Lecture 6: The Pigeonhole Principle and Probability Spaces

Anup Rao

January 17, 2018

We discuss the pigeonhole principle and probability spaces.

Pigeonhole Principle

The pigeonhole principle is an extremely simple yet powerful tool to prove combinatorial facts. It says that if you try to put $n+1$ pigeons in n holes, then some hole must get at least 2 pigeons.

Example: Intersecting Families of Sets

How large can a family of subsets of $[n]$ be, if every two sets in the family must intersect?

If we want to find lots of subsets of $[n]$ that all intersect each other, one option is to pick all the sets that contains 1. This gives 2^{n-1} different sets, and all of them intersect each other. Another option is to pick all the sets of size bigger than $n/2$. If n is odd, this also gives 2^{n-1} sets that intersect each other. It turns out there are no larger families of sets that all intersect each other!

To prove this, we use the pigeonhole principle. We will define 2^{n-1} holes, where each hole consists of a pair of sets where every set A is paired with its complement A^c. There are 2^n options for A, and each pair contains exactly two sets, so this gives $2^n/2 = 2^{n-1}$ pairs. Now suppose we have m sets that all intersect each other. These m sets are our pigeons. Put each pigeon in the hole that contains it. If $m > 2^{n-1}$, there are more pigeons than holes, so there will be a hole that contains 2 pigeons. But this cannot happen, because then there will be a pigeon A and a pigeon A^c, which means our family of sets do not all intersect each other, since $A \cap A^c = \emptyset$.

Example: The Erdös-Szekeres Theorem

What is the length of the longest increasing or decreasing subsequence in a given sequence of n distinct numbers?

We did not discuss this in class, by I include it here because it is cool.

If the sequence of n distinct numbers is $1, 2, 3, \ldots, n$, then there is an increasing subsequence of length n, but the longest decreasing subsequence has length 1. If the sequence is $n, n-1, \ldots, 1$, then there is a decreasing subsequence of length n, but the longest increasing subsequence has length at most 1.

If $n = m^2$, then consider the sequence

$$x = m, m-1, \ldots, 1, 2m, 2m-1, \ldots, m+1, 3m, 3m-1, \ldots, 2m+1, \ldots .$$

The longest increasing subsequence of x has length at most $m = \sqrt{n}$, and the longest decreasing sequence of x has length at most $m = \sqrt{n}$.

The Erdös-Szekeres theorem proves that this is the worst case if you want to minimize the length of both the longest increasing subsequence and the longest increasing subsequence.

It is not really important that the numbers be distinct. If they are not distinct, we still get something about the length of the longest non-increasing subsequence and the length of the longest non-decreasing subsequence.

Theorem 1. *If $n = m^2 + 1$, then every sequence of numbers either has an increasing subsequence of length $m+1$ or a decreasing subsequence of length $m+1$.*

Proof. We use the pigeonhole principle. Our holes will be the points of $[m] \times [m]$. Our pigeons will be the elements of $[n]$. Say we are given a sequence $x = x_1, \ldots, x_n$. For every $i \in [n]$, let a_i be the length of the longest increasing subsequence that ends at x_i, and let b_i be the length of the longest decreasing sequence that begins at x_i. Put the pigeon i in the hole (a_i, b_i).

Now, if there is no increasing subsequence in x of length bigger than m, and no decreasing subsequence in x of length bigger than m, then for every i, $a_i \leq m$ and $b_i \leq m$. So there are m^2 holes, and $n = m^2 + 1$ pigeons that are each placed in those holes. By the pigeonhole principle, two pigeons must end up in the same hole. Suppose $(a_j, b_j) = (a_i, b_i)$, for some $i < j$.

There are two cases. If $x_i < x_j$, then the longest increasing sequence ending at x_i has length a_i. But this means that there is an increasing sequence of length $a_i + 1$ ending at x_j—just add x_j to the increasing sequence that ends at x_i. So, we cannot have $a_i = a_j$.

If $x_i > x_j$, then the longest decreasing sequence ending at x_i has length b_i. But this means that there is a decreasing sequence of length $b_i + 1$ ending at x_j—just add x_j to the decreasing sequence that ends at x_i. So, we cannot have $b_i = b_j$.

In either case, we get a contradiction, so it must be that x contains either an increasing sequence of length $m+1$ or a decreasing subsequence of length $m+1$. □

Example: Dirichlet's Theorem

How accurately can we approximate a real number by a rational number?

We will not have time to discuss Dirichlet's theorem in class, and you will not be tested on it. I include it here because it is another cool application of the pigeonhole principle.

The real numbers are all the numbers on the real line. The rational numbers are numbers that can be expressed as p/q where p, q are integers and $q \neq 0$. There are certainly real numbers, like $\sqrt{2}$, that are not rational. If x is a real number, then for every q, we can certainly

find p so that $|x - p/q| < 1/(2q)$, since all of the multiples of $1/q$ are only $1/q$ apart. But can we do better? Is it possible to approximate x by a rational number with denominator q with a smaller error than $1/q$?

For example, consider the number $x = 0.55555555\ldots$. This is a real number. But if you try to approximate it with a rational number where $q = 1000$ then the best approximation is 0.556 upto k decimal places. The difference $|x - 0.556|$ is $0.00044444\ldots$ which, after a little bit of calculation, turns out to be the same as $\frac{4}{9000} = \frac{4}{9} \cdot \frac{1}{q}$. No matter how high the power of 10 you pick for q, you would end up with an error of $\frac{4}{9} \cdot \frac{1}{q}$.

Dirchlet proved that a much smaller error is possible to all numbers, as long as you choose the right denominator q! Moreover, the proof is a simple use of the pigeonhole principle:

Theorem 2. *Let x be a positive real number. Then for every positive integer n, there is a rational number p/q such that $1 \leq q \leq n$ and*

$$|x - p/q| < \frac{1}{qn} \leq \frac{1}{q^2}.$$

Proof. To use the pigeonhole principle, we need to identify $n + 1$ pigeons and n holes. Consider the n holes

$$\left[0, \frac{1}{n}\right), \left[\frac{1}{n}, \frac{2}{n}\right), \ldots, \left[\frac{n-1}{n}, 1\right).$$

Recall that $[u, v)$ is the interval containing all numbers x with $u \leq x < v$.

The $n + 1$ pigeons will correspond to the numbers

$$x, 2x, \ldots, (n+1)x.$$

For each of the numbers $i \cdot x$ in this list, let $i \cdot x = y_i + \epsilon_i$, where y_i is an integer, and $\epsilon_i \in [0, 1]$ is the fractional part of $i \cdot x$. Put the pigeon $i \cdot x$ in the hole that contains ϵ_i.

By the pigeonhole principle, we must have i, j such that $i \cdot x, j \cdot x$ are in the same holes. Assume without loss of generality that $i > j$. Then

$$(i - j) \cdot x = (y_i - y_j) + (\epsilon_i - \epsilon_j),$$

so

$$|(i - j)x - (y_i - y_j)| < 1/n.$$

Dividing through by $(i - j)$, we get

$$|x - (y_i - y_j)/(i - j)| < 1/(n(i - j)).$$

Set $p = y_i - y_j$ and $q = i - j$ to complete the proof. □

Probability

A PROBABILITY SPACE GIVES a very useful way to generalize the idea of counting the size of sets. A *probability space* is defined by a set Ω, usually called the *domain* or *sample space*, and a *distribution*. A distribution is a function $p : \Omega \to \mathbb{R}$ mapping the elements of Ω to real numbers so that

There are many commonly used notations for distributions. Sometimes people write Pr or *Prob* instead of p.

- For all $x \in \Omega$, $p(x) \geq 0$.
- $\sum_{x \in \Omega} p(x) = 1$.

A probability space captures the concept of a random process happening. The elements x encode all the possible outcomes of the process, and $p(x)$ represents the chance that x is the outcome.

How would you encode a single coin toss as a porbability space?

An *event* in the probability space is a subset $E \subseteq \Omega$. The probability of the event is $p(E) = \sum_{x \in E} p(x)$.

For example, suppose we toss a fair coin twice. Then the sample space is $\Omega = \{HH, TT, HT, TH\}$. The distribution puts equal weight on all outcomes, so we have $p(x) = 1/4$ for every $x \in \Omega$. If we consider the subset $E \subseteq \Omega$ where the first coin toss is heads, then it is of size 2 so $p(E) = 2/4 = 1/2$.

A very common situation is when the distribution is uniform over the sample space, meaning that $p(x) = p(y)$ for all $x, y \in \Omega$. In this case, the probability of an event E is exactly $p(E) = \frac{|E|}{|\Omega|}$—it is just the ratio of the size of E to the size of Ω. However, the nice thing that probabilities make sense even when the sets Ω and E are of infinite size.

Example

Suppose we toss a fair coin n times. What is the probability that the number of heads is even?

Here the sample space consists of all 2^n possible coin tosses. If E denotes the event that the number of heads is even, then we have

$$|E| = \binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \dots$$

So, the probability that the number of heads is even is

$$\frac{|E|}{|\Omega|} = \frac{\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \dots}{2^n}.$$

This looks like a complicated expression, but we can simplify it using facts that we have proved about binomial coefficients. First, we know that $2^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots$, and we also know that all the even

See Lecture 4 from January 10.

coefficients sum to exactly the same value as all the odd coefficients sum to. So, we have

$$\frac{|E|}{|\Omega|} = \frac{\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \dots}{\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots} = \frac{\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \dots}{2(\binom{n}{0} + \binom{n}{2} + \dots)} = \frac{1}{2}.$$

You can use a very similar argument to prove that the probability that the number of heads is odd is $1/2$.

Example

Suppose you take a uniformly random walk in a grid, starting from the point $(0, 0)$ and ending at the point (n, n), and always moving either up or right, what is the probability that you cross the diagonal?

In previous lectures, we calculated the total number of such walks to be $|\Omega| = \binom{2n}{n}$. We also calculated the number that cross the diagonal as $|E| = \frac{n}{n+1} \cdot \binom{2n}{n}$. So, the probability of crossing the diagonal is $\frac{\frac{n}{n+1} \cdot \binom{2n}{n}}{\binom{2n}{n}} = \frac{n}{n+1}$.

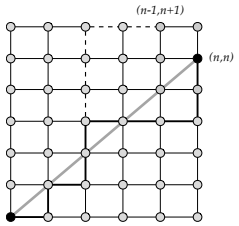

Figure 1: We discussed walking on the grid in Lecture 3 on January 8.