2 Counting

From a mathematical viewpoint, there is and there cannot be any mysticism in probability. The easiest entry to this field is via discrete mathematics, which is the topic we study during the first few weeks. We begin by introducing some of the basic concepts of discrete mathematics: sets, lists, and functions. They are used to model a large variety of circumstances and relations and are fundamental to mathematics and beyond.

Sets. A set is an unordered collection of distinct elements. The union, intersection, difference, and symmetric difference of two sets are defined as:

\[
A \cup B = \{ x \mid x \in A \text{ or } x \in B \}, \quad (1)
\]

\[
A \cap B = \{ x \mid x \in A \text{ and } x \in B \}, \quad (2)
\]

\[
A - B = \{ x \mid x \in A \text{ and } x \not\in B \}, \quad (3)
\]

\[
A \oplus B = (A - B) \cup (B - A). \quad (4)
\]

Look at Figure 1 for a visual description of the sets that result from the four types of operations. We say that \(A\) and \(B\) are disjoint if \(A \cap B = \emptyset\). The number of elements in a set \(A\) is denoted as \(|A|\), or sometimes as \(#A\). It is referred to as the size or the cardinality of \(A\). The number of elements in the union of two sets cannot be larger than the sum of the two sizes:

\[
|A \cup B| \leq |A| + |B|, \quad (5)
\]

with equality if \(A\) and \(B\) are disjoint. A subset \(X \subseteq A\) is a set such that every element in \(X\) is also an element of \(A\). Note that \(A\) may or may not have additional elements. It follows that \(|X| \leq |A|\). The power set of \(A\) is the set of all subsets: \(2^A = \{ X \mid X \subseteq A \}\). It includes the empty subset, which is denoted as \(\emptyset\). As suggested by the notation, the cardinality of the power set is

\[
|2^A| = 2^{|A|}. \quad (6)
\]

Indeed, when we select a subset, we have two choices for each element: to use it or not to use it. There are \(|A|\) elements and therefore \(2^{|A|}\) different ways to resolve the choices.

Lists and functions. A list is an ordered collection of not necessarily distinct elements. Instead of studying them directly, we express lists in terms of another mathematical concept in which we map elements of one set to elements of another set. A function \(f\) from a domain \(D\) to a range \(R\), denoted as \(f : D \rightarrow R\), associates exactly one element in \(R\) to each element \(x \in D\). A list of \(k\) elements is a function \(\{1, 2, \ldots, k\} \rightarrow R\); see Figure 2 corresponds to the list \(abc\)\(b\)\(z\)\(133\). To count the functions from a finite domain to a finite range, we note that we have \(|R|\) choices for each element in \(D\). Since the choices are independent, we have a total of \(|D|^{|R|}\) functions.

Besides modeling lists, functions are generally useful to express relationships. For example, every gene has a length, measured as the number of nucleotides, so we have a function from the set of genes to the positive integers. It is possible that two genes have the same length, and there are surely positive integers that are not the lengths of any gene. We will refine our language about functions to make these distinctions. The function \(f : D \rightarrow R\) is injective or one-to-one if \(f(x) \neq f(y)\) for all \(x \neq y\). It is surjective or onto if for every \(r \in R\), there exists some \(x \in D\) with \(f(x) = r\). The function is bijective or a one-to-one correspondence if it is both injective and surjective. Observe that two sets \(D\) and \(R\) have the same size if and only if there exists a bijection \(f : D \rightarrow R\). For finite sets, this is indeed an observation, while for infinite sets, it is a definition.

Asking how many bijections there are from \(D\) to \(R\) only makes sense if they have the same size. For finite sets, be-
ing injective is then the same as being bijective. To count the number of bijections, we assign elements of $R$ to elements of $D$, in sequence. Assuming $|D| = |R| = n$, we have $n$ choices for the first element in the domain, $n - 1$ choices for the second, $n - 2$ for the third, and so on. Hence the number of different bijections from $D$ to $R$ is $n \cdot (n - 1) \cdot \ldots \cdot 1 = n!$, pronounced $n$ factorial.

**Permutations.** A permutation is an ordered collection of distinct elements. In the original form, they use all the elements, so they can be defined as a bijection from a finite set $D$ to itself, $f : D \rightarrow D$. For example, the permutations of $\{1, 2, 3\}$ are: 123, 132, 213, 231, 312, and 321. Here we list the permutations in lexicographic order, same as they would appear in a dictionary. Assuming $|D| = k$, there are $k!$ permutations or, equivalently, orderings of the set. To see this, we note that there are $k$ choices for the first element, $k - 1$ choices for the second, $k - 2$ for the third, and so on. The total number of choices is therefore $k(k-1) \cdot \ldots \cdot 1 = k!$.

To shorten notation, we write $[k] = \{1, 2, \ldots, k\}$. For $k \leq n$, a $k$-element permutation is an injection $[k] \rightarrow [n]$. In other words, a $k$-element permutation is a list of $k$ distinct elements from $[n]$. For example, the $3$-element permutations of $\{1, 2, 3, 4\}$ are

$$
123, 124, 132, 134, 142, 143,
213, 214, 231, 234, 241, 243,
312, 314, 321, 324, 341, 342,
412, 413, 421, 423, 431, 432.
$$

This list has length $24$. In general, we have

$$
n \cdot (n - 1) \cdot \ldots \cdot (n - k + 1) = \frac{n!}{(n-k)!} \quad (7)
$$

$k$-element permutations of $n$ elements. Note that the above list contains six orderings of $\{1, 2, 3\}$, and indeed six orderings of any subset of $3$ elements. In general, we have $k!$ orderings of every $k$-element subset of the $n$ elements. It follows that we get the number of such subsets by dividing (7) through $k!$.

**Subsets.** The binomial coefficient $\binom{n}{k}$, pronounced $n$ choose $k$, is by definition the number of $k$-element subsets of a size $n$ set. As noted above, we have

$$
\binom{n}{k} = \frac{n!}{(n-k)!k!} \quad (8)
$$

We fill out the following tables with values of $\binom{n}{k}$, where the row index is the values of $n$ and the column index is the value of $k$. Values of $(\binom{n}{k})$ for $0 > k$ and $k > n$ are set to zero and are omitted from the table:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

We notice several patterns, namely $\binom{n}{0} = \binom{n}{n} = 1$ (meaning there is only one way to choose no element, and also only one way to choose all element), and $\binom{n}{k} = \binom{n}{n-k}$. This table is also known as Pascal’s Triangle. If we draw it symmetrically, then we see that each entry in the triangle is the sum of the two closest entries in the previous row:

$$
1
1 \quad 1
1 \quad 2 \quad 1
1 \quad 3 \quad 3 \quad 1
1 \quad 4 \quad 6 \quad 4 \quad 1
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1
$$

We express the above recipe of constructing an entry as the sum of two previous entries more formally. For convenience, we define $\binom{n}{k} = 0$ whenever $k < 0$, $n < 0$, or $n < k$.

**Pascal’s Relation.** $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$.

**PROOF.** We give two arguments for this identity. The first works by algebraic manipulations. We get

$$
\binom{n}{k} = \frac{(n-k)(n-1)! + k(n-1)!}{(n-k)!k!} = \frac{(n-1)!}{(n-k-1)!k!} + \frac{(n-1)!}{(n-k)!k!} = \binom{n-1}{k-1} + \binom{n-1}{k}.
$$

For the second argument, we partition the subsets of size $k$ into two collections. Let $|S| = n$ and let $a$ be an arbitrary but fixed element from $S$. Then $\binom{n}{k}$ counts the number of $k$-element subsets of $S$. To get the number of subsets that contain $a$, we count the $(k-1)$-element subsets of $S - \{a\}$, and to each such subset, we add $a$ to get a $k$-element subset. To get the number of subsets that do not contain $a$, we count the $k$-element subsets of $S - \{a\}$. The former is $\binom{n-1}{k-1}$ and the latter is $\binom{n-1}{k}$. Since the subsets
that contain a are different from the subsets that do not contain a, we get the number of k-element subsets of S equal to \( \binom{n-1}{k-1} + \binom{n-1}{k} \), as required.

One of the advantages of the binomial coefficients over the powers of the integers is that they permit simple formulas for interesting sums. As an example, we take a sum of binomial coefficients in which the k is fixed.

**LEMMA.** \( \sum_{j=k}^{n} \binom{j}{k} = \binom{n+1}{k+1} \).

**PROOF.** We use Pascal’s Relation to prove this identity. It is instructive to trace our steps graphically, as shown in Figure 3. In the first step, we replace \( \binom{n+1}{k+1} \) by the two binomial coefficients above it in Pascal’s triangle. Keeping the first term, we replace the second term by the two elements above it in the triangle, and so forth. When we leave the triangle, the term is zero and we can stop. The remaining terms are the ones in the claimed sum.

![Figure 3](image)

**Binomial coefficients.** We use binomial coefficients to find a formula for \((x+y)^n\). First, let us look at an example:

\[
(x + y)^2 = (x + y)(x + y) = xx + yx + xy + yy = x^2 + 2xy + y^2.
\]

Notice that the coefficients in the last line are the same as in the second line of Pascal’s Triangle. This is more generally the case:

**BINOMIAL THEOREM.** \((x + y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i}y^i\).

**PROOF.** If we write each term of the result before combining like terms, we list every possible way to select one x or one y from each factor. Thus, the coefficient of \(x^{n-i}y^i\) is equal to \(\binom{n}{i}\). In words, it is the number of ways we can select \(n - i\) factors to be \(x\) and have the remaining \(i\) factors to be \(y\). This is equivalent to selecting \(i\) factors to be \(y\) and have the remaining factors be \(x\).

The Binomial Theorem has a number of interesting consequences. The first states something we already know, namely that the cardinality of the power set is 2 to the cardinality of the set. Specifically, if we add the numbers of subsets of size \(k\), for all possible choices of \(k\), we get the total number of subsets:

**COROLLARY 1.** \(\sum_{i=0}^{n} \binom{n}{i} = 2^n\).

**PROOF.** Let \(x = y = 1\).

To find out how many subsets of a finite set have odd cardinality and how many have even cardinality, we consider the alternating sum of binomial coefficients:

**COROLLARY 2.** \(\sum_{i=0}^{n} (-1)^i \binom{n}{i} = 0\).

**PROOF.** Set \(x = 1\) and \(y = -1\).

This implies that there are equally many odd subsets as there are even subsets. Hence, the number of odd subsets of a set of cardinality \(n\) is \(2^{n-1}\).

**Multisets.** The difference between a set and a multiset is that the latter may contain the same element multiple times. In other words, a multiset is an unordered collection of **not necessarily distinct** elements. We can list the repetitions,

\(\circ G, T, T, A, T, G\)

or we can specify the multiplicities,

\(m(A) = 1, m(C) = 0, m(G) = 2, m(T) = 3\).

The size of a multiset is the sum of the multiplicities. For example, if we grid up a gene of \(k\) nucleotides, we get a multiset of size \(k\). We show how to count the possible multisets by considering the problem of distributing \(k\) (identical) books among \(n\) (different) shelves. The number of ways is equal to

- the number of size-\(k\) multisets of the \(n\) shelves;
- the number of ways to write \(k\) as a sum of \(n\) non-negative integers.

To get the formula, we line up our \(k\) books, then place \(n - 1\) dividers between them. The number of books between the \(i\)-th and the \((i - 1)\)-st dividers is equal to the number of books on the \(i\)-th shelf. We thus have \(n + k - 1\)
objects, $k$ books plus $n - 1$ dividers. The number of ways to choose $n - 1$ dividers from $n + k - 1$ objects is $\binom{n + k - 1}{n - 1} = \binom{n + k - 1}{k}$. Returning to the gene example, there are only $\binom{k + 3}{3} = \binom{k + 3}{k}$ different multisets we can get by grinding up genes of length $k$. This is a much smaller number than the number of genes of length $k$ we can form.