Basic Combinatorics

Carl G. Wagner

Department of Mathematics

The University of Tennessee

Knoxville, TN 37996-1300

Contents

List of Figures

List of Tables

Chapter 1

The Fibonacci Numbers From a Combinatorial Perspective

1.1 A Simple Counting Problem

Let n be a positive integer. A *composition of n* is a way of writing n as an ordered sum of one or more positive integers (called *parts*). For example, the compositions of 3 are $1+1+1, 1+2, 2+1$, and 3. Let $f(n) :=$ the number of compositions of n in which all parts belong to the set $\{1, 2\}$

Problem. Determine $f(n)$.

What does it mean to "determine" the solution to a counting problem? In elementary combinatorics, such solutions are usually given by a *formula* (a.k.a. a *closed form solution*) involving the variable n. To find the solution for a particular value of n, one simply "plugs in" that value for n in the formula. Many of the counting problems that we encounter will have closed form solutions. But there are other ways to determine the solution to counting problems. The present problem was chosen to illustrate one such way, the *recurrence relation method*, which yields a *recursive formula*, or *recursive solution* to a counting problem. This approach is described below.

Let us determine the value of $f(n)$ for some small values of n by "brute force," i.e., by listing and counting all of the acceptable compositions. We have:

$$
\begin{aligned}
&f(1) = 1, \text{ the acceptable compositions being} &&: 1 \\
&f(2) = 2, \text{ the acceptable compositions being} &&: 1 + 1; 2 \\
&f(3) = 3, \text{ the acceptable compositions being} &&: 1 + 1 + 1, 1 + 2; 2 + 1 \\
&f(4) = 5, \text{ the acceptable compositions being} &&: 1 + 1 + 1 + 1, 1 + 1 + 2, 1 + 2 + 1; \\
& &&\quad 2 + 1 + 1, 2 + 2
\end{aligned}
$$

Further listing and counting reveals that $f(5) = 8$ and $f(6) = 13$. On the basis of this, we might conjecture, that for all $n \geq 3$,

$$(*) \qquad\qquad f(n) = f(n - 1) + f(n - 2).$$

i.e., that $f(n)$ is the nth Fibonacci number .

But how can we prove this if we don't have a formula for $f(n)$? We prove it by a *combinatorial argument*, i.e., an argument that makes use of the *combinatorial* (i.e., counting) *interpretation* of the quantities $f(n), f(n-1)$), and $f(n-2)$. Specifically, we argue that *each side of* $(*)$ *counts the same thing*, namely the collection of all compositions of n in which all parts belong to $\{1, 2\}$. That the LHS (left hand side) of $(*)$ counts this collection follows from the very definition of $f(n)$. Now the RHS of $(*)$ counts this same collection, but in two disjoint, exhaustive categories:

(i) The category of acceptable compositions with initial part equal to 1. There are $f(n-1)$ such compositions, for they all consist of 1, followed by a composition of $n-1$ in which all parts belong to $\{1, 2\}$.

(ii) The category of acceptable compositions with initial part equal to 2. There are $f(n-2)$ such compositions, for they all consist of 2, followed by a composition of $n-2$ in which all parts belong to $\{1, 2\}$.

The *recursive solution* to our problem thus consists of the *initial conditions* $f(1) = 1, f(2) = 2$, along with the *recurrence relation*

$$f(n) = f(n-1) + f(n-2), \quad \forall n \geq 3.$$

Using this recurrence relation we can calculate $f(n)$ for any value of n that we like. Of course it is not a "one-shot calculation." We wind up calculating *all* of the values $f(1), f(2), f(3), \ldots f(n)$ in order to get $f(n)$, and there is no way around this if we must rely on a recursive formula.

1.2 A Closed Form Expression for $f(n)$

Despite the simplicity of our recursive formula for $f(n)$, you may be wondering if there is a closed form expression for $f(n)$. There is indeed such an expression, and it is given by the rather formidable formula

$(**)$
$$f(n) = \frac{1}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{2} \right)^{n+1} - \left(\frac{1 - \sqrt{5}}{2} \right)^{n+1} \right).$$

It is pretty mind-boggling that the simplest formula for $f(n)$, which we know is a positive integer, involves the *irrational* real number $\sqrt{5}$. How might you go about proving $(**)$? Well, you could show that the RHS of $(**)$ takes the value 1 when $n = 1$ and 2 when $n = 2$ and satisfies the same recurrence relation as $f(n)$ [try it]. But that would be an unenlightening proof (like inductive proofs of identities that don't explain how those identities were *discovered*, but merely *verify* them in plodding, bookkeeping fashion).

In what follows, we derive $(**)$ from $(*)$ and the initial conditions $f(1) = 1$ and $f(2) = 2$. In the process we'll use a result from the theory of infinite series, which shows that although elementary combinatorics may be categorized as "discrete math," advanced combinatorics draws on continuous as well as discrete mathematics.

1.3 The Method of Generating Functions

To get a closed form expression for $f(n)$ from $(*)$ we use the so- called method of generating functions . We'll go into this in much more detail later in the course, so don't be concerned if you don't follow all the details. This chapter is designed to be flashy and entertaining, to pique your interest, not to intimidate you. What we do is to consider the "generating function" $F(x)$ defined by

$$(***) \qquad F(x) = f(0) + f(1)x + f(2)x^2 + f(3)x^3 + \cdots$$

where we define $f(0) = 1$ to fit in with the general recurrence $(*)$. How do we know that this infinite series converges for some nonzero x? Well, it's easy to prove by induction using $(*)$ that $f(n) \le 2^n$ for all $n \ge 0$ [try it] and since $\sum_{n=0}^{\infty} 2^n x^n = \sum_{n=0}^{\infty}(2x)^n$ converges absolutely for $|x| < 1/2$, and dominates $\sum_{n=0}^{\infty} f(n)x^n$, it follows that $\sum_{n=0}^{\infty} f(n)x^n$ converges absolutely on this interval.

Now we use a little trick:
Rewrite

$$(\text{i}) \qquad F(x) = f(0) + f(1)x + f(2)x^2 + \cdots + f(n)x^n + \cdots$$

Multiply by x, then x^2, getting

$$(\text{ii}) \qquad xF(x) = f(0)x + f(1)x^2 + \cdots + f(n-1)x^n + \cdots$$
$$(\text{iii}) \qquad x^2F(x) = f(0)x^2 + \cdots + f(n-2)x^n + \cdots$$

Take (i) - (ii) - (iii), getting

$$(1 - x - x^2)F(x) = f(0) + (f(1) - f(0))\,x + (f(2) - f(1) - f(0))\,x^2$$
$$(\text{iv}) \qquad\qquad + \cdots + (f(n) - f(n-1) - f(n-2))\,x^n$$
$$+ \cdots = 1,$$

since $f(0) = 1, f(1) - f(0) = 1 - 1 = 0$, and $f(n) - f(n-1) - f(n-2) = 0$ for $n \ge 2$. So

$$(\text{v}) \qquad F(x) = \frac{1}{1 - x - x^2}$$

Now we just find the Taylor series of $1/1 - x - x^2$; the coefficient of x^n in this series will be $f(n)$. Finding this Taylor series is just tedious algebra involving partial fractions. We'll look at the details later when we study generating functions more carefully. For now let's just note that one can write

$$(\text{vi}) \qquad \frac{1}{1 - x - x^2} = \frac{A}{1 - \alpha x} + \frac{B}{1 - \beta x}$$

where $\alpha = \frac{1+\sqrt{5}}{2}$ (the Greeks' "golden ratio") $\beta = \frac{1-\sqrt{5}}{2}, A = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)$, and $B = -\frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)$. Now from (vi) we get (using $\frac{1}{1-u} = \sum_{n=0}^{\infty} u^n$ for $|u| < 1$)

$$f(x) = \frac{1}{1 - x - x^2} = A\sum_{n=0}^{\infty}(\alpha x)^n + B\sum_{n=0}^{\infty}(\beta x)^n$$
$$(\text{vii})$$
$$= \sum_{n=0}^{\infty}(A\alpha^n + B\beta^n)x^n.$$

So

(vii)

$$f(n) = A\alpha^n + B\beta^n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right)$$

by plugging in the above values of α, β, A, and B.

This formula was first derived (in the first application of generating functions) by Abraham De Moivre (1667-1754) , and independently by Daniel Bernoulli (1700-1782)

1.4   Approximation of $f(n)$

Note that $\frac{1+\sqrt{5}}{2} \cong 1.618$ and $\frac{1-\sqrt{5}}{2} \cong -0.618$, so $\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} \to \infty$ as $n \to \infty$ and $\left(\frac{1-\sqrt{5}}{2}\right)^{n+1} \to 0$ as $n \to \infty$ Two easily proved consequences of this are

(viii)

$$f(n) \sim \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}$$

and

(ix)

$$\lim_{n\to\infty} \frac{f(n)}{f(n-1)} = \frac{1+\sqrt{5}}{2}.$$

In fact, if $g(n)$ is any sequence satisfying $g(0) = a$ and $g(1) = b$, where, say, a and b are positive reals, and $g(n) = g(n-1) + g(n-2)$ for all $n \geq 2$, then

(x)

$$\lim_{n\to\infty} \frac{g(n)}{g(n-1)} = \frac{1+\sqrt{5}}{2}$$

Try to prove this using (ix).

Chapter 2

Functions, Sequences, Words, and Distributions

2.1 Multisets and sets

A *multiset* is a collection of objects, taken without regard to order, and with repetitions of the same object allowed. For example, $M = \{1, 1, 1, 2, 2, 2, 2, 3\}$ is a multiset. Since the order in which objects in M are listed is immaterial, we could also write, among many other possibilities, $M = \{1, 3, 2, 2, 1, 2, 2, 1\}$. One sometimes encounters multisets written in "exponential" notation, where the "exponent" indicates the frequency of occurrence of an object in the multiset. With this notation, one would write $M = \{1^3, 2^4, 3^1\}$. The list of objects belonging to a multiset is always enclosed by a pair of curly brackets. The *cardinality* (i.e., number of elements) of a multiset takes account of repetitions. So, for example, the multiset M has cardinality 8.

A *set* is simply a multiset in which there are no repetitions of the same object. For example, $S = \{1, 2, 3\} = \{1, 3, 2\} = \{2, 1, 3\} = \{2, 3, 1\} = \{3, 1, 2\} = \{3, 2, 1\}$ is a set. As an alternative to listing the elements of a set, one can specify them by a characterizing property. So, for example, one could write $S = \{x : x^3 - 6x^2 + 11x - 6 = 0\}$.

The following sets occur with such frequency as to warrant special symbols:

(a) $\mathbb{P} = \{1, 2, 3, \ldots\}$, the set of *positive integers*

(b) $\mathbb{N} = \{0, 1, 2, \ldots\}$, the set of *nonnegative integers*

(c) $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, the set of *integers*

(d) $\mathbb{Q} = \{m/n : m, n \in \mathbb{Z} \text{ and } n \neq 0\}$, the set of *rational numbers*

(e) $\mathbb{R}$, the set of *real numbers*

(f) $\mathbb{C}$, the set of *complex numbers*

As usual, the empty set is denoted by $\emptyset$. Following notation introduced by Richard Stanley, which has been adopted by many combinatorists, we shall also write $[0] = \emptyset$ and $[n] = \{1, 2, \ldots, n\}$ for $n \in \mathbb{P}$. Note that for all $n \in \mathbb{N}$, $[n]$ contains n elements.

Remark. It is advisable to avoid using the term "natural number," since this term is used by some to denote elements of $\mathbb{P}$ and by others to denote elements of $\mathbb{N}$.

Remark. In elementary mathematics courses one writes $A \subseteq B$ when A is a subset of B and $A \subset B$ when A is a proper subset of B (i.e., when $A \subseteq B$, but $A \neq B$). Most mathematicians, however, write $A \subset B$ when A is a subset of B and $A \subsetneq B$ when A is a proper subset of B. We shall follow the latter practice.

2.2 Functions

If A and B are sets, a *function f from A to B* is a rule which assigns to each element $a \in A$ a single element $f(a) \in B$. In such a case, one writes $f : A \to B$, reading this symbolic expression as "f is a function from A to B" or "f maps A to B". The set A is called the *domain of f* and the set B is called the *codomain of f*. Functions are also called *maps, mappings, transformations,* or *functionals* in certain contexts.

The function $f : A \to B$ is *injective* (in older terminology, *one-to-one*) if $a_1 \neq a_2 \Rightarrow f(a_1) \neq f(a_2)$, i.e., if distinct elements of A are always mapped by f to distinct elements of B. Taking the contrapositive of the above definition, we get an occasionally useful equivalent formulation of injectivity of f, namely, that $f(a_1) = f(a_2) \Rightarrow a_1 = a_2$.

The function $f : \mathbb{R} \to \mathbb{R}$ defined for all $x \in \mathbb{R}$ by $f(x) = x^2$ is not injective since distinct real numbers need not have distinct squares (e.g., $f(1) = f(-1) = 1$). On the other hand, the function $g : [0, \infty) \to \mathbb{R}$ defined for all $x \in [0, \infty)$ by $g(x) = x^2$ is injective, since distinct *nonnegative* real numbers have distinct squares. This example illustrates the general principle that one can always "repair" a non-injective function by cutting down its domain to a set on which the resulting function is injective. This can in general by done in many ways. In the foregoing case, for example, we might have cut down the domain of f to $(-\infty, 0]$. In the case of a *constant function* (a function mapping every element of its domain to the same element of its codomain) one needs, of course, to cut down the domain to a single element!

If $f : A \to B$ the *range of f* (also called the *image of A under f* - an expression that is often confusingly abbreviated by careless textbook authors to "the image of f") consists of all members of B that occur as values of $f(a)$ as a runs through A, i.e., range$(f) = \{f(a) : a \in A\}$. Of course, range$(f) \subset B$. If range$(f) = B$, i.e., if the range and codomain of f coincide, then f is *surjective* (in older terminology, *onto*). Equivalently, $f : A \to B$ is surjective if, for every $b \in B$, there exists at least one $a \in A$ such that $f(a) = b$.

The function $f : \mathbb{R} \to \mathbb{R}$ defined for all $x \in \mathbb{R}$ by $f(x) = x^2$ is not surjective, since range$(f) = [0, \infty)$ is a proper subset of the codomain $\mathbb{R}$. On the other hand, the function $h : \mathbb{R} \to [0, \infty)$ defined for all $x \in \mathbb{R}$ by $h(x) = x^2$ is surjective. This example illustrates the general principle that one can always "repair" a non-surjective function by cutting down its codomain to include only elements of its range. If f is a constant function one needs, of course, to cut down the codomain to a single element.

The domain of a function is often fixed by practical or theoretical considerations. If the function in question is not injective, we may not wish to "repair" it by cutting down its domain, because we want to know the behavior of the function on the entire domain. On the other hand, there would appear to be no reason for putting up with nonsurjective functions. So why do we ever employ codomains that properly contain the ranges of the functions we are considering?

The answer is that it can be difficult to determine the range of a function given by a complicated formula. In the case of the following function, we would need to know whether Goldbach's conjecture (which asserts that every even number greater than 2 is the sum of two primes) is true in order to determine its range. Let $A = \{4, 6, 8, 10, ...\}$ and $B = \{0, 1\}$, and define $f : A \rightarrow B$ by the rule

$$f(a) = \begin{cases} 1, & \text{if } a \text{ is the sum of two primes} \\ 0, & \text{otherwise} \end{cases}.$$

No one presently knows what the range of f is.

The function $f : A \rightarrow B$ is *bijective* (in older terminology, *one-to-one onto*, or a *one-to-one correspondence*) if it is both injective and surjective. That is, f is bijective if, for each $b \in B$, there exists a *unique* $a \in A$ (denoted $f^{-1}(b)$) such that $f(a) = b$. The function $f^{-1} : B \rightarrow A$ is called the *inverse of f*, and is itself bijective. In this situation $f^{-1} \circ f$ is the identity function on A ($f^{-1}(f(a)) = a$ for all $a \in A$) and $f \circ f^{-1}$ is the identity function on B ($f(f^{-1}(b)) = b$ for all $b \in B$). For example, the function $f : [0, \infty) \rightarrow [0, \infty)$ given by $f(x) = x^2$ is bijective, with inverse $f^{-1}(y) = \sqrt{y}$, and the function $g : (-\infty, 0] \rightarrow [0, \infty)$ given by $g(x) = x^2$ is bijective, with inverse $g^{-1}(y) = -\sqrt{y}$.

2.3 Sequences and words

Let $f : A \rightarrow B$. If $A \subset \mathbb{Z}$, f is called a *sequence in B*. Usually, $A = \mathbb{P}$, $\mathbb{N}$, or $[k]$, for some $k \in \mathbb{P}$. Sequences are usually denoted using ordered lists, enclosed in parentheses, instead of using functional notation. Thus, with $t_i = f(i)$, a sequence $f : \mathbb{P} \rightarrow B$ is usually denoted $(t_1, t_2, \ldots)$, a sequence $f : \mathbb{N} \rightarrow B$ is usually denoted $(t_0, t_1, t_2, \ldots)$, and a sequence $f : [k] \rightarrow B$ is usually denoted $(t_1, t_2, \ldots, t_k)$. The latter is called a *sequence of length k in B*.

A sequence of length k in B is also called a *word of length k in the "alphabet" B*. When a sequence is construed as a word it is typically written without enclosing parentheses and with no commas. Thus, the sequence $(t_1, t_2, \ldots, t_k)$ is written $t_1 t_2 \cdots t_k$ when construed as a word. Also, whereas one refers to t_i as the ith *term* in the sequence $(t_1, t_2, \ldots, t_k)$, it is called the ith *letter* of the word $t_1 t_2 \cdots t_k$.

The function $f : [k] \rightarrow B$ associated with the sequence $(t_1, t_2, \ldots, t_k)$ by the rule $f(i) = t_i$ is injective if and only if all the terms in the sequence are distinct. Similarly, injectivity of the function associated with the word $t_1 t_2 \cdot t_k$ is equivalent to there being no repeated letters in the word. Sequences and words associated with injective functions from $[k]$ to B are also called *permutations of length k in B*, or *permutations of the members of B taken k at a time*.

The function $f : [k] \rightarrow B$ associated with the sequence $(t_1, \ldots, t_k)$ by the rule $f(i) = t_i$ is surjective if and only if each element of B occurs at least once as a term in the sequence. Similarly, surjectivity of the function associated with the word $t_1 t_2 \cdots t_k$ is equivalent to the fact that each element of B occurs at least once as a letter of the word.

2.4 Distributions

Let A be a set of apples and B be a set of bowls. Assume that the apples bear the labels $a_1, \ldots, a_k$ and the bowls the labels $b_1, \ldots, b_n$. Consider a function $f : A \rightarrow B$. Associated with

any such function is a *distribution* of the apples in A among the bowls of B according to the rule: apple a_i is placed in bowl $f(a_i)$. Along with sequences and words, distributions provide another nicely concrete interpretation of functions.

In the above context $f : A \to B$ is injective if and only if at most one apple is placed in each bowl and surjective if and only if at least one apple is placed in each bowl.

Of course one can employ instead of a set of apples any set of labeled (also called "distinct" or "distinguishable") objects and instead of a set of bowls any set of labeled "containers." The usual models employed in texts on combinatorics and probability take A to be a set of labeled balls and B to be a set of labeled urns, and are termed "urn models." We shall use the language of balls and urns whenever we construe functions as distributions.

It is perhaps worth noting in conclusion that there is a subtle conceptual difference between sequences and distributions. In the case of a sequence $f : [k] \to B$, object $f(i)$ from the codomain B is placed in position (or "slot") i of the sequence. In the case of a distribution $f : A \to B$, object a_i from the domain is placed in "container" $f(a_i)$ of the codomain.

Careful readers may wonder why we have made a point of speaking of, for example, a set of *labeled* (a.k.a., *distinct* or *distinguishable*) balls or urns. Aren't members of a set distinct by definition, making such modifiers redundant? In fact, such modifiers are redundant, but it is traditional in combinatorics to use them for extra emphasis, to distinguish this case from the case where the balls or urns are indistinguishable. In the latter case it has been traditional to refer to a "set" of indistinguishable balls or urns, even though such collections are best conceptualized as multisets. Distributions with indistinguishable balls (resp., urns) will be treated in Chapter 5 (resp., 10).

2.5 The cardinality of a set

Let A be a set and let $n \in \mathbb{P}$. If there exists a bijection $f : [n] \to A$, we say that A is finite and has *cardinality n*, symbolizing this by $|A| = n$. Less formally, if $|A| = n$, we say that A has n elements, or that A is an n-set. Note that a bijection $f : [n] \to A$ amounts simply to a way of counting the elements in A, with $f(i)$ being the element of A that we count as we say "i."

The empty set is considered to be finite, with $|\emptyset| = 0$. It is the only set of cardinality 0. (Incidentally, if we had made sense of functions with empty domain, something we shall do in a later section, we could have included this case in our original definition of cardinality, by allowing $n \in \mathbb{N}$. For it turns out that the so-called empty function is a bijection from $[0] = \emptyset$ to $\emptyset$.)

The following theorems are intuitively obvious, although their rigorous proof would require a careful study of the sets $[n]$ based on an axiomatic analysis of the nonnegative integers. (See, for example, *Modern Algebra*, vol I, Chapter III, by Seth Warner , Prentice Hall 1965. Both volumes I and II of this outstanding text have been reprinted by Dover in a single volume paperback.)

Theorem 2.1. *If A and B are finite sets, then*

1° *there exists a bijective function $f : A \to B$ if and only if $|A| = |B|$,*

2° *there exists an injective function $f : A \to B$ if and only if $|A| \leq |B|$, and*

3° *there exists a surjective function $f : A \to B$ if and only if $|A| \geq |B|$.*

Theorem 2.2. *If A and B are finite sets and $|A| = |B|$, then a function $f : A \to B$ is injective if and only if it is surjective.*

It is perhaps worth noting here that the notion of cardinality has been extended by Georg Cantor (1845-1918) to the realm of infinite sets. In particular, one writes $|A| = \aleph_0$ (aleph-null, or aleph nought) if there exists a bijection $f : \mathbb{P} \to A$, and $|A| = c$ (for "continuum") if there exists a bijection $f : \mathbb{R} \to A$. There is an infinite number of "infinite cardinals" such as $\aleph_0$ and c. In the infinite realm, Theorem 2.1 defines an ordering of the infinite cardinals. As for Theorem 2.2, it is no longer true for infinite sets. For example, $|\mathbb{P}| = |E| = \aleph_0$ where E is the set of even positive integers, since the map $f : \mathbb{P} \to E$ defined by $f(i) = 2i$ is a bijection. On the other hand, the map $g : E \to \mathbb{P}$ defined by $g(i) = i$ is injective, but not surjective.

2.6 The addition and multiplication rules

Enumerative combinatorics deals with the theory and practice of determining the cardinalities of finite sets or certain natural classes thereof. Though many counting problems appear daunting when viewed in their entirety, the following two rules frequently enable one to decompose such problems into manageable parts.

The *addition rule* simply asserts that if A and B are finite, disjoint sets ($A \cap B = \emptyset$) then $|A \cup B| = |A| + |B|$. Using this rule it is easy to prove by induction that if $(A_1, \ldots, A_k)$ is any sequence of pairwise disjoint sets ($i \neq j \Rightarrow A_i \cap A_j = \emptyset$) then $|A_1 \cup \cdots \cup A_k| = |A_1| + \cdots + |A_k|$. We shall make extensive use of these rules, which enable us to count a set by "partitioning" it into a collection of pairwise disjoint, exhaustive subsets, counting these subsets, and adding the results. The addition rule is used extensively to establish recurrence relations. Recall, for example, our proof (in Chapter 1) of the recurrence relation $f(n) = f(n-1) + f(n-2)$, where $f(n)$ is the number of compositions of n with all parts in $\{1, 2\}$.

The addition rule can also be used to prove that for any finite sets A and B

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

and this result can in turn be used to prove that

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

for any finite sets A, B, C, that

$$\begin{aligned}
|A \cup B \cup C \cup D| = {} & |A| + |B| + |C| + |D| - |A \cap B| - |A \cap C| \\
& - |A \cap D| - |B \cap C| - |B \cap D| - |C \cap D| \\
& + |A \cap B \cap C| + |A \cap B \cap D| + |A \cap C \cap D| \\
& + |B \cap C \cap D| - |A \cap B \cap C \cap D|,
\end{aligned}$$

etc., etc. These generalizations of the addition rule are special cases of what is called the *principle of inclusion and exclusion*, a powerful counting technique that will be developed in Chapter 14.

The *multiplication rule* is a rule for counting functions (and, hence, sequences, words, and distributions) subject to various restrictions on the values assigned by the functions to elements

9

in their domains. Specifically, suppose that $A = \{a_1, \ldots, a_k\}$ and $B = \{b_1, \ldots, b_n\}$. Suppose that there is a family $\mathcal{F}$ of functions from A to B and that in constructing a typical function $f \in \mathcal{F}$, there are n_1 choices for $f(a_1)$, and (whatever the value chosen for $f(a_1)$) n_2 choices for $f(a_2)$, and (whatever the values chosen for $f(a_1)$ and $f(a_2)$) n_3 choices for $f(a_3)$, and $\ldots$, and (whatever the values chosen for $f(a_1), \ldots, f(a_{k-1})$) n_k choices for $f(a_k)$. Then the family $\mathcal{F}$ contains $n_1 \times n_2 \times \cdots \times n_k$ functions.

Theorem 2.3. *Let $A = \{a_1, \ldots, a_k\}$ and $B = \{b_1, \ldots, b_n\}$, and let $\mathcal{F}$ be the family of all functions from A to B. Then $|\mathcal{F}| = n^k$. In particular, letting $A = [k]$, it follows that there are n^k sequences of length k in B or, equivalently, n^k words of length k in the "alphabet" B. Letting A be a set of k labeled balls and B a set of n labeled urns, it follows that there are n^k possible distributions of k labeled balls among n labeled urns.*

Proof. Apply the multiplication rule with $n_1 = n_2 = \cdots = n_k = n$. $\qquad\square$

Remark. If A and B are any sets, the family of all functions from A to B is often denoted B^A. This notation is a nice mnemonic for Theorem 2.3, for this theorem asserts that when A and B are finite, $|B^A| = |B|^{|A|}$.

Next, we use the multiplication rule to count the injective functions from a k-set to an n-set. The following new notation will prove to be very useful throughout the course. Let x be any real number and define

$$x^{\underline{0}} = 1 \qquad \text{(in particular, } 0^{\underline{0}} = 1\text{)}$$
$$x^{\underline{k}} = x(x-1) \cdot (x-k+1) \text{ if } k \in \mathbb{P}$$

The expression $x^{\underline{k}}$ is called the kth *falling factorial power of x*, and is often read as "x to the k falling." Note that when $x = n$, a nonnegative integer, and $n < k$, then $n^{\underline{k}} = 0$, since one of the terms in the product $n(n-1)\cdots(n-k+1)$ is equal to zero in such a case (e.g., $3^{\underline{5}} = 3 \cdot 2 \cdot 1 \cdot 0 \cdot (-1) = 0$, and $0^{\underline{2}} = 0 \cdot (-1) = 0$). Also, for all $n \in \mathbb{N}$, $n^{\underline{n}} = n!$, if we define $0! = 1$.

Theorem 2.4. *Let $A = \{a_1, \ldots, a_k\}$ and $B = \{b_1, \ldots, b_n\}$. There are $n^{\underline{k}} = n(n-1)\cdots(n-k+1)$ injective functions from A to B. In particular, letting $A = [k]$ it follows that there are $n^{\underline{k}}$ permutations of the n things in B taken k at a time. Letting A be a set of k labeled balls and B a set of n labeled urns, it follows that there are $n^{\underline{k}}$ possible distributions of k labeled balls among n labeled urns with at most one ball per urn.*

Proof. Suppose first that $n \geq k$. In constructing an injective function $f : A \to B$, there are n choices for $f(a_1)$, $n-1$ choices for $f(a_2)$, $n-2$ choices for $f(a_3)$, $\ldots$, and $n-(k-1) = n-k+1$ choices for $f(a_n)$. The theorem follows by the multiplication rule in this case.

Suppose that $n < k$. By Theorem 2.1 ($2°$) there are then *no* injective functions from A to B. But by the remarks preceding the statement of this theorem $n^{\underline{k}} = 0$ when $n < k$. So $n^{\underline{k}}$ is the correct formula for the number of injections from a k-set to an n-set for any positive integers k and n. $\qquad\square$

The following theorem combines Theorem 2.2 with the case $k = n$ of Theorem 2.4.

Theorem 2.5. *Let $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$. There are $n^{\underline{n}} = n!$ bijections from A to B. In particular, letting $A = [n]$, it follows that there are $n!$ permutations of the n things in B taken n at a time. Letting A be a set of n labeled balls and B a set of n labeled urns, it follows that there are $n!$ possible distributions of n labeled balls among n labeled urns with exactly one ball per urn.*

Proof. There are $n^{\underline{n}} = n!$ injective functions from A to B by Theorem 2.4 with $k = n$. But by Theorem 2.2, when $|A| = |B| = n$, as is the case here, the family of injective functions, the family of surjective functions, and the family of bijective functions from A to B coincide. $\square$

It would be nice at this point to enumerate the surjections from a finite set A to a finite set B. Unfortunely, we must defer this problem until Chapter 9. For now, we shall have to be content with the following partial results, based on Theorems 2.2 and 2.4:

1° If $|A| < |B|$, there are no surjections from A to B.

2° If $|A| = |B| = n$, there are $n^{\underline{n}} = n!$ surjections from A to B.

Of course, 1° and 2° above can be specialized to the case of sequences, words, or distributions.

2.7 Useful counting strategies

In counting a family of functions f with domain $A = \{a_1, \ldots, a_k\}$, it is useful to begin with k slots, (optionally) labeled $a_1, \ldots, a_k$:

$$\overline{a_1} \quad \overline{a_2} \quad \overline{a_3} \quad \cdots \quad \overline{a_k}$$

One then fills in the slot labeled a_i with the number n_i of possible values of $f(a_i)$, given the restrictions of the problem, and multiplies:

$$\frac{n_1}{a_1} \quad \times \quad \frac{n_2}{a_2} \quad \times \quad \frac{n_3}{a_3} \quad \times \quad \cdots_{\cdots} \quad \times \quad \frac{n_k}{a_k}$$

Example 1. There are 3 highways from Knoxville to Nashville, and 4 from Nashville to Memphis. How many roundtrip itineraries are there from Knoxville to Memphis via Nashville? How many itineraries are there if one never travels the same highway?

Solution. A round trip itinerary is a sequence of length 4, the ith term of which designates the highway taken on the ith leg of the trip. The solution to the first problem is thus $\underline{3} \times \underline{4} \times \underline{4} \times \underline{3} = 144$, and the solution to the second problem is $\underline{3} \times \underline{4} \times \underline{3} \times \underline{2} = 72$.

Remark. If there are 2 highways from Knoxville to Asheville and 5 from Asheville to Durham, the number of itineraries one could follow in making a round trip from Knoxville to Memphis via Nashville *or* from Knoxville to Durham via Asheville would be calculated, using both the addition and multiplication rules as

$$\underline{3} \times \underline{4} \times \underline{4} \times \underline{3} \quad + \quad \underline{2} \times \underline{5} \times \underline{5} \times \underline{2} \quad = \quad 244$$

In employing the multiplication rule, one need not fill in the slots from left to right (the labeling of elements of A as $\{a_1, \ldots, a_k\}$ in the statement of the rule was completely arbitrary). In fact *one should always fill in the number of alternatives so that the slot subject to the most restrictions is filled in first, the slot subject to the next most restrictions is filled in next*, etc.

Example 2. How many odd, 4-digit numbers are there having no repeated digits?

Solution. Fill in the slots below in the order indicated and multiply:

$$\underset{(2)}{8} \quad \times \quad \underset{(3)}{8} \quad \times \quad \underset{(4)}{7} \quad \times \quad \underset{(1)}{5}$$

(The position marked (1) can be occupied by any of the 5 odd digits; the position marked (2) can be occupied by any of the 8 digits that are different from 0 and from the digit in position (1); etc.; etc.).

Note what happens if we try to work the above problem left to right.

$$\underset{(1)}{9} \quad \times \quad \underset{(2)}{9} \quad \times \quad \underset{(3)}{8} \quad \times \quad \underset{(4)}{?}$$

We get off to a flying start, but cannot fill in the last blank, since the number of choices here depends on how many odd digits have been chosen to occupy slots (1), (2), and (3).

Sometimes, no matter how cleverly we choose the order in which slots are filled in, a counting problem simply must be decomposed into two or more subproblems.

Example 3. How many even, 4 digit numbers are there having no repeated digits?

Solution. Following the strategy of the above example, we can easily fill in slot (1) below, but the entry in slot (2) depends on whether the digit 0 or one of the digits 2, 4, 6, 8 is chosen as the last digit:

$$\underset{(2)}{?} \quad \underset{(3)}{} \quad \underset{(4)}{} \quad \underset{(1)}{5}.$$

So we decompose the problem into two subproblems. First we count the 4 digit numbers in question with the last digit equal to zero:

$$\underset{(2)}{9} \quad \times \quad \underset{(3)}{8} \quad \times \quad \underset{(4)}{7} \quad \times \quad \underset{(1)}{1} \quad = \quad 504.$$

Then we count those with last digit equal to 2, 4, 6, or 8:

$$\underset{(2)}{8} \quad \times \quad \underset{(3)}{8} \quad \times \quad \underset{(4)}{7} \quad \times \quad \underset{(1)}{4} \quad = \quad 1792.$$

The total number of even, 4-digit numbers with no repeated digits is thus $504 + 1792 = 2296$.

Problems involving the enumeration of sequences with prescribed or forbidden adjacencies may be solved by "pasting" together objects that must be adjacent, so that they form a single object.

Example 4. In how many ways may Graham, Knuth, Stanley, and Wilf line up for a photograph if Graham wishes to stand next to Knuth?

Solution. There are 3! permutations of GK, S, and W and 3! permutations of KG, S, and W. So there are $3! + 3! = 12$ ways to arrange these individuals (all of whom are famous combinatorists) under the given restriction.

Remark. To count arrangements with forbidden adjacencies, simply count the arrangements in which those adjacencies occur, and subtract from the total number of arrangements. Thus, the number of ways in which the 4 combinatorists can line up for a photograph with Graham and Knuth *not* adjacent is $4! - 12 = 12$.

2.8 The pigeonhole principle

Theorem 2.1, 2°, asserts in part that, given finite sets A and B, if there is an injection $f : A \to B$, then $|A| \leq |B|$. The contrapositive of this assertion is called the *pigeonhole principle*: *Given finite sets A and B, if $|A| > |B|$, then there is no injection $f : A \to B$.* Equivalently, if $|A| > |B|$, then for every function $f : A \to B$, there exists $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $f(a_1) = f(a_2)$. In less abstract language, if we have a set of pigeons and a set of pigeonholes and there are more pigeons than pigeonholes, then, for every way of distributing the pigeons among the pigeonholes, there will always be a pigeonhole occupied by at least two pigeons.

This simple observation can be used to prove a number of interesting "existence" theorems (theorems that assert the existence of some type of combinatorial configuration, but do not specify the exact number of such configurations).

Here is a trivial example: Among any set of 13 people there must be at least two with the same astrological sign.

Here is a less trivial example:

Theorem 2.6. *Among any set of n people ($n \geq 2$) there must be at least two having the same number of acquaintances within that set of people.*

Proof. (Acquaintanceship is taken to be an irreflexive, symmetric relation.) *Case 1.* Somebody knows everybody else. Hence, by symmetry, everybody knows somebody else. Denoting the set of n people by S, it follows that the function f defined for all $s \in S$ by $f(s) =$ the number of acquaintances of s in S takes its values in the codomain $[n-1]$. Since $|S| = n > n - 1 = |[n-1]|$, there exist, by the pigeonhole principle, s_1 and $s_2 \in S$ with $s_1 \neq s_2$ and $f(s_1) = f(s_2)$.

Case 2. Nobody knows everybody. In this case, the function f defined above takes its values in the codomain $\{0, 1, \ldots, n-2\}$, an $(n-1)$-set, and the desired result once again follows by the pigeonhole principle. □

Remark. The above theorem often appears expressed in the language of graph theory. A *graph* on a set of points is simply an irreflexive, symmetric relation on that set, usually represented geometrically by connecting two related points with an *edge*. The *degree* is the number of edges connecting that point to other points

For example, degree$(d) = 4$ in Figure 2.1. Theorem 2.6 asserts that for any graph on $n \geq 2$ points there must be at least two points having the same degree.
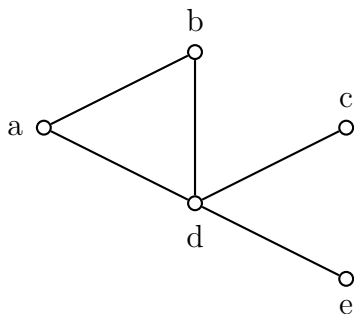
Figure 2.1: Graph (degree(d) = 4)

2.9   Functions with empty domain and/or codomain

We have shown that if $|A| = k$ and $|B| = n$, where $k, n \in \mathbb{P}$, then there are n^k functions from A to B, of which $n^{\underline{k}}$ are injective. By making sense of functions with empty domain and/or codomain, we shall show in this section that the aforementioned formulas in fact hold for all $k, n \in \mathbb{N}$, given that we define

$$0^0 = 0^{\underline{0}} = 0! = 1.$$

We have defined a function from A to B (where A and B are nonempty sets) as a rule which assigns to each element $a \in A$ a single element $f(a) \in B$. Given such a function f, the *graph of* f, graph (f), is defined as follows:

$$\mathrm{graph}(f) := \{(a, f(a)) : a \in A\}.$$

That is, the graph of f consists of all ordered pairs comprised of elements of A and their images in B under f. Note that graph$(f) \subset A \times B$, the *Cartesian product* of A and B, where

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

When $A, B \subset \mathbb{R}$, graph(f) is the graph of f in the familiar sense, i.e., a certain set of "points" in the Cartesian plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$.

When is a subset $G \subset A \times B$ the graph of a function from A to B? This will be the case if and only if *for each $a \in A$, there is exactly one $b \in B$ such that $(a, b) \in G$.*

Since a function is fully characterized by its domain, codomain, and graph, many authors identify a function with its graph, using instead of the "intensional" concept of function (function-as-rule) the following "extensional" concept of function (function-as-set-of-ordered-pairs):

($*$) A function from A to B is a subset $G \subset A \times B$ such that for all $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in G$.

Under the extensional concept of function, injectivity and surjectivity are defined as follows:

($**$) A function $G \subset A \times B$ is injective if for all $a_1, a_2 \in A$ with $a_1 \neq a_2$, $(a_1, b_1) \in G$ and $(a_2, b_2) \in G \Rightarrow b_1 \neq b_2$.

($* * *$) A function $G \subset A \times B$ is surjective if for all $b \in B$, there exists at least one $a \in A$ such that $(a, b) \in G$.

Note that there is no assumption in ($*$), ($**$), or ($* * *$) that A and B are nonempty.

Before considering the case where A and/or B is empty we need to review a concept known among mathematicians and logicians as vacuous truth. The statement

"All things with property P have property Q."

is said to be *vacuously true* when there are no things with property P. The case for regarding such a statement as true is that a universally quantified statement is true so long as one cannot find a counterexample, i.e., so long as one cannot find a thing that has property P but does not have property Q. Of course, if $P = \emptyset$, one cannot find a thing with property P at all, and hence one cannot find a thing with property P that does not have property Q. So, for example, if I have no daughters, then I am making a (vacuously) true statement if I assert that all my daughters have red hair. Similarly, it is vacuously true that all even prime numbers greater than 10 are perfect squares.

Now let us consider the case of functions with empty domain.

Theorem 2.7. *If B is any set, empty or nonempty, there is exactly one function from $\emptyset$ to B, namely the empty set $\emptyset$ (also called the "empty function"). This function is injective. It is surjective if and only if $B = \emptyset$.*

Proof. Since $\emptyset \times B = \emptyset$, $\emptyset \subset \emptyset \times B$. Moreover, it is vacuously true that for all $a \in \emptyset$, there is exactly one $b \in B$ such that $(a, b) \in \emptyset$. Similarly, ($**$) holds vacuously when $G = \emptyset$ and $A = \emptyset$, so $\emptyset$ is injective. Also ($* * *$) is (vacuously) true when $G = \emptyset$ if and only if $B = \emptyset$. So $\emptyset$ is surjective if and only if $B = \emptyset$. □

Next we consider functions with empty codomain.

Theorem 2.8. *If $A \neq \emptyset$, there are no functions (hence, no injective or surjective functions) from A to $\emptyset$. There is just one function from $\emptyset$ to $\emptyset$, namely $\emptyset$, and it is bijective.*

Proof. Since $A \times \emptyset = \emptyset$ and the only subset of $\emptyset$ is $\emptyset$, the only candidate for a function from A to $\emptyset$ is $\emptyset$. But if $A \neq \emptyset$, $\emptyset$ does not qualify as a function from A to $\emptyset$. For given $a \in A$ (and there *are* such a's if $A \neq \emptyset$) there fails to exist exactly one $b \in \emptyset$ such that $(a, b) \in \emptyset$ (because there fails to exist *any* $b \in \emptyset$). The second assertion is included in the case $b = \emptyset$ of Theorem 2.7. □

Combining Theorem 2.3, 2.4, 2.5, 2.7, and 2.8, we get the following summary theorem.

Theorem 2.9. *Define $0^0 = 0^{\underline{0}} = 0! = 1$. For all $k, n \in \mathbb{N}$, if $|A| = k$ and $|B| = n$, then there are n^k functions from A to B, of which $n^{\underline{k}}$ are injective. If $|A| = |B| = n$, there are $n!$ bijections from A to B.*

Proof. Check, using the aforementioned theorems, that these formulas work for all $n, k \in \mathbb{N}$. □

Remark. When Theorems 2.1 and 2.2 were stated it was implicit that A and B were nonempty. In fact, by Theorems 2.7 and 2.8, it is clear that Theorems 2.1 ($1°$ and $2°$, but not $3°$) and 2.2 hold for arbitrary finite sets.

Remark. It is a worthwhile exercise to cast Theorem 2.7, 2.8, and 2.9 in the language of sequences, words, and distributions. In particular, there is just one sequence (word) of length 0 in an arbitrary set B, the empty sequence (empty word). And there is just one way to distribute an empty set of balls among an arbitrary set of urns (the empty distribution).

Chapter 3

Subsets with Prescribed Cardinality

3.1 The power set of a set

The power set of a set A, denoted 2^A (we'll see why shortly) is the set of all subsets of A, i.e.,

$$2^A = \{B : B \subset A\}.$$

For example, $2^{\{a,b,c\}} = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\}$.

Theorem 3.1. *If $|A| = n$, where $n \in \mathbb{N}$, then A has 2^n subsets, i.e., $|2^A| = 2^{|A|}$.*

Proof. The assertion is true for $n = 0$, since $2^{\emptyset} = \{\emptyset\}$. Suppose $n \in \mathbb{P}$. Let $A = \{a_1, a_2, \ldots, a_n\}$. Define a function $f : 2^A \to \{0,1\}^{[n]}$ by

$$f(B) = (t_1, t_2, \ldots, t_n)$$

where $t_i = 0$ if $a_i \notin B$ and $t_i = 1$ if $a_i \in B$ (For example $f(\emptyset) = (0,0,\ldots,0)$, $f(\{a_1, a_3\}) = (1,0,1,0,\ldots,0)$, $f(A) = (1,1,\ldots,1)$, etc.]. This function f is clearly a bijection. So by Theorem 2.1 (1°), $|2^A| = |\{0,1\}^{[n]}|$. But by Theorem 2.3, $|\{0,1\}^{[n]}| = 2^n$. $\square$

The sequence $f(B) = (t_1, t_2, \ldots, t_n)$ is called the *bit string representation* (or *binary word representation*) of the subset B. Subsets are represented in a computer by means of such sequences in the set of "bits" $\{0,1\}$.

Note that just as the notation B^A furnishes a mnemonic for the combinatorial result $|B^A| = |B|^{|A|}$, the notation 2^A furnishes a mnemonic for the result $|2^A| = 2^{|A|}$.

3.2 Binomial coefficients

Let $|A| = n$, where $n \in \mathbb{N}$. For every $k \in \mathbb{N}$, let $\binom{n}{k}$ denote the number of subsets of A that have k elements, i.e.,

$$\binom{n}{k} = |\{B : B \subset A \text{ and } |B| = k\}|.$$

The symbol $\binom{n}{k}$ is read " n choose k," or "the kth binomial coefficient of order n." Certain values of $\binom{n}{k}$ are immediately obvious. For example, $\binom{n}{0} = 1$, since the only subset of an n-set

having cardinality 0 is the empty set. Also, $\binom{n}{n} = 1$, since the only subset of an n-set A having cardinality n is A itself. Moreover, it is clear that $\binom{n}{k} = 0$ if $k > n$, for the subsets of an n-set must have cardinality less than or equal to n.

Theorem 3.2. *For all $n \in \mathbb{N}$ and all $k \in \mathbb{N}$,*

$$\binom{n}{k} = \frac{n^{\underline{k}}}{k!},$$

where $n^{\underline{0}} = 1$ and $0! = 1$.

Proof. Let A be an n-set. All of the $n^{\underline{k}}$ permutations of the n things in A taken k at a time arise from (i°) choosing a subset $B \subset A$ with $|B| = k$ and (ii°) permuting the elements of B in one of the $k!$ possible ways. Hence

$$n^{\underline{k}} = \binom{n}{k} k!.$$

Solving for $\binom{n}{k}$ yields $\binom{n}{k} = \frac{n^{\underline{k}}}{k!}$. $\qquad\square$

Remark. A subset of cardinality k of the n-set A is sometimes called a *combination of the n things in A taken k at a time.* The essence of the above proof is that every combination of n things taken k at a time gives rise to $k!$ permutations of n things taken k at a time. Hence there are $k!$ times as many permutations of n things taken k at a time as there are combinations of n things taken k at a time.

Since $n^{\underline{k}} = 0$ if $k > n$, it follows that $\binom{n}{k} = \frac{n^{\underline{k}}}{k!} = 0$ if $k > n$, as we previously observed. If $0 \le k \le n$, there is an alternative formula for $\binom{n}{k}$. We have in this case

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

Let us make a partial table of the binomial coefficients, the so-called *Pascal's Triangle* (Blaise Pascal, 1623-1662).

$n\backslash k$	0	1	2	3	4	5	6
0	1	0	0	0	0	0	0
1	1	1	0	0	0	0	0
2	1	2	1	0	0	0	0
3	1	3	3	1	0	0	0
4	1	4	6	4	1	0	0
5	1	5	10	10	5	1	0
6	1	6	15	20	15	6	1

Table 3.1: Binomial Coefficients

Theorem 3.3. *For all $n, k \in \mathbb{N}$ such that $0 \le k \le n$, $\binom{n}{k} = \binom{n}{n-k}$.*

Proof. There is an easy algebraic proof of this, using the formula $\binom{n}{k} = n!/k!(n-k)!$. But we prefer a combinatorial proof. Let $|A| = n$, and let $\mathcal{E}$ be the set of all k-element subsets of A and $\mathcal{F}$ the set of all $(n-k)$-element subsets of A. By the addition rule, the function f that maps each $B \in \mathcal{E}$ to B^c, the complement of B in A, is a function from $\mathcal{E}$ to $\mathcal{F}$. It is easy to prove that f is a bijection. Hence $|\mathcal{E}| = |\mathcal{F}|$. But $|\mathcal{E}| = \binom{n}{k}$ and $|\mathcal{F}| = \binom{n}{n-k}$. $\qquad\square$

Pascal's triangle is generated by a simple recurrence relation.

Theorem 3.4. *For all $n \in \mathbb{N}$, $\binom{n}{0} = 1$ and for all $k \in \mathbb{P}$, $\binom{0}{k} = 0$. For all $n, k \in \mathbb{P}$,*

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

Proof. Let $A = \{a_1, \ldots, a_n\}$. The k-element subsets of A belong to one of two disjoint, exhaustive classes (i) the class of k-element subsets which contain a_1 as a member, and (ii) the class of k-element subsets, which do not contain a_1 as a member. There are clearly $\binom{n-1}{k-1}$ subsets in the first class (choose $k-1$ additional elements from $\{a_2, \ldots, a_n\}$ to go along with a_1) and $\binom{n-1}{k}$ subsets in the second class (since a_1 is excluded as a member, choose all k elements of the subset from $\{a_2, \ldots, a_n\}$). By the addition rule it follows that $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$. $\qquad\square$

Theorem 3.5. *For all $n \in \mathbb{N}$,*

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

Proof. You are probably familiar with the proof of Theorem 3.5 based on the binomial theorem, which we shall prove later. Actually, there is a much simpler combinatorial proof. By Theorem 3.1, 2^n counts the total number of subsets of an n-set. But so does $\sum_{k=0}^{n} \binom{n}{k}$, counting these subsets in $n+1$ disjoint exhaustive classes, the class of subsets of cardinality $k = 0, k = 1, \ldots, k = n$. By the extended addition rule we must have, therefore, $\sum_{k=0}^{n} \binom{n}{k} = $ the total number of subsets of an n-set $= 2^n$. $\qquad\square$

The following is an extremely useful identity rarely mentioned in combinatorics texts.

Theorem 3.6. *For all $n, k \in \mathbb{P}$,*

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

Proof. By Theorem 3.2,

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

19

Here is a combinatorial proof. It is equivalent to prove that for all $n, k \in \mathbb{P}$

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

But each side of the above counts the number of ways to choose from n people a committee of k people, with one committee member designated as chair. The LHS counts such committees by first choosing the k members ($\binom{n}{k}$ ways) and then one of these k as chair (k ways). The RHS counts them by first choosing the chair (n ways), then $k-1$ additional members from the $n-1$ remaining people ($\binom{n-1}{k-1}$ ways). $\qquad\square$

Here is an application of the above theorem.

Corollary (3.6.1). *For all $n \in \mathbb{N}$,*

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

Proof. One can prove this identity by induction on n, but that is not very enlightening. We give a proof that discovers the result as well as proving it. First note that the identity holds for $n = 0$. So assume $n \geq 1$. Then

$$\sum_{k=0}^{n} k \binom{n}{k} = \sum_{k=1}^{n} k \binom{n}{k} = \sum_{k=1}^{n} k \cdot \frac{n}{k} \binom{n-1}{k-1}$$

$$= n \sum_{k=1}^{n} \binom{n-1}{k-1} \underset{\text{let } j=k-1}{=} n \sum_{j=0}^{n-1} \binom{n-1}{j} = n \cdot 2^{n-1}.$$

$\qquad\square$

Remark. One can obviously iterate Theorem 3.6, proving, for example, that if $n, k \in \mathbb{P}$ with $n, k \geq 2$, $\binom{n}{k} = \frac{n(n-1)}{k(k-1)} \binom{n-2}{k-2}$. This observation is relevant to Problem 5.

Remark. There is also a combinatorial proof of Corollary 3.6.1 in the spirit of the combinatorial proof of Theorem 3.6.

Our next theorem provides a formula for the sum of a "vertical" sequence of binomial coefficients.

Theorem 3.7. *For all $n \in \mathbb{N}$ and all $k \in \mathbb{N}$,*

$$\sum_{j=0}^{n} \binom{j}{k} = \sum_{j=k}^{n} \binom{j}{k} = \binom{n+1}{k+1}.$$

#1. By Theorem 3.4, with $j + 1$ replacing n and $k + 1$ replacing k, we have $\binom{j+1}{k+1} = \binom{j}{k} + \binom{j}{k+1}$ for all $j, k \in \mathbb{N}$, i.e., $\binom{j}{k} = \binom{j+1}{k+1} - \binom{j}{k+1}$. Substituting this expression for $\binom{j}{k}$ we get

$$\sum_{j=0}^{n} \binom{j}{k} = \sum_{j=0}^{n} \left[\binom{j+1}{k+1} - \binom{j}{k+1} \right] = \left[\binom{1}{k+1} - \binom{0}{k+1} \right] + \left[\binom{2}{k+1} - \binom{1}{k+1} \right]$$

$$+ \left[\binom{3}{k+1} - \binom{2}{k+1} \right] + \cdots + \left[\binom{n+1}{k+1} - \binom{n}{k+1} \right]$$

$$= \binom{n+1}{k+1} - \binom{0}{k+1} = \binom{n+1}{k+1}, \qquad \text{by "telescoping."}$$

$\square$

#2. Clearly $\binom{n+1}{k+1}$ counts the class of all $(k + 1)$-element subsets of $[n + 1]$. But this class of subsets may be partitioned into subclasses corresponding to $j = k, k + 1, \ldots, n$ as follows. The subclass of subsets with largest element equal to $k + 1$ is counted by $\binom{k}{k}$ (why?), the subclass of subsets with largest element equal to $k + 2$ is counted by $\binom{k+1}{k}$ (why?), ..., and the subclass of subsets with largest element equal to $n + 1$ is counted by $\binom{n}{k}$ (why?). The identity in question follows by the extended addition rule. $\square$

Theorem 3.7 may seem rather esoteric, but it yields a discrete analogue of the integration formula $\int_0^b x^k dx = b^{k+1}/k + 1$ that plays a basic role in the theory of finite summation. The following corollary elaborates this point.

Corollary (3.7.1). *For all $n \in \mathbb{N}$ and for all $k \in \mathbb{N}$*

$$\sum_{j=0}^{n} j^{\underline{k}} = \sum_{j=k}^{n} j^{\underline{k}} = \frac{1}{k+1}(n+1)^{\underline{k+1}}.$$

Proof. Multiply the result of Theorem 3.7 by $k!$. $\square$

Remark. If we set $k = 1$ in the above corollary, we get

$$\sum_{j=0}^{n} j^{\underline{1}} = 1 + 2 + \cdots + n = \frac{1}{2}(n+1)^{\underline{2}} = \frac{n(n+1)}{2},$$

the familiar formula for the sum of the first n positive integers. If we set $k = 2$, we get

$$\sum_{j=0}^{n} j^{\underline{2}} = \sum_{j=0}^{n} j^2 - j = \sum_{j=0}^{n} j^2 - \sum_{j=0}^{n} j = \frac{1}{3}(n+1)^{\underline{3}}.$$

Thus

$$\sum_{j=0}^{n} j^2 = 1^2 + 2^2 + \cdots + n^2 = \frac{1}{3}(n+1)^{\underline{3}} + \frac{n(n+1)}{2} = \frac{n(n+1)(2n+1)}{6},$$

the familiar formula for the sum of the squares of the first n positive integers. One can continue in this fashion and find formulas for sums of cubes, fourth powers, etc. of the first n positive integers. See §11.2 (3°) for an elaboration of this idea.

Combining Theorem 3.7 with Theorem 3.3 yields the following corollary, which provides a formula for the sum of a "diagonal" sequence of binomial coefficients.

Corollary (3.7.2). *For all* $n \in \mathbb{N}$ *and* $r \in \mathbb{N}$

$$\sum_{i=0}^{r} \binom{n+i}{i} = \binom{n}{0} + \binom{n+1}{1} + \cdots + \binom{n+r}{r} = \binom{n+r+1}{r}.$$

Proof. By Theorem 3.3,

$$\sum_{i=0}^{r} \binom{n+i}{i} = \sum_{i=0}^{r} \binom{n+i}{n} \underset{j=n+i}{=} \sum_{j=n}^{n+r} \binom{j}{n} \underset{\text{Th. 3.7}}{=} \binom{n+r+1}{n+1} \underset{\text{Th. 3.3}}{=} \binom{n+r+1}{r}.$$

$\square$

We conclude this section with a famous binomial coefficient identity known as Vandermonde's identity (Abnit-Theophile Vandermonde, 1735-1796).

Theorem 3.8. *For all* m, n, *and* $r \in \mathbb{N}$,

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

Proof. Given a set of m men and n women, there are $\binom{m+n}{r}$ ways to select a committee with r members from this set. The RHS of the above identity counts these committees in subclasses corresponding to $j = 0, \ldots, r$, when j denotes the number of men on a committee. $\square$

Remark. A noncombinatorial proof of Theorem 3.8, based on the binomial theorem, expands $(1 + x)^m$, $(1 + x)^n$, and multiplies, comparing the coefficient of x^r in the product with the coefficient of x^r in the expansion of $(1 + x)^{m+n}$. This proof is considerably more tedious than the combinatorial proof.

Corollary (3.8.1). *For all* $n \in \mathbb{N}$

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

Proof. Let $m = r = n$ in Theorem 3.8, and use Theorem 3.3:

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

$\square$

Chapter 4

Sequences of Two Sorts of Things with Prescribed Frequency

4.1 A special sequence counting problem

In how many ways may 3 (indistinguishable) x's and 4 (indistinguishable) y's be arranged in a row? Equivalently, how many words of length 7 in the alphabet $\{x, y\}$ are there in which x appears 3 times and y appears 4 times? The multiplication rule is not helpful here. One can fill in the number of possible choices in the first 3 slots of the word

$$\underset{(1)}{2} \times \underset{(2)}{2} \times \underset{(3)}{2} \quad \underset{(4)}{?} \quad \underset{(5)}{} \quad \underset{(6)}{} \quad \underset{(7)}{},$$

but the number of choices for slot (4) depends on how many x's were chosen to occupy the first 3 slots. Going back and decomposing the problem into subcases results in virtually listing all the possible arrangements, something we want to avoid.

The solution to this problem requires a different approach. Consider the slots

$$\underset{(1)}{} \quad \underset{(2)}{} \quad \underset{(3)}{} \quad \underset{(4)}{} \quad \underset{(5)}{} \quad \underset{(6)}{} \quad \underset{(7)}{}.$$

Once we have chosen the three slots to be occupied by x's, we have completely specified a word of the required type, for we must put the y's in the remaining slots. And how many ways are there to choose 3 of the 7 slots in which to place x's? The question answers itself, "7 choose 3", i.e., $\binom{7}{3}$ ways. So there are $\binom{7}{3}$ linear arrangements of 3 x's and 4 y's. Of course, we could also write the answer as $\binom{7}{4}$ or as 7!/3!4!. The general rule is given by the following theorem.

Theorem 4.1. *The number of words of length n, consisting of n_1 letters of one sort and n_2 letters of another sort, where $n = n_1 + n_2$, is $\binom{n}{n_1} = \binom{n}{n_2} = \frac{(n_1+n_2)!}{n_1! n_2!}$.*

Proof. Obvious. □

Remark. Theorem 4.1 may be formulated as a result about distributions, as follows: The number of ways to distribute n labeled balls among 2 urns, labeled u_1 and u_2, so that n_i balls are placed in u_i, $i = 1, 2$ $(n_1 + n_2 = n)$ is $\binom{n}{n_1} = \binom{n}{n_2} = \frac{(n_1+n_2)!}{n_1! n_2!}$. This sort of distribution is called a *distribution with prescribed occupancy numbers.*

Remark. Theorem 4.1 may also be formulated as a result about functions. Given a function $f : A \to B$, and an arbitrary $b \in B$, let

$$f^{\leftarrow}(b) = \{a \in A : f(a) = b\}.$$

The set $f^{\leftarrow}(b)$ is called the *preimage* (or *inverse image*) of b under f. If $b \notin$ range(f), then $f^{\leftarrow}(b) = \emptyset$. The mapping $f^{\leftarrow}$ is a function with domain B and codomain 2^A, and is defined for any $f : A \to B$. It should not be confused with f^{-1}, which is only defined if f is bijective.

The *function-with-prescribed-preimage cardinalities* version of Theorem 4.1 is as follows: If $n = n_1 + n_2$, the number of functions $f : \{a_1, a_2, \ldots, a_n\} \to \{b_1, b_2\}$ such that $|f^{\leftarrow}(b_i)| = n_i$, $i = 1, 2$, is $\binom{n}{n_1} = \binom{n}{n_2} = \frac{(n_1+n_2)!}{n_1! n_2!}$.

4.2 The binomial theorem

If we expand $(x + y)^2 = (x + y)(x + y)$, we get, before simplification,

$$(x + y)^2 = xx + xy + yx + yy,$$

the sum of all 4 words of length 2 in the alphabet $\{x, y\}$. Similarly

$$(x + y)^3 = xxx + \underline{xxy} + \underline{xyx} + xyy + \underline{yxx} + yxy + yyx + yyy,$$

the sum of all 8 words of length 3 in the alphabet $\{x, y\}$. After simplification, we get the familiar formulas

$$(x + y)^2 = x^2 + 2xy + y^2, \qquad \text{and}$$
$$(x + y)^3 = x^3 + \underline{3}x^2y + 3xy^2 + y^3.$$

Where did the coefficient 3 of x^2y come from? It came from the 3 underlined words in the presimplified expansion of $(x + y)^3$. And how could we have predicted that this coefficient would be 3, without writing out the presimplified expansion? Well, the number of words of length 3 comprised of 2 x's and 1 y is $\binom{3}{1} = \binom{3}{2} = 3$, that's how. This leads to a proof of the binomial theorem:

Theorem 4.2. *For all $n \in \mathbb{N}$,*

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

Proof. Proof. Before simplification the expansion of $(x + y)^n$ consists of the sum of all 2^n words of length n in the alphabet $\{x, y\}$. The number of such words comprised of k x's and $n - k$ y's is $\binom{n}{k}$ by Theorem 4.1. $\square$

The formula

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = y^n + \binom{n}{1} xy^{n-1} + \binom{n}{2} x^2 y^{n-2} + \cdots + x^n$$

is an expansion in *ascending powers of x*. We can of course also express the binomial theorem as

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = x^n + \binom{n}{1} x^{n-1}y + \binom{n}{2} x^{n-2}y^2 + \cdots + y^n,$$

an expansion in *descending powers of x*.

By substituting various numerical values for x and y in the binomial theorem, one can generate various binomial coefficient identities:

(i) Setting $x = y = 1$ in Theorem 4.2 yields

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

which yields an alternative proof of Theorem 3.5.

(ii) Setting $x = -1$ and $y = 1$ yields, for $n \in \mathbb{P}$,

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = \binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^n \binom{n}{n} = 0,$$

i.e.,

$$\binom{n}{0} + \binom{n}{2} + \cdots = \binom{n}{1} + \binom{n}{3} + \cdots,$$

i.e.,

$$\sum_{\substack{0 \le k \le n \\ k \text{ even}}} \binom{n}{k} = \sum_{\substack{0 \le k \le n \\ k \text{ odd}}} \binom{n}{k}. \quad \text{(See Problem 3)}$$

It is important to learn to recognize when a sum is in fact a binomial expansion, or nearly so. The following illustrates a few tricks of the trade

(iii) Simplify $\sum_{k=0}^{n} \binom{n}{k} a^k$. Solution:

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

(iv) Simplify $\sum_{k=1}^{17} (-1)^k \binom{17}{k} 13^k$. Solution:

$$\sum_{k=1}^{17} (-1)^k \binom{17}{k} 13^k = \sum_{k=1}^{17} \binom{17}{k} (-13)^k (1)^{17-k}$$

$$= \sum_{k=0}^{17} \binom{17}{k} (-13)^k (1)^{17-k} - \binom{17}{0} (-13)^0 (1)^{17}$$

$$= (-13 + 1)^{17} - 1 = (-12)^{17} - 1.$$

(v) Simplify $\sum_{k=1}^{17}(-1)^k\binom{17}{k}13^{17-k}$. Solution:

$$\sum_{k=1}^{17}(-1)^k\binom{17}{k}13^{17-k} = \sum_{k=0}^{17}\binom{17}{k}(-1)^k(13)^{17-k} - \binom{17}{0}(-1)^0(13)^{17}$$

$$= (-1+13)^{17} - 13^{17} = 12^{17} - 13^{17}.$$

4.3 Counting lattice paths in the plane

How many paths from A to B are there on the "lattice" shown in Figure 4.3 if we must always move east or north?
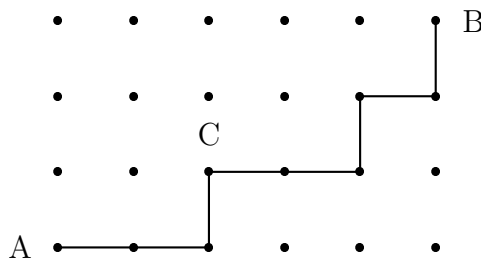
Figure 4.1: Path EENEENEN

Solution: Any lattice path from A to B involves 3 moves north and 5 moves east in some sequence. Indeed the lattice paths in question are in one-to-one correspondence with all possible words of length 8 consisting of 3 N's and 5 E's. By Theorem 4.1, there are $\binom{8}{3} = \binom{8}{5} = 8!/3!5!$ such words, and thus that number of lattice paths.

If asked for the number of lattice paths A to B that pass through C, we would simply count paths from A to C and paths from C to B and multiply:

$$\binom{3}{1} \times \binom{5}{2}.$$

In general, the number of lattice paths in $\mathbb{R}^2$ from $(0,0)$ to (p,q) is

$$\binom{p+q}{p} = \binom{p+q}{q} = \frac{(p+q)!}{p!q!}.$$

The solution to the above problem is easy, once you look at it the right way. But one doesn't always have an immediate clever insight into the simplest solution to a problem. One of the nice things about elementary enumerative combinatorics is that one can often approach problems by making a table of values for small values of the relevant parameters by brute force enumeration. To approach the above problem we could make the table below, where the entry in row p and column q is the number of lattice paths from $(0,0)$ to (p,q), determined by listing and counting. The numbers we are getting support the conjecture that the solution is a binomial coefficient of

$p \backslash q$	0	1	2	3
0	1	1	1	1
1	1	2	3	4
2	1	3	6	10
3	1	4	10	20

Table 4.1: Lattice Path Enumeration

some sort, and it is not too hard to guess that it is $\binom{p+q}{p}$. Of course, one then has to give a proof. But in the process of listing and counting, one may well have discovered the key representation of a lattice path as a word in the alphabet $\{N, E\}$.

Another approach to this problem uses recursion. From the table it appears that, letting $L(p, q)$ denote the number of lattice paths from $(0, 0)$ to (p, q), we have $L(p, 0) = L(0, q) = 1$ for all $p, q \in \mathbb{N}$ and

$$L(p, q) = L(p, q - 1) + L(p - 1, q)$$

if $p, q \in \mathbb{P}$. There is an easy combinatorial proof of this recurrence: We categorize the lattice paths in question in two classes (i) those paths in which the last point we visit before reaching (p, q) is $(p, q - 1)$ and (ii) those in which the last point visited before (p, q) is $(p - 1, q)$. It is easy to check that $L(p, q) = \binom{p+q}{p}$ satisfies the above boundary values and recurrence. Since there is clearly a unique function of p and q satisfying those boundary values and that recurrence, it must be that $L(p, q) = \binom{p+q}{p}$.

Using the above strategy, you should be able to solve Honors Problem III.

Chapter 5

Sequences of Integers with Prescribed Sum

5.1   Urn problems with indistinguishable balls

We know from Theorem 2.3 that there are 5^{12} ways to distribute 12 labeled balls among 5 labeled urns. Suppose, however, that the balls are unlabeled and thus indistinguishable. If such balls are distributed among urns $u_1, u_2, \ldots, u_5$, we cannot tell *which* balls occupy urn u_i, but only *how many* balls, n_i, occupy u_i, for $i = 1, \ldots, 5$. It is thus clear that the following are two different ways of asking the same question:

  (i) In how many ways may 12 indistinguishable balls be distributed among 5 urns labeled $u_1, \ldots, u_5$?

  (ii) How many sequences $(n_1, n_2, \ldots, n_5)$ in $\mathbb{N}$ are there satisfying $n_1 + n_2 + \cdots + n_5 = 12$, i.e., in how many ways may we write 12 as an ordered sum of 5 nonnegative integers?

A pictorial representation of several distributions suggests how to solve this problem.

$$| \, 0 \, 0 \, 0 \, | \, 0 \, 0 \, | \quad | \, 0 \, 0 \, 0 \, 0 \, | \, 0 \, 0 \, 0 \, |$$

$$3 + 2 + 0 + 4 + 3 = 12$$

$$| \quad | \, 0 \, 0 \, 0 \, 0 \, 0 \, | \, 0 \, 0 \, | \, 0 \, 0 \, 0 \, 0 \, 0 \, | \quad |$$

$$0 + 5 + 2 + 5 + 0 = 12$$

We see that distributions correspond in one-to-one fashion with sequences of 6 vertical lines and 12 0's, with the stipulation that the first and last entry of the sequence is a vertical line. Ignoring these two vertical lines, we see that distributions correspond in one-to-one fashion with arbitrary sequences of 4 vertical lines and 12 0's, of which there are $\binom{16}{4}$ by Theorem 4.1.

More generally the distributions of n indistinguishable balls among k labeled urns correspond in one-to-one fashion with sequences of $k-1$ vertical lines and n 0's, which leads to the following theorem.

Theorem 5.1. *For all $n \in \mathbb{N}$ and all $k \in \mathbb{P}$, there are $\binom{n+k-1}{k-1}$ distributions of n indistinguishable balls among k labeled urns. Equivalently, there are $\binom{n+k-1}{k-1}$ sequences $(n_1, n_2, \ldots, n_k)$ in $\mathbb{N}$ such that $n_1 + n_2 + \cdots + n_k = n$.*

Proof. Use above remarks and Theorem 4.1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

A sequence $(n_1, n_2, \ldots, n_k)$ in $\mathbb{N}$ such that $n_1 + n_2 + \cdots + n_k = n$ is called a *weak composition of n with k parts*. If all of the parts n_i are positive, then $(n_1, \ldots, n_k)$ is called a *composition of n with k parts*. A composition of n with k parts corresponds to a distribution of n indistinguishable balls among k labeled urns with at least one ball per urn.

Theorem 5.2. *There are $\binom{n-1}{k-1}$ distributions of n indistinguishable balls among k labeled urns with at least one ball per urn, and hence $\binom{n-1}{k-1}$ compositions of n with k parts.*

#1. Place one ball in each of the k urns, fulfilling the requirement of at least one ball per urn. Since balls are indistinguishable, there is one way to do this. Next, distribute the remaining $n-k$ balls among the k urns. By Theorem 5.1, there are $\binom{(n-k)+k-1}{k-1} = \binom{n-1}{k-1}$ ways to do this. $\qquad$ $\square$

#2. Insert $k-1$ vertical lines below, choosing from the $n-1$ spaces between balls for their position

$$\underline{\mid 0 \ \ 0 \ \ 0 \ \ 0 \quad \cdots\cdots\cdots \quad 0 \ \ 0 \mid}.$$

$$n-1 \text{ possible locations of}$$
$$k-1 \text{ vertical lines}$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

In the preceding theorem we counted distributions subject to the requirement that at least one ball be placed in each urn. Obviously this can be generalized, with lower bounds of a general sort placed on the "occupancy numbers" of each urn.

Theorem 5.3. *If $b_1 + b_2 + \cdots + b_k \leq n$, there are*

$$\binom{n - b_1 - b_2 - \cdots - b_k + k - 1}{k - 1}$$

distributions of n indistinguishable balls among k urns labeled $u_1, \ldots, u_k$ such that at least b_i balls are placed in urn u_i, $i = 1, \ldots, k$.

Proof. Place b_i balls in urn u_i, $i = 1, \ldots, k$ and proceed as in the proof of Theorem 5.2. $\qquad$ $\square$

Remark. It is natural in the above formulation to think of the b_i as nonnegative integers. But the result is actually more general. This general result is best expressed in terms of sequences of integers with prescribed sum and integer lower bounds on the individual terms.

Theorem 5.4. *If $n \in \mathbb{N}$, $k \in \mathbb{P}$, and $(b_1, b_2, \ldots, b_k)$ is a sequence in $\mathbb{Z}$ such that $b_1 + b_2 + \cdots + b_k \leq n$, then there are*

$$\binom{n - b_1 - b_2 - \cdots - b_k + k - 1}{k - 1}$$

sequences $(n_1, n_2, \ldots, n_k)$ in $\mathbb{Z}$ such that $n_i \geq b_i$, $i = 1, \ldots, k$, and
$n_1 + n_2 + \cdots + n_k = n$.

Proof. Corresponding to each such sequence $(n_1, n_2, \ldots, n_k)$ is a sequence $(m_1, \ldots, m_k)$ in $\mathbb{N}$, where $m_i = n_i - b_i$, satisfying $m_1 + m_2 + \cdots + m_k = n - b_1 - b_2 - \cdots - b_k$. Since this correspondence is clearly one-to-one and there are $\binom{n - b_1 - \cdots - b_k + k - 1}{k - 1}$ sequences $(m_1, \ldots, m_k)$ in $\mathbb{N}$ summing to the nonnegative integer $n - b_1 - \cdots - b_k$ by Theorem 5.1, the desired result follows. $\square$

Remark. Given Theorem 5.4, Theorem 5.3 is superfluous, since it merely represents the special case where $(b_1, \ldots, b_k)$ is a sequence in $\mathbb{N}$, phrased in the language of distributions. Theorem 5.3 was included for pedagogical, not logical reasons.

5.2   The family of all compositions of n

Given $n \in \mathbb{P}$, there are by Theorem 5.2 $\binom{n-1}{0}$ compositions of n with 1 part, $\binom{n-1}{1}$ compositions of n with 2 parts, ..., and $\binom{n-1}{n-1}$ compositions of n with n parts. There are no compositions of n with k parts if $k > n$ (then $k - 1 > n - 1$ and so $\binom{n-1}{k-1} = 0$, so our formula works for all $n, k \in \mathbb{P}$).

Theorem 5.5. *There are 2^{n-1} compositions of the positive integer n.*

Proof. Sum the number of compositions of n with k parts for $k = 1, \ldots, n$, getting

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

$\square$

Remark. There is an infinite number of weak compositions of the nonnegative integer n, since $\binom{n+k-1}{k-1} > 0$ for all $k \in \mathbb{P}$.

Remark. One can also prove Theorem 5.5 without using Theorem 5.2. Let $C(n)$ denote the number of compositions of n. Then clearly $C(1) = 1$ and for all $n \geq 2$

$$C(n) = C(n - 1) + C(n - 2) + \cdots + C(1) + 1,$$

as one sees by categorizing compositions according to the size $(1, 2, \ldots, n - 1$, or $n)$ of the first term of the composition. Using this recurrence relation, it is straightforward to prove by (course-of-values) induction on n that $C(n) = 2^{n-1}$ for all $n \in \mathbb{P}$.

5.3 Upper bounds on the terms of sequences with prescribed sum

In contrast to the case of lower bounds, determining the number of sequences with prescribed sum subject to upper bounds on individual terms is much more difficult. In Chapter 14 we will exhibit a general solution to this problem. In the mean time, however, we wish to introduce a powerful method for solving concrete cases of enumeration problems of this sort, indeed a method that accomodates virtually *any* restrictions that we like on the individual terms.

We begin with a simple example. If a red die and a clear die are tossed, in how many ways may the sum 9 be attained? Elementary probability texts usually solve this problem by brute force, deriving the answer 4. Here is another approach, which not only solves the problem for the sum 9, but for all possible sums.

Simply multiply the polynomials

$$(x + x^2 + x^3 + x^4 + x^5 + x^6)(x + x^2 + x^3 + x^4 + x^5 + x^6)$$
$$= x^2 + 2x^3 + 3x^4 + 4x^5 + 5x^6 + 6x^7$$
$$+ 5x^8 + 4x^9 + 3x^{10} + 2x^{11} + x^{12}.$$

Note that the coefficient of x^9 is 4, the answer to our problem. More generally, the coefficient of x^n (for $n = 2, \ldots, 12$) is the number of ways to get sum n, i.e., the number of sequences (r, c) in $\mathbb{N}$ such that $1 \le r, c \le 6$ and $r + c = n$.

If there were a third (say, blue) die we would expand

$$(x + \cdots + x^6)^3 = x^3 + 3x^4 + \cdots + x^{18},$$

where the coefficient of x^n is the number of ways to get sum n with 3 dice. If k dice were rolled, the "generating function" of the solution would be $(x + \cdots + x^6)^k$.

In general to determine the number of sequences $(n_1, n_2, \ldots, n_k)$ such that $n_1 + \cdots + n_k = n$, with various restrictions on the n_i's, we form k polynomials, each corresponding to a term n_i, with the ith polynomial consisting of a sum of powers of x, the powers being the allowable values of the term n_i. We multiply these k polynomials. The coefficient of x^n is the solution. Of course, we wind up solving a lot of other counting problems by this method, not just the problem of counting the ways to get a particular sum n.

Example. How many sequences (n_1, n_2, n_3) of integers are there such that $n_1 + n_2 + n_3 = 12$, subject to the restrictions that n_1 is odd and $1 \le n_1 \le 5$, n_2 is even and $0 \le n_2 \le 8$, and n_3 is prime and $3 \le n_3 \le 11$?

Solution: Multiply the polynomials

$$(x + x^3 + x^5)(1 + x^2 + x^4 + x^6 + x^8)(x^3 + x^5 + x^7 + x^{11}).$$

The answer is the coefficient of x^{12}. (If no term involving the power x^{12} appears in the expansion of this product, the answer is 0.)

Remark. Polynomial multiplication can be done on a computer using any of a number of mathematical packages.

Remark. One can use negative powers of x with the above method when negative integers are allowed as values of the terms in the integer sequences under consideration. It is also possible to have an infinite number of nonnegative powers appearing in a sum of powers of x. One can often derive general formulas by this technique. For example, the number of weak compositions of n with k parts is the coefficient of x^n in the expansion

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

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

by Newton's generalization of the binomial theorem. But

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

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

$$= \frac{(n+k-1)^{\underline{n}}}{n!} = \binom{n+k-1}{n} = \binom{n+k-1}{k-1},$$

which provides an alternative proof of Theorem 5.1.

Chapter 6

Sequences of k Sorts of Things with Prescribed Frequency

6.1 Trinomial Coefficients

Recall from Chapter 4 (Theorem 4.1 ff.) that if $n_1 + n_2 = n$, then $\binom{n}{n_1} = \binom{n}{n_2} = \frac{n!}{n_1! n_2!}$ counts

(i) the number of sequences of n_1 x's and n_2 y's,

(ii) the number of distributions of n labeled balls among 2 urns, u_1 and u_2, such that n_i balls are placed in u_i, $i = 1, 2$,

(iii) the number of functions $f : \{a_1, \ldots, a_n\} \rightarrow \{b_1, b_2\}$ such that $|f^{\leftarrow}(b_i)| = n_i$, $i = 1, 2$.

We wish to generalize these results in this chapter. This generalization will preserve the parallels between sequences with prescribed frequency of types of terms, distributions with prescribed occupancy numbers, and functions with prescribed preimage cardinalities.

We start with a concrete example: In how many ways may 2 x's, 3 y's, and 5 z's be arranged in a sequence? Using the strategy of Chapter 4, consider the problem of filling the 10 slots below with appropriate letters.

$$\overline{(1)}\ \overline{(2)}\ \overline{(3)}\ \overline{(4)}\ \overline{(5)}\ \overline{(6)}\ \overline{(7)}\ \overline{(8)}\ \overline{(9)}\ \overline{(10)}$$

First we choose 2 of the slots (in any of the $\binom{10}{2}$ possible ways) in which to place the 2 x's. Then, from the remaining 8 slots we choose 3 slots (in any of the $\binom{8}{3}$ possible ways) in which to place the 3 y's. The 5 z's go in the remaining 5 slots. The solution is therefore

$$\binom{10}{2}\binom{8}{3} = \frac{10!}{2!8!} \cdot \frac{8!}{3!5!} = \frac{10!}{2!3!5!}.$$

Clearly, this is also the number of ways to distribute 10 labeled balls among 3 urns, u_1, u_2 and u_3, such that 2 balls are placed in u_1, 3 balls in u_2, and 5 balls in u_3. It is also the number of functions $f : \{a_1, \ldots, a_{10}\} \rightarrow \{b_1, b_2, b_3\}$ with $|f^{\leftarrow}(b_1)| = 2$, $|f^{\leftarrow}(b_2)| = 3$, and $|f^{\leftarrow}(b_3)| = 5$.

The number $\frac{10!}{2!3!5!}$ is a trinomial coefficient of order 10, often denoted $\binom{10}{2,3,5}$. In general, if $n, n_1, n_2, n_3 \in \mathbb{N}$ and $n_1 + n_2 + n_3 = n$,

$$\binom{n}{n_1, n_2, n_3} := \binom{n}{n_1}\binom{n - n_1}{n_2} = \frac{n!}{n_1! n_2! n_3!}$$

is termed a *trinomial coefficient of order n.*

Theorem 6.1. *If $n, n_1, n_2, n_3 \in \mathbb{N}$ and $n_1 + n_2 + n_3 = n$, then $\binom{n}{n_1, n_2, n_3}$ counts*

(i) *the number of sequences of n_1 x's, n_2 y's, and n_3 z's,*

(ii) *the number of distributions of n labeled balls among 3 urns, labeled u_1, u_2 and u_3, such that n_i balls are placed in u_i, $i = 1, 2, 3$, and*

(iii) *the number of functions $f : \{a_1, \ldots, a_n\} \to \{b_1, b_2, b_3\}$ such that $|f^{\leftarrow}(b_i)| = n_i$, $i = 1, 2, 3$.*

Proof. Choose n_1 of the n slots (balls; elements of $\{a_1, \ldots, a_n\}$) in which to place the n_1 x's (which will be placed in u_1; which will be mapped to b_1) in any of the $\binom{n}{n_1}$ possible ways. Choose n_2 of the remaining $n - n_1$ slots (balls, elements of $\{a_1, \ldots, a_n\}$) in which to place the n_2 y's (which will be placed in u_2, which will be mapped to b_2) in any of the $\binom{n - n_1}{n_2}$ possible ways. Place the n_3 z's in the remaining $n - n_1 - n_2 = n_3$ slots (place the remaining $n - n_1 - n_2 = n_3$ balls in u_3; map the remaining $n - n_1 - n_2 = n_3$ elements of $\{a_1, \ldots, a_n\}$ to b_3.) Multiply

$$\binom{n}{n_1}\binom{n - n_1}{n_2} = \frac{n!}{n_1!(n - n_1)!}\frac{(n - n_1)!}{n_2!(n - n_1 - n_2)!}$$
$$= \frac{n!}{n_1! n_2! n_3!} = \binom{n}{n_1, n_2, n_3}.$$

$\square$

Remark. Statisticians often use the abbreviated notation $\binom{n}{n_1, n_2}$ for the above trinomial coefficient, in analogy with the binomial coefficient notation $\binom{n}{n_1}$. We shall *not* follow this practice. If we write $\binom{n}{n_1, n_2}$, it will be the case that $n_1 + n_2 = n$, with $\binom{n}{n_1, n_2}$ simply being a more elaborate notation for $\binom{n}{n_1} = \binom{n}{n_2} = \frac{n!}{n_1! n_2!}$.

There are $n + 1$ *notationally distinct* binomial coefficients of order n, namely $\binom{n}{0} = \binom{n}{0, n}$, $\binom{n}{1} = \binom{n}{1, n-1}, \ldots, \binom{n}{n} = \binom{n}{n, 0}$. Of course notationally distinct binomial coefficients need *not* take *distinct values*. We know, for example, that $\binom{n}{k} = \binom{n}{n-k}$, which could also be expressed in more elaborate notation $\binom{n}{n_1, n_2} = \binom{n}{n_2, n_1}$, where $n_1 + n_2 = n$. This symmetry property of binomial coefficients accounts for all of the repeated values in a given row of Pascal's triangle, i.e., $\binom{n}{i} = \binom{n}{j}$ for $1 \leq i, j \leq n$ if and only if $j = i$ or $j = n - i$ (It is an interesting exercise to prove this. Try it.)

How many notationally distinct trinomial coefficients of order n are there? There is one such coefficient $\binom{n}{n_1, n_2, n_3}$ corresponding to each sequence (n_1, n_2, n_3) in $\mathbb{N}$ with $n_1 + n_2 + n_3 = n$. By Theorem 5.1, the answer to this question is therefore

$$\binom{n + 3 - 1}{3 - 1} = \binom{n + 2}{2} = \frac{n^2 + 3n + 2}{2}.$$

Of course, notationally distinct trinomial coefficients need not take distinct values. Corresponding to the symmetry property of binomial coefficients, we have

$$\binom{n}{n_1, n_2, n_3} = \binom{n}{m_1, m_2, m_3}$$

whenever (m_1, m_2, m_3) is a rearrangement of (n_1, n_2, n_3).

Question. Does the above symmetry property of trinomial coefficients account for all repeated values among trinomial coefficients of order n?

The sum of all binomial coefficients of order n is 2^n. The sum of all trinomial coefficients of order n is given by the following theorem.

Theorem 6.2. *For all $n \in \mathbb{N}$,*

$$\sum_{\substack{n_1+n_2+n_3=n \\ n_i \in \mathbb{N}}} \binom{n}{n_1, n_2, n_3} = 3^n.$$

Proof. Each side of the above equation counts the class of distributions of n labeled balls among 3 urns u_1, u_2 and u_3. The RHS counts this family of distributions by Theorem 2.3. The LHS counts this family in $\binom{n+2}{2}$ subclasses, one for each sequence (n_1, n_2, n_3) of occupancy numbers for u_1, u_2, and u_3. For by Theorem 6.1, there are $\binom{n}{n_1, n_2, n_3}$ distributions with occupancy numbers (n_1, n_2, n_3). □

Remark. One can also give a sequence counting or function counting argument for the above identity.

6.2 The trinomial theorem

If we expand $(x + y + z)^2 = (x + y + z)(x + y + z)$, we get, before simplification,

$$(x + y + z)^2 = xx + xy + xz + yx + yy + yz + zx + zy + zz,$$

the sum of all $3^2 = 9$ words of length 2 in the alphabet $\{x, y, z\}$. Similarly,

$$\begin{aligned}
(x + y + z)^3 = {} & xxx + xxy + xxz + xyx + xyy + xyz + xzx + xzy + xzz + yxx \\
& + yxy + yxz + yyx + yyy + yyz + yzx + yzy + yzz + zxx + zxy \\
& + zxz + zyx + zyy + zyz + zzx + zzy + zzz,
\end{aligned}$$

the sum of all $3^3 = 27$ words of length 3 in the alphabet $\{x, y, z\}$. After simplification, we get

$$(x + y + z)^2 = x^2 + 2xy + 2xz + y^2 + 2yz + z^2$$

and

$$(x + y + z)^3 = x^3 + 3x^2y + 3x^2z + 3xy^2 + 6xyz + 3xz^2 + y^3 + 3y^2z + 3yz^2 + z^3.$$

More generally, before simplification, the expansion of $(x + y + z)^n$ involves the sum of all 3^n words of length n in the alphabet $\{x, y, z\}$. The following "trinomial theorem" describes the situation after simplification.

35

Theorem 6.3. *For all $n \in \mathbb{N}$,*

$$(x+y+z)^n = \sum_{\substack{n_1+n_2+n_3=n \\ n_i \in \mathbb{N}}} \binom{n}{n_1, n_2, n_3} x^{n_1} y^{n_2} z^{n_3}.$$

Proof. By Theorem 6.1, there are $\binom{n}{n_1, n_2, n_3}$ words of length n in the alphabet $\{x, y, z\}$ in which x appears n_1 times, y appears n_2 times, and z appears n_3 times. $\qquad\square$

Remark. As noted above, before simplification, the expansion of $(x+y+z)^n$ involves a sum of 3^n terms. After simplification, there are as many terms as there are weak compositions of n with 3 parts, namely $\binom{n+2}{2}$.

In the foregoing, we have written the expansions of $(x+y+z)^2$ and $(x+y+z)^3$ in a particular order. But that order is not dictated by the formula

$$(x+y+z)^n = \sum_{\substack{n_1+n_2+n_3=n \\ n_i \in \mathbb{N}}} \binom{n}{n_1, n_2, n_3} x^{n_1} y^{n_2} z^{n_3},$$

which simply tells us to find all weak compositions (n_1, n_2, n_3) of n, form the associated terms $\binom{n}{n_1, n_2, n_3} x^{n_1} y^{n_2} z^{n_3}$, and sum these terms in whatever order we like.

In contrast, writing the binomial theorem in the form

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

dictates that the expansion be written in the order $y^n + \binom{n}{1} xy^{n-1} + \cdots + \binom{n}{n-1} x^{n-1} y + x^n$. We could write the binomial theorem, in such a way that the ordering of terms is not dictated, as

$$(x+y)^n = \sum_{\substack{n_1+n_2=n \\ n_i \in \mathbb{N}}} \binom{n}{n_1, n_2} x^{n_1} y^{n_2},$$

although this is rarely done.

Can we write the trinomial theorem in such a way that the ordering of the terms is dictated? The answer is affirmative, but we require double summation, as follows:

$$\begin{aligned}
(x+y+z)^n &= \sum_{i=0}^{n} \sum_{j=0}^{n-i} \binom{n}{i, j, n-i-j} x^i y^j z^{n-i-j} \\
&= \left[\binom{n}{0, 0, n} x^0 y^0 z^n + \binom{n}{0, 1, n-1} x^0 y^1 z^{n-1} + \cdots + \binom{n}{0, n, 0} x^0 y^n z^0 \right] \\
&\quad + \left[\binom{n}{1, 0, n-1} x^1 y^0 z^{n-1} + \binom{n}{1, 1, n-2} x^1 y^1 z^{n-2} + \cdots + \binom{n}{1, n-1, 0} x^1 y^{n-1} z^0 \right] \\
&\quad + \left[\binom{n}{2, 0, n-2} x^2 y^0 z^{n-2} + \cdots \right] + \cdots + \left[\binom{n}{n, 0, 0} x^n y^0 z^0 \right]
\end{aligned}$$

In this expansion we first set $i = 0$ and run j from 0 to n, then set $i = 1$ and run j from 0 to $n - 1$, then set $i = 2$ and run j from 0 to $n - 2$, etc.

Just as we can generate various binomial coefficient identities from the binomial theorem by substituting particular values for x and y, we can generate various identities involving trinomial

coefficients from the trinomial theorem. For example, setting $x = y = z = 1$ in Theorem 6.3 yields

$$\sum_{\substack{n_1+n_2+n_3=n \\ n_i \in \mathbb{N}}} \binom{n}{n_1, n_2, n_3} = 3^n,$$

which provides an alternative to the combinatorial proof of this identity offered in Theorem 6.2.

6.3 Multinomial coefficients and the multinomial theorem

For all $n \in \mathbb{N}$, all $k \in \mathbb{P}$, and for every weak composition $(n_1, n_2, \ldots, n_k)$ of n with k parts, the k-nomial coefficient of order n, $\binom{n}{n_1, n_2, \ldots, n_k}$ is defined by

$$\binom{n}{n_1, n_2, \ldots, n_k} := \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3} \cdots \binom{n-n_1-n_2-\cdots-n_{k-2}}{n_{k-1}}$$
$$= \frac{n!}{n_1! n_2! \cdots n_k!}.$$

The following three theorems generalize Theorems 6.1, 6.2, and 6.3.

Theorem 6.4. *For all $n \in \mathbb{N}$, all $k \in \mathbb{P}$, and for every weak composition $(n_1, n_2, \ldots, n_k)$ of n with k parts, the k-nomial coefficient $\binom{n}{n_1, n_2, \ldots, n_k}$ counts*

(i) the number of sequences of n_1 x_1's, n_2 x_2's, $\ldots$, and n_k x_k's,

(ii) the number of distributions of n labeled balls among k urns, labeled $u_1, u_2, \ldots, u_k$, such that n_i balls are placed in u_i, $i = 1, \ldots, k$, and

(iii) the number of functions $f : \{a_1, a_2, \ldots, a_n\} \to \{b_1, b_2, \ldots, b_k\}$ such that $|f^{\leftarrow}(b_i)| = n_i$, $i = 1, \ldots, k$.

Proof. Straightforward generalization of the proof of Theorem 6.1. $\square$

Theorem 6.5. *For all $n \in \mathbb{N}$ and all $k \in \mathbb{P}$,*

$$\sum_{\substack{n_1+n_2+\cdots+n_k=n \\ n_i \in \mathbb{N}}} \binom{n}{n_1, n_2, \ldots, n_k} = k^n.$$

That is, the sum of all k-nomial coefficients of order n is k^n.

Proof. Straightforward generalization of the proof of Theorem 6.2. $\square$

Theorem 6.6 (Multinomial Theorem). *For all $n \in \mathbb{N}$ and all $k \in \mathbb{P}$,*

$$(x_1 + x_2 + \cdots + x_k)^n = \sum_{\substack{n_1+n_2+\cdots+n_k=n \\ n_i \in \mathbb{N}}} \binom{n}{n_1, n_2, \ldots, n_k} x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}.$$

Proof. The proof is straightforward generalization of the proof of Theorem 6.3. $\qquad\square$

Remark. Before simplification, the expansion of $(x_1 + x_2 + \cdots + x_k)^n$ consists of the sum of all k^n words of length n in the alphabet $\{x_1, x_2, \ldots, x_k\}$. After simplification there are as many terms as there are weak compositions of n with k parts, namely $\binom{n+k-1}{k-1}$.

Remark. Theorem 6.6 does not dictate the order in which the terms of the simplified expansion of $(x_1 + x_2 + \cdots + x_k)^n$ appear. One way to dictate that order is to use $(k-1)$-fold summation, e.g.,

$$(x_1 + x_2 + \cdots + x_k)^n =$$

$$\sum_{n_1=0}^{n} \sum_{n_2=0}^{n-n_1} \sum_{n_3=0}^{n-n_1-n_2} \cdots \sum_{n_{k-1}=0}^{n-n_1-\cdots-n_{k-2}} \binom{n}{n_1, n_2, \ldots, n_{k-1}, n-n_1-\cdots-n_{k-1}} x_1^{n_1} x_2^{n_2} \cdots x_{k-1}^{n_{k-1}} x_k^{n-n_1-\cdots-n_{k-1}}.$$

Chapter 7

Combinatorics and Probability

Enumerative combinatorics provides one of the basic foundations of probability theory. For if S is a finite set with subset E and one chooses an element of S at random, the probability that the element chosen actually belongs to E is given by the formula $|E|/|S|$. The determination of the cardinalities of various sets thus underlies this probability model (the so-called "uniform model") in a crucial way.

The applications of enumerative combinatorics are by no means limited to uniform models. Indeed, they permeate all of discrete probability theory. This chapter outlines a few particularly basic applications.

7.1   The Multinomial Distribution

We begin by reviewing the binomial distribution, which arises in the case of n independent trials, each having two possible outcomes, called arbitrarily "success" and "failure." On each trial the probability of success is p and the probability of failure is q, where $p + q = 1$. If the "random variable" X records the total number of successes on n trials, X is said to be *binomially distributed with parameters n and p* (abbreviated $X \sim \text{binomial}\,(n, p)$). For each $k = 0, 1, \ldots, n$

$$P(X = k) = \text{the probability of exactly } k \text{ successes in } n \text{ trials}$$

$$= \binom{n}{k} p^k q^{n-k}.$$

The proof of this assertion is essentially that of the binomial theorem: Any sequence of k successes and (hence) $n - k$ failures has probability $p^k q^{n-k}$ by the independence of outcomes from trial to trial. By Theorem 4.1, there are $\binom{n}{k}$ such sequences. Of course,

$$\sum_{k=0}^{n} P(x = k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1,$$

as one would expect from probabilistic considerations.

Suppose that on each trial there are *three* possible outcomes o_1, o_2, and o_3, with respective probabilities p_1, p_2, and p_3, where $p_1 + p_2 + p_3 = 1$. Trials are again independent. Let X_i record

the number of occurrences of outcome o_i in n trials, $i = 1, 2, 3$. Then, clearly, for each weak composition (n_1, n_2, n_3) of n,

$$P(X_1 = n_1, X_2 = n_2, X_3 = n_3) = \binom{n}{n_1, n_2, n_3} p_1^{n_1} p_2^{n_2} p_3^{n_3}.$$

The "jointly distributed" random variables X_1, X_2, and X_3 are said to be trinomially distributed with parameters n, p_1, p_2, p_3 (abbreviated $(X_1, X_2, X_3) \sim \text{trinomial}(n; p_1, p_2, p_3)$). Of course,

$$\sum_{\substack{n_1+n_2+n_3=n \\ n_i \in \mathbb{N}}} P(X_1 = n_1, X_2 = n_2, X_3 = n_3) = \sum_{\substack{n_1+n_2+n_3=n \\ n_i \in \mathbb{N}}} \binom{n}{n_1, n_2, n_3} p_1^{n_1} p_2^{n_2} p_3^{n_3}$$
$$= (p_1 + p_2 + p_3)^n = 1,$$

as one would expect on probabilistic grounds.

Remark. Statisticians often delete the random variable X_3, since $X_3 = n - X_1 - X_2$, writing

$$P(X_1 = n_1, X_2 = n_2) = \binom{n}{n_1, n_2} p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n-n_1-n_2}$$

for all $n_1, n_2 \in \mathbb{N}$ with $n_1 + n_2 \le n$, where $\binom{n}{n_1, n_2} := n!/n_1! n_2! (n - n_1 - n_2)!$. They refer to X_1 and X_2 as being trinomially distributed with parameters n, p_1, and p_2.

Example. Suppose that we toss a pair of dice 12 times, with o_1 denoting the outcome that the sum on a given toss is even, o_2 the outcome that the sum is odd and ≤ 3, and o_3 the outcome that the sum is odd and ≥ 5.

By expanding $(x + \cdots + x^6)^2$ we can readily determine the probabilities p_i of outcomes o_i, $i = 1, 2, 3$. We get $p_1 = 18/36 = 1/2$, $p_2 = 2/36 = 1/18$, and $p_3 = 16/36 = 4/9$. Consequently, for example,

$$P(X_1 = 2, X_2 = 3, X_3 = 7) = \binom{12}{2, 3, 7} \left(\frac{1}{2}\right)^2 \left(\frac{1}{18}\right)^3 \left(\frac{4}{9}\right)^7,$$

where X_i records the number of occurrences of outcome o_i, $i = 1, 2, 3$.

If, on each of n independent trials, there are k possible outcomes $o_1, \ldots, o_k$, with respective probabilities $p_1, \ldots, p_k$ ($p_1 + \cdots + p_k = 1$), and if X_i records the number of occurrences of o_i in these n trials, then for every weak composition $(n_1, \ldots, n_k)$ of n,

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \binom{n}{n_1, n_2, \ldots, n_k} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}.$$

Here the jointly distributed random variables $X_1, \ldots, X_k$ are k-nomially distributed with parameters $n; p_1, \ldots, p_k$.

Remark. To test the hypothesis that $X_1, \ldots, X_k$ are k-nomially distributed with parameters $n; p_1, \ldots, p_k$, one uses a so-called chi-squared test (see Larsen & Marx, *An Introduction to Mathematical Statistics and its Applications*, Prentice-Hall, Chapter 9). Multinomial distributions also underlie other "goodness of fit" tests in statistics. To test the hypothesis that a continuous random variable Y has density function f, for example, one can choose values $y_1 < y_2 < \cdots < y_{k-1}$,

let o_1 denote the outcome $Y < y_1$, o_i the outcome $y_{i-1} < Y < y_i$, $i = 2, \ldots, k-1$, and o_k the outcome $Y > y_{k-1}$. If we make n independent observations of the value of Y, letting X_i record the number of occurrences of o_i, then $(X_1, \ldots, X_k)$ should be k-nomially distributed with parameters $n; p_1, \ldots, p_k$, where

$$p_1 = \int_{-\infty}^{y_1} f(y)dy, \quad p_i = \int_{y_{i-1}}^{y_i} f(y)dy, \quad i = 2, \ldots, k-1,$$

and $p_k = \int_{y_{k-1}}^{\infty} f(y)dy$. If this hypothesis regarding $(X_1, \ldots, X_k)$ is rejected by the appropriate chi-squared test, doubt is accordingly cast on the hypothesis that Y has density function f.

7.2 Moments and factorial moments

Generalizing Corollary 3.6.1, you showed in Problem 5 that

$$\sum_{k=0}^{n} k^{\underline{r}}\binom{n}{k} = n^{\underline{r}}2^{n-r} \text{ for all } n, r \in \mathbb{N}.$$

In particular,

$$\sum_{k=0}^{n} k^{\underline{2}}\binom{n}{k} = n^{\underline{2}}2^{n-2} \text{ for all } n \in \mathbb{N}.$$

In general sums involving falling factorial powers are easier to evaluate than sums involving ordinary powers. Fortunately, the former can be used to evaluate the latter. For example, to evaluate

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

one notes that $k^2 = k^{\underline{2}} + k^{\underline{1}}$ $(k^{\underline{1}} = k)$, so

$$\begin{aligned}
s &= \sum_{k=0}^{n} (k^{\underline{2}} + k^{\underline{1}})\binom{n}{k} \\
&= \sum_{k=0}^{n} k^{\underline{2}}\binom{n}{k} + \sum_{k=0}^{n} k^{\underline{1}}\binom{n}{k} \\
&= n^{\underline{2}}2^{n-2} + n^{\underline{1}}2^{n-1} \\
&= n(n-1)2^{n-2} + n2^{n-1} \\
&= n(n+1)2^{n-2}
\end{aligned}$$

Problem 15 yields to a similar strategy.

With the above nonprobabilistic introduction to sums involving powers and falling factorial powers, we now consider the calculation of moments and factorial moments of a random variable.

If X is a random variable taking the values $x_1, x_2, \ldots, x_n$, then the *expected value* (also called the *mean* or the *first moment*) of X, $E(X)$, is defined by

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i).$$

If h is any function of X and $Y = h(X)$, then, by a basic result of probability theory,

$$E(Y) = E(h(X)) = \sum_{i=1}^{n} h(x_i)P(X = x_i).$$

In particular for all $r \in \mathbb{N}$

$$E(X^r) = \sum_{i=1}^{n} x_i^r P(X = x_i)$$

and

$$E(X^{\underline{r}}) = E(X(X-1)\cdots(X-r+1))$$
$$= \sum_{i=1}^{n} x_i^{\underline{r}} P(X = x_i).$$

The quantity $E(X^r)$ is called the rth *moment of X* , and the quantity $E(X^{\underline{r}})$ is called the rth (falling) *factorial moment of X*. The *variance of X*, $\text{var}(X)$, is defined by

$$\text{Var}(X) = E((X - E(X))^2),$$

but most easily calculated by the equivalent formula

$$\text{Var}(X) = E(X^2) - (E(X))^2.$$

The variance is a measure of how widely dispersed about its mean the values of a random variable are.

7.3 The mean and variance of binomial and hypergeometric random variables

If X is a binomial random variable with parameters n and p, then, as is well known, $E(X) = np$ and $\text{Var}(X) = np(1-p)$. The slickest proofs of these results decompose X into a sum of n "indicator random variables" and make use of the linearity of E on arbitrary sums of random variables and the linearity of Var on sums of independent random variables: For $i = 1, \ldots, n$, define X_i by

$$X_i = \begin{cases} 1 & \text{if success occurs on trial } i \\ 0 & \text{if failure occurs on trial } i \end{cases}.$$

Clearly $X = \sum_{i=1}^{n} X_i$, $E(X_i) = E(X_i^2) = p$ for all i, and so $\text{Var}(X_i) = p - p^2 = p(1-p)$. Hence $E(X) = \sum_{i=1}^{n} E(X_i) = np$ and $\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i) = np(1-p)$.

We wish to derive these results by more elementary proofs that do not depend on linearity properties of E and Var.

Theorem 7.1. *If $X \sim binomial(n, p)$, then $E(X) = np$.*

Proof.

$$E(X) = \sum_{k=0}^{n} kP(X = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \frac{n}{k} \binom{n-1}{k-1} p \cdot p^{k-1} (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}$$

$$\underset{(j=k-1)}{=} np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j}$$

$$= np(p + (1-p))^{n-1} = np.$$

$\square$

In Problem 16 you are asked to find the variance of a binomial random variable without invoking any linearity properties of the variance.

Consider now the following sampling problem. There is a finite population U and disjoint subsets M and W of U such that $M \cup W = U$, where $|M| = m$ and $|W| = w$ (so $|U| = m + w$). Select n members of U at random, letting X record the number of members of M in this sample. Depending on how the sampling is done, the random variable X will have one of two distributions.

(i) If we sample *with replacement*, then clearly X is binomially distributed with parameters n and $p = m/(m + w)$. So $E(X) = np = nm/(m + w)$ and $\text{Var}(X) = np(1 - p) = nmw/(m + w)^2$.

(ii) If we sample *without replacement*, then X is a *hypergeometric random variable* with parameters n, m, and w ($X \sim \text{hypergeom}(n; m, w)$). Clearly,

$$P(X = k) = \frac{\binom{m}{k} \binom{w}{n-k}}{\binom{m+w}{n}}, \quad k = 0, 1, \ldots, n$$

Note that by Vandermonde's identity,

$$\sum_{k=0}^{n} P(X = k) = \frac{1}{\binom{m+w}{n}} \sum_{k=0}^{n} \binom{m}{k} \binom{w}{n-k} = \frac{\binom{m+w}{n}}{\binom{m+w}{n}} = 1,$$

as one would expect on probabilistic grounds.

When the sample size n is "small" relative to the population size $m + w$, a binomial random variable with parameters n and $p = m/(m + w)$ furnishes a good approximation of the hypergeometric random variable with parameters n, m, and w. In fact, in all cases, even those where n is not small relative to $m + w$, these two random variables have the same mean.

Theorem 7.2. *If X is hypergeometric with parameters n, m, and w, then $E(X) = nm/(m+w)$.*

Proof.

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{n} k \frac{\binom{m}{k}\binom{w}{n-k}}{\binom{m+w}{n}} \\
&= \frac{1}{\binom{m+w}{n}} \sum_{k=1}^{n} k \binom{m}{k}\binom{w}{n-k} = \frac{1}{\binom{m+w}{n}} \sum_{k=1}^{n} k \cdot \frac{m}{k} \binom{m-1}{k-1}\binom{w}{n-k} \\
&= \frac{m}{\binom{m+w}{n}} \sum_{k=1}^{n} \binom{m-1}{k-1}\binom{w}{n-k} \underset{(j=k-1)}{=} \frac{m}{\binom{m+w}{n}} \sum_{j=0}^{n-1} \binom{m-1}{j}\binom{w}{(n-1)-j} \\
&= \frac{m}{\binom{m+w}{n}} \binom{m+w-1}{n-1} = \frac{m\binom{m+w-1}{n-1}}{\frac{(m+w)}{n}\binom{m+w-1}{n-1}} \\
&= nm/(m+w).
\end{aligned}
$$

$\square$

On the other hand, if $n \geq 2$, the variance of a hypergeometric random variable with parameters n, m, and w is strictly less than the variance of a binomial random variable with parameters n and $p = m/(m+w)$.

Theorem 7.3. *Under the hypotheses of Theorem 7.2,*

$$
Var(X) = \frac{nmw}{(m+w)^2}\left(\frac{m+w-n}{m+w-1}\right).
$$

Proof.

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2$$

$$= \sum_{k=0}^{n} k^2 \frac{\binom{m}{k}\binom{w}{n-k}}{\binom{m+w}{n}} - \left(\frac{nm}{m+w}\right)^2$$

$$= \sum_{k=0}^{n} [k(k-1) + k] \frac{\binom{m}{k}\binom{w}{n-k}}{\binom{m+w}{n}} - \left(\frac{nm}{m+w}\right)^2$$

$$= \sum_{k=0}^{n} k(k-1) \frac{\binom{m}{k}\binom{w}{n-k}}{\binom{m+w}{n}} + \sum_{k=0}^{n} k \frac{\binom{m}{k}\binom{w}{n-k}}{\binom{m+w}{n}} - \left(\frac{nm}{m+w}\right)^2$$

$$= \frac{1}{\binom{m+w}{n}} \sum_{k=2}^{n} k(k-1) \frac{m^{\underline{2}}}{k^{\underline{2}}} \binom{m-2}{k-2}\binom{w}{n-k} + \frac{nm}{m+w} - \left(\frac{nm}{m+w}\right)^2$$

$$= \frac{m^{\underline{2}}}{\binom{m+w}{n}} \sum_{k=2}^{n} \binom{m-2}{k-2}\binom{w}{n-k} + \frac{nm}{m+w} - \left(\frac{nm}{m+w}\right)^2$$

$$\underset{(j=k-2)}{=} \frac{m^{\underline{2}}}{\binom{m+w}{n}} \sum_{j=0}^{n-2} \binom{m-2}{j}\binom{w}{n-2-j} + \frac{nm}{m+w} - \left(\frac{nm}{m+w}\right)^2$$

$$= \frac{m^{\underline{2}}}{\binom{m+w}{n}} \binom{m+w-2}{n-2} + \frac{nm}{m+w} - \left(\frac{nm}{m+w}\right)^2$$

$$= \frac{n^{\underline{2}} m^{\underline{2}}}{(m+w)^{\underline{2}}} + \frac{nm}{m+w} - \left(\frac{nm}{m+w}\right)^2$$

$$\underset{\text{(algebra)}}{=} \cdots = \frac{nmw}{(m+w)^2} \left(\frac{m+w-n}{m+w-1}\right).$$

$\square$

Remark. As n increases, $\mathrm{Var}(X)$ decreases, with $\mathrm{Var}(X) = 0$ when $n = m + w$ (Why?). In any case, if $n \geq 2$, $\mathrm{Var}(X) < \frac{nmw}{(m+w)^2}$, the variance of a binomial random variable with parameters n and $p = m/(m+w)$.

Chapter 8

Binary Relations

8.1   Introduction

If S is any set, a *binary relation ρ on S* is a property possessed by certain ordered pairs of elements of S. If $x, y \in S$ and x stands in the relation ρ to y, one writes $x\rho y$ (if x does *not* stand in the relation ρ to y, one writes $\sim x\rho y$ or $x \not\rho y$). If S is any set of real numbers, for example, then $<, \leq, >, \geq$, and $=$ are all binary relations on S. If S is a set of geometric figures, then "is congruent to," and "is similar to" are binary relations on S.

The *graph* of a binary relation ρ on S consists of all ordered pairs $(x, y) \in S \times S$ such that $x\rho y$, i.e.,

$$\text{graph}(\rho) = \{(x, y) \in S \times S : x\rho y\}.$$

Since the graph of any binary relation on S is a subset of $S \times S$, and since any subset of $S \times S$ can be construed as the graph of some binary relation on S (how?), some authors simply define a binary relation on S to be any subset of $S \times S$. Thus, just as in the case of functions one can construe a binary relation intensionally (as a property) or extensionally (as a set of ordered pairs). In the latter case, a binary relation is identified with its graph. We shall have occasion to construe binary relations both intensionally and extensionally.

The following three relations are so basic that they have been given special names:

(i) The *empty relation on S* is the subset $\phi \subset S \times S$. The relation ρ on S is the empty relation if for all $x, y \in S$, $\sim x\rho y$, where, we recall, the latter notation indicates that x does not stand in the relation ρ to y. If, for example, S is a set of real numbers and one defines $x\rho y$ to mean that $x < y$ and $x > y$, then ρ is the empty relation.

(ii) The *identity relation, $=$, on S* is defined extensionally as the subset $\{(x, x) : x \in S\}$ of $S \times S$, and intensionally by stipulating that $x = y$ if and only if x and y are the same element of S.

(iii) The *universal relation on S* is the subset of $S \times S$ consisting of $S \times S$ itself. The relation ρ on S is the universal relation on S if and only if $x\rho y$ for all $x, y \in S$. For example, if S is a set of real numbers and one defines $x\rho y$ to mean that $x < y$ or $x \geq y$, then ρ is the universal relation on S.

Remark. If $S = \emptyset$, the empty relation, the identity relation, and the universal relation on S all coincide. If $|S| = 1$, the identity relation and the universal relation coincide, but they are distinct from the empty relation. If $|S| \geq 2$, these three relations are all distinct.

The following theorem is a straightforward corollary of Theorem 3.1.

Theorem 8.1. *For all $n \in \mathbb{N}$, if $|S| = n$, then there are $2^{(n^2)}$ binary relations on S.*

Proof. A binary relation on S may be construed as a subset of $S \times S = \{(x, y) : x, y \in S\}$. Since $|S \times S| = n^2$ by the multiplication principle, $S \times S$ has $2^{(n^2)}$ subsets by Theorem 3.1. $\square$

8.2 Representations of binary relations

We have already seen examples of binary relations given by descriptive phrases or symbols abbreviating descriptive phrases (binary relations construed intensionally). And of course one can also represent a binary relation simply by listing the set of ordered pairs which comprise it (binary relations construed extensionally).

Two other useful representations of binary relations are (i) the digraph representation and (ii) the matrix representation. These representations are typically restricted to the case where S is finite.

(i) The digraph (short for "directed graph") representation of the binary relation ρ on $S = \{x_1, \ldots, x_n\}$ consists of a set of points in the plane, arbitrarily labeled $x_1, \ldots, x_n$, with an arrow from x_i to x_j whenever $x_i \rho x_j$ and a "loop" at x_i whenever $x_i \rho x_i$. Figure 8.1 represents the digraph representations of (I) the empty relation (II) the identity relation and (III) the universal relation on $S = \{x_1, x_2, x_3\}$:

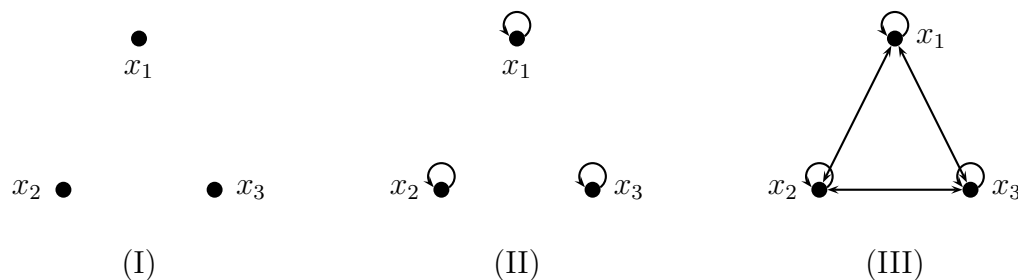
Figure 8.1: Binary Relations

(ii) The matrix representation of the binary relation ρ on $S = \{x_1, \ldots, x_n\}$ is the $n \times n$ matrix $M = (m_{ij})$ where

$$m_{ij} = \begin{cases} 1 & \text{if } x_i \rho x_j \\ 0 & \text{if } \sim x_i \rho x_j \end{cases}$$

47

Here are the matrix representations of the aforementioned three binary relations:

$$
\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
\qquad
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\qquad
\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.
$$
$$
\text{(I)} \qquad\qquad\qquad \text{(II)} \qquad\qquad\qquad \text{(III)}
$$

Special properties of binary relations (e.g., reflexivity and symmetry), to be discussed in the next section, are particularly salient in their digraph and matrix representations.

8.3   Special properties of binary relations

A binary relation ρ on S is

(1) *reflexive* if, for all $x \in S$, $x\rho x$;

(2) *irreflexive* if, for all $x \in S$, $\sim x\rho x$;

(3) *symmetric* if, for all $x, y \in S$, if $x\rho y$ then $y\rho x$;

(4) *asymmetric* if, for all $x, y \in S$, if $x\rho y$, then $\sim y\rho x$;

(5) *antisymmetric* if, for all $x, y \in S$, if $x\rho y$ and $y\rho x$, then $x = y$ (equivalently, if $x \neq y$ and $x\rho y$, then $\sim y\rho x$);

(6) *transitive* if, for all $x, y, z \in S$, if $x\rho y$ and $y\rho z$, then $x\rho z$; and

(7) *complete* if, for all $x, y \in S$, if $x \neq y$, then $x\rho y$ or $y\rho x$.

Theorem 8.2. *A binary relation ρ is asymmetric if and only if it is irreflexive and antisymmetric.*

Proof. Necessity. Suppose ρ is asymmetric. If there exists an x such that $x\rho x$, then by asymmetry, $\sim x\rho x$, a contradiction. Hence for all x, $\sim x\rho x$, i.e., ρ is irreflexive. Antisymmetry (in its parenthetical reformulation) is clearly a consequence of asymmetry.

Sufficiency. Suppose ρ is irreflexive and antisymmetric. We need to show that for all x, y, if $x\rho y$, then $\sim y\rho x$. If $x = y$, $\sim y\rho x$ asserts $\sim x\rho x$, which we know to be true for all x by irreflexivity. If $x \neq y$, then $x\rho y$ implies $\sim y\rho x$ by antisymmetry. □

Remark. The empty relation is not reflexive (unless $S = \emptyset$, in which case reflexivity holds vacuously). It is always irreflexive. The empty relation also vacuously satisfies the defining conditions of symmetry, asymmetry, antisymmetry and transitivity. It is (vacuously) complete if $S = \emptyset$ or $|S| = 1$, and not complete if $|S| \geq 2$.

Remark. Let $M = (m_{ij})$ be the matrix representation of ρ on $S = \{x_1, \ldots, x_n\}$. Then, clearly,

1) ρ is reflexive if and only if $m_{ii} = 1$, $i = 1, \ldots, n$;

2) ρ is irreflexive if and only if $m_{ii} = 0$, $i = 1, \ldots, n$; and

3) ρ is symmetric if and only if M is symmetric, i.e., $m_{ij} = m_{ji}$ for $i, j = 1, \ldots, n$.

Readers are encouraged to formulate the other four basic properties of binary relations in matrix terms, and also to formulate all seven of the basic properties in digraph terms. This will be helpful in the solution of Problems 17 and 18.

Chapter 9

Surjections and Ordered Partitions

9.1 Enumerating surjections

We saw in Chapter 2 that there are k^n functions $f : [n] \to [k]$, and that $k^{\underline{n}} = k(k-1)\cdots(k-n+1)$ of these functions are injective. In this chapter we determine the number of surjective functions $f : [n] \to [k]$. Let us denote this number by $\sigma(n,k)$. Then, of course, $\sigma(n,k)$ is also the number of

(i) surjective functions $f : A \to B$, when A and B are *any* sets with $|A| = n$ and $|B| = k$;

(ii) distributions of n labeled balls among k labeled urns with at least one ball per urn (i.e., with no urn left empty); and

(iii) words of length n in an alphabet $B = \{b_1, \ldots, b_k\}$ such that each letter of the alphabet appears at least once in the word.

From our discussion in Chapter 2 we can make a partial table of the numbers $\sigma(n,k)$:

$n\backslash k$	0	1	2	3	4	5
0	1	0	0	0	0	0
1	0	1	0	0	0	0
2	0	1	2	0	0	0
3	0	1		6	0	0
4	0	1			24	0
5	0	1				120

Table 9.1: Partial Table of $\sigma(n,k)$

Remark. Recall that:

$$\sigma(0,0) = 1$$
$$\sigma(0,k) = \sigma(n,0) = 0 \quad \text{for all } n, k \in \mathbb{P}$$
$$\sigma(n,k) = 0 \quad \text{if } n < k$$
$$\sigma(n,n) = n! \quad \text{for all } n \in \mathbb{N}$$
$$\sigma(n,1) = 1 \quad \text{for all } n \in \mathbb{P}$$

We now fill in some other values of $\sigma(n,k)$, construed as the number of distributions of n labeled balls among k labeled urns with no urn left empty. We'll use "modified brute force" to do this, i.e., we won't list and count every single such distribution, but we will partition the class of such distributions into what in some cases is a fairly large number of subclasses (in fact, $\binom{n-1}{k-1}$ subclasses) whose cardinalities we already know from Chapter 6.

The subclasses consist of all distributions with a fixed sequence $(n_1, \ldots, n_k)$ of occupancy numbers, where $n_1 + \cdots + n_k = n$ and each $n_i \in \mathbb{P}$.

Theorem 9.1. *For all $n \in \mathbb{P}$ and all $k \in \mathbb{P}$ with $k \leq n$,*

$$\sigma(n,k) = \sum_{\substack{n_1 + \cdots + n_k = n \\ n_i \in \mathbb{P}}} \binom{n}{n_1, \ldots, n_k}.$$

Proof. Each side of the above formula counts the distributions of n labeled balls among k labeled urns with no urn left empty. The LHS does this by definition of $\sigma(n,k)$. The RHS counts this class of distributions in $\binom{n-1}{k-1}$ pairwise disjoint, exhaustive subclasses, one for each of the compositions $(n_1, \ldots, n_k)$ of n, with $\binom{n}{n_1, \ldots, n_k}$ enumerating all distributions of the n balls among urns $u_1, \ldots, u_k$, where n_i balls are placed in urn u_i, $i = 1, \ldots, k$ (Theorem 6.4). It is instructive to compare Theorem 9.1 with Theorem 6.4. If we sum all k-nomial coefficients $\binom{n}{n_1, \ldots, n_k}$ of order n, we get k^n. If we sum only those for which each $n_i \in \mathbb{P}$, we get $\sigma(n,k)$. $\square$

Using Theorem 9.1, we get, for example

(1°) $$\sigma(3,2) = \binom{3}{2,1} + \binom{3}{1,2} = 6$$

(2°) $$\sigma(4,2) = \binom{4}{1,3} + \binom{4}{2,2} + \binom{4}{3,1} = 14$$

(3°) $$\sigma(5,2) = \binom{5}{1,4} + \binom{5}{2,3} + \binom{5}{3,2} + \binom{5}{4,1} = 30$$

(4°) $$\sigma(4,3) = \binom{4}{1,1,2} + \binom{4}{1,2,1} + \binom{4}{2,1,1} = 36$$

(5°) $$\sigma(5,3) = \binom{5}{1,1,3} + \binom{5}{1,2,2} + \binom{5}{1,3,1} + \binom{5}{2,1,2}$$
$$+ \binom{5}{2,2,1} + \binom{5}{3,1,1} = 150$$

(6°) $$\sigma(5,4) = \binom{5}{1,1,1,2} + \binom{5}{1,1,2,1} + \binom{5}{1,2,1,1}$$
$$+ \binom{5}{2,1,1,1} = 240$$

So a more complete table of $\sigma(n,k)$ is: This table suggests that $\sigma(n,k)$ satisfies a simple

$n\backslash k$	0	1	2	3	4	5	
0	1	0	0	0	0	0	$\cdots$
1	0	1	0	0	0	0	$\cdots$
2	0	1	2	0	0	0	$\cdots$
3	0	1	6	6	0	0	$\cdots$
4	0	1	14	36	24	0	$\cdots$
5	0	1	30	150	240	120	$\cdots$

Table 9.2: More Complete Table of $\sigma(n,k)$

recurrence relation, which is confirmed by the proof of the following theorem.

Theorem 9.2. *For all $n, k \in \mathbb{P}$,*

$$\sigma(n,k) = k(\sigma(n-1,k-1) + \sigma(n-1,k)).$$

Proof. It is equivalent to show that

$$\sigma(n,k) = k\sigma(n-1,k-1) + k\sigma(n-1,k),$$

and this is true since each side of this equation counts the class of distributions of n balls labeled, let us say, $1, \ldots, n$ among k urns labeled $u_1, \ldots, u_k$, with no urn left empty. The LHS counts this class by definition of $\sigma(n,k)$. The RHS counts this class in two disjoint exhaustive subclasses,

(i) the subclass of distributions for which ball 1 is placed in an urn all by itself,

(ii) the subclass of distributions for which ball 1 is placed in an urn containing at least one other ball.

Since there are $k\sigma(n-1, k-1)$ distributions in subclass (i) and $k\sigma(n-1, k)$ distributions in subclass (ii) [why?] the proof is complete. $\square$

9.2 Ordered partitions of a set

If A is any set and $(A_1, \ldots, A_k)$ is a sequence of subsets of A satisfying

(i) each $A_i \neq \emptyset$

(ii) $A_i \cap A_j = \emptyset$ whenever $i \neq j$

(iii) $A_1 \cup \cdots \cup A_k = A$

then $(A_1, \ldots, A_k)$ is called an *ordered partition of A with k blocks* (the subsets A_i being of course the blocks). In words, an ordered partition of A with k blocks is a *sequence of nonempty, pairwise disjoint subsets of A, with union A.*

Theorem 9.3. *If $|A| = n$ and $(n_1, \ldots, n_k)$ is a composition of n with k parts, then there are $\binom{n}{n_1, \ldots, n_k}$ ordered partitions $(A_1, \ldots, A_k)$ of A such that $|A_i| = n_i$, $i = 1, \ldots, k$. Hence there are*

$$\sum_{\substack{n_1 + \cdots + n_k = n \\ n_i \in \mathbb{P}}} \binom{n}{n_1, \ldots, n_k} = \sigma(n, k)$$

ordered partitions of A with k blocks.

Proof. Choose n_1 elements of A to comprise A_1, n_2 of the remaining $n - n_1$ elements to comprise A_2, etc. (just as in the proof of Theorem 6.4). There are thus

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3} \cdots \binom{n-n_1-n_2-\cdots-n_{k-2}}{n_{k-1}}$$
$$= \frac{n!}{n_1! n_2! \cdots n_k!} = \binom{n}{n_1, \ldots, n_k}$$

ways to construct an ordered partition $(A_1, \ldots, A_k)$ of A with the prescribed block cardinalities. The second assertion follows from Theorem 9.1. $\square$

Remark. The notion of an ordered partition $(A_1, \ldots, A_k)$ of a set A is just an abstraction of the idea of a distribution of a set of labeled balls among urns labeled $u_1, \ldots, u_k$. Instead of distributing balls, we distribute the elements of A, whatever they may be. And instead of placing these elements in actual urns, we enclose them in curly brackets, and arrange the resulting sets in a sequence.

With $\sigma(n,k)$ viewed as the number of surjections from an n-set to a k-set, it isn't obvious that the row sum

$$P_n := \sum_{k=0}^{n} \sigma(n,k)$$

is of any combinatorial interest. But now that we have an alternative interpretation of $\sigma(n,k)$ as the number of ordered partitions of an n-set with k blocks, it is clear that P_n is the *total number of ordered partitions of an n-set with all possible numbers of blocks.*

Ordered partitions of a set A, where A is a set of alternative policies or commodity bundles, are of interest in economics. The ordered partition $(A_1, \ldots, A_k)$ of A is viewed as a *preferential ordering* of the alternatives in A. All things in A_1 are tied for first place, all things in A_2 are tied for second place, etc. Economists call the blocks A_i "indifference classes," since a decisionmaker is indifferent between the alternatives comprising a given block A_i.

The numbers P_n are sometimes called the "horse-race numbers," since P_n is the number of ways that n horses can finish a race, with ties of every conceivable type possible (including an n-way tie for first, corresponding to an ordered partition of the n horses comprised of a single block.)

Here is a partial table of values of P_n:

n :	0	1	2	3	4	5
P_n :	1	1	3	13	75	541

Although it is hard to guess from this table, or even from an extended version of this table, the numbers P_n are generated by a fairly simple recursive formula.

Theorem 9.4. *If P_n is the number of ordered partitions of an n-set, then $P_0 = 1$ and, for all $n \in \mathbb{P}$,*

$$P_n = \sum_{k=1}^{n} \binom{n}{k} P_{n-k} = \sum_{j=0}^{n-1} \binom{n}{j} P_j$$

Proof. Like the LHS of the above identity, the middle sum counts the class of all ordered partitions of an n-set A, but in n pairwise disjoint, exhaustive subclasses, one for each $k = 1, \ldots, n$, where k denotes the size of the initial block A_1 of an ordered partition. For given a fixed size k, we can construct all ordered partitions of A having an initial block of size k by

(i) choosing k elements of A (in $\binom{n}{k}$ ways) to comprise the initial block A_1

(ii) following A_1 with any ordered partition of the remaining $n - k$ elements of A, which, by definition, can be done in P_{n-k} ways.

Finally

$$\sum_{k=1}^{n} \binom{n}{k} P_{n-k} = \sum_{k=1}^{n} \binom{n}{n-k} P_{n-k}$$

$$= \binom{n}{n-1} P_{n-1} + \binom{n}{n-2} P_{n-2} + \cdots + \binom{n}{1} P_1 + \binom{n}{0} P_0$$

$$= \binom{n}{0} P_0 + \binom{n}{1} P_1 + \cdots + \binom{n}{n-2} P_{n-2} + \binom{n}{n-1} P_{n-1}$$

$$= \sum_{j=0}^{n-1} \binom{n}{j} P_j.$$

$\square$

Remark. In a more advanced course in combinatorics, it is usually proved that

$$\sum_{n=0}^{\infty} P_n \frac{x^n}{n!} = \frac{1}{2 - e^x} \quad \text{for } |x| < \log 2.$$

Using this result, one can show that

$$P_n = \sum_{k=0}^{\infty} \frac{k^n}{2^{k+1}} = \frac{0^n}{2} + \frac{1^n}{4} + \frac{2^n}{8} + \frac{3^n}{16} + \cdots$$

and also that

$$P_n \sim \frac{n!}{2(\log 2)^{n+1}}.$$

Since there are $\sigma(n,k)$ surjective functions $f : A \to B$, where $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_k\}$, and also $\sigma(n,k)$ ordered partitions $(A_1, \ldots, A_k)$ of the n-set A, there exists, by Theorem 2.1, at least one bijection from the class $\mathcal{S}$ of all surjections $f : A \to B$ to the class $\mathcal{P}$ of all ordered partitions of A with k blocks. Indeed, by Theorem 2.5, there are $(\sigma(n,k))!$ such bijections. Here is a particularly natural bijection $\beta : \mathcal{S} \to \mathcal{P}$. Given $f \in \mathcal{S}$, let

$$\beta(f) = (f^{\leftarrow}(b_1), f^{\leftarrow}(b_2), \ldots, f^{\leftarrow}(b_k)),$$

where $f^{\leftarrow}(b_j)$ is the preimage of b_j under f, for $j = 1, \ldots, k$. We leave it as an exercise to show that β is a bijection (what, by the way, is β^{-1}?). We say that the ordered partition $\beta(f)$ of A is *induced by the surjection* $f : A \to B$ *relative to the labeling* $\{b_1, \ldots, b_k\}$ *of* B.

Remark. A binary relation ρ on A is called a *weak order* if it is reflexive, transitive, and complete. In more advanced courses it is proved that there is a bijection from the class of all weak orders on A to the class of all ordered partitions of A. It follows that there are P_n weak orders on an n-set A.

Chapter 10

Partitions and Equivalence Relations

10.1   Partitions of a set

Recall that an *ordered partition of a set A with k blocks* is a *sequence* $(A_1, \ldots, A_k)$ of nonempty, pairwise disjoint subsets of A, with union A. By contrast, a *partition of A with k blocks* is a *set* $\{A_1, \ldots, A_k\}$ of nonempty, pairwise disjoint subsets of A, with union A.

For all $n, k \in \mathbb{N}$, let $S(n, k)$ denote the number of partitions of $[n]$ (or any other n-set) with k blocks. The numbers $S(n, k)$ are called *Stirling numbers of the second kind* (James Stirling, 1692-1770).

There is a simple connection between $S(n, k)$ and $\sigma(n, k)$. Since every partition of $[n]$ with k blocks gives rise to $k!$ ordered partitions of $[n]$, one for each permutation of the blocks of the original partition, we have

$$\sigma(n, k) = k! S(n, k),$$

i.e.,

$$S(n, k) = \frac{\sigma(n, k)}{k!}.$$

From the table of values of $\sigma(n, k)$ we therefore derive immediately the following partial table of values of $S(n, k)$:

$n \backslash k$	0	1	2	3	4	5
0	1	0	0	0	0	0
1	0	1	0	0	0	0
2	0	1	1	0	0	0
3	0	1	3	1	0	0
4	0	1	7	6	1	0
5	0	1	15	25	10	1

Table 10.1: Stirling's Triangle

As indicated by the following theorem, the array of numbers $S(n, k)$ is generated by a simple recursive formula.

Theorem 10.1. *For all $n, k \in \mathbb{N}$, let $S(n, k)$ denote the number of partitions of $[n]$ with k blocks. Then $S(0,0) = 1$, $S(n, 0) = S(0, k) = 0$ for all $n, k \in \mathbb{P}$, and for all $n, k \in \mathbb{P}$*

$$S(n, k) = S(n - 1, k - 1) + kS(n - 1, k).$$

Proof #1. By Theorem 9.2 we have

$$\sigma(n, k) = k\sigma(n - 1, k - 1) + k\sigma(n - 1, k).$$

Dividing each side of this equation by $k!$ yields

$$\frac{\sigma(n, k)}{k!} = \frac{\sigma(n - 1, k - 1)}{(k - 1)!} + k\frac{\sigma(n - 1, k)}{k!},$$

i.e.,

$$S(n, k) = S(n - 1, k - 1) + kS(n - 1, k).$$

$\square$

Proof #2. Both sides of the equation

$$S(n, k) = S(n - 1, k - 1) + kS(n - 1, k)$$

count the class of partitions of $[n]$ with k blocks, the L.H.S. by definition. The R.H.S. counts this class in two disjoint, exhaustive subclasses,

  (i) the subclass of partitions for which $\{1\}$ is one of the blocks; and

  (ii) the subclass of partitions for which 1 belongs to a block of cardinality strictly greater than 1.

There are $S(n - 1, k - 1)$ partitions in subclass (i) and $kS(n - 1, k)$ partitions in subclass (ii) [why?], which proves the desired result. $\square$

Remark. As we noted in Chapter 9, $\sigma(n, k)$ counts the distributions of n labeled balls among k labeled urns with no urn left empty. Correspondingly, $S(n, k)$ *counts the distributions of n labeled balls among k unlabeled* (i.e., *indistinguishable*) *urns with no urn left empty.* (When urns are indistinguishable, we can only tell which balls are together in an urn, but not which urn they occupy.)

10.2 The Bell numbers

For all $n \in \mathbb{N}$, let

$$B_n := \sum_{k=0}^{n} S(n, k).$$

The numbers B_n are called *Bell numbers* (Eric Temple Bell, 1883-1960). Obviously, B_n counts the class of all partitions of $[n]$, or any other n-set. B_n also counts the class of all distributions of n labeled balls among n indistinguishable urns (This is a bit subtle. Why is this true, although

it is not true that P_n counts the class of distributions of n labeled balls among n labeled urns, the latter class having not P_n but n^n members?)

Summing the rows of Stirling's triangle we get the following partial table of Bell numbers

$$n: \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5$$

$$B_n: \quad 1 \quad 1 \quad 2 \quad 5 \quad 15 \quad 52$$

Here, for example, are the 15 partitions of the set $[4] = \{1, 2, 3, 4\}$

a. *partitions with 1 block* $[S(4, 1) = 1$ *of these*$]$

    1. $\{\{1, 2, 3, 4\}\}$

b. *partitions with 2 blocks* $[S(4, 2) = 7$ *of these*$]$

    2. $\{\{1\}, \{2, 3, 4\}\}$
    3. $\{\{1, 2\}, \{3, 4\}\}$
    4. $\{\{1, 3\}, \{2, 4\}\}$
    5. $\{\{1, 4\}, \{2, 3\}\}$
    6. $\{\{1, 2, 3\}, \{4\}\}$
    7. $\{\{1, 2, 4\}, \{3\}\}$
    8. $\{\{1, 3, 4\}, \{2\}\}$

c. *partitions with 3 blocks* $[S(4, 3) = 6$ *of these*$]$

    9. $\{\{1\}, \{2\}, \{3, 4\}\}$
    10. $\{\{1\}, \{2, 3\}, \{4\}\}$
    11. $\{\{1\}, \{2, 4\}, \{3\}\}$
    12. $\{\{1, 2\}, \{3\}, \{4\}\}$
    13. $\{\{1, 3\}, \{2\}, \{4\}\}$
    14. $\{\{1, 4\}, \{2\}, \{3\}\}$

d. *partitions with 4 blocks* $[S(4, 4) = 1$ *of these*$]$

    15. $\{\{1\}, \{2\}, \{3\}, \{4\}\}$.

Corresponding to Theorem 9.4 for $P(n)$, we have the following recursive formula for the Bell numbers.

Theorem 10.2. *For all $n \in \mathbb{N}$, let B_n denote the number of partitions of $[n]$. Then $B_0 = 1$ and for all $n \in \mathbb{P}$*

$$B_n = \sum_{k=1}^{n} \binom{n-1}{k-1} B_{n-k} = \sum_{j=0}^{n-1} \binom{n-1}{j} B_j.$$

Proof. Each side of the first equation counts the partitions of $[n]$, the L.H.S. by definition. The R.H.S. counts these partitions in n pairwise disjoint, exhaustive subclasses, one for each $k = 1, \ldots, n$, where k is the size of the block containing the element 1. For given such a k, we construct all partitions for which 1 belongs to a block of size k by

(i) choosing, in one of $\binom{n-1}{k-1}$ possible ways, $k - 1$ additional elements from the set $\{2, \ldots, n\}$ to comprise the block containing the element 1, and

(ii) partitioning the remaining $n - k$ elements in any of the B_{n-k} possible ways.

The second way of expressing the above recurrence follows from

$$
\begin{aligned}
B_n &= \sum_{k=1}^{n} \binom{n-1}{k-1} B_{n-k} = \sum_{k=1}^{n} \binom{n-1}{n-k} B_{n-k} \\
&= \binom{n-1}{n-1} B_{n-1} + \binom{n-1}{n-2} B_{n-2} + \cdots + \binom{n-1}{1} B_1 + \binom{n-1}{0} B_0 \\
&= \binom{n-1}{0} B_0 + \binom{n-1}{1} B_1 + \cdots + \binom{n-1}{n-1} B_{n-1} \\
&= \sum_{j=0}^{n-1} \binom{n-1}{j} B_j.
\end{aligned}
$$

One often sees the latter recurrence written with $n + 1$ replacing n. The complete recursive formula then specifies $B_0 = 1$ and

$$
B_{n+1} = \sum_{j=0}^{n} \binom{n}{j} B_j \quad \text{for all } n \in \mathbb{N}.
$$

$\square$

The latter recurrence is used in the proof of next theorem, which is my candidate for one of the prettiest formulas in mathematics. We first recall the following lemma from analysis:

Lemma. *If* $f(x) = \sum_{n=0}^{\infty} a_n x^n / n!$ *and* $g(x) = \sum_{n=0}^{\infty} b_n x^n / n!$ *for* $|x| < r$, *then* $f(x)g(x) = \sum_{n=0}^{\infty} c_n x^n / n!$ *for* $|x| < r$, *where* $c_n = \sum_{j=0}^{n} \binom{n}{j} a_j b_{n-j}$.

Proof. See any analysis text (e.g. Apostol, *Mathematical Analysis*; Wade, *An Introduction to Analysis*). $\square$

Theorem 10.3. *For all* $x \in \mathbb{R}$,

$$
\sum_{n=0}^{\infty} B_n \frac{x^n}{n!} = e^{e^x - 1}.
$$

Partial Proof. We shall prove this result based on the assumption that $\sum_{n=0}^{\infty} B_n x^n / n!$ converges for all $x \in \mathbb{R}$. Let

$$
\psi(x) = \sum_{n=0}^{\infty} B_n x^n / n!.
$$

By the Lemma

$$\psi(x)e^x = \left(\sum_{n=0}^{\infty} B_n \, x^n/n!\right)\left(\sum_{n=0}^{\infty} 1 \, x^n/n!\right)$$

$$= \sum_{n=0}^{\infty}\left[\sum_{j=0}^{n}\binom{n}{j}B_j\right] x^n/n!$$

$$= \sum_{n=0}^{\infty} B_{n+1} \, x^n/n!,$$

by (the last line in the proof of) Theorem 10.2.

On the other hand,

$$\psi'(x) = \sum_{n=0}^{\infty} n \, B_n \, x^{n-1}/n! = \sum_{n=1}^{\infty} B_n \, x^{n-1}/(n-1)!$$

$$= \sum_{n=0}^{\infty} B_{n+1} \, x^n/n!.$$

So

$$\psi'(x) = e^x \psi(x),$$

i.e., with $y = \psi(x)$,

$$\frac{dy}{dx} = e^x y.$$

Solving this differential equation by separation of variables, we get

$$\frac{dy}{y} = e^x dx,$$

so

$$\int \frac{dy}{y} = \int e^x dx,$$

i.e.,

$$\log y = e^x + c,$$

or

$$\psi(x) = y = e^{e^x + c}.$$

Since $\psi(0) = 1$, $e^{1+c} = 1$ and so $c = -1$, i.e.,

$$\psi(x) = \sum_{n=0}^{\infty} B_n \, x^n/n! = e^{e^x - 1}.$$

$\square$

Remark. Using Theorem 10.3, one can establish the "Dobinski formula"

$$B_n = \frac{1}{e}\sum_{k=0}^{\infty} \frac{k^n}{k!}$$

and also various asymptotic formulas which we shall not pursue here.

10.3   Equivalence relations

Recall that a binary relation on a set A is an *equivalence relation* if it is reflexive, symmetric, and transitive. Examples of equivalence relations are

(i) "is similar to" and "is congruent to" on any set of geometric figures;

(ii) "is the same height (weight, age) as" on any set of people;

(iii) "equals" on any set; and

(iv) "is congruent to (mod m)" on any set of integers, where, for $m \in \mathbb{P}$ and $x, y \in \mathbb{Z}$ we say "x is congruent (mod m) to y," symbolized $x \equiv y \pmod{m}$, if $m|x - y$, i.e., if x and y each leave the same remainder when divided by m.

Let ρ be an equivalence relation on A and let $x \in A$. Define

$$\underline{|x|}_\rho = \{y \in A: \ x\rho y\}.$$

The set $\underline{|x|}_\rho$ is called the *ρ-equivalence class determined by x*. For example, let $A = \{1, 2, 3, 6, 7, 8, 9, 47\}$ and let $x\rho y$ mean that $x \equiv y \pmod 4$. Then

$$\underline{|\,1\,|}_\rho = \{1, 9\}$$
$$\underline{|\,2\,|}_\rho = \{2, 6\}$$
$$\underline{|\,3\,|}_\rho = \{3, 7, 47\}$$
$$\underline{|\,6\,|}_\rho = \{2, 6\}$$
$$\underline{|\,7\,|}_\rho = \{3, 7, 47\}$$
$$\underline{|\,8\,|}_\rho = \{8\}$$
$$\underline{|\,9\,|}_\rho = \{1, 9\}$$
$$\underline{|\,47\,|}_\rho = \{3, 7, 47\}.$$

Note that for each $x \in A$, $\underline{|x|}_\rho \neq \emptyset$, since $x \in \underline{|x|}_\rho$. Also $\underline{|x|}_\rho = \underline{|y|}_\rho$ if and only if $x\rho y$. Finally, given any ρ-equivalence classes $\underline{|x|}_\rho$ and $\underline{|y|}_\rho$, we have either $\underline{|x|}_\rho \cap \underline{|y|}_\rho = \emptyset$ or $\underline{|x|}_\rho = \underline{|y|}_\rho$. In particular, the set of *distinct* ρ-equivalence classes $\{\{1, 9\}, \{2, 6\}, \{3, 7, 47\}, \{8\}\}$ is a partition of A. These observations hold in fact for any equivalence relation.

Theorem 10.4. *Let ρ be an equivalence relation on A. Then*

$1°$ *for all $x \in A$, $x \in \underline{|x|}_\rho$, so every ρ-equivalence class is nonempty;*

$2°$ *for all $x, y \in A$, $\underline{|x|}_\rho = \underline{|y|}_\rho$ if and only if $x\rho y$; and*

$3°$ *for all $x, y \in A$, $\underline{|x|}_\rho \cap \underline{|y|}_\rho = \emptyset$ or $\underline{|x|}_\rho = \underline{|y|}_\rho$.*

Hence the set of distinct ρ-equivalence classes is a partition of A.

Proof. 1°. By reflexivity of ρ, $x\rho x$ for all $x \in A$, and so by the definition of $\lfloor x \rfloor_\rho$, $x \in \lfloor x \rfloor_\rho$.

2°. *Sufficiency.* Suppose $x\rho y$, so that $y\rho x$ by symmetry of ρ. To show $\lfloor x \rfloor_\rho \subset \lfloor y \rfloor_\rho$, let $z \in \lfloor x \rfloor_\rho$. Then $x\rho z$, which, with $y\rho x$ and transitivity of ρ, implies $y\rho z$. Hence, $z \in \lfloor y \rfloor_\rho$. The proof that $\lfloor y \rfloor_\rho \subset \lfloor x \rfloor_\rho$ is similar.

Necessity. If $\lfloor x \rfloor_\rho = \lfloor y \rfloor_\rho$, then since $y \in \lfloor y \rfloor_\rho$ by 1°, $y \in \lfloor x \rfloor_\rho$. Hence by definition of $\lfloor x \rfloor_\rho$, $x\rho y$.

3°. It is equivalent to show that if $\lfloor x \rfloor_\rho \cap \lfloor y \rfloor_\rho \neq \emptyset$, then $\lfloor x \rfloor_\rho = \lfloor y \rfloor_\rho$. Suppose $z \in \lfloor x \rfloor_\rho \cap \lfloor y \rfloor_\rho$. Then $z \in \lfloor x \rfloor_\rho$, so $x\rho z$; and $z \in \lfloor y \rfloor_\rho$, so $y\rho z$, and hence $z\rho y$. Combining $x\rho z$ with $z\rho y$ yields $x\rho y$, whence $\lfloor x \rfloor_\rho = \lfloor y \rfloor_\rho$ by 2°.

It follows that the set of distinct ρ-equivalence classes is a partition of A, for these equivalence classes are nonempty (1°) pairwise disjoint (3°), and their union is equal to A (1°). □

The partition of A arising from ρ as above is denoted Π_ρ and called *the partition of A induced by the equivalence relation* ρ. If $\mathcal{E}_A$ denotes the class of all equivalence relations on A and $\mathcal{P}_A$ the class of all partitions of A, then the mapping $\varphi : \mathcal{E}_A \to \mathcal{P}_A$ defined for all $\rho \in \mathcal{E}_A$ by

$$\varphi(\rho) = \Pi_\rho$$

is in fact a bijection, the proof of which we leave as an exercise. The inverse function $\varphi^{-1} : \mathcal{P}_A \to \mathcal{E}_A$ is defined for every partition $\Pi \in \mathcal{P}_A$ by $\varphi^{-1}(\Pi) = \rho_\Pi$, where $x\rho_\Pi y$ if and only if x and y belong to the same block of the partition Π. The equivalence relation ρ_Π is called the *equivalence relation on A induced by the partition* Π.

Since the blocks of a partition correspond to distinct equivalence classes of the equivalence relation induced by that partition, we get yet another combinatorial interpretation of $S(n,k)$ and B_n:

I. $S(n,k)$ counts not only the partitions of $[n]$ with k blocks and the distributions of n labeled balls among k unlabeled urns with no urn empty, but also the equivalence relations on $[n]$, or any other n-set, having k distinct equivalence classes.

II. B_n counts not only all partitions of $[n]$ and all distributions of n labeled balls among n unlabeled urns, but also all equivalence relations on $[n]$, or any other n-set.

So the formula developed in Theorem 10.3,

$$\sum_{n=0}^{\infty} B_n x^n/n! = e^{(e^x-1)},$$

is even more striking than it appears at first glance. For it relates the most important type of binary relation in mathematics, the equivalence relation, with what is arguably the most important function in analysis, the exponential function.

10.4 The twelvefold way

A great deal of what we have done so far (and a few things that we haven't done) can be neatly summarized in a table that the eminent combinatorist Gian-Carlo Rota has called "the twelvefold way." The point is that elementary combinatorics, when construed in terms of counting distributions of balls among urns, can be decomposed into twelve basic problems, depending on

whether the balls and urns are labeled or unlabeled, and on whether there are restrictions on the number of balls per urn.

Table 10.2: The Twelvefold Way: Distributions of n balls among k urns, $n, k \in \mathbb{P}$

	I	II	III
Labeled balls + Labeled Urns	k^n	$k^{\underline{n}}$	$\sigma(n, k)$
Unlabeled balls + Labeled Urns	$\binom{n+k-1}{k-1}$	$\binom{k}{n}$	$\binom{n-1}{k-1}$
Labeled balls + Unlabeled Urns	$\sum_{j=1}^{k} S(n, j)$	$\begin{cases} 1 \text{ if } n \le k \\ 0 \text{ if } n > k \end{cases}$	$S(n, k)$
Unlabeled balls + Unlabeled Urns	$\sum_{j=1}^{k} p(n, j)$	$\begin{cases} 1 \text{ if } n \le k \\ 0 \text{ if } n > k \end{cases}$	$p(n, k)^*$

Column I: no restrictions on distribution
Column II: at most one ball per urn (Note: in every case, one gets 0 if $n > k$)
Column III: at least one ball per urn (Note: in every case, one gets 0 if $n < k$)

* We have not studied the numbers $p(n, k)$, which also count the "partitions of the integer n with k parts," i.e., the multisets of positive integers that have cardinality k (counting repetitions of the same number) and sum to n. These numbers are often studied in courses in number theory, as well as in more advanced combinatorics courses.

Chapter 11

Factorial Polynomials

11.1 Stirling numbers as connection constants

Recall that $x^{\underline{0}} := 1$ and $x^{\underline{k}} = x(x-1)\cdots(x-k+1)$ for $k \in \mathbb{P}$. Since for all $k \in \mathbb{N}$ $x^{\underline{k}}$ is a polynomial of degree k, the sequence of polynomials $(x^{\underline{0}}, x^{\underline{1}}, \ldots)$ is an ordered basis of the infinite dimensional vector space of polynomials over $\mathbb{R}$ (or $\mathbb{C}$), and for all $n \in \mathbb{N}$, the sequence of polynomials $(x^{\underline{0}}, x^{\underline{1}}, \ldots, x^{\underline{n}})$ is an ordered basis of the $(n+1)$-dimensional vector space of polynomials of degree $\leq n$ over $\mathbb{R}$ (or $\mathbb{C}$). In particular, for every $n \in \mathbb{N}$, there exists a unique sequence of constants $(A(n,0), A(n,1), \ldots, A(n,n))$ such that

$$x^n = \sum_{k=0}^{n} A(n,k)x^{\underline{k}} = A(n,0)x^{\underline{0}} + A(n,1)x^{\underline{1}} + \cdots + A(n,n)x^{\underline{n}}.$$

Let us make a partial table of the numbers $A(n,k)$, setting $A(n,k) = 0$ if $k > n$. We get

$n\backslash k$	0	1	2	3	4
0	1	0	0	0	0
1	0	1	0	0	0
2	0	1	1	0	0
3	0	1	3	1	0
4	0	1	7	6	1

Table 11.1: $A(n,k)$ for $0 \leq n, k \leq 4$

It is easy to determine $A(n,k)$ by inspection for $n = 0, 1, 2$. For $n \geq 3$ we find the numbers $A(n,k)$ by solving some simple simultaneous equations. Suppose, for example, that $n = 3$. We have

$$x^3 = A(3,0)x^{\underline{0}} + A(3,1)x^{\underline{1}} + A(3,2)x^{\underline{2}} + A(3,3)x^{\underline{3}}$$
$$= A(3,0) + A(3,1)x + A(3,2)x(x-1) + A(3,3)x(x-1)(x-2).$$

Substituting $x = 0, 1, 2,$ and 3 above yields the simultaneous linear equations

$$0 = A(3,0)$$
$$1 = A(3,0) + A(3,1)$$
$$8 = A(3,0) + 2A(3,1) + 2A(3,2)$$
$$27 = A(3,0) + 3A(3,1) + 6A(3,2) + 6A(3,3)$$

which have the unique solution $A(3,0) = 0$, $A(3,1) = 1$, $A(3,2) = 3$, and $A(3,3) = 1$.

Of course it is very tedious to find the values of $A(n,k)$ in this way. But note that, at least for the values in our partial table, $A(n,k) = S(n,k)$. In fact, this is true for all $n, k \in \mathbb{N}$.

Theorem 11.1. *For all $n \in \mathbb{N}$,*

$$x^n = \sum_{k=0}^{n} S(n,k) x^{\underline{k}}.$$

Proof #1. To show that $A(n,k) = S(n,k)$, note first that $A(0,0) = 1 = S(0,0)$, and, more generally, that $A(n,n) = 1 = S(n,n)$ for all $n \in \mathbb{N}$. Also $A(n,k) = 0 = S(n,k)$ if $0 \le n < k$. Moreover, $A(n,0) = 0 = S(n,0)$ for all $n \in \mathbb{P}$ as we see by setting $x = 0$ in the equation $x^n = \sum_{k=0}^{n} A(n,k) x^{\underline{k}}$. Since $A(n,k)$ and $S(n,k)$ agree on this extended set of boundary values, our proof will be complete if we can show that $A(n,k) = A(n-1,k-1) + kA(n-1,k)$ whenever $1 \le k < n$. To do this, note that for $n \in \mathbb{P}$

$$x^n = x \cdot x^{n-1} = x \sum_{j=0}^{n-1} A(n-1,j) x^{\underline{j}} = \sum_{j=0}^{n-1} A(n-1,j) x \cdot x^{\underline{j}}.$$

Since $x^{\underline{j+1}} = x(x-1)\cdots(x-j+1)(x-j) = x^{\underline{j}}(x-j) = x \cdot x^{\underline{j}} - jx^{\underline{j}}$,

$$x \cdot x^{\underline{j}} = x^{\underline{j+1}} + jx^{\underline{j}}.$$

So

$$x^n = \sum_{j=0}^{n-1} A(n-1,j)[x^{\underline{j+1}} + jx^{\underline{j}}]$$

$$= \sum_{j=0}^{n-1} A(n-1,j) x^{\underline{j+1}} + \sum_{j=0}^{n-1} jA(n-1,j) x^{\underline{j}}$$

$$= \sum_{k=1}^{n} A(n-1,k-1) x^{\underline{k}} + \sum_{k=1}^{n-1} kA(n-1,k) x^{\underline{k}}$$

$$= \sum_{k=1}^{n-1} (A(n-1,k-1) + kA(n-1,k)) x^{\underline{k}} + A(n-1,n-1) x^{\underline{n}}.$$

But by definition of the numbers $A(n,k)$, we also have

$$x^n = \sum_{k=1}^{n-1} A(n,k) x^{\underline{k}} + A(n,n) x^{\underline{n}}.$$

Comparing coefficients of $x^{\underline{k}}$ in the last two lines yields the desired result. $\qquad \square$

Proof #2. Recall that if $p(x)$ and $q(x)$ are polynomials of degree $\leq n$ and $p(r) = q(r)$ for at least $n+1$ distinct real numbers r, then $p(x) = q(x)$. Take $p(x) = x^n$ and $q(x) = \sum_{k=0}^{n} S(n,k) x^{\underline{k}}$.

We'll show that for all $r \in \mathbb{P}$, $p(r) = q(r)$, i.e.,

$$r^n = \sum_{k=0}^{n} S(n,k) r^{\underline{k}}.$$

Since $S(n,k) = \sigma(n,k)/k!$ and $r^{\underline{k}}/k! = \binom{r}{k}$, this is equivalent to

$$r^n = \sum_{k=0}^{n} \sigma(n,k) \binom{r}{k}.$$

Now each side of the above equation counts the class of functions $f : [n] \to [r]$, the LHS by the multiplication rule. The RHS counts these functions in $n+1$ disjoint, exhaustive subclasses, one for each $k = 0, \ldots, n$ where k denotes the cardinality of the range of f [Why?]. $\square$

Remark. Combinatorists have studied a number of connection constants in addition to Stirling numbers of the second kind. For example, the *Stirling numbers of the first kind* $s(n,k)$ are defined by the equations

$$x^{\underline{n}} = \sum_{k=0}^{n} s(n,k) x^k \quad \text{for all } n \in \mathbb{N},$$

the *signless Stirling numbers of the first kind* $c(n,k)$, so called because $c(n,k) = |s(n,k)|$, by

$$x^{\overline{n}} = \sum_{k=0}^{n} c(n,k) x^k \quad \text{for all } n \in \mathbb{N},$$

where $x^{\overline{0}} := 1$ and $x^{\overline{n}} := x(x+1)\cdots(x+n-1)$ if $n \in \mathbb{P}$, and the *Lah numbers* $L(n,k)$ by

$$x^{\overline{n}} = \sum_{k=0}^{n} L(n,k) x^{\underline{k}} \quad \text{for all } n \in \mathbb{N}.$$

It turns out that $L(n,k)$ counts the distributions of n balls, labeled $1, \ldots, n$, among k unlabeled urns, with no urn left empty and with the balls placed in each urn being arranged in some order. As for $c(n,k)$, it counts distributions of this type in which the left-most ball in each urn has the smallest number of all the balls in that urn.

11.2 Applications to evaluating certain sums

We have previously noted that it is often easier to evaluate sums involving falling factorial powers than to evaluate sums involving ordinary powers. Stirling numbers of the second kind enable us to convert sums of the latter type to sums of the former type and thus provide a powerful tool for evaluating otherwise difficult sums. Here are a few examples.

1° *Generalizing problem 15.* In Problem 5 you showed that

$$(*) \qquad \sum_{k=0}^{n} k^{\underline{j}}\binom{n}{k} = n^{\underline{j}}2^{n-j} \quad \text{for all } n, j \in \mathbb{N}.$$

In Problem 15 you evaluated $\sum_{k=0}^{n} k^3\binom{n}{k}$ by determining by a bit of algebra that $k^3 = k^{\underline{3}}+3k^{\underline{2}}+k^{\underline{1}}$, so that

$$\sum_{k=0}^{n} k^3\binom{n}{k} = \sum_{k=0}^{n} k^{\underline{3}}\binom{n}{k} + 3\sum_{k=0}^{n} k^{\underline{2}}\binom{n}{k} + \sum_{k=0}^{n} k^{\underline{1}}\binom{n}{k},$$

the latter three sums being easily evaluated by $(*)$. Now that we know about Stirling numbers we can dispense with the algebra, and get

$$k^3 = S(3,0)k^{\underline{0}} + S(3,1)k^{\underline{1}} + S(3,2)k^{\underline{2}} + S(3,3)k^{\underline{3}}$$
$$= k^{\underline{1}} + 3k^{\underline{2}} + k^{\underline{3}},$$

completely mechanically. Similarly, for any $r \in \mathbb{N}$, we could evaluate

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

by rewriting

$$k^r = S(r,0)k^{\underline{0}} + S(r,1)k^{\underline{1}} + S(r,2)k^{\underline{2}} + \cdots + S(r,r)k^{\underline{r}}$$

and using $(*)$ for $j = 0, \ldots, r$.

2° *Generalizing problem 16.* Recall that if $X \sim \text{binomial}(n, p)$, then for all $r \in \mathbb{N}$

$$E(X^{\underline{r}}) = \sum_{k=0}^{n} k^{\underline{r}}\binom{n}{k}p^k(1-p)^{n-k}, \quad \text{and}$$

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

In Problem 16, you used a bit of algebra to see that $k^2 = k + k^{\underline{2}}$ and thus evaluated $E(X^2)$ as follows:

$$E(X^2) = \sum_{k=0}^{n} k^2\binom{n}{k}p^k(1-p)^{n-k}$$

$$= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} + \sum_{k=0}^{n} k^{\underline{2}}\binom{n}{k}p^k(1-p)^{n-k}$$

$$= E(X) + E(X^{\underline{2}}),$$

the latter two expressions being easily evaluated.

In Problem 21 you are asked to evaluate $E(X^r)$ for arbitrary $r \in \mathbb{N}$ using the same approach, replacing k^r, as in 1° above, by $S(r,0)k^{\underline{0}} + S(r,1)k^{\underline{1}} + S(r,2)k^{\underline{2}} + \cdots + S(r,r)k^{\underline{r}}$. Using this method you will see that

$$E(X^r) = \sum_{j=0}^{r} S(r,j)E(X^{\underline{j}}).$$

As a matter of fact, *the above formula holds not just for binomial random variables but for arbitrary random variables X, given that the moments* (which may be infinite sums or improper integrals in general) *exist.*

$3°$ *Evaluating power sums.* We have previously seen the formulas

$$1 + 2 + \cdots + n = n(n+1)/2$$

and

$$1^2 + 2^2 + \cdots + n^2 = n(n + \frac{1}{2})(n+1)/3$$
$$= n(n+1)(2n+1)/6.$$

These are examples of power sums. In general for fixed $r, n \in \mathbb{N}$, the rth *power sum* $S_r(n)$ is defined by

$$S_r(n) = \sum_{k=0}^{n} k^r = 0^r + 1^r + \cdots + n^r.$$

In particular $S_0(n) = 0^0 + 1^0 + \cdots + n^0 = n + 1$ (since $0^0 := 1$ in combinatorics). Of course if r and n are both positive integers then one can also write $S_r(n) = 1^r + \cdots + n^r$, since $0^r = 0$ if $r \in \mathbb{P}$. It is not so easy to evaluate $S_r(n)$ for arbitrary r. But it is easy to evaluate the falling factorial power sums

$$S_{\underline{j}}(n) := \sum_{k=0}^{n} k^{\underline{j}} = 0^{\underline{j}} + 1^{\underline{j}} + \cdots + n^{\underline{j}}.$$

As we showed in Corollary 3.7.1, we have, for fixed but arbitrary $j, n \in \mathbb{N}$

$$S_{\underline{j}}(n) := \sum_{k=0}^{n} k^{\underline{j}} = \frac{1}{j+1} (n+1)^{\underline{j+1}}$$
$$= \frac{1}{j+1} (n+1)(n)(n-1) \cdots (n-j+1).$$

Combining Theorem 11.1 and the above result yields the following evaluation of $S_r(n)$.

Theorem 11.2. *For fixed $r, n \in \mathbb{N}$,*

$$S_r(n) := \sum_{k=0}^{n} k^r = \sum_{j=0}^{r} S(r,j) \frac{1}{j+1} (n+1)^{\underline{j+1}}$$
$$= \sum_{j=0}^{r} \sigma(r,j) \binom{n+1}{j+1}.$$

Proof. By Theorem 11.1, with $x = k$,

$$S_r(n) = \sum_{k=0}^{n} S(r,0)k^{\underline{0}} + S(r,1)k^{\underline{1}} + \cdots + S(r,r)k^{\underline{r}}$$

$$= S(r,0)\sum_{k=0}^{n} k^{\underline{0}} + S(r,1)\sum_{k=0}^{n} k^{\underline{1}} + \cdots + S(r,r)\sum_{k=0}^{n} k^{\underline{r}}$$

$$= \sum_{j=0}^{r} S(r,j)S_{\underline{j}}(n) = \sum_{j=0}^{r} S(r,j)\frac{1}{j+1}(n+1)^{\underline{j+1}}$$

$$= \sum_{j=0}^{r} \sigma(r,j)\binom{n+1}{j+1},$$

where the second-to-last equality follows from the above formula for $S_{\underline{j}}(n)$, and the last inequality from the fact that $S(r,j) = \sigma(r,j)/j!$. $\qquad\square$

For fixed $n \in \mathbb{N}$, there is also a nice recursive formula for $S_r(n)$.

Theorem 11.3. *For fixed $n \in \mathbb{N}$, $S_0(n) = n+1$ and for all $r \in \mathbb{P}$,*

$$S_r(n) = \frac{1}{r+1}\left\{(n+1)^{r+1} - \sum_{j=0}^{r-1}\binom{r+1}{j}S_j(n)\right\}.$$

Proof. For all $r, k \in \mathbb{N}$, by the binomial theorem,

$$(k+1)^{r+1} - k^{r+1} = \sum_{j=0}^{r}\binom{r+1}{j}k^j.$$

Summing each side of this equation for $k = 0, \ldots, n$ yields

$$\sum_{k=0}^{n}\left\{(k+1)^{r+1} - k^{r+1}\right\} = \sum_{k=0}^{n}\sum_{j=0}^{r}\binom{r+1}{j}k^j,$$

i.e.,

$$(n+1)^{r+1} = \sum_{j=0}^{r}\binom{r+1}{j}\sum_{k=0}^{n} k^j$$

$$= \sum_{j=0}^{r}\binom{r+1}{j}S_j(n).$$

Solving the above equation for $S_r(n)$ yields the desired recurrence. $\qquad\square$

Applying the above for $r = 1$ yields, e.g.,

$$S_1(n) := 0 + 1 + \cdots + n = \frac{1}{2}\left\{(n+1)^2 - \binom{2}{0}S_0(n)\right\}$$

$$= \frac{1}{2}\left((n+1)^2 - (n+1)\right) = \frac{n(n+1)}{2}.$$

Chapter 12

The Calculus of Finite Differences

12.1 Introduction

1. If f is a real valued function of a real variable, and x and h are real numbers, with $h \neq 0$, then
$$\Delta_h f(x) := \frac{f(x+h) - f(x)}{h}$$
is called the difference of f at x with increment h (or the difference quotient of f at x with increment h). Recall that
$$Df(x) = \lim_{h \to 0} \Delta_h f(x), \quad \text{if this limit exists.}$$

2. We shall study the special case of Δ_h corresponding to $h = 1$, and just write Δ instead of Δ_1, so that
$$\Delta f(x) := f(x+1) - f(x).$$

3. The operator Δ has several properties in common with the differentiation operator D:

   (a) $\Delta c = 0$, for every constant function c

   (b) $\Delta(cf(x)) = c\Delta f(x)$

   (c) $\Delta(f(x) + g(x)) = \Delta f(x) + \Delta g(x)$.

   Properties b. and c. express the fact that Δ is a *linear operator*. *Exercise:* Work out Δ-analogues of the product and quotient rules for D.

4. Recall that for all $n \in \mathbb{N}$, $Dx^n = nx^{n-1}$. Δx^n is not so nice:

$$\Delta x^n = (x+1)^n - x^n = \left(\sum_{k=0}^{n} \binom{n}{k} x^k \right) - x^n = \sum_{k=0}^{n-1} \binom{n}{k} x^k$$

   . Note, however, that, like D, Δ is *degree-reducing* on x^n, i.e. $\text{degree}(\Delta x^n) = n - 1$. Along with the linearity of Δ (b. & c. above), this implies that Δ *is degree-reducing on any polynomial $p(x)$*, i.e., if $\text{degree}(p(x)) = n$, then $\text{degree}(\Delta p(x)) = n - 1$.

5. (a) Δ does behave nicely on $x^{\underline{n}}$. In fact $\Delta x^{\underline{n}} = n x^{\underline{n-1}}$, as the following calculation shows:

$$\Delta x^{\underline{n}} = (x+1)^{\underline{n}} - x^{\underline{n}}$$
$$= (x+1)(x)(x-1)\cdots \underbrace{(x+1-n+1)}_{x-n+2} - x(x-1)\cdots(x-n+2)(x-n+1)$$
$$= x(x-1)\cdots(x-n+2)\underbrace{[(x+1)-(x-n+1)]}_{n}$$
$$= n x^{\underline{n-1}}.$$

In terms of analogies, we might write $\Delta : x^{\underline{n}} \ :: \ D : x^n$ (in general $A : B \ :: \ C : D$ is read "A is to B as C is to D").

(b) Let us fill in the blank in the analogy $\Delta : \underline{\quad} \ :: \ D : e^x$. We need to find a function $f(x)$ such that $\Delta f(x) = f(x)$. The solution: 2^x, for

$$\Delta 2^x = 2^{x+1} - 2^x = 2^x(2-1) = 2^x.$$

(c) A variant on formula 5a. Define the "generalized binomial coefficient" $\binom{x}{n}$ for $x \in \mathbb{R}$ and $n \in \mathbb{N}$ by:

$$\binom{x}{n} := \frac{x^{\underline{n}}}{n!} = \frac{x(x-1)\cdots(x-n+1)}{n!},$$

with $\binom{x}{0} := 1$. By 12.5a, and 12.3b,

$$\boxed{\Delta\binom{x}{n}} = \Delta \frac{x^{\underline{n}}}{n!} = \frac{1}{n!}\Delta x^{\underline{n}}$$
$$= \frac{1}{n!}\cdot n x^{\underline{n-1}} = \frac{1}{(n-1)!}x^{\underline{n-1}}$$
$$= \boxed{\binom{x}{n-1}}.$$

12.2   Higher differences

Just as we can have 2nd, 3rd, etc. order derivatives, by appropriately repeating the operation D, we can have differences of various orders. We define

$$\Delta^\circ f(x) = f(x)$$
$$\Delta f(x) = f(x+1) - f(x)$$
$$\Delta^2 f(x) = \Delta(\Delta f(x))$$
$$\Delta^3 f(x) = \Delta(\Delta^2 f(x)), \quad \text{etc.}$$

Recall that $Dx^n = nx^{n-1}$, $D^2 x^n = n(n-1)x^{n-2} = n^{\underline{2}}x^{n-2}$, and, in general,

$$D^k x^n = n^{\underline{k}} x^{n-k}.$$

In particular, $D^n x^n = n \underline{n} x^0 = n!$ and $D^k x^n = 0$ if $k > n$.

Similarly $\Delta x^{\underline{n}} = n x^{\underline{n-1}}$, $\Delta^2 x^{\underline{n}} = n^{\underline{2}} x^{\underline{n-2}}$, and, in general,

$$\Delta^k x^{\underline{n}} = n^{\underline{k}} x^{\underline{n-k}}.$$

In particular, $\Delta^n x^{\underline{n}} = n^{\underline{n}} \cdot x^{\underline{0}} = n!$ and $\Delta^k x^{\underline{n}} = 0$ if $k > n$.

12.3   A Δ-analogue of Taylor's Theorem for Polynomials

Recall that if $p(x) = \sum_{k=0}^n c_k x^k$, then by Taylor's theorem

$$c_k = \frac{D^k p(x)\big|_{x=0}}{k!} = \frac{D^k p(0)}{k!},$$

i.e., if $p(x)$ is a polynomial of degree n, then

$$p(x) = \sum_{k=0}^n \frac{D^k p(0)}{k!} x^k.$$

The Δ-analogue if this result is: If $p(x) = \sum_{k=0}^n a_k x^{\underline{k}}$, then

($\star$)
$$a_k = \frac{\Delta^k p(x)\big|_{x=0}}{k!} = \frac{\Delta^k p(0)}{k!},$$

i.e., if $p(x)$ is a polynomial of degree n, then

$$p(x) = \sum_{k=0}^n \frac{\Delta^k p(0)}{k!} x^{\underline{k}} = \sum_{k=0}^n \Delta^k p(0) \binom{x}{k}.$$

Let us prove ($\star$). To do this, we need to change the index of summation from k to j, writing $p(x) = \sum_{j=0}^n a_j x^{\underline{j}}$. Then for all $k \in \mathbb{N}$,

$$\Delta^k p(x) \underset{\text{by linearity of } \Delta^k}{=} \sum_{j=0}^n a_j \Delta^k x^{\underline{j}} \underset{\text{from 12.6 above}}{=} \sum_{j=0}^n a_j \cdot j^{\underline{k}} x^{\underline{j-k}}$$

$$= \sum_{j=k}^n a_j \cdot j^{\underline{k}} x^{\underline{j-k}} \qquad \text{(since } j^{\underline{k}} = 0 \text{ if } j < k)$$

$$= a_k \cdot k^{\underline{k}} x^{\underline{0}} + a_{k+1} \cdot (k+1)^{\underline{k}} x^{\underline{1}} + a_{k+2} \cdot (k+2)^{\underline{k}} x^{\underline{2}} + \cdots + a_n n^{\underline{k}} x^{\underline{n-k}}$$

so $\Delta^k p(0) = a_k \cdot k^{\underline{k}} = a_k \cdot k!$ and $a_k = \Delta^k p(0)/k!$, as asserted.

12.4 Polynomial Interpolation

Recall that if $(x_0, x_1, \ldots, x_n)$ is a sequence of $n + 1$ distinct reals and $(y_0, y_1, \ldots, y_n)$ is any sequence of reals, then, among the polynomials of degree $\leq n$, there is a unique polynomial $p(x)$ such that $p(x_k) = y_k$, $k = 0, \ldots, n$. This unique polynomial may be determined by "Lagrange's Interpolation Formula":

$$p(x) = \sum_{k=0}^{n} y_k \, \ell_k(x), \quad \text{where}$$

$$\ell_k(x) = \frac{\prod_{j \neq k}(x - x_j)}{\prod_{j \neq k}(x_k - x_j)}$$

Note that $\text{degree}(\ell_k(x)) = n$ for each k and $\ell_k(x_i) = \begin{cases} 0 & \text{if } k \neq i \\ 1 & \text{if } k = i \end{cases}$

For example, if $(x_0, x_1, x_2) = (-2, \frac{1}{2}, \sqrt{2})$ and $(y_0, y_1, y_2) = (\pi, e, 13)$, the unique polynomial $p(x)$ of degree ≤ 2 such that $p(-2) = \pi$, $p(\frac{1}{2}) = e$, and $p(\sqrt{2}) = 13$ is:

$$p(x) = \pi \cdot \frac{(x - \frac{1}{2})(x - \sqrt{2})}{(-2 - \frac{1}{2})(-2 - \sqrt{2})} + e \cdot \frac{(x + 2)(x - \sqrt{2})}{(\frac{1}{2} + 2)(\frac{1}{2} - \sqrt{2})} + 13 \cdot \frac{(x + 2)(x - \frac{1}{2})}{(\sqrt{2} + 2)(\sqrt{2} - \frac{1}{2})}.$$

The above method always works, but it can be tedious to use. An alternative method of "constructing-to-order" the polynomial $p(x)$ is available if $(x_0, x_1, \ldots, x_n) = (0, 1, 2, \ldots, n)$, i.e., if the values of $p(x)$ are specified on the special set of $n + 1$ reals consisting of the 1st $n + 1$ nonnegative integers. We use the formula

$$p(x) = \sum_{k=0}^{n} \Delta^k p(0) \binom{x}{k},$$

calculating $\Delta^k p(0)$ from the values $p(i) = y_i$, as in the following example:

$$
\begin{array}{ccccccccc}
& & & & p(0) & & & & \\
& & & \nearrow & & \Delta p(0) & & & \\
p(0) & = & y_0 & = & 5 & & \nearrow & & \Delta^2 p(0) \\
& & & & -7 & & \nearrow & & \Delta^3 p(0) \\
p(1) & = & y_1 & = & -2 & & 9 & & \nearrow \\
& & & & 2 & & & -1 & \\
p(2) & = & y_2 & = & 0 & & 8 & & \\
& & & & 10 & & & & \\
p(3) & = & y_3 & = & 10 & & & & \\
\end{array}
$$

$$p(x) = 5 \binom{x}{0} - 7 \binom{x}{1} + 9 \binom{x}{2} - 1 \binom{x}{3}$$

• Here is another example, which shows that we may be able to get away with using a polynomial of degree $< n$, even when $n + 1$ values of the polynomial are specified:

$$
\begin{array}{llrrrrrr}
p(0) & = & y_0 & = & 1 \\
 & & & & & -2 \\
p(1) & = & y_1 & = & -1 & & -4 \\
 & & & & & -6 & & 6 \\
p(2) & = & y_2 & = & -7 & & 2 & & 0 \\
 & & & & & -4 & & 6 & & 0 \\
p(3) & = & y_3 & = & -11 & & 8 & & 0 & & 0 \\
 & & & & & 4 & & 6 & & 0 \\
p(4) & = & y_4 & = & -7 & & 14 & & 0 \\
 & & & & & 18 & & 6 \\
p(5) & = & y_5 & = & 11 & & 20 \\
 & & & & & 38 \\
p(6) & = & y_6 & = & 49
\end{array}
$$

$$
p(x) = 1\binom{x}{0} - 2\binom{x}{1} - 4\binom{x}{2} + 6\binom{x}{3},
$$

a polynomial of degree 3 that takes specified values at 7 values of x.

• This method can also be used to guess formulas: For example, suppose we wanted to guess a formula for $S_2(n) = \sum_{k=0}^{n} k^2$. Make a partial table of values

$$
\begin{array}{llrrrrrr}
S_2(0) & = & 0 \\
 & & & 1 \\
S_2(1) & = & 1 & & 3 \\
 & & & 4 & & 2 \\
S_2(2) & = & 5 & & 5 & & 0 \\
 & & & 9 & & 2 & & 0 \\
S_2(3) & = & 14 & & 7 & & 0 & & 0 \\
 & & & 16 & & 2 & & 0 \\
S_2(4) & = & 30 & & 9 & & 0 \\
 & & & 25 & & 2 \\
S_2(5) & = & 55 & & 11 \\
 & & & 36 \\
S_6(6) & = & 91
\end{array}
$$

The presence of 3 zeros is a hopeful sign - it means that a relatively simple polynomial may be "fitted" to this "data" namely $p(x) = 0\binom{x}{0} + 1\binom{x}{1} + 3\binom{x}{2} + 2\binom{x}{3}$. We are only interested in the values $x = n$, so we guess

$$
\begin{aligned}
S_2(n) = p(n) &= \binom{n}{1} + 3\binom{n}{2} + 2\binom{n}{3} \\
&= n + \frac{3}{2}n(n-1) + \frac{1}{3}n(n-1)(n-2) \\
&= \frac{n(n+\frac{1}{2})(n+1)}{3},
\end{aligned}
$$

which may be proved by induction.

12.5 Antidifferences and Summation

If $\Delta F(x) = f(x)$ we say that F is *an antidifference of f*, and write $F = \Delta^{-1}f$ or $F(x) = \Delta^{-1}f(x)$. If $F(x)$ is an antidifference of $f(x)$, so is $F(x) + c$ for any "periodic constant" c, i.e., any function c such that $c(x + 1) = c(x)$. There are extensive tables of antidifferences available (e.g., in the excellent Schaum's Outline of Finite Differences). Here are some basic results, gotten by "reversing" earlier results on differences

	$f(x)$	$F(x) = \Delta^{-1}f(x)$
1.	0	c
2.	$x^{\underline{n}}, \ n \in \mathbb{N}$	$\dfrac{1}{n+1}x^{\underline{n+1}} + c$
3.	$\binom{x}{n}, \ n \in \mathbb{N}$	$\binom{x}{n+1} + c$
4.	2^x	$2^x + c$
5.	$x^n = \displaystyle\sum_{k=0}^{n} S(n,k)x^{\underline{k}}$	$\displaystyle\sum_{k=0}^{n} \dfrac{S(n,k)}{k+1}x^{\underline{k+1}} + c$

Table 12.1: Basic Difference Results

Note. By linearity of Δ, we also have

$$\Delta^{-1}(af(x)) = a\Delta^{-1}f(x)$$
$$\Delta^{-1}(f(x) + g(x)) = \Delta^{-1}f(x) + \Delta^{-1}g(x).$$

Just as antiderivatives are used to evaluate definite integrals, antidifferences are used to evaluate finite sums:

Theorem 12.1 (Fundamental Theorem of the Calculus of Finite Differences). *If $F = \Delta^{-1}f$, then for any integers a and b, with $a \leq b$*

$$\boxed{\sum_{k=a}^{b} f(k) = F(b+1) - F(a).}$$

Proof. Since $\Delta F = f$, for each k we have $\Delta F(k) = F(k+1) - F(k) = f(k)$, so

$$\sum_{k=a}^{b} f(k) = \sum_{k=a}^{b} F(k+1) - F(k)$$
$$= (F(a+1) - F(a)) + (F(a+2) - F(a+1)) + (F(a+3) - F(a+2))$$
$$+ \cdots + (F(b+1) - F(b)),$$

which "telescopes" to $F(b+1) - F(a)$. $\qquad\square$

Some applications of FTCFD:

(1) Summing the geometric progression $\sum_{k=0}^{n} 2^k$. Of course, a simple expression for this sum, $2^{n+1} - 1$, is well known and derivable by elementary algebra; but here is the derivation by FTCFD:

$$\sum_{k=0}^{n} 2^k = \sum_{k=0}^{n} f(k) \quad \text{where} \quad f(x) = 2^x.$$

We know $F(x) = 2^x$ is an antidifference of $f(x)$ so by FTCFD, the sum equals

$$F(n+1) - F(0) = 2^{n+1} - 2^0 = 2^{n+1} - 1.$$

(2) We have previously showed that

$$\text{(a)} \quad \sum_{k=0}^{n} \binom{k}{j} = \sum_{k=j}^{n} \binom{k}{j} = \binom{n+1}{j+1},$$

and using (a), that

$$\text{(b)} \quad S_{\underline{j}}(n) := \sum_{k=0}^{n} k^{\underline{j}} = \frac{1}{j+1}(n+1)^{\underline{j+1}}.$$

Here is a derivation of these results by FTCFD:

$$\text{(a)} \quad \sum_{k=0}^{n} \binom{k}{j} = \sum_{k=0}^{n} f(k),$$

where $f(x) = \binom{x}{j}$. We know $F(x) = \binom{x}{j+1}$ is an antidifference if $f(x)$ so the sum equals

$$F(n+1) - F(0) = \binom{n+1}{j+1} - \binom{0}{j+1} = \binom{n+1}{j+1}.$$

$$\text{(b)} \quad \sum_{k=0}^{n} k^{\underline{j}} = \sum_{k=0}^{n} f(k),$$

where $f(x) = x^{\underline{j}}$. We know $F(x) = \frac{1}{j+1}x^{\underline{j+1}}$ is an antidifference of $f(x)$, so the sum equals

$$F(n+1) - F(0) = \frac{1}{j+1}(n+1)^{\underline{j+1}} - \underbrace{\frac{1}{j+1}0^{\underline{j+1}}}_{\substack{=0 \text{ as } j+1 \geq 1 \\ (\text{since } j \in \mathbb{N})}}$$

$$= \frac{1}{j+1}(n+1)^{\underline{j+1}}.$$

(3) Finally, we can get a formula for the power sum $S_r(n) = \sum_{k=0}^{n} k^r$, $r \in \mathbb{P}$:

$$\sum_{k=0}^{n} k^r = \sum_{k=0}^{n} f(k), \quad \text{where } f(x) = x^r.$$

By our earlier table of antidifferences $F(x) = \sum_{k=0}^{r} \frac{S(r,k)}{k+1} x^{\underline{k+1}}$ is an antidifference of $f(x)$, so

$$\boxed{\sum_{k=0}^{n} k^r} = F(n+1) - F(0) = \sum_{k=0}^{r} \frac{S(r,k)}{k+1}(n+1)^{\underline{k+1}} - \underbrace{\sum_{k=0}^{r} \frac{S(r,k)}{k+1} 0^{\underline{k+1}}}_{=0 \text{ since } k+1 \geq 1}$$

$$= \boxed{\sum_{k=0}^{r} \frac{S(r,k)}{k+1}(n+1)^{\underline{k+1}}}.$$

12.6 A formula for $\Delta^k f(x)$

We have defined higher differences simply by iteration of Δ, but there is actually a nice formula for $\Delta^k f(x)$ that shows directly how this quantity depends on the functional values $f(x)$, $f(x+1), \ldots, f(x+k)$. One can discover this formula simply by brute force calculation: we have

$$\Delta^1 f(x) = f(x+1) - f(x)$$
$$\Delta^2 f(x) = \Delta\left(f(x+1) - f(x)\right)$$
$$= [f(x+2) - f(x+1)] - [f(x+1) - f(x)]$$
$$= f(x+2) - 2f(x+1) + f(x)$$
$$\Delta^3 f(x) = \Delta\left(f(x+2) - 2f(x+1) + f(x)\right)$$
$$= f(x+3) - 3f(x+2) + 3f(x+1) - f(x)$$

which leads to the conjecture

$$\boxed{\Delta^k f(x) = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(x+j)},$$

which may be proved by induction, or by the following slick "operator proof":
Define operators E and I by $Ef(x) = f(x+1)$ and $If(x) = f(x)$ so $E^j f(x) = f(x+j)$ and $I^j f(x) = f(x)$. Note that $\Delta = E - I$. So

$$\Delta^k = (E - I)^k = \sum_{j=0}^{k} \binom{k}{j} E^j (-I)^{k-j}$$

$$= \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} E^j I^{k-j}$$

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

Thus,

$$\Delta^k f(x) = \left(\sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} E^j \right) f(x)$$

$$= \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} E^j f(x)$$

$$= \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(x+j).$$

12.7 A formula for $\sigma(n, k)$

Recall the identity

$$x^n = \sum_{k=0}^{n} \sigma(n, k) \binom{x}{k},$$

which we established combinatorially by showing that it was true for $x = r$, $r \in \mathbb{P}$. By our Δ-analogue of Taylor's Theorem (12.7 above) we must have

$$\boxed{\sigma(n, k)} = \Delta^k x^n \big|_{x=0}$$

$$\underset{\text{(by formula in 12.6 above)}}{=} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} (x+j)^n \bigg|_{x=0}$$

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

We will prove this formula in another way later in the course.

Chapter 13

Deriving Closed Form Solutions From Recursive Formulas

13.1   Introduction: First Order Linear Recurrences

In this chapter we wish to develop methods for deriving a closed form expression for the nth term of a sequence $(f(0), f(1), \ldots)$ defined by a recursive formula. The type of recursive formula that we consider will be of "order" d for some fixed positive integer d, which means that

    $1°$ there are d fixed "initial values" $v_0, v_1, \ldots, v_{d-1}$ of f, i.e., we are given that $f(0) = v_0$, $f(1) = v_1$, …, and $f(d-1) = v_{d-1}$,

and

    $2°$ there is a function $\phi : \mathbb{R}^d \to \mathbb{R}$ such that

$$f(n) = \phi(f(n-1), f(n-2), \ldots, f(n-d))$$

    for all $n \geq d$.

Example. The Fibonacci sequence is given by a second order recursive formula with $v_0 = v_1 = 1$ and $\phi(x_1, x_2) = x_1 + x_2$.

Remark 1. Many interesting recursive formulas do not have a fixed finite order. For example, the Bell numbers B_n of Chapter 10 and the "horse race" numbers P_n of Chapter 9 are generated by recurrences involving *all* previous terms in the sequence, rather than a fixed number d of previous terms.

Remark 2. One can write the recurrence $2°$ above equivalently as

$(2a°)$ $f(n+d) = \phi(f(n+d-1), f(n+d-2), \ldots, f(n))$

for all $n \geq 0$. As we shall see, writing a recurrence in form $(2°)$, with n the largest index, is preferable when using generating function methods to solve for $f(n)$, as in §13.2. On the other hand, writing a recurrence in form $(2a°)$, with n the smallest index, is preferable when using characteristic polynomials to solve for $f(n)$, as in §13.4.

    The simplest recursive formulas, the *first order, linear recursive formulas*, are trivial to solve.

Theorem 13.1. *Let v_0, m, and b be fixed constants. There is a unique sequence f satisfying $f(0) = v_0$ and*

$$f(n+1) = mf(n) + b \quad \text{for all } n \geq 0,$$

which is given by the closed form expression

$$f(n) = m^n v_0 + (1 + m + \cdots + m^{n-1})b$$

for all $n \geq 0$. If $m \neq 1$, this may be written equivalently as

$$f(n) = m^n v_0 + \left(\frac{m^n - 1}{m - 1}\right)b.$$

Proof. Existence. Simply verify that the sequence f given by the formula $f(n) = m^n v_0 + (1 + \cdots + m^{n-1})b$ satisfies $f(0) = v_0$ and $f(n+1) = mf(n) + b$ for all $n \geq 0$.

Uniqueness. Show by induction that if f and g are sequences satisfying $f(0) = g(0) = v_0$ and $f(n+1) = mf(n) + b$ and $g(n+1) = mg(n) + b$ for all $n \geq 0$, then $f(n) = g(n)$ for all $n \geq 0$. $\square$

Remark 3. Recursive formulas for sequences are sometimes called "discrete dynamical systems." I regard this as a pompous term for a simple notion. It is used to associate the latter with the trendy glamour of chaos and fractals. (Similarly, the homely differential equation is now termed a "dynamical system.")

Remark 4. It is appropriate (and not pompous) to call recursive formulas "difference equations." Here's why: Rewrite the recurrence

$$f(n+1) = mf(n) + b$$

by subtracting $f(n)$ from each side, getting

$$f(n+1) - f(n) = (m-1)f(n) + b,$$

i.e.,

$$\Delta f(n) = (m-1)f(n) + b,$$

which explains the terminology "difference equation," in the first order linear case. More generally one can rewrite the first order recurrence

$$f(n+1) = \phi(f(n))$$

as

$$\Delta f(n) = \psi(f(n))$$

where $\psi(x) = \phi(x) - x$.

Higher order recurrences can similarly be reformulated as difference equations involving $\Delta^1, \Delta^2, \ldots, \Delta^d$.

13.2   Application of Theorem 13.1 to financial mathematics

Let

$$v_0 = P, \text{ the "principle" or initial amount invested,}$$
$$f(n) = S(n), \text{ the amount accumulated at time } t = n,$$
$$m = (1 + r), \ r \text{ the rate of interest per unit time period (usually } 0 < r < 1),$$
$$b = p, \text{ the "payment" made at } t = 1, 2, \dots \text{ (can be negative).}$$

Theorem 13.1 asserts that the unique S satisfying $S(0) = P$ and $S(n + 1) = (1 + r)S(n) + p$ for all $n \geq 0$ is given by the formula

$$S = S(n) = (1 + r)^n P + \left(\frac{(1 + r)^n - 1}{r} \right) p \text{ for all } n \geq 0.$$

This formula has a large number of applications. In addition to the obvious application of calculating S, given r, n, P, and p, one can solve for any of these variables in terms of the others. In a number of such cases the resulting formulas have been given special names. For example, if r and n are given and $P = 0$, and we solve for p in terms of S, we get the "Uniform payment sinking fund formula"

$$p = \left[\frac{r}{(1 + r)^n - 1} \right] S,$$

which tells us what our periodic payments p at $t = 1, \dots, n$ must be under interest rate r in order to accumulate S at $t = n$.

A table of the most basic formulas, all of which involve one of the variables S, P, or p taking the value zero, appears below.

Name of Formula	Zero Variable	Given Variable	Unknown Variable	Formula
Single payment \ compound amount:	p	P	S	$S = [(1+r)^n]P$
Single payment \ present value:	p	S	P	$P = \left[\frac{1}{(1+r)^n}\right]S$
Uniform payment \ compound amount:	P	p	S	$S = \left[\frac{(1+r)^n-1}{r}\right]p$
Uniform payment \ sinking fund:	P	S	p	$p = \left[\frac{r}{(1+r)^n-1}\right]S$
Uniform payment \ capital recovery:	S	P	p	$p = -\left[\frac{r(1+r)^n}{(1+r)^n-1}\right]P$
Uniform payment \ present value:	S	p	P	$P = -\left[\frac{(1+r)^n-1}{r(1+r)^n}\right]p$

Table 13.1: Basic Formulae of Financial Mathematics

13.3 The Generating Function Method

We have already seen an application of this method in Chapter 1, where we derived the formula

$$f(n) = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}$$

for the Fibonacci sequence f, defined recursively by $f(0) = f(1) = 1$ and $f(n) = f(n-1)+f(n-2)$ for all $n \geq 2$.

The generating function method works on a wide variety of recursive formulas, though it sometimes requires some cleverness to see just how. It can, however, be applied in perfectly mechanical fashion to "d^{th} order, linear, homogeneous recursive formulas with constant coefficients," i.e., those satisfying

1° $f(0) = v_0, f(1) = v_1, \ldots, f(d-1) = v_{d-1}$

and

2° $f(n) = c_1 f(n-1) + c_2 f(n-2) + \cdots + c_d f(n-d)$ for all $n \geq d$,

where $v_0, v_1, \ldots, v_{d-1}$ and $c_1, c_2, \ldots, c_d$ are fixed constants. Let us illustrate the method on another second order recursion.

Problem. Find a closed-form expression for $f(n)$ if $f(1) = -11/4$, $f(2) = 133/16$ and $f(n) = -\frac{11}{4}f(n-1) + \frac{3}{4}f(n-2)$ for all $n \geq 3$.

Solution: It is advisable to "extend" f to $\mathbb{N}$, defining $f(0)$ in such a way that the recurrence

$$f(n) = -\frac{11}{4}f(n-1) + \frac{3}{4}f(n-2)$$

holds for $n = 2$. To see what $f(0)$ must be, plug in $n = 2$ above and solve for $f(0)$:

$$f(2) = -\frac{11}{4}f(1) + \frac{3}{4}f(0)$$

$$f(0) = \frac{4}{3}f(2) + \frac{11}{3}f(1) = \frac{4}{3} \cdot \frac{133}{16} + \frac{11}{3}\left(\frac{-11}{4}\right) = 1.$$

So we may rephrase the problem as follows: Find a closed-form expression for $f(n)$ if $f(0) = 1$, $f(1) = \frac{-11}{4}$ and, for all $n \geq 2$, $f(n) = \frac{-11}{4}f(n-1) + \frac{3}{4}f(n-2)$, i.e., $f(n) + \frac{11}{4}f(n-1) - \frac{3}{4}f(n-2) = 0$. Let

(i) $$F(x) = f(0) + f(1)x + f(2)x^2 + \cdots + f(n)x^n + \cdots , \quad \text{so}$$

(ii) $$\frac{11}{4}xF(x) = \frac{11}{4}f(0)x + \frac{11}{4}f(1)x^2 + \cdots + \frac{11}{4}f(n-1)x^n + \cdots , \quad \text{and}$$

(iii) $$-\frac{3}{4}x^2F(x) = -\frac{3}{4}f(0)x^2 + \cdots + -\frac{3}{4}f(n-2)x^n + \cdots .$$

Adding (i), (ii), and (iii), we get

$$(1 + \frac{11}{4}x - \frac{3}{4}x^2)F(x) = f(0) + \left[f(1) + \frac{11}{4}f(0)\right]x = 1.$$

Hence

$$F(x) = \frac{1}{1 + \frac{11}{4}x - \frac{3}{4}x^2} = \frac{1}{(1 + 3x)(1 - \frac{1}{4}x)}$$

$$= \frac{A}{1 + 3x} + \frac{B}{1 - \frac{1}{4}x}$$

$$= A\sum_{n=0}^{\infty}(-3x)^n + B\sum_{n=0}^{\infty}\left(\frac{1}{4}x\right)^n$$

(recall: $\sum_{n=0}^{\infty} u^n = \frac{1}{1-u}$ if $|u| < 1$)

$$= \sum_{n=0}^{\infty}\left[A(-3)^n + B\left(\frac{1}{4}\right)^n\right]x^n$$

$$= \sum_{n=0}^{\infty} f(n)x^n.$$

Comparing coefficients of x^n yields

$$f(n) = A(-3)^n + B(\frac{1}{4})^n.$$

To determine A and B just do the algebra for the partial fraction decomposition

$$\frac{1}{(1+3x)(1-\frac{1}{4}x)} = \frac{A}{1+3x} + \frac{B}{1-\frac{1}{4}x}.$$

Multiply each side by $(1+3x)(1-\frac{1}{4}x)$

$$1 = A(1 - \frac{1}{4}x) + B(1 + 3x).$$

Let $x = 4$, getting $B = \frac{1}{13}$ and let $x = -\frac{1}{3}$, getting $A = \frac{12}{13}$, so

$$f(n) = \left(\frac{12}{13}\right)(-3)^n + \left(\frac{1}{13}\right)\left(\frac{1}{4}\right)^n.$$

One can of course check to see that this formula yields the correct values for $n = 0$ and $n = 1$ and that it satisfies the recurrence $f(n) = -\frac{11}{4}f(n-1) + \frac{3}{4}f(n-2)$.

13.4   Combinatorial Factorizations

As part of the generating function solution to the problem of finding a formula for the nth Fibonacci number, we needed to find the factorization

$$1 - x - x^2 = \left(1 - \left(\frac{1+\sqrt{5}}{2}\right)x\right)\left(1 - \left(\frac{1-\sqrt{5}}{2}\right)x\right).$$

In the problem treated in the above section we needed to find the factorization

$$1 + \frac{11}{4}x - \frac{3}{4}x^2 = (1 + 3x)(1 - \frac{1}{4}x).$$

More generally, in combinatorial applications we wish to factor a polynomial

(1) $$p(x) = c_n x^n + \cdots + c_0 = c_0 + c_1 x + \cdots + c_n x^n,$$

where $c_n \neq 0$ and $c_0 \neq 0$, in the form

(2) $$p(x) = c_0(1 - R_1 x)(1 - R_2 x)\cdots(1 - R_n x)$$

rather than in the form desired in algebra, namely,

(3) $$p(x) = c_n(x - r_1)(x - r_2)\cdots(x - r_n).$$

To get the "combinatorial factorization" (2) of $p(x)$ from its "algebraic factorization" (3), note that the constant term of $p(x)$ is c_0 by (1) and $c_n(-r_1)(-r_2)\cdots(-r_n)$ by (3). Hence $c_0 = c_n(-r_1)(-r_2)\cdots(-r_n)$, i.e.,

(4) $$c_n = c_0\left(-\frac{1}{r_1}\right)\left(-\frac{1}{r_2}\right)\cdots\left(-\frac{1}{r_n}\right).$$

Substituting this expression for c_n in (3) yields

$$p(x) = c_0 \left(-\frac{1}{r_1} \right) \cdots \left(-\frac{1}{r_n} \right) (x - r_1) \cdots (x - r_n)$$

$$= c_0 \left(-\frac{1}{r_1}(x - r_1) \right) \cdots \left(-\frac{1}{r_n}(x - r_n) \right)$$

$$= c_0 \left(1 - \frac{1}{r_1}x \right) \cdots \left(1 - \frac{1}{r_n}x \right).$$

So to get a combinatorial factorization

$$p(x) = c_0(1 - R_1 x) \cdots (1 - R_n x)$$

of $p(x)$, simply find the roots $r_1, \ldots, r_n$ of $p(x)$ and set $R_i = \frac{1}{r_i}$, $i = 1, \ldots, n$.

In fact, there is a shortcut to finding the numbers $R_1, \ldots, R_n$. For the reciprocals of the roots of $p(x) = c_0 + c_1 x + \cdots + c_n x^n$ are just the roots of the "reciprocal polynomial" of $p(x)$,

$$\hat{p}(x) := c_0 x^n + c_1 x^{n-1} + \cdots + c_n, \quad \text{for}$$

since $\hat{p}(x) = x^n p \left(\frac{1}{x} \right)$,

$$\hat{p}\left(\frac{1}{r_i} \right) = r_i^{-n} p(r_i) = 0 \quad \text{if } p(r_i) = 0.$$

So the most efficient way to find the combinatorial factorization of $p(x) = c_0 + c_1 x + \cdots + c_n x$ proceeds as follows:

1° Find the roots $R_1, \ldots, R_n$ of $\hat{p}(x) := c_0 x^n + c_1 x^{n-1} + \cdots + c_n$

2° Write
$$p(x) = c_0(1 - R_1 x) \cdots (1 - R_n x).$$

Example. $p(x) = 1 - x - x^2$, so $\hat{p}(x) = x^2 - x - 1$. The roots of $\hat{p}(x)$, by the quadratic formula, are $R_1 = \frac{1 + \sqrt{5}}{2}$ and $R_2 = \frac{1 - \sqrt{5}}{2}$. So

$$1 - x - x^2 = \left(1 - \left(\frac{1 + \sqrt{5}}{2} \right) x \right) \left(1 - \left(\frac{1 - \sqrt{5}}{2} \right) x \right).$$

13.5 The Characteristic Polynomial Method

The following is a brief exposition of the characteristic polynomial method for solving for $f(n)$, where f satisfies the initial conditions

(7) $$f(0) = v_0, \quad f(1) = v_1, \quad \ldots, \quad f(d-1) = v_{d-1}$$

and the (implicit) recurrence relation

(8) $$a_d f(n+d) + a_{d-1} f(n+d-1) + \cdots + a_0 f(n) = 0$$

for all $n \geq 0$.

The *characteristic polynomial* $q(x)$ of (8) is defined by

$$(9) \qquad q(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0.$$

The technique for solving (8), subject to (7), is similar to a technique used in differential equations. One finds the "general solution" of (8), which involves arbitrary parameters $\lambda_1, \cdots, \lambda_d$, and then determines the values of these parameters from (7), thereby finding the "particular solution" of (8), subject to (7).

The key fact underlying this method is given by the following theorem.

Theorem 13.2. *If $q(r) = 0$, where $q(x)$ is given by (9), then the geometric progression $f(n) = r^n$ satisfies the implicit recurrence (8).*

Proof. We must show that

$$(11) \qquad a_d r^{n+d} + a_{d-1} r^{n+d-1} + \cdots + a_0 r^n = 0$$

for all $n \geq 0$. But if $q(r) = 0$, then by (9)

$$(12) \qquad a_d r^d + a_{d-1} r^{d-1} + \cdots + a_0 = 0.$$

Multiplying (12) by r^n yields (11). $\qquad \square$

Combined with some linear algebra that we shall not pursue here, the above theorem yields the following procedure for finding the general solution of the implicit recurrence relation (8):

I. If the roots $r_1, r_2, \ldots, r_d$ of $q(x)$ are distinct, the general solution of (8) is

$$(13) \qquad f(n) = \lambda_1 r_1^n + \lambda_2 r_2^n + \cdots + \lambda_d r_d^n,$$

where $\lambda_1, \ldots, \lambda_d$ are arbitrary complex numbers.

II. If there are repeated roots, say m_1 roots equal to r_1, m_2 roots equal to r_2, ..., and m_s roots equal to r_s, where $m_1 + m_2 + \cdots + m_s = d$, with $r_1, \ldots, r_s$ distinct, the general solution of (8) is

$$
\begin{aligned}
f(n) = {} & \lambda_1 r_1^n + \lambda_2 n r_1^n + \cdots + \lambda_{m_1} n^{m_1 - 1} r_1^n \\
& + \lambda_{m_1+1} r_2^n + \lambda_{m_1+2} n r_2^n + \cdots + \lambda_{m_1+m_2} n^{m_2 - 1} r_2^n \\
& + \cdots \\
(14) \qquad & + \lambda_{m_1+\cdots+m_{s-1}+1} r_s^n + \lambda_{m_1+\cdots+m_{s-1}+2} n r_s^n + \cdots + \lambda_{m_1+\cdots+m_s} n^{m_s - 1} r_s^n,
\end{aligned}
$$

where $\lambda_1, \lambda_2, \ldots, \lambda_{m_1+\cdots+m_s}$ are any complex numbers.

To find the particular solution of (8), subject to (7) plug in the values $n = 0, 1, \ldots, d-1$ in (13) or (14), as appropriate, and set the resulting expressions equal, respectively, to $v_0, v_1, \ldots, v_{d-1}$. This will give you d simultaneous linear equations in the unknowns $\lambda_1, \ldots, \lambda_d$, and it turns out that these equations always have a unique solution.

Example I. Rabbits redux. Find $f(n)$, where $f(0) = f(1) = 1$ and $f(n+2) = f(n+1) + f(n)$ for all $n \geq 0$. Rewrite this as an implicit recurrence

$$(15) \qquad f(n+2) - f(n+1) - f(n) = 0 \quad \forall n \geq 0.$$

The characteristic polynomial $q(x)$ of (15) is given by

(16) $$q(x) = x^2 - x - 1,$$

and its roots are

(17) $$r_1 = \frac{1 + \sqrt{5}}{2} \quad \text{and} \quad r_2 = \frac{1 - \sqrt{5}}{2}.$$

So the general solution to (15) is

(18) $$f(n) = \lambda_1 \left(\frac{1 + \sqrt{5}}{2} \right)^n + \lambda_2 \left(\frac{1 - \sqrt{5}}{2} \right)^n.$$

Plugging in $n = 0$ and $n = 1$ in (18) yields

$$\lambda_1 + \lambda_2 = 1 \quad (= f(0))$$

(19) $$\left(\frac{1 + \sqrt{5}}{2} \right) \lambda_1 + \left(\frac{1 - \sqrt{5}}{2} \right) \lambda_2 = 1 \quad (= f(1)).$$

Solving system (19) yields

(20) $$\lambda_1 = \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right) \qquad \lambda_2 = \frac{-1}{\sqrt{5}} \left(\frac{1 - \sqrt{5}}{2} \right).$$

Substituting these values for λ_1 and λ_2 in (18) yields the formula

(21) $$f(n) = \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^{n+1} - \frac{1}{\sqrt{5}} \left(\frac{1 - \sqrt{5}}{2} \right)^{n+1}.$$

Example II. Find $f(n)$, if $f(0) = 8$, $f(1) = 35$, $f(2) = 171$, and $f(n) = 9f(n-1) - 24f(n-2) + 20f(n-3)$, $\forall n \geq 3$.

In the above recurrence replace n by $n+3$ so that n is the smallest argument, and then write the recurrence implicitly:

(22) $$f(n+3) - 9f(n+2) + 24f(n+1) - 20f(n) = 0$$

for all $n \geq 0$.

The characteristic polynomial of (22) is

(23) $$\begin{aligned} q(x) &= x^3 - 9x^2 + 24x - 20 \\ &= (x-2)^2(x-5), \end{aligned}$$

so the general solution to (22) is

(24) $$f(n) = \lambda_1 2^n + \lambda_2 n 2^n + \lambda_3 5^n.$$

Further,

$$f(0) = 8 = \lambda_1 + \qquad \lambda_3$$
$$f(1) = 35 = 2\lambda_1 + 2\lambda_2 + 5\lambda_3$$
(25)
$$f(2) = 171 = 4\lambda_1 + 8\lambda_2 + 25\lambda_3,$$

which has the unique solution $\lambda_1 = 1$, $\lambda_2 = -1$, $\lambda_3 = 7$.

Hence,

$$f(n) = 2^n - n2^n + 7 \cdot 5^n.$$

Clearly, if a recurrence can be put in the form (8), it is easier to determine $f(n)$ by the characteristic polynomial method than by the generating function method. As we see in the next section, however, the generating function method can be useful outside the realm of linear recurrences.

13.6 A Nonlinear Recurrence

For all $n \geq 1$, let $f(n)$ be the number of ways to "parenthesize" the sum $x_1 + \cdots + x_n$. A partial table of $f(n)$ follows:

n	$f(n)$	Parenthesizations of $x_1 + \cdots + x_n$
1	1	x_1
2	1	$x + x_2$
3	2	$x_1 + (x_2 + x_3)$, $(x_1 + x_2) + x_3$
4	5	$x_1 + (x_2 + (x_3 + x_4))$, $x_1 + ((x_2 + x_3) + x_4)$, $(x_1 + x_2) + (x_3 + x_4)$, $(x_1 + (x_2 + x_3)) + x_4$, $((x_1 + x_2) + x_3) + x_4$

Table 13.2: Parenthesizations of $x_1 + \cdots + x_n$

A parenthesization of $x_1 + \cdots + x_n$ provides a way of summing $x_1 + \cdots + x_n$. The standard left-to-right sum corresponds to the parenthesization $(\cdots(((x_1 + x_2) + x_3) + x_4) + \cdots) + x_n$.

Theorem 13.3. *If $f(n)$ denotes the number of parenthesizations of $x_1 + \cdots + x_n$, then for all $n \geq 1$*

$$f(n) = \frac{1}{n}\binom{2n-2}{n-1}.$$

Proof. From the above table, $f(1) = 1$. We claim that if $n \geq 2$, then

$$f(n) = f(1)f(n-1) + f(2)f(n-2) + \cdots + f(n-1)f(1)$$
(26)
$$= \sum_{k=1}^{n-1} f(k)f(n-k).$$

88

Here k is the number of x_i's to the left of last-performed addition. Set

$$(27) \qquad F(x) = f(1)x + f(2)x^2 + \cdots = \sum_{n=1}^{\infty} f(n)x^n.$$

Then

$$
\begin{aligned}
F^2(x) &= [f(1) \cdot f(1)]x^2 + [f(1)f(2) + f(2)f(1)]x^3 \\
&\quad + \cdots [f(1)f(n-1) + \cdots + f(n-1)f(1)]x^n + \cdots \\
&= f(2)x^2 + f(3)x^3 + \cdots + f(n)x^n + \cdots \\
&= F(x) - f(1)x = F(x) - x,
\end{aligned}
$$

by (26), i.e.,

$$(28) \qquad F^2(x) - F(x) + x = 0.$$

Formula (28) is an example of a *functional equation*, i.e., an equation for which the unknown is a function. It can be solved by an extension of the quadratic formula, with $F(x)$ regarded as unknown. We get

$$(29) \qquad F(x) = \frac{1 \pm \sqrt{1 - 4x}}{2}$$

Since by (27), $F(0) = 0$, we must have

$$(30) \qquad F(x) = \frac{1 - \sqrt{1 - 4x}}{2}.$$

By Newton's generalization of the binomial theorem,

$$(31) \qquad (1 + t)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} t^n, \quad \alpha \in \mathbb{R}, \quad |t| < 1.$$

So

$$
\begin{aligned}
(32) \qquad F(x) &= \frac{1}{2} - \frac{1}{2}(1 - 4x)^{\frac{1}{2}} \\
&= \frac{1}{2} - \frac{1}{2}\sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n}(-4x)^n \\
&= \frac{1}{2} - \frac{1}{2} + \sum_{n=1}^{\infty} \left(-\frac{1}{2}\right)\binom{\frac{1}{2}}{n}(-4)^n x^n
\end{aligned}
$$

Comparing coefficients of x^n in (27) and (32) yields

$$
\begin{aligned}
f(n) &= \left(-\frac{1}{2}\right)\binom{\frac{1}{2}}{n}(-4)^n \\
&= \left(-\frac{1}{2}\right)\frac{\left(\frac{1}{2}\right)\left(\frac{1}{2} - 1\right) \cdots \left(\frac{1}{2} - n + 1\right)}{n!}(-4)^n \\
&= \cdots = \frac{1}{n}\binom{2n - 2}{n - 1}.
\end{aligned}
$$

$\square$

The numbers $f(n)$ above are closely related to the *Catalan numbers* $c(n)$, defined for all $n \geq 0$ by

$$c(n) = \frac{1}{n+1}\binom{2n}{n}.$$

Specifically, $f(n) = c(n-1)$ for all $n \geq 1$. There are around 40 different combinatorial problems whose solutions involve Catalan numbers. Here are a few examples.

A Recall that there are $\binom{2n}{n}$ lattice paths from $(0,0)$ to (n,n) for all $n \geq 0$. It turns out that the number of such paths that never rise above the diagonal line from $(0,0)$ to (n,n) is given by $c(n)$.

Exercise. There are $2n$ people who have lined up at random at a movie ticket window. Of these people n have \$0.50 and n have a dollar bill. Admission costs \$0.50 and the cashier has no change. What is the probability that the cashier will be able to give change to each dollar holder?

B For $n \geq 3$ let us record the number of ways to triangulate an n-gon using nonintersecting diagonals. Call this number $t(n)$
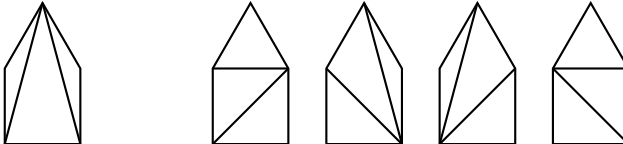
n	$t(n)$	triangulations
3	1	
4	2	
5	5	

Figure 13.1: n-gon Triangulations

In general it may be shown that $t(n) = c(n-2)$ for all $n \geq 3$.

Chapter 14

The Principle of Inclusion and Exclusion

14.1 A generalization of the addition rule

Recall from §2.6 that if A and B are finite, disjoint sets, then $A \cup B$ is finite, and

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

For arbitrary finite sets A and B, $|A \cup B|$ may be calculated by the following theorem.

Theorem 14.1. *If A and B are finite sets, then $|A \cup B| = |A| + |B| - |A \cap B|$.*

Proof. It is clear from the Venn diagram shown in Figure 14.1.
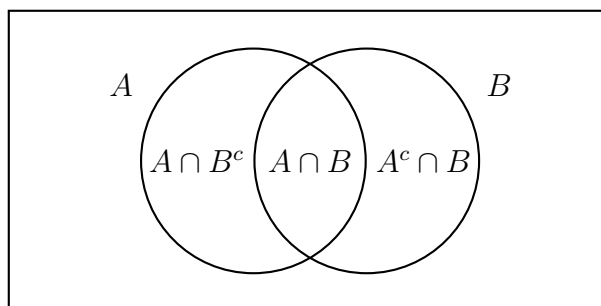
Figure 14.1: Venn Diagram

that $A \cap B^c$ and B are disjoint and $(A \cap B^c) \cup B = A \cup B$. So by the addition rule

(1) $$|A \cup B| = |A \cap B^c| + |B|.$$

Also, $A \cap B^c$ and $A \cap B$ are disjoint and $(A \cap B^c) \cup (A \cap B) = A$. So by the addition rule

(2) $$|A| = |A \cap B^c| + |A \cap B|, \quad \text{i.e.,}$$
(3) $$|A \cap B^c| = |A| - |A \cap B|.$$

Substituting this expression for $|A \cap B^c|$ into (1) yields the desired result. $\square$

The above theorem states the simplest case of the *principle of inclusion and exclusion.* The general case, which may be proved by induction on n using Theorem 4.1, is stated in the following theorem.

Theorem 14.2 (Principle of inclusion and exclusion). *If $(A_1, A_2, \ldots, A_k)$ is any sequence of finite sets, then*

$$|A_1 \cup A_2 \cup \cdots \cup A_k| = \sum_{1 \le i_1 \le k} |A_{i_1}| - \sum_{1 \le i_1 < i_2 \le k} |A_{i_1} \cap A_{i_2}| + \sum_{1 \le i_1 < i_2 < i_3 \le k} |A_{i_1} \cap A_{i_2} \cap A_{i_3}|$$
$$- \cdots + (-1)^{k-1} |A_1 \cap A_2 \cap \cdots \cap A_k|.$$

Special cases of Theorem 14.2, corresponding to $k = 2$ and $k = 3$, are

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|$$

and

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|.$$

Remark. Theorem 14.2 may be expressed in more compact (though perhaps less understandable) form as

$$\left| \cup_{i=1}^k A_i \right| = \sum_{\substack{I \subset [k] \\ I \ne \phi}} (-1)^{|I|-1} \left| \cap_{i \in I} A_i \right|.$$

14.2 The complementary principle of inclusion and exclusion

Theorem 14.2, as stated above, is not in the form most useful for applications. The following theorem states a variant of Theorem 14.2, the "complementary" principle of inclusion and exclusion, that is used most often in solving enumeration problems.

Theorem 14.3. *If $(A_1, A_2, \ldots, A_k)$ is a sequence of subsets of the finite set A, then*

$$|A - (A_1 \cup A_2 \cup \cdots \cup A_k)| = |A_1^c \cap A_2^c \cap \cdots \cap A_k^c|$$
$$= |A| - \sum_{1 \le i_1 \le k} |A_i| + \sum_{1 \le i_1 < i_2 \le k} |A_{i_1} \cap A_{i_2}|$$
$$- \sum_{1 \le i_1 < i_2 < i_3 \le k} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| + \cdots + (-1)^k |A_1 \cap A_2 \cap \cdots \cap A_k|$$
$$= \sum_{I \subset [k]} (-1)^{|I|} \left| \cap_{i \in I} A_i \right|,$$

where $\cap_{i \in \phi} A_i := A$.

Proof. The sets $A - (A_1 \cup \cdots \cup A_k)$ and $A_1 \cup \cdots \cup A_k$ are disjoint and $\{A - (A_1 \cup \cdots \cup A_k)\} \cup (A_1 \cup \cdots \cup A_k) = A$. By the addition rule,

$$|A| = |\{A - (A_1 \cup \cdots \cup A_k)\} \cup (A_1 \cup \cdots \cup A_k)|$$
$$= |A - (A_1 \cup \cdots \cup A_k)| + |A_1 \cup \cdots \cup A_k|,$$

i.e., $|A - (A_1 \cup \cdots \cup A_k)| = |A| - |A_1 \cup \cdots \cup A_k|$, which, with Theorem 14.2, yields the desired result. $\square$

14.3 Counting surjections

In Chapter 12 we used finite difference methods to derive the formula

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

for the number $\sigma(n,k)$ of surjections from $[n]$ to $[k]$. We now derive this formula using the complementary principle of inclusion and exclusion.

Let $A = [k]^{[n]} = \{f : [n] \rightarrow [k]\}$ and for $i = 1, \ldots, k$, let $A_i = \{f \in A :$ is *not* in the range of $f\}$. Then $A_1 \cup A_2 \cup \cdots \cup A_k$ is the set of nonsurjective functions from $[n]$ to $[k]$ and so $A - (A_1 \cup \cdots \cup A_k)$ is the set of surjective functions from $[n]$ to $[k]$. Hence

$$\begin{aligned}
\sigma(n,k) &= |A - (A_1 \cup \cdots \cup A_k)| \\
&= |A| - \sum_{1 \le i_1 \le k} |A_{i_1}| + \sum_{1 \le i_1 < i_2 \le k} |A_{i_1} \cap A_{i_2}| \\
&\quad - \sum_{1 \le i_1 < i_2 < i_3 \le k} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| + \cdots + (-1)^k |A_1 \cap A_2 \cap \cdots \cap A_k| \\
&= k^n - \binom{k}{1}(k-1)^n + \binom{k}{2}(k-2)^n - \binom{k}{3}(k-3)^n + \cdots + (-1)^k \binom{k}{k}(k-k)^n \\
&= \sum_{j=0}^{k}(-1)^j \binom{k}{j}(k-j)^n = \sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}j^n.
\end{aligned}$$

Remark 1. Note that in specifying that $A_i = \{f \in A : i$ is *not* in the range of $f\}$ we are *not* saying that every member of $[k]$ except i is in the range of f. Rather, A_i consists of all functions $f : [n] \rightarrow [k]$ such that i (and possibly other members of $[k]$) is not in the range of f.

Remark 2. In the above analysis, all of the terms $|A_{i_1}|$ were identically equal to $(k-1)^n$, all of the terms $|A_{i_1} \cap A_{i_2}|$ were identically equal to $(k-2)^n$, etc. Many counting problems have this feature. In general, if in the formula

$$|A - (A_1 \cup \cdots \cup A_k)| = \sum_{I \subset [k]}(-1)^{|I|} |\cap_{i \in I} A_i|$$

it is the case that $|\cap_{i \in I} A_i| = c_j$ whenever $|I| = j$, for $j = 0, \ldots, k$, then it will be the case that

$$|A - (A_1 \cup \cdots \cup A_k)| = \sum_{j=0}^{k}(-1)^j \binom{k}{j}c_j.$$

Combinatorial formulas of this type are very common and often indicative of the existence of a straightforward justification by the complementary principle of inclusion and exclusion.

14.4 Sequences with prescribed sum revisited

How many sequences (n_1, n_2, n_3) of integers are there satisfying

(i) $n_1 + n_2 + n_3 = 30$

subject to the constraints

(ii) $-3 < n_1 \leq 10$

(iii) $4 \leq n_2 \leq 8$

(iv) $0 < n_3 < 19$.

As we saw in §5.3, this problem can be solved by generating functions. Simply multiply the expressions
$$(x^{-2} + x^{-1} + 1 + x + \cdots + x^{10})(x^4 + \cdots + x^8)(x + \cdots + x^{18})$$
and read off the coefficient of x^{30} as the answer.

One can also solve this problem using Theorem 14.3. First, rewrite the constraints using inclusive (i.e., nonstrict) inequalities

(ii*) $-2 \leq n_1 \leq 10$

(iii*) $4 \leq n_2 \leq 8$

(iv*) $1 \leq n_3 \leq 18$,

rewriting these with uniform lower bounds of zero,

(ii**) $0 \leq n_1 + 2 \leq 12$

(iii**) $0 \leq n_2 - 4 \leq 4$

(iv**) $0 \leq n_3 - 1 \leq 17$.

Change variables to $m_1 := n_1 + 2$, $m_2 := n_2 - 4$, and $m_3 := n_3 - 1$. In terms of the m_i's, the original problem is equivalent to counting the sequences (m_1, m_2, m_3) of integers satisfying

(1) $m_1 + m_2 + m_3 = 27$ (why?)

subject to the constraints

(2) $0 \leq m_1 \leq 12$

(3) $0 \leq m_2 \leq 4$

(4) $0 \leq m_3 \leq 17$.

Now, let $A = \{(m_1, m_2, m_3) : m_1 + m_2 + m_3 = 27, \text{ with each } m_i \in \mathbb{N}\}$. Let

$$A_1 = \{(m_1, m_2, m_3) \in A : m_1 \geq 13\}$$
$$A_2 = \{(m_1, m_2, m_3) \in A : m_2 \geq 5\}$$
$$A_3 = \{(m_1, m_2, m_3) \in A : m_3 \geq 18\}$$

The set $A_1 \cup A_2 \cup A_3$ consists of those sequences in A that violate at least one of the constraints (2), (3), or (4). Hence, the set $A - (A_1 \cup A_2 \cup A_3)$ consists of those sequences satisfying (1), (2), (3), and (4). By Theorem 14.3

$$|A - (A_1 \cup A_2 \cup A_3)| = |A| - |A_1| - |A_2| - |A_3| + |A_1 \cap A_2| + |A_1 \cap A_3|$$
$$+ |A_2 \cap A_3| - |A_1 \cap A_2 \cap A_3|.$$

By Theorem 5.1, $|A| = \binom{29}{2}$. By Theorem 5.3, $|A_1| = \binom{16}{2}$, $|A_2| = \binom{24}{2}$, $|A_3| = \binom{11}{2}$, $|A_1 \cap A_2| = \binom{11}{2}$, and $|A_2 \cap A_3| = \binom{6}{2}$. Obviously $|A_1 \cap A_3| = |A_1 \cap A_2 \cap A_3| = 0$.

Hence, the number of solutions to (1)-(4) is

$$\binom{29}{2} - \binom{16}{2} - \binom{24}{2} - \binom{11}{2} + \binom{11}{2} + \binom{6}{2} = 25.$$

14.5 A problem in elementary number theory

Recall that if $x \in \mathbb{R}$, the *floor function* (also called the *greatest integer function*) of x, denoted $\lfloor x \rfloor$, is defined to be the largest integer n such that $n \le x$. Similarly, the *ceiling function* (also called the *least integer function*) of x, denoted $\lceil x \rceil$, is defined to be the smallest integer m such that $x \le m$. Of course, if $n \in \mathbb{Z}$, then $\lfloor n \rfloor = \lceil n \rceil = n$.

How many numbers in the set $[n] := \{1, \ldots, n\}$ are exactly divisible by the positive integer d? A little thought shows that the numbers in question are just the multiples $1 \cdot d$, $2 \cdot d$, $3 \cdot d$, $\ldots$, $\lfloor \frac{n}{d} \rfloor d$. So there are $\lfloor \frac{n}{d} \rfloor$ such numbers.

We symbolize the fact that n is exactly divisible by d (i.e., that n is an integer multiple of d) by $d \mid n$, reading this as "d divides n." Given integers d_1 and d_2, it is the case that $d_1 \mid n$ and $d_2 \mid n$ if and only if $\text{lcm}(d_1, d_2) \mid n$. More generally, given integers $d_1, d_2, \ldots, d_k$, it is the case that $d_1 \mid n$, $d_2 \mid n$, $\ldots$, and $d_k \mid n$ if and only if $\text{lcm}(d_1, \ldots, d_k) \mid n$.

With the above introductory remarks we are now prepared to solve the following problem:

How many numbers in the set $[337]$ are divisible neither by 8, nor by 12, nor by 20?

Solution: Let $A = [337]$, and for every positive integer d, let $A_d = \{n \in A : d \mid n\}$. The set $A_8 \cup A_{12} \cup A_{20}$ consists of those numbers in $[337]$ divisible by at least one of the numbers 8, 12, and 20. Hence the set $A - (A_8 \cup A_{12} \cup A_{20})$ consists of those numbers in $[337]$ that are divisible by none of the numbers 8, 12, and 20. By Theorem 14.3,

$$|A - (A_8 \cup A_{12} \cup A_{20})| = |A| - |A_8| - |A_{12}| - |A_{20}| + |A_8 \cap A_{12}| + |A_8 \cap A_{20}|$$
$$+ |A_{12} \cap A_{20}| - |A_8 \cap A_{12} \cap A_{20}|$$
$$= |A| - |A_8| - |A_{12}| - |A_{20}| + |A_{24}| + |A_{40}| + |A_{60}| - |A_{120}|$$
$$= 337 - \lfloor \frac{337}{8} \rfloor - \lfloor \frac{337}{12} \rfloor - \lfloor \frac{337}{20} \rfloor + \lfloor \frac{337}{24} \rfloor$$
$$+ \lfloor \frac{337}{40} \rfloor + \lfloor \frac{337}{60} \rfloor - \lfloor \frac{337}{120} \rfloor$$
$$= 337 - 42 - 28 - 16 + 14 + 8 + 5 - 2 = 276.$$

14.6   Problème des rencontres

Recall that the number of bijections $f : [n] \to [n]$, also called permutations of $[n]$, is given by $n!$. An element $i \in [n]$ such that $f(i) = i$ is called a *fixed point* (or, in French, *rencontre*) of f. For example, the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 1 & 4 & 2 \end{pmatrix},$$

sometimes written simply as the word 53142 has just one fixed point, namely, $i = 4$. A permutation having no fixed points is called a *dérangement*.

   Problème des rencontres: Determine d_n, the number of dérangements of $[n]$, where $n \in \mathbb{N}$. By brute force enumeration, one may check that $d_0 = 1$ (why?), $d_1 = 0$, $d_2 = 1$, and $d_3 = 2$. The general formula is given by the following theorem.

Theorem 14.4. *For all $n \in \mathbb{N}$,*

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

Proof. Let $A = \{f : [n] \to [n], \ f \text{ bijective}\}$ and for all $i \in [n]$, let
$A_i = \{f \in A : f(i) = i\}$. Then $A_1 \cup \cdots \cup A_n$ is the set of all bijections $f : [n] \to [n]$ having at least one fixed point and so $A - (A_1 \cup \cdots \cup A_n)$ is the set of all dérangements of $[n]$. By Theorem 14.3,

$$d_n = |A - (A_1 \cup \cdots \cup A_n)| = |A| - \sum_{1 \le i_1 \le n} |A_{i_1}| + \sum_{1 \le i_1 < i_2 \le n} |A_{i_1} \cap A_{i_2}| - \cdots + (-1)^n |A_1 \cap \cdots \cap A_n|$$

$$= n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! - \binom{n}{3}(n-3)! + \cdots + (-1)^n \binom{n}{n}(n-n)!$$

$$= n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right)$$

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

$\square$

   There is an amusing application of the above theorem, namely, the "hat check" problem. It seems that n gentlemen check their hats at the opera, and the attendant misplaces the checks on the hats and simply hands them back at random. What is the probability p_n that no one receives his own hat back? Abstractly, this amounts to asking for the probability that a randomly chosen permutation of $[n]$ is in fact a dérangement. Clearly

$$p_n = \frac{d_n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \to \frac{1}{e} \text{ as } n \to \infty.$$

In particular, therefore, we get the asymptotic formula

$$d_n \sim \frac{n!}{e}.$$

The formula for d_n given in Theorem 14.4 is tedious to use. Fortunately, there are several simple recursive formulas for d_n.

Theorem 14.5. *The numbers d_n satisfy the initial condition $d_0 = 1$ and the recurrence*

$$d_n = nd_{n-1} + (-1)^n \quad \text{for all } n \in \mathbb{P}.$$

Proof.

$$
\begin{aligned}
nd_{n-1} + (-1)^n &= n \left[(n-1)! \sum_{k=0}^{n-1} \frac{(-1)^k}{k!} \right] + (-1)^n \\
&= n! \sum_{k=0}^{n-1} \frac{(-1)^k}{k!} + n! \frac{(-1)^n}{n!} \\
&= n! \left(\sum_{k=0}^{n-1} \frac{(-1)^k}{k!} + \frac{(-1)^n}{n!} \right) \\
&= n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} = d_n, \text{ by Theorem 14.4.}
\end{aligned}
$$

$\square$

Theorem 14.6. *The numbers d_n satisfy the initial conditions $d_0 = 1$, $d_1 = 0$, and the recurrence*

$$d_n = (n-1)[d_{n-2} + d_{n-1}] \quad \text{for all } n \geq 2.$$

Proof. If $n \geq 2$, by Theorem 14.5, applied twice, once for d_n, then for d_{n-1}

$$
\begin{aligned}
d_n &= nd_{n-1} + (-1)^n \\
&= (n-1)d_{n-1} + d_{n-1} + (-1)^n \\
&= (n-1)d_{n-1} + [(n-1)d_{n-2} + (-1)^{n-1}] + (-1)^n \\
&= (n-1)[d_{n-1} + d_{n-2}].
\end{aligned}
$$

$\square$

Remark. There are combinatorial proofs of the recurrences of Theorems 14.5 and 14.6, but they are somewhat subtle, particularly in the former case.

The number of bijections $f : [n] \to [n]$ having no fixed points is given by d_n, as specified by the above theorems. More generally for $k = 0, 1, \ldots, n$, the number of such bijections having exactly k fixed points is clearly given by

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

Theorem 14.7. *For all $n \in \mathbb{N}$*

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

Proof. Each side of the above identity counts the class of bijections $f : [n] \to [n]$, the RHS by Theorem 2.5. The LHS counts this class in $n + 1$ disjoint, exhaustive subclasses, one for each $k = 0, 1, \ldots, n$, with the kth subclass comprised of all permutations having exactly k fixed points. $\qquad\square$

Let X record the number of fixed points of a randomly chosen bijection $f : [n] \to [n]$.

Theorem 14.8. *For all $n \in \mathbb{P}$, $E(X) = 1$.*

Proof.

$$E(X) = \sum_{k=0}^{n} kP(X = k) = \sum_{k=1}^{n} kP(X = k)$$

$$= \sum_{k=1}^{n} k\frac{\binom{n}{k}d_{n-k}}{n!} = \sum_{k=1}^{n} \binom{n-1}{k-1}\frac{d_{n-k}}{(n-1)!}$$

$$\underset{j=k-1}{=} \sum_{j=0}^{n-1} \binom{n-1}{j}\frac{d_{(n-1)-j}}{(n-1)!} = 1$$

by Theorem 14.7 with n replaced by $n - 1$, and k replaced by j. $\qquad\square$

Remark. Theorem 14.8 can also be proved by writing X as a sum of indicator random variables $X_1, \ldots, X_n$, where

$$X_i = \begin{cases} 1 & \text{if } i \text{ is a fixed point of } f \\ 0 & \text{if } i \text{ is not a fixed point of } f. \end{cases}$$

For clearly

$$E(X_i) = P(X_i = 1) = \frac{\# \text{ of bijections with } f(i) = i}{n!}$$

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

Hence $E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{1}{n} = 1$.

In problem 32 you are asked to consider a variation of the Problème des Rencontres in which one considers fixed points of arbitrary (not necessarily bijective) functions $f : [n] \to [n]$.

14.7 Problème des ménages

How many visually distinct ways are there to seat n married couples in $2n$ chairs placed around a circular table, with sexes alternating and spouses non-adjacent? This is the ménage problem (problème des ménages) , first posed by Lucas in 1891. The solution which we give below was first published (without proof) by Touchard in 1934. The first published proof of this result, due to Kaplansky, appeared in 1943. The proof that we give appeared in an elegant paper by Bogart and Doyle ("A non-sexist solution of the ménage problem," American Mathematical Monthly **43** (1986), 514-518).

We first establish two preliminary theorems about the placement of nonoverlapping dominos.

Theorem 14.9. *Let $dom(line\ m, k)$ denote the number of ways to place k indistinguishable, nonoverlapping dominos on a linear sequence of numbers $1, 2, \ldots, m$, where each domino covers two numbers. Then*

$$dom(line\ m, k) = \binom{m - k}{k}.$$

Proof. Each such placement corresponds in one-to-one fashion with a sequential arrangement of k 2's and $m - 2k$ 1's, as illustrated by the following example:

$$1 \quad \boxed{2 \quad 3} \quad 4 \quad 5 \quad \boxed{6 \quad 7} \quad \boxed{8 \quad 9} \quad 10$$

(This placement corresponds to the sequence 1211221.) By Theorem 4.1, there are $\binom{m-k}{k}$ sequential arrangements of k 2's and $m - 2k$ 1's. $\qquad\square$

Theorem 14.10. *Let $dom(circle\ m, k)$ denote the number of ways to place k indistinguishable, nonoverlapping dominos on a circular sequence of numbers $1, 2, \ldots, m$, where each domino covers two numbers. Then*

$$dom(circle\ m, k) = \frac{m}{m - k} \binom{m - k}{k}.$$

Proof. We count the placements in two disjoint, exhaustive classes, (i) the class of placements where the number 1 is covered by a domino and (ii) the class of placements where the number 1 is not covered by a domino.
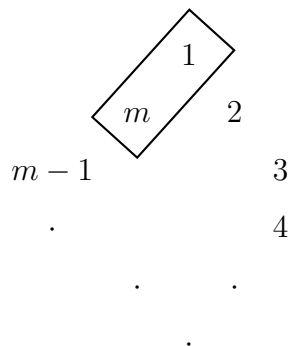
The number of placements in which, as above, m and 1 are covered by a single domino is clearly equal to dom(line $m - 2, k - 1$). Similarly, the number of placements in which 1 and 2 are covered by single domino is dom(line $m - 2, k - 1$). On the other hand, the number of placements in which 1 is not covered by any domino is clearly dom(line $m - 1, k$). By Theorem 14.9,

$$\text{dom(circle } m, k) = 2\,\text{dom(line } m - 2, k - 1) + \text{dom(line } m - 1, k)$$
$$= 2\binom{m - k - 1}{k - 1} + \binom{m - k - 1}{k}$$
$$= \frac{m}{m - k}\binom{m - k}{k}.$$

$\qquad\square$

We may now prove the main theorem of this section. Let M_n denote the number of seatings of the type described above in the statement of the ménage problem. The number M_n is called the nth *ménage number*.

Theorem 14.11. $M_1 = 0$, and for all $n \geq 2$,

$$M_n = 2(n!) \sum_{k=0}^{n} (-1)^k \frac{2n}{2n-k} \binom{2n-k}{k} (n-k)!.$$

Proof. Label the couples $1, 2, \ldots, n$. Let A be the set of all sex-alternating seatings of the n couples, and for $i = 1, \ldots, n$, let A_i be the set of seatings in A where the spouses of couple i are adjacent. Clearly,

$$M_n = |A - (A_1 \cup \cdots \cup A_n)|$$

$$= |A| - \sum_{1 \leq i_1 \leq n} |A_{i_1}| + \sum_{1 \leq i_1 < i_2 \leq n} |A_{i_1} \cap A_i| - \cdots + (-1)^n |A_1 \cap \cdots \cap A_n|$$

$$= |A| - \binom{n}{1} W_1 + \binom{n}{2} W_2 - \cdots + (-1)^k \binom{n}{k} W_k + \cdots + (-1)^n \binom{n}{n} W_n,$$

where $W_k =$ the number of seatings in A in which spouses of a fixed set of k couples are adjacent. Now $|A| = 2(n!)^2$, for in constructing a seating in A we decide whether to seat the women in odd or even numbered seats (2 possibilities) then place the women in those seats ($n!$ possibilities) and the men in the remaining seats ($n!$ possibilities). As for the numbers W_k, we have

$$W_k = 2 \cdot \text{dom}(\text{circle } 2n, k) \cdot k! \cdot (n-k)!(n-k)!$$

by the following argument. In constructing a seating in which spouses of a fixed set of k couples are adjacent, we

1° decide whether to seat women in odd or even numbered seats (2 possibilities),

2° choose k pairs of adjacent seats in which to seat the k couples ($\text{dom}(\text{circle } 2n, k)$ possibilities),

3° assign the couples to those pairs (i.e., dominos) ($k!$ possibilities),

4° seat the remaining $n - k$ women ($(n-k)!$ possibilities), and

5° seat the remaining $n - k$ men ($(n-k)!$ possibilities).

Using the formula $\text{dom}(\text{circle } 2n, k) = \frac{2n}{2n-k} \binom{2n-k}{k}$ from Theorem 14.10 then yields the formula

$$W_k = \frac{4n}{2n-k} \binom{2n-k}{k} k! [(n-k)!]^2.$$

Noting that $W_0 = 2(n!)^2 = |A|$, we get

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

$$= \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{4n}{2n-k} \binom{2n-k}{k} k! [(n-k)!]^2$$

$$= 2(n!) \sum_{k=0}^{n} (-1)^k \frac{2n}{2n-k} \binom{2n-k}{k} (n-k)!,$$

as asserted. $\qquad\square$

Note the above formula yields $M_2 = 0$, which makes sense. The first 10 values of M_n are tabulated below.

n	M_n
1	0
2	0
3	12
4	96
5	3120
6	$115,200$
7	$5,836,320$
8	$382,072,320$
9	$31,488,549,120$
10	$3,191,834,419,200$

Table 14.1: The Ménage Numbers M_n, $\quad 1 \leq n \leq 10$

Chapter 15

Problems and Honors Problems

Be patient towards all that is unsolved in your heart. Try to love the questions them-selves.

<div align="right">Rainer Maria Rilke</div>

15.1 Problems

Problem 1. For every positive integer n, let $w(n) = $ the number of binary words of length n (i.e., sequences of length n, each entry of which is either 0 or 1) having no adjacent zeros. Find a recursive formula for $w(n)$.

Problem 2. Determine the number of (a) injective functions $f : [98] \to [100]$, (b) injective functions $f : [100] \to [98]$, (c) injective functions $f : [100] \to [100]$, (d) surjective functions $f : [98] \to [100]$, (e) surjective functions $f : [100] \to [100]$.

Problem 3. (a) How many 5-element subsets of $[10]$ contain at least one of the members of $[3]$? (b) How many 5-element subsets of $[10]$ contain two odd and three even numbers? (c) Show that $[10]$ has as many subsets of even cardinality as it has subsets of odd cardinality. (d) [hard: extra credit]. Given $n \in \mathbb{P}$, exhibit a bijection from the class of subsets of $[n]$ of even cardinality to the class of subsets of $[n]$ of odd cardinality. (e) Assuming that a bijection of the aforementioned type exists (even if you couldn't find one) find a simple formula for the number of subsets of $[n]$ of odd cardinality, when $n \in \mathbb{P}$.

Problem 4. Let $n \geq 2$ and $k \geq 2$. The set of all k-element subsets of $[n]$ may be partitioned into four classes: (I) the class of subsets containing both 1 and 2; (II) the class of subsets containing 1, but not 2; (III) the class of subsets containing 2, but not 1; (IV) the class of subsets containing neither 1 nor 2. (a) How many k-element subsets of $[n]$ fall into class I? class II? class III? class IV? (b) What recurrence relation follows from the answers to (a)?

Problem 5. Find a simple expression for the sum $\sum_{k=0}^{n} k(k-1)\binom{n}{k}$ for $n \geq 0$ by using Theorem 3.6. Then give a second, *combinatorial*, proof of your result. Can you generalize the foregoing to find a simple expression for $\sum_{k=0}^{n} k^{\underline{r}}\binom{n}{k}$ for arbitrary $r \in \mathbb{N}$?

Problem 6. Find a simple expression for the sum $\sum_{k=0}^{n} \frac{1}{k+1}\binom{n}{k}$ for $n \geq 0$ by using Theorem 3.6.

Problem 7. Simplify the following sums, giving a combinatorial argument for your answer to part (d)

(a) $\displaystyle\sum_{j=1}^{513}(-1)^j\binom{513}{j}a^j$

(b) $\displaystyle\sum_{r=0}^{13}(-1)^r\binom{13}{r}5^{13-r}$

(c) $\displaystyle\sum_{k=0}^{18}\binom{18}{k}e^{-k}$

(d) $\displaystyle\sum_{k=0}^{25}\binom{37}{k}\binom{63}{25-k}$

Problem 8. How many lattice paths from $(-5,-3)$ to $(7,4)$ pass through

(a) both $(-1,1)$ and $(3,2)$

(b) neither $(-1,1)$ nor $(3,2)$?

Problem 9. (a) We have 20 identical blocks each to be painted red, white, or blue. In how many ways may this be done if at least 4 blocks must be painted red and at least 5 blocks must be painted white? (b) In how many ways may 100 indistinguishable balls be placed in 30 distinguishable urns if at least 3 balls must be placed in each urn?

Problem 10. I have 5 identical pennies, 8 identical nickels, and 10 identical dimes. In how many ways may I select a prime number of these coins, if I must select at least 2 pennies, at least 4 nickels, and at least 6 dimes? Use the method of generating functions, simplifying your answer to a single number.

Problem 11. (a) Given 10 balls, labeled $1,\ldots,10$, and three urns, labeled $1,2,3$, the trinomial coefficient $\binom{10}{3,2,5}$ counts a certain category of distributions of the 10 balls among the three urns. Describe this category of distributions in *a complete, succinct, properly punctuated sentence.* (b) Give a combinatorial proof of the identity

$$\binom{10}{3,2,5}=\binom{9}{2,2,5}+\binom{9}{3,1,5}+\binom{9}{3,2,4}$$

Problem 12. How many sequential arrangements of (all) of the letters a,A,b,B,c,C,d,D are there (a) in which the lower case letters appear in alphabetical order - not necessarily adjacent to each other - and similarly for the upper case letters? (b) in which the lower case letter of each type precedes - not necessarily directly - the upper case letter of the same type?

Problem 13. (a) Simplify the sum

$$\sum_{\substack{n_1+n_2+n_3+n_4+n_5=3\\ n_i\in\mathbb{N}}}\binom{3}{n_1,n_2,n_3,n_4,n_5}.$$

(b) Simplify the sum

$$\sum_{\substack{i+j+k=10\\ i,j,k\in\mathbb{N}}}(-1)^i\frac{10!}{i!j!k!}2^j.$$

(c) How many terms are there in the expansion of $(x+y+z+w)^{88}$ (i) before simplification by the multinomial theorem (ii) after simplification by that theorem?

Problem 14. Three dice are rolled 10 times, and the sum of the numbers showing on the top faces of these dice is recorded each of the 10 times. Let X_1 denote the number of times that the sum is a prime number, X_2 the number of times that the sum is composite (i.e., not prime), and odd, and X_3 the number of times that the sum is composite and even. Determine $P(X_1 = 7, X_2 = 2, X_3 = 1)$, expressing your final answer as a decimal rounded to 3 decimal places. Use the method of generating functions to do the counting relevant to this problem.

Problem 15. In Problem 5, you developed a simple expression for $\sum_{k=0}^{n} k^r \binom{n}{k}$. Using the cases $r = 1, 2,$ and 3 of that result, find a simple expression for $\sum_{k=0}^{n} k^3 \binom{n}{k}$.

Problem 16. Let X record the number of successes in a binomial experiment with n trials and prob(success) $= p$ on each trial.

  (a) Evaluate the second factorial moment of X, $E(X^{\underline{2}})$

  (b) Using the result of (a) and the fact that $E(X) = np$ (Theorem 7.1), evaluate the second moment of X, $E(X^2)$

  (c) Using the above results, evaluate $\text{Var}(X) = E(X^2) - [E(X)]^2$.

Problem 17. Determine the number of binary relations on $[n]$ that are (a) symmetric and asymmetric (b) reflexive, symmetric, and complete (c) reflexive, symmetric, and antisymmetric.

Problem 18. Determine the number of binary relations on $[n]$ that are (a) asymmetric (b) antisymmetric (c) complete.

Problem 19. (a) Find a simple expression for $\sigma(n, 2)$ when $n \geq 2$. (b) Do the same for $\sigma(n, n-1)$ when $n \geq 2$.

Problem 20. (a) An ordered partition of $[8]$ is selected at random. What is the probability that it contains 7 or more blocks? (b) A partition of $[8]$ is selected at random. What is the probability that it contains 7 or more blocks?

Problem 21. Let X record the number of successes in a binomial experiment with n trials and prob(success) $= p$ on each trial. (a) Evaluate the rth factorial moment of X, $E(X^{\underline{r}})$. (b) Using the results of (a), evaluate the rth moment of X, $E(X^r)$, and the special case $E(X^4)$.

Problem 22. Using the result $\sum_{k=0}^{n} k^{\underline{j}} = \frac{1}{j+1}(n+1)^{\underline{j+1}}$, evaluate $S_r(n) = \sum_{k=0}^{n} k^r$ for the special cases $r = 3, 4,$ and 5. Express answers as polynomials in n, in factored form (e.g., $S_1(n) = n(n+1)/2$).

Problem 23. (a) Determine $\Delta^{-1}\alpha^x$ for fixed real $\alpha > 0$, where $\alpha \neq 0$. (b) Show that $\Delta^{-1}\sin \alpha x = -\cos(\alpha x - \frac{\alpha}{2})/(2 \sin \frac{\alpha}{2}) + c(x)^*$ for fixed $\alpha \in \mathbb{R}^{**}$. (c) Evaluate $\sum_{k=0}^{101} \sin \frac{k\pi}{2}$ in two different ways. (d) Determine $\Delta^{-1}\cos \alpha x$ for fixed $\alpha \in \mathbb{R}^{**}$.
    [* c is any function of period 1, i.e., $c(x+1) = c(x)$ for all x]
    [** where $\alpha \notin \{2\pi m : m \in \mathbb{Z}\}$]

Problem 24. Recall that the product rule for differentiation, $(fg)' = f'g + fg'$, yields the integration-by-parts formula $\int f'g = fg - \int fg'$. (a) Prove that $\Delta(f(x)g(x)) = (\Delta f(x))g(x) + f(x+1)\Delta g(x)$ and thus that (b) $\Delta^{-1}((\Delta f(x))g(x)) = f(x)g(x) - \Delta^{-1}(f(x+1)\Delta g(x))$. (c) Determine $\Delta^{-1}(x2^x)$.

Problem 25. Use the generating function method to find a formula for $f(n)$ if $f(0) = 2$, $f(1) = 10$, and $f(n) = 10f(n-1) - 21f(n-2)$, $\forall n \geq 2$.

Problem 26. Let $f(n)$ denote the number of words of length n in the alphabet $\{a, b, c\}$ in which a and b are never adjacent. So $f(1) = 3$ (the acceptable words of length 1 being a, b, and c) and $f(2) = 7$ (the acceptable words of length 2 being aa, bb, cc, ac, bc, ca, and cb). By a rather subtle argument one may establish the recurrence relation $f(n) = 2f(n-1) + f(n-2)$ for all $n \geq 3$. Find a formula for $f(n)$.

Problem 27. Find a formula for $f(n)$ if $f(0) = -1$, $f(1) = 15$, $f(2) = 171$, and $f(n) = 9f(n-1) - 27f(n-2) + 27f(n-3)$, $\forall n \geq 3$.

Problem 28. For all $n \geq 1$, let $f(n) =$ the number of words of length n in the alphabet $\{a, b, c, d, e\}$ with no adjacent a's. Determine as much information about the numbers $f(n)$ as you can, including, at a minimum, the values of $f(n)$ for $n = 1, 2, \ldots, 10$. This problem is deliberately open-ended. It is up to you to decide how to explore these numbers.

Problem 29. How many elements of the set $[15,000]$ are divisible neither by 6, nor by 9, nor by 15?

Problem 30. How many sequences (n_1, n_2, n_3, n_4) of nonnegative integers are there satisfying all of the following: (1) $n_1 + n_2 + n_3 + n_4 = 50$, (2) $n_1 \leq 10$, (3) $n_2 \leq 20$, (4) $n_3 \leq 30$, (4) $n_4 \leq 40$? Express your final answer as a single number, but show details of your derivation.

Problem 31. (a) How many permutations of $[10]$ have neither 1 nor 2 as fixed points? (b) Eight letters are removed from their envelopes, read, and replaced at random. What is the probability that at least one letter will be returned to its own envelope? What is the probability that at least two letters will be returned to their own envelopes?

Problem 32. As in the special case of permutations, we say of an arbitrary function $f : [n] \to [n]$ that k is a fixed point of f if $f(k) = k$. For all $n \in \mathbb{N}$, let D_n denote the number of functions $f : [n] \to [n]$ having no fixed points. Investigate D_n, determining in particular the probability (call it π_n) that a randomly chosen function $f : [n] \to [n]$ has no fixed points, and also $\lim_{n \to \infty} \pi_n$.

15.2 Honors Problems

Honors Problem I. Determine appropriate initial values and a recurrence relation for each of the following sequences, where $n \in \mathbb{P}$ and $s, r \in \mathbb{P}$ with $s, r \geq 2$:

   1a. $w_2(n) = \#$ of words of length n in the alphabet $\{1, 2\}$ with no two 1's adjacent.

   1b. $w_r(n) = \#$ of words of length n in the alphabet $\{1, 2, \ldots, r\}$ with no two 1's adjacent.

   1c. $w_{r,s}(n) = \#$ of words of length n in the alphabet $\{1, 2, \ldots, r\}$ with no s 1's adjacent (fewer than s adjacent 1's OK).

   2a. $c_2(n) = \#$ of sequences (necessarily of length $\leq n$), whose entries belong to the set $\{1, 2\}$ and sum to n.

2b. $c_r(n) = \#$ of sequences (necessarily of length $\leq n$), whose entries belong to the set $\{1, \ldots, r\}$ and sum to n.

3a. $s_2(n) = \#$ of subsets $A \subset [n]$ such that $|x - y| \geq 2$ for all $x, y \in A$ with $x \neq y$.

3b. $s_r(n) = \#$ of subsets $A \subset [n]$ such that $|x - y| \geq r$ for all $x, y \in A$ with $x \neq y$.

Honors Problem II. Prove that for all $n \in \mathbb{N}$, $\sum_{k=0}^{n} \binom{n-k}{k} = F_n$, the nth Fibonacci number. There is of course a straightforward inductive proof (which is acceptable), but there is also a combinatorial proof, which you are encouraged to find.

Honors Problem III. For $r, s \in \mathbb{N}$, let $Q(r, s) =$ the number of lattice paths, with diagonals allowed, from $(0, 0)$ to (r, s). Determine $Q(r, 0)$ and $Q(0, s)$ for all $r, s \in \mathbb{N}$. By partitioning the family of all such paths into a few natural categories, find a recurrence relation for $Q(r, s)$ when $r, s \geq 1$. Can you find a closed form expression for $Q(r, 1)$ and $Q(r, 2)$? What can be said about the parity of $Q(r, s)$?

Honors Problem IV. The following problem arose in acoustics: Find a simple formula for

a. $\binom{n}{0}\binom{n}{1} + \binom{n}{1}\binom{n}{2} + \cdots + \binom{n}{n-1}\binom{n}{n}$,

b. $\binom{n}{0}\binom{n}{2} + \binom{n}{1}\binom{n}{3} + \cdots + \binom{n}{n-2}\binom{n}{n}$,

and, more generally, for

c. $\sum_{k=0}^{n-r} \binom{n}{k}\binom{n}{k+r}$ where $0 \leq r \leq n$.

Honors Problem V. A real function f is *nondecreasing* if $x \leq y \Rightarrow f(x) \leq f(y)$ and *increasing* if $x < y \Rightarrow f(x) < f(y)$. (a) How many $f : [n] \to [k]$ are nondecreasing? (b) How many $f : [n] \to [k]$ are increasing?

Honors Problem VI. For $n \in \mathbb{N}$ and $k \in \mathbb{P}$, find the number of solutions of the inequality $n_1 + \cdots + n_k \leq n$, with each $n_i \in \mathbb{N}$. There is an obvious, complicated answer to this question. Try to find a simple answer.

Honors Problem VII. Our formula for the unique polynomial $p(x)$ of degree $\leq n$ taking the preassigned values $p(0), p(1), \ldots, p(n)$ at $x = 0, 1, \ldots, n$ was

$$p(x) = \sum_{k=0}^{n} \Delta^k p(0) \binom{x}{k}$$

a. Develop a more general formula of this type based on knowledge of the values $p(r), p(r+1), \ldots, p(r+n)$, where r is some fixed but arbitrary real number.

b. Develop an even more general formula of this type based on knowledge of the values $p(r), p(r+h), p(r+2h), \ldots, p(r+nh)$, where r is a fixed but arbitrary real number and h is a fixed but arbitrary positive real number. (Hint: Use the operator Δ_h here.)

Honors Problem VIII. An *ordered k-cover* of [n], for fixed $k \in \mathbb{P}$ and $n \in \mathbb{N}$, is a sequence $(S_1, \ldots, S_k)$ of subsets of [n] such that $S_1 \cup \cdots \cup S_k = [n]$. Let $c(n, k)$ denote the number of ordered k-covers of [n]. Develop a formula for $c(n, k)$ using the principle of inclusion and exclusion, simplify the formula, and then give a quick justification for the simplified formula.

Honors Problem IX. Solve the "relaxed ménage problem," i.e., determine $m_n :=$ the number of visually distinct ways of seating n couples around a circular table with no spouses adjacent (sexes need not alternate).

References

The following books and web sites furnish additional sources of information about combinatorics. Books are listed in ascending degree of difficulty. Web Sites are listed at the end of the list.

[1] G. Berman and K. D. Fryer. *Introduction to Combinatorics*. Academic Press, New York, 1972.

[2] N. Vilenkin. *Combinatorics*. Academic Press, New York, 1971.

[3] K. Bogart. *Introductory Combinatorics*. Harcourt Brace Jovanovitch, Orlando, FL, 2nd edition, 1990.

[4] D. Cohen. *Basic Techniques of Combinatorial Theory*. Wiley, New York, 1978.

[5] R. A. Brualdi. *Introductory Combinatorics*. Prentice Hall, Englewood Cliffs, NJ, 4th edition, 2004.

[6] C. L. Liu. *Introduction to Combinatorial Mathematics*. McGraw-Hill, New York, 1968.

[7] F. S. Roberts and B. Tesman. *Applied Combinatorics*. Pearson Prentice Hall, Upper Saddle River, NJ, 2004.

[8] R. Lidl and G. Pilz. *Applied Abstract Algebra*. Springer-Verlag, New York, 2nd edition, 1998.

[9] H. J. Ryser. *Combinatorial Mathematics*, volume 14 of *Carus Mathematical Monographs*. Mathematical Association of America, 1963.

[10] C. Berge. *Principles of Combinatorics*. Academic Press, New York, 1971.

[11] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics*. Addison-Wesley, Reading, MA, 1990.

[12] H. S. Wilf. *generatingfunctionology*. Academic Press, New York, 1990.

[13] R. P. Stanley. *Enumerative Combinatorics*, volume 1. Cambridge University Press, Cambridge, England, 1999.

[14] The On-line Encyclopedia of Integer Sequences. `http://www.research.att.com/~njas/sequences/`.

[15] Journal of Integer Sequences,. `http://www.cs.uwaterloo.ca/journals/JIS/`.

[16] The Electronic Journal of Combinatorics. `http://www.combinatorics.org/`.

[17] Annals of Combinatorics. `http://www.combinatorics.net/`.

[18] Mathworld. `http://mathworld.wolfram.com/`.

Index

Warner, Seth, 8
weak composition of n with k parts, 29
weak order, 55
word of length k, 7