Double Counting

Victoria Krakovna
vkrakovna@gmail.com

1 Introduction

In many combinatorics problems, it is useful to count a quantity in two ways. Let's start with a simple example.

Example 1. (Iran 2010 #2) There are n points in the plane such that no three of them are collinear. Prove that the number of triangles, whose vertices are chosen from these n points and whose area is 1, is not greater than $\frac{2}{3}(n^2 - n)$.

Solution. Let the number of such triangles be k. For each edge between two points in the set we count the number of triangles it is part of. Let the total number over all edges be T.

On the one hand, for any edge AB, there are at most 4 points such that the triangles they form with A and B have the same area. This is because those points have to be the same distance from line AB, and no three of them are collinear. Thus, $T \leq 4\binom{n}{2}$. On the other hand, each triangle has 3 edges, so $T \geq 3k$. Thus,

$$k \leq \frac{T}{3} \leq \frac{4}{3}\binom{n}{2} = \frac{2}{3}(n^2 - n).$$

□

It's a good idea to consider double counting if the problem involves a pairing like students and committees, or an array of numbers; it's also often useful in graph theory problems.

2 Some tips for setting up the double counting

1. Look for a natural counting. (In the above example, the most apparent things we can sum over are points, edges and triangles. In this case edges are better than points because it is easier to bound the number of triangles that an edge is involved in. Thus, we sum over edges and triangles.)

2. Consider counting ordered pairs or triples of things. (For instance, in Example 1 we counted pairs of the form (edge, triangle).)

3. If there is a desired unknown quantity in the problem, try to find two ways to count some other quantity, where one count involves the unknown and the other does not. (This is done almost trivially in Example 1, where the expression that involves the unknown k is $T = 3k$.)

To illustrate, here are some more examples:

Example 2. (IMO 1987 #1) Let $p_n(k)$ be the number of permutations of the set $\{1, 2, 3, \ldots, n\}$ which have exactly k fixed points. Prove that

$$\sum_{k=0}^{n} kp_n(k) = n!.$$

Solution. The first idea that might occur here would be to find $p_n(k)$, then multiply it by k, sum it up... probably resulting in a big expression. However, if we look at the required result, we see that it suggests a natural counting - the left hand side is the total number of fixed points over all permutations. Another way to obtain that is to consider that each element of $\{1, 2, 3, \ldots, n\}$ is a fixed point in $(n-1)!$ permutations, so the total is $n(n-1)! = n!$. (Note that we are counting pairs of the form (point, permutation) such that the point is a fixed point of the permutation.) □

Example 3. (China Hong Kong MO 2007) In a school there are 2007 girls and 2007 boys. Each student joins no more than 100 clubs in the school. It is known that any two students of opposite genders have joined at least one common club. Show that there is a club with at least 11 boys and 11 girls.

Solution. Since we are only given information about club membership of students of different gender, this suggests that we should consider triples of the form (boy, girl, club), where the boy and girl both attend the club. Let the total number of such triples be T.

For each pair (boy, girl), we know that they attend at least one club together, so since there are 2007^2 such pairs,

$$T \geq 2007^2 \cdot 1.$$

Assume that there is no club with at least 11 boys and 11 girls. Let X be the number of triples involving clubs with at most 10 boys, and Y be the number of triples involving clubs with at most 10 girls. Since any student is in at most 100 clubs, the number of (girl, club) pairs is at most $2007 \cdot 100$, so $X \leq 2007 \cdot 100 \cdot 10$. Similarly, $Y \leq 2007 \cdot 100 \cdot 10$. Then,

$$2007^2 \leq T \leq X + Y \leq 2 \cdot 2007 \cdot 1000 = 2007 \cdot 2000$$

which is a contradiction. □

Example 4. (IMO SL 2003) Let $x_1, x_2, ..., x_n$ and $y_1, y_2, ..., y_n$ be real numbers. Let $A = \{a_{ij}\}$ (with $1 \leq i, j \leq n$) be an $n \times n$ matrix with entries

$$a_{ij} = \begin{cases} 1 & \text{if} \quad x_i + y_j \geq 0; \\ 0 & \text{if} \quad x_i + y_j < 0. \end{cases}$$

Let B be an $n \times n$ matrix with entries 0 or 1 such that the sum of the elements of each row and each column of B equals to the corresponding sum for the matrix A. Show that $A = B$.

Solution. Unlike in the previous problems, here it is not at all obvious what quantity we should count in two ways. We want it to involve the x_i and y_i, as well as the a_{ij} and b_{ij}. It makes sense to consider something that is symmetric in both of the above pairs of variables, and equals to zero if $A = B$. Let

$$S = \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i + y_j)(a_{ij} - b_{ij}).$$

On the one hand,

$$S = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i(a_{ij} - b_{ij}) + \sum_{i=1}^{n} \sum_{j=1}^{n} y_j(a_{ij} - b_{ij}) = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} (a_{ij} - b_{ij}) + \sum_{j=1}^{n} y_j \sum_{i=1}^{n} (a_{ij} - b_{ij}) = 0,$$

since the column and row sums in A and B are the same.

On the other hand, using the definition of a_{ij} we have the following:

- when $x_i + y_j \geq 0$, $a_{ij} - b_{ij} = 1 - b_{ij} \geq 0$;
- when $x_i + y_j < 0$, $a_{ij} - b_{ij} = -b_{ij} \leq 0$;

so $(x_i + y_j)(a_{ij} - b_{ij}) \geq 0 \ \forall i, j$. Since $S = 0$, it follows that $(x_i + y_j)(a_{ij} - b_{ij}) = 0 \ \forall i, j$. From this it is easy to derive that $a_{ij} = b_{ij} \ \forall i, j$. □

3 Cool proofs using double counting

The following is a proof of Cayley's formula, which counts the number of trees on n distinct vertices. There are several other proofs of this fact (using bijection, linear algebra, and recursion), but the double counting proof is considered the most beautiful of them all. [1]

Theorem 1. ***(Cayley's Formula)*** *The number of different unrooted trees that can be formed from a set of n* distinct *vertices is $T_n = n^{n-2}$.*

Proof. We count the number S_n of sequences of $n - 1$ directed edges that form a tree on the n distinct vertices.

Firstly, such a sequence can be obtained by taking a tree on the n vertices, choosing one of its nodes as the root, and taking some permutation of its edges. Since a particular sequence can only be obtained from one unrooted tree, the number of sequences is

$$S_n = T_n \cdot n(n-1)! = T_n n!.$$

Secondly, we can start with the empty graph on n vertices, and add in $n - 1$ directed edges one by one. After k edges have been added, the graph consists of $n - k$ rooted trees (an isolated vertex is considered a tree). A new edge can go from any vertex to a root of any of the trees (except the tree this vertex belongs to). This is necessary and sufficient to preserve the tree structure. The number of choices for the new edge is thus $n(n - k - 1)$, and thus the number of choices for the whole sequence is

$$S_n = \prod_{k=1}^{n-1} n(n-k-1) = n^{n-1}(n-1)! = n^{n-2} n!$$

The desired conclusion follows. □

[1] The four proofs are given in Aigner and Ziegler's book "Proofs from THE BOOK"

A more unexpected use of double counting is the following proof of Fermat's Little Theorem.

Theorem 2. ***(Fermat's Little Theorem)*** *If a is an integer and p is a prime, then*

$$a^p \equiv a \pmod{p}.$$

Proof. Consider the set of strings of length p using an alphabet with a different symbols. Note that these strings can be separated into equivalence classes, where two strings are equivalent if they are rotations of each other. Here is an example of such an equivalence class (called a"necklace") for $p = 5$: $\{BBCCC, BCCCB, CCCBB, CCBBC, CBBCC\}$.

Let's call a string with at least two distinct symbols in it *non-trivial*. All the rotations of a non-trivial string are distinct - since p is prime, a string cannot consist of several identical substrings of size greater than 1. Thus, all the equivalence classes have size p, except for those formed from a trivial string, which have size 1.

Then, there are two ways to count the number of non-trivial strings. Since there are a^p strings in total, a of which are trivial, we have $a^p - a$ non-trivial strings. Also, the number of non-trivial strings is p times the number of equivalence classes formed by non-trivial strings. Therefore, p divides $a^p - a$. □

4 Problems

Roughly in order of difficulty.

1. Prove the following identity:

$$\binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \cdots + n\binom{n}{n} = n2^{n-1}$$

2. (Iran 2010 #6) A school has n students, and each student can take any number of classes. Every class has at least two students in it. We know that if two different classes have at least two common students, then the number of students in these two classes is different. Prove that the number of classes is not greater that $(n-1)^2$.

3. (IMO SL 2004) There are 10001 students at a university. Some students join together to form several clubs (a student may belong to different clubs). Some clubs join together to form several societies (a club may belong to different societies). There are a total of k societies. Suppose that the following conditions hold:

   (a) Each pair of students is in exactly one club.

   (b) For each student and each society, the student is in exactly one club of the society.

   (c) Each club has an odd number of students. In addition, a club with $2m + 1$ students (m is a positive integer) is in exactly m societies.

   Find all possible values of k.

4. (IMO 1998 #2) In a competition there are m contestants and n judges, where $n \geq 3$ is an odd integer. Each judge rates each contestant as either "pass" or "fail". Suppose k is a number such that for any two judges their ratings coincide for at most k contestants. Prove that

$$\frac{k}{m} \geq \frac{n-1}{2n}$$

5. (MOP practice test 2007) In a $n \times n$ array, each of the numbers in $\{1, 2, \ldots, n\}$ appears exactly n times. Show that there is a row or a column in the array with at least $\sqrt{n}$ distinct numbers.

6. (USAMO 1995 #5) In a group of n people, some pairs of people are friends and the other pairs are enemies. There are k friendly pairs in total, and it is given that no three people are all friends with each other. Prove that there exists a person whose set of enemies has at most $k(1 - \frac{4k}{n^2})$ friendly pairs in it.

7. Consider an undirected graph with n vertices that has no cycles of length 4. Show that the number of edges is at most $\frac{n}{4}(1 + \sqrt{4n-3})$.

8. (IMO 1989 #3) Let n and k be positive integers, and let S be a set of n points in the plane such that no three points of S are collinear, and for any point P of S there are at least k points of S equidistant from P. Prove that $k < \frac{1}{2} + \sqrt{2n}$.

5 References

1. Mathematical Excalibur (Volume 13, Number 4), "Double Counting".

   `http://www.math.ust.hk/excalibur/v13_n4.pdf`

2. M. Aigner, G. Ziegler, "Proofs from THE BOOK".

3. Wikipedia page on Double Counting.

   `http://en.wikipedia.org/wiki/Double_counting_(proof_technique)`

4. Yufei Zhao, "Double Counting using Incidence Matrices".

   `http://web.mit.edu/yufeiz/www/doublecounting_mop.pdf`

5. MathLinks forum posts.

   `http://www.artofproblemsolving.com/Forum/index.php`