Probability and Statistics (ENM 503)

Michael A. Carchidi

September 9, 2018

Chapter 1 - Introduction to Combinatorics (The Art of Counting)

The following notes are based on the textbook entitled: *An Introduction to Discrete Mathematics* by Steven Roman (2nd edition) and these notes can be viewed at

https://canvas.upenn.edu/

after you log in using your PennKey user name and Password.

1. Introduction

In this chapter we shall study the art of counting the number of elements in a specified *finite* set S, known also as the *size* of the set S, which is denoted by $|S|$. If the size of the set is small, such as the ordered outcomes obtained when two fair 6-sided die are rolled, then we may simply list the elements of S and count them directly. In this case we would have

$$S = \{(a,b)|a = 1,2,3,4,5,6 \text{ and } b = 1,2,3,4,5,6\},$$

which would give

$$\begin{aligned} S \;=\; & \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6), \\ & (2,1),(2,2),(2,3),(2,4),(2,5),(2,6), \\ & (3,1),(3,2),(3,3),(3,4),(3,5),(3,6), \\ & (4,1),(4,2),(4,3),(4,4),(4,5),(4,6), \\ & (5,1),(5,2),(5,3),(5,4),(5,5),(5,6), \\ & (6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\} \end{aligned}$$

and then $|S| = 36$. Similarly if T lists the totals obtained when two fair 6-sided die are rolled, then adding the coordinates of the elements of S gives

$$T = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

as the possible elements of T, so that $|T| = 11$.

On the other hand, if the set S contains many elements such as the number of 5-card poker hands that are possible out of a deck of 52 distinct playing cards, then determining $|S|$ by listing the elements such as

$$S = \{(2\text{C}, 2\text{D}, 2\text{H}, 2\text{S}, 3\text{C}), (2\text{C}, 2\text{D}, 2\text{H}, 2\text{S}, 4\text{C}), (2\text{C}, 2\text{D}, 2\text{H}, 2\text{S}, 5\text{C}), ..., \}$$

is out of the question since the number of elements in S is large. In fact, we shall see later that $|S| = 2,598,960$, which would take about

$$\frac{2,598,960}{24 \times 3,600} \simeq 30.1 \text{ days}$$

or *one month* to list out at a rate of one element per second.

Combinatorics is that branch of mathematics which is concerned with the question of determining the number of elements in a finite set S, *without having to list them out one-by-one*. You will note that the title of this chapter refers to combinatorics as the "art" of counting and not the "mathematics" of counting because, in my opinion, the formulas for counting the number of elements in a set are quite simple, but being able to apply the right formula (or formulas) for a given problem is more of an art since it comes from much practice.

A very good book that discusses the beginnings of combinatorics (at a nice level) is called *An Introduction to Discrete Mathematics* by Steven Roman (2nd edition) and this chapter, along with its many examples are taken from this text. Let's begin with a listing of some problems which we shall solve later in the chapter and the next.

Example #1: Naming Computer Chips

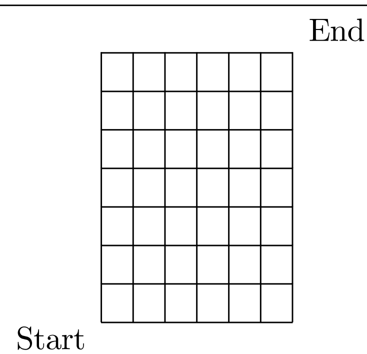
A certain company manufactures computer chips. Each type of chip must be given a name so that it can be identified by customers. The company has decided that it can be more organized in its bookkeeping if the names of these chips consists of 3 letters follows by 3 digits, rather that simply naming the chips: #1, #2, #3, and so on. For example, one possible name is CDX154. If S is the set of all possible names for these computer chips, determine the size of the set S. We shall see in Example #8 that the answer is $S = 17,576,000$.■

Example #2: The Early Days of Computer Programming

In the early days of computer programming (1970s), computer programs would be written by students using punch cards and then these cards would be submitted to a computer operator who would feed these through a card reader for the mainframe computer so that this computer could run the program. The output would then be printed on large sheets of paper and given back to the student sometime later that day (or week). Suppose that every morning, a computer operator receives one computer program from each of 9 students and must decide in which order to run these programs. In order to be fair, the operator wants to assign a different order to the programs each day. For how many days can he do this? In other words, how many different orderings are there of 9 programs? We shall see (between Examples #17 and #18) that the answer is $362,880$. ■

Example #3: Taking A Walk

There are 13 one-way streets arranged in a rectangular grid with 6 of the streets going from West to East and 7 of the streets going from South to North, as shown in the figure below.



East is to the right and North is up

A person wants to walk the 13 blocks from the most southwest point on this map (labeled by Start) to the most northeast point (labeled by End), always heading either east or north. The problem is to count the number of ways that this can be done. For example, one path is simply to go east for 6 blocks and then north for the remaining 7 blocks, which we represent by

$$E \rightarrow E \rightarrow E \rightarrow E \rightarrow E \rightarrow E \rightarrow N \rightarrow N \rightarrow N \rightarrow N \rightarrow N \rightarrow N \rightarrow N.$$

or (without the arrows) $EEEEEENNNNNNN$. Another path is

$$N \rightarrow E \rightarrow N \rightarrow E \rightarrow N \rightarrow E \rightarrow N \rightarrow E \rightarrow N \rightarrow E \rightarrow N \rightarrow E \rightarrow N,$$

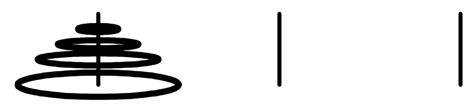
or (without the arrows) $NENENENENENEN$, and so on. Note that these paths can also be viewed as the *minimum distance traveled paths* from "Start" to "End". We shall see in Example #29 that the answer to this is $1,716$. ■

Example #4: What a Mess

Suppose that your mathematics class has 12 students. On the day that your midterm exams are returned, your professor makes a total mess of things and returns the exams in such a way that every student gets someone else's exam, instead of his or her own exam. How many ways can this be done? We shall see in Example #25 of Chapter #2 that the answer to this problem is $176,214,841$, and we shall see in that chapter that this is an example of a problem in *derangements*. ■

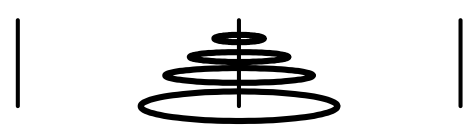
Example #5: The Towers of Hanoi

A game known as the Towers of Hanoi is pictured below. In the version shown below, there are 3 pegs (labeled 1, 2, 3 from left to right) and 4 circular disks of increasing diameter, with each of them having a small hole through their centers thereby allowing the pegs to go through the centers of the disks.

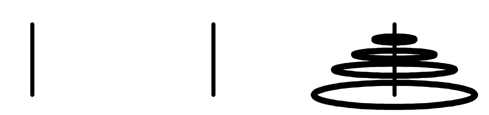


The game “begins” with all four disks stacked in decreasing size on the first peg

The game begins with all 4 disks placed on the first (left-most) peg so that the largest disk is at the bottom of the stack and the other disks (in decreasing diameter) are placed on top of this one. The object of the game is to transfer all of the four disks from the first (left-most) peg onto one of the other two pegs, so that the largest disk is at the bottom and the other disks (in decreasing diameter) are placed on top of this as shown in the two figures below.



The game “ends” when all four disks are stacked in decreasing size on the 2nd peg or



The game "ends" when all four disks are stacked in decreasing size on the 3rd peg

A disk may be moved from off of any one peg to any other peg with the condition that a larger disk may not be placed on top of a smaller disk. The problem here is to count the *minimum number* of moves required to transfer the 4-high "tower" of disks from the first peg to the third peg. More generally, we want to count the minimum number of moves required to transfer an n-high "tower" of disks from the first peg to the third peg. We shall see in Examples #31 and #36 that the answer to this is 15.■

The Tower of Hanoi: History

As taken from Wikipedia: The puzzle was invented by the French mathematician Édouard Lucas in 1883. There is a story about an Indian temple in Kashi Vishwanath which contains a large room with three time-worn posts in it surrounded by 64 golden disks. Brahmin priests, acting out the command of an ancient prophecy, have been moving these disks, in accordance with the immutable rules of the Brahma, since that time. The puzzle is therefore also known as the Tower of Brahma puzzle. According to the legend, when the last move of the puzzle will be completed, the world will end. It is not clear whether Lucas invented this legend or was inspired by it. If the legend were true, and if the priests were able to move disks at a rate of *one per second*, using the smallest number of moves, it would take them $2^{64} - 1$ seconds or roughly 585 billion years or $18,446,744,073,709,551,615$ turns to finish, or about 127 times the current age of the sun. There are many variations on this legend. For instance, in some tellings, the temple is a monastery and the priests are monks. The temple or monastery may be said to be in different parts of the world including Hanoi, Vietnam, and may be associated with any religion. In some versions, other elements are introduced, such as the fact that the tower was created at the beginning of the world, or that the priests or monks may make only one move per day.

Example #6: Giving Out Change for a Dollar

How many ways are there to give change for a dollar bill using pennies, nickels, dimes, and quarters? For example, two possible ways are: (a) 5 pennies, 2 dimes, and 3 quarters, and (b) 10 pennies, 1 nickel, 6 dimes, and 1 quarter. We shall see in Example #27 of Chapter #2 that the answer to this is 242. ■

Example #7: Inclusion-Exclusion

In a group of 200 people, 100 like Coke, 149 like Pepsi, 83 like Seven-Up, 80 like Coke and Pepsi, 66 like Coke and Seven-Up, 45 like Pepsi and Seven-Up, and 12 like Coke, Pepsi, and Seven-Up. How many of these people like none of the soft drinks? We shall see in Example #20 of Chapter #2 that the answer to this is 47. ■

Example #8: Integer Solutions To Algebraic Equations

Determine the number of integer solutions to the equation $x + y + z = 25$, with the property that

$$2 \leq x \leq 8 \quad , \quad 3 \leq y \leq 12 \quad \text{and} \quad 7 \leq z \leq 10.$$

Note that one such solution is $x = 5$, $y = 10$ and $z = 10$. We shall see in Example #23 of Chapter #2 that the answer to this is 18. ■

Hopefully, these few examples of counting problems, as well as many more examples that will be introduced later, will show you that such problems do occur frequently in various contexts. Now let us proceed to a discussion of how to solve counting problems. We begin with the most powerful and yet fundamental principle used in counting problems, known as the multiplication principle

2. The Multiplication and Addition Principles

We will begin our study of counting techniques with two very simple, yet very powerful rules, which we will discuss in this and the next section. In fact, much of what we will do in this chapter is based on these two simple rules.

The Multiplication Principle

It turns out that many counting problems amount to counting the number of ways to perform a certain "sequence" of tasks, one followed by the next, which is followed by the next, and so on, where each task can be performed in several different ways. As a simple example, consider the problem of counting names for computer chips as described in Example #1 above. Deciding on a name for a particular computer chip amounts to performing a sequence of 6 tasks. The first task is to choose the first letter in the name, and this task can be performed in 26 different ways (A, B, C, ..., X, Y, Z), since there are 26 letters in the English alphabet. The second task is to choose the second letter, and it too can be performed in 26 different ways. The third task is to choose the third letter and it too can be performed in 26 different ways, the fourth task is to choose the first digit in the name and this can be performed in 10 different ways (0, 1, 2, 3, 4, 5, 6, 7, 8, 9), the fifth task is to choose the second digit in the name and this can also be performed in 10 different ways, and finally the sixth task is to choose the third digit in the name and this can also be performed in 10 different ways. Clearly, the number of possible names is simply the number of ways to perform this sequence of 6 tasks.

Our first counting rule, called the multiplication rule, will tell us the number of ways to perform such a sequence of k tasks, say T_1, T_2, ..., T_k. By the phrase "performing a sequence T_1, T_2, ..., T_k of k tasks," we mean of course, that task T_1 is performed first, then task T_2, and so on. Before we can state the multiplication rule, however, we must discuss one very important point. The basic (or simpliest) multiplication rule applies only under the assumption that the *number of ways* to perform any one of the tasks in the sequence, say T_i, does not depend on how the previous tasks T_1, T_2, ..., T_{i-1} were performed. In a sense, this is an assumption about the independence of the tasks in the sequence. If a sequence of tasks does not satisfy this type of independence condition, then we cannot apply the basic multiplication rule. We will see later how to extend the rule for when the tasks are not independent, but for now, let us assume that independence holds and now let us state the multiplication rule.

Suppose that T_1, T_2, ..., T_k is an ordered sequence of k tasks with the property that the number of ways to perform any task in the sequence does not depend on how the previous tasks in the sequence were performed. Then, if there are n_1 ways to perform the 1st task T_1, n_2 ways to perform the 2nd task T_2, n_3 ways to perform the 3rd task T_3, and so on up to there being n_k ways to perform the last (kth) task T_k, then the number of ways (N) to perform the entire sequence of tasks in the order T_1 followed by T_2 followed by T_3 and so one up to the last task T_k, is the product

$$N = n_1 \times n_2 \times n_3 \times \cdots \times n_k. \tag{1a}$$

More generally, if the number of ways to perform any task in the sequence does depend on how the previous tasks in the sequence were performed, then we may refer to $n_1 = n(T_1)$ as the number of ways to perform the initial task T_1, $n_2 = n(T_2|T_1)$ as the number of ways to perform the next task T_1, given that the previous task T_1 has already been performed (as represented by the vertical bar $|$), $n_3 = n(T_3|T_1T_2)$ as the number of ways to perform the next task T_3, given that the previous two tasks T_1 and T_2 have already been performed, and so on, up to

$$n_k = n(T_k|T_1T_2T_3 \cdots T_{k-1})$$

as the number of ways to perform the last or k^{th} task T_k, given that the previous $k-1$ tasks T_1, T_2, T_3, ..., T_{k-1} have already been performed. Then Equation (1a), using these values of n_1, n_2, n_3,..., n_k, gives the number of ways the sequence of k tasks $T_1, T_2, T_3, ..., T_k$ can be performed in the order $T_1, T_2, T_3, ..., T_k$.

To understand why this is the case, suppose we wanted to perform a sequence of two tasks T_1 and T_2. Task T_1 is to choose a digit 0, 1, 2, ..., 9 and the second task T_2 is to choose a letter A, B, C, ..., X, Y, Z from the alphabet and we want to determine how many ways this can be done. We could begin by creating the list of the 10 possibilities:

0A, 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 9A

if the letter is always an A. Then we would create the list of the 10 possibilities:

0B, 1B, 2B, 3B, 4B, 5B, 6B, 7B, 8B, 9B

if the letter is always a B, and so on until we create the list of the 10 possibilities:

0Z, 1Z, 2Z, 3Z, 4Z, 5Z, 6Z, 7Z, 8Z, 9Z,

if the letter is always a Z. Since there are 26 such lists (26 letters in the alphabet) and each list contains 10 possibilities, we see that there are

$$10 + 10 + 10 + \cdots + 10 = \sum_{j=1}^{26} 10 = 10 \times 26 = 260$$

ways of performing the sequence of the two task T_1 and T_2, in accordance with the multiplication principle.

The Addition Principle

The addition principle (as stated here) is a special case of a more general principle known as the principle of inclusion/exclusion, which we shall discuss in the next chapter. For these discussions, suppose that the number of ways to perform a sequence of k tasks can be divided into a sequence of m *disjoint* subtasks, then the number of ways of performing the sequence of these k tasks is simply the sum of the number of ways in performing the sequence of m disjoint subtasks.

Note that two tasks are said to be *disjoint* if the list of possibilities that makes up the first task has nothing in common with the list of possibilities that makes up the second task.

Mathematically, if a set A can be partitioned into m disjoint subsets A_1, A_2, A_3, ..., A_m, so that

$$A = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_m$$

with $A_i \cap A_j = \emptyset$ (the empty set) for all $i, j = 1, 2, 3, ..., m$ and $i \neq j$, then

$$|A| = |A_1| + |A_2| + |A_3| + \cdots + |A_m|. \tag{1b}$$

Two sets B and C are said to be *disjoint* if they have no elements in common which says that their intersection, defined by

$$B \cap C = \{x | x \in B \text{ and } x \in C\},$$

is empty, *i.e.*, $B \cap C = \emptyset$.

It should be noted that the addition principle was already used above where we wanted to perform a sequence of $k = 2$ two tasks T_1 and T_2 where the first task T_1 is to choose a digit 0, 1, 2, ..., 9 and the second task T_2 is to choose a letter A, B, C, ..., X, Y, Z from the alphabet. We did this by considering the subtask A_1, which created the list of the 10 possibilities:

0A, 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 9A

where the letter is always an A. Then we created the subtask A_2 which gives the list of the 10 possibilities:

0B, 1B, 2B, 3B, 4B, 5B, 6B, 7B, 8B, 9B

where the letter is always a B, and so on until we create the subtask A_{26} which gives the list of the 10 possibilities:

0Z, 1Z, 2Z, 3Z, 4Z, 5Z, 6Z, 7Z, 8Z, 9Z,

where the letter is always a Z. Since there are 26 such *disjoint* subtasks (26 letters in the alphabet) and each subtask contains 10 possibilities, we see that there are

$$|A| = |A_1| + |A_2| + |A_3| + \cdots + |A_{25}| + |A_{26}| = 10 + 10 + 10 + \cdots + 10 \text{ (26 terms)}$$

or $|A| = 260$ ways of performing the sequence of the two task T_1 and T_2. One sees from this that quite often the addition principle leads to the multiplication principle.

Example #9: The Solution to Example #1

Example #1 states that a certain company manufactures computer chips. Each type of chip must be given a name so that it can be identified by customers. The company has decided that it is more impressive if the names consists of 3 letters follows by 3 digits, rather that simply naming the chips: Chip #1, Chip #2, Chip #3, and so on. For example, one possible name is Chip CDX154. If S is the set of all possible names for these computer chips, then we see that there are six task to perform. The first three involve choosing a letter from the 26 in the alphabet set {A, B, C, ..., X, Y, Z}, and the last three involve choosing one of 10 digits from the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Using the multiplication principle, we then see that

$$|S| = 26 \times 26 \times 26 \times 10 \times 10 \times 10$$

or $|S| = 17,576,000$, a result that would have taken a long time to list out one at a time. In fact, listing these out at a rate of one per second would take

$$17,576,000 \text{ sec} \times \frac{1 \text{ min}}{60 \text{ sec}} \times \frac{1 \text{ hour}}{60 \text{ min}} \times \frac{1 \text{ day}}{24 \text{ hour}}$$

or 203.4 days. ■

Quite often, both the multiplication and addition principles can be jointly used to solve a counting problem. We illustrate this with the next example.

Example #10: An Alphabet and Its Words

An alphabet is a set of distinct symbols and an alphabet having m distinct symbols is called an alphabet of size m. Suppose that Σ is an alphabet of size $m \geq 1$. Forming a word of length n over Σ amounts to performing a sequence of n tasks. Task 1 is to choose a symbol from Σ for first symbol in the word, task 2 is to choose another (not necessarily different) symbol from Σ for second symbol in the word, and so on. Since each task can be performed in m different ways, the *multiplication* principle tells us that there are $m \times m \times \cdots \times m$ with n factors, or m^n different ways to perform the entire sequence of n tasks. The set of all these words that can be formed from the alphabet Σ is denoted by Σ_n and so there are m^n different words of length n over the set Σ. In summary, we may say that

$$\text{If} \quad |\Sigma| = m, \quad \text{then} \quad |\Sigma_n| = m^n. \tag{2}$$

For example, if $\Sigma = \{0, 1\}$, then there are 2^n different binary words of length n that can be constructed. In the case when $n = 3$, we have $2^3 = 8$ and the 8 binary words of length 3 are as follows.

$$\Sigma_3 = \{000, 001, 010, 011, 100, 101, 110, 111\}.$$

We can use this information and the *addition* principle to now compute the size of the set Γ_n, of all words of length *at most* n, which are words of length n or less, including the empty word that has no letters. To do this, we observe that

$$\Gamma_n = \Sigma_0 \cup \Sigma_1 \cup \Sigma_2 \cup \cdots \cup \Sigma_{n-1} \cup \Sigma_n$$

and since each of these sets are disjoint, (*i.e.*, $\Sigma_i \cap \Sigma_j = \emptyset$) for all $i \neq j$, we conclude from the addition principle that

$$|\Gamma_n| = |\Sigma_0| + |\Sigma_1| + |\Sigma_2| + \cdots + |\Sigma_{n-1}| + |\Sigma_n|$$

and since $\Sigma_0 = \{\emptyset\}$ and thereby has one element (namely the empty set), then we have

$$|\Gamma_n| = 1 + m + m^2 + \cdots + m^{n-1} + m^n.$$

Using the algebraic identity

$$1 + x + x^2 + \cdots + x^{n-1} + x^n = \begin{cases} n+1, & \text{when} \quad x = 1 \\ \\ (x^{n+1} - 1)/(x - 1), & \text{when} \quad x \neq 1 \end{cases}$$

for $n = 0, 1, 2, 3, \ldots$, we find that

$$|\Gamma_n| = \begin{cases} n+1, & \text{when} \quad m = 1 \\ \\ (m^{n+1} - 1)/(m - 1), & \text{when} \quad m > 1 \end{cases}. \tag{3}$$

Note that $\Sigma_0 = \{\emptyset\} \neq \emptyset$ which gives $|\Sigma_0| = 1$. ■

Example #11: Why Order Could Be Important

How many *even* integers with distinct digits are there between 1 and 99 where we shall agree to write the integers between 1 and 9, inclusive as 01, 02, ... , 09? By distinct digits, we mean that the two digits must be different. For example, the integer 22 does not have distinct digits. To solve this, let T_1 be the task of choosing the units digit from among the 5 possibilities 0, 2, 4, 6, and 8, since the number must be even. Let T_2 be the task of choosing the tens digit. Now the number of ways in performing the task T_2 is 9 since we may choose any of the $(10 - 1 = 9)$ remaining digits 0, 1, 2, 3, 4, 5, 6, 7, 8 or 9 that was not chosen in T_1. Note that the particular choice for T_2 does depend on the particular choice for T_1, so that if 6 is chosen for T_1, it cannot be chosen for T_2, but the *number of choices* (namely 9) for T_2 is not dependent on the choice made in task T_1. There are 5 ways to perform the first task T_1, and there are 9 ways to perform the second task T_2. Thus, according to the multiplication principle, there are $5 \times 9 = 45$ even integers with distinct digits between 01 and 99.

The previous example points out a very important fact, namely, the order in which we pick our tasks can be crucial to the analysis. For if, in the previous example, we had let T_1 be the task of choosing the tens digit, and T_2, be the task of choosing the units digit, then we have to be more careful in the analysis because would be 10 ways to perform task T_1, since there are 10 possible digits for the tens place. However, the number of ways to perform task T_2 would then depend on how we performed task T_1 and hence not be independent of T_1. For if we had chosen an odd digit for the tens place (1, 3, 5, 7, 9), then we would have 5 possibilities for the units digit (0, 2, 4, 6, 8), but if we had chosen an even digit for the tens digit (0, 2, 4, 6, 8), then we would have only 4 possibilities for the units digit, namely 0, 2, 4, 6, 8 minus the digit that was chosen for the tens place. Thus, if we had chosen our tasks in this order, the independence condition would not hold, and we could not apply the multiplication rule in the same way that we did above. However, since the sets $\{1, 3, 5, 7, 9\}$ and $\{0, 2, 4, 6, 8\}$ are disjoint, we could still solve the problem by computing

$$\# \text{ of even digits of the form odd-even} + \# \text{ of even digits of the form even-even}$$

which equals $(5 \times 5) + (5 \times 4) = 25 + 20 = 45$. ■

The multiplication rule is really a very simple idea. The trick is in recognizing when it applies, that is, in recognizing whether or not a given problem can be thought of as a problem involving a sequence of tasks. As you will see, a great many problems can be expressed in this way, and so this rule does apply in a surprisingly large number of problems. From now on, whenever we use the multiplication rule, we will assume that the required independence condition has been verified.

3. One-To-One Correspondences

Quite often when trying to count the number of elements in a set, it might be easier to count the number of elements in a different set that can be shown to have the same size as the original set. Showing that two sets have the same size is done through the ideal of one-to-one correspondence. Because of its importance, let use review the ideal behind one-to-one correspondence. Suppose we have two *finite* sets

$$A = \{a_1, a_2, a_3, ..., a_n\} \qquad \text{and} \qquad B = \{b_1, b_2, b_3, ..., b_m\}$$

and we would like to determine which set is larger in size *without knowing anything about how to count*. We let $|A|$ be the size of A (which is just the number of elements in A), and we let $|B|$ be the size of B (which is just the number of elements in B). We could determine which set is larger in size *without actually counting the elements in A and B* by taking an element a_1 out of A and an element b_1 out of B and placing them side by side.

$$a_1 \leftrightarrow b_1$$

Then we take another element a_2 out of A and another element b_2 out of B and place them side by side.

$$\begin{array}{ccc} a_1 & \leftrightarrow & b_1 \\ a_2 & \leftrightarrow & b_2 \end{array}$$

Then we take another element a_3 out of A and another element b_3 out of B and place them side by side,

$$\begin{array}{ccc} a_1 & \leftrightarrow & b_1 \\ a_2 & \leftrightarrow & b_2 \\ a_3 & \leftrightarrow & b_3 \end{array}$$

and so on until we either: (i) run out of elements in A (with still some elements left in B), (ii) run out of elements in B (with still some elements left in A), or (iii) run out of elements in A and B at the same time. Of course: (i) if we run out of elements in A with still some elements left in B, then we say that B is larger in size than A (written as $|A| < |B|$), (ii) if we run out of elements in B with still some elements left in A, then we say that A is larger in size than B (written as $|A| > |B|$), and (iii) if we run out of elements in A and B at the same time, then we say that A and B have the same size (written as $|A| = |B|$). The process in which elements are placed side-by-side, as described above, uses the mathematical idea of a mapping (or function) from one set to another.

A Mapping (or Function) From One Set to Another Set

Given two sets A and B, a mapping (or function) from A to B (written as $f : A \to B$) is a rule that assigns to each element in A a *unique* element in B. This means that assigning an element a_1 in A to both b_1 and b_2 (with $b_1 \neq b_2$) in B is not permitted. However, if a_1 and a_2 are two different elements in A, assigning each of these to the same element b, in B is allowed. If the element a in A is assigned to the element b in B, then we write $f(a) = b$. Mathematically we then see that $f(a_1) = b_1$ and $f(a_1) = b_2$ with $b_1 \neq b_2$, is not allowed but $f(a_1) = b_1$ and $f(a_2) = b_1$ with $a_1 \neq a_2$ is allowed.

A function is like a "machine" that takes as input, a single element in A and returns as output, a single element in B, and this may be represented as follows.

$$\boxed{(a \in A) \to \boxed{\text{Function Machine } (f)} \to (b = f(a) \in B)}$$

There are special types of functions that will be used in counting problems and we now discuss these.

Special Types of Functions

A function $f : A \to B$ is called a *one-to-one* function from A to B if the expression $f(a_1) = f(a_2)$ leads only to $a_1 = a_2$. A function $f : A \to B$ is called an *onto* function from A to B if for *every* element b in B, there is an element a in A such that $f(a) = b$. A function $f : A \to B$ is called a *one-to-one correspondence* between the sets A and B if it is both one-to-one and onto. It should be noted that if $f : A \to B$ is called a one-to-one correspondence between the sets A and B, then the inverse function $f^{-1} : B \to A$ is also a one-to-one correspondence between the sets A and B.

One-to-One Correspondence and the Size of Sets

Consider two sets A and B, it should be evident that A and B have the same size (*i.e.*, same number of elements) *if and only if* there exist a one-to-one correspondence between A and B. Of course this implies that two sets A and B are not the same size if there does not exist a one-to-one correspondence between the two sets.

Example #12: The Number of Subsets of a Set of Size n

The multiplication rule can be used to determine the number of subsets of a set containing n elements, but it takes a little bit of "finesse" to use it properly. This example will show some of the "art" that is involved in combinatorics and how one-to-one correspondence can be used. Starting with a simple specific example, which is sometimes good to do just to organize one's thought process, suppose that P_{A_5} is the set of all subsets of a 5-element set

$$A_5 = \{a_1, a_2, a_3, a_4, a_5\}$$

and suppose we define the sets

$$S_{m,5} = \{x_1, x_2, x_3, x_4, x_5\}$$

as other 5-element sets, where each x_i is either 0 or 1. Since there are 2 choices for x_1, 2 choices for x_2, 2 choices for x_3, 2 choices for x_4 and 2 choices for x_5, the multiplication principle tells us that there are

$$2 \times 2 \times 2 \times 2 \times 2 = 2^5 = 32$$

such possible $S_{m,5}$ sets, which we may label as $S_{0,5}$, $S_{1,5}$, $S_{2,5}$, ..., $S_{31,5}$ by setting

$$m = 2^4 x_1 + 2^3 x_2 + 2^2 x_3 + 2x_4 + x_5$$

and then we define

$$S_5 = \{S_{0,5}, S_{1,5}, S_{2,5}, ..., S_{31,5}\}$$

and we note that $|S_5| = 2^5 = 32$. For example, the set $\{1, 0, 0, 1, 1\}$ would be $S_{19,5}$, since

$$m = 2^4(1) + 2^3(0) + 2^2(0) + 2(1) + (1) = 19.$$

Now consider a subset B of $A_5 = \{a_1, a_2, a_3, a_4, a_5\}$, e.g.,

$$B = \{a_1, a_4, a_5\}.$$

Since each subset of A either contains the element a_1 or does not contain the element a_1, we may capture this ideal by assigning x_1 (in $S_{m,5}$) the value of 1 if $a_1 \in B$ and assigning x_1 (in $S_{m,5}$) the value of 0 if $a_1 \notin B$. We may also set x_2

(in $S_{m,5}$) equal to 1 if $a_2 \in B$ and set x_2 (in $S_{m,5}$) equal to 0 if $a_2 \notin B$, and so on. In other words, we set

$$x_i = \begin{cases} 0, & \text{if} \quad a_i \notin B \\ \\ 1, & \text{if} \quad a_i \in B \end{cases},$$

for all $i = 1, 2, 3, 4, 5$. With this, we may now match with the subset $B = \{a_1, a_4, a_5\}$ of A_5, the set

$$S_{19,5} = \{x_1, x_2, x_3, x_4, x_5\} = \{1, 0, 0, 1, 1\}$$

since $a_1 \in B$, $a_2 \notin B$, $a_3 \notin B$, $a_4 \in B$, and $a_5 \in B$, and this matching is a *one-to-one correspondence* between the subsets of A_5, (namely P_{A_5}) and the set of all $S_{m,5}$. In the above example, we have

$$\{a_1, a_4, a_5\} \Leftrightarrow \{1, 0, 0, 1, 1\},$$

and note that $\{a_5, a_1, a_4\}$, which is the same as $\{a_1, a_4, a_5\}$, will have

$$\{a_5, a_1, a_4\} = \{a_1, a_4, a_5\} \Leftrightarrow \{1, 0, 0, 1, 1\}$$

as well. For this reason we may always order the elements in B using the same order as those in A_5. We also have, for example

$$\emptyset \Leftrightarrow \{0, 0, 0, 0, 0\} = S_{0,5} \quad \text{and} \quad \{a_1, a_2, a_3, a_4, a_5\} \Leftrightarrow \{1, 1, 1, 1, 1\} = S_{31,5} = A_5.$$

Since there are $m = 2^5 = 32$ possible sets $S_{m,5}$ in S_5, there must also be 32 possible elements in P_{A_5} and hence 32 possible subsets of $A_5 = \{a_1, a_2, a_3, a_4, a_5\}$. Of course, this example using 5 elements in A_5 easily generalizes to

$$A_n = \{a_1, a_2, a_3, ..., a_n\} \quad \text{and} \quad S_{m,n} = \{x_1, x_2, x_3, ..., x_n\}$$

with

$$m = 2^{n-1}x_1 + 2^{n-2}x_2 + \cdots + 2^2 x_{n-2} + 2x_{n-1} + x_n$$

and

$$S_n = \{S_{m,n} | m = 0, 1, 2, ..., 2^n - 1\}$$

and so $|P_{A_n}| = |S_n| = 2^n = 2^{|A_n|}$. ■

A Very Powerful Result

This example shows that if we want to count the elements in some set A and if we could construct another set B and a one-to-one correspondence between A and B, then we could count the number of elements in A by first counting the number of elements in B (which might be easier to do) and then setting $|A| = |B|$. Other examples of this idea will appear throughout this course.

4. The Pigeonhole Principle

Let us now consider the second of our simple, yet very powerful rules, known as the Dirichlet pigeonhole principle. This rule is named after the French mathematician Peter Gustav Lejeune Dirichlet (1805–1859), but we will refer to it simply as the pigeonhole principle.

The Pigeonhole Principle

If m balls are placed in n boxes, then one of the boxes must receive at least

$$\lceil m/n \rceil$$

of the balls, where

$$C(x) = \lceil x \rceil \tag{4}$$

is the *ceiling* of x, which is the *smallest integer greater than or equal to* x. Of course, the pigeonhole principle can be worded in many different ways. To understand the reason for the principle, consider $m = 17$ balls and $n = 6$ boxes and suppose we start placing the balls in the boxes so that each box gets the smallest number of possible balls. We would do this by first placing one ball in each box,

$$\boxed{\text{o}} \quad \boxed{\text{o}} \quad \boxed{\text{o}} \quad \boxed{\text{o}} \quad \boxed{\text{o}} \quad \boxed{\text{o}}$$

thereby using up 6 balls. Then we place another ball in each box

$$\boxed{\text{o,o}} \quad \boxed{\text{o,o}} \quad \boxed{\text{o,o}} \quad \boxed{\text{o,o}} \quad \boxed{\text{o,o}} \quad \boxed{\text{o,o}}$$

thereby using up another 6 balls. Then we place another ball in each box

$$\boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o}}$$

thereby using up the remaining 5 balls. We note that one of the boxes (in this case one of the first 5 boxes) must contain at least

$$C(17/6) = \lceil 17/6 \rceil = \lceil 2.833333 \rceil = 3$$

balls. If, in the above example, we have $m = 19$ and $n = 6$, then we have

$$\boxed{\text{o}} \quad \boxed{\text{o}} \quad \boxed{\text{o}} \quad \boxed{\text{o}} \quad \boxed{\text{o}} \quad \boxed{\text{o}}$$

and

$$\boxed{\text{o,o}} \quad \boxed{\text{o,o}} \quad \boxed{\text{o,o}} \quad \boxed{\text{o,o}} \quad \boxed{\text{o,o}} \quad \boxed{\text{o,o}}$$

and

$$\boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}}$$

and

$$\boxed{\text{o,o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}} \quad \boxed{\text{o,o,o}}$$

showing that one of the boxes (in this case the first box) must contain at least

$$C(19/6) = \lceil 19/6 \rceil = \lceil 3.166667 \rceil = 4$$

balls.

The pigeonhole principle seems like a very simple rule, almost not even worth mentioning. But, as you will soon see, it can be used in an amazingly large variety of ways to solve complicated counting problems. Let us begin with some simple examples.

Example #13: Birthdays

In any group of 367 people, at least two people must have the same birthday. In this case, we label 366 boxes with the 366 possible birthdays (including February 29), and label 367 balls with the names of the 367 people. Then we place each ball into the box that bears the birth date of the person whose name is on the ball. According to the pigeonhole principle, one of the boxes must receive at least

$$\lceil 367/366 \rceil = \lceil 1.002732 \rceil = 2$$

balls so that there must be at least two people with the same birthday. ■

Example #14: Hairs and New York City

As of 2012, the population of New York City is around 8.337 million people. If we assume that no person has more than 500,000 hairs on his or her head (a reasonable assumption, by the way), then according to the pigeonhole principle, there must be at least

$$\left\lceil \frac{8.337 \times 10^6}{500,000} \right\rceil = \lceil 16.674 \rceil = 17$$

people in New York City with the same number of hairs on their heads. ■

The pigeonhole principle can be used to solve certain problems from a branch of mathematics known as number theory, which can be described as the study of the positive integers. Let us consider a few examples of this type.

Example #15: Number Theory

Given any subset of $n+1$ integers from the set $\{1, 2, ..., 2n\}$, there must be two integers from the subset with the property that one of them divides the other. An integer k divides an integer j if, when we divide j by k, the remainder is equal to 0. Put another way, k divides j if j is a multiple of k, that is, if $j = qk$, where q is some integer. To solve this problem, let us write each of the $n+1$ integers in the form of a power of 2 times an odd factor. For example, we would write

$$84 = 2^2 \times 21 \quad , \quad 26 = 2^1 \times 13 \quad , \quad 7 = 2^0 \times 7.$$

Of course, all of the $n+1$ odd factors obtained in this way are between 1 and $2n$. But there are only n odd numbers between 1 and $2n$, and so there are only n possibilities for these $n+1$ odd factors. Therefore, we can apply the pigeonhole principle, which tells us that at least two of the odd factors must be equal. (Here we have n "boxes" and $n+1$ "balls.") This means that two of the integers in the subset must have the form $y = 2^a c$ and $z = 2^b c$ where c is the common odd factor. But either $a < b$, in which case

$$\frac{z}{y} = \frac{2^b c}{2^a c} = 2^{b-a} = \text{ an integer}$$

or $a \geq b$, in which case

$$\frac{y}{z} = \frac{2^a c}{2^b c} = 2^{a-b} = \text{ an integer}$$

which shows either y divides evenly into z or z divides evenly into y. In either case, one of the numbers y or z divides the other, which is what we wanted to show.■ Let us consider another example of the same type as the previous one.

Example #16: More Number Theory

Given any collection of n integers, not necessarily distinct, let us show that there is some subcollection of these integers whose sum is divisible by n. Toward this end, let us denote the integers by a_1, a_2, a_3, ..., a_n, and let us consider the sums

$$\begin{aligned} s_1 &= a_1 \\ s_2 &= a_1 + a_2 \\ s_3 &= a_1 + a_2 + a_3 \\ &\vdots \\ s_n &= a_1 + a_2 + a_3 + \cdots + a_n \end{aligned}$$

Note that these are not all of the possible sums of subcollections of the integers a_1, a_2, a_3, ..., a_n, but these are the only ones that we need. Let us divide each of these sums by n and denote the quotients by q_1, q_2, q_3, ..., q_n, so that

$$\begin{aligned} s_1 &= a_1 = q_1 n + r_1 \\ s_2 &= a_1 + a_2 = q_2 n + r_2 \\ s_3 &= a_1 + a_2 + a_3 = q_3 n + r_3 \\ &\vdots \\ s_n &= a_1 + a_2 + a_3 + \cdots + a_n = q_n n + r_n \end{aligned}$$

where the remainders satisfy $0 \leq r_1, r_2, ..., r_n \leq n - 1$. Now, if one of the remainders, say r_j, is equal to zero, then the corresponding sum s_j would satisfy $s_j = q_j n$ and hence be divisible by n. Thus we would have found a sum, namely s_j, which is divisible by n. If none of the remainders are equal to zero, then we can still find a sum that is divisible by n. To see this we note that in this case, all of the remainders $r_1, r_2, r_3, ..., r_n$ would be strictly between 0 and n so that $1 \leq r_1, r_2, ..., r_n \leq n - 1$ But there are only $n - 1$ integers strictly between 0 and n, and so we can use the pigeonhole principle to conclude that at least two of the n remainders must be equal. Let us suppose without any lost in generality that

$r_j = r_k$, where j and k are integers between 1 and n, and $j < k$. Then since

$$s_j = a_1 + a_2 + a_3 + \cdots + a_j = q_j n + r_j$$

and

$$s_k = a_1 + a_2 + a_3 + \cdots + a_j + a_{j+1} + \cdots + a_k = q_k n + r_k$$

we have

$$s_k - s_j = a_{j+1} + \cdots + a_k = (q_k - q_j)n + (r_k - r_j) = (q_k - q_j)n$$

since $r_j = r_k$, and this shows that the sum

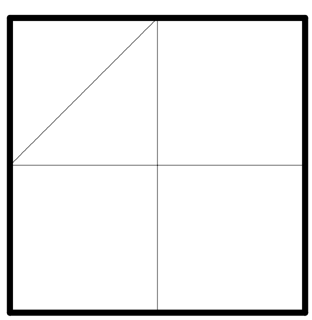
$$a_{j+1} + \cdots + a_k$$

is divisible by n. In either case, that is, whether some remainder is equal to zero or not, there is a subcollection of a_1, a_2, a_3, ..., a_n, whose sum is divisible by n, which is what we wanted to show. ■

The pigeonhole principle can also be used to solve certain types of geometric problems. Let us consider an example of this type.

Example #17: A Geometry Problem

What is the largest number of points that can be placed in a square whose side has length 2, in such a way that no two points are a distance of $\sqrt{2}$ or less from each other? Consider the square shown below,



A Square of side length 2

which we have divided into 4 smaller squares, each with side length equal to 1. According to the Pythagorean Theorem, the diagonal of each of the small squares has length $\sqrt{2}$. Thus, if we are going to place points in the large square in such a way that no 2 points are within a distance of $\sqrt{2}$, we cannot place more than 1 point in each of the smaller squares. Therefore, according to the pigeonhole principle, we cannot place more than 4 points in the large square. Notice that the pigeonhole principle does not tell us that we can place 4 points in the square. It only tells us that we cannot place more than 4 points. However, if we place 1 point in each of the 4 corners of the large square, then no 2 of these points will be within a distance of $\sqrt{2}$ from each other. Hence, the answer to the question is 4. ■

5. Permutations

Let us consider the counting problem in Example #2. Recall that the problem is to count the number of possible ways in which a computer operator can order 9 computer programs. Now, if we denote the programs by the symbols p_1, p_2, p_3, ..., p_9 , then any order of the programs corresponds to an ordered arrangement of these symbols. For example, the arrangement

$$p_6p_2p_1p_3p_9p_4p_8p_7p_5$$

is one possible order, and the arrangement

$$p_2p_9p_1p_3p_8p_4p_5p_7p_6$$

is another possible order. Ordered arrangements of objects occur very frequently in combinatorics, and they deserve a special name.

An *ordered* arrangement of a set of objects is called a *permutation* of the objects. Note that an *unordered* arrangement of a set of objects is just the set itself. If there are n objects in the permutation, we say that the permutation has size n, or is an n-permutation. Thus, in above example, each order of the programs is simply a permutation of the 9 programs

$$p_1p_2p_3p_4p_5p_6p_7p_8p_9,$$

and the question in Example #2 is "How many permutations are there of 9 programs?" It would be nice to have a formula for the number of permutations of any set of n objects, but before deriving such a formula, let us consider a few more examples of permutations.

Example #18: Simple Examples

Given a set of one object $\{a\}$, there is only one permutation, namely $\{a\}$. Given a set of two objects $\{a, b\}$, there is are two permutations, namely $\{ab, ba\}$. Given a set of three objects $\{a, b, c\}$, there is are six permutations, namely $\{abc, acb, bac, bca, cab, cba\}$. Given a set of four objects $\{a, b, c, b\}$, there is are 24 permutations, namely

$$\begin{gathered}\{abcd, abdc, acbd, acdb, adbc, adcb. \\ bacd, badc, bcad, bcda, bdac, bdca, \\ cabd, cadb, cbad, cbda, cdab, cdba, \\ dabc, dacb, dbac, dbca, dcab, dcba\},\end{gathered}$$

showing that there are 24 permutations of size 4. ■

It is not hard to derive a formula for the number of permutations of n objects, if we use the multiplication rule. In fact, the number of permutations of a set of n objects is the product of the first n positive integers, that is

$$n(n-1)(n-2)\cdots(2)(1) = n!$$

and the proof of this is as follows. Forming a permutation of n objects can be thought of as performing a sequence of n tasks. The first task is to choose the first object for the permutation, and this can be done in n different ways. The second task is to choose the second object for the permutation, and this can be done in $n-1$ different ways, since there are only $n-1$ objects left after the first task has been performed. The third task is to choose the third object for the permutation, and this can be done in $n-2$ different ways since there are only $n-2$ objects left after the first and second tasks has been performed. This continues until we reach the nth task, which is to choose the nth object for the permutation, and this can be done in only one way, since there is only one object left at this point. Now, the *number* of ways to perform the ith task in the sequence does not depend on how the previous $i-1$ tasks were performed. We emphasize the word "number" here because the outcome for the ith task does depend on how the previous tasks were performed. It is only the number of possible outcomes that does not. Therefore, we can apply the multiplication rule, which tells us that there are $n(n-1)(n-2)\cdots(2)(1) = n!$ different ways to perform the entire sequence of n tasks. That is,

there are $n!$ different permutations of n distinct objects.

Using this result, we can now quickly solve the problem (Example #2) of ordering the 9 computer programs as $9! = 362,880$.

We see then that there are $n!$ *ordered* arrangement of a set of n objects. Note that there is only 1 *unordered* arrangement of a set of n objects since this is just the set itself.

Example #19: Permutations

Suppose that we want to arrange 6 different math books, 4 different computer science books, and 3 different chemistry books on a single bookshelf. (a) In how many ways can this be done? (b) In how many ways can this be done if the math books must come first, then the computer science books, and finally the chemistry books? (c) In how many ways can this be done if all books of the same subject must be kept together?

To answer part (a), we simply observe that each arrangement of the 13 books on the bookshelf is a permutation of the books. Hence there are $13! = 6,227,020,800$ possible arrangements of the books. As to part (b), there are 6! different ways to arrange the math books, there are 4! different ways to arrange the computer science books, and there are 3! different ways to arrange the chemistry books. Hence, according to the multiplication rule, there are $6! \times 4! \times 3! = 103,680$ different ways to arrange the books on the bookshelf, under the conditions given in part (b).

Finally, we can use the results of part (b), along with the multiplication rule, to solve part (c). For if we let T_1 be the task of deciding the order in which the three subjects-math, computer science, and chemistry-appear on the shelf, and if we let T_2 be the task of ordering the actual books, once the order of the subjects has been decided, then arranging the books on the bookshelf is the same as performing the sequence of tasks T_1 and T_2. Now, there are $3! = 6$ ways of performing task T_1 and there are $103,680$ ways of performing task T_2 as shown in part (b). Hence, according to the multiplication rule, there are

$$6 \times 103,680 = 622,080$$

different ways to arrange the books on the bookshelf, under the conditions given in part (c). ■

Counting What You Don't Want Might Be Easier

For many problems where one is counting the number of ways in which something can be done, with certain restrictions, it might be easier to count the number of ways that same thing can be done when these restrictions are violated, and then subtract that from the number of ways that same thing can be done without any of the restrictions. In other words, when considering a counting problem, it is sometimes easier to count what you do not want, and then subtract that from some total, rather than to count what you do want directly. Mathematically, if A is a set and $\bar{A}$ is the compliment of A in S, *i.e.*,

$$\bar{A} = \{x \in S | x \notin A\},$$

then

$$A \cap \bar{A} = \emptyset \qquad \text{and} \qquad A \cup \bar{A} = S$$

which says that

$$|A| + |\bar{A}| = |S| \qquad \text{or} \qquad |A| = |S| - |\bar{A}|.$$

Let us now look at some examples of this.

Example #20: Counting What You Don't Want Might Be Easier

How many ways can we order 9 computer programs p_1, p_2, p_3, ..., p_9, if program p_2 cannot immediately follow program p_1? This is an example of a type of problem where it is much easier to count what we do not want and subtract that from the total. In this case, there are a total of 9! unrestricted orderings and so if we subtract from 9! the number of orderings in which program p_2 does immediately follow program p_1, then we will have the answer to our question. In order to compute the number of orderings in which program p_2 does follow program p_1, we reason as follows. As long as program p_2 must follow program p_1 we can think of these two programs as "tied together" into one program $[p_1p_2]$. In effect then, there are only 8 different programs, $[p_1p_2]$, p_3, p_4, ..., p_9, and so there are 8! orderings in which p_2 follows p_1. Thus, there are

$$9! - 8! = 362,880 - 40,320 = 322,560$$

orderings in which program p_2 does not follow program p_1.■

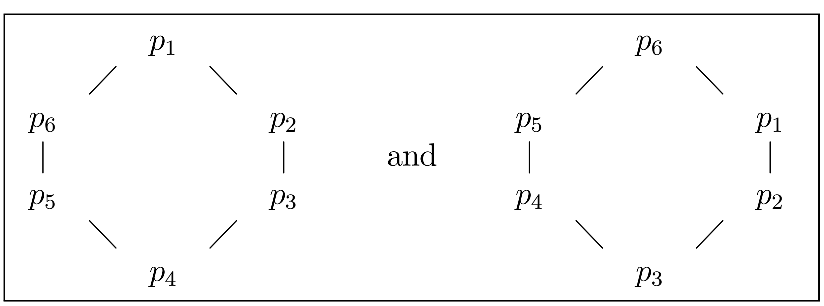
Example #21: Counting What You Don't Want Might Be Easier

How many bytes are there that contain at least two 1's? A byte is a binary word of length 8. Rather than try to count this number directly, it is much easier in this case to count the number of bytes that contain less than two 1's, which are the bytes we do not want, and subtract that number from $2^8 = 256$, which is the total number of bytes. Now, it is easy to count the number of bytes that have less than two 1's. After all, there is only one byte that contains no 1's, namely the byte with all 0's (00000000). Also, there are only 8 bytes that contain exactly one 1, namely: 10000000, 01000000, 00100000, 00010000, 00001000, 00000100, 00000010, and 00000001. Hence, we see that there are $1 + 8 = 9$ bytes that contain less than two 1's, and so there are $2^8 - 9 = 247$ bytes that contain at least two 1's. ■

Circular Permutations

We know that there are $n!$ different ways to arrange n people in a line, since each arrangement corresponds to a permutation of the n people. But how many ways are there to arrange n people in a circle?

The answer is not $n!$. The reason is that two circular arrangements are considered to be the same if one can be obtained from the other by a rigid rotation. For example, the two circular arrangements shown below



are the same since the second one is just the first one rotated clockwise by 60°. Our problem is to figure out a way to avoid counting these two arrangements as being different. To do this, let us denote the n people by p_1, p_2, p_3, ..., p_n, and imagine that we are putting these n people in n seats around a round table. Then we can avoid the problem of rotations by simply agreeing to always put the first person, p_1, at the head of the table, as pictured below.

$$\begin{array}{ccc} & p_1 & \\ / & & \backslash \\ \backslash & & / \\ & \_ & \end{array}$$

Then we can fill the remaining $n-1$ seats with the remaining $n-1$ people and there are $(n-1)!$ ways of doing this. Circular arrangements such as these are known as *circular permutations*, and the previous example shows that there are $(n-1)!$ circular permutations of n objects. The figure below give the $3! = 6$ circular permutations of 4 objects

$$\begin{array}{ccccccccccc} & & p_1 & & & & & & p_1 & & \\ & / & & \backslash & & & & / & & \backslash & \\ p_4 & & & & p_2 & \text{and} & p_3 & & & & p_2 \\ & \backslash & & / & & & & \backslash & & / & \\ & & p_3 & & & & & & p_4 & & \end{array}$$

and

$$\begin{array}{ccccccccccc} & & p_1 & & & & & & p_1 & & \\ & / & & \backslash & & & & / & & \backslash & \\ p_4 & & & & p_3 & \text{and} & p_2 & & & & p_3 \\ & \backslash & & / & & & & \backslash & & / & \\ & & p_2 & & & & & & p_4 & & \end{array}$$

and

$$\begin{array}{ccccccccccc} & & p_1 & & & & & & p_1 & & \\ & / & & \backslash & & & & / & & \backslash & \\ p_3 & & & & p_4 & \text{and} & p_2 & & & & p_4 \\ & \backslash & & / & & & & \backslash & & / & \\ & & p_2 & & & & & & p_3 & & \end{array}$$

6. More on Permutations

It often happens that we have a set S of size n, but that we wish to form permutations using only k of the objects at a time, where $k \leq n$. In this section, we want to find a formula for the number of permutations of size k, taken from a set of size n.

Example #22: Some Simple Examples

The permutations of size 2, taken from the set of 4 letters $\{a, b, c, d\}$ are listed as

ab	ac	ad
ba	bc	bd
ca	cb	cd
da	db	dc

and so there are 12 permutations of size 2 from a set of 4 objects.

The 24 permutations of size 3, taken from this same set are listed as

abc	acb	bac	bca	cab	cba
abd	adb	bad	bda	dab	dba
adc	acd	dac	dca	cad	cda
bcd	bdc	cbd	cdb	dcb	dbc

For convenience, we have organized these permutations into four groups according to which letters they involve. ■

Words of Length k Over a Finite Alphabet - No Repeated Letters

As you can see from these examples, if Σ is a finite set, then a permutation of size k, taken from the set Σ, is exactly the same as a word of length k over Σ that has no repeated letters. Therefore, when we count the number of permutations of size k, we will also be counting the number of words of length k that have no repeated letters. Using the multiplication rule, we can easily obtain a formula for the number of permutations of size k, taken from a set of size n. Let us denote this number by $P(n, k)$ and to determine a formula for $P(n, k)$, we note that forming a permutation of size k from a set of size n, can be thought of as performing a sequence of k tasks. The first task is to choose the first element for the permutation, and there are n ways to do this. The second task is to choose the second element for the permutation, and there are $n-1$ ways to do this, and so on. The kth task is to choose the kth element for the permutation, and since there are $n-(k-1)$ objects left to choose from at this point, there are $n-(k-1)=n-k+1$ ways to perform this last task. Hence, according to the multiplication rule, we have

$$P(n,k)=n(n-1)(n-2)\cdots(n-k+1) \tag{5a}$$

which is denoted by $(n)_k$, and we may write as

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

or

$$P(n,k)=(n)_k=\frac{n!}{(n-k)!} \tag{5b}$$

for $0\leq k\leq n$. We should note that

$$(n)_k=n\times(n-1)_{k-1} \tag{6}$$

which is useful when computing $(n)_k$ for large n and k.

Example #23: Sending Messages

In order to send messages from one boat to another, flags are sometimes used. Suppose that a certain boat has 10 different flags and 1 flagpole. If each ordered arrangement of 3 flags on the flagpole represents a different message, how many messages are possible? Since there is 1 message for each permutation of size 3, taken from the 10 flags, there are

$$P(10,3)=(10)_3=10\times9\times8=720$$

possible messages. ■

Example #24: License Plates

Suppose that a certain state has license plates consisting of 3 letters followed by 3 digits. How many license plates are there in this state that do not have a repeated letter or a repeated digit? Forming such a license plate can be thought of as performing two tasks. The first task is to determine the 3 letters, and there are $P(26,3)$ ways of doing this, since there are $P(26,3)$ different permutations of size 3, taken from the 26 letters of the alphabet. The second task is to choose the 3 digits, and there are $P(10,3)$ ways to do this. Hence, according to the multiplication rule, there are

$$P(26,3)P(10,3) = (26)_3(10)_3 = (26)(25)(24) \times (10)(9)(8) = 15600 \times 720$$

or $11,232,000$ such license plates. It is interesting to compare this number with the total number of license plates possible when repetitions are allowed. According to the multiplication rule, there are

$$26 \times 26 \times 26 \times 10 \times 10 \times 10 = 17,576,000$$

license plates consisting of 3 letters followed by 3 digits. This says that

$$\frac{11,232,000}{17,576,000} = \frac{108}{169} \simeq 0.64$$

or 64% of all license plates will not have either a repeated letter or a repeated digit! Of course this also says that $100\% - 64\% = 36\%$ of all license plates will have either a repeated letter or a repeated digit! ■

7. Combinations

In the previous two sections, we have been studying ordered arrangements of objects. Now, we want to study *unordered* arrangements. Actually, we have already studied some unordered arrangements, since an unordered arrangement of the elements of a set S is nothing more than a subset of S. In this section, we want to consider certain combinatorial questions relating to subsets that we did not discuss. For example, we want to determine how many subsets of size k can be formed from a set of size n. We do know that a set of size n has 2^n subsets of all sizes. Let us begin with a definition.

Let S be a set. Then an unordered arrangement of k elements of S, that is, a subset of S of size k, is also called a *combination* of size k, or a k-combination taken from S. We will denote the number of combinations of size k taken from a set of size n, by $C(n, k)$. Thus, $C(n, k)$ is the number of subsets of size k taken from a set of size n.

Example #25: A Simple Listing

For purposes of comparison, let us make a list of all of the 3-combinations and 3-permutations of the set $S = \{1, 2, 3, 4\}$. These are shown in the table below.

3-permutations of $S = \{1, 2, 3, 4\}$	3-combinations of $S = \{1, 2, 3, 4\}$
$\{123, 132, 213, 231, 312, 321\}$	$\{123\}$
$\{124, 142, 214, 241, 412, 421\}$	$\{124\}$
$\{134, 143, 314, 341, 413, 431\}$	$\{134\}$
$\{234, 243, 324, 342, 423, 432\}$	$\{234\}$

Notice that we have arranged the permutations into four groups, each of which consists of those permutations that use the same three integers. Also, we have placed each combination next to the group of permutations that involves the same three integers, and since order does not matter in combinations, we simply choose to order the digits from smallest to largest. ■

As it turns out, there is a simple relationship between the numbers $C(n, k)$ and $P(n, k)$, and since we already have a formula for $P(n, k)$, we can use this relationship to obtain a formula for $C(n, k)$. This relationship simply states that

$$P(n, k) = k! \times C(n, k) \qquad \text{or} \qquad C(n, k) = \frac{P(n, k)}{k!} \tag{7}$$

for $0 \leq k \leq n$. To prove this let us first consider the case $k = 0$. Since $C(n, 0)$ is the number of subsets of size 0 of a set of size n, we know that $C(n, 0) = 1$ since only the empty set is a subset with size 0. We also have $P(n, 0) = 1$, and since $0! = 1$, we do indeed have

$$C(n, 0) = \frac{P(n, 0)}{0!}$$

Now let us consider the case $1 \leq k \leq n$. To make the argument easier to follow, we will determine the number of k-combinations of the set $S = \{1, 2, ..., n\}$. Certainly, this number will be the same as the number of k-combinations of any set of size n. Let us imagine that we have made a list, similar to the one in the previous example, of the $P(n,k)$ permutations and the $C(n,k)$ combinations of size k, taken from the set S. Thus, the permutations are grouped together according to which integers they involve, and each combination is placed next to that group of permutations that involves the same integers. Now, each group of permutations in this list contains exactly $k!$ permutations, since it contains all of the permutations of a particular choice of k of the integers $1, 2, 3, ..., n$. For instance, the first group in the example above contains the $3! = 6$ permutations of the three integers 1, 2, and 3 and the second group contains the $3! = 6$ permutations of the three integers 1, 2, and 4. Therefore, and this is the key to the proof, each of the $C(n,k)$ combinations in the list corresponds to exactly $k!$ permutations. This means that if we multiply $C(n,k)$ by $k!$, we must get the total number of permutations $P(n,k)$. In symbols,we have $k!C(n,k) = P(n,k)$. Dividing by $k!$, we arrive at Equation (7).

Putting in the expression for $P(n,k)$ in Equation (7), we have

$$C(n,k) = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}$$

which we denote as

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

and is known as the binomial coefficient "n choose k". Thus we find that

$$C(n,k) = \binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{8}$$

and $C(n,k)$ equals the number of subsets of size k of a set of size n, which is also the number of *unordered* arrangements of k objects taken from a set of n objects. It should be noted that

$$C(n,k) = C(n,n-k) \qquad \text{or} \qquad \binom{n}{k} = \binom{n}{n-k} \tag{9a}$$

and $C(n,k) = C(n-1,k) + C(n-1,k-1)$, or

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}. \tag{9b}$$

Example #26: Number of 5-Card Poker Hands

We can now answer the question posed at the beginning of this chapter. The number of 5-card poker hands that are possible out of a set of 52 distinct playing cards is now

$$C(52,5) = \binom{52}{5} = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} = 2,598,960.$$

Note that of these, a royal flush is either of the four hands

$$\{(10\text{C}, \text{JC}, \text{QC}, \text{KC}, \text{AC}), (10\text{D}, \text{JD}, \text{QD}, \text{KD}, \text{AD}),$$
$$(10\text{H}, \text{JH}, \text{QH}, \text{KH}, \text{AH}), (10\text{S}, \text{JS}, \text{QS}, \text{KS}, \text{AS})\}.$$

We shall see later that this implies that the probability of being dealt a royal flush is

$$\frac{4}{2,598,960} = \frac{1}{649,740},$$

showing that your chances of getting a royal flush is 1 in every $649,740$ hands. If you play poker every day of the week and get an average of 100 hands per night, this amounts to playing for $6,497.40$ days or 17.8 years before you will (on average) get a royal flush! ■

Example #27: Number of Committees

A certain club has 5 male and 7 female members. How many ways are there to form a 7-person committee consisting of 3 men and 4 women? There are $C(5,3)$ ways of choosing 3 men to serve on the committee, and there are $C(7,4)$ ways of choosing 4 women to serve on the committee. Using the multiplication principle, there are

$$C(5,3) \times C(7,4) = \binom{5}{3} \times \binom{7}{4} = 10 \times 35 = 350$$

ways to form the 7-person committee. ■

Example #28: Computing What You Don't Want

How many subsets of size 2 are there of the set $S = \{1, 2, 3, ..., n\}$ that do not consist of two consecutive integers? For example, the set $\{5, 6\}$ does consist of two consecutive integers. The simplest way to answer this question is to count the number of subsets of size 2 that do consist of two consecutive integers and then subtract that number from $C(n, 2)$, which is the total number of subsets of size 2. Clearly, there are $n-1$ subsets of size 2 consisting of consecutive integers, namely, $\{1,2\}, \{2,3\}, \{3,4\}, ..., \{n-1, n\}$. Hence there are

$$C(n,2)-(n-1)=\binom{n}{2}-(n-1)=\frac{n(n-1)}{2}-(n-1)=\frac{1}{2}(n-1)(n-2)$$

subsets of $\{1, 2, 3, ..., n\}$ that do not consist of two consecutive integers. ■

Example #29: It's Easy To Make A Mistake

How many ways can 12 people be partitioned into 3 groups of 4 people each? Let us denote the people by p_1, p_2, p_3, ..., p_{12}. We might try to reason as follows. The number of ways of choosing 4 people for the first group, call it G_1, is $C(12, 4)$; the number of ways of choosing 4 people for the second group, G_2 , is $C(8, 4)$ since there are only 8 people left after the first group is chosen; and the number of ways of choosing 4 people for the last group, G_3, is $C(4, 4)$ since there are only 4 people left after the first and second groups are chosen. Hence, according to the multiplication rule, there are

$$C(12,4)\times C(8,4)\times C(4,4)=\binom{12}{4}\times\binom{8}{4}\times\binom{4}{4}=34,650$$

ways of partitioning the 12 people into 3 groups of equal size. Unfortunately, there is a flaw in this reasoning. What we have counted is the number of ways to partition 12 people into 3 distinct groups, G_1, G_2 and G_3 of equal size. Thus, for instance, the partition

$$\mathrm{G}_1=\{1,2,3,4\} \quad , \quad \mathrm{G}_2=\{5,6,7,8\} \quad , \quad \mathrm{G}_3=\{9,10,11,12\}.$$

is counted separately from the partition

$$\mathrm{G}_1=\{5,6,7,8\} \quad , \quad \mathrm{G}_2=\{1,2,3,4\} \quad , \quad G_3=\{9,10,11,12\}$$

even though, from the way that the question is worded, we do not want to count these partitions as being different. If we did, then the answer would be

$$C(12,4)\times C(8,4)\times C(4,4)=\binom{12}{4}\times\binom{8}{4}\times\binom{4}{4}=34,650$$

as computed above. But all is not lost. We have simply over-counted the number we are looking for and all we have to do is determine how much we have over-counted. To do this, we simply observe that each of the partitions of the 12 people into 3 groups of equal size is counted 3! times in the number $C(12,4) \times C(8,4) \times C(4,4)$. For example, the partition

$$\{1,2,3,4\} \quad , \quad \{5,6,7,8\} \quad , \quad \{9,10,11,12\}$$

is counted $3! = 6$ times as follows

$G_1 = \{1,2,3,4\}$	$G_2 = \{5,6,7,8\}$	$G_3 = \{9,10,11,12\}$
$G_1 = \{1,2,3,4\}$	$G_2 = \{9,10,11,12\}$	$G_3 = \{5,6,7,8\}$
$G_1 = \{5,6,7,8\}$	$G_2 = \{1,2,3,4\}$	$G_3 = \{9,10,11,12\}$
$G_1 = \{5,6,7,8\}$	$G_2 = \{9,10,11,12\}$	$G_3 = \{1,2,3,4\}$
$G_1 = \{9,10,11,12\}$	$G_2 = \{1,2,3,4\}$	$G_3 = \{5,6,7,8\}$
$G_1 = \{9,10,11,12\}$	$G_2 = \{5,6,7,8\}$	$G_3 = \{1,2,3,4\}$

Since each of the ways of partitioning 12 people into 3 equal size groups is counted exactly $3! = 6$ times in the number $C(12,4) \times C(8,4) \times C(4,4)$, the actual number of ways to partition the 12 people is

$$\frac{C(12,4) \times C(8,4) \times C(4,4)}{3!} = \frac{34,650}{6}$$

or $5,775$. ■

8. Properties of the Binomial Coefficients - For Reading

There are perhaps more identities known about the binomial coefficients than other coefficients. In this section, we just present some of these and leave the derivations to the interested student. The first is known as Pascal's identity which says that

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \tag{10a}$$

for $0 \le k \le n-1$. For convenience, we define

$$\binom{n}{k} = 0$$

when $k < 0$ and when $k > n$, and by writing

$$\binom{n}{k} \qquad \text{as} \qquad \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!},$$

it allows us to define

$$\binom{a}{k} = \frac{a(a-1)(a-2)\cdots(a-k+1)}{k!} \tag{10b}$$

for a not an integer. We also have the binomial theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \tag{10c}$$

for $n = 0, 1, 2, ...$ and as a special case of this, we set $x = y = 1$, we have

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n. \tag{10d}$$

By setting $x = -1$ and $y = 1$, we also have

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0. \tag{10d}$$

Another useful identity is

$$\sum_{j=k}^{n} \binom{j}{k} = \binom{n+1}{k+1} \tag{10e}$$

for $0 \le k \le n$. The student is encouraged to prove these and to explore other binomial coefficient identities.

9. The Multinomial Coefficients

The binomial formula tells us how to expand the powers of the binomial $x + y$. In this section, we want to develop a similar formula for expanding the powers of any multinomial

$$x_1 + x_2 + \cdots + x_k.$$

Let us begin with the case of a trinomial $x + y + z$. As we did in the previous section, the first step is to write out the product

$$(x + y + z)^n = (x + y + z)(x + y + z) \cdots (x + y + z)$$

where there are n factors on the right side. Now, if we imagine that the product on the right is completely expanded, then all of the terms will have the form $x^i y^j z^k$, where i, j, and k are nonnegative integers with the property that $i + j + k = n$. After all, each term in this expansion is formed by choosing either an x, y, or z from each of the n factors and multiplying these variables together. Whenever we choose i x's, j y's, and k z's (where $i + j + k = n$), then the resulting term is $x^i y^j z^k$. The problem we must face now is to determine, for each possible choice of i, j, and k, the number of times the term $x^i y^j z^k$ appears in the expansion. But this is not hard to do. For there are

$$\binom{n}{i}$$

ways to choose i of the n factors to contribute an x, and once this has been done, there are

$$\binom{n-i}{j}$$

ways to choose j of the remaining factors to contribute a y and finally, there are

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

ways to choose k of the remaining factors to contribute a z. Thus, according to the multiplication rule, the number of times that the term $x^i y^j z^k$ appears in the expansion of $(x + y + z)^n$ is

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

which reduces to

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

or, since $i + j + k = n$, we have $n - i - j - k = 0$, and so

$$\binom{n}{i} \times \binom{n-i}{j} \times \binom{n-i-j}{k} = \frac{n!}{i!j!k!}$$

which we denote by

$$\frac{n!}{i!j!k!} = \binom{n}{i, j, k}.$$

Hence, if we collect like terms in the expansion of $(x + y + z)^n$, we will get the term

$$\binom{n}{i, j, k} x^i y^j z^k.$$

Therefore, the expansion of $(x + y + z)^n$ is just the sum of terms of this form, as i, j, and k range over all nonnegative integers with $i + j + k = n$. Let us put this into a theorem, which is known as the *trinomial* theorem, which says that

$$(x + y + z)^n = \sum_{\substack{0 \leq i,j,k \\ i+j+k=n}} \binom{n}{i, j, k} x^i y^j z^k. \tag{11}$$

Note that the binomial theorem can be stated as

$$(x + y)^n = \sum_{\substack{0 \leq i,j \\ i+j=n}} \binom{n}{i, j} x^i y^j \tag{12}$$

where

$$\binom{n}{i, j} = \frac{n!}{i!j!}.$$

It is now not hard to generalize this to the multinomial theorem

$$(x_1 + x_2 + \cdots + x_k)^n = \sum_{\substack{0 \leq n_1, n_2, \ldots, n_k \\ n_1 + n_2 + \cdots + n_k = n}} \binom{n}{n_1, n_2, \ldots, n_k} x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} \tag{13}$$

where

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}. \tag{14}$$

A Combinatorial Interpretation of the Multinomial Coefficient

In order to get a feeling for what multinomial coefficients count, let us look at what binomial coefficients count in a slightly different way. The binomial coefficient

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

can be thought of as counting the number of ways of dividing a set S of size n into two disjoint subsets A_1 and A_2, where A_1 has size k and A_2 has size $n-k$. After all, choosing a subset of S amounts to the same thing as dividing S into two disjoint subsets which are the subset of "chosen" elements and the subset of "leftover" elements. Now, multinomial coefficients also count the number of ways of dividing a set into disjoint subsets, but in general more than just two subsets. In fact, the multinomial coefficient

$$\binom{n}{n_1, n_2, \ldots, n_k}$$

where $n_1, n_2, \ldots, n_k$ are nonnegative integers satisfying $n_1 + n_2 + \cdots + n_k = n$ counts the number of ways of dividing a set S of size n into k *mutually disjoint* subsets A_1, A_2, A_3, ..., A_k, where A_j has size n_j, for all $j = 1, 2, 3, ..., k$. Note that mutually disjoint subsets A_1, A_2, A_3, ..., A_k means that $A_i \cap A_j = \emptyset$ for $i \neq j$.

To prove this, we note that dividing a set S of size n into k mutually disjoint subsets A_1, A_2, A_3, ..., A_k, where A_j has size n_j, for all $j = 1, 2, 3, ..., k$ amounts to performing a sequence of k tasks. The first task is to choose n_1 of the elements of S to form the subset A_1, the second task is to choose n_2 of the remaining $n - n_1$ elements of S to form the second subset A_2, and so on. The kth task is to choose n_k, of the remaining $n - n_1 - n_2 - \cdots - n_{k-1}$ elements of S to form the set A_k. According to the multiplication rule, the number of ways to perform this sequence of k tasks is

$$\binom{n}{n_1} \times \binom{n-n_1}{n_2} \times \binom{n-n_1-n_2}{n_3} \times \cdots \times \binom{n-n_1-n_2-\cdots-n_{k-1}}{n_k}$$

and this reduces to just

$$\binom{n}{n_1, n_2, n_3, \ldots, n_k} = \frac{n!}{n_1! n_2! n_3! \cdots n_k!}$$

In other words, there are

$$\binom{n}{n_1, n_2, n_3, \ldots, n_k}$$

ways to divide the set S into subsets A_1, A_2, A_3, ..., A_k of the desired size. This completes the proof.

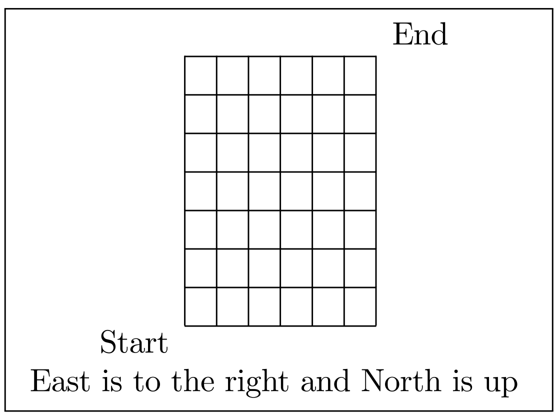
The student should now show that if $\Sigma = \{a_1, a_2, .., a_k\}$ is an alphabet, there are

$$\binom{n}{n_1, n_2, n_3, \ldots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!} \tag{15}$$

words of length n that contains exactly n_1 a_1's, n_2 a_2's, n_3 a_3's and so on up to n_k a_k's.

Example #30: Example #3 Revisited

Recall that Example #3 asked the following question. There are 13 one-way streets arranged in a rectangular grid with 6 of the streets going from west to east and 7 of the streets going from south to north. A person wants to walk the 13 blocks from the most southwest point on this map to the most northeast point, always heading either east or north as illustrated in the following figure.



The problem is to count the number of ways that this can be done. To answer this, we use a one-to-one correspondence. We may map every walk to a sequence of E's (east) and N's (north). For example, if you walk the 13 blocks from the most southwest point on this map to the most northeast point, by walking 6 blocks directly east and then walk 7 blocks directly north, we have the walk

$$EEEEEENNNNNNN.$$

Another walk would be

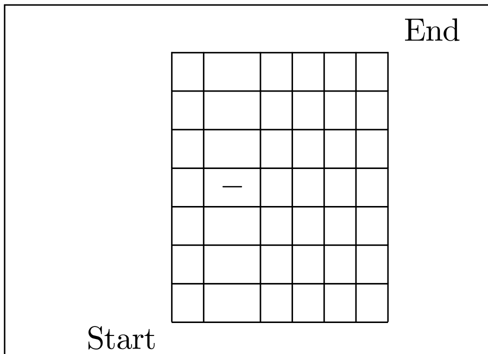
$$ENENENENENENN.$$

We see then that each walk along the streets correspond to a word of length 13 from the alphabet $\Sigma = \{E, N\}$ with 6 E's and 7 N's. This leads to

$$\binom{13}{6,7} = \frac{13!}{6!7!} = 1,716$$

possible walks. Note that if these streets were two-way, then the paths we are interested in can be referred to as the number of *shortest* paths that a person can take starting from "Start" and ending at "End", since shortest paths involve no back-tracking and hence must consist of a total of 13 blocks.

As a slight variation in this problem, suppose that the person wants to walk from "Start" to "End" with the condition that they must pass through the block shows (by the dash) in the figure below.



East is to the right and North is up

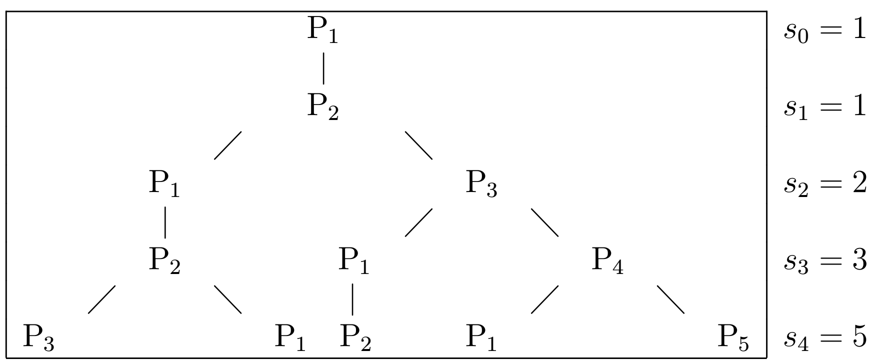
Treating the "Start" as the point $(0,0)$ and the "End" as the point $(6,7)$, then, we may simply express this as two smaller paths, one that takes you from $(0,0)$, the "Start" to the point $(1,3)$, then from $(1,3)$ to $(2,3)$ and one that takes you from the point $(2,3)$ to the "End", which is the point $(6,7)$. This then leads to

$$\binom{4}{1,3} \times 1 \times \binom{8}{4,4} = \frac{4!}{1!2!} \times 1 \times \frac{8!}{4!4!} = 840$$

such paths. Of course, if we want to walk to avoid the path from $(1,3)$ to $(2,3)$, then there are $1716 - 840 = 876$ such paths. ■

10. An Introduction to Recurrence Relations

In the year 1202, a mathematician named Leonardo Fibonacci (1170?-1250?) posed the following simple counting problem. Let us assume that pairs of rabbits do not produce offspring during their first month of life, but after their first month, they produce a new pair of offspring each month. If we start with one pair of newborn rabbits, and if we assume that no rabbits die, how many pairs of rabbits will there be after n months? To solve this problem, let us denote the number of pairs of rabbits at the *end of* the nth month by s_n. Then, of course, $s_0 = 1$ and $s_1 = 1$, since rabbits cannot reproduce until they have been alive for 2 months. After 2 months, the first pair of rabbits produces a pair of offspring, and so $s_2 = 2$. This includes the original pair and the pair that was produced. At the end of the third month, there will be 3 pairs of rabbits, so that $s_3 = 3$. This includes the original pair, the first pair that was produced and a new pair that is produced from the original pair. We may represent this as follows, where P_k represents a pair of rabbits that are k months old.



For example, at the end of the 3rd month, there are three pairs of rabbits so that $s_3 = 3$, one pair that is 2 months old (P_2), one pair that is one month old (P_1) and one pair that is 4 months old (P_4). In general, we can obtain information about the number s_n by reasoning as follows. At the end of the nth month, the s_{n-1} pairs of rabbits that were alive at the end of the previous month will still be alive since we are assuming that no rabbits die. This contributes s_{n-1} pairs of rabbits to the total number of pairs for the nth month. But, there will also be some newborn pairs. In fact, each of the s_{n-2} pairs of rabbits that were alive 2 months prior to the nth month, being at least 2 months old themselves, will bear a new pair of rabbits. This contributes s_{n-2} additional pairs of rabbits to the total for the nth month. Hence, we have

$$s_n = s_{n-1} + s_{n-2} \tag{16a}$$

for $n = 2, 3, 4, ...$ In words, Equation (16a) says that any term in the sequence s_0, s_1, s_2, ..., s_n, ..., (from the third one on) is equal to the sum of the two preceding terms.

This information, together with the knowledge that

$$s_0 = 1 \qquad \text{and} \qquad s_1 = 1, \tag{16b}$$

completely determines the entire sequence. By this we mean that, given this information, and enough time, we can find any term in the sequence. For example, we can easily compute the first few terms in the sequence, which are

$$\{1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ...\}$$

However, finding a general formula for the nth term of this sequence is another matter! Equation (16a) is an example of what is known as a *recurrence relation* for the sequence s_0, s_1, s_2, s_3, ... It is a formula that describes each member of a sequence in terms of previous members. Equation (16b) give the *initial conditions* of the recurrence relation in Equation (16a). Recurrence relations such as Equation (16a) occur frequently in many different branches of mathematics and the sciences. For example, they occur in such diverse areas as physics, computer science, statistics, genetics, botany, economics, psychology, sociology, and many others. Recurrence relations are also known as *difference equations*.

In this section, we will consider some other examples of counting problems that lead to recurrence relations, and we will discuss a very simple method for solving certain recurrence relations. By solving a recurrence relation, we mean finding a general formula for the nth term s_n. In the next two sections, we will consider two methods for solving recurrence relations. The sequence s_0, s_1, s_2, s_3, ... , described by Equations (16a,b), is called the *Fibonacci sequence*, and the numbers in the sequence are called the *Fibonacci numbers*. We will see the Fibonacci numbers again when we count certain types of binary words. Also, these numbers occur in the most remarkable places in nature. For example, on some plants, thorns and leaves grow in a spiral pattern, and the number of thorns per revolution about the stalk is a ratio of two Fibonacci numbers. For instance, the apple or oak tree has 5 growths for every 2 turns around the stalk, the pear tree has 8 growths for every 3 turns, and the willow tree has 13 growths for every 5 turns. If you are interested in exploring these matters further, a good reference is Peter Stevens' book *Patterns in Nature*, published by Little, Brown and Co., Boston.

Let us now turn to some examples of counting problems that can lead to recurrence relations. We will not attempt to solve any of these recurrence relations until after we have discussed some methods for obtaining a solution.

Example #31: Subsets Again

One of the simplest recurrence relations arises from the problem of determining the number of subsets of a set of size n. Let s_n be this number. For the sake of argument, consider the set $S = \{1, 2, 3, ..., n\}$. Then the subsets of S can be divided into two groups, those that contain the element n and those that do not. Clearly, there are s_{n-1} subsets of S that do not contain the element n, since these are just the subsets of the set $\{1, 2, 3, ..., n-1\}$, which has size $n-1$. But, there are also s_{n-1} subsets of S that do contain n, since these subsets can be formed by taking each of the s_{n-1} subsets of $\{1, 2, ..., n-1\}$ and including the element n. Thus, the total number of subsets of S is equal to $s_{n-1} + s_{n-1} = 2s_{n-1}$, and we get the recurrence relation

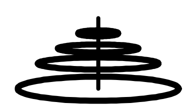
$$s_n = 2s_{n-1} \tag{17}$$

valid for all $n = 1, 2, 3, ...$. As for initial conditions, since a set of size 0 has exactly 1 subset (namely the empty set) and so we have $s_0 = 1$.

Notice that, in this case, we need only one initial condition, rather than two. The reason is that only the first term, s_0, in the sequence is needed in order to be able to compute the others from Equation (17). On the other hand, in the Fibonacci example, the first two terms, s_0 and s_1 are needed in order to use Equation (16a). ■

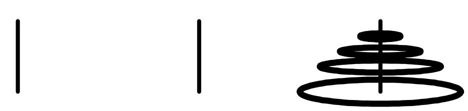
Example #32: The Towers of Hanoi

The Towers of Hanoi game, which we discussed in Example #5, is a counting problem that naturally leads to a recurrence relation. Recall that the object of this game is to transfer a tower of n disks from the first peg, as shown below for the case of $n = 4$



The game "begins" with all four disks stacked in decreasing size on the first peg

to the third peg, as pictured in the next figure,

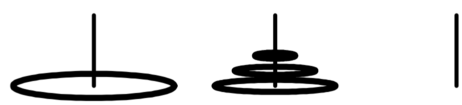


The game "end" with all four disks stacked in decreasing size on the third peg

under the conditions that only one disk can be moved at a time and the no disk can be placed on top of another disk of smaller outside diameter. The problem is to determine the minimum number of moves required to perform this operation. Suppose that we let s_n be the minimum number of moves required to transfer a tower of n disks from the first peg to the third peg under these conditions. Notice that in this case, we start the sequence with s_1, rather than s_0, since there is no such thing as a tower of zero disks. This is typical of recurrence relations.

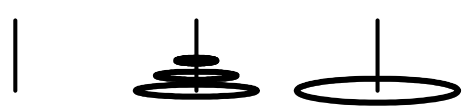
Sometimes it is more convenient to start with the term s_0, and other times it is more convenient to start with the term s_1. Let us see if we can find a recurrence relation for the sequence s_1, s_2, s_3, ..., s_n, which is an expression for s_n in terms of the previous terms s_1, s_2, s_3, ..., s_{n-1}.

For illustration, imagine moving a tower of 4 disks as shown in the figure below. Since the bottom disk can only be placed on an empty space, we first have to move the top three disks to the middle peg, as shown in the next figure and this takes s_3 moves.



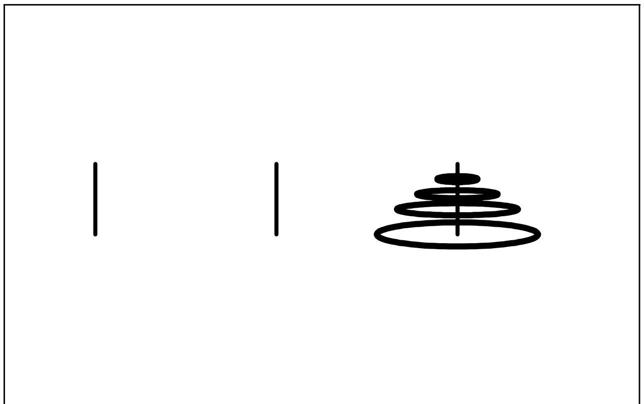
The position after s_3 moves

Then one addition move is made to place the largest disk on the empty peg, as illustrated in the next figure.



The position one move later

Then, an addition s_3 moves later places the other three disks on the largest one as illustrated next.



The position s_3 more moves later

It should be clear that $s_4 = s_3 + 1 + s_3 = 2s_3 + 1$.

Imagine now moving a tower of n disks from the first peg to the third peg in as few moves as possible. We saw in the $n = 4$ example above, at some point during this operation, the largest disk (after s_{n-1} previous moves) must be taken off the first peg and placed on the third peg that is empty (which is one move). Then an additional s_{n-1} moves is required to place the other smaller $n - 1$ disks that are on the middle peg on the largest disk that is on the third peg. Hence, the minimum number of moves required to complete the entire transfer is

$$s_n = s_{n-1} + 1 + s_{n-1} = 2s_{n-1} + 1,$$

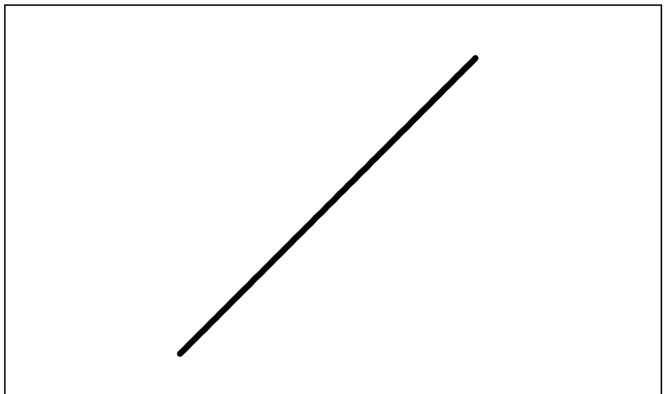
and this gives us the recurrence relation

$$s_n = 2s_{n-1} + 1 \tag{18}$$

which is valid for all $n = 2, 3, 4, ...$ As for initial conditions, since it takes only one move to transfer a tower consisting of one disk, we have $s_1 = 1$.

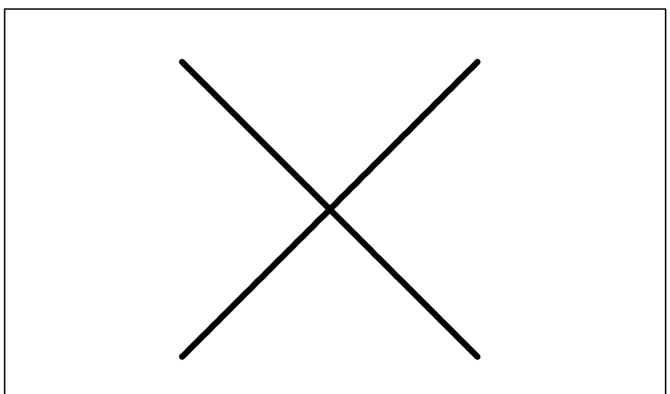
Example #33: A Geometry Problem - Lines on a Plane

A set of straight lines in the plane is said to be in *general position* if none of the lines are parallel, and if no three of the lines go through the same point. Into how many distinct regions does a set of n lines in general position divide the plane? To get a feel for this problem, let us consider a few simple cases. Of course, a single line divides the plane into 2 distinct regions as shown below.



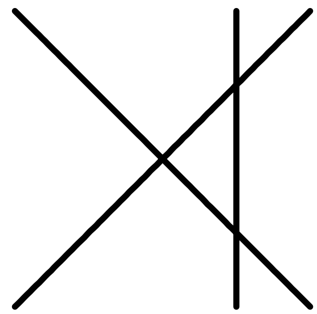
One Line Divides The Plane Into 2 Regions

Two lines divide the plane into 4 distinct regions as shown below.



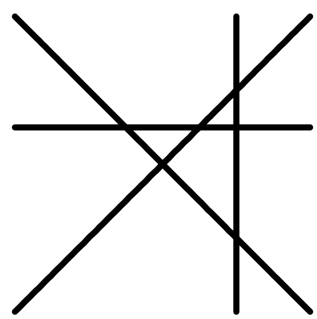
Two Lines Divide The Plane Into 4 Regions

Three lines divide the plane into 7 distinct regions as shown below.



Three Lines Divide The Plane Into 7 Regions

Four lines divide the plane into 11 distinct regions as shown below.



Four Lines Divide The Plane Into 11 Regions

Now, if we let p_n be the number of distinct regions into which n lines in general position divide the plane, then we have $p_1 = 2$, $p_2 = 4$, $p_3 = 7$ and $p_4 = 11$.

Let us try to find a recurrence relation for the sequence p_2, p_3, ..., p_n. Consider a set of $n-1$ lines in general position. These lines divide the plane into p_{n-1} regions. Now let us imagine adding another line to these in such a way that the new set of lines is still in general position. Thus, the new line must intersect all of the old lines, but not at any points where two of the old lines intersect. Our plan is to determine what happens to the p_{n-1} old regions, that is, the regions formed by the old lines, when the new line is added.

Perhaps the easiest way to see what happens is to imagine moving along the new line, starting at a point so far out on the line that it has not yet met any of the old lines. Every time the new line crosses an old line, it enters into a new region and divides that region into two parts. Since the new line meets $n-1$ old lines, no two at a time, it actually goes through n regions (counting the ones before the new line meets the first old line, and after it has met the last old line). Thus, the new line divides n of the old regions into two regions each, for a net gain of n regions, and so, after the new line is added, there are a total of $p_{n-1}+n$ regions. This gives us the recurrence relation

$$p_n = p_{n-1} + n \tag{19}$$

valid for all $n = 2, 3, 4, ...$ The initial condition is $p_1 = 2$. ■

Example #34: Binary Words

Recurrence relations occur naturally in trying to count words that have some restrictions placed on the pattern of letters. For example, suppose that we want to count the number of binary words that do not contain two (or more) consecutive 0's. In other words, we want to count binary words that do not contain the bit pattern 00. A bit is a binary digit, thus it is either a 0 or a 1. If s_n is the number of binary words of length n that do not contain the bit pattern 00, then we can obtain a recurrence relation for s_n as follows. Such binary words can be classified according to their first letter. Those that start with a 1 are of the form $1x......x$. where $x......x$ is any binary word of length $n-1$ that does not contain the bit pattern 00, and those that start with a 0 are of the form $01x......x$, where $x......x$ is a binary word of length $n-2$ that does not contain the bit pattern 00. Such a word cannot start with 00, for that is the forbidden bit pattern. Now, there are precisely s_{n-1} binary words of length $n-1$ that do not contain the bit pattern 00, and so there are s_{n-1} binary words of the type $1x......x$. Also, there are precisely s_{n-2} binary words of length $n-2$ that do not contain the bit pattern 00. Hence, there are s_{n-2} binary words of the type in $01x......x$. Therefore, there are $s_{n-1}+s_{n-2}$ binary words of length n that do not contain the bit pattern 00, and so we have

$$s_n = s_{n-1} + s_{n-2}, \tag{20}$$

which happens to be the Fibonacci recurrence relation. Also, it is easy to see that $s_0 = 1$ and $s_1 = 2$. Notice that the initial conditions in this case are a bit different than in the case of Fibonacci's rabbits. This simply means that the sequence starts with 1, 2, 3, 5, 8, ..., rather than 1, 1, 2, 3, 5, 8, ... ■

Example #35: More Binary Words

Let us find a recurrence relation for the number s_n of binary words of length n that do not contain the bit pattern 001. Such words are of one of the following three types. Those that begin with a 1 are of the form $1x......x$, where $x......x$ is a binary word of length $n-1$ that does not contain the bit pattern 001. Those that begin with a 0 are of two possible types. Either they begin with 01, in which case they are of the form $01x......x$, where $x......x$ is a binary word of length $n-2$ that does not contain the bit pattern 001, or they begin with 00. But, there is only one word of length n that begins with 00 and does not contain the bit pattern 001, namely, the word 000......0.

Now we can reason as in the previous example. Since there are s_{n-1} words of the type $1x......x$, s_{n-2} words of the type $01x......x$, and 1 word of the type 000......0, we get the recurrence relation

$$s_n = s_{n-1} + s_{n-2} + 1 \tag{21}$$

Also, we have the initial conditions $s_0 = 1$ and $s_1 = 2$. ■

Solving Recurrence Relations - The Method of Iterations

Now that we have considered some examples of recurrence relations, let us discuss how we might solve them. One of the simplest methods available to us is called *iteration*. The idea behind this method is simply to use the given recurrence relation over and over again, each time for a different value of n, in the hopes of seeing a pattern. A few examples will make the method clear.

Example #36: Number of Subsets

In Example #31, we found that

$$s_n = 2s_{n-1}$$

for all $n = 1, 2, 3, ...$ The method of iteration says that

$$s_1 = 2s_0 \quad , \quad s_2 = 2s_1 = 2(2s_0) = 2^2 s_0$$

and

$$s_3 = 2s_2 = 2(2^2 s_0) = 2^3 s_0.$$

It should be clear that

$$s_n = 2^n s_0$$

and with the initial condition $s_0 = 1$, we find that $s_n = 2^n$.

Of course, this value for s_n is only a guess, and so it is important to make sure that it really is a solution to the recurrence relation by substituting it into the

$$s_n = 2s_{n-1}$$

and see if it works. When doing this, we find that

$$2^n = 2 \times 2^{n-1}$$

which is true. We must not forget to check the initial conditions and we find that $s_0 = 2^0 = 1$ which is also true. ■

Each of the substitutions that we made in the last example is called an iteration, hence the name of the method. Let us try another example of iteration.

Example #37: The Towers of Hanoi

In Example #32, we found that

$$s_n = 2s_{n-1} + 1$$

for all $n = 2, 3, 4, ...$ along with the initial condition $s_1 = 1$. Using the method of iteration we have

$$s_2 = 2s_1 + 1 \quad , \quad s_3 = 2s_2 + 1 = 2(2s_1 + 1) + 1 = 2^2 s_1 + 2 + 1$$

and

$$s_4 = 2s_3 + 1 = 2(2^2 s_1 + 2 + 1) + 1 = 2^3 s_1 + 2^2 + 2 + 1.$$

It should be clear that

$$s_n = 2^{n-1} s_1 + 2^{n-2} + 2^{n-3} + \cdots + 2 + 1 = 2^{n-1} s_1 + \frac{2^{n-1} - 1}{2 - 1}$$

or

$$s_n = 2^{n-1}s_1 + 2^{n-1} - 1 = 2^{n-1}(1 + s_1) - 1.$$

Putting in the initial condition $s_1 = 1$, we find that

$$s_n = 2^n - 1.$$

Of course, the value for s_n is only a guess, and so it is important to make sure that it really is a solution to the recurrence relation by substituting it into the

$$s_n = 2s_{n-1} + 1$$

and see if it works. When doing this, we find that

$$2^n - 1 = 2(2^{n-1} - 1) + 1 = 2^n - 1$$

which is true. We also have $s_1 = 2^1 - 1 = 1$, which is also true.

Note that we may also solve $s_n = 2s_{n-1} + 1$ by replacing n by k and writing it as

$$\frac{1}{2}s_k - s_{k-1} = \frac{1}{2} \qquad \text{or} \qquad \frac{1}{2^k}s_k - \frac{1}{2^{k-1}}s_{k-1} = \frac{1}{2^k}$$

for $k = 2, 3, 4, \ldots$. Summing this from $k = 2$ to $k = n$ gives

$$\sum_{k=2}^{n}\left(\frac{1}{2^k}s_k - \frac{1}{2^{k-1}}s_{k-1}\right) = \sum_{k=2}^{n}\frac{1}{2^k} = \frac{1}{2} - \frac{1}{2^n}$$

or

$$\frac{1}{2^n}s_n - \frac{1}{2^1}s_1 = \frac{1}{2} - \frac{1}{2^n}$$

which says that

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

and this agrees with our earlier calculation. Note that the sum

$$\sum_{k=2}^{n}\left(\frac{1}{2^k}s_k - \frac{1}{2^{k-1}}s_{k-1}\right)$$

is called a telescopic sum and can be simplified using

$$\sum_{k=a}^{b}(T_k - T_{k-1}) = (T_a - T_{a-1}) + (T_{a+1} - T_a) + \cdots + (T_{b-1} - T_{b-2}) + (T_b - T_{b-1})$$

which reduces to

$$\sum_{k=a}^{b}(T_k - T_{k-1}) = T_b - T_{a-1}.$$

Thus, we see that the minimum number of moves required to complete the Towers of Hanoi puzzle, with a tower consisting of n disks, is $2^n - 1$. In particular, with 8 disks, the minimum number of moves is $2^8 - 1 = 255$. ■

A General Result for $s_n = as_{n-1} + f(n)$

The method used in the previous problem suggest a general formula for the recurrence relation

$$s_n = as_{n-1} + f(n)$$

in which $a \neq 0$ is a constant and given s_1. We need only note that

$$\frac{1}{a}s_n = s_{n-1} + \frac{1}{a}f(n)$$

or

$$\frac{1}{a^n}s_n - \frac{1}{a^{n-1}}s_{n-1} = \frac{1}{a^n}f(n)$$

for $n = 2, 3, \ldots.$ Setting $n = 2$, we have

$$\frac{1}{a^2}s_2 - \frac{1}{a}s_1 = \frac{1}{a^2}f(2)$$

and setting $n = 3$, we have

$$\frac{1}{a^3}s_3 - \frac{1}{a^2}s_2 = \frac{1}{a^3}f(3)$$

and setting $n = 4$, we have

$$\frac{1}{a^4}s_4 - \frac{1}{a^3}s_3 = \frac{1}{a^4}f(4)$$

and so on up to

$$\frac{1}{a^n}s_n - \frac{1}{a^{n-1}}s_{n-1} = \frac{1}{a^n}f(n).$$

Adding all of these leads to

$$\frac{1}{a^n}s_n - \frac{1}{a}s_1 = \sum_{j=2}^{n} \frac{1}{a^j} f(j)$$

which says that

$$s_n = a^n \left(\frac{1}{a}s_1 + \sum_{j=2}^{n} \frac{1}{a^j} f(j) \right)$$

or

$$s_n = a^{n-1}s_1 + \sum_{j=2}^{n} a^{n-j} f(j),$$

which gives a general formula for solving the recurrence relation $s_n = as_{n-1} + f(n)$, for when a is a non-zero constant and f is a known function of n.

Example #38: The General Line Intersection Problem

In Example #33, we found that the number of regions that n general intersecting lines divides the plane into is p_n, with

$$p_n = p_{n-1} + n$$

and $p_1 = 2$. Using the idea of a telescopic sum, we note that

$$p_n - p_{n-1} = n$$

which says that

$$\begin{aligned} p_2 - p_1 &= 2 \\ p_3 - p_2 &= 3 \\ p_4 - p_3 &= 4 \\ &\vdots \\ p_n - p_{n-1} &= n \end{aligned}$$

and adding these we find that

$$(p_2 - p_1) + (p_3 - p_2) + (p_4 - p_3) + \cdots + (p_n - p_{n-1}) = 2 + 3 + 4 + \cdots + n$$

which leads to

$$p_n - p_1 = \sum_{j=2}^{n} j \qquad \text{or} \qquad p_n = p_1 + \sum_{j=2}^{n} j = 2 + \sum_{j=2}^{n} j = 1 + \sum_{j=1}^{n} j$$

or simply

$$p_n = 1 + \frac{1}{2}n(n+1)$$

for $n = 1, 2, 3, ...$

Of course, this value for p_n is only a guess, and so it is important to make sure that it really is a solution to the recurrence relation by substituting it into the

$$p_n = p_{n-1} + n$$

and see if it works, along with the initial conditions. When doing this, we find that

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

which is true, and

$$p_1 = 1 + \frac{1}{2}(1)(1+1) = 2$$

which is also true. ■

11. Second-Order Homogeneous Recurrence Relations

Unfortunately, there is no general method that can be used to solve all types of recurrence relations. However, there is a very important class of recurrence relations that we can solve. This class includes the Fibonacci recurrence relation and are of the following form. A second order linear homogeneous recurrence relation with constant coefficients is a recurrence relation of the form

$$s_n = as_{n-1} + bs_{n-2} \tag{22}$$

for $n = 2, 3, 4, ...$, where a and b are constants. Let us make a comment about the terms used in this definition. Second order refers to the fact that s_n depends on s_{n-1} and s_{n-2}. First order (such as $s_n = 2s_{n-1} + 1$) is when s_n depends only on s_{n-1}. The recurrence relation is linear because it does not involve any powers or products of members of the sequence s_n . For example, the recurrence relation $s_n = s_{n-1}^2 + s_{n-2}$ is not linear. Equation (22) is homogeneous because it is satisfied by the sequence $s_n = 0$ for all n. For example, the recurrence relation $s_n = s_{n-1} + s_{n-2} + 1$ is not homogeneous.

The first thing that we should notice is that a recurrence relation of the form of Equation (22), for $n = 2, 3, 4, ...$, together with initial conditions of the form $s_0 = u$ and $s_1 = v$ (where u and v are constants) completely determines a sequence s_n. Therefore, once we find a solution to Equation (22) that satisfies the given initial conditions, we will know that our solution is the only one possible. As you can see, the Fibonacci sequence

$$s_n = s_{n-1} + s_{n-2}$$

for $n = 2, 3, 4, ...$, with $s_0 = 1$ and $s_1 = 1$, is a second order linear recurrence relation with constant coefficients, where $a = b = 1$.

Solving Equation (22)

The method for solving Equation (22) is to "guess" a solution of the form $s_n = r^n$, where r is a constant non-zero parameter to be determine. To determine r, we require that $s_n = r^n$ satisfies Equation (22) and this leads to

$$r^n = ar^{n-1} + br^{n-2}$$

or

$$r^2 - ar - b = 0$$

which has the solutions

$$r_{\pm} = \frac{a \pm \sqrt{a^2 + 4b}}{2}. \tag{23}$$

There are two cases to consider.

The Case When $a^2 + 4b \neq 0$

The first case is when $a^2 + 4b \neq 0$ and in this case $r_+ \neq r_-$ and both

$$r_+^n \qquad \text{and} \qquad r_-^n$$

are solutions to Equation (22) and hence so is

$$s_n = A_+ r_+^n + A_- r_-^n$$

for arbitrary constants A_+ and A_-. Using the fact that

$$s_0 = A_+ + A_- = u \quad \text{and} \quad s_1 = A_+ r_+ + A_- r_- = v$$

then gives A_+ and A_- as

$$\begin{bmatrix} A_+ \\ A_- \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ r_+ & r_- \end{bmatrix}^{-1} \begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{r_- - r_+} \begin{bmatrix} r_- & -1 \\ -r_+ & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

or

$$\begin{bmatrix} A_+ \\ A_- \end{bmatrix} = \frac{1}{r_- - r_+} \begin{bmatrix} ur_- - v \\ v - ur_+ \end{bmatrix}.$$

Since

$$r_- - r_+ = \frac{a - \sqrt{a^2+4b}}{2} - \frac{a + \sqrt{a^2+4b}}{2} = -\sqrt{a^2+4b}$$

we find that

$$\begin{bmatrix} A_+ \\ A_- \end{bmatrix} = \frac{1}{\sqrt{a^2+4b}} \begin{bmatrix} v - ur_- \\ ur_+ - v \end{bmatrix}.$$

To summarize the case when $a^2 + 4b \neq 0$, we have

$$s_n = \frac{(v - ur_-)r_+^n + (ur_+ - v)r_-^n}{\sqrt{a^2+4b}} \tag{24a}$$

where $s_0 = u$, $s_1 = v$ and

$$r_\pm = \frac{a \pm \sqrt{a^2+4b}}{2}. \tag{24b}$$

Example #39: The Fibonacci Sequence

The Fibonacci sequence has $a = b = 1$ and $u = v = 1$, and so

$$r_\pm = \frac{1 \pm \sqrt{5}}{2}$$

and then

$$s_n = \frac{\left(1 - \frac{1-\sqrt{5}}{2}\right)\left(\frac{1+\sqrt{5}}{2}\right)^n + \left(\frac{1+\sqrt{5}}{2} - 1\right)\left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}},$$

which reduces to

$$s_n = \frac{\left(\frac{1+\sqrt{5}}{2}\right)\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)\left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}},$$

or simply

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

for $n = 0, 1, 2, 3, \ldots$. As a check, we note that

$$s_0 = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{0+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{0+1} = 1$$

and

$$s_1 = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{1+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{1+1} = 1$$

and

$$s_2 = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{2+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{2+1} = 2$$

and

$$s_3 = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{3+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{3+1} = 3$$

and

$$s_4 = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{4+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{4+1} = 5$$

and so on. It is remarkable that all the square roots of 5 cancel to give an integer for s_n in the end. ■

The Case When $a^2 + 4b = 0$

Going back to Equation (23), we see that when $a^2 + 4b = 0$, we have

$$r_{\pm} = \frac{1}{2}a = r_0$$

showing that $s_n = r_0^n$ is a solution to

$$s_n = as_{n-1} + bs_{n-2}.$$

Another solution is nr_0^n, since this yields

$$nr_0^n = a(n-1)r_0^{n-1} + b(n-2)r_0^{n-2}$$

or

$$nr_0^2 - a(n-1)r_0 - b(n-2) = n\left(\frac{a}{2}\right)^2 - a(n-1)\left(\frac{a}{2}\right) - b(n-2)$$

which reduces to

$$nr_0^2 - a(n-1)r_0 - b(n-2) = -\frac{(a^2+4b)(n-2)}{4} = 0.$$

Thus we have

$$s_n = A_0 r_0^n + B_0 n r_0^n$$

as a general solution to $s_n = as_{n-1} + bs_{n-2}$ under the condition that $a^2 + 4b = 0$. To determine A_0 and B_0, we have

$$s_0 = A_0 = u \qquad \text{and} \qquad s_1 = A_0 r_0 + B_0 r_0 = v$$

which says that $A_0 = u$ and $B_0 = (v - ur_0)/r_0$. To summarize, we see that

$$s_n = ur_0^n + \left(\frac{v - ur_0}{r_0}\right) nr_0^n$$

or

$$s_n = ur_0^n + (v - ur_0)nr_0^{n-1} \tag{24c}$$

is the solution to $s_n = as_{n-1} + bs_{n-2}$ under the condition that $a^2 + 4b = 0$ along with $s_0 = u$ and $s_1 = v$.

Example #40: A Case When $a^2 + 4b \neq 0$

Let us solve the recurrence relation

$$s_n = 2s_{n-1} + 3s_{n-2}$$

for $n = 2, 3, 4, \ldots$, with initial conditions $s_0 = 0$ and $s_1 = 8$. Here we have $u = 0$, $v = 8$, $a = 2$ and $b = 3$ so that $a^2 + 4b = 2^2 + 4(3) = 16 \neq 0$ and

$$r_\pm = \frac{2 \pm \sqrt{2^2 + 4(3)}}{2} = 1 \pm 2$$

so using Equation (24a), we find that

$$s_n = \frac{(8-(0)(-1))(3)^n + ((0)(3)-8)(-1)^n}{\sqrt{2^2+4(3)}}$$

which reduces to

$$s_n = \frac{8(3)^n - 8(-1)^n}{4} = 2(3)^n - 2(-1)^n.$$

You should verify this by showing that $s_n = 2s_{n-1} + 3s_{n-2}$ along with $s_0 = 0$ and $s_1 = 8$. ■

Example #41: A Case When $a^2 + 4b = 0$

Let us solve the recurrence relation

$$s_n = 4s_{n-1} - 4s_{n-2}$$

for $n = 2, 3, 4, \ldots$, with initial conditions $s_0 = 1$ and $s_1 = 1$. Here we have $u = 1$, $v = 1$, $a = 4$ and $b = -4$ so that $a^2 + 4b = 4^2 + 4(-4) = 0$ and

$$r_0 = \frac{4 \pm \sqrt{4^2 + 4(-4)}}{2} = 2$$

so using Equation (24c), we find that

$$s_n = (1)(2)^n + (1-(1)(2))n(2)^{n-1}$$

which reduces to

$$s_n = 2^n - n2^{n-1} = (2-n)2^{n-1}.$$

You should verify this by showing that $s_n = 4s_{n-1} - 4s_{n-2}$ along with $s_0 = 1$ and $s_1 = 1$. ■

Before concluding this section, we should make one more remark about recurrence relations. There are many other more sophisticated methods for solving recurrence relations, and we will briefly discuss one of them in the next section. However, as we mentioned at the beginning of this section, there is no general method that will always work to solve any recurrence relation, and there are many important recurrence relations that have never been solved.

Actually, this situation is not as bad as it may seem. For from a strictly computational point of view, the recurrence relation itself may turn out to be much more valuable than its solution! As an example, if we wanted to compute the first 1000 Fibonacci numbers, with the aid of a computer, it might be preferable to have the computer calculate these numbers directly from the recurrence relation

$$s_n = s_{n-1} + s_{n-2} \tag{25a}$$

rather than from direct result

$$s_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}. \tag{25b}$$

12. Second-Order Non-Homogeneous Recurrence Relations

In the previous section, we discussed a special type of homogeneous recurrence relation. In this section, we want to remove the restriction that the recurrence relation is homogeneous and consider the second-order linear recurrence relation of the form

$$s_n = as_{n-1} + bs_{n-2} + f(n) \tag{26}$$

for $n = 2, 3, 4, ...$, where a and b are constants and $f(n)$ is an expression that depends only on n, and not on the sequence s_n.

When $f(n) = 0$ for all n, Equation (26) is a homogeneous recurrence relation of the type that we discussed in the previous section. However, when $f(n)$ is not always equal to 0, then Equation (26) is called a non-homogeneous recurrence relation. For example, the recurrence relations

$$s_n = s_{n-1} + 2n \qquad \text{and} \qquad s_n = s_{n-2} + 1$$

and

$$s_n = 2s_{n-1} + 3s_{n-2} + n - 1$$

are non-homogeneous..

If the recurrence relation

$$s_n = as_{n-1} + bs_{n-2} + f(n)$$

for $n = 2, 3, 4, \ldots$, is non-homogeneous, that is, if $f(n) \neq 0$ for some n, then the recurrence relation obtained by dropping the tern $f(n)$, namely

$$s_n = as_{n-1} + bs_{n-2}$$

is called the associated homogeneous recurrence relation.

In general, there is no method for solving non-homogeneous recurrence relations of the form in Equation (26). However, we can take a big step forward by relating the solutions of Equation (26) to the solutions of the associated homogeneous recurrence relation, which we can solve. To see how this can be done, suppose that p_n is a "particular" solution to the non-homogeneous recurrence relation

$$s_n = as_{n-1} + bs_{n-2} + f(n)$$

so that

$$p_n = ap_{n-1} + bp_{n-2} + f(n)$$

Then it is easy to see that

$$s_n - p_n = a(s_{n-1} - p_{n-1}) + b(s_{n-2} - p_{n-2})$$

which says that $h_n = s_n - p_n$ is the solution to the associated homogeneous recurrence relation because

$$h_n = ah_{n-1} + bh_{n-2}.$$

Thus we need only find the *general* solution to the associated homogeneous recurrence relation and *any particular* solution to the non-homogeneous recurrence relation.

Example #42: More Binary Words - Example #35 Revisited

Recall that in Example #35, we wanted to find the number s_n of binary words of length n that do not contain the bit pattern 001 and we found that

$$s_n = s_{n-1} + s_{n-2} + 1$$

along with the initial conditions $s_0 = 1$ and $s_1 = 2$. We already know that

$$s_n = A_+ \left(\frac{1+\sqrt{5}}{2}\right)^n + A_- \left(\frac{1-\sqrt{5}}{2}\right)^n$$

is a solution to the associated homogeneous equation

$$s_n = s_{n-1} + s_{n-2}.$$

To determine a particular solution to the non-homogeneous equation

$$s_n = s_{n-1} + s_{n-2} + 1$$

we note that $f(n) = 1$ is a constant and so let us "guess" that $p_n = A$, also a constant. To determine A, we place $p_n = A$ into

$$p_n = p_{n-1} + p_{n-2} + 1$$

resulting in

$$A = A + A + 1 \qquad \text{or} \qquad A = -1.$$

Thus we find that

$$s_n = A_+ \left(\frac{1+\sqrt{5}}{2}\right)^n + A_- \left(\frac{1-\sqrt{5}}{2}\right)^n - 1$$

is a general solution to $s_n = s_{n-1} + s_{n-2} + 1$, and to determine A_+ and A_-, we use

$$s_0 = A_+ \left(\frac{1+\sqrt{5}}{2}\right)^0 + A_- \left(\frac{1-\sqrt{5}}{2}\right)^0 - 1 = 1$$

and

$$s_1 = A_+ \left(\frac{1+\sqrt{5}}{2}\right)^1 + A_- \left(\frac{1-\sqrt{5}}{2}\right)^1 - 1 = 2$$

which both reduce to

$$A_+ + A_- = 2 \qquad \text{and} \qquad A_+\left(1+\sqrt{5}\right) + A_-\left(1-\sqrt{5}\right) = 6$$

which leads to

$$A_+ = 1 + \frac{2}{\sqrt{5}} \qquad \text{and} \qquad A_- = 1 - \frac{2}{\sqrt{5}}$$

and hence

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

As a check we note that

$$s_0 = \left(1 + \frac{2}{\sqrt{5}}\right)\left(\frac{1+\sqrt{5}}{2}\right)^0 + \left(1 - \frac{2}{\sqrt{5}}\right)\left(\frac{1-\sqrt{5}}{2}\right)^0 - 1 = 1$$

and

$$s_1 = \left(1 + \frac{2}{\sqrt{5}}\right)\left(\frac{1+\sqrt{5}}{2}\right)^1 + \left(1 - \frac{2}{\sqrt{5}}\right)\left(\frac{1-\sqrt{5}}{2}\right)^1 - 1 = 2$$

and

$$s_2 = \left(1 + \frac{2}{\sqrt{5}}\right)\left(\frac{1+\sqrt{5}}{2}\right)^2 + \left(1 - \frac{2}{\sqrt{5}}\right)\left(\frac{1-\sqrt{5}}{2}\right)^2 - 1 = 4$$

which is also obtained from

$$s_2 = s_1 + s_0 + 1 = 2 + 1 + 1 = 4.$$

You should verify that

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

satisfies $s_n = s_{n-1} + s_{n-2} + 1$ for all $n = 2, 3, 4, ...$, along with $s_0 = 1$ and $s_1 = 2$.
■

Example #43: $f(n)$ Is A Polynomial in n

Let us solve the recurrence relation

$$s_n = 3s_{n-1} + 4s_{n-2} + n + 1$$

for $n = 2, 3, 4, ...$, along with the initial conditions $s_0 = 1$ and $s_1 = 0$. First we note that the associated homogeneous equation is

$$s_n = 3s_{n-1} + 4s_{n-2}$$

in which $a = 3$ and $b = 4$ so that $a^2 + 4b = (3)^2 + 4(4) = 25 \neq 0$ and hence

$$r_{\pm} = \frac{a \pm \sqrt{a^2 + 4b}}{2} = \frac{3 \pm \sqrt{3^2 + 4(4)}}{2} = \frac{3 \pm 5}{2}$$

or $r_- = -1$ and $r_+ = 4$. This leads to

$$s_n = A_+(4)^n + A_-(-1)^n$$

as the general solution to $s_n = 3s_{n-1} + 4s_{n-2}$. To determine a particular solution to

$$s_n = 3s_{n-1} + 4s_{n-2} + n + 1$$

we use the fact that $f(n) = n + 1$ is a first degree polynomial in n and so we "guess" a solution of the form $p_n = An + B$, where A and B are constants to be determined. In other words, we "guess" a solution for p_n which has the *same form* as that of $f(n)$. Putting $p_n = An + B$ into

$$p_n = 3p_{n-1} + 4p_{n-2} + n + 1$$

we find that

$$An + B = 3(A(n-1) + B) + 4(A(n-2) + B) + n + 1$$

or

$$(6A + 1)n + (6B - 11A + 1) = 0$$

which is true for all $n = 2, 3, 4, ...$, provided that

$$6A + 1 = 0 \qquad \text{and} \qquad 6B - 11A + 1 = 0$$

which says that $A = -1/6$ and $B = -17/36$. Thus we find that

$$p_n = -\frac{1}{6}n - \frac{17}{36}$$

is a particular solution to

$$p_n = 3p_{n-1} + 4p_{n-2} + n + 1$$

and hence

$$s_n = A_+(4)^n + A_-(-1)^n - \frac{1}{6}n - \frac{17}{36}$$

is a general solution to

$$s_n = 3s_{n-1} + 4s_{n-2} + n + 1.$$

Finally, to determine A_+ and A_-, we have

$$s_0 = A_+(4)^0 + A_-(-1)^0 - \frac{1}{6}(0) - \frac{17}{36} = A_+ + A_- - \frac{17}{36} = 1$$

and

$$s_1 = A_+(4)^1 + A_-(-1)^1 - \frac{1}{6}(1) - \frac{17}{36} = 4A_+ - A_- - \frac{23}{36} = 0$$

resulting in

$$A_+ + A_- = \frac{53}{36} \qquad \text{and} \qquad 4A_+ - A_- = \frac{23}{36}$$

or $A_+ = 19/45$ and $A_- = 21/20$. Thus we find that

$$s_n = \frac{19}{45}(4)^n + \frac{21}{20}(-1)^n - \frac{1}{6}n - \frac{17}{36}$$

is the solution to the problem. As a check we note that

$$s_0 = \frac{19}{45}(4)^0 + \frac{21}{20}(-1)^0 - \frac{1}{6}(0) - \frac{17}{36} = 1$$

and

$$s_1 = \frac{19}{45}(4)^1 + \frac{21}{20}(-1)^1 - \frac{1}{6}(1) - \frac{17}{36} = 0$$

and

$$s_2 = \frac{19}{45}(4)^2 + \frac{21}{20}(-1)^2 - \frac{1}{6}(2) - \frac{17}{36} = 7$$

which is also obtained using

$$s_2 = 3s_1 + 4s_0 + 2 + 1 = 3(0) + 4(1) + 3 = 7.$$

You should verify that

$$s_n = \frac{19}{45}(4)^n + \frac{21}{20}(-1)^n - \frac{1}{6}n - \frac{17}{36}$$

satisfies $s_n = 3s_{n-1} + 4s_{n-2} + n + 1$ for all $n = 2, 3, 4, \ldots$, along with $s_0 = 1$ and $s_1 = 0$. ■

Example #44: $f(n)$ Is of the Form $f(n) = Ac^n$

Let us solve the recurrence relation

$$s_n = 3s_{n-1} + 4s_{n-2} + 2(3)^n$$

for $n = 2, 3, 4, ...$, along with the initial conditions $s_0 = 1$ and $s_1 = 0$. First we note that the associated homogeneous equation is

$$s_n = 3s_{n-1} + 4s_{n-2}$$

in which $a = 3$ and $b = 4$ so that $a^2 + 4b = (3)^2 + 4(4) = 25 \neq 0$ and hence

$$r_{\pm} = \frac{a \pm \sqrt{a^2 + 4b}}{2} = \frac{3 \pm \sqrt{3^2 + 4(4)}}{2} = \frac{3 \pm 5}{2}$$

or $r_- = -1$ and $r_+ = 4$. This leads to

$$s_n = A_+(4)^n + A_-(-1)^n$$

as the general solution to $s_n = 3s_{n-1} + 4s_{n-2}$. To determine a particular solution to

$$s_n = 3s_{n-1} + 4s_{n-2} + 2(3)^n$$

we use the fact that $f(n) = 2(3)^n$ and we "guess" a solution of the form $p_n = A(3)^n$, where A is a constant to be determined. In other words, we "guess" a solution for p_n which has the *same form* as that of $f(n)$. Putting $p_n = A(3)^n$ into

$$p_n = 3p_{n-1} + 4p_{n-2} + 2(3)^n$$

we find that

$$A(3)^n = 3(A(3)^{n-1}) + 4(A(3)^{n-2}) + 2(3)^n$$

or

$$9A = 13A + 18$$

which says that $A = -9/2$. Thus we find that

$$p_n = -\frac{9}{2}(3)^n = -\left(\frac{1}{2}\right)3^{n+2}$$

is a particular solution to

$$p_n = 3p_{n-1} + 4p_{n-2} + 2(3)^n$$

and hence

$$s_n = A_+(4)^n + A_-(-1)^n - \left(\frac{1}{2}\right) 3^{n+2}$$

is a general solution to

$$s_n = 3s_{n-1} + 4s_{n-2} + 2(3)^n.$$

Finally, to determine A_+ and A_-, we have

$$s_0 = A_+(4)^0 + A_-(-1)^0 - \left(\frac{1}{2}\right) 3^{0+2} = A_+ + A_- - \frac{9}{2} = 1$$

and

$$s_1 = A_+(4)^1 + A_-(-1)^1 - \left(\frac{1}{2}\right) 3^{1+2} = 4A_+ - A_- - \frac{27}{2} = 0$$

resulting in

$$A_+ + A_- = \frac{11}{2} \qquad \text{and} \qquad 4A_+ - A_- = \frac{27}{2}$$

$A_+ = 19/5$ and $A_- = 17/10$. Thus we find that

$$s_n = \frac{19}{5}(4)^n + \frac{17}{10}(-1)^n - \left(\frac{1}{2}\right) 3^{n+2}$$

is the solution to the problem. As a check we note that

$$s_0 = \frac{19}{5}(4)^0 + \frac{17}{10}(-1)^0 - \left(\frac{1}{2}\right) 3^{0+2} = 1$$

and

$$s_1 = \frac{19}{5}(4)^1 + \frac{17}{10}(-1)^1 - \left(\frac{1}{2}\right) 3^{1+2} = 0$$

and

$$s_2 = \frac{19}{5}(4)^2 + \frac{17}{10}(-1)^2 - \left(\frac{1}{2}\right) 3^{2+2} = 22$$

which is also obtained using

$$s_2 = 3s_1 + 4s_0 + 2(3)^n = 3(0) + 4(1) + 2(3)^2 = 22.$$

You should verify that

$$s_n = \frac{19}{5}(4)^n + \frac{17}{10}(-1)^n - \left(\frac{1}{2}\right) 3^{n+2}$$

does satisfy $s_n = 3s_{n-1} + 4s_{n-2} + 2(3)^n$ for all $n = 2, 3, 4, ...$, along with $s_0 = 1$ and $s_1 = 0$. ■

More General Forms For $f(n)$

It should be clear that if $f(n)$ is of the form

$$f(n) = \sum_{i=0}^{m} a_i n^i + \sum_{j=1}^{k} b_k (c_k)^j \tag{27a}$$

then guessing a particular solution of the form

$$p_n = \sum_{i=0}^{m} A_i n^i + \sum_{j=1}^{k} B_k (c_k)^j \tag{27b}$$

should help with solving the recurrence relation

$$s_n = a s_{n-1} + b s_{n-2} + f(n). \tag{27c}$$

The student interested in these ideas should consider taking (or reading a textbook) a course on *Difference Equations*. Two good books that I would highly recommend are: *Difference Equations, An Introduction with Applications* by Walter G. Kelley and Allan C. Peterson and *Difference Equations, Theory and Applications*, by Ronald E. Mickens.