STIRLING NUMBERS

by

Gian-Carlo Rota

Stephen W. Smoliar

Technical Report No. 27

May, 1973
This monograph is based on a set of notes taken from lectures on combinatorics given by Professor Gian-Carlo Rota at the Massachusetts Institute of Technology in 1969. It is indeed a pleasure to finally gather this material under one cover - not so much for the originality of the results as for the originality of the methodology. Professor Rota's constructivist approach enabled me to formulate a general modus operandi which may, indeed, serve combinatorics in a far broader sense than is offered here. It is all the more fitting that this material be issued through a Computer Science Department, as our methods of abstract reasoning are much more akin to those of the computer programmer than to the traditional "Satz-Beweis" style of classical mathematics. Such methods may, hopefully one day extend beyond combinatorics to other areas of mathematics.

Stephen W. Smoliar

Haifa, Israel
Stirling Numbers

1. Preliminaries and Notations

The Stirling numbers first appeared in James Stirling’s Methodus Differentialis, Sive Tractatus de Summationae et Interpolatione Serierum, one of the earliest writings on the calculus of finite differences. Stirling observed that if the numbers \( \text{C}_n^r \) were defined to be the coefficients of the expression:

\[
x(x+1)(x+2) \ldots (x+n-1) = \text{C}_n^0 x^n + \text{C}_n^1 x^{n-1} + \ldots + \text{C}_n^{n-1} x,
\]

then

\[
\frac{1}{n!} = \sum_{r=0}^{\infty} \frac{\text{C}_n^{r-1}}{z(z+1)\ldots(z+r)}
\]

and if \( \Gamma_n^s \) were defined to be the coefficients of the infinite series:

\[
\frac{1}{x(x+1)\ldots(x+n-1)} = \sum_{s=0}^{\infty} \frac{(-1)^s \Gamma_n^s}{x+n+s},
\]

then

\[
z^n = \Gamma_n^{n-1} z + \Gamma_n^{n-2} z(z-1) + \ldots + \Gamma_n^0 z(z-1) \ldots (z-n+1).
\]

Professor Niels Nielsen proposed that the numbers \( \text{C}_n^r \) be called Stirling Numbers of the First Species and the numbers \( \Gamma_n^s \) be called Stirling Numbers of the Second Species.

The current definition of the Stirling numbers was formulated by Charles Jordan. We use the following notations for factorial, lower factorial, and upper factorial, respectively:

\[
x! = x(x-1)(x-2) \ldots 3 \cdot 2 \cdot 1
\]

\[
(x)_n = x(x-1)(x-2) \ldots (x-n+1)
\]

\[
x^{(n)} = x(x+1)(x+2) \ldots (x+n-1).
\]
Then the **Stirling numbers of the first kind** are the coefficients of the expression:

\[ (x)_n = \sum_{k=0}^{n} s(n,k)x^k \tag{8} \]

and the **Stirling numbers of the second kind** are the coefficients of the expression:

\[ x^n = \sum_{k=0}^{n} S(n,k)(x)_k \tag{9} \]

The notations \( s(n,k) \) and \( S(n,k) \) are those of John Riordan.

Observe that the numbers \( c^r_n \) are related to the \( s(n,k) \) by

\[ c^r_n = (-1)^{n-k} s(n,k) \tag{10} \]

We shall use Riordan's notation and denote these numbers \( c(n,r) \) and call them the **signed Stirling numbers**.

We propose to develop a theory of Stirling numbers which is independent of the calculus of finite differences. We shall formulate definitions of the Stirling numbers which are based purely on set-theoretic, combinatorial notions. We shall then show - again through combinatorial techniques - that our definitions coincide with the classical ones. Furthermore, we shall investigate analogies between the behavior of the Stirling numbers and that most fundamental function of combinatorics, the binomial coefficient.

Most of our results will arise from the study of functions between two finite sets. In the sequel we shall denote the domain of such functions by the letter \( S \) and the range by the letter \( X \). The **size** of a set \( S \), namely the number of elements of \( S \), will be written \( \nu(s) \). We shall use the letter \( n \) for the size of the set \( S \) and the letter \( x \) for the size of the set \( X \). Let \( X^S \) denote the family of all functions from \( S \) to \( X \); we shall not recall the precise definition of such a function \( f:S \to X \), nor will we prove the elementary fact that the total number of functions whose domain is \( S \) and whose range is \( X \) is the integer \( x^n \).
After the notion of function, the next most fundamental notion is that of a partition. A partition of a set $S$ is a family of subsets \( \{B_1, B_2, \ldots, B_k\} \) called blocks whose union is $S$ (i.e., $B_1 \cup B_2 \cup \ldots \cup B_k = S$), such that any two distinct blocks are disjoint. We shall denote partitions by letters of the Greek alphabet such as $\pi, \sigma, \tau, \ldots$.

A relation $R$ on a set $S$ is a subset of $S \times S$; given two elements of $S$, $x$ and $y$, if $x$ is related to $y$ by $R$, i.e. if $(x, y) \in R$, we write $xRy$. $R$ is an equivalence relation if it satisfies the following three properties:

(a) $xRx, \forall x \in S$
(b) $xRy \Rightarrow yRx, \forall x, y \in S$
(c) $xRy \land yRz \Rightarrow xRz, \forall x, y, z \in S$.

We shall generally use the symbol $\sim$ to denote equivalence relations. Given $x$, an element of $S$, we define the equivalence class of $x$ with respect to $\sim$, denoted $[x]_\sim$, to be the following subset of $S$:

$$[x]_\sim = \{y : y \sim x\}.$$

The family of all equivalence classes of elements of $S$ is a partition, which we shall call the partition induced by $\sim$. Similarly, if $\pi$ is a partition of $S$, define a relation $R$ as follows: $xRy$ whenever $x$ and $y$ lie in the same block of $\pi$. It is easily verified that $R$ is an equivalence relation. Therefore, the notion of partition and equivalence relation are mathematically equivalent, although psychologically different.

Suppose $f$ is a function from $S$ to $X$; we define an equivalence relation on the domain $S$. Say $a \sim b$ whenever $f(a) = f(b)$; this is clearly an equivalence relation. The partition of $S$ which it induces is called the kernel of $f$.

We define the number of subsets of size $k$ out of a set of size $n$ to be the binomial coefficient:

$$\binom{n}{k},$$

read "binomial $n,k$". This is not the conventional definition of the binomial
coefficient, but as we develop its properties we shall see that this definition is equivalent to other formulations.

Similarly, the numbers of partitions with k blocks out of a set of size n is written

\[ S(n,k) \]

We shall use this as our definition of Stirling numbers of the second kind.

2. Sets and Partitions

Our major concern is with the enumeration of certain classes of functions with interesting properties. We shall see that the same general enumeration techniques will be employed in most of the problems we attack. To begin with, we have the elementary sum and product law.

**Sum and Product Law:** Let \( E_1, E_2, \ldots, E_m \) be \( m \) independent events; that is, the outcome of any one of these events is independent of the outcome of any other. Suppose \( E_1 \) has \( n_1 \) possible outcomes, \( E_2 \) has \( n_2 \) possible outcomes, etc. Then there are \( \prod_{i=1}^{m} n_i \) possible outcomes if \( E_1 \) or \( E_2 \) or \( \ldots \) or \( E_m \) occurs; if all \( m \) events occur together, there are \( \prod_{i=1}^{m} n_i \) possible outcomes.

Observe that this law is very closely related to the sum and product laws of probability theory. The most elementary application of this law is the computation of the number of permutations of a set \( S \) of \( n \) elements. We imagine an ordered array of \( n \) boxes, and the event \( E_i \) consists of placing one element in the \( i \)th box. Performing the events sequentially, \( E_1, E_2, \ldots \) etc., we see that there are \( n \) ways to fill the first box. \( E_2 \) then has \( n-1 \) possible outcomes regardless of the outcome of \( E_1 \); likewise, \( E_3 \) has \( n-2 \) possible outcomes, etc. Hence, we see that the number of permutations of \( S \) is

\[ \prod_{i=1}^{n} i = n! \]

In general, we shall apply the following modus operandi of enumeration:

(a) We specify the collection \( F \) of elements to be enumerated.
(b) Using an equivalence relation, we partition \( F \) into blocks \( F_1, F_2, \ldots, F_n \), so that
\[
\nu(F) = \sum_{i=1}^{m} \nu(F_i).
\]

(c) We characterize the enumeration of the \( F_i \) by a sequence of independent events \( E_1, E_2, \ldots, E_m \), so that
\[
\nu(F_i) = \prod_{j=1}^{m} \nu(E_j).
\]

(d) Then
\[
\nu(F) = \prod_{i=1}^{n} \prod_{j=1}^{m} \nu(E_j).
\]

We shall see that there is a close analogy between identities satisfied by binomial coefficients and identities satisfied by Stirling numbers of the second kind. Although this analogy has never been made precise, it often provides a powerful heuristic method for guessing new identities.

Many of the identities proved below also have algebraic proofs; it is, however, instructive to derive them by purely set-theoretic arguments. The main reason for this preference is that a set-theoretic proof leads to fruitful generalizations and to better understanding. It is easy to pass from a set-theoretic proof to an algebraic one, but much harder to unravel the set-theoretic background that may lie hidden in an algebraic identity.

We have already observed that the total number of functions from a set \( S \) to a set \( X \) is \( x^n \). What is the number of one-to-one functions, or monomorphisms from \( S \) to \( X \)? The answer is very simple: it is the lower factorial power:
\[
(x)_n = x(x-1)(x-2) \ldots (x-n+1).
\]

The proof is also simple. Order the set \( S \) linearly, so that its elements may be named \( 1, 2, \ldots, n \). Every one-to-one function from \( S \) to \( X \) can be uniquely generated by the following process: Map the element 1 of \( S \) to any of the \( x \) elements of \( X \). This can be done in \( x \) ways. Next, map the element 2 of \( S \) to any one of the elements of \( X \) which is not equal to the image of 1. This can be done in \( x-1 \) ways. Proceeding in this way, we find that the element \( k+1 \) can be mapped to \( X \) in \( x-k \) different ways. Therefore, the total number of monomorphisms from \( S \) to \( X \) is obtained by multiplying these numbers and this gives \( (x)_n \). q.e.d.
We may express the number of monomorphisms in another manner by the following expressions:

\[
\binom{x}{n} n! \]

This is verified by the fact that since every one-to-one function must have a range of size \(n\), we may enumerate them by considering all possible ranges \(\binom{x}{n}\) and for each range counting the number of bijections (i.e. permutations) onto it: \((n!)\). Hence, we have the following equality.

\[ (1) \quad (x)_n = \binom{x}{n} n! \]

On the other hand, we see from the definition of lower factorial power that

\[ (2) \quad (x)_n = \frac{x!}{(x-n)!} \]

We thus obtain the classical formula for the binomial coefficient,

\[ (3) \quad \binom{x}{n} = \frac{x!}{n!(x-n)!} \]

After counting one-to-one functions it is natural to ask about onto functions, or epimorphisms. The enumeration is very easily obtained in terms of Stirling numbers of the second kind, and has a striking resemblance to expression (1). The number of epimorphisms whose domain is \(S\) and whose range is \(X\) is

\[ xl S(n,x) \]

Note that this expression is meaningless unless \(n\) exceeds \(x\), just as it is meaningless to consider monomorphisms unless \(x\) exceeds \(n\).

The proof is exceedingly simple. There are \(S(n,x)\) possible partitions of \(S\) into \(X\) blocks, i.e. \(S(n,x)\) possible kernels for an epimorphism. For each kernel, there are \(xl\) possible assignments of blocks to members of \(X\). Therefore, these are \(xl S(n,x)\) onto functions. Naturally enough, this proof closely resembles the verification of (1).
We next give a set-theoretic proof of the well-known classical binomial theorem:

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

Partition the set \( X \) into two blocks, \( A \) and \( B \), of sizes \( a \) and \( b \), respectively; hence, there are \((a+b)^n\) functions from the set \( S \) to the set \( X \). We now enumerate these functions by our modus operandi. (1) \( F \) is the set of all functions from \( S \) to \( X \). (2) Partition \( F \) by the equivalence relation \( \sim \), defined so that \( f \sim g \) whenever \( \nu(f^{-1}(A)) = \nu(g^{-1}(A)) \). The equivalence classes \( F_i \) are then those functions which map exactly \( i \) elements into \( A \). (3) On the other hand, \( F_i \) is enumerated in terms of three independent events. First, select a subset of \( S \) of size \( i \); this can be done in \( \binom{n}{i} \) ways. Next, map that subset into \( A \); this can be done in \( a^i \) ways. Finally, map the remaining elements of \( S \) into \( B \); this can be done in \( b^{n-i} \) ways. Hence, \( \nu(F_i) = \binom{n}{i} a^i b^{n-i} \). Therefore, the total number of functions from \( S \) to \( X \) is \( \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i} \), which gives the desired result.

Suppose we set \( b=1 \) in formula (3). Then we obtain

\[ (a+1)^n = \sum_{k=0}^{n} \binom{n}{k} a^k. \]

This provides us with a definition of the binomial coefficients which is analogous to Jordan's definitions of the Stirling numbers. Let us now prove that our definition of \( S(n,k) \) is equivalent to that of Jordan, namely, that

\[ \sum_{n=0}^{n} S(n,k)(x)_k = x^n. \]

The proof of this identity gives us another excellent opportunity to demonstrate our modus operandi. \( x^n \) is simply the number of functions of \( S \) to \( X \). (1) We wish to enumerate \( F \), the family of all functions from \( S \) to \( X \). (2) Let \( f \) and \( g \) be two elements of \( F \); say \( f \sim g \) whenever \( f \) and \( g \) both have the same number of blocks in their kernels. \( \sim \) is clearly an equivalence relation, so it induces a partition of \( F \) into blocks \( F_0, F_1, F_2, \ldots, F_n \), where
F_i is the set of those functions from S to X with i blocks in their kernels.

(3) The size of F_i is determined by first enumerating the ways of partitioning S into i blocks (S(n,i)), and then by enumerating the number of one-to-one functions from each of these families of i blocks into X((x)_i). Therefore, v(F_i) = S(n,i)(x)_i. (4) Then v(F) = \sum_{i=0}^{n} S(n,i)(x)_i; note that S(n,0) = 0.

We next establish the Vandermonde formula:

\[(a+b)^n = \sum_k (a^k)(b^{n-k}),\]

where the sum on the right ranges over those integers k for which the binomial coefficients are non-zero. Note that we consider \(\binom{n}{k}\) to be equal to zero unless \(k \leq n\).

The proof of (6) is again extremely simple: the left side counts the subsets of size n out of a set S of a+b elements. Partition S into two blocks A and B of sizes a and b; let C be a subset of size n. Then if the size of \(C \cap A\) is k, the size of \(C \cap B\) is n-k. Conversely, every subset C of size n is obtained uniquely by picking first a subset of size k out of A, then a subset of size n-k out of B, and then taking their union. For each k, this can be done in

\(\binom{a}{k}\binom{b}{n-k}\)

ways; summing over all k's we complete the proof.

Let us now derive a Stirling-number analog of the Vandermonde formula. It turns out there is a variety of them, and we shall begin with the simplest:

\[(7) S(n+1,k) = \sum_i (i^n)S(n-i,k-1).\]

Proof: (1) By the expression on the left side, we see that we must enumerate the family F of all partitions with k blocks of a set T of size n+1. (2) Select one element from T, call it a, and let S = T-{a}, a set of size n. We define an equivalence relation on the family F of all partitions of T into k blocks; if \(\pi\) and \(\sigma\) are elements of F, we say \(\pi \sim \sigma\) whenever the size of the block of a
which contains a is the same as the size of the block of \( c \) which contains a. This equivalence relation induces a partition of \( F \) into blocks \( \{F_1, F_2, \ldots, F_{n+1}\} \) such that every partition in \( F_i \) places a in a block of size \( i \). (3) Next we compute \( \nu(F_i) \). To obtain a partition such that a belongs to a block of size \( i+1 \), it suffices to choose a subset of \( S \) of size \( i \), add the element a to it, and then partition the remaining \( n-i \) elements of \( S \) into \( k-1 \) blocks. Since there are \( \binom{n}{i} \) ways to choose a subset of \( S \) of size \( i \), and \( S(n-1,k-1) \) ways to properly partition the remainder of \( S \), we see that \( F_{i+1} \) has size

\[
\nu(F_{i+1}) = \binom{n}{i} S(n-i,k-1).
\]

Taking the sum

\[
S(n+1,k) = \sum_{i=1}^{n+1} \nu(F_i).
\]

Now gives us the derived formula (7), q.e.d. Note that the Stirling number \( S(n,k) \) is zero unless \( 1 \leq k \leq n \).

The idea of the preceding proof can be extended by taking, instead of a single element a, two elements a and b. Let \( T \) be a set of size \( n+2 \) containing two distinguished elements, a and b; and let \( S = T - \{a,b\} \). Proceeding as before, we obtain the formula for the Stirling number \( S(n+2,k) \) as follows:

\[
S(n+2,k) = \sum_{i=1}^{n+1} \binom{n}{i} 2! S(i,2) S(n-i,k-2) + \sum_{j=1}^{k} \binom{n}{j} S(n-j,k-1)
\]

Proof(1) We are to count the partitions of a set \( T \) of size \( n+2 \) into \( k \) blocks. Partition the set \( T \) into a set \( S \) of size \( n \) and two additional sets \( \{a\} \) and \( \{b\} \). Let \( F \) be the family of all partitions of \( T \) into \( k \) blocks. (2) Partition the family \( F \) into two classes of blocks, \( F(1,1) \) and \( F(2,j) \). The partition \( \pi \) belongs to the block \( F(1,1) \) whenever the union of the blocks containing the elements a and b is a subset of \( T \) of size \( i+2 \). On the other hand, \( F(2,j) \) consists of those partitions which place a and b in a single block of size \( j+2 \). (3) The size of \( F(1,1) \) is determined as follows: A partition belonging to \( F(1,1) \) is obtained by first choosing a subset of \( S \) of size \( i \), then partitioning this subset \( l \) into two blocks, then assigning each of these blocks to a or to b, and finally partitioning the complement of \( l \) into \( k-2 \) blocks. The first operation
can be performed in \( \binom{n}{1} \) ways, the second in \( S(1,2) \) ways, the third in 2! ways and the last in \( S(2n-1;k) \) ways. Furthermore, these operations can be performed independently of each other.

Similarly, a partition \( \sigma \) will belong to \( F(2,j) \) if and only if both elements \( a \) and \( b \) belong to one and the same block of \( \sigma \). Every such partition can be obtained as follows: first choose a subset \( J \) of \( S \) of size \( j \); second, add to it the elements \( a \) and \( b \); Thirdly, partition the complement of the subset \( J \) in \( S \) into \( k-1 \) blocks. It is clear that these operations are independent of each other. The first can be performed in \( \binom{n}{j} \) ways and the second in \( S(n-j,k-1) \) ways, so that \( \nu(F(2,j)) = \binom{n}{j}S(n-j,k-1) \). (4) Summing over the blocks of \( F(1,1) \) and \( F(2,j) \), we obtain formula (8). Of course, the indices \( i \) and \( j \) only take those values for which each term in the sum is non-zero. The proof is therefore complete.

Pursuing this theme, one can obtain more and more complex identities for the Stirling numbers. The simplest generalization consists of partitioning the set \( T \) into a subset \( S \), together with several blocks each containing one element, and proceeding in much the same way as we have proved (8). There are of course innumerable variations, leading to formulas of great complexity and occasionally of beauty. Of these formulas we shall only prove two here, because of limitations of space, and because the method should already be fairly clear to the reader. The first one is the recursion formula: For the binomial coefficients this formula reads as follows:

\[
\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k},
\]

and can be proved set-theoretically as follows: Let \( T \) be a set of size \( n+1 \). Partition \( T \) into a block \( S \) of size \( n \), and the single-element block \( a \). The family \( F \) of subsets of size \( k \) can be partitioned into two blocks as follows: the first block contains those subsets which contain the element \( a \), the second contains all those that do not contain the element \( a \). To every subset in the first block there corresponds uniquely a subset of \( S \) of size \( k-1 \); similarly, every subset in the second block corresponds to a subset of \( S \) of
size $k$. This accounts for the two terms on the right side and completes the proof.

Now to the analogous recursion formula for Stirling numbers, which is

\begin{equation}
S(n+1,k) = S(n,k-1) + kS(n,k).
\end{equation}

The proof is entirely analogous to the proof of (9). In the same notation, the family $F$ of all partitions into $k$ blocks of the set $T$ is partitioned into two blocks $F_1$ and $F_2$ as follows: $F_1$ contains all those partitions for which the element $a$ is a block of size 1, and $F_2$ contains all those partitions for which $a$ is not in a block of size 1. The block $F_1$ accounts for the first term of the right side of (10), and $F_2$ accounts for the second term. The factor $k$ appears because the element $a$ can become attached to any one of the $k$ blocks partitioning the set $S$. q.e.d.

From the recursion formula (10) for the Stirling numbers one can obtain a Pascal triangle:

\begin{align*}
1 \\
1 & 1 \\
1 & 3 & 1 \\
1 & 7 & 6 & 1 \\
1 & 15 & 25 & 10 & 1 \\
1 & 31 & 90 & 65 & 15 & 1
\end{align*}

whose definition and manner of construction are left for the reader.

We have already mentioned the fact that to "every" algebraic identity involving binomial coefficients one can consider a set-theoretic identity, which in turn can be used to prove the algebraic version. As a last instance of this phenomenon we want to establish the identity

\begin{equation}
\sum_{k=1}^{n} \binom{n}{k} k^{x} k^{-1} = n(x+1)^{n-1},
\end{equation}

as well as its Stirling-analog. The proof goes as follows. We enumerate
functions from a set S of size n to a set T, which is in turn partitioned into a block X and a one-element block a. (1) The right side accounts for the set F of all functions whose domain is some subset of S of size n-1 and whose range is T. There are \( n \) ways of choosing a subset of S of size n-1, and for any such choice there are \((x+1)^{n-1}\) functions from S to T. This accounts for the right side. (2) As for the left side, we partition the set F of all such functions into blocks \( \{F_1, F_2, \ldots, F_n\} \). The block \( F_k \) contains all those functions in F which map n-k elements into the element a and k-1 elements into X. (3) The size of \( F_k \) is calculated as follows. A function in \( F_k \) is determined by first choosing k elements out of S (this can be done in \( \binom{n}{k} \) ways), then discarding an element (this can be done in k ways), then mapping the remaining subset of size k-1 into the set X arbitrarily. The last operation can be done in \( x^{k-1} \) ways. The operations are independent, and therefore the total number of functions in \( F_k \) is the product of the number of ways of performing each operation. (4) This accounts for the \( k^{th} \) term of the left side, completing the proof.

Now to the Stirling-analog, which is:

\[
\sum_{k=1}^{n} S(n,k)k(x)^{k-1} = (x+1)^n - x^n.
\]

In the same notation as the preceding proof, this formula is easily obtained as follows. (1) The right side enumerates those functions from the set S to the set X which contain element a in their range; indeed, the second term of the right side subtracts the size of the set of those functions which do not contain the element a in the range. (2) The family F of such functions is partitioned into blocks \( \{F_1, F_2, \ldots, F_n\} \), where \( F_k \) consists of all functions whose range contains the element a and whose kernel contains exactly k blocks. (3) The size of \( F_k \) is found as follows: first partition the set S into k blocks in an arbitrary way - this can be done in \( S(n,k) \) ways -; second, pick one of these blocks at random and map every element in this block to the element a - this can be done in k ways -; lastly, assign to each one of the remaining blocks one of the elements of X - this can be done in \( (x)^k \) ways.
(4) This accounts for the $k^{th}$ term on the left side - since the operations are independent - and completes the proof.

Before leaving this subject, we warn the reader that the analogies between binomial coefficients and Stirling numbers are barely scratched by the examples and ideas we have hinted at in this section. We shall see how the analogy extends to the signless Stirling numbers and the Stirling numbers of the first kind, but first we need to introduce a set-theoretic interpretation to the upper factorial power.

3. Placings

Unfortunately, the idea of function is often too weak a tool for many combinatorial enumerations; and we shall have to strengthen it in order to solve several important problems. A major strengthening is that of action by a permutation group. We consider not just the family of functions from some domain $S$ to a range $X$, but also a permutation group of bijections from the domain onto itself. Hence, we have broadened our operation from a single function to a function pair. This brings us to the notion of placing. Definition. A placing of a set $S$ into a set $X$ is a function $f:S \to X$, together with a linear ordering for each of the blocks of the kernel of $f$.

The typical example of placing is the problem of assigning a set of $n$ flags to a set of $x$ masts, since the flags placed on each of the masts are linearly ordered by height.

Suppose $S$ is the set of integers $N = \{1, 2, \ldots, n\}$; then a placing of $N$ into $X$ may be regarded as a function $f$ together with a permutation of the elements in each block of the kernel of $f$. Hence, a placing may be expressed by the notion of a permutation acting on the domain of a function; since this notion will be particularly vital in the sequel, we review a few basic facts about permutations.

A permutation is simply a bijection from a finite set onto itself; in general, our finite sets will be sets of integers. A cyclic permutation of $n$
elements, sometimes denoted \((a_1 a_2 \ldots a_n)\), maps \(a_1\) into \(a_2\), \(a_2\) into \(a_3\), ..., and \(a_n\) into \(a_1\). A basic theorem about permutations, which can be read in any algebra text, states that any permutation can be written as a product of disjoint cyclic permutations. The cyclic permutations of such a decomposition are called the cycles of the permutation. A placing thus corresponds to a permutation of the domain whose cycles correspond to the blocks of the kernel.

Our first task is to enumerate the placings from \(S\) to \(X\). Recall our definition of upper factorial power:

\[
x(n) = x(x+1)(x+2) \ldots (x+n-1).
\]

It turns out that \(x(n)\) equals the number of placings from a set \(S\) of size \(n\) to a set \(X\) of size \(x\). This can be seen by either of the following two methods:

**Proof(1):** By identifying \(S\) with \(N\), we may assume that \(S\) has a linear order. We want to place \(n\) elements into \(x\) boxes such that within each box, the elements are linearly ordered. The first element has \(x\) possible places; we simply have to choose a box. However, the second element has \(x+1\) possible places; we may place it in one of the \(x-1\) other boxes, or we may place it in the same box as the first and above the first, or we may place it in the same box as the first and below the first. Similarly, the third element has \(x+2\) possible places, etc.

**Proof(2):** As above, we associate with \(S\) the set \(N\) of natural numbers. Add \(x-1\) elements, called "stoppers", to the set \(N\), and call the resulting set \(N'\). Then \(\nu(N') = n+x-1\). Now, to any permutation of the set \(N'\) there corresponds a unique placing, as follows: the \(x-1\) stoppers partition the elements of \(N\) into \(x\) blocks (some of which may be empty - for example, if two stoppers lie between the same pair of integers). Furthermore, the linear ordering on \(N'\) induces a linear ordering on each of these blocks, thus producing what we shall call a linear partition of \(N\) into at most \(x\) blocks. Finally, we consider \(X\) as linearly ordered, and map the subset of \(N\) between the \(k^\text{th}\) and the
(k+1)st stoppers to the (k+1)st element of X. In this way, a linear ordering of N' uniquely defines a placing. On the other hand, a placing determines exactly (x-1)! linear orderings on N', since the stoppers may be arbitrarily permuted without altering the placing (i.e. they are "unlabeled"). The number of linear orderings on N' is (x+n-1)!; hence, there are \( \frac{(x+n-1)!}{(x-1)!} \) placings from N to X, q.e.d.

The next natural question us to ask the number of onto placings from S to X. This number is:

\[
n! \binom{n-1}{x-1}
\]

This can also be proved using the method of stoppers. This time, assume that the x elements of the range have a linear order. Now permute the n elements of the domain, and line them up. There are n-1 spaces between the n elements; select x-1 of them, and insert stoppers there, thus partitioning the n elements into x blocks. Since the blocks have a linear order as induced by the permutation, this represents a unique onto placing - mapping the first block to the first element of the range, the second block to the second element, and so on. Since there are n! permutations of the domain and \( \binom{n-1}{x-1} \) ways to insert the stoppers, we obtain the desired formula, q.e.d.

A placing is distinguished from a function by the linear ordering placed on the blocks of the kernel. Hence, the kernel of a placing is not just a partition of the domain, but a partition together with a linear ordering of the blocks. As we have seen such a partition is called a linear partition, and the kernel of a placing is called a linear kernel.

The number of linear partitions of a set S of size n into x blocks is given by the formula:

\[
\ell_{n,x} = \frac{n!}{x!} \binom{n-1}{x-1}
\]

The numbers \( \ell_{n,x} \) are called the Lah numbers.
The proof is as follows: There are \( n! \binom{n-1}{x-1} \) onto placings from \( S \) to \( X \). Since the \( x \) blocks can be permuted in \( x! \) ways, each giving a different placing, there are \( \frac{n!}{x!} \binom{n-1}{x-1} \) linear partitions of \( S \) having \( x \) blocks. q.e.d.

The relation between Lah numbers and placings is analogous to the relation between Stirling numbers and partitions. For example, consider the following formula which strongly resembles formula (5) of the preceding section:

\[
x^{(n)} = \sum_{k=1}^{n} \frac{n!}{k!} \binom{n-1}{k-1} (x)_k = \sum_{k=1}^{n} \ell_{n,k}(x)_n.
\]

For a proof, we appeal, once again, to our reliable modus operandi.

1. We wish to enumerate the family \( F \) of all placings of \( N \) to \( X \), which we know to be of size \( x^{(n)} \).
2. Partition \( F \) by the equivalence relation under which \( f \sim g \) whenever the linear kernels of \( f \) and \( g \) have the same number of blocks. \( F_i \), then, is the set of all placings whose linear kernel has \( i \) blocks.
3. The number of functions in \( F_i \) is now easily computed. First, enumerate the number of linear partitions of \( N \) having \( i \) blocks \( (\ell_{n,i}) \), and then, for each linear partition, enumerate the number of one-to-one functions from the \( i \) blocks to \( X \) \((x)_i\). Hence, \( \nu(F_i) = \ell_{n,i}(x)_i \).
4. Summing over all values of \( i \), we obtain the desired formula.

We are now ready for Stirling's expansion of the upper factorial power in terms of coefficients of \( x^k \) (formula (1) of Section 1). To this end, we introduce the signless Stirling numbers, denoted \( c(n,k) \), which are defined to be the number of permutations of a set of size \( n \) having \( k \) cycles. Recall that every permutation of a set can be uniquely decomposed into disjoint cycles, corresponding to the transitivity classes of the set \( S \) under the action of the permutation and its powers. The underlying partition into disjoint cycles will be called the cycle decomposition of the permutation.

The identity we wish to verify is:

\[
x^{(n)} = \sum_{k=1}^{n} c(n,k) x^k.
\]
Proof: (1) The left side is simply the number of placings from the set $N$ to the set $X$, as usual. (2) We proceed to partition this set $F$ of placings into $n$ blocks $F_1, F_2, \ldots, F_n$. To every placing we can assign a unique permutation of the set $N$, which is composed of the various permutations of each of the blocks of the linear kernel of the placing. In other words, every linear partition of $N$ induces a permutation of $N$. $F_k$ then consists of those placings whose induced permutations have exactly $k$ cycles. It is important to note that this integer $k$ is not necessarily equal to, but in general greater than, the number of blocks in the linear kernel of the placing.

(3) Next we show that $V(F_k) = c(n, k) \cdot x^k$; this entails showing that to any given permutation $p$ whose cycle decomposition has $k$ blocks, there corresponds exactly $x^k$ placings in the block $F_k$. Once $p$ is given, any function mapping the cycles of $p$ to the set $X$ will give a placing. Indeed, given such a function, the inverse image of any element of a $x$ $X$ is a set of disjoint cycles which together determine a unique linear order on the union of the underlying blocks. Hence, it determines one block in the linear kernel of the underlying placing. Since there are exactly $x^k$ such functions, and since there are exactly $c(n, k)$ permutations with $k$ cycles, we have verified the size of $F_k$. (4) Adding over all values of $k$ from 1 to $n$, we obtain the desired formula.

If now we place $(x)_n$ on the left-hand side of formula (3), we obtain a sign change, and the formula becomes:

$$ (x)_n = \sum_{k=1}^{n} (-1)^{n-k} c(n, k) \cdot x^k $$

The coefficients $(-1)^{n-k} c(n, k)$ are defined to be the Stirling numbers of the first kind and are denoted $s(n, k)$. Formula (4) then reduces to Jordan's formulation given in Section 1 (formula (8)).

While there is no set-theoretic interpretation of Stirling numbers of the first kind which is as elegant as those of the second kind and signless Stirling numbers, we may relate the Stirling numbers of both the first and
second kind to the problem of colouring a graph. This involves the process of chromatic reduction which was introduced by Ronald Read.

By colouring a graph we mean assigning a colour to each vertex such that no two vertices connected by an edge have the same colour. In general, the process of colouring a graph with \( n \) vertices using \( x \) colours may be regarded as a function from \( S \), the set of vertices, to \( X \), the set of colours.

Chromatic polynomials provide an expression for the number of ways a graph may be coloured in \( x \) colours; the quantity is given as a polynomial expanded either in terms of \( x^i \) or in terms of \( (x)_i \). As motivation, let us consider the two simplest colouring problems.

Let \( S \) be the empty graph with \( n \) vertices; that is, no two vertices are connected by an edge. Then any function from \( S \) to \( X \) constitutes a colouring since any vertex may be assigned any colour. Hence, there are \( x^n \) possible colourings of \( S \).

Now suppose \( S \) is the complete graph with \( n \) vertices; that is, every pair of vertices is connected by an edge. What are the functions which constitute colourings? Clearly, they are the monomorphisms, since no two vertices may be assigned the same colour. Hence, there are \( (x)_n \) ways of colouring \( S \).

The process of chromatic reduction enables us to consider the problem of colouring an arbitrary graph in terms of colouring either a collection of empty graphs or a collection of full graphs. The reduction process relies on a single theorem.

**Theorem:** Let \( S \) be a graph with two non-adjacent vertices \( A \) and \( B \). Let \( S' \) be the graph obtained from \( S \) by connecting \( A \) and \( B \) with an edge. Let \( S'' \) be the graph obtained from \( S \) by identifying \( A \) and \( B \) as a single vertex and eliminating any multiple edges. If \( M_S(x) \) denotes the number of ways of colouring \( S \) in \( x \) colours, then we have the identity:

\[
M_S(x) = M_{S'}(x) + M_{S''}(x)
\]
Proof: All colourings of $S$ may be partitioned into two classes: (1) those colourings in which $A$ and $B$ are assigned different colours, and (2) those colourings in which $A$ and $B$ are assigned the same colour. The number of colourings of the first type is equal to the number of colourings of $S'$, and the number of colourings of the second type is equal to the number of colourings of $S''$. q.e.d.

Given a particular graph, this identity may be represented symbolically as follows:

Each of the graphs on the right can be likewise expanded, and we may repeat the process until the resulting graphs are full (i.e. there no longer exist two non-adjacent vertices). Let us carry out this procedure in detail starting with the empty graph of four vertices:

$$
\begin{align*}
A & \quad \quad \quad B \\
0 & \quad \quad \quad 0 \\
\end{align*}
$$

$$
= A + B \\
\begin{align*}
A & \quad 0 \\
B & \quad 0 \\
\end{align*}
$$

$$
= A'' + B'' \\
\begin{align*}
A & \quad 0 \\
B & \quad 0 \\
\end{align*}
$$

$$
= A''' + B''' \\
\begin{align*}
A & \quad 0 \\
B & \quad 0 \\
\end{align*}
$$

$$
= A'''' + B'''' \\
\begin{align*}
A & \quad 0 \\
B & \quad 0 \\
\end{align*}
$$

...
Since we know the number of colourings for full and empty graphs, we may express this as an identity in $x$, the number of colours:

$$x^4 = (x)_4 + 6(x)_3 + 7(x)_2 + (x)_1.$$

This is just a special case of Jordan's definition of Stirling numbers of the second kind.

To repeat the procedure for Stirling numbers of the first kind, we rewrite identity (5) as follows:

$$M_{S_1}(x) = M_S(x) - M_{S_1}(x).$$

Now we take two adjacent vertices $A$ and $B$ and form two new graphs, one in which the connecting edge is deleted and one in which $A$ and $B$ are identified as a single vertex. Symbolically:

$$\begin{array}{c}
\begin{array}{c}
\text{full graph}
\end{array}
\end{array} = \begin{array}{c}
\begin{array}{c}
\text{graph with edge deleted}
\end{array}
\end{array} - \begin{array}{c}
\begin{array}{c}
\text{graph with $A$ and $B$ identified}
\end{array}
\end{array}$$

This time we expand the full graph:

$$\begin{array}{c}
\begin{array}{c}
\text{full graph}
\end{array}
\end{array} = \begin{array}{c}
\begin{array}{c}
\text{graph with edge deleted}
\end{array}
\end{array} - \begin{array}{c}
\begin{array}{c}
\text{graph with $B$ and $B'$ identified}
\end{array}
\end{array}$$
Again, we write this as a polynomial in the number of colours $x$:

\[(x)_4 = x^4 - 6x^3 + 11x^2 - 6x.\]

This is a special case of Jordan's definition of Stirling numbers of the first kind. We now have a somewhat intuitive, "effective" procedure for computing Stirling numbers, although it is rather clumsy.

It would be far more desirable to have a recursion formula such as formulas (9) and (10) in Section 2. We have such a recursion formula, but it will be easier to present it first as an identity of signed Stirling numbers. We may then convert this into an identity for Stirling numbers of the first kind. The identity we shall prove is:

\[c(n+1, k) = c(n, k-1) + nc(n, k).\]

The quantity on the left-hand side represents the number of permutations of a set of size $n+1$ having $k$ cycles. Let $T$ be a set of size $n+1$, let $a$ be a distinguished element of $T$, and let $S = T - \{a\}$. The $k$-cycle permutations of $T$ divide up into two classes: (1) those in which $\{a\}$ is an element of a singleton cycle, and (2) those in which $\{a\}$ is an element of a cycle of size at least two. The number of permutations in the first class is simply equal to the number of permutations of $S$ having $k-1$ classes, and this is just $c(n, k-1)$. To enumerate the second class we simply enumerate all $k$-cycle partitions of $S$. The element $a$ may then be added to the cycle decomposition immediately after any of the $n$ elements of $S$. Hence, there are $nc(n, k)$ permutations in the second class. We now have a recursion formula.

Interpreting this in terms of Stirling numbers of the first kind, we obtain another recursion formula:

\[s(n+1, k) = s(n, k-1) - ns(n, k).\]

This gives us the following Pascal triangle:
Returning now to the notion of placings, we wish to proceed along the same lines: as in our study of functions. If we define a one-to-one placing to be a placing whose function is a monomorphism, we see that the one-to-one placings are simply the one-to-one functions. However, the notion of an onto placing is distinct from that of an onto function, and we have seen that there are
\[ n! \left( \frac{n-1}{k-1} \right) \]
on-onto placings from \( S \) to \( X \).

Now let us consider a special class of onto placings, namely those in which each block of the linear kernel is determined by a cyclic permutation. Let us call these into placings. Since we are considering cycle decompositions of permutations, it will be more convenient to consider into placings from \( N \) to \( X \). The number of these is simply
\[ x! \ c(n,x) \ . \]

The proof is very simple. There are \( c(n,x) \) permutations of \( N \) consisting of \( x \) cycles, and for each permutation there are \( x! \) ways to assign the cycles one-to-one to the elements of \( X \). Hence, there are \( x!c(n,x) \) into placings.

Now we give an analog of the Vandermonde formula for signless Stirling numbers. Proceeding exactly as we did for Stirling numbers of the second kind, we prove the formula
\[ (11) \quad c(n+1,k) = \sum_{i=0}^{n} \binom{n}{i} i! \ c(n-i,k-1) \ . \]
Proof: (1) We wish to enumerate the family $F$ of all permutations having $k$ cycles of a set $T$ of size $n+1$. (2) Select one element from $T$, call it $a$, and let $S = T - \{a\}$, a set of size $n$. We define an equivalence relation on $T$: if $p$ and $q$ are elements of $F$, we say $p \sim q$ whenever the size of the cycle of $p$ which contains $a$ is the same as the size of the cycle of $q$ which contains $a$. This equivalence relation induces a partition of $F$ into blocks $\{F_0, F_1, \ldots, F_n\}$ such that every permutation in $F_i$ places $a$ in a cycle of size $i+1$. (3) Next we compute $\nu(F_i)$. To obtain a permutation such that $a$ belongs to a cycle of size $i+1$, we first choose a subset of $S$ of size $i$ (which can be done in $\binom{n}{i}$ ways), then add $a$ and arrange the $i+1$ elements to define a cyclic permutation (which can be done in $(i+1)!$ ways; since each cyclic permutation has $i+1$ representations there are $\frac{(i+1)!}{i+1} = i!$ cyclic permutations), and finally permute the remaining elements with $k-1$ cycles (which can be done in $c(n-i, k-1)$ ways). Hence:

$$\nu(F_i) = \binom{n}{i} i! c(n-i, k-1).$$

(4) Taking the sum:

$$c(n+1, k) = \sum_{i=0}^{n} \nu(F_i),$$

gives us formula (11). q.e.d.

We may now simply $\binom{n}{i} i!$ to $(n)_i$ to obtain:

$$c(n+1, k) = \sum_{i=0}^{n} (n)_i c(n-i, k-1).$$  \hspace{1cm} (12)

We may, of course, transform (11) and (12) to identities of Stirling numbers of the first kind:

$$s(n+1, k) = \sum_{i=0}^{n} (-1)^i \binom{n}{i} i! s(n-i, k-1)$$  \hspace{1cm} (13)

$$s(n+1, k) = \sum_{i=0}^{n} (-1)^i (n)_i s(n-i, k-1).$$  \hspace{1cm} (14)
Note, by the way, that \((n)_0 = 1\) if \(n > 0\).

Next, we derive an analog for formulas (11) and (12) from the preceding section:

\[
(x+1)^{(n)} - x^{(n)} = \sum_{k=1}^{n} c(n,k) k(x+1)^{k-1}
\]

(1) Let \(T = X \cup \{a\}\) where \(X\) is a set of size \(x\). Then \((x+1)^{(n)} - x^{(n)}\) is equal to the number of placings from \(N\) to \(T\) which have \(a\) in their range.

(2) As in the proof of formula (3), we partition the family \(F\) of these placings into blocks \(F_1, F_2, \ldots, F_n\) where \(F_k\) consists of those placings whose induced permutations have exactly \(k\) cycles. (3) Next we enumerate \(v(F_k)\). There are \(c(n,k)\) ways to choose an underlying permutation for such a placing. Next, choose a cycle which will be mapped to \(a\); this can be done in \(k\) ways. The remaining cycles can be mapped to any element of \(T\), so this can be done in \((x+1)^{k-1}\) ways. Hence

\[
v(F_k) = c(n,k) k(x+1)^{k-1}.
\]

(4) Summing over \(k\), we obtain formula (15), q.e.d.

In terms of Stirling numbers of the first kind, this formula reads:

\[
(x+1)^{(n)} - x^{(n)} = \sum_{k=1}^{n} (-1)^{n-k} s(n,k) k(x+1)^{k-1}.
\]

We may also pass to Stirling numbers of the first kind as follows:

\[
(x+1)^{(n)} - (x)^{(n)} = \sum_{k=1}^{n} (-1)^{n-k} c(n,k) k(x+1)^{k-1};
\]

i.e.

\[
(x+1)^{(n)} - (x)^{(n)} = \sum_{k=1}^{n} s(n,k) k(x+1)^{k-1}.
\]
4. Combinations

Having strengthened the notion of function to the more powerful one of placing, we next consider a special subclass of functions, namely those for which the elements of the domain are indistinguishable. If \( f \) is a function from \( S \) to \( X \) and the elements of \( S \) are indistinguishable, then \( f \) is indistinguishable from another function \( g \) from \( S \) to \( X \) if there exists a permutation \( p \) of \( S \) such that \( f \circ p = g \). If this is the case, we write \( f \circ \equiv g \). It is easy to see that \( \equiv \) is an equivalence relation. If \( v(S) = n \) and \( v(X) = x \), we call an equivalence class with respect to \( \equiv \) a combination of size \( n \) out of a set of size \( x \).

A combination is uniquely determined, once we have assigned to each element \( a \) of the range the non-negative integer \( s(a) \) equal to the number of elements of the domain which are mapped into \( a \) by any function in the given equivalence class. In other words, we can associate to every combination from \( S \) to \( X \) a function \( s \) from \( X \) to the non-negative integers, which uniquely determines the combination. The function \( s \) is often read "the number of occurrences of" and called the indicator function, or characteristic function, of the combination. The most familiar case of such an indicator function occurs for combinations which are equivalence classes of one-to-one functions. Such a combination is simply identifiable as a subset of the set \( X \). In this case the indicator function is the classical function taking the value 1 on a subset of \( X \) and zero elsewhere. Observe that a function \( s \) from \( X \) to the non-negative integers is an indicator function of some combination of size \( n \) if and only if

\[
(1) \quad \sum_{a \in X} s(a) = n,
\]
where \( n \) is, of course, the size of \( S \). Our task will be to enumerate various kinds of combinations. The total number of combinations of \( X \) of size \( n \) is given by

\[
(\binom{n}{x} \cdot \frac{x}{n!} = \binom{x + n - 1}{x - 1} = \binom{x + n - 1}{n})
\]

We give two proofs of this important result. We begin by giving a wrong proof, which will lead us to the correct guess. The number of functions from \( S \) to \( X \) is \( x^n \); there are \( n! \) permutations of the set \( S \); hence \( \frac{x^n}{n!} \) is the total number of combinations. What is wrong with this argument? There is one thing that is particularly wrong and that is that the result is not an integer! The trouble is that by dividing by \( n! \) we have counted some functions too many times, because some functions may remain invariant under certain families of permutation.

To make this argument work we must use placings instead of functions. We note that if \( F \) and \( G \) are placings of \( S \) to \( X \), and if we define \( F \) to be equivalent to \( G \) whenever there is a permutation \( p \) of \( S \) such that \( F \circ p = G \) then the equivalence classes of placings are again combinations. (In fact we could have defined combinations in this way, using placings instead of functions). Furthermore, placings are never invariant under any permutation of their domain other than the identity permutation, so their equivalence classes under arbitrary permutations have \( n! \) members. Hence, the number of combinations equals the number of placings divided by \( n! \), that is, (2). Since \( x^{(n)} = \frac{(x+n-1)!}{(x-1)!} \), we may rewrite this expression as \( \frac{(x+n-1)!}{n!(x-1)!} = \binom{x+n-1}{x-1} \), etc.

For a second proof take \( x+n-1 \) objects and linearly order them. Select \( x-1 \) of them to be stoppers - this may be done in \( \binom{x+n-1}{x-1} \) ways; we now show how
this procedure uniquely defines the indicator function of a sampling. If \( a_1 \) is the first element of \( X \), define \( s(a_1) \) to be the number of objects above, but not including, the first stopper. Let \( a_2 \) be the number of objects between the first and second stoppers, counting neither of the stoppers, and so forth. Clearly the result is a combination of size \( n \).

Next, we consider one-to-one combinations, from \( S \) to \( X \), that is, subsets of \( X \). As we have already seen there an exactly

\[
\frac{x}{n!} = \frac{x!}{(x-n)!n!}
\]

subsets of size \( n \) out of the set \( X \). Lastly we enumerate onto combinations from \( S \) to \( X \). These are often called compositions of the set \( X \); if the set \( X \) is visualized as the set of integers \( \{1,2,\ldots,x\} \), then an onto combination of size \( n \) is simply the assignment of a summand to each of the integers \( 1,2,\ldots,x \) in such a way that the sum of all such summands is precisely \( n \).

The total number of onto combinations from \( S \) to \( X \) is given by

\[
\binom{n-1}{x-1}
\]

This follows immediately from taking the number of onto placings, as given in section 3, and dividing by \( n! \) Alternatively, one can derive this formula by the method of stoppers.
Now let us consider another enumeration problem. Suppose our set S is a cabinet with n drawers and suppose that each drawer is partitioned into an infinite number of sets. We wish to assign elements of X to these drawers in such a way that each slot contains at most one element. Furthermore, any element of X may be assigned to any number of slots.

How do we enumerate the number of possible assignments? Suppose, first, that we only have one drawer. Then the drawer may be left empty; this is one possible assignment. If only one slot is not empty, there are \(x\) choices for its contents; this accounts for \(x\) more assignments. Similarly, there are \(x^2\) assignments in which two slots are non-empty and, in general, \(x^i\) assignments in which \(i\) slots are non-empty. NOTE THAT THE SLOTS ARE INDISTINGUISHABLE. The total number of assignments is then

\[
\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}
\]

If there are \(n\) drawers, the process is repeated independently \(n\) times, and the number of assignments is

\[
\left(\frac{1}{1-x}\right)^n
\]

Now let us enumerate these assignments using combinations. We partition the set \(F\) of all such assignments into an infinite number of blocks \(F_0, F_1, F_2, \ldots\), where \(F_k\) consists of those assignments in which \(k\) slots throughout the entire cabinet are left nonempty. We enumerate \(\nu(F_k)\) as
follows: there are \( \frac{n(k)}{k!} \) ways to select \( k \) slots from the \( n \) drawers since this is just taking a combination of size \( k \) from a set of size \( n \); then, there are \( x^k \) ways to assign elements of \( x \) to these slots. Hence,

\[
v(F_k) = \frac{n(k)}{k!} \cdot x^k
\]

Taking the infinite sum over all values of \( k \), we get the identity:

\[
\left( \frac{1}{1-x} \right)^n = \sum_{k=0}^{\infty} \frac{n(k)}{k!} \cdot x^k
\]

5. **Summation Formulas**

In this section, we wish to evaluate the following four sums:

(a) \[
\sum_{k=0}^{n} \binom{n}{k}
\]

(b) \[
\sum_{k=1}^{n} c(n,k)
\]

(c) \[
\sum_{k=1}^{n} s(n,k)
\]

(d) \[
\sum_{k=1}^{n} S(n,k)
\]

Summation (a) is fundamental set theory. The sum is the number of all possible subsets of a set \( S \) of size \( n \). This is simply \( 2^n \). It may be viewed in terms of the number of functions from \( S \) to \( \{0,1\} \). A function
f represents a subset A by the rule: \( x \in A \) if and only if \( f(x) = 1 \). Hence, our first formula is

\[
\sum_{k=0}^{n} \binom{n}{k} = 2^n.
\]

In a similar manner, summation (b) gives us the total number of permutations of a set of size \( n \). Not only does this follow from the fact that \( c(n,k) \) is the number of permutations of a set of size \( n \) having \( k \) cycles, but also if we substitute \( x = 1 \) in equation (3) of Section 3, we obtain:

\[
\sum_{k=1}^{n} c(n,k) = 1^{(n)} = n!.
\]

To settle summation (c) we may make the same substitution in equation (4) of section 3. This gives us:

\[
\sum_{k=1}^{n} s(n,k) = (1)_n = \begin{cases} 1 & \text{if } n=1 \\ 0 & \text{if } n>2 \end{cases}
\]

Finally, summation (d) gives us the total number of partitions of a set of size \( n \). This number is sometimes called the Bell number, after Eric Temple Bell, and is denoted \( B_n \). While there is no simple closed form for the Bell numbers, we may readily establish the following recursion formula:

\[
B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k
\]
This formula may be proved in the tradition of the preceding proofs. \( B_{n+1} \) gives us the number of partitions of a set, which we shall call \( A \), of \( n+1 \) elements. Let us write \( A = B \cup x \) where \( B \) is a set of size \( n \). In each partition of \( A \), \( x \) is in a block with from zero to \( n \) other elements. Therefore, we divide the partitions of \( A \) into the \( n+1 \) equivalence classes \( F_0, F_1, \ldots, F_n \), where \( F_i \) consists of those partitions in which \( x \) is in a block with \( i \) other elements. To calculate \( v(F_i) \) we see that there are \( \binom{n}{i} \) ways to choose the \( i \) other elements in the block and \( B_{n-i} \) ways to partition the remaining elements.

Therefore:

\[
B_{n+1} = \sum_{i=0}^{n} \binom{n}{i} B_{n-i}
\]

However, since

\[
\binom{n}{i} = \binom{n}{n-i},
\]

we obtain

\[
B_{n+1} = \sum_{i=0}^{n} \binom{n}{n-i} B_{n-i} = \sum_{k=0}^{n} \binom{n}{k} B_k.
\]