The Pigeonhole Principle

2

2.1 The Pigeonhole Principle

The **pigeonhole principle** is one of the most used tools in combinatorics, and one of the simplest ones. It is applied frequently in graph theory, enumerative combinatorics and combinatorial geometry. Its applications reach other areas of mathematics, like number theory and analysis, among others. In olympiad combinatorics problems, using this principle is a golden rule and one must always be looking for a way to apply it. The first use of the pigeonhole principle is said to be by Dirichlet in 1834, and for this reason it is also known as the Dirichlet principle.

Proposition 2.1.1 (Pigeonhole principle) *If $n+1$ objects are arranged in n places, there must be at least two objects in the same place.*

The proof of this proposition is almost immediate. If in every place there would be at most one object, we would have at most n objects, which contradicts the hypothesis. However, with the same reasoning we can prove a stronger version.

Proposition 2.1.2 (Pigeonhole principle, strong version) *If n objects are arranged in k places, there are at least $\lceil \frac{n}{k} \rceil$ objects in the same place.*

$\lceil x \rceil$ represents the smallest integer that is greater than or equal to x. Even though this result seems rather inconspicuous, it has very strong applications and is used frequently in olympiad problems.

Example 2.1.3 Given a triangle in the plane, prove that there is no line that does not go through any of its vertices but intersects all three sides.

Solution Any line divides the plane into two parts. By the pigeonhole principle, since there are three vertices, there must be at least two on the same side. The triangle side formed by those two vertices does not intersect the line. □

P. Soberón, *Problem-Solving Methods in Combinatorics*,
DOI 10.1007/978-3-0348-0597-1_2, © Springer Basel 2013

Example 2.1.4 Prove that given 13 points with integer coordinates, one can always find 4 of them such that their center of gravity[1] has integer coordinates.

Solution Let us regard the coordinates modulo 2. There are only 4 possibilities: $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$. So, by the pigeonhole principle, out of every 5 there must be two that have the same coordinates modulo 2. Let us take 2 such points and separate them from the others. We can continue this process and remove pairs with the same parity modulo 2 until we only have 3 points left. At that moment we have 5 different pairs which sum to 0 modulo 2 in both entries. Each sum modulo 4 can be 0 or 2, which gives only 4 possibilities for the sums modulo 4. Since we have 5 pairs, there must be 2 of them whose sums have the same entries modulo 4. These 4 points are the ones we were looking for. □

Exercise 2.1.5 Find 12 points in the plane with integer coordinates such that the center of gravity of any 4 of them does not have integer coordinates.

Example 2.1.6 (Russia 2000) A 100×100 board is divided into unit squares. These squares are colored with 4 colors so that every row and every column has 25 squares of each color. Prove that there are 2 rows and 2 columns such that their 4 intersections are painted with different colors.

Solution Let us count the number P of pairs of squares (a_1, a_2) such that a_1 and a_2 are in the same row and have different colors. In every row there are 25 squares of each color, so there are $\binom{4}{2}25 \cdot 25 = 6 \cdot 25 \cdot 25$ such pairs in each row. Thus $P = 100 \cdot 6 \cdot 25 \cdot 25$. We know there are $\binom{100}{2}$ pairs of columns and each of the P pairs must be in one of these pairs of columns. By the pigeonhole principle, there is a pair of columns that has at least

$$\frac{P}{\binom{100}{2}} = \frac{100 \cdot 6 \cdot 25 \cdot 25}{\frac{100 \cdot 99}{2}} = \frac{12 \cdot 25 \cdot 25}{99} = \frac{100 \cdot 75}{99} > 75$$

of the P pairs. From now on we only consider the pairs of squares in the same row, with different colors and that use these two columns.

If these are not the columns we are looking for, then for any two of the pairs of squares we just mentioned there is at least one color they share. Take one of these pairs, and suppose it uses colors black and blue. Since there are more than 50 of these pairs, there is at least one that does not have color black. If it did not have

[1]The center of gravity of the set S of points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ in the plane is defined as the point

$$\left(\frac{x_1 + x_2 + \cdots + x_n}{n}, \frac{y_1 + y_2 + \cdots + y_n}{n} \right).$$

That is, it is the average of the set of points.

color blue, we would be done, so it must have blue and some other color (suppose it is green). In the same way there must be a pair that does not use color blue. Since it must share at least one color with the first pair, it must have black, and since it must share at least one color with the second pair it must have green. So the other pair is black and green.

Once we know we have these 3 pairs, any other pair must be either black and blue, blue and green or green and black. Since we have more than 75 of these pairs, these colors are used more than 150 times. However, each color is used only 25 times in each of the columns, so they can be used at most 150 times in total, which contradicts the previous statement. Thus, there are two rows of the kind we are looking for in the problem. □

A different version of the pigeonhole principle can also be used for infinite sets.

Proposition 2.1.7 (Infinite pigeonhole principle) *Given an infinite set of objects, if they are arranged in a finite number of places, there is at least one place with an infinite number of objects.*

The proof is analogous to the one for the pigeonhole principle: if in every place there is a finite number of objects, in total there would be a finite number of objects, which is not true.

Example 2.1.8 A 100×100 board is divided into unit squares. In every square there is an arrow that points up, down, left or right. The board square is surrounded by a wall, except for the right side of the top right corner square. An insect is placed in one of the squares. Each second, the insect moves one unit in the direction of the arrow in its square. When the insect moves, the arrow of the square it was in moves 90 degrees clockwise. If the indicated movement cannot be done, the insect does not move that second, but the arrow in its squares does move. Is it possible that the insect never leaves the board?

Solution We are going to prove that regardless of how the arrows are or where the insect is placed, it always leaves the board. Suppose this is not true, i.e., the insect is trapped. In this case, the insect makes an infinite number of steps in the board. Since there are only 100^2 squares, by the infinite pigeonhole principle, there is a square that is visited an infinite number of times.

Each time the insect goes through this square, the arrow in there moves. Thus, the insect was also an infinite number of times in each of the neighboring squares. By repeating this argument, the insect also visited an infinite number of times each of the neighbors of those squares. In this way we conclude that the insect visited an infinite number of times each square in the board, in particular the top right corner. This is impossible, because when that arrow points to the right the insect leaves the board. □

If one is careful enough, one can solve the previous example using only the finite version of the pigeonhole principle. However, doing it with the infinite version is much easier.

2.2 Ramsey Numbers

Example 2.2.1 Prove that in a party of 6 persons there are always three of them who know each other, or three of them such that no two of them know each other.

Solution Consider 6 points in the plane, one for each person in the party. We are going to draw a blue line segment between two points if the persons they represent know each other and a green line segment if they do not. We want to prove that there is either a blue triangle or a green triangle with its vertices in the original points.

Let v_0 be one of the points. From v_0 there are 5 lines going out to the other points, which are painted with two colors. By the pigeonhole principle, there are at least three of these lines painted in the same color (suppose it is blue). Let v_1, v_2, v_3 be three points that are connected with v_0 with a blue segment. If there is a blue segment between any two of them, those two vertices and v_0 form a blue triangle. If there is no blue segment between any two of them, then v_1, v_2, v_3 form a green triangle. □

By solving this problem, we have proved that if we place 6 points in the plane and we join them with lines of two colors, there are three of them that form a triangle of only one color. The question now is if we can generalize this to bigger sets, when we are no longer looking for triangles of one color. That is, given two positive integers l and s, is there a number n large enough such that by placing n points in the plane and joining them with blue or green lines, there are always l of them such that all lines between two of them are blue or there are s of them such that all lines between two of them are green? In the previous example we saw that if $l = s = 3$, then $n = 6$ works. If there are such numbers n, we are interested in finding the smallest one that satisfies this property. If such number exists, it is denoted by

$$r(l, s)$$

and it is called the "**Ramsey's number** of (l, s)".

Exercise 2.2.2 Prove that if $l = s = 3$, then $n = 5$ is not enough.

Exercise 2.2.3 Let l be a positive integer. Prove that $r(2, l)$ exists and that $r(2, l) = l$.

Proposition 2.2.4 *For each pair* (l, s) *of positive integers the Ramsey number* $r(l, s)$ *exists, and if* $l, s \geq 2$, *then* $r(l, s) \leq r(l - 1, s) + r(l, s - 1)$.

Proof We prove this by induction on $l + s$. By Exercise 2.2.3, we know that if one of l, s is at most 2, $r(l, s)$ exists. This covers all cases with $l + s \leq 5$.

Suppose that if $l + s = k - 1$ then $r(l, s)$ exists. We want to prove that if $l + s = k$ then $r(l, s)$ exists. If any of l, s is at most 2, we have already done those cases, so we can suppose $l, s \geq 3$. Notice that $l + (s - 1) = (l - 1) + s = k - 1$, so $r(l, s - 1)$ and $r(l - 1, s)$ exist.

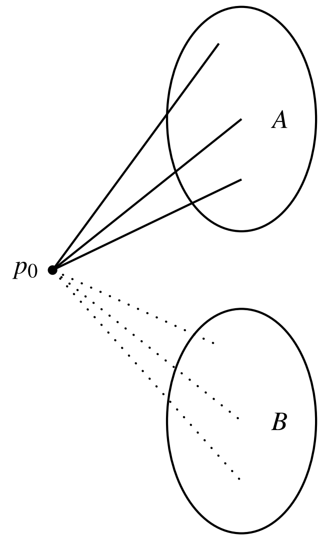

Fig. 2.1 p_0 is joined with *blue lines* with A and with *green lines* with B

Consider a set of $r(l-1,s)+r(l,s-1)$ points in the plane joined with blue or green line segments. We want to prove that there are l of them joined only by blue segments or there are s of them joined only by green segments. Since $r(l,s)$ would be the minimum number that satisfies this, we would obtain that $r(l,s)$ exists and that $r(l,s) \le r(l-1,s)+r(l,s-1)$.

Let p_0 be one of the points. We claim that p_0 must be joined with at least $r(l-1,s)$ points with blue segments or at least with $r(l,s-1)$ points with green segments. If this does not happen, p_0 would be joined with at most $r(l-1,s)-1$ points with blue segments and at most with $r(l,s-1)-1$ points with green segments, so it would not be joined with the other $r(l-1,s)+r(l,s-1)-1$ points. (See Fig. 2.1.)

In other words, using the notation of the figure, we showed that either $|A| \ge r(l-1,s)$ or $|B| \ge r(l,s-1)$. Suppose that p_0 is joined with $r(l-1,s)$ points with blue segments. Among these points there are $l-1$ joined only with blue segments or s joined only with green segments. In the first case, these $l-1$ points and p_0 are only joined with blue segments, and in the second case, we have s points joined only with green segments. If p_0 is connected with $r(l,s-1)$ points with green segments, the way to find the sets is analogous. Thus $r(l,s)$ exists and $r(l,s) \le r(l-1,s)+r(l,s-1)$. $\square$

With this we have proven that the Ramsey numbers exist. However, finding specific Ramsey numbers turns out to be incredibly difficult. Up to this date no Ramsey number $r(l,s)$ is known with both $l,s \ge 5$.

2.3 The Erdős-Szekeres Theorem

Another important application of the pigeonhole principle is to number sequences. The question we are interested in is: If we have a long enough sequence $(c_1, c_2, \ldots, c_k)$ of different numbers, when is there a large increasing subsequence or a large decreasing subsequence? More precisely, given positive integers a and b, is there a k large enough such that any sequence of k different numbers contains either an increasing subsequence of $a + 1$ numbers or a decreasing subsequence of $b + 1$ numbers?

Once we know the Ramsey numbers exist, it is easy to see that there is such a k. This is because if our sequence has at least $r(a+1, b+1)$ numbers, we can associate each of them with a point in the plane. Any two points are going to be joined with a blue segment if the sequence they form is increasing and with a green segment if the sequence they form is decreasing. Then, there must be $a + 1$ points that are joined only with blue segments or $b + 1$ points that are joined only with green segments. This set represents the subsequence we were looking for.

The bound we obtained in this process is $r(a + 1, b + 1)$, which is much larger than we needed. The Erdős-Szekeres theorem gives us the best possible bound for this result.

Theorem 2.3.1 (Erdős, Szekeres 1935) *Given any sequence of $ab+1$ different numbers, there is always an increasing subsequence of at least $a + 1$ numbers or a decreasing subsequence of at least $b + 1$ numbers.*

Proof Consider a sequence $(c_1, c_2, \ldots, c_{ab+1})$ of $ab+1$ different numbers. To each c_j of the sequence we assign a pair (a_j, b_j) of positive integers, where a_j is the length of the longest increasing subsequence that ends in c_j and b_j is the length of the longest decreasing subsequence that ends in c_j.

Given two numbers c_i and c_j in the sequence with $i < j$, we prove that their pairs (a_i, b_i) and (a_j, b_j) cannot be equal. If $c_i < c_j$, we can add c_j to the largest increasing subsequence that ends in c_i, so we have an increasing subsequence of length $a_i + 1$ that ends in c_j. This gives $a_j \geq a_i + 1$. If $c_i > c_j$, we can add c_j to the largest decreasing subsequence that ends in c_i, so we have a decreasing subsequence of length $b_i + 1$ that ends in c_j. This gives $b_j \geq b_i + 1$.

If there were no subsequences of the lengths we were looking for, then for each $1 \leq j \leq ab + 1$ we would have $a_j \leq a$ and $b_j \leq b$. This would give us at most ab different pairs. Since we have $ab + 1$ pairs, by the pigeonhole principle at least two must be equal, which is a contradiction. □

Exercise 2.3.2 Given any two positive integers a and b, find a sequence of ab different real numbers with no increasing subsequences of length $a + 1$ or more and no decreasing subsequences of length $b + 1$ or more.

2.4 An Application in Number Theory

Besides its importance in combinatorics, the pigeonhole principle has various strong applications in number theory. One of the most known ones is in showing that the decimal representation of any rational number[2] is periodic after some point. In other words, after some point is begins to repeat itself.

In this section we show a different application, establishing that every prime number of the form $4k + 1$ can be written as the sum of two squares. To see this we need the following proposition:

Proposition 2.4.1 *For any integers n and u, there are integers x and y not both* 0 *such that $-\sqrt{n} \leq x \leq \sqrt{n}$, $-\sqrt{n} \leq y \leq \sqrt{n}$ and $x - uy$ is divisible by n.*

Proof Let $k + 1 = \lfloor\sqrt{n}\rfloor$ be the largest integer that is smaller than or equal to $\sqrt{n}$, that is, $k \leq \sqrt{n} < k + 1$. Consider the numbers of the form $x - uy$ with x and y in $\{0, 1, 2, \ldots, k\}$. Each has $k + 1$ options, so there are $(k + 1)^2 > n$ possible numbers. Thus (by the pigeonhole principle!), there are two that leave the same remainder when divided by n. If they are $x_1 - uy_1$ and $x_2 - uy_2$, their difference $(x_1 - x_2) - u(y_1 - y_2)$ is divisible by n. If we take $x = x_1 - x_2$ and $y = y_1 - y_2$, we have that they are not both 0, since the pairs (x_1, y_1) and (x_2, y_2) were different. Also, x and y are in the desired intervals. □

With this we are ready to prove that every prime number of the form $4k + 1$ is the sum of two squares. The only thing we need to know about these numbers is that for every prime p of the form $4k + 1$ there is a u such that $u^2 + 1$ is divisible by p. This last result can also be proven in a combinatorial way, counting how many pairs of numbers a, b in $\{0, 1, 2, \ldots, p - 1\}$ satisfy $ab \equiv -1 \pmod{p}$ and how many pairs of different numbers a, b satisfy $ab \equiv 1 \pmod{p}$. However, we will not do it in this book.

Theorem 2.4.2 (Fermat) *Every prime p of the form $4k + 1$ can be written as the sum of two squares.*

Proof Let u be an integer such that $u^2 + 1$ is divisible by p. Using Proposition 2.4.1 we know that there are integers x, y not both 0 such that $x - uy$ is divisible by p and $-\sqrt{p} \leq x \leq \sqrt{p}$, $-\sqrt{p} \leq y \leq \sqrt{p}$. The condition can be translated to $x^2 \leq p$ and $y^2 \leq p$. Since p is prime, it is not a perfect square, so the inequalities are strict. Since $x \equiv uy \pmod{p}$, we have that $x^2 \equiv u^2y^2 \equiv -y^2 \pmod{p}$, so $x^2 + y^2$ is divisible by p. However,

$$0 < x^2 + y^2 < 2p.$$

With this we have that $x^2 + y^2 = p$. □

[2] A rational number is one that can be written as the quotient of two integers.

It is also known that a prime p of the form $4k + 1$ can be written as the sum of two squares in a unique way, so this application of the pigeonhole principle finds the only pair that satisfies this. This shows that even though the technique seems elementary, it gives very precise results.

2.5 Problems

Problem 2.1 Show that given 13 points in the plane with integer coordinates, there are three of them whose center of gravity has integer coordinates.

Problem 2.2 Show that in a party there are always two persons who have shaken hands with the same number of persons.

Problem 2.3 (OIM 1998) In a meeting there are representatives of n countries ($n \geq 2$) sitting at a round table. It is known that for any two representatives of the same country their neighbors to their right cannot belong to the same country. Find the largest possible number of representatives in the meeting.

Problem 2.4 (OMM 2003) There are n boys and n girls in a party. Each boy likes a girls and each girl likes b boys. Find all pairs (a, b) such that there must always be a boy and a girl that like each other.

Problem 2.5 For each n show that there is a Fibonacci[3] number that ends in at least n zeros.

Problem 2.6 Show that given a subset of $n + 1$ elements of $\{1, 2, 3, \dots, 2n\}$, there are two elements in that subset such that one is divisible by the other.

Problem 2.7 (Vietnam 2007) Given a regular 2007-gon, find the smallest positive integer k such that among any k vertices of the polygon there are 4 with the property that the convex quadrilateral they form shares 3 sides with the polygon.

Problem 2.8 (Cono Sur Olympiad 2007) Consider a 2007×2007 board. Some squares of the board are painted. The board is called "charrúa" if no row is completely painted and no column is completely painted.

- What is the maximum number k of painted squares in a charrúa board?
- For such k, find the number of different charrúa boards.

Problem 2.9 Show that if 6 points are placed in the plane and they are joined with blue or green segments, then at least two monochromatic triangles are formed with vertices in the 6 points.

[3] Fibonacci numbers are defined by the formulas (6.3) and (6.4).

Problem 2.10 (OMM 1998) The sides and diagonals of a regular octagon are colored black or red. Show that there are at least 7 monochromatic triangles with vertices in the vertices of the octagon.

Problem 2.11 (IMO 1964) 17 people communicate by mail with each other. In all their letters they only discuss one of three possible topics. Each pair of persons discusses only one topic. Show that there are at least three persons that discussed only one topic.

Problem 2.12 Show that if l, s are positive integers, then

$$r(l,s) \leq \binom{l+s-2}{l-1}.$$

Problem 2.13 Show that $r(3,4)=9$.

Problem 2.14 Show that if an infinite number of points in the plane are joined with blue or green segments, there is always an infinite number of those points such that all the segments joining them are of only one color.

Problem 2.15 (Peru 2009) In the congress, three disjoint committees of 100 congressmen each are formed. Every pair of congressmen may know each other or not. Show that there are two congressmen from different committees such that in the third committee there are 17 congressmen that know both of them or there are 17 congressmen that know neither of them.

Problem 2.16 (IMO 1985) We are given 1985 positive integers such that none has a prime divisor greater than 23. Show that there are 4 of them whose product is the fourth power of an integer.

Problem 2.17 (Russia 1972) Show that if we are given 50 segments in a line, then there are 8 of them which are pairwise disjoint or 8 of them with a common point.

Problem 2.18 (IMO 1972) Show that given 10 positive integers of two digits each, there are two disjoint subsets A and B with the same sum of elements.

Problem 2.19 There are two circles of length 420. On one 420 points are marked and on the other some arcs of circumference are painted red such that their total length adds up less than 1. Show that there is a way to place one of the circles on top of the other so that no marked point is on a colored arc.

Problem 2.20 (Romania 2004) Let $n \geq 2$ be an integer and X a set of n elements. Let $A_1, A_2, \dots, A_{101}$ be subsets of X such that the union of any 50 of them has more than $\frac{50n}{51}$ elements. Show that there are three of the A_j's such that the intersection of any two is not empty.

Problem 2.21 (Tournament of towns 1985) A class of 32 students is organized in 33 teams. Every team consists of three students and there are no identical teams. Show that there are two teams with exactly one common student.

Problem 2.22 (Great Britain 2011) Let G be the set of points (x, y) in the plane such that x and y are integers in the range $1 \leq x, y \leq 2011$. A subset S of G is said to be *parallelogram-free* if there is no proper parallelogram with all its vertices in S. Determine the largest possible size of a parallelogram-free subset of G.

Note: A proper parallelogram is one whose vertices do not all lie on the same line.

Problem 2.23 (Italy 2009) Let n be a positive integer. We say that k is n-square if for every coloring of the squares of a $2n \times k$ board with n colors there are two rows and two columns such that the 4 intersections they make are of the same color. Find the minimum k that is n-square.

Problem 2.24 (Romania 2009) Let n be a positive integer. A board of size $N = n^2 + 1$ is divided into unit squares with N rows and N columns. The N^2 squares are colored with one of N colors in such a way that each color was used N times. Show that, regardless of the coloring, there is a row or a column with at least $n + 1$ different colors.

http://www.springer.com/978-3-0348-0596-4

Problem-Solving Methods in Combinatorics
An Approach to Olympiad Problems
Soberón, P.
2013, IX, 174 p. 65 illus., 10 illus. in color., Softcover
ISBN: 978-3-0348-0596-4
A product of Birkhäuser Basel