CHAPTER 4

Sieving Methods

Introduction

A "sieving method" is a technique that allows us to count or list some things indirectly. After a few words about organization and difficulty, we'll introduce the two sieving methods discussed in this chapter.

- The sections of this chapter are independent of each other. Thus, if your instructor assigns only the material on the Principle of Inclusion and Exclusion, you need not read the sections on structures with symmetries. You may also read the material on counting structures with symmetries without reading the material on listing them.
- The material in this chapter is more difficult than the first three chapters in this part. Since the material here is not needed until Part IV, it may be postponed.

Structures Lacking Things

In Section 4.1 we look at the problem of counting structures that lack certain things; e.g., lists with no repeated elements or permutations with no fixed points. Sometimes, as in the case of lists with no repeated elements, it is easy to count the structures directly. That situation is not of interest here. Instead, we'll examine what happens when it's fairly easy to count structures which have some of the properties but hard to count those which have none of the properties. For example, consider permutations of $\underline{n}$ and a set $\{F_1, \ldots, F_n\}$ of n properties where F_i is the property that the permutation fixes i; that is, maps i to i. Suppose our problem is to count permutations with none of the properties; that is, permutations with no fixed points. This is hard. However, it is fairly easy to count permutations whose fixed points include some specified set S; that is, permutations that have at least some the properties $\{F_i \mid i \in S\}$. These counts can be used to indirectly solve the original problem by using the "Principle of Inclusion and Exclusion."

The Principle of Inclusion and Exclusion can be extended in various ways. We briefly indicate some of these at the end of Section 4.1.

Structures with Symmetries

At the end of Example 1.12 (p. 13), we asked how many ways we could form a six long circular sequence using ones and twos and found that we could not solve it. In this section we'll develop the necessary tools.

The circular sequence problem is difficult because "symmetries induce equivalences." What does this mean? The sequence 121212 looks the same if it is circularly shifted two positions. This is a symmetry of the sequence. Several sequences correspond to the same circular sequence of ones and twos. We say these lists are "equivalent." Thus, as we saw in Example 1.12, the three sequences 112112, 121121 and 211211 are equivalent. We can find a sequence equivalent to a given one by reading the given sequence "circularly:" Start reading at any point. At the end of the sequence jump to the start and continue until you return to where you began reading.

To list the circular sequences, we need a list C of sequences such that every sequence is equivalent to exactly one sequence in C. Thus, exactly one of the sequences 112112, 121121 and 211211 would appear in C. Counting the circular sequences means finding $|C|$. We'll discuss listing first and then counting.

We have already dealt with one important case of symmetries, namely, when our structures are lists and we are allowed to permute the items in the list in any fashion whatsoever. In other words, two lists are the same if they can be made identical by permuting the elements in one of them. In fact, this case is so important that it has a name: multisets. (Remember that a multiset is simply a list where order is irrelevant.) If the elements of the multiset can be ordered, then we can take our representatives C to be a collection of nondecreasing functions. This was discussed in Section 2.3.

In Section 4.2 we'll look at the problem of listing structures when symmetries are present. This is much like the nonmathematical notion of a sieve: all that comes through the sieve are "canonical" representations of the structures. Decision trees play an important role.

In Section 4.3 we'll look at the problem of counting, rather than listing, these structures. "Burnside's Lemma" provides us with an indirect method for doing this.

4.1 The Principle of Inclusion and Exclusion

Imagine that a professor on the first day of class wants to obtain information on the course background of the students. He wants to know what number of students have had Math 21, what number have had Comp Sci 13 and various combinations such as "Comp Sci 13 but not Math 21." For some reason, to calculate these numbers the professor asks just the following three questions.

"How many of you have had Math 21?"
"How many of you have had Comp Sci 13?"
"How many of you have had Comp Sci 13 and Math 21?"

Suppose the number of students is 15, 12 and 8, respectively.

Can the professor now determine answers to all other possible questions concerning having taken or not taken these courses? Let's look at a couple of possibilities.

How many have had Comp Sci 13 but not Math 21? Of the 12 students who have had the first course, 8 have had the second and so $12 - 8 = 4$ of them have not had the second.

How many students have had neither course? That will depend on the total number of students in the class. Suppose there are 30 students in the class. We might think that $30 - 15 - 12 = 3$ of them have had neither course. This is not correct because the students who had both courses were subtracted off twice. To get the answer, we must add them back in once. The result is that there are $3 + 8 = 11$ students who have had neither course.

We can rephrase the previous discussion in terms of sets. Let S be the set of students in the class, S_1 the subset who have had Math 21 and S_2 the subset who have had Comp Sci 13. The information that was obtained by questioning the class can be written as

$$|S| = 30, \quad |S_1| = 15, \quad |S_2| = 12 \quad \text{and} \quad |S_1 \cap S_2| = 8,$$

where $S_1 \cap S_2$ denotes the intersection of the sets S_1 and S_2. We saw that the number of students who had neither course is given by

$$|S| - \Big(|S_1| + |S_2|\Big) + |S_1 \cap S_2|. \tag{4.1}$$

How can this result be extended to more than two classes? The answer is provided by the following theorem. After stating it, we'll see how it can be applied before proving it.

Theorem 4.1 Principle of Inclusion and Exclusion *Let $S_1, S_2, \ldots, S_m$ be subsets of a set S. Let $N_0 = |S|$ and, for $r > 0$, let*

$$N_r \;=\; \sum |S_{i_1} \cap \cdots \cap S_{i_r}|, \tag{4.2}$$

where the sum is over all r-long strictly increasing sequences chosen from $\underline{m}$; that is, $\{i_1, \ldots, i_r\}$ ranges over all r-subsets of $\underline{m}$. The number of elements in S that are not in any of $S_1, \ldots, S_m$ is

$$\sum_{i=0}^{m} (-1)^i N_i \;=\; N_0 - N_1 + N_2 \cdots + (-1)^m N_m. \tag{4.3}$$

When $m = 2$, the one long sequences are 1 and 2, giving $N_1 = |S_1| + |S_2|$. The only two long sequence is 1,2 and so $N_2 = |S_1 \cap S_2|$. Thus (4.3) reduces to (4.1) in this case.

As we saw in an earlier chapter, strictly increasing sequences are equivalent to subsets. Also, the order in which we do intersections of sets does not matter. (Just as the order of addition does not matter and the order of multiplication does not matter.) This explains why we could have said that the sum defining N_r was over all r-subsets $\{i_1, \ldots, i_r\}$ of $\underline{m}$.

One can rewrite (4.3) in a somewhat different form. To begin with, the Rule of Sum tells us that $|S|$ equals the number of things not in any of the S_i's plus the number of things that are in at least one of the S_i's. The latter equals

$$|S_1 \cup \cdots \cup S_m|.$$

Using (4.3) and noting that $N_0 = |S|$, we have

$$N_0 \;=\; |S_1 \cup \cdots \cup S_m| + N_0 - N_1 + N_2 \cdots + (-1)^m N_m.$$

Rearranging leads to

Corollary *With the same notation as in Theorem 4.1 (p. 95),*

$$|S_1 \cup \cdots \cup S_m| \;=\; \sum_{i=1}^{m} (-1)^{i-1} N_i. \tag{4.4}$$

In this form, the Principle of Inclusion and Exclusion can be viewed as an extension of the Rule of Sum: The Rule of Sum tells us that if $T = S_1 \cup \cdots \cup S_m$ and if each structure in T appears in exactly one of the S_i, then

$$|T| \;=\; |S_1| + |S_2| + \ldots + |S_m|.$$

The left hand side of this equation is the left hand side of (4.4). The right hand side of this equation is N_1, the first term on the right hand side of (4.4). The remaining terms on the right hand side of (4.4) can be thought of as "corrections" to the Rule of Sum due to the fact that elements of T can appear in more than one S_i.

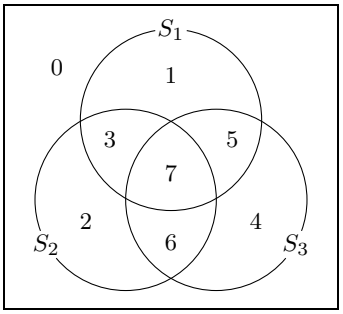


Figure 4.1 A Venn diagram for three subsets S_1, S_2 and S_3.

Example 4.1 Venn diagrams When m is quite small in Theorem 4.1, it is possible to draw a picture, called a *Venn diagram* that illustrates the theorem. Figure 4.1 shows such a diagram for $m = 3$. The interior of the box should be thought of as containing points which are the elements of S. (These points are not actually shown in the diagram.) Similarly, the interior of the circle labeled S_1 contains the elements of S_1 and its exterior contains the points not in S_1. Altogether, the three circles for S_1, S_2 and S_3 divide the box into eight regions which we have numbered 0 through 7.

In the figure, region 7 corresponds to $S_1 \cap S_2 \cap S_3$. Region 0 corresponds to those elements of S that are not in any S_i. Region 3 corresponds to those elements of S that are in S_1 and S_2 but not in S_3. You should be able to describe all of the other regions in a similar manner. The elements of S_1 are those in the four regions numbered 1, 3, 5 and 7. The elements of $S_1 \cap S_3$ are those in regions 5 and 7. You should be able to describe all the intersections in the Principle of Inclusion and Exclusion in this manner. You can then determine how often each region is counted in N_r and thereby obtain a proof of (4.3) for $m = 3$. It is possible to generalize this argument to prove (4.3), but we will give a slightly different proof of (4.3) later. □

Example 4.2 Using the theorem Many of Alice's 16 friends are athletic—they cycle, jog or swim on a regular basis. In fact we know that 6 of them cycle, 6 of them jog, 6 of them swim, 4 of them cycle and jog, 2 of them cycle and swim, 3 of them jog and swim and 2 of them engage in all three activities. How many of Alice's friends do none of these things on a regular basis?

Let S be the set of all friends, S_1 the set that cycle, S_2 the set that jog and S_3 the set that swim. We will apply (4.3) with $m = 3$. The information we were given can be rewritten as follows:

$$\begin{aligned}
N_0 &= 16 \quad \text{since} \quad |S| = 16 \\
N_1 &= 18 \quad \text{since} \quad |S_1| = 6 \quad |S_2| = 6 \quad |S_3| = 6; \\
N_2 &= 9 \quad \text{since} \quad |S_1 \cap S_2| = 4 \quad |S_1 \cap S_3| = 2 \quad |S_2 \cap S_3| = 3; \\
N_3 &= 2 \quad \text{since} \quad |S_1 \cap S_2 \cap S_3| = 2.
\end{aligned}$$

Thus the answer to our question is that $16 - 18 + 9 - 2 = 5$ of her friends neither cycle nor jog nor swim regularly. □

At this point you may well object that this method is worse than useless because there are much easier ways to get the answer. For example, to find out how many students took neither Math 21 nor Comp Sci 13, it would be easier to simply ask "How many of you have had neither Math 21 nor Comp Sci 13?" This is true. So far we've just been getting familiar with what the Principle of Inclusion and Exclusion means. We now turn to some examples where it is useful.

Example 4.3 Counting surjections How many surjections are there from $\underline{n}$ to $\underline{k}$?

This problem is closely related to $S(n,k)$, the Stirling numbers of the second kind, which we studied previously but couldn't find a formula for. In fact, a surjection f defines a partition of the domain into k blocks where the ith block is $f^{-1}(i)$. Since the blocks are all distinct and $S(n,k)$ does not care about the order of the blocks, the number of surjections is $k!S(n,k)$. Our attention will be devoted to the surjections—we don't need $S(n,k)$ here—but we pointed out the connection because it will allow us to get a formula for $S(n,k)$, too.

Let S be the set of all functions from $\underline{n}$ to $\underline{k}$ and let S_i be the set of those functions that never take on the value i. In this notation, the set of surjections is the subset of S that does not belong to any of $S_1, \ldots, S_k$ because a surjection takes on all values in its range. This suggests that we use (4.3) with $m = k$.

We found long ago that $|S| = k^n$. It is equally easy to find $|S_{i_1} \cap \cdots \cap S_{i_r}|$: The set whose cardinality we are taking is just all functions from $\underline{n}$ to $\underline{k} - \{i_1, \ldots, i_r\}$ and so equals $(k-r)^n$. This tells us that each of the terms in the sum (4.2) defining N_r equals $(k-r)^n$. Consequently, N_r is $(k-r)^n$ times the number of terms. Since there $\binom{k}{r}$ subsets of $\underline{k}$ of size r, the sum contains $\binom{k}{r}$ terms and $N_r = \binom{k}{r}(k-r)^n$. It follows from (4.3) that the number of surjections is

$$\sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n. \tag{4.5}$$

Combining this with the remarks at the start of this example, we have

$$S(n,k) \;=\; \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n \;=\; \sum_{i=0}^{k} \frac{(-1)^i}{i!} \frac{(k-i)^n}{(k-i)!}. \tag{4.6}$$

Because of the possibility of considerable cancellation due to alternating signs, numerical evaluation of this expression for large values of n and k can be awkward. □

In learning to apply the Principle of Inclusion and Exclusion, it can be difficult to decide what the sets $S, S_1, \ldots$ should be. It is often helpful to think in terms of

- a larger problem S that is easier to solve and
- conditions C_i that *all* do NOT hold for precisely those structures in S that are solutions of the original problem.

What's the connection between all this and the sets in Theorem 4.1? The set S_i is the set of structures in the larger problem S that satisfy C_i. Note that NOT appears in our description because (4.3) counts those elements of S that are NOT in any of the S_i's.

Let's look at the previous example in these terms. Our larger problem is the set of all functions for $\underline{n}$ to $\underline{k}$; that is, $S = \underline{k}^{\underline{n}}$. Since we want those functions that do NOT omit any of the values $1, \ldots, k$, we take C_i to be the condition that $f : \underline{n} \to \underline{k}$ omits the value i; that is, $i \notin \text{Image}(f)$.

Sometimes people talk about properties instead of conditions. In this case, they speak of "having a property" instead of "satisfying a condition."

Example 4.4 Counting solutions to equations How many different solutions are there to the equation

$$x_1 + x_2 + x_3 + x_4 + x_5 = n, \tag{4.7}$$

where the x_i's must be positive integers, none of which exceeds k?

Since it is easier to solve the equations without the constraint that the x_i's not exceed k, we'll use Theorem 4.1 as follows:

- Let S be the set of all positive integer solutions $(x_1, \ldots, x_5)$ of the equation.
- Let the ith condition be $x_i > k$ for $i = 1$, 2, 3, 4 and 5. The answer to the original problem is the number of elements of S (i.e., positive integer solutions) that satisfy none of the conditions.

For this to be useful, we must be able to easily determine, for example, how many solutions to (4.7) have $x_1 > k$ and $x_3 > k$. This is simply the number of solutions to $y_1 + \cdots + y_5 = n - 2k$ because we can take $x_1 = y_1 + k$, $x_3 = y_3 + k$ and $x_j = y_j$ for $j = 2$, 4 and 5.

We are ready to apply (4.3) with $m = 5$. To begin with, what is $|S|$? A solution in S can be obtained by inserting commas and plus signs in the $n - 1$ spaces between the n ones in $(1\ 1\ 1 \ldots 1)$ in such a way that either a plus or a comma, but not both, is inserted in each space and exactly 4 commas are used. Thus $|S| = \binom{n-1}{4}$. By this and the end of the previous paragraph, it follows that

$$|S_{i_1} \cap \ldots \cap S_{i_r}| \;=\; \binom{n - kr - 1}{4}, \tag{4.8}$$

where the binomial coefficient is taken to be zero if $n - kr - 1 < 4$. Since there are $\binom{5}{r}$ choices for the set $\{i_1, \ldots, i_r\}$, the number of solutions to (4.7) is

$$\begin{aligned}\binom{n-1}{4} &- \binom{5}{1}\binom{n-k-1}{4} + \binom{5}{2}\binom{n-2k-1}{4} \\ &- \binom{5}{3}\binom{n-3k-1}{4} + \binom{5}{4}\binom{n-4k-1}{4} \\ &- \binom{5}{5}\binom{n-5k-1}{4}.\end{aligned}$$

This formula is a bit tricky. If we blindly replace the binomial coefficients using the falling factorial formula

$$\binom{m}{4} \;=\; \frac{m(m-1)(m-2)(m-3)}{24} \tag{4.9}$$

and use algebra to simplify the result, we will discover that the number of solutions is zero! How can this be? The definition of binomial coefficient that we used for (4.8) gives $\binom{m}{4} = 0$ when $m < 0$, which does not agree with the falling factorial formula (4.9). Thus (4.9) cannot be used when $m < 0$.

The problem that we have been considering can be interpreted in other ways:

- How many compositions of n are there that consist of five parts, none of which exceed k?
- How many ways can n unlabeled balls be placed into five labeled boxes so that no box has more than k balls?

You should easily be able to see that these problems are all equivalent. □

Finally, the proof of Theorem 4.1:

Proof: Suppose that $s \in S$. Let $X \subseteq \underline{k}$ be such that $x \in X$ if and only if $s \in S_x$; that is, X is the set consisting of the indices of those S_i's that contain s. How much does s contribute to the sum in (4.3)? For the theorem to be true, it must contribute 1 if $X = \emptyset$ and 0 otherwise.

Clearly s contributes 1 to the sum when $X = \emptyset$, but what happens when $X \neq \emptyset$?

To begin with, what does s contribute to N_r when $r > 0$? It contributes nothing to some terms and contributes 1 to those terms in N_r for which $s \in S_{i_1} \cap \cdots \cap S_{i_r}$. By the definition of X, this happens if and only if $\{i_1, \ldots, i_r\} \subseteq X$ Thus s contributes to precisely those terms of N_r that correspond to subsets of X. Since there are $\binom{|X|}{r}$ such terms, s contributes $\binom{|X|}{r}$ to N_r. This is 0 if $r > |X|$. Thus the contribution of s to (4.3) is

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

By the binomial theorem, this sum is $(1-1)^{|X|} = 0^{|X|}$, which is zero when $|X| > 0$.

A different proof using "characteristic functions" is given in Exercise 4.1.10. □

Example 4.5 Derangements Recall that a derangement of $\underline{n}$ is a permutation f such that $f(x) = x$ has no solutions; i.e., the permutation has no cycles of length 1. A cycle of length 1 is also called a "fixed point." Let D_n be the number of derangements of $\underline{n}$. What is the value of D_n?

Let the set S of objects be all permutations of $\underline{n}$ and, for $1 \leq i \leq n$, let S_i be those permutations having i as a fixed point. In other words, the larger problem is counting all permutations. If σ is a permutation, condition C_i states that $\sigma(i) = i$.

The set $S_{i_1} \cap \cdots \cap S_{i_r}$ consists of those permutations for which the r elements of $I = \{i_1, \ldots, i_r\}$ are fixed points. Since such permutations can be thought of as permutations of $\underline{n} - I$, there are $(n-r)!$ of them. Thus $N_r = \binom{n}{r}(n-r)! = n!/r!$. By (4.3),

$$D_n \;=\; n! \sum_{i=0}^{n} \frac{(-1)^i}{i!}. \tag{4.10}$$

We will use the following theorem from calculus to obtain a simple approximation to D_n.

Theorem 4.2 Alternating series *Suppose that $|b_0| \geq |b_1| \geq |b_2| \geq \ldots$, that the values of b_k alternate in sign and that $\lim_{k\to\infty} b_k = 0$. Then $\sum_{k=0}^{\infty} b_k$ converges and*

$$\left| \sum_{k=0}^{\infty} b_k - \sum_{k=0}^{n} b_k \right| \;\leq\; |b_{n+1}|.$$

The terms in (4.10) alternate in sign and decrease in magnitude. Since

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \;=\; \frac{1}{e},$$

it follows that D_n differs from $n!/e$ by at most $\frac{n!}{(n+1)!} = \frac{1}{n+1}$. Hence, for $n > 1$, D_n is the closest integer to $n!/e$. □

Exercises

4.1.1. Let us define a "typical four letter word" to be a string of four letters $L_1L_2L_3L_4$ where L_1 and L_4 are consonants and at least one of L_2 and L_3 is a vowel.

(a) For $i = 2$ and $i = 3$, let V_i be the set of sequences $L_1L_2L_3L_4$ where L_1 and L_4 are consonants, L_i is a vowel and the remaining letter is arbitrary. Draw a Venn diagram for the two sets V_2 and V_3. Indicate what part of the diagram corresponds to typical four letter words and calculate the number of such words by using the Rule of Product and the Principle of Inclusion and Exclusion.

(b) For $i = 2$ and $i = 3$, let C_i be the set of sequences $L_1L_2L_3L_4$ where L_1, L_i and L_4 are consonants and the remaining letter is arbitrary. Draw a Venn diagram for the two sets C_2 and C_3. Indicate what part of the diagram corresponds to typical four letter words and calculate the number of such words.

4.1.2. How many ways can n married couples be paired up to form n couples so that each couple consists of a man and a woman and so that no couple is one of the original married couples?

4.1.3. Charled Dodgson (Lewis Carroll) Speaks of a battle among 100 combatants in which 80 lost an arm, 85 a leg, 70 an eye and 75 an ear. (Yes, it's gruesome, but that's the way he stated it.) Some number p of people lost all four.

(a) It is possible that p could be as large as 70? Why?

*(b) Find a lower bound for p and explain how your lower bound could actually be achieved.
Hint. A key to getting a lower bound is to realize that there are only 100 people.

4.1.4. How many ways can we make an n-card hand that contains at least one card from each each of the 4 suits?
Hint. Let a property of a hand be the absence of a suit.

4.1.5. Let $\varphi(N)$ be the number of integers between 1 and N inclusive that have no factors in common with N. Thus $\varphi(1) = \varphi(2) = 1$, $\varphi(3) = \varphi(4) = \varphi(6) = 2$ and $\varphi(5) = 4$. φ is called the *Euler phi function*. Let $p_1, \ldots, p_n$ be the primes that divide N. For example, when $N = 300$, the list of primes is $2, 3, 5$. Let S_j be the set of $x \in \underline{N}$ such that x is a divisible by p_j, or, equivalently, the jth property is that p_j divides the number.

(a) Prove that (4.3) determines $\varphi(N)$.

(b) Prove

$$|S_{i_1} \cap \cdots \cap S_{i_r}| = \frac{N}{p_{i_1} \cdots p_{i_r}}.$$

(c) Use this to prove $\varphi(N) = N \prod_{k=1}^{n} \left(1 - \frac{1}{p_k}\right)$.

4.1.6. Call an $n \times n$ matrix A of zeroes and ones *bad* if there is an index k such that $a_{k,i} = a_{i,k} = 0$ for $1 \le i \le n$. In other words, the row and column passing through (k, k) consist entirely of zeroes. Let $g(n)$ be the number of $n \times n$ matrices of zeroes and ones which are *not* bad.

(a) For any subset K of $\underline{n}$, let $z(K)$ be the number of $n \times n$ matrices A of zeroes and ones such that $a_{i,j} = 0$ if either i or j or both belong to K. Explain why $z(K)$ depends only on $|K|$ and obtain a simple formula for $z(K)$. Call it z_k where $k = |K|$.

(b) Express $g(n)$ as a fairly simple sum in terms of z_k.

4.1.7. Let C be the multiset $\{c_1, c_1, c_2, c_2, \ldots, c_m, c_m\}$ containing two copies each of m distinct symbols. How many ways can the elements of S be arranged in an ordered list so that adjacent symbols are distinct.
Hint. A list in which c_i and c_i are adjacent can be thought of as a list made from the multiset

$$C \cup \{c_i^2\} - \{c_i, c_i\},$$

where c_i^2 is a new symbol that stands for c_ic_i in the list.

*4.1.8. This is the same as the previous exercise, except that now each of c_1 through c_m appears in C three times instead of twice. The constraint is still the same: Adjacent symbols must be distinct.
Hint. There are now two types of properties, namely c_ic_i appearing in the list and $c_ic_ic_i$ appearing in the list. Call the corresponding sets S_i^2 and S_i^3. In computing N_r you need to consider how many times you require an S_i^3 and how many times you require an S_j^2 as well as S_j^3.

4.1.9. Let $(S, \Pr)$ be a probability space and let $S_1, \ldots, S_m$ be subsets of S. Prove that

$$\Pr((S_1 \cup \cdots \cup S_m)^c) \;=\; \sum_{i=0}^{m} (-1)^i N_i \quad \text{where} \quad N_r \;=\; \sum \Pr(S_{i_1} \cap \cdots \cap S_{i_r}),$$

the sum ranging over all r-long strictly increasing sequences chosen from $\underline{m}$.

4.1.10. The goal of this exercise is to use "characteristic functions" to prove Theorem 4.1 (p. 95). Let $\chi_i : S \to \{0, 1\}$ be the characteristic function of S_i; that is,

$$\chi_i(s) \;=\; \begin{cases} 1 & \text{if } s \in \S_i \\ 0 & \text{if } s \notin S_i. \end{cases}$$

(a) Explain why the number we want in the theorem is $\displaystyle\sum_{s \in S} \prod_{i=1}^{m} \bigl(1 - \chi_i(s)\bigr)$.

(b) Prove that

$$\prod_{i=1}^{m} \bigl(1 - \chi_i(s)\bigr) \;=\; \sum_{I \subseteq \underline{m}} (-1)^{|I|} \prod_{i \in I} \chi_i(s).$$

(c) Complete the proof of Theorem 4.1.

4.1.11. We want to count the number of elements in exactly k of the S_i. Let K^c be the complement of K relative to $\underline{m}$; that is, $K^c = \underline{m} \backslash K$.

(a) Explain why the number we want is

$$\sum_{s \in S} \sum_{\substack{K \subseteq \underline{m} \\ |K| = k}} \Biggl(\prod_{i \in K} \chi_i(s) \Biggr) \Biggl(\prod_{i \in K^c} \bigl(1 - \chi_i(s)\bigr) \Biggr).$$

(b) Show that this expression is

$$\sum_{s \in S} \sum_{\substack{K \subseteq \underline{m} \\ |K| = k}} \sum_{J \subseteq J^c} (-1)^{|J|} \prod_{i \in J \cup K} \chi_i(s).$$

(c) Show that this equals

$$\sum_{s \in S} \sum_{L \subseteq \underline{m}} \sum_{\substack{K \subseteq L \\ |K| = k}} (-1)^{|L|-k} \Biggl(\prod_{i \in L} \chi_i(s) \Biggr) \;=\; \sum_{s \in S} \sum_{L \subseteq \underline{m}} (-1)^{|L|-k} \Biggl(\prod_{i \in L} \chi_i(s) \Biggr) \binom{|L|}{k}.$$

(d) Conclude that the number of elements in S that belong to exactly k of the S_i is

$$\sum_{\ell=k}^{m} (-1)^{\ell-k} \binom{\ell}{k} N_\ell.$$

*Bonferroni's Inequalities

We conclude this section by looking briefly at two more advanced topics related to the Principle of Inclusion and Exclusion: Bonferroni's Inequalities and partially ordered sets.

Theroem 4.1 can sometimes be a bit of a problem to use even after we've formulated our problem and know exactly what we must count. There are two reasons for this. First, there will be a lot of addition and subtraction to do if m is large. Second, it may be difficult to actually compute values of N_r so we may have to be content with estimating them for small values of r and ignoring them when r is large. Because of these problems, we may prefer to obtain a quick approximation to (4.3). A method that is frequently useful for doing this is provided by the following theorem.

Theorem 4.3 Bonferroni's inequalities *Let the notation be the same as in Theorem 4.1 (p. 95) and let E be the number of elements of S not in any of the S_i. Then*

$$-N_t \;\le\; E - \sum_{r=0}^{t-1}(-1)^r N_r \;\le\; N_t;$$

i.e., truncating the sum gives an error which is no larger than the first term that was neglected. Furthermore, the sum is either an overestimate or an underestimate according as t is odd or even, respectively.

We can't prove this simply by appealing to Theorem 4.2 (p. 99) because the terms may be increasing in size. The proof of Bonferroni's Inequalities is left as an exercise.

Example 4.6 Using the theorem Let $r(n,k)$ be the fraction of those functions in $\underline{k^n}$ which are surjections. Using Bonferroni's inequalities and the ideas in Example 4.3 (p. 97), we'll estimate $r(n,k)$.

Let's begin with $t = 2$ in the theorem. In that case we simply need to divide the $i = 0$ and $i = 1$ terms in (4.5) by the total number of functions. Thus

$$r(n,k) \;\ge\; \frac{k^n - k(k-1)^n}{k^n} \;=\; 1 - k(1-1/k)^n.$$

With $k = 10$ and $n = 40$, we see that at least 85.2% of the functions in $\underline{10^{40}}$ are surjections.

If we set $t = 3$ in Bonferroni's inequalities, we obtain the upper bound

$$r(n,k) \le \frac{k^n - k(k-1)^n + \binom{k}{2}(k-2)^n}{k^n} \;=\; 1 - k(1-1/k)^n + \binom{k}{2}(1-2/k)^n.$$

With $k = 10$ and $n = 40$, we see that at most 85.8% of the functions in $\underline{10^{40}}$ are surjections. □

*Partially Ordered Sets

There is an important generalization of the Principle of Inclusion and Exclusion (Theorem 4.1 (p. 95)) which we'll just touch on. It requires some new concepts.

A *binary relation* ρ on a set S is a subset of $S \times S$. Instead of writing $(x, y) \in \rho$, people write $x\rho y$. For example, if S is a set of integers, then we can let ρ be the set of all $x, y \in S$ with x less than y. Thus, $x\rho y$ if and only if x is less than y. People usually use the notation $<$ for this binary relation. As another example, we can let S be the set of all subsets of $\underline{n}$ and let $\subseteq$ be the binary relation.

We can describe equivalence relations as binary relations: Let $\sim$ be an equivalence relation on S. Those pairs (x, y) for which $x \sim y$ form a subset of $S \times S$.

We now define another important binary relation.

Definition 4.1 Partially Ordered Set *A set P and a binary relation ρ satisfying*

(P-1) *$x\rho x$ for all $x \in P$;*

(P-2) *if $x\rho y$ and $y\rho x$, then $x = y$; and*

(P-3) *if $x\rho y$ and $y\rho z$, then $x\rho z$*

is called a **partially ordered set**, *also called a poset. The binary relation ρ is called a* **partial order**.

The real numbers with $x\rho y$ meaning "x is less than or equal to y" is a poset. The subsets of a set with $x\rho y$ meaning $x \subseteq y$ is a poset. Because of these examples, people often use the symbol $\leq$ or the symbol $\subseteq$ in place of ρ, even when the partial order does not involve numbers or subsets.

We now return to Theorem 4.1 and begin by rewriting the terms in (4.2) as functions of sets: $f(\{i_1, \ldots, i_r\}) = |S_{i_1} \cap \cdots \cap S_{i_r}|$. How should we define $f(\emptyset)$? It should be the size of the empty intersection. In most situations, the best choice for the empty intersection is everything. Thus we should probably take $f(\emptyset) = |S|$, the size of the set that contains everything.

Many people find this choice for the empty intersection confusing, so we digress briefly to explain it. (If it does not confuse you, skip to the next paragraph.) Let's look at something a bit more familiar—summations. As you know, the value of

$$g(A) \;=\; \sum_{i \in A} h(i) \tag{4.11}$$

is defined to be the sum of $f(i)$ over all $i \in A$. You should easily see that if A and B are disjoint nonempty sets, then

$$g(A \cup B) \;=\; g(A) + g(B). \tag{4.12}$$

If we want this to be true for $B = \emptyset$, we must have

$$g(A) + g(\emptyset) \;=\; g(A \cup \emptyset) \;=\; g(A), \tag{4.13}$$

and so $g(\emptyset)$ must equal 0, the identity element for addition. Suppose we replace the sum in (4.11) with a product. Then (4.12) becomes $g(A \cup B) = g(A)g(B)$ and the parallel to (4.13) gives $g(A)g(\emptyset) = g(A)$. Thus $g(A)$ should be 1, the identity for multiplication. Instead of g and h being numerically valued functions, they could be set valued functions and we could replace the summation in (4.11) with either a set union or a set intersection. (In terms of the previous notation, we would write A_i instead of $h(i)$.) Then $g(\emptyset)$ would be taken to be the identity for set union or set intersection respectively; that is, either $g(A) \cup g(\emptyset) = g(A)$ or $g(A) \cap g(\emptyset) = g(A)$. This leads to $g(\emptyset) = \emptyset$ and $g(\emptyset) = S$, respectively.

Let's recap where we were before our digression: We defined

$$f(A) \;=\; \left|\bigcup_{i\in A} S_i\right|$$

for $A \neq \emptyset$ and $f(\emptyset) = |S|$. Now, N_r is simply the sum of $f(R)$ over all subsets R of $\underline{m}$ of size r; i.e.,

$$N_r \;=\; \sum_{\substack{R\subseteq \underline{m}\\ |R|=r}} f(R).$$

We can rewrite (4.3) in the form

$$\sum_{R\subseteq \underline{m}} (-1)^{|R|} f(R).$$

In words, we can describe $f(\{i_1,\ldots,i_r\})$ as the number of things that satisfy conditions $i_1,\ldots,i_r$ and possibly others. In a similar manner, we could define a function $e(\{i_1,\ldots,i_r\})$ to the number of things that satisfy condition $i_1,\ldots,i_r$ and none of the other conditions in the collection $1,\ldots,m$. In these terms, (4.3) is a formula for $e(\emptyset)$. Also, by the definitions of e and f, we have $f(R) = \sum_{Q\supseteq R} e(Q)$.

We state without proof a generalization of the Principle of Inclusion and Exclusion.

Theorem 4.4 *Let P be the partially ordered set of subsets of $\underline{k}$. For any two functions e and f with domain P*

$$f(x) \;=\; \sum_{y\supseteq x} e(y) \quad \textit{for all } x\in P \tag{4.14}$$

if and only if

$$e(x) \;=\; \sum_{y\supseteq x} (-1)^{|y|-|x|} f(y) \quad \textit{for all } x\in P. \tag{4.15}$$

This result can be extended to any finite partially partially ordered set P if $(-1)^{|y|-|x|}$ is replaced by a function $\mu(x,y)$, called the Möbius function of the partially ordered set P. We will not explore this.

Exercises

4.1.12. This exercise extends the Principle of Inclusion and Exclusion (4.3). Let E_k be the number of elements of S that lie in exactly k of the sets $S_1,\ S_2,\ \ldots,\ S_m$. Prove that

$$E_k \;=\; \sum_{i=0}^{m-k} (-1)^i \binom{k+i}{i} N_{k+i}. \tag{4.16}$$

4.1.13. The purpose of this exercise is to prove Bonferroni's inequalities.

(a) Prove the inequalities are equivalent to the statement that sums

$$c_t(X) \;=\; \binom{|X|}{0} - \binom{|X|}{1} + \binom{|X|}{2} - \cdots \pm \binom{|X|}{t}$$

alternate in sign until eventually becoming 0 for $t \geq |X|$.

(b) Prove $c_t(X) \;=\; (-1)^t \binom{|X|-1}{t}$ and so complete the proof.

4.1.14. Find a formula that bears the same relation to Bonferroni's inequalities that (4.4) bears to (4.3); i.e., find inequalities for approximations to $|S_1\cup\cdots\cup S_m|$ rather than for approximations to E.

4.1.15. Consider the following algorithm due to H. Wilf.

Initialize: Let the N_i be defined as in (4.3).
Loop: Execute the following code.

```
For j = 0, 1, ..., m − 1
    For i = m − 1, m − 2, ..., j
        N_i = N_i − N_{i+1}
    End for
End for
```

The loop on i means that i starts at $m-1$ and decreases to j.

(a) By carrying out the algorithm for $m = 2$ and $m = 3$, prove that N_i is replaced by E_i, where E_i is given by Exercise 4.1.12 for these values of m.

(b) We can rephrase the algorithm in a set theoretic form. Replace N_r by N_r^*, the multiset which contains each $s \in S$ as many times as there are $1 \le i_1 < i_2 < \ldots < i_r \le m$ such that

$$s \in S_{i_1} \cap \ldots \cap S_{i_r}. \tag{4.17}$$

By the definition of N_r, it follows immediately that $N_r = |N_r^*|$. Similarly, replace E_k by E_k^*, the set of those elements of S that belong to exactly k of the sets $S_1, \ldots, S_m$. Replace N_i by N_i^* in the algorithm and interpret $N_i^* - N_{i+1}^*$ to be the multiset that contains $s \in S$ as many times as it appears in N_i^* minus the number of times it appears in N_{i+1}^*. We claim that the algorithm now stops with N_i^* replaced by E_i^*. Prove this for $m = 2$ and $m = 3$.

*(c) Using induction on t, prove the set theoretic form of the algorithm by proving that after t iterations of the loop on j; i.e., $j = 0, \ldots, t-1$ the following is true. If $s \in S$ appears in exactly p of the sets $S_1, \ldots, S_m$, then it appears in N_r^* with multiplicity

$$\mu(p,r,t) \;=\; \begin{cases} \binom{p-t}{r-t}, & \text{if } t \le p; \\ 1, & \text{if } t > p \text{ and } r = p; \\ 0, & \text{if } t > p \text{ and } r \ne p. \end{cases} \tag{4.18}$$

Also prove that no $s \in S$ ever appears more times in an N_{r+1}^* than it does in an N_r^* when we are calculating $N_r^* - N_{r+1}^*$.

(d) Prove that the validity of the set theoretic form of the algorithm implies the validity of the numerical form of the algorithm.
Hint. Use the last sentence in (c).

4.1.16. Let $D_n(k)$ be the number of permutations of $\underline{n}$ that have exactly k fixed points. Thus $D_n(0) = D_n$, the number of derangements of $\underline{n}$.

(a) Use Exercise 4.1.12 to obtain a formula for $D_n(k)$.

(b) Give a simple, direct combinatorial proof that $D_n(k) = \binom{n}{k} D_{n-k}$.

(c) Using algebra and (4.10), prove the answers in (a) and (b) are equal.

4.1.17. Let $A = \{a_1, \ldots, a_m\}$ be a set of m integers, all greater than 1. Let $d(n,k,A)$ be the number of integers in $\underline{n}$ that are divisible by exactly k of the integers in A.

(a) Assuming that the elements of A are distinct primes all dividing n, obtain a formula for $d(n,k,A)$ by using Exercise 4.1.12. Specialize this formula to obtain a formula for the Euler phi function $\varphi(n)$ discussed in Exercise 4.1.5.

(b) Relax the constraints in (a) by replacing the assumption that the elements in A are primes by the assumption that no two elements in A have a common factor.

(c) Relax the constraints in (a) further by not requiring that the elements of A divide n.

(d) Can you relax the constraints in (a) still further by making no assumptions about A and n except that A consists of m integers greater than 1?

4.1.18. Explain why the real numbers with $x \rho y$ meaning "x is less than y" is not a poset.

4.1.19. Prove that the following are posets.

(a) The real numbers with $x\rho y$ meaning "x is less than or equal to y."

(b) The real numbers with $x\rho y$ meaning "x is greater than or equal to y."

(c) The subsets of a set with $x\rho y$ meaning $x \subseteq y$.

(d) The positive integers with $x\rho y$ meaning y/x is an integer.

4.1.20. Prove that if (S, ρ) is a poset then so is (S, τ) where $x\tau y$ if and only if $y\rho x$.

4.1.21. Let S be the set of all partitions of $\underline{n}$. If $x, y \in S$, write $x\rho y$ if and only if every block of y is a union of one or more blocks of x. For example, $\{1,2\},\{3\},\{4\}\rho\{\{1,2,4\},\{3\}\}$. Prove that this is a poset.

4.1.22. Suppose that (R, ρ) and (T, τ) are posets. Prove that $(R \times T, \pi)$ is a poset if $(r, s)\pi(r', s')$ means that both $r\rho r'$ and $t\tau t'$ are true.

4.1.23. We will deduce the result in Exercise 4.1.12 as a consequence of the partially ordered set extension of the Principle of Inclusion and Exclusion.

(a) We look at subsets y of $\{1, 2, \ldots, m\}$. Let $e(y)$ be the number of elements in S that belong to every S_i for which $i \in y$ and to none of the S_j for which $j \notin y$. Prove that E_k is the sum of $e(y)$ over all y of size k.

(b) Prove that if $f(x)$ is defined by (4.14), then

$$f(x) \;=\; \Bigl|\bigcap_{i\in x} S_i\Bigr|.$$

(c) Conclude that (4.15) implies (4.16).

4.2 Listing Structures with Symmetries

By using decision trees, introduced in Chapter 3, we can produce our list C of canonical representatives. There are many ways to go about it. We'll illustrate this by some examples. Many of the examples are based on the Ferris wheel problem of Example 1.12 (p. 13): How many distinct six long circular sequences of ones and twos are there?

Example 4.7 A straightforward method One approach to the Ferris wheel problem is to simply generate all sequences and reject those that are equivalent to an earlier one in the lex order. For example, we would reject both 121121 and 211211 because they are equivalent to 112112, which occurs earlier in lex order.

We can reduce the size of the decision tree by being careful; e.g., the sequence that starts 1211 . . . can never be lexically least because we could shift it two positions to get 11 . . . 12.

Even with these ideas, the decision tree is rather large. Hence, we'll shorten the problem we've been considering to sequences of length 4. The decision tree is shown in Figure 4.2. It is simply the tree for generating all functions from $\underline{4}$ to $\underline{2}$ with those functions which have a (lexically) smaller circular shift removed. How did we do the removal? When we decided to begin with 2, there was no possibility of ever choosing a 1—a circular shift would begin with 1 and so be smaller. Also, if any 2's are present, we can never end with a 1 because a circular shift that moved it to the front would produce a smaller sequence. This rule was applied to determine the possible decisions at 112, 121 and 122. This explains everything that's missing from the full tree for $\underline{2}^{\underline{4}}$.

This approach can get rather unwieldy when doing larger problems by hand. Try using it for the six long Ferris wheel. □

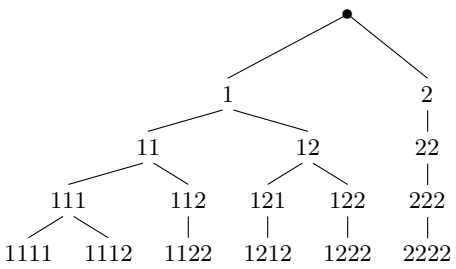


Figure 4.2 A Ferris wheel decision tree.

Example 4.8 Another problem Four identical spheres are glued together so that three of them lie at the vertices of an equilateral triangle and the fourth lies at the center. That is, the centers of the spheres lie in a plane and three of the centers are at the corners of an equilateral triangle while the fourth is in the center. Thus, the sphere arrangement remains unchanged in appearance if it is flipped over about any of three axes or if it is rotated 120 degrees about an axis that passes through the center of the center sphere and is perpendicular to the plane of the centers. Draw yourself a picture to illustrate this—it is very useful to get into the habit of drawing pictures to help visualize problems like this.

We have four tiny identical red balls and four tiny identical green balls. The balls are to be placed in the spheres so that each sphere contains exactly two balls. How many arrangements are possible?

The calculation can be done with the help of a decision tree. The first decision could be the number of red balls to be placed in the center sphere. If no red balls are placed in the center sphere, then two green balls must be placed there and two in the outer spheres. Those two in the outer spheres can either be placed in the same sphere or in different spheres. Proceeding in this sort of way, we can construct a decision tree. You should do this, verifying that exactly six distinct arrangements are possible. □

Example 4.9 A subtler method Another approach to the Ferris wheel problem is to take into account some of the effects of the symmetry when designing the decision tree. Let's look at our six long Ferris wheel.

The basic idea is to look at properties that depend only on the circular sequence rather than on how we have chosen to write it as a list. Unlike the simpler approach of listing things in lex order, there are a variety of choices for constructing the decision tree. As a result, different people may construct different decision trees. Constructing a good tree may take a fair bit of thought. Is it worth the effort? Yes, because a good decision tree may be considerably smaller than one obtained by a more simplistic approach.

Before reading further in this example, construct a simple lex order decision tree like the one in Figure 4.2, but for the six long Ferris wheel.

Since the number of ones in a sequence remains the same, we can partition the problem according to the number of ones that appear in the 6 long sequence. Thus our list of possible first decisions could be

more 1's than 2's, three of each, more 2's than 1's.

We can save ourselves some work right away by noting that all the sequences that arise from the third choice can be obtained from those of the first choice by replacing 1's with 2's and 2's with 1's.

What should our next decisions be? We'll do something different. Define a function s on sequences with $s(x)$ equal to the *minimal* amount the sequence x must be circularly shifted to obtain x again. (This is called the "period" of the circular sequence.) Thus $s(111111) = 1$, $s(121121) = 3$ and $s(122221) = 6$. Note that if x and y are equivalent, then $s(x) = s(y)$. You should convince yourself

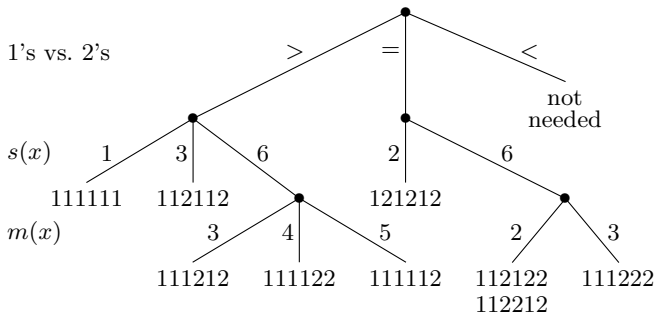


Figure 4.3 Another Ferris wheel decision tree.

that x consists of $6/s(x)$ copies of the first $s(x)$ elements of x. As a result, the only possible values of s are 1, 2, 3 and 6, and the ratio of 1's to 2's in x is the same as in the first $s(x)$ entries.

In this paragraph, we consider the case in which the number of 1's and 2's in x are equal. By the above, $s(x)$ must be even. If $s(x) = 2$, we need three repeats of a two long pattern that contains a 1 and a 2. We can take x to be either 121212 or 212121. We'll adopt our usual convention of using the lexically least equivalent sequence for the canonical sequence, so $x = 121212$. If $s(x) = 6$, the situation is more complex and so another decision will be used to break this case down further. Again, many choices are possible. We'll use $m(x)$, the length of the longest string of consecutive ones in the sequence x. Remember to read the list circularly, so $m(121211) = 3$. Put this sequence at the start of the list. A little thought should convince you that $m(x) = 1$ implies $x = 121212$, which does not have $s(x) = 6$. Since there are three 1's and three 2's, $m(x)$ must be 2 or 3. If $m(x) = 3$, we have $x = 111222$. If $m(x) = 2$, we have $x = 112??2$. The questionable entries must be a one and a two. Either order works giving us the two sequences 112122 and 112212.

A similar analysis to that in the previous paragraph can be used for the case in which there are more 1's than 2's. We leave it to you to carry out the analysis for this case.

The decision tree we have developed is shown in Figure 4.3. Compare its size with that of the simple lex order tree you were asked to construct. Construct another decision tree using a different sequence of decisions than we did in this example. Your goal should be to come up with something different from Figure 4.3 that is about the same size as it is. □

It is helpful to understand better the difference between the two methods we've used for the Ferris wheel problem. Our first method is a straightforward pruning of the decision tree for listing all functions in lex order. When several functions correspond to the same structure, we retain only the lexicographically least one of them. The simplicity of the method makes it fairly easy to program. Unfortunately it can lead to a rather large tree, often containing many decisions that lead to no canonical solutions. Thus, although it is straightforward, using it for hand calculation may lead to errors because of the amount of work involved.

Our second method requires some ingenuity. The basic idea is to select some feature of the problem that lets us break it into smaller problems of the same basic type but with additional conditions. Let's look at what we did. First we divided the problem into three parts depending on the number of ones versus the number of twos. Each part was a problem of the same type; e.g., how many different arrangements are there *where each arrangement has more ones than twos.* Isn't this what we did in the first method when we made a decision like "the first element of the sequence is 1?" No! This is not a problem of the same type because the condition is not invariant under rotation of the sequence. On the other hand, the condition that there be more ones than twos is invariant under rotation.

In our second method, we next chose another property that is invariant under rotation of the sequence: how much we had to rotate the sequence before it looked the same. Next we looked at the

longest consecutive string of ones, with the sequence read circularly so that the first entry follows the last. Again, this is invariant under rotation. Sometimes we did not need to go that far because it was easy to see the solutions; e.g., after $s(x) = 1$, it was clear that 111111 was the only solution. On the other hand, after the decision sequence $=, 6, 2$ in Figure 4.3, it was still not obvious what the answer was. At this time we decided it was easier to shift back to our first method rather than find another property that was invariant under rotation. We did this on scratch paper and simply wrote the result as two solutions in the figure.

We might call the second method the *symmetry invariant* method. Why is symmetry invariance better than the first method? When done cleverly, it leads to smaller decision trees and hence less chance for computational errors. On the other hand, you may make mistakes because it is less mechanical or you may make poor selections for the decision criteria. If you are applying symmetry invariance, how do you decide what properties to select as decision criteria? Also, how do you decide when to switch back to the first method? There are no rules for this. Experience is the best guide.

Example 4.10 Listing necklaces We'll work another example using symmetry invariance. How many ways can the corners of a regular hexagon be labeled using the labels B, R and W, standing for the colors blue, red and white? Note that an unlabeled hexagon can be rotated 60° and/or flipped over and still look the same. You could imagine this as a hexagon made from wire with a round bead to be placed at each corner. We impose a condition on the finished hexagon of beads:

Adjacent beads must be different colors.

Our first decision will be the number of colors that actually appear. If only two are used, there are only three solutions: BRBRBR, BWBWBW and RWRWRW since adjacent colors must be different. (We've used the same sort of notation we used for the Ferris wheel.) If three colors are used, we decide how many of each actually appear. The possibilities are 2,2,2 and all six permutations of 1,2,3. To do the latter, we need only consider the case of 1 blue, 2 red and 3 white and then permute the colors in our solutions in all six possible ways. (To count, we simply multiply this case by 6.)

Let's do the 1 blue, 2 red and 3 white case by the first method. A canonical sequence must start with B and be followed somehow by 2 R's and 3 W's so that adjacent letters are different. Call the sequence associated with the hexagon $\mathrm{B}x_2x_3x_4x_5x_6$. There are no solutions with $x_2 = \mathrm{R}$ or with $x_6 = \mathrm{R}$ because we are not allowed to have two W's adjacent. Thus $x_2 = x_6 = \mathrm{W}$. We easily obtain the single solution BWRWRW. Remember that this gives six solutions through permutation of colors.

The case in which each color is used twice remains to be done. We make a decision based on whether the two B's are opposite each other or not on the hexagon. We use the first method now. The case of opposite B's leads to the sequence $\mathrm{B}x_2x_3\mathrm{B}x_5x_6$ and the other case leads to $\mathrm{B}y_2\mathrm{B}y_4y_5y_6$. In the first case, x_2 and x_3 are different. This leads to two lexically least sequences: BRWBRW and BRWBWR. (The sequence BWRBWR is just BRWBRW flipped over.) In the second case, choosing y_2 determines the remaining y's. The two results are BRBWRW and BWBRWR.

Adding up our results, there are $3 + 6 \times 1 + (2 + 2) = 13$ solutions. □

Exercises

4.2.1. Redo Example 4.10 using only the first (mechanical) method.

4.2.2. How many ways can the eight corners of a regular octagon be labeled using the labels B and W. Note that an unlabeled octagon can be rotated 45° and/or flipped over and still look the same.

4.2.3. Let $F(r)$ the number of ways to place beads at the vertices of a square when we are given r different types of round beads. Let $f(r)$ be the same number except that at least one bead of each of the r types must be used. Rotations and reflections of the square are allowed as with the hexagon in Example 4.10.

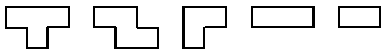
(a) Prove that $F(r) = \binom{r}{1}f(1) + \binom{r}{2}f(2) + \binom{r}{3}f(3) + \binom{r}{4}f(4)$

(b) By evaluating $f(r)$ for $r \leq 4$, obtain an explicit formula for $F(r)$.

4.2.4. State and prove a generalization of the formula in the previous exercise that expresses $F(r)$ in terms of the function f. Possible generalizations are to the hexagon and the n-gon. Can you find a generalization that has little or no connection with symmetries?

4.2.5. We want to list the coverings of a 4×4 board by 8 dominoes, where solutions that differ only by a rotation and/or reflection of the board are considered to be the same. For example, Figure 3.15 shows 11 ways to cover a 3×4 board. With rotation and/or reflection, only 5 are distinct. The lex order minimal descriptions of the distinct ones are *hhhhhh*, *hhhvvh*, *hhvhvh*, *hhvvv* and *hvvvvh*. List the lexically least coverings of the 4×4 board; i.e., our standard choices for canonical representatives.

4.2.6. Draw a decision tree for covering the 4×4 board using one each of the shapes shown below. Two boards are equivalent if one can be transformed to the other by rotations and/or reflections.
Hint. First place the "T" shaped piece.



4.2.7. This problem is concerned with listing colorings of the faces of a cube. Unless you are very good at visualizing in three dimensions, we recommend that you have a cube available to manipulate. (Even a sugar cube could be used.) Also, when listing solutions, you may find it convenient to represent the cube in the plane by "unfolding" it as shown and writing the colors in the squares. The line of four faces can be thought of as the four sides and the other two faces can be thought as the top and bottom.

(a) List and count the ways to color the faces using at most the 2 colors black and white.

(b) List and count the ways to color the faces using at most the two colors black and white, with the added condition that we do not distinguish between a cube and its color negative (interchanging black and white).

(c) List and count the ways to color the faces using at most the two colors black and white, with the added condition we cannot distinguish between a cube and its mirror image.

(d) List and count the ways to color the faces using all of the colors black, red and white; i.e., every color must appear on each cube.

(e) Count the ways to color the faces using the colors black, red and white. On any given cube, all colors need not appear.

(f) Find a formula $F(r)$ for the number of ways to color the faces of a cube using r colors so that whenever two faces are opposite one another they are colored differently.
Hint. See Exercise 4.2.4.

4.2.8. We say that f is a "Boolean" function if $f\colon \{0,1\}^n \to \{0,1\}$.

(a) Prove that a Boolean function with $n = 2$ can be thought of as placing zeroes and ones at the corners of a 1×1 square with lower left corner at the origin. Give a similar interpretation for $n = 3$ using a cube.

(b) We want to count the number of "different" Boolean functions with $n = 2$ and $n = 3$. Two functions will be considered equivalent if one can be obtained from the other by permuting the arguments and/or complementation. We can describe this precisely in an algebraic fashion by saying that $f, g\colon \{0,1\}^n \to \{0,1\}$ are equivalent if and only if there is a permutation σ of $\underline{n}$ and $c, d_1, \ldots, d_n \in \{0,1\}$ such that

$$f(x_1, \ldots, x_n) \;=\; c \oplus g(x_{\sigma(1)} \oplus d_1, \ldots x_{\sigma(n)} \oplus d_n),$$

where $u \oplus v$ is $u + v$ unless $u = v = 1$ in which case $u \oplus v = 0$. ("Exclusive or" and "mod 2 sum" are other names for $u \oplus v$.) For $n = 2$, there are four different Boolean functions:

$$\big(f(0,0), f(0,1), f(1,0), f(1,1)\big) \;=\; \begin{cases} (0,0,1,1), & (0,1,0,1), \\ (0,1,1,1), & (1,1,1,,1). \end{cases}$$

Interpret the equivalence of Boolean functions in terms of symmetries involving the square and cube when $n = 2, 3$.

(c) List the different Boolean functions when $n = 3$.

4.3 Counting Structures with Symmetries

We've been using "equivalent" rather loosely without saying what it means. Since ambiguous terms provide an easy way to make errors, we should define it.

Definition 4.2 Equivalence *An* **equivalence relation** *on a set S is a partition of S. We say that $s, t \in S$ are* **equivalent** *if and only if they belong to the same block of the partition. If the symbol $\sim$ denotes the equivalence relation, then we write $s \sim t$ to indicate that s and t are equivalent. An* **equivalence class** *is a subset of S consisting of all objects in the set that are equivalent to some object; i.e., an equivalence class is a block of the partition.*

Returning to our circular sequence problem, what do the equivalence classes look like? First, 111111 is in a class by itself because all rotations give us the same sequence again. Likewise, 222222 is in a class by itself. The sequences {121212, 212121} is a third equivalence class. The sequences 112112 and 122122 are in different equivalence classes, each of which contains 3 sequences. So far, we have 5 equivalence classes containing a total of 10 sequences. What about the remaining $2^6 - 10 = 54$ sequences? It turns out that they fall into 9 equivalence classes of 6 sequences each. Thus there are $5 + 9 = 14$ equivalence classes; that is, the answer to our circular sequence problem is 14.

This method is awkward for larger problems. You might try to do 12 long circular sequences of ones, twos and threes, where the answer is 44,368. Burnside's Lemma allows us to do such problems more easily. In order to state and prove it we need some observations about the symmetries.

In our problem the symmetries are rotations of the Ferris wheel through 0°, 60°, 120°, 180°, 240° and 300°. These correspond to reading a sequence circularly starting with the first, second, ... and sixth positions, respectively. Let S be the set of all six long sequences of zeroes and ones. The six symmetries correspond to six permutations of S by means of the circular reading. For example, if g_i is the permutation that starts reading in position $i + 1$, then $g_1(111122) = 111221$ and $g_3(111122) = 122111$. Note that $g_0(x) = x$ for all x. Alternatively, we can think of g_i as shifting the sequence "circularly" to the left by i positions. The set $G = \{g_0, \ldots, g_5\}$ has some important properties, namely

(G-1) There is an $e \in G$ such that $e(x) = x$ for all $x \in S$.

(G-2) If $f \in G$, then the inverse of f exists and $f^{-1} \in G$.

(G-3) If $f, g \in G$, then the composition fg is in G.

The function e is called the "identity" and e is reserved for its name. You should be able to verify that $g_i^{-1} = g_{5-i}$ and $g_i g_j = g_k$, where $k = i + j$ if this is less than 6 and $k = i + j - 6$ otherwise. Any set of permutations with properties (G-1), (G-2) and (G-3) is called a *permutation group*. Group theory is an important subject that is part of the branch of mathematics called algebra. We barely touch on it here.

Symmetries always lead to permutation groups. Why? First, recall that a symmetry of some thing is a rearrangement that leaves the thing looking the same (in our case, the thing is the empty Ferris wheel). Taking the inverse corresponds to reversing the motion of the symmetry, so it again leaves the thing looking the same. Taking a product corresponds to one symmetry followed by another and so leaves the thing looking the same.

What is the connection between the equivalence classes and the permutation group for the sequences? It is simple: Two sequences $x, y \in S$ are equivalent if and only if $y = g(x)$ for some $g \in G$. In general, a group G of permutations on a set S defines an equivalence relation in this way. That requires a bit of proof, which we give at the end of the section. We can now state Burnside's Lemma, but we defer its proof until the end of the section. In this theorem the expression $\sum_{g \in G} N(g)$ appears. For those unfamiliar with such notation, it means that we must add up the values of $N(g)$ for all $g \in G$. The order in which we add them does not matter since order is irrelevant in addition.

Theorem 4.5 Burnside's Lemma *Let S be a set with a permutation group G. The number of equivalence classes that G defines on S is*

$$\frac{1}{|G|} \sum_{g \in G} N(g),$$

where $N(g)$ is the number of $x \in S$ such that $g(x) = x$.

Example 4.11 The Ferris wheel generalized We'll redo the Ferris wheel problem and generalize it.

Burnside's Lemma tells us that the answer to the Ferris wheel problem is

$$\tfrac{1}{6}\big(N(g_0) + N(g_1) + N(g_2) + N(g_3) + N(g_4) + N(g_5)\big).$$

Let's compute the terms in the sum. $N(g_0) = 2^6$ since $g_0 = e$, the identity, and there are two choices for each of the six positions in the sequence. What is $N(g_1)$? If $x = x_1x_2x_3x_4x_5x_6$, then $g_1(x) = x_2x_3x_4x_5x_6x_1$. Since we want $g_1(x) = x$, we need $x_1 = x_2$, $x_2 = x_3$, ... and $x_6 = x_1$. In other words, all the x_i's are equal. Thus $N(g_1) = 2$. Since $g_2(x) = x_3x_4x_5x_6x_1x_2$, we find that $x_1 = x_3 = x_5$ and $x_2 = x_4 = x_6$. Thus $N(g_2) = 2^2 = 4$. You should be able to prove that $N(g_3) = 8$, $N(g_4) = 4$ and $N(g_5) = 2$. Thus the number of equivalence classes is $\frac{1}{6}(64 + 2 + 4 + 8 + 4 + 2) = 14$.

Now suppose that instead of placing just ones and twos in circular sequences, we have k symbols to choose from. Our work in the previous paragraph makes it easy for us to write down a formula. Note that for $g_2(x) = x$ we found that $x_1 = x_3 = x_5$ and $x_2 = x_4 = x_6$. Thus we can choose one symbol as the value for $x_1 = x_3 = x_5$ and another (possibly the same) symbol as the value for $x_2 = x_4 = x_6$. Thus $N(g_2) = k^2$. The other $N(g_i)$ values can be determined in a similar manner giving us

$$\tfrac{1}{6}(k^6 + k + k^2 + k^3 + k^2 + k) \;=\; \frac{k(k^5 + k^2 + 2k + 2)}{6}$$

different arrangements. When $k = 2$, we obtain the result from the previous paragraph.

Now let's modify what we just did by adding the requirement that adjacent symbols must be different. In this case, $N(g_1) = 0$ because $g_1(x) = x$ requires that $x_1 = x_2$, which is forbidden. For

$g_3(x) = x$, we have $x_1 = x_4$, $x_2 = x_5$ and $x_3 = x_6$. Since the symbols assigned to $x_1 = x_4$, x_2 and x_3 must all be different, $N(g_3) = k(k-1)(k-2)$. With a bit more work, we find that there are

$$\tfrac{1}{6}\big(N(g_0) + 0 + k(k-1) + k(k-1)(k-2) + k(k-1) + 0\big)$$

equivalence classes. The determination of $N(g_0)$ is a bit more difficult. We'll discuss this type of problem in Section 6.2. The answer is $N(g_0) = (k-1)^6 + k - 1$. Thus the number of equivalence classes is

$$\tfrac{1}{6}(k-1)\big((k-1)^5 + k^2 + 1\big).$$

We can check these calculations a bit by noting that when $k = 1$ there should be no solutions and when $k = 2$ there should be 1. Substitution into our formula does indeed give 0 and 1. □

We can shorten some of the work in the previous example by thinking of permutations a bit differently. Instead of looking at permutations of 6 long sequences, we can look at the way the permutation rearranges the positions of the sequence. For example, $g_2(x_1x_2x_3x_4x_5x_6) = x_3x_4x_5x_6x_1x_2$ can be interpreted as saying position 1 is replaced by position 3, position 2 is replaced by position 4, and so forth. This is a new permutation: Instead of permuting the set of 6 long sequences it permutes the set $\underline{6}$. To emphasize both the difference and the relationship, we use γ, the Greek letter corresponding to g. In cycle form, $\gamma_2 = (1,3,5)(2,4,6)$. Thinking of a sequence as a function $f\colon \underline{6} \to \underline{k}$, the function will be counted by $N(\gamma_2) = N(g_2)$ if and only if it is constant on the cycles of γ_2. This is a general result:

> **Principle** *In the set of allowed functions, $N(\gamma)$ counts precisely those allowed functions which are constant on the cycles of γ; i.e., those functions f such that $f(x) = f(y)$ whenever x and y are in the same cycle of γ.*

For example, if all functions $f\colon A \to B$ are allowed, $N(\gamma)$ is simply $|B|^c$ where c is the number of cycles of γ. In the concluding problem of the last example, not all functions were allowed because the adjacency constraint requires that $f(i) \neq f(i+1)$ for all i. In that case, computing $N(\gamma)$ is a bit harder, but the principle can still be used.

Example 4.12 Counting necklaces How many ways can 4 identical round green beads and 4 identical round red beads be arranged to form a necklace of 8 beads? Due to the nature of the beads, two necklaces will be the same if one can be obtained from the other by rotation or flipping over. The symmetries are like those in Example 4.10 (p. 109) except that we now have 8 positions instead of 6.

We can imagine a necklace as an 8 long sequence and use the idea we just discussed to describe the permutations. Obviously all that matters for counting purposes is the size of the cycles—not what they contain. Altogether there are 16 permutations of $\underline{8}$, which are shown here in 5 classes, where P_k is the set of permutations having k cycles. We've omitted commas separating entries in the cycles. We leave it to you to check that the list is correct and complete.

P_8	(1)(2)(3)(4)(5)(6)(7)(8)		
P_5	(1)(2,8)(3,7)(4,6)(5)	(1,3)(2)(4,8)(5,7)(6)	(1,5)(2,4)(3)(6,8)(7)
	(1,7)(2,6)(3,5)(4)(8)		
P_4	(1,5)(2,6)(3,7)(4,8)	(1,2)(3,8)(4,7)(5,6)	(1,4)(2,3)(5,8)(6,7)
	(1,6)(2,5)(3,4)(7,8)	(1,8)(2,7)(3,6)(4,5)	
P_2	(1,3,5,7)(2,4,6,8)	(1,7,5,3)(2,8,6,4)	
P_1	(1,2,3,4,5,6,7,8)	(1,4,7,2,5,8,3,6)	(1,8,7,6,5,4,3,2)
	(1,6,3,8,5,2,7,4)		

Suppose that $\gamma \in P_8$; i.e., $\gamma = (1)\cdots(8)$. Since each cycle has length 1, we can simply choose 4 cycles to be the green beads. This can be done in $\binom{8}{4} = 70$ ways. Thus $N(\gamma) = 70$. Suppose $\gamma \in P_5$. We can either choose 2 of the 2 cycles to be green beads OR choose both 1 cycles and one of the 2 cycles. Thus $N(\gamma) = \binom{3}{2} + \binom{3}{1} = 6$. Similarly, for P_4, P_2 and P_1 we have the values $\binom{4}{2} = 6$, 2 and 0, respectively, for $N(\gamma)$. Thus the number of necklaces is

$$\tfrac{1}{16}(70 + 6 \times 4 + 6 \times 5 + 2 \times 2) \;=\; 8.$$

Since there were only a few solutions, it probably would have been easier to count them by first listing them. Do it. Unfortunately, it is usually not easy to tell in advance that there will be so few solutions that listing them is easier than using Burnside's Lemma. One approach is to start listing them. If there seem to be too many, your time will probably not have been wasted because you will have gotten a better feel for the problem. □

*Proofs

We'll conclude this section with the two proofs that we put off:

1. A permutation group G on a set S gives an equivalence relation on S.
2. Burnside's Lemma is true.

Our proofs will be fairly heavy in notation and manipulation. If this causes difficulties for you, it may help if you consider what is happening in a simple case. For example, you might look the permutations associated with the Ferris wheel problem. You may also need to reread the proofs.

Proof: (Permutation groups give equivalence relations.) To prove this, we must prove that defining $x, y \in S$ to be equivalent if and only if $y = g(x)$ for some $g \in G$ does indeed give an equivalence relation. In other words, we must prove that there is a partition of S such that x and y are in the same block of S if and only if $y = g(x)$ for some $g \in G$.

Let

$$B_x \;=\; \{y \in S | y = g(x) \text{ for some } g \in G\}. \tag{4.19}$$

We need to know that the set of B_x's form a partition of S.

(a) We have $x \in B_x$ because $x = e(x)$. Thus every x in S is in at least one block.

(b) We must prove that the blocks are disjoint; that is, if $B_x \cap B_y \neq \emptyset$, then $B_x = B_y$. Suppose that $z \in B_x \cap B_y$ and $w \in B_x$, then, by (4.19), there are permutations f, g and h such that $z = f(x)$, $z = g(y)$ and $w = h(x)$. Thus

$$w \;=\; h(x) \;=\; h(f^{-1}(z)) \;=\; h(f^{-1}(g(y))) \;=\; (hf^{-1}g)(y).$$

By (G-2) and (G-3), $hf^{-1}g \in G$. Thus $w \in B_y$. We proved that $B_x \subseteq B_y$. Similarly, $B_y \subseteq B_x$ and so $B_x = B_y$. □

Proof: (Burnside's Lemma) Before proving Burnside's Lemma, we'll prove something that will be needed later in the proof. Let x be some element of S and let I_x be the set of all $g \in G$ such that $g(x) = x$. We will prove that

$$|I_x| \cdot |B_x| \;=\; |G|, \qquad 4.20 \quad \text{where } B_x \text{ is defined by (4.19).}$$

To illustrate this, consider our Ferris wheel problem with $x = 121212$, we have $B_x = \{121212, 212121\}$ and $I_x = \{g_0, g_2, g_4\}$ and $|G| = 6$. You should look at some other examples to convince yourself that (4.20) is true in general.

How can we prove (4.20)? We use a trick. Let $F\colon G \to S$ be defined by $F(g) = g(x)$. Be careful: x is *fixed* and g is the variable. Note that $\text{Image}(F) = B_x$ and $F^{-1}(x) = I_x$. We claim that $|F^{-1}(y)| = |F^{-1}(x)|$ for all $y \in B_x$. The validity of this claim is enough to prove (4.20) because the claim proves that the coimage of F, which is a partition of G, consists of $|B_x|$ blocks each of size $|F^{-1}(x)|$.

We now prove the claim. Now both x and y are fixed. Since $y \in B_x$, there is some $h \in G$ such that $y = h(x)$. Then

$$\begin{aligned}
F^{-1}(y) &= \{g \in G \mid g(x) = y\} && \text{by the definition of } F^{-1};\\
&= \{g \in G \mid g(x) = h(x)\} && \text{since } y = h(x);\\
&= \{g \in G \mid (h^{-1}g)(x) = x\} && \text{by (G-2);}\\
&= \{hk \mid k \in G \text{ and } k(x) = x\} && \text{by setting } k = h^{-1}g;\\
&= \{hk \mid k \in F^{-1}(x)\} && \text{by the definition of } F;\\
&= hF^{-1}(x).
\end{aligned}$$

This gives us a bijection between $F^{-1}(y)$ and $F^{-1}(x)$. Thus $|F^{-1}(y)| \;=\; |F^{-1}(x)|$.

We now prove Burnside's Lemma. The number of equivalence classes is simply the number of distinct B_x's. Unfortunately we can't easily get our hands on an entire equivalence class or a canonical representative. The following observation will let us look at all the elements in each equivalence class; i.e., all the elements in S.

$$\text{For any set } T, \quad 1 \;=\; \sum_{t \in T} \frac{1}{|T|}.$$

You should be able to prove this easily.

Let $\mathcal{E}$ be the set of equivalence classes of S. Then

$$|\mathcal{E}| \;=\; \sum_{B \in \mathcal{E}} 1 \;=\; \sum_{B \in \mathcal{E}} \sum_{y \in B} \frac{1}{|B|} \;=\; \sum_{B \in \mathcal{E}} \sum_{y \in B} \frac{1}{|B_y|},$$

since $B_y = B$. The last double sum is just $\sum_{y \in S} 1/|B_y|$ because each $y \in S$ belongs to exactly one equivalence class. Let $\chi(P)$ be 1 if the statement P is true and 0 if it is false. This is called a *characteristic function*. Using the above and (4.20),

$$\begin{aligned}
|\mathcal{E}| &= \sum_{y \in S} \frac{1}{|B_y|} \;=\; \sum_{y \in S} \frac{|I_y|}{|G|} \;=\; \frac{1}{|G|} \sum_{y \in S} |I_y|\\
&= \frac{1}{|G|} \sum_{y \in S} \Big(\sum_{g \in G} \chi(g(y) = y) \Big) && \text{by the definition of } I_y;\\
&= \frac{1}{|G|} \sum_{g \in G} \Big(\sum_{y \in S} \chi(g(y) = y) \Big) && \text{by interchanging summation;}\\
&= \frac{1}{|G|} \sum_{g \in G} N(g) && \text{by the definition of } N(g).
\end{aligned}$$

This completes the proof of Burnside's Lemma. □

Exercises

4.3.1. Suppose that you can count only ordered lists and you would like a formula for $C(n,k)$, the number of k element subsets of $\underline{n}$. Let A be the set of all k-lists without repeated elements that can be formed from $\underline{n}$. Let B be all subsets of $\underline{n}$. We define $F\colon A \to B$ as follows. For $a \in A$, let $F(a)$ be the set whose elements are the items in the list a. By studying the image of F and $|F^{-1}(x)|$ for $x \in \text{Image}(F)$, obtain a formula for $C(n,k)$.

4.3.2. How many 8-long circular sequences can be made using the ten digits $0, 1, 2, \ldots, 9$ if no digit can appear more than once? Can you generalize your answer to n-long circular sequences when k things are available instead of just ten?

4.3.3. Redo Example 4.12 with the numbers of beads changed from 4 and 4 to 3 and 5. Use Burnside's Lemma.

4.3.4. Redo Example 4.12 where there are k colors of beads and there are no constraints on how often a color may be used.

4.3.5. Label the vertices of a regular n-gon clockwise using the numbers 1 to n in order. We can describe a symmetry of the n-gon by a permutation of $\underline{n}$. The set of n symmetries of the n-gon that involve just rotating it in the plane about its center is called the *cyclic group* on $\underline{n}$. The set that also allows flipping the n-gon over is called the *dihedral group* on $\underline{n}$. It contains $2n$ symmetries, including the original n from the cyclic group.

(a) Describe the elements of the cyclic group as permutations in two line form. (There is a very simple description of the second line. You should be able to find it by drawing a picture and rotating it.)

(b) Describe the elements of the dihedral group as permutations in two line form. (There is a simple description of the second line of the additional permutations not in the cyclic group.)

(c) Describe the cycles of the elements of the dihedral group that are not in the cyclic group.

4.3.6. How many ways can 8 squares be colored green on a 4×4 board of 16 squares?

(a) Assume that the only symmetries that are allowed are rotations of the board.

(b) Assume that the board can be rotated and flipped over.

4.3.7. Starting with the observation

$$\sum_{y\in S}\Big(\sum_{g\in G}\chi(g(y)=y)\Big) \;=\; \sum_{g\in G}\Big(\sum_{y\in S}\chi(g(y)=y)\Big),$$

use (4.20) to prove Burnside's Lemma. (This is just a rearrangement of the proof in the text.)

4.3.8. Let $D(n)$ be the number of ways to arrange n dominoes to cover a $2 \times n$ board with no symmetries allowed. Let $d(n)$ be the number of ways to arrange them when rotations and reflections are allowed.

(a) List the coverings that give $D(5) = 8$ and $D(6) = 13$. Describe the coverings in general.

(b) Prove that $D(n)$ is the number of compositions of n where the only allowed parts are 1 and 2.

(c) Prove that $d(n)$ is the number of equivalence classes of compositions of n into ones and twos, where two compositions are equivalent if they are the same or if reading one from left to right is the same as the other read from right to left.

(d) Prove that

$$d(n) \;=\; \begin{cases} \frac{1}{2}\Big(D(n)+D(k)\Big) & \text{if } n = 2k+1; \\ \frac{1}{2}\Big(D(n)+D(k)+D(k-1)\Big) & \text{if } n = 2k. \end{cases}$$

Notes and References

Alternative discussions of the Principle of Inclusion and Exclusion can be found in the texts by Bogart [3; Ch.3], Stanley [6; Ch.2] and Tucker [7; Ch.8]. Stanley [6; Sec.2.6] uses the "Involution Principle" to give a "bijective" proof of the Principle of Inclusion and Exclusion. A bijective proof of a formula first interprets both sides of a formula as counting something in a simple manner (in particular, no minus signs are normally present). The proof consists of a bijection between the two sets of objects being counted. The Involution Principle is a fairly new technique for proving bijections that was introduced by Garsia and Milne [4]. Exercise 4.1.15 was adapted from [8].

Gian-Carlo Rota [5] introduced Möbius inversion to combinatorialists. A less advanced discussion of Möbius inversion and its applications has been given by Bender and Goldman [1]. Möbius inversion is only one of the many aspects of partially ordered sets which have become important in combinatorial theory. See Stanley [6; Ch.3] for an introduction. In turn, partially ordered sets are only one of the tools of modern algebraic mathematics that are important in combinatorics. This explosive growth of algebraic methods in combinatorics began in the late 1960's.

We'll return to the study of objects with symmetries in Section 11.3, where we connect a special case of Burnside's Lemma with generating functions. Among the texts that discuss enumeration with symmetries are those by Biggs [2; Chs.13,14] and Tucker [7; Ch.9]. Williamson [9; Ch.4] goes deeper into some aspects related to computer applications. The study of objects with symmetries is inevitably tied to the theory of permutation groups, which we have attempted to minimize. See Biggs [2] for more background on group theory.

1. Edward A. Bender and Jay R. Goldman, On the applications of Möbius inversion in combinatorial analysis, *American Math. Monthly* **82** (1975), 789–803.
2. Norman L. Biggs, *Discrete Mathematics*, 2nd ed., Oxford Univ. Press (2003).
3. Kenneth P. Bogart, *Introductory Combinatorics*, 3rd ed., Brooks/Cole (2000).
4. Adriano M. Garsia and Stephen C. Milne, A Rogers-Ramanujan bijection, *J. Combinatorial Theory, Series A* **31** (1981), 289–339.
5. Gian-Carlo Rota, On the foundations of combinatorial theory I. Theory of Möbius functions, *Zeitschrift für Wahrscheinlichkeitstheorie* **2** (1964), 340–368.
6. Richard P. Stanley, *Enumerative Combinatorics*, vols. 1 and 2, Cambridge Univ. Press (1999, 2001).
7. Alan C. Tucker, *Applied Combinatorics*, 4th ed., John Wiley (2001).
8. Herbert S. Wilf, Two algorithms for the sieve method, *J. of Algorithms* **12** (1991), 179–182.
9. S. Gill Williamson, *Combinatorics for Computer Science*, Dover (2002).