Lecture 14 - The Pigeonhole Principle

6.042 - April 3, 2003

1 Injections

Recall the definition of an injective function.

Definition 1 *A function $f : A \to B$ is an* injection *if for all $a, a' \in A$, $(a \neq a') \Rightarrow (f(a) \neq f(a'))$.*

Informally, f is an injection if no two items in the range A are mapped to the same item in the domain B. Note that f need not take on every value in the range B; the only restriction is that f cannot take on the same value twice.

For example, let A be a set of balls, and let B be a set of boxes. Then an injection from A to B is a way of assigning balls to boxes so that no box gets multiple balls. If such an assignment exists, then intuitively there can be no more balls than boxes. We can state this as a theorem.

Theorem 1 *Let A and B be finite sets. If $f : A \to B$ is an injection, then $|A| \leq |B|$.*

Proof. Suppose that A is the n-element set $\{a_1, a_2, \ldots, a_n\}$. Since f maps A to B, the set $\{f(a_1), f(a_2), \ldots f(a_n)\}$ is a subset of B. Because f is an injection, all the $f(a_i)$ are different and so B also contains at least n elements. Therefore, $|A| \leq |B|$. $\square$

The condition in the theorem that sets A and B are finite is important. For example, suppose $A = \{0, 1, 2, 3, \ldots\}$ and $B = \{0, 2, 4, 6, \ldots\}$. The function $f : A \to B$ defined by $f(x) = 2x$ is an injection. This would seem to imply that A is no larger than B. However, in fact A contains all the natural numbers, but B contains only the even natural numbers! We will resolve this apparent paradox later in the course; for now we'll avoid the problem by sticking with finite sets.

2 The Pigeonhole Principle

The theorem about injections, Theorem 1, has an important corollary called the *pigeonhole principle*. The name derives from the following colorful statement of the result:

Theorem 2 (Pigeonhole Principle) *If $n + 1$ pigeons are put into n holes, then at least one hole contains two or more pigeons.*

Proof. Let A be a set of $n+1$ "pigeons", let B be a set of n "holes", and let the function $f : A \to B$ define an assignment of pigeons to holes. The contrapositive of Theorem 1 says that f can not an injection, because $|A| > |B|$. Therefore, f must assign at least two pigeons to the same hole. $\Box$

As a trivial application of the pigeonhole principle, suppose that there are three people in a room. The pigeonhole principle implies that there are at least two people with the same gender. In this case, the "pigeons" are the three people and the "pigeonholes" are the two possible genders, male and female. Since there are more pigeons than pigeonholes, two pigeons must be in the same pigeonhole; that is, two people must have the same gender.

More generally, if p pigeons are assigned to h holes and $p > h$, then some hole must get two or more pigeons. Many other useful variations on the pigeonhole principle are possible, as we shall see. The underlying intuition is always the same, however: if you stick lots of balls into few bins, some bin must end up with a lot of balls.

The pigeonhole principle seems too obvious to be really useful, but the next few examples show that it can gives short proofs of results that are difficult to obtain by other means.

3 Disjoint Subset Sums

List 10 different numbers between 1 and 99. Now, if you look carefully, you'll find two disjoint subsets of those numbers with the same sum. Those subsets might not be easy to find, but we can be certain that they exist!

Theorem 3 *For every set S of 10 distinct integers between 1 and 99, there are two disjoint, nonempty subsets of S whose elements sum to the same quantity.*

Proof. The number of nonempty subsets of the 10 numbers is $2^{10} - 1$. (We can either include or exclude each of the ten elements, which gives 2^{10} possibilities. But this counts the empty set, so we must subtract 1.) These subsets are the "pigeons".

The least possible sum of the elements in a nonempty subset is 1. The greatest possible sum is $90 + 91 + 92 + \ldots + 99 = 945$. Thus, the sum of the elements in every subset must lie between 1 and 945. These values are the "holes".

Therefore, there are 1023 pigeons and 945 holes. Now we assign each subset (a pigeon) to the sum of its elements (a hole). By the pigeonhole priciple, two pigeons must be assigned to the same hole; that is, there must exist two distinct, nonempty subsets A and B with the same sum.

There is one problem remaining: the sets A and B might not be disjoint. We can obtain disjoint sets A' and B' by deleting the shared elements from both A and B:

$$\begin{array}{rcl} A' & = & A - (A \cap B) \\ B' & = & B - (A \cap B) \end{array}$$

Since the sets A and B have the same sum and we've subtracted the same elements from each, sets A' and B' must also have the same sum.

Finally, we must check that throwing out shared elements did not empty out A' or B'. If, say, A' were empty, then that would imply that A was a subset of B. But this is impossible, since sets A and B have the same sum. $\square$

So what is the largest set of numbers in the range 1 to 99 with distinct subset sums? Replacing 99 by n gives an unsolved problem that is considered extremely difficult!

4 Marching Band

The players in a marching band are arranged in tidy rows and columns. Here are the heights of the players in the band, in inches:

$$\begin{bmatrix} 66 & 72 & 68 & 64 \\ 67 & 65 & 71 & 61 \\ 69 & 62 & 70 & 63 \\ 61 & 73 & 66 & 64 \end{bmatrix}$$

Suppose that, within each row, band members are sorted from left to right by height:

$$\begin{bmatrix} 64 & 66 & 68 & 72 \\ 61 & 65 & 67 & 71 \\ 62 & 63 & 69 & 70 \\ 61 & 64 & 66 & 73 \end{bmatrix}$$

Now suppose that, in addition, we sort the players within each column.

$$\begin{bmatrix} 64 & 66 & 69 & 73 \\ 62 & 65 & 68 & 72 \\ 61 & 64 & 67 & 71 \\ 61 & 63 & 66 & 70 \end{bmatrix}$$

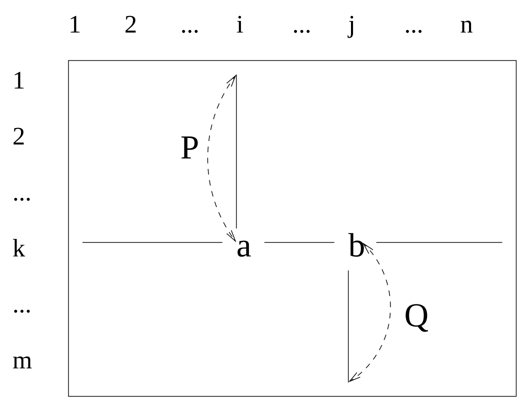
This last operation rearranged the band players within every column. But, surprisingly, the players in each row are still sorted by height! This is no coincidence, as we can prove using the pigeonhole principle.

Theorem 4 *Given a matrix of numbers A with m rows and n columns, perform the following steps:*

1. *Rearrange the numbers in each of the m rows of A so that they are in nondecreasing order from left to right. Call the resulting matrix A'.*
2. *Rearrange the numbers in each of the n columns of A' so that they are in nondecreasing order from bottom to top. Call the resulting matrix A''.*

Then the numbers in each of the m rows of A'' are still in nondecreasing order from left to right.

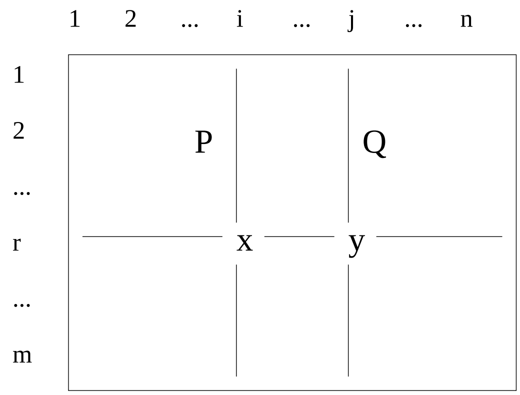
Proof. The proof is by contradiction. Suppose that the rows of the final matrix A'' are not all sorted. Then A'' must contain a row in which a bigger number a is placed to the left of a smaller number b as shown below.



As suggested in the diagram, let k be the index of the row containing a and b. Let i be the index of a's column, and let j be the index of b's column. Furthermore, define P to be the sequence of numbers in column i in rows 1 through k, and let Q be the sequence of numbers in column j in rows k through m. Note that the sequences P and Q contain a total of $m+1$ numbers: two from row k and one from every other row of the matrix.

Because the columns of A'' are freshly sorted, all numbers in P are greater than or equal to a and all numbers in Q are less than or equal to b. Since $a > b$ by assumption, every number in P is strictly greater than every number in Q.

Now consider the earlier matrix A', obtained from the original matrix by sorting the rows. In this matrix, the elements of P appear in some order in column i, and the elements of Q appear in some order in column j. Regard the $m+1$ elements of P and Q as pigeons, and regard the m rows of the matrix as holes. By the pigeonhole principle, some two pigeons must be assigned to the same hole; that is, two numbers from the sequences P and Q must both appear in some row of the matrix A'. One of these numbers must be from P and one must be from Q, since the elements of P are in distinct rows and the elements of Q are also in distinct rows. Specifically, suppose that $x \in P$ and $y \in Q$ both appear in row r as shown below:



We know that $x > y$, since all elements of P are greater than all elements of Q. But this implies that row r is not sorted, contradicting the definition of matrix A'. Therefore, our initial assumption that the rows of the final matrix are not all sorted is false, and the theorem is proved. □

5 Strong Pigeonhole Principle

Here is one possible generalization of the pigeonhole principle.

Theorem 5 (Strong Pigeonhole Principle) *If p pigeons are placed in h holes, then some hole contains at least $\lceil p/h \rceil$ pigeons.*

We'll prove this theorem somewhat less formally than before.

Proof. Suppose that the claim is false; that is, every hole contains fewer than $\lceil p/h \rceil$ pigeons. Then, since the number of pigeons in each hole is an integer, every hole actually contains fewer than p/h pigeons. Thus, altogether, the holes contain fewer than $h \cdot (p/h) = p$ pigeons. But this is a contradiction, because p pigeons were placed in the holes. Therefore, the claim must be true. □

With the strong pigeonhole principle in hand, you can swagger around parties shouting outragous claims such as: **There are 1000 Americans who have *exactly* the same number of hairs on their heads! Even discounting bald people!**

Let's prove this claim using the strong pigeonhole principle. There are at least 200,000,000 Americans who are not bald. Regard them as "pigeons". Every non-bald person has between 1 and 200,000 hairs on his or her head. Regard these numbers as "holes". Assign each pigeon (American) to a pigeonhole (number between 1 and 200,000) according to the number of hairs on his or her head. By the strong pigeonhole principle, at least $\lceil 200,000,000/200,000 \rceil =$ 1000 pigeons end up in the same hole; that is, at least 1000 Americans have exactly the same number of hairs on their heads, excluding bald people.

6 Ramsey Theory

Ramsey theory is a notoriously difficult branch of mathematics. On an intuitive level, all the theorems in this area say roughly the same thing: every sufficiently large mess must contain highly ordered components. For example, suppose that there is a group of people, where each pair are either friends or enemies. (This is the mess.) Ramsey's theorem says that if the group is sufficiently large, then there must exist either five people who are all friends of one another or else five people who are all enemies. (These are the highly ordered components.) Precisely how large the group must be is a tricky question. For now, let's solve an easier problem in Ramsey theory using the pigeonhole principle.

Theorem 6 *In every group of six people, where each pair are either friends or enemies, there are either three mutual friends or three mutual enemies.*

Proof. Let A be one of the six people. Break the remaining five people into two sets, F_A, comprising the friends of A, and E_A, comprising the enemies of A.

The five people are the "pigeons". The two sets are the "holes". Invoking the Strong Pigeonhole Principle, we see that either F_A or E_A has at least $\lceil 5/2 \rceil = 3$ elements. That is, $|F_A| \geq 3$ or $|E_A| \geq 3$.

Case 1: Assume $|F_A| \geq 3$. Without loss of generality, assume that $\{B, C, D\} \subseteq F_A$. If any two of B, C and D are friends, then these two and A form a set of three mutual friends. Otherwise, B, C and D form a set of three mutual enemies.

Case 2: Assume $|E_A| \geq 3$. Without loss of generality, assume that $\{B, C, D\} \subseteq E_A$. If any two of B, C and D are enemies, then these two and A form a set of three mutual enemies. Otherwise, B, C and D form a set of three mutual friends.

This completes the proof. □

An analogous claim for a group of five people does not hold; if friendships and hatreds between five people are arranged just right, then there are neither three mutual friends nor three mutual enemies.

What is the smallest group for which there must be either *four* mutual friends or *four* mutual enemies? The answer turns out to be 18. Amazingly, the answer to the same question for five friends or five enemies is unknown. And specialists in Ramsey theory consider finding the answer for six friends or enemies to be completely hopeless!