Discrete Mathematics 04/03/2019

Lecture 10: Pigeon Hole Principle

Instructor: Goutam Paul *Scribe: Anirban Nath*

Contents

10.1 Introduction

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 *Number Theory* in 1834, and for this reason it is also known as the Dirichlet's *Box Principle*.

Proposition 10.1.1. (Pigeon Hole principle) *If* $(n+1)$ *pigeons are put into* n *holes, then at least one hole has more than one pigeon.*

It is known as the simplest version of **Pigeon-Hole Principle** . Now the very first thing that comes to our mind is the following:

- ***Question*** : How to recognize if the box principle is to be used ?

- ***Answer*** : Every existence problem about finite and, sometimes infinite sets is usually solved by Pigeon Hole Principle **(PHP)**.The principle is a pure existence assertion. It gives no help in finding a multiply occupied box.

Our **objective** and obviously main difficulty is the identification of the pigeons and the holes. The only way to construct ideas and become habituated to the application, is solving problems.

10.2 Problems with most simplified version of PHP

Problem 10.2.1. *There are n married couples. How many persons are to be selected randomly (without considering gender) to guaranty that there is at least one married couple?*

Answer. As stated above, our objective is to find the pigeons and the holes so that we can apply **PHP**. Here the problem urges to find the *existence* of a **couple**. Hence, we take the couples as our "**holes**" and the individuals as our "**pigeons**".

Now, let us suppose that the minimum number of individuals to be selected is x. Then, clearly from the proposition stated above we can find x to be $n + 1$.

So, the minimum number of persons to be selected to guarantee a couple is $n + 1$.

Now, we focus on a special type of problem. These problems deals with some **succession of events**.

Problem 10.2.2. *Given m integers $a_1, a_2, \dots, a_m$, show that there exist integers k, s with $0 \leq k < s \leq m$, such that $a_{k+1} + a_{k+2} + \cdots + a_s$ is divisible by m.*

Answer. Consider the sequence,

$$\begin{gathered} a_1 \\ a_1 + a_2 \\ a_1 + a_2 + a_3 \\ \cdots \\ a_1 + a_2 + a_3 + \cdots + a_m \end{gathered}$$

If any one these m sums is divisible by m, then we are done. Otherwise, suppose none of them are divisible by m. So, each of them leaves a non-zero remainder between $1, 2, 3, \ldots, m-1$. since there are m sums and $(m-1)$ possible values of remainders, then by **PHP**, two of them leaves the same remainder after division by m.
Here, we use remainders as "**holes**" and terms of sequences modulo m as "**pigeons**".
So, let

$$a_1 + a_2 + \cdots + a_k = bm + r$$

$$a_1 + a_2 + \cdots + a_s = cm + r$$

This gives (if $k < s$), $a_{k+1} + a_{k+2} + \cdots + a_s = m(c - b)$. Thus, m divides $a_{k+1} + a_{k+2} + \cdots + a_s$.

Problem 10.2.3. *A chess-master who has 11 weeks to prepare for a chess-tournament decides to play at least one game everyday, but no more than 12 games in a week. Show that there exists a succession of days during which the chess-master plays exactly 21 games.*

Answer. Observe that the question wants us to prove only the **existence** of such successive days, not trying to specify which days are those. And this is an absolute chance to apply **PHP**.
Let, a_r denote the total number of games played on the first r days for $1 \leq r \leq 77$.
Hence, $1 \leq a_1 < a_2 \cdots < a_{77}$. Now the chess-master plays at least one games per day. So we have $a_{i+1} \geq a_i + 1$. Also, since at most 12 games are played during a week, $a_{77} \leq (12 \times 11) = 132$. Thus $a_{77} + 21 \leq 153$.
Now, consider the sequence of 154 positive integers,

$$a_1, a_2, \ldots, a_{77}, a_1 + 21, a_2 + 21, \ldots, a_{77} + 21$$

where both a_i and $a_i + 21 \leq 153 \ \forall \ i$. So by **PHP** at least two of the elements in the sequence are equal. But, $a_1 < a_2 < \cdots < a_{77}$ and $a_1 + 21 < a_2 + 21 < \cdots < a_{77} + 21$. Hence $a_i = a_j + 21$ for some $j < i$.
Hence $a_j - a_i = (j+1)^{th}$day games $+ \cdots + i^{th}$day games $= 21$.

Remark 10.2.1. We can generalise the above problem as follows:

Suppose a chess master decides to play for d *consecutive days, playing at least 1 game per day and a total of no more than* b *games where* $d < b < 2d$. *Then for each* $i \leq 2d - b - 1$ *there is a succession of days on which, in total, the chess-master plays exactly* i *games.*

We list some similar problems here as an exercise:

Problem 10.2.4. *A storekeeper's list contains 115 items each marked "available" or "unavailable" of which 60 are available. Show that, there are at least 2 "available" items in the list exactly 4 items apart.*

Problem 10.2.5. *Consider a number* m *such that* $gcd(10, m) = 1$. *Then* m *divides one of the integers in the sequence* $1, 11, 111, 1111, \ldots$.

And some more problems of simple **PHP** is given below:

Problem 10.2.6. *Prove that if 100 integers are chosen from the integers* $1, 2, 3, \ldots 200$ *such that at least one of them is smaller than 15, then there exist two of them such that one divides the other.*

Answer. Let us suppose that we have chosen 100 positive integers, not exceeding 200, none of which divides the other. We will show that none of the numbers from 1 to 15 is contained among those 100 numbers.
Let us consider all the *greatest* odd divisors of the chosen numbers. It is obvious that these divisors form the set of all odd numbers not exceeding 200. In particular, these odd divisors include the numbers $1, 3, 9, 27, 81$. Since among the numbers corresponding to these odd divisors there are no two numbers one of which is divisible by one another, the number containing the odd factor 27 must be divisible by a power of 2 with exponent at least 1,the number containing the odd factor 9 must be divisible by a power of 2 with exponent at least 12, the number containing the odd factor 3 must be divisible by a power of 2 with exponent at least 3, the number containing the odd factor 21 must be divisible by a power of 2 with exponent at least 4. This means that the numbers 1, 2, 3, 4, 6, 8, 9, 12 are not contained in those numbers.

In the same way, we can consider the numbers 5, 15 and 45 and prove that the numbers 5, 10 and 15 are not contained in those numbers. Considering 7, 21 and 63 we can exclude 7 and 14. We can exclude 11 by considering 11 and 33, and finally we can exclude 13 by considering 13 and 39.

Problem 10.2.7. *Among the integers* $1, 2, \ldots 200$ *if 101 integers are chosen, show that there exists two such that one divides the other.*

Hint. Write any number as $m = 2^d.q$ where q is odd. Apply PHP.

Problem 10.2.8. *Let* $(x_i, y_i), 1 \leq i \leq 5$*, be a set of five distinct points with integer co-ordinates in* $x - y$ *plane. Show that the mid-point of the line joining at least one pair of points has integer co-ordinates.*

10.3 Generalized Pigeon Hole Principle and Problems

Proposition 10.3.1. (Generalized PHP) *If we put* $Nk + 1$ *or more pigeons in* N *pigeon-holes, then some pigeon-hole must contain at least* $k + 1$ *pigeons.*

Proof. Suppose not, then each of the holes contains less than or equal to k pigeons. Then, the total number of pigeons is at most kN. But we had $Nk + 1$ pigeons. Hence contradiction. □

Now we look into some problems concerned with this general version of PHP.

Problem 10.3.1. *Given 8 different natural numbers, none greater than 15, show that at least three pairs of them has same positive difference.(The pairs need not to be disjoint as sets)*

Answer. We have here 14 possible differences between the 8 given numbers(the values of differences being 1 through 14). These are the 14 pigeonholes. But what are the pigeons? They must be the differences between pairs of given numbers. However there are 28 pairs, and we can fit them in our 14 holes in such a way that there are exactly two pigeons in each hole (and therefore no hole containing three). Here, we must consider additional constraints because we cannot put more than one pigeon in hole numbered 14 because 14 can be obtained only in one way : $15 - 1$. This means that the remaining 27 pigeons must be put in 13 holes. The generalized PHP gives us the result.

Problem 10.3.2. *What is the largest number of squares on an 8×8 chessboard which can be coloured green, so that in any arrangement of L-trominos, at least one square is not coloured green?*

Answer. The answer is **32**. Indeed, suppose that 33 or more squares are coloured green. Then, we divide the board into sixteen 2×2 squares. Now, our pigeons are green squares and the holes are 2×2 squares.Then, PHP guarantees that there is at least one 2×2 square that contains 3 green square. These 3 forms room for the $forbidden$ tromino, that is a contradiction. An example of a optimal design with required property is putting all the black squares green.

Problem 10.3.3. *Let $a_1, a_2, \ldots, a_n$ ($n \geq 5$) be any sequence of positive integers. Prove that, it is always possible to select a subsequence and add or subtract its elements such that the sum is a multiple of n^2.*

Answer. Consider all subsets $\{i_1, i_2, \ldots, i_k\}$ of the set $\{1, 2, \ldots, n\}$. Let, $S(i_1, i_2, \ldots, i_k) = \sum_{j=1}^{k} a_{i_j}$. The number of such terms is $2^n - 1 > n^2$ for $n \geq 5$. Thus, two of these sums will have the same remainder upon division by n^2. Their difference will be divisible by n^2. This difference has the form $\pm a_{s_1} \pm a_{s_2} \pm \cdots \pm a_{s_t}$ for some $t \geq 1$ and some selection of indices $s_1, \ldots, s_t$.

Problem 10.3.4. *There are 15 computers and 10 printers. At most 10 computers can request for printing at a time. Find the minimum number of connections necessary such that all the requests can be served without delay. [Each link from a particular computer to a particular printer must be treated as a connection]*

Answer. First, we try to find the number of connections that are sufficient for such a configuration.

Claim. *60 is a sufficient number of connection.*

Proof. Let, we call the computers as $C_1, C_2, \ldots, C_{15}$ and the printers as $P_1, P_2, \ldots, P_{10}$. We construct one connection each from C_i to P_i for $i = 1, 2, \ldots, 10$ and then connect all printers to each of C_j for $j = 11, \ldots, 15$. It is easy to observe that for any 10 request, we have a feasible option for distributing those to the printers. □

Claim. *60 is the minimum requirement.*

Proof. Suppose not, there exist a solution with less than 60 connections. There are 10 printers. Thus by generalised PHP we can claim that, there exist one printers that is connected to at most 5 computers. WLOG, let that printer be P_1 and it is connected to $C_1, C_2, \ldots, C_5$. Thus, the remaining 9 printers are connected to the remaining 10 computers. Now, one possible option is the 10 requests are coming from those 10 computers namely $C_6, C_7, \ldots, C_{15}$. So, any number of connections less than 60 is not sufficient. Hence, our minimum requirement is 60. □

Problem 10.3.5. *Ten students solved a total number of 35 problems in a math Olympiad. Each problem was solved by exactly one student. There is at least one student who solved exactly 1 problem, at least one student who solved exactly 2 problems and at least one student who solved exactly 3 problems. Prove that there is also at least one student who has solved at least 5 problems.*

Hint. At least $1+2+3 = 6$ problems were solved by student mentioned in the problem statement.

Problem 10.3.6. *On a certain planet in the solar system Tau Cetus, more than half the surface of the planet is dry land. Show that the Tau Cetans can dig a tunnel straight through the centre of the planet, beginning and ending on a dry land.*

Hint. Colour each point on the dry land *red* and the point diametrically opposite to it *green*.

10.4 Another form of Pigeonhole Principle

Proposition 10.4.1. *If the average of n positive numbers is t, then at least one of the numbers is greater than or equal to t. Further, at least one of the numbers is less than or equal to t.*

Proof. Let, $a_1, a_2, \ldots, a_n$ be the numbers. Then by data,

$$\frac{a_1 + a_2 + \cdots + a_n}{n} = t \qquad a_1 + a_2 + \cdots + a_n = tn.$$

Hence, if each of the n numbers $a_1, a_2, \ldots, a_n$ is less than t, then the sum of these numbers will be less than tn contradicting the above equation.

A similar argument shows that at least one of the numbers is less than or equal to t. □

Remark 10.4.1. *If the numbers $a_1, a_2, \ldots, a_n$ are integers, then by the above form of PHP we can say that at least one of them is $\geq t_0$. where t_0 is the smallest integer not less than t; and at least one is $\leq [t]$ where $[t]$ is the integral part of t.*

Now, let us discuss the strong form of PHP and then we will discuss problems and lead our discussions to Erdös-Szekeres theorem.

10.5 Strong form of Pigeon Hole Principle

Proposition 10.5.1. *Let $q_1, q_2, \ldots, q_n$ be positive integers. If $(q_1 + q_2 + \cdots + q_n - n + 1)$ objects are put into n boxes, then the first box contains at least q_1 objects, or the second box contains at least q_2 objects or ... or n^{th} box contains at least q_n objects.*

Proof. Suppose we distribute $(q_1 + q_2 + \cdots + q_n - n + 1)$ objects into n boxes and if for each $i = 1, 2, \ldots, n$, the i^{th} box contains less than q_i elements, then the total number of objects into n boxes is

$$\leq (q_1 - 1) + (q_2 - 1) + \cdots + (q_n - 1) = q_1 + q_2 + \cdots + q_n - n.$$

But this number is less than the total number of objects in n boxes. Hence, for at least one i, i^{th} box must contain at least q_i objects. □

10.6 Various problems with application of PHP

Problem 10.6.1. *Let* $a_1, a_2, \ldots, a_{100}$ *and* $b_1, b_2, \ldots, b_{100}$ *be any two permutations of the integers 1 to 100. Prove that among the hundred products* $a_1b_1, a_2b_2, \ldots, a_{100}b_{100}$*, there are two products whose difference is divisible by 100.*

Answer. Suppose that the 100 products a_ib_i leave 100 different remainders when divided by 100. Then 50 of them must be odd and the remaining 50 must be even since their remainders are just a permutation of $1, 2, \ldots, 100$. The 50 odd products will use up all the odd a_i and odd b_i, Hence the even products are products of even numbers and are therefore divisible by 4. But then none of the products will be of the form $4k + 2$, which is a contradiction.

Problem 10.6.2. *Prove that no 7 integers, not exceeding 24, can have sums of all subsets different.*

Answer. Suppose S be any $7 - subset$ of $\{1, 2, \ldots, 24\}$.The number of non-empty subsets of S having at most 4 elements is $\binom{7}{1} + \binom{7}{2} + \binom{7}{3} + \binom{7}{4} = 98$.

If T is one of those subsets, then the sum of elements in T is in between 1 and (21 +22 +23 +24) = 90. Since $90 < 98$, by PHP it follows that the sum corresponding of the above 98 subsets cannot all be different.

Note that for the 6-element subset $\{11, 17, 20, 22, 23, 24\}$ sums corresponding to all subsets are different.

Problem 10.6.3. *There are 1958 computers which can communicate among themselves in 6 different languages with the provision that any two computers can communicate in 1 languages out if those 6. Prove that, there exist at least 3 computers whose mutual language, two by two, is the same.*

Hint. $1958 = 326 \times 6 + 1 \qquad 326 = 65 \times 5 + 1 \qquad 65 = 16 \times 4 + 1$ $16 = 5 \times 3 + 1$. Use generalized PHP successively.

Problem 10.6.4. *Two of 70 distinct positive integers less than 200 have differences of 4, 5 or 9.*

Hint. Consider the sets $\mathcal{A} = \{a_i | i = 1, 2, \ldots, 70\}$, $\mathcal{B} = \{a_i + 4 | i = 1, 2, \ldots, 70\}$ and $\mathcal{C} = \{a_i + 9 | i = 1, 2, \ldots, 70\}$

Problem 10.6.5. *The length of each side of a convex quadrilateral $ABCD$ is less than 24. Let, P be any point inside $ABCD$. Prove that there exist a vertex, say A, such that $|PA| < 17$.*

Answer. Suppose that all 4 distances of P from the 4 vertices are ≥ 17. Join P to A, B, C, D. Then at least one of the 4 angles at P is $\geq 90^o$. Suppose it is $\angle APB$. Then,

$$24^2 = 576 \geq |AB|^2 \geq |PA|^2 + |PB|^2 \geq 17^2 + 17^2 = 578$$

which is a contradiction.

Problem 10.6.6. *Prove that among any seven real numbers $y_1, y_2, \ldots, y_7$, there exists two of them such that,*

$$0 \leq \frac{y_i - y_j}{1 + y_i y_j} \leq \frac{1}{\sqrt{3}}$$

Hint. Take $y_i = tan(v_i)$ and identify that $tan\frac{\pi}{6} = \frac{1}{\sqrt{3}}$

Problem 10.6.7. [1] *There are 650 points inside a circle of radius 16. Prove that there exists a ring with inner radius 2 and outer radius 3 covering 10 of these points.*

Answer. The most crucial observation is that ***the point P belongs to a ring with centre at O iff the point O belongs to a congruent ring with centre at P***. Thus it is sufficient to prove that, If we consider all such rings with centres in the given points, then, one of these points will be covered by 10 rings. These rings lie inside a circle of radius $16 + 3 = 19$ with area $19^2\pi = 361\pi$. Now, observe that $9 \times 361\pi = 3249\pi$, but the sum of the areas of the rings is $(3^2\pi - 2^2\pi) \times 650 = 3250\pi$.

[1] This is a quite hard problem. It is included only to show a beautiful application of PHP

10.7 Erdös-Szekeres Theorem

Theorem 10.7.1. *Every sequence $a_1, a_2, \ldots, a_{mn+1}$ of $mn+1$ distinct reals contains either an increasing sub-sequence of length $m+1$ or a decreasing sub-sequence of length $n+1$.*

Proof. Suppose not, thus any increasing sub-sequence has length at most m and any decreasing sub-sequence has length at most n.
Now, WLOG, the list is ordered and call it $\mathcal{S}$. We start with a_1, and list the decreasing sub-sequence starting from a_1. Suppose, it is as follows:

$$a_1 = b_{11} > b_{12} > \cdots > b_{1j_1} \qquad \mathcal{A}_1 = \{b_{1k} | k = 1, 2, \ldots, j_1\}.$$

Now consider $\mathcal{S}_1 = \mathcal{S} \setminus \mathcal{A}_1$. Consider the first element in $\mathcal{S}_1$ and call it b_{21} and consider the largest sub-sequence starting from it. Suppose it is as follows:

$$b_{21} > b_{22} > \cdots > b_{2j_2} \qquad \mathcal{A}_2 = \{b_{2k} | k = 1, 2, \ldots, j_2\}.$$

Now consider $\mathcal{S}_2 = \mathcal{S}_1 \setminus \mathcal{A}_2$. Consider the first element in $\mathcal{S}_2$ and call it b_{31} and consider the largest sub-sequence starting from it and so on until the list is exhausted.
Finally we get something like this:

$$\begin{aligned} &b_{11} > b_{12} > \cdots > b_{1j_1} \\ &b_{21} > b_{22} > \cdots > b_{2j_2} \\ &b_{31} > b_{32} > \cdots > b_{3j_3} \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &b_{p1} > b_{p2} > \cdots > b_{pj_p} \end{aligned}$$

where each of $j_1, j_2, \ldots, j_p \leq n$. Now, observe that,

$$b_{11} < b_{21} < \cdots < b_{p1}$$

because if $b_{r1} > b_{(r-1)1}$, then $b_{(r-1)1} \in \mathcal{A}_r$ which is not the case.

So,the total number of elements is $\leq np$. But we have a total of $mn+1$ elements. Thus by PHP, p must be strictly grater than m since the listing is ***exhaustive***. Thus, we have obtained an increasing sequence of minimum $m+1$ elements. Hence, contradiction! □

For any further reference, you can check the following website that covers several proofs of the aforesaid theorem.

http://www.cs.umd.edu/~gasarch/BLOGPAPERS/subseq4proofs.pdf

Further discussions about this theorem will be provided in the following lecture.