INCLUSION–EXCLUSION
(COUNTING WITH VENN DIAGRAMS)

If a school has 100 students with 50 students taking French, 40 students taking Latin, and 20 students taking both languages, how many students take no language?

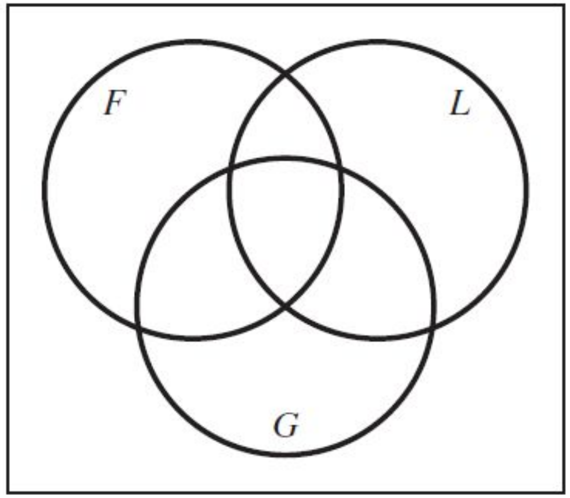
$$N(\overline{F} \cap \overline{L}) = N - N(F \cup L) = N - N(F) - N(L) + N(F \cap L) \tag{2}$$

Formula (2) is the 2-set version of the general n-set inclusion–exclusion formula that we present in the next section. It is called an inclusion–exclusion formula because first we include the whole set, N; then we exclude (subtract) the single sets F and L; and then we include (add) the 2-set intersection $F \cap L$. With more sets, this alternating inclusion and exclusion process will continue several rounds.

How many arrangements of the digits 0, 1, 2, . . . , 9 are there in which the first digit is greater than 1 and the last digit is less than 8?

For the inclusion–exclusion formula, we need to define sets that represent the complements of the original constraints we are given.

What would be $N(\overline{F} \cap \overline{L} \cap \overline{G})$



$$N(\overline{F} \cap \overline{L} \cap \overline{G}) = N - [N(F) + N(L) + N(G)] + [N(F \cap L) + N(L \cap G) + N(F \cap G)] - N(F \cap L \cap G) \quad (4)$$

For general sets A_1, A_2, A_3, we rewrite (4) as

$$N(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3) = N - \sum_i N(A_i) + \sum_{ij} N(A_i \cap A_j) - N(A_1 \cap A_2 \cap A_3) \quad (5)$$

How many n-digit ternary (0, 1, 2) sequences are there with at least one 0, at least one 1, and at least one 2? How many n-digit ternary sequences with at least one void (missing digit)?

Now we turn to the second part of this problem involving n-digit ternary sequences with at least one void.

The phrase "at least" is used in a very different way here than it was used in the first part. At least one void means a void of the digit 0 or a void of the digit 1 or a void of digit 2.

In terms of the A_i defined above, we want to count the union of the A_i's—namely, $N(A_0 \cup A_1 \cup A_2)$.

$N(A_0 \cup A_1 \cup A_2) = N$ $N(\overline{A}_0 \cap \overline{A}_1 \cap \overline{A}_2) = 3^n - (3^n - 3 \times 2^n + 3)$

$$= 3 \times 2^n - 3 \blacksquare$$

Here is an example of a counting problem that cannot be solved by methods developed in the three previous chapters.

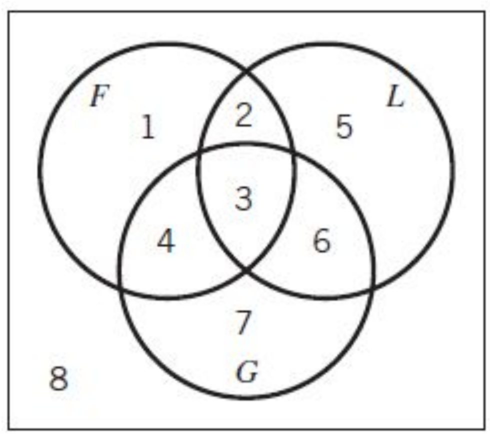
How many positive integers ≤70 are relatively prime to 70?
(“relatively prime to 70” means “have no common divisors with 70.”)

Let U be the set of integers between 1 and 70.
The prime divisors of 70 are 2, 5, and 7. We want to count the number of integers ≤70 that do not have 2 or 5 or 7 as divisors.
What would be A_1 A_2 A_3
Let A_1 be the set of integers in U that are evenly divisible by 2, or equivalently, integers in U that are multiples of 2; A_2 be integers evenly divisible by 5; and A_3 be integers evenly divisible by 7.
Then the number of positive integers ≤70 that are relatively prime to 70 equals $N(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3)$

Suppose there are 100 students in a school and there are 40 students taking each language, French, Latin, and German. Twenty students are taking only French, 20 only Latin, and 15 only German. In addition, 10 students are taking French and Latin. How many students are taking all three languages? No language?

Here Inclusion-exclusion formula cannot be applied directly.

We draw the Venn diagram for this problem and number each region as shown in Figure below.
Let N_i denote the number of students in region i, for i = 1, 2, . . . , 8.



Given

$N_1 = 20$, $N_5 = 20$, $N_7 = 15$, $N_2 + N_3 = 10$.

The set of students taking French is F, which consists of regions 1, 2, 3, and 4. So N_4?

$40 = N(F) = N_1 + (N_2 + N_3) + N_4 = 20 + (10) + N_4$, or $N_4 = 10$

Similarly, set L consists of regions 2, 3, 5, and 6, and so N_6?

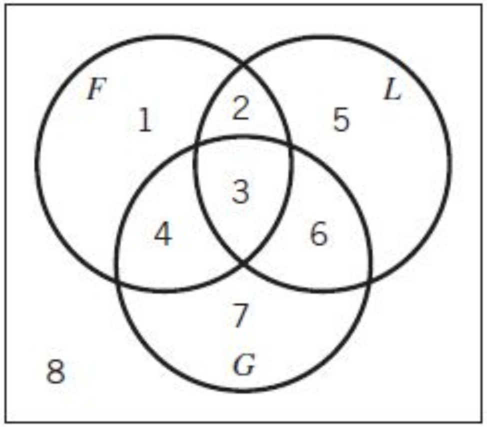
$40 = N(L) = (N_2 + N_3) + N_5 + N_6 = (10) + 20 + N_6$, implying $N_6 = 10$.

Since G consists of regions 3, 4, 6, and 7, and since we were given $N_7 = 15$ and have just found that $N_4 = N_6 = 10$, then $40 = N(G) = N_3 + N_4 + N_6 + N_7 = N_3 + 10 + 10 + 15$ or $N_3 = 5$.

But region 3 is the subset $F \cap L \cap G$ of students taking all 3 languages.

Thus, there are five trilingual students.

Now $N_2 + N_3 = 10$ and N_3 was found to be 5; thus $N_2 = 5$.



$$N_8 = N(\overline{F \cap L \cap G}) \text{ is the number of students taking no language.}$$

Since all regions total to N, then

$$\begin{aligned} N_8 = N - \sum_{i=1}^{7} N_i &= 100 - (20 + 5 + 5 + 10 + 20 + 10 + 15) \\ &= 100 - 85 = 15 \quad \blacksquare \end{aligned}$$