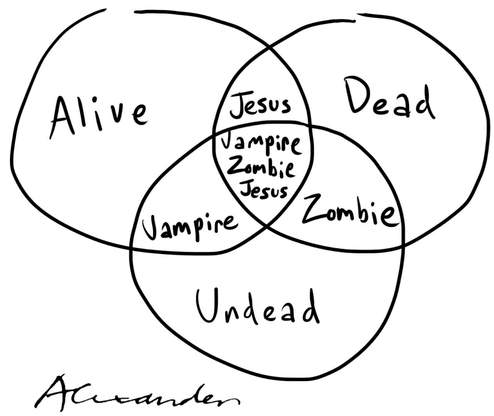
LESSON 5 : INTRO TO PRINCIPLE OF INCLUSION & EXCLUSION (PIE)



- PIE allows us to calculate the number of objects that are members of **at least** one group when there are many groups of interest.

5.1: PRIMER: WHAT IS A SET?

A **set** is just a collection of objects. A set can have numbers, letters, or even colors in it. There can be any non-negative integral *number of elements* in a set (after all, you can't have -3, π, $2+4i$, or $^5/_8$ths of an element in a set). Some examples of sets you may see in math problems are the set of all even numbers less than 100, the set of all families in a town with dogs, and the set of students who are passing Mr. Newton's class.

Most of the time, in math, we only look at sets of numbers. If S is the set of even numbers from 1 to 10, we denote it as $S = \{2, 4, 6, 8, 10\}$.

The order of items in a set do not matter, so sets like $\{1, 2, 3\}$, $\{3, 2, 1\}$ and $\{2, 1, 3\}$ are all the same set. In addition, an element can only appear in a set one time, it is not possible to have an element appear twice or more.

We denote the number of elements in a set S, aka its size, or more formally its **cardinality** as $|S|$.

- Ex: $|\{1, 2, 4, 8\}| = 4$.

The **union** of two sets is the set that contains all the elements from both the sets. Remember that because elements can not appear in a set twice, the union of two sets will not have any duplicate elements. The union of sets A and B is noted as $A \cup B$.

- Ex: $\{1, 2, 3, 5, 10\} \cup \{2, 3, 6, 7, 10, 11\} = \{1, 2, 3, 5, 6, 7, 10, 11\}$

The **intersection** of two sets contains all the elements they share in common. The intersection of sets A and B is noted as $A \cap B$.

- Ex: $\{1,2,3,5,10\} \cap \{2,3,6,7,10,11\} = \{2,3\}$.

5.2: TWO SETS

Deriving PIE for Two Sets:

If we have two sets, A and B. If we want to count the number of elements in $A \cup B$ (that is, $|A \cup B|$):

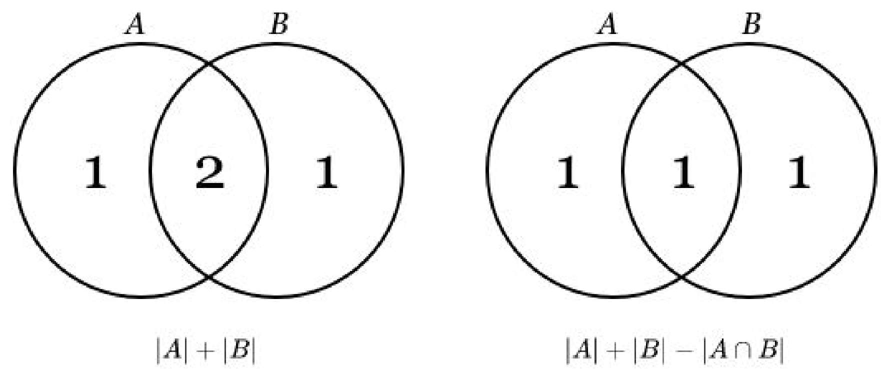
- we can start by saying the size is at most $|A| + |B|$, which means "the number of elements in A" plus "the number of elements in B".
- But, if you notice, this double counts the elements that are contained in both A and B, or $A \cap B$. (assuming that these events can occur concurrently - not mutually exclusive)
- To correct for this, we subtract $|A \cap B|$.

So, we end up with the following formula:
$|A \cup B| = |A| + |B| - |A \cap B|$

This formula ensures that for each of the following cases, an element is only counted once:

1. Element is in neither A nor B
   - not counted at all
2. Element is in A but not B
   - counted only in term 1
3. Element is in B but not A
   - counted only in term 2
4. Element is in both A and B
   - added twice in terms 1 and 2, subtracted once in term 3

This graphic is really helpful for understanding PIE with two sets:



The number in each subset represents how many times that subset is counted.
Note that by adding the elements in sets A and B, their intersection is counted twice.

Note that this formula holds true only when there are two sets. In general, as the number of sets grow, the formula becomes more complicated because we have to account for many intersections to avoid overcounting.

Ex 1: How many numbers from 1 to 100 are either divisible by 3 or 7 (or both)?

Firstly, what are our sets?

A is the set of numbers between 1 and 100 divisible by 3. We know that $|A| = \lfloor \frac{100}{3} \rfloor = \lfloor 33.\overline{33} \rfloor = 33$.

B is the set of numbers between 1 and 100 divisible by 7. We know that $|B| = \lfloor \frac{100}{7} \rfloor = \lfloor 14.\overline{285714} \rfloor = 14$.

$A \cup B$ is the set of numbers between 1 and 100 divisible by either 3 or 7. This is what we are trying to find.

$A \cap B$ is the set of numbers between 1 and 100 divisible by 3 and 7, i.e. divisible by 21. We know that $|A \cap B| = \lfloor \frac{100}{21} \rfloor = \lfloor 4.76... \rfloor = 4$.

So by our formula, $\boxed{|A \cup B| = |A| + |B| - |A \cap B| = 43}$.

5.3: THREE SETS

Deriving PIE for three sets:

We've already examined PIE for two sets. This case follows the same logic, but there is more room for overcounting! Let's say we have 3 sets, A, B, and C. If we want to find the number of elements in all three sets, or $|A \cup B \cup C|$:

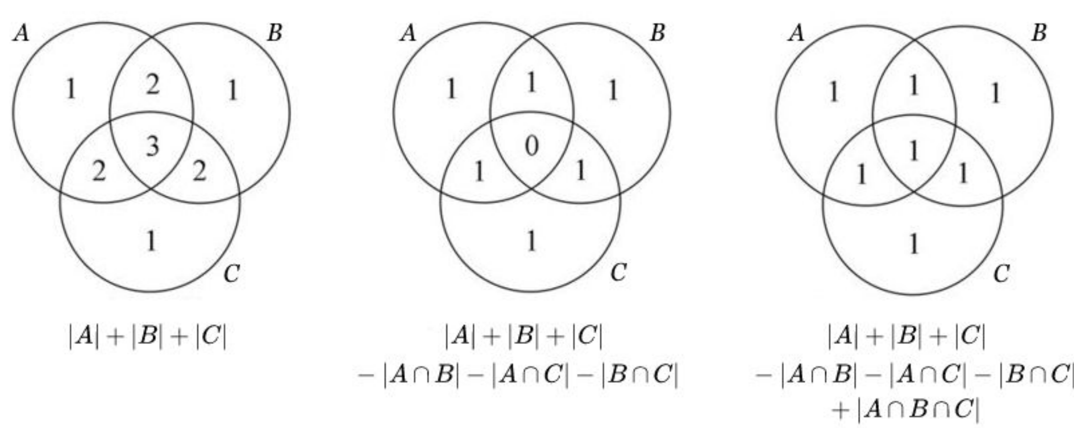
- Just like last time, we know that the size is a maximum of $|A|+|B|+|C|$.
- But now, we have double counted the elements that are found in two sets, and triple counted the ones that are in all three!
- To correct for this, we can subtract $|A \cap B|$, $|B \cap C|$, $|A \cap C|$. This makes sure all the elements that are in two sets are counted only once.$|A \cap B|$, $|B \cap C|$, $|A \cap C|$. This makes sure all the elements that are in two sets are counted only once.
- One last thing before we have included all the elements. If you look at the above subtractions, we have also subtracted elements that occur in **all the sets** 3 times, causing them to not be counted at all. To correct for this, we simply add $|A \cap B \cap C|$.

So, we end up with the following formula:

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

This formula ensures that each element is counted once, with no overcounting of elements.

Here's another great graphic for three sets:



Again, the number in each subset represents how many times that subset is counted. Here, the overlap of all three subsets must be added back in due to it being removed after the second step.

Note that this formula holds true only when there are three sets. In general, as the number of sets grow, the formula becomes more complicated because we have to account for many intersections to avoid overcounting.

Ex 2: SEM offers three types of math for seniors: Calculus CD, Statistics 2, and ATTAM. 40 students are taking Calculus, 37 are taking Statistics, and 34 are taking ATTAM. 29 are taking both Calculus and Statistics, 27 are taking both Calculus and ATTAM, and 20 are taking both Statistics and ATTAM. The only student crazy enough to be taking all three classes is Pranay. How many students took at least one of these courses?

A is the set of all students taking Calculus. $|A| = 40$.
B is the set of all students taking Statistics. $|B| = 37$.
C is the set of all students taking ATTAM. $|C| = 34$.

$|A \cap B| = 29$. $|A \cap C| = 27$. $|B \cap C| = 20$.
Finally, $|A \cap B \cap C| = 1$.
So our answer is $\boxed{|A \cup B \cup C| = 40 + 37 + 34 - 29 - 27 - 20 + 1 = 36}$

5.4: THE GENERAL CASE

You may have noticed some patterns when working through the derivation of PIE for three sets. In general, the number of distinct elements in a collection of n sets is given by:

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{k=1}^{n}\left[(-1)^{k-1} \cdot \sum_{i_1<i_2<\cdots<i_k}\left|\bigcap_{i}^{i_1,i_2,\cdots,i_k} A_j\right|\right]$$
$$= \sum_{i}^{n}|A_i| - \sum i<j^n\,|A_i \cap A_j| + \sum_{i<j<k}^{n}|A_i \cap A_j \cap A_k|$$
$$- \cdots + (-1)^{n-1}\,|A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n|$$

However, it is is very unlikely that you ever need to remember this formula. In fact, we recommend not to memorize it. The significance of this formula is the idea it represents. Here’s what the formula is saying:

1. First, take all the elements in all n sets.
2. Second, subtract out all elements that appear in two sets. Basically, subtract out all possible intersections of two sets.
3. Then add all the elements which appear in three sets.
4. Next subtract all the elements that appear in four sets.
5. Keep doing this until you reach the intersection of all sets.

We will examine this more in the next PIE lecture.

5.5: EXTENSION TO PROBABILITY

Just as PIE works for counting the number of elements in a group, it can work just as well for probability.

First, some vocabulary:
- If two events ARE **mutually exclusive**, they CANNOT occur at the same time.

  - If they are NOT mutually exclusive, however, the two events CAN occur at the same time.

  - This idea is very important in PIE, and introduces us to the idea of **overcounting** in probability.

When two events can occur together, the probability of either event occurring is given by the probability of them both occurring separately minus the the probability of them occurring together. Algebraically, if we let the events be A and B:

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

Similarly, for three events:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$$

- It is <u>very</u> important to know how PIE applies to probability, as many questions deal with probability instead of sets

- But as you see, these formulas are almost identical to what we saw for PIE with sets earlier. So if you understood that, you should be good here.

- In questions, when they ask "what is the probability of X OR Y"... this formula should come to mind

- In relation to probability, ∪ means "or" and ∩ means "and"

Problems:

1. There are exactly three types of students in a school: the geeks, the wannabees, and the athletes. Each student is classified into at least one of these categories. And the total number of students in the school is 1000. Suppose that the following is given:

The total number of students who are geeks is 310.
The total number of students who are wannabees is 650.
The total number of students who are athletes is 440.
The total number of students who are both geeks and wannabees is 170.
The total number of students who are both geeks and athletes is 150.
The total number of students who are both wannabees and athletes is 180.

What is the total number of students who fit into all 3 categories?

2. Many states use a sequence of three letters followed by a sequence of three digits as their standard license-plate pattern. Given that each three-letter three-digit arrangement is equally likely, the probability that such a license plate will contain at least one palindrome (a three-letter arrangement or a three-digit arrangement that reads the same left-to-right as it does right-to-left) is $\frac{m}{n}$, where m and n are relatively prime positive integers. Find $m+n$. (2002 AIME I #1)

3. There are 20 students participating in an after-school program offering classes in yoga, bridge, and painting. Each student must take at least one of these three classes, but may take two or all three. There are 10 students taking yoga, 13 taking bridge, and 9 taking painting. There are 9 students taking at least two classes. How many students are taking **all three** classes? (2017 AMC 10B #13)

4. At the Winter Sochi Olympics Press Conference, there are 200 foreign journalists. Out of them,

175 people can speak German,
150 people can speak French,
180 people can speak English,
160 people can speak Japanese.

What is the minimum number of foreigners that can speak all the four languages?

5. How many positive integers less than 180 are **not** relatively prime to 180.

6. My school now offers 3 new foreign languages: Arabic, Japanese, and Russian. There are 50 students enrolled in at least one of the classes. Suppose that 18 are taking both Arabic and Japanese. 15 are taking both Arabic and Russian, 13 are taking both Japanese and Russian, and 7 are taking all three languages. How many students are taking **at least two** languages?

Answers: 100, 059, 003, 065, 131, 32