Olympiad Combinatorics

Pranav A. Sriram

August 2014

Copyright notices

About the Author

Pranav Sriram graduated from high school at The International School Bangalore, India, and will be a Freshman at Stanford University this Fall.

6. Counting in Two Ways

Introduction

Several combinatorics problems ask us to count something – for example, the number of permutations of the numbers from 1 to n without fixed points, or the number of binary strings of length n with more 1s than 0s. What's interesting is that the techniques used to solve counting or enumeration problems can be applied to problems that *don't* ask us to count anything. Problems in fields such as combinatorial geometry, graph theory, extremal set theory and even number theory can be solved by clever applications of counting – *twice*.

The basic idea underlying this chapter is to compute or estimate some quantity Q (which will depend on the problem and information given to us) by counting in two different ways. We hence obtain two different expressions or bounds for Q. For instance, we may obtain $E_1 \geq Q$ and $E_2 = Q$. This allows us to conclude that $E_1 \geq E_2$, which may have been very difficult to prove directly. The role of counting in this approach is thus to allow us to convert complicated combinatorial information into convenient algebraic statements. The main challenge lies in choosing Q appropriately, so that we use all the information given to us and derive an algebraic conclusion relevant to what we are trying to prove.

Incidence Matrices

Let $A_1, A_2, \ldots, A_n$ be subsets of $\boldsymbol{S} = \{1, 2, \ldots, m\}$. A convenient way to express this information is by drawing an $n \times m$ matrix, with the n rows representing $A_1, A_2, \ldots, A_n$ and the m columns representing the elements of $\boldsymbol{S}$. Entry $a_{ij} = 1$ if and only if element j belongs to A_i. Otherwise, $a_{ij} = 0$. The idea of counting the total number of 1s in an incidence matrix is very useful.

Example 1
Let $A_1, A_2, \ldots, A_6$ be subsets of $\boldsymbol{S} = \{1, 2, \ldots, 8\}$. Suppose each set A_i has 4 elements and each element in $\boldsymbol{S}$ is in m of the A_i's. Find m.

Answer:
We draw an incidence matrix with six rows, representing the subsets $A_1, A_2, \ldots, A_6$ and eight columns representing the elements of $\boldsymbol{S}$. The entry in the ith row and jth column is 1 if and only if the element j belongs to A_i. Otherwise the entry is 0. Since $|A_i| = 4$, each row contains four 1s. There are 6 rows, so the total number of 1s in our matrix is $6 \times 4 = 24$.

Now each element of $\boldsymbol{S}$ is in m of the A_i's. Thus each column of our matrix contains m 1s. So the total number of 1s in the matrix is $8m$, since there are 8 columns. Thus $24 = 8m$, so $m = 3$. ■

1	1	0	1	1	0	0	0

Figure 6.1: A sample 6 x 8 incidence matrix with one row filled in, illustrating set A_3 containing elements 1, 2, 4 and 5.

Counting Pairs and Triples

What we are actually doing in the above proof is counting pairs of the form (*element*, *set*) where the set contains the element. Each 1 in the matrix corresponds to such a pair. If we choose the set first and then the element, there are 6 choices for the set and then 4 for the element, for a total of 24 pairs. We can also choose the element first (8 choices), and then choose the set (m choices, since each element belongs to m sets) for a total of $8m$ pairs. Equating the two answers, $8m = 24$, so $m = 3$.

More generally, we have the following result: If $A_1, A_2, \ldots, A_m$ are subsets of $\{1, 2, \ldots, n\}$ and each element j belongs to d_j of the subsets, then

$$\sum_{i=1}^{m} |A_i| = \sum_{j=1}^{n} d_j.$$

Both sides count the total number of 1s in the matrix, which is the number of pairs (*set*, *element*). The left side counts this quantity by picking the set first and the right side counts it by picking the element first. Note that both sides are also the sum of 1s in the incidence matrix.

In the first example, we counted pairs of the form (*set*, *element*) where the element belongs to the set. There are a few important variations of this technique:

(i) Count *triples* of the form (*set*, *set*, *element*) where the two sets are distinct both contain the element. This is especially useful if we are given information about the **intersection size** of any two sets. These triples can be counted either by first fixing the two sets and then picking the element from their intersection, or by fixing the element and then picking two sets to which it belongs.

Note that counting triples of the form (*set*, *set*, *element*) is equivalent to counting the number of **pairs of 1s that are in the same column** in the incidence matrix representation.

(ii) Count triples of the form (*element*, *element*, *set*) where the two elements both belong to the set. This is useful if we are given information about how many sets two elements appear together in. These triples can be counted in two ways: you can either fix the two elements first or you can fix the set first.

Note that counting triples of the form (*element*, *element*, *set*) is equivalent to counting the number of **pairs of 1s in the same row** in the incidence matrix representation.

The next example demonstrates (ii) in part (a), and the original idea of counting pairs (*set*, *element*) in part (b).

Example 2 [Balanced block designs]
Let $\boldsymbol{X} = \{1, 2, \ldots, v\}$ be a set of elements. A (v, k, λ) *block design* over $\boldsymbol{X}$ is a collection of distinct subsets of $\boldsymbol{X}$ (called blocks) such that:
(i) Each block contains exactly k elements of $\boldsymbol{X}$
(ii) Every pair of distinct elements of $\boldsymbol{X}$ is contained in exactly λ blocks

Let b be the number of blocks. Prove that:
(a) Each element of $\boldsymbol{X}$ is contained in exactly $r = \lambda\,(v - 1)/(k - 1)$ blocks. (In particular, this means that each element is in the same number of blocks, which is initially not obvious)
(b) $r = b\,k\,/\,v$

Answer:
(a) Consider an element s in $\boldsymbol{X}$. We count in two ways the number of triples $(s, u, \boldsymbol{B})$ where u is an element (different from s) and $\boldsymbol{B}$ is a block containing s and u. The first way we count will be to fix $\boldsymbol{B}$ and then u, and the second way will do the reverse. If s is in r blocks, then there are r ways to choose $\boldsymbol{B}$, and then $(k - 1)$ ways to choose u from B. This gives a total of $r\,(k - 1)$. On

the other hand, there are $(v - 1)$ ways to choose u first, and then λ ways to choose $\boldsymbol{B}$ such than $\boldsymbol{B}$ contains both u and s (by condition (ii)). This gives a total of $\lambda (v - 1)$. Hence $r (k - 1) = \lambda (v - 1)$, which is what we wanted.

(b) We count in two ways the number of pairs $(x, \boldsymbol{B})$ where x is an element in a block $\boldsymbol{B}$. There are v ways to choose x, and then r ways to choose $\boldsymbol{B}$. This gives $v\,r$ pairs. On the other hand, there are b ways to choose $\boldsymbol{B}$ first, and then k ways to choose x since $|\boldsymbol{B}| = k$. This gives $b\,k$ pairs. Hence $b\,k = v\,r$. ■

The real power of counting in two ways lies in proving *inequalities*. Typically, we count the number of pairs P (or triples T) of some objects in two ways. At least one of the two counting procedures should give us a bound on P (or T, as may be the case). To do this, we need to cleverly exploit information given to us in the problem statement. The next example is fairly simple, as we use ideas we have already seen in preceding examples.

Example 3 [USA TST 2005]
Let n be an integer greater than 1. For a positive integer m, let $\boldsymbol{S}_m = \{1, 2, ..., mn\}$. Suppose that there exists a $2n$-element set $\boldsymbol{T}$ such that

(a) each element of $\boldsymbol{T}$ is an m-element subset of $\boldsymbol{S}_m$
(b) each pair of elements of $\boldsymbol{T}$ shares at most one common element; and
(c) each element of $\boldsymbol{S}_m$ is contained in exactly two elements of $\boldsymbol{T}$.

Determine the maximum possible value of m in terms of n.

Remark: Make sure you understand the problem – the “elements” of $\boldsymbol{T}$ are actually sets, that is, $\boldsymbol{T}$ is actually a family of subsets of $\boldsymbol{S}_m$.

Answer:
Let $A_1, A_2, ..., A_{2n}$ be the elements of $\boldsymbol{T}$. Let S be the number of triples (x, A_i, A_j) where x is an element of S_m belonging to both sets A_i and A_j. If we choose x first, there is only one choice for the pair

(A_i, A_j) since x belongs in exactly two elements of $\boldsymbol{T}$ by (c). This gives $S = mn$ (the number of choices for x). If we select A_i and A_j first, there is at most one choice for x by (b). Thus $S \leq \binom{2n}{2}$, the number of ways of choosing the pair (A_i, A_j). Hence

$$mn = S \leq \binom{2n}{2} \;\Rightarrow\; m \leq 2n - 1.$$

To give a construction when $m = 2n$-1, simply take $2n$ lines in the plane, no three of which concur and no two of which are parallel. There will be $\binom{2n}{2} = mn$ intersection points formed. The $2n$ lines are the $2n$ elements of $\boldsymbol{T}$, and the mn points are the elements of $\boldsymbol{S}_n$. The conditions of the problem are satisfied, since each point lies on exactly two lines, each two lines meet at exactly one point and each line contains $m = 2n - 1$ points since it meets the other $2n$-1 lines once each. ■

Slightly harder problems require a clever choice of *what pairs or triples to count,* and how to use the information in the problem to get the bounds we want. This comes with practice. One general principle to note is to pay attention to key phrases in the problem like "at most" and "at least". These pieces of information often give a good idea of what we should count.

Example 4 [IMO 1998, Problem 2]

In a competition, there are a contestants and b judges, where $b \geq 3$ 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 $\frac{k}{a} \geq \frac{b-1}{2b}$.

Answer:

Let T be the number of triples (judge, judge, contestant) where the two judges both gave the same rating to the contestant. We can select the two judges in $\binom{b}{2}$ ways, and then select the contestant in **at most k** ways by the condition of the problem. Hence $T \leq k\binom{b}{2}$.

Now take any individual contestant, and suppose the number of judges who rated her "pass" is p and the number who rated her "fail" is $b - p$. The number of triples containing this candidate is $\binom{p}{2} + \binom{b-p}{2} \geq \binom{(b+1)/2}{2} + \binom{(b-1)/2}{2} = (b - 1)^2/4$. Here we used convexity and the fact that b is odd.

Thus each candidate is in at least $(b - 1)^2/4$ triples, so $T \geq a(b - 1)^2/4$. Combining this with our earlier estimate,

$$a(b - 1)^2/4 \leq kb(b - 1)/2 \Rightarrow \frac{k}{a} \geq \frac{b-1}{2b} \blacksquare$$

Unlike the previous example, the next problem offers us no clues that lead us to guess what we should count. However, we can exploit the geometry of the problem to our advantage.

Example 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 $\frac{2}{3}(n^2 - n)$.

Answer:
Let the number of such triangles be k. We count pairs (*edge*, *triangle*) such that the triangle contains the edge. If the number of such pairs is P, then clearly $P = 3k$, since each triangle has 3 edges.

On the other hand, for any edge AB, there are at most four 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. Hence P is at most 4 times the number of edges, which is at most $\binom{n}{2}$. Thus $P \leq 4\binom{n}{2}$. This gives

$$3k \leq 4\binom{n}{2} \Rightarrow k \leq \frac{2}{3}(n^2 - n) \blacksquare$$

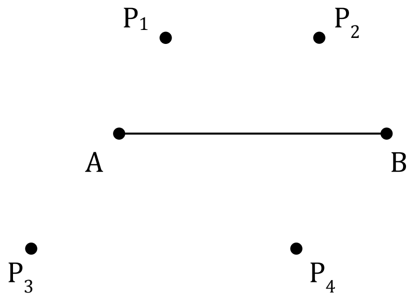

Figure 6.2: At most four points P_1, P_2, P_3, P_4 can form a triangle of unit area with segment AB

Remark: Whenever I'm faced with a combinatorial geometry problem that involves proving an inequality, like the above problem, I use the following principle: use the geometry of the situation to extract some combinatorial information. After that, ignore the geometry completely and use the combinatorial information to prove the inequality. We use this principle in the next example as well.

Example 6 [IMO 1987]

Let n and k be positive integers and let $\boldsymbol{S}$ be a set of n points in the plane such that:

(i) No three points of $\boldsymbol{S}$ are collinear
(ii) For every point P in $\boldsymbol{S}$, there are at least k points in $\boldsymbol{S}$ equidistant from P.

Prove that $k < \frac{1}{2} + \sqrt{2n}$.

Answer:

Condition (ii) implies that for each point P_i in $\boldsymbol{S}$, there exists a circle C_i with center P_i and passing through at least k points of $\boldsymbol{S}$.

Now we count pairs (P_i, P_j) such that P_i and P_j are points in $\boldsymbol{S}$.

Obviously the number of such pairs is $\binom{n}{2}$. On the other hand, each circle C_i has k points on its circumference, which give rise to $\binom{k}{2}$ pairs of points. Thus the n circles in total give us $n\binom{k}{2}$ points. However, there is over counting, since some pairs of points may belong to two circles. Since any two circles meet in at most 2 points, the number of pairs of points that we have counted twice is at most equal to the number of pairs of circles, which is $\binom{n}{2}$. Hence the total number of pairs of points in $\boldsymbol{S}$ is at least $n\binom{k}{2} - \binom{n}{2}$. This implies

$$n\binom{k}{2} - \binom{n}{2} \le \binom{n}{2} \Rightarrow n\binom{k}{2} \le 2\binom{n}{2} \Rightarrow k^2 + k - (n-1) \le 0.$$

Solving this quadratic inequality, noting that k and n are integers, gives us the desired result. ■

The next example again requires a good choice of what to count, in order to capture all the given information.

Example 7 [IMO Shortlist 2004, C1]

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:

(i) Each pair of students is 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.

Answer:

In order to use all the information in the question, we count triples (a, C, S), where a is a student, C is a club and S is a society, where $a \in C$ and $C \in S$. Let the number of such triples be T.

Suppose we first fix a, then S, then C. We can choose a in 10001 ways, S in k ways and then finally C in only one way (by condition (ii)). Hence $T = 10001\,k$.

Now suppose we fix C first. There are $|C|$ ways of doing this. Then by condition (iii), there are $(|C|\ -1)/2$ ways to choose S. Finally there is only one way to choose a, by (ii). This gives

$$T = \sum_{C \text{ is a club}} |C|(|C| - 1)/2 = \sum_{C \text{ is a club}} \binom{|C|}{2}$$

On the other hand, the sum $\sum_{C \text{ is a club}} \binom{|C|}{2}$ is actually equal to the number of pairs of students. This is because each pair of students is in exactly one club by (i), so each pair of students is counted exactly once. Hence this sum is equal to $\binom{10001}{2}$, so putting everything together

$$\binom{10001}{2} = T = 10001\,k \Rightarrow \boldsymbol{k = 5000}.$$

Finally, to construct a configuration for k = 5000, let there be only one club C containing all students and 5000 societies all containing only one club (C). It's easy to see that this works. ■

Counting with Graphs

In the next few examples, we show how to use counting in two ways to solve some problems on graphs. Modeling situations using graphs is very useful, since graphs are very convenient to work with while counting in two ways. For example, suppose we want to count pairs of people such that the two people are friends. If we draw a graph with vertices representing people and an edge between two people if and only if they are friends, then the problem is equivalent to counting the number of edges in the graph.

Some useful properties of graphs

Let $\boldsymbol{G}$ be a graph with n vertices $v_1, v_2, \ldots, v_n$. Let d_i be the degree of v_i, $\boldsymbol{E}$ be the set of edges and $|\boldsymbol{E}| = k$. All summations without indices are assumed to be from 1 to n. We have the following useful properties:

Lemma 1: $\sum d_i = 2k$ (this is because the LHS counts each edge of the graph twice)

Lemma 2: $\sum d_i^2 \geq \frac{(\sum d_i)^2}{n}$ (By Cauchy-Schwarz)

$$\Rightarrow \sum d_i^2 \geq \frac{4k^2}{n}$$

Lemma 3: $\sum \binom{d_i}{2} \geq \frac{2k^2}{n} - k$

Proof: $\binom{d_i}{2} = \frac{d_i^2 - d_i}{2}$. Using lemma 1 and lemma 2 produces the result.

Lemma 4: $\sum_{v_i v_j \in E}(d_i + d_j) = \sum_{i=1}^{n} d_i^2$

Proof: Each term d_i appears in the sum on the LHS d_i times (once for each of the neighbors of v_i). Thus the total sum will be the sum of $d_i \times d_i = d_i^2$ for each i, which is the RHS.

There are also some important results on directed graphs, especially *tournaments.* A tournament on n vertices is a directed graph such that between any two vertices u and v, there is either a directed edge from u to v or a directed edge from v to u. One can interpret these graphs as follows: the n vertices stand for participants in a tournament, and each two players play a match.

There are no ties. If v beats u, then there is a directed edge from v to u.

Let $P_1, P_2, \ldots, P_n$ be the n participants. Let w_i and l_i denote the number of wins and losses of P_i. Clearly $w_i + l_i = (n - 1)$ for each i, because each person plays against $(n - 1)$ others. Also, $\sum w_i = \sum l_i$ since each match has a winner and a loser, and so contributes 1 to both sides. Hence in fact both sides are equal to $\binom{n}{2}$, the total number of matches. We have another interesting but less obvious result:

Lemma 5: $\sum w_i^2 = \sum l_i^2$.

Proof: Define a *noncyclic triple* to be a set of 3 players A, B and C such that A beat both B and C and B beat C. Call A the *winner* of the triplet and C the *loser* of the triplet. If we count noncyclic triplets by winners, the sum would be $\sum\binom{w_i}{2}$, since after choosing the winner there are $\binom{w_i}{2}$ ways to choose the other two players who he beat. If we count by losers, the sum is $\sum\binom{l_i}{2}$, since after choosing the loser there are $\binom{l_i}{2}$ ways to choose the other two players. Hence $\sum\binom{w_i}{2} = \sum\binom{l_i}{2}$. Combining this with $\sum w_i = \sum l_i$ we get the result.

Remark: Whenever you see expressions of the type in this lemma, like a sum of squares, try to interpret them combinatorially. For instance, it is often useful to convert x^2 to $2\binom{x}{2} + 2x$. Allow these sums to give you hints as to what to count. In the proof of lemma 5, the term $\binom{w_i}{2}$ gives us a hint to count triples of the form (X, Y, Z) such that X beat both Y and Z. This almost automatically leads us to the solution.

Example 8 [APMO 1989] (U*)

Show that a graph with n vertices and k edges has at least $k(4k - n^2)/3n$ triangles.

Note: The symbol **U*** in brackets next to a problem indicates that it is a useful result and should be remembered.

Answer:
We count pairs (*edge*, *triangle*) where the triangle contains the edge. Consider an edge v_iv_j. How many triangles have v_iv_j as an edge? v_i is joined to $(d_i - 1)$ vertices other than v_j, and v_j is joined to d_j vertices other than v_i. There are only $n - 2$ vertices other than v_i and v_j. Hence at least $(d_i - 1) + (d_j - 1) - (n - 2) = (d_i + d_j - n)$ vertices are joined to *both* v_i and v_j. Each of these gives one triangle. Hence each edge v_iv_j is in at least max $\{0, (d_i + d_j - n)\}$ triangles.

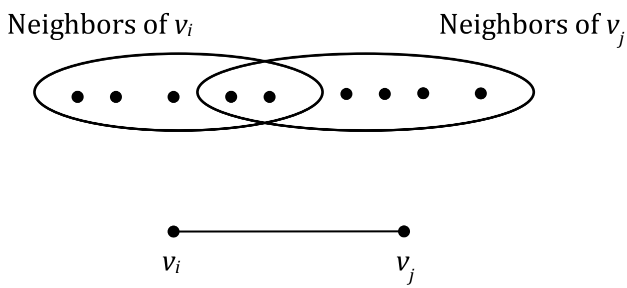

Figure 6.3: The set of vertices neighboring both v_i and v_j must contain at least $(d_i + d_j - n)$ vertices

Thus the total number of triangles is at least

$$\frac{1}{3}\sum_{v_iv_j \in E}(d_i + d_j - n) = \frac{1}{3}\sum_{i=1}^{n} d_i^2 - \frac{nk}{3} \quad \text{(Using lemma 4)}$$

$$\geq \frac{1}{3} \times \frac{4k^2}{n} - \frac{nk}{3} = \frac{k(4k-n^2)}{3n}. \quad \text{(Using lemma 3)}$$

Note that we divided by 3 because otherwise each triangle would be counted thrice (once for each edge). ■

Corollary 1 (U*)
A graph with no triangles has at most $\lfloor n^2/4 \rfloor$ edges. Equality is achieved only by bipartite graphs with an equal or almost equal number of vertices in each part. This is an extremely useful result, and is a special case of **Turan's theorem**, which will be discussed in the exercises of chapter 8.

Example 9 [Indian TST 2001] (U*)
Let $\boldsymbol{G}$ be a graph with E edges, n vertices and no 4-cycles. Show that $E \leq \frac{n}{4}(1 + \sqrt{4n-3})$.

Answer:
Let the vertices be $\{v_1, \ldots, v_n\}$ and let the degree of v_i be d_i. Let T be the number of "*V-shapes*": that is, triples of vertices (u, v, w) such that v and w are both adjacent to u. The vertices v and w may or may not be adjacent and triples $\{u, v, w\}$ and $\{u, w, v\}$ are considered the same.

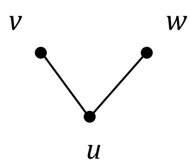

Figure 6.4: A "V shape"

The reason for this choice of T is that if we first select v and w, then there is at most one u such that $\{u, v, w\}$ is a triple in T. Otherwise there would be a 4 cycle. Hence we get $\boldsymbol{T} \leq \binom{\boldsymbol{n}}{\boldsymbol{2}}$, since for each of the $\binom{n}{2}$ ways of choosing v and w, there is at most one way to choose u.

If we choose u first, there are $\binom{d_u}{2}$ ways of choosing v and w, where d_u is the degree of u. Summing over all choices for u, and then using lemma 3, we get

$$T = \sum_{i=1}^{n} \binom{d_i}{2} \geq \frac{2E^2}{n} - E$$

Combining this with $T \leq \binom{n}{2}$,

$$\frac{2E^2}{n} - E \leq \frac{n(n-1)}{2}.$$

This reduces to a quadratic inequality in E, which yields the desired bound. ■

Sometimes, when we need to bound or count the number of objects satisfying some property, it is easier or more convenient to count the number of objects *not* satisfying the property. Then we can subtract this from the total number of objects to get the result.

Example 10 [USAMO 1995]
Suppose that in a certain society, each pair of persons can be classified as either *amicable* or *hostile*. We shall say that each member of an amicable pair is a *friend* of the other, and each member of a hostile pair is a *foe* of the other. Suppose that the society has n people and q amicable pairs, and that for every set of three persons, at least one pair is hostile. Prove that there is at least one member of the society whose foes include $q\,(1 - 4q/n^2)$ or fewer amicable pairs.

Answer:
We naturally rephrase the problem in graph theoretic terms, with vertices representing people and an edge joining two vertices if and only if they form an amicable pair. The graph has no triangles by assumption, n vertices and q edges. We wish to estimate the number of edges containing 2 foes of X, where X is a vertex. To do this, we first count P, the number of pairs (E, X), where E is an edge containing X or a friend of X.

First we count the number of pairs (X, E) where E is an edge containing a neighbor of X but not containing X. This quantity will be equal to T, the number of triples (X, Y, Z) such that XY and YZ are edges. (X, Y, Z) is considered different from (Z, Y, X). Note that XZ cannot be an edge by the condition that there are no triangles. To compute T, we count by Y. The number of triples containing Y is $d_Y(d_Y - 1)$, so the total number of triples is $\sum_{i=1}^{n} d_i(d_i - 1)$.

Now clearly the number of pairs (X, E) where E is an edge containing X is given by the sum $\sum_{i=1}^{n} d_i$. If we add this summation to the previous summation, we would have counted the number of pairs (X, E) where E is an edge containing X OR a friend of X but not X. Thus the total number of such pairs is

$$P = \sum_{i=1}^{n} d_i(d_i - 1) + \sum_{i=1}^{n} d_i = \sum_{i=1}^{n} d_i^2 \geq 4q^2/n,$$

by lemma 2.

Hence by averaging, there is some X such that there at least $4q^2/n^2$ pairs (X, E), where E is an edge X or at least one neighbor of X. Thus the number of edges joining two foes of X is at most $q - 4q^2/n^2 = q(1 - 4q/n^2)$.

Example 11 [Generalization of Iran TST 2008]

In a tournament with n players, each pair of players played exactly once and there were no ties. Let j, k be integers less than n such that $j < 1 + \frac{n\binom{(n-1)/2}{k}}{\binom{n}{k}}$. Show that there exist sets $\boldsymbol{A}$ and $\boldsymbol{B}$ of k players and j players respectively, such that each player in $\boldsymbol{A}$ beat each player in $\boldsymbol{B}$.

Answer

Count $(k + 1)$–tuples of the form $(P_1, P_2, \ldots, P_k, L)$ where L lost to each of the players $P_1, P_2, \ldots, P_k$. Let T be the total number of such tuples. If we fix L, we get $\binom{d_L}{k}$ tuples containing L, where d_L is the number of players L lost to. Summing over all n choices of L, $T \geq$

$\sum_{i=1}^{n}\binom{d_i}{k}$, where d_i is the number of losses of the ith player. Hence by Jensen's inequality, $T \geq n \times \binom{\sum_{i=1}^{n} d_i/n}{k} = n \times \binom{(n-1)/2}{k}$, since $\sum_{i=1}^{n} d_i = n\,(n-1)/2$.

Now assume to the contrary that there do not exist such sets ***A*** and ***B***. Then for any choice of $P_1, P_2, \ldots, P_k$, there are at most $(j-1)$ choices for L. Hence $T \leq \binom{n}{k}(j-1)$.

Combining these estimates gives $(j-1) \geq \frac{n\binom{(n-1)/2}{k}}{\binom{n}{k}}$, which contradicts the condition of the problem. Thus, our assumption in the second paragraph is false, and such sets ***A*** and ***B*** indeed exist.

Example 12 [IMO Shortlist 2010 C5]
$n \geq 4$ 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 these three players formed a cyclic triple (a set (A, B, C) such that A beat B, B beat C and C beat A). 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 ith player. Prove that

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

Answer:
Note that

$$\sum_{i=1}^{n}(w_i - l_i)^3 = \sum_{i=1}^{n}(w_i^3 - l_i^3) + 3\sum_{i=1}^{n}(w_i l_i^2 - w_i^2 l_i)$$

We will show that

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

(ii) $\sum_{i=1}^{n} w_i l_i^2 \geq \sum_{i=1}^{n} w_i^2 l_i$

From now on, any summation without indices is assumed to be from 1 to n. Note that by using lemma 5, we can reduce **(i)** to the "more combinatorial" form

(iii) $\sum\binom{w_i}{3} \geq \sum\binom{l_i}{3}$

Let us define a "chained quadruple" as a set of 4 players with no cyclic triple amongst them. It is easy to see that a chained quadruple has

(a) A person who won against all the other three players, called the *winner*

(b) A person who lost against all the other three players, called the *loser*

The converse of (a) is not true, since the other three players may form a cyclic triple. However, the converse of (b) holds, since by assumption there is no bad quadruple. Let Q be the number of chained quadruples. If we count Q by picking the loser first, we get

$$Q = \sum\binom{l_i}{3}$$

If we count Q by picking the winner first, noting that the converse of (a) doesn't hold, then

$$Q \leq \sum\binom{w_i}{3}$$

Hence $\sum\binom{w_i}{3} \geq \sum\binom{l_i}{3}$, which proves **(iii)** and hence **(i)**.

To prove **(ii)**, subtract $\sum w_i l_i$ from both sides and divide by 2, to write it as:

(iv) $\sum w_i\binom{l_i}{2} \geq \sum l_i\binom{w_i}{2}$

Observe that the LHS of this expression counts pairs of the form (*quadruple*, *person*) such that the person won exactly one game against the other three in the quadruple. Similarly, the RHS counts pairs such that the person won exactly two games.

Now let us look at the types of quadruples we can have. If in a certain quadruple the number of games won by each person against the other three are a, b, c, d in non-increasing order, we say that this quadruple is of type (a, b, c, d). The only types we can have are:

(3, 1, 1, 1) – Note that this refers to a quadruple in which one person beat the other three, and the other three each won one game. This type of quadruple contributes 3 to the LHS of **(iv)** (three people won one game) and 0 to the RHS (no one won two games).
(2, 2, 1, 1) – This contributes 2 to both sides of **(iv)**
(3, 2, 1, 0) – This contributes 1 to both sides of **(iv)**
(2, 2, 2, 0) – This is not allowed: this is a bad company.

Thus we see that every allowed quadruple contributes at least as much to the LHS of **(iv)** as it does to the RHS. Hence **(iv)** indeed holds, which proves **(ii)**. Hence **(i)** and **(ii)** together give us the desired result and we are done.

Remark 1: This example shows the true power of "interpreting things combinatorially".

Remark 2: This problem was the first relatively hard (rated above C2 or C3) combinatorics problem I ever solved, and my solution was essentially the one above. The thought process behind this solution is fairly natural – keep expressing things "combinatorially", let these expressions guide what you choose to count, and exploit the fact that there is no "(2, 2, 2, 0)". Also note that it is not essential to prove **(i)** and **(ii)** separately: one can directly show that $\sum\binom{w_i}{3} + \sum w_i\binom{l_i}{2} \geq \sum\binom{l_i}{3} + \sum l_i\binom{w_i}{2}$ by comparing the contributions to each side by each type of quadruple.

Miscellaneous Applications

In this section we look at some unexpected applications of counting in two ways.

Example 13 [IMO 2001, Problem 4]
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 $S(a) = \sum_{i=1}^{n} a_i c_i$. Prove that there exist permutations $a \neq b$ of such that $n!$ is a divisor of $S(a) - S(b)$.

Answer:
Suppose to the contrary that all the $S(a)$'s are distinct modulo $(n!)$. Since there are $n!$ possibilities for a, this means that $S(a)$ takes each value in $\{1, 2, \ldots, n!\}$ modulo $n!$ Consider the sum of all the $S(a)$'s modulo $n!$. If the sum is S, then

$$S \equiv 1 + 2 + \ldots + n! \equiv n!\,(n! + 1)/2 \bmod n! \equiv n!/2 \bmod n!.$$

On the other hand, the coefficient of each c_i in S is

$$(n-1)!\,(1 + 2 + \ldots + n) = n!\,(n+1)/2 \equiv 0 \bmod n!,$$

since n is odd and 2 divides $n + 1$. Thus the coefficient of each c_i in S is divisible by $n!$, so $S \equiv 0 \bmod n!$. This is a contradiction to the result in the first paragraph. ■

Example 14 [IMO Shortlist 2003, C4]
Given n real numbers $x_1, x_2, \ldots, x_n$ and n further real numbers $y_1, y_2, \ldots, y_n$. The entries a_{ij} (with $1 \leq i, j \leq n$) of an $n \times n$ matrix A are defined as follows:

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

Further, let B be an $n \times n$ matrix whose elements are numbers from the set {0, 1} satisfying the following condition: The sum of all elements of each row of B equals the sum of all elements of the corresponding row of A; the sum of all elements of each column of B equals the sum of all elements of the corresponding column of A. Show that in this case $A = B$.

Answer:
Let b_{ij} denote the entry in the ith row and jth column of B. Define

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

On the one hand,

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

since $\sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} b_{ij}$ and $\sum_{i=1}^{n} a_{ij} = \sum_{i=1}^{n} b_{ij}$ by the conditions of the problem.

On the other hand, note that if $x_i + y_j \ge 0$, then $a_{ij} = 1$ so $(a_{ij} - b_{ij}) \ge 0$. If $x_i + y_j < 0$, then $a_{ij} = 0$ so $a_{ij} - b_{ij} \le 0$. Thus in both cases, $(x_i + y_j)(a_{ij} - b_{ij}) \ge 0$. Hence each term in the summation is nonnegative, but the total sum is 0. Thus each term is 0. Hence whenever $(x_i + y_j) \ne 0$, we must have $a_{ij} = b_{ij}$. Whenever $(x_i + y_j) = 0$, then $a_{ij} = 1$. In these cases we must have $b_{ij} = 1$ since the sum of all the entries in both matrices is the same. Hence in all cases $a_{ij} = b_{ij}$, and we are done. ■

Remark: Where on earth does the expression

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

come from?!?! Note that one way of proving that several different real numbers are 0 is to show that their squares sum to 0, since no square is negative. Thus, a first approach to the problem may be to show that the sum

$$S' = \sum_{1 \le i,j \le n}(a_{ij} - b_{ij})^2$$

is 0. This doesn't work as it doesn't utilize information about the x's and y's. Instead we try the following modification: we seek to weight each term $(a_{ij} - b_{ij})$ by some other quantity that still ensures that each term in the summation is nonnegative, and additionally enables us to use the information about the x's and y's to show that the entire sum is 0.

Example 15 [Indian TST 2010]
Let $A = (a_{jk})$ be a 10 x 10 array of positive real numbers such that the sum of numbers in each row as well as in each column is 1. Show that there exists $j < k$ and $l < m$ such that

$$a_{jl}a_{km} + a_{jm}a_{kl} \ge \frac{1}{50}$$

Answer:
To make things more intuitive, let us interpret the algebraic expression $a_{jl}a_{km} + a_{jm}a_{kl}$visually. The centers of the squares containing entries a_{jl}, a_{jm}, a_{km} and a_{kl} form a rectangle with sides parallel to grid lines. Define the *value* of this rectangle to be $a_{jl}a_{km} + a_{jm}a_{kl}$. Assuming to the contrary that the *value* of any such rectangle is strictly less than 1/50.

Observe that as j, k, l, m vary within the bounds $1 \le j < k \le 10$ and $1 \le l < m \le 10$, we obtain $\binom{10}{2}^2 = 45^2$ such rectangles. Let S be the sum of *values* of these 45^2 rectangles. By our earlier assumption, we obtain $S < 45^2/50 = 40.5$. We will now compute S in a different way to yield a contradiction.

Note that a_{jl} and a_{km} lie diagonally opposite and a_{jm} and a_{kl} lie diagonally opposite each other. Thus in each rectangle the diagonally opposite pairs of entries are multiplied. Hence, when the sum of *values* is taken over all rectangles, each entry a_{ij} occurs in products with every other entry in the array *except those in its own row or column*, since two entries in the same row or column can never be diagonally opposite in a rectangle. Therefore,

$$S = \frac{1}{2}\sum_{1\le i,j\le 10} a_{ij}\, S_{ij},$$

where S_{ij} is the sum of all entries except those in the ith row and jth column. Note that we have divided by two since if we simply sum the terms $a_{ij}S_{ij}$, we will be counting each product $a_{ij}a_{kl}$ twice.

Observe that $S_{ij} = (10 - 1 - 1 + a_{ij}) = (8 + a_{ij})$, since the sum of all entries is 10 and the sum in each row and column is 1. Note that the " $+ a_{ij}$" occurs since when we subtract all elements in row i and in column j, a_{ij} is subtracted twice. Thus the total sum is

$$S = \frac{1}{2}\sum_{1\le i,j\le 10} a_{ij}\, S_{ij} = \frac{1}{2}\sum_{1\le i,j\le 10} a_{ij}(a_{ij} + 8)$$
$$= 4\sum_{1\le i,j\le 10} a_{ij} + \frac{1}{2}\sum_{1\le i,j\le 10} a_{ij}^2$$

Now $\sum_{1\le i,j\le 10} a_{ij} = 10$ and $\sum_{1\le i,j\le 10} a_{ij}^2 \ge \frac{(\sum_{1\le i,j\le 10} a_{ij})^2}{100} = 1$, using Cauchy Schwarz. Thus

$S \ge 4\text{x}10 + 0.5 = 40.5$, a contradiction. ■

Remark: The visual interpretation as "diagonally opposite entries in rectangles" is by no means essential (and entails a little abuse of notation as well, for which I apologize). Simply taking a suitable double summation would lead to a significantly shorter but equivalent proof. However, I felt the basic intuition behind the problem may have been lost in a sea of symbols that would have mysteriously spat out the solution, so I chose to write the proof this way.

Exercises

1. **[Due to Grigni and Sipser]**
Consider an $m \times n$ table (m rows, n columns), in which each cell either contains a 0 or a 1. Suppose the entire table contains at least αmn 1s, where $0 < \alpha < 1$. Show that at least one of the following must be true:
(i) There exists a row containing at least $n\sqrt{\alpha}$ 1s
(ii) There exist at least $m\sqrt{\alpha}$ rows containing at least αn 1s.

2. **[Italy TST 2005, Problem 1]**
A class is attended by n students ($n > 3$). The day before the final exam, each group of three students conspire against another student to throw him/her out of the exam. Prove that there is a student against whom there are at least $\sqrt[3]{(n-1)(n-2)}$ conspirators.

3. **[Important Lemmas on incident matrices] (U*)**
Let A be an $r \times c$ matrix with row sums R_i (that is, the sum of the elements in the ith row is R_i) and column sums C_j. Suppose R_i and C_j are positive for all $1 \leq i \leq r$ and $1 \leq j \leq c$.

(i) Show that $\sum_{i,j} \frac{a_{ij}}{R_i} = r$ and $\sum_{i,j} \frac{a_{ij}}{C_j} = c$
(ii) Suppose $C_j \geq R_i$ whenever $a_{ij} = 1$. Using (i), show that $r \geq c$.
(iii) Suppose instead of the condition in (ii) we were given that $0 < R_i < c$ and $0 < C_j < r$ for each i and each j, and furthermore, $C_j \geq R_i$ whenever $a_{ij} = 0$. Prove that $r \geq c$.

4. **[IMO 1987, Problem 1]**
Let $p(n, k)$ denote the number of permutations of $\{1, 2, \ldots, n\}$ with exactly k fixed points. Show that $\sum_{k=1}^{n} kp(n, k) = n!$

5. [Corradi's Lemma] (U*)
Let $A_1, A_2, \ldots, A_n$ be r-element subsets of a set $\boldsymbol{X}$. Suppose that $A_i \cap A_j \le k$ for all $1 \le i < j \le n$. Show that $|\boldsymbol{X}| \ge \frac{nr^2}{r+(n-1)k}$.

6. [Erdos-Ko-Rado] (U*)
Let $\boldsymbol{F}$ be a family of k-element subsets of $\{1, 2, \ldots, n\}$ such that every two sets in $\boldsymbol{F}$ intersect in at least one element. Show that $|\boldsymbol{F}| \le \binom{n-1}{k-1}$.

7. [Indian Postal Coaching 2011]
In a lottery, a person must select six distinct numbers from $\{1, 2, \ldots, 36\}$ to put on a ticket. The lottery committee will then draw six distinct numbers randomly from $\{1, 2, \ldots, 36\}$. Any ticket not containing any of these 6 numbers is a winning ticket. Show that there exists a set of nine tickets such that at least one of them will certainly be a winning ticket, whereas this statement is false if 9 is replaced by 8.

8. [Hong Kong 2007]
In a school there are 2007 girls and 2007 boys. Each student joins at most 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.

9. [IMO Shortlist 1995, C5]
At a meeting of $12k$ people, each person exchanges greetings with exactly $(3k + 6)$ others. For any two people, the number of people who exchange greetings with both of them is the same. How many people are at the meeting?

10. [Based on Furedi's result on maximal intersecting families]
Let n and k be positive integers with $n > 2k - 1$, and let $\boldsymbol{F}$ be a family of subsets of $\{1, 2, \ldots, n\}$ such that each set in $\boldsymbol{F}$ contains k elements, and every pair of sets in $\boldsymbol{F}$ has nonzero

intersection. Suppose further that for any k-element subset $\boldsymbol{X}$ of $\{1, 2, \ldots, n\}$ not in $\boldsymbol{F}$, there exists a set $\boldsymbol{Y}$ in $\boldsymbol{F}$ such that $\boldsymbol{X} \cap \boldsymbol{Y} = \emptyset$. Show that there are at least $\frac{\binom{n}{k}}{\binom{n-k}{k}+1}$ sets in $\boldsymbol{F}$.

11. [IMO Shortlist 2000, C3]

Let $n > 3$ be a fixed positive integer. Given a set $\boldsymbol{S}$ of n points P_1, P_2, …, P_n in the plane such that no three are collinear and no four concyclic, let a_t be the number of circles $P_i P_j P_k$ that contain P_t in their interior, and let $m(\boldsymbol{S}) = a_1 + a_2 + \ldots + a_n$. Prove that there exists a positive integer $f(n)$ depending only on n such that the points of $\boldsymbol{S}$ are the vertices of a convex polygon if and only if $m(\boldsymbol{S}) = f(n)$.

12. [Iran 2010]

There are n students in a school, and each student can take any number of classes. There are at least two students in each class. Furthermore, if two different classes have two or more students in common, then these classes have a different number of students. Show that the number of classes is at most $(n-1)^2$.

13. [IMO Shortlist 2004, C4]

Consider a matrix of size $n \times n$ whose entries are real numbers of absolute value not exceeding 1. The sum of all entries of the matrix is 0. Let n be an even positive integer. Determine the least number C such that every such matrix necessarily has a row or a column with the sum of its entries not exceeding C in absolute value.

14. [Generalization of USAMO 2011, Problem 6]

Let $A_1, A_2, \ldots, A_n$ be sets such that $|A_i| = \binom{n-1}{2}$ for each $1 \leq i \leq n$ and $|A_i \cap A_j| = (n-2)$ for each $1 \leq i < j \leq n$. Show that $|A_1 \cup A_2 \cup \ldots \cup A_n| \geq \binom{n}{3}$, and show that it is possible for equality to occur.

15. [Iran 1999]

Suppose that $C_1, C_2, \ldots, C_n$ ($n \geq 2$) are circles of radius one in the plane such that no two of them are tangent, and the subset of the plane formed by the union of these circles is connected.

Let $\boldsymbol{S}$ be the set of points that belong to at least two circles. Show that $|\boldsymbol{S}| \geq n$.

16. [IMO Shortlist 2000, C5]

Suppose n rectangles are drawn in the plane. Each rectangle has parallel sides and the sides of distinct rectangles lie on distinct lines. The rectangles divide the plane into a number of regions. For each region R let $v(R)$ be the number of vertices. Take the sum of $v(R)$ over all regions which have one or more vertices of the rectangles in their boundary. Show that this sum is less than $40n$.

17. [Indian TST 1998]

Let $\boldsymbol{X}$ be a set of $2k$ elements and $\boldsymbol{F}$ a family of subsets of $\boldsymbol{X}$ each of cardinality k such that each subset of $\boldsymbol{X}$ of cardinality $(k-1)$ is contained in exactly one member of $\boldsymbol{F}$. Show that (k + 1) is a prime.

18. [IMO Shortlist 1988]

For what values of n does there exist an $n \times n$ array of entries -1, 0 or 1 such that the $2n$ sums obtained by summing the elements of the rows and the columns are all different?

19. [IMO 2001, Problem 3]

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.

20. [IMO 2005, Problem 6]
In a mathematical competition 6 problems were posed to the contestants. Each pair of problems was solved by more than 2/5 of the contestants. Nobody solved all 6 problems. Show that there were at least 2 contestants who each solved exactly 5 problems.

21. Let $\boldsymbol{A}$ be a set with n elements, and let $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$ be subsets of $\boldsymbol{A}$ such that $|\boldsymbol{A}_i| \geq 2$ for each $1 \leq \mathrm{i} \leq n$. Suppose that for each 2-element subset $\boldsymbol{A}'$ of $\boldsymbol{A}$, there is a unique i such that $\boldsymbol{A}'$ is a (not necessarily proper) subset of $\boldsymbol{A}_i$. Show that for all pairs (i, j) such that $1 \leq i < j \leq n$, $\boldsymbol{A}_i \cap \boldsymbol{A}_j > 0$.

22. [USAMO 1999 proposal]
Let n, k and m be positive integers with $n > 2k$. Let $\boldsymbol{S}$ be a nonempty set of k-element subsets of $\{1, 2, \ldots, n\}$ such that every $(k + 1)$-element subset of $\{1, 2, \ldots, n\}$ contains exactly m elements of $\boldsymbol{S}$. Prove that $\boldsymbol{S}$ must contain every k-element subset of $\{1, 2, \ldots, n\}$.

23. [Based on Zarankeiwicz' problem]
At a math contest there were m contestants and n problems. It turned out that there were numbers $a < m$ and $b < n$ such that there did not exist a set of a contestants and b problems such that all a contestants solved all b problems. Define the *score* of each contestant to be the number of problems he solved, and let S denote the sum of the scores of all m contestants. Show that $S \leq (a - 1)^{1/b} n m^{1\text{-}1/b} + (b - 1)m$.

24. [IMO Shortlist 2007, C7]
Let $\alpha < \frac{3-\sqrt{5}}{2}$ be a positive real number. Prove that there exist positive integers n and $p > \alpha 2^n$ for which one can select $2p$ pairwise distinct subsets $S_1, S_2, \ldots, S_p, T_1, T_2, \ldots, T_p$ of the set $\{1, 2, \ldots, n\}$ such that $S_i \cap T_j \neq \emptyset$ for all $1 \leq i, j \leq p$.