31

Inclusion-Exclusion Counting

THE SUM RULE SAYS that **if** A **and** B **are disjoint sets, then** $|A \cup B| = |A| + |B|$. If the sets are not disjoint, then this formula over counts the number of elements in the union of A and B. For example, if $A = \{a, b, c\}$ and $B = \{c, d, e\}$, then

$$|A \cup B| = |\{a, b, c\} \cup \{c, d, e\}| = |\{a, b, c, d, e\}| = 5.$$

So, we see that $|A \cup B| \neq 3 + 3 = |A| + |B|$.

31.1 Inclusion-Exclusion principle

The correct way to count the number of elements in $|A \cup B|$ when A and B might not be disjoint is via the **inclusion-exclusion** formula. To derive this formula, notice that $A \cup B = (A - B) \cup B$, and that the sets $A - B$ and B are disjoint. So we can apply the sum rule to conclude

$$|A \cup B| = |(A - B) \cup B| = |A - B| + |B|.$$

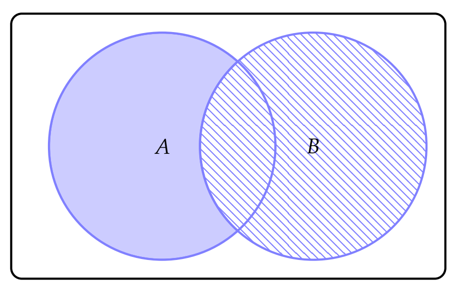

Figure 31.1: $A \cup B = (A - B) \cup B$

Next, notice that $A = (A - B) \cup (A \cap B)$, and the two sets on the right are disjoint. So, using the sum rule, we get

$$|A| = |(A - B) \cup (A \cap B)| = |A - B| + |A \cap B|,$$

which we can rearrange as

$$|A - B| = |A| - |A \cap B|.$$

So, replacing $|A - B|$ by $|A| - |A \cap B|$ in the formula $|A \cup B| = |(A - B) \cup B| = |A - B| + |B|$, we end up with the **inclusion-exclusion** formula:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

In words, to count the number of items in the union of two sets, include one for everything in the first set, and include one for everything in the second set, then exclude one for each element in the overlap of the two sets (since those elements will have been counted twice).

Example 31.1. *How many students are there in a discrete math class if 15 students are computer science majors, 7 are math majors, and 3 are double majors in math and computer science?*

Solution. *Let C denote the subset of computer science majors in the class, and M denote the math majors. Then* $|C| = 15$, $|M| = 7$ *and* $|C \cap M| = 3$. *So by the principle of inclusion-exclusion there are* $|C| + |M| - |C \cap M| = 15 + 7 - 3 = 19$ *students in the class.* ♣

Example 31.2. *How many integers between* 1 *and* 1000 *are divisible by either* 7 *or* 11*?*

Solution. *Let S denote the set of integers between* 1 *and* 1000 *divisible by* 7*, and E denote the set of integers between* 1 *and* 1000 *divisible by* 11*. We need to count the number of integers in* $S \cup E$*. By the principle of inclusion-exclusion, we have*

$$\begin{aligned}|S \cup E| = |S| + |E| - |S \cap E| &= \left\lfloor \frac{1000}{7} \right\rfloor + \left\lfloor \frac{1000}{11} \right\rfloor - \left\lfloor \frac{1000}{77} \right\rfloor \\ &= 142 + 90 - 12 = 120.\end{aligned}$$

♣

31.2 *Extended inclusion-exclustion principle*

The inclusion-exclusion principle can be extended to the problem of counting the number of elements in the union of three sets. The trick is the think of the union of three sets as the union of two sets. It goes as follows:

$$\begin{aligned}
|A \cup B \cup C| &= |(A \cup B) \cup C| \\
&= |A \cup B| + |C| - |(A \cup B) \cap C| \\
&= |A| + |B| + |C| - |A \cap B| - |(A \cup B) \cap C| \\
&= |A| + |B| + |C| - |A \cap B| - |(A \cap C) \cup (B \cap C)| \\
&= |A| + |B| + |C| - |A \cap B| - |(A \cap C) \cup (B \cap C)| \\
&= |A| + |B| + |C| - |A \cap B| - (|A \cap C| + |B \cap C| - |(A \cap C) \cap (B \cap C)|) \\
&= |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.
\end{aligned}$$

This might more appropriately be named the inclusion-exclusion-inclusion formula, but nobody calls it that. In words, the formula says that to count the number of elements in the union of three sets, first, include everything in each set, then exclude everything in the overlap of each pair of sets, and finally, re-include everything in the overlap of all three sets.

Example 31.3. *How many integers between* 1 *and* 1000 *are divisible by at least one of* 7, 9, *and* 11*?*

Solution. *Let S denote the set of integers between* 1 *and* 1000 *divisible by* 7, *let N denote the set of integers between* 1 *and* 1000 *divisible by* 9, *and E denote the set of integers between* 1 *and* 1000 *divisible by* 11. *We need to count the number of integers in $S \cup N \cup E$. By the principle of inclusion-exclusion,*

$$\begin{aligned}
|S \cup N \cup E| &= |S| + |N| + |E| - |S \cap N| - |S \cap E| - |N \cap E| + |S \cap N \cap E| \\
&= \left\lfloor \frac{1000}{7} \right\rfloor + \left\lfloor \frac{1000}{9} \right\rfloor + \left\lfloor \frac{1000}{11} \right\rfloor - \left\lfloor \frac{1000}{63} \right\rfloor - \left\lfloor \frac{1000}{77} \right\rfloor - \left\lfloor \frac{1000}{99} \right\rfloor + \left\lfloor \frac{1000}{693} \right\rfloor \\
&= 142 + 111 + 90 - 15 - 12 - 10 + 1 = 307.
\end{aligned}$$

There are similar inclusion-exclusion formulas for the union of four, five, six, $\cdots$ sets. The formulas can be proved by induction with the inductive step using the trick we used above to go from two sets to three. However, there is a much neater way to prove the formula based on the Binomial Theorem.

Theorem 31.4. *Given finite sets $A_1, A_2, ..., A_n$*

$$\left|\bigcup_{k=1}^{n} A_k\right| = \sum_{k=1}^{n}|A_k| - \sum_{1\le j<k\le n}|A_j \cap A_k| + \cdots + (-1)^{n-1}\left|\bigcap_{k=1}^{n} A_k\right|.$$

Proof. *Suppose $x \in \bigcup_{k=1}^{n} A_k$. We need to show that x is counted exactly once by the right-hand side of the promised formula. Say $x \in A_i$ for exactly p of the sets A_i, where $1 \le p \le n$.*

The key to the proof is being able to count the number of intersections in each summation on the right-hand side of the offered formula that contain x since we will account for x once for each such term. The number of such terms in the first sum is $n = \binom{p}{1}$, the number in the second term is $\binom{p}{2}$, and, in general, the number of terms in the j^{th} sum will be $\binom{p}{j}$ provided $j \le p$. If $j > p$ then x will not be any of the intersections of j of the sets, and so will not contribute any more to the right side of the formula.

So the total number of times x is accounted for on the right hand side is

$$\begin{aligned}\binom{p}{1} - \binom{p}{2} - ... + (-1)^{p-1}\binom{p}{p}&\\ &= 1 - \left(\binom{p}{0} - \binom{p}{1} + \binom{p}{2} - ... + (-1)^{p}\binom{p}{p}\right)\\ &= 1 - (1-1)^p = 1.\end{aligned}$$

Just as we hoped. ♣

Example 31.5. *How many students are in a calculus class if* 14 *are math majors,* 22 *are computer science majors,* 15 *are engineering majors, and* 13 *are chemistry majors, if* 5 *students are double majoring in math and computer science,* 3 *students are double majoring in chemistry and engineering,* 10 *are double majoring in computer science and engineering,* 4 *are double majoring in chemistry and computer science, none are double majoring in math and engineering and none are double majoring in math and chemistry, and no student has more than two majors?*

Solution. *Let A_1 denote the math majors, A_2 denote the computer science majors, A_3 denote the engineering majors, and A_4 the chemistry majors. Then the information given is*

$$|A_1| = 14, \quad |A_2| = 22, \quad |A_3| = 15, \quad |A_4| = 13,$$
$$|A_1 \cap A_2| = 5, \quad |A_1 \cap A_3| = 0, \quad |A_1 \cap A_4| = 0,$$
$$|A_2 \cap A_3| = 10, \quad |A_2 \cap A_4| = 4, \quad |A_3 \cap A_4| = 3,$$
$$|A_1 \cap A_2 \cap A_3| = 0, \quad |A_1 \cap A_2 \cap A_4| = 0,$$
$$|A_1 \cap A_3 \cap A_4| = 0, \quad |A_2 \cap A_3 \cap A_4| = 0,$$

and

$$|A_1 \cap A_2 \cap A_3 \cap A_4| = 0.$$

So, by inclusion-exclusion, the number of students in the class is

$$14 + 22 + 15 + 13 - 5 - 10 - 4 - 3 = 42.$$

Example 31.6. *How many ternary strings (using* 0*'s,* 1*'s and* 2*'s) of length* 8 *either start with a* 1*, end with two* 0*'s or have* 4*th and* 5*th positions* 12*, respectively?*

Solution. *Let A_1 denote the set of ternary strings of length* 8 *which start with a* 1*, A_2 denote the set of ternary strings of length* 8 *which end with two* 0*'s, and A_3 denote the set of ternary strings of length* 8 *which have* 4*th and* 5*th positions* 12*. By inclusion-exclusion, the answer is* $3^7 + 3^6 + 3^6 - 3^5 - 3^5 - 3^4 + 3^3$.

31.3 Inclusion-exclusion with the Good=Total-Bad trick

The inclusion-exclusion formula is often used along with the *Good=Total-Bad* trick.

Example 31.7. *How many integers between* 1 *and* 1000 *are divisible by none of* 7, 9, *and* 11*?*

Solution. *There are* 1000 *numbers between* 1 *and* 1000 *(assuming* 1 *and* 1000 *are included). As counted before, there are* 307 *of those that are divisi-*

ble by at least one of 7, 9, *and* 11. *That means there are* $1000 - 307 = 693$ *that are divisible by none of* 7, 9, *or* 11.

31.4 Exercises

Exercise 31.1. *At a certain college no student is allowed more than two majors. How many students are in the college if there are* 70 *math majors,* 160 *chemistry majors,* 230 *biology majors,* 56 *geology majors,* 24 *physics majors,* 35 *anthropology majors,* 12 *double math-physics majors,* 10 *double math-chemistry majors,* 4 *double biology-math majors,* 53 *double biology-chemistry majors,* 5 *double biology-anthropology majors, and no other double majors?*

Exercise 31.2. *How many bit strings of length 15 start with the string* 1111*, end with the string* 1000 *or have* 4^{th} *through* 7^{th} *bits* 1010*?*

Exercise 31.3. *How many positive integers between* 1000 *and* 9999 *inclusive are not divisible by any of* $4, 10$ *or* 25 *(careful!)?*

Exercise 31.4. *How many permutations of the digits* $1, 2, 3, 4, 5$*, have at least one digit in its own spot? In other words, a* 1 *in the first spot, or a* 2 *in the second, etc. For example,* 35241 *is OK since it has a* 4 *in the fourth spot, and* 14235 *is OK, since it has a* 1 *in the first spot (and also a* 5 *in the fifth spot). But* 31452 *is no good. Hint: Let* A_1 *be the set of permutations that have* 1 *in the first spot, let* A_2 *be the set of permutations that* 2 *in the second spot, and so on.*

Exercise 31.5. *How many permutations of the digits* $1, 2, 3, 4, 5$ *have no digit in its own spot?*

31.5 Problems

Problem 31.1. *The membership of a language club consists of seven people who speak only English, eight speak only French, five speak only Spanish, seven speak only English and Spanish, two speak only French and Spanish, there are none who speak only English and French, and there are four who speak all three languages. How many members are in the club?*

Problem 31.2. *How many integers between* 1 *and* 10000 *(inclusive) are divisible by at least one of* 9*,* 10*, or* 11*?*

Problem 31.3. *Suppose* p, q *are two different primes. How many integers between* 1 *and the product* pq *are relatively prime to* pq *(or, same thing, how* *many are divisible by neither of p and q)? (The correct answer will factor neatly.)*

Problem 31.4. *Suppose p, q, r are three different primes. How many integers between* 1 *and the product pqr are relatively prime to pqr (or, same thing, how many are divisible by none of p, q and r)? (The correct answer will factor neatly.)*

Problem 31.5. *Suppose p, q, r, s are four different primes. How many integers between* 1 *and the product pqrs are relatively prime to pqrs (or, same thing, how many are divisible by none of p, q, r and s)? (The correct answer will factor neatly.)*

Problem 31.6. *Based on the results of the previous three problems, can you guess the neat formula for five, six, seven, and so on, different primes?*

Problem 31.7. *Of the words of length ten using the alphabet $\Sigma = \{a, b, c\}$, how many either begin abc or end cba or have cccccc as the middle six letters?*

Problem 31.8. *There are* 6! *permutations of the numbers* $1, 2, 3, 4, 5, 6$*. In some of these there is a run of three (or more) consecutive numbers that increase (left to right) such as* 514632 *which has the increasing run* 146*. Others do not have any increasing runs of length three such as (a cheap example)* 654321 *and (not quite as cheap)* 615243*. How many of the* 6! *permutations contain no increasing runs of length three (or more)? (Hint: runs of length three can start with the first, second, third, or fourth spot in the permutation.)*

32
The Pigeonhole Principle

The pigeonhole principle, like the sum and product rules, is another one of those absolutely obvious counting facts. The statement is simple: If $n+1$ objects are divided into n piles (some piles can be empty), then at least one pile must have two or more objects in it. Or, more colorfully, if $n+1$ pigeons land in n pigeonholes, then at least one pigeonhole has two or more pigeons. What could be more obvious? The pigeonhole principle is used to show that no matter how a certain task is carried out, some specific result must always happen.

As a simple example, suppose we have a drawer containing ten identical black socks and ten identical white socks. How many socks do we need to select to be sure we have a matching pair? The answer is three. Think of the pigeonholes as the colors black and white, and as each sock is selected put it in the pigeonhole of its color. After we have placed the third sock, one of the two pigeonholes must have at least two socks in it, and we will have a matching pair. Of course, we may have been lucky and had a pair after picking the second sock, but the pigeonhole principle *guarantees* that with the third sock we will have a pair.

As another example, suppose license plates are made consisting of four digits followed by two letters. Are there enough license plates for a state with seven million cars? No, since there are only $10^4 \cdot 26^2 = 6760000$ possible license plates, and so, by the pigeonhole principle, at least two of the seven million plates assigned would have to be the same.

32.1 General pigeonhole principle

A slightly fancier version of the pigeonhole principle says that if N objects are distributed in k piles, then there must be a least one pile with $\left\lceil \frac{N}{k} \right\rceil$ objects in it.

That formula looks impressive, but actually is easy to understand. For example, if there are 52 people in a room, we can be absolutely certain that there are at least eight born on the same day of the week. Think of it this way: with 49 people, it would be possible to have seven born on each of the seven days of the week. But when the 50^{th} one is reached, it must boost one day up to an eighth person. That is really about all there is to it. The general proof of the fancy pigeonhole principle uses this same sort of reasoning. It is a proof by contradiction, and goes as follows:

Avoidance principle: how long can we go before our hand is forced?

Theorem 32.1 (Pigeonhole Principle)**.** *If N objects are distributed in k piles, then there must be a least one pile with $\left\lceil \frac{N}{k} \right\rceil$ objects in it.*

Proof. *Suppose we have N objects distributed in k piles, and suppose that every pile has fewer than $\left\lceil \frac{N}{k} \right\rceil$ objects in it. That means that the piles each contain $\left\lceil \frac{N}{k} \right\rceil - 1$ or fewer objects. We will use the fact that $\left\lceil \frac{N}{k} \right\rceil < \frac{N}{k} + 1$ to complete the proof. The total number of object will be at most $k\left(\left\lceil \frac{N}{k} \right\rceil - 1\right) < k\left(\left(\frac{N}{k} + 1\right) - 1\right) = N$. That is a contradiction since we know there is a total of N objects in the k piles.* ♣

32.2 Examples

Even though the pigeonhole principle sounds very simple, clever applications of it can produce totally unexpected results.

Example 32.2. *Five misanthropes move to a perfectly square deserted island that measures two kilometers on a side. Of course, being misanthropes, they want to live as far from each other as possible. Show that, no matter where they build on the island, some two will be no more than $\sqrt{2}$ kilometers of each other.*

Solution. *Divide the island into four one kilometer by one kilometer squares by drawing lines joining the midpoints of opposite sides. Since there are five* *people and four squares, the pigeonhole principle guarantees there will be two people living in one of those four squares. But people in one of those squares cannot be further apart than the length of the diagonal of the square which is, according to Pythagoras,* $\sqrt{2}$. ♣

Example 32.3. *For any positive integer* n*, there is a positive multiple of* n *made up of a number of* 1*'s followed by a number of* 0*'s. For example, for* $n = 1084$*, we see* $1084 \cdot 1025 = 1111100$.

Solution. *Consider the* $n + 1$ *integers* $1, 11, 111, \cdots, 11 \cdots 1$*, where the last one consists of* 1 *repeated* $n + 1$ *times. Some two of these must be the same modulo* n*, and so* n *will divide the difference of some two of them. But the difference of two of those numbers is of the required type.* ♣

Example 32.4. *Bill has* 20 *days to prepare his tiddledywinks title defense. He has decided to practice at least one hour every day. But, to avoid burn-out, he will not practice more than a total of* 30 *hours. Show there is a sequence of consecutive days during which he practices exactly* 9 *hours.*

Solution. *For* $j = 1, 2, \cdots 20$*, let* t_j = *the total number of hours Bill practices up to and including day* j*. Since he practices at least one hour every day, and the total number of hours is no more than* 30*, we see*

$$0 < t_1 < t_2 < \cdots < t_{20} \leq 30.$$

Adding 9 *to each term we get*

$$9 < t_1 + 9 < t_2 + 9 < \cdots < t_{20} + 9 \leq 39.$$

So we have 40 *integers* $t_1, t_2, \cdots, t_{20}, t_1 + 9, \cdots t_{20} + 9$*, all between* 1 *and* 39*. By the pigeonhole principle, some two must be equal, and the only way that can happen is for* $t_i = t_j + 9$ *for some* i *and* j*. It follows that* $t_i - t_j = 9$*, and since the difference* $t_i - t_j$ *is the the total number of hours Bill practiced from day* $j + 1$ *to day* i*, that shows there is a sequence of consecutive days during which he practiced exactly* 9 *hours.* ♣

32.3 Exercises

Exercise 32.1. *Show that in any group of eight people, at least two were born on the same day of the week.*

Exercise 32.2. *Show that in any group of* 100 *people, at least* 15 *were born on the same day of the week.*

Exercise 32.3. *How many cards must be selected from a deck to be sure that at least six of the selected cards have the same suit?*

Exercise 32.4. *Show that in any set of* n *integers, where* $n \geq 2$*, there must be a pair with a difference that is a multiple of* $n - 1$*.*

Exercise 32.5. *Al has* 75 *days to master discrete mathematics. He decides to study at least one hour every day, but no more than a total of* 125 *hours. Show there must be a sequence of consecutive days during which he studies exactly* 24 *hours.*

Exercise 32.6. *Show that in any set of* 217 *integers, there must be a pair with a difference that is a multiple of* 216*.*

32.4 Problems

Problem 32.1. *Show that in a town with population* $18,000$*, there must be at least two people with the same three initials.*

Problem 32.2. *What is the smallest town population that will guarantee there will be at least two people with the same three initials?*

Problem 32.3. *What is the smallest town population that will guarantee there will be at least five people with the same three initials?*

Problem 32.4. *How many cards have to be selected from a* 52 *card deck to be sure there will be two cards of the same suit?*

Problem 32.5. *How many cards have to be selected from a* 52 *card deck to be sure there will be two cards of the same rank?*

Problem 32.6. *Five misanthropes buy a six mile by twelve mile rectangular plot in the arctic. Show that no matter where they build their igloos, there will be at least two people that are no more than five miles apart. (You can assume the ice sheet they buy is perfectly flat.)*

Problem 32.7. *In any list of n integers, there will be a chunk of consecutive entries from the list that add up to a multiple of n. For example: in the list $-8, 4, 22, -11, 7$, we have $4 + 22 - 11 = 15$ is a multiple of 5.*

Problem 32.8. *Suppose $a_1, a_2, a_3 \ldots, a_{99}$ is a permutation of $1, 2, \ldots, 99$. Show that the product*

$$(a_1 + 1)(a_2 + 2)(a_3 + 3) \ldots (a_{99} + 99)$$

is even.

Problem 32.9. *In a rematch, Bill has 30 days to train for a new defense of his tiddledywinks title. He plans to practice at least one hour every day, but no more than 45 hours total. Show there is a sequence of consecutive days during which he practices exactly 14 hours.*