Double Counting Using Incidence Matrices

Yufei Zhao

1. Introduction.

Combinatorics problems appear often on mathematics competitions, and they frequently involve scenarios where individuals are associated with organizations, following a set of rules. Here is one such scenario.

Example 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 is equal to the number of subcommittees.

To investigate problems like this, we need a method of representing and visualizing the setup. We employ incidence matrices for this purpose. In our incidence matrices, each row represents an individual, and each column an organization. A matrix entry is set to 1 if the individual corresponding to its row belongs to the organization corresponding to its column; otherwise, the entry is set to 0. (Of course, the roles of rows and columns could be interchanged.) Two possible incidence matrices for Example 1 are shown.

$$\begin{pmatrix} 1&1&1&0&0&0&0\\ 1&0&0&1&1&0&0\\ 1&0&0&0&0&1&1\\ 0&1&0&1&0&1&0\\ 0&1&0&0&1&0&1\\ 0&0&1&1&0&0&1\\ 0&0&1&0&1&1&0 \end{pmatrix} \qquad \begin{pmatrix} 1&1&1&0&0&0&0&0&0\\ 0&0&0&1&1&1&0&0&0\\ 0&0&0&0&0&0&1&1&1\\ 1&0&0&1&0&0&1&0&0\\ 0&1&0&0&1&0&0&1&0\\ 0&0&1&0&0&1&0&0&1\\ 1&0&0&0&1&0&0&0&1\\ 0&1&0&0&0&1&1&0&0\\ 0&0&1&1&0&0&0&1&0 \end{pmatrix}$$

Let us define the notation that we will be using for incidence matrices. Let r and c denote the number of rows and columns, respectively, let M denote the number of 1s, and let R_i and C_j denote the number of 1s in the i^{th} row and j^{th} column, respectively.

In most of our examples, we will look at an incidence matrix from two perspectives—by rows and by columns. This will allow us to obtain either an identity or an inequality that can be used to prove certain properties.

2. Counting the number of 1s.

When presented with an incidence matrix, one might ask, *how many* 1s *are there*? That is, what is the value of M?

If we count the 1s row-by-row, we see that M is the sum of R_i over all rows i. On the other hand, counting the 1s column-by-column yields M as the sum of C_j over all columns j. We have proved the following:

Proposition 1. If A is an incidence matrix with r rows and c columns having row sums R_i, for $i = 1, 2, \ldots, r$ and column sums C_j, for $j = 1, 2, \ldots, c$, then

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

We now apply this proposition to Example 1, where the incidence matrix has $R_i = C_j = 3$ for all i and j. The equation in the proposition yields $3r = 3c$. Thus, $r = c$, which is the desired result.

3. Counting pairs of 1s.

Often a restriction is imposed that applies to every *pair* of organizations (or individuals). For example, it may be that every two organizations share exactly one common member. Such problems can usually be approached by counting pairs of 1s. Specifically, we are interested in the number of pairs of 1s that lie in the same column (or row).

Proposition 2. Let A be an $r \times c$ incidence matrix with column sums C_j. Suppose that, for every two rows, there exist exactly t columns that contain 1s 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 1s that lie in the same column. We count the elements of $\mathcal{T}$ in two different ways.

Counting by rows: For any two rows, there are t pairs of 1s among these rows that belong to $\mathcal{T}$; thus, $|\mathcal{T}| = t\binom{r}{2}$.

Counting by columns: In the j^{th} column, there are C_j 1s and thus $\binom{C_j}{2}$ pairs of 1s. 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. ■

3.1. Inequalities.

Sometimes we are not given enough information to produce a combinatorial identity. Instead, we have to work with inequalities and bounds.

Many incidence-matrix problems are concerned with the existence of a certain subconfiguration. Such problems are often solved by contradiction. Assuming that the opposite result holds, we can count a particular set (for example, the set of all pairs of 1s that belong to the same column) in two different ways, once by rows and once by columns. If we can establish an upper bound in one count and a lower bound in the other count such that the upper bound is less than the lower bound, then a contradiction is reached.

The above idea is illustrated in the following problem, given in the 2002 International Mathematics Competition for University Students [3].

Example 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.

Proof: Assume that the contrary is true; that is, for every two students, there is some problem that neither of them solved. Consider the 6×200 incidence matrix for this configuration, where an entry in the matrix is 1 if the student corresponding to the column did *not* solve the problem corresponding to the row, and is 0 otherwise. The setup is illustrated below.

$$\begin{array}{l} \text{Problem 1} \\ \text{Problem 2} \\ \text{Problem 3} \\ \text{Problem 4} \\ \text{Problem 5} \\ \text{Problem 6} \end{array} \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 0 \end{pmatrix}$$

Let $\mathcal{T}$ denote the set of pairs of 1s that belong to the same row. We now 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 1s among these two columns that belong to the same row. Hence, 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 eighty 1s in each row. Thus, each row contains at most $\binom{80}{2}$ pairs of 1s. Since there are six rows, we have $|\mathcal{T}| \leq 6\binom{80}{2} = 18{,}960$.

The above two inequalities are clearly contradictory. The desired conclusion follows. ∎

3.2. Convexity of $\binom{n}{2}$.

Because we are often interested in counting pairs of 1s, the function $f(n) = \binom{n}{2}$ appears 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.

Lemma 1. Let $a_1, a_2, \ldots, a_n$ be positive integers, and let $s = \sum_{k=1}^{n} a_k$. Then

$$\binom{a_1}{2} + \binom{a_2}{2} + \binom{a_3}{2} + \cdots + \binom{a_n}{2} \geq \frac{s(s-n)}{2n}.$$

Proof: Since $f(x) = \frac{1}{2}x(x-1)$ is convex, we have, by Jensen's Inequality,

$$\frac{f(a_1) + f(a_2) + \cdots + f(a_n)}{n} \geq f\left(\frac{s}{n}\right),$$

from which the result follows easily. ∎

In fact, we can tighten the bound in Lemma 1.

Lemma 2. Let a_1, a_2, $\ldots$, a_n be positive integers, and let $s = \sum_{k=1}^{n} a_k$. If $s = nk + r$, where k and r are integers such that $0 \leq r < n$, then

$$\binom{a_1}{2} + \binom{a_2}{2} + \binom{a_3}{2} + \cdots + \binom{a_n}{2} \;\geq\; r\binom{k+1}{2} + (n-r)\binom{k}{2}.$$

Proof: Without loss of generality, we may assume that $a_1 \geq a_2 \geq \cdots \geq a_n$. Since a_1, a_2, $\ldots$, a_n are integers, the vector $\langle a_1, a_2, \ldots, a_n\rangle$ must majorize the vector $\langle \underbrace{k+1, \ldots, k+1}_{r \text{ times}}, \underbrace{k, \ldots, k}_{n-r \text{ times}} \rangle$. Since $f(n) = \binom{n}{2}$ is convex, we have, by Karamata's Majorization Inequality [1],

$$f(a_1) + f(a_2) + \cdots + f(a_n) \;\geq\; rf(k+1) + (n-r)f(k)\,,$$

and the result follows immediately.

We may also use an optimization argument. We want to minimize the value of $\sum_{i=1}^{n} \binom{a_i}{2}$. Suppose that there exist some indices i and j such that $a_j - a_i > 1$. Then

$$\begin{aligned}
&\binom{a_i}{2} + \binom{a_j}{2} - \binom{a_i+1}{2} - \binom{a_j-1}{2} \\
&\qquad = \frac{a_i(a_i-1)}{2} + \frac{a_j(a_j-1)}{2} - \frac{a_i(a_i+1)}{2} - \frac{(a_j-1)(a_j-2)}{2} \\
&\qquad = a_j - a_i - 1 \;>\; 0\,.
\end{aligned}$$

Thus, by replacing any such (a_i, a_j) by (a_i+1, a_j-1) in the sum $\sum_{i=1}^{n} \binom{a_i}{2}$, we decrease the sum. By repeating this process, we can transform any initial sequence into one where no two terms differ by more than 1. When the process terminates, the sequence must consist of the term $k+1$ repeated r times and the term k repeated $n-r$ times. Since we never increased the sum, the initial sum must be at least as great as the final sum, which is $r\binom{k+1}{2} + (n-r)\binom{k}{2}$. The result follows.

Now we present a problem from the 1998 International Mathematical Olympiad [4] that can be solved using this idea.

Example 3. (IMO 1998) 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 $k/a \geq (b-1)/2b$.

Proof: 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" or "fail", respectively.

Let $\mathcal{T}$ denote the set of pairs of entries that belong to the same column and 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 $|\mathcal{T}| \leq k\binom{b}{2} = \frac{1}{2}\,kb(b-1)$.

Counting by columns: If a column has p 1s and q 0s, then it contributes $\binom{p}{2} + \binom{q}{2}$ pairs to $\mathcal{T}$. Note that $p + q = b$ is odd. By Lemma 2,

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

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

Combining the inequalities for $\mathcal{T}$, we get $\frac{1}{4}\,a(b-1)^2 \leq \frac{1}{2}\,kb(b-1)$. Thus, $k/a \geq (b-1)/2b$. ∎

4. Counting with weights.

Let us revisit the idea of counting 1s. However, this time, we will assign a "weight" to each 1 in such a way that the weights of all the 1s in each row sum to 1. Then the sum of all the weights in the matrix is equal to r. The following proposition comes from this idea.

Proposition 3. Let $A = (a_{ij})$ be an $r \times c$ incidence matrix with row sums R_i and column sums C_j.

(a) If $R_i > 0$ for $1 \leq i \leq r$, then $\sum\limits_{i,j} \dfrac{a_{ij}}{R_i} = r$.

(b) If $C_j > 0$ for $1 \leq j \leq c$, then $\sum\limits_{i,j} \dfrac{a_{ij}}{C_j} = c$.

Proof: To prove (a), we calculate

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

The proof of (b) is similar. ∎

The following proposition leads to an application of this idea.

Proposition 4. Let $A = (a_{ij})$ be an $r \times c$ incidence matrix with row sums R_i and column sums C_j such that $R_i > 0$ and $C_j > 0$ for $1 \leq i \leq r$ and $1 \leq j \leq c$. If $C_j \geq R_i$ whenever $a_{ij} = 1$, then $r \geq c$.

Proof: If $C_j \geq R_i$ whenever $a_{ij} = 1$, then $\frac{a_{ij}}{R_i} \geq \frac{a_{ij}}{C_j}$ for all $1 \leq i \leq r$ and $1 \leq j \leq c$. From Proposition 3, we have

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

This completes the proof. ∎

Note that $r = c$ if $R_i = C_j$ whenever $a_{ij} = 1$. This equality version of Proposition 4 is somewhat stronger than that used in the solution of Example 1. It is worth noting that when this equality case of Proposition 4 is extended to real matrices, it provides an immediate solution to a problem which appeared recently on the Canadian Mathematical Olympiad [6]. The problem is included below, and the solution is left as an exercise.

Example 4. (CMO 2006) In a rectangular array of non-negative 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 $m = n$.

Now, we will use this technique to solve a problem from the Third Round of the 16th Iranian Mathematical Olympiad 1998–1999 [2].

Example 5. (Iran 1998/1999) Suppose that $C_1, \ldots, C_n$ $(n \geq 2)$ are circles of radius 1 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 $|S| \geq n$.

Proof: Let us set up a matrix with n columns, each representing a 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, and since no two circles are tangent, every column contains at least two 1s. As well, by definition, each row contains at least two 1s. We are required to show that $|S| \geq n$. In light of Proposition 4, we will show that $R_i \leq C_j$ whenever $a_{ij} = 1$.

Suppose that $a_{ij} = 1$. Each 1 in row i distinct from a_{ij} 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. We will associate each 1 in row i distinct from a_{ij} with a 1 from column j different from a_{ij} that represents the second intersection. Note that no 1 in column j is associated with two different 1s in row i, as this would mean that three different unit circles are passing through the same two points, which is not possible. Thus, there is an injection from the 1s in row i to the 1s in column j, thereby implying that $R_i \leq C_j$.

$$\begin{pmatrix} & & & \vdots & & & \\ & 1 & \longrightarrow & 1 & & & \\ & \uparrow & & \vdots & & & \\ \cdots & 1 & \cdots & a_{ij}=1 & \cdots & 1 & \cdots \\ & & & \vdots & & \downarrow & \\ & & & 1 & \longleftarrow & 1 & \\ & & & \vdots & & & \end{pmatrix}$$

By Proposition 4, the number of rows is greater than or equal to the number of columns, implying that $|S| \geq n$. ∎

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

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

Proof: Suppose, on the contrary, that $r < c$. Then, whenever $a_{ij} = 0$, we have $0 < r - C_j < c - R_i$, and hence, $\dfrac{R_i}{c - R_i} < \dfrac{C_j}{r - C_j}$. Recalling that M denotes the number of 1s in A, we have

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

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

As an application, the following example is a special case of Fisher's Inequality for block designs [5].

Example 6. 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 $m \le n$.

Proof: The result holds trivially if the collection is empty ($m = 0$) or if $m = 1$. Thus, we may assume that $m \ge 2$. It is easy to see that none of the sets S_i are empty. Hence, we will assume that 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_{ij} is 1 if element j belongs to set S_i, and 0 otherwise.

Now let us show that the hypotheses of Proposition 5 are satisfied. If any row has all 1s, 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 $n \ge 2$ because the sets are distinct. If any column has all 0s, then that element belongs to none of the sets and we may simply remove that column. We may do this until every column satisfies $C_j \ge 1$ because if the result holds for this reduced matrix, it certainly holds for the original matrix A. Finally, if any column has all 1s, say the first column, then $|S_i \cap S_j| = 1$ implies that no other column may contain two 1s. As well, at most one row may contain a single 1 (in the first column), and each of the other $r - 1$ rows must have the second 1 in distinct columns. Hence, the number of columns must be greater than or equal to the number of rows, giving $m \le n$ in this case as well. We are now ready to employ Proposition 5.

Let us consider any $a_{ij} = 0$. By the given condition, for every 1 in column j, its corresponding subset must intersect with A_i. Thus, we may associate each 1 on C_j with a 1 in row i such that the element represented by the 1 on R_i also belongs to the subset represented by the 1 on C_j. Note that this association is injective, since having two 1s on C_j both associated with the same 1 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 1s in the i^{th} row as there are in the j^{th} column.

$$\begin{pmatrix}
 & & & \vdots & & & \\
 & & & 1 & \longrightarrow & 1 & \\
 & & & \vdots & & \downarrow & \\
\cdots & 1 & \cdots & a_{ij} = 0 & \cdots & 1 & \cdots \\
 & \uparrow & & \vdots & & & \\
 & 1 & \longleftarrow & 1 & & & \\
 & & & \vdots & & &
\end{pmatrix}$$

Therefore, $R_i \geq C_j$ for any $a_{ij} = 0$. It follows from Proposition 5 (with the roles of rows and columns interchanged) that $m \leq n$.

5. Final Remarks.

A large number of combinatorial contest problems can be solved by counting in two ways. Incidence matrices help us to visualize a situation and find the set that should be counted. But it is often easier to bypass the use of a matrix in the final presentation of a solution. The direct use of set theory notation, for example, may give a cleaner presentation at the inconvenience of leaving the reader clueless as to where the ideas came from.

References

[1] S. Gueron, Substitutions, Inequalities, and History, *Crux Mathematicorum with Mathematical Mayhem*, 28 (2002), 88–90.

[2] Problem 3, Third Round of Iranian Mathematical Olympiad 1998–1999, *Crux Mathematicorum with Mathematical Mayhem*, 30 (2004), 161.

[3] `http://www.imc-math.org/`

[4] `http://imo.math.ca/Exams/1998imo.html`

[5] `http://mathworld.wolfram.com/FishersBlockDesignInequality.html`

[6] `http://cms.math.ca/Competitions/CMO/solutions/sol_2006.pdf`

Yufei Zhao
3 Ames Street, Box 332
Cambridge, MA 02142, USA
yufeizhao@hotmail.com