TOPIC 1: The Pigeon Hole or Box Principle

The pigeon hole principle is a simple combinatorial principle first formulated by Dirichlet (1805-1859):

> If you distribute n+1 pigeons into n holes then at least one hole contains more than one pigeon.

Despite its simplicity it has a number of useful and unexpected applications. It is possible to prove deep theorems with it. The principle is only an existence assertion and does not help to identify the hole with more than one pigeon. However every existence assertion about finite sets is provable this way.

The main difficulty lies in identifying the pigeons and the holes. Let's start with some warm-up problems:

- A school has 400 students. Then there are two students who share the same birthday.
- For what minimum number of students does the statement remain valid?
- How many students does the school have to have so that there are three students who share a common birthday?

The last problem uses the following easy generalization of the principle:

> If you distribute more than $k \cdot n$ pigeons into n holes then at least one hole contains more than k pigeons.

Let's look at some worked examples.

Example 1:
There are n people at a meeting. Show that there are two people who have the same number of acquaintances.

Note that in mathematical problems about acquaintances it is always understood that knowing someone is reflexive which means if person A knows person B then automatically person B knows person A.

Solution:
A person (pigeon) is assigned hole number i, if he/she has i acquaintances. So we have the n pigeons and holes with number 0, 1, 2, … , n-1. But, obviously, holes 0 and n-1 cannot, at the same time, have any pigeons in it. Because a person in hole 0 does not

know anyone else at the meeting and a person in hole n-1 would know everyone else at the meeting (since there are a total of n people) and this cannot happen simultaneously. Hence we have more pigeons than holes, and by the principle the conclusion follows.

Example 2 (Erdös):
We are given not necessarily distinct integers $a_1, a_2, \ldots, a_n$. Then there is a subset of all these numbers whose sum is divisible by n.

Solution:
Consider the n numbers:

$s_1 = a_1,$
$s_2 = a_1 + a_2,$
….
$s_n = a_1 + a_2 + \ldots + a_n$

If one of these numbers is divisible by n then we are done. Otherwise their mod n remainders are in {1, 2, … , n-1}. Since there are only n-1 of these (we have excluded 0 in the meantime), two of the sums, say s_p and s_q with $p < q$, have the same remainder mod n. But then

$$s_q - s_p = a_{p+1} + \ldots + a_q$$

is divisible by n.

Example 3 (Erdös):
Prove that among n+1 numbers from the set {1, 2, … , 2n} there are always two, so that one is divisible by the other.

Solution:
If the n+1 numbers are $a_1, \ldots, a_{n+1}$ we write them as $a_i = 2^{k_i} b_i$ where b_i is odd. This way we have generated n+1 odd numbers from the interval [1, … , 2n-1]. There are only n odd numbers in this interval. Hence there are $p<q$ so that $b_p = b_q$. Then one of a_p is divisible by a_q (since one power of two always divides the other).

Example 4 (Erdös):
Prove that among n+1 integers from the set {1, 2, … , 2n} there are always two which are relatively prime.

Solution: It is not true if only n numbers are given (consider all the even numbers for example). But among n+1 numbers there must be a pair of consecutive numbers which will be relatively prime.

Example 5:
Prove that the sequence of the unit digits of the Fibonacci sequence is periodic.

Solution: Look at the pairs (f_i, f_{i+1}) mod 10. Among the first 101 = 10·10 + 1 pairs there must be two which are identical. The recursion of the Fibonacci numbers now implies the desired periodicity and one full period must be between those two pairs. This also means that the period length is less than 100. Note that looking at f_i mod 10 is not enough.

Example 6:
Among 5 lattice points in the plane you can always choose two so that the mid point of the corresponding line segment is also a lattice point.

Solution:
The mid point of two points A (a, b) and B (c, d) is given by ((a+c)/2, (b+d)/2). These coordinates are both integers if (a, c) and (b, d) have the same parity, i.e. are both even or are both odd. Since there are only four parity patterns, namely (even, even), (even, odd), (odd, even) and (odd, odd) among five points one can find two with the same parity pattern.

Example 7 (USAMTS 2007/8, Round 3, Problem 4):
Prove that 101 divides infinitely many of the numbers in the set {2007, 20072007, 200720072007, …}.

Solution:
Remember one of Polya's principles: Can you solve an easier problem? So let's try to find at least one number of the given set which is divisible by 101. This suggests the pigeon hole principle - there must be two among the first 102 numbers 2007, 20072007, ... that are congruent mod 101 and so their difference is divisible by 101. Since that difference is of the form $2007...2007 \cdot 10^b$ (say a times 2007's where 1 <= a <= 101) and since gcd (101, 10) = 1, it follows 2007...2007 (a times 2007's) is divisible by 101. That was fairly straightforward. The question is how to inductively create one after the other. Starting with the next 102 bigger ones from the one found would not work as you may end up with the same one. But there is an easy way around it: Look at the 102 numbers of the form 2007...2007, the first consisting of a+1 times 2007's, the next of 2(a+1) times 2007's and the last of 102(a+1) times 2007's. Now as in the first step you can generate a number of the desired form with c·(a+1) 2007's, 1<= c <= 101. This new number is bigger than the first one (could have used instead of a+1 any number bigger than a). Induction will finish the job. Obviously 2007 can be replaced by any four digit number.

Example 8:
A chess master has only another 77 days left to prepare for a tournament. He wants to play at least one game per day but not more than 132 in total. Prove that there is a sequence of days during which he always played exactly 21 games in total.

Solution:
Let a_i be the number of games he played in total up to and including day i. We want to show that there are indices i and j so that $a_i = a_j + 21$ (then the chess master has played exactly 21 days in total during the days j+1, j+2, … , i). We note:

$$1 \le a_1 < a_2 < \ldots < a_{77} \le 132 \quad \text{and} \quad 22 \le a_1 + 21 < a_2 + 21 < \ldots < a_{77} + 21 \le 153$$

Among the 154 numbers $a_1, a_2, \ldots, a_{77}, a_1 + 21, a_2 + 21, \ldots, a_{77} + 21$ from the set $\{1, 2, 3, \ldots, 153\}$ there must be two equal. Since all the a_i are different and all the $a_i + 21$ are different there must be indices i and j so that $a_i = a_j + 21$.

In general the box principle applies when there is not enough "room" for a number of objects to squeeze in. Sometimes, it helps to argue by contradiction. Here is an example.

Example 9:
Each of ten line segments is longer than 1 but shorter than 55. Prove that you can select three which one can form a triangle with.

Solution: Suppose the line segments are $1 < a_1 \le a_2 \le \ldots \le a_{10} < 55$. Assume no triangle can be constructed. Then $a_3 \ge a_2 + a_1 > 1 + 1 = 2$; $a_4 \ge a_3 + a_2 > 2 + 1 = 3$. We get the Fibonacci numbers with $a_{10} > 55$ which is a contradiction.

The following problem is a classic. It appeared at the 1953 Putnam exam and in the problem section of the American Mathematical Monthly in 1958. Generalizations appeared as IMO problems in the 70's and 80's. It also founded a new branch of combinatorics called Ramsey theory which is an active research area.

Example 10:
Among six persons there are always three who all know each other or three none of which knows each other.

Solution:
Choose any of the six persons and call him/her P. Among the other five persons, P either knows three (say A, B, C) or does not know three (otherwise there would be only 2+2=4 other persons). In the first case, if any two of the three people A, B, C know each other then those two together with P form a set of three people who all know each other. Otherwise none of the three people A, B, C know each other and we are done as well. One argues similarly in the second case. It may help to represent the people as points in the plane and color an edge green if the two people know each other and red if they don't. Then we can rephrase the statement as: There are six points in the plane and their edges are colored either green or red. Then the edges from either a green or a red triangle.

The principle is very powerful. As mentioned above it is possible to prove deep results with it. A very difficult example, which also requires a fair amount of technique, is:

> Given any string of digits one can find a power of two that starts in the decimal expansion exactly with the given digits.

Practice Problems:

1) Let $a_1, a_2, \dots, a_n$ be a permutation of the numbers 1, 2, … , n where n is odd. Show that the product $(a_1-1)(a_2-2)\dots(a_n-n)$ is always even.

2) Prove that in a convex 2n-gon there is always a diagonal not parallel to any side.

3) Color all points of the plane with n colors. Show that there are two points of distance 1 having the same color. Do the problem for n=2 and n=3. (It is an unsolved problem whether the conclusion holds for n=4, 5 or 6. For n>6 the conclusion is unknown to be false).

4) Inside a room of area 5 you place 9 rugs each of area 1 and of arbitrary shape. Prove that there are two rugs which overlap by at least 1/9.

5) If n is a positive integer then it has a multiple which consists in the decimal system only of digits 0s and 1s. If gcd (n, 10) = 1 then n has a multiple which consists only of 1s.