Lecture 3: Miscellaneous Techniques

Rajat Mittal*

IIT Kanpur

In this document, we will take a look at few diverse techniques used in combinatorics, exemplifying the fact that combinatorics is a collection of different counting techniques. Specifically, we will look at inclusion-exclusion, pigeonhole principle and application of linear algebra in combinatorics.

1 Inclusion exclusion

First technique is called inclusion-exclusion. You might have used it in some small cases in high-school. We will generalize the principle and show some of its applications.

Let us take an example first. Suppose, we have 100 students in an institute using various social-media websites. Say 44 use Facebook, 50 use Twitter and 56 use Whatsapp.

It is also known that how many of them use multiple websites. 27 use Facebook and Twitter, 31 are on Facebook and Whatsapp, also 34 use Whatsapp and Twitter. There are 19 who use all three websites.

How many students are there who do not use any website?

This kind of question can be visualized easily using a *Venn-diagram*.

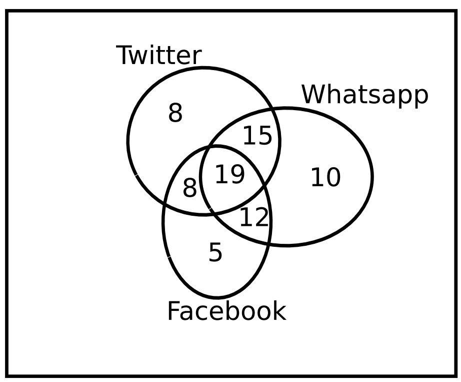

Fig. 1. Website Users

To draw the Venn-diagram, the idea is to start from the inner-most part of the Venn-diagram, and gradually deduce the numbers in the outer-parts. It is given that the innermost part contains 19 people. Now, going outwards, we need number of people who use Facebook and Twitter but not all three websites. From the data given, that number should be $27 - 19 = 8$. Once we have determined all such *exclusive* intersection of two sets, the number of users for Twitter not using any other website is,

$$50 - (8 + 19 + 15) = 8.$$

We can similarly find number of students exclusively using Facebook and exclusively using Whatsapp. Taking the next step, it is quite clear that 23 $(= 100 - (56 + 8 + 8 + 5))$ students do not use any social media website.

But, if there were twenty websites (or a lot of websites) then this calculation strategy is quite tedious!

Let us try another strategy. Notice that the number of students not using any website is: the total number of students minus the number of students using at least one website. We try to count the number of students using at least one website in a different way.

* Thanks to Nitin Saxena for his notes from the previous iteration of the course.

Basically, we move from the outer-part towards the inner-parts (of an undrawn Venn-diagram). Our first guess would be to sum up students using any website $(44 + 50 + 56 = 150)$.

Exercise 1. This is wrong, because there are only 100 students. What is the problem?

Clearly, this counts a student using two websites twice, so let us subtract the students who are in the intersection of two websites. Hence, the next guess would be $150 - (27 + 31 + 34) = 58$.

Again, students using all three websites were counted 3 times in the beginning, then subtracted thrice, were not counted at all. Adding them, our true count is $58 + 19 = 77$.

So, there are $100 - 77 = 23$ students who do not use any of the three websites.

Let us generalize this argument for more than three sets. Suppose there is a universe U with subsets $A_1, A_2, \cdots, A_n$. For our previous example, U will be the set of all students ($|U| = 100$) and A_1 will be the set of students using Facebook and so on.

In the problem, we are given their intersections $A_I := \cap_{i \in I} A_i$, for every subset $I \subseteq [n]$ (verify this). The question is to find the number of elements of U not present in any set A_i.

According to the example given above: it seemed that we first subtracted the cardinality of A_I with $|I| = 1$, then added $|A_I|$ for I's of size 2 and so on. Hence, the number of elements not in any of the sets A_i should be given by,

$$\left| U - \bigcup_{i \in [n]} A_i \right| = \sum_{I \subseteq [n]} (-1)^{|I|} \cdot |A_I| \ .$$

Note 1. $A_\emptyset$ is the universe U itself. Why?

The guess given above is correct and is called the *principle of inclusion and exclusion.*

Exercise 2. Why is it called inclusion-exclusion?

Let us formally prove the principle now.

Proof of the principle. We will show the equality by finding the count of an element u of U in RHS. There are two cases,

– either u is not contained in any A_i (then we should count it exactly once),
– or u is contained in at least one A_i (then it should not be counted).

If an element u is not contained in any of the sets A_I then it will be counted exactly once (namely, by the term $|A_\emptyset|$).

So, we only need to show that every *other* element is counted 0 times (overall in the RHS). Let J_u be the set of indices j such that u is contained in A_j.

$$J_u = \{j : u \in A_j\}.$$

Notice, J_u is the maximal nonempty subset of $[n]$. That means, u is contained in A_j if and only if j is in J_u.

u will be counted in A_I iff I is contained in J_u. Hence, the number of times u gets counted in RHS is,

$$c_u := \sum_{I \subseteq J_u} (-1)^{|I|} \cdot 1 \, .$$

Suppose $|J_u| = k$. We will be done if we can show that:

Exercise 3. $c_u = \sum_{i=0}^{k} (-1)^i \binom{k}{i} = 0$.

Consider the binomial expansion of $(1 - 1)^k$.

This proves that every element of U is counted exactly once if it is not in any of the A_i's, and not counted otherwise. This proves the inclusion-exclusion principle. $\square$

Let us take a standard application of inclusion exclusion known as *Derangements*.

Example 1 (Derangements). Suppose we have n distinct letters and n distinct envelopes with one envelope marked for one particular letter. Think of the letters being numbered from 1 to n and every envelope having a number (again between 1 and n written over it).

In how many ways could you place letter in envelopes (one letter goes to exactly one envelope), s.t., no letter goes to the correct envelope?

We want to solve it using inclusion exclusion.

Exercise 4. What is the universe U and what is its size?

The universe should be the set of all permutations of letters. So, there are $n!$ ways to put letters into envelopes and that is our universe U.

Exercise 5. What is A_i now?

Denote A_i to be the set of ways when letter i goes to its *correct* envelope. Convince yourself, we are interested in $\left|U - \cup_{i\in[n]} A_i\right|$, number of ways when no letter goes to its correct location.

First, to apply the inclusion-exclusion formula, we need to calculate A_I. That is a simple combinatorial exercise. After placing $|I|$ letters in the correct position, we have $(n-|I|)!$ ways to place remaining letters. From inclusion exclusion, the number of derangements is

$$\sum_{I\subseteq[n]} (-1)^{|I|}(n-|I|)!.$$

There are $\binom{n}{i}$ subsets of size i giving the same value. Hence, the number of derangements is,

$$\sum_{i=0}^{n}(-1)^i\binom{n}{i}(n-i)! = n!\sum_{i=0}^{n}\frac{(-1)^i}{i!}.$$

2 Pigeonhole principle

This is one of the simplest principles (to state and prove) in mathematics which has numerous applications (difficult to prove otherwise).

Theorem 1. *If there are $n+1$ pigeons and n pigeonholes then at least one pigeonhole will have more than 1 pigeon.*

Proof. Let the pigeon distribution be $a_1 \le \cdots \le a_n$. We have $a_n \ge (\sum_i a_i)/n = \frac{n+1}{n} > 1$. $\square$

This seemingly obvious theorem has many nice applications. Let us look at a few of them.

Example 2. To start with, it implies that, in a group of 367 friends there are at least two whose birthdays coincide. Here, friends are the 'pigeons' and birth-dates are the 'pigeonholes'.

Example 3. Let us say that there are n users of Facebook. Show that there exist at least two people who have same number of friends.

Like in some previous proofs, we can assume that $n \ge 3$. For small cases you can easily check that the theorem holds.

Since there are n users, the number of friends a user can have will only range from $\{0, 1, 2, \cdots, n-1\}$.

First try: there are n people (pigeons) and n possible number of friends (pigeonholes), so we cannot apply the pigeonhole principle directly here.

Second try: There are two cases,

– there is a person who is friends with everybody (has $n-1$ friends),
– there is no such person (no user has $n-1$ friends).

For the first case, if someone is friends with $n-1$ people then everyone has at least one friend. That means, 0 cannot appear in the possible number of friends. So there are n pigeons (users) and $n-1$ pigeonholes (number of friends), thus, at least two people will have the same number of friends.

For the second case, if no one has $n-1$ friends, then again there are n pigeons (users) and $n-1$ pigeonholes (number of friends).

Exercise 6. Suppose there is an island in the shape of an equilateral triangle with side 2 km. Is it possible to assign spaces for five houses on the island, such that, no two houses are within a distance of 1 km ?

Divide the triangle into four equal regions by considering the triangle formed by the midpoints of the sides.

Let us consider the question we asked in the first class.

Example 4. Show that given any integer n there exists a number, containing only 0 and 1 in its decimal expansion, divisible by n.

The difficulty lies in figuring out what are pigeons and what are pigeonholes. Suppose, we consider all possible stings of $0, 1$ as pigeons. What are the pigeonholes? They should be the remainders possible when divided by n.

The remainder can only be $\{0, 1, \cdots, n-1\}$. We can take any $n+1$ numbers with 0 and 1 in their decimal expansion, there are infinite number of them. By pigeonhole principle, at least 2 of them will have the same remainder when divided by n. So, their difference will be divisible by n.

Exercise 7. What is the problem with this proof?

This only implies that there are two $0, 1$ strings whose difference is divisible by n. But the difference need not be a $0, 1$ strings.

The question is, can we have $n+1$ numbers whose differences only contain 0 and 1? Consider the numbers $1, 11, 111, \cdots$, we can pick $n+1$ such numbers. If we subtract the smaller from the bigger one, we will get a $0, 1$ string.

Note 2. The string will actually be very special. It will be all 1's followed by all 0's.

Exercise 8. Can you extend the argument to show that there are infinitely many strings of $0, 1$ divisible by n.

Let us take a final example before we generalize the pigeonhole principle.

Example 5. Show that for any irrational number x and any integer n, you can find a rational approximation $\frac{p}{q}$, s.t.,

– $1 \leq q \leq n$.
– $\left|x - \frac{p}{q}\right| \leq \frac{1}{nq}$.

To simplify, first multiply both sides by q,

$$|qx - p| \leq \frac{1}{n}.$$

In other words, we need a q in $\{1, 2, \cdots, n\}$, such that, the difference between qx and an integer is smaller than $\frac{1}{n}$.

Look at the fractional part of qx for q between 1 and n and put them into pigeonholes of kind $(\frac{i}{n}, \frac{i+1}{n})$ for different i. If any fractional part falls between $(0, \frac{1}{n})$, we are already done. If not, there should be two fractional parts falling in the same pigeonhole by pigeonhole principle.

Exercise 9. Show that you are done if two fractional parts fall in same pigeonhole.

Take their difference.

The pigeonhole principle can be extended slightly, with a very similar proof (show it).

Theorem 2. *If there are $rn+1$ pigeons and n pigeonholes then at least one pigeonhole will have more than r pigeons.*

Let us look at an application of this generalization.

Example 6. Given 6 vertices of a hexagon, join all pairs of vertices by a line. Color every line segment between two vertices by either red or blue. Show that there is at least one monochromatic triangle formed (all lines of the triangle have the same color).

Choose a vertex v. There are 5 edges going from it. Since there are 2 colors, at least three edges are of the same color by the new pigeonhole principle. Suppose these 3 edges are of red color (you can pick blue, it doesn't matter) and they go to vertices v_1, v_2, v_3.

We know that there are 3 edges between v_1, v_2, v_3. If all of them are blue then we have a monochromatic blue triangle (v_1, v_2, v_3). Otherwise, at least one pair is colored red, say v_1, v_2. In this case, the triangle formed by v_1, v_2 and the original vertex v is all red. Hence, there always exists a monochromatic triangle.

Exercise 10. Color the edges of a pentagon, with every vertex connected to every other vertex, such that there is no monochromatic triangle.

3 Combinatorics through linear algebra

For the last technique, we will give an example where linear algebra comes to the rescue of combinatorics. This example and its exposition is taken from the book by Babai and Frankl with the title *Linear algebra methods in combinatorics*. The entire book is about problems in combinatorics which can be solved using linear algebra.

Suppose you are a student in an institute. To make a club, you need to get permission from the head of your institute.

Every club gets funding from the university. So, your motivation is to make as many clubs as possible. On the other hand, the head wants to minimize the number of clubs.

The head really likes even numbers and puts the following conditions.

- All clubs should have even number of members.
- All pairwise intersection of clubs should be even.

Note 3. We assume that no two clubs can have exactly the same members, otherwise you can have infinite clubs.

Assume that there are n students.

Exercise 11. Show that you can make at least $2^{\lfloor n/2 \rfloor}$ clubs.

Put people in pairs and keep them together in every club possible.

This is a pretty big number! Since head's strategy is not very successful, one of the mathematicians suggests to the head that *even* should be replaced by *odd* in the first rule.

Exercise 12. What if you replace even by odd in both the rules?

Very much against the wishes of the head, rules are changed to,

- All clubs should have odd number of members.
- All pairwise intersection of clubs should be even.

Let us call this rule *size-odd-intersection-even.* Will this reduce the number of possible clubs?

Exercise 13. Remember that the number of clubs were exponential in n previously. Can you give a better upper/lower bound with this *size-odd-intersection-even* rule?

Surprisingly, linear algebra gives a linear bound on number of clubs. Now, we will show that the you can form at most n clubs under the *size-odd-intersection-even* rule.

The idea is to associate a vector in $\mathbb{Q}^n$ for every club (a vector space of dimension n). It will be shown that these vectors are linearly independent. Hence, there can be at most n clubs.

Exercise 14. Try to find such vectors.

The vectors will be in dimension n, where every co-ordinate corresponds to a student. The vectors are simply the indicator vectors, i.e., the i-th entry is 1 if i-th student is in the club, otherwise it is 0. For example, if only students 1 and 2 are present in a club, the associated vector is,

$$v_{\{1,2\}} = (1100 \cdots 0)$$

Suppose we have m clubs, there are m vectors associated with them. Call them $v_1, v_2, \cdots, v_m$.

Exercise 15. What is the inner product of v_i and v_j?

The inner product $v_i^T v_j$ is the number of members present in both clubs i and j. The *size-odd-intersection-even* rule translates to conditions,

$$v_i^T v_j = \begin{cases} \text{even}, & \text{if } i \neq j \\ \text{odd}, & \text{if } i = j. \end{cases}$$

The final step is to show that these vectors are linearly independent. If not, there exist $\lambda_1, \lambda_2, \cdots, \lambda_m$, s.t.,

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m = 0.$$

Assume that coefficients for only vectors $i_1, i_2, \cdots, i'_m$ are non-zero. Then,

$$\lambda_{i_1} v_{i_1} + \lambda_{i_2} v_{i_2} + \cdots + \lambda_{i'_m} v_{i'_m} = 0.$$

Exercise 16. Show that, without loss of generality, we can assume λ_{i_j} are integers and not all even.

Again, we can assume that λ_{i_1} is odd. Taking inner product with v_{i_1}, we get,

$$odd + even + even + \cdots + even = 0.$$

This is a contradiction. Hence, the vectors $v_1, v_2, \cdots, v_m$ are linearly independent. Since they reside in an n-dimensional space, $m < n$.

This proves that there can be at most n clubs.

References

1. K. H. Rosen. Discrete Mathematics and Its Applications. *McGraw-Hill*, 1999.
2. N. L. Biggs. Discrete Mathematics. *Oxford University Press*, 2003.