Basic Principles of Enumeration.

In this section and the next we'll see several "basic principles" of enumeration and apply them to a variety of problems, of both mathematical and general interest. The basic principles themselves are phrased in the language of sets. For a (finite) set A we will use the notation $|A|$ or $\#A$ to denote the number of elements of A, also called the *cardinality* of A. The empty set is denoted by $\varnothing$ – it is the unique set with no elements.

1. Choices – "AND" versus "OR".

In the next few pages we will often be constructing an object of some kind by repeatedly making a sequence of choices. In order to count the total number of objects we could construct we must know how many choices are available at each step, *but we must know more*: we also need to know how to combine these numbers correctly. A generally good guideline is to look for the words "AND" and "OR" in the description of the sequence of choices available. Here are a few simple examples. On a table before you are 7 books, 8 magazines, and 5 DVDs.

- *Choose a book and a DVD.* There are 7 choices for book AND 5 choices for DVD: $7 \times 5 = 35$ choices in all.
- *Choose a book or a magazine.* There are 7 choices for book OR 8 choices for magazine: $7 + 8 = 15$ choices in all.
- *Choose a book and either a magazine or a DVD.* There are $7 \times (8+5) = 91$ possible choices.
- *Choose either a book and a magazine, or a DVD.* There are $(7 \times 8) + 5 = 61$ possible choices.

Generally, "AND" corresponds to multiplication and "OR" corresponds to addition. The last two of the above examples show that it is important to determine exactly how the words "AND" and "OR" combine in the description of the problem.

From a mathematical point of view, "AND" corresponds to the Cartesian product of sets. If you choose one element of the set A AND you choose one element of the set B, then this is equivalent to choosing one element of the *Cartesian product of A and B*:

$$A \times B = \{(a, b) : \ a \in A \text{ and } b \in B\},$$

which is the set of all ordered pairs of elements (a, b) with $a \in A$ and $b \in B$. In general, the cardinalities of these sets are related by the formula

$$|A \times B| = |A| \cdot |B|.$$

Similarly, from a mathematical point of view, "OR" corresponds to the union of sets. If you choose one element of the set A OR you choose one element of the set B, then this is equivalent to choosing one element of the *union of A and B*:

$$A \cup B = \{c : \ c \in A \text{ or } c \in B\},$$

which is the set of all elements c which are either in A or in B.

It is NOT always true that $|A \cup B| = |A| + |B|$, because any elements in both A and B would be counted twice by $|A| + |B|$. The *intersection of A and B* is the set

$$A \cap B = \{c : \ c \in A \text{ and } c \in B\},$$

which is the set of all elements c which are both in A and in B. What is generally true is that

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

(This is the first instance of the Principle of Inclusion/Exclusion, which will be discussed in Section 8.) In particular, if $A \cap B = \varnothing$ then $|A \cup B| = |A| + |B|$. Thus, in order to interpret "OR" as addition, it is important to check that the sets of choices A and B have no elements in common. Such a union of sets A and B for which $A \cap B = \varnothing$ is called a *disjoint union* of sets.

When you solve enumeration problems on your own it is usually very useful to describe a choice sequence for constructing the set of objects of interest, paying close attention to the words "AND" and "OR".

2. Permutations.

A *permutation* of a set S is a list of the elements of S exactly once each, in some order. For example, the permutations of the set $\{1, a, X, g\}$ are:

$$\begin{array}{llll} 1aXg & a1Xg & X1ag & g1aX \\ 1agX & a1gX & X1ga & g1Xa \\ 1Xag & aX1g & Xa1g & ga1X \\ 1Xga & aXg1 & Xag1 & gaX1 \\ 1gaX & ag1X & Xg1a & gX1a \\ 1gXa & agX1 & Xga1 & gXa1 \end{array}$$

To construct a permutation of S we can choose any element v of S to be the first element in the permutation and follow this with any permutation of the set $S \smallsetminus \{v\}$. That is how the table above is arranged – each of the four columns corresponds to one choice of an element of $\{1, a, X, g\}$ to be the first element of the permutation. Within each column the permutations of the remaining elements are listed after the first element.

Let p_n denote the number of permutations of an n–element set S. The first sentence of the previous paragraph is translated into the equation

$$p_n = n \cdot p_{n-1},$$

provided that n is positive. (In this equation there are n choices for the first element v of the permutation, AND p_{n-1} choices for the permutation of $S \setminus \{v\}$ which follows it.) It is important to note here that each permutation of S will be produced exactly once by this construction.

Since it is easy to see that $p_1 = 1$ (and $p_2 = 2$), a simple proof by induction on n shows the following:

Theorem 2.1. *For every $n \geq 1$, the number of permutations of an n–element set is*

$$n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1.$$

The term n *factorial* is used for the number $n(n-1)\cdots 3 \cdot 2 \cdot 1$, and it is denoted by $n!$ for convenience.

We also define $0!$ to be the number of permutations of the 0–element (empty) set $\varnothing$. Since we want the equation $p_n = n \cdot p_{n-1}$ to hold when $n = 1$, and since $p_1 = 1! = 1$, we conclude that $0! = p_0 = 1$ as well.

3. Subsets.

A *subset* of a set S is a collection of some (perhaps none or all) of the elements of S, at most once each and in no particular order.

To specify a particular subset A of S, one has to decide for each element v of S whether v is in A or v is not in A. Thus we have two choices – $v \in A$ OR $v \notin A$ – for each element v of S. If $S = \{v_1, v_2, ..., v_n\}$ has n elements then the total number of choices is 2^n since we have 2 choices for v_1 AND 2 choices for v_2 AND ... AND 2 choices for v_n. Therefore....

Theorem 3.1. *For every $n \geq 0$, the number of subsets of an n–element set is 2^n.*

4. Partial permutations.

A *partial permutation* of a set S is a permutation of a subset of S. More directly, it is a list of some (perhaps none or all) of the elements of S, at most once each and listed in some particular order. We are going to count partial permutations of length k of an n–element set.

First think about the particular case $n = 6$ and $k = 3$, and the set $S = \{a, b, c, d, e, f\}$. A partial permutation of S of length 3 is a list (x, y, z) of elements of S, which must all be different. There are:
6 choices for x (since x is in S), AND
5 choices for y (since $y \in S$ but $y \neq x$), AND
4 choices for z (since $z \in S$ but $z \neq x$ and $z \neq y$).
Altogether there are $6 \cdot 5 \cdot 4 = 120$ partial permutations of $\{a, b, c, d, e, f\}$ of length 3.

This kind of reasoning works just as well in the general case. If S is an n–element set and we want to construct a partial permuation $(v_1, v_2, ..., v_k)$ of elements

of S of length k, then there are:
n choices for v_1, AND
$n-1$ choices for v_2, AND
....
$n-(k-2)$ choices for v_{k-1}, AND
$n-(k-1)$ choices for v_k.

This proves the following result.

Theorem 4.1. *For $n, k \geq 0$, the number of partial permutations of length k of an n–element set is $n(n-1)\cdots(n-k+2)(n-k+1)$.*

Notice that if $k > n$ then the number 0 will appear as one of the factors in the product $n(n-1)\cdots(n-k+2)(n-k+1)$. This makes sense, because if $k > n$ then there are no partial permutations of length k of an n–element set. On the other hand, if $0 \leq k \leq n$ then we could also write this product as

$$n(n-1)\cdots(n-k+2)(n-k+1) = \frac{n!}{(n-k)!}.$$

5. k–element subsets.

We refine the result of Section 3 by counting subsets of an n–element set S which have a particular cardinality k. So for $n, k \geq 0$ let $\binom{n}{k}$ denote the number of k–element subsets of an n–element set S. Notice that if $k < 0$ or $k > n$ then $\binom{n}{k} = 0$ because in these cases it is impossible for S to have a k–element subset. Thus we need only consider k in the range $0 \leq k \leq n$.

To count k–element subsets of S we consider another way of constructing a partial permuation of length k of S. Specifically, we can choose a k–element subset A of S AND a permutation of A. The result will be a permutation of a subset of S of length k. Since every partial permutation of length k of S is constructed exactly once in this way, this translates into the equation

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

In summary, we have proved the following result.

Theorem 5.1. *For $0 \leq k \leq n$, the number of k–element subsets of an n–element set is*

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

The numbers $\binom{n}{k}$ are read as "n choose k" and are called *binomial coefficients.*

6. Think of what the numbers mean.

Usually, when faced with a formula to prove, one's first thought is to prove it by algebraic calculations, or perhaps with an induction argument, or maybe with a combination of the two. But often that is not the easiest way, nor is it the most informative. A much better strategy is one which gives some insight into the "meaning" of all of the parts of the formula. If we can interpret all the numbers as counting things, addition as "OR", and multiplication as "AND", then we can hope to find an explanation of the formula by constructing some objects in the correct way.

For example, consider the equation, for any $n \geq 0$:

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n.$$

This could be proved by induction on n, but many more details would have to be given and the proof would not address the true "meaning" of the formula. Instead, let's interpret everything combinatorially:

- 2^n is the number of subsets of an n–element set, $\{1, 2, ..., n\}$ say;
- $\binom{n}{k}$ is the number of k–element subsets of $\{1, 2, ..., n\}$, for each $0 \leq k \leq n$;
- addition corresponds to "OR" (that is, disjoint union of sets).

So, this formula is saying that choosing a subset of $\{1, 2, ..., n\}$ (in one of 2^n ways) is equivalent to choosing a k–element subset of $\{1, 2, ..., n\}$ (in one of $\binom{n}{k}$ ways) for exactly one value of k in the range $0 \leq k \leq n$. Said that way the formula becomes self–evident, and there is nothing more to prove.

As another example, consider the equation

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

where we are using the fact that $\binom{m}{j} = 0$ if $j < 0$ or $j > m$. This equation can be proven algebraically from the formula of Theorem 5.1, and that is a good exercise which I encourage you to try. But a more informative proof interprets these numbers combinatorially as follows:

- $\binom{n}{k}$ is the number of k–element subsets of $\{1, 2, ..., n\}$;
- $\binom{n-1}{k-1}$ is the number of $(k-1)$–element subsets of $\{1, 2, ..., n-1\}$;
- $\binom{n-1}{k}$ is the number of k–element subsets of $\{1, 2, ..., n-1\}$;
- addition corresponds to disjoint union of sets.

So, this equation is saying that choosing a k–element subset A of $\{1, 2, ..., n\}$ is equivalent to either choosing a $(k-1)$–element subset of $\{1, 2, ..., n-1\}$ or a k–element subset of $\{1, 2, ..., n-1\}$. This is perhaps not as clear as the previous example, but the two cases depend upon whether the chosen k–element subset A

of $\{1, 2, ..., n\}$ is such that $n \in A$ OR $n \notin A$. If $n \in A$ then $A \smallsetminus \{n\}$ is a $(k-1)$–element subset of $\{1, 2, ..., n-1\}$, while if $n \notin A$ then A is a k–element subset of $\{1, 2, ..., n-1\}$. This construction explains the correspondence, proving the formula.

This principle – interpreting equations combinatorially and proving the formulas by describing explicit correspondences between sets of objects – is one of the most important and powerful ideas in enumeration. We'll have a lot of practice using this way of thinking in the next few weeks.

Incidentally, the equation in the second example above is a very useful recurrence relation for computing binomial coefficients quickly. Together with the facts

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

and $\binom{n}{0} = \binom{n}{n} = 1$ it can be used to grind out any number of binomial coefficients without difficulty. The resulting table is known as *Pascal's Triangle*:

$n \backslash k$	0	1	2	3	4	5	6	7	8
0	1								
1	1	1							
2	1	2	1						
3	1	3	3	1					
4	1	4	6	4	1				
5	1	5	10	10	5	1			
6	1	6	15	20	15	6	1		
7	1	7	21	35	35	21	7	1	
8	1	8	28	56	70	56	28	8	1

7. Multisets.

Imagine a bag which contains a large number of marbles of three colours – red, green, and blue, say. The marbles are all indistinguishable from one another except for their colours. There are N marbles of each colour, where N is very, very large (more precisely we should be considering the limit as $N \to \infty$). If I reach into the bag and pull out a handful of 11 marbles, I will have r red marbles, g green marbles, and b blue marbles, for some nonnegative integers (r, g, b) such that $r + g + b = 11$. How many possibile outcomes are there?

The word "multiset" is meant to suggest a set in which the objects can occur more than once. For example, the outcome $(4, 5, 2)$ in the above situation corresponds to the "set" $\{R, R, R, R, G, G, G, G, G, B, B\}$ in which R is a red marble, G is a green marble, and B is a blue marble. This is an 11–element multiset with elements of three types. The number of these multisets is the solution to the above problem.

In general, if there are t types of element then a *multiset of size n with elements of t types* is a sequence of nonnegative integers $(m_1, ..., m_t)$ such that

$$m_1 + m_2 + \cdots + m_t = n.$$

The interpretation is that m_i is the number of elements of the multiset which are of the i–th type, for each $1 \leq i \leq t$.

Theorem 7.1. *For any $n \geq 0$ and $t \geq 1$, the number of n–element multisets with elements of t types is*

$$\binom{n+t-1}{t-1}.$$

Proof. Think of what that number means! By Theorem 5.1 $\binom{n+t-1}{t-1}$ is the number of $(t-1)$–element subsets of an $(n+t-1)$–element set. So, let's write down a row of $(n+t-1)$ circles from left to right:

O O O O O O O O O O O O O

and cross out some $t-1$ of these circles to choose a $(t-1)$–element subset:

O O O O X O O O O O X O O

Now the $t-1$ crosses chop the remaining sequence of n circles into t segments of consecutive circles. (Some of these segments might be empty, which is to say of length zero.) Let m_i be the length of the i–th segment of consecutive O-s in this construction. Then $m_1 + m_2 + \cdots + m_t = n$, so that $(m_1, m_2, ..., m_t)$ is an n–element multiset with t types. Conversely, if $(m_1, m_2, ..., m_t)$ is an n–element multiset with t types then write down a sequence of m_1 O-s, then an X, then m_2 O-s, then an X, and so on, finishing with an X and then m_t O-s. The positions of the X-s will indicate a $(t-1)$–element subset of the positions $\{1, 2, ..., n+t-1\}$.

The construction of the above paragraph shows how to translate between $(t-1)$–element subsets of $\{1, 2, ..., n+t-1\}$ and n–element multisets with t types of element. This one–to–one correspondence completes the proof of the theorem. □

To answer the original question of this section, the number of 11–element multisets with elements of 3 types is $\binom{11+3-1}{3-1} = \binom{13}{2} = 78$.

8. The Principle of Inclusion/Exclusion.

In a vase is a bouquet of flowers. Each flower is (at least one of) fresh, fragrant, or colourful:

(a) 11 flowers are fresh;
(b) 7 flowers are fragrant;
(c) 8 flowers are colourful;
(d) 6 flowers are fresh and fragrant;

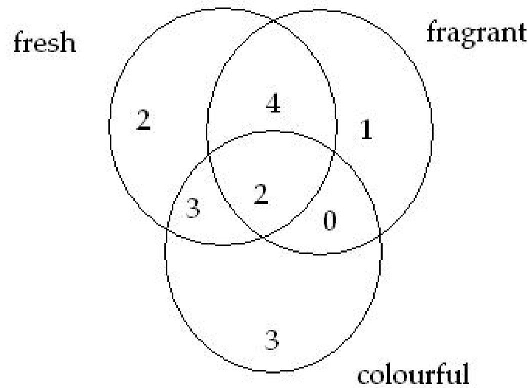

FIGURE 1. A Venn diagram for three sets.

(e) 5 flowers are fresh and colourful;
(f) 2 flowers are fragrant and colourful;
(g) 2 flowers are fresh, fragrant, and colourful.
How many flowers are in the bouquet?

The Principle of Inclusion/Exclusion is a systematic method for answering such questions, which involve overlapping conditions which can be satisfied (or not) in various combinations.

For a small problem like the one above we can work backwards as follows:
(h) from (g) and (f) there are 0 flowers which are fragrant and colourful but not fresh;
(i) from (g) and (e) there are 3 flowers which are fresh and colourful but not fragrant;
(j) from (g) and (d) there are 4 flowers which are fresh and fragrant but not colourful;
(k) from (c)(g)(h)(i) there are 3 flowers which are colourful but neither fresh not fragrant;
(ℓ) from (b)(g)(h)(j) there is 1 flower which is fragrant but neither fresh nor colourful;
(m) from (a)(g)(i)(j) there are 2 flowers which are fresh but neither fragrant nor colourful. The total number of flowers is counted by the disjoint union of the cases (g) through (m); that is $2 + 0 + 3 + 4 + 3 + 1 + 2 = 15$.

A *Venn diagram* is extremely useful for organizing this calculation. Figure 1 is a Venn diagram for the three sets involved in this question. Item (g) in the original data gives the number of flowers (2) counted in the central triangle. The subsequent

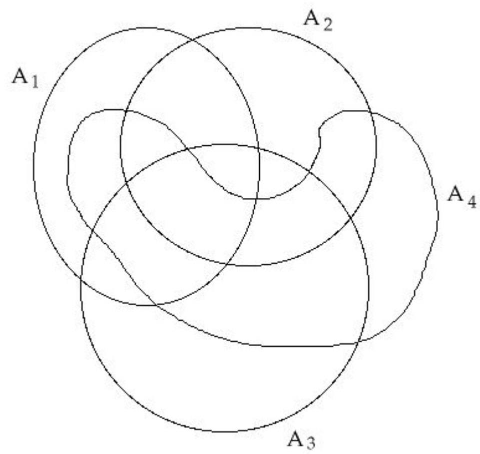

FIGURE 2. A Venn diagram for four sets.

steps (h) to (m) calculate the rest of the numbers in the diagram, moving outwards from the center.

The above works very well for three properties (fresh, fragrant, colourful) but becomes increasingly difficult to apply as the number of properties increases. Consider this alternative formula:

$$(a) + (b) + (c) - (d) - (e) - (f) + (g) = 11 + 7 + 8 - 6 - 5 - 2 + 2 = 15.$$

This looks much easier to apply, and it gives the right answer, always. Why? That is the Principle of Inclusion/Exclusion, which we now explain in general.

Let A_1, A_2,... A_t be finite sets. We want a formula for the cardinality of the union of these sets $A_1 \cup A_2 \cup \cdots \cup A_t$. First a bit of notation: if S is a nonempty subset of $\{1, 2, ..., t\}$ then let A_S denote the intersection of the sets A_i for all $i \in S$. So, for example, with this notation we have $A_{\{2,3,5\}} = A_2 \cap A_3 \cap A_5$.

Theorem 8.1. *Let A_1, A_2,..., A_t be finite sets. Then*

$$|A_1 \cup A_2 \cup \cdots \cup A_t| = \sum_{\varnothing \neq S \subseteq \{1,2,...,t\}} (-1)^{|S|-1} |A_S|.$$

(In this formula the summation is over all nonempty subsets of $\{1, 2, .., t\}$.)

We'll prove this theorem a little bit later on.

9. Combinatorial Probabilities.

We can reinterpret counting problems in terms of probabilities by making one additional hypothesis. That hypothesis is that **every possible outcome is equally likely**. The exact definition of what is an "outcome" depends on the particular problem. If Ω denotes a (finite) set of all possible outcomes, then any subset E of Ω is what a probabilist calls an *event*. The probability that a randomly chosen outcome from Ω is in the set E is $|E|/|\Omega|$ exactly because every outcome has likelyhood $1/|\Omega|$ of being chosen, and there are $|E|$ elements in E. Here are a few examples to illustrate these ideas.

• *What is the probability that a random subset of* $\{1, 2, ..., 8\}$ *has at most* 3 *elements?*
Here an outcome is a subset of $\{1, 2, ..., 8\}$, and there are $2^8 = 256$ such subsets. The number of subsets of $\{1, 2, ..., 8\}$ with at most 3 elements is

$$\binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \binom{8}{3} = 1 + 8 + 28 + 56 = 93.$$

So the probability in question is

$$\frac{93}{256} = 0.363281...$$

to six decimal places.

• *What is the probability that a random permutation of the set* $\{a, b, c, d, e, f\}$ *contains the letters* fad *as a consecutive subsequence?*
Here an outcome is a permutation of $\{a, b, c, d, e, f\}$, and there are $6! = 720$ such permutations. Those permutations of this set which contain fad as a consecutive subsequence can be constructed uniquely as the permutations of the set $\{b, c, e, fad\}$, so there are $4! = 24$ of these. Thus, the probability in question is

$$\frac{24}{720} = \frac{1}{30} = 0.03333...$$

• *What is the probability that a randomly chosen* 2*–element multiset with* t *types of element has both elements of the same type?*
The outcomes are the 2–element multisets with t types, numbering

$$\binom{2+t-1}{t-1} = \binom{t+1}{t-1} = \binom{t+1}{2} = \frac{(t+1)t}{2}$$

in total. Of these, exactly t of them have both elements of the same type – choose one of the t types and take two elements of that type. Thus, the probability in question is

$$\frac{2t}{(t+1)t} = \frac{2}{t+1}.$$

The values for the first few t are given in the following table to four decimal places:

t	1	2	3	4	5	6	7
	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500

Examples and Applications.

Here is a collection of examples and applications of the Basic Principles of Enumeration presented in the previous section. Our explanation of the Binomial Theorem in particular is meant as a warm–up towards the general definition of a "generating function", which is the last main topic of the first part of this course.

1. The Vandermonde Convolution Formula.

For natural numbers m, n, and k:

$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}.$$

For example, with $m = 4$ and $n = 2$ and $k = 3$ this says that

$$\binom{6}{3} = \binom{4}{0}\binom{2}{3} + \binom{4}{1}\binom{2}{2} + \binom{4}{2}\binom{2}{1} + \binom{4}{3}\binom{2}{0}.$$

In general, this can be proven algebraically by induction on $m + n$ but the proof is detailed and doesn't give much insight into what the formula "means". (The formula can also be deduced easily from the Binomial Theorem, as we shall see below.)

Here is a direct **combinatorial** proof, illustrating the strategy of thinking about what the numbers mean. On the LHS, $\binom{m+n}{k}$ is the number of k–element subsets S of the set $\{1, 2, ..., m + n\}$. On the RHS, the number can be produced as follows:

- choose a value of j in the range $0 \le j \le k$, and
- choose a j–element subset A of $\{1, 2, ..., m\}$, and
- choose a $(k - j)$–element subset of $\{m + 1, ..., m + n\}$.

(Notice that the set $\{m+1, ..., m+n\}$ has n elements, so it has $\binom{n}{k-j}$ subsets of size $k - j$.)

Now the formula is proved by showing a one–to–one correspondence between the k–element subsets S of $\{1, 2, ..., m+n\}$ counted on the LHS, and the pairs (A, B) of subsets counted on the RHS. This correspondence is easy to describe: given a k–element subset S of $\{1, 2, ..., m + n\}$ we let

$$A = S \cap \{1, 2, ..., m\}$$

and

$$B = S \cap \{m + 1, m + 2, ..., m + n\}.$$

Conversely, given a pair of subsets (A, B) satisfying the conditions on the RHS of the formula, we let $S = A \cup B$. After some thought you'll see that these constructions match the objects on the LHS with the objects on the RHS in a one–to–one correspondence $S \leftrightarrow (A, B)$. Since there are the same number of objects on each side, the formula is proved.

2. Common Birthdays.

Let $p(n)$ denote the probability that in a randomly chosen group of n people, at least two of them are born on the same day of the year. What does the function $p(n)$ look like?

To simplify the analysis, we will ignore the existence of leap years and assume that every year has exactly 365 days. (This introduces a tiny error but does not change the qualitative "shape" of the answer.) Moreover, we will also assume that people's birthdays are independently and uniformly distributed over the 365 days of the year, so that we can use the ideas of combinatorial probability theory. (This is a reasonable approximation, since twins are relatively rare.)

To begin with, $p(1) = 0$ since there is only $n = 1$ person in the group. Also, if $n > 365$ then $p(n) = 1$ since there are more people in the group than days in a year, so at least two people in the group must have the same birthday.

For n in the range $2 \leq n \leq 365$ it is quite complicated to analyze the probability $p(n)$ directly. However, the complementary probability $1-p(n)$ is relatively easy to compute. From the definition of $p(n)$ we see that $1 - p(n)$ is the probability that in a randomly chosen group of n people, **no two of them** are born on the same day of the year. This model is equivalent to rolling "no pair" when throwing n independent dice each with 365 sides. If we list the people in the group as P_1, P_2,..., P_n in any arbitrary order, then their birthdays must form a partial permutation of the 365 days of the year, of length n. There are $365!/(365 - n)!$ such partial permutations. Since the total number of outcomes is 365^n, we have derived the formula

$$1 - p(n) = \frac{365!}{(365 - n)!365^n}.$$

Therefore

$$p(n) = 1 - \frac{365!}{(365 - n)!365^n}.$$

To give you some feeling for what this probability looks like, here is a table of $p(n)$ (rounded to six decimal places) for selected values of $2 \leq n \leq 365$. Figure 3 gives a graph of this function.

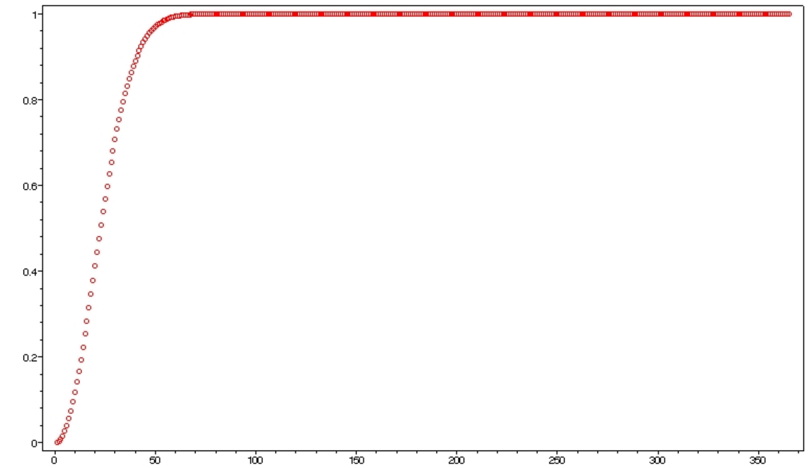

FIGURE 3. The probability of a common birthday among n random people

n	$p(n)$	n	$p(n)$	n	$p(n)$
2	0.002740	25	0.568700	70	0.999160
3	0.008204	30	0.706316	80	0.999914
4	0.016356	35	0.814383	100	1.
5	0.027136	40	0.891232	150	1.
10	0.116948	45	0.940976	200	1.
15	0.252901	50	0.970374	250	1.
20	0.411438	60	0.994123	300	1.

It is a rather surprising fact that $p(23) = 0.507297$ – so that if you randomly choose a set of 23 people on earth then there is a slightly better than 50% chance that at least two of them will have the same birthday. (Actually, since we have ignored leap years the true probability is slightly smaller – if you're going to make a bet it would be better to wager on the existence of common birthdays among a random group of $n = 30$ people. In this case you are better than a $7 : 3$ favourite to win the bet.)

3. An Example with Multisets.

A packet of *Maynard's Wine Gums* consists of a roll of 10 candies, each of which has one of five "flavours" – *Green, Yellow, Orange, Red*, or *Purple*. I especially like the purple ones. What is the chance that when I buy a packet of Wine Gums I get exactly k purple ones (for each $0 \leq k \leq 10$)?

This example is designed to illustrate the fact that the probabilities depend on which model is used to analyze the situation. There are two reasonable possibilities for this problem, which I will call the **dice model** and the **multiset model**.

In the "dice model" we keep track of the fact that the candies are stacked up in the roll from bottom to top, so there is a natural sequence $(c_1, c_2, ..., c_{10})$ of flavours one sees when the packet is opened. For example, the sequences

$$(G, P, R, Y, Y, G, O, R, Y, O)$$

and

$$(Y, G, O, P, R, R, Y, G, O, Y)$$

count as different outcomes in this model. We have a sequence of 10 candies, and a choice of one of 5 flavours for each candy, giving a total of $5^{10} = 9765625$ outcomes. (This is equivalent to rolling a sequence of 10 5–sided dice, hence the name for the model.)

In the "multiset model" we disregard the order in which the candies occur in the packet as being an inessential detail. The only important information about the packet is the number of candies of each type that it contains. For example, both of the outcomes in the previous paragraph reduce to the same multiset

$$\{G, G, Y, Y, Y, O, O, R, R, P\}$$

or $(2, 3, 2, 2, 1)$ in this model. Thus we are regarding the packet as a multiset of size 10 with 5 types of element, giving a total of $\binom{10+5-1}{5-1} = \binom{14}{4} = 1001$ outcomes.

Notice that the number of outcomes in the dice model is much larger than in the multiset model. It should come as no surprise, then, that the probabilities we compute will depend strongly on which of these two models we consider. (The *true values* for the probabilities depend on the details of the manufacturing process by which the packets are made. These cannot be calculated, but must be measured instead.)

Let us consider the dice model first, and let $d(k)$ denote the probability of getting exactly k purple candies in a packet. There are $\binom{10}{k}$ choices for the positions of these k purple candies, and $(5-1)^{10-k}$ choices for the sequence of (non–purple) flavours of the other $10-k$ candies. This gives a total of $\binom{10}{k}4^{10-k}$ outcomes with exactly k purple candies in this model. Therefore,

$$d(k) = \binom{10}{k}\frac{4^{10-k}}{5^{10}}$$

for each $0 \le k \le 10$. Here is a table of these probabilities (rounded to six decimal places).

k	$d(k)$	k	$d(k)$
0	0.107374		
1	0.268435	6	0.005505
2	0.301990	7	0.000786
3	0.201327	8	0.000074
4	0.088080	9	0.000004
5	0.026424	10	0.000000

Next let's consider the multiset model, and let $m(k)$ denote the probability of getting exactly k purple candies in a packet. If we have k purple candies then the rest of the candies form a multiset of size $10-k$ with elements of 4 types, so there are $\binom{10-k+4-1}{4-1} = \binom{13-k}{3}$ such outcomes in this model. Therefore,

$$m(k) = \frac{\binom{13-k}{3}}{\binom{14}{4}}$$

for each $0 \le k \le 10$. Here is a table of these probabilities (rounded to six decimal places).

k	$m(k)$	k	$m(k)$
0	0.285714		
1	0.219780	6	0.034965
2	0.164835	7	0.019980
3	0.119880	8	0.009990
4	0.083916	9	0.003996
5	0.055944	10	0.000999

The differences between the two models are clearly seen in these tables.

In closing, I'd like to make two points about these models.

First, given a multiset $(m_1, ..., m_t)$ of size n with elements of t types, the number of outcomes in the dice model which "reduce" to this multiset is

$$\binom{n}{m_1, ..., m_t} = \frac{n!}{m_1! \cdot m_2! \cdots m_t!},$$

called a *multinomial coefficient*. This can be seen intuitively by arranging the n elements of the multiset in a line in one of $n!$ ways, and noticing that since we can't tell the m_i elements of type i apart we can freely rearrange them in $m_i!$ ways without changing the arrangement. (A more careful argument goes by induction on t using the case $t = 2$ of binomial coefficients.)

For the second point, the above analysis of the multiset model can be generalized to prove the following identity: for any integers $n \geq 1$ and $t \geq 2$:

$$\binom{n+t-1}{t-1} = \sum_{k=0}^{n} \binom{n-k+t-2}{t-2}.$$

(This is a good exercise – think about what the numbers mean!)

4. Poker Hands.

Poker is played with a standard deck of 52 cards, divided into four *suits* (Spades $\spadesuit$, Hearts $\heartsuit$, Diamonds $\diamondsuit$, and Clubs $\clubsuit$) with 13 cards in each suit:
A (Ace), 2, 3, 4, 5, 6, 7, 8, 9, 10, J (Jack), Q (Queen), K (King).
An Ace can be *high* (above K) or *low* (below 2) at the player's choice. Many variations on the game exist, but the common theme is to make the best 5–card hand according to the ranking of poker hands. This ranking is determined by how unlikely it is to be dealt such a hand. The types of hand are as follows:

• Straight Flush: this is five cards of the same suit with consecutive values.
For example, $8\heartsuit\ 9\heartsuit\ 10\heartsuit\ J\heartsuit\ Q\heartsuit$.
• Four of a Kind (or Quads): this is four cards of the same value, with any fifth card.
For example, $7\spadesuit\ 7\heartsuit\ 7\diamondsuit\ 7\clubsuit\ 4\diamondsuit$.
• Full House (or Tight, or Boat): this is three cards of the same value, and a pair of cards of another value.
For example, $9\spadesuit\ 9\heartsuit\ 9\diamondsuit\ A\diamondsuit\ A\clubsuit$.
• Flush: this is five cards of the same suit, but not with consecutive values.
For example, $3\heartsuit\ 7\heartsuit\ 10\heartsuit\ J\heartsuit\ K\heartsuit$.
• Straight: this is five cards with consecutive values, but not of the same suit.
For example, $8\heartsuit\ 9\clubsuit\ 10\spadesuit\ J\heartsuit\ Q\diamondsuit$.
• Three of a Kind (or Trips): this is three cards of the same value, and two more cards not of the same value.
For example, $8\spadesuit\ 8\heartsuit\ 8\diamondsuit\ K\diamondsuit\ 5\clubsuit$.
• Two Pair: this is self–explanatory.
For example, $J\heartsuit\ J\clubsuit\ 6\diamondsuit\ 6\clubsuit\ A\spadesuit$.
• One Pair: this is also self–explanatory.
For example, $Q\spadesuit\ Q\diamondsuit\ 8\diamondsuit\ 7\clubsuit\ 2\spadesuit$.
• Busted Hand: this is anything not covered above.
For example, $K\spadesuit\ Q\diamondsuit\ 8\diamondsuit\ 7\clubsuit\ 2\spadesuit$.
Of course there are $\binom{52}{5} = 2598960$ possible 5–element subsets of a standard deck of 52 cards, so this is the total number of possible poker hands. How many of these hands are of each of the above types? The answers are easily availabe on the WWWeb, so there's no sense trying to keep them secret. Here they are: N is the

number of outcomes of each type, and $p = N/\binom{52}{5}$ is the probability of each type of outcome.

Hand	N	p
Straight Flush	40	0.000015
Quads	624	0.000240
Full House	3744	0.001441
Flush	5108	0.001965
Straight	10200	0.003925
Trips	54912	0.021128
Two Pair	123552	0.047539
One Pair	1098240	0.422569
Busted	1302540	0.501177

The derivation of these numbers is an excellent exercise, so we will do only three of the cases – Flush, Straight, and Busted – as illustrations.

• To construct a Flush hand there are 4 choices for suit, and $\binom{13}{5}$ choices for 5 cards from that suit. However, 10 of these choices ($A2345$, 23456, ... up to $10JQKA$) for card values lead to straight flushes and must be discounted. Hence the total number of flushes is $4[\binom{13}{5} - 10] = 5108$.

• To construct a Straight hand there are 10 choices for the consecutive ranks of the cards ($A2345$, 23456, ... up to $10JQKA$), and 4^5 choices for the suits on the cards. However, 4 of these choices for suits give all five cards the same suit – these lead to straight flushes and must be discounted. Hence the total number of straights is $10 \cdot [4^5 - 4] = 10200$.

• To construct a Busted hand there are $\binom{13}{5} - 10$ choices for 5 values of cards which are not consecutive (no straight) and have no pairs. Having chosen these values there are $4^5 - 4$ choices for the suits on the cards which do not give all five cards the same suit (no flush). Hence the total number of busted hands is $[\binom{13}{5} - 10] \cdot [4^5 - 4] = 1302540$.

Here is another example problem from a poker variation called *Texas Hold'em.* (Since the 1970s this has become the most popular form of "serious poker".) In this game the cards are dealt as follows:

• each player is dealt two cards, and there is a round of betting;

• three cards are dealt in the middle of the table (the "flop"), and there is a round of betting;

• another card is dealt in the middle of the table (the "turn"), and there is a round of betting;

• a fifth card is dealt in the middle of the table (the "river"), and there is a final round of betting.

After this, on the showdown each player can choose the best 5–card hand from among the seven cards (s)he can see (the two in hand and the five on the table).

Flushes are strong hands in Texas Hold'em – especially Ace-high flushes. Let's suppose that you have been dealt the $A\diamondsuit 9\diamondsuit$. *What is the chance that the flop comes with at least two diamonds?* Since you hold two cards, so there are 50 cards remaining in the deck. Three of these cards will appear on the flop, so there are $\binom{50}{3} = 19600$ possible flops. (Some of these 50 cards are in your opponents' hands and thus can not appear on the flop – but you have no information about which cards they hold, so that *given the information you know* each of the $\binom{50}{3}$ flops is equally likely.) Of the 50 remaining cards 11 are diamonds (since you hold two diamonds). The Vandermonde convolution formula with $m = 11$ and $n = 39$ and $k = 3$ implies that

$$\binom{50}{3} = \binom{11}{0}\binom{39}{3} + \binom{11}{1}\binom{39}{2} + \binom{11}{2}\binom{39}{1} + \binom{11}{3}\binom{39}{0}.$$

These four terms correspond to the cases that the flop comes with 0, 1, 2, or 3 of the remaining 11 diamonds. Working out the numbers gives

$$19600 = 9139 + 8151 + 2145 + 165.$$

That is, there are $2145 + 165 = 2310$ flops with at least two diamonds. The probability of the flop "hitting" your hand in this way is $2310/19600 = 0.117857$, a little better than one chance in nine, or very close to $8:1$ odds against.

If you want to continue this example further, here are some good questions with relevance to the actual play of the game. Suppose that you hold $A\diamondsuit 9\diamondsuit$ as above and the flop has in fact come with exactly two diamonds.

- What is the probability that the turn card is a diamond?
- If the turn card is not a diamond, then what is the probability that the river card is a diamond?

(In the above two cases you make your flush, and have a very strong hand.)

- Assume that the turn card is a diamond, and that after the turn there is no pair showing among the six cards you can see. What is the probability that the river card will pair one of the four cards already on the table?

(This *could* result with your flush being "rivered" by someone else making a full house or quads with the last card.)

5. Derangements.

A group of eight people meet for dinner at a fancy restaurant and check their coats at the door. After a delicious gourmet meal the group leaves, and on the way out the eight coats are returned to the eight people completely at random by an incompetent clerk. What is the probability that no-one gets the correct coat? This is a "classical" example, known as the *derangement problem.*

Of course, we want to solve the derangement problem not just for $n = 8$ people, but for any value of n. To state the problem in mathematically precise language, imagine that the people are listed P_1, P_2, ..., P_n in any arbitrary order. We can record who gets whose coat by a sequence of numbers $(c_1, c_2, ..., c_n)$ in which $c_i = j$ means that P_i was given the coat belonging to P_j. The sequence $(c_1, c_2, ..., c_n)$ will thus contain each of the numbers $1, 2, ..., n$ exactly once in some order. In other words, $(c_1, ..., c_n)$ is a permutation of the set $\{1, 2, ..., n\}$, and we assume that this permutation is chosen randomly by the incompetent clerk. Person i gets the correct coat exactly when $c_i = i$. Thus, in general the derangement problem is to determine – for a random permutation $(c_1, ..., c_n)$ of $\{1, 2, ..., n\}$ – the probability that $c_i \neq i$ for all $1 \leq i \leq n$.

For small values of n the derangement problem can be analyzed directly, but complications arise as n gets larger. In fact, this example is perfectly designed to illustrate the Principle of Inclusion/Exclusion. To see how this applies, for each $1 \leq i \leq n$ let A_i be the set of permutations of $\{1, ..., n\}$ such that $c_i = i$. That is, A_i is the set of ways in which the coats are returned and person i gets the correct coat. From that interpretation, the union of sets $A_1 \cup A_2 \cup \cdots \cup A_n$ is the set of ways in which the coats are returned and **at least one person** gets the correct coat. Therefore, the complementary set of permutations gives those ways of returning the coats so that **no-one** gets the correct coat. The number of these *derangements of n objects* is thus

$$n! - |A_1 \cup A_2 \cup \cdots \cup A_n|.$$

It remains to apply the Principle of Inclusion/Exclusion to determine $|A_1 \cup \cdots \cup A_n|$. To do this we need to determine $|A_S|$ for every nonempty subset $\varnothing \neq S \subseteq \{1, 2, ..., n\}$. Consider the example $n = 8$ and $S = \{2, 3, 6\}$. In this case $A_{\{2,3,6\}} = A_2 \cap A_3 \cap A_6$ is the set of those permutations of $\{1, ..., 8\}$ such that $c_2 = 2$ and $c_3 = 3$ and $c_6 = 6$. Such a permutation looks like $\square\ 2\ 3\ \square\ \square\ 6\ \square\ \square$ where the boxes are filled with the numbers $\{1, 4, 5, 7, 8\}$ in some order. Since there are $5!$ permutations of $\{1, 4, 5, 7, 8\}$ it follows that $|A_{\{2,3,6\}}| = 5! = 120$ in this case. The general case is similar. If $\varnothing \neq S \subseteq \{1, 2, ..., n\}$ is a k–element subset then the permutations of $\{1, 2, ..., n\}$ in A_S are obtained by fixing $c_i = i$ for all $i \in S$, and permuting the remaining $n - k$ elements of $\{1, ..., n\} \smallsetminus S$ among themselves. Since there are $(n - k)!$ such permutations we see that $|A_S| = (n - k)!$.

Since $|A_S| = (n-k)!$ for every k–element subset of $\{1, 2, ..., n\}$ – and there are $\binom{n}{k}$ such k–element subsets – the Principle of Inclusion/Exclusion implies that

$$\begin{aligned} |A_1 \cup \cdots \cup A_n| &= \sum_{\varnothing \neq S \subseteq \{1,...,n\}} (-1)^{|S|-1} |A_S| \\ &= \sum_{k=1}^{n} \binom{n}{k} (-1)^{k-1} (n-k)! = n! \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!}. \end{aligned}$$

Finally, it follows that the number of derangements of n objects is

$$n! - n!\sum_{k=1}^{n}\frac{(-1)^{k-1}}{k!} = n!\sum_{k=0}^{n}\frac{(-1)^k}{k!}.$$

Since the total number of permutations of n objects is $n!$, the probability that a randomly chosen permutation of $\{1, 2, ..., n\}$ is a derangement is

$$D_n = \sum_{k=0}^{n}\frac{(-1)^k}{k!}.$$

The following table lists the first several values of the function D_n (with the decimals rounded to six places).

n	D_n	D_n
0	1/1	1.
1	0/1	0.000000
2	1/2	0.500000
3	1/3	0.333333
4	3/8	0.375000
5	11/30	0.366667
6	53/144	0.368056
7	103/280	0.367857
8	2119/5760	0.367882
9	16687/45630	0.367879
10	16481/44800	0.367879

Notice that for $n \geq 7$ the value of D_n changes very little. If you recall the Taylor series expansion of the exponential function

$$e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!}$$

then it is easy to see that as n tends to infinity the probability D_n approaches the limiting value of $e^{-1} = .3678794412....$

In summary, for the original $n = 8$ derangement problem, the probability that no-one gets their own coat is very close to 36.79%.

6. The Binomial Theorem.

For any natural number n,

$$(x + y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}.$$

This formula is an identity between two polynomials in the variables x and y. You probably have seen a proof of it by induction on n, but we are going to prove it here in a completely different way. This way of thinking is a first step toward understanding the method of **generating functions** which will be the last major concept in this half of the course.

To begin with, look at what the formula specializes to when $x = 1$ and $y = 1$. We get

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

which we proved by counting the subsets of $\{1, 2, ..., n\}$ in two different ways. This suggests that we use the same idea here, but keep track of more information.

When we showed that $\{1, 2, ..., n\}$ has 2^n subsets we argued as follows. In order to construct a subset $S \subseteq \{1, ..., n\}$ we have to make a choice for each element $1 \leq i \leq n$: either $i \in S$ or $i \notin S$. Since there are 2 choices for each element and there are n elements of $\{1, ..., n\}$, the result is a total of 2^n possible choices for S. Now let's record these choices by a sequence $(a_1, a_2, ..., a_n)$ in which $a_i = x$ if $i \in S$ and $a_i = y$ if $i \notin S$. So, for example, with $n = 9$ the sequence $(x, y, y, x, y, x, x, y, x)$ records the choices for the subset $\{1, 4, 6, 7, 9\}$.

To prove the Binomial Theorem we multiply the entries of such a sequence together, and sum the result over all subsets of $\{1, 2, ..., n\}$. For each $1 \leq i \leq n$ we can choose either $a_i = x$ or $a_i = y$, giving a contribution of $x + y$. Doing this for each $1 \leq i \leq n$ multiplies these contributions together, yielding $(x + y)^n$. Each of the 2^n subsets of $\{1, 2, ..., n\}$ corresponds to one way of choosing a term from this product of binomials, explaining the LHS of the formula. To interpret the RHS notice that each of the k–element subsets S of $\{1, 2, ..., n\}$ will contribute $x^k y^{n-k}$ to the summation (since we choose an x for every $i \in S$ and a y for every $i \notin S$). There are $\binom{n}{k}$ k–element subsets of $\{1, ..., n\}$. As k ranges from 0 to n this counts every subset of $\{1, ..., n\}$ exactly once. Since every subset of $\{1, ..., n\}$ contributes the same amount to both the LHS and the RHS, the two sides of the formula are equal.

Once you are comfortable with the idea of the above argument, you can see that it is summarized nicely by the following "one–line calculation":

$$(x + y)^n = \sum_{S \subseteq \{1,...,n\}} x^{|S|} y^{n-|S|} = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

As mentioned above, the Binomial Theorem easily implies the Vandermonde Convolution formula. To see this, begin with the obvious identity of polynomials

$$(x + 1)^{m+n} = (x + 1)^m \cdot (x + 1)^n$$

and use the Binomial Theorem (with $y = 1$) to expand each of the factors.

$$\begin{aligned}
\sum_{k=0}^{m+n} \binom{m+n}{k} x^k &= \left(\sum_{j=0}^{m} \binom{m}{j} x^j\right) \left(\sum_{i=0}^{n} \binom{n}{i} x^i\right) \\
&= \sum_{j=0}^{m} \sum_{i=0}^{n} \binom{m}{j} \binom{n}{i} x^{j+i} \\
&= \sum_{k=0}^{m+n} \sum_{j=0}^{k} \binom{m}{j} \binom{n}{k-j} x^k.
\end{aligned}$$

(The last step is accomplished by re–indexing the double summation.) Since the polynomials on the LHS and on the RHS are equal, they must have the same coefficients. By comparing the coefficients of x^k on both sides we see that

$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j} \binom{n}{k-j},$$

as claimed.

The Binomial Theorem can be generalized to more than two variables, and the result is known as the *Multinomial Theorem*:

$$(x_1 + x_2 + \cdots + x_t)^n = \sum_{m_1+\cdots+m_t=n} \binom{n}{m_1, \ldots, m_t} x_1^{m_1} \cdots x_t^{m_t}.$$

The summation on the RHS is over all multisets of size n with elements of t types. It is a good exercise to generalize the above proof of the Binomial Theorem to prove this.