Pigeon Hole Problem

Johan Vonk, Alex Tholen, Adi Mittal, Aditi Bonthu *

February 2017

1 Introduction

Theorem 1.1 (Pigeon Hole Principle)**.** *With n things and m boxes, you must have put at least $\lceil \frac{n}{m} \rceil$ in a box.*

Proof. Let us suppose that total "n" number of pigeons are to be put in "m" number of pigeonholes and $n > m$. Let us assume there is no pigeonhole with at least $\frac{n}{m}$ pigeons. In this case, every pigeonhole will have less than $\frac{n}{m}$ pigeons. Therefore the number of pigeons in each pigeon hole $< \frac{n}{m}$ and the total numbers of pigeons is smaller than the number of pigeonholes. There are m pigeonholes so the total number of pigeons is $< m \times \frac{n}{m}$ and therefore $< n$. This can't be because our definition is that n is the number of pigeons. □

1.1 Basic uses of Pigeonhole

Example 1.1. *Prove that in any group of six people there are either three mutual friends or three mutual strangers, assuming that the friendship is always reciprocated.*

Proof. Represent the people as vertices on a graph, and denote friendships with red edges and "stranger-ship" with blue edges. We will show that there exists a monochromatic triangle. Consider the relationship of P_1 to the 5 others. By the pigeonhole principle, 3 of the others must have the same relationship to P_1. Without loss of generality, say P_2, P_3, P_4 are connected to P_1 by red edges. Consider the edges between P_2, P_3, and P_4. If any of them are red, then we have a red triangle. If not, we have a blue triangle. □

Example 1.2. *Consider any five points $P_1, ...P_5$ in the interior of a square S of side length* 1*. Show that one can find two of the points at distance at most $\frac{\sqrt{2}}{2}$ apart.*

Proof. Consider partitioning the square into 4 sub-squares. By the pigeonhole principle, 2 of the points must be in one "sub-box." The distance between those

*Peter Ruhm, Simon Rubinstein-Salzedo, and the rest of Euler Circle two points must be less than the diameter of the sub-box: $\frac{\sqrt{2}}{2}$ (the length of the sub-box's diagonal). This proves the desired result. □

Example 1.3. *A chess-master has 77 days to prepare for a tournament. He wants to play at least one game per day, but not more than 132 games. Prove that there is a sequence of successive days on which he plays exactly 21 games.*

Proof. Define S_i $(1 \leq i \leq 77)$ as the total number games the chess-master plays from day 1 up to day i. Because she plays at least one game a day, $1 \leq S_1 < S_2 < \ldots < S_{77} \leq 132$ (i.e. the $S_i\,'s$ are distinct). Define $T_i = S_i + 21$. Note that the $T_i\,'s$ are all distinct. Now, out of the S_i's and T_i's, there are $77 \times 2 = 154$ numbers, but these numbers can take at most $132 + 21 = 153$ possible values. By the pigeonhole principle, two of the numbers are equal. This implies that for some i, j, $S_j = T_i = S_i + 21$. Hence, the chess-master plays exactly 21 games in the consecutive block from day $i + 1$ to day j. □

2 Number Theory

Theorem 2.1. *If a is irrational, there exists infinite $\frac{p}{q}$ such that*

$$\left| a - \frac{p}{q} \right| < \frac{1}{q^2}$$

Proof. Let M be any positive integer. Define $\{x\} = x - \lfloor x \rfloor$. We have the sequence $\{0\}, \{a\}, \{2a\}, ..., \{Ma\}$ that is in $[\frac{j-1}{M}, \frac{j}{M})$ for $M \leq j > 0$. So, there exists j and k such that $|aj - ak| < \frac{1}{M}$. Expanding, that is $a(j-k) - \lfloor ja \rfloor - \lfloor ka \rfloor ka = a(j-k) + INT$. So, we have $a(j-k) < \frac{1}{M}$. let $p = \lfloor j - k \rfloor$ and $q = j - k$. So, we have

$$\left| a - \frac{p}{q} \right| = |qa - p| = \left| \frac{j - ka - \lfloor j - k \rfloor}{q} \right| = \left| \frac{< \frac{1}{M}}{< M} \right| < \frac{1}{M^2} < \frac{1}{q^2}.$$

□

Theorem 2.2. *If there is a sequence of $mn + 1$ numbers*

$$a_1, a_2, a_3, a_4, \ldots, a_{mn+1}$$

then there exists either an increasing subsequence of length $m+1$ or a decreasing subsequense of size $n + 1$.

Proof. Let t_i be the longest increasing subsequence starting at a_i. If there exists a $t_i > m$, then we have an increasing subsequence of lenght $m + 1$. If not, $t_i \in [1, m]$. Then by pigeonhole there must be $n + 1$ t_{j_i}'s such that $t_{j_1} = t_{j_2} = \cdots = t_{j_{n+1}}$ for some t_js with $j_i < j_{i+1}$. Let them be equal to s. Let's look at the relationship between a_{j_1} and a_{j_2}. If $a_{j_2} > a_{j_1}$ then there is an increasing subsequence of length $s + 1$ satrting at a_{j_1}. That is impossible, so $a_{j_2} < a_{j_1}$. So, for any $i < k$ $a_{j_i} > a_{j_k}$. So, we have a decreasing subsequence of length $n + 1$ starting at a_{j_1}. Q.E.D. □

3 Counting and Probability

Example 3.1 (Turkeys)**.** *On thanksgiving, two or more of the consumed turkeys will have the same weight when rounded to the nearest millionth of a pound.*

Proof. Turkeys weigh roughly 15 pounds, with the largest on record at 37 pounds. If we could weigh all the turkeys to a millionth of a pound, then there are 37 million possible values. There are 46 million turkeys consumed on Thanksgiving. By the pigeonhole principle, two of those turkeys must have the same weight to the nearest millionth of a pound because $47 \times 10^6 > 37 \times 10^6$. We can be even more precise. There must be at least 10×10^6 turkeys with the same weight to a millionth of a pound. $\square$

Example 3.2. *Gary is training for a triathlon. Over a 30 day period, he pledges to train at least once per day, and 45 times in all. Then there will be a period of consecutive days where he trains exactly 14 times.*

Proof. Let S_i indicate the cumulative number of workouts by day i. Since each day contains one workout, and the total number of workouts is 45, we know that:

$$S_1 < S_2 < \cdots < S_{30} = 45$$

We want to prove there is some place with $i < j$ such that $S_i + 14 = S_j$. Start by adding 14 to every term in the inequality:

$$S_1 + 14 < S_2 + 14 < \cdots < S_{30} + 14 = 59$$

The two inequalities imply there are 60 numbers ($S_1, S_2, \ldots, S_{30}$ and $S_1 + 14, S_2 + 14, \ldots, S_{30} + 14$) that can assume any of the 59 integer values from 1 to 59. By the pigeonhole principle, two of the numbers must be the same. Which two? Notice that none of the numbers $S_1 < S_2 < \cdots < S_{30}$ could possibly be equal to one another (Rick takes at least one workout every day, so the sequence is strictly increasing). The same logic is true for the group $S_1 + 14, S_2 + 14, \ldots, S_{30} + 14$. Therefore, we must have one value from the group $S_1 < S_2 < \cdots < S_{30}$ equal to one of the values from the group $S_1 + 14, S_2 + 14, \ldots, S_{30} + 14$, which is exactly what we wanted to prove. $\square$

Example 3.3. *For any 5 points placed on a sphere, some hemisphere must contain 4 of the points?*

Proof. Consider the great circle through any two of the points. This partitions the sphere into two hemispheres. By the pigeonhole principle, 2 of the remaining 3 points must lie in one of the hemispheres. These two points, along with the original two points, lie in a closed semi-sphere. $\square$

Example 3.4 (Birthday Problem)**.** *What is the approximate probability that in a group of $n > 2$ people, at least 2 of them have the same birthday?*

Proof.

$$1 - p(n) = \overline{p}(n) = \prod_{k=1}^{n-1}(1 - \frac{k}{365})$$

because it is the product of the chances that it doesn't happen. What about the number of people needed for a fifty percent chance? □

Example 3.5. *What is the minimum number of people such that $\overline{p}(n) < \frac{1}{2}$?*

Proof.

$$\overline{p}(n) < \frac{1}{2}$$

so

$$\prod_{k=1}^{n-1}(1 - \frac{k}{365}) < \frac{1}{2}$$

because $1 - k < e^{-x}$

$$e^{-(n(n-1))/(2\times 365)} < \frac{1}{2}$$

solve for n to get

$$n^2 - n > 2 \times 365 \ln 2$$

$$730 \ln 2 \approx 505.997$$

so when $n = 23$, $n^2 - n = 506$ which is barely above 505.997 So to get more than a 50% chance for 23 people. For 22 people, $22^2 - 22 < 2 \times 365 \ln 2$, so it won't work. □

4 Real Life Applications

4.1 Computer Science

Example 4.1 (Lossless Compression Algorithm)**.** *No universally lossless compression algorithm exist.*

Proof. Let us call the input sequence L and the output O. Because this is a compression algorithm, the length of O (o) must be less than the length of L (l). The number of possible combinations for the input sequence is (2^l) and for the output sequence it is (2^o). By the pigeon hole principle, multiple inputs must go to one output. If this happens, then their is no way of getting one of the values back when you decompress so it can't be lossless. □

Example 4.2 (Collisions in Hash Tables)**.** *Collisions are inevitable in a finite hash table.*

Proof. The number of possible keys exceeds the number of indices in the array. The number of possible combinations for the input is (n^l) and for the output it is (n^o). Since $l < o$, $n^l < n^o$. By the pigeonhole principle there must be a collision because multiple values must go to the same hash number. □

4.2 Formal Language Theory

Lemma 4.1 (Pumping Lemma)**.** *For any regular language L, any sufficiently long word w which is in L can be split into three parts such as $w = xyz$, such that all the strings xy^kz for $k \geq 0$ are also in L.*

Proof. In Formal Language Theory, for every regular language there is a finite state automation (FSA) that accepts a word only if it is in the language. The number of states in the FSA is called the pumping length p. We take a word w with 1 more state than the pumping length p. Let $q_0 \cdots q_p$ be the sequence of the next p states during the FSA for a word length greater than the pumping length. Because the FSA only has p states the $p + 1$ visited states (0 to p) according to the pigeonhole principle it must include a repetition. Let say this state is q_s. The transitions that take the FSA from the first instance of q_s to the second match some string y. If we repeat this than it will still be part of the language. $\square$

4.3 Art Gallery Problem

Theorem 4.2 (Chvátal's art gallery theorem)**.** $\lfloor \frac{n}{3} \rfloor$ *guards are always sufficient and sometimes necessary to guard a simple polygon with no holes and n vertices.*

Proof. This proof is due to Steve Fisk. First, triangulate the polygon (without adding vertices). Then 3-color the graph. To show that is possible, we observe the dual graph which is an graph having a vertex on every triangle and an edge for every pair of adjacent triangles. This dual graph is a tree because any cycle would form the boundary of a hole in the polygon, contrary to our assumption of there being no holes. When there are multiple triangles, the dual graph must have a vertex with only one neighbor which corresponds to a triangle that is adjacent to the other triangles along only one side. The polygon formed by removing this triangle has a 3-coloring by induction and is easily extended to the additional vertex of the removed triangle. Once a 3-coloring is found, every triangle has three color. The vertices with any one color form a valid guard set. Since the three colors divide the n vertices of the polygon, the color with the fewest vertices forms a guard set with at more $\lfloor \frac{n}{3} \rfloor$ guards. $\square$