

The Pigeonhole Principle

Focus Learning

July 26, 2020

Abstract

The goal of this handout is to let the reader become familiar with the Pigeonhole Principle, an idea used in all levels of competitive mathematics. The Pigeonhole Principle is also known as **Dirichlet's Box Principle**.

Contents

1	Introduction	2
2	Basic Examples	2
3	More Difficult Examples	3
3.1	Connection with the Mean and Expected Value	5
3.2	Pigeons and Points (Geometry)	6
4	Problems	6
4.1	Challenges from the Olympiads	7

1 Introduction

Going with the spirit of the name of the principle, consider 8 pigeons and 7 pigeonholes. It is not hard to realize that if all the pigeons went to a pigeonhole, at least one hole would have more than one pigeon. We can easily generalize this.

Proposition 1.1. *If n objects are placed into k boxes, where $n > k$, then at least 1 box will have more than 1 object.*

Now, what if I had 16 pigeons and 7 pigeonholes. Rather than only saying 1 pigeonhole would have more than 1 pigeon, we can make a stronger claim. In this case, at least one pigeonhole would contain at least 2 pigeons. We can create a better Pigeonhole Principle. That is...

Theorem 1.2. *If there are n boxes and more than $k \cdot n$ objects are being distributed among the boxes, at least one box will contain more than k objects.*

Proof. We proceed by contradiction. Assume that all of the n boxes contain k or less objects. Then, there are a maximum of $k \cdot n$ objects, which contradicts the assumption that there are more than $k \cdot n$ objects, so we are done. \square

We present one last idea for this section.

Proposition 1.3. *If n objects are placed into k boxes, where $n > k$, then at least one box must contain $\left\lfloor \frac{n-1}{k} \right\rfloor + 1$ balls.*

Although we have this formula, it is usually not necessary to memorize it. Typically, common sense will suffice when the problem requires you to use this formula. The proof is left as an exercise for the reader.

For simplicity, from now on, the “pigeons” will be known as balls and the “pigeonholes” will be known as boxes.

2 Basic Examples

Let's start with a simple example.

Example 2.1. There exists a group of 4 friends. What is the minimum amount of apples I should give them such that at least one person will receive more than 1 apple.

Solution. By the Pigeonhole Principle, when there are 5 apples, at least person will receive more than one apple, so our answer is at most 5. However, notice that if there were 4 apples, each child could get 1 apple, so no child would get more than 1 apple. Hence, our answer is $\boxed{5}$.

In some scenarios, the Pigeonhole Principle cannot be directly used, but the idea behind it is the same. Here is an example.

Example 2.2. (2019 AMC) A box contains 28 red balls, 20 green balls, 19 yellow balls, 13 blue balls, 11 white balls, and 9 black balls. What is the minimum number of balls that must be drawn from the box without replacement to guarantee that at least 15 balls of a single color will be drawn?

Solution. If we had an infinite number of balls for each color, then it's easy to see that by the Pigeonhole Principle, the answer would be $14 \cdot 6 + 1 = 85$, since in the worst case scenario we have to draw 14 balls of each color. For the original problem, we can change it so that for the colors that have less than 15 balls, we draw all of them (rather than 14). We can draw all the black, white, and blue balls, while also drawing 14 of the yellow, green, and red balls. The next ball we draw forces some color to have 15 balls, so our answer is 76.

We are not only limited to objects as the balls and boxes; they could be anything. A classic example involves modular arithmetic.

Example 2.3. Prove that if we are given any 11 integers, there exists a pair of integers such that their difference is divisible by 10.

Solution. Here, the balls and boxes aren't that clear. We are only working with integers. However, the words "difference" and "divisible" hint at modular arithmetic. We can express two integers being congruent modulo 10 as

$$a \equiv b \pmod{10}$$

and we are finding the numbers a and b . Note that there are 10 choices for the residue of a and b . But, we are given 11 total numbers. We can let our balls be the 11 numbers and the boxes be the residues. By the Pigeonhole Principle, at least 2 numbers will be congruent mod 10, which implies that their difference is divisible by 10. Those are the two numbers a and b , and we are done.

Our key idea here is determining our balls in boxes. We had to cleverly figure out that using the residues was the crux in this problem. Determining what our balls and boxes are is often the most important part in Pigeonhole problems. In the next example, we will see how we can use this idea in geometry as well.

Example 2.4. Prove that for any 8 points chosen in a unit circle, there exists 2 points that are at most 1 unit away from each other.

Solution. We want to somehow create the boxes so that each box will have more than 1 ball. The balls seem to be the 8 points. We are able to make the boxes by dividing the circle into 7 congruent sectors. By the Pigeonhole Principle, at least one sector will have more than 1 point. In a sector, two points can be at most 1 unit apart, so the two points in the same sector are the two points that are at most 1 unit away from each other ¹.

3 More Difficult Examples

Although in the last section the boxes weren't given to us, it was usually quite easy to identify them. Also, the number of balls was always directly to us. In more advanced problems, however, we often have to create the balls and boxes on our own.

As usual, let's start with an example.

Example 3.1 (1972 IMO). Prove that from a set of ten distinct two-digit numbers (in the decimal system), it is possible to select two disjoint subsets whose members have the same sum.

¹Note that this does not mean 8 is the minimum number of points such that the condition is satisfied.

Solution. We first try to create the balls. Since we want to find two subsets that both satisfy a condition, the subsets seem to have the highest chance at being the balls (In other words, the subsets go into the same box). Consider any set of ten distinct two-digit numbers. There are $2^{10} = 1024$ subsets. However, note that the empty subset and the subset with all the elements will never be able to satisfy the condition we want to prove, so we have to only consider 1022 subsets.

Now, we find the possible sum of the subsets, which will act of the boxes. The smallest sum is simply the subset $\{10\}$ and the largest subset that is not the entire set is $\{91, 92, \dots, 99\}$, which has sum

$$91 + 92 + 93 + \dots + 99 = 770.$$

Each of the 1022 subsets takes on exactly one value between 10 and 770. By the Pigeonhole Principle, two subsets will have the same sum. Further, if those two sets have overlapping elements, we can subtract those elements from both sets². Since we have created the two disjoint subsets, we are done. \square

The crux of this problem was identifying what the balls and boxes were. After that, the Pigeonhole Principle was sufficient. Since the problem statement wanted us to find two subsets that both satisfied a condition, we conjectured that the subsets were the balls. Then, since the problem wanted subsets that had the same sum, we made an educated guess that the value of the sum were the boxes.

Example 3.2. Prove that there exists a multiple of 2020 that consists solely of the digits 1 and 0.

Solution. Though we could bash it, that doesn't seem fun. Since we are dealing with digits, we can try to use modular arithmetic. Consider the set of positive integers that consists only of the digit 1; that is,

$$1, 11, 111, 1111, 11111 \dots$$

Call this set S . We can take each of the numbers in $S \pmod{2020}$. Although we don't know exactly what each number $\pmod{2020}$ is, we know that it will be an integer between 0 and 2019, inclusive. Consider the first 2021 integers in S . Since each of them $\pmod{2020}$ is an integer between 0 and 2019, at least two numbers will be the same $\pmod{2020}$. Call these two numbers a and b . Since $a \equiv b \pmod{2020}$, by definition, taking the positive difference of a and b will result in a number that is a multiple of 2020. Since a and b both consist only of the digit 1, their difference will consist only of the digits 1 and 0, so we are done.

This problem was similar to Example 2.3 in that we considered modular arithmetic. We were able to solve the problem without finding the two numbers. The key idea in this problem was recognizing that there are limited residues, which gives us the idea to let them be the "boxes" in the Pigeonhole Principle.

Example 3.3 (1978 Putnam). Let A be a set formed by choosing 20 numbers arbitrarily from the arithmetic sequence $1, 4, 7, \dots, 100$. Prove that there must be two numbers in A such that their sum is 104.

Solution. It is easy to identify that the 20 numbers are the balls, since we want two of them to satisfy some property. However, the trickiness lies in identifying the boxes. Since we are given an arithmetic sequence, let's write everything in terms of a variable n . We get that the numbers are in the form $3n + 1$ for $0 \leq n \leq 33$. Now, we realize that we want to choose two numbers $3p + 1$ and $3q + 1$ such that they sum to $3(34) + 2$. This can be rewritten as $p + q = 34$, where p and q are non-negative integers less than 34. The values of p and q that work are

$$(33, 1), (32, 2), (31, 3), \dots, (18, 16),$$

so there are 16 pairs in total. These pairs seem like they are our boxes! However, we first have to remove the numbers 0 and 34, since they do not form pairs and would not work if we applied the Pigeonhole Principle. Now, since there are 16 pairs and 18 numbers to choose, 2 numbers will be from the same pair, and those two numbers will be our p and q . Since we have found the two numbers, we are done³. \square

²Note that this action cannot create an empty subset because one of the subsets we have chosen cannot be a subset of the other

³In this problem, 19 numbers works as well.

3.1 Connection with the Mean and Expected Value

We begin this subsection with two theorems.

Theorem 3.4. *If the arithmetic mean of a set of n values is x , then there exists one element that is greater than or equal to x and one element that is less than or equal to x*

Proof. We proceed by contradiction. To prove that there exists an element that is greater than or equal to x , we assume that all elements are less than x . We have

$$\sum_{i=1}^n a_i = nx.$$

However, all the a_i s are less than x , so the equation cannot hold, and we have a contradiction. The proof is similar for proving there exists an element that is less than or equal to x , so we are done. \square

Theorem 3.5. *If the expected value of a variable x is $E(x)$, then there exists a case of x where x is greater than or equal to $E(x)$ and a case of x where x is less than or equal to $E(x)$.*

The proof is very similar to Theorem 3.4 and the theorem is very intuitive, so we will not be proving it. It will be left as an exercise if the reader wishes to prove it.

An immediate corollary very similar to Theorem 3.5 follows.

Corollary 3.6. *If the expected value of x is $E(x)$, and x must be an integer, then there exists a value of x that is greater than or equal to $\lceil E(x) \rceil$ and a value of x that is less than or equal to $\lfloor E(x) \rfloor$.*

Let's see an instructive example of this.

Example 3.7 (2012 USAMO). A circle is divided into 432 congruent arcs by 432 points. The points are colored in four colors such that some 108 points are colored Red, some 108 points are colored Green, some 108 points are colored Blue, and the remaining 108 points are colored Yellow. Prove that one can choose three points of each color in such a way that the four triangles formed by the chosen points of the same color are congruent.

Solution. WLOG, let's consider the red points. Notice that there are 431 rotations that allow the points to overlap. For each of the red points, a randomly chosen rotation will have a $\frac{108}{431}$ chance of coinciding with a point that was previously green. Since there are 108 red points, by Linearity of Expectation, in a single rotation, the amount of red points that land in a spot that was previously green is expected to be

$$\frac{108}{431} \cdot 108,$$

which is slightly greater than 27. By Corollary 3.6, there exists a rotation such that 28 red points are rotated to a position previously held by a green point. This means that on the original circle there exists 28 green points that are rotations from 28 red points.

We now do something similar. For each of the 28 red points already selected, a random rotation has a $\frac{108}{431}$ chance of overlapping with a yellow point. By Linearity of Expectation,

$$\frac{108}{431} \cdot 28$$

red points will overlap with a yellow point. Since this number is slightly greater than 7, there exists a rotation such that 8 of the 28 already selected red points all land on a yellow point. We have now created 8 red, green, and yellow points that are all rotations of each other.

We can finally consider the last color, blue. For each of the 8 red points that we already selected, a random rotation has a $\frac{108}{431}$ chance of overlapping with a blue point. By Linearity of Expectation, the amount of red points that will overlap with a blue point is

$$\frac{108}{431} \cdot 8.$$

This number is slightly greater than 2, so there exists a rotation such that 3 of the red points will overlap with 3 of the blue points.

From the 3 points of each color that we selected, we can construct a triangle of each color. Since they are rotations of each other, they must be congruent, and we are done.

In this problem, we used Corollary 3.6 to figure out the *existence* of a rotation we wanted. Then, we applied the same idea 3 times in order to get the result we wanted. The Theorems in this section are especially useful when the problem looks like a Pigeonhole Problem but cannot be directly solved with it.

3.2 Pigeons and Points (Geometry)

The Pigeonhole Principle is also occasionally seen with geometry. In most cases, it deals with points or lines that satisfy a condition. Some clever thinking on what the boxes will be is usually the main difficulty when solving these types of problems.

Example 3.8 (USAMTS). Every point in a plane is either red, green, or blue. Prove that there exists a rectangle in the plane such that all of its vertices are the same color.

Solution. Choose a rectangle in the plane consisting of 4 columns and 19 rows. Consider the first row. Since there are four points and 3 colors, at least 2 points will be the same color.

Now consider the 19 rows. Since for each row there are 3 ways to choose the color that is repeated and $\binom{4}{2} = 6$ ways to choose which points are repeated, there are 18 possible arrangements of the 3 colors. Since there are 19 rows and 18 possible arrangements of the 3 colors, by the Pigeonhole Principle, 2 rows will be exactly the same and contain at least 2 points of the same color. The points that are the same color in those two rows will form the rectangle, and we are done.

Our next one uses a bit of number theory, but deals with points within a plane.

Example 3.9. Prove that if I choose any 5 distinct lattice points on the coordinate plane and compute the midpoints of all the segments formed by those 5 lattice points, there exists a midpoint that lies on a lattice point.

Solution. It seems quite obvious that the initial 5 points are the balls, so now we just have to determine the boxes. With a little motivation from parity and the fact that there are 5 points, it wouldn't be hard to guess that the parity of the points is the boxes. Since there are 2 choices of parity for the x-coordinate and 2 choices of parity for the y-coordinate of the points, there are $2^2 = 4$ coordinate points with different parities. By the Pigeonhole Principle, two points will have the same parity for both the x and y coordinates. Since the sum of 2 odd numbers and the sum of 2 even numbers is both even, and the midpoint of two points is $\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$, the two points with the same parity for both the x and y coordinates will have a midpoint that is an integer, and is therefore a lattice point.

4 Problems

Exercise 4.1 (Classic). If I have an infinite number of red socks, yellow socks, pink socks, green socks, and blue socks, how many socks do I have to draw I order to guarantee I have a pair of socks that are the same color.

Exercise 4.2. In Foclandia, there exists 6 types of cars, labeled A through F . I want to create a group of people such that at least 4 people have the same type of car. How many people should I gather?

Exercise 4.3. Solve Example 3.2, but 2020 is replaced by n .

Exercise 4.4 (Chattanooga Math Circle). There is an $n \times n$ grid divided into unit squares. Each unit square is filled with either -1, 0, or 1. Is it possible to fill in the grid such that the sum of each row, column, and main diagonal is different?

Exercise 4.5 (Classic). Prove that in a group of n people where $n > 1$, there exists two people that have the same number of friends. (Assume that if person a is friends with person b , then person b is friends with person a as well.)

Exercise 4.6 (2006 AMC). Six distinct positive integers are randomly chosen between 1 and 2006, inclusive. What is the probability that some pair of these integers has a difference that is a multiple of 5?

Exercise 4.7. Prove Theorem 3.5.

Exercise 4.8. In a standard 8×8 chessboard, two opposite corner squares are removed. Prove that it is impossible to tile the remaining chessboard with 2×1 tiles.

Exercise 4.9 (Brilliant). Ten points are placed within a unit equilateral triangle. Show that there exists two points with distance at most $\frac{1}{3}$ apart.

Exercise 4.10 (1996 India RMO). If A is a fifty element subset of the set $1, 2, \dots, 100$ such that no two numbers from A add up to 100, show that A contains a square

Exercise 4.11 (University of Victoria). Prove that if $n + 1$ integers are selected from among $\{1, 2, \dots, 2n\}$, then the selection includes integers a and b such that $\gcd(a, b) = 1$

Exercise 4.12 (2003 Manhattan Math Olympiad). Prove that from any set of one hundred whole numbers, one can choose either one number which is divisible by 100, or several numbers whose sum is divisible by 100.

Exercise 4.13 (1999 AIME). Forty teams play a tournament in which every team plays every other team exactly once. No ties occur, and each team has a 50% chance of winning any game it plays. The probability that no two teams win the same number of games is $\frac{m}{n}$, where m and n are relatively prime positive integers. Find $\log_2 n$.

4.1 Challenges from the Olympiads

This section contains examples of the Pigeonhole Principle found in various Olympiads. Solutions for most of the problems can be found online.

Exercise 4.14 (1997 Japan MO). Take 10 points inside the circle with diameter 5. Prove that for any these 10 points there exist two points whose distance is less than 2.

Exercise 4.15 (2017 India RMO Maharashtra and Goa⁴). Let $P(X)$ and $Q(X)$ be polynomials of degree 6 and degree 3, respectively such that

$$P(X) > Q(x)^2 + Q(X) + x^2 - 6$$

for all real values of X . If all the roots of $P(X)$ are real numbers, then prove that there exists two roots of $P(X)$, say α and β , such that $|\alpha - \beta| < 1$.

⁴Strangely, I wasn't able to easily find this problem online. [This](#) source explains the solution pretty well. Credit goes to **Prithwijiit De** for the linked document.

Exercise 4.16 (1985 IMO). Given a set M of 1985 distinct positive integers, none of which has a prime divisor greater than 23, prove that M contains a subset of 4 elements whose product is the 4th power of an integer.

Exercise 4.17 (2011 ISL). Consider a polynomial $P(x) = \prod_{j=1}^9 (x + d_j)$, where d_1, d_2, \dots, d_9 are nine distinct integers. Prove that there exists an integer N , such that for all integers $x \geq N$ the number $P(x)$ is divisible by a prime number greater than 20.

Exercise 4.18 (2004 China Math Olympiad). Let M be a set consisting of n points in the plane that satisfy the following conditions:

1. There exist 7 points in M which constitute the vertices of a convex heptagon
2. If for any 5 points in M which constitute the vertices of a convex pentagon, then there is a point in M which lies in the interior of the pentagon.

Find the minimum value of n .