Pigeonhole principle for mathletes

David S. Gunderson

Fall 2019

Sorry, references and dates are not all filled in.

1 Pigeonhole Principle (Schubfach Prinzip, Box principle)

If r + 1 pigeons were to roost in r holes, then two pigeons would just have to get acquainted. In general, if r(m-1) + 1 pigeons were to roost in r holes, then there would be at least one hole with at least m pigeons in it.

Define $[r] = \{1, 2, 3, ..., r\}$. Partitioning problems are often written in terms of colourings. For example, a partition of a set X into three parts, say $X = X_1 \cup X_2 \cup X_3$ can be seen as a function $f: X \to [3]$, where for each $i \in [3]$, $f^{-1}(i) = X_i$, sometimes called the *i*-th colour class, and can be thought of as the *i*-th hole or box.

Written in the notation of colourings, the (finite) pigeonhole principle says:

Lemma 1.1 (Pigeonhole principle (PHP), finite) For each $r \ge 1$ and $m \ge 1$, and any coloring

$$\Delta: [r(m-1)+1] \longrightarrow [r]$$

there exist m elements of $\{0, 1, 2, ..., r(m-1)\}$ which are monochromatic with respect to Δ (coloured the same), that is, there is some $i \in [r]$ so that $|\Delta^{-1}(i)| \geq m$.

Proof: Let r and m be fixed, and suppose that

$$\Delta: [r(m-1)+1] \longrightarrow [r]$$

is a given r-colouring. If the conclusion were false, that is, if there are no m elements monochromatic, then each colour class would have at most m-1 elements, giving at most r(m-1) elements in all, a contradiction.

Another version of PHP says that if kn + 1 elements are partitioned into k classes, one class contains n + 1 elements.

Similarly, if one were to divide an infinite set into two 'smaller' sets, then one of them must be infinite also (if they were both finite, putting them back together gives just a finite set, obviously not the whole set!). The same argument works for dividing an infinite set into any finite number of pieces. This idea can be given in the form of a lemma, also called the *pigeon hole principle*; the symbol ω is the first infinite ordinal, often written $\omega = \{0, 1, 2, \ldots\}$.

Lemma 1.2 (Pigeonhole principle, infinite) For any positive integer r and $\Delta : \omega \longrightarrow [r] = \{1, 2, ..., r\}$, there exists an $i \in [r]$ so that $\Delta^{-1}(i)$ is infinite.

There are many interesting applications of the PHP, of which only a few are mentioned.

Lemma 1.3 Any rational number has either a finite decimal expansion or a repeating decimal expansion.

Proof idea: In the division algorithm, dividing q into p is done with at most q remainders. As soon as one remainder is reused, the algorithm repeats itself. If 0 is a remainder, the algorithm stops.

Lemma 1.4 Let n > 1 be an integer. If n + 1 distinct numbers are chosen from $\{1, 2, ..., 2n - 1\}$, then two of the numbers sum to 2n.

Proof: Consider the *n* sets $H_1 = \{1, 2n-1\}, H_2 = \{2, 2n-2\}, \ldots, H_{n-1} = \{n-1, n+1\}, H_n = \{n\}$. By the PHP, two of the chosen numbers must be in some H_i , and since H_n consists of a single element, this H_i must be of the form $H_i = \{i, 2n-i\}$.

Lemma 1.5 For any positive integer n, there exist positive integers s and t with $1 \le s < t \le 11$ so that $10 \mid (n^t - n^s)$.

Proof: Fix n and examine the numbers $n, n^2, n^3, \ldots, n^{11}$. By the PHP, two of them must have the same last digit, say n^s and n^t . Subtracting, arrive at a number with 0 as the last digit. \Box

Lemma 1.6 Let m > 1 be a fixed positive integer, and let $W = \{w_1, w_2, \ldots, w_n\} \subset \{1, 2, \ldots, m\}$. If $2^n - 2 > (n - 1)m$, then there exist nonempty disjoint subsets W_1 , W_2 of W so that $\sum_{w_i \in W_1} w_i = \sum_{w_i \in W_2}$.

Proof: The number of nonempty proper subsets X of W is $2^n - 2$ and the sum of the numbers in each such X is contained in $\{1, 2, \ldots, (n-1)m$, so by the PHP, there are two distinct subsets X and Y so that the sum of the numbers in X is equal to the sum of the numbers in Y. If $X \cap Y = \emptyset$, there is nothing to show. If $X \cap Y \neq \emptyset$, then remove the common numbers from each creating X_0 and Y_0 . Since the same sum was removed from each, the sums of numbers in X_0 and Y_0 still agree. Neither of X_0 nor Y_0 can be void since otherwise one of X or Y would have been contained in the other as a proper subset, and in this case the sums can not agree.

The above lemma can be reformulated in many ways. For example, if ten weights are between 1 and 99 grams, two groups of these weights can balance a scale. Another example is if ten items priced from 1 cent to 99 cents, two people can always make *independent* purchases that are priced the same.

Here is another (surprising?) result that follows by the PHP.

Lemma 1.7 For every odd positive integer m, there is an n such that m divides $2^n - 1$.

For each $k \in \mathbb{N}$ with $2^k - 1 \ge m$, examine $f(k) = 2^k - 1 \pmod{m}$. By the PHP, fix k_1 and k_2 with $k_1 > k_2$ so that $f(k_1) = f(k_2)$. Then there exists $\ell_1 > \ell_2$ and $0 \le r < m$ so that

$$2^{k_1} - 1 = \ell_1 m + r,$$

$$2^{k_2} - 1 = \ell_2 m + r.$$

Subtracting these two equations gives $(\ell_1 - \ell_2)m = 2^{k_1} - 2^{k_2} = 2^{k_2}(2^{k_1-k_2} - 1)$. Since *m* is odd, *m* is relatively prime to 2^{k_2} and so *m* must divide $2^{k_1-k_2} - 1$; then $n = k_1 - k_2$ is as desired. \Box

This next lemma uses precisely the same trick, but yields a surprisingly strong statement. This comes from a German mathematical contest a few years ago (Bundesweltbewerb Mathematik 1995, 2. Runde) and was first shared with me by Martin Stein. Recall that any rational number between 0 and 1 has a decimal expansion which either terminates or repeats. If the period of repetition is r, then a standard exercise shows that the number can be written as p/q where q is a string of r 9's. (For example, .347347347...=347/999.) This implies that any prime not equal to either 2 or 5 has a multiple whose digits are all 9's. The following result extends this notion.

Lemma 1.8 For every k > 1 there exists $m < 12k^4$ which is a multiple of k, so that m contains at most 4 different digits in its decimal representation.

Proof: Fix k. Let S be the set of natural numbers consisting of only 0's and 1's in its decimal representation. For every $s \in S$, consider $f(s) = s \pmod{k}$. By the PHP, there exist s_1 and s_2 so that $f(s_1) = f(s_2)$, that is, there exist $\ell_1 > \ell_2$ and $0 \le r < k$ so that

$$s_1 = \ell_1 k + r$$

$$s_2 = \ell_2 k + r.$$

Subtracting, $s_1 - s_2 = (\ell_1 - \ell_2)k$. It is not hard to check that $m = s_1 - s_2$ now has only the digits 0,1,8,9 in its representation.

Next, it is shown that m can be bounded above by $11.2 \times k^{\log_2(10)}$, less than a constant times k^4 . This is since there are 2^{ℓ} elements in S with $\ell + 1$ digits, and so if $2^{\ell} > k$, restrict S to those elements with at most $\ell + 1$ digits (in fact, one can restrict to a set with slightly smaller numbers, but this makes stating the result more complicated). It suffices then to have $\ell > \log_2(k)$; in this case, the maximum element in S need be $\sum_{i=0}^{\ell+1} 10^i < 11.2 \times 10^{\ell} < 11.2k^{\log_2(10)}$.

The proof yields a slightly stronger statement: for each integer k > 1, there exists a multiple m of k which uses only the digits 0,1,8,9. By the way, if one repeats the above proof with 0's and 5's playing the role of 0's and 1's in S, one can always get a multiple of k that uses only the digits 0,4,5,9 (with a slightly different bound on m).

The PHP yields some very classic results regarding partitioning integers.

Lemma 1.9 If n + 1 integers are chosen from $\{1, 2, ..., 2n\}$, one of these integers must divide the other.

Proof: Write each of the chosen integers as $2^t s$ where s is odd. Since there are only n values the odd component can take, two of them must agree, that is, there is some s' so that two of the chosen numbers are of the form $2^{t_1}s'$ and $2^{t_2}s'$, in which case one divides the other.

Theorem 1.10 (Erdős, Szekeres, 1935) For a fixed positive integer n, in any sequence of $n^2 + 1$ distinct integers there is a strictly monotone subsequence of length (at least) n + 1.

Proof: Let $x_0, x_1, x_2, \ldots, x_{n^2}$ be a sequence of distinct integers.

First make a simple (but essential) observation: Examine two elements of the sequence, say x_i and x_j , where $0 \le i < j \le n^2$. If $x_i < x_j$ then the longest increasing subsequence beginning with x_i is longer than the longest increasing subsequence beginning with x_j ; similarly, if $x_i > x_j$ then the longest decreasing subsequence beginning with x_i is longer than the longest decreasing subsequence beginning with x_i is longer than the longest decreasing subsequence beginning with x_i .

Label each x_i with a pair (a_i, b_i) indicating the lengths of the longest increasing/decreasing subsequences beginning with x_i ; no two pairs can be identical. If all monotone subsequences have length at most n, that is, if for each $i, 1 \leq a_i, b_i \leq n$, then by the PHP, two pairs must agree—but that is impossible. So at least one of the a_i or b_i must be larger than n.

The next theorem is known by some as the Bolzano-Weirstrass theorem; one half of its proof uses the infinite version of PHP (Lemma 1.2) in a rather subtle way.

Theorem 1.11 Let $S \subset \mathbb{R}^n$. S is closed and bounded if and only if every infinite subset X of S, at least one limit point of X is contained in S.

Proof: Let $S \subset \mathbb{R}^n$ be closed and bounded and let $X \subset S$ be infinite. By the Heine-Borel theorem, S is compact. If S contained no limit points of X, then for each $s \in S$, there exists a neighbourhood N_s containing s so that $|N_s \cap X| \leq 1$, that is, N_s contains at most one point from X; the set $\{N_s : s \in S\}$ is a cover for S, and no finite subcover covers X, and no finite subcover covers S, contradicting S being compact.

Now suppose that every infinite subset of S has a limit point in S. If S is not bounded, picking successively larger points creates an infinite sequence with no limit points, so S is bounded. It remains to show that S is closed.

In hopes of a contradiction, suppose that S is not closed, that is, S does not contain one of its limit points, say, $p \in \mathbb{R}^n \setminus S$. For each $k = 1, 2, 3, \ldots$, there is a point $s_k \in S$ with $|p - s_k| < 1/k$. Put $S' = \{s_1, s_2, s_3, \ldots\}$. If S' is finite, then by the PHP, there exists some $s' \in S'$ so that |p - s'| < 1/kfor infinitely many values of k—an impossibility—so S' is infinite. Observe that p is a limit point of S'; in fact, p is the only limit point of S', since for any $q \neq p$, if k is large enough,

$$|s_k - q| \ge |p - q| - |s_k - p| \ge |p - q| - \frac{1}{k} \ge \frac{1}{2}|p - q|$$

showing that q is not a limit point. Since S' is infinite, by assumption, S' has a limit point in S, and p is the only limit point of S', contradicting that $p \notin S$. So conclude that S is closed.

This next result was proved by Dirichlet in 1879, and is why the PHP is sometimes called *Dirichlet's principle*, even though he was not the first to use the PHP. The statement is a bit complicated, but I give it here for historical reasons.

Theorem 1.12 (Dirichlet, 1879) Let x be any real number. For any positive integer n, there is a rational number p/q (where p and q are integers) such that $1 \le q \le n$ and

$$\left|x - \frac{p}{q}\right| < \frac{1}{nq}.$$

Proof: Let $\{x\}$ denote the fractional part of x; for example, if x = 3.12, then $\{x\} = .12$.

If x is already rational, then there is nothing to prove. So suppose that x is not rational (called irrational) and consider the n + 1 numbers

$$\{x\}, \{2x\}, \{3x\}, \dots, \{nx\}, \{(n+1)x\}.$$

Putting these numbers in the pigeonholes

$$(0, 1/n), (1/n, 2/n), \dots, (\frac{n-1}{n}, 1),$$

two must be in the same pigeonhole, say $\{ax\}$ and $\{bx\}$ with a < b, and so differ by at most 1/n. Put q = b - a; since q is the difference between two different numbers in 1, 2, ..., n + 1, it follows that $1 \le q \le n$.

It remains only to see that there is an integer p so that |qx - p| < 1/n, and then division by q finishes the theorem.

The PHP has some standard applications in geometry.

Lemma 1.13 If five points are chosen from an equilateral triangle of side length 2, then two of these points must be within distance 1.

Proof: By joining midpoints of each side, divide the equilateral triangle into 4 equilateral triangles each of side length 1, and apply the PHP. \Box

Similarly, one can ensure that for any five points in a square of side length 2, two must be at distance at most $\sqrt{2}$ (partition the square into four equal subsquares and apply the PHP).

A *lattice point* in the Euclidean plane is one with integer coordinates.

Lemma 1.14 For any five lattice points in the plane, one pair has a midpoint that is also a lattice point.

Proof: For i = 1, ..., 5, let $P_i = (x_i, y_i)$ be a lattice point. Each is of one of the following four types: odd-odd, odd-even, even-odd, even-even, so by the PHP, two must be of the same type, say P_i and P_j . The midpoint of $\overline{P_iP_j}$ is $(\frac{x_i+x_j}{2}, \frac{y_i+y_j}{2})$, a lattice point since odd plus odd is even and even plus even is even.

Similarly, for any nine lattice points in the plane, no three collinear, there is a midpoint that is a lattice point (argue modulo 3, with $2^3 = 8$ types).

One can discuss partitioning 'subtrees of an infinite tree' in order to give a very useful result.

A partially ordered set (or simply a poset) (P, \leq) is a set P together with a relation, \leq that is reflexive $(p \leq p)$, antisymmetric $(p \leq q \text{ and } q \leq p \text{ implies } p = q)$, and transitive $(p \leq q \leq r \text{ implies } p \leq r)$, that is, \leq is a partial order. The relation \leq is a total order, or linear order if for any two elements $p, q \in P$, either $p \leq q$ or $q \leq p$ holds. A poset (T, \leq) is a tree if for every $t \in T$, the set $\{x \in T : x \leq t\}$ is a totally ordered set with no infinite descending subsequence. A tree (T, \leq) is said to be rooted if there exists a unique vertex (called the root) $v \in T$ with the property that $v \leq x$ for every $x \in T$. A vertex $y \neq x$ is a successor of x if $x \leq y$ and for any $z \neq x$ satisfying $x \leq z \leq y$, y = z holds. A tree is locally finite if every vertex has finitely many successors. A branch of a tree is a maximal linearly ordered subset.

Lemma 1.15 (König's infinity lemma, 1927) A locally finite rooted infinite tree has an infinite branch.

Proof: Let (T, \leq) be a locally finite tree with root v, and let $v_0^1, v_1^1, \ldots, v_i^1, \ldots, (i \in I^1$ be a labelling of the successors of v. By Lemma ??, one of the trees

$$\{(T_i^1, \leq) : T_i^1 = \{x \in T : v_i^1 \leq x\} : i \in I^1\}$$

is infinite, say T_0^1 (with root v_0^1) is one such. Repeat this idea using trees having roots that are successors of v_0^1 to obtain another infinite tree T_0^2 . Continue in this manner to get an infinite number of trees, $T_0^1, T_0^2, T_0^3, \ldots$ Then the vertices v, v_0^1, v_0^2, \ldots determine an infinite path. \Box

2 A few common problem themes

In this section are three problems (with solutions), that might reveal tricks needed in the exercises below. The first is relatively easy, but the last two are challenging. Try to understand the solutions, and see how they can be adapted for some of the exercises.

Problem 2.1 Let T be an equilateral triangle with side length 1. Show that if five points are selected on the triangle (or its interior), there are two whose distance apart is at most $\frac{1}{2}$.

Solution: Find the midpoints of each of the three sides, join them to form four triangles, each equilateral with side length $\frac{1}{2}$. In order to get four 'boxes', be careful to assign which border points 'belong' to which of the four new triangles (this can be done in many ways). Five points into four 'boxes' implies that two must be in one box. Any two points in an equilateral triangle with side length $\frac{1}{2}$ are at distance at most $\frac{1}{2}$ apart.

[Similarly, one can ensure that for any five points in a square of side length 2, two must be at distance at most $\sqrt{2}$ (partition the square into four equal subsquares and apply the PHP).]

Problem 2.2 Karen has 28 days to prepare for a chess tournament, and she wishes to play at least one match a day, but no more than 40 during this time. Prove that there is a sequence of consecutive days that she plays exactly 15 matches.

Solution: Let t_i denote the number of matches she played up to and including the *i*-th day. Then $1 \leq t_1 < t_2 < \cdots < t_{28} \leq 40$, and so also $16 \leq t_1 + 15 < t_2 + 15 < \cdots < t_{28} + 15 \leq 55$. These 56 numbers can take on only 55 values, so two must be the same. These two can not be from the same list, so must be of the form $t_j = t_i + 15$ for some j > i. Then on days numbered $i + 1, i + 2, \ldots, j$, she played exactly 15 matches.

Before discussing the next problem, recall how to count subsets. If $S = \{1, 2, 3\}$ is a set, how many subsets does S have? The subsets are:

$$\{ \}, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\},$$

eight in all. Notice that $8 = 2^3$. In general, an *n*-element set has 2^n subsets.

Problem 2.3 Let $S = \{z_1, z_2, ..., z_6\}$ be a set of positive integers, each at most 14 and no two equal. Show that the sums of the elements in all non-empty subsets of S cannot all be different.

Solution: There are $2^6 - 1 = 63$ non-empty subsets of *S*. The sum of all elements in *S* is at most 9 + 10 + 11 + 12 + 13 + 14 = 69. There are 63 pigeons and maybe 69 holes, not enough pigeons to apply PHP. So, instead, look at only the subsets with at most five elements. Then the maximum sum is at most 10 + 11 + 12 + 13 + 14 = 60, and since only one set (the one with all six) has been eliminated, there are 62 possible subsets. Now apply PHP, and conclude that two subsets have the same sum. (See Lemma ?? for general case.)

2.1 Warm up exercises

Exercise 1 Show that among any 13 people, at least two have birthdays in the same month. What are the "holes" or "colours"?

Exercise 2 How many people do you need to have to guarantee that three have the same birthday? Can you generalize this to p people, not just p = 3?

Exercise 3 You have five different pairs socks, one pair each of red, blue, green, black, and white. It is dark outside and you haven't turned on any lights. If the socks are loose in the drawer, how many would you have to take out before you were guaranteed a pair?

Exercise 4 A list of 85 different positive integers is given. Prove that there are two that have the same remainder upon division by 84.

Exercise 5 No one has more than 400,000 hairs on their head. Show that there are two people in Winnipeg that have precisely the same numbers of hairs on their heads.

Exercise 6 How many times must one roll a single die to guarantee that the same score is obtained 10 times?

Exercise 7 You pack a lunch for mathletics practice, and you bring along some chocolate bars to share: 25 Snickers, 8 Coffee Crisps, 13 KitKats, and 30 Aero bars. It is late in the dorms, and your den mother tells you that if you share, all 20 students must get the same kind of bar. How many bars do you need to pull from your knapsack to guarantee that there will be 20 all the same?

Exercise 8 Show that among any 11 numbers chosen from 1–100, there are at least two whose square roots differ by less than 1.

Exercise 9 Write down any 12 two digit numbers. Show that there are two whose difference is of the form aa.

Exercise 10 Let T be an equilateral triangle with side length 1. Show that if ten points are selected on the triangle (or its interior), there are two whose distance apart is at most $\frac{1}{2}$.

Exercise 11 If a target is an equilateral triangle with side length 2, and the target is hit by 17 arrows, what is the minimum distance between arrows?

Exercise 12. Given 5 points on a sphere, show that there is a hemisphere containing at least 4.

Exercise 13 Show that in any collection of six different numbers from 1, 2, ..., 9, there must be two whose sum is 10.

2.2 More challenging problems

Exercise 14 Christine has 77 days to prepare for a golf tournament, and wants to play at least one round (of nine holes) each day, but no more than 132 rounds in total. Prove that there is a sequence of consecutive days that she plays exactly 21 games.

Exercise 15 Pick your favourite positive integer n, and write down a list of n integers, a_1, a_2, \ldots, a_n —they need not be all different. Show that there is always a subset of these numbers whose sum is divisible by n. [In fact, this sum can be chosen from consecutive numbers in your list!]

Exercise 16 You and a friend attend a garage sale, and on one table, there are ten items all priced at less than a dollar. Show that both you and your friend can each buy items (at least one) so that each of you spend exactly the same amount. Hint: look at how many different purchases one can make.

Exercise 17 If 33 rooks are placed on a chess board, show that there are always at least 5 that are mutually non-attacking.

Exercise 18 Let a_1, \ldots, a_n be n positive integers with $a_1 \le a_2 \le \cdots \le a_n \le 2n$. The least common multiple of any two of them is greater than 2n. Prove that $a_1 > 2n/3$.

Exercise 19 Ten points are placed on a circular disk of diameter 5. Show that there are two points with a distance apart of at most 2.

Exercise 20 The numbers from 1 to 81 are written on the squares of a 9×9 checkerboard. Show that there exist two neighbouring squares (adjacent in same row or same column) that have numbers differing by at least 6.

Exercise 21 You are at a math party at that there are 500 people (it can happen!). Show that there are always two people who have exactly the same number of acquaintances. [For the purpose of this question, if A is acquainted with B, then B is acquainted with A.] Does the same hold for a party of 20?

2.3 Solutions

1. Each of the 12 months is like a pigeonhole (or box).

2. There are 366 different possible birthdays (including February 29) so $2 \cdot 366 + 1 = 733$. In general, if there are $(p-1) \cdot 366 + 1$ people, p of them will have the same birthday.

3. Six. Each hole corresponds to a colour, and if you try to put six socks into five holes, two must go in the same hole.

4. Classify each positive integer according to what remainder r is obtained upon division by 84. Since the possible choices for r are $0, 1, 2, \ldots, 83$ (of which there are 84), by the PHP, two must be in the same remainder class.

5. There are more than 400,001 people in Winnipeg, and there are 400,001 choices (including zero) for numbers of hairs, so two numbers must agree.

6. One could possibly roll every score 9 times first, using $9 \times 6 = 54$ rolls. So 55 rolls will guarantee the same scores 10 times.

7. You could conceivably pull out 19 Snickers, 8 Coffee Crisps, 13 KitKats, and 19 Aero bars first. Then any one more will complete a set. So, in all, pulling out 60 will guarantee happiness.

8. Let (x, y] denote the set of numbers n satisfying $x < n \le y$. Consider the ten boxes containing 1, (1, 4], (4, 9], (9, 16], (16, 25], ..., (81, 100]. By the PHP, among any 11 numbers, two must be in the same box.

9. Group 10–99 into eleven categories:

$$S_{10} = \{10, 21, 32, 43, 54, 65, 76, 87, 98\}$$

$$S_{11} = \{11, 22, 33, 44, 55, 66, 77, 88, 99\}$$

$$S_{12} = \{12, 23, 34, 45, 56, 67, 78, 89\}$$

$$\vdots$$

$$S_{20} = \{20, 31, 42, 53, 64, 75, 86, 97\}.$$

Note that in each category, any two numbers differ by a multiple of 11. By the PHP, there must be some S_i that contains two numbers.

10. Divide the sides into thirds, and make nine equilateral triangles, each of side length $\frac{1}{3}$. In order to get nine 'boxes', be careful to assign which border points 'belong' to which of the nine new triangles (this can be done in many ways). Ten points into nine 'boxes' implies that two must be in one box. Any two points in an equilateral triangle with side length $\frac{1}{3}$ are at distance at most $\frac{1}{3}$ apart.

11. First divide the triangle into four triangles each of side length 1, then further subdivide each into four more, each of side length $\frac{1}{2}$, giving 16 small triangles in all. By the PHP, two arrows must have hit the same small triangle, and so can have distance at most $\frac{1}{2}$ apart.

12. Pick any two points and form the great circle C passing through them. There are three remaining points. If any one of these three is also on C, then either of the last two points will complete the set of four. If none of the three are on C, then by the PHP, two must lie on one side; these two together with the original two form the desired set of four.