Double Counting

Thomas Guo

January 4, 2022

1 Introduction

A common technique in combinatorics is to calculate a specified quantity, usually the size of a set, in different ways. Typically, we then set the expressions equal and proceed from there. This is similar to a bijection, which proves that the sizes of two different sets are equal to one expression.

A classic example of double counting can be seen in the computation of the area of a staircase made up of 1 by 1 unit cells. In this case, the two different ways of calculating the area would be to sum by rows, or sum by columns.

Algebraically, for a sequence of positive integers $a_1 \leq a_2 \leq \cdots \leq a_n$, we have

$$a_1 + a_2 + \cdots + a_n = b_1 + b_2 + \cdots + b_{a_n},$$

where b_i denotes the number of terms in the sequence $a_1, a_2, \ldots, a_n$ that are greater than or equal to i (hence why the right sum ends at index a_n, as all later terms are zero). In this case, the staircase we are summing has n columns with heights $a_1, a_2, \ldots, a_n$ respectively, and b_i is simply the length of the ith row.

The condition that the a_i's be sorted is only required to match the visual representation of a staircase, and is not important for the algebra. Thus, we can generalize our notion of a staircase to become an m by n incidence matrix filled with 0s and 1s, and the two ways of counting are again by row and by column.

There are many ways to formulate a double counting argument, but the use of incidence matrices is a systematic approach that works on many kinds of double counting problems, such as ones involving sums of positive integer sequences, sets and elements, and bipartite graphs. While incidence matrices often help you structure and visualize the problem configuration, when it comes to solution writing, set theory notation (such as counting number of pairs or triples) is often cleaner and thus preferred.

2 Examples

1. Show that for an integer $n \geq 1$, $v_p(n!) = \lfloor \frac{n}{p} \rfloor + \lfloor \frac{n}{p^2} \rfloor + ... + \lfloor \frac{n}{p^k} \rfloor$, where $k = \lfloor \log_p n \rfloor$. Here, $v_p(m)$ denotes the highest exponent of a prime p that divides m, so $p^{v_p(m)} | m$ but $p^{v_p(m)+1} \nmid m$.

   We will construct a $k \times n$ incidence matrix where entry $(i, j) = 1$ if $p^i \mid j$ and entry $(i, j) = 0$ otherwise, for $1 \leq i \leq k$ and $1 \leq j \leq n$. Now, consider the sum of all entries in column j. This corresponds to the number of powers of p which divide j, which is precisely equal to $v_p(j)$ (as we must have $v_p(j) \leq \lfloor \log_p j \rfloor \leq k$). Thus, the sum of all entries in our incidence array is equal to $\sum_{j=1}^{n} v_p(j) = v_p(\prod_{j=1}^{n} j) = v_p(n!)$. On the other hand, consider the sum of all entries in row i. This corresponds to the number of multiples of p^i from 1 to n, which is precisely equal to $\lfloor \frac{n}{p^i} \rfloor$. Therefore, the sum of all entries in our incidence array is also equal to $\lfloor \frac{n}{p} \rfloor + \lfloor \frac{n}{p^2} \rfloor + ... + \lfloor \frac{n}{p^k} \rfloor$, proving the desired equality.

2. Show that
$$\sum_{k=0}^{n} \binom{n}{k}\binom{k}{d} = 2^{n-d}\binom{n}{d}.$$

   Sometimes, the quantity to be counted is more combinatorial. Suppose we have n distinguishable people. We will count the number of ways to pick a subset of these n people, with exactly d people (in the subset) designated as the executive committee. One way to compute this quantity is by first picking the executive committee in $\binom{n}{d}$ ways, and then deciding whether the remaining $n - d$ people are in the subset in 2^{n-d} ways, yielding the right hand side. Another way to compute this quantity is by summing over all possible subset sizes k, and for each k we have $\binom{n}{k}$ possible subsets and $\binom{k}{d}$ ways to pick the executive committee from the subset, yielding the left hand side.

3. (Russia 1996) In the Duma there are 1600 delegates, who have formed 16000 committees of 80 people each. Prove that one can find two committees having no fewer than four common members.

   We will construct a $1600 \times \binom{16000}{2}$ incidence matrix, where each column corresponds to an unordered pair of distinct committees (j, k). Now, we set the entry $(i, (j, k))$ to equal 1 if delegate i is in both committees j and k, and 0 otherwise. Consider the sum of the entries in row i. If delegate i is part of c_i committees, then the sum in row i is equal to $\binom{c_i}{2}$. Also, a simple double counting argument tells us that $\sum_{i=1}^{1600} c_i$ is equal to the total size of all of the committees, which equals $80 \cdot 16000$. Hence, the average of all of the c_i's is equal to 800. Now, either by applying Jensen's inequality or completing the square, we see that $\sum_{i=1}^{1600} c_i$ is minimized when all of the c_i's are equal to 800, so $\sum_{i=1}^{1600} \binom{c_i}{2} \geq 1600 \cdot \binom{800}{2}$. Therefore, the total sum of all entries in our incidence array is at least $1600 \cdot \binom{800}{2}$, and since we have $\binom{16000}{2}$ columns, by Pigeonhole, some column must have at least $\lceil \frac{1600 \cdot \binom{800}{2}}{\binom{16000}{2}} \rceil = 4$ entries equal to 1. This corresponds to a pair of committees (j, k) that have no fewer than four common members.

4. In an n x n matrix where $n = m^2$ for some positive integer m, each of the numbers $1, 2, \ldots, n$ appear exactly n times. Show that there is a row or a column in the matrix with at least m distinct numbers.

   We will construct an $n \times 2n$ incidence matrix, where each of the $2n$ columns corresponds to a row or column of the given $n \times n$ matrix. Now, entry (i, j) is equal to 1 if element i appears in row/column j, and is equal to 0 otherwise. We wish to show that some column j in our incidence matrix sums to at least m. Thus, it suffices to show that the total sum of all entries in our incidence array is at least $2mn$. Now, we claim that the sum of the entries in each row i is at least $2m$. Suppose that element i appears in a rows and b columns. Then, there are at most ab cells that contain i, so $n \le ab$. Thus, by AM-GM, we have $a + b \ge 2\sqrt{ab} = 2m$, so element i appears in at least $2m$ rows and columns. Hence, the sum of the entries in each row i is at least $2m$, so the total sum of all entries in our incidence array is at least $2mn$, and therefore there is a row or column in the $n \times n$ matrix with at least m distinct numbers.

5. (Iran 2010) 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 $\frac{2}{3}(n^2 - n)$.

   We will count the number of pairs (i, j), where triangle i (of area 1) contains edge j. Clearly, the total number of pairs is three times the number of triangles. Also, for each edge j, suppose its two endpoints are A and B. Then, the locus of points C for which the area of $\triangle ABC$ is equal to 1 is two lines parallel to AB, one on each side of it (as the distance from C to line AB must be exactly $\frac{2}{|AB|}$). By assumption, each of these two lines contains at most two of the n given points as no three given points are collinear. Thus, there are at most 4 such points C among the n given points, so there are at most 4 pairs (i, j) for each edge j. Hence, in total, we have at most $4 \cdot \binom{n}{2} = 2n(n-1)$ pairs, so the number of triangles of area 1 is at most $\frac{2}{3}(n^2 - n)$.

6. On the first day, there are a hundred students who are divided into five groups. On the second day, the same one hundred students are divided into four groups. Prove that there exists a student who belongs to a larger group on the second day than the first.

   Let g_i be the size of student i's group. If we directly try calculating the sum of g_i, double counting shows that it is equal to $\sum_j G_j^2$, where the G_j's are the sizes of the groups. However, on the second day, there is no clear relationship to the new sum $\sum_j G_j^2$, as this value can increase or decrease depending on the distribution of students. Instead, we need to be more creative. Since weighing each student with 1 gives a sum of $\sum_j G_j = 100$ and weighing each student with g_i gives a sum of $\sum_j G_j^2$, we may seek to eliminate G_j entirely by weighing each student with $\frac{1}{g_i}$. Indeed, we have

$$\sum_{i=1}^{100} \frac{1}{g_i} = \sum_j 1.$$

   The right sum is equal to the number of groups, which changes from five on the first day to four on the second day. Thus, the left sum decreases from the first day to the second day, so some term $\frac{1}{g_i}$ must also decrease, which means that some student i is in a larger group on the second day than the first.

3 Tips

1. Start by identifying a (somewhat easily) computable quantity.
2. In general, we want to view the problem structure in two different ways.
3. While we may often switch the order of summation, sometimes the second way of counting involves a greater restructuring of the problem, such as viewing it person by person, vertex by vertex, value by value, element by element, etc.
4. It often helps to see how the sum of the quantities calculated locally compares to the quantity calculated globally.
5. Sometimes, it may be difficult to express the exact value of a quantity, but approximations and bounds can prove just as useful!
6. Specifically when dealing with sets and elements, we can use the fact that the sum of the sizes of the sets is equal to the sum of c_i, where c_i is the number of sets that element i appears in.

4 Problems

1. (IMC 2002) Two hundred students participated in a mathematical contest. They had 6 problems to solve. It is known that each problem was correctly solved by at least 120 participants. Prove that there must be two participants such that every problem was solved by at least one of these two students.
2. (IMO 2001) Let n be an odd integer greater than 1 and let $c_1, c_2, \ldots, c_n$ be integers. For each permutation $a = (a_1, a_2, \ldots, a_n)$ of $\{1, 2, \ldots, n\}$, define $S(a) = \sum_{i=1}^{n} c_i a_i$. Prove that there exist permutations $a \neq b$ of $\{1, 2, \ldots, n\}$ such that $n!$ is a divisor of $S(a) - S(b)$.
3. (IMO 1987) 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
$$\sum_{k=0}^{n} k p_n(k) = n!$$
4. Let $S_1, S_2, \ldots, S_m$ be distinct subsets of $\{1, 2, \ldots, n\}$ such that $|S_i \cap S_j| = 1$ for all $i \neq j$. Prove that $m \leq n$.
5. (China) 16 students took part in a competition. All problems were multiple choice questions with four choices. Students were required to choose one and only one of the choices for each question. It was said that any two students had at most one answer in common. Find the maximum number of problems.
6. (APMO 1989) Let S be a set consisting of m pairs (a, b) of positive integers with the property that $1 \leq a < b \leq n$. Show that there are at least
$$4m \cdot \frac{(m - \frac{n^2}{4})}{3n}$$
triples (a, b, c) such that (a, b), (a, c), and (b, c) belong to S.

7. (ISL 2004) There are 10001 students at an 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:

   i.) Each pair of students are in exactly one club.

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

   iii.) 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.

8. (IMO 1998) In a contest, there are m candidates and n judges, where $n \geq 3$ is an odd integer. Each candidate is evaluated by each judge as either pass or fail. Suppose that each pair of judges agrees on at most k candidates. Prove that

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

9. (IMO 1989) Let n and k be positive integers and let S be a set of n points in the plane such that

   i.) no three points of S are collinear, and

   ii.) for every point P of S there are at least k points of S equidistant from P.

   Prove that:

$$k < \frac{1}{2} + \sqrt{2 \cdot n}$$

10. (ISL 2010) $n \geq 4$ players participated in a tennis tournament. Any two players have played exactly one game, and there was no tie game. We call a company of four players *bad* if one player was defeated by the other three players, and each of these three players won a game and lost another game among themselves. Suppose that there is no bad company in this tournament. Let w_i and l_i be respectively the number of wins and losses of the i-th player. Prove that

$$\sum_{i=1}^{n} (w_i - l_i)^3 \geq 0.$$

11. (IMO 2001) Twenty-one girls and twenty-one boys took part in a mathematical competition. It turned out that each contestant solved at most six problems, and for each pair of a girl and a boy, there was at least one problem that was solved by both the girl and the boy. Show that there is a problem that was solved by at least three girls and at least three boys.

12. (IMO 2021) Let $m \geq 2$ be an integer, A a finite set of integers (not necessarily positive) and $B_1, B_2, ..., B_m$ subsets of A. Suppose that, for every $k = 1, 2, ..., m$, the sum of the elements of B_k is m^k. Prove that A contains at least $\frac{m}{2}$ elements.