1 The Pigeonhole Principle

The pigeonhole principle is simply the assertion that if you have \( k \) boxes and \( k + 1 \) objects to be put in the boxes, than there will be a box with more than 1 object in it. The ‘pigeonhole’ name comes from applying the definition to the case where the boxes are pigeonholes, and the objects, pigeons.

A more mathematical formulation:

**Theorem 1.** Let \( f \) be a function \( f : A \rightarrow B \) with \( |A| = n + 1 \) and \( |B| = n \). Then \( f \) is not 1-1.

It is easy to prove this by induction on \( n \) (Do it!)

It should be clear that this formalization is equivalent to the intuitive description above.

There is an easy generalization:

**Theorem 2.** If we want to put \( O \) objects in \( B \) boxes, we’ll have a box with \( \geq \lceil \frac{O}{B} \rceil \) objects.

For example, if we want to put 7 objects in 3 boxes, we’ll have to put at least 3 in some box, since if we put 2 or less, we can take care of at most 6 objects.

There are some simple applications:

- In a group of 8 people, two will have their birthday on the same day of the week
• Among 27 English words, there are two that start with the same letter

• In a course where grades can be 0, 1, 2, \ldots 10, among 12 students, two will have the same grade

• If you grab 4 single socks from a drawer that has white, black, and red socks, you will have a pair with matching colors

• Consider standard playing cards (4 suits, 13 cards in each suit.) If you draw 9 cards, your are guaranteed to get 3 of the same suit (but if you want 3 of a particular suit, say hearts, you may need to draw 42 cards.)

We can answer some less obvious questions. For example, how many area codes are needed to accommodate 25 million users (in North America)?

Phone numbers obey the following scheme (NANP): a phone number is of the form

\[(a_1a_2a_3) \ o_1o_2o_3 − s_1s_2s_3s_4\]

where the \(a_1a_2a_3\) is the area code, \(o_1o_2o_3\) is the office code (a remainder of the time where neighborhoods had their own exchanges (google PEnnsylvania 6-5000 – earlier the scheme was 2 letters, each denoting a number – P was 7, E was 3), and the final 4 digits are the station code. There are restrictions on the digits allowed: the first and the fourth digit cannot be a 0 or a 1 in the \(o_1o_2o_3 − s_1s_2s_3s_4\) part. It follows from the previous lecture that the number of such phone numbers is \(8 \times 10^2 \times 8 \times 10^6 = 6.4 \times 10^9\) We need at least 4 different area codes to come up to 25 million distinct phone numbers. (This is Example 8 in Rosen)

Your parents may remember a time when all Chicago had area code 312. Now you know why there are 773 and 872 area codes! (Some people and companies were upset with the change.)

1.1 Some Clever Applications

Example 10 in Rosen: in a 30-day month a baseball team plays at least one game every day, and no more than 45 games in the period. Then there exists a set of consecutive days \(d_i, d_{i+1} \cdots d_{i+m}\), during which they played exactly 14 games.

\[Proof.\] Let \(a_i\) be the number of games the team played up to and including day \(i\): this yields a strictly increasing sequence \(a_1, \cdots a_{30} \subset \{1, 2, \cdots 45\}\). In
particular, this implies that the $a_j$ are distinct. Now consider the sequence $a_1 + 14, \ldots, a_{30} + 14$. This sequence has also distinct elements. The union of the two sets of numbers is a subset of $\{1, 2, \ldots, 59\}$, so there is a number that appears twice. Since each sequence is of distinct numbers, it must be that the pair belongs to different sequences. But then $\exists i, j : a_i = a_j + 14$, which means that from day $j + 1$ to day $i$ the team played exactly 14 games.

Read also example 11 and Theorem 3 in Rosen. They were covered in class, with proofs very similar to Rosen.

### 1.2 A taste of Ramsey Theory

**Theorem 3.** In a party with 6 people there are either

- a group of 3 people who have been introduced to each other, or
- a group of 3 people no two of whom were introduced to each other

*Note: Rosen talk of ‘friends’ and ‘enemies’ which needs clarification—in real life these relations are not necessarily symmetric.*

**Proof.** (see Rosen) is a relatively simple argument. Take a man, say A. Among the remaining 5, there must be a group of 3, say B, C, and D, who have either been all introduced to A, or none of whom were introduced (by Generalized Pigeonhole: 2 groups 5 people, one group must have 3 people.) WLOG, assume B, C, and D were introduced to A. If any two of them were introduced to each other, we are done. If not, we found a group of 3 no pair of whom were introduced to each other—and we are also done. \(\square\)

This is the first step in a subject called Ramsey Theory. In general we have the following

**Theorem 4.** \(\forall n, m > 0 \exists R(n, m)\) such that in any group of \(R(n, m)\) people, there is either

- a group of \(n\) people who have been introduced to each other, or
- a group of \(m\) people no two of whom were introduced to each other

We will not cover the proof.