15 Inclusion-Exclusion

Today, we introduce basic concepts in probability theory and we learn about one of its fundamental principles.

Throwing dice. Consider a simple example of a probabilistic experiment: throwing two dice and counting the total number of dots. Each die has six sides with 1 to 6 dots. The result of a throw is thus a number between 2 and 12. There are 36 possible outcomes, 6 for each die, which we draw as the entries of a matrix; see Figure 16.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

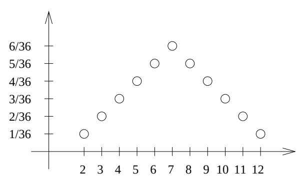

Figure 16: Left: the two dice give the row index and the column index of the entry in the matrix. Right: the most likely sum is 7, with probability $\frac{1}{6}$, the length of the diagonal divided by the size of the matrix.

Basic concepts. The set of possible outcomes of an experiment is the *sample space*, denoted as Ω. A possible outcome is an *element*, $x \in \Omega$. A subset of outcomes is an *event*, $A \subseteq \Omega$. The *probability* or *weight* of an element x is $P(x)$, a real number between 0 and 1. For finite sample spaces, the *probability* of an event is $P(A) = \sum_{x \in A} P(x)$.

For example, in the two dice experiment, we set $\Omega = \{2, 3, \ldots, 12\}$. An event could be to throw an even number. The probabilities of the different outcomes are given in Figure 16 and we can compute

$$P(\text{even}) = \frac{1+3+5+5+3+1}{36} = \frac{1}{2}.$$

More formally, we call a function $P : \Omega \to \mathbb{R}$ a *probability distribution* or a *probability measure* if

(i) $P(x) \geq 0$ for every $x \in \Omega$;

(ii) $P(A \dot\cup B) = P(A) + P(B)$ for all disjoint events $A \cap B = \emptyset$;

(iii) $P(\Omega) = 1$.

A common example is the *uniform probability distribution* defined by $P(x) = P(y)$ for all $x, y \in \Omega$. Clearly, if Ω is finite then

$$P(A) = \frac{|A|}{|\Omega|}$$

for every event $A \subseteq \Omega$.

Union of non-disjoint events. Suppose we throw two dice and ask what is the probability that the outcome is even or larger than 7. Write A for the event of having an even number and B for the event that the number exceeds 7. Then $P(A) = \frac{1}{2}$, $P(B) = \frac{15}{36}$, and $P(A \cap B) = \frac{9}{36}$. The question asks for the probability of the union of A and B. We get this by adding the probabilities of A and B and then subtracting the probability of the intersection, because it has been added twice,

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

which gives $\frac{6}{12} + \frac{5}{12} - \frac{3}{12} = \frac{2}{3}$. If we had three events, then we would subtract all pairwise intersections and add back in the triplewise intersection, that is,

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) \\ &\quad - P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

Principle of inclusion-exclusion. We can generalize the idea of compensating by subtracting to n events.

PIE THEOREM (FOR PROBABILITY). The probability of the union of n events is

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n} (-1)^{k+1} \sum P(A_{i_1} \cap \ldots \cap A_{i_k}),$$

where the inner sum is over all subsets of k events.

PROOF. Let x be an element in $\bigcup_{i=1}^{n} A_i$ and H the subset of $\{1, 2, \ldots, n\}$ such that $x \in A_i$ iff $i \in H$. The contribution of x to the sum is $P(x)$ for each odd subset of H and $-P(x)$ for each even subset of H. If we include $\emptyset$ as an even subset, then the number of odd and even subsets is the same. We can prove this using the Binomial Theorem:

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

But in the claimed equation, we do not account for the empty set. Hence, there is a surplus of one odd subset and therefore a net contribution of $P(x)$. This is true for every element. The PIE Theorem for Probability follows. $\square$

Checking hats. Suppose n people get their hats returned in random order. What is the chance that at least one gets the correct hat? Let A_i be the event that person i gets the correct hat. Then

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Similarly,

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = \frac{(n-k)!}{n!}.$$

The event that at least one person gets the correct hat is the union of the A_i. Writing $P = P(\bigcup_{i=1}^n A_i)$ for its probability, we have

$$\begin{aligned} P &= \sum_{k=1}^n (-1)^{k+1} \sum P(A_{i_1} \cap \ldots \cap A_{i_k}) \\ &= \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} \\ &= \sum_{k=1}^n (-1)^{k+1} \frac{1}{k!} \\ &= 1 - \frac{1}{2} + \frac{1}{3!} - \ldots \pm \frac{1}{n!}. \end{aligned}$$

Recall from Taylor expansion of real-valued functions that $e^x = 1 + x + x^2/2 + x^3/3! + \ldots$. Hence,

$$P = 1 - e^{-1} = 0.6\ldots$$

Inclusion-exclusion for counting. The principle of inclusion-exclusion generally applies to measuring things. Counting elements in finite sets is an example.

PIE THEOREM (FOR COUNTING). For a collection of n finite sets, we have

$$|\bigcup_{i=1}^n A_i| = \sum_{k=1}^n (-1)^{k+1} \sum |A_{i_1} \cap \ldots \cap A_{i_k}|,$$

where the second sum is over all subsets of k events.

The only difference to the PIE Theorem for Probability is that for each x, we count 1 instead of $P(x)$.

Counting surjective functions. Let M and N be finite sets, and $m = |M|$ and $n = |N|$ their cardinalities. Counting the functions of the form $f : M \to N$ is easy. Each $x \in M$ has n choices for its image, the choices are independent, and therefore the number of functions is n^m. How many of these functions are surjective? To answer this question, let $N = \{y_1, y_2, \ldots, y_n\}$ and let A_i be the set of functions in which y_i is not the image of any element in M. Writing A for the set of all functions and S for the set of all surjective functions, we have

$$S = A - \bigcup_{i=1}^n A_i.$$

We already know $|A|$. Similarly, $|A_i| = (n-1)^m$. Furthermore, the size of the intersection of k of the A_i is

$$|A_{i_1} \cap \ldots \cap A_{i_k}| = (n-k)^m.$$

We can now use inclusion-exclusion to get the number of functions in the union, namely,

$$|\bigcup_{i=1}^n A_i| = \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} (n-k)^m.$$

To get the number of surjective functions, we subtract the size of the union from the total number of functions,

$$|S| = \sum_{k=0}^n (-1)^k \binom{n}{k} (n-k)^m.$$

For $m < n$, this number should be 0, and for $m = n$, it should be $n!$. Check whether this is indeed the case for small values of m and n.