MAT 307: Combinatorics

Lecture 2: The pigeonhole principle and double counting

Instructor: Jacob Fox

1 The pigeonhole principle

If n objects are placed in k boxes, $k < n$, then at least one box contains more than one object.

This is so obvious, one might think that nothing non-trivial can be derived from this "principle". And yet, this principle is very useful.

1.1 Two equal degrees

The following consequence is quite easy.

Theorem 1. *In any graph, there are two vertices of equal degree.*

Proof. For any graph on n vertices, the degrees are between 0 and $n-1$. Therefore, the only way all degrees could be different is that there is exactly one vertex of each possible degree. In particular, there is a vertex v of degree 0 and a vertex w of degree $n-1$. However, if there is an edge (v, w), then v cannot have degree 0, and if there is no edge (v, w) then w cannot have degree $n-1$. This is a contradiction. □

1.2 Subsets without divisors

Let $[2n] = \{1, 2, \ldots, 2n\}$. Suppose you want to pick a subset $S \subset [2n]$ so that no number in S divides another. How many numbers can you pick? Obviously, you can take $S = \{n+1, n+2, \ldots, 2n\}$ and no number divides another. Can you pick more then n numbers? The answer is negative.

Theorem 2. *For any subset $S \subset [2n]$ of size $|S| > n$, there are two numbers $a, b \in S$ such that $a|b$.*

Proof. For each odd number $a \in [2n]$, let $C_a = \{2^k a : k \geq 0, 2^k a \leq 2n\}$. The number of these classes is n and every element $b \in [2n]$ belongs to exactly one of them, for a obtained by dividing b by the highest possible power of 2. Consider $S \subset [2n]$ of size $|S| > n$. By the pigeonhole principle, there is a class C_a that contains at least two elements of S. □

1.3 Rational approximation

Theorem 3. *For any $x \in R$ and $n > 0$, there is a rational number p/q, $1 \leq q \leq n$, such that*

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

Note that it is easy to get an approximation whose error is at most $\frac{1}{n}$, by fixing the denominator to be $q = n$. The improved approximation uses the pigeonhole principle and is due to Dirichlet (1879).

Proof. Let $\{x\}$ denote the fractional part of x. Consider $\{ax\}$ for $a = 1, 2, \ldots, n+1$ and place these $n+1$ numbers into n buckets $[0, 1/n)$, $[1/n, 2/n)$, ..., $[(n-1)/n, 1)$. There must be a bucket containing at least two numbers $\{ax\} \leq \{a'x\}$. We set $q = a' - a$ and we get $\{qx\} = \{a'x - ax\} < 1/n$. This means that $qx = p + \epsilon$ where p is an integer and $\epsilon = \{qx\} < 1/n$. Hence,

$$x = \frac{p}{q} + \frac{\epsilon}{q}.$$

□

1.4 Monotone subsequences

Finally, we give an application which is less immediate. Given an arbitrary sequence of distinct real numbers, what is the largest monotone subsequence that we can always find? It is easy to construct sequences of mn numbers such that any increasing subsequence has length at most m and any decreasing subsequence has length at most n. We show that this is an extremal example.

Theorem 4. *For any sequence of $mn+1$ distinct real numbers $a_0, a_1, \ldots, a_{mn}$, there is an increasing subsequence of length $m+1$ or a decreasing subsequence of length $n+1$.*

Proof. Let t_i denote the maximum length of an increasing subsequence starting with a_i. If $t_i > m$ for some i, we are done. So assume $t_i \in \{1, 2, \ldots, m\}$ for all i; i.e. we have $mn+1$ numbers in m buckets. By the pigeonhole principle, there must be a value $s \in \{1, 2, \ldots, m\}$ such that $t_i = s$ for at least $n+1$ indices, $i_0 < i_1 < \ldots < i_n$. Now we claim that $a_{i_0} > a_{i_1} > \ldots > a_{i_n}$. Indeed, if there were a pair such that $a_{i_j} < a_{i_{j+1}}$, we could extend the increasing subsequence starting at $a_{i_{j+1}}$ by adding a_{i_j}, to get an increasing subsequence of length $s+1$. However, this contradicts $t_{i_j} = s$. □

2 Double counting

Another elementary trick which often brings surprising results is *double counting*. As the name suggests, the trick involves counting a certain quantity in two different ways and comparing the results.

2.1 Sum of degrees in a graph

The following observation is due to Leonard Euler (1736).

Lemma 1. *For any graph G, the sum of degrees over all vertices is even.*

Proof. For a vertex v and edge e, let $i(v, e) = 1$ if $v \in e$ and 0 otherwise. We count all the incidences between vertices and edges in two ways:

- $\sum_{v \in V, e \in E} i(v, e) = \sum_{v \in V} \sum_{e \in E} i(v, e) = \sum_{v \in V} d(v)$,
  because $d(v)$ is exactly the number of edges incident with v.
- $\sum_{v \in V, e \in E} i(v, e) = \sum_{e \in E} \sum_{v \in V} i(v, e) = 2|E|$,
  because every edge is incident with exactly two vertices.

Thus we have proved that the sum of all degrees is exactly twice the number of edges. □

2.2 Average number of divisors

Let $t(n)$ denote the number of divisors of n. E.g., for a prime n, $t(n) = 2$, while for a power of 2, $t(2^k) = k + 1$. We would like to know what is the *average number of divisors*,

$$\bar{t}(n) = \frac{1}{n} \sum_{j=1}^{n} t(j).$$

This seems to be a complicated question; however, double counting gives a simple answer. Let $d(i, j) = 1$ if $i|j$ and 0 otherwise. I.e., $t(j) = \sum_{i=1}^{n} d(i, j)$. We count the total number of dividing pairs $i|j$, $\sum_{i,j=1}^{n} d(i, j)$, in two different ways.

- $\sum_{i,j=1}^{n} d(i, j) = \sum_{j=1}^{n} t(j) = n \cdot \bar{t}(n)$.
- $\sum_{i,j=1}^{n} d(i, j) \simeq \sum_{i=1}^{n} \frac{n}{i} = n \cdot H_n$, where H_n is the n-th harmonic number.

In the second case, we have been somewhat sloppy and neglected some roundoff errors, but these add up to at most n overall. We can conclude that $t(n) \simeq H_n \simeq \ln n$, within an error of 1.