INCLUSION-EXCLUSION PRINCIPLE

KIM, SUNGJIN

Abstract. We prove a generalized version of inclusion-exclusion principle including Bonferroni inequalities.

1. Notations

- $A_1, A_2, \cdots, A_n$ denote n events.
- $P_1, P_2, \cdots, P_n$ denote sums of j-fold intersection probabilities:

$$P_j = \sum_{1\le i_1 < \cdots < i_j \le n} P(A_{i_1} \cap \cdots \cap A_{i_j}),$$

and we set $P_0 = 1$ for convention.

- $E_1, E_2, \cdots, E_n$ denote the events that exactly j events occur out of n events $A_1, \cdots, A_n$.

2. Identities and Inequalities

We will prove the following identities using generating functions.

Theorem 1. *We have*

$$P_j = \sum_{k=j}^{n} \binom{k}{j} P(E_k), \tag{1}$$

and

$$P(E_j) = \sum_{k=0}^{n-j} (-1)^k \binom{j+k}{k} P_{j+k}. \tag{2}$$

Proof. The following two functions are the generating functions for P_j and $P(E_j)$ respectively,

$$f(x) := \mathbf{E}\left[\prod_{i=1}^{n}(x1_{A_i} + 1)\right],$$

$$g(x) := \mathbf{E}\left[\prod_{i=1}^{n}(x1_{A_i} + 1_{A_i^c})\right].$$

In fact, the coefficients of x^j in $f(x)$, $g(x)$ are P_j and $P(E_j)$ respectively.

Since $1_{A^c} = 1 - 1_A$ for any event A, we have

$$f(x) = g(x+1), \quad g(x) = f(x-1). \tag{3}$$

Then (1), (2) follow from binomial theorem and the above identities (3). The identity (2) is the generalized inclusion-exclusion principle, and the usual inclusion-exclusion principle is when $j = 0$. □

We focus on the alternating sum (2). It is natural to guess that the partial sums with even number of terms is $\le P(E_j)$, and the partial sums with odd number of terms is $\ge P(E_j)$. We need the following lemma.

Lemma 1. *Let $n \ge 1$ and $0 \le m \le n$. Then we have*

$$\sum_{j=0}^{m} (-1)^j \binom{n}{j} = (-1)^m \binom{n-1}{m}. \tag{4}$$

Proof. This follows from Pascal's triangle. In fact,

$$\sum_{j=0}^{m}(-1)^j\binom{n}{j} = \sum_{j=0}^{m}(-1)^j\left(\binom{n-1}{j-1}+\binom{n-1}{j}\right)$$
$$= -\sum_{j\le m-1}(-1)^j\binom{n-1}{j}+\sum_{j\le m}(-1)^j\binom{n-1}{j}$$
$$= (-1)^m\binom{n-1}{m}.$$

□

Now, we can prove the following version of Bonferroni inequalities:

Theorem 2. *Let $0 \le j \le n$ and $0 \le 2m < 2m+1 \le n-j$. Then we have*

$$\sum_{k=0}^{2m+1}(-1)^k\binom{j+k}{k}P_{j+k} \le P(E_j) \le \sum_{k=0}^{2m}(-1)^k\binom{j+k}{k}P_{j+k}. \tag{5}$$

By (1) and (4), we obtain for any $0 \le m \le n-j$,

$$\sum_{k=0}^{m}(-1)^k\binom{j+k}{k}P_{j+k} = \sum_{k=0}^{m}(-1)^k\binom{j+k}{k}\sum_{s=j+k}^{n}\binom{s}{j+k}P(E_s)$$
$$= \sum_{k=0}^{m}(-1)^k\sum_{s=j+k}^{n}\binom{s}{j}\binom{s-j}{k}P(E_s)$$
$$= \sum_{s=j}^{n}\binom{s}{j}\sum_{k\le\min(s-j,\ m)}(-1)^k\binom{s-j}{k}P(E_s)$$
$$= P(E_j)+(-1)^m\sum_{s=j+1}^{n}\binom{s}{j}\binom{s-j-1}{m}P(E_s).$$

Therefore, (5) follows, and the usual Bonferroni inequalities are when $j=0$.