Chapter 12

The Pigeonhole Principle

12.1 Statements of the principle

In we put more than n pigeons into n pigeonholes, at least one pigeonhole will house two or more pigeons. This trivial observation is the basis of ingenious combinatorial arguments, and is the subject of this chapter. Let's begin with the various guises of the pigeonhole principle that are encountered in combinatorics.

Basic form. If m objects are put in n boxes and $n < m$, then at least one box contains at least two objects. The one-line proof is by contradiction: If every box contains at most one object, there are at most $n \cdot 1 = n$ objects. A more rigorous formulation of the principle is as follows: Given two sets A and B, with $|A| = m > n = |B|$, for any function $f : A \to B$ there exists $b \in B$ such that

$$\big|\{x \in A : f(x) = b\}\big| > 1.$$

General form. If m objects are put in n boxes, then at least one box contains at least $\lceil m/n \rceil$ objects. The proof is again by contradiction: If every box contains at most $\lceil m/n \rceil - 1 < m/n$ objects, there are less than $n(m/n) = m$ objects. The more rigorous formulation is: Given two sets A and B, for any function $f : A \to B$ there exists $b \in B$ such that

$$\big|\{x \in A : f(x) = b\}\big| \geq \left\lceil \frac{m}{n} \right\rceil.$$

Dijkstra's form. For a nonempty finite collection of integers (not necessarily distinct), the maximum value is at least the average value. It is a good exercise to verify that this is equivalent to the general form above.

12.2 Simple applications

Let's begin with some easy applications of the pigeonhole principle.

First application. There are two San Franciscans with the exact same number of hairs on their heads. Indeed, according to P&G Hair Facts, the average person's head has about 100,000 hairs, while "some people have as many as 150,000." So it seems safe to bet that every San Franciscan has at most 700,000 hairs on his or her head. On the other hand, the year 2000 US Census counted 776,733 San Francisco residents. The pigeonhole principle implies that at least two of them have the exact same number of hairs.

Second application. At a cocktail party with six or more people, there are three mutual acquaintances or three mutual strangers. Indeed, pick an arbitrary person a. By the pigeonhole principle, among the other five or more people, either there are three of a's acquaintances, or three people who are strangers to a. Let's say there are three that are a's acquaintances, the other case is analogous. If those three are mutual strangers we are done. Otherwise there are two among them, call them b and c, who know each other. Then a, b and c are mutual acquaintances and we are done.

Third application. Consider an infinite two-dimensional plane, every point of which is colored either red or blue; then there are two points one yard apart that are the same color. Indeed, take an arbitrary equilateral triangle with a side length of one yard. By the pigeonhole principle, two of its vertices have the same color.

Fourth application. Consider the numbers $1, 2, \ldots, 2n$, and take any $n + 1$ of them; then there are two numbers in this sample that are coprime. Indeed, consider the pairs $\{1, 2\}, \{3, 4\}, \ldots, \{2n - 1, 2n\}$. By the pigeonhole principle, both numbers from one of these pairs are in the sample. These numbers differ by 1 and are thus coprime. (This follows from the same argument as in Euclid's proof of the infinity of primes.)

12.3 Advanced applications

The following lemma comes from a classical 1935 paper by Paul Erdös and George Szekeres titled "A combinatorial problem in geometry":

Lemma 12.3.1. *In any ordered sequence of n^2+1 distinct real numbers $a_1, a_2, \ldots, a_{n^2+1}$, there is either a monotone increasing subsequence of length $n+1$ or a monotone decreasing subsequence of length $n+1$. Namely, there is a set of indices $1 \le i_1 < i_2 < \cdots < i_{n+1} \le n^2+1$, such that either $a_{i_1} > a_{i_2} > \cdots > a_{i_{n+1}}$ or $a_{i_1} < a_{i_2} < \cdots < a_{i_{n+1}}$.*

Proof. For $1 \le i \le n^2 + 1$, let η_i be the length of the longest monotone increasing subsequence that starts at a_i. If some $\eta_i > n$, we are done. Otherwise, by the pigeonhole principle, there exists $1 \le j \le n$, and some set $i_1 < i_2 < \cdots < i_m$ of size $m \ge \lceil (n^2 + 1)/n \rceil = n + 1$, such that $\eta_{i_1} = \eta_{i_2} = \cdots = \eta_{i_m} = j$. Now, consider two numbers a_{i_k} and $a_{i_{k+1}}$. If $a_{i_k} < a_{i_{k+1}}$, we get an increasing subsequence starting at a_{i_k} of length $j + 1$, which is a contradiction. Hence $a_{i_k} > a_{i_{k+1}}$ in particular, and

$a_{i_1} > a_{i_2} > \cdots > a_{i_m}$ in general, giving us a decreasing subsequence of length at least $n+1$. $\square$

Here is another pigeonhole gem, the last one for today:

Proposition 12.3.2. *Given a sequence of n not necessarily distinct integers $a_1, a_2, \ldots, a_n$, there is a nonempty consecutive subsequence $a_i, a_{i+1}, \ldots, a_j$ whose sum $\sum_{m=i}^{j} a_m$ is a multiple of n. (The subsequence might consist of a single element.)*

Proof. Consider the collection

$$\left(\sum_{i=1}^{0} a_i, \sum_{i=1}^{1} a_i, \sum_{i=1}^{2} a_i, \ldots, \sum_{i=1}^{n} a_i\right).$$

This collection has size $n+1$ and its first element is the empty sum $\sum_{i=1}^{0} a_i = 0$. There are only n possible remainders modulo n, thus by the pigeonhole principle, there are two numbers in the above collection of size $n+1$ that leave the same remainder. Let these be $\sum_{i=1}^{l} a_i$ and $\sum_{i=1}^{k} a_i$, with $l < k$. By a lemma we once proved, it follows that

$$n \Big| \left(\sum_{i=1}^{k} a_i - \sum_{i=1}^{l} a_i\right),$$

which implies

$$n \Big| \sum_{i=l+1}^{k} a_i.$$

$\square$