5 Counting and Probability

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

**Probability measure.** We begin with 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 4.

Figure 4: 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 $\frac{6}{36}$, the length of the diagonal divided by the size of the matrix.

We need definitions to construct a general framework that formalizes this and similar examples. The set of possible outcomes of a probabilistic experiment is the sample space, denoted as $\Omega$. 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 4, and we can compute

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

More formally, we call a function $P : 2\Omega \rightarrow \mathbb{R}$ a probability measure if

(i) $P(x) \geq 0$ for every $x \in \Omega$;
(ii) $P(A \cup B) = P(A) + P(B)$ for all disjoint events $A \cap B = \emptyset$;
(iii) $P(\Omega) = 1$.

A common example is the uniform probability measure defined by $P(x) = P(y)$ for all $x, y \in \Omega$. Clearly, if $\Omega$ is finite then $P(A) = |A|/|\Omega|$ 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 $P(A) = \frac{1}{2}$, $P(B) = \frac{12}{36}$, and $P(A \cap B) = \frac{8}{36}$. 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 $\frac{1}{2} + \frac{12}{36} - \frac{8}{36} = \frac{2}{3}$. If we had three events, then we would subtract all pairwise intersections and add back in the triplewise intersection:

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

The reason for adding $P(A \cap B \cap C)$ is that it has been added three times, as part of $P(A)$, $P(B)$, and $P(C)$, but it has also been subtracted three times, as part of $P(A \cap B)$, $P(A \cap C)$, and $P(B \cap C)$. We can generalize the idea.

**PIE Theorem (for probability).** The probability of the union of not necessarily disjoint events $A_1$ to $A_n$ is the alternating sum of probabilities over all non-empty subcollections of events:

$$P(\bigcup_{i=1}^n A_i) = \sum_{k=1}^n (-1)^{k+1} \sum_{\mathcal{H}} P(A_{i_1} \cap \ldots \cap A_{i_k}).$$

**Proof.** Let $x$ be an element in $\bigcup_{i=1}^n A_i$ and $H$ the subset of $\{1, 2, \ldots, n\}$ such that $x \in A_i$ if $i \in H$. 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, as we have proved using the Binomial Theorem. In this application, we do not include the empty set, so we get

$$\sum_{k=1}^n (-1)^{k+1} \binom{n}{k} = 1.$$

In words, there is a surplus of one odd subset and therefore a net contribution of $P(x)$. This is true for every element. The claimed equation follows.
**Checking hats.** Next, we use the Principle of Inclusion-Exclusion to compute the probability of getting hats returned correctly. Suppose \( n \) people check their hats before going to the rock concert, and after the concert, they 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,
\[
P(A_i \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \frac{(n - k)!}{n!}.
\]
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
\[
P = \sum_{k=1}^{n} (-1)^{k+1} \sum P(A_{i_1} \cap \ldots \cap A_{i_k})
\]
\[
= \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n - k)!}{n!}
\]
\[
= \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!}
\]
\[
= 1 - \frac{1}{2} + \frac{1}{3!} - \ldots \pm \frac{1}{n!}.
\]
Recall from the Taylor expansion of real-valued functions that \( e^x = 1 + x + x^2/2 + x^3/3! + \ldots \). Hence,
\[
P = 1 - e^{-1} = 0.632 \ldots
\]
in the limit, when \( n \) goes to infinity. For finite numbers, the probability oscillates; that is: the chance that one hat gets returned correctly for an odd number \( n \) of people is higher than for \( n - 1 \) people but it is also higher than for \( n + 1 \) people. Similarly, for odd \( n \), the probability decreases with increasing \( n \), while for even \( n \), the probability increases with increasing \( n \).

**Counting surjective functions.** The principle of inclusion-exclusion generally applies to measuring things. Counting elements in finite sets is an example.

**PIE Theorem (for counting).** The cardinality of the union of not necessarily disjoint sets \( A_1 \) to \( A_n \) is the alternating sum of cardinalities over all non-empty subcollections of the sets:
\[
| \bigcup_{i=1}^{n} A_i | = \sum_{k=1}^{n} (-1)^{k+1} \sum |A_{i_1} \cap \ldots \cap A_{i_k}|.
\]
The only difference to the PIE Theorem for Probability is that for each \( x \), we count 1 instead of \( P(x) \), so we do not repeat the proof.

We use the counting version of the principle to count surjective functions. Counting the functions of the form \( f : [m] \rightarrow [n] \) is easy. Each \( j \in [m] \) 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 \( A_i \) be the set of functions in which \( i \in [n] \) 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
\[
|A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}| = (n - k)^m.
\]
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 \( n! \). Is this really the case?