NOTES ON INCLUSION-EXCLUSION AND INDICATOR FUNCTIONS

1. Indicator functions

Recall from Calculus I operations with real valued functions. Fix a set U, which will serve as our universe. For any set U, two functions, $f, g : U \to \mathbb{R}$, have a sum and a product, such that

$$f \cdot g : U \to \mathbb{R} \quad \text{and} \qquad f + g : U \to \mathbb{R},$$

and the actions of these functions is described that for all $x \in U$,

$$(f \cdot g)(x) = f(x) \cdot g(x) \quad \text{and} \qquad (f + g)(x) = f(x) + g(x).$$

For every $c \in \mathbb{R}$, the *constant function* $c : U \to \mathbb{R}$ is defined by $c(x) = c$ for all $x \in U$. Constant times a function is defined as product of the function with the constant function corresponding to the constant.

For any $A \subseteq U$, the *indicator function* χ_A is defined as a $\chi_A : U \to \mathbb{R}$ function, with the action $\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$

Observe that $\chi_A \cdot \chi_B = \chi_{A \cap B}$, and furthermore

$$\chi_{A_1} \cdot \chi_{A_2} \cdot \dots \cdot \chi_{A_n} = \chi_{A_1 \cap A_2 \cap \dots \cap A_n}. \tag{1}$$

Notice that the constant function 1 is equal to χ_U. From now on we focus on universes U that are *finite sets*. Observe that the operation that assigns to for every $f : U \to \mathbb{R}$ function the value

$$\sum_{x \in U} f(x) \tag{2}$$

shows analogy with the definite integral of a function. For example, this operation can be applied for the sum of functions termwise, and a constant multiplier of the function can brought out to the outside of the operation. Furthermore, if $f(x) \leq g(x)$ for every x, then $\sum_{x \in U} f(x) \leq \sum_{x \in U} g(x)$. Observe that

$$\sum_{x \in U} \chi_A(x) = |A|. \tag{3}$$

2. Inclusion-exclusion formula

Let us be given sets $A_1, A_2, ..., A_n \subseteq U$. To avoid using double subscripts, we abbreviate χ_{A_i} to χ_i. We claim the following identity:

$$\chi_{\overline{A_1 \cup A_2 \cup \dots \cup A_n}} = (1 - \chi_1) \cdot (1 - \chi_2) \cdot \dots \cdot (1 - \chi_n). \tag{4}$$

Indeed, both sides are 0, if $x \in A_1 \cup A_2 \cup ... \cup A_n$, and 1 otherwise. Expand the RHS of (4) by the distributive rule, and change the product of indicator functions of sets to the indicator function of the intersection set as in (1). We obtain the *indicator function version of the inclusion-exclusion formula*:

$$\chi_{\overline{A_1 \cup A_2 \cup \ldots \cup A_n}} = 1 - \sum_i \chi_i + \sum_{i<j} \chi_{A_i \cap A_j} - \sum_{i<j<k} \chi_{A_i \cap A_j \cap A_k} + \cdots + (-1)^n \chi_{A_1 \cap A_2 \cap \ldots \cap A_n}. \tag{5}$$

As $\chi_{\overline{A_1 \cup A_2 \cup \ldots \cup A_n}} = 1 - \chi_{A_1 \cup A_2 \cup \ldots \cup A_n}$, formula (5) provides

$$\chi_{A_1 \cup A_2 \cup \ldots \cup A_n} = \sum_i \chi_i - \sum_{i<j} \chi_{A_i \cap A_j} + \sum_{i<j<k} \chi_{A_i \cap A_j \cap A_k} + \cdots + (-1)^{n-1} \chi_{A_1 \cap A_2 \cap \ldots \cap A_n}. \tag{6}$$

Apply the operation (2) to the two equal functions in (5), both sides must give the same result. Using the observation (3), we obtain the usual inclusion-exclusion formula for the complement of the union:

$$\begin{aligned} (7)\quad & |\overline{A_1 \cup A_2 \cup \ldots \cup A_n}| \\ = \quad & |U| - \sum_i |A_i| + \sum_{i<j} |A_i \cap A_j| - \sum_{i<j<k} |A_i \cap A_j \cap A_k| + \cdots + (-1)^n |A_1 \cap A_2 \cap \ldots \cap A_n|. \end{aligned}$$

The formula (7) above is often called *Poincaré's formula*. Similarly, applying the operation (2) to the two equal functions in (6), both sides must give the same result. Using the observation (3), we obtain the union version of inclusion-exclusion formula:

$$\begin{aligned} (8)\quad & |A_1 \cup A_2 \cup \ldots \cup A_n| \\ = \quad & \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| + \cdots + (-1)^{n-1} |A_1 \cap A_2 \cap \ldots \cap A_n|. \end{aligned}$$

The moral is that the inclusion-exclusion formula holds since a version of it already holds for the indicator functions. Such arguments extends to cases when an *inequality* holds between some linear expression of indicator functions. Applying operation (3) will keep the direction of the inequality. In fact, the inequality results are more basic, as any equality is just two simultaneous inequalities in the opposite directions.

Let us find some more examples. To make sure that our formulae fit into a page, we introduce the notation

$$\sigma_t = \sum_{i_1 < i_2 < \ldots < i_t} |A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_t}|.$$

Let us figure out what is σ_1. In sum above, $t = 1$, so we take only one index i_1. For simplicity, just call it i. For the intersection of t sets, we have just one set, the intersection is this set itself. So $\sigma_1 = \sum_i |A_i|$. A harder question is what is σ_0? We do not involve *any* set in the intersection. In general,

$$A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_t} = \Big\{ x \in U : x \in A_{i_1} \wedge x \in A_{i_2} \wedge \ldots \wedge x \in A_{i_t} \Big\},$$

and if no sets are present in the intersection, the righthandside set is $\{x \in U :$ no condition $\} = U$, as no condition is always true. So we convinced ourselves that $\sigma_0 = |U|$. (Still I would not base the proof of a very important result on this convention!) In this new notation, (8) reads as

$$|A_1 \cup A_2 \cup \ldots \cup A_n| = \sigma_1 - \sigma_2 + \sigma_3 - \ldots + (-1)^{n-1}\sigma_n; \tag{9}$$

while (7) reads as

$$|\overline{A_1 \cup A_2 \cup \ldots \cup A_n}| = \sigma_0 - \sigma_1 + \sigma_2 - \sigma_3 - \ldots + (-1)^n \sigma_n. \tag{10}$$

3. Bonferroni inequalities

The *Bonferroni inequalities* (11) assert that in (9), if we stop computing the righthandside after an odd number of terms, then we have an upper bound for the size of the union, and if if we stop computing the righthandside after an even number of terms, then we have a lower bound for the size of the union. Formally, for every positive integer k,

$$\sum_{t=1}^{2k} (-1)^{t+1}\sigma_t \leq |A_1 \cup A_2 \cup \ldots \cup A_n| \leq \sum_{t=1}^{2k-1} (-1)^{t+1}\sigma_t. \tag{11}$$

Now we spell out the related indicator function inequality, of which the application of the (3) operation yields (11), proving the Bonferroni inequalities:

$$\sum_{t=1}^{2k} (-1)^{t+1} \sum_{i_1 < i_2 < \ldots < i_t} \chi_{A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_t}} \leq \chi_{A_1 \cup A_2 \cup \ldots \cup A_n} \leq \sum_{t=1}^{2k-1} (-1)^{t+1} \sum_{i_1 < i_2 < \ldots < i_t} \chi_{A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_t}}. \tag{12}$$

We have to show the inequalities in (12) for all $x \in U$. We do it by cases. If x is not an element of any of the n sets A_i, (12) evaluated in x says $0 \leq 0 \leq 0$, which is true. Assume now that x is an element of *exactly* m of the sets A_i, and $m \geq 1$. Then (12) evaluated in x boils down to

$$\sum_{t=1}^{2k} (-1)^{t+1} \binom{m}{t} \leq 1 \leq \sum_{t=1}^{2k-1} (-1)^{t+1} \binom{m}{t}. \tag{13}$$

The reason is that $\chi_{A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_t}}$ can be nonzero (namely 1) if and only if x is an element of each of the sets $A_{i_1}, A_{i_2}, \ldots, A_{i_t}$. For this we have to select the $i_1 < i_2 < \cdots < i_t$ indices from the index set of the m sets that contain x.

To verify (13), we use a lemma about binomial coefficients. For $n \geq 1$, we have

$$\sum_{i=0}^{\ell} (-1)^i \binom{n}{i} = (-1)^\ell \binom{n-1}{\ell}. \tag{14}$$

Hence the expression in (14) is non-negative for even ℓ and non-positive for odd ℓ. Roughly speaking, the sign of the sum is the sign of its last term. The proof of (14) is based on how we compute the n^{th} row of the Pascal triangle from row $n-1$, for $n \geq 1$.

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

Writing 1 as $\binom{m}{0}$ in the middle of formula (13) and making comparison with (14) verifies formula (13).

4. Charles Jordan's formula

Formula (7) can be understood as expessing how many elements of the universe belong to exactly 0 of the sets $A_1, A_2, ..., A_n$ as element. Charles Jordan's theorem tells how many elements of U belong to exactly q $(q \geq 1)$ of the sets $A_1, A_2, ..., A_n$ as element:

$$\sum_{j=q}^{n} (-1)^{j+q} \binom{j}{q} \sigma_j. \tag{15}$$

Charles Jordan's formula is actually valid for $q = 0$ as well, and boils down to the Poincaré formula (7) in this instance, as shown in (10).

Charles Jordan's formula also follows from a related identity of the indicator functions. Let χ denote the indicator function of the set, whose elements belong to exactly q of the sets $A_1, A_2, ..., A_n$. The indicator function identity is

$$\chi = \sum_{j=q}^{n} (-1)^{j+q} \binom{j}{q} \sum_{i_1 < i_2 < \ldots < i_j} \chi_{A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_j}}. \tag{16}$$

Applying the operation (3) to (16) we obtain (15). To verify that the two functions in (16) are equal, we have to show that they are equal in every $x \in U$. Assume that x belongs to exactly m of the sets $A_1, A_2, ..., A_n$. Then we have

$$\chi(x) = \begin{cases} 1 & \text{if } m = q \\ 0 & \text{otherwise.} \end{cases}$$

From the other side,

$$\begin{aligned} & \sum_{j=q}^{n} (-1)^{j+q} \binom{j}{q} \sum_{i_1 < i_2 < \ldots < i_j} \chi_{A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_j}}(x) = \sum_{j=q}^{n} (-1)^{j+q} \binom{j}{q} \binom{m}{j} \\ = \ & \binom{m}{q} \sum_{j=q}^{n} (-1)^{j-q} \binom{m-q}{j-q} = \binom{m}{q} = \begin{cases} 1 & \text{if } m = q \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

This completes the proof of (16), and in turn, of (15). (For details of the calculation, the identity $\binom{j}{q}\binom{m}{j} = \binom{m}{q}\binom{m-q}{j-q}$ can be easily verified from the fraction of factorials form of the binomial coefficient. You can even find a combinatorial proof to this identity. $\sum_{j=q}^{n}(-1)^{j-q}\binom{m-q}{j-q} = \sum_{\ell=0}^{m-q}(-1)^{\ell}\binom{m-q}{\ell}$ is true by changing the summation variable by $\ell = j - q$. The result is the alternating sum of binomial coefficients in the row $m - q$ of the Pascal Triangle. We know this alternating sum is zero—except when $m - q = 0$, i.e., $m = q$. In this case the row 0 of the Pascal Triangle is just $\binom{0}{0} = 1$, and the alternating sum is 1. In this case the coefficient $\binom{m}{q} = \binom{m}{m} = 1$.)

5. A custom-taylored inequality

Kai Lai Chung discovered that for any natural number ℓ, the inequality

$$\binom{\ell+1}{2}|A_1 \cup A_2 \cup \ldots \cup A_n| \geq \ell \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j|.$$

(This formula is actually useful in statistics.) In the special case of $\ell = 1$, this is one of the Bonferroni inequalities. Again, this inequality follows by the operation (3) from the corresponding inequality for indicator functions:

$$\binom{\ell+1}{2}\chi_{A_1 \cup A_2 \cup \ldots \cup A_n} \geq \ell \sum_i \chi_i - \sum_{i<j} \chi_{A_i \cap A_j}.$$

We have to show the inequality of the functions in every $x \in U$. If x does not belong to any of the sets $A_1, A_2, ..., A_n$, the inequality is $0 \geq 0$. Assume that x belongs to exactly m of the sets $A_1, A_2, ..., A_n$ for some $m \geq 1$. In this case the indicator function inequality boils down to

$$\binom{\ell+1}{2} \geq \ell m - \binom{m}{2},$$

which, by simple algebra, is equivalent to

$$(\ell - m) + (\ell - m)^2 \geq 0.$$

Actually the parabola $y = x + x^2$ takes negative values in the interval $(-1, 0)$ but never for an integer x. Therefore the integrality of ℓ and m is important.