Mathematical Foundations of Computer Science

This Lecture: Combinatorics - Pigeonhole Principle

Department of Mathematical and Computational Sciences
National Institute of Technology Karnataka, Surathkal

MA714 (Odd Semester [2025-26])

Introduction

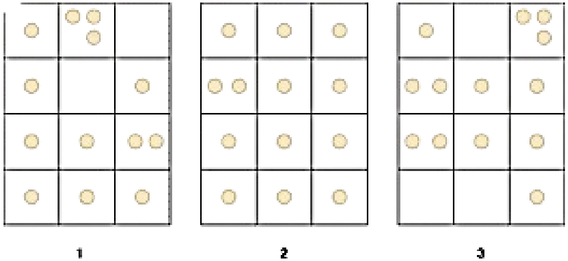
The pigeonhole principle (also sometimes called the Dirichlet drawer principle) is a simple yet powerful idea in mathematics that can be used to show some surprising things.

What is it informally?

If you place more than n items into n containers, then at least one container must contain more than one item.

It all started with the pigeons and pigeonholes and hence the name, "*Pigeonhole Principle*". Suppose that a set of pigeons fly into a set of pigeonholes. Then if there are more pigeons than pigeonholes, there must be at least one pigeonhole that has more than one pigeon in it.

Shown below are three groups of twelve pigeonholes. Each group has thirteen pigeons (represented by the circles), so there is at least one pigeonhole in each group that has at least two pigeons.



We can, of course, apply this principle to other things besides pigeons and pigeonholes. Let us first define the basic form of the Pigeonhole Principle formally:

Formal Statement:

Basic form (Simple Pigeonhole Principle)

If $n + 1$ or more objects are placed into n boxes, then at least one box contains two or more objects.

Example 1

Suppose we have a group of 27 English words. Then there must be at least two words in the group that begin with the same letter. In this case, the 27 English words are the pigeons and we can think of the pigeonholes as buckets that store words that start with a specific letter of the English alphabet.
In this case, there are 26 pigeonholes (one for each letter in the English alphabet) but 27 pigeons, so we can immediately see that there must be at least one bucket with more than one word in it. But a bucket with more than one word implies that there are at least two words that start with the same letter.

Example 2

In any group of 367 people, there must be at least two people with the same birthday. Here we can think of a person as a pigeon and a room that has all the people who have a specific birthday as a pigeonhole. Since there are only 366 possible birthdays, we have 366 rooms. But if we put 367 people in 366 rooms, then we know that there must be at least one room with at least two people in it. This means that there must be at least two people with the same birthday.

Example 3

Let us say, we know the maximum number of hairs that a person can have on his head. Let k be this number. Then in any group of more than k people, there must be at least two people with the same number of hairs on their heads.

A person represents a pigeon and a room with people that have a specific number of hairs on their heads represents a pigeonhole. Since we have more people than rooms, there must be at least one room with at least two people in it.

Now, the average number of hairs on a person's head is about 1,00,000; so, this means that the maximum is probably at most a few lakhs. This means that in any group of say 3,00,000 people, it is guaranteed that there are at least two people with the same number of hairs on their heads!

We can say even more when the number of objects is more than a multiple of the number of boxes. For example, in any set of 21 decimal digits, there must be 3 that are the same. This is because if 21 objects are put into 10 boxes, there must be at least one box with at least 3 objects in it.

Generalized Pigeonhole Principle

If n objects are placed into k boxes, then at least one box contains at least $\left\lceil \frac{n}{k} \right\rceil$ objects.

Suppose that n objects are placed into k boxes, but every box contains at most $\lceil \frac{n}{k} \rceil - 1$ objects. Then the total number of objects is at most $k(\lceil \frac{n}{k} \rceil - 1)$; but this value is less than $k((\frac{n}{k} + 1) - 1) = n$. This is because we know that $\lceil \frac{n}{k} \rceil < (\frac{n}{k}) + 1$; so, if we substitute the expression $(\frac{n}{k}) + 1$ in the place of $\lceil \frac{n}{k} \rceil$ in the expression $k(\lceil \frac{n}{k} \rceil - 1)$, then we will get $k((\frac{n}{k} + 1) - 1) = n$ which must be bigger in value.
Thus, we get a contradiction since there are n objects in total.

Example 4

In any group of 100 people there is at least $\lceil \frac{100}{12} \rceil = 9$ who were born in the same month. Here the objects we are placing are the people and the boxes can be imagined as rooms that have people who were born in a specific month. Since we have 100 people and 12 rooms (one for each month of the year), there must be a room with at least 9 people in it.

Example 5

You have 11 pairs of socks in a drawer, all mixed up. What is the minimum number of socks you must pick to guarantee getting a matching pair?
Answer: There are only 11 different kinds of socks. If you pick 12 socks, by the Pigeonhole Principle, at least one kind of sock must occur more than once. Hence, 12 is the answer. .

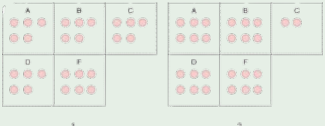
Example 6

To guarantee that at least three cards of the same suit are chosen, we must select at least nine cards from a standard deck of 52 cards. In this case, we can think of the boxes as marked with a particular suit, and as we select cards from the deck we place them in the box with the same suit as the selected card. The generalized pigeonhole principle tells us that if n cards are selected, there is at least one box containing at least $\lceil \frac{n}{4} \rceil$ cards.
So, if $\lceil \frac{n}{4} \rceil \geq 3$, then we know that at least three cards of the same suit must have been selected. The smallest possible value that n can take for this condition to be true is 9, so that at least nine cards must be selected.

Example 7

The minimum number of students needed in a class to ensure that at least six students receive the same grade is 26 if there are five possible grades (e.g. A, B, C, D, and F). We can think of each object as a person and each box as a room with people that have a specific grade. The minimum number of students needed to satisfy the condition is equal to the smallest integer n such that $\lceil \frac{n}{5} \rceil = 6$, and the smallest such integer $n = 26$.

Two instances of this example are shown below. n both cases, we have 5 boxes that represent the five possible grades and 26 red circles representing 26 students that are assigned some grade. In both cases, there is at least one box containing at least six red circles which means that there are at least six students who got the same grade.



Example 8: General form Application

Suppose 100 students are assigned to 9 project groups. What is the minimum number of students that must be in at least one group?
Answer: Using the generalized principle: $\lceil \frac{100}{9} \rceil = 12$. So, at least one group must have 12 or more students.

Example 9

Among any $n+1$ integers such that each integer is at most $2n$, there must be an integer that divides one of the other integers. We can write each of the $n+1$ integers $x_1, x_2, \cdots, x_{n+1}$ as a power of 2 times an odd integer. So, let $x_j = 2^{k_j} q_j$ for $j = 1, 2, \cdots, n+1$, where k_j is a non-negative integer and q_j is odd. The integers $q_1, q_2, \cdots, q_{n+1}$ are all odd positive integers less than $2n$.

It follows from the pigeonhole principle that two of the integers $q_1, q_2, \cdots, q_{n+1}$ must be equal, since there are only n odd positive integers less than $2n$. So, there must exist integers i and j such that $q_i = q_j$. Let q be the common value of q_i and q_j. Then we have that $x_i = 2^{k_i} q$ and $x_j = 2^{k_j} q$. Therefore, if $k_i < k_j$ then x_i divides x_j, whereas if $k_i > k_j$, then x_j divides x_i.

Example 10

Suppose that there is a sports team that plays at least one game a day for 30 days, but no more than 45 games in total. Then there must be a period of some consecutive days during which the team must play exactly 14 games. Prove or disprove this.

Solution: Let x_j be the number of games played on or before the j^{th} day of the 30 day period. Then $x_1, x_2, \cdots, x_{30}$ is an increasing sequence of distinct positive integers, with $1 \leq x_j \leq 45$. Furthermore, $x_1 + 14, x_2 + 14, \cdots, x_{30} + 14$ is also an increasing sequence of distinct positive integers, with $15 \leq x_j + 14 \leq 59$.

The 60 positive integers $x_1, x_2, \cdots, x_{30}, x_1 + 14, x_2 + 14, \cdots, x_{30}$ are all less than or equal to 59. The pigeonhole principle then tells us that two of these integers are equal. Since the integers $x_1, x_2, \cdots, x_{30}$ are all distinct and the integers $x_1 + 14, x_2 + 14, \cdots, x_{30} + 14$ are also all distinct, there must be indices i and j with $x_i = x_j + 14$.

But this means that exactly 14 games were played from day $j + 1$ to day i.

1. **Hash Functions & Hash Collisions**: In hashing, data is mapped to a fixed-size array. If the number of inputs exceeds the number of outputs, collisions are inevitable. E.g., hashing 1000 keys into 100 buckets leads to some buckets having ≥ 10 keys.
   Cryptography Application: Collision-free hash functions for arbitrarily long inputs are impossible, leading to the need for probabilistic security assumptions.
2. **Pigeonhole in Proofs of Non-Invertibility**: The principle is used to prove that some functions are non-injective. E.g., mapping 1024 inputs to 256 outputs ensures that some outputs must correspond to multiple inputs.

3. **Data Compression**: Lossless compression of every file is impossible. Some files must share compressed versions, making unique recovery impossible.
4. **Networking & Resource Allocation**: If there are more tasks than servers, at least one server must handle multiple tasks. Used in worst-case analysis.
5. **Cache Memory and Page Replacement**: If cache can hold n pages and more than n pages are accessed, some must be evicted. This underpins page replacement algorithms.

The Pigeonhole Principle is deceptively simple, but powerful. It is used in proofs of existence in theoretical computer science where direct construction may be hard, but existence is guaranteed logically. It is a tool for reasoning about limitations in algorithms, storage, and computation.

1. Kenneth Rosen, *Discrete Mathematics and its Applications*, 5^{th} edition, McGraw Hill, NY, 2003.
2. C.L. Liu, Elements of Discrete Mathematics, 2^{nd} edition, McGraw Hill.
3. D'Angelo and West, Mathematical Thinking: Problem-Solving and Proofs, 2^{nd} edition, Pearson.
4. Graham, Knuth, and Patashnik, Concrete Mathematics, 2^{nd} edition, Addison-Wesley.