Incidence matrices

CJ Quines

May 14, 2022

Warmup

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 the number of subcommittees.

   Sketch: Count the number of (subcommittee, member) pairs such that the member is in the subcommittee. There are n subcommittees with 3 members each, and m members in 3 subcommittees each. Thus $3n = 3m$ so $n = m$.

2. Let $A_1, A_2, \ldots, A_6$ be 4-element subsets of $S = \{1, 2, \ldots, 8\}$. Suppose each element in S is in exactly r of the A_is. Find r.

   Sketch: Count the number of (subset, element) pairs such that the element is in the subset. There are six subsets with four elements each, and eight elements in r subsets each. Thus $6 \times 4 = 8r$ so $r = 3$.

Definitions

Here's the setup. Let $A_1, A_2, \ldots, A_r$ be subsets of $\{1, 2, \ldots, c\}$. (A set of subsets of S is called a *family of subsets* of S.) Consider an $r \times c$ matrix, where the ith row and jth column has 1 if $j \in A_i$ and 0 otherwise. This is the *incidence matrix* of the subsets.

For example: $\{1, 2, 3, 4, 5, 6\}$ has subsets $\{1, 2, 3\}$, $\{2, 4, 6\}$, $\{3, 4, 5\}$. Its incidence matrix is

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

A family of subsets is also known as an *incidence structure*. This is because we can interpret this geometrically, where the elements of the base set are points, and the subsets are "lines" that contain these points. Here's the above incidence matrix, represented geometrically:
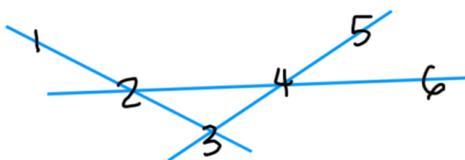
Incidence structures can also be interpreted as edges in a bipartite graph. On one side, your vertices are the elements. On the other side, your vertices are the subsets. An edge joins an element to each subset it belongs to.

The incidence matrix is not a method, just a tool. The method we'll explore is double counting: first **counting by rows**, then **counting by columns**.

Examples

1. (Baltic Way 2001) A set of 8 problems was prepared for an exam. Each student was given 3 of them. No two students received more than one common problem. What is the maximum number of students?

   Sketch: Construct an incidence matrix where columns are problems and rows are students. We want to count the total number of ones.

   Count by rows: If there are s students, there are s rows with 3 each, giving $3s$ in total.

   Count by columns: We want the number of ones in each column, but we don't know that, so instead we bound. Suppose a column had 4 ones. Then the rows that had those ones must have 2 other ones, so there are 8 more ones among those rows. But no two of these ones can be in the same column, otherwise there would be two students who got two or more common problems. There are only 7 remaining columns, so this is impossible. Thus, there are at most 3 ones in each column. There are 8 columns with at most 3 each, giving at most 24 in total.

   Thus $3s \le 24$, so $s \le 8$. It remains to find a construction.

   Remark 1: Please remember how we bounded the number of ones in each column. We won't have another example with this trick, but it will appear several times in the problems.

   Remark 2: Here's a cool idea that *almost* works. Construct an incidence matrix where columns are *pairs of problems* and rows are students. An entry would be 1 if the student was given that pair of problems, and 0 otherwise. Then each pair of problems could've been given to at most one student, as otherwise two students would've had two common problems. This gives us $3s \le \binom{8}{2}$, or $s \le 9$, which isn't a strong enough bound.

2. (China 1996) Eight singers participate in an art festival where several songs are performed. Each song is performed by 4 singers, and each pair of singers performs together in λ songs. Determine the minimum value of λ.

   Sketch 1: Construct an incidence matrix where columns are singers and rows are songs. The trick is to count the number of "row pairs" of ones: that is, the number of pairs of ones that lie in the same row. In this case, rows represent songs, so pairs of ones in the same row are pairs of singers performing a song.

   Count by rows: Suppose there are b rows. Each row has 4 ones, giving 6 row pairs, and $6b$ pairs in total.

   Count by columns: Pick a pair of columns; there are 28 of them. Each pair of columns gives λ row pairs. This gives 28λ pairs in total.

   Thus $6b = 28\lambda$, so $\lambda \ge 3$. It remains to find a construction; this is actually the painful part.

   Sketch 2: We still count row pairs of ones, but this time, we fix a column and count only the pairs that have a one in this column. In other words, we're counting the pairs of singers who sing in the same song, with one of the singers fixed. Suppose this given singer sang r songs.

   Count by rows: Suppose the given singer sang r songs. Each of the r songs gives 3 row pairs, and $3r$ pairs in total.

   Count by columns: Each of the other 7 singers sang λ songs with this singer, giving 7λ pairs in total.

   Thus $3r = 7\lambda$ and $\lambda \ge 3$; it remains to find a construction.

Remark: For these problems, the write-ups are easier to understand if we don't mention incidence matrices (or rows, or columns) at all. They're just a tool to help us find what to double count.

3. (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.

Sketch: Construct an incidence matrix where columns are problems and rows are students. Assume the opposite. That means each pair of students didn't solve at least one problem. In other words, each pair of rows has at least one column where they're both zero! That means we want to count column pairs of zeroes.

Count by rows: Pick a pair of rows; there are $\binom{200}{2} = 19\,900$ of them. Each pair of rows gives at least one column pair, so at least $19\,900$ in total.

Count by columns: Each column has at least 120 ones, so there are at most 80 zeroes in each column. There are six columns, so at most $6\binom{80}{2} = 18\,960$ in total.

Combining the two counts gives a contradiction.

Remark: The way I'd write this solution would be as follows. Suppose the contrary. Let T be the number of tuples (student, student, problem), such that there are two distinct students and a problem that neither one solved. Then [...] so $T \geq 19\,900$. Also [...] so $T \leq 18\,960$. Thus $19\,900 \leq T \leq 18\,960$, contradiction.

4. In a chess club, n people gathered to play chess against each other, as they pleased. No two people played against each other more than once. At the end of the day, it was observed that a total of $3n$ games had been played. Moreover, if we choose any two players, say A and B, there would be at most one other player who had played with both A and B. Prove that $n > 30$.

Sketch: Construct an incidence matrix where columns and rows are people. We count the column pairs of ones.

Count by rows: Pick a pair of rows; there are $\binom{n}{2}$ of them. Each pair of rows gives at most one column pair, as at most one person played with the people represented by these two rows. There are at most $\binom{n}{2}$ column pairs in total.

Count by columns: Suppose the ith column had c_i ones. Because there are a total of $3n$ games, $\sum c_i = 6n$. Applying Jensen's, we get

$$\sum_{i=1}^{n} \binom{c_i}{2} \geq n\binom{\left(\sum c_i\right)/n}{2} = n\binom{6}{2} = 15n.$$

Combining the two counts, $15n \leq \binom{n}{2}$, thus $n \geq 31$ as desired.

Remark: When we can't get good column bounds, we can sometimes assign variables and use Jensen's. Using Jensen's like this is too weak, however, because equality isn't always possible. For example, it's true that $\binom{a}{2} + \binom{b}{2} \geq 2\binom{\frac{a+b}{2}}{2}$. But if $a + b$ is odd, you can get a better lower bound of $\binom{\frac{a+b-1}{2}}{2} + \binom{\frac{a+b+1}{2}}{2}$.

Problems

1. (Balanced block designs) Let $A_1, A_2, \ldots, A_b$ be k-element subsets of $S = \{1, 2, \ldots, v\}$. Suppose each pair of elements in S is in exactly λ of the A_is. Prove that each element in S is contained in exactly r of the A_is, for $r = \lambda(v-1)/(k-1) = bk/v$. **Hint:** 4

2. (USA TST 2005) Let n be an integer greater than 1. For a positive integer m, let $S_m = \{1, 2, \ldots, mn\}$. Suppose there exists a $2n$-element set T such that

   a) each element of T is an m-element subset of S_m,

   b) each pair of elements of T shares at most one common element, and

   c) each element of S_m is contained in exactly two elements of T.

   Determine the maximum possible value of m in terms of n. **Hint:** 9

3. Let $S_1, S_2, \ldots, S_{5n}$ be subsets of $A = \{1, 2, \ldots, 100\}$ such that every element of A appears in exactly $4n$ of the S_is. Prove there exists i, j, k such that $S_i \cup S_j \cup S_k = A$.

4. (China 1994?) Let $A_1, A_2, \ldots, A_k$ be 5-element subsets of $S = \{1, 2, \ldots, 10\}$ such that $|A_i \cap A_j| \leq 2$ for $1 \leq i < j \leq k$. Find the maximum value of k. **Hint:** 5

5. (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. **Hint:** 5

6. (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

$$\frac{k}{a} \geq \frac{b-1}{2b}.$$

7. (USA 2011) Let X be a set with $|X| = 225$. Suppose further there are eleven subsets $A_1, \ldots, A_{11}$ of X such that $|A_i| = 45$ for $1 \leq i \leq 11$ and $|A_i \cap A_j| = 9$ for $1 \leq i < j \leq 11$. Prove that $|A_1 \cup \cdots \cup A_{11}| \geq 165$, and give an example for which equality holds. **Hint:** 9

Harder problems

8. (Singapore 2010) Let n be a positive integer. Find the smallest positive integer k with the property that for any coloring of the squares of a $2n \times k$ chessboard with n colors, there are 2 columns and 2 rows such that the 4 squares in their intersections have the same color. **Hint:** 3

9. (USA 1979) An organization has n members, and it has $n+1$ distinct three-member committees. Prove that there are two committees that share exactly one member. **Hint:** 6

10. (Singapore 2006) Let n be a positive integer. Let $S_1, S_2, \ldots, S_k$ be a collection of $2n$-element subsets of $\{1, 2, \ldots, 4n-1, 4n\}$ so that $S_i \cap S_j$ contains at most n elements for all $1 \leq i < j \leq k$. Show that $k \leq 6^{(n+1)/2}$. **Hint:** 1

11. (USA 1984) A math exam has two papers, where each paper consists of at least one question and both papers have 28 questions altogether. Each pupil attempted 7 questions. Each pair of questions was attempted by exactly two pupils. Show that one pupil attempted either zero or at least four questions in the first paper. **Hint:** 2

12. (China 1993) A group of 10 people went to a bookstore. It is known that everyone bought exactly 3 books, and for every two people, 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? **Hint:** 7

13. (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. **Hint:** 8

14. (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 solved exactly 5 problems.

Hints

1. Use Remark 2 from the first example.

2. How many students are there? How many problems did each solve?

3. The case $n = 4$ is similar to the China TST 1992 problem.

4. This generalizes the second example.

5. For the construction, try to put things into sets greedily, and that'll probably work.

6. Induct on n.

7. First show the answer is at least 4. Then show it can't be 4.

8. Bound triplets of matching problems in rows and columns.

9. For the construction, represent something geometrically, like the figure on page 1.

Sketches

1. We count row pairs of ones, where one of the ones is in a given column. **Count by rows:** If the given column has r ones, each gives $k-1$ pairs along its row. **Count by columns:** If we pick one of the other $v-1$ columns, there are λ rows that contain a pair. Thus $r(k-1) = \lambda(v-1)$. The $bk = vr$ follows from counting all the ones.

    Remark: Note that v, k, and λ determine b and r. We call the family of subsets a 2-balanced design, because the subsets are balanced with respect to the 2-element subsets of S. We want to prove that every 2-balanced design is also 1-balanced. In fact, for every $s < t$, every t-balanced design is also s-balanced; the proof is similar.

2. We count column pairs of ones. **Count by rows:** There are $\binom{2n}{2}$ pairs of rows, and each pair of row has at most one pair. **Count by columns:** Each of the mn columns has exactly two ones. Thus $mn \le \binom{2n}{2}$, so $m \le 2n - 1$. It remains to find a construction; consider $2n$ lines in general position.

3. Assume otherwise; we count column triplets of zeroes. **Count by rows:** There are $\binom{5n}{3}$ triplets of rows, each giving at least one column pair. **Count by columns:** Each column has n zeroes, giving $100\binom{n}{3}$ triplets. Thus $\binom{5n}{3} \le 100\binom{n}{3}$, contradiction after some algebra.

4. We count column pairs of ones. **Count by rows:** There are $\binom{k}{2}$ pairs of rows, each giving at most two column pairs. **Count by columns:** The total number of ones is $5k$, so by Jensen, there are $\sum \binom{c_i}{2} \geq 10\binom{k/2}{2}$ column pairs. Thus $10\binom{k/2}{2} \leq 2\binom{k}{2}$, so $k \leq 6$. It remains to find a construction, e.g. 01234, 01567, 02589, 13689, 24679, 34578.

5. Construct the incidence matrix where columns are problems, rows are students, and entries are the students' answers. We count column pairs of matching answers. **Count by rows:** There are $\binom{16}{2}$ pairs of rows, each giving at most one pair. **Count by columns:** Suppose there are p columns. By Jensen, each column gives at least $4\binom{4}{2}$ pairs. Thus $4p\binom{4}{2} \leq \binom{16}{2}$, so $p \leq 5$. It remains to find a construction; consider families of parallel lines in $\mathbb{F}_4 \times \mathbb{F}_4$.

6. Construct the incidence matrix where columns are contestants and rows are judges. We count column pairs of matching ratings. **Count by rows:** There are $\binom{b}{2}$ pairs of rows, each giving at most k pairs. **Count by columns:** By Jensen, each of the a columns gives at least $\binom{(b-1)/2}{2} + \binom{(b+1)/2}{2} = (b-1)^2/4$ pairs. Combining, we get $a(b-1)^2/4 \leq k\binom{b}{2}$, as desired.

7. We count column pairs of ones. **Count by rows:** There are $\binom{11}{2}$ pairs of rows, each giving 9 column pairs. **Count by columns:** Suppose there are n columns, and that the ith column had c_i ones. There are a total of 45×11 ones. By Jensen, we get $\sum \binom{c_i}{2} \geq n\binom{495/n}{2}$. Combining, we get $9\binom{11}{2} \geq n\binom{495/n}{2}$, or $n \geq 165$. It remains to find a construction; consider 11 planes in general position.

References

Problems were mostly lifted from Loh's Combinatorics of Sets ☑ handout, Shi-Jie's Incidence Matrix ☑ handout, and Zhao's Double Counting ☑ handout.