, x > y$.

Sometimes it helps (in proofs) to consider the negation of a proposition. Verify the following equivalence.

$$\begin{aligned} \neg(\forall x \in D, P(x)) &\equiv \exists x \in D, \neg P(x) \\ \neg(\exists x \in D, P(x)) &\equiv \forall x \in D, \neg P(x) \end{aligned}$$

Proofs

We will illustrate some proof techniques by proving some properties about numbers. Before we do that let's go through some basic definitions given below.

An integer n is *even* iff $n = 2k$ for some integer k. An integer is *odd* iff $n = 2k+1$ for some integer k. Symbolically,

$$\begin{aligned} n \text{ is even} &\leftrightarrow \exists \text{ an integer } k \text{ s.t. } n = 2k \\ n \text{ is odd} &\leftrightarrow \exists \text{ an integer } k \text{ s.t. } n = 2k + 1 \end{aligned}$$

An integer n is *prime* iff $n > 1$ and for all positive integers r and s, if $n = r \cdot s$, then $r = 1$ or $s = 1$. Otherwise n is *composite*.

Given any real number x, the *floor of* x, denoted by $\lfloor x \rfloor$, is defined as follows

$$\lfloor x \rfloor = n \leftrightarrow n \leq x < n + 1, \text{ where } n \text{ is an integer}$$

Given any real number x, the *ceiling of* x, denoted by $\lceil x \rceil$, is defined as follows

$$\lceil x \rceil = n \leftrightarrow n - 1 < x \leq n, \text{ where } n \text{ is an integer}$$

A real number is *rational* iff it can be expressed as a ratio of two integers with a non-zero denominator. A real number that is not rational is *irrational*. More formally,

$$r \text{ is rational} \leftrightarrow \exists \text{ integers } a \text{ and } b \text{ such that } r = a/b \text{ and } b \neq 0.$$

Example. Prove the following: If the sum of two integers is even then so is their difference.

Solution. Let m and n be particular but arbitrarily chosen integers such that $m+n$ is even. By definition of even, we have $m+n=2k$, for some integer k. Then

$$m = 2k - n$$

Now $m-n$ can be written as follows.

$$\begin{aligned} m-n &= 2k-n-n \\ &= 2(k-n) \end{aligned}$$

Since k and n are integers, $k-n$ is an integer, $2(k-n)$ is even and hence $m-n$ is even.

Example. Prove that, for all integers n, if n is odd then n^2+n+1 is odd.

Solution. Since n is odd $n=2k+1$ for some integer k. Then,

$$\begin{aligned} n^2+n+1 &= (2k+1)^2+2k+1+1 \\ &= 4k^2+4k+1+2k+2 \\ &= 4k^2+6k+2+1 \\ &= 2(2k^2+3k+1)+1 \end{aligned}$$

Since k is an integer, $p=2k^2+3k+1$ is an integer and n^2+n+1 is odd, since $n^2+n+1=2p+1$ where p is an integer.

Example. Let x be an integer. If $x>1$, then x^3+1 is composite.

Solution. Let x be an arbitrary but specific integer such that $x>1$. We can rewrite x^3+1 as $(x+1)(x^2-x+1)$. Note that since x is an integer both $(x+1)$ and (x^2-x+1) are integers. Hence $(x+1)|x^3+1$ and $(x^2-x+1)|x^3+1$. We now need to show that $x+1>1$ and $x^2-x+1>1$. Since $x>1$, clearly, $x+1>1$. $x^2-x+1>1$ by the following reasoning.

$$\begin{aligned} x &> 1 & \\ x^2 &> x & \text{(Multiplying both sides by } x.) \\ x^2-x &> 0 & \text{(Subtracting both sides by } x.) \\ x^2-x+1 &> 1 & \text{(Adding 1 to both sides.)} \end{aligned}$$

We can also argue that $x^2-x+1>1$ by showing that $x+1<x^3+1$. Since $x>1$ we have $x^2>x$ and hence $x^2>1$. Multiplying both sides by x again we get $x^3>x$. This means that $x+1<x^3+1$ and since $(x+1)|x^3+1$, we conclude that x^3+1 is composite.

Note: One student asked the question that why can't we write x^3+1 as $x^3(1+\frac{1}{x^3})$. The reason is that for an integer $x>1$, $(1+\frac{1}{x^3})$ is not an integer and the proof breaks down.

Example. Prove that, for all real numbers x and all integers m,

$$\lfloor x + m \rfloor = \lfloor x \rfloor + m$$

Solution. Let $x = y + \epsilon$, where y is the largest integer with value at most x and $0 \leq \epsilon < 1$. Then,

$$\begin{aligned} x + m &= y + \epsilon + m \\ \lfloor x + m \rfloor &= \lfloor y + m + \epsilon \rfloor \\ &= y + m \\ &= \lfloor x \rfloor + m \end{aligned}$$

Example. Prove that if x and y are integers where $x + y$ is even, then x and y are both odd or both even.

Solution. To prove the above claim we will prove its contrapositive which is "if exactly one of x or y is even then $x + y$ is odd". Without loss of generality, for some integers k and l, let $x = 2k$ be even and $y = 2l + 1$ be odd. Then,

$$\begin{aligned} x + y &= 2k + 2l + 1 \\ &= 2(k + l) + 1 \end{aligned}$$

Since k and l are integers so is $k + l$ and $2(k + l)$ is even and hence $x + y$ is odd.

Example. Show that at least three of any 25 days chosen must fall in the same month of the year.

Solution. Assume for contradiction that the proposition "at least three of any 25 days chosen must fall in the same month of the year" is not true. This means that each month can have at most two of the 25 days chosen. Since there are 12 months, there can be at most 24 days that must have been chosen. This contradicts the premise that we chosen 25 days. In other words, by assuming that the proposition in the question is false, we have proved that (25 days are chosen) and (at most 24 days are chosen), which is clearly a contradiction.

Example. If $3n + 2$ is odd then n is odd.

Solution. We will show the above claim is true by giving a proof by contradiction. Thus assume that $3n + 2$ is odd and n is even. Since n is even, there exists an integer k such that $n = 2k$. Thus $3n + 2$ can be written as

$$3(2k) + 2 = 2(3k + 1)$$

Since k is an integer, clearly $3k+1$ is an integer. Thus $3n+2$ is even. Note that our premise is that $3n + 2$ is odd and we have shown that $3n + 2$ is even. This is a contradiction. This proves the claim.

Example. Prove that for all real numbers a and b, if the product ab is an irrational number, then either a or b, or both must be irrational.

Solution. We will prove the above claim by proving the contrapositive. That is, we will show that if both a and b are rational numbers then their product ab is a rational number. Let $a = p/q$ and $b = r/s$, where p, q, r, and s are integers and $q \neq 0$ and $s \neq 0$. The product ab can be expressed as follows.

$$ab = \frac{p}{q} \cdot \frac{r}{s} = \frac{pr}{qs}$$

Note that the numerator pr is an integer and so is the denominator qs. Also, since $q \neq 0$ and $s \neq 0$, the denominator $qs \neq 0$. Thus ab is a rational number.

A Brief Detour: Set Operations.

We will make a small detour to understand operations on sets. Below are some definitions.

- Let A and B be sets. The *union* of the sets A and B, denoted by $A \cup B$, is the set that contains those elements that are either in A or in B, or in both. As an example, if $A = \{\text{Ron, Bob, Kelly}\}$ and $B = \{\text{Tim, Ryan, Bob}\}$ then $A \cup B = \{\text{Ron, Bob, Kelly, Tim, Ryan}\}$.

- Let A and B be sets. The *intersection* of the sets A and B, denoted by $A \cap B$, is the set that contains those elements that are in both A and B. For example, if $A = \{\text{Ron, Bob, Kelly}\}$ and $B = \{\text{Tim, Ryan, Bob}\}$ then $A \cap B = \{\text{Bob}\}$.

- Two sets are called *disjoint* if their intersection is an empty set.

- A collection of nonempty sets $\{A_1, A_2, \ldots, A_n\}$ is a *partition* of a set A if, and only if, (i) $A = \bigcup_{i=1}^{n} A_i$ and (ii) $A_1, A_2, \ldots, A_n$ are mutually (pairwise) disjoint.

- Let A and B be two sets. The *difference* of A and B, denoted by $A \setminus B$ (or $A - B$) is the set containing those elements that are in A but not in B. For example, if $A = \{1, 2, 3, 4\}$ and $B = \{2, 3, 4, 6, 8\}$ then $A \setminus B = \{1\}$.

- The *complement* of a set A is the set of elements not in A. It is denoted by $\overline{A}$. Thus, if U is the universe of elements in consideration, then the complement of set A is given by
$$\overline{A} = U \setminus A$$
As an example, if $U = \mathbb{N}$ and A is the set of non-negative even integers, then $\overline{A}$ is the set of all positive odd integers.

- Let A and B be sets. The *cartesian product* of A and B, denoted by $A \times B$, is the set of all ordered pairs formed by taking an element from A together with an element from B in all possible ways. That is, $A \times B = \{(a, b) \mid a \in A, b \in B\}$.

Example. Let $A = \{2^1, 2^2, 2^3, \ldots\}$ and $B = \{2, 4, 6, \ldots\}$. Prove that $A \subseteq B$.

Solution. Let x be an arbitrary but particular element in A. Element x is of the form 2^j, for some positive integer j. Note that an element in B is of the form $2 \cdot i$, for some $i \in \{1, 2, 3, \ldots\}$. Clearly, $x = 2^j = 2 \cdot i$, where $i = 2^{j-1}$. Since j is positive, $j - 1 \geq 0$ and hence $i \geq 1$. Thus $x \in B$ and hence we conclude that $A \subseteq B$.

Example. Let A and B be sets. Then, $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

Solution. ($\Rightarrow$): If $A = B$ then since every set is a subset of itself, we have $A \subseteq B$ and $B \subseteq A$.
($\Leftarrow$): Let x be any element in A. Since $A \subseteq B$, x is also an element in B. Similarly, if an element $y \in B$, since $B \subseteq A$, y is also an element in A. Thus there is no element in A that is not in B and there is no element in B that is not in A, that is, A and B have the same elements. By definition, $A = B$.

Example. Let $A = \{n \mid n = 2k + 5 \text{ for some } k \in \mathbb{N}\}$ and $B = \{n \mid n = 2j + 1 \text{ for some } j \in \mathbb{N}\}$. Is $A \subseteq B$?

Solution. Let x be any arbitrary but particular element in A. Then,

$$\begin{aligned} x &= 2k + 5, \quad \text{for some integer } k. \\ &= 2(k + 2) + 1 \end{aligned}$$

Since $k \in \mathbb{N}$, $k + 2 \in \mathbb{N}$, and hence we have proved that any arbitrary element $x \in A$ also belongs to the set B. Thus $A \subseteq B$.

Example. Let $A = \{n \in \mathbb{N} \,|\, n = 2k^2 - 3, \text{ for some } k \in \mathbb{N}\}$ and $B = \{n \in \mathbb{N} \,|\, n = j^2 + 3 \text{ for some } j \in \mathbb{N}\}$. Prove that $A \not\subseteq B$.

Solution. Note that $5 \in A$, since $5 = 2 \cdot 2^2 - 3$. Observe that for 5 to be an element of B, $5 = j^2 + 3$, that is, $j^2 = 2$, which is impossible since j must be a natural number. Thus we have found an element of A that does not belong to B and hence $A \not\subseteq B$.

Example. Let $A = \{n \in \mathbb{N} \,|\, n \geq 2 \text{ and } n = 4j - 5, \text{ for some } j \in \mathbb{N}\}$ and $B = \{n \in \mathbb{N} \,|\, n \geq 0 \text{ and } n = 2k + 1 \text{ for some } k \in \mathbb{N}\}$. Prove that $A \subset B$.

Solution. Let x be an arbitrary but particular element of A. We know that x is of the form $4j - 5$, where $j \in \mathbb{N}$. Thus we get

$$\begin{aligned} x &= 4j - 5 \\ &= 2 \cdot 2j - 6 + 1 \\ &= 2(2j - 3) + 1 \end{aligned}$$

Since $x \geq 2$, it must be that $4j - 5 \geq 2$. Solving for j gives us $j \geq 7/4$. Since $j \in \mathbb{N}$, we have $j \geq 2$. Thus the integer $2j - 3 \geq 1$. Thus $x \in B$ and hence $A \subseteq B$.

Note that the element $1 \in B$, but it does not belong to A. Hence $A \subset B$.

Example. Recall the *cartesian product* of A and B, denoted by $A \times B$, is the set of all ordered pairs formed by taking an element from A together with an element from B in all possible ways. That is, $A \times B = \{(a, b) \,|\, a \in A, b \in B\}$. Prove that if A and B are non-empty sets then $A \times B = B \times A$ iff $A = B$.

Solution. First we will prove that if $A = B$ then $A \times B = B \times A$. Since $A = B$, $A \times B = A \times A = B \times A$.

Now assume that $A \times B = B \times A$. We will show that $A = B$. Let x be any arbitrary but particular element in A. Consider any element $y \in B$. Note that y must exist since $B \neq \emptyset$. Since $A \times B = B \times A$, the element (x, y) is in $A \times B$ as well as $B \times A$. Hence $x \in B$, which means $A \subseteq B$. The proof for $B \subseteq A$ is along the same lines.

Do you see why the condition that A are B are non-empty is necessary? Suppose that one of the sets is empty and the other is not. Then $A \times B = B \times A = \emptyset$, but $A \neq B$.

DeMorgan's Laws Let A, B, and C be sets. Then

$$\begin{aligned} A - (B \cup C) &= (A - B) \cap (A - C) \\ A - (B \cap C) &= (A - B) \cup (A - C) \end{aligned}$$

Example. Prove that the product of two odd numbers is an odd number.

Solution. Let x and y be particular but arbitrarily chosen odd numbers. Then, $x = 2k+1$ and $y = 2l + 1$, for some integers k and l. We have

$$x \cdot y = (2k + 1) \cdot (2l + 1) = 4kl + 2(k + l) + 1 = 2(2kl + k + l) + 1$$

Let $p = 2kl + k + l$. Since k and l are integers, p is an integer and $x \cdot y = 2p + 1$ is odd.

Example. Prove that $\sqrt{2}$ is irrational.

Solution. For the purpose of contradiction, assume that $\sqrt{2}$ is a rational number. Then there are integers a and b ($b \neq 0$) with no common factors such that

$$\sqrt{2} = \frac{a}{b}$$

Squaring both sides of the above equation gives

$$\begin{aligned} 2 &= \frac{a^2}{b^2} \\ a^2 &= 2b^2 \end{aligned} \tag{10}$$

From (10) we conclude that a^2 is even. This fact combined with the result of previous example implies that a is even. Then, for some integer k, let

$$a = 2k \tag{11}$$

Combining (10) and (11) we get

$$\begin{aligned} 4k^2 &= 2b^2 \\ 2k^2 &= b^2 \end{aligned}$$

The above equation implies that b^2 is even and hence b is even. Since we know a is even this means that a and b have 2 as a common factor which contradicts the assumption that a and b have no common factors.

We will now give a very elegant proof for the fact that "$\sqrt{2}$ is irrational" using the *unique factorization theorem* which is also called the *fundamental theorem of arithmetic*.

The unique factorization theorem states that every positive number can be uniquely represented as a product of primes. More formally, it can be stated as follows.

> Given any integer $n > 1$, there exist a positive integer k, distinct prime numbers $p_1, p_2, \ldots, p_k$, and positive integers $e_1, e_2, \ldots, e_k$ such that
>
> $$n = p_1^{e_1} p_2^{e_2} p_3^{e_3} \cdots p_k^{e_k}$$
>
> and any other expression of n as a product of primes is identical to this except, perhaps, for the order in which the factors are written.

Example. Prove that $\sqrt{2}$ is irrational using the unique factorization theorem.

Solution. Assume for the purpose of contradiction that $\sqrt{2}$ is rational. Then there are integers a and b ($b \neq 0$) such that

$$\sqrt{2} = \frac{a}{b}$$

Squaring both sides of the above equation gives

$$\begin{aligned} 2 &= \frac{a^2}{b^2} \\ a^2 &= 2b^2 \end{aligned}$$

Let $S(m)$ be the sum of the number of times each prime factor occurs in the unique factorization of m. Note that $S(a^2)$ and $S(b^2)$ is even. Why? Because the number of times that each prime factor appears in the prime factorization of a^2 and b^2 is exactly twice the number of times that it appears in the prime factorization of a and b. Then, $S(2b^2) = 1 + S(b^2)$ must be odd. This is a contradiction as $S(a^2)$ is even and the prime factorization of a positive integer is unique.

Example. Prove or disprove that the sum of two irrational numbers is irrational.

Solution. The above statement is false. Consider the two irrational numbers, $\sqrt{2}$ and $-\sqrt{2}$. Their sum is $0 = 0/1$, a rational number.

Example. Show that there exist irrational numbers x and y such that x^y is rational.

Solution. We know that $\sqrt{2}$ is an irrational number. Consider $\sqrt{2}^{\sqrt{2}}$.

Case I: $\sqrt{2}^{\sqrt{2}}$ is rational.
In this case we are done by setting $x = y = \sqrt{2}$.

Case II: $\sqrt{2}^{\sqrt{2}}$ is irrational.
In this case, let $x = \sqrt{2}^{\sqrt{2}}$ and let $y = \sqrt{2}$. Then, $x^y = \left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = (\sqrt{2})^2 = 2$, which is an integer and hence rational.

Example. Prove that for all positive integers n,

$$n \text{ is even} \leftrightarrow 7n + 4 \text{ is even}$$

Solution. Let n be a particular but arbitrarily chosen integer.
Proof for n is even $\rightarrow$ $7n + 4$ is even. Since n is even, $n = 2k$ for some integer k. Then,

$$7n + 4 = 7(2k) + 4 = 2(7k + 2)$$

Hence, $7n + 4$ is even.

Proof for $7n + 4$ is even $\rightarrow$ n is even. Since $7n + 4$ is even and n is a positive integer, let $7n + 4 = 2l$ for some integer $l \geq 6$. Then,

$$7n = 2l - 4 = 2(l - 2)$$

Clearly, $7n$ is even. Combining the fact that 7 is odd with the result of the Example 1, we conclude that n is even.

We can also prove the latter by proving its contrapositive, i.e., we can prove

$$\text{if } n \text{ is odd then } 7n + 4 \text{ is odd.}$$

Since n is a positive odd integer, we have $n = 2k + 1, \text{for some integer} k \geq 0$. Thus we have

$$\begin{aligned} 7n + 4 &= 7(2k + 1) + 4 \\ &= 14k + 10 + 1 \\ &= 2(7k + 5) + 1 \\ &= 2k' + 1, \text{where } k' = 7k + 5 \text{ is an integer.} \end{aligned}$$

Example. Prove that there are infinitely many prime numbers.

Solution. Assume, for the sake of contradiction, that there are only finitely many primes. Let p be the largest prime number. Then all the prime numbers can be listed as

$$2, 3, 5, 7, 11, 13, \ldots, p$$

Consider an integer n that is formed by multiplying all the prime numbers and then adding 1. That is,

$$n = (2 \times 3 \times 5 \times 7 \times \cdots p) + 1$$

Clearly, $n > p$. Since p is the largest prime number, n cannot be a prime number. In other words, n is composite. Let q be any prime number. Because of the way n is constructed, when n is divided by q the remainder is 1. That is, n is not a multiple of q. This contradicts

the Fundamental Theorem of Arithmetic.

Alternate Proof by Filip Saidak. Let n be an arbitrary positive integer greater than 1. Since n and $n+1$ are consecutive integers, they must be relatively prime. Hence, the number $N_2 = n(n+1)$ must have at least two different prime factors. Similarly, since the integers $n(n+1)$ and $n(n+1)+1$ are consecutive, and therefore relatively prime, the number

$$N_3 = n(n+1)[n(n+1)+1]$$

must have at least three different prime factors. This process can be continued indefinitely, so the number of primes must be infinite.

Mathematical Induction

Example. Prove that for all integers $n \geq 1$,

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

Solution. We will prove the claim using induction on n.
Induction hypothesis: Assume that the claim is true when $n = k$, for some $k \geq 1$. In other words assume that

$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2}$$

Base Case: $n = 1$. The claim is true for $n = 1$ as both sides of the equation equal to 1.
Induction step: To prove that the claim is true when $n = k+1$. That is, we want to show that

$$\sum_{i=1}^{k+1} = \frac{(k+1)(k+2)}{2}$$

We can do this as follows.

$$\begin{aligned}
\sum_{i=1}^{k+1} i &= \sum_{i=1}^{k} i + (k+1) \\
&= \frac{k(k+1)}{2} + k + 1 \qquad \text{(using induction hypothesis)} \\
&= \frac{k(k+1) + 2(k+1)}{2} \\
&= \frac{(k+1)(k+2)}{2}
\end{aligned}$$

Example. Prove that the sum of the first n positive odd numbers is n^2.

Solution. We want to prove that $\forall$ positive integers $n, P(n)$ where $P(n)$ is the following property.

$$\sum_{i=0}^{n-1} 2i + 1 = n^2$$

Base Case: We want to show that $P(1)$ is true. This is clearly true as

$$\sum_{i=0}^{0} 2i + 1 = 1 = 1^2$$

Induction Hypothesis: Assume $P(k)$ is true for some $k \geq 1$.
Induction Step: We want to show that $P(k+1)$ is true, i.e., we want to show that

$$\sum_{i=0}^{k} 2i + 1 = (k+1)^2$$

We can do this as follows.

$$\begin{aligned} \sum_{i=0}^{k} 2i + 1 &= \sum_{i=0}^{k-1} 2i + 1 + 2k + 1 \\ &= k^2 + 2k + 1 \qquad \text{(using induction hypothesis)} \\ &= (k+1)^2 \end{aligned}$$

Example. Show that for all integers $n \geq 0$, if $r \neq 1$,

$$\sum_{i=0}^{n} ar^i = \frac{a(r^{n+1} - 1)}{r - 1}$$

Solution. Let r be any real number that is not equal to 1. We want to prove that $\forall$ integers $n, P(n)$, where $P(n)$ is given by

$$\sum_{i=0}^{n} ar^i = \frac{a(r^{n+1} - 1)}{r - 1}$$

Base Case: We want to show that $P(0)$ is true.

$$\sum_{i=0}^{0} ar^i = a = \frac{a(r-1)}{r-1}$$

Induction Hypothesis: Assume that $P(k)$ is true for some $k \geq 0$.
Induction Step: We want to show that $P(k+1)$ is true, i.e., we want to prove that

$$\sum_{i=0}^{k+1} ar^i = \frac{a(r^{k+2}-1)}{r-1}$$

We can do this as follows.

$$\begin{aligned}
\text{L.H.S.} &= \sum_{i=0}^{k+1} ar^i \\
&= \sum_{i=0}^{k} ar^i + ar^{k+1} \\
&= \frac{ar^{k+1}-a}{r-1} + ar^{k+1} \\
&= \frac{a(r^{k+1}-1)}{r-1} + \frac{ar^{k+1}(r-1)}{r-1} \\
&= \frac{a}{r-1}\left(r^{k+1}(1+r-1)-1\right) \\
&= \frac{a}{r-1}\left(r^{k+2}-1\right) \\
&= \frac{a(r^{k+2}-1)}{r-1}
\end{aligned}$$

Example. Prove that $\forall$ non-negative integers n,

$$\sum_{i=0}^{n} 2^i = 2^{n+1} - 1$$

Solution. By setting $a = 1$, $r = 2$ in the result of the previous problem, the claim follows.

Example. Prove that $\forall$ non-negative integers n, $2^{2n} - 1$ is a multiple of 3.

Solution. We want to prove that $\forall$ non-negative integers n, $P(n)$, where $P(n)$ is

$$2^{2n} - 1 = 3k, \text{ for some non-negative integer } k$$

Base Step: $P(0)$ is true as shown below.

$$2^0 - 1 = 0 = 3 \cdot 0.$$

Induction Hypothesis: Assume that $P(x)$ is true for some $x \geq 0$, i.e., $2^{2x} - 1 = 3 \cdot k'$, for some $k' \geq 0$.
Induction Step: We want to prove that $P(x+1)$ is true, i.e., we want to show that

$$2^{2(x+1)} - 1 = 3l, \text{ for some non-negative integer } l.$$

We can show this as follows.

$$\begin{aligned}
\text{L.H.S.} &= 2^{2(x+1)} - 1 \\
&= 2^{2x+2} - 1 \\
&= 2^{2x} \cdot 2^2 - 1 \\
&= 2^{2x} \cdot 4 - 1 \\
&= 2^{2x} \cdot (3+1) - 1 \\
&= 3 \cdot 2^{2x} + 2^{2x} - 1 \\
&= 3 \cdot 2^{2x} + 3 \cdot k' \qquad \text{(using induction hypothesis)} \\
&= 3(2^{2x} + k') \\
&= 3l, \qquad \text{where } l = 2^{2x} + k'
\end{aligned}$$

Since x and k' are integers l is also an integer. Hence, $P(x+1)$ is true.

Example. Prove that $\forall n \in \mathbb{N}, n > 1 \rightarrow n! < n^n$.

Solution. Below is a simple direct proof for this inequality.

$$\begin{aligned}
n! &= 1 \times 2 \times 3 \times \cdots \times n \\
&< n \times n \times n \times \cdots \times n \\
&= n^n
\end{aligned}$$

We now give a proof using induction. Let $P(n)$ denote the following property.

$$n! < n^n$$

Induction Hypothesis: Assume that $P(k)$ is true for some $k > 1$.
Base Case: We want to prove $P(2)$. $P(2)$ is the proposition that $2! < 2^2$, or $2 < 4$, which is true.
Induction Step: We want to prove $P(k+1)$, i.e., we want to prove that $(k+1)! < (k+1)^{k+1}$.

$$\begin{aligned}
\text{L.H.S.} &= (k+1)! \\
&= k! \times (k+1) \\
&< k^k \times (k+1) \qquad \text{(using induction hypothesis)} \\
&< (k+1)^k \times (k+1) \quad \text{(since } k > 1\text{)} \\
&= (k+1)^{k+1}
\end{aligned}$$

Example. Recall that for any set A, $\mathcal{P}(A)$ denotes the power set of A. Let $S = \{x_1, x_2, \ldots, x_n\}$. Prove using induction that for all positive integers n, $|\mathcal{P}(S)| = 2^n$.

Solution. We will prove the claim using induction on n.
Induction Hypothesis: Assume that the claim is true when $n = k$, for some $k \geq 1$. In other words, assume that if $S = \{x_1, x_2, \ldots, x_k\}$, then $|\mathcal{P}(S)| = 2^k$.
Base Case:$n = 1$. When $S = \{x_1\}$, there are exactly two subsets of S, namely $\emptyset$ and S, itself. Thus the claim is true when $n = 1$.
Induction Step: We want to prove that the claim is true when $n = k + 1$. In other words, we want to show that if $S = \{x_1, x_2, \ldots, x_k, x_{k+1}\}$, then $|\mathcal{P}(S)| = 2^{k+1}$. Let $S' = \{x_1, x_2, \ldots, x_k\}$. The set of all subsets of S can be partitioned into S_1 and S_2, where $S_1 \subset \mathcal{P}(S)$ contains subsets of S that does not contain x_{k+1}, and $S_2 \subset S$ contains subsets of $\mathcal{P}(S)$ that contains x_{k+1}. Thus we have

$$|\mathcal{P}(S)| = |S_1| + |S_2| \tag{12}$$

Note that S_1 contains all subsets of $\mathcal{P}(S')$. By the induction hypothesis, we have $|S_1| = |\mathcal{P}(S')| = 2^k$. We will now compute $|S_2|$. Observe that each set in S_2 is of the form $\{x_{k+1}\} \cup X$, where X is a subset of S'. By induction hypothesis, we know that there are 2^k subsets of S' and hence $|S_2| = 2^k$. Plugging in the values for $|S_1|$ and $|S_2|$ in (12), we get

$$|\mathcal{P}(S)| = 2^k + 2^k = 2^{k+1}$$

Example Let $A_1, A_2, \ldots, A_n$ be sets (where $n \geq 2$). Suppose for any two sets A_i and A_j either $A_i \subseteq A_j$ or $A_j \subseteq A_i$. Prove by induction that one of these n sets is a subset of all of them.

Solution. We will prove the claim using induction on n.
Induction Hypothesis: Assume that the claim is true when $n = k$, for some $k \geq 2$. In other words, assume that if we have sets $A_1, A_2, \ldots, A_k$, where for any two sets A_i and A_j, either $A_i \subseteq A_j$ or $A_j \subseteq A_i$ then one of the k sets is a subset of all of the k sets.
Base Case: $n = 2$. We have two sets A_1, A_2 and we know that $A_1 \subseteq A_2$ or $A_2 \subseteq A_1$. Without loss of generality assume that $A_1 \subseteq A_2$. Then A_1 is a subset of A_1 and is also a subset of A_2, so the claim holds when $n = 2$.
Induction Step: We want to prove the claim when $n = k + 1$. That is, we are given a set $S = \{A_1, A_2, \ldots, A_{k+1}\}$ of with the property that for every pair of sets $A_i \in S$ and $A_j \in S$, either $A_i \subseteq A_j$ or $A_j \subseteq A_i$. We want to show that there is a set in S that is a subset of all $k + 1$ sets in S. Let $S' = S \setminus \{A_{k+1}\}$. By induction hypothesis, there is a set $A_p \in S'$ that is a subset of all sets in S'. We now consider the following two cases.
Case 1: $A_p \subseteq A_{k+1}$. Then it follows that A_p is a subset of all sets in S.
Case 2: $A_{k+1} \subseteq A_p$. Since A_p is a subset of all sets in S' and $A_{k+1} \subseteq A_p$, it follows that A_{k+1} is a subset of all sets in S.

Example. For all $n \geq 1$, prove that n lines separate the plane into $(n^2 + n + 2)/2$ regions. Assume that no two of these lines are parallel and no three pass through a common point.

Solution. Let $P(n)$ be the property that n lines, such that no two of them are parallel and no three of them pass through a common point, separate the plane into $(n^2+n+2)/2$ regions. We will prove the claim by induction on n.

Induction Hypothesis: Assume that $P(k)$ is true for some $k > 0$.
Base Case: $P(1)$ is true since one line divides the plane into 2 regions which is also given by $(1^2+1+2)/2$.
Induction Step: To prove that $P(k+1)$ is true. Consider a set S of $k+1$ lines such that no two of them are parallel and no three of them pass through a common point. Remove any line ℓ from the set S. Let S' be the resulting set of k lines. By induction hypothesis, the k lines in S' divide the plane into $(k^2+k+2)/2$ regions. Now we add the line ℓ to the set S' to obtain the set S. Line ℓ intersects exactly once with each of the k lines in S'. These intersections divide the line ℓ into $k+1$ line segments. Each of these line segments passes through a region and hence $k+1$ additional regions are created. Hence, the total number of regions formed by $k+1$ lines is given by

$$\frac{k^2+k+2}{2} + k + 1 = \frac{k^2+3k+4}{2} = \frac{k^2+2k+1+k+3}{2} = \frac{(k+1)^2+(k+1)+2}{2}$$

Thus $P(k+1)$ is correct and this completes the proof.

Example. Let n be a non-negative integer. Show that any $2^n \times 2^n$ region with one central square removed can be tiled using L-shaped pieces, where the pieces cover three squares at a time (Figure 4).

Solution. (Attempt 1) Let R_n denote a $2^n \times 2^n$ region. Let $P(n)$ be the property that R_n with one central square removed can be tiled using L-shaped pieces.

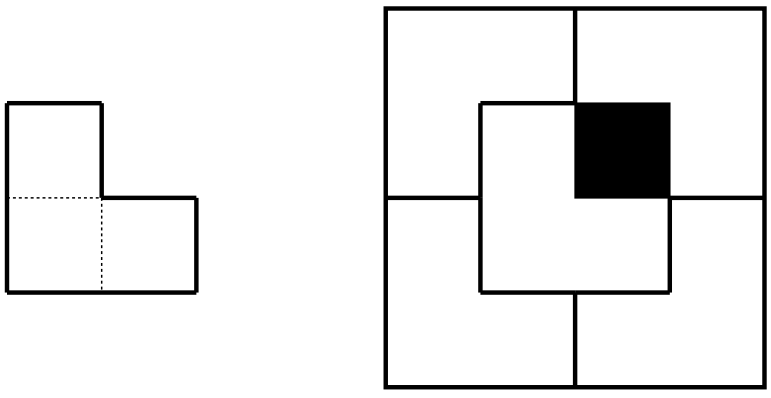


Figure 4: A L-tile and an L-tiling of a $2^2 \times 2^2$ region without a square.

Induction Hypothesis: Assume that $P(k)$ is true for some $k \geq 0$.
Base Case: We want to prove that $P(0)$ is true. This is true because a 1×1 region with one central square removed requires 0 tiles.
Induction Step: We want to prove that $P(k+1)$ is true, i.e., region R_{k+1} with one central

square removed can be tiled using L-shaped pieces.
R_{k+1} can be divided into four regions of size $2^k \times 2^k$. Note that the four central corners of R_{k+1} can be covered using one L-shaped tile and one square hole (Figure 5). Each of the four remaining regions has one hole and is of the size $2^k \times 2^k$. By induction hypothesis, these regions can be covered using L-shaped pieces. Thus, since the four disjoint regions can be covered using L-shaped tiles, R_{k+1} without a central square can also be covered using L-shaped tiles.

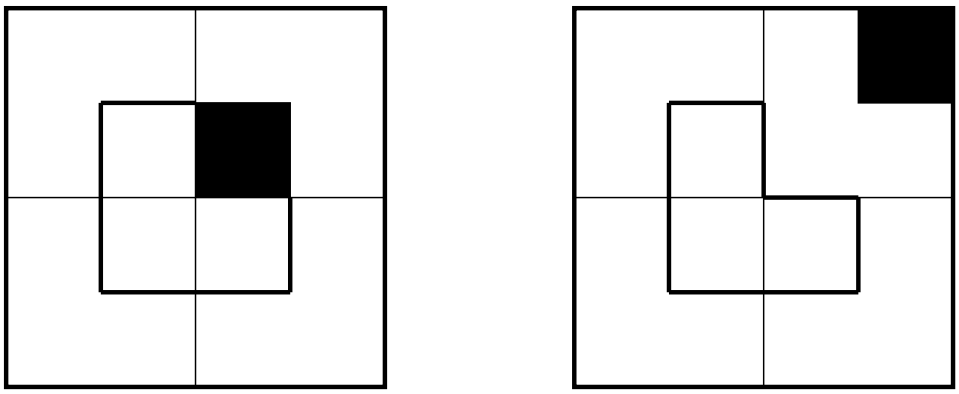


Figure 5: Illustration of the two proof attempts.

Our use of induction hypothesis is incorrect as we have assumed that region R_k without a *central* square (not a *corner* square) can be covered using L-shaped tiles.

Surprisingly, we can get around this obstacle by proving the following stronger claim.

"For all positive integers n, any R_n region with *any* one square removed can be L-tiled."

Let $P(n)$ be the property that R_n without one square can be L-tiled.
Induction Hypothesis: Assume that $P(k)$ is true for some k.
Base Case: We want to prove that $P(0)$ is true. This is true because a 1×1 region with one square removed requires 0 tiles.
Induction Step: We want to prove that $P(k+1)$ is true, i.e., region R_{k+1} wthout one square that is located anywhere can be L-tiled. Divide R_{k+1} into four R_k regions. One of the four R_k regions that does not have one square can be L-tiled (using induction hypothesis). Each of the other three R_k regions without the corner square that is located at the center of R_{k+1} can be L-tiled (using induction hypothesis). By using one more L-tile we can cover the three central squares of R_{k+1}.

Strong Induction.

For any property P, if $P(0)$ and $\forall n \in \mathbb{N}, P(0) \wedge P(1) \wedge P(2) \wedge \cdots \wedge P(k) \rightarrow P(k+1)$, then $\forall n \in \mathbb{N}, P(n)$.

Example. Prove that if n is an integer greater than 1 then either n is a prime or it can be written as a product of primes.

Solution. Let $P(n)$ be "n can be written as a product of primes".
Induction Hypothesis: Assume that $P(j)$ is true for $1 < j \leq k$.
Base Case: We want to show that $P(2)$ is true. This is clearly true as 2 is a prime.
Induction Step: We want to show that $P(k+1)$ is true.
Case I: $k+1$ is prime. In this case we are done.
Case II: $k+1$ is composite. Then,

$$k+1 = a \times b, \quad \text{for some } a \text{ and } b \text{ s.t. } 2 \leq a \leq b < k+1$$

By induction hypothesis, a is a prime or it can be written as a product of primes. The same applies to b. Since $k+1 = a \times b$, it can be written as a product of primes, namely those primes in the factorization of a and those in the factorization of b.

Example. Prove that, for any positive integer n, if $x_1, x_2, \ldots, x_n$ are n distinct real numbers, then no matter how the parenthesis are inserted into their product, the number of multiplications used to compute the product is $n-1$.

Solution. Let $P(n)$ be the property that "If $x_1, x_2, \ldots, x_n$ are n distinct real numbers, then no matter how the parentheses are inserted into their product, the number of multiplications used to compute the product is $n-1$".
Induction Hypothesis: Assume that $P(j)$ is true for all j such that $1 \leq j \leq k$.
Base Case: $P(1)$ is true, since x_1 is computed using 0 multiplications.
Induction Step: We want to prove $P(k+1)$. Consider the product of $k+1$ distinct factors, $x_1, x_2, \ldots, x_{k+1}$. When parentheses are inserted in order to compute the product of factors, some multiplication must be the final one. Consider the two terms, of this final multiplication. Each one is a product of at most k factors. Suppose the first and the second term in the final multiplication contain f_k and s_k factors. Clearly, $1 \leq f_k, s_k \leq k$. Thus, by induction hypothesis, the number of multiplications to obtain the first term of the final multiplication is $f_k - 1$ and the number of multiplications to obtain the second term of the final multiplication is $s_k - 1$. It follows that the number of multiplications to compute the product of $x_1, x_2, \ldots, x_k, x_{k+1}$ is

$$(f_k - 1) + (s_k - 1) + 1 = f_k + s_k - 1 = k + 1 - 1 = k$$

Example. The game of NIM is played as follows: Some positive number of sticks are placed on the ground. Two players take turns, removing one, two or three sticks. The player to remove the last stick loses.

A winning strategy is a rule for how many sticks to remove when there are n left. Prove that the first player has a winning strategy iff the number of sticks, n, is not $4k+1$ for any $k \in \mathbb{N}$.

Solution. We will show that if $n = 4k + 1$ then player 2 has a strategy that will force a win for him, otherwise, player 1 has a strategy that will force a win for him.

Let $P(n)$ be the property that if $n = 4k + 1$ for some $k \in N$ then the first player loses, and if $n = 4k, 4k + 2$, or $4k + 3$, the first player wins. This exhausts all possible cases for n.
Induction Hypothesis: Assume that for some $z \geq 1$, $P(j)$ is true for all j such that $1 \leq j \leq z$.
Base Case: $P(1)$ is true. The first player has no choice but to remove one stick and lose.
Induction Step: We want to prove $P(z + 1)$. We consider the following four cases.
Case I: $z + 1 = 4k + 1$, for some k. We have already handled the base case, so we can assume that $z + 1 \geq 5$. Consider what the first player might do to win: he can remove 1, 2, or 3 sticks. If he removes one stick then the remaining number of sticks $n = 4k$. By strong induction, the player who plays at this point has a winning strategy. So the player who played first loses. Similarly, if the first player removes two sticks or three sticks, the remaining number of sticks is $4(k - 1) + 3$ and $4(k - 1) + 2$ respectively. Again, the first player loses (using induction hypothesis). Thus, in this case, the first player loses regardless of what move he/she makes.
Case II: $z + 1 = 4k$, or $z + 1 = 4k + 2$, or $z + 1 = 4k + 3$. If the first player removes three sticks in the first case, one stick in the second case, and two sticks in the third case then the second player sees $4(k - 1) + 1$ sticks in the first case and $4k + 1$ sticks in the other two cases. By induction hypothesis, in each case the second player loses.

Example. Prove that the two forms of induction, weak induction and strong induction, are equivalent. In other words, prove that any statement that admits a strong induction proof can be proved using weak induction and vice-versa.

Solution. Suppose we want to show that a $P(n)$ is true for all positive integers $n \geq n_0$. The two forms of inductive proofs are as follows.

Weak Induction: Assume that

(a_w) $P(n_0)$ is true
(b_w) For any $k \geq n_0$, $P(k) \implies P(k + 1)$ is true.

Then, $P(n)$ is true for all positive integers $n \geq n_0$.

Strong Induction: Assume that

(a_s) $P(n_0)$ is true
(b_s) For any $k \geq n_0$, $P(n_0) \wedge P(n_0 + 1) \wedge \cdots \wedge P(k) \implies P(k + 1)$ is true.

Then, $P(n)$ is true for all positive integers $n \geq n_0$.

We will show that it is always possible to convert a strong induction proof into a weak induction proof and vice-versa.

The conversion from a weak induction proof to a strong induction proof is trivial, since (b_s) implies (b_w).

We now show that a strong induction proof can be converted to a weak induction proof. Let

$$Q(n) \doteq P(n_0) \wedge P(n_0+1) \wedge \cdots \wedge P(n)$$

Induction Hypothesis: Assume that $Q(k)$ is true for some $k \geq n_0$.
Base Case: Since $Q(n_0) = P(n_0)$ and we know that $P(n_0)$ is true from (a_s), $Q(n_0)$ is true.
Induction Step: We want to show that $Q(k) \implies Q(k+1)$. We have

$$\begin{aligned} Q(k) &\implies P(k+1) && \text{(from } (b_s)\text{)} \\ \therefore Q(k) &\implies Q(k) \wedge P(k+1) \\ \therefore Q(k) &\implies Q(k+1) \end{aligned}$$

Thus we have converted a strong induction proof in P to a weak induction proof in Q.

Graphs

A *graph* consists of two sets, a non-empty set, V, of vertices or nodes, and a possibly empty set, E, of 2-element subsets of V. Such is graph is denoted by $G = (V, E)$. Each element of E is called an *edge*. We say that an edge $\{u, v\} \in E$ *connects* vertices u and v. Two nodes u and v are *adjacent* if $\{u, v\} \in E$. Nodes adjacent to a vertex u are called *neighbors* of u. The number of neighbors of a vertex v is called the *degree* of v and is denoted by $deg(v)$. The value $\delta(G) = \min_{v \in V}\{deg(v)\}$ is the *minimum degree* of G, the value $\Delta(G) = \max_{v \in V}\{deg(v)\}$ is the *maximum degree* of G. An edge that connects a node to itself is called a *loop* and multiple edges between the same pair of nodes are called *parallel* edges. Graphs without loops and parallel edges are called *simple* graphs, otherwise they are called *multigraphs*. Unless specified otherwise, we will only deal with simple graphs.

Example. Prove that the sum of degrees of all nodes in a graph is twice the number of edges.

Solution. Since each edge is incident to exactly two vertices, each edge contributes two to the sum of degrees of the vertices. The claim follows.

Example. In any graph there are an even number of vertices of odd degree.

Solution. Let V_e and V_o be the set of vertices with even degree and the set of vertices with odd degree respectively in a graph $G = (V, E)$. Then,

$$\sum_{v \in V} deg(v) = \sum_{v \in V_e} deg(v) + \sum_{v \in V_o} deg(v)$$

The first term on R.H.S. is even since each vertex in V_e has an even degree. From the previous example, we know that L.H.S. of the above equation is even. Thus the second term on the R.H.S. must be even. Let $|V_o| = \ell$. We want to show that ℓ is even. Since each vertex in V_o has odd degree, we have

$$\begin{aligned}(2k_1 + 1) + (2k_2 + 1) + \cdots + (2k_\ell + 1) &\text{ is an even number}\\ 2(k_1 + k_2 + \cdots + k_\ell) + \ell &\text{ is an even number}\\ \therefore \ell &\text{ is an even number}\end{aligned}$$

This proves the claim.

A *walk* in G is a non-empty sequence $v_0e_0v_1e_1 \ldots e_{k-1}v_k$ of vertices and edges in G such that $e_i = \{v_i, v_{i+1}\}$ for all $i < k$. If the vertices in a walk are all distinct, we call it a *path* in G. Thus, a *path* in G is a sequence of distinct vertices $v_0, v_1, v_2, \ldots v_k$ such that for all i, $0 \leq i < k$, $\{v_i, v_{i+1}\} \in E$. The *length* of the walk (path) is k, the number of edges in the walk (resp. path). Note that the length of the walk (path) is one less than the number of vertices in the walk (path) sequence. If $v_o = v_k$, the walk (path) is *closed*. A closed path is called a *cycle*.
The graph $H = (V', E')$ is a *subgraph* of $G = (V, E)$ if $V' \subseteq V$ and $E' \subseteq E$. A graph G is *connected* if there is a path in G between its every pair of vertices. A graph H is a *connected component*("island") of G if (a) H is a subgraph of G, (b) H is connected, and (c) H is maximal, i.e., H is not contained in any other connected subgraph of G. In short, H is a connected component of G if H is a maximal subgraph of G that is connected.

We say that H is an *induced subgraph* of a graph G if the vertex set of H is a subset of the vertex set of G, and if u and v are vertices in H, then (u, v) is an edge in H iff (u, v) is an edge in G.

Example. Prove that every graph with n vertices and m edges has at least $n - m$ connected components.

Solution. We will prove this claim by doing induction on m.
Induction Hypothesis: Assume that for some $k \geq 0$, every graph with n vertices and k edges has at least $n - k$ connected components.
Base Case: $m = 0$. A graph with n vertices and no edges has n connected components as each vertex itself is a connected component. Hence the claim is true for $m = 0$.
Induction Step: We want to prove that a graph, G, with n vertices and $k + 1$ edges has at least $n-(k+1) = n-k-1$ connected components. Consider a subgraph G' of G obtained by removing any arbitrary edge, say $\{u, v\}$, from G. The graph G' has n vertices and k edges. By induction hypothesis, G' has at least $n - k$ connected components. Now add $\{u, v\}$ to G' to obtain the graph G. We consider the following two cases.
Case I: u and v belong to the same connected component of G'. In this case, adding the edge $\{u, v\}$ to G' is not going to change any connected components of G'. Hence, in this case the number of connected components of G is the same as the number of connected

components of G' which is at least $n - k > n - k - 1$.
CaseII: u and v belong to different connected components of G'. In this case, the two connected components containing u and v become one connected component in G. All other connected components in G' remain unchanged. Thus, G has one less connected component than G'. Hence, G has at least $n - k - 1$ connected components.

Example. Prove that every connected graph with n vertices has at least $n - 1$ edges.

Solution. We will prove the contrapositive, i.e., a graph G with $m \leq n-2$ edges is disconnected. From the result of the previous problem, we know that the number of components of G is at least

$$n - m \geq n - (n - 2) = 2$$

which means that G is disconnected. This proves the claim.

One could also have proved the above claim directly by observing that a connected graph has exactly one connected component. Hence, $1 \geq n - m$. Rearranging the terms gives us $m \geq n - 1$.

Review of Definitions: Walk, path, cycle, connected graph, subgraph, induced subgraph, connected component.

Trees

A graph with no cycles is *acyclic*. A *tree* is a connected acyclic graph. A vertex of degree greater than 1 in a tree is called an *internal vertex*, otherwise it is called a *leaf*. A *forest* is an acyclic graph.

Example. Prove that every tree with at least two vertices has at least two leaves and deleting a leaf from an n-vertex tree produces a tree with $n - 1$ vertices.

Solution. A connected graph with at least two vertices has an edge. In an acyclic graph, an endpoint of a maximal non-trivial path (a path that is not contained in a longer path) has no neighbors other than its only neighbor on the path. Hence, the endpoints of such a path are leaves.

Let v be a leaf of a tree T and let $T' = T - v$. A vertex of degree 1 belongs to no path connecting two vertices other than v. Hence, for any two vertices $u, w \in V(T')$, every path from u to w in T is also in T'. Hence T' is connected. Since deleting a vertex cannot create a cycle, T' is also acyclic. Thus, T' is a tree with $n - 1$ vertices.

Example. For a n-vertex graph G, the following are equivalent and characterize trees with n vertices.

(1) G is a tree.
(2) G is connected and has exactly $n - 1$ edges.

(3) G is minimally connected, i.e., G is connected but $G - \{e\}$ is disconnected for every edge $e \in G$.
(4) G contains no cycle but $G + \{x, y\}$ does, for any two non-adjacent vertices $x, y \in G$.
(5) Any two vertices of G are linked by a unique path in G.

Solution. **$(1 \to 2)$.** We can prove this by induction on n. The property is clearly true for $n = 1$ as G has 0 edges. Assume that any tree with k vertices, for some $k \geq 0$, has $k-1$ edges. We want to prove that a tree G with $k+1$ vertices has k edges. From the example we did in last class we know that G has a leaf, say v, and that $G' = G - \{v\}$ is connected. By induction hypothesis, G' has $k-1$ edges. Since $deg(v) = 1$, G has k edges.

$(2 \to 3)$. Note that $G - \{e\}$ has n vertices and $n-2$ edges. We know that such a graph has at least 2 connected components and hence is disconnected.

$(3 \to 4)$. We are assuming that removing *any* edge in G disconnects G. If G contains a cycle then removing any edge, say $\{u, v\}$, that is part of the cycle does not disconnect G as any path that uses $\{u, v\}$ can now use the alternate route from u to v on the cycle. Since G is connected there is a path from x to y in G. Let $G' = G + \{x, y\}$. G' consists of a cycle formed by the edge $\{x, y\}$ and the path from x to y in G.

$(4 \to 5)$. Note that since $G + \{x, y\}$ creates a cycle for for any two non-adjacent vertices in G, it must be that there must be a path between x and y in G. We will now show that there is exactly one path between any two vertices in G. We will prove this by showing that if there are two vertices that have two different paths between them then G contains a cycle. Assume that there are two paths from u to v. Beginning at u, let a be the first vertex at which the two paths separate and let b be the first vertex after a where the two paths meet. Then, there are two simple paths from a to b with no common edges. Combining these two paths gives us a cycle.

$(5 \to 1)$. Since there is a path between any two vertices in G, G must be connected. Now we want to show that G is acyclic. Assume otherwise. Then, any two vertices on the cycle can reach each other by two disjoint, simple paths that consist of edges of the cycle. This proves that not every pair of vertices in G has a unique path between them. We have thus proved the claim by proving the contrapositive.

Spanning Trees

A *spanning subgraph* of a graph G is a subgraph with vertex set $V(G)$. A *spanning tree* is a spanning subgraph that is a tree.

Example. Every connected graph $G = (V, E)$ contains a spanning tree.

Solution. Let $T' = (V, E')$ be a minimally connected spanning subgraph of G. For a moment assume that such a T' always exists. Then, by the equivalence of statements (1) and (3), T' is a tree. Since T' is also a spanning subgraph of G, it is a spanning tree of G.

We now show that T', a minimally connected subgraph of G always exists. We will show this by actually constructing a minimally connected subgraph of G as follows. For each edge $e \in E$, remove e from E if its removal does not disconnect the graph. Let T' be the resulting subgraph obtained after each edge has been processed once. By construction and because G is connected, T' is connected. Also, by construction, no edge in T' can be removed without disconnecting T'. Hence, T' is minimally connected.

Rooted Trees

A *rooted tree* is a tree in which one vertex is distinguished from the others and is called the *root*. The *level* of a vertex, say u, is the number of edges along the unique path between u and the root. The *height* of a rooted tree is the maximum level of any vertex in the tree. Given any vertex of a rooted tree, the *children* of v are neighbors of v that are one level away from the root than v. If a vertex v is a child of u, then u is called the *parent* of v. Two vertices that are both children of the same parent are called *siblings*. Given vertices v and w, if v lies on the unique path between w and the root, then v is an *ancestor* of w and w is a *descendant* of v. A vertex in a rooted tree is called a *leaf* if it has no children. Vertices that have children are called *internal* vertices. The root is an internal vertex unless it is the only vertex in the graph, in which case it is a leaf. These definitions are illustrated in Figure 6. A *binary tree* is a rooted tree in which every internal vertex has at most two

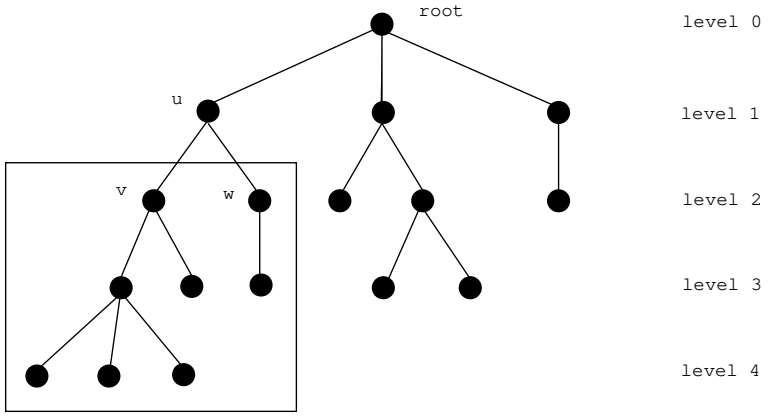


Figure 6: A rooted tree of height 4. In this tree v is a child of u, u is a parent of v, and v and w are siblings. All vertices in the marked portion of the tree descendants of u, which is an ancestor of each vertex.

children. Each child in the binary tree is designated either a left child or a right child (but not both). A *full binary tree* is a binary tree in which each internal vertex has exactly two children.

Given an internal vertex v of a binary tree T, the left subtree of v is the binary tree whose root is the left child of v, whose vertices consists of the left child of v and all its

descendants, and whose edges consist of all those edges of T that connect the vertices of the left subtree together. The right subtree of v is defined analogously.

Example. Prove the following. If k is a positive integer and T is a full binary tree with k internal vertices then T has a total of $2k+1$ vertices and has $k+1$ leaves.

Solution. Suppose T is a full binary tree with k internal vertices. Observe that the set of all vertices of T can be partitioned into two disjoint subsets: the set of all vertices in T that have a parent and the set of vertcies in T that do not have a parent. The root of T is the only vertex in T that does not have a parent. Also, every internal vertex of a full binary tree has exactly two children. Thus, the total number of children of all internal vertices equals $2k$. This is also the number of vertices that have a parent. Adding one for the root to this number gives us the total number of vertices in T to be $2k+1$.

Also, the total number of vertices in T is the sum of the number of internal vertices in T and the number of leaves in T. Hence, the number of leaves in T equals $2k+1-k = k+1$.

Example. Any binary tree of height at most h has at most 2^h leaves.

Solution. We will prove the claim by doing induction on h. Let $P(h)$ be the property that a binary tree of height at most h has at most 2^h leaves.

Induction Hypothesis: Assume that $P(k)$ is true for some $k \geq 0$.
Base Case: $P(0)$ is clearly true as there is only one tree of height at most zero. This tree has only one vertex which is a leaf. This equals $2^0 = 1$.
Induction Step: We want to prove that $P(k+1)$ is true. Consider any binary tree T having height at most $k+1$, and root r. Since we have already proven the claim for trees with height zero in the base case, we will assume that the height of T is at least one. The root r has at least one and at most two children. Each subtree rooted at a child of r is a rooted binary tree of height at most k. By induction hypothesis, the number of leaves in these subtrees is at most 2^k. Since r has at most two subtrees rooted at its children, the total number of leaves in T is at most $2 \times 2^k = 2^{k+1}$. This proves that $P(k+1)$ is true and hence completes the proof.

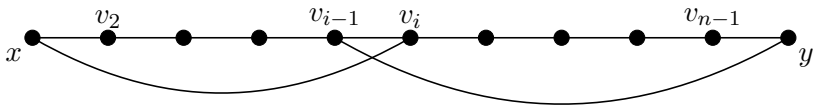
Hamiltonian Graphs

A *Hamiltonian cycle* in a graph G is a cycle in which each vertex of G appears exactly once. A graph is *Hamiltonian* if it contains a Hamiltonian cycle.

To determine whether a graph is Hamiltonian or not is a very hard problem.

Example. For any integer $n \geq 3$, let G be a simple graph on n vertices, and assume that all vertices in G are of degree at least $n/2$. Prove that G has a Hamiltonian cycle.

Solution. Assume for contradiction that G does not have a Hamiltonian cycle. Add new edges to G one-by-one, until we come to a point where adding an edge, say (x, y), creates a Hamiltonian cycle. Let G' be the graph in which all vertices have degree at least $n/2$ and G' does not have a Hamiltonian cycle, but adding (x, y) will make G' Hamiltonian. Since adding edge (x, y) creates a Hamiltonian cycle in G', it must be that G' has a Hamiltonian path that begins at x and ends at y. Let the path be $x = v_1, v_2, \ldots, v_{n-1}, v_n = y$. We now apply the pigeon-hole principle as follows. Let the pigeons be the edges incident on the vertices x and y, and let the holes be the $(n-1)$ edges of the form (v_i, v_{i+1}), where $1 \leq i \leq n-1$. An edge (pigeon) of the form (x, v_i) is assigned to the "hole" (v_{i-1}, v_i) and an edge (pigeon) of the form (y, v_i) is assigned to the "hole" (v_i, v_{i+1}). Since $\deg(x) \geq n/2$ and $\deg(y) \geq n/2$ and at most one edge incident on x (or y) is assigned to a hole, by the pigeon-hole principle, there must be i such that $3 \leq i \leq n-1$ and there is an edge (x, v_i) and an edge (y, v_{i-1}) (see figure below). Note that since the edge (x, y) does not exist in G', the hole corresponding to (v_1, v_2) only has one edge, namely (x, v_2). Similarly, the hole (v_{n-1}, v_n) will only contain the edge (y, v_{n-1}). But this would mean that $xv_2v_3 \cdots v_{i-1}yv_{n-1}v_{n-2} \cdots v_i$ is a Hamiltonian cycle, a contradiction.



Example. If $\delta(G) \geq 2$ then G contains a cycle.

Solution. Let P be a longest path (actually, any *maximal* path suffices) in G and let u be an endpoint of P. Since P cannot be extended, every neighbor of u is a vertex in P. Since $deg(u) \geq 2$, u has a neighbor $v \in P$ via an edge that is not in P. The edge $\{u, v\}$ completes the cycle with the portion of P from v to u.

Eulerian Graphs

An *Eulerian circuit* is a closed walk in which each edge appears exactly once. A connected graph is *Eulerian* if it contains an Eulerian circuit. Recall that a *Hamiltonian cycle* in a graph G is a cycle in which each vertex of G appears exactly once. A graph is *Hamiltonian* if it contains a Hamiltonian cycle.

To determine whether a graph is Hamiltonian or not is significantly harder than determining whether a graph is Eulerian or not. The following theorem gives us a necessary and sufficient condition for a connected graph to be Eulerian.

Example. Prove that a connected graph G is Eulerian iff every vertex in G has even degree.

Solution. *Necessity:* To prove that "if G is Eulerian then every vertex in G has even degree". Let C denote the Eulerian circuit in G. Each passage of C through a vertex uses two incident edges and the first edge is paired with the last at the first vertex. Hence every vertex has even degree.

Sufficiency: To prove that "if every vertex in G has even degree then G is Eulerian". We will prove this using induction on the number of edges, m.
Induction Hypothesis: Assume that the property holds for any graph G with j edges, for all j such that $0 \leq j \leq k$.
Base Case: $m = 0$. In this case G has only one vertex and that itself forms a Eulerian circuit.
Induction Step: We want to prove that the property holds when G has n vertices and $k+1$ edges. Since G has at least one edge and because G is connected and every vertex of G has even degree, $\delta(G) \geq 2$. From the result of the previous problem, G contains a cycle, say C. Let G' be the graph obtained from G by removing the edges in $E(C)$. Since C has either 0 or 2 edges at every vertex of G, each vertex in G' also has even degree. However, G' may not be connected. By induction hypothesis, each connected component of G' has an Eulerian circuit. We can now construct an Eulerian circuit of G as follows. Traverse C, but when a component of G' is entered for the first time, we detour along the Eulerian circuit of that component. The circuit ends at the vertex where we began the detour. When we complete the traversal of C, we have completed an Eulerian circuit of G.

Alternative Proof for the Sufficiency Condition: Let G be a graph with all degrees even and let

$$W = v_0 e_0 \dots e_{l-1} v_l$$

be the longest walk in G using no edge more than once. Since W cannot be extended all edges incident on v_l are part of W. Since all vertices in G have even degree it must be that $v_l = v_0$. Thus W is a closed walk. If W is Eulerian then we are done. Otherwise, there must be an edge in $E[G] \setminus E[W]$ that is incident on some vertex in W. Let this edge be $e = \{u, v_i\}$. Then the walk

$$u e v_i e_i \dots e_{l-1} v_l e_0 v_0 e_1 \dots e_{i-1} v_i$$

is longer than W, a contradiction.

Graph Coloring

Consider the following scenario. There are n courses for which final exams need to be scheduled. Each exam needs a two hour slot. Since each student may be in more than one course, the exams need to be scheduled such that two courses that have common students don't have their final exams at the same time. The objective is to find minimum number of time slots that would be required to schedule all the exams.

A graph is *k-colorable* if each vertex can be colored using one of the k colors so that adjacent vertices are colored using different colors. The *chromatic number* of a graph G, $\chi(G)$, is the smallest value of k for which G is k-colorable.

The problem of scheduling exams can be modeled as a graph coloring problem. Construct a graph in which there is a vertex for each course and two vertices u and v are connected by an edge if there is a student who is taking both the courses corresponding to u and v. The chromatic number of the graph will provide the required solution to the problem.

Finding the chromatic number of a graph "quickly" is a very hard problem. Even finding a reasonable approximate solution is very hard!!

A *bipartite graph* is a graph that is 2-colorable.

Example. Prove that a graph with maximum degree at most k is $(k+1)$-colorable.

Solution. Let $P(n)$ be the property that a graph with n vertices and maximum degree at most k is $(k+1)$-colorable. We will now prove the claim by doing induction on n.
Base Case: $P(1)$ is clearly true as a graph with just one vertex has maximum degree zero and can be colored using one color.
Induction Hypothesis: Assume that $P(h)$ is true for some $h \geq 1$.
Induction Step: We want to prove that $P(h+1)$ is true. Let G be a graph with maximum degree at most k and having $h+1$ vertices. Let G' be the graph obtained from G by removing a vertex v along with the edges incident on v. G' has h vertices and has a maximum degree at most k. By induction hypothesis, G' is $(k+1)$-colorable. Now insert v along with its incident edges. Since we have a palette of $k+1$ colors and $deg(v) \leq k$, we can always color v using a color that is not used by any of its neighbors. Thus, $P(h+1)$ is true. This completes the proof.

Matchings

A *matching* in a graph is a set of edges with no shared end-points. The vertices incident on the edges of a matching M are called *M-saturated*, the others are called *M-unsaturated*. A *perfect matching* in a graph is a matching that saturates every vertex in the graph.

A *maximal matching* in a graph is a matching that is not contained in a larger matching. A *maximum matching* is a matching of maximum size among all matchings in the graph. Every maximum matching is a maximal matching, but the converse is not true. Figure 7 illustrates some of these definitions.

Given a matching, M, an *M-alternating path* is a path that alternates between edges in M and edges not in M. An M-alternating path whose endpoints are M-unsaturated is called an *M-augmenting* path. Given an M-augmenting path P, we can replace the edges of M in P with the edges in $E(P) \setminus M$ to obtain a new matching with one more edge. Thus, when M is a maximum matching there is no M-augmenting path.

For graphs G and H, the *symmetric difference* $G \oplus H$ is a subgraph of $G \cup H$ whose edges are the edges of $G \cup H$ that appear in either G or H, but not both. We also use the notation for

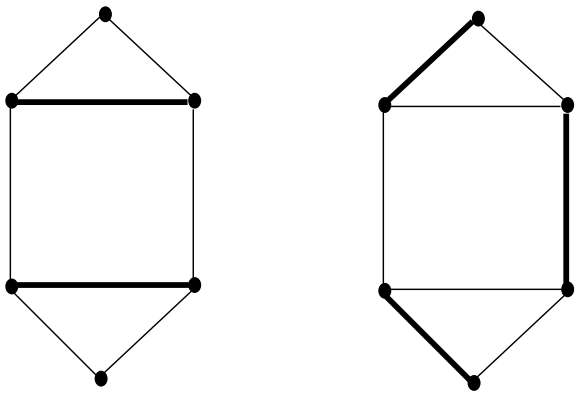


Figure 7: (a) a graph G with the bold edges representing a maximal matching, (b) the bold edges represent a maximum matching in G that is also perfect.

set of edges; in particular, if M and M' are matchings then $M \oplus M' = (M \setminus M') \cup (M' \setminus M)$.

Example. Prove that a matching M in G is maximum iff G contains no M-augmenting path.

Solution. We will prove the necessary condition by proving its contrapositive, i.e., we will prove that if G contains an M-augmenting path then M is not a maximum matching. Suppose that G contains a M-augmenting path $v_0v_1v_2 \dots v_{2m+1}$ (Note that an M-augmenting path must be of odd length). Define $M' \subseteq E$ by

$$M' = M \setminus \{(v_1, v_2), (v_3, v_4), \dots, (v_{2m-1}, v_{2m})\} \cup \{(v_0, v_1), (v_2, v_3), \dots, (v_{2m}, v_{2m+1})\}$$

Then M' is a matching in G and $|M'| = |M| + 1$. Thus M is not a maximum matching.

We will prove the converse by proving the contraposition. Assume that M is not a maximum matching. Let M' be a maximum matching in G. Then $|M'| > |M|$. Set $H = G[M \oplus M']$. Figure 9 illustrates this operation. Observe that every vertex in H has either degree one or degree two in H, since it can be incident with at most one edge of M and one edge of M'. Thus each component of H is either an even cycle with edges alternating in M and M' or else a path with edges alternating in M and M'. Since $|M'| > |M|$, H contains more edges of M' than of M, and H must contain a component which is a path, P, that starts and ends with edges in M'. Since the start vertex and end vertex of P are M'-saturated in H they must be M-unsaturated in G. Thus, P is an M-augmenting path in G. This completes the proof.

Example. Prove that a matching M in G is maximum iff G contains no M-augmenting path.

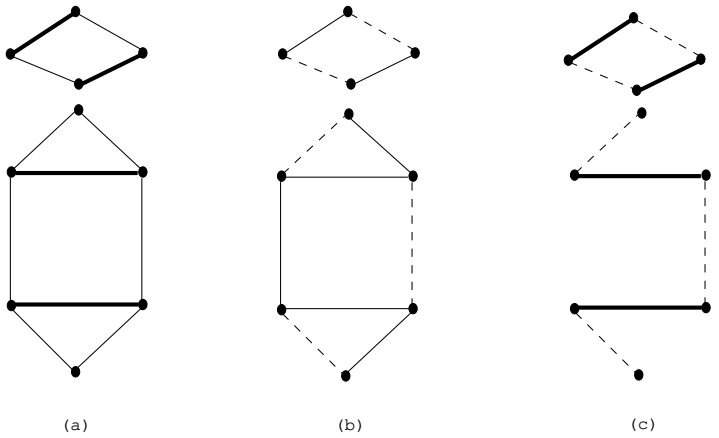


Figure 8: (a) a graph G with a matching M represented by the bold edges, (b) the dashed edges represent a matching M' in G, (c) $G[M \oplus M']$

Solution. We will prove the necessary condition by proving its contrapositive, i.e., we will prove that if G contains an M-augmenting path then M is not a maximum matching. Suppose that G contains a M-augmenting path $v_0v_1v_2 \dots v_{2m+1}$ (Note that an M-augmenting path must be of odd length). Define $M' \subseteq E$ by

$$M' = M \setminus \{(v_1, v_2), (v_3, v_4), \dots, (v_{2m-1}, v_{2m})\} \cup \{(v_0, v_1), (v_2, v_3), \dots, (v_{2m}, v_{2m+1})\}$$

Then M' is a matching in G and $|M'| = |M| + 1$. Thus M is not a maximum matching.

We will prove the converse by proving the contraposition. Assume that M is not a maximum matching. Let M' be a maximum matching in G. Then $|M'| > |M|$. Set $H = G[M \oplus M']$. Figure 9 illustrates this operation. Observe that every vertex in H has either degree one or degree two in H, since it can be incident with at most one edge of M and one edge of M'. Thus each component of H is either an even cycle with edges alternating in M and M' or else a path with edges alternating in M and M'. Since $|M'| > |M|$, H contains more edges of M' than of M, and H must contain a component which is a path, P, that starts and ends with edges in M'. Since the start vertex and end vertex of P are M'-saturated in H they must be M-unsaturated in G. Thus, P is an M-augmenting path in G. This completes the proof.

Matching in Bipartite Graphs

An *independent set* of a graph is a set of pair-wise non-adjacent vertices. A *bipartite graph*, (U, V, E), is a graph whose vertex set is $U \cup V$ and for each edge $e = (u, v) \in E$, $u \in U$ and $v \in V$. In other words, U and V are independent sets and each edge in E connects a vertex in U to a vertex in V.

Now consider the following scenario. There is a set of girls and a set of boys. Each girl likes some boys and dislikes others. What conditions would guarantee that each girl

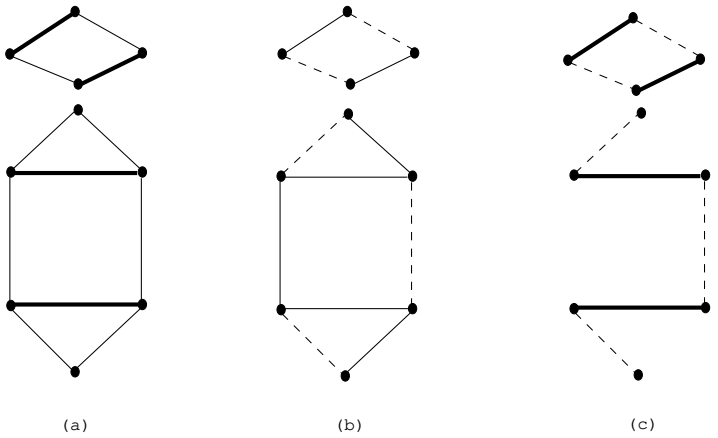


Figure 9: (a) a graph G with a matching M represented by the bold edges, (b) the dashed edges represent a matching M' in G, (c) $G[M \oplus M']$

is paired-up with a boy that she likes and that no two girls are paired-up with the same boy.

We can model this situation using a bipartite graph, (X, Y, E), where each vertex in X represents a girl, each vertex in Y represents a boy and and edge $(g, b) \in E$ means that girl g likes boy b. We are interested in the conditions that would guarantee a matching that saturates every vertex in X.
Hall's theorem gives the necessary and sufficient conditions for the existence of such matchings in bipartite graphs.

Example. [Hall's Theorem] Let $G = (X, Y, E)$ be a bipartite graph. For any set S of vertices, let $N_G(S)$ be the set of vertices adjacent to vertices in S. Prove that G contains a matching that saturates every vertex in X iff $|N_G(S)| \geq |S|, \forall S \subseteq X$. The condition "For all $S \subseteq X, |N(S)| \geq |S|$" is called Hall's condition.

Solution. We prove that Hall's condition is necessary as follows. Suppose G contains a matching M that saturates every vertex in X. Let S be a subset of X. Since each vertex in S is matched under M to a distinct vertex in $N_G(S)$, $|N_G(S)| \geq |S|$.

We will now prove the sufficiency of Hall's conditon, i.e., if $|N_G(S)| \geq |S|, \forall S \subseteq X$ then G contains a matching that saturates every vertex in X. We prove this by induction on the size of X.

Base Case: $|X| = 1$. If the only vertex in X is connected to at least one vertex in Y then clearly a matching exists.
Induction Hypothesis: Assume that Hall's condition is sufficient when $|X| = j$, for all j such that $1 \leq j \leq k$.

Induction Step: We want to prove that the sufficiency of Hall's condition when $|X| = k+1$. Let $G = (X, Y, E)$ be a graph with $k+1$ vertices in X such that $\forall S \subseteq X, |N_G(S)| \geq |S|$. We consider the following two cases.

Case I: For every non-empty proper subset $W \subset X$, $|N_G(W)| > |W|$. In this case, we pair-up an arbitrary vertex $x \in X$ with one of its neighbors, say $y \in Y$. Now consider the subgraph $G' = (X', Y', E')$, where $X' = X \setminus \{x\}$, $Y' = Y \setminus \{y\}$, and $E' = E \setminus \{(x, y)\}$). After the removal of y, the neighborhood of any subset, $S' \subseteq X'$ in G' is at most one less than its neighborhood in G. But since $|N_G(S')| > |S'|$, after removal of y, it must be that $|N_{G'}(S')| \geq |S'|$. Thus, Hall's condition holds for G'. By induction hypothesis, G' contains a matching M' that saturates every vertex in X'. Hence, $M' \cup \{(x, y)\}$ is a matching that saturates every vertex in X.

Case II: For some non-empty proper subset $W \subset X$, $|N(W)| = |W|$. For all $S' \subseteq W$, we have $N_G(S') \subseteq N_G(W)$. Hence, Hall's condition holds for the subgraph induced by $W \cup N(W)$. By induction hypothesis, there is a matching M_1 that matches every vertex in W to a vertex in $N_G(W)$. Note that M_1 is a perfect matching. Consider the subgraph $G' = (X', Y', E')$, where $X' = X \setminus W$, $Y' = Y \setminus N(W)$, and E' consists of all edges between X' and Y'. If we can prove that Hall's condition holds for G' then by induction hypothesis, G' has a matching M_2 that saturates every vertex in X'. Then, $M_1 \cup M_2$ is clearly a matching in G that saturates every vertex in X. It now remains to prove that $\forall T \subseteq X', |N_{G'}(T)| \geq |T|$. Note that $N_G(W \cup T) = N_G(W) \cup N_{G'}(T)$, $|N_G(W)| = |W|$, W and T are disjoint, and $N_G(W)$ and $N_{G'}(T)$ are disjoint. Then,

$$\begin{aligned} |N_G(W \cup T)| &\geq |W \cup T| \text{ (follows because } \forall S \subseteq X, |N_G(S)| \geq |S|) \\ |N_G(W)| + |N_{G'}(T)| &\geq |W| + |T| \\ |W| + |N_{G'}(T)| &\geq |W| + |T| \\ |N_{G'}(T)| &\geq |T| \end{aligned}$$

This proves the sufficiency of Hall's condition.

Relations and Functions

Relations

A *binary relation* is a set of ordered pairs. For example, let $R = \{(1,2),(2,3),(5,4)\}$. Then since $(1,2) \in R$, we say that 1 is related to 2 by relation R. We denote this by $1\,R\,2$. Similarly, since $(4,7) \notin R$, 4 is not related to 7 by relation R, denoted by $4 \not R\ 7$.

A binary relation R from set A to set B is a subset of the cartesian product $A \times B$. When $A = B$, we say that R is a relation on set A.

Example. Let $A = \{1,2,3,4\}$ and $B = \{a,b,c\}$. Consider the following relations.

$$R_1 = \{(1,1),(1,2),(2,2),(2,3)\}$$
$$R_2 = \{(1,2),(2,3),(3,4),(4,1),(4,4)\}$$
$$R_3 = \{(1,a),(2,a),(3,b),(4,c)\}$$
$$R_4 = \{(a,1),(a,3),(a,4),(c,1)\}$$
$$R_5 = \{(a,a),(a,b),(1,c)\}$$

R_1 and R_2 are relations on A. R_3 is a relation from A to B. R_4 is a relation from B to A. R_5 is not a relation on sets A and B and it is neither a relation from A to B nor a relation from B to A.

Below are some more examples of relations.

- If S is a set then "is a subset of ", $\subseteq$ is a relation on $\mathcal{P}(S)$, the power set of S.
- "is a student in" is a relation from the set of students to the set of courses.
- "=" is a relation on $\mathbb{Z}$.
- "has a path in G to" is a relation on $V(G)$, the set of vertices in G.

Example. How many relations are there on a set of n elements?

Solution. Note that $|A \times A| = n^2$. Since any subset of $A \times A$ is a relation on A, the number of possible relations is the cardinality of the power set of $A \times A$, which is 2^{n^2}.

Properties of Relations

Let R be a relation defined on set A. We say that R is

- *reflexive*, if for all $x \in A$, $(x,x) \in R$.
- *irreflexive*, if for all $x \in A$, $(x,x) \notin R$.
- *symmetric*, if for all $x,y \in A$, $(x,y) \in R \implies (y,x) \in R$.

- *antisymmetric*, if for all $x, y \in A$, $x\,R\,y$ and $y\,R\,x \implies x = y$.
- *transitive*, if for all $x, y, z \in A$, $x\,R\,y$ and $y\,R\,z \implies x\,R\,z$.

Note that the terms *symmetric* and *antisymmetric* are not opposites. A relation may be both symmetric and antisymmetric or can neither be symmetric nor be antisymmetric.

Example. What are the properties of the following relations?

R_1 : equality relation on $\mathbb{Z}$.
R_2 : "is a sibling of" relation on the set of all people.
R_3 : " $\leq$ " relation on $\mathbb{Z}$.
R_4 : " $<$ " relation on $\mathbb{Z}$.
R_5 : "$|$" relation on $\mathbb{Z}^+$.
R_6 : "$|$" relation on $\mathbb{Z}$.
R_7 : " $\subseteq$ " relation on the power set of a set S.
$R_8 : \{(x, y) \in \mathbb{R}^2 \,:\, |x - y| < \epsilon\}$, where $\epsilon = 0.001$

Solution.

$$\begin{aligned} \text{Reflexive} &: R_1, R_3, R_5, R_7, R_8 \\ \text{Irreflexive} &: R_2, R_4 \\ \text{Symmetric} &: R_1, R_2, R_8 \\ \text{Antisymmetric} &: R_1, R_3, R_4, R_5, R_7 \\ \text{Transitive} &: R_1, R_3, R_4, R_5, R_6, R_7 \end{aligned}$$

Note that R_6 is not reflexive because $(0, 0) \notin R_6$; it is not antisymmetric because for any integer a, $a|-a$ and $-a|a$, but $a \neq -a$. R_2 is not transitive because x and z could be the same person. Observe that R_6 is an example of a relation that is neither symmetric nor antisymmetric. R_1 is an example of a relation that is symmetric and antisymmetric.

Example. How many reflexive relations are there on a set A of size n?

Solution. We know that $R \subseteq A \times A$. The procedure of constructing a reflexive relation R is as follows:

> Step 1: From $A \times A$, include in R all ordered pairs of the form (a, a).
> Step 2: For every ordered pair in $A \times A$ of the form (a, b), where $a \neq b$, choose whether to include it in R or not.

There is one way to do Step 1 and $2^{n(n-1)}$ ways to do Step 2. By the multiplication rule, the number of reflexive relations on a set n elements is $2^{n(n-1)}$.

Equivalence Relations

A relation R on a set A is an *equivalence relation* if and only if it is reflexive, symmetric and transitive.

Example Let m be a positive integer. Show that the *congruent modulo m* relation

$$R = \{(a, b) : a \equiv b \pmod{m}\}$$

is an equivalence relation on the set of integers.
(If m is a positive integer then integers x and y are *congruent modulo m*, written as $x \equiv y \pmod{m}$, if $m|(x - y)$).

Solution. To show that R is an equivalence relation we need to show that it is reflexive, symmetric, and transitive. R is reflexive because $a - a = 0$, and $0 = m \cdot 0$. R is symmetric because if $a \equiv b \pmod{m}$, it means that $a-b = m \cdot k$, for some integer k. Thus $b-a = m(-k)$ and hence $(b, a) \in R$. To show that R is transitive, suppose that that $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$. Thus, for some integers q_1 and q_2, we have $a-b = m(q_1)$ and $b-c = m(q_2)$. Adding these two equations, we get $a - c = m(q_1 + q_2)$ and thus $a \equiv c \pmod{m}$. Hence R is transitive.

Example. Suppose that R is the relation on the set of strings of English letters such that $a\,R\,b$ if and only if $l(a) = l(b)$, where $l(x)$ is the length of the string x. Is R an equivalence relation?

Solution. R is reflexive as $l(a) = l(a)$, for any string a, and hence $a\,R\,a$. Next, suppose that $a\,R\,b$. This means that $l(a) = l(b)$ and hence $l(b) = l(a)$. Thus $b\,R\,a$ and hence R is symmetric. Finally, suppose that $a\,R\,b$ and $b\,R\,c$. Thus $l(a) = l(b)$ and $l(b) = l(c)$, which implies that $l(a) = l(c)$. Hence $a\,R\,c$ and R is transitive. Since R is reflexive, symmetric, and transitive, it is an equivalence relation.

Equivalence Classes

Let R be an equivalence relation on a set A and let $a \in A$. The *equivalence class of* a, denoted by $[a]_R$[2], is the set of all elements of A related (by R) to a; that is

$$[a]_R = \{x \in A \mid a\,R\,x\}$$

If $b \in [a]_R$, then b is called the *representative* of this equivalence class. Any element in a class can be used as a representative of the class.

Example. Let R be an equivalence relation on a set A. Then the following statements for elements $a, b \in A$ are equivalent

(i) $b \in [a]$ (ii) $[a] = [b]$ (iii) $[a] \cap [b] \neq \emptyset$

[2]The subscript R in $[a]_R$ is dropped when the relation in reference is clear from the context.

Solution. We will prove (i) $\implies$ (ii), (ii) $\implies$ (iii), and (iii) $\implies$ (i).
(i) $\implies$ (ii): We will prove the claim by showing that when $b \in [a]$, $[a] \subseteq [b]$ and $[b] \subseteq [a]$. Let c be any arbitrary but particular element in $[a]$. By definition, $a\,R\,c$. Since $b \in [a]$, it means that $a\,R\,b$, which further implies $b\,R\,a$ (since R is symmetric). Since R is transitive and we know that $b\,R\,a$ and $a\,R\,c$, we have $b\,R\,c$ and thus $c \in [b]$. We have thus proved that $[a] \subseteq [b]$.

Let $d \in [b]$. By definition, $b\,R\,d$. We also know that $a\,R\,b$. Since R is transitive, $a\,R\,b$ and $b\,R\,d$, we have $a\,R\,d$. Thus, by definition, $d \in [a]$. We have thus proved that $[b] \subseteq [a]$.

(ii) $\implies$ (iii): To prove this we just need to show that $[a] \neq \emptyset$. Since R is reflexive, we know that $a \in [a]$. Since $[a] = [b]$ and $[a]$ is non-empty, it follows that $[a] \cap [b] \neq \emptyset$.

(iii) $\implies$ (i): Let $c \in [a] \cap [b]$. Thus $a\,R\,c$ and $b\,R\,c$. Since R is symmetric, we have $c\,R\,b$. Since R is transitive, $a\,R\,c$ and $c\,R\,b$, we have $a\,R\,b$. By, definition $b \in [a]$.

Example. Let R be an equivalence relation on a set A. Then the set $\{[a]_R \mid a \in A\}$ is a partition of the set A. Each element of the set is called an *equivalence class* of R. Conversely, given a partition $\{A_i\}$ of the set A, there is an equivalence relation R that has sets A_i as its equivalence classes.

Solution. Since each element $a \in A$ is in its own equivalent class $[a]$, each equivalent class is non-empty and $\bigcup_{a \in A}[a] = A$. From the claim in the previous example (example we did in last class), for any two elements a and b in A, $[a]$ and $[b]$ are either equal or disjoint. Thus the equivalent classes partition the set A.

We now prove the converse. Let R be the relation on A that contains all possible pairs (x, y), where x and y belong to the same subset A_i in the partition. We want to show that R is reflexive, symmetric and transitive. R is reflexive as any element $a \in A$ is in the same subset of the partition as itself. Next suppose that $a\,R\,b$. This means that a and b are in the same subset of the partition of A. Thus, we have $b\,R\,a$ and hence R is symmetric. Finally, suppose that $a\,R\,b$ and $b\,R\,c$. This means that a and b are in the same subset of the partition and so are b and c. This means that a and c are in the same subset of the partition and hence we have $a\,R\,c$. Thus R is transitive.

Example. If an equivalence relation R is defined by the following set partition on A, then express R as a set of ordered pairs.

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

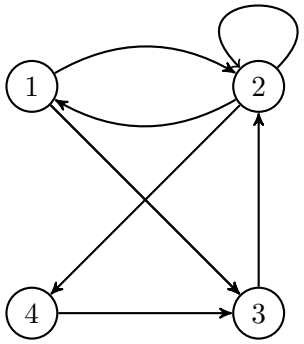
Solution.

$$R = \{(1,1), (2,2), (3,3), (4,4), (1,3), (1,4), (3,1), (3,4), (4,3), (4,1)\}$$

Representing Relations Using Directed Graphs

A *directed graph*, or *digraph* $G = (V, E)$ consists of a set V of vertices and a subset $E \subseteq V \times V$ of edges or arcs. An edge of the form (u, u) is represented as an arc from u to itself.

A binary relation R on a set A can be represented as a directed graph in which the vertices represent the elements of A and for every ordered pair $(a, b) \in R$, there is an edge from vertex a to vertex b. For example, the digraph corresponding to the relation $R = \{(1,2),(1,3),(2,1),(2,2),(2,4),(3,2),(4,3)\}$ on the set $\{1,2,3,4\}$ is shown below.



The directed graph G representing a relation R can be used to determine properties of the relation R. R is reflexive iff G contains a self-loop at every vertex. R is symmetric iff for each edge (a, b) $(a \neq b)$ in G, there is also an edge (b, a) in G. R is antisymmetric iff for any two distinct vertices a, b there are no edges between them or exactly one of (a, b) or (b, a) is in G. Thus R is antisymmetric iff for any two distinct vertices a and b, both (a, b) and (b, a) are not present in G. The relation R is transitive iff edge (u, w) always exists whenever there is an edge (u, v) and (v, w), for some vertex v.

Operations on Relations

We can take a relation or a pair of relations and produce a new relation. Since a relation R from set A to set B is a subset of $A \times B$, operations that apply to sets apply to relations.

Example. Let $A = \{1,2,3\}$ and $B = \{a,b,c,d\}$. Let $R_1 = \{(1,a),(1,c),(2,c),(3,a)\}$. Let $R_2 = \{(1,b),(1,c),(1,d),(2,b)\}$. Then we have

$$R_1 \cup R_2 = \{(1,a),(1,b),(1,c),(1,d),(2,b),(2,c),(3,a)\}$$
$$R_1 \cap R_2 = \{(1,c)\}$$
$$R_1 \setminus R_2 = \{(1,a),(2,c),(3,a)\}$$
$$R_2 \setminus R_1 = \{(1,b),(1,d),(2,b)\}$$

Example. Let A and B be the set of all students and the set of all courses at a school, respectively. Suppose R_1 consists of all ordered pairs (a, b), where a is a student who has taken course b, and R_2 consists of all ordered pairs (a, b), where a is a student who requires course b to graduate. What are the relations $R_1 \cup R_2$, $R_1 \cap R_2$, $R_1 \oplus R_2$, $R_1 \setminus R_2$, and $R_2 \setminus R_1$?

Solution. $R_1 \cup R_2$ consists of all ordered pairs (a, b), where a is a student who has taken course b or requires course b to graduate.
$R_1 \cap R_2$ consists of all ordered pairs (a, b), where a is a student who has taken course b and requires course b to graduate.
$R_1 \oplus R_2$ consists of all ordered pairs (a, b), where a is a student who has taken course b or requires course b to graduate, but not both.
$R_1 \setminus R_2$ consists of all ordered pairs (a, b), where a is a student who has taken course b but does not require it to graduate.
$R_2 \setminus R_1$ consists of all ordered pairs (a, b), where a is a student who required course b to graduate but has not taken it.

Inverse Relation

Let R be a relation from A to B. Then the *inverse* of R, written R^{-1}, is the relation from B to A defined by

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

Example. Let $A = \{a, b, c\}$ and let $R = \{(a, a), (a, b), (b, a), (c, a)\}$. Then

$$R^{-1} = \{(a, a), (b, a), (a, b), (a, c)\}$$

Note that R and R^{-1} are almost equal.

Example. A relation R on a set A is symmetric iff $R = R^{-1}$.

Solution. ($\Longrightarrow$) Suppose R is symmetric on A. We will prove that $R = R^{-1}$ by showing that $R \subseteq R^{-1}$ and $R^{-1} \subseteq R$. We will prove $R \subseteq R^{-1}$ by showing that an arbitrary element $(a, b) \in R$ is also in R^{-1}. Since R is symmetric, $(b, a) \in R$. By definition of R^{-1}, since $(b, a) \in R$, it must be that $(a, b) \in R^{-1}$. To prove $R^{-1} \subseteq R$, we will show that an arbitrary element $(a, b) \in R^{-1}$ is also in R. By definition of R^{-1}, it must be that $(b, a) \in R$. Since R is symmetric, (a, b) must also be in R.
($\Longleftarrow$) Suppose that $R = R^{-1}$. Let (a, b) be an arbitrary ordered pair in R. To prove that R is symmetric we need to show that $(b, a) \in R$. By definition of R^{-1}, $(b, a) \in R^{-1}$. Since $R = R^{-1}$, R must contain (b, a).

Composition of Relations

Let R be a relation from A to B and S be a relation from B to C. The *composition of S with R* is the relation from A to C:

$$S \circ R = \{(x, z) \mid \text{ there exists a } y \in B \text{ such that } x\, R\, y \text{ and } y\, S\, z\}$$

Example. Let $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$, and $C = \{a, b, c\}$. Let R and S be relations from A to B and from B to C, respectively, where

$$R = \{(1, 3), (3, 3), (3, 4), (4, 5), (4, 6)\}$$
$$S = \{(3, b), (4, a), (4, c), (5, a), (5, b), (6, c)\}$$

What is the composite of the relations R and S?

Solution. $S \circ R = \{(1, b), (3, a), (3, b), (3, c), (4, a), (4, b), (4, c)\}$

Let R be a relation on a set A. The powers $R^n, n = 1, 2, 3, \ldots$, are defined recursively by

$$R^1 = R \qquad \text{and} \qquad R^{n+1} = R^n \circ R$$

Observe that $R^2 = R \circ R, R^3 = R^2 \circ R = (R \circ R) \circ R$, and so on.

Example. Let R be a relation on a set A. Then R is transitive iff $R^n \subseteq R$, for all $n \geq 1$.

Solution. We first show that if $R^n \subseteq R$, for all $n \geq 1$, then R is transitive. Note that if $(a, b) \in R$ and $(b, c) \in R$ then $(a, c) \in R^2$. Since $R^2 \subseteq R$, it must be that $(a, c) \in R$, which means that R is transitive.

We will prove R is transitive $\implies R^n \subseteq R$, for all $n \geq 1$, using induction on n.
Induction hypothesis: Assume that if R is transitive then $R^k \subseteq R$, for some $k \geq 1$.
Base Case: The claim holds tivially when $n = 1$, since $R^1 = R$.
Induction Step: We want to prove the claim when $n = k + 1$. In other words, we want to prove that if R is transitive then $R^{k+1} \subseteq R$. We will prove this by showing that an arbitrary but particular ordered pair (a, b) in R^{k+1} is also present in R. By definition, $R^{k+1} = R^k \circ R$. Since $(a, b) \in R^{k+1}$, there must be a c, such that $(a, c) \in R$ and $(c, b) \in R^k$. We know by induction hypothesis that $R^k \subseteq R$, which means that $(c, b) \in R$. Since R is transitive, and $(a, c) \in R$ and $(c, b) \in R$, we have $(a, b) \in R$. This completes the proof.

Functions

Let A and B be sets. A *function* from A to B is a relation, f, from A to B such that for all $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in f$. If $(a, b) \in f$, then we write $b = f(a)$. A function from A to B is also called a *mapping* from A to B and we write it as $f : A \to B$. The set A is called the *domain* of f and the set B the codomain. If $a \in A$ then the element $b = f(a)$ is called the *image* of a under f. The *range* of f, denoted by $\text{Ran}(f)$ is the set

$$Ran(f) = \{b \in B \mid \exists a \in A \text{ such that } b = f(a)\}$$

Two functions are *equal* if they have the same domain, have the same codomain, and map each element of the domain to the same element in the codomain.

Example. Let A and B be finite sets of size a and b, respectively. How many functions are there from A to B?

Solution. The procedure of forming a function is as follows: in Step i choose the image of the ith element in A. There are a steps and there are b ways to perform each step. Thus the total number of ways to create a function from A to B is b^a.

Let $f : A \to B$ be a function.

- f is said to be *one-to-one* or *injective*, iff for every $x, y \in A$ such that $x \neq y, f(x) \neq f(y)$.
- f is called *onto* or *surjective*, iff for every element $b \in B$ there is an element $a \in A$ with $f(a) = b$.
- f is a *one-to-one correspondence* or *bijection*, if it is both one-to-one and onto.

Example. Classify the following functions.

- $f_1(x) = x^2$ from the set of integers to the set of integers.
- $f_2(x) = x^2$ from the set of non-negative real numbers to the set of non-negative real numbers.
- $f_3(x) = x + 1$ from the set of integers to the set of integers.
- $f_4(x) = x$ from a set A to A. This function is called the identity function.

Solution.

$$\begin{aligned} \text{injective} &: f_2, f_3, f_4 \\ \text{surjective} &: f_2, f_3, f_4 \\ \text{bijective} &: f_2, f_3, f_4 \end{aligned}$$

Inverse and Composition

Let f be a one-to-one correspondence from the set A to the set B. The *inverse* function of f is the function that maps an element $b \in B$ to the unique element $a \in A$ such that $f(a) = b$. The inverse function of f is denoted by f^{-1}. Hence $f^{-1}(b) = a$ when $f(a) = b$.

Note that if f is not bijective then its inverse does not exist.

Let $f : A \to B$ and $g : B \to C$ be functions. The *composition* of the function g with f is the function $g \circ f : A \to C$, defined by

$$(g \circ f)(x) = g(f(x)), \forall x \in A$$

Example. Let g be the function from the set $\{a, b, c\}$ to itself such that $g(a) = b, g(b) = c$, and $g(c) = a$. Let f be the function from the set $\{a, b, c\}$ to the set $\{1, 2, 3\}$ such that $f(a) = 3, f(b) = 2$, and $f(c) = 1$. What is the composition of f with g and what is the composition of g with f?

Solution. The composition function $f \circ g$ is as follows: $(f \circ g)(a) = f(g(a)) = f(b) = 2$, $(f \circ g)(b) = f(g(b)) = f(c) = 1$, and $(f \circ g)(c) = f(g(c)) = f(a) = 3$.

$(g \circ f)$ is not defined as the range of f is not a subset of the domain of g.

Example. Let f and g be the functions from the set of integers to the set of integers defined by $f(x) = 2x + 3$ and $g(x) = 3x + 2$. What is the composition of f and g? What is the composition of g and f?

Solution. $(f \circ g)(x) = f(g(x)) = 2(3x+2)+3 = 6x+7$. Similarly, $(g \circ f)(x) = g(f(x)) = 3(2x + 3) + 2 = 6x + 11$. This example shows that commutative law does not apply to the composition of functions.

Example. Let $f : A \to B$ and $g : B \to C$ be two functions. Then

i. if f and g are surjective then so is $g \circ f$.

ii if f and g are injective then so is $g \circ f$.

iii if f and g are bijective then so is $g \circ f$.

Solution. Let $c \in C$. Since g is surjective there must be a $b \in B$ such that $g(b) = c$. Since f is surjective there must be a $a \in A$ such that $f(a) = b$. Thus $(g \circ f)(a) = g(f(a)) = g(b) = c$. This proves that $g \circ f$ is surjective.

Let $a, a' \in A$ such that $(g \circ f)(a) = (g \circ f)(a')$. This means that $g(f(a)) = g(f(a'))$. Since g is injective we have $f(a) = f(a')$. Then since f is injective, we have $a = a'$.

The bijectivity of $(g \circ f)$ follows from the injectivity and surjectivity of $(g \circ f)$.