Chapter 4

Some Counting Problems; Multinomial Coefficients, The Inclusion-Exclusion Principle, Sylvester's Formula, The Sieve Formula

4.1 Counting Permutations and Functions

In this short section, we consider some simple counting problems.

Let us begin with permutations. Recall that a *permutation* of a set, A, is any bijection between A and itself.

If A is a finite set with n elements, we mentioned earlier (without proof) that A has $n!$ permutations, where the *factorial function*, $n \mapsto n!$ $(n \in \mathbb{N})$, is given recursively by:

$$0! = 1$$
$$(n+1)! = (n+1)n!.$$

The reader should check that the existence of the function, $n \mapsto n!$, can be justified using the Recursion Theorem (Theorem 2.5.1).

Proposition 4.1.1 *The number of permutations of a set of n elements is $n!$.*

Let us also count the number of functions between two finite sets.

Proposition 4.1.2 *If A and B are finite sets with $|A| = m$ and $|B| = n$, then the set of function, B^A, from A to B has n^m elements.*

As a corollary, we determine the cardinality of a finite power set.

Corollary 4.1.3 *For any finite set, A, if $|A| = n$, then $|2^A| = 2^n$.*

Computing the value of the factorial function for a few inputs, say $n = 1, 2 \ldots, 10$, shows that it grows very fast. For example,

$$10! = 3,628,800.$$

Is it possible to quantify how fast factorial grows compared to other functions, say n^n or e^n?

Remarkably, the answer is yes. A beautiful formula due to James Stirling (1692-1770) tells us that

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n,$$

which means that

$$\lim_{n \to \infty} \frac{n!}{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n} = 1.$$
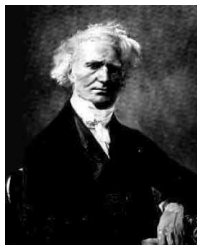
Figure 4.1: Jacques Binet, 1786-1856

Here, of course,

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} + \cdots$$

the base of the natural logarithm.

It is even possible to estimate the error. It turns out that

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\lambda_n},$$

where

$$\frac{1}{12n + 1} < \lambda_n < \frac{1}{12n},$$

a formula due to Jacques Binet (1786-1856).

Let us introduce some notation used for comparing the rate of growth of functions.

We begin with the *"Big oh" notation*.

Given any two functions, $f \colon \mathbb{N} \to \mathbb{R}$ and $g \colon \mathbb{N} \to \mathbb{R}$, we say that *$f$ is $O(g)$ (or $f(n)$ is $O(g(n))$)* iff there is some $N > 0$ and a constant $c > 0$ such that

$$|f(n)| \leq c|g(n)|, \quad \text{for all} \quad n \geq N.$$

In other words, for n large enough, $|f(n)|$ is bounded by $c|g(n)|$. We sometimes write $n >> 0$ to indicate that n is "large."

For example λ_n is $O(\frac{1}{12n})$. By abuse of notation, we often write $f(n) = O(g(n))$ even though this does not make sense.

The *"Big omega" notation* means the following: *f is $\Omega(g)$ (or $f(n)$ is $\Omega(g(n))$)* iff there is some $N > 0$ and a constant $c > 0$ such that

$$|f(n)| \geq c|g(n)|, \quad \text{for all} \quad n \geq N.$$

The reader should check that $f(n)$ is $O(g(n))$ iff $g(n)$ is $\Omega(f(n))$.

We can combine O and Ω to get the *"Big theta" notation*: f *is* $\Theta(g)$ *(or* $f(n)$ *is* $\Theta(g(n)))$ iff there is some $N > 0$ and some constants $c_1 > 0$ and $c_2 > 0$ such that

$$c_1|g(n)| \leq |f(n)| \leq c_2|g(n)|, \quad \text{for all} \quad n \geq N.$$

Finally, the *"Little oh" notation* expresses the fact that a function, f, has much slower growth than a function g.

We say that f *is* $o(g)$ *(or* $f(n)$ *is* $o(g(n)))$ iff

$$\lim_{n \to \infty} \frac{f(n)}{g(n)} = 0.$$

For example, $\sqrt{n}$ is $o(n)$.

4.2   Counting Subsets of Size k; Binomial and Multinomial Coefficients

Let us now count the number of subsets of cardinality k of a set of cardinality n, with $0 \leq k \leq n$.

Denote this number by $\binom{n}{k}$ (say "n choose k"). Actually, in the proposition below, it will be more convenient to assume that $k \in \mathbb{Z}$.

Proposition 4.2.1 *For all $n \in \mathbb{N}$ and all $k \in \mathbb{Z}$, if $\binom{n}{k}$ denotes the number of subsets of cardinality k of a set of cardinality n, then*

$$\binom{0}{0} = 1$$
$$\binom{n}{k} = 0 \quad if \quad k \notin \{0, 1, \ldots, n\}$$
$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \quad (n \geq 1,\ 0 \leq k \leq n).$$

The numbers $\binom{n}{k}$ are also called *binomial coefficients*, because they arise in the expansion of the binomial expression $(a+b)^n$, as we will see shortly.

The binomial coefficients can be computed inductively using the formula

$$\binom{\color{red}n}{\color{red}k} = \binom{\color{blue}n-1}{\color{blue}k} + \binom{\color{cyan}n-1}{\color{cyan}k-1}$$

(sometimes known as *Pascal's recurrence formula*) by forming what is usually called *Pascal's triangle*, which is based on the recurrence for $\binom{n}{k}$:

n	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\binom{n}{9}$	$\binom{n}{10}$	$\cdots$
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			
9	1	9	36	84	126	126	84	36	9	1		
10	1	10	45	120	210	252	210	120	45	10	1	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Figure 4.2: Blaise Pascal, 1623-1662

We can also give the following explicit formula for $\binom{n}{k}$ in terms of the factorial function:

Proposition 4.2.2 *For all $n, k \in \mathbb{N}$, with $0 \le k \le n$, we have*

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

Then, it is very easy to see that

$$\binom{n}{k} = \binom{n}{n-k}.$$

Remarks:

(1) The binomial coefficients were already known in the twelfth century by the Indian Scholar Bhaskra. Pascal's triangle was taught back in 1265 by the Persian philosopher, Nasir-Ad-Din.

(2) The formula given in Proposition 4.2.2 suggests generalizing the definition of the binomial coefficients to upper indices taking *real* values.

Indeed, for all $r \in \mathbb{R}$ and all integers, $k \in \mathbb{Z}$, we can set

$$\binom{r}{k} = \begin{cases} \dfrac{r^{\underline{k}}}{k!} = \dfrac{r(r-1)\cdots(r-k+1)}{k(k-1)\cdots 2 \cdot 1} & \text{if } k \geq 0 \\ 0 & \text{if } k < 0. \end{cases}$$

Note that the expression in the numerator, $r^{\underline{k}}$, stands for the product of the k terms

$$\overbrace{r(r-1)\cdots(r-k+1)}^{k \text{ terms}}.$$

By convention, the value of this expression is 1 when $k = 0$, so that $\binom{r}{0} = 1$.

The expression $\binom{r}{k}$ can be viewed as a polynomial of degree k in r. The generalized binomial coefficients allow for a useful extension of the binomial formula (see next) to real exponents.

However, beware that the symmetry identity fails when r is not a natural number and that the formula in Proposition 4.2.2 (in terms of the factorial function) only makes sense for natural numbers.

We now prove the "binomial formula" (also called "binomial theorem").

Proposition 4.2.3 *(Binomial Formula) For all $n \in \mathbb{N}$ and for all reals, $a, b \in \mathbb{R}$, (or more generally, any two commuting variables a, b, i.e., satisfying $ab = ba$), we have the formula:*

$$(a + b)^n = a^n + \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 + \cdots$$
$$+ \binom{n}{k} a^{n-k}b^k + \cdots + \binom{n}{n-1} ab^{n-1} + b^n.$$

The above can be written concisely as

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

Remark: The binomial formula can be generalized to the case where the exponent, r, is a real number (even negative). This result is usually known as the *binomial theorem* or *Newton's generalized binomial theorem*.

Formally, the binomial theorem states that

$$(a+b)^r = \sum_{k=0}^{\infty} \binom{r}{k} a^{r-k} b^k, \qquad r \in \mathbb{N} \quad \text{or} \quad |b/a| < 1.$$

Observe that when r is not a natural number, the right-hand side is an infinite sum and the condition $|b/a| < 1$ insures that the series converges.

For example, when $a = 1$ and $r = 1/2$, if we rename b as x, we get

$$
\begin{aligned}
(1+x)^{\frac{1}{2}} &= \sum_{k=0}^{\infty} \binom{\frac{1}{2}}{k} x^k \\
&= 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{1}{2} \left(\frac{1}{2} - 1\right) \cdots \left(\frac{1}{2} - k + 1\right) x^k \\
&= 1 + \sum_{k=1}^{\infty} (-1)^{k-1} \frac{1 \cdot 3 \cdot 5 \cdots (2k-3)}{2 \cdot 4 \cdot 6 \cdots 2k} x^k \\
&= 1 + \sum_{k=1}^{\infty} \frac{(-1)^{k-1}(2k)!}{2^{2k}(2k-1)(k!)^2} x^k, \\
&= 1 + \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{2^{2k}(2k-1)} \binom{2k}{k} x^k \\
&= 1 + \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{2^{2k}} \frac{1}{k} \binom{2k-2}{k-1} x^k,
\end{aligned}
$$

which converges if $|x| < 1$.

The first few terms of this series are

$$(1 + x)^{\frac{1}{2}} = 1 + \frac{1}{2} x - \frac{1}{8} x^2 + \frac{1}{16} x^3 - \frac{5}{128} x^4 + \cdots$$

For $r = -1$, we get the familiar geometric series

$$\frac{1}{1 + x} = 1 - x + x^2 - x^3 + \cdots + (-1)^k x^k + \cdots ,$$

which converges if $|x| < 1$.

Remark: The numbers,

$$C_n = \frac{1}{n + 1} \binom{2n}{n},$$

are the *Catalan numbers*. They are the solution of many counting problems in combinatorics.

Proposition 4.2.4 *The number of injections between a set, A, with m elements and a set, B, with n elements, where $m \leq n$, is given by*
$$\frac{n!}{(n-m)!} = n(n - 1) \cdots (n - m + 1).$$

Counting the number of surjections between a set with n elements and a set with p elements, where $n \geq p$, is harder.

We state the following formula without giving a proof right now. Finding a proof of this formula is an interesting exercise.

We will give a quick proof using the Inclusion-Exclusion Principle in Section 4.4.

Proposition 4.2.5 *The number of surjections, S_{np}, between a set, A, with n elements and a set, B, with p elements, where $n \geq p$, is given by*

$$S_{np} = p^n - \binom{p}{1}(p-1)^n + \binom{p}{2}(p-2)^n + \cdots$$
$$+ (-1)^{p-1}\binom{p}{p-1}.$$

Remarks:

1. It can be shown that S_{np} satisfies the following peculiar version of Pascal's recurrence formula:

$$S_{np} = p(S_{n-1\,p} + S_{n-1\,p-1}), \qquad p \geq 2,$$

   and, of course, $S_{n\,1} = 1$ and $S_{np} = 0$ if $p > n$.

   Using this recurrence formula and the fact that $S_{nn} = n!$, simple expressions can be obtained for $S_{n+1\,n}$ and $S_{n+2\,n}$.

2. The numbers, S_{np}, are intimately related to the so-called *Stirling numbers of the second kind*, denoted $\left\{ {n \atop p} \right\}$, $S(n, p)$, or $S_n^{(p)}$, which count the number of partitions of a set of n elements into p nonempty pairwise disjoint blocks (see Section 5.5). In fact,

$$S_{np} = p!\left\{ {n \atop p} \right\}.$$

The Stirling numbers, $\left\{{n \atop p}\right\}$, satisfy a recurrence equation which is another variant of Pascal's recurrence formula:

$$\left\{{n \atop 1}\right\} = 1$$

$$\left\{{n \atop n}\right\} = 1$$

$$\left\{{n \atop p}\right\} = \left\{{n-1 \atop p-1}\right\} + p\left\{{n-1 \atop p}\right\} \qquad (1 \le p < n).$$

The total numbers of partitions of a set with $n \ge 1$ elements is given by the *Bell number*,

$$b_n = \sum_{p=1}^{n} \left\{{n \atop p}\right\}.$$

There is a recurrence formula for the Bell numbers but it is complicated and not very useful because the formula for b_{n+1} involves all the previous Bell numbers.

A good reference for all these special numbers is Graham, Knuth and Patashnik [9], Chapter 6.

Figure 4.3: Eric Temple Bell, 1883-1960 (left) and Donald Knuth, 1938- (right)

The binomial coefficients can be generalized as follows. For all $n, m, k_1, \ldots, k_m \in \mathbb{N}$, with $k_1 + \cdots + k_m = n$ and $m \geq 2$, we have the *multinomial coefficient*,

$$\binom{n}{k_1 \cdots k_m},$$

which counts the number of ways of splitting a set of n elements into an ordered sequence of m disjoint subsets, the ith subset having $k_i \geq 0$ elements.

Such sequences of disjoint subsets whose union is $\{1, \ldots, n\}$ itself are sometimes called *ordered partitions*.

Beware that some of the subsets in an ordered partition may be empty, so we feel that the terminology "partition" is confusing since as will see in Section 5.5, the subsets that form a partition are never empty.

Note that when $m = 2$, the number of ways of splitting a set of n elements into two disjoint subsets where the first subset has k_1 elements and the second subset has $k_2 = n - k_1$ elements is precisely the number of subsets of size k_1 of a set of n elements, that is

$$\binom{n}{k_1 \; k_2} = \binom{n}{k_1}.$$

Observe that the order of the m subsets matters.

For example, for $n = 5$, $m = 4$, $k_1 = 2$ and $k_2 = k_3 = k_4 = 1$, the sequences of subsets
$(\{1, 2\}, \{3\}, \{4\}, \{5\})$, $(\{1, 2\}, \{3\}, \{5\}, \{4\})$,
$(\{1, 2\}, \{5\}, \{3\}, \{4\})$, $(\{1, 2\}, \{4\}, \{3\}, \{5\})$,
$(\{1, 2\}, \{4\}, \{5\}, \{3\})$, $(\{1, 2\}, \{5\}, \{4\}, \{3\})$
are all different and they correspond to the same partition, $\{\{1, 2\}, \{3\}, \{4\}, \{5\}\}$.

Proposition 4.2.6 *For all $n, m, k_1, \ldots, k_m \in \mathbb{N}$, with $k_1 + \cdots + k_m = n$ and $m \geq 2$, we have*

$$\binom{n}{k_1 \cdots k_m} = \frac{n!}{k_1! \cdots k_m!}.$$

As in the binomial case, it is convenient to set

$$\binom{n}{k_1 \cdots k_m} = 0$$

if $k_i < 0$ or $k_i > n$, for any i, with $1 \leq i \leq m$. Then, Proposition 4.2.1 is generalized as follows:

Proposition 4.2.7 *For all $n, m, k_1, \ldots, k_m \in \mathbb{N}$, with $k_1 + \cdots + k_m = n$, $n \geq 1$ and $m \geq 2$, we have*

$$\binom{n}{k_1 \cdots k_m} = \sum_{i=1}^{m} \binom{n-1}{k_1 \cdots (k_i - 1) \cdots k_m}.$$

Remark: Proposition 4.2.7 shows that Pascal's triangle generalizes to "higher dimensions", that is, to $m \geq 3$.

Indeed, it is possible to give a geometric interpretation of Proposition 4.2.7 in which the multinomial coefficients corresponding to those $k_1, \ldots, k_m$ with $k_1 + \cdots + k_m = n$ lie on the hyperplane of equation $x_1 + \cdots + x_m = n$ in $\mathbb{R}^m$, and all the multinomial coefficients for which $n \leq N$, for any fixed N, lie in a generalized tetrahedron called a *simplex*.

When $m = 3$, the multinomial coefficients for which $n \leq N$ lie in a tetrahedron whose faces are the planes of equations, $x = 0$; $y = 0$; $z = 0$; and $x + y + z = N$.

We have also the following generalization of Proposition 4.2.3:

Proposition 4.2.8 *(Multinomial Formula) For all* $n, m \in \mathbb{N}$ *with* $m \geq 2$, *for all pairwise commuting variables* $a_1, \ldots, a_m$, *we have*

$$(a_1 + \cdots + a_m)^n = \sum_{\substack{k_1, \ldots, k_m \geq 0 \\ k_1 + \cdots + k_m = n}} \binom{n}{k_1 \cdots k_m} a_1^{k_1} \cdots a_m^{k_m}.$$

How many terms occur on the right-hand side of the multinomial formula?

After a moment of reflexion, we see that this is the number of finite multisets of size n whose elements are drawn from a set of m elements, which is also equal to the number of m-tuples, $k_1, \ldots, k_m$, with $k_i \in \mathbb{N}$ and

$$k_1 + \cdots + k_m = n.$$

Proposition 4.2.9 *The number of finite multisets of size* $n \geq 0$ *whose elements come from a set of size* $m \geq 1$ *is*

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

4.3   Some Properties of the Binomial Coefficients

The binomial coefficients satisfy many remarkable identities.

If one looks at the Pascal triangle, it is easy to figure out what are the sums of the elements in any given row

It is also easy to figure out what are the sums of $n-m+1$ consecutive elements in any given column (starting from the top and with $0 \leq m \leq n$).

What about the sums of elements on the diagonals? Again, it is easy to determine what these sums are.

Here are the answers, beginning with sums of the elements in a column.

(a) Sum of the first $n - m + 1$ elements in column m $(0 \leq m \leq n)$.

For example, if we consider the sum of the first 5 (non-zero) elements in column $m = 3$ (so, $n = 7$), we find that

$$1 + 4 + 10 + 20 + 35 = 70,$$

where 70 is the entry on the next row and the next column.

n	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\cdots$
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	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Thus, we conjecture that

$$\binom{m}{m} + \binom{m+1}{m} + \cdots + \binom{n-1}{m} + \binom{n}{m} = \binom{n+1}{m+1},$$

which is easily proved by induction.

The above formula can be written concisely as

$$\sum_{k=m}^{n} \binom{k}{m} = \binom{n+1}{m+1},$$

or even as

$$\sum_{k=0}^{n} \binom{k}{m} = \binom{n+1}{m+1},$$

since $\binom{k}{m} = 0$ when $k < m$.

It is often called the *upper summation formula* since it involves a sum over an index, k, appearing in the upper position of the binomial coefficient, $\binom{k}{m}$.

(b) Sum of the elements in row n.

For example, if we consider the sum of the elements in row $n = 6$, we find that

$$1 + 6 + 15 + 20 + 15 + 6 + 1 = 64 = 2^6.$$

n	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\cdots$
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	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Thus, we conjecture that

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

This is easily proved by induction of by setting $a = b = 1$ in the binomial formula for $(a + b)^n$.

Unlike the columns for which there is a formula for the partial sums, there is no closed form formula for the partial sums of the rows.

However, there is a closed form formula for partial alternating sums of rows. Indeed, it is easily shown by induction that

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

if $0 \le m \le n$. For example

$$1 - 7 + 21 - 35 = -20.$$

Also, for $m = n$, we get

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

(c) Sum of the first $n + 1$ elements on the descending diagonal starting from row m.

For example, if we consider the sum of the first 5 elements starting from row $m = 3$ (so, $n = 4$), we find that

$$1 + 4 + 10 + 20 + 35 = 70,$$

the elements on the next row below the last element, 35.

n	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\cdots$
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	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Thus, we conjecture that

$$\binom{m}{0} + \binom{m+1}{1} + \cdots + \binom{m+n}{n} = \binom{m+n+1}{n},$$

which is easily shown by induction.

The above formula can be written concisely as

$$\sum_{k=0}^{n}\binom{m+k}{k}=\binom{m+n+1}{n},$$

It is often called the *parallel summation formula* since it involves a sum over an index, k, appearing both in the upper and in the lower position of the binomial coefficient, $\binom{m+k}{k}$.

(d) Sum of the elements on the ascending diagonal starting from row n.

n	F_{n+1}	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\cdots$
0	1	1									
1	1	1	1								
2	2	1	2	1							
3	3	1	3	3	1						
4	5	1	4	6	4	1					
5	8	1	5	10	10	5	1				
6	13	1	6	15	20	15	6	1			
7	21	1	7	21	35	35	21	7	1		
8	34	1	8	28	56	70	56	28	8	1	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

For example, the sum of the numbers on the diagonal starting on row 6 (in cyan), row 7 (in blue) and row 8 (in red) are:

$$1 + 6 + 5 + 1 = 13$$
$$4 + 10 + 6 + 1 = 21$$
$$1 + 10 + 15 + 7 + 1 = 34.$$

We recognize the *Fibonacci numbers*, F_7, F_8 and F_9, what a nice surprise!

Recall that $F_0 = 0$, $F_1 = 1$ and

$$F_{n+2} = F_{n+1} + F_n.$$

Thus, we conjecture that

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

The above formula can indeed be proved by induction, but we have to distinguish the two case where n is even or odd.

We now list a few more formulae which are often used in the manipulations of binomial coefficients.

They are among the "top ten binomial coefficient identities" listed in Graham, Knuth and Patashnik [9], see Chapter 5.

(e) The equation

$$\binom{n}{i}\binom{n-i}{k-i} = \binom{k}{i}\binom{n}{k},$$

holds for all n, i, k, with $0 \le i \le k \le n$.

This is because, we find that after a few calculations,

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

Observe that the expression in the middle is really the trinomial coefficient

$$\binom{n}{i \ \ k-i \ \ n-k}.$$

For this reason, the equation (e) is often called *trinomial revision*.

For $i = 1$, we get

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

So, if $k \neq 0$, we get the equation

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}, \qquad k \neq 0.$$

This equation is often called the *absorption identity*.

(f) The equation

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

holds for $m, n, p \geq 0$ such that $m + p \geq n$.

This equation is usually known as *Vandermonde convolution*.

An interesting special case of Vandermonde convolution arises when $m = p = n$. In this case, we get the equation

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

However, $\binom{n}{k} = \binom{n}{n-k}$, so we get

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

that is, the sum of the squares of the entries on row n of the Pascal triangle is the middle element on row $2n$.

A summary of the top nine binomial coefficient identities is given in Figure 4.4.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \qquad 0 \le k \le n \qquad \textit{factorial expansion}$$

$$\binom{n}{k} = \binom{n}{n-k}, \qquad 0 \le k \le n \qquad \textit{symmetry}$$

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}, \qquad k \ne 0 \qquad \textit{absorption}$$

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}, \qquad 0 \le k \le n \qquad \textit{addition/induction}$$

$$\binom{n}{i}\binom{n-i}{k-i} = \binom{k}{i}\binom{n}{k}, \qquad 0 \le i \le k \le n \qquad \textit{trinomial revision}$$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k, \qquad n \ge 0 \qquad \textit{binomial formula}$$

$$\sum_{k=0}^{n} \binom{m+k}{k} = \binom{m+n+1}{n}, \qquad m,n \ge 0 \qquad \textit{parallel summation}$$

$$\sum_{k=0}^{n} \binom{k}{m} = \binom{n+1}{m+1}, \qquad 0 \le m \le n \qquad \textit{upper summation}$$

$$\binom{m+p}{n} = \sum_{k=0}^{m} \binom{m}{k}\binom{p}{n-k} \qquad \begin{array}{l} m+p \ge n \\ m,n,p \ge 0 \end{array} \qquad \textit{Vandermonde convolution}$$

Figure 4.4: Summary of Binomial Coefficient Identities

Remark: Going back to the generalized binomial coefficients, $\binom{r}{k}$, where r is a real number, possibly negative, the following formula is easily shown:

$$\binom{r}{k} = (-1)^k \binom{k-r-1}{k},$$

where $r \in \mathbb{R}$ and $k \in \mathbb{Z}$.

If $r < 0$ and $k \geq 1$ then $k - r - 1 > 0$, so the formula shows how a binomial coefficient with negative upper index can be expessed as a binomial coefficient with positive index.

For this reason, this formula is known as *negating the upper index*.

Next, we would like to better understand the growth pattern of the binomial coefficients.

Looking at the Pascal triangle, it is clear that when $n = 2m$ is even, the central element, $\binom{2m}{m}$, is the largest element on row $2m$ and when $n = 2m + 1$ is odd, the two central elements, $\binom{2m+1}{m} = \binom{2m+1}{m+1}$, are the largest elements on row $2m + 1$.

Furthermore, $\binom{n}{k}$ is strictly increasing until it reaches its maximal value and then it is strictly decreasing (with two equal maximum values when n is odd).

The above facts are easy to prove by considering the ratio

$$\binom{n}{k} \bigg/ \binom{n}{k+1} = \frac{k+1}{n-k},$$

where $0 \leq k \leq n - 1$.

It would be nice to have an estimate of how large is the maximum value of the largest binomial coefficient, $\binom{n}{\lfloor n/2 \rfloor}$.

Since the sum of the elements on row n is 2^n and since there are $n + 1$ elements on row n, some rough bounds are

$$\frac{2^n}{n + 1} \leq \binom{n}{\lfloor n/2 \rfloor} < 2^n$$

for all $n \geq 1$.

Thus, we see that the middle element on row n grows very fast (exponentially).

We can get a sharper estimate using Stirling's formula (see Section 4.1). We give such an estimate when $n = 2m$ is even, the case where n is odd being similar.

We have

$$\binom{2m}{m} \sim \frac{2^{2m}}{\sqrt{\pi m}}.$$

The next question is to figure out how quickly $\binom{n}{k}$ drops from its maximum value, $\binom{n}{\lfloor n/2 \rfloor}$.

Let us consider the case where $n = 2m$ is even, the case when n is odd being similar and left as an exercise.

We would like to estimate the ratio

$$\binom{2m}{m-t} \bigg/ \binom{2m}{m},$$

where $0 \leq t \leq m$.

Actually, it will be more convenient to deal with the inverse ratio,

$$r(t) = \binom{2m}{m} \bigg/ \binom{2m}{m-t} = \frac{(m-t)!(m+t)!}{(m!)^2}.$$

Observe that

$$r(t) = \frac{(m+t)(m+t-1)\cdots(m+1)}{m(m-1)\cdots(m-t+1)}.$$

The above expression is not easy to handle but if we take its (natural) logarithm, we can use basic inequalities about logarithms to get some bounds.

We will make use of the following proposition:

Proposition 4.3.1 *We have the inequalities*

$$1 - \frac{1}{x} \leq \ln x \leq x - 1,$$

for all $x \in \mathbb{R}$ with $x > 0$.

We are now ready to prove the following inequalities:

Proposition 4.3.2 *For every $m \geq 0$ and every t, with $0 \leq t \leq m$, we have the inequalities*

$$e^{-t^2/(m-t+1)} \leq \binom{2m}{m-t} \bigg/ \binom{2m}{m} \leq e^{-t^2/(m+t)}.$$

This implies that

$$\binom{2m}{m-t} \bigg/ \binom{2m}{m} \sim e^{-\frac{t^2}{m}},$$

for m large and $0 \leq t \leq m$.

What is remarkable about Proposition 4.3.2 is that it shows that $\binom{2m}{m-t}$ varies according to the *Gaussian curve* (also known as *bell curve*), $t \mapsto e^{-\frac{t^2}{m}}$, which is the probability density function of the *normal distribution* (or *Gaussian distribution*).

If we make the change of variable, $k = m - t$, we see that if $0 \leq k \leq 2m$, then

$$\binom{2m}{k} \sim e^{-\frac{(m-k)^2}{m}} \binom{2m}{m}.$$

If we plot this curve, we observe that it reaches its maximum for $k = m$ and that it decays very quickly as k varies away from m.

It is an interesting exercise to plot a bar chart of the binomial coefficients and the above curve together, say for $m = 50$. One will find that the bell curve is an excellent fit.

Given some number, $c > 1$, it sometimes desirable to find for which values of t does the inequality

$$\binom{2m}{m} \Big/ \binom{2m}{m - t} > c$$

hold. This question can be answered using Proposition 4.3.2.

Proposition 4.3.3 *For every constant, $c > 1$, and every natural number, $m \geq 0$, if $\sqrt{m \ln c} + \ln c \leq t \leq m$, then*

$$\binom{2m}{m} \bigg/ \binom{2m}{m - t} > c$$

and if $0 \leq t \leq \sqrt{m \ln c} - \ln c \leq m$, then

$$\binom{2m}{m} \bigg/ \binom{2m}{m - t} \leq c.$$

As an example, if $m = 1000$ and $c = 100$, we will have

$$\binom{1000}{500} \bigg/ \binom{1000}{500 - (500 - k)} > 100$$

or equivalently

$$\binom{1000}{k} \bigg/ \binom{1000}{500} < \frac{1}{100}$$

when $500 - k \geq \sqrt{500 \ln 100} + \ln 100$, that is, when

$$k \leq 447.4.$$

It is also possible to give an upper on the partial sum

$$\binom{2m}{0} + \binom{2m}{1} + \cdots + \binom{2m}{k-1},$$

with $0 \leq k \leq m$, in terms of the ratio, $c = \binom{2m}{k} / \binom{2m}{m}$.

The following proposition is taken from Lovász, Pelikán and Vesztergombi [12] (Lemma 3.8.2, Chapter 3):

Proposition 4.3.4 *For any natural numbers m and k with $0 \leq k \leq m$, if we let $c = \binom{2m}{k} / \binom{2m}{m}$, then we have*

$$\binom{2m}{0} + \binom{2m}{1} + \cdots + \binom{2m}{k-1} < c \, 2^{2m-1}.$$

This proposition implies an important result in (discrete) probability theory as explained in [12] (see Chapter 5).

Observe that 2^{2m} is the sum of all the entries on row $2m$.

As an application, if $k \leq 447$, the sum of the first 447 numbers on row 1000 of the Pascal triangle makes up less than 0.5% of the total sum and similarly for the last 447 entries.

Thus, the middle 107 entries account for 99% of the total sum.

4.4 The Inclusion-Exclusion Principle, Sylvester's Formula, The Sieve Formula

We close this chapter with the proof of a poweful formula for determining the cardinality of the union of a finite number of (finite) sets in terms of the cardinalities of the various intersections of these sets.

This identity variously attributed Nicholas Bernoulli, de Moivre, Sylvester and Poincaré has many applications to counting problems and to probability theory.

Figure 4.5: Abraham de Moivre, 1667-1754 (left) and Henri Poincaré, 1854-1912 (right)

We begin with the "baby case" of two finite sets.

Proposition 4.4.1 *Given any two finite sets, A, and B, we have*

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

We would like to generalize the formula of Proposition 4.4.1 to any finite collection of finite sets, $A_1, \ldots, A_n$.

A moment of reflexion shows that when $n = 3$, we have

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

One of the obstacles in generalizing the above formula to n sets is purely notational: *We need a way of denoting arbitrary intersections of sets belonging to a family of sets indexed by $\{1, \ldots, n\}$.*

We can do this by using indices ranging over subsets of $\{1, \ldots, n\}$, as opposed to indices ranging over integers.

So, for example, for any nonempty subset, $I \subseteq \{1, \ldots, n\}$, the expression $\bigcap_{i \in I} A_i$ denotes the intersection of all the subsets whose index, i, belongs to I.

Theorem 4.4.2 *(Inclusion-Exclusion Principle) For any finite sequence, $A_1, \ldots, A_n$, of $n \geq 2$ subsets of a finite set, X, we have*

$$\left| \bigcup_{k=1}^{n} A_k \right| = \sum_{\substack{I \subseteq \{1, \ldots, n\} \\ I \neq \emptyset}} (-1)^{(|I|-1)} \left| \bigcap_{i \in I} A_i \right|.$$

As an application of the Inclusion-Exclusion Principle, let us prove the formula for counting the number of surjections from $\{1, \ldots, n\}$ to $\{1, \ldots, p\}$, with $p \leq n$, given in Proposition 4.2.5.

Recall that the total number of functions from $\{1, \ldots, n\}$ to $\{1, \ldots, p\}$ is p^n.

The trick is to count the number of functions that are *not* surjective.

Any such function has the property that its image misses one element from $\{1, \ldots, p\}$.

So, if we let

$$A_i = \{f \colon \{1, \ldots, n\} \to \{1, \ldots, p\} \mid i \notin \operatorname{Im}(f)\},$$

we need to count $|A_1 \cup \cdots \cup A_p|$.

But, we can easily do this using the Inclusion-Exclusion Principle.

We find that

$$\left| \bigcap_{i \in I} A_i \right| = (p - k)^n.$$

From this, the Inclusion-Exclusion Principle yields

$$|A_1 \cup \cdots \cup A_p| = \sum_{k=1}^{p-1} (-1)^{k-1} \binom{p}{k} (p-k)^n,$$

and so, the number of surjections, $S_{n\,p}$, is

$$
\begin{aligned}
S_{n\,p} &= p^n - |A_1 \cup \cdots \cup A_p| \\
&= p^n - \sum_{k=1}^{p-1} (-1)^{k-1} \binom{p}{k} (p-k)^n \\
&= \sum_{k=0}^{p-1} (-1)^k \binom{p}{k} (p-k)^n \\
&= p^n - \binom{p}{1}(p-1)^n + \binom{p}{2}(p-2)^n + \cdots \\
&\qquad + (-1)^{p-1} \binom{p}{p-1},
\end{aligned}
$$

which is indeed the formula of Proposition 4.2.5.

Another amusing application of the Inclusion-Exclusion Principle is the formula giving the number, p_n, of permutations of $\{1, \ldots, n\}$ that leave no element fixed (i.e., $f(i) \neq i$, for all $i \in \{1, \ldots, n\}$). Such permutations are often called *derangements*.

We get

$$
\begin{aligned}
p_n &= n! \left(1 - \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{(-1)^k}{k!} + \cdots + \frac{(-1)^n}{n!} \right) \\
&= n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! + \cdots + (-1)^n.
\end{aligned}
$$

Remark: We know (using the series expansion for e^x in which we set $x = -1$) that

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

Consequently, the factor of $n!$ in the above formula for p_n is the sum of the first $n + 1$ terms of $\frac{1}{e}$ and so,

$$
\lim_{n \to \infty} \frac{p_n}{n!} = \frac{1}{e}.
$$

It turns out that the series for $\frac{1}{e}$ converges very rapidly, so $p_n \approx \frac{1}{e}n!$.

The ratio $p_n/n!$ has an interesting interpretation in terms of *probabilities*.

Assume n persons go to a restaurant (or to the theatre, etc.) and that they all check their coats. Unfortunately, the cleck loses all the coat tags.

Then, $p_n/n!$ is the probability that nobody will get her or his own coat back!

As we just explained, this probability is roughly $\frac{1}{e} \approx \frac{1}{3}$, a *surprisingly large number*.

The Inclusion-Exclusion Principle can be easily generalized in a useful way as follows:

Given a finite set, X, let m be any given function, $m \colon X \to \mathbb{R}_+$, and for any nonempty subset, $A \subseteq X$, set

$$m(A) = \sum_{a \in A} m(a),$$

with the convention that $m(\emptyset) = 0$ (Recall that $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$).

For any $x \in X$, the number $m(x)$ is called the *weight* (or *measure*) of x and the quantity $m(A)$ is often called the *measure of the set A*.

For example, if $m(x) = 1$ for all $x \in A$, then $m(A) = |A|$, the cardinality of A, which is the special case that we have been considering.

For any two subsets, $A, B \subseteq X$, it is obvious that

$$
\begin{aligned}
m(A \cup B) &= m(A) + m(B) - m(A \cap B) \\
m(X - A) &= m(X) - m(A) \\
m(\overline{A \cup B}) &= m(\overline{A} \cap \overline{B}) \\
m(\overline{A \cap B}) &= m(\overline{A} \cup \overline{B}),
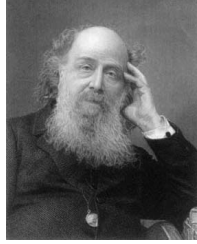\end{aligned}
$$

where $\overline{A} = X - A$.

Figure 4.6: James Joseph Sylvester, 1814-1897

Then, we have the following version of Theorem 4.4.2:

Theorem 4.4.3 *(Inclusion-Exclusion Principle, Version 2) Given any measure function, $m\colon X \to \mathbb{R}_+$, for any finite sequence, $A_1, \ldots, A_n$, of $n \geq 2$ subsets of a finite set, X, we have*

$$m\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ I \neq \emptyset}} (-1)^{(|I|-1)}\, m\left(\bigcap_{i \in I} A_i\right).$$

A useful corollary of Theorem 4.4.3 often known as Sylvester's formula is:

Theorem 4.4.4 *(Sylvester's Formula) Given any measure, $m \colon X \to \mathbb{R}_+$, for any finite sequence, $A_1, \ldots, A_n$, of $n \geq 2$ subsets of a finite set, X, the measure of the set of elements of X that do not belong to any of the sets A_i is given by*

$$
m \left(\bigcap_{k=1}^{n} \overline{A}_k \right) = m(X) + \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ I \neq \emptyset}} (-1)^{|I|}\, m \left(\bigcap_{i \in I} A_i \right).
$$

Note that if we use the convention that when the index set, I, is empty then

$$
\bigcap_{i \in \emptyset} A_i = X,
$$

then the term $m(X)$ can be included in the above sum by removing the condition that $I \neq \emptyset$ and this version of Sylvester's formula is written:

$$
m \left(\bigcap_{k=1}^{n} \overline{A}_k \right) = \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|}\, m \left(\bigcap_{i \in I} A_i \right).
$$

Sometimes, it is also convenient to regroup terms involving subsets, I, having the same cardinality and another way to state Sylvester's formula is as follows:

$$m\left(\bigcap_{k=1}^{n} \overline{A}_k\right) = \sum_{k=0}^{n} (-1)^k \sum_{\substack{I \subseteq \{1,\dots,n\} \\ |I|=k}} m\left(\bigcap_{i\in I} A_i\right).$$

$$\text{(Sylvester's Formula)}$$

Finally, Sylvester's formula can be generalized to a formula usually known as the "Sieve Formula":

Theorem 4.4.5 *(Sieve Formula) Given any measure, $m\colon X \to \mathbb{R}_+$, for any finite sequence, $A_1, \ldots, A_n$, of $n \geq 2$ subsets of a finite set, X, the measure of the set of elements of X that belong to exactly p of the sets A_i $(0 \leq p \leq n)$ is given by*

$$T_n^p = \sum_{k=p}^{n} (-1)^{k-p} \binom{k}{p} \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} m\left(\bigcap_{i \in I} A_i\right).$$

Observe that Sylvester's Formula is the special case of the Sieve Formula for which $p = 0$.

The Inclusion-Exclusion Principle (and its relatives) plays an important role in combinatorics and probability theory as the reader will verify by consulting any text on combinatorics.

A classical reference on combinatorics is Berge [1]; a more recent is Cameron [3].

More advanced references are van Lint and Wilson [19], and Stanley [17].

Another great (but deceptively tough) reference covering discrete mathematics and including a lot of combinatorics is Graham, Knuth and Patashnik [9].

Conway and Guy [4] is another beautiful book that presents many fascinating and intriguing geometric and combinatorial properties of numbers in a very untertaining manner.

For readers interested in geometry with a combinatriol flavor, Matousek [13] is a delightful (but more advanced) reference.

We are now ready to study special kinds of relations: Partial orders and equivalence relations.