Lecture outline: Inclusion-Exclusion Principle

Inclusion/exclusion is a relatively simple counting technique that allows to count the number of elements of a set that satisfy none from a list of given properties.

It is a very general technique, that has applications in many fields, from theoretical ones (computing the permanent of a matrix) to applied ones (estimating the reliability of a communication network).

The precise topics we will study are:

- The Inclusion/Exclusion: motivation, notation, statement, proof, examples.
- The Inclusion/Exclusion: proof and generalization
- The example of derangements.

Introductory example We want to count the number of integers in $[1, 30]$ that are not divisible by 2 and not divisible by 3 and not divisible by 5.

Think a little bit about it and it is easy to find there are 8 such numbers: $\{1, 7, 11, 13, 17, 19, 23, 29\}$.

What if we had started by substracting from 30 the number of integers divisible by 2, the number of integers divisible by 3 and finally the number of integers divisible by 5:

- there are 15 numbers divisible by 2
- there are 10 numbers divisible by 3
- there are 6 numbers divisible by 5

This would give $30 - (15 + 10 + 6) = -1$, which is certainly **not** the correct answer.

What happened is that we **over-counted** (*i.e.* considered more than once) some numbers are divisible by two of $\{2, 3, 5\}$: $\{6, 12, 18, 24, 30\}$ for 2 and 3, $\{10, 20, 30\}$ for 2 and 5, $\{15, 30\}$ for 3 and 5, and we removed each of them twice, so we need to add $5 + 3 + 2 = 10$ to our result, which gives $-1 + 10 = 9$ which is wrong again.

Indeed we counted 30 twice, as it is divisible by 2 and 3 and 5, again some over-counting, that we need to correct by removing 1, to obtain $9 - 1 = 8$ which is correct.

Inclusion-exclusion is only a formalisation of this principle to correct over-counting, which is better visualized on a **Venn Diagram**.

Principle of Inclusion-Exclusion

1. Let $\mathcal{S}$ be a set of N elements (often for us combinatorial objects): $|\mathcal{S}| = N$.

2. Let

$$C_1, \ldots, C_t$$

be t properties that the elements of this set could satisfy.

3. We denote the complement of a property C_i by $\overline{C_i}$: an element of $\mathcal{S}$ satisfies property $\overline{C_i}$ if it does not satisfy property C_i.

4. For a given subset $\{i_1, \ldots, i_k\}$ of $[t]$, we denote by

$$N(C_{i_1} C_{i_2} \cdots C_{i_k})$$

the number of elements of $\mathcal{S}$ that satisfy **all properties** $C_{i_1}, C_{i_2}, \cdots, C_{i_k}$.

5. We denote by $\overline{N}$ the number of elements of $\mathcal{S}$ that satisfy **none of the properties** $C_1, C_2, \cdots, C_t$, i.e.

$$\overline{N} = N(\overline{C_1}\,\overline{C_2} \cdots \overline{C_t})$$

i.e. the set of all elements

6. For $0 \leq k \leq t$, we define S_k as follows:

$$S_0 = N$$

$$S_k = \sum_{i_1 < i_2 < \cdots < i_k} N(C_{i_1} C_{i_2} \cdots C_{i_k})$$

Be careful: S_k is **not** the number of elements from $\mathcal{S}$ that satisfy **at least** k properties as the elements that satisfy more than k properties are counted more than once.

Inclusion-Exclusion Theorem.

$$\overline{N} = \sum_{k=0}^{t} (-1)^k S_k$$

Inclusion-Exclusion Theorem.

$$\overline{N} = \sum_{k=0}^{t} (-1)^k S_k$$

Illustration with our introductory example.

$\mathcal{S} = \{1, 2, \ldots, 30\}$, $S_0 = N = 30$

C_1 is the property "is divisible by 2"
C_2 is the property "is divisible by 3"
C_3 is the property "is divisible by 5"

Using basic arithmetic we obtain:
$N(C_1) = \lfloor 30/2 \rfloor = 15$
$N(C_2) = \lfloor 30/3 \rfloor = 10$
$N(C_1) = \lfloor 30/5 \rfloor = 6$
$S_1 = 31$

$N(C_1, C_2) = \lfloor 30/6 \rfloor = 5$
$N(C_1, C_3) = \lfloor 30/10 \rfloor = 3$
$N(C_2, C_3) = \lfloor 30/15 \rfloor = 2$
$S_2 = 10$

$S_3 = N(C_1, C_2, C_3) = \lfloor 30/30 \rfloor = 1$

So, applying the Inclusion-Exclusion Theorem, we obtain very quickly:

$$\overline{N} = S_0 - S_1 + S_2 - S_3 = 30 - 31 + 10 - 1 = 8$$

Example: rearranging a string with forbidden substrings

How many permutations of the letters of

$$CTAGCGAAAT$$

contain neither of the substrings CAC, GAG, $TAAT$?

1. $\mathcal{S}$ is the set of all strings with content composed of four symbols A, two symbol C, two symbols G and two symbol T. From our study of counting sequences in week 1, we know that

$$N = |\mathcal{S}| = \frac{10!}{4!2!2!2!} =$$

2. We consider three properties of the strings in $\mathcal{S}$:
C_1: the string contains CAC
C_2: the string contains GAG
C_3: the string contains $TAAT$
And we want to compute $\overline{N} = N(\overline{C}_1\,\overline{C}_2\,\overline{C}_3)$

3. $S_0 = N$ is known.

$S_1 = N(C_1) + N(C_2) + N(C_3)$, so we will compute each of the three terms:
$N(C_1)$ is the number of strings that contain the substring CAC. So let's consider CAC as a single symbol X, remove its letters from the pool of symbols and count the number of strings containing X, A, A, A, G, G, T, T:

$$N(C_1) = \frac{8!}{1!3!2!2!} =$$

The same idea applies to GAG and gives the same result: $N(C_2) = N(C_1) = \quad$).
For C_3 we have

$$N(C_3) = \frac{7!}{1!2!2!2!} =$$

We now turn to $S_2 = N(C_1C_2) + N(C_1C_3) + N(C_2C_3)$.

$N(C_1C_2)$ is the number of strings of content $\{X, Y, A, A, T, T\}$ where X stands for CAC and Y for GAG. So

$$N(C_1C_2) = \frac{6!}{1!1!2!2!}$$

$N(C_1C_3)$ is the number of strings of content $\{X, Y, A, G, G\}$ where X stands for CAC and Y for $TAAT$. So

$$N(C_1C_2) = \frac{5!}{1!1!1!2!}$$

It is obvious that $N(C_1C_3) = N(C_2C_3)$.

Finally, $S_3 = N(C_1C_2C_3)$, is the number of strings that contain the substrings CAC, GAG and $TAAT$. This is the number of strings of $\{X, Y, Z\}$ where X stands for CAC and Y for GAG and Z for $TAAT$, which is 3!.

So, putting all together, we have that

$$\overline{N} = \frac{10!}{4!2!2!2!} - \left(2\frac{8!}{2!2!2!2!} + \frac{7!}{1!2!2!2!}\right) + \left(\frac{6!}{1!1!2!2!} + 2\frac{5!}{1!1!1!2!}\right) - 3!$$

Important remark. The argument above works because the three patterns CAC, GAG and $TAAT$ do not overlap and any forbidden pattern can occur at most once. If we had patterns CAC, GCA and $TAAT$ for example, then we could not simply consider each pattern as a symbol, as a string of the letters that contains the first two patterns could be $GCACTATAAG$ for example and the calculation would be quite messy.

Example: constrained solutions to an equation

Find the number of solutions of the equation:

$$x_1 + x_2 + x_3 + x_4 = 18$$

such that

$$0 \leq x_i \leq 7, \; i = 1, 2, 3, 4.$$

First we need to define $\mathcal{S}$: $\mathcal{S}$ is the set of all **non-negative integer solutions** to $x_1 + x_2 + x_3 + x_4 = 18$, with no constraints on the x_is.

Next we need to define properties on the elements of $\mathcal{S}$: C_i is the property that the value of x_i is greater or equal to 8: $x_i \geq 8$.

So now, we have defined all we need to apply the Inclusion/Exclusion, and we want to compute

$$\overline{N} = N(\overline{C}_1 \, \overline{C}_2 \, \overline{C}_3 \, \overline{C}_4).$$

Summary

When should Inclusion/Exclusion be considered to solve a counting problem?

When we want to count the size of the subset R of a larger set $\mathcal{S}$ composed of all elements of $\mathcal{S}$ that do **NOT** satisfy a list of properties.

To apply the technique of Inclusion/Exclusion to a counting problem, we then need to do the following:

1. Define what is $\mathcal{S}$, and what is its size.
2. Define what are $C_1, \ldots, C_t$, the properties to avoid.
3. Compute the size of each S_k (this is the hard part).
4. Apply the Inclusion/Exclusion Theorem.

Derangements

Definition. A derangement p of $[n]$ is a permutation of $[n]$ of length n in which $p[i] \neq i$ for all $i \in [n]$.

The question we are interested in is the following: for a given n, how many derangements of n are-there? We denote this number by d_n.

Which permutations of $\{1, 2, 3\}$ are derangements?

1 2 3

1 3 2

2 1 3

2 3 1

3 1 2

3 2 1

We want to count the number of derangements by Inclusion/Exclusion.

1. The set $\mathcal{S}$ is the set of all permutations of $\{1, 2, \ldots, n\}$: $|\mathcal{S}| = n!$

2. We define n properties $C_1, \cdots, C_n$ as follows: C_i is the property that a permutation P satisfies $P[i] = i$.

3. Counting the numbers $S_0, S_1, \ldots, S_n$ is the hard part.

Claim. $S_k = n!/k!$ (to be proved on the next slide).

4. Apply Inclusion/Exclusion.

Theorem.

$$d_n = \sum_{k=0}^{n} (-1)^k \frac{n!}{k!} \simeq \frac{n!}{e}$$

Claim. $S_k = n!/k!$

Proof of the Inclusion-Exclusion Theorem

Why is

$$\overline{N} = \sum_{k=0}^{t} (-1)^k S_k$$

true?

Take an arbitrary element r of $\mathcal{S}$. As

$$S_k = \sum_{\{i_1,\ldots,i_k\}\subseteq[t]} N(C_{i_1} C_{i_2} \cdots C_{i_k})$$

we are going to say that "S_k counts r p times" if there are exactly p subsets $\{i_1, \ldots, i_k\} \subseteq [t]$ such that r satisfies all of the properties $C_{i_1}, C_{i_2}, \ldots, C_{i_k}$. We denote this by

$$N(r, k) = p$$

with the convention that

$$N(r, 0) = 1$$

So we want to show that, for every $r \in \mathcal{S}$:

- if r does not satisfy any of the properties C_i, $\sum_{k=0}^{t} (-1)^k N(r, k) = 1$
- if r satisfies at least one of the properties C_i, $\sum_{k=0}^{t} (-1)^k N(r, k) = 0$

The first point is immediate: if r does not satisfy any of the properties C_i, then $N(r,k) = 0$ for $k > 0$, so r is counted only by S_0 and $\sum_{k=0}^{t}(-1)^k N(r,k) = N(r,0) = 1$.

For the second point, assume that r satisfies exactly c properties, $c > 0$.

$$N(r,0) = 1$$

$$N(r,1) = c$$

$$N(r,2) = \binom{c}{2}$$

$$N(r,3) = \binom{c}{3}$$

and in general

$$N(r,k) = \binom{c}{k},\ k \geq 0$$

So we can apply the Binomial Theorem:

$$\sum_{k=0}^{c}(-1)^k N(r,k) = \sum_{k=0}^{c}\binom{c}{k}(-1)^k(1)^{c-k} = (1-1)^c = 0$$

Moreover, as

$$N(r,k) = 0, k > c$$

$$\sum_{k=0}^{t}(-1)^k N(r,k) = \sum_{k=0}^{c}(-1)^k N(r,k) = 0$$

The Inclusion/Exclusion theorem: A generalization

Now we are interested in a more general question:

How many elements of $\mathcal{S}$ satisfy **exactly** m properties among $C_1, \ C_2, \ \cdots, \ C_t$?

Let's denote by E_m the number of elements of $\mathcal{S}$ that satisfy exactly m properties.

Theorem.

$$E_m = \sum_{k=0}^{t-m} (-1)^k \binom{m+k}{k} S_{m+k}$$

If now we are interested in counting the number L_m of elements of $\mathcal{S}$ that satisfy **at least** m properties among $C_1, \ C_2, \ \cdots, \ C_t$, we have.

Corollary.

$$L_m = \sum_{k=0}^{t-m} (-1)^k \binom{m+k-1}{m-1} S_{m+k}$$

We will not see the proofs of these results, as they are following very closely the principles of the proof of the Inclusion-Exclusion Theorem, and are only slightly more technical in terms of manipulation of binomial numbers.