New Zealand Mathematical Olympiad Committee

The Principle of Inclusion-Exclusion

Chris Tuffley

1 Introduction

In *Basic Counting Principles* [1] you learnt that

$$|A \cup B| = |A| + |B| - |A \cap B|. \tag{1}$$

In words, to count the objects that are in A *or* in B, you should: count the number of objects in A; add it to the number of objects in B; and *subtract* the number of objects in A *and* B, because you're counted them twice. For example, there are $15+10-5 = 20$ numbers between 1 and 60 inclusive that are divisible by four or six, because there are $60/4 = 15$ that are divisible by four; $60/6 = 10$ that are divisible by 6; and $60/12 = 5$ that are divisible by four *and* six.

The *Principle of Inclusion-Exclusion* tells us how to generalise this formula to a union of arbitrarily many sets. The name comes from the fact that we systematically include and then exclude elements until everything has been counted just the right number of times. You can see this going on in equation (1) above: the count *includes* everything in A and everything in B, and then *excludes* everything in $A \cap B$. With more sets there are more intersections to consider, so we have to alternate including and excluding a few more times to get things just right.

Let's see how this works when finding the size of a union of three sets, $A_1 \cup A_2 \cup A_3$. First, we include everything in A_1, A_2 and A_3, to get $|A_1| + |A_2| + |A_3|$ as our first guess at the size of the union. This counts everything in $A_1 \cap A_2$ twice, once in A_1 and once in A_2, so we must exclude elements of $A_1 \cap A_2$ to correct for this, just as above. The same applies to elements of $A_1 \cap A_3$ and $A_2 \cap A_3$, so we subtract $|A_1 \cap A_2| + |A_1 \cap A_3| + |A_2 \cap A_3|$. But now elements of $A_1 \cap A_2 \cap A_3$ have been included three times and excluded three times, so in total they haven't been counted at all! Correcting for this, we finally arrive at the formula

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|.$$

Using this we can find for example that there are

$$\frac{100}{4} + \frac{100}{5} + \left\lfloor \frac{100}{6} \right\rfloor - \frac{100}{20} - \left\lfloor \frac{100}{12} \right\rfloor - \left\lfloor \frac{100}{30} \right\rfloor + \left\lfloor \frac{100}{60} \right\rfloor$$
$$= 25 + 20 + 16 - 5 - 8 - 3 + 1 = 46$$

numbers between 1 and 100 inclusive that are divisible by 4, 5 or 6.

To state the general case, let A be a set, and let $A_1, \dots, A_n$ be n subsets of A. Specifying the "superset" A lets us state the Principle in two forms: one giving the size of the union

$A_1 \cup \cdots \cup A_n$, and a second giving the size of its complement $\overline{A_1 \cup \cdots \cup A_n}$. Of course, knowing one immediately tells us the other; but both forms can be useful, and to my mind the second has a more aesthetic proof. But I'll let you make your own mind up on that one in Exercise 1!

Theorem 1 (The Principle of Inclusion-Exclusion)**.** *The size of the union $A_1 \cup \cdots \cup A_n$ is given by*

$$|A_1 \cup \cdots \cup A_n| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| \\ - \cdots + (-1)^{n-1} |A_1 \cap A_2 \cap \cdots \cap A_n|,$$

and the size of its complement in A is given by

$$\left|\overline{A_1 \cup \cdots \cup A_n}\right| = |A| - \sum_i |A_i| + \sum_{i<j} |A_i \cap A_j| - \sum_{i<j<k} |A_i \cap A_j \cap A_k| \\ + \cdots + (-1)^n |A_1 \cap A_2 \cap \cdots \cap A_n|. \tag{2}$$

Exercise 1. Prove the second form (2) of the Principle of Inclusion-Exclusion as follows:

1. Suppose that $x \in A$ belongs to exactly m of the sets $A_1, \ldots, A_n$. Show that the right-hand side of (2) counts x a total of $\sum_{k=0}^{m} \binom{m}{k}(-1)^k$ times.
2. Recognise the sum above as the binomial expansion of $(1 + (-1))^m$, and conclude that x has been counted a total of zero times if $m \neq 0$, and exactly once if $m = 0$.

2 Examples

Let's look at a couple more examples before I abandon you to solve the problems in Section 3 by yourself.

Example 1. *How many integer solutions are there to the system*

$$x_1 + x_2 + x_3 = 12, \qquad 0 \leq x_i \leq 5?$$

Solution: If you've read *Basic Counting Principles* [1] you'll know that

$$x_1 + x_2 + \cdots + x_k = n$$

has $\binom{n+k-1}{k-1}$ solutions in non-negative integers. So let's let A be the set of integer solutions to the system

$$x_1 + x_2 + x_3 = 12, \qquad 0 \leq x_i;$$

there are $\binom{14}{2} = 91$ of these. We want to throw out the ones with $x_i \geq 6$ for some i, so for $i = 1, 2, 3$ let $A_i = \{(x_1, x_2, x_3) \in A : x_i \geq 6\}$. The question is now asking us for the size of $\overline{A_1 \cup A_2 \cup A_2}$, which is exactly what the second form of the Principle of Inclusion-Exclusion tells us how to find.

The first step is to find the size of the sets A_i. Now, $|A_1|$ is the number of integer solutions to $x_1 + x_2 + x_3 = 12$ with $x_1 \geq 6$ and x_2, x_3 non-negative, which (letting $x_1' = x_1 - 6$) is the same as the number of solutions to $x_1' + x_2 + x_3 = 6$ in non-negative integers. We know from above that there are $\binom{6+2}{2}$ of these, so by symmetry $|A_1| = |A_2| = |A_3| = \binom{8}{2} = 28$.

Next, we need to know the size of the pairwise intersections $A_i \cap A_j$. Of course, the only solution in A satisfying (for example) $x_1, x_2 \geq 6$ is $x_1 = x_2 = 6$, $x_3 = 0$, so these intersections all have size 1; but if the relevant solutions weren't so easy to count directly, we could take a similar approach to above, by letting $x_1' = x_1 - 6$ and $x_2' = x_2 - 6$. We'd then find ourselves counting the non-negative integer solutions to $x_1' + x_2' + x_3 = 0$, and our formula above would give us the answer $\binom{3+0-1}{2} = \binom{2}{2} = 1$.

Lastly, we need to know the size of $A_1 \cap A_2 \cap A_3$. Clearly this set is empty (there are no solutions to our equation with each variable at least six!), so equation (2) finally tells us the answer we want is

$$91 - 3 \times 28 + 3 \times 1 - 0 = 10.$$

□

Example 2 (Derangements)**.** *There are n guests at a Secret Santa party. Each guest brings a present, and the presents are re-distributed randomly so that each guest receives exactly one gift. What's the probability that no-one ends up with the gift they brought?*

Solution: There are $n!$ possible distributions, so we need to count the ones in which no guest ends up with their own gift. This number is known as the number of *derangements* of n objects, and we'll denote it by D_n. We'll think of a derangement as a permutation of $1, \ldots, n$ in which i is in its natural position for no i. The first few values are easily found by hand: check that $D_0 = 1$, $D_1 = 0$, $D_2 = 1$, and $D_3 = 2$.

To find D_n using the Principle of Inclusion-Exclusion, let A be the set of all permutations of $1, \ldots, n$, and let A_i be the set of permutations in which i is in the ith place. We then want to find the size of $\overline{A_1 \cup \cdots \cup A_n}$ using equation (2). To do this, we need to find the size of a k-fold intersection $A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}$. A permutation lying in this set has $i_1, i_2, \ldots, i_k$ in their natural positions and the remaining $n - k$ numbers scrambled arbitrarily, which gives $(n-k)!$ possibilities. Since there are $\binom{n}{k}$ possibilities for the sets involved in the intersection, equation (2) gives

$$\begin{aligned} D_n &= \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)! \\ &= \sum_{k=0}^{n} (-1)^k \frac{n!}{k!(n-k)!} (n-k)! \\ &= n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}. \end{aligned}$$

The probability we want is then $\displaystyle\sum_{k=0}^{n} \frac{(-1)^k}{k!}$. Table 1 calculates this for some small values of n. □

n	probability	approx. decimal value
0	1/1	1
1	0/1	0
2	1/2	0.5
3	2/6	0.333333...
4	9/24	0.375
5	44/120	0.366666...
6	265/270	0.368055...
7	1854/5040	0.367857142
8	14833/40320	0.367881944

Table 1: The probability found in Example 2, for small values of n.

For another approach to derangement numbers, see *Recurrence Relations* [2].

An aside

Table 1 suggests that for large n, the probability that no-one receives the gift they brought is approximately 0.3679. This idea is borne out by considering the formula found above: at each step we are alternately adding or subtracting $1/n!$, and since this quantity is rapidly shrinking to zero, we expect the probability to make smaller and smaller oscillations about some limiting value. In fact, the difference between the limiting value and the last value calculated should never be more than the next term added or subtracted—so the $n = 8$ value in the table must be within $1/9! \approx 0.000003$ of the limit.

The observation in the last paragraph is formalised in a theorem known as the *Alternating Series Test*. This theorem tells us that a sum with the same properties as ours *does* approach a limiting value, but not what that value actually is. However, in our case we can actually do a bit better and find the limit. If you've met logarithms, you may have met the *natural logarithm*, sometimes written ln rather than log. This is the logarithm to base e, where $e \approx 2.71828182845$. It turns out that

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!},$$

so our limiting value is $e^{-1} = 1/e$, or about 0.367879441.

Thus, whether there are ten people at the party or a hundred, there's about a 63.2% chance (nearly two in three!) that *someone* will end up with the gift he or she brought. Of course, the chance that this is actually *you* is smaller at a larger party: the probability the ith guest gets their own gift is $|A_i|/n!$, or $1/n$.

3 Problems

1. (a) Count the *odd* positive integers less than 120 that are *not* divisible by 3, 5 or 7.

(b) Use your answer to (a) to determine the number of primes less than or equal to 120.

2. How many integer solutions are there to the system

$$x_1 + x_2 + x_3 + x_4 = 30, \qquad 0 \le x_i \le 10?$$

3. A bakery sells three types of bagels, namely plain, poppy, and sesame, and it has nine, three and five respectively of each type left. How many ways are there to order a bag of a dozen bagels?

4. In a certain gambling game, a six-sided die is rolled five times; the roller wins if the last roll is the same as one of the previous rolls. What is the probability of winning this game?

   Before you work out the answer, guess a ballpark figure for it. Were you close?

5. (a) How many ways are there to re-arrange the characters of IMOCamp09 so that none of the "words" IMO, Camp and 09 occur as consecutive characters?

   (b) How does the answer change if we must count the re-arrangements of IMOCAMP09 that do not contain IMO, CAMP or 09 as consecutive characters?

6. How many ways are there to arrange the integers $1, 2, 3, \ldots, 10$ so that none of $1, 2, 3, 4, 5$ is in its natural position?

7. How many ways are there to distribute m distinguishable balls among n distinguishable boxes so that each box gets at least one ball?

   ("Distinguishable" means that we can tell for example which ball is which, perhaps because each has been painted a different colour. "Distinguishable" vs. "indistinguishable" can make a big difference in combinatorics!)

8. A class of $2n$ students has been grouped into n pairs. How many ways are there to re-pair the students so that every pair gets broken up?

   Hint: first show that there are $\dfrac{(2k)!}{2^k k!}$ *ways to pair up* $2k$ *students.*

9. (a) How many permutations are there of $1, 2, \ldots, n$, in which i is never followed by $i+1$?

   (b) How does the answer change, if in addition n may not be followed by 1?

10. The *Euler ϕ function* is defined on the positive integers by

$$\phi(n) = |\{k : 1 \le k \le n, \gcd(k, n) = 1\}| .$$

   For example, $\phi(12) = 4$, because there are four positive integers (namely 1, 5, 7 and 11) that are less than or equal to 12 and co-prime to it.

Use the Principle of Inclusion-Exclusion to verify that

$$\phi(n) = n \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right),$$

where $p_1, \ldots, p_k$ are the distinct primes dividing n.

References

[1] Michael Albert. *Basic Counting Principles*. NZ Mathematical Olympiad Committee, January 2009. Available from www.mathsolympiad.org.nz.

[2] Chris Tuffley. *Recurrence Relations*. NZ Mathematical Olympiad Committee, January 2009. Available from www.mathsolympiad.org.nz.

2nd February 2009

www.mathsolympiad.org.nz