1 Inclusion-Exclusion

Definition Boolean polynomials. (The definition is recursive. Boolean polynomial is what we can build with iteration of the procedure below.)

- Variable names like $A_1, A_2, ...$ are Boolean polynomials.
- If B_1 and B_2 are Boolean polynomials, then so are $(B_1) \cap (B_2)$, $(B_1) \cup (B_2)$, $\overline{B_1}$.

For example, $\overline{A_1}$, $\overline{A_1} \cap A_2$, $(\overline{A_1} \cap A_2) \cup A_1$, etc. are all Boolean polynomials. Given sets (or events in a probability space) $S_1, S_2, ...$, it makes sense to evaluate the Boolean polynomial by substituting $A_i \leftarrow S_i$, and completing the operations we obtain a set (event) as *value* of the polynomial with this substitution. We say that a Boolean polynomial is n-variable, if it uses n distinct variable names. A variable name is a 1-variable Boolean polynomial.
Definition Indicator function
Given a probability space $(\Omega, \mathcal{A}, p)$ (for events) or a universe U (for sets), the indicator function of an $A \in \mathcal{A}$ or $A \subseteq U$ is a function $\chi_A : \Omega \to \mathbb{R}$ (or a function $\chi_A : U \to \mathbb{R}$) such that

$$\chi_A(a) = \begin{cases} 1, & \text{if } a \in A, \\ 0, & \text{if } a \notin A. \end{cases}$$

(The definition for sets can be considered as special case of the definition for probability spaces, if we associate with a universe U the probability space $(\Omega, \mathcal{A}, p)$ with $\Omega = U$, $\mathcal{A} = 2^U$ and for $A \in \mathcal{A}$ and $p(A) = |A|/|\Omega|$.)

Lemma 1 *For any n-variable Boolean polynomial $B(A_1, ..., A_n)$ there exists an an n-variable real function $f_B : \mathbb{R}^n \to \mathbb{R}$, such that for any substitution events $S_1, ..., S_n \in \mathcal{A}$ from any probability space into the variables $A_1, ..., A_n$, the following $\Omega \to \mathbb{R}$ functions are equal:*

$$\chi_{B(S_1,...,S_n)} = f_B(\chi_{S_1}, ..., \chi_{S_n}).$$

Proof. As Boolean polynomials were defined recursively, we have to prove the statement by induction that mimicks the recursive definition. If A is a variable name, then we define $f_A : \mathbb{R} \to \mathbb{R}$ by $f_A(x) = x$. Clearly for any $S \in \mathcal{A}$ and any $\omega \in \omega$, $\chi_S(\omega) = (x \circ \chi_s)(\omega)$ as required. If $B(A_1, ..., A_n)$ was defined as $B = (B_1) \cap (B_2)$ from already defined Boolean polynomials, then we must have that the variable sets of B_1 and B_2 are subsets of the variable set $A_1, ..., A_n$, and some real functions f_{B_1} and f_{B_2} satisfy the requirements of the Lemma for them. Rewrite f_{B_1} and f_{B_2} in a form with n variables such that their variables correspond in order to the variables $A_1, ..., A_n$, possibly not using at all some variables. Define $f_B(x_1, ..., x_n) := f_{B_1}(x_1, ..., x_n) f_{B_2}(x_1, ..., x_n)$. It clearly suffices. Similarly, if $B = \overline{B_1}$, set $f_B = 1 - f_{B_1}$; and if $B = (B_1) \cup (B_2)$, set $f_B(x_1, ..., x_n) := f_{B_1}(x_1, ..., x_n) + f_{B_2}(x_1, ..., x_n) - f_{B_1}(x_1, ..., x_n) f_{B_2}(x_1, ..., x_n)$, after rewriting f_{B_1} and f_{B_2} as in the case of intersection. □

Let us consider the following two statements. For a fixed $(\Omega, \mathcal{A}, p)$ probability space, Boolean polynomials $B_j(A_1, ..., A_n)$ and constants $c_j \in \mathbb{R}$ $(j = 1, 2, ..., m)$ and events $S_1, ..., S_n \in \mathcal{A}$,

- (a) $\forall \omega \in \Omega \quad \sum_{j=1}^m c_j \chi_{B_j(S_1,...,S_n)}(\omega) \geq 0$
- (b) $\sum_{j=1}^m c_j p(B_j(S_1, ..., S_n)) \geq 0$.

It is clear the (a) implies (b): compute the expectation of both sides of (a), use that expectation of a non-negative random variable is non-negative, and use the linearity of expectation on the left-hand side.

Theorem 2 *Let us be given an $(\Omega, \mathcal{A}, p)$ probability space, Boolean polynomials $B_j(A_1, ..., A_n)$ and constants $c_j \in \mathbb{R}$ $(j = 1, 2, ..., m)$. The following statements are equivalent:*

- *(i)* $\forall \omega \in \Omega$ and substitution $A_i \leftarrow S_i$ $(S_i \in \mathcal{A})$, we have $\sum_{j=1}^m c_j \chi_{B_j(S_1,...,S_n)}(\omega) \geq 0$.
- *(ii)* for all substitution $A_i \leftarrow S_i$ $(S_i \in \mathcal{A})$, we have $\sum_{j=1}^m c_j p(B_j(S_1, ..., S_n)) \geq 0$.

- *(iii)* for all substitution $A_i \leftarrow S_i = \emptyset$ *or* $A_i \leftarrow S_i = \Omega$, we have $\sum_{j=1}^{m} c_j p(B_j(S_1, ..., S_n)) \geq 0$.

Proof. It is obvious that (i) implies (ii) and that (ii) implies (iii). We show (iii) implies (i) by proving the contrapostive. Assume that for some events $S_1, ..., S_n$ and an $\omega \in \Omega$, we have $\sum_{j=1}^{m} c_j \chi_{B_j(S_1,...,S_n)}(\omega) < 0$. Define the following new events:

$$S_i^* = \begin{cases} \Omega, & \text{if } \omega \in S_i, \\ \emptyset, & \text{if } \omega \notin S_i. \end{cases}$$

Next, using the lemma for the first equation,

$$\chi_{B_j(S_1,...,S_n)}(\omega) = f_{B_j}(\chi_{S_1}(\omega), ..., \chi_{S_n}(\omega)) = f_{B_j}(\chi_{S_1^*}(\omega), ..., \chi_{S_n^*}(\omega)) = f_{B_j}(\chi_{S_1^*}(\omega^*), ..., \chi_{S_n^*}(\omega^*)),$$

and the last equation holds for *every* $\omega^* \in \Omega$. Observe that

$$p(B_j(S_1^*, ..., S_n^*)) = E(\chi_{B_j(S_1^*,...,S_n^*)}) = E(f_{B_j}(\chi_{S_1^*}, ..., \chi_{S_n^*})) = \chi_{B_j(S_1,...,S_n)}(\omega).$$

Repeating this for $j = 1, 2, ..., m$, we obtain the expected conclusion:

$$\sum_{j=1}^{m} c_j p(B_j(S_1^*, ..., S_n^*)) = \sum_{j=1}^{m} c_j \chi_{B_j(S_1,...,S_n)}(\omega) < 0.$$

□

As every equality $a = b$ can be written as simultaneous inequalities $a \leq b$ and $b \leq a$, we obtain:

Corollary 3 *The Theorem applies if ≥ 0 is changed to $= 0$ in all three clauses.*

One easily obtains the inclusion-exclusion formula as an application for the corollary. First, for events S_i ($i \in \{1, 2, ..., n\}$) and $I \subseteq \{1, 2, ..., n\}$, define $S_I := \cap_{i \in I} S_i$. The inclusion-exclusion formula says:

$$p\Big(\bigcup_{i=1}^{n} S_i\Big) = \sum_{i=1}^{n} \sum_{\substack{|I|=i \\ I \subseteq \{1,2,...,n\}}} (-1)^{i-1} p(S_I). \tag{1.1}$$

inclexcl

The proof, according the the theorem, only requires checking the identity for events where for every i, $S_i = \emptyset$ or $S_i = \Omega$. If every $S_i = \emptyset$, both sides in (1.1) are equal to 0. So assume that we have $\ell \geq$ events $= \Omega$, the other $n - \ell$ events are $= \emptyset$. Now the lefthand side in (1.1) is $= 1$, while the righthand side is $\binom{\ell}{1} - \binom{\ell}{2} + \binom{\ell}{3} - ... + (-1)^{\ell-1}\binom{\ell}{\ell}$. These are equal as the expansion of $(1-1)^{\ell} = 0$ according to binomial theorem shows.

Theorem 4 (Bonferroni Inequalities)

$$\sum_{i=1}^{2t} \sum_{\substack{|I|=i \\ I \subseteq \{1,2,...,n\}}} (-1)^{i-1} p(S_I) \leq p\Big(\bigcup_{i=1}^{n} S_i\Big) \leq \sum_{i=1}^{2t+1} \sum_{\substack{|I|=i \\ I \subseteq \{1,2,...,n\}}} (-1)^{i-1} p(S_I).$$

Proof. The result is trivial when all are $S_i = \emptyset$. Otherwise, assume that exactly $\ell \geq 1$ of them are $= \Omega$. With this substitution, the theorem boils down to

$$\sum_{i=1}^{2t} (-1)^{i-1} \binom{\ell}{i} \leq 1 \leq \sum_{i=1}^{2t+1} (-1)^{i-1} \binom{\ell}{i}.$$

This follows from the identity $\sum_{i=0}^{k} (-1)^i \binom{\ell}{i} = 1 + \sum_{i=1}^{k} (-1)^i \Big(\binom{\ell-1}{i-1} + \binom{\ell-1}{i}\Big) = (-1)^k \binom{\ell-1}{k}$. □