Counting in Two Ways

Incidence Matrices

June 26, 2007

Yufei Zhao

yufeiz@mit.edu

Problems to be discussed in lecture

Problem 1. In a certain committee, each member belongs to exactly three subcommittees, and each subcommittee has exactly three members. Prove that the number of members equals to the number of subcommittees.

Problem 2. (IMC 2002) Two hundred students participated in a mathematical contest. They had six 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.

Problem 3. (IMO 1998) In a competition, there are *a* contestants and *b* judges, where $b \ge 3$ 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 $k/a \ge (b-1)/2b$.

Problem 4. (Iran 1999) Suppose that C_1, \ldots, C_n $(n \ge 2)$ 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 S be the set of points that belong to at least two circles. Show that $|S| \ge n$.

Problem 5. 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 $m \leq n$.

Counting in Two Ways

Incidence Matrices

June 26, 2007

Yufei Zhao

yufeiz@mit.edu

A large number of combinatorics problems involve looking at a quantity in at least two different ways. This technique is often called "double counting." In this note, we will mainly focus on how to use incidence matrices to help us set up the counting. While the incidence matrix is a powerful tool, there are many problems which cannot be tackled using incidence matrices—you will encounter some of this in the problems section.

Incidence Matrices

Let's start with a really simple (and silly) example.

Problem 1. In a certain committee, each member belongs to exactly three subcommittees, and each subcommittee has exactly three members. Prove that the number of members equals to the number of subcommittees.

Here's how we usually set up the incidence matrix. In our incidence matrix, each row represents an individual, and each column represents an organization. An entry is 1 if the individual corresponding to its row belongs to the organization corresponding to its column; otherwise, the entry is 0. Of course, the roles of rows and columns may be interchanged. Two examples of the configuration described in Problem 1 are shown below.

	(1	1	1	0	0	0	0	0	0 \
$(1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0)$	0	0	0	1	1	1	0	0	0
1 0 0 1 1 0 0	0	0	0	0	0	0	1	1	1
1 0 0 0 0 1 1	1	0	0	1	0	0	1	0	0
0 1 0 1 0 1 0	0	1	0	0	1	0	0	1	0
0 1 0 0 1 0 1	0	0	1	0	0	1	0	0	1
0 0 1 1 0 0 1	1	0	0	0	1	0	0	0	1
	0	1	0	0	0	1	1	0	0
`````	0	0	1	1	0	0	0	1	0 /

**Counting the number of 1's.** To solve Problem 1, we could ask ourselves, how many 1's are there in the incidence matrix? Suppose that there are n subcommittees and m members. Then the incidence matrix is a  $m \times n$  matrix. The given conditions tell us that each row contains 3 ones, so there are 3m ones in total. On the other hand, each column contains 3 ones, so there are 3m ones in total. Equating the two counts, we see that 3m = 3n, so m = n, which is what we wanted to prove.

The following result follows from the same idea.

**Proposition 1.** If  $A = (a_{i,j})$  is a  $r \times c$  matrix with row sums  $R_i$ , i = 1, 2, ..., r, and column sums  $C_j$ , j = 1, 2, ..., c, then

$$\sum_{i=1}^r R_i = \sum_{j=1}^c C_j.$$

#### Counting pairs of 1's.

Here's an approach that appears frequently in combinatorial problems. Oftentimes, we are given some restriction that applies to every *pair* of organizations (or individuals). For example, it may be that every two organizations share exactly one common member. In this case, counting the number of 1's as we did above does not incorporate all the given information, and thus would likely be unsuccessful. Fortunately, such problems can usually be approached by counting pairs of 1's. Specifically, we are interested in the number of pairs of 1's that lie on the same column (or row).

**Proposition 2.** Let  $A = (a_{i,j})$  be a  $r \times c$  (0,1)-matrix with column sums  $C_j$ . Suppose that for every two rows, there exist exactly t columns that contain 1's from both rows, then

$$t\binom{r}{2} = \sum_{j=1}^{c} \binom{C_j}{2}.$$

*Proof.* Let  $\mathcal{T}$  denote the set of all unordered pairs of 1's that lie in the same column. Let us count the elements of  $\mathcal{T}$  in two different ways.

Counting by rows: For any two rows, there are t pairs of 1's among these rows that belong to  $\mathcal{T}$ , so  $|\mathcal{T}| = t {r \choose 2}$ .

Counting by columns: In the  $j^{\text{th}}$  column, there are  $C_j$  1's, and thus  $\binom{C_j}{2}$  pairs. Counting over all the columns gives  $|\mathcal{T}| = \sum_{i=1}^{c} \binom{C_j}{2}$ .

The result follows by equating the above two expressions.

**Inequalities.** In most problems of the type, we are not given enough information to directly produce a combinatorial identity. Instead, we have to work with inequalities and bounds.

Many incidence matrix problems strive for the existence of a certain subconfiguration. These problems can usually be approached through contradiction. Under the assumption that the opposite result holds, we can count a particular set (e.g., the set of all pairs of 1's that belong in the same column) in two different ways, once by rows and once by columns. Furthermore, we would want to establish an upper bound in one of our counts, and a lower bound in the other count. If the upper bound turns out to be less than the lower bound, then a contradiction is reached.

**Problem 2.** (IMC 2002) Two hundred students participated in a mathematical contest. They had six 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.

Solution. Let us assume that the contrary is true. That is, for every two students, there is some problem that neither of them solved. This prompts us to count the pairs of students with their unsolved problem.

Let us consider the incidence matrix of this configuration. We have six rows, each representing a problem, and 200 columns, each representing a student. In light of the above remark, we make an entry of the matrix 1 if the student corresponding to the column did *not* solve the problem corresponding to the row, and make the entry 0 otherwise. The setup is illustrated below.

 $\begin{array}{c} \text{Problem 1} \\ \text{Problem 2} \\ \text{Problem 3} \\ \text{Problem 4} \\ \text{Problem 5} \\ \text{Problem 6} \end{array} \left( \begin{array}{ccccc} 0 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 0 \end{array} \right)$ 

Let  $\mathcal{T}$  denote the set of pairs of 1's that belong in the same row. Let us consider the cardinality of  $\mathcal{T}$  from two different perspectives.

Counting by columns: We assumed that for every two students, there was a problem that neither of them solved. Thus, for every two columns, there is at least one pair of 1's among these two columns that belong in the same row. So we can find an element of  $\mathcal{T}$  in every pair of columns. Since there are  $\binom{200}{2}$  pairs of columns, we have  $|\mathcal{T}| \geq \binom{200}{2} = 19,900$ .

Counting by rows: We are told that each problem was solved by at least 120 students. This means that there are at most 80 ones in each row. So each row contains at most  $\binom{80}{2}$  pairs of 1's. Since there are six rows, we have  $|\mathcal{T}| \leq 6\binom{80}{2} = 18,960$ .

Combining the above two inequalities, we get  $19,900 \le |\mathcal{T}| \le 18,960$ , which is clearly absurd. Therefore, our initial assumption must be false. So there must be two students such that every problem was solved by at least one of these two students.

**Convexity.** As we are often interested in counting pairs of 1's, the function  $f(n) = \binom{n}{2}$  appears quite frequently. Let us extend this function to the real numbers in the obvious way:  $f(x) = \frac{1}{2}x(x-1)$ . Note that f is a convex function. Using Jensen's theorem, we can get the following inequality.

$$\binom{a_1}{2} + \binom{a_2}{2} + \binom{a_3}{2} + \dots + \binom{a_n}{2} \ge \frac{s(s-n)}{2n},$$

where  $a_1, a_2, \dots, a_n$  are positive integers and  $s = a_1 + a_2 + \dots + a_n$ .

However, this bound is not always the best possible, since the equality is attained at  $a_i = s/n$  for all i, which may not be achieved as  $a_i$  needs to be an integer. Using the fact that the  $a_i$ 's must be integers, we can get the following tight bound, either using Karamata's majorization inequality, or more simply through discrete smoothing.

$$\binom{a_1}{2} + \binom{a_2}{2} + \binom{a_3}{2} + \dots + \binom{a_n}{2} \ge r\binom{k+1}{2} + (n-r)\binom{k}{2}$$

where  $a_1, a_2, \dots, a_n$  are positive integers, and  $s = a_1 + a_2 + \dots + a_n = nk + r$ , where k and r are integers such that  $0 \le r < n$ .

**Problem 3.** (IMO 1998) In a competition, there are *a* contestants and *b* judges, where  $b \ge 3$  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  $k/a \ge (b-1)/2b$ .

Solution. Let us form an incidence matrix as usual. Let there be b rows, each representing a judge, and a columns, each representing a contestant. Make the entries 1 or 0, representing "pass" and "fail", respectively.

Let  $\mathcal{T}$  denote denote the set of pairs of entries in the same column that are either both 0 or 1. Again, we will count  $\mathcal{T}$  in two different ways.

Counting by rows: Since the ratings of any two judges coincide for at most k contestants, for every two rows, at most k pairs belong in  $\mathcal{T}$ . Since there are  $\binom{b}{2}$  ways to choose two rows, we have  $|\mathcal{T}| \leq k\binom{b}{2} = \frac{kb(b-1)}{2}$ .

Counting by columns: For a particular column, suppose there are p ones and q zeros, then there are  $\binom{p}{2} + \binom{q}{2}$  pairs in  $\mathcal{T}$ . Note that p + q = b is odd, using smoothing, we get

$$\binom{p}{2} + \binom{q}{2} \ge \binom{\frac{b+1}{2}}{2} + \binom{\frac{b-1}{2}}{2} = \frac{(b-1)^2}{4}.$$

Since there are *a* rows, we must have  $|\mathcal{T}| \ge \frac{a(b-1)^2}{4}$ .

Combining the two inequalities, we get  $\frac{a(b-1)^2}{4} \leq |\mathcal{T}| \leq \frac{kb(b-1)}{2}$ . It follows that  $\frac{a(b-1)^2}{4} \leq \frac{kb(b-1)}{2}$  and thus  $\frac{k}{a} \geq \frac{b-1}{2b}$ .

#### Counting with weights

**Problem 4.** (Iran 1999) Suppose that  $C_1, \ldots, C_n$   $(n \ge 2)$  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 S be the set of points that belong to at least two circles. Show that  $|S| \ge n$ .

Let us set up a matrix with n columns, each representing an unit circle, and |S| rows, each representing an intersection point. An entry is 1 if the corresponding point lies on the corresponding circle and 0 otherwise. Since no circle is disjoint from the rest, every column contains at least two 1's as no two circles are tangent. As well, by definition, each row must contain at least two 1's.

Let one focus one 1 in the incidence matrix, say  $a_{ij} = 1$ . Each one on row *i* distinct from  $a_{i,j}$  corresponds to a circle that goes through the point represented by row *i*. Any such circle meets the circle  $C_j$  at exactly two points as no tangency is allowed. So we will associate each one in row *i* distinct from  $a_{i,j}$  with a one from column *j* different from  $a_{i,j}$  that represents the second intersection. Note that no one in column *j* is associated with two different 1's on row *i*, as this would mean that three different unit circles are passing through the same two points, which is impossible. Hence, there is an injection from the 1's in row *i* to the 1's in column *j*.



How can we use this information? Well, this is where the weights come in.

Let us revisit the idea of counting 1's. However, this time, we will assign a "weight" to each 1. For example, if an incidence matrix has three 1's on each row, and we assign a weight of  $\frac{1}{3}$  to each 1, then the sum of all the weights is r, the number of rows. We will see momentarily why this might be useful.

Using the same idea, if we associate each one with a weight, in such a way the weights of all the 1's in each row sum to 1, then the sum of the weights of all the 1's in the matrix equals to r. The following proposition comes from this idea.

**Proposition 3.** Let  $A = (a_{i,j})$  be an  $r \times c$  matrix with row sums  $R_i$ , and column sums  $C_j$ . If  $R_i > 0$  for  $1 \le i \le r$ , then

$$\sum_{i,j} \frac{a_{i,j}}{R_i} = r.$$

Similarly, if  $C_j > 0$  for  $1 \le j \le c$ , then

$$\sum_{i,j} \frac{a_{i,j}}{C_j} = c$$

*Proof.* We have

$$\sum_{i,j} \frac{a_{i,j}}{R_i} = \sum_{i=1}^r \left( \frac{1}{R_i} \sum_{j=1}^c a_{i,j} \right) = \sum_{i=1}^r \left( \frac{1}{R_i} R_i \right) = \sum_{i=1}^r 1 = r.$$

The proof of the second part is done in a similar manner.

The following proposition leads to an application of this idea.

**Proposition 4.** Let  $A = (a_{i,j})$  be a  $r \times c$  (0,1)-matrix with row sums  $R_i$  and column sums  $C_j$  such that  $R_i > 0$  and  $C_j > 0$  for  $1 \le i \le r$  and  $1 \le j \le c$ . If  $C_j \ge R_i$  whenever  $a_{i,j} = 1$ , then  $r \ge c$ .

*Proof.* When  $a_{i,j} = 1$ ,  $R_i \leq C_j$  implies that  $\frac{1}{R_i} \geq \frac{1}{C_j}$ . It follows that  $\frac{a_{i,j}}{R_i} \geq \frac{a_{i,j}}{C_j}$  for all  $1 \leq i \leq r$  and  $1 \leq j \leq c$ . From Proposition 3, we have

$$r = \sum_{i,j} \frac{a_{i,j}}{R_i} \ge \sum_{i,j} \frac{a_{i,j}}{C_j} = c.$$

Back to the problem. Since there is an injective mapping from the 1's in row *i* to the 1's in column *j*, we see that  $C_j \ge R_i$  whenever  $a_{i,j} = 1$ . Therefore, therefore  $r \ge c$ , and the result follows.  $\Box$ .

We will play one more variation on this technique. Sometimes we may not be able to compare  $R_i$  and  $C_j$  when  $a_{i,j} = 1$ , but we may be able to make the comparison when  $a_{i,j} = 0$ . The next proposition is an analogue of Proposition 4.

**Proposition 5.** Let  $A = (a_{i,j})$  be an  $r \times c$  (0,1)-matrix with row sums  $R_i$ , and column sums  $C_j$ . If  $0 < R_i < c$  and  $0 < C_j < r$  for  $1 \le i \le r$  and  $1 \le j \le c$ , and  $C_j \ge R_i$  whenever  $a_{i,j} = 0$ , then  $r \ge c$ .

*Proof.* Suppose on the contrary that r < c. Then  $0 < r - C_j < c - R_i$  whenever  $a_{i,j} = 0$ . Hence,  $\frac{1}{c - R_i} < \frac{1}{r - C_j}$ , which implies

$$\frac{R_i}{c-R_i} < \frac{C_j}{r-C_j}.$$

Let M denote the number of 1's in A, we have

$$M = \sum_{i=1}^{r} R_i = \sum_{i=1}^{r} (c - R_i) \frac{R_i}{c - R_i} = \sum_{i=1}^{r} \left( \sum_{i=1}^{c} (1 - a_{i,j}) \right) \frac{R_i}{c - R_i} = \sum_{i,j} \frac{(1 - a_{i,j})R_i}{c - R_i}$$
$$< \sum_{i,j} \frac{(1 - a_{i,j})C_i}{r - C_j} = \sum_{j=1}^{c} \left( \sum_{i=1}^{r} (1 - a_{i,j}) \right) \frac{C_j}{r - C_j} = \sum_{j=1}^{c} (r - C_j) \frac{C_j}{r - C_j} = M$$

This is clearly impossible. Therefore,  $r \ge c$ .

**Problem 5.** 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  $m \leq n$ .

This problem is a special case of Fisher's inequality. It has a very simple and element proof using linear algebra. However, in this note, we give the combinatorial solution to the problem following the models that we have developed so far.

*Proof.* The result holds trivially if the collection is empty (m = 0) or m = 1. So we may assume that  $m \ge 2$ . It is easy to see that none of the sets  $S_i$  are empty. So assume  $m \ge 2$  and all of the sets are non-empty.

As usual, we consider the incidence matrix A for the collection of sets. The m rows of A correspond to sets and the n columns correspond to the elements, where  $a_{i,j}$  is 1 if element j belongs to set  $S_i$ , and is 0 otherwise.

Now let us show that the hypotheses of Proposition 5 are satisfied. If any row has all 1's, say the first row, then the constraint  $|S_1 \cap S_i| = 1$  for all  $i \neq 1$  forces  $|S_i| = 1$ , which, along with  $|S_i \cap S_j| = 1$ , implies that m = 2, and  $n \geq 2$  because the sets are distinct. If any column has all zeros, then that element belongs to none of the sets and we may simply remove that column. We may do this until every column satisfies  $C_j \geq 1$  because if the result holds for this reduced matrix, it certainly holds for the original A. Finally, if any column has all 1's, say the first column, then  $|S_i \cap S_j| = 1$  implies that no other column may contain two 1's. As well, at most one row may contain a single one (on the first column), and each of the other r - 1rows must have the second one on distinct columns. So the number of columns must be greater than or equal to the number of rows, giving  $m \leq n$  in this case as well. We are now ready to employ Proposition 5.

$$\begin{pmatrix} & \vdots & & \\ & 1 & \to & 1 \\ & \vdots & & \downarrow \\ \cdots & 1 & \cdots & a_{i,j} = 0 & \cdots & 1 & \cdots \\ & \uparrow & & \vdots & & \\ & 1 & \leftarrow & 1 & & \\ & & \vdots & & & \end{pmatrix}$$

Let us consider any  $a_{i,j} = 0$ . By the given condition, for every one on column j, its corresponding subset must intersect with  $A_i$ . So we may correspond each one on  $C_j$  with an one on row i such that the element represented by the one on  $R_i$  also belongs to the subset represented by the one on  $C_j$ . Note that this correspondence is injective, since having two 1's on  $C_j$  both corresponding to the same one in  $R_i$  implies that some two subsets intersect in at least two elements. The injective mapping implies that there must be at least as many 1's on  $i^{\text{th}}$  row as there are on the  $j^{\text{th}}$  column.

Thus  $R_i \ge C_j$  for any  $a_{i,j} = 0$ . It follows from Proposition 5 (with the roles of rows and columns interchanged) that  $m \le n$ .

**Final remarks** Incidence matrices can be very useful for visualizing the combinatorial configuration. However, when writing up a solution, it's usually easier to avoid the incidence matrix and simply stick with set theory notation instead.

#### Problems

- 1. (China 1993) A group of 10 people went to a bookstore. It is known that
  - (1) Everyone bought exactly 3 books;
  - (2) For every two persons, there is at least one book that both of them bought.

What is the least number of people that could have bought the book purchased by the greatest number of people?

- 2. (IMO 2004 Shortlist) 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.

- 3. Let X be a finite set with |X| = n, and let  $A_1, A_2, \ldots, A_m$  be three-element substs of X such that  $|A_i \cap A_j| \leq 1$  for all  $i \neq j$ . Show that there exists a subset A of X with at least  $\lfloor \sqrt{2n} \rfloor$  elements containing none of the  $A_i$ 's.
- 4. Let  $A_1, A_2, \ldots, A_7$  be subsets of  $M = \{1, 2, \ldots, 7\}$ , such that each pair of elements of M belongs to exactly one of the subsets, and  $|A_i| \ge 3$  for each i. Show that  $|A_i \cap A_j| = 1$  for all  $i \ne j$ .
- 5. An organization has n members, and it has n + 1 three-member committees, no two of which have identical membership. Prove that there are two committees that share exactly one member.

- 6. (China TST 1992) Sixteen students took part in a math competition where every problem was a multiple choice question with four choices. After the contest, it is found that any two students had at most one answer in common. Determine the maximum number of questions.
- 7. (China TST 1995) Twenty-one people took a test with 15 true and false questions. It is known that for every two people, there is at least one question that both have answered correctly. Determine the minimum possible number of people that could have correctly answered the question that most number of people are correct on.
- 8. (China 1996) Eight singers participate in an art festival where m songs are performed. Each song is performed by 4 singers, and each pair of singers performs together in the same number of songs. Find the smallest m for which this is possible.
- 9. (Canada 2006) In a rectangular array of nonnegative real numbers with m rows and n columns, each row and each column contains at least one positive element. Moreover, if a row and a column intersect in a positive element, then the sums of their elements are the same. Prove that m = n.
- 10. (Iberoamerican Olympiad 2001) Let X be a set with n elements. Given k > 2 subsets of X, each with at least r elements, show that we can always find two of them whose intersection has at least  $r \frac{nk}{4(k-1)}$  elements.
- 11. (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 any point P of S there are at least k points of S equidistant from P.

Prove that:

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

12. Let  $A_1, A_2, ..., A_k$  be subsets of  $S = \{1, 2, ..., 10\}$  such that

(1) 
$$|A_i| = 5, i = 1, 2, \dots, k;$$

(2)  $|A_i \cap A_j| \le 2, \ 1 \le i < j \le k.$ 

Determine the maximum possible value of k.

- 13. (Burnside's Lemma) Let G be a finite group that acts on the set A. For  $g \in G$ , let Fix(g) denote the number of elements in A that are fixed by g. Show that the number of orbits of G on A is equal to  $\frac{1}{|G|} \sum_{g \in G} Fix(g)$ .
- 14. Let A be a set with |A| = n, and let  $A_1, A_2, \ldots, A_n$  be subsets of A with  $|A_i| \ge 2, 1 \le i \le n$ . Suppose that for each two-element subsets A' of A there is a unique i such that  $A' \subseteq A_i$ . Prove that  $A_i \cap A_j \ne \emptyset$  for any  $i \ne j$ .
- 15. (IMO 2001) Twenty-one girls and twenty-one boys took part in a mathematical competition. It turned out that
  - (a) each contestant solved at most six problems, and
  - (b) 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.

Prove that there is a problem that was solved by at least three girls and at least three boys.

16. (IMO 2005) In a mathematical competition 6 problems were posed to the contestants. Each pair of problems was solved by more than  $\frac{2}{5}$  of the contestants. Nobody solved all 6 problems. Show that there were at least 2 contestants who each solved exactly 5 problems.