Math408: Combinatorics

University of North Dakota Mathematics Department

Spring 2011

Table of Contents

Chapter 0: Overview

There are three essential problems in combinatorics. These are the *existence problem*, the *counting problem*, and the *optimization problem.* This course deals primarily with the first two in reverse order.

The first two chapters are preparatory in nature. Chapter 1 deals with basic counting. Since Math208 is a prerequisite for this course, you should already have a pretty good grasp of this topic. This chapter will normally be covered at an accelerated rate. Chapter 2 is a short introduction to graph theory - which serves as a nice tie-in between the counting problem, and the existence problem. Graph theory is also essential for the optimization problem. Not every instructor of Math208 covers graph theory beyond the basics of representing relations on a set via digraphs. This short chapter should level the playing field between those students who have seen more graph theory and those who have not.

Chapter 3 is devoted to intermediate counting techniques. Again, some of this material will be review for certain, but not all, students who have successfully completed Math208. The material on generating functions requires some ability to manipulate power series in a formal fashion. This explains why Calculus II is a prerequisite for this course.

Counting theory is crowned by the so-called Pólya Counting, which is the topic of Chapter 4. Pólya Counting requires some basic group theory. This is not the last topic where abstract algebra rears its head.

The terminal chapters are devoted to combinatorial designs and a short introduction to coding theory. The existence problem is the main question addressed here. The flavor of these notes is to approach the problems from an algebraic perspective. Thus we will spend considerable effort investigating finite fields and finite geometries over finite fields.

My personal experience was that seeing these algebraic structures in action before taking abstract algebra was a huge advantage. I've also encountered quite a few students who took this course after completing abstract algebra. Prior experience with abstract algebra did not necessarily give them an advantage in this course, but they did tend to come away with a much improved opinion of, and improved respect for, the field of abstract algebra.

Chapter 1: Basic Counting

We generally denote sets by capital English letters, and their elements as lowercase English letters. We denote the cardinality of a finite set, A, by $|A|$. A set with $|A| = n$ is called an n-set. We denote an arbitrary universal set by $\mathcal{U}$, and the complement of a set (relative to $\mathcal{U}$) by $\overline{A}$. Unless otherwise indicated, all sets mentioned in this chapter are finite sets.

§1.1 Counting Principles

The basic principles of counting theory are the *multiplication principle*, the *principle of inclusion/exclusion*, the *addition principle*, and the *exclusion principle*.

The multiplication principle states that the cardinality of a Cartesian product is the product of the cardinalities. In the most basic case we have $|A \times B| = |A| \cdot |B|$. An argument for this is that $A \times B$ consists of all ordered pairs (a, b) where $a \in A$ and $b \in B$. There are $|A|$ choices for a and then $|B|$ choices for b. A common rephrasing of the principle is that if a task can be decomposed into two sequential subtasks, where there are n_1 ways to complete the first subtask, and then n_2 ways to complete the second subtask, then altogether there are $n_1 \cdot n_2$ ways to complete the task.

Notice the connection between the multiplication principle and the logical connective AND. Also, realize that this naturally extends to general Cartesian products with finitely many terms.

Example: The number of binary strings of length 10 is 2^{10} since it is $|\{0, 1\}|^{10}$.

Example: The number of ternary strings of length n is 3^n.

Example: The number of functions from a k-set to an n-set is n^k.

Example: The number of strings of length k using n symbols with repetition allowed is n^k.

Example: The number of 1-1 functions from a k-set to an n-set is $n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-(k-1))$.

The basic principle of inclusion/exclusion states that $|A \cup B| = |A| + |B| - |A \cap B|$. So we include elements when either in A, or in B, but then have to exclude the elements in $A \cap B$, since they've been included twice each.

Example: How many students are there in a discrete math class if 15 students are computer science majors, 7 are math majors, and 3 are double majors in math and computer science?

Solution: Let A denote the subset of computer science majors in the class, and B denote the math majors. Then $|A| = 15$, $|B| = 7$ and $|A \cap B| = 3 \neq 0$. So by the principle of inclusion/exclusion there are $15 + 7 - 3 = 19$ students in the class.

The general principle of inclusion/exclusion will be discussed in a later section.

The addition principle is a special case of the principle of inclusion/exclusion. If $A \cap B = \emptyset$, then $|A \cup B| = |A| + |B|$. In general the cardinality of a finite collection of pairwise disjoint finite sets is the sum of their cardinalities. That is, if $A_i \cap A_j = \emptyset$ for $i \neq j$, and $|A_i| < \infty$ for all i, then $\left| \bigcup_{i=1}^{n} A_i \right| = \sum_{i=1}^{n} |A_i|$.

The exclusion principle is a special case of the addition principle. A set and its complement are always disjoint, so $|A| + |\overline{A}| = |\mathcal{U}|$, or equivalently $|A| = |\mathcal{U}| - |\overline{A}|$.

§1.2 Permutations and Combinations

Given an n-set of objects, an r-string from the n-set is a sequence of length r. We take the convention that the string is identified with its output list. So the string $a_1 = a, a_2 = b, a_3 = c, a_4 = b$ is denoted $abcb$.

The number of r-strings from a set of size n is n^r as we saw in the previous section. As a string we see that order matters. That is, the string $abcd$ is not the same as the string $bcad$. Also repetition is allowed, since for example aaa is a 3-string from the set of lowercase English letters.

An r-permutation from an n-set is an ordered selection of r distinct objects from the n-set. We denote the number of r-permutations of an n-set $P(n,r)$. By the multiplication principle $P(n,r) = n(n-1)\cdot \ldots \cdot (n-(r-1)) = n(n-1)\cdot \ldots \cdot (n-r+1) = \dfrac{n!}{(n-r)!}$.

The number $P(n,r)$ is the same as the number of one-to-one functions from a set of size r to a set of size n.

An r-combination from an n-set is an unordered collection of r distinct elements from the set. In other words an r-combination of an n-set is a r-subset. We denote the number of r-combinations from an n-set by $C(n,r)$ or $\dbinom{n}{r}$.

Theorem 1 $r!\dbinom{n}{r} = P(n,r)$

Proof: For each r-combination from an n-set, there are $r!$ ways for us to order the set without repetition. Each ordering gives rise to exactly one r-permutation from the n-set. Every r-permutation from the n-set arises in this fashion. ■

Corollary 1 $\dbinom{n}{r} = \dfrac{n!}{r!(n-r)!}$

Since $n-(n-r) = r$, we also have

Corollary 2 $\dbinom{n}{r} = \dbinom{n}{n-r}$

Example: Suppose we have a club with 20 members. If we want to select a committee of 5 members, then there are $C(20,5)$ ways to do this since the order of people on the committee doesn't matter. However if the club wants to elect a board of officers consisting of a president, vice president, secretary, treasurer, and sergeant-at-arms, then there are $P(20,5)$ ways to do this. In each instance, repetition is not allowed. What makes the difference between the two cases is that the first is an unordered selection without repetition, whereas the second is an ordered selection without repetition.

§1.3 Combinatorial Arguments and the Binomial Theorem

One of the most famous combinatorial arguments is attributed to Blaise Pascal and bears his name. The understanding we adopt is that any number of the form $\dbinom{m}{s}$, where m and s are integers, is zero, if either $s > m$, or $s < 0$ (or both).

Theorem (Pascal's Identity) *Let n and k be non-negative integers, then*

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

Proof: Let S be a set with $n+1$ elements, and let $a \in S$. Put $T = S - \{a\}$ so $|T| = n$. On the one hand S has $\binom{n+1}{k}$ subsets of size k. On the other hand, S has $\binom{n}{k}$ k-subsets which are subsets of T and $\binom{n}{k-1}$ k-subsets consisting of a together with a $(k-1)$-subset of T. Since these two types of subsets are disjoint, the result follows by the addition principle. ■

You may be more familiar with Pascal's Identity through Pascal's Triangle

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
\iddots & & & & & \vdots & & & & & \ddots
\end{array}$$

The border entries are always 1. Each inner entry of the triangle is the sum of the two entries diagonally above it.

A nice application of Pascal's Identity is in the proof of the following theorem. We first state one lemma, without proof.

Lemma 1 *When m is a non-negative integer* $\binom{m}{0} = 1 = \binom{m}{m}$.

Theorem 2 (The Binomial Theorem) *When n is a non-negative integer and $x, y \in \mathbb{R}$*

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

Proof by induction on n When $n = 0$ the result is clear. So suppose that for some $n \geq 0$

we have $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$, for any $x, y \in \mathbb{R}$. Then

$$\begin{aligned}
(x+y)^{n+1} &= (x+y)^n(x+y), \text{ by recursive definition of integral exponents}\\
&= \Big[\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}\Big](x+y), \text{ by inductive hypothesis}\\
&= \Big[\sum_{k=0}^{n}\binom{n}{k}x^{k+1} y^{n-k}\Big] + \Big[\sum_{k=0}^{n}\binom{n}{k}x^k y^{n+1-k}\Big]\\
&= \binom{n}{n}x^{n+1} + \Big[\sum_{k=0}^{n-1}\binom{n}{k}x^{k+1} y^{n-k}\Big] + \Big[\sum_{k=1}^{n}\binom{n}{k}x^k y^{n+1-k}\Big] + \binom{n}{0}y^{n+1}\\
&= \binom{n}{n}x^{n+1} + \Big[\sum_{l=1}^{n}\binom{n}{l-1}x^{l} y^{n-(l-1)}\Big] + \Big[\sum_{k=1}^{n}\binom{n}{k}x^k y^{n+1-k}\Big] + \binom{n}{0}y^{n+1}\\
&= \binom{n}{n}x^{n+1} + \Big[\sum_{l=1}^{n}\binom{n}{l-1}x^{l} y^{n+1-l}\Big] + \Big[\sum_{k=1}^{n}\binom{n}{k}x^k y^{n+1-k}\Big] + \binom{n}{0}y^{n+1}\\
&= \binom{n}{n}x^{n+1} + \Big[\sum_{k=1}^{n}\Big[\binom{n}{k-1}+\binom{n}{k}\Big]x^k y^{n+1-k}\Big] + \binom{n}{0}y^{n+1}\\
&= \binom{n+1}{n+1}x^{n+1} + \Big[\sum_{k=1}^{n}\binom{n+1}{k}x^k y^{n+1-k}\Big] + \binom{n+1}{0}y^{n+1}, \text{by Pascal's identity}\\
&= \sum_{k=0}^{n+1}\binom{n+1}{k}x^k y^{n+1-k} \qquad \blacksquare
\end{aligned}$$

From the binomial theorem we can derive facts such as

Theorem 3 *A finite set with n elements has 2^n subsets*

Proof: By the addition principle the number of subsets of an n-set is

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

By the binomial theorem $\sum_{k=0}^{n}\binom{n}{k}1^k 1^{n-k} = (1+1)^n = 2^n$ ■

§1.4 General Inclusion/Exclusion

In general when we are given n finite sets $A_1, A_2, ..., A_n$ and we want to compute the cardinality of their generalized union we use the following theorem.

Theorem 1 *Given finite sets* $A_1, A_2, ..., A_n$

$$\Big|\bigcup_{k=1}^{n} A_k\Big| = \Big[\sum_{k=1}^{n}|A_k|\Big] - \Big[\sum_{1\le j<k\le n}|A_j\cap A_k|\Big] + \Big[\sum_{1\le i<j<k\le n}|A_i\cap A_j\cap A_k|\Big] + ... + (-1)^{n+1}\Big|\bigcap_{k=1}^{n} A_k\Big|$$

Proof: We draw this as a corollary of the next theorem. ■

Let $\mathcal{U}$ be a finite universal set which contains the general union of $A_1, A_2, ..., A_n$. To compute the cardinality of the general intersection of complements of $A_1, A_2, ..., A_n$ we use the general version of DeMorgan's laws and the principle of exclusion. That is

$$\Big|\bigcap_{k=1}^{n} \overline{A_k}\Big| = |\mathcal{U}| - \Big|\overline{\bigcap_{k=1}^{n} \overline{A_k}}\Big| = |\mathcal{U}| - \Big|\bigcup_{k=1}^{n} A_k\Big|.$$

So equivalent to theorem 1 is

Theorem 2 *Given finite sets* $A_1, A_2, ..., A_n$

$$\Big|\bigcap_{k=1}^{n} \overline{A_k}\Big| = |\mathcal{U}| - \Big[\sum_{k=1}^{n}|A_k|\Big] + \Big[\sum_{1\le j<k\le n}|A_j\cap A_k|\Big] - ... + (-1)^{n}\Big|\bigcap_{k=1}^{n} A_k\Big|$$

Proof: Let $x \in \mathcal{U}$. Then two cases to consider are 1) $x \notin A_i$ for all i and 2) $x \in A_i$ for exactly p of the sets A_i, where $1 \le p \le n$.

In the first case, x is counted once on the left hand side. It is also counted only once on the right hand side in the $|\mathcal{U}|$ term. It is not counted in any of the subsequent terms on the right hand side.

In the second case, x is not counted on the left hand side, since it is not in the general intersection of the complements.

Denote the term $|\mathcal{U}|$ as the 0th term, $\sum_{k=1}^{n}|A_k|$ as the 1st term, etc. Since x is a member of exactly p of the sets $A_1, ..., A_n$, it gets counted $\binom{p}{m}$ times in the mth term. (Remember that $\binom{n}{k} = 0$, when $k > n$)

So the total number of times x is counted on the right hand side is

$$\binom{p}{0} - \binom{p}{1} + \binom{p}{2} - ... + (-1)^p\binom{p}{p}.$$

All terms of the form $\binom{p}{k}$, where $k > p$ do not contribute. By the binomial theorem

$$0 = 0^p = (1 + (-1))^p = \binom{p}{0} - \binom{p}{1} + \binom{p}{2} - ... + (-1)^p\binom{p}{p}.$$

So the count is correct. ■

Example 1 How many students are in a calculus class if 14 are math majors, 22 are computer science majors, 15 are engineering majors, and 13 are chemistry majors, if 5 students are double majoring in math and computer science, 3 students are double majoring in chemistry and engineering, 10 are double majoring in computer science and engineering, 4 are double majoring in chemistry and computer science, none are double majoring in math and engineering and none are double majoring in math and chemistry, and no student has more than two majors?

Solution: Let A_1 denote the math majors, A_2 denote the computer science majors, A_3 denote the engineering majors, and A_4 the chemistry majors. Then the information given is
$|A_1| = 14, |A_2| = 22, |A_3| = 15, |A_4| = 13, |A_1 \cap A_2| = 5, |A_1 \cap A_3| = 0, |A_1 \cap A_4| = 0,$
$|A_2 \cap A_3| = 10, |A_2 \cap A_4| = 4, |A_3 \cap A_4| = 3, |A_1 \cap A_2 \cap A_3| = 0, |A_1 \cap A_2 \cap A_4| = 0$
$|A_1 \cap A_3 \cap A_4| = 0, |A_2 \cap A_3 \cap A_4| = 0, |A_1 \cap A_2 \cap A_3 \cap A_4| = 0.$

So by the general rule of inclusion/exclusion, the number of students in the class is $14 + 22 + 15 + 13 - 5 - 10 - 4 - 3 = 32$.

Example 2 How many ternary strings (using 0's, 1's and 2's) of length 8 either start with a 1, end with two 0's or have 4th and 5th positions 12?

Solution: Let A_1 denote the set of ternary strings of length 8 which start with a 1, A_2 denote the set of ternary strings of length 8 which end with two 0's, and A_3 denote the set of ternary strings of length 8 which have 4th and 5th positions 12. By the general rule of inclusion/exclusion our answer is

$$3^7 + 3^6 + 3^6 - 3^5 - 3^5 - 3^4 + 3^3$$

§1.5 Novice Counting

All of the counting exercises you've been asked to complete so far have not been realistic. In general it won't be true that a counting problem fits neatly into a section. So we need to work on the bigger picture.

When we start any counting exercise it is true that there is an underlying exercise at the basic level that we want to consider first. So instead of answering the question immediately we might first want to decide on what type of exercise we have. So far we have seen three types which are distinguishable by the answers to two questions.

1) In forming the objects we want to count, is repetition or replacement allowed?

2) In forming the objects we want to count, does the order of selection matter?

The three scenarios we have seen so far are described in the table below.

Order	Repetition	Type	Form
Y	Y	r-strings	n^r
Y	N	r-permutations	$P(n,r)$
N	N	r-combinations	$\binom{n}{r}$

There are two problems to address. First of all the table above is incomplete. What about, for example, counting objects where repetition is allowed, but order doesn't matter. Second of all, there are connections among the types which make some solutions appear misleading. But as a general rule of thumb, if we correctly identify the type of problem we are working on, then all we have to do is use the principles of addition, multiplication, inclusion/exclusion, or exclusion to decompose our problem into subproblems. The solutions to the subproblems often have the same form as the underlying problem. The principles we employed direct us on how the sub-solutions should be recombined to give the final answer.

As an example of the second problem, if we ask how many binary strings of length 10 contain exactly three 1's, then the underlying problem is an r-string problem. But in this case the answer is $\binom{10}{3}$. Of course this is really $\binom{10}{3}1^3 1^7$ from the binomial theorem. In this case the part of the answer which looks like n^r is suppressed since it's trivial. To see the difference we might ask how many ternary strings of length 10 contain exactly three 1's. Now the answer is $\binom{10}{3}1^3 2^7$, since we choose the three positions for the 1's to go in, and then fill in each of the 7 remaining positions with a 0 or a 2.

To begin to address the first problem we introduce

<u>The Donut Shop Problem</u> If you get to the donut shop before the cops get there, you will find that they have a nice variety of donuts. You might want to order several dozen. They will put your order in a box. You don't particularly care what order the donuts are put into the box. You do usually want more than one of several types. The number of ways for you to complete your order is therefore a counting problem where order doesn't matter, and repetition is allowed.

In order to answer the question of how many ways you can complete your order, we first recast the problem mathematically. From among n types of objects we want to select r objects. If x_i denotes the number of objects of the ith type selected, we have $0 \leq x_i$, (since we cannot choose a negative number of chocolate donouts), also $x_i \in \mathbb{Z}$, (since we cannot select fractional parts of donuts). So the different ways to order are in one-to-one correspondence with the solutions in non-negative integers to $x_1 + x_2 + ... + x_n = r$.

Next, in order to compute the number of solutions in non-negative integers to $x_1 + x_2 + ... + x_n = r$, we model each solution as a string (possibly empty) of x_1 1's followed by a $+$, then a string of x_2 1's followed by a $+$, ... then a string of x_{n-1} 1's followed by a $+$, then a string of x_n 1's. So for example, if $x_1 = 2, x_2 = 0, x_3 = 1, x_4 = 3$ is a solution to $x_1 + x_2 + x_3 + x_4 = 6$ the string we get is $11 + +1 + 111$. So the total number of solutions in non-negative integers to $x_1 + ... + x_n = r$, is the number of binary strings of length $r + n - 1$ with exactly r 1's. From the remark above, this is $\binom{n+r-1}{r}$.

The donut shop problem is not very realistic in two ways. First it is common that some of your order will be determined by other people. You might for example canvas the people in your office before you go to see if there is anything you can pick up for them. So whereas you want to order r donuts, you might have been asked to pick up a certain number of various types.

<u>The More Realistic Donut Shop Problem</u> Now suppose that we know that we want to select r donuts from among n types so that at least $a_i (a_i \geq 0)$ donuts of type i are selected. In terms of our equation, we have $x_1 + x_2 + ... + x_n = r$, where $a_i \leq x_i$, and $x_i \in \mathbb{Z}$. If we set

$y_i = x_i - a_i$ for $i = 1, ..., n$, and $a = \sum_{i=1}^{n} a_i$, then $0 \leq y_i$, $y_i \in \mathbb{Z}$ and

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} (x_i - a_i) = [\sum_{i=1}^{n} x_i] - [\sum_{i=1}^{n} a_i] = r - a$$

So the number of ways to complete our order is $\binom{n + (r-a) - 1}{(r-a)}$.

Still, we qualified the donut shop problem by supposing that we arrived before the cops did.

The Real Donut Shop Problem If we arrive at the donut shop after canvassing our friends, we want to select r donuts from among n types. The problem is that there are probably only a few left of each type. This may place an upper limit on how often we can select a particular type. So now we wish to count solutions to $a_i \leq x_i \leq b_i$, $x_i \in \mathbb{Z}$, and $x_1 + x_2 + ... + x_n = r$. We proceed by replacing r by $s = r - a$, where a is the sum of lower bounds. We also replace b_i by $c_i = b_i - a_i$ for $i = 1, ..., n$. So we want to find the number of solutions to $0 \leq y_i \leq c_i$, $y_i \in \mathbb{Z}$, and $y_1 + y_2 + ... + y_n = s$. There are several ways to proceed. We choose inclusion/exclusion. Let us set $\mathcal{U}$ to be all solutions in non-negative integers to $y_1 + ... + y_n = s$. Next let A_i denote those solutions in non-negative integers to $y_1 + ... + y_n = r$, where $c_i < y_i$. Then we want to compute $|\overline{A_1} \cap \overline{A_2} \cap \overline{A_3} \cap ... \cap \overline{A_n}|$, which we can do by general inclusion/exclusion, and the ideas from the more realistic donut shop problem.

Example 1 Let us count the number of solutions to $x_1 + x_2 + x_3 + x_4 = 34$ where $0 \leq x_1 \leq 4, 0 \leq x_2 \leq 5, 0 \leq x_3 \leq 8$ and $0 \leq x_4 \leq 40$. So as above we have $c_1 = 4, c_2 = 5, c_3 = 8$, and $c_4 = 40$. Also A_i will denote the solutions in non-negative integers to $x_1 + x_2 + x_3 + x_4 = 34$, with $x_i > c_i$, $i = 1, 2, 3, 4$. So $|\mathcal{U}| = \binom{34+4-1}{34}$. Next realize that $A_4 = \emptyset$, so $\overline{A_4} = \mathcal{U}$ and $\overline{A_1} \cap \overline{A_2} \cap \overline{A_3} \cap \overline{A_4} = \overline{A_1} \cap \overline{A_2} \cap \overline{A_3}$ Now to compute A_1, we must first rephrase $x_1 > 4$ as a non-strict inequality, i.e. $5 \leq x_1$. So $|A_1| = \binom{29+4-1}{29}$. Similarly $|A_2| = \binom{28+4-1}{28}$, and $|A_3| = \binom{25+4-1}{25}$. Next we have that $A_1 \cap A_2$ is all solutions in non-negative integers to $x_1 + x_2 + x_3 + x_4 = 34$ with $5 \leq x_1$ and $6 \leq x_2$. So $|A_1 \cap A_2| = \binom{23+4-1}{23}$. Also $|A_1 \cap A_3| = \binom{20+4-1}{20}$ and $|A_2 \cap A_3| = \binom{19+4-1}{19}$. Finally $|A_1 \cap A_2 \cap A_3| = \binom{14+4-1}{14}$. So the final answer is

$$\binom{34+4-1}{34} - \binom{29+4-1}{29} - \binom{28+4-1}{28} - \binom{25+4-1}{25} + \binom{23+4-1}{23} + \binom{20+4-1}{20} + \binom{19+4-1}{19} - \binom{14+4-1}{14}$$

We can now solve general counting exercises where order is unimportant and repetition is restricted somewhere between no repetition, and full repetition.

To complete the picture we should be able to also solve counting exercises where order is important and repetition is partial. This is somewhat easier. It suffices to consider the subcases in the next example.

Example 2 Let us take as initial problem the number of quaternary strings of length 15. There are 4^{15} of these. Now if we ask how many contain exactly two 0's, the answer is $\binom{15}{2}3^{13}$. If we ask how many contain exactly two 0's and four 1's, the answer is $\binom{15}{2}\binom{13}{4}2^9$. And if we ask how many contain exactly two 0's, four 1's and five 2's, the answer is $\binom{15}{2}\binom{13}{4}\binom{9}{5}\binom{4}{4}$.

So in fact many types of counting are related by what we call the multinomial theorem.

Theorem 1 *When r is a non-negative integer and $x_1, x_2, ..., x_n \in \mathbb{R}$*

$$(x_1 + x_2 + ... + x_n)^r = \sum_{\substack{e_1+e_2+...+e_n=r \\ 0 \leq e_i}} \binom{r}{e_1, e_2, ...e_n} x_1^{e_1} x_2^{e_2} ... x_n^{e_n},$$

where $\binom{r}{e_1, e_2, ...e_n} = \dfrac{r!}{e_1!e_2!...e_n!}$.

To recap, when we have a basic counting exercise, we should first ask whether order is important and then ask whether repetition is allowed. This will get us into the right ballpark as far as the form of the solution. We must use basic counting principles to decompose the exercise into sub-problems. Solve the sub-problems, and put the pieces back together. Solutions to sub-problems usually take the same form as the underlying problem, though they may be related to it via the multinomial theorem. The table below synopsizes six basic cases.

Order	Repetition	Form
Y	Y	n^r
Y	N	$P(n, r)$
N	Y	$\binom{r+n-1}{r}$
N	N	$\binom{n}{r}$
Y	some	$\binom{r}{k_1, k_2, ..., k_n}$
N	some	$\binom{r+n-1}{r}$ w/ I-E

§1.6 Occupancy Problems

The purpose of this ultimate section is to show that some basic counting exercises can be re-phrased as so-called occupancy problems. A consequence will be that we can easily introduce occupancy problems which are not amenable to the elementary tactics we have dealt with so far. It's in order to solve these types of problems that we will be generating more counting tactics in chapters 3 and 4.

The basic occupancy problem has us placing n objects into k containers/boxes. To classify the type of occupancy problem we have, we must answer three yes/no questions. There will therefore be $8 = 2^3$ basic occupancy problems. The three questions are:

1) Are the objects distinguishable from one another?
2) Are the boxes distinguishable from one another?
3) Can a box remain unoccupied?

If the answer to all three questions is yes, then the number of ways to place the n objects into the k boxes is clearly the number of functions from an n-set to a k-set, which is k^n.

If, on the other hand the answer to the first question is no, but the other two answers are yes, then we have the basic donut shoppe problem. So the number of ways to distribute n identical objects among k distinguishable boxes is the number of solutions in non-negative whole numbers to $x_1 + x_2 + ... + x_k = n$, where x_i is the number of objects placed in the ith box.

If we change the answer to the third question to no, then we have the more realistic donut shoppe problem. Now we need the number of solutions in positive integers to $x_1 + ... + x_k = n$, or equivalently the number of solutions in non-negative integers to $y_1 + ... + y_k = n - k$. This is $C((n-k)+k-1, n-k) = C(n-1, n-k) = C(n-1, k-1)$.

So it might appear that there is nothing really new here. That every one of our occupancy problems can be solved by an elementary counting technique. However if we define $S(n,k)$ to be the number of ways to distribute n distinguishable objects into k indistinguishable boxes we will derive in chapter 3 that

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

Upon making this definition we can answer three more of our eight basic occupancy problems. This is summarized in the table.

Objects Distinguished	Boxes Distingushed	Empty boxes allowed	Number of ways to complete
Y	Y	Y	k^n
Y	Y	N	$k!S(n,k)$
N	Y	Y	$\binom{k+n-1}{k}$
N	Y	N	$\binom{n-1}{k-1}$
Y	N	Y	$\sum_{i=1}^{k} S(n,i)$
Y	N	N	$S(n,k)$

The numbers $S(n,k)$ are called the Stirling numbers of the second kind. The table indicates their relative importance for counting solutions to occupancy problems.

We close this section by pointing out that two occupancy problems remain. A partition of a positive integer n is a collection of positive integers which sum to n.

Example: The partitions of 5 are $\{5\}, \{4,1\}, \{3,2\}, \{3,1,1\}, \{2,2,1\}, \{2,1,1,1\}, \{1,1,1,1,1\}$.

So the number of ways to place n indistinguishable objects into k indistinguishable boxes if no box is empty is the number of partitions of n into exactly k parts. If we denote this by $p_k(n)$, then we see that $p_2(5) = 2$. Also $p_1(n) = p_n(n) = 1$ for all positive integers n. Meanwhile $p_k(n) = 0$ is $k > n$. Finally $p_2(n) = \lceil (n-1)/2 \rceil$.

The final occupancy problem is to place n indistinguishable objects into k indistinguishable boxes if some boxes may be empty. The number of ways this can be done is $\sum_{i=1}^{k} p_i(n)$. This is the number of partitions of n into k or fewer parts.

A more complete discussion of the partitions of integers into parts can be found in most decent number theory books, or a good combinatorics reference book like Marshall Hall's.

Chapter 1 Exercises

1. How many non-negative whole numbers less than 1 million contain the digit 2?

2. How many bit strings have length 3, 4 or 5?

3. How many whole numbers are there which have five digits, each being a number in $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, and either having all digits odd or having all digits even?

4. How many 5-letter words from the lowercase English alphabet either start with f or do not have the letter f?

5. In how many ways can we get a sum of 3 or a sum of 4 when two dice are rolled?

6. List all permutations of $\{1, 2, 3\}$. Repeat for $\{1, 2, 3, 4\}$.

7. How many permutations of $\{1, 2, 3, 4, 5\}$ begin with 5?

8. How many permutations of $\{1, 2, 3, ..., n\}$ begin with 1 and end with n?

9. Find a) $P(3, 2)$, b) $P(5, 3)$, c) $P(8, 5)$, d) $P(1, 3)$.

10. Let $A = \{0, 1, 2, 3, 4, 5, 6\}$.

    a) Find the number of strings of length 4 using elements of A.

    b) Repeat part a, if no element of A can be used twice.

    c) Repeat part a, if the first element of the string is 3

    d) Repeat part c, if no element of A can be used twice.

11. Enumerate the subsets of $\{a, b, c, d\}$.

12. If A is a 9-set, how many nonemepty subsets does A have?

13. If A is an 8-set, how many subsets with more than 2 elements does A have?

14. Find a) $C(6, 3)$, b) $C(7, 4)$, c) $C(n, 1)$, d) $C(2, 5)$.

15. In how many ways can 8 blood samples be divided into 2 groups to be sent to different laboratories for testing, if there are four samples per group.

16. Repeat exercise 15, if the laboratories are not distinguishable.

17. A committee is to be chosen from a set of 8 women and 6 men. How many ways are there to form the committee if

    a) the committee has 5 people, 3 women and 2 men?

    b) the committee has any size, but there are an equal number of men and women?

    c) the committee has 7 people and there must be more men than women?

18. Prove that $\binom{n}{m}\binom{m}{k} = \binom{n}{k}\binom{n-k}{m-k}$.

19. Give a combinatorial argument to prove Vandermonde's identity

$$\binom{n+m}{r} = \binom{n}{0}\binom{m}{r} + \binom{n}{1}\binom{m}{r-1} + ... + \binom{n}{k}\binom{m}{r-k} + ... + \binom{n}{r}\binom{m}{0}$$

20. Prove that $\binom{n}{0}+\binom{n+1}{1}+\binom{n+2}{2}+...+\binom{n+r}{r}=\binom{n+r+1}{r}$.

21. Calculate the probabilty that when a fair 6-sided die is tossed, the outcome is

a) an odd number.

b) a number less than or equal to 2.

c) a number divisible by 3,

22. Calculate the probability that in 4 tosses of a fair coin, there are at most 3 heads.

23. Calculate the probability that a family with three children has

a) exactly 2 boys.

b) at least 2 boys.

c) at least 1 boy and at least 1 girl.

24. What is the probability that a bit string of length 5, chosen at random, does not have two consecutive zeroes?

25. Suppose that a system with four independent components, each of which is equally likely to work or to not work. Suppose that the system works if and only if at least three components work. What is the probability that the system works?

26. In how many ways can we choose 8 bottles of soda if there are 5 brands to choose from?

27. Find all partitions of a) 4, b) 6, c) 7.

28. Find all partitions of 8 into four or fewer parts.

29. Compute a) $S(n,0)$, b) $S(n,1)$, c) $S(n,2)$, d) $S(n,n-1)$, e) $S(n,n)$

30. Show by a combinatorial argument that $S(n,k)=kS(n-1,k)+S(n-1,k-1)$.

31. How many solutions in non-negative integers are there to $x_1+x_2+x_3+x_4=18$ which satisfy $1\leq x_i\leq 8$ for $i=1,2,3,4$?

32. Expand a) $(x+y)^5$, b) $(a+2b)^3$, c) $(2u+3v)^4$

33. Find the coefficient of x^{11} in the expansion of

a) $(1+x)^{15}$, b) $(2+x)^{13}$, c) $(2x+3y)^{11}$

34. What is the coefficient of x^{10} in the expansion of $(1+x)^{12}(1+x)^4$?

35. What is the coefficient of a^3b^2c in the expansion of $(a+b+c+2)^8$?

36. How many solutions in non-negative integers are there to $x_1+x_2+x_3+x_4+x_5=47$ which satisfy $x_1\leq 6$, $x_2\leq 8$, and $x_3\leq 10$?

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

38. Find $\sum_{k=0}^{n}4^k\binom{n}{k}$.

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

40. An octapeptide is a chain of 8 amino acids, each of which is one of 20 naturally occurring amino acids. How many octapeptides are there?

41. In an RNA chain of 15 bases, there are 4 A's, 6 U's, 4 G's, and 1 C. If the chain begins with AU and ends with UG, how many chains are there?

42. An ice cream parlor offers 29 different flavors. How many different triple cones are possible if each scoop on the cone has to be a different flavor?

43. A cigarette company surveys 100,000 people. Of these 40,000 are males, according to the company's report. Also 80,000 are smokers and 10,000 of those surveyed have cancer. However, of those suveyed, there are 1000 males with cancer, 2000 smokers with cancer, and 3000 male smokers. Finally there are 100 male smokers with cancer. How many female nonsmokers without cancer are there? Is there something wrong with the company's report?

44. One hundred water samples were tested for traces of three different types of chemicals, mercury, arsenic, and lead. Of the 100 samples 7 were found to have mercury, 5 to have arsenic, 4 to have lead, 3 to have mercury and arsenic, 3 to have arsenic and lead, 2 to have mercury and lead, and 1 to have mercury, arsenic, but no lead. How many samples had a trace of at least one of the three chemiclas?

45. Of 100 cars tested at an inspection station, 9 had defective headlights, 8 defective brakes, 7 defective horns, 2 defective windshield wipers, 4 defective headlights and brakes, 3 defective headlights and horns, 2 defective headlights and windshield wipers, 1 defective horn and windshield wipers, 1 had defective headlights, brakes and horn, 1 had defective headlights, horn, and windshield wipers, and none had any other combination of defects. Find the number of cars which had at least one of the defects in question.

46. How many integers between 1 and 10,000 inclusive are divisible by none of $5, 7$, and 11?

47. A multiple choice test contains 10 questions. There are four possible answers for each question.
a) How many ways can a student answer the questions if every question must be answered?
b) How many ways can a student answer the questions if questions can be left unanswered?

48. How many positive integers between 100 and 999 inclusive are divisible by 10 or 25?

49. How many strings of eight lowercase English letters are there

a) if the letters may be repeated?

b) if no letter may be repeated?

c) which start with the letter x, and letters may be repeated?

d) which contain the letter x, and the letters may be repeated?

e) which contain the letter x, if no letter can be repeated?

f) which contain at least one vowel (a, e, i, o or u), if letters may be repeated?

g) which contain exactly two vowels, if letters may be repeated?

h) which contain at least one vowel, where letters may not be repeated?

50. How many bit strings of length 9 either begin “00”, or end “1010”?

51. In how many different orders can six runners finish a race if no ties occur?

52. How many subsets with an odd number of elements does a set with 10 elements have?

53. How many bit strings of length 9 have

a) exactly three 1’s?

b) at least three 1’s?

c) at most three 1’s?

d) more zeroes than ones?

54. How many bit strings of length ten contain at least three ones and at least three zeroes?

55. How many ways are there to seat six people around a circular table where seatings are considered to be equivalent if they can be obtained from each other by rotating the table?

56. Show that if n is a positive integer, then $\binom{2n}{2} = 2\binom{n}{2} + n^2$

a) using a combinatorial argument.

b) by algebraic manipulation.

57. How many bit strings of length 15 start with the string 101, end with the string 1001 or have 3rd through 6 bits 1010?

58. How many positive integers between 1000 and 9999 inclusive are not divisible by any of $4, 10$ and 25?

59. How many quaternary strings of length n are there (a quaternary string uses 0’s, 1’s, 2’s, and 3’s)?

60. How many solutions in integers are there to $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 54$, where $3 \leq x_1$, $4 \leq x_2$, $5 \leq x_3$, and $6 < x_4, x_5, x_6$?

61. How many strings of twelve lowercase English letters are there

a) which start with the letter x, if letters may be repeated?

b) which contain the letter x, if letters can be repeated?

c) which contain the letters x and y, if letters can be repeated?

d) which contain at least one vowel, where letters may not be repeated?

62. How many bit strings of length 19 either begin “00”, or have 4th, 5th and 6th digits “101”, or end “1010”?

63. How many pentary strings of length 15 consist of two 0’s, four 1’s, three 2’s, five 3’s and one 4?

64. How many ternary strings of length 9 have

a) exactly three 1’s?

b) at least three 1’s?

c) at most three 1’s?

Chapter 2: Introduction to Graphs

You should already have some experience representing set-theoretic objects as digraphs. In this chapter we introduce some ideas and uses of undirected graphs, those whose edges are not directed.

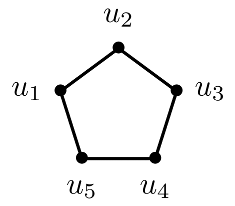
§2.1 Graph Terminology

Loosely speaking, an undirected graph is a doodle, where we have a set of points (called vertices). Some of the points are connected by arcs (called edges). If our graph contains loops, we call it a pseudograph. If we allow multiple connections between vertices we have a multigraph. Clearly in order to understand pseudographs and multigraphs it will be necessary to understand the simplest case, where we do not have multiple edges, directed edges, or loops. Such an undirected graph is called a simple graph if we need to distinguish it from a pseudograph or a multigraph. Henceforth in this chapter, unless specified otherwise, graph means undirected, simple graph.

Formally a graph, $G = (V, E)$ consists of a set of vertices V and a set E of edges, where any edge $e \in E$ corresponds to an unordered pair of vertices $\{u, v\}$. We say that the edge e is incident with u and v. The vertices u and v are adjacent, when $\{u, v\} \in E$. otherwise they are not. We often write $u \sim v$ to denote that u and v are adjacent. We also call u and v neighbors in this case. Of course $u \not\sim v$ denotes that $\{u, v\} \notin E$. All of our graphs will have finite vertex sets, and therefore finite edge sets.

Most often we won't want to deal with the set-theoretic version of a graph, we will want to work with a graphical representation, or a $0,1$-matrix representation. This presents a problem since there is possibly more than one way of representing the graph either way. We will deal with this problem formally in the next section. To represent a graph graphically we draw a point for each vertex, and use arcs to connect those points corresponding to adjacent vertices. To represent a graph as a $0,1$-matrix we can either use an adjacency matrix or an incidence matrix. In the first case we use the vertex set in some order to label rows and columns (same order) of a $|V| \times |V|$ matrix. The entry in the row labeled u and column labeled v is 1 if $u \sim v$ and 0 if $u \not\sim v$. In the second case we use V to index the rows of a $|V| \times |E|$ matrix, and E to index the columns. The entry in the row labeled u and column labeled e is 1 if u is incident with e, and 0 otherwise.

Example: Let $G = (\{u_1, u_2, u_3, u_4, u_5\}, \{\{u_1, u_2\}, \{u_2, u_3\}, \{u_3, u_4\}, \{u_4, u_5\}, \{u_5, u_1\}\})$. We represent G graphically, using an adjacency matrix A_G, and using an incidence matrix M_G.



$$A_G = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad M_G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

All of these represent the same object. When doing graph theory we usually think of the graphic object.

Now that we know what graphs are, we are naturally interested in their behavior. For example, for a vertex v in a graph $G = (V, E)$ we denote the number of edges incident with v as $deg(v)$, the degree of v. For a digraph it would then make sense to define the in-degree of a vertex as the number of edges into v, and similarly the out-degree. These are denoted by $id(v)$ and $od(v)$ respectively.

Theorem (Hand-Shaking Theorem) *In any graph* $G = (V, E)$, $\sum_{v \in V} deg(v) = 2|E|$.

Proof: Every edge is incident with two vertices (counting multiplicity for loops). ∎

Corollary *In an undirected graph there are an even number of vertices of odd degree.*

Corollary *In a digraph* D, $\sum_{v \in V} id(v) = \sum_{v \in V} od(v) = |E|$.

In order to explore the more general situations it is handy to have notation to describe certain special graphs. The reader is strongly encouraged to represent these graphically.

1) For $n \geq 0$, K_n denotes the simple graph on n vertices where every pair of vertices is adjacent. K_0 is of course the empty graph. K_n is the complete graph on n vertices.

2) For $n \geq 3$, C_n denotes the simple graph on n vertices, $v_1, ..., v_n$, where $E = \{\{v_i, v_j\} | j - i \equiv \pm 1 \pmod{n}\}$. C_n is the n-cycle.

3) For $n \geq 2$, L_n denotes the n-link. $L_2 = K_2$, and for $n > 2$ L_n is the result of removing any edge from C_n.

4) For $n \geq 3$, W_n denotes the n-wheel. To form W_n add one vertex to C_n and make it adjacent to every other vertex.

5) For $n \geq 0$, the n-cube, Q_n, is the graph whose vertices are all binary strings of length n. Two vertices are adjacent only if they differ in exactly one position.

6) A graph is bipartite if there is a partition $V = V_1 \cup V_2$ so that any edge is incident with one vertex from each part of the partition. In case every vertex of V_1 is adjacent to every vertex of V_2 and $|V_1| = m$ with $|V_2| = n$, the result is the complete bipartite graph $K_{m,n}$.

As you might guess from the constructions of L_n and W_n from C_n it makes sense to discuss the union and intersection of graphs. For a simple graph on n vertices, G, it even makes sense to discuss the complement $\overline{G}$ (relative to K_n).

As far as subgraphs are concerned we stress that a subgraph $H = (W, F)$ of a graph $G = (V, E)$, has $W \subseteq V$, $F \subseteq E$, and if $f = \{u, v\} \in F$, then both u and v are in W. Finally we define the induced subgraph on a subset W of V to be the graph with vertex set W, and all edges $f = \{u, v\} \in E$, where $u, v \in W$.

§2.2 Graph Isomorphism

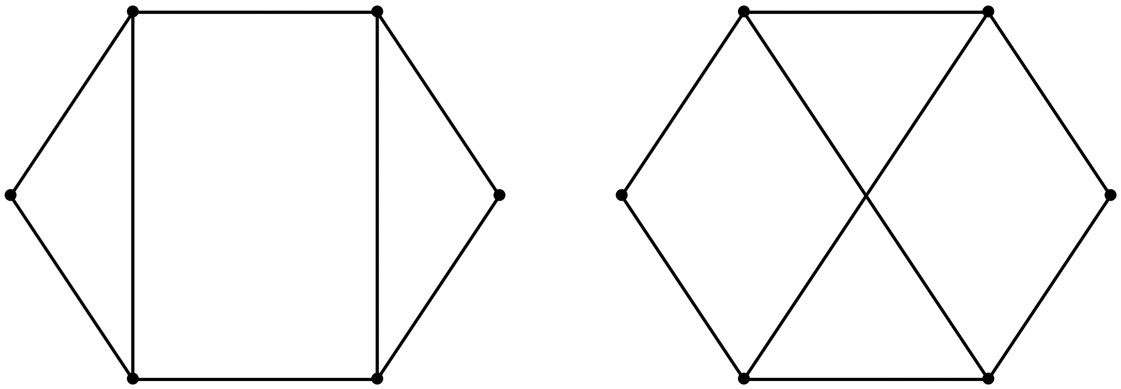
In this section we deal with the problem that there is more than one way to present a graph. Informally two graphs $G = (V, E)$ and $H = (W, F)$ are isomorphic, if one can be redrawn to be identical to the other. Thus the two graphs would represent equivalent set-theoretic objects. Clearly it is necessary that $|V| = |W|$ and $|E| = |F|$.

Formally graphs $G = (V, E)$ and $H = (W, F)$ are isomorphic if there exists a bijective function $\varphi : V \longrightarrow W$, called a graph isomorphism, with the property that $\{u, v\} \in E$ iff $\{\varphi(u), \varphi(v)\} \in F$. We write $G \cong H$ in case such a function exists. $G \ncong H$ signifies that G and H are not isomorphic.

The property in the definition is called the adjacency-preserving property. It is absolutely essential since for example L_4 and $K_{1,3}$ are both graphs with 4 vertices and 3 edges, yet they are not isomorphic. In fact the adjacency-preserving property of a graph isomorphism guarantees that $deg(u) = deg(\varphi(u))$ for all $u \in V$. In particular if $G \cong H$ and the degrees of the vertices of G are listed in increasing order, then this list must be identical to the sequence formed when the degrees of the vertices of H are listed in increasing order. The list of degrees of a graph, G, in increasing order is its degree sequence, and is denoted $ds(G)$. Thus $G \cong H$ implies $ds(G) = ds(H)$. Equivalently $ds(G) \neq ds(H)$ implies $G \ncong H$.

However $ds(G) = ds(H)$ is not sufficient for $G \cong H$ as the following example indicates.

Example:

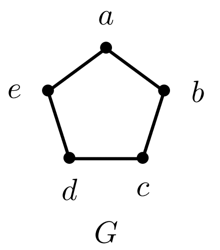


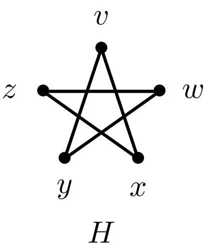
G H

The graph H is bipartite, the graph G is not. Since H can be re-drawn as a two-part graph, and G cannot, G cannot be isomorphic to H.

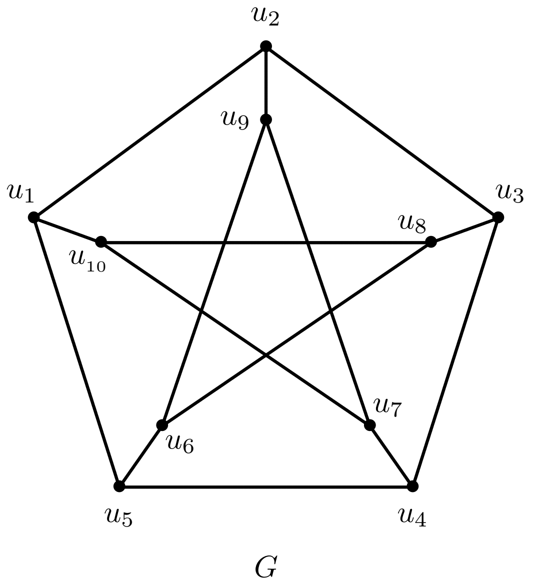
So given two graphs with identical degree sequences and a bijective function φ between their vertex sets which preserves degree, we must still show that φ preserves adjacency before we can conclude that the two graphs are isomorphic. This is most efficiently accomplished by representing G via an adjacency matrix A_G with respect to an ordering $v_1, v_2, ..., v_n$ of its vertex set, and comparing it to the representation of H via an adjacency matrix A_H with respect to the ordering $\varphi(v_1), \varphi(v_2), ..., \varphi(v_n)$. φ is a graph isomorphism iff $A_G = A_H$ iff $A_G \oplus A_H = 0$.

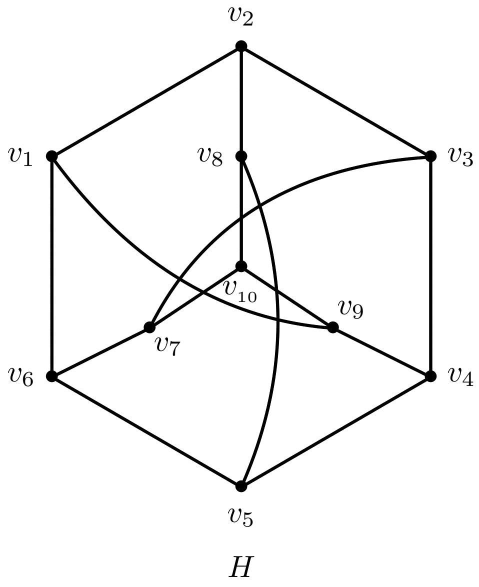
Example: Let G be a 5-cycle on a, b, c, d, e drawn as a regular pentagon with vertices arranged clockwise, in order, at the corners. Let H have vertex set v, w, x, y, z and graphical presentation as a pentagram (five-pointed star), where the vertices of the graph are the ends of the points of the star, and are arranged clockwise, in order. Then $\varphi = \{(a, v), (b, x), (c, z), (d, w), (e, y)\}$ is a graph isomorphism from G to H.





Example: The two graphs below are isomorphic under the map $\varphi = \{(u_1, v_1), (u_2, v_2), (u_3, v_3), (u_4, v_4), (u_5, v_9), (u_6, v_{10}), (u_7, v_5), (u_8, v_7), (u_9, v_8), (u_{10}, v_6)\}$.



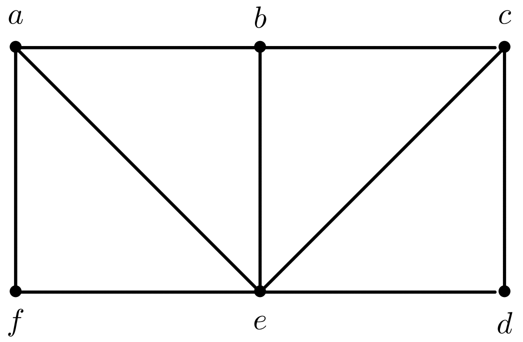


The graph G is the standard graphical presentation of what is called Petersen's Graph. Notice that it could be described as the graph whose vertex set is all 2-sets of a 5-set, and where $u \sim v$ iff $|u \cap v| = 0$.

§2.3 Paths

A path of length n in a graph is an alternating sequence of vertices and edges of the form $v_0, e_1, v_1, e_2, ..., v_{n-1}, e_n, v_n$, where $e_i = \{v_{i-1}, v_i\}$, for $i = 1, ..., n$. A path is simple if no edge is repeated. A circuit is a path with $v_0 = v_n$. A simple circuit is a cycle. In a digraph we require $e_i = (v_{i-1}, v_i)$, for $i = 1, ..., n$. In a simple graph we can and will suppress the edges and therefore consider any string of pairwise incident vertices as a path.

Example: In the graph below a, b, e, d, c, d, c is a path of length 6. a, f, b, d, c, d, e, a is not a path, since neither of $\{b, f\}$ or $\{b, d\}$ is an edge. a, b, e, b, a is a circuit, but not a cycle since $\{a, b\}$ and $\{b, e\}$ are used more than once. a, b, c, d, e, f, a is a cycle of length 6.



A graph is connected if there is a path in the graph between any two vertices. As a matter of fact, one can prove:

Theorem *If G is an undirected, connected graph, then there is a simple path between any two vertices.*

Proof: Exercise.

In a graph the distance between two vertices is defined as the length of the shortest simple path between them. For example, in the graph from exercise 1, the distance from a to d is 2. When a graph is not connected, it's maximal connected subgraphs are called components. If two vertices in a graph are in different components our convention is that their distance is ∞.

A vertex in a graph is a cutvertex, if removal of the vertex and its incident edges results in a graph with more components. Similarly a bridge is an edge whose removal yields a graph with more components.

We close this section with a discussion of two special types of paths. The description of the paths is remarkably similar. The point of the discussion is that in the area of discrete mathematics one can turn an easy problem into a hard one, just by changing a few words.

An Eulerian path in a graph is a simple path which uses every edge of the graph. An Eulerian cycle is an Eulerian path which is also a cycle. This type of path is interesting in that if a graph is Eulerian (has an Eulerian path or cycle) then it can be drawn completely without lifting one's writing utensil from the writing surface and without retracing any edges.

Example: The graph K_5 is Eulerian, in fact it has an Eulerian cycle.

Example: The graph L_n is Eulerian, but does not have an Eulerian cycle.

Example: The graph K_4 is not Eulerian. Try it.

A Hamiltonian path in a graph is a simple path which uses every vertex exactly once. A Hamiltonian cycle is one of the form $v_0, v_1, ..., v_n, v_0$, where $v_0, ..., v_n$ is a Hamiltonian path.

Example: K_n is Hamiltonian for $n \geq 0$, and has a Hamiltonian cycle for $n \geq 3$.

Example: W_n has a Hamiltonian cycle for $n \geq 3$.

Example: L_n has a Hamiltonian path, but no Hamiltonian cycle for $n \geq 2$

These two types of path are similar, in that there is a list of necessary conditions which a graph must satisfy, if it is to possess either type of cycle. If G is a graph with either an Eulerian or Hamiltonian cycle, then

1) G is connected.
2) every vertex has degree at least 2.
3) G has no bridges.
4) If G has a Hamiltonian cycle, then G has no cutvertices.

These types of path are different in that Leonhard Euler completely solved the problem of which graphs are Eulerian. Moreover the criteria is surprising simple. In contrast, no one has been able to find a similar solution for the problem of which graphs are Hamiltonian.

Spurred by the Sunday afternoon pastime of people in Kaliningrad, Russia Euler proved the following theorem.

Theorem *A connected multigraph has an Eulerian cycle iff every vertex has even degree.*

Proof: Let G be a connected multigraph with an Eulerian cycle and suppose that v is a vertex in G with $deg(v) = 2m + 1$, for some $m \in \mathbb{N}$. Let i denote the number of times the cycle passes through v. Since every edge is used exactly once in the cycle, and each time v is visited 2 different edges are used, we have $2i = 2m + 1 \longrightarrow\!\!\!\!\!\times\!\!\!\longleftarrow$.

Conversely, let G be a connected multigraph where every vertex has even degree. Select a vertex u and build a simple path P starting at u. Each time a vertex is reached we add any edge not already used. Any time a vertex $v \neq u$ is reached its even degree guarantees a new edge out, since we used one edge to arrive there. Since G is finite, we must reach a vertex where P cannot continue. And this vertex must be u by the preceding remark. Therefore P is a cycle.

If this cycle contains every edge we are done. Otherwise when these edges are removed from G we obtain a set of connected components $H_1, ..., H_m$ which are subgraphs of G and which each satisfy that all vertices have even degree. Since their sizes are smaller, we may inductively construct an Eulerian cycle for each H_i. Since each G is connected, each H_i contains a vertex of the initial cycle, say v_j. If we call the Eulerian cycle of H_i, C_i, then $v_0, ...v_j, C_i, v_j, ..., v_n, v_0$ is a cycle in G. Since the H_i are disjoint, we may insert each Eulerian subcycle thus obtaining an Eulerian cycle for G. ■

As a corollary we have

Theorem *A connected multigraph has an Eulerian path, but no Eulerian cycle iff it has exactly two vertices of odd degree.*

The following theorem is an example of a sufficient condition for a graph to have a Hamiltonian cycle. This condition is clearly not necessary by considering C_n for $n \geq 5$.

Theorem *Let G be a connected, simple graph on $n \geq 3$ vertices. If $deg(v) \geq n/2$ for every vertex v, then G has a Hamiltonian cycle.*

Proof: Suppose that the theorem is false. Let G satisfy the conditions on vertex degree, connectivity, and simplicity. Moreover suppose that of all counterexamples on n vertices, G is maximal with respect to the number of edges.

G is not complete, since K_n has a Hamiltonian cycle, for $n \geq 3$. Therefore G has two vertices v_1 and v_n with $v_1 \not\sim v_n$. By maximality the graph $G_1 = G \cup \{v_1, v_n\}$ has a Hamiltonian cycle. Moreover this cycle uses the edge $\{v_1, v_n\}$, else G has a Hamiltonian cycle. So we may suppose that the Hamiltonian cycle in G_1 is of the form $v_1, v_2, ..., v_n, v_1$. Thus $v_1, ..., v_n$ is a Hamiltonian path in G.

Let $k = deg(v_1)$. So $k = |S| = |\{v \in V | v_1 \sim v\}|$. If $v_{i+1} \in S$, then $v_i \not\sim v_n$, else $v_1, ..., v_i, v_n, v_{n-1}, ..., v_{i+1}, v_1$ is a Hamiltonian cycle in G. Therefore

$$deg(v_n) \leq (n-1) - k \leq n - 1 - n/2 = n/2 - 1. \text{ —×—} \qquad \blacksquare$$

§2.4 Trees

Trees are one of the most important classes of graphs. A tree is a connected, undirected graph, with no cycles. Consequently a tree is a simple graph. Moreover we have

Theorem *A graph G is a tree iff there is a unique simple path between any two vertices.*

Proof: Suppose that G is a tree, and let u and v be two vertices of G. Since G is connected, there is a simple path P of the form $u = v_0, v_1, ..., v_n = v$. If Q is a different simple path from u to v, say $u = w_0, w_1, ..., w_n = v$ let i be the smallest subscript so that $w_i = v_i$, but $v_{i+1} \neq w_{i+1}$. Also let j be the next smallest subscript where $v_j = w_j$. By construction $v_i, v_{i+1}, ..., v_j, w_{j-1}, w_{j-2}, ..., w_i$ is a cycle in G —×—.

Conversely, if G is a graph where there is a unique simple path between any pair of vertices, then by definition G is connected. If G contained a cycle, C, then any two vertices of C would be joined by two distinct simple paths.—×— Therefore G contains no cycles, and is a tree. ∎

A consequence of theorem 1 is that given any vertex r in a tree, we can draw T with r at the top and the other vertices in levels below. The neighbors of r thus appear at the first level and are called r's children. The neighbors of r's children are put in the second level, and are r's grandchildren. In general the ith level consists of those vertices in the tree which are at distance i from r. The result is called a rooted tree. A rooted tree is by default directed, but we suppress the arrows on edges since every edge is drawn downwards. The height of a rooted tree is the maximum level number.

Naturally, besides child and parent, many geneological terms apply to rooted trees, and are suggestive of the structure. For example if $T = (V, E, r)$ is a rooted tree with root r, and $v \in V - \{r\}$, the ancestors of v are all vertices on the path from r to v, including r, but excluding v. The descendants of a vertex, w consist of all vertices which have w as one of their ancestors. The subtree rooted at w is the rooted tree consisting of w, its descendants, and all requisite paths. A vertex with no children is a leaf, and a vertex with at least one child is called an internal vertex.

To distinguish rooted trees by breadth, we use the term m-ary to mean that any internal vertex has at most m children. An m-ary tree is full if every internal vertex has exactly m children. When $m = 2$, we use the term binary.

As an initial application of rooted trees we prove the following theorem.

Theorem *A tree on n vertices has $n - 1$ edges.*

Proof: Let $T = (V, E)$ be a tree with n vertices. Let $u \in V$ and form the rooted tree $T = (V, E, u)$ rooted at u. Any edge $e \in E$ joins two vertices v and w where v is the parent of w. This allows us to define a function $f : E \longrightarrow V - \{u\}$ by $f(e) = w$. f is one-to-one by uniqueness of simple path from u to w. f is onto by connectivity. Therefore $|E| = |V - \{u\}| = |V| - 1 = n - 1$. ■

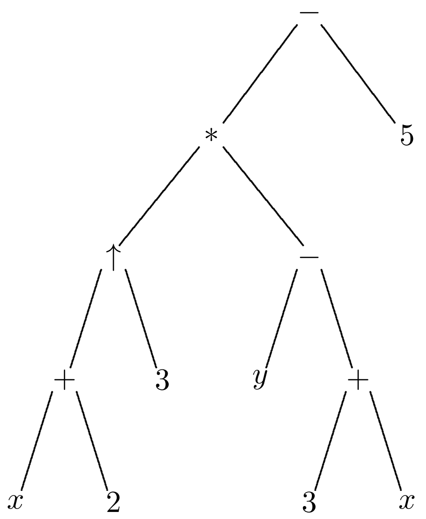
We draw as corollary

Corollary A full m-ary tree with i internal vertices has $n = mi + 1$ vertices.

Since every vertex in a rooted tree is either internal or a leaf, we know that a full m-ary tree with i internal vertices has $l = (m-1)i + 1$ leaves. In short, if we know any two of the three quantities n, i and l for a full m-ary tree, we can deduce the third.

A very important application of full, rooted, binary trees is their use to model arithmetic expressions. In this case the last operation performed acts as the root. Call this operation $\star$. $\star$ is usually one of addition, subtraction, multiplication, or division. Since order of evaluation matters when we subtract or divide, we need to also order the tree distinguising each pair of children as left child and right child. Our expression is then modeled as $T_1 \star T_2$, where T_1 is the left child, and each of T_1, and T_2 may be constructed recursively.

Example: The expression $((x+2) \uparrow 3) * (y - (3+x)) - 5$ is modeled by the tree below.



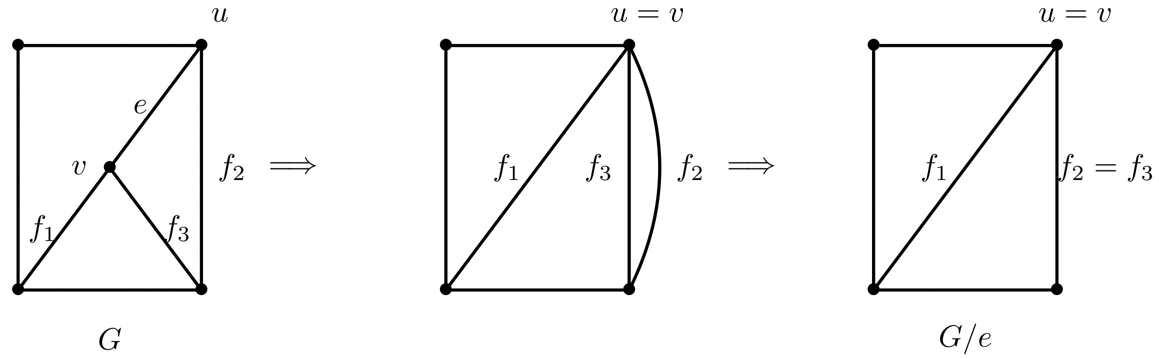
§2.5 Graph Coloring

Let C be a set of colors. A coloring of a graph $G = (V, E)$ is a function $f : V \longrightarrow C$. A coloring is proper in case $f(u) \neq f(v)$, whenever $u \sim v$. For the remainder of this chapter all colorings will be proper colorings. Clearly we take G to be simple.

Two important questions arise. First, what is the minimum number of colors required to color a given graph G. This number is denoted by $\chi(G)$, and is called the chromatic number of G. The second question is, if we are given a set of colors, C, of size m, how many ways can we color G using the colors from C? We denote the answer by $P(G, m)$. We realize that m is variable, so we call the function $P(G, x)$ the chromatic polynomial of G. To prove that this is always a polynomial we need several definitions, and a lemma.

Given a graph $G = (V, E)$, and an edge $e = \{u, v\} \in E$, the edge-deleted subgraph is $G - e = (V, E - \{e\})$. Meanwhile the contraction of G by e, denoted G/e, is the graph obtained from $G - e$ by identifying the endpoints u and v, and any resulting multiple edges identified to a single edge.

Example:



Fundamental Reduction Lemma *Let $G = (V, E)$ be a simple graph, and $e = \{u, v\} \in E$. Then $P(G - e, x) = P(G, x) + P(G/e, x)$*

Proof: Any proper coloring of $G - e$ either has $f(u) = f(v)$, in which case it gives a proper coloring of G/e, or $f(u) \neq f(v)$, in which case it gives a proper coloring of G. ■

Corollary (Fundamental Reduction Theorem) *If $G = (V, E)$ is a simple graph, and $e \in E$, then $P(G, x) = P(G - e, x) - P(G/e, x)$.*

The graph $G = (V, \emptyset)$, with $|V| = n$ is denoted by I_n. Clearly $P(I_n, x) = x^n$. This is the basis step for an induction proof of

Corollary *If $G = (V, E)$ is a simple graph, then $P(G, x)$ is a polynomial in x.*

Proof: We induct on $|E|$. The base case is above.

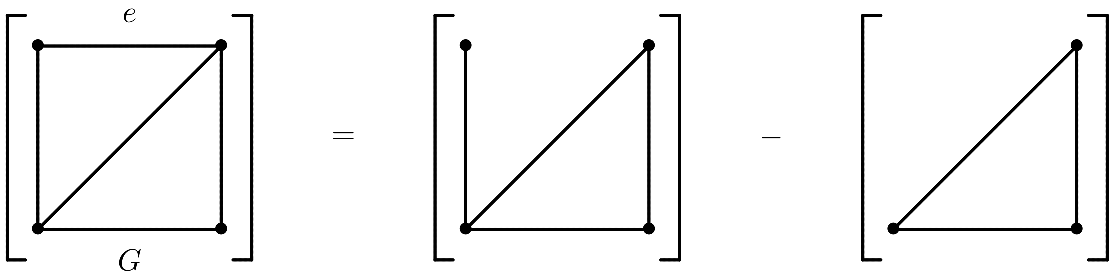
For the inductive step we suppose that the theorem holds for all graphs with fewer than k edges. We let G be a graph with $|E| = k$. Thus both $G - e$, and G/e have fewer than k edges. Therefore $P(G - e, x)$ and $P(G/e, x)$ are polynomials in x. Therefore, by the fundamental reduction theorem, $P(G, x)$ is a polynomial in x. ■

We state without proof

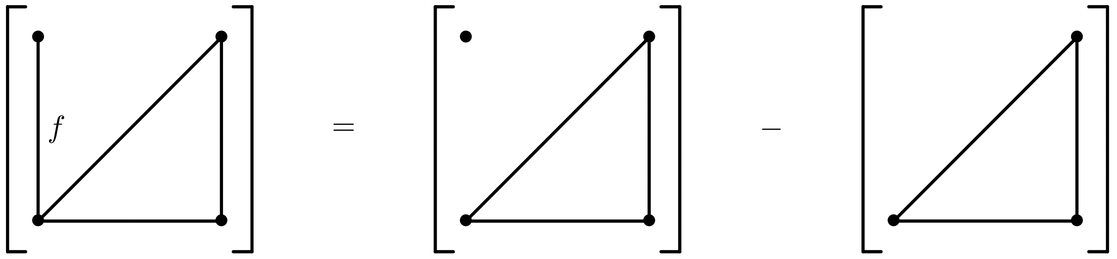
Theorem *If $G_1 \cap G_2 = \emptyset$, then $P(G_1 \cup G_2, x) = P(G_1, x) \cdot P(G_2, x)$.*

Before we proceed with an example, we observe that
$P(K_n, x) = x(x-1)(x-2)(....)(x-n+1)$, which we will denote by $x^{(n)}$. Also, in practice we will denote $P(G, x)$ by placing large square brackets around G.

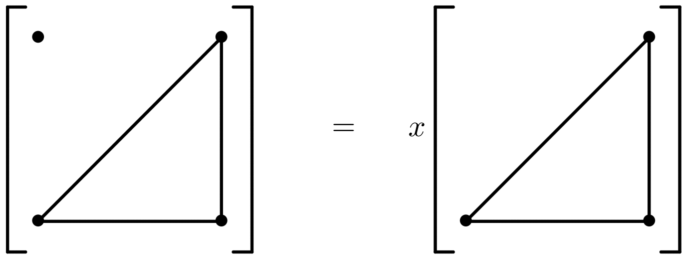
Example:



Now we apply the reduction theorem to $G - e$ to see



Since



We have that $P(G - e, x) = (x-1)P(K_3, x)$. Therefore

$$P(G, x) = (x-1)P(K_3, x) - P(K_3, x) = (x-2)P(K_3, x) = x(x-1)(x-2)^2.$$

Similar to the previous theorems of this section we have

Theorem: (Second Reduction Theorem) *If* G_1*, and* G_2 *are simple graphs with* $G_1 \cap G_2 = K_m$*,*

$$P(G_1 \cup G_2, x) = \frac{P(G_1, x) \cdot P(G_2, x)}{x^{(m)}}$$

By employing the reduction theorems, we have a fairly efficient procedure to compute $P(G, x)$.

This in turn allows us to compute $\chi(G)$. We observe that the value of $P(G, x)$ will be zero whenever x is a non-negative whole number strictly smaller than $\chi(G)$. So $\chi(G)$ is characterized as being the smallest non-negative integer for which $P(G, x) \neq 0$.

In addition, by the factor theorem from basic algebra, if $\chi(G) = k$ we can always write $P(G, x)$ in the form $x^{e_1}(x-1)^{e_2}(x-2)^{e_3}...(x-(k-1))^{e_k}g(x)$, where the exponents e_i are positive and $g(x)$ is a polynomial with no integral roots. Conversely, writing $P(G, x)$ in this form allows us to deduce $\chi(G) = k$.

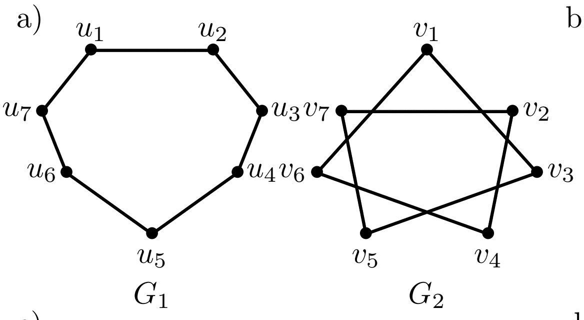
Warning: A common mistake occurs when someone finds a coloring of G using k colors and deduces that $\chi(G) = k$. The correct deduction is $\chi(G) \leq k$. To show equality we must either use the idea above, or perhaps the last theorem.

Theorem: *If G is a simple graph with an induced subgraph isomorphic to K_m, then $\chi(G) \geq m$.*

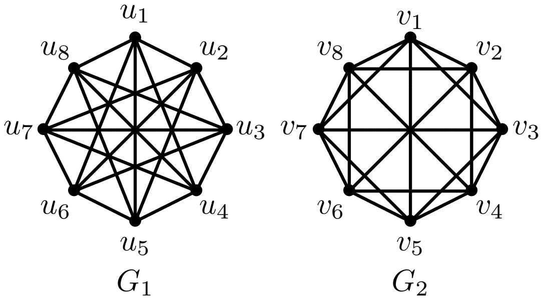
Chapter 2 Exercises

1. For each pair of graphs find a graph isomorphism $\varphi : G_1 \longrightarrow G_2$, and confirm φ is edge-preserving using adjacency matrices, or prove that $G_1 \not\cong G_2$.

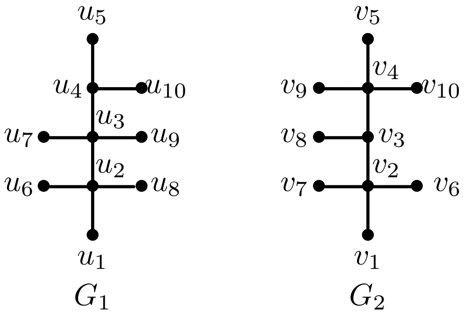
a)



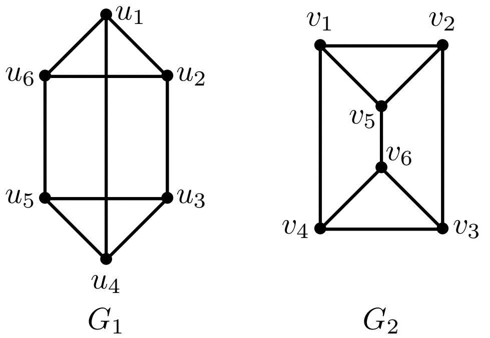
b)



c)



d)

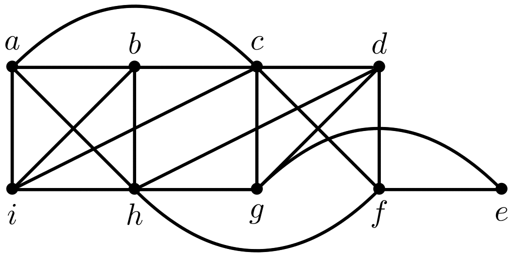


2. For which values of n is C_n bipartite? Q_n?

3. Prove the first theorem from section 2.3.

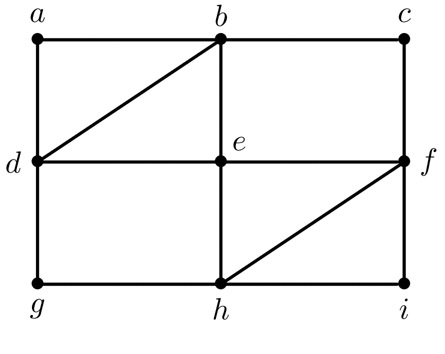
4. For each graph below i) find an Eulerian path, or prove that none exists, and ii) find a Hamiltonian cycle or prove that none exists.

a)



c) Q_3, the 3-cube

b)



d) the Petersen graph

5. Answer the following questions about the rooted tree.

a) Which vertex is the root?

b) Which vertices are internal?

c) Which vertices are leaves?

d) Which vertices are children of b?

e) Which vertices are grandchildren of b?

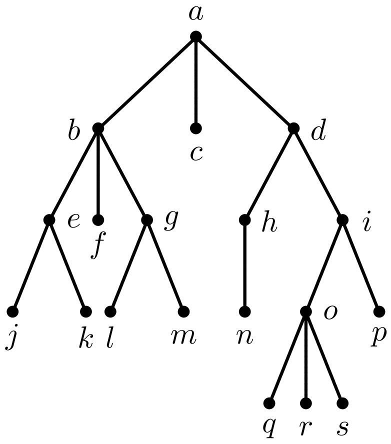
f) Which vertex is the parent of m?

g) Which vertices are siblings of q?

h) Which vertices are ancestors of p?

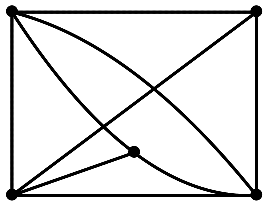
i) Which vertices are descendants of d?

j) What level is i at?

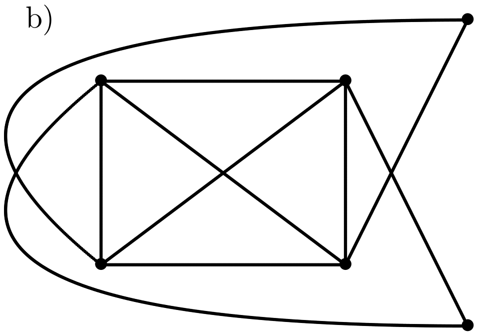


6. For each graph determine its chromatic number $\chi(G)$. Justify your answers.

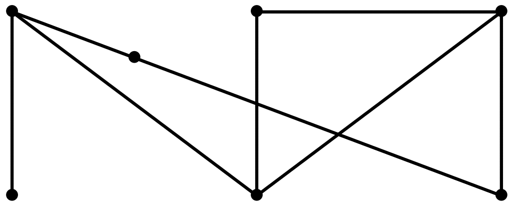
a)



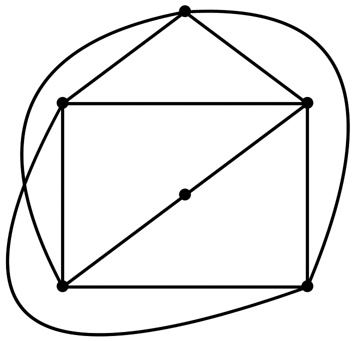
b)



c)

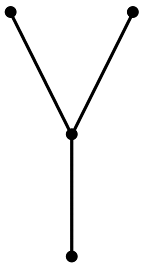


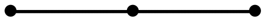
d)

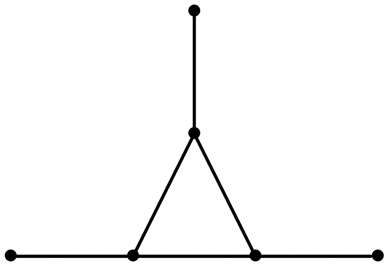


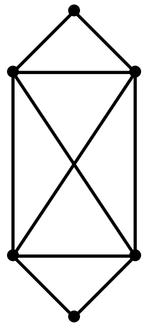
7. In assigning frequencies to mobile radio telephones, a zone gets a frequency to be used by all vehicles in the zone. Two zones that interfere (because of proximity or meteorological reasons) must get different frequencies. How many different frequencies are required if there are 6 zones, a, b, c, d, e, and f, where zone a interferes with zone b only; b interferes with a, c, and d; c with b, d, and e; d with b, c, and e; e with c, d, and f; and f with e only? Justify your answer.

8. Find the chromatic polynomial of each graph. Use the chromatic polynomial to find the number of ways of coloring the graph in at most 3 colors. repeat for at most 4 colors.


a)


b)

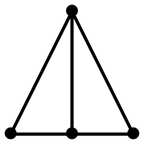

c)

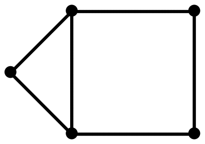

d)

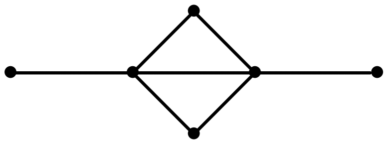
9. Let L_n be the graph consisting of a simple chain of n vertices. Find a formula for $P(L_n, x)$.

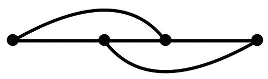
10. Let C_n denote the simple cycle on n vertices. Find a formula for $P(C_{2m}, x)$. Repeat for C_{2k+1}.

11. Use reduction theorems to compute the chromatic polynomials of each graph. Use the chromatic polynomial to compute the graph's chromatic number. Find a coloring using the minimal number of colors.


a)


b)


c)


d)

Chapter 3: Intermediate Counting

We saw at the end of chapter 1 that there are problems which can easily be posed, which do not admit solutions by the tactics of basic counting. In this chapter we develop further tactics, in part to rectify this apparent short-fall of basic counting.

§3.1 Generating Functions

If f is a smooth enough function near $x = 0$ it can be expanded (maybe via Taylor's Theorem) in terms of a Maclaurin Series. That is, in some neighborhood of $x = 0$, there is no difference between the function $f(x)$, and the values of the power series $\sum_{k=0}^{\infty} a_k x^k$. In fact Taylor's Theorem tells us that the coefficients $a_k = f^{(k)}(0)/k!$, where $f^{(k)}(x)$ is the kth derivative of $f(x)$.

If $f(x) = \sum_{k=0}^{\infty} a_k x^k$, for all values of x in some neighborhood of $x = 0$, we say that $f(x)$ is the (ordinary) generating function for the sequence of coefficients $\Big(a_k\Big)_{k=0}^{\infty}$.

Example: $f(x) = \dfrac{1}{1-x} = 1 + x + x^2 + x^3 + ... = \sum_{k=0}^{\infty} x^k$, for $|x| < 1$, so $f(x)$ generates the constant sequence $\Big(1, 1, 1, 1, 1, ...\Big) = \Big(1\Big)_{k=0}^{\infty}$.

Example: $f(x) = (1 + x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4$, for all $x \in \mathbb{R}$, so $f(x)$ generates the sequence $\Big(1, 4, 6, 4, 1, 0, 0, 0....\Big) = \left(\binom{4}{k}\right)_{k=0}^{\infty}$.

Our first challenge is to develop a set of tools which allow us to build a library of basic generating functions.

Theorem: *If $f(x)$ generates the sequence $\Big(a_n\Big)_{n=0}^{\infty}$, and $g(x)$ generates the sequence $\Big(b_n\Big)_{n=0}^{\infty}$ then*

1) $f(x) \pm g(x)$ generates the sequence $\Big(a_n \pm b_n\Big)_{n=0}^{\infty}$.

2) $f(x)g(x)$ generates the sequence $\Big(c_n\Big)_{n=0}^{\infty}$, where $c_k = \sum_{l=0}^{k} a_l b_{k-l}$.

3) $f'(x)$ generates the sequence $\Big(a_1, 2a_2, 3a_3, ...\Big) = \Big((n + 1)a + n + 1\Big)_{n=0}^{\infty}$.

4) The function $F(x)$, with $F(0) = 0$ and $F'(x) = f(x)$, generates $\left(0, a_0, \dfrac{a_1}{2}, \dfrac{a_2}{3}, ...\right)$.

Proof: See your favorite calculus II textbook.

Notice that we can also deduce that $xf'(x)$ generates $\Big(na_n\Big)_{n=0}^{\infty}$. Also if F is as in part 4, then $F(x)/x$ generates $\left(\dfrac{a_n}{n+1}\right)_{n=0}^{\infty}$.

Finally we remark that we may replace the symbol x with all sorts of things formally, and deduce new and fun-filled facts.

Example: Since $\dfrac{1}{1-x}$, generates $\left(1^n\right)_{n=0}^{\infty}$, $\dfrac{1}{1-2x}$ generates $\left(2^n\right)_{n=0}^{\infty}$.

Also $\dfrac{1}{1+x}$ generates $\left((-1)^n\right)_{n=0}^{\infty}$.

Example: From the previous example, $\displaystyle\int \frac{dx}{1+x} = \int \left[\sum_{n=0}^{\infty}(-1)^n x^n\right] dx$

Now technically we can only interchange the order of integration and summation if we have the right kind of convergence of our power series. Since we are only interested in formal manipulation of power series, we'll not worry about this subtlely here, nor henceforth. Thus

$$\ln(1+x) = \int \frac{dx}{1+x} = \sum_{n=0}^{\infty}\left[(-1)^n \int x^n\ dx\right] = \left[\sum_{n=0}^{\infty}\frac{(-1)^n x^{n+1}}{n+1}\right] + C$$

The value of C is found to be 0 by substituting in $x = 0$ and evaluating $\ln(1) = 0$. So $\ln(1+x)$ generates the sequence $\left(\dfrac{(-1)^n}{n+1}\right)_{n=0}^{\infty}$.

Example: Similar to the previous example we can start with $\displaystyle\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$. We differentiate both sides of the equation with respect to x, interchanging the order of differentiation and summation on the right-hand side. We arrive at $\displaystyle\frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} nx^{n-1} = \sum_{m=0}^{\infty}(m+1)x^m$. Thus $\displaystyle\frac{x}{(1-x)^2} = \sum_{m=0}^{\infty}(m+1)x^{m+1} = \sum_{k=0}^{\infty} kx^k$ generates the sequence $\left(k\right)_{k=0}^{\infty}$.

§3.2 Applications to Counting

Consider the generating function $(1+x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4 + 0x^5 + 0x^6 + ...$, which generates the sequence $\left(\binom{4}{k}\right)_{k=0}^{\infty}$, whose terms are the number of k-subsets of a 4-set. This function is the result of evaluating the expression $(1+ax)(1+bx)(1+cx)(1+dx)$ at $a = b = c = d = 1$. The expansion of this function gives

$$1 + (a+b+c+d)x + (ab+ac+ad+bc+bd+cd)x^2 + (abc+abd+acd+bcd)x^3 + (abcd)x^4.$$

The coefficients of x^i in this expansion clearly describe all of the subsets of $\{a, b, c, d\}$ of size i.

More generally it occurs to us that the coefficient of x^k in the expansion of $(1+x)^n$ is the number of k-subsets of an n-set. If we write the expansion of $(1+a_1x)(1+a_2x)(1+a_3x)...(1+a_nx)$, then the coefficient of x^k is an ordered list describing all k-subsets of $\{a_1, a_2, ..., a_n\}$. In fact $(1+x)^n$ results if we set $a_i = 1$, for $i = 1$ to n, in the expression $(1+a_1x)(1+a_2x)(1+a_3x)...(1+a_nx)$.

Now what about $(1 + ax + a^2x^2 + a^3x^3)(1 + bx + b^2x^2)(1 + cx)$? Upon expansion we find this is $1 + (a+b+c)x + (a^2 + ab + ac + b^2 + bc)x^2 + (a^3 + a^2b + a^2c + ab^2 + abc +$

$b^2c)x^3 + (a^3b + a^c + a^2b^2)x^4 + (a^3b^2 + a^3bc)x^5 + (a^3b^2c)x^6$. Setting $a = b = c = 1$ we have $(1 + x + x^2 + x^3)(1 + x + x^2)(1 + x) = 1 + 3x + 5x^2 + 6x^3 + 3x^4 + 2x^5 + x^6$. So what is the combinatorial significance of this sequence?

After a little thought, and considering the coefficients in the first expansion, we realize this counts the number of solutions in non-negative integers to $y_1+y_2+y_3 = i$, where $y_1 \leq 3, y_2 \leq 2$, and $y_3 \leq 1$. Which is to say that the first expression generates all multisets of size i from $\{a, b, c\}$, using at most 3 a's, at most 2 b's, and at most one c.

In general we wish to compute the number of integral solutions to $y_1 + y_2 + ... + y_k = r$, where $a_i \leq y_i \leq b_i, i = 1, ...k$, and the a_i's and b_i's are integral lower and upper bounds. To do this we can now use a generating function approach. We simply compute the coefficient of x^r in the expansion of $\prod_{i=1}^{k}(x^{a_i} + x^{a_i+1} + x^{a_i+2} + ... + x^{b_i}) = \prod_{i=1}^{k}[\sum_{j=a_i}^{b_i} x^j]$. Probably we get a computer algebra system to do this for us. Again we would use a computer algebra system if we used place-holders d_i, $i = 1, ..., k$ to generate the actual solutions, i.e. expanded $\prod_{i=1}^{k}[\sum_{j=a_i}^{b_i}(d_ix)^j]$.

Naturally we could re-index the problem to count the number of solutions in non-negative integers to $y_1 + ... + y_k = s$, where $y_i \leq c_i = b_i - a_i$, and $s = r - a_1 - a_2 - ... - a_k$. Now we need the coefficient of x^s in the expansion of $\prod_{i=1}^{k}(1 + x + x^2 + ... + x^{c_i}) = \prod_{i=1}^{k}[\sum_{j=0}^{c_i} x^j]$.

And, as might happen in an ideal world, we might have $s \leq c_i$ for all i, so that there are effectively no upper bounds. Here we want to realize that the coefficient of x^s in the expansion of $\prod_{i=1}^{k}(1 + x + x^2 + ... + x^{c_i}) = \prod_{i=1}^{k}[\sum_{j=0}^{c_i} x^j]$, is the same as the coefficient of x^s in the expansion of $\prod_{i=1}^{k}(1 + x + x^2 + ...) = \prod_{i=1}^{k}[\sum_{j=0}^{\infty} x^j] = \prod_{i=1}^{k}\frac{1}{1-x} = (1 - x)^{-k}$. So apparently $(1 - x)^{-k}$ generates the sequence $\left(\binom{s+k-1}{s}\right)_{s=0}^{\infty}$??!!

Amazingly this little gem of information can also be derived from the general version of the Binomial Theorem discovered by Isaac Newton.

Theorem(Newton) *For any real number* $u \neq 0, (1 + x)^u = \sum_{k=0}^{\infty}\binom{u}{k}x^k$, *where* $\binom{u}{k} = u(u - 1)(u - 2)...(u - k + 1)/k!$.

When we set $u = -p$, where p is a positive integer we find that

$$\begin{aligned}\binom{-p}{k} &= \frac{(-p)(-p-1)(-p-2)...(-p-k+1)}{k!}\\ &= (-1)^k\frac{(p+k-1)(p+k-2)...(p+1)p}{k!}\\ &= (-1)^k\binom{p+k-1}{k}\end{aligned}$$

So replacing x by $-x$ in the theorem we arrive at

$$\begin{aligned}(1-x)^{-p} &= \sum_{k=0}^{\infty}\binom{-p}{k}(-x)^k \\ &= \sum_{k=0}^{\infty}(-1)^k\binom{p+k-1}{k}(-1)^k x^k \\ &= \sum_{k=0}^{\infty}\binom{p+k-1}{k}x^k\end{aligned}$$

since $(-1)^k(-1)^k = (-1)^{2k} = 1$. Thus $(1-x)^{-p}$ generates the number of solutions to the donut shoppe problem.

Before we leave this topic we remark that there are other directions in which we might generalize. For example, the number of ways to make n cents change using pennies, nickels, dimes and quarters is the number of solutions in non-negative integers to

$$y_1 + 5y_2 + 10y_3 + 25y_4 = n.$$

This is also the coefficient of x^n in the expansion of $(1-x)^{-1}(1-x^5)^{-1}(1-x^{10})^{-1}(1-x^{25})^{-1}$. So we can use generating functions to compute the number of, and enumerate the solutions to a great variety of linear diophantine equations.

§3.3 Exponential Generating Functions

As we saw in chapter 1, Stirling Numbers of the second kind are important for counting solutions to occupancy problems. In order to derive the formula given for $S(n,k)$ in chapter 1, we will use exponential generating functions.

If $f(x) = \sum_{k=0}^{\infty}\frac{a_k}{k!}x^k$, for all values of x in some neighborhood of $x=0$, we say that $f(x)$ is the exponential generating function for the sequence of coefficients $\left(a_k\right)_{k=0}^{\infty}$.

So, for example, $f(x) = e^x$ is the exponential generating function for $\left(1\right)_{k=0}^{\infty}$. And in general $g(x) = e^{\alpha x}$ is the exponential generating function for $\left(\alpha^k\right)_{k=0}^{\infty}$.

Next we recall that combinations and permutations are related by $P(n,k) = k!C(n,k)$, or $P(n,k)/k! = C(n,k)$. Similar formulas apply when some repetition is allowed, ala the MISSISSIPPI problem. So we consider the expansion of

$$(1+\frac{a}{1!}x+\frac{a^2}{2!}x^2+\frac{a^3}{3!})((1+\frac{b}{1!}x+\frac{b^2}{2!}x^2)((1+\frac{c}{1!}x)$$

which comes out to

$$1+(\frac{a}{1!}+\frac{b}{1!}+\frac{c}{1!})x+(\frac{a^2}{2!}+\frac{ab}{1!1!}+\frac{ac}{1!1!}+\frac{b^2}{2!}+\frac{bc}{1!1!})x^2+$$
$$(\frac{a^3}{3!}+\frac{a^2b}{2!1!}+\frac{a^2c}{2!1!}+\frac{ab^2}{1!2!}+\frac{abc}{1!1!1!}+\frac{b^2c}{2!1!})x^3+(\frac{a^3b}{3!1!}+\frac{a^3c}{3!1!}+$$
$$\frac{a^2b^2}{2!2!}+\frac{a^2bc}{2!1!1!}+\frac{ab^2c}{1!2!1!})x^4+(\frac{a^3b^2}{3!2!}+\frac{a^3bc}{3!1!1!}+\frac{a^2b^2c}{2!2!1!})x^5+(\frac{a^3b^2c}{3!2!1!})x^6$$

Next we multiply and divide the coefficient of x^k by $k!$ to get

$$1+1!(\frac{a}{1!}+\frac{b}{1!}+\frac{c}{1!})\frac{x}{1!}+2!(\frac{a^2}{2!}+\frac{ab}{1!1!}+\frac{ac}{1!1!}+\frac{b^2}{2!}+\frac{bc}{1!1!})\frac{x^2}{2!}+$$
$$3!(\frac{a^3}{3!}+\frac{a^2b}{2!1!}+\frac{a^2c}{2!1!}+\frac{ab^2}{1!2!}+\frac{abc}{1!1!1!}+\frac{b^2c}{2!1!})\frac{x^3}{3!}+4!(\frac{a^3b}{3!1!}+\frac{a^3c}{3!1!}+$$
$$\frac{a^2b^2}{2!2!}+\frac{a^2bc}{2!1!1!}+\frac{ab^2c}{1!2!1!})\frac{x^4}{4!}+5!(\frac{a^3b^2}{3!2!}+\frac{a^3bc}{3!1!1!}+\frac{a^2b^2c}{2!2!1!})\frac{x^5}{5!}+6!(\frac{a^3b^2c}{3!2!1!})\frac{x^6}{6!}$$

Now, for example, the coefficient on a "sub-multi-set" term like a^2b^2c, is the number of 5-strings over $\{a,b,c\}$ using 2 a's, 2 b's and 1 c. So setting $a=b=c=1$, we can generate the number of permutations with partial replacement over a given alphabet.

Theorem: *Given r types of objects, with e_i indistinguishable objects of type i, $i=1,2,...,r$, the number of distinguishable permutations of length k using up to e_i objects of type i is the coefficient of $x^k/k!$ in the exponential generating function*

$$\left(1+\frac{x}{1!}+\frac{x^2}{2!}+...+\frac{x^{e_1}}{e_1!}\right)\left(1+\frac{x}{1!}+...+\frac{x^{e_2}}{e_2!}\right)....\left(1+\frac{x}{1!}+...+\frac{x^{e_r}}{e_r!}\right)$$

As a final application of exponential generating functions we derive the formula $S(n,k)=\frac{1}{k!}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^n$.

We find $T(n,k)=k!S(n,k)=\#$ onto functions from an n-set to a k-set $=$ the number of n-permutations of a k-set using each set element at least once. So $T(n,k)$ is the coefficient of $x^k/k!$ in the expansion of

$$\left(x+\frac{x^2}{2!}+\frac{x^3}{3!}+....\right)^k=(e^x-1)^k.$$

By the binomial theorem

$$\begin{aligned}(e^x-1)^k&=\sum_{i=0}^{k}\binom{k}{i}(-1)^ie^{(k-i)x}\\&=\sum_{i=0}^{k}\binom{k}{i}(-1)^i\sum_{n=0}^{\infty}(k-i)^n\frac{x^n}{n!}\\&=\sum_{n=0}^{\infty}\frac{x^n}{n!}\sum_{i=0}^{k}\binom{k}{i}(-1)^i(k-i)^n\end{aligned}$$

where the second equation is from the Maclaurin series for e^x, and the order of summation can be reversed since the Maclaurin series for e^x converges absolutely on the entire real line.

So we have $T(n,k) = \sum_{i=0}^{k} \binom{k}{i}(-1)^i(k-i)^n$, and thus $S(n,k)$ is as stated.

§3.4 Recurrence Relations

Given a sequence, a, with domain $D = \{n \in \mathbb{Z} | n \geq m\}$ and codomain $\mathbb{R}$, a recurrence relation is a formula which for all $n \in \{l \in \mathbb{Z} | l \geq k\}$ (where $k \geq m$) relates a_n, in some manner, to a finite number of preceding terms of the sequence and possibly a function of n.

We will almost exclusively be interested in recurrence relations which take the form

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + ... + c_k a_{n-k} + f(n)$$

where $c_1, c_2, ..., c_k \in \mathbb{R}$, and $c_k \neq 0$. Such a recurrence relation is a linear recurrence relation with constant coefficients. When the function $f(n) = 0$, we call the recurrence relation homogeneous. If $f(n)$ is not identically zero, the recurrence relation is non-homogeneous. The number k is called the degree of the relation.

WARNING: A formula which recursively defines a sequence does not completely determine the sequence. Observe, for example, that geometric sequences with common ratio r all satisfy $a_n = r \cdot a_{n-1}$, for $n \geq 1$. What picks out a single sequence in this case is its initial term. In general we may need to know several initial terms, which are called the initial conditions of the sequence. Given a sufficient number of initial conditions, and a recurrence relation, we get exactly one sequence of real numbers.

Example: Suppose that we deposit A_0 dollars in an account drawing t percent interest per annum compounded yearly, and that no withdrawals occur. Then if A_n denotes the money in the account after n years, we have $A_n = (1 + \frac{t}{100})A_{n-1}$, when $n \geq 1$.

In this simplest of all cases, it is clear that the initial condition is of paramount importance. It's also true that we can find an explicit formula for A_n in this case since the sequence is geometric. In short $A_n = r^n \cdot A_0$, where $r = (1 + \frac{t}{100})$. We call this process solving the recurrence relation. In this example we solved it by inspection. The general case is more difficult and is the topic of the next section.

Example: A canonical example relates the story of the Towers of Hanoi. A group of monks wished a magical tower to be constructed from 1000 stone rings. The rings were to be of 1000 different sizes. The size and composition of the rings was to be designed so that any ring could support the entire weight of all of the rings smaller than itself, but each ring would be crushed beneath the weight of any larger ring.

The monks hired the lowest bidder to construct the tower in a clearing in the dense jungle nearby. Upon completion of construction the engineers brought the monks to see their work. The monks admired the exquisite workmanship, but informed the engineers that the tower was not in the proper clearing.

In the jungle there were only three permanent clearings. The monks had labelled them A, B and C. The engineers had labelled them in reverse order. The monks instructed the engineers to move the tower from clearing A to clearing C!

Because of the massive size of the rings, the engineers could only move one per day. No ring could be left anywhere in the jungle except one of A, B, or C. Finally each clearing was only large enough so that rings could be stored there by stacking them one on top of another.

The monks then asked the engineers how long it would take for them to fix the problem.

Before they all flipped a gasket, the most mathematically talented engineer came upon the following solution.

Let H_n denote the minimum number of days required to move an n ring tower from A to C under the constraints given. Then $H_1 = 1$, and in general an n ring tower can be moved from A to C by first moving the top $(n-1)$ rings from A to B leaving the bottom ring at A, then moving the bottom ring from A to C, and then moving the top $(n-1)$ rings from clearing B to clearing C. So $H_n = 2 \cdot H_{n-1} + 1$, for $n \geq 2$.

By unwinding this sequence similar to the one at the end of section 2.4 we get

$$\begin{aligned}
H_n &= 2 \cdot H_{n-1} + 1 \\
&= 2 \cdot [2 \cdot H_{n-2} + 1] + 1 = 2^2 \cdot H_{n-2} + 2 + 1 \\
&= 2^2 \cdot [2 \cdot H_{n-3} + 1] + 2 + 1 = 2^3 \cdot H_{n-3} + 2^2 + 2 + 1 \\
&\vdots
\end{aligned}$$

At the kth iteration

$$\begin{aligned}
H_n &= 2^k H_{n-k} + 2^{k-1} + 2^{k-2} + ... + 2^3 + 2^2 + 2 + 1 \\
&\vdots \\
&= 2^{n-1} H_{n-(n-1)} + 2^{n-2} + ... + 2^2 + 2 + 1 \\
&= 2^{n-1} H_1 + 2^{n-2} + ... + 2^2 + 2 + 1 \\
&= \sum_{l=0}^{n-1} 2^l, \text{ since } H_1 = 1 \\
&= 2^n - 1, \text{ by the geometric sum formula}
\end{aligned}$$

So the problem would be fixed in $2^{1000} - 1$ days, or approximately 2.93564×10^{296} centuries. Hence the term job security!

Example: A more realistic example might be to count the number of binary strings of length $n \geq 0$ which contain a pair of consecutive 0's.

Here we let b_n denote the binary strings which have length n and contain a pair of consecutive zeroes. We can find $b_0 = b_1 = 0$ by inspection, as well as $b_2 = 1$.

To find b_3 we write down all bit strings of length 3 and cross out those which do not have a pair of consecutive zeroes. What's left is $000, 001$ and 100. So $b_3 = 3$.

Similarly $b_4 = 8$ since the strings of length 4 which contain a pair of consecutive zeroes are $0000, 0001, 0010, 0011, 0100, 1000, 1001$, and 1100. We might continue this time-consuming and tedious process hoping to discover a pattern by sheer luck, or we can attempt to use a combinatorial approach.

Let S_k denote the bit strings of length k which contain a pair of consecutive zeroes. Notice that $|S_k| = b_k$.

If $x \in S_n$, then either $x = 1y$, where $y \in S_{n-1}$, or $x = 0w$, where w is a bit string of length $n-1$. In case $x = 0w$, either $x = 01z$, where $z \in S_{n-2}$, or $w = 00v$, where v is any binary string of length $n-2$. Since these sub-cases are exclusive, by the addition principle

$$b_n = b_{n-1} + b_{n-2} + 2^{n-2}, \text{ for } n \geq 2.$$

Together with the initial conditions $b_0 = b_1 = 0$, this completely determines the sequence. A good check is that the terms we generated actually satisfy this relation. Of course we probably would not want to use the recursive definition to find b_{128}. This motivates the next section.

§3.5 The Method of Characteristic Roots

In the last section we noted that any homogeneous linear recurrence relation of degree 1 with constant coefficient corresponds to a geometric sequence. These can therefore be solved by inspection. Also from the Towers of Hanoi story one might guess correctly that most non-homogeneous linear recurrence relations with constant coefficients can be solved by unwinding. Neither of these methods is powerful enough in general. So in this section we introduce a basic method for solving homogeneous linear recurrence relations with constant coefficients.

We begin by considering the case $k = 2$. So we have a recurrence relation of the form $a_n = c_1 a_{n-1} + c_2 a_{n-2}$, for $n \geq m$, where c_1 and c_2 are real constants. We must also have two initial conditions a_{m-1} and a_{m-2} to get started at $n = m$. We will dispense with the general case here and suppose that $m = 2$. That is, we are given a_0 and a_1 and the formula $a_n = c_1 a_{n-1} + c_2 a_{n-2}$, for $n \geq 2$. Notice that $c_2 \neq 0$ or else we have a linear recurrence relation with constant coefficients and degree 1. What we seek is a closed form expression for a_n, which is a function of n alone, and which is therefore independent of the previous terms of the sequence.

Of fundamental importance to this method is the characteristic polynomial, denoted $\chi(x)$ of the recurrence relation. For the case above we define $\chi(x) = x^2 - c_1 x - c_2$. Notice that the degree of $\chi(x)$ coincides with the degree of the recurrence relation. Notice also that the non-leading coefficients of $\chi(x)$ are simply the negatives of the coefficients of the recurrence relation. This allows us to generalize the definition so that the characteristic polynomial of $a_n = c_1 a_{n-1} + ... + c_k a_{n-k}$ is $\chi(x) = x^k - c_1 x^{k-1} - ... - c_{k-1} x - c_k$. A number r (possibly complex) is a characteristic root if $\chi(r) = 0$. From basic algebra we know that r is a root of a polynomial iff $(x - r)$ is a factor of the polynomial. When $\chi(x)$ is a degree 2 polynomial by the quadratic formula, either $\chi(x) = (x - r_1)(x - r_2)$, where $r_1 \neq r_2$, or $\chi(x) = (x - r)^2$, for some r.

Theorem: *Let c_1 and c_2 be real numbers. Suppose that the polynomial $x^2 - c_1 x - c_2$ has two distinct roots r_1 and r_2. Then a sequence $a : \mathbb{N} \longrightarrow \mathbb{R}$ is a solution of the recurrence relation $a_n = c_1 a_{n-1} + c_2 a_{n-2}$, for $n \geq 2$ iff $a_m = \alpha r_1^m + \beta r_2^m$, for all $m \in \mathbb{N}$, for some constants α and β.*

Proof: If $a_m = \alpha r_1^m + \beta r_2^m$ for all $m \in \mathbb{N}$, where α and β are some constants, then since $r_i^2 - c_1 r_i - c_2 = 0, i = 1, 2$, we have $r_i^2 = c_1 r_i + c_2$, $i = 1, 2$. So for $n \geq 2$

$$\begin{aligned} c_1 a_{n-1} + c_2 a_{n-2} &= c_1(\alpha r_1^{n-1} + \beta r_2^{n-1}) + c_2(\alpha r_1^{n-2} + \beta r_2^{n-2}) \\ &= \alpha r_1^{n-2}(c_1 r_1 + c_2) + \beta r_2^{n-2}(c_1 r_2 + c_2), \text{ distributing and combining} \\ &= \alpha r_1^{n-2} \cdot r_1^2 + \beta r_2^{n-2} \cdot r_2^2, \text{ by the remark above} \\ &= \alpha r_1^n + \beta r_2^n = a_n \end{aligned}$$

Conversely, if a is a solution of the recurrence relation and has initial terms a_0 and a_1, then one checks that the sequence $a_m = \alpha r_1^m + \beta r_2^m$ with

$$\alpha = \frac{a_1 - a_0 \cdot r_2}{r_1 - r_2}, \text{ and } \beta = \frac{a_0 r_1 - a_1}{r_1 - r_2}$$

also satisfies the relation and has the same initial conditions. The equations for α and β come from solving the system of linear equations

$$\begin{aligned} a_0 &= \alpha r_1^0 + \beta r_2^0 = \alpha + \beta \\ a_1 &= \alpha r_1^1 + \beta r_2^1 = \alpha r_1 + \beta r_2 \end{aligned}$$

Thus we will want in general to be able to solve systems of linear equations. ■

Example: Solve the recurrence relation $a_0 = 2, a_1 = 3$ and $a_n = a_{n-2}$, for $n \geq 2$.

Solution: The recurrence relation is a linear homogeneous recurrence relation of degree 2 with constant coefficients $c_1 = 0$ and $c_2 = 1$. The characterisitic polynomial is

$$\chi(x) = x^2 - 0 \cdot x - 1 = x^2 - 1.$$

The characteristic polynomial has two distinct roots since $x^2 - 1 = (x-1)(x+1)$. So say $r_1 = 1$ and $r_2 = -1$. Then

$$\begin{aligned} 2 = a_0 &= \alpha 1^0 + \beta(-1)^0 = \alpha + \beta \\ 3 = a_1 &= \alpha 1^1 + \beta(-1)^1 = \alpha + \beta(-1) = \alpha - \beta \end{aligned}$$

Adding the two equations eliminates β and gives $5 = 2\alpha$, so $\alpha = 5/2$. Substituting this into the first equation, $2 = 5/2 + \beta$, we see that $\beta = -1/2$. So $a_n = \frac{5}{2} \cdot 1^n + \frac{-1}{2}(-1)^n = \frac{5}{2} - \frac{1}{2} \cdot (-1)^n$.

Example: Solve the recurrence relation $a_1 = 3$, $a_2 = 5$, and $a_n = 5a_{n-1} - 6a_{n-2}$, for $n \geq 3$.

Solution: Here the characteristic polynomial is $\chi(x) = x^2 - 5x + 6 = (x-2)(x-3)$. So we suppose that $a_m = \alpha 2^m + \beta 3^m$, for all $m \geq 1$. The initial conditions give rise to the system of equations

$$\begin{aligned} 3 = a_1 &= \alpha 2^1 + \beta 3^1 = 2\alpha + 3\beta \\ 5 = a_2 &= \alpha 2^2 + \beta 3^2 = 4\alpha + 9\beta \end{aligned}$$

If we multiply the top equation through by 2 we get

$$\begin{aligned} 6 &= 4\alpha + 6\beta \\ 5 &= 4\alpha + 9\beta \end{aligned}$$

Subtracting the second equation from the first eliminates α and gives $1 = -3\beta$. So $\beta = -1/3$. Substitution into the first equation yields $3 = 2\alpha + 3 \cdot (-1/3)$, so $\alpha = 2$. Thus

$$a_m = 2 \cdot 2^m - \frac{1}{3} \cdot 3^m = 2^{m+1} - 3^{m-1}, \text{ for all } m \geq 1.$$

The other case we mentioned had a characteristic polynomial of degree two with one repeated root. Since the proof is similar we simply state

Theorem: *Let c_1 and c_2 be real numbers with $c_2 \neq 0$ and suppose that the polynomial $x^2 - c_1x - c_2$ has a root r with multiplicity 2, so that $x^2 - c_1x - c_2 = (x-r)^2$. Then a sequence $a : \mathbb{N} \longrightarrow \mathbb{R}$ is a solution of the recurrence relation $a_n = c_1a_{n-1} + c_2a_{n-2}$, for $n \geq 2$ iff $a_m = (\alpha + \beta m)r^m$, for all $m \in \mathbb{N}$, for some constants α and β.*

Example: Solve the recurrence relation $a_0 = -1, a_1 = 4$ and $a_n = 4a_{n-1} - 4a_{n-2}$, for $n \geq 2$.

Solution: In this case we have $\chi(x) = x^2 - 4x + 4 = (x-2)^2$. So we suppose that $a_m = (\alpha + \beta m)2^m$ for all $m \in \mathbb{N}$. The initial conditions give rise to the system of equations

$$\begin{aligned} -1 = a_0 &= (\alpha + \beta \cdot 0)2^0 = (\alpha) \cdot 1 = \alpha \\ 4 = a_1 &= (\alpha + \beta \cdot 1)2^1 = 2(\alpha + \beta) \cdot 2 \end{aligned}$$

Substituting $\alpha = -1$ into the second equation gives $4 = 2(\beta - 1)$, so $2 = \beta - 1$ and $\beta = 3$. Therefore $a_m = (3m-1)2^m$ for all $m \in \mathbb{N}$.

Finally we state without proof the theorem which governs the general method of characteristic roots.

Theorem: *Let $c_1, c_2, ..., c_k \in \mathbb{R}$ with $c_k \neq 0$. Suppose that*

$$\chi(x) = x^k - c_1x^{k-1} - c_2x^{k-2} - ... - c_{k-1}x - c_k = (x-r_1)^{j_1}(x-r_2)^{j_2} \cdot ... \cdot (x-r_s)^{j_s}$$

where $r_1, r_2, ..., r_s$ are distinct roots of $\chi(x)$, and $j_1, j_2, ..., j_s$ are positive integers so that $j_1 + j_2 + j_3 + ... + j_s = k$. Then a sequence $a : \mathbb{N} \longrightarrow \mathbb{R}$ is a solution of the recurrence relation $a_n = c_1a_{n-1} + c_2a_{n-2} + ... + c_ka_{n-k}$, for $n \geq k$ iff $a_m = p_1(m)r_1^m + p_2(m)r_2^m + ... + p_s(m)r_s^m$ for all $m \in \mathbb{N}$, where $p_i(m) = \alpha_{0,i} + \alpha_{1,i}m + \alpha_{2,i}m^2 + ... + \alpha_{j_i-1,i}m^{j_i-1}$, $1 \leq i \leq s$ and the $\alpha_{l,i}$'s are constants.

The problem with the general case is that given the recurrence relation we can simply write down the characteristic polynomial. However it can be quite a challenge to factor it as per the theorem. Even if we succeed in factoring it we are faced with the tedious task of setting up and solving a system of k linear equations in k unknowns (the $\alpha_{l,i}$'s). The basic methods of elimination, substitution, or graphing which are covered in a prerequisite course will often not be up to the task. This motivates better notation and more advanced methods for solving systems of equations. This has been a paid advertisement for a course in linear algebra.

Perhaps more to the point is the fact that recurrence relations are discrete versions of differential equations. So as the final examples indicate, we can have systems of recurrence relations, or even analogues of partial differential equations. The method of characteristic roots will not necessarily apply to these. We will therefore be motivated for the ultimate section of this chapter.

Example: An example of a discrete partial differential equation is given by the two variable function $A(m,n)$ which is called Ackerman's Function. The function is recursively defined via

$$A(m,n) = \begin{cases} n+1, & \text{if } m = 0 \\ A(m-1,1), & \text{if } m > 0 \text{ and } n = 0 \\ A(m-1, A(m,n-1)), & \text{if } m,n > 0 \end{cases}$$

Example: Another example, as one can show by combinatorial argument, is that the Stirling Numbers of the second kind satisfy

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

Example: Pascal's identity for binomial coefficients is a discrete partial differential equation which has an analytic solution.

Example: Let a_n be the number of ternary strings of length n which have an even number of 0's and an odd number of 1's.

We can find a system of recurrence relations which the terms of the sequence (a_n) satisfies:
For each natural number n, let $\mathcal{U}_n = \{0,1,2\}^n$,
$A_n = \{z \in \mathcal{U}_n | z$ has an even number of 0's, and an odd number of 1's$\}$,
$B_n = \{z \in \mathcal{U}_n | z$ has an even number of 0's, and an even number of 1's$\}$,
$C_n = \{z \in \mathcal{U}_n | z$ has an odd number of 0's, and an odd number of 1's$\}$, and
$D_n = \{z \in \mathcal{U}_n | z$ has an odd number of 0's, and an even number of 1's$\}$.
Finally let $b_n = |B_n|, c_n = |C_n|$, and $d_n = |D_n|$.
1) By the addition principle $a_n + b_n + c_n + d_n = 3^n$, for $n \geq 0$.
2) If $w \in A_{n+1}$, then either $w = 2x$, for some $x \in A_n$, or $w = 1y$, for some $y \in B_n$, or $w = 0z$, for some $z \in C_n$. So again by the addition principle $a_{n+1} = a_n + b_n + c_n$, for $n \geq 0$.
3) Similarly we have $b_{n+1} = a_n + b_n + d_n$, for $n \geq 0$.
4) Finally $c_{n+1} = a_n + c_n + d_n$, for $n \geq 0$.

Now we have four sequences which satisfy the four recurrence relations

$$\begin{aligned} 3^n &= a_n + b_n + c_n + d_n \\ a_{n+1} &= a_n + b_n + c_n \\ b_{n+1} &= a_n + b_n + d_n \\ c_{n+1} &= a_n + c_n + d_n \end{aligned}$$

The initial conditions are $a_0 = c_0 = d_0 = 0$, and $b_0 = 1$, since the empty string λ contains zero 1's and zero 0's

§3.6 Solving Recurrence Relations by the Method of Generating Functions

We begin with an example.

Example: Suppose that we are given the generating function $G(x) = \dfrac{2 - x + x^2}{1 - 2x - x^2 + 2x^3}$. Now let's find a closed formula for the sequence it generates.

This involves Partial Fraction Expansion, another subject you're supposed to know from Calculus II. In this case we factor the denominator. $G(x) = \dfrac{2 - x + x^2}{(1+x)(1-x)(1-2x)}$. So the rules for partial fraction expansion tell us that

$$G(x) = \frac{2 - x + x^2}{(1+x)(1-x)(1-2x)} = \frac{A}{1+x} + \frac{B}{1-x} + \frac{C}{1-2x}, \text{ for all } x \in \mathbb{R} - \{-1, 1, \frac{1}{2}\}.$$

Now multiply both sides by $(1+x)(1-x)(1-2x)$ to clear denominators. Thus

$$2-x+x^2 = A(1-x)(1-2x)+B(1+x)(1-2x)+C(1+x)(1-x), \text{ for all } x \in \mathbb{R}-\{-1,1,\frac{1}{2}\}.$$

Since polynomials are continuous we deduce that

$$2-x+x^2 = A(1-x)(1-2x)+B(1+x)(1-2x)+C(1+x)(1-x), \text{ for all } x \in \mathbb{R}.$$

In particular we can choose three values of x to generate 3 equations in the three unknowns A, B and C in order to solve for them. Of course some values of x give rise to easier systems of equations, especially the roots of the denominator of $G(x)$. For example, if $x = 1$ we have $2-x+x^2 = 2-1+1 = 2$ on the left-hand side of our equation, while the right-hand side is $A(1-1)(1-2)+B(1+1)(1-2)+C(1+1)(1-1) = A\cdot 0\cdot(-1)+B\cdot 2\cdot(-1)+C\cdot 2\cdot 0 = -2B$. So $B = -1$. Similarly evaluating our expression at $x = -1$ gives $A = 2/3$, and evaluating at $x = 1/2$ gives $C = 7/3$. So

$$G(x) = \frac{2-x+x^2}{(1+x)(1-x)(1-2x)} = \frac{\frac{2}{3}}{1+x} - \frac{1}{1-x} + \frac{\frac{7}{3}}{1-2x}, \text{ for all } x \in \mathbb{R}-\{-1,1,\frac{1}{2}\}.$$

By the methods of section 3.1 we can now easily write a Maclaurin series for $G(x)$,

$$G(x) = \sum_{k=0}^{\infty}\left(\frac{2}{3}(-1)^k - 1\cdot 1^k + \frac{7}{3}2^k\right)x^k$$

in a neighborhood of $x = 0$. Thus $G(x)$ generates the sequence $\left(\frac{2}{3}(-1)^k - 1\cdot 1^k + \frac{7}{3}2^k\right)_{k=0}^{\infty}$. The last theorem of the previous section indicates that this sequence (a_n), satisifes a linear homogeneous recurrence relation with characteristic polynomial

$$\chi(x) = (x-1)(x+1)(x-2) = x^3 - 2x^2 - x + 2,$$

which one might recognize as $x^3 \cdot d(\frac{1}{x})$, where d is the denominator of G. We say that χ and d are reciprocal polynomials. So our sequence is the solution of the linear homogeneous recurrence relation $a_n = 2a_{n-1} + a_{n-2} - 2a_{n-3}$, for $n \geq 3$, with initial conditions $a_0 = 2, a_1 = 3, a_3 = 9$.

If we can find a way to recapture $G(x)$ from this recurrence relation, we will have found another method for solving this recurrence relation.

We begin by realizing that we have infinitely many equations

$$\begin{aligned} a_3 &= 2a_2 + a_1 - 2a_0 \\ a_4 &= 2a_3 + a_2 - 2a_1 \\ a_5 &= 2a_4 + a_3 - 2a_2 \\ a_6 &= 2a_5 + a_4 - 2a_3 \\ \vdots &= \vdots \\ a_{k+3} &= 2a_{k+2} + a_{k+1} - 2a_k \\ \vdots &= \vdots \end{aligned}$$

We multiply each term of the equation ending with a_n by x^n, for all $n \geq 0$.

$$\begin{aligned}
a_3x^0 &= 2a_2x^0 + a_1x^0 - 2a_0x^0 \\
a_4x^1 &= 2a_3x^1 + a_2x^1 - 2a_1x^1 \\
a_5x^2 &= 2a_4x^2 + a_3x^2 - 2a_2x^2 \\
a_6x^3 &= 2a_5x^3 + a_4x^3 - 2a_3x^3 \\
\vdots &= \vdots \\
a_{k+3}x^k &= 2a_{k+2}x^k + a_{k+1}x^k - 2a_kx^k \\
\vdots &= \vdots
\end{aligned}$$

Now we add them all up collecting terms in columns

$$\sum_{n=0}^{\infty} a_{n+3}x^n = 2\Bigg[\sum_{n=0}^{\infty} a_{n+2}x^n\Bigg] + \Bigg[\sum_{n=0}^{\infty} a_{n+1}x^n\Bigg] - 2\Bigg[\sum_{n=0}^{\infty} a_nx^n\Bigg]$$

By definition of generating function we have

$$\sum_{n=0}^{\infty} a_{n+3}x^n = 2\Bigg[\sum_{n=0}^{\infty} a_{n+2}x^n\Bigg] + \Bigg[\sum_{n=0}^{\infty} a_{n+1}x^n\Bigg] - 2G(x)$$

And in fact if we concentrate on the sum on the left-hand side we see that

$$\begin{aligned}
\sum_{n=0}^{\infty} a_{n+3}x^n &= a_3x^0 + a_4x^1 + a_5x^2 + ... \\
&= \frac{x^3}{x^3}\Bigg[a_3x^0 + a_4x^1 + a_5x^2 + ...\Bigg] \\
&= \frac{1}{x^3}\Bigg[a_3x^3 + a_4x^4 + a_5x^5 + ...\Bigg] \\
&= \frac{1}{x^3}\Bigg[-a_0 - a_1x - a_2x^2 + a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 + ...\Bigg] \\
&= \frac{1}{x^3}\Bigg[-a_0 - a_1x - a_2x^2 + G(x)\Bigg] \\
&= \frac{1}{x^3}\Bigg[-2 - 3x - 9x^2 + G(x)\Bigg]
\end{aligned}$$

Similarly

$$2\Bigg[\sum_{n=0}^{\infty} a_{n+2}x^n\Bigg] = \frac{2}{x^2}\Bigg[-2 - 3x + G(x)\Bigg]$$

and

$$\Bigg[\sum_{n=0}^{\infty} a_{n+1}x^n\Bigg] = \frac{1}{x}\Bigg[-2 + G(x)\Bigg]$$

So substitution yields

$$\frac{1}{x^3}\Big[-2-3x-9x^2+G(x)\Big]=\frac{2}{x^2}\Big[-2-3x+G(x)\Big]+\frac{1}{x}\Big[-2+G(x)\Big]-2G(x)$$

Now multiply through by x^3 to give

$$-2-3x-9x^2+G(x)=2x\Big[-2-3x+G(x)\Big]+x^2\Big[-2+G(x)\Big]-2x^3G(x)$$

Distribute products and collect all terms with $G(x)$ on the left hand-side

$$-2-3x-9x^2+G(x)=-4x-6x^2+2xG(x)+-2x^2+x^2G(x)-2x^3G(x)$$

$$G(x)-2xG(x)-x^2G(x)+2x^3G(X)=2+3x+9x^2-4x-6x^2-2x^2$$

$$G(x)\Big[1-2x-x^2+2x^3\Big]=2-x+x^2$$

$$G(x)=\frac{2-x+x^2}{1-2x-x^2+2x^3}$$

In general to solve $a_{n+k}=\left[\sum_{i=1}^{k}c_ia_{n+k-i}\right]+f(n)$, for $n\geq 0$ given the initial conditions $a_0,a_1,...,a_{k-1}$, we let $G(x)$ denote the generating function for the sequence (a_n) and $H(x)$ denote the generating function for the sequence $(f(n))$. Then working as above we can write

$$\frac{1}{x^k}\left[G(x)-\sum_{i=0}^{k-1}a_ix^i\right]=\frac{c_1}{x^{k-1}}\left[G(x)-\sum_{i=0}^{k-2}a_ix^i\right]+...+\frac{c_{k-1}}{x}\Big[G(x)-a_0\Big]+c_kG(x)+H(x)$$

This expression can be solved for $G(x)$ if we can find $H(x)$. Then partial fraction expansion will yield (a_n).

Some remarks are in order. First, realize that this method still requires us to solve a system of linear equations of the same order as required by the method of characterisitic roots. However, we have no choice in the equations generated when we employ the method of characteristic roots. With the method of generating functions, we often get to select equations corresponding to eigenvalues of the linear system. These equations are therefore diagonal, or un-coupled. Also the method of generating functions is more amenable to changes in the forcing term $f(n)$. So if we are interested in stabililty, or the qualitative study of a particular kind of recurrence relation, the method of generating functions is the way to go. Finally, the methods extends nicely to systems of linear recurrence relations.

Example: Let us solve $h_{n+2}=6h_{n+1}-9h_n+2(n+2)$, for $n\geq 0$ given $h_0=1$, and $h_1=0$, by the method of generating functions.

Let $H(x)$ generate (h_n). Then

$$\sum_{n=0}^{\infty}h_{n+2}x^n=6\left[\sum_{n=0}^{\infty}h_{n+1}x^n\right]-9\left[\sum_{n=0}^{\infty}h_nx^n\right]+2\left[\sum_{n=0}^{\infty}(n+2)x^n\right]$$

$$\frac{1}{x^2}\Big[-1+H(x)\Big] = \frac{6}{x}\Big[-1+H(x)\Big] - 9H(x) + \frac{2}{x^2}\Big[-0x^0 - 1x^1 + \sum_{n=0}^{\infty} nx^n\Big]$$

$$-1+H(x) = -6x + 6xH(x) - 9x^2H(x) - 2x + \frac{2x}{(1-x)^2}$$

$$H(x)[1-6x+9x^2] = 1 - 8x + \frac{2x}{(1-x)^2}$$

$$H(x)[1-6x+9x^2] = \frac{1-8x+17x^2-8x^3}{(1-x)^2}$$

$$H(x) = \frac{1-8x+17x^2-8x^3}{(1-3x)^2(1-x)^2}$$

For the partial fraction expansion

$$H(x) = \frac{1-8x+17x^2-8x^3}{(1-3x)^2(1-x)^2} = \frac{A}{1-x} + \frac{B}{(1-x)^2} + \frac{C}{1-3x} + \frac{D}{(1-3x)^2}$$

$$1-8x+17x^2-8x^3 = A(1-x)(1-3x)^2 + B(1-3x)^2 + C(1-x)^2(1-3x) + D(1-x)^2$$

When $x = 1$, $1-8x+17x^2-8x^3 = 2$, while the right-hand side reduces to $B(-2)^2 = 4B$, so $B = 1/2$.
When $x = 1/3$, $1-8x+17x^2-8x^3 = -2/27$, while the right-hand side reduces to $D(2/3)^2 = 4D/9$, so $D = -1/6$.
Having exhausted the eigenvalues, we choose $x = 0$, which generates the equation $1 = A + B + C + D$, so $A + C = 2/3$ Secondly we set $x = -1$ which gives $34 = 32A + 16B + 16C + 4D$, which becomes $2A + C = 5/3$. We solve these two linear equations for $A = 1$, and $C = -1/3$. Thus

$$\begin{aligned} H(x) &= \frac{1}{1-x} + \frac{\frac{1}{2}}{(1-x)^2} - \frac{\frac{1}{3}}{1-3x} - \frac{\frac{1}{6}}{(1-3x)^2} \\ &= \sum_{n=0}^{\infty}\left(1^n + \frac{1}{2}(n+1)1^n - \frac{1}{3}3^n - \frac{1}{6}(n+1)3^n\right)x^n \end{aligned}$$

So

$$\begin{aligned} h_n &= 1 + \frac{1}{2}(n+1) - \frac{1}{3}3^n - \frac{1}{6}(n+1)3^n \\ &= \frac{3}{2} + \frac{n}{2} - 3^{n-1} - \frac{n}{2}3^{n-1} - \frac{1}{2}3^{n-1} \\ &= \frac{1}{2}[3 + n - 3^n - n3^{n-1}], \text{ for } n \geq 0 \end{aligned}$$

Example: As an example of using generating functions to solve a system of linear recurrence relations, let us consider the example from the end of the previous section.

$$\begin{aligned} 3^n &= a_n + b_n + c_n + d_n \\ a_{n+1} &= a_n + b_n + c_n \\ b_{n+1} &= a_n + b_n + d_n \\ c_{n+1} &= a_n + c_n + d_n \end{aligned}$$

where the initial conditions are $a_0 = c_0 = d_0 = 0$, and $b_0 = 1$.

Solving the first equation for d_n in terms of 3^n and the other terms, and substituting gives

$$\begin{aligned} a_{n+1} &= a_n + b_n + c_n \\ b_{n+1} &= 3^n - c_n \\ c_{n+1} &= 3^n - b_n \end{aligned}$$

Converting to generating functions gives

$$\begin{aligned} \frac{1}{x}\Big[A(x) - a_0\Big] &= A(x) + B(x) + C(x) \\ \frac{1}{x}\Big[B(x) - b_0\Big] &= \frac{1}{1-3x} - C(x) \\ \frac{1}{x}\Big[C(x) - c_0\Big] &= \frac{1}{1-3x} - B(x) \end{aligned}$$

Clear denominators and input the initial conditions

$$\begin{aligned} A(x) &= xA(x) + xB(x) + xC(x) \\ B(x) - 1 &= \frac{x}{1-3x} - xC(x) \\ C(x) &= \frac{x}{1-3x} - xB(x) \end{aligned}$$

From the first equation we have $(1-x)A(x) = x[B(x) + C(x)]$, so we can find $A(x)$ if we can find $B(x)$, and $C(x)$.

Substitute the third equation into the second

$$\begin{aligned} B(x) &= 1 + \frac{x}{1-3x} - xC(x) \\ &= \frac{1-3x+x}{1-3x} - x\left[\frac{x}{1-3x} - xB(x)\right] \\ &= \frac{1-2x-x^2}{1-3x} + x^2B(x) \end{aligned}$$

So $B(x) = \dfrac{1-2x-x^2}{(1-3x)(1-x)(1+x)} = \dfrac{\frac{1}{2}}{1-x} + \dfrac{\frac{1}{4}}{1+x} + \dfrac{\frac{1}{4}}{1-3x}$ which means $b_n = \dfrac{1}{2} + \dfrac{1}{4}(-1)^n + \dfrac{1}{4}3^n$, for $n \geq 0$.

Next,

$$\begin{aligned} C(x) &= \frac{x}{1-3x} - xB(x) = \frac{x(1-x^2) - (x - 2x^2 - x^3)}{(1-3x)(1-x^2)} \\ &= \frac{2x^2}{(1-3x)(1-x)(1+x)} = \frac{\frac{-1}{2}}{1-x} + \frac{\frac{1}{4}}{1+x} + \frac{\frac{1}{4}}{1-3x}. \end{aligned}$$

So $c_n = -\dfrac{1}{2} + \dfrac{1}{4}(-1)^n + \dfrac{1}{4}3^n$, for $n \geq 0$.

Now $B(x) + C(x) = \dfrac{1 - 2x - x^2 + 2x^2}{(1-3x)(1-x)(1+x)} = \dfrac{(1-x)^2}{(1-3x)(1-x)(1+x)} = \dfrac{1-x}{(1-3x)(1+x)}$. So $A(x) = \dfrac{x}{1-x}\Big[B(x) + C(x)\Big] = \dfrac{x}{(1-3x)(1+x)} = \dfrac{\frac{1}{4}}{1-3x} - \dfrac{\frac{1}{4}}{1+x}$. Consequently $a_n = \dfrac{1}{4}\Big[3^n - (-1)^n\Big]$, for $n \geq 0$. By symmetry we see $d_n = a_n$.

Chapter 3 Exercises

1. For each of the following functions, find its Maclaurin expansion be computing the derivatives $f^{(k)}(0)$.

a) $f(x) = \cos x$ b) $f(x) = e^{2x}$ c) $f(x) = \sin(3x)$

d) $f(x) = x + e^x$ e) $f(x) = xe^x$ f) $f(x) = \ln(1 + 5x)$

2. For each of the following functions use known Maclaurin expansions to find the Maclaurin expansion.

a) $f(x) = x^2 + \dfrac{1}{1-x}$ b) $f(x) = \dfrac{x^3}{1-x}$ c) $f(x) = \sin(x^3)$

d) $f(x) = \dfrac{1}{(x-1)^3}$ e) $f(x) = 7e^x + e^{8x}$ f) $f(x) = \ln(1 + 3x)$

g) $f(x) = x^5 \sin(x^2)$ h) $f(x) = \dfrac{1}{1-2x}\, e^x$

3. For the following sequences indexed by $(0, 1, 2, 3, ...)$ find the ordinary generating function. Simplify if possible.

a) $(1, 1, 1, 0, 0, 0, 0, 0, ...)$ b) $(1, 0, 2, 3, 4, 0, 0, 0, 0, 0, ...)$

c) $(3, 3, 3, ...)$ d) $(1, 0, 1, 1, 1, 1, ...)$

e) $(0, 0, 0, 1, 1, 1, 1, 1, 1, ...)$ f) $(0, 0, 4, 4, 4, 4, 4, ...)$

g) $(1, 1, 1, 2, 1, 1, 1, 1, 1, ...)$ h) $(a_k) = (\frac{2}{k!})$

i) $(a_k) = (\frac{2^k}{k!})$ j) $(0, 0, \frac{1}{2!}, \frac{1}{3!}, \frac{1}{4!}, ...)$

k) $(1, -1, \frac{1}{2!}, \frac{-1}{3!}, \frac{1}{4!}, ...)$ l) $(1, 0, 1, 0, 1, 0, 1, ...)$

m) $(2, 0, \frac{-2}{3!}, 0, \frac{2}{5!}, 0, \frac{-2}{7!}, ...)$ n) $(3, \frac{-3}{2}, \frac{3}{3}, \frac{-3}{4}, \frac{3}{5}, ...)$

4. Find the sequence whose ordinary generating function is given.

a) $f(x) = (x + 5)^2$ b) $f(x) = (1 + x)^4$ c) $f(x) = \frac{x^5}{1-x}$

d) $f(x) = \frac{1}{1-3x}$ e) $f(x) = \frac{1}{1+8x}$ f) $f(x) = e^{4x}$

g) $f(x) = 1 + \frac{1}{1-x}$ h) $f(x) = 1 + e^x$ i) $f(x) = xe^x$

j) $f(x) = x^3 + x^4 + e^x$ k) $f(x) = \frac{1}{1-x^2}$ l) $f(x) = e^{-2x}$

m) $f(x) = \sin(3x)$ n) $f(x) = \frac{1}{1+x^2}$ o) $f(x) = \frac{1}{(1+x)^2}$

p) $f(x) = \dfrac{1}{1-3x} + \dfrac{4x}{1-x}$ q) $f(x) = \dfrac{e^x + e^{-x}}{2}$

5. Find a simple expression for the ordinary generating function for each sequence.

a) $a_k = k + 2$ b) $a_k = 7k$ c) $a_k = k^2$

d) $a_k = k(k + 1)$ e) $a_k = (k + 1)\frac{1}{k!}$

6. In each of the following set up the appropriate generating function. DO NOT CALCULATE AN ANSWER, BUT INDICATE WHAT YOU ARE LOOKING FOR eg the coefficient of x^9.

a) An athletics director wants to pick at least 3 small college teams as football opponents for a particular season, at least 3 teams from medium-sized colleges, and at least 2 large-college teams. She has limited the choice to 7 small college teams, 6 medium-sized college teams, and

4 teams from large colleges. In how many ways can she select 11 opponents assuming that she is not distinguishing 2 teams from a particular sized college as different?

b) In how many ways can 8 binary digits be selected if each must be selected an even number of times?

c) How many ways are there to choose 10 voters from a group of 5 republicans, 5 democrats and 7 independents if we want at least 3 independents and any two voters of the same political persuasion are considered indistinguishable?

d) A Geiger counter records the impact of five different kinds of radioactive particles over a period of five minutes. How many ways are there to obtain a count of 20?

e) In checking the work of a proofreader we look for 4 types of proofreading errors. In how many ways can we find 40 errors?

f) How many ways are there to distribute 15 identical balls into 10 distinguishable cells?

g) Repeat f) if no cell may be empty.

h) How many solutions in integers are there to $x_1 + x_2 + x_3 = 12$, where $0 \leq x_i \leq 6$?

7. Use the binomial theorem to find the coefficient of x^4 in the expansion of

a) $f(x) = \sqrt[3]{1+x}$ b) $f(x) = (1+x)^{-2}$ c) $f(x) = (1-x)^{-5}$ d) $f(x) = (1+4x)^{\frac{1}{2}}$

8. Find the coefficient of x^7 in the expansion of

a) $f(x) = (1-x)^{-6}x^4$ b) $f(x) = (1-x)^{-4}x^{11}$ c) $f(x) = (1+x)^{\frac{1}{2}}x^3$

9. If $f(x) = (1+x)^{\frac{1}{3}}$ is the ordinary generating function for (a_k), find a_k.

10. Each of the following function is the exponential generating function for a sequence (a_k). Find the sequence.

a) $f(x) = 3 + 3x + 3x^2 + 3x^3 + ...$ b) $f(x) = \frac{1}{1-x}$ c) $f(x) = x^2 + 3x$

d) $f(x) = e^{6x}$ e) $f(x) = e^x + e^{4x}$ f) $f(x) = (1+x^2)^n$

11. Another mythical tale tells of magical pairs of rabbits. The pairs behave in the following fashion. Any newborn pair of rabbits are one male and one female. A pair mates for life, and inbreeding doesn't introduce any problems. Once a pair is two months old they reproduce issuing exactly one new pair each month. Let f_n denote the number of pairs of rabbits after n months. Suppose that $f_0 = 0$ and a newborn pair is given as a gift, so $f_1 = 1$. Find a recurrence relation for f_n.

12. For each of the following sequences find a recurrence relation satisfied by the sequence. Include a sufficient number of initial conditions to specify the sequence.

a) $a_n = 2n + 2$, $n \geq 0$ b) $a_n = 2 \cdot 3^n, n \geq 1$

c) $a_n = n^2, n \geq 0$ d) $a_n = n + (-1)^n, n \geq 0$

13. Find a recurrence relation for the number of binary strings of length n which do not contain a pair of consecutive zeroes.

14. Find a recurrence relation for the number of trinary strings of length n which contain a pair of consecutive zeroes.

15. Find a recurrence relation for the number of binary strings of length n which do not contain the substring 01. Try again with 010 in place of 01.

16. A codeword over $\{0, 1, 2\}$ is considered legitimate iff there is an even number of 0's and an odd number of 1's. Find simultaneous recurrence relations from which it is possible to compute the number of legitimate codewords of length n.

17. Suppose that we have stamps of denominations 4,8, and 10 cents each in unlimited supply. Let $f(n)$ be the number of ways to make n cents postage assuming the ordering of the stamps used matters. So a six cent stamp followed by a four cent stamp is different from a four cent stamp followed by a six cent stamp. Find a recurrence relation for $f(n)$. Check your recurrence by computing $f(14)$ and enumerating all possible ways to make fourteen cents postage.

18. Find the characteristic equation of each of the following recurrences.

a) $a_n = 2a_{n-1} - a_{n-2}$

b) $b_k = 10b_{k-1} - 16b_{k-2}$

c) $c_n = 3c_{n-1} + 12c_{n-2} - 18c_{n-3}$

d) $d_n = 8d_{n-4} + 16d_{n-5}$

e) $e_k = e_{k-2}$

f) $f_{n+1} = -f_n + 2f_{n-1}$

g) $g_n = 15g_{n-1} + 12g_{n-2} + 11g_{n-3} - 33g_{n-8}$

h) $h_n = 4h_{n-2}$

i) $i_n = 6i_{n-1} - 11i_{n-2} + 6i_{n-3}$

j) $j_n = 2j_{n-1} + j_{n-2} - 2j_{n-3}$

19. Find the characteristic roots of each recurrence from exercise 18

20. Solve the recurrence relations using the method of characteristic roots.

a) $a_0 = 1, a_1 = 6$ and $a_n = 6a_{n-1} - 9a_{n-2}$, for $n \geq 2$.

b) $a_0 = 3, a_1 = 6$ and $a_n = a_{n-1} + 6a_{n-2}$, for $n \geq 2$.

c) $a_2 = 5, a_3 = 13$ and $a_n = 7a_{n-1} - 10a_{n-2}$, for $n \geq 4$.

d) $a_0 = 6, a_1 = -3$ and $a_n = -4a_{n-1} + 5a_{n-2}$, for $n \geq 2$.

e) $a_0 = 0, a_1 = 1$ and $a_n = a_{n-1} + a_{n-2}$, for $n \geq 2$.

f) $a_0 = 2, a_1 = 5, a_2 = 15$, and $a_n = 6a_{n-1} - 11a_{n-2} + 6a_{n-3}$, for $n \geq 3$.

21. Use generating functions to solve each of the recurrences.

a) $a_n = 2a_{n-1} - a_{n-2} + 2^{n-2}, n \geq 2$, where $a_0 = 3$, $a_1 = 5$

b) $b_k = 10b_{k-1} - 16b_{k-2}, k \geq 2$, where $b_0 = 0$, and $b_1 = 1$

c) $c_m = -c_{m-1} + 2c_{m-2}, m \geq 2$, where $c_0 = c_1 = 1$

d) $d_n = 6d_{n-1} - 11d_{n-2} + 6d_{n-3}$, where$d_0 = 0$, $d_1 = 1$, and $d_2 = 2$.

22. Find a generating function for$C_{n+1} = 2nC_n + 2C_n + 2, n \geq 0$, where $C_0 = 1$. Then find an exponential generating function for the same recurrence.

23. In each case suppose that $G(x)$ is the ordinary generating function for (a_n). Find a_n.

a) $G(x) = \dfrac{1}{(1-2x)(1-4x)}$

b) $G(x) = \dfrac{2x+1}{(1-3x)(1-4x)}$

c) $G(x) = \dfrac{x^2}{(1-4x)(1-5x)(1-6x)}$

d) $G(x) = \dfrac{1}{8x^2 - 6x + 1}$

e) $G(x) = \dfrac{x}{x^2 - 3x + 2}$

f) $G(x) = \dfrac{1}{6x^3 - 5x^2 + x}$

g) $G(x) = \dfrac{2 - 3x}{(1 - x)^2(1 - 2x)}$

24. Solve simultaneously the recurrences

$$\begin{aligned} a_{n+1} &= a_n + b_n + c_n, n \geq 1 \\ b_{n+1} &= 4^n - c_n, n \geq 1 \\ c_{n+1} &= 4^n - b_n, n \geq 1 \end{aligned}$$

subject to $a_1 = b_1 = c_1 = 1$.

Chapter 4: Pólya Counting

Abstract algebra, or modern algebra, is an area of mathematics where one studies general algebraic structures. In contrast the algebra that one studies in, for example, college algebra is very specific and concrete - dealing (almost) exclusively with algebra using real or complex number systems.

Basic counting requires only basic algebra. Intermediate counting requires algebra which is a bit more generalized, namely the formal manipulation of power series. It therefore makes sense that advanced counting requires abstract algebra.

The roots of the counting techniques we study are in the work Burnside and Frobenius. George Pólya is credited with the most fundamental theorem pertinent to this approach to counting. Thus this type of counting is named after him.

We start with a large collection of objects. We then decide upon a framework to help us classify the objects (uniquely) by various types. The goal is to develop a theory to determine the number of distinct types. The theory we develop uses group theory in deciding which objects are of the same type.

We therefore begin with a section devoted to developing the notion of when two objects are of the same type. This is followed by some basic group theory in the second section. The Fundamental Lemma in section three is often attributed to Burnside. There are arguments that it should be attributed elsewhere. We will therefore not specifically attribute it to him. Still if the student should find themselves reading an older text, they'll recognize the lemma by the moniker "Burnside's Lemma".

§4.1 Equivalence Relations

Given a large set, in order to classify it's members into distinct types, the classification scheme must meet certain criteria. First, for any element of the set there should be a class that the element fits into. Second, no element should fall into two classes simultaeously. Otherwise we cannot claim that the types actually classify the elements. Finally, in part because we want to be able to determine the number of distinct classes, we shall want to require that no class is empty. The point is that we could arbitrarily define classes which will be empty and not really be counting the number of necessary classes.

For such a classification scheme the distinct classes $C_1, C_2, ..., C_r$ form a partition of the set A. A partition of a set , A, is a collection of pairwise-disjoint, non-empty subsets, whose union is A.

Given a partition of A into the parts $C_1, C_2, ..., C_r$ we can relate two elements of A when they are in the same part. This relation, $R \subseteq A \times A$, has the property that it is reflexive, since every element of A is in some part. R is also symmetric, since if a and b are in the same part, so are b and a. Finally R is transitive since if a and b are in the same part C_i, and b and c are in the same part C_j, then $b \in C_i \cap C_j$. From the fact that the parts are pairwise disjoint we conclude that $C_i = C_j$. Whence a and c are in the same part. So a partition of a set defines an equivalence relation on the set.

Conversely given a relation R on A which is an equivalence relation, so it's simultaneously reflexive, symmetric and transitive, we can define the equivalence class of any $a \in A$, denoted by $[a]$, as the set of all elements of A in the relation with a.

Theorem: *If R is an equivalence relation on A, and $a, b \in A$, then $[a] \cap [b] \neq \emptyset$ implies that $[a] = [b]$*

Proof: Let $c \in [a] \cap [b]$. Then $(a,c),(b,c) \in R$. Since R is symmetric $(a,c),(c,b) \in R$. Because R is transitive we get $(a,b) \in R$. So now if $d \in [a]$, we know (b,a) and (a,d) are in R. Therefore $(b,d) \in R$, which is equivalent to $d \in [b]$. Thus $[a] \subseteq [b]$. The result follows by appealing to symmetry of argument. ■

As a corollary we draw that the distinct equivalence classes of an equivalence relation form a partition of the set.

Example: A standard deck of cards contains 52 cards. Each card has a rank $2,3,4,5,6,7,8,9$, 10, $J =$ jack,$Q =$ queen, $K =$ king, or $A =$ ace. Each card also has a suit $\heartsuit, \spadesuit, \clubsuit,$, or $\diamondsuit$.

So we could choose to classify the cards in a standard deck by considering two cards "the same" if they had the same suit. In this case there would be four distinct types of objects.

Of course, we might also define two cards to be the same if they were of the same rank. Now there would be thirteen distinct types of objects.

So we sometimes might need to use a subscript $[a]_R$ to distinguish the equivalence class of an element a with respect to R as opposed to $[a]_T$ which would denote the equivalence class of a with respect to T.

Example: The relation $\pmod{m}$ on $\mathbb{Z}$.

Example: 2-Colorings of C_4

§4.2 Permutation Groups

A binary operation on a set S is a function from $S \times S$ into S. Standard examples are addition and multiplication. We will use multiplicative notation. So the image of (g,h) is written gh.

A group is a set with a binary operation which is associative, i.e. $a(bc) = (ab)c$ for all $a, b, c, \in G$, has identity, i.e. there exists $e \in G$ with $eg = g = ge$ for all $g \in G$, and inverses, so for each $g \in G$, there is h, denoted g^{-1}, with $gh = e = hg$. If in addition the operation is commutative ($ab = ba$ for all $a, b \in G$) we call the group abelian. Denote the cardinality of G by $|G|$. This is called the order of G.

Example: $\mathbb{Z}, +$. Here the identity is additive $0 + n = n + 0 = n$, and the inverse is too: $n + (-n) = (-n) + n = 0$.

Example: $\mathbb{R}^* = \mathbb{R} - \{0\}, \cdot$. Here we suppress the symbol and use juxtaposition. The identity is 1, and the inverse is the reciprocal.

Notice that in a group we have the cancellation property: $ac = ab$ implies $c = b$.

Example: Let A be a set and put $Sym(A) = \{f : A \longrightarrow A | f \text{ is bijective }\}$. Then $Sym(A)$ is a group with operation being composition of functions. The identity is the identity function on A, and inverses are inverse functions.

If $|A| = |B| = n < \infty$, there is a bijective function $f : A \longrightarrow B$. Then Sym $(A) \cong$ Sym (B) where $\pi \in$ Sym (A) corresponds to $f \pi f^{-1} \in$ Sym (B) one must check the correspondence is 1-1, onto and preserves operations, f is a group isomorphism, which of course behaves somewhat like a graph isomorphosm. The only difference is graph isomorphisms preserve adjacency and group isomorphisms preserve operation.

Standardly, if $|A| = n$, we therefore take $A = \{1, 2, 3, ..., n\}$ and denote Sym (A) by S_n, the symmetric group on n letters. This is a permutation group because its elements are n-permutations of an n-set.

Elements of S_n may be described by two-row tables, T-tables, bipartite graphs, etc.

For computational ease and typesetting efficiency we use cycle notation. $(a_1 a_2 ... a_l)$ denotes the permutation on n letters where a_i gets mapped to a_{i+1}, with subscripts computed modulo l, and where any letter not in $\{a_1, ..., a_l\}$ is fixed. This is a rather important convention.

Two permutations, π and ρ are disjoint, if $\pi(a) \neq a$ implies $\rho(a) = a$ and vice versa.

Theorem: *Every nonidentity permutation $\pi \in S_n$ can be written (uniquely up to order) as a product of disjoint cycles.*

Sketch of proof: The proof uses the second form of mathematical induction, and can be compared to the standard proof of the Fundamental Theorem of Arithmetic. In the inductive step we pick x with $\pi(x) \neq x$. Form the cycle $\gamma = (x, \pi(x), \pi^2(x), ..., \pi^{l-1}(x))$ (so l is the least positive integer with $\pi^l(x) = x$). If $\pi = \gamma$, we're done. Otherwise, since γ fixes any element of $\{1, ..., n\} - \{x, \pi(x), ..., \pi^{l-1}(x)\}$, $\pi\gamma^{-1}$ fixes $\{x, \pi(x), ..., \pi^{l-1}(x)\}$ and can therefore be considered as an element of S_{n-l}. Induction kicks in and we're done. ∎

The order of the cycles is not fixed in the previous theorem because

Fact: *If γ and π are disjoint, then $\gamma\pi = \pi\gamma$.*

Sketch of proof: Check that the two compositions have the same value at each $i \in \{1, 2, ..., n\}$. There are essentially three cases to consider.

Notice that the converse is not true. Any non-identity permutation will commute with its powers, but will not be disjoint from them.

A subgroup of a group is a subset which is also a group. We write $H \leq G$ to mean H is a subgroup of G.

Theorem: *Let $\emptyset \neq H \subseteq G$, where G is a group. H is a subgroup iff $ab^{-1} \in H$ for all $a, b \in H$.*

Proof: The forward implication is trivial.

For the converse, since $H \neq \emptyset$, there is $a \in H$. Hence $e = aa^{-1} \in H$. And if $b \in H$, $b^{-1} = eb^{-1} \in H$. H inherits associativity from G. ∎

Theorem: *Let $\emptyset \neq H \subseteq G$, where G is a group and $\#H = n < \infty$. H is a subgroup iff $ab \in H$ for all $a, b \in H$.*

Proof: The forward implication is trivial.

Let $a \in H$. Among the elements $a, a^2, a^3, ..., a^n, a^{n+1}$, all must be in H by closure. Since $|H| = n$, these elements cannot all be distinct. Therefore there exists superscripts i and j with $j > i$ so that $a^j = a^i$. Therefore $a^{j-i} = e \in H$. In fact we see that $a^{-1} = a^{j-i-1}$. So for $a, b \in H$, $a, b^{-1} \in H$. By closure $ab^{-1} \in H$. And we're done by the previous theorem. ∎

Notice that a cycle γ of length l satisifies that l is the smallest positive integer for which $\gamma^l = id$. We say the the cycle has order l, since the set of distinct powers of $\{id = \gamma^0, \gamma, ...\gamma^{l-1}\}$ is a group or order l, since it satisfies the previous theorem. We write $\langle\gamma\rangle$ for this group.

Example: $< (1234) >= \{id, (1234), (13)(24), (1432)\}$. Notice that a cycle's order is the same as its length.

For general permutations we have that when $\gamma\pi = \pi\gamma$, by induction $(\gamma\pi)^n = \gamma^n\pi^n$. Together with the fact that the order (as a group element) of an l-cycle is l when γ and π commute $o(\gamma\pi) = lcm(o(\gamma), o(\pi))$. Thus if we write a permutation as a product of disjoint cycles, we can easily compute its order.

When π is the product of disjoint cycles which includes exactly e_a cycles of length $a > 2$, we define the <u>type</u> of π to be type $(\pi) = \prod_a a^{e_a}$, and extend to the identity by saying the identity has type 1. The order of π as a group element is the least common multiple of lengths a, where $e_a > 0$.

<u>§4.3 Group Actions</u>

If S is a set and G is a group, <u>G acts on S</u> if there is a map $* : G \times S \longrightarrow S$ so that
1) $g_1 * (g_2 * s) = (g_1g_2) * s$ and 2) $e * s = s$ for all $s \in S$.

Example: Every group acts on itself by left multiplication.

Remark: We sometimes will find it useful to use functional notation for a group action. So instead of writing $g * s$ we will write $g(s)$.

If G acts on S define the relation $\sim$ on S by $x \sim y$ iff there exists $g \in G$ with $g * x = y$.

Theorem: *$\sim$ is an equivalence relation on G.*

Proof: 1) For $s \in S$, $s \sim s$ since $e * s = s$. Therefore $\sim$ is reflexive.

2) If $x \sim y$, there is $g \in G$ so that $g * x = y$. Then $g^{-1} * y = g^{-1} * (g * x) = (g^{-1}g) * x = e * x = x$, so $y \sim x$. Therefore $\sim$ is symmetric.

3) If $x \sim y$ and $y \sim z$, then there are $g, h \in G$ with $g * x = y$, and $h * y = z$. So $(hg) * x = h * (g * x) = h * y = z$. Hence $\sim$ is transitive. ■

Example: $\langle(1234)\rangle$ acts on the 2-colored C_4's.

Example: $\{id, \text{reverse}\}$ acts on the set of open necklaces of length k.

General example: G a group acts on itself by conjugation $g * h = ghg^{-1}$. Here the equivalence classes are the conjugacy classes.

For $x \in S$, the <u>orbit</u> of x under G is denoted $G * x = \{g * x | g \in G\}$. Also $G_x = \{g \in G | g * x = x\}$ is the <u>stabilizer</u> of x in G.

Fact: $G_x \leq G$. (the proof is straightforward)

Theorem: $|G| = |G * a| \cdot |G_a|$.

Proof: Suppose that $G * a = \{b_1, ..., b_r\}$, where the b_i's are all distinct. Then there is an r-subset $P = \{\pi_1, \pi_2, ..., \pi_r\}$ of G with $\pi_i * a = b_i$, for $i = 1, 2, ..., r$ (the π's are distinct since $\pi * a$ is unique).

Now for $\gamma \in G$, if $\gamma * a = b_k = \pi_k * a$, then $\pi_k^{-1} * \gamma * a = a$. So $\pi_k^{-1}\gamma \in G_a$. Say $\pi_k^{-1}\gamma = \sigma \in G_a$. Then $\gamma = id\gamma = (\pi_k\pi_k^{-1})\gamma = \pi_k(\pi_k^{-1}\gamma) = \pi_k\sigma$. So every element of G can be written as something in P times something in G_a. Therefore $|G| \leq |P||G_a| = |G * a||G_a|$.

To show equality suppose that $\gamma = \pi_k\sigma = \pi_m\tau$, where $\pi_k, \pi_m \in P$ and $\sigma, \tau \in G_a$. Thus $\pi_k\sigma * a = \pi_k * a = b_k$, while $\pi_m\tau * a = \pi_m * a = b_m$. So $b_k = b_m$, and thus $\pi_k = \pi_m$. Thus by the cancellation property of a group $\sigma = \tau$. ■

When G is finite, the theorem gives a useful divisibility condition.

General Example: If $H \leq G$, for any $g \in G$, $gH = \{gh | h \in H\}$ is the left coset of H in G labelled by g. The group G acts on $S = \{gH | g \in G\}$, the set of left cosets of H in G, by left multiplication i.e. $g * xH = gxH$. In this case $G_H = H$ (i.e. $aH = H$ iff $a \in H$, and generally $aH = bH$ iff $ab^{-1} \in H$). We denote the number of distinct left cosets of H in G by $[G : H]$ and call it the index of H in G. The previous theorem says that $|G| = |H| \cdot [G : H]$, which is Lagrange's Theorem.

If G acts on S and there is $s \in S$ so that $G * s = S$, then G acts transitively on S. In the previous general example G acts transitively on the left cosets of H by left multiplication. It happens that G also acts transitively on the right cosets of H in G, by $g * Hk = Hkg$. So the theorem allows us to deduce that the number of distinct right cosets of H in G is the same as the number of distinct left cosets of H in G.

When $H \leq G$ sometimes the set of left cosets is different from the set of right cosets. For example the cosets of $\langle(12)\rangle$ in S_3. However sometimes the set of left cosets is the same as the set if right cosets, i.e. $gH = Hg$ for all $g \in G$. We call such a subgroup normal, and write $N \triangle G$.

The condition that $gN = Ng$ means that for every $n \in N$, $gn \in Ng$. Therefore there exists $n' \in N$ with $gn = n'g$. So $gN = Ng$ iff for all $n \in N, gng^{-1} = n' \in N$ iff $gNg^{-1} = N$ for all $g \in G$.

If $N \triangle G$ and $n \in N$, then C_n the conjugacy class of n must be in N. Therefore $N \triangle G$ implies that N is a union of conjugacy classes. So computing the conjugacy classes of G can facilitate finding normal subgroups of G.

By generalizing the example of $\langle(123)\rangle \triangle S_3$ we conclude that if $[G : H] = 2$, then $H \triangle G$.

An element $a \in A$ is invariant under $\pi \in G$ if $\pi * a = a$.

Example: The four bead necklace with all beads white is invariant under (1234).

For a permutation π we define $\text{Inv}(\pi) = |\{a \in A | \pi * a = a\}|$. So $\text{Inv}(\pi)$ is the number of elements in A which π leaves invariant. We call the set $\{a \in A | \pi * a = a\} =$ Fix (π). So $\text{Inv}(\pi)$ is the size of $\text{Fix}(\pi)$. We also say that $\pi \in G$ stabilizes a, if $\pi * a = a$.

THE LEMMA: *Let S be the equivalence relation on A induced by the action of a group G. Then the number of distinct equivalence classes is $\frac{1}{|G|} \sum_{\pi \in G} Inv(\pi)$.*

Proof: Let $F = \{(\pi, x) \in G \times A | \pi * x = x\}$. Define $\mathbb{1}(g, x)$ to be 1 if $(\pi, x) \in F$, and 0 otherwise. Finally let $S = \{G * x_1, G * x_2, ..., G * x_r\}$ be the distinct orbits of A under G.

$$\begin{aligned} |F| &= \sum_{x \in A} \sum_{\pi \in G} \mathbb{1}(\pi, x) = \sum_{x \in A} |G_x| \\ &= \sum_{\pi \in G} \sum_{x \in A} \mathbb{1}(\pi, x) = \sum_{\pi \in G} |\text{Fix}(\pi)| = \sum_{\pi \in G} \text{ Inv } (\pi) \end{aligned}$$

Hence

$$\begin{aligned}\frac{1}{|G|}\sum_{\pi\in G}\text{Inv}(\pi) &= \frac{1}{|G|}\sum_{x\in A}|G_x| = \sum_{x\in A}\frac{|G_x|}{|G|}\\ &= \sum_{x\in A}\frac{|G_x|}{|G*x||G_x|} = \sum_{x\in A}\frac{1}{|G*x|}\\ &= \sum_{i=1}^{r}\sum_{x\in G*x_i}\frac{1}{|G*x_i|}\\ &= \sum_{i=1}^{r}\left[\frac{1}{t}+\frac{1}{t}+...+\frac{1}{t}\right], t \;\text{ terms}\\ &= \sum_{i=1}^{r}1 = r = |S|\end{aligned}$$

So the number of orbits is the average of the fix set sizes. ■

Example: There are six equivalence classes when $\langle(1234)\rangle$ acts on the 2-colored necklaces with 4 beads.

§4.4 Colorings

Let D be a set. A coloring of D is an assignment of a color to each element of D. That is a coloring corresponds to a function $f : D \longrightarrow R$, where R is a set of colors. When $|D| = k$, and $|R| = m$, there are m^k colorings of D using the colors from R.

Let $C(D, R)$ denote the set of all colorings of D using the colors from R.

If G is a permutation group on D and $\pi \in G$, then there is a corresponding permutation π^* of $C(D, R)$. If f is a coloring, then $\pi^*(f)$ is another coloring where $\pi^*f(d) = f(\pi(d))$, for all d in D.

As will be seen in our examples, we might also define this more naturally as $\pi^*f(d) = f(\pi^{-1}(d))$, for all $d \in D$. The point is that in our fundamental lemma, we are summing over all group elements - which is equivalent to summing over all of the inverses of group elements.

Example: Label the vertices of a 4-cycle clockwise 1,2,3, and 4. Color these Red, Blue, Yellow and Green. We can think of f as the output string RBYG. Let $\pi = (1234)$. Then π^*fhas output string BYGR.

As an obvious fact we have

Fact: *The set of induced permutations $G^* = \{\pi^*|\pi \in G\}$ is a group under composition of functions, and $|G^*| = |G|$.*

More to the point is the following fact.

Fact: *If G induces an equivalence relation S on D, then G^* induces an equivalence relation S^* on $C(D, R)$ by fS^*g iff there exists $\pi \in G$ with $g = \pi^*f$.*

The proof is left to the reader. Notice here that $g = \pi^*f$ iff $f = (\pi^{-1})^*g$. So our previous remark is well-founded.

The point is that we may concentrate on D and groups acting on D, since really all that is important about R is its cardinality.

§4.5 The Cycle Index and the Pattern Inventory

When $\pi \in S_n$ is written as a product of disjoint cycles, the result is called its cycle decomposition. We will write $cd(\pi)$ for the cycle decomposition of π. Given π we can order the terms of $cd(\pi)$ with all the 1-cycles first, followed by any 2-cycles, etc. So $\pi = \prod_{i=1}^{l}\prod_{j=1}^{e_i}\gamma_{i,j}$, where $\gamma_{i,j}$ is a cycle of length i and there are e_i cycles of length i appearing.

Next we define $type(\pi) = \prod_{i=1}^{l} i^{e_i}$. Notice that $\sum_{i=1}^{l} ie_i = n$. Also we define $cyc(\pi) = \sum_{i=1}^{l} e_i$ which is the number of cycles in $cd(\pi)$.

Example: $(1234)(56)(78) \in S_8$ has type $2^2 4^1$ and $cyc((1234)(56)(78)) = 3$.

Theorem: (Pólya Version 1) *Suppose that G is a group acting on D. Let R be an m-set of colors. Then the number of distinct colorings in $C(D,R)$ which is the number of distinct equivalence classes for S^* induced by G on $C(D,R)$ equals*$\frac{1}{|G|}\sum_{\pi\in G} m^{cyc(\pi)}$.

Proof: By the fundamental lemma it suffices to show that $m^{cyc(\pi)} = Inv(\pi^*)$ for all $\pi \in G$. But any element of $C(D,R)$ is left invariant under π^* iff all elements of D in a cycle of π are colored the same color. So for each cycle in $cd(\pi)$ we have m choices for color. The multiplication principle now gives the result. ■

So as promised the induced group G^*, and anything about the set of colors R, except its size, are immaterial.

If G is a permutation group where k is the length of the longest cycle occuring in the cycle decomposition of any element of G, the cycle index of G is defined as

$$P_G[x_1, x_2, ..., x_k] = \frac{1}{|G|}\sum_{\pi\in G}\left(\prod_{a\geq 1} x_a^{e_a}\right)$$

where the exponents in the products come from the cycle decompositions of the corresponding permutations.

Example: $G = \langle(1234)\rangle$ consists of $\{(1)(2)(3)(4), (1234), (13)(24), (1432)\}$. So

$P_G(x_1, x_2, x_3, x_4) = \frac{1}{4}[x_1^4 + x_2^2 + 2x_4]$.

Notice that $P_G(m,m,m,m) = \frac{1}{4}[m^4 + m^2 + 2m] = \frac{1}{|G|}\sum_{\pi\in G} m^{cyc(\pi)}$.

A weight function $w : R \longrightarrow S$, is any function from a set, R, of colors into the set S.

Theorem: (Pólya Version 2) *If a group G acts on a set D whose elements are colored by elements of R, which are weighted by w, then the expression*

$$P_G\left[\sum_{r\in R} w(r), \sum_{r\in R}(w(r)^2), ..., \sum_{r\in R}(w(r)^k)\right]$$

generates the pattern inventory of distinct colorings by weight, where $P_G[x_1, x_2, ..., x_k]$ is the cycle index of G.

Proof: See appendix 1.

Most especially the constant function $w(r) = 1$ for all $r \in R$ gives the number of distinct colorings.

By singling out a particular color, say $w(r) = 1$ for $r \neq b$, and $w(b) = b \neq 1$. we can generate the distinct colorings enumerated by how many times the color b is used.

Chapter 4 Exercises

1. In each case determine whether S is an equivalence relation on A. If it is not, determine which properties of an equivalence relation fail to hold.

a) A is the power set of $\{1, 2, 3, ..., n\}$, aSb iff a and b have the same number of elements.

b) A is the power set of $\{1, 2, 3, ..., n\}$, aSb iff a and b are disjoint.

c) A is all people in Grand Forks, aSb iff a and b have the same blood type.

d) A is all people in North Dakota, aSb iff a and b live within 10 miles of each other.

e) A is all bit strings of length 10, aSb iff a and b have the same number of ones.

2. For each equivalence relation from exercise 1, identify all equivalence classes.

3. In each case find $\pi_1 \circ \pi_2$

a) $\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$, $\quad \pi_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}$

b) $\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$, $\quad \pi_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}$

c) $\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 4 & 5 \end{pmatrix}$, $\quad \pi_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 3 & 5 & 4 \end{pmatrix}$

d) $\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 1 & 5 & 2 & 4 & 3 \end{pmatrix}$, $\quad \pi_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 1 \end{pmatrix}$

4. In each case determine if $\circ$ is a binary operation on X. If it is, determine which of the three group axioms hold for $X, \circ$.

a) $X = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 \end{pmatrix} \right\}$, where $\circ$ is composition of functions.

b) $X = \mathbb{Q}$, $\circ$ is addition.

c) $X = \mathbb{Q}$, $\circ$ is multiplication.

d) $X = \mathbb{N}$, $\circ$ is addition.

e) $X = \mathbb{R} - \{0\}$, $\circ$ is multiplication.

f) X is all 2×2 matrices with real entries, $\circ$ is matrix multiplication.

5. In each case the group G induces an equivalence relation on the set A, find all of the distinct equivalence classes.

a) $A = \{1, 2, 3, 4, 5\}$, $G = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 2 & 3 & 4 & 1 \end{pmatrix} \right\}$, $\circ$ is composition.

b) $A = \{1, 2, 3, 4, 5, 6\}$, $G = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 4 & 5 & 6 \end{pmatrix}, \right.$
$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 4 & 3 & 5 & 6 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 4 & 6 & 5 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 4 & 3 & 5 & 6 \end{pmatrix},$
$\left. \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 4 & 6 & 5 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 4 & 3 & 6 & 5 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 4 & 3 & 6 & 5 \end{pmatrix} \right\}$, $\circ$ is composition.

c) $A = \{1, 2, 3, 4, 5\}$, $G = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 1 & 2 \end{pmatrix}, \right.$
$\left. \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 2 & 3 & 4 \end{pmatrix} \right\}$, $\circ$ is composition.

6. Check your answers to exercise 5 by using THE LEMMA to compute the number of distinct equivalence classes.

7. Check your answers to exercise 6 by computing $\displaystyle\sum_{a\in A} \frac{1}{|G * a|}$.

8. How many allowable colorings (not necessarily distinct) are there for the vertices of a cube if the allowable set of colors is {red, green, blue}?

9. How many allowable colorings (not necessarily distinct) are there for the vertices of a regular tetrahedron if the allowable set of colors is {red, green, blue, yellow}?

10. We make open necklaces using two colors of beads (red and blue), and consider two the same if they are identical, or if one is the reversal of the other. If a necklace consists of 3 beads

a) Find G^*

b) Find the number of distinct necklaces using THE LEMMA.

c) Check your answer by enumerating the distinct necklaces.

11. Repeat exercise 10, if we use 4 beads instead of 3.

12. Repeat exercise 10, if we use 5 beads per necklace.

13. Repeat exercise 10, if we use three colors of beads (red, white, and blue).

14. Suppose that we make closed necklaces using two colors of beads, (red and blue), and consider two the same if they are identical, or one can be formed from the other by rotation. If a necklace consists of 3 beads, use THE LEMMA to compute the number of distinct necklaces.

15. Repeat exercise 14 where necklaces have 4 beads each.

16. Repeat exercise 15 where we have three colors of beads.

17. Compute $cyc(\pi)$ for every permutation from exercise 5.

18. Encode every permutation from exercise 17 as $x_1^{b_1}x_2^{b_2}...x_k^{b_k}$.

19. Use the first Version of Polya's Theorem to compute the number of nonisomorphic graphs on 3 vertices.

20. Repeat exercise 19 for graphs on 4 vertices.

21. For the real glutton, repeat exercise 19 for graphs on 5 vertices.

22. Consider a cube in 3-space. There are eight vertices. The following symmetries correspond to permutations of these vertices.

a) the identity symmetry

b) rotations by $\pi/2$ radians around lines connecting the centers of opposite faces.

c) rotations by $\pi/4$ or $3\pi/4$ radians around the lines connecting the centers of opposite faces.

d) rotations by $\pi/2$ radians around lines connecting the midpoints of opposite edges.

e) rotations by $2\pi/3$ radians around lines connecting opposite vertices.

Encode each of these types of symmetries in the form $x_1^{b_1}x_2^{b_2}...x_8^{b_8}$. Determine the number of each type of symmetry, and write down the cycle index of this group of symmetries.

23. The number of permutations of $\{1, 2, ..., n\}$ with code $x_1^{b_1} x_2^{b_2} ... x_n^{b_n}$ is given by

$$\frac{n!}{b_1! b_2! ... b_n! 1^{b_1} 2^{b_2} 3^{b_3} ... n^{b_n}}$$

Verify this formula for $n = 5, b_1 = 1, b_2 = 2, b_3 = b_4 = b_5 = 0$ by enumerating all permutations of the proper type.

24. How many open four bead necklaces are there in which each bead is one of the colors b, r, or p, there is at least one p, and two necklaces are considered the same if they are identical or if one is the reversal of the other.

25. Repeat exercise 24 if the necklaces are closed and two are considered the same if one is a rotation of the other.

26. Repeat exercise 24 if the necklaces are closed and two are considered the same if one is a rotation, or a reflection of the other.

Chapter 5: Combinatorial Designs

This chapter begins our foray into the realm governed by the existence question. Given certain conditions, can we find a configuration of finite sets which meet the conditions? If we cannot, we wish to prove so. If we can, we'd like to be able to demonstrate that the configuration satisfies the conditions. If possible, we might even want to classify all possible configurations which meet the requirements.

For non-existence proofs, we will not necessarily need the power of abstract algebra. For constructions, and to aid in discussing classification we will. So in this chapter we start with two sections dedicated to finite fields.

This is followed by a section on the most basic configurations we will consider, Latin squares. The initial motivation for considering these configurations was to remove possible ambiguities which might negatively affect the collection of data from agricultural experiments.

The third section is devoted to introducing so-called balanced incomplete block designs, which were also used in the design of experiments.

In the last two sections we give some basic construction techniques for balanced incomplete block designs, and discuss connections with the classification of finite simple groups.

§5.1 Finite Prime Fields

Recall that $\mathbb{Z}$ denotes the integers, aka whole numbers. With respect to the operations of addition and multiplication the integers satisfy the following algebraic axioms:

1) and 5) Addition and Multiplication are associative, eg. $(ab)c = abc = a(bc)$ always.

2) and 6) Addition and Multiplication are commutative, eg. $a + b = b + a$ always.

3) and 7) There is an additive identity (0) and a multiplicative identity (1).

4) Every element has an additive inverse.

8) Multiplication distributes over addition.

The function mod $n : \mathbb{Z} \longrightarrow \{0, 1, 2, 3, ..., n-1\} := \mathbb{Z}_n$ takes as input any whole number a and outputs r, where $a = qn + r$, and $0 \leq r < n$. We can use this to define and addition and a multiplication on $\mathbb{Z}_n$ where $a +_n b = (a + b)\text{mod } n$, and $a \times_n b = (a \times b)\text{mod } n$.

Technically we're operating on the equivalence classes modulo n. So $[a] +_n [b] = [a+b]$, and $[a] \times_n [b] = [ab]$, where the convention is that we always use the standard system of distinct representatives for equivalence classes $\{[0], [1], [2], ...[n-1]\}$. However the equivalence class notation is a little cumbersome, so we"ll dispense with the technicalities to save ourselves a little grief. Also when the context is clear we'll drop the subscipts on the operations.

With these definitions $\mathbb{Z}_n = \{0, 1, 2, ..., n-1\}, +_n, \times_n$ also satisfies axioms 1-8 by inheritance and the fact that mod n is a function.

<u>Definition</u> A (commutative) ring is a set R with operations addition and multiplication which satisfies axioms 1)-8).

<u>Definition</u> A field is a set $\mathbb{F}$ with operations addition and multiplication which satisfy axioms 1-8 above, has $1 \neq 0$ and which satisfies the further axiom

9) Every nonzero element has a multiplicative inverse.

Equivalently $\mathbb{F}^* := \mathbb{F} - \{0\}$ is a group under multiplication.

Note that the integers do not form a field since, for example, $2^{-1} = \frac{1}{2} \notin \mathbb{Z}$.

Fact: If $\mathbb{F}$ is a field and $a \neq 0$, then $ab = ac$ implies $b = c$.

Proof: Since $a \neq 0$, $a^{-1} \in \mathbb{F}$. Thus $ab = ac$ means $a^{-1}ab = a^{-1}ac$. Whence $b = c$. ■

Fact: If $\mathbb{F}$ is a field and $ab = 0$, then $a = 0$, or $b = 0$.

Proof: If $a \neq 0$ and $ab = 0$, then $b = a^{-1}ab = a^{-1}0 = 0$. ■

Corollary If R is a set with addition and multiplication satisfying axioms $1 - 5, 7$ and 8 and there are nonzero elements $a, b \in R$ with $ab = 0$, then R is not a field.

Theorem: Let $n > 1$ be an integer. $\mathbb{Z}_n$ is a field iff n is prime.

Proof: For the reverse implication it suffices to show that every nonzero element has a multiplicative inverse. So let $a \in \mathbb{Z}_n - \{0\}$, where n is prime. Because n is prime $gcd(a, n) = 1$. There are therefore integers s and t so that $as + tn = 1$. Therefore $a \times_n s = 1$.

Conversely, since $n > 1$, if n is not prime, then it is composite. Therefore there exist integers a, b with $1 < a, b < n$ and $n = ab$. Therefore $a \times_n b = 0$. Therefore $\mathbb{Z}_n$ is not a field by the corollary above. ■

Fact: If $a, b \in \mathbb{F}$ a field and $a \neq 0$, then there is a unique solution in $\mathbb{F}$ to $ax = b$, namely $x = a^{-1}b$.

Corollary: If $a \in \mathbb{F} - \{0\}$, where $\mathbb{F}$ is a field, then a^{-1} is unique.

Fact: If $\mathbb{F}$ is a field then $x^2 - 1 = 0$ has at most 2 solutions in $\mathbb{F}$.

Proof: In any field $x^2 - 1 = (x - 1)(x + 1)$, so if $x^2 - 1 = 0$, either $x + 1 = 0$ or $x - 1 = 0$. Since $1 = -1$ in $\mathbb{Z}_2$ these solutions do not need to be distinct. ■

Corollary: For p an odd prime $\mathbb{Z}_p - \{-1, 0, 1\}$ is the disjoint union of $\frac{p-3}{2}$ sets of the form $\{a, a^{-1}\}$.

Proof: Any $a \in \mathbb{Z}_p - \{-1, 0, 1\}$ has a unique multiplicative inverse. a^{-1} satisfies that $(a^{-1})^{-1} = a$. Finally $a = a^{-1}$ is equivalent to $a^2 = 1$. Which is equivalent to a being a solution to $x^2 - 1 = 0$. ■

Wilson's Theorem: If p is prime, then $(p - 1)! \equiv -1 (\text{mod } p)$.

Proof: The theorem is trivial for $p = 2$ so suppose that p is an odd prime. Let I be a subset of $\mathbb{Z}_p$ so that $\{1\} \cup \{-1\} \cup_{a \in I} \{a, a^{-1}\}$ is a partition of $\mathbb{Z}_p^*$. Then

$$(p-1)! = (p-1)(p-2)... \cdot 2 \cdot 1 \equiv -1 \cdot 1 \cdot \prod_{a \in I} aa^{-1} \equiv -1 \cdot 1 \cdot 1^{\frac{p-3}{2}} \equiv -1 (\text{mod } p)$$ ■

Fact: If $a \in \mathbb{Z}_p^*$, then for each $k \in \mathbb{Z}_p^*$ there is a unique $j \in \mathbb{Z}_p^*$ with $ak \equiv j (\text{mod } p)$.

Proof: If $ak \equiv al (\text{mod } p)$, then multiplying both sides by a^{-1} gives $k \equiv l (\text{mod } p)$ ■

Fermat's Little Theorem: If p is prime and $p \not| a$, then $a^{p-1} \equiv 1(\text{mod } p)$.

Proof: WLOG $p \neq 2$. Now $a \cdot 2a \cdot 3a \cdot ... \cdot (p-1)a \equiv 1 \cdot 2 \cdot 3 \cdot ... \cdot (p-1)(\text{mod } p)$ by the previous fact after rearranging terms. We rewrite this as $a^{p-1}[(p-1)!] \equiv (p-1)!(\text{mod } p)$. We deduce the conclusion by cancelling $-1 \equiv (p-1)!(\text{mod } p)$ from both sides. ∎

Definition: For integers n, a $ord_n a$ is the least positive integer k so that $a^k \equiv 1(\text{mod } n)$. Notice that $ord_n a$ exists only when $gcd(a, n) = 1$

Theorem: $a^m \equiv 1(\text{mod } n)$ iff $k = ord_n a | m$.

Proof: ($\Longrightarrow$) If $a^m \equiv 1(\text{mod } n)$ there are unique integers q, r with $m = kq + r$ and $0 \leq r < k$. Now $1 \equiv a^m \equiv a^{kq+r} \equiv a^{kq}a^r \equiv (a^k)^q a^r \equiv 1^q a^r \equiv a^r(\text{mod } n)$. By the minimality of k we must have $r = 0$.
($\Longleftarrow$) If $k|m$, then $m = kq$ for some $q \in \mathbb{Z}$. Thus $a^m = a^{kq} \equiv (a^k)^q \equiv 1^q \equiv 1(\text{mod } n)$ ∎

Corollary: If p is prime $ord_p a|(p-1)$ for all $a \in \mathbb{Z}_p^*$.

Definition: An integer a is a primitive root modulo a prime p if $ord_p a = p - 1$.

Theorem:(Lagrange) If p is a prime integer and f is a polynomial with integral coefficients of degree n and p does not divide f's lead coefficient, then $f(x) \equiv 0(\text{mod } p)$ has at most n incongruent solutions mod p.

Proof: We use induction on n. If $f(x) = a_1 x + a_0$ where $p \not| a_1$ then $f(x) \equiv 0(\text{mod } p)$ is equivalent to $a_1 x \equiv -a_0(\text{mod } p)$. We are done by previous theorems. So suppose that the theorem is true for all polynomials with integral coefficients of degree less than or equal to $n-1$ for which p does not divide the lead coefficient. Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + .. + a_1 x + a_0$, where $p \not| a_n$.

If $f(x) \equiv 0(\text{mod } p)$ has no solutions, then we're done.

Else if a is a solution write $f(x) = (x-a)q(x) + r(x)$, where $r(x) = 0$, or the degree of $r(x)$ is less than the degree of $x - a$ (which is one). Thus r is a constant polynomial. Notice that the degree of $q(x) = n - 1$.

Now $f(a) = (a-a)q(a) + r = r \equiv 0(\text{mod } p)$. So $f(x) \equiv (x-a)q(x)(\text{mod } p)$.

So if b is another solution to $f(x) \equiv 0(\text{mod } p)$, then either $(b-a) \equiv 0(\text{mod } p)$ or $q(b) \equiv 0(\text{mod } p)$. Thus any solution $b \not\equiv a(\text{mod } p)$ is a solution to $q(x) \equiv 0(\text{mod } p)$. We are now done by inductive hypothesis. ∎

Theorem: Let p be prime and d a positive divisor of $p - 1$. $x^d - 1 \equiv 0(\text{mod } p)$ has exactly d incongruent solutions modulo p.

Proof: Since $d|(p-1)$, there is an integer e with $p - 1 = de$. So $x^{p-1} - 1 = (x^d - 1)(x^{d(e-1)} + x^{d(e-2)} + ... + x^d + 1)$. Call the second polynomial $q(x)$. By Fermat's Little Theorem $x^{p-1} - 1 \equiv 0(\text{mod } p)$ has exactly $p - 1$ incongruent solutions modulo p. Each of these is either a solution of $x^d - 1 \equiv 0(\text{mod } p)$, or $q(x) \equiv 0(\text{mod } p)$.

By Lagrange's Theorem the number of incongruent solutions to $x^d - 1 \equiv 0(\text{mod } p)$ is less than or equal to d. Also the number of incongruent solutions to $q(x) \equiv 0(\text{mod } p)$ is less than or equal to $(p-1) - d = d(e-1) = deg(q(x))$. Therefore the number of incongruent solutions to $x^d - 1 \equiv 0(\text{mod } p)$ is at least $(p-1) - [(p-1) - d] = d$. Thus the number of incongruent solutions to $x^d - 1 \equiv 0(\text{mod } p)$ is exactly d. ∎

Euler's φ-function counts the number of positive whole numbers not greater than a given one which are also relatively prime to it.

Theorem: Let p be prime and d be a positive divisor of $p-1$, there are exactly $\varphi(d)$ incongruent integers with order d modulo p.

Proof: First we need a lemma due to K.F. Gauss.

Lemma: For a positive integer n, $\sum\limits_{\substack{d|n \\ d>0}} \varphi(d) = n$

Proof of lemma: Let $d|n, d>0$ and $S_d = \{m \in \mathbb{Z} | 1 \leq m \leq n \text{ and } gcd(m,n) = d\}$. Recall that $gcd(m,n) = d$ iff $gcd(\frac{m}{d}, \frac{n}{d}) = 1$. Thus $|S_d| = \varphi(\frac{n}{d})$.

Every integer m with $1 \leq m \leq n$ is in exactly one set S_d where d is a positive divisor of n, therefore $\sum\limits_{\substack{d|n \\ d>0}} \varphi(\dfrac{n}{d}) = n$.

But since $d|n$ and $d > 0, n = dk$, for some positive integer k which is also a divisor of n. Therefore $\sum\limits_{\substack{d|n \\ d>0}} \varphi(\dfrac{n}{d}) = \sum\limits_{\substack{k|n \\ k>0}} \varphi(k) = n$. ■

Now if we let $f(d)$ denote the number of integers between 1 and $p-1$ inclusive which have order d modulo p we have $p-1 = \sum\limits_{\substack{d|(p-1) \\ d>0}} f(d) = \sum\limits_{\substack{d|(p-1) \\ d>0}} \varphi(d)$.

To show $f(d) = \varphi(d)$ for all positive divisors d of $p-1$ it suffices to show $f(d) \leq \varphi(d)$ for all positive divisors d of $p-1$, since we could not have strict inequality anywhere.

So if $f(d) = 0$ for some positive divisor d of $p-1$, we are done since $\varphi(d) > 0$ for $d > 0$.

Else $f(d) > 0$ and there is $a \in \mathbb{Z}_p^*$ with $ord_p a = d$. The definition of order implies that $a, a^2, ..., a^d$ are all incongruent (else $ord_p a < d$). Finally for $1 \leq j \leq d$ $(a^j)^d \equiv a^{jd} \equiv a^{dj} \equiv (a^d)^j \equiv 1^j \equiv 1(\text{mod } p)$. So the d powers of a above are all d incongruent solutions to $x^d - 1 \equiv 0(\text{mod } p)$.

The proof of the theorem is completed by the following lemma and corollary.

Lemma: For $i \in \mathbb{Z}$ $ord_n(a^i) = \dfrac{ord_n a}{gcd(ord_n a, i)}$.

Proof: Put $d = gcd(ord_n a, i)$ and $k = ord_n a$. Write $k = db$ and $i = dc$ where $b, c \in \mathbb{Z}$. Notice that $gcd(b,c) = 1$ and $b = \dfrac{k}{d} = \dfrac{ord_n a}{gcd(ord_n a, i)}$.

Now $(a^i)^b \equiv (a^{dc})^b \equiv a^{bcd} \equiv (a^{bd})^c \equiv (a^k)^c \equiv 1^c \equiv 1(\text{mod } n)$. Therefore $ord_n(a^i)|b$.

Also $a^{i \cdot ord_n(a^i)} \equiv (a^i)^{ord_n a^i} \equiv 1(\text{mod } n)$, so $k|i \cdot ord_n(a^i)$. Which is to say that $db|dc \cdot ord_n(a^i)$. Therefore $b|c \cdot ord_n(a^i)$. Since $gcd(b,c) = 1$ we have $b|ord_n(a^i)$.

Since $b|ord_n(a^i)$ and vice versa, and they are both positive integers, they are equal. ■

Corollary: $ord_n(a^i) = ord_n a$ iff $gcd(ord_n a, i) = 1$.

The theorem has now been proved. ■

Corollary: (Primitive Root Theorem) There are exactly $\varphi(p-1)$ incongruent primitive roots modulo a prime p.

We wish to generalize the above set of theorems to a larger class of objects. So far what we have done is prove a body of theorems for the so-called finite prime fields, which are those which have a prime number of elements. In general we can construct a finite field of order $q = p^n$ for any prime p and positive integer n.

§5.2 Finite Fields

To begin with the characteristic of a ring (or field) is zero if $m \cdot \mathbf{1} = 0$ implies $m = 0$, otherwise it is the least positive integer m so that $m \cdot \mathbf{1} = \mathbf{1} + \mathbf{1} + ... + \mathbf{1} = 0$ in the ring.

Theorem: If $\mathbb{F}$ is a finite field then its characteristic is prime, and $\mathbb{F}$ contains a subfield isomorphic to $\mathbb{Z}_p$.

Proof: Let $\mathbf{1} \in \mathbb{F}$, and $n = |\mathbb{F}|$. Then $\mathbf{1}, 2 \cdot \mathbf{1}, ..., (n+1) \cdot \mathbf{1}$ are not all distinct. Therefore there are positive integers i and j with $i < j$ and $i \cdot \mathbf{1} = j \cdot \mathbf{1}$. Thus $(j-i) \cdot \mathbf{1} = 0$ with $j - i > 0$. Thus a finite field does not have characteristic zero.

Let m be the characteristic of $\mathbb{F}$. So $\mathbb{F}_0 = \{\mathbf{1}, 2 \cdot \mathbf{1}, 3 \cdot \mathbf{1}, ..., m \cdot \mathbf{1} = 0\}$ are all distinct in $\mathbb{F}$ (if not, m is not be the least positive integer with $m \cdot \mathbf{1} = 0$). Notice that $\mathbb{F}_0$ is closed under addition and multiplication. Also $\mathbb{F}_0$ satisfies the field axioms by inheritance from $\mathbb{F}$. Therefore $\mathbb{F}_0$ is a subfield of $\mathbb{F}$.

Define a ring homomorphism (preserves + and $\cdot$) from $\mathbb{F}_0$ to $\mathbb{Z}_m$ by $f(k \cdot \mathbf{1}) = k$. Since f is clearly bijective and preserves operations, $\mathbb{Z}_m$ must be a field. Thus m is prime and we are done. ■

Corollary: If $\mathbb{F}$ is a finite field with characteristic p, then $p \cdot x = 0$ for all $x \in \mathbb{F}$.

Proof: $p \cdot x = x + x + ... + x = x(\mathbf{1} + \mathbf{1} + ... + \mathbf{1}) = x \cdot p \cdot \mathbf{1} = x \cdot 0 = 0$ ■

Corollary: If $\mathbb{F}$ is a finite field with q elements and characteristic p, then $q = p^n$ for some $n \in \mathbb{N}$.

Proof: Since $\mathbb{F}$ is finite we can form a set $S = \{x_1, x_2, ..., x_n\}$ with a minimal number of elements so that every element $x \in \mathbb{F}$ is a linear combination $a_1x_1 + a_2x_2 + ... + a_nx_n$, where $a_1, a_2, ..., a_n \in \mathbb{Z}_p$ (from the previous corollary). No element of S can be written as a linear combination of the others, or else S does not have a minimal number of elements.

For some $x \in \mathbb{F}$ suppose that $x = a_1x_1 + a_2x_2 + ... + a_nx_n$ and $x = b_1x_1 + b_2x_2 + ... + b_nx_n$ where $a_i, b_i \in \mathbb{Z}_p$. Suppose that there exists i so that $a_j = b_j$ for all $j > i$, but $a_i \neq b_i$. Then $0 = (a_1 - b_1)x_1 + (a_2 - b_2)x_2 + ... + (a_i - b_i)x_i$, with $c = a_i - b_i \neq 0 \in \mathbb{Z}_p$. Put $d = c^{-1}$ in $\mathbb{Z}_p$. Then $x_i = -d[(a_1 - b_1)x_1 + (a_2 - b_2)x_2 + ... + (a_{i-1} - b_{i-1})x_{i-1}]$ a contradiction. Thus every $x \in \mathbb{F}$ is writable in exactly one way as a $\mathbb{Z}_p$-linear combination of the x_k's. Every $\mathbb{Z}_p$-linear combination of elements in S is in $\mathbb{F}$ because $\mathbb{F}$ is closed under addition. We therefore have that $|\mathbb{F}| = p^n$. ■

In order to construct finite fields of order p^n where p is prime and $n > 1$ we must first consider polynomials.

A natural power function is of the form x^m where $m \in \mathbb{N} = \{0, 1, 2, ...\}$. For a ring, R, a polynomial over R is an R-linear combination of a finite number of natural power functions. For a nonzero polynomial the largest natural number n for which the coefficient of x^n is nonzero is the degree of the polynomial. When a polynomial of degree n has 1 as the coefficient on x^n it is called a monic polynomial. The degree of the zero polynomial is not really defined, but can be taken to be -1 or even $-\infty$. The set of all polynomials over R is denoted $R[x]$. We may add two polynomials by adding the respective coefficients and multiply polynomials by convoluting the coefficients.

Fact: $R[x]$ is a ring with respect to the addition and multiplication described above. Moreover $R[x]$ is commutative if R is.

Proof: Omitted for esthetic reasons. ■

When the ring of coefficients R is a field special things happen. For example

Theorem: (The Division Algorithm for Polynomials) If $\mathbb{F}$ is a field, $f, g \in \mathbb{F}[x]$ and $g \neq 0$ then there are unique polynomials q and r so that $f = qg + r$ and either $r = 0$ or the degree of r is strictly less than the degree of g.

Proof: Use induction on the degree of f.

We write $g|f$ and say g <u>divides</u> f exactly when $r = 0$ from the division algorithm.

If $c \in \mathbb{F}$ and $f(x) = \sum_{k=0}^{n} a_i x^i$, then $f(c) = \sum_{k=0}^{n} a_i c^i$ is the <u>value</u> of f at c. We say that c is a <u>root</u> of f when $f(c) = 0$.

Theorem: (Factor Theorem) If $f \in \mathbb{F}[x]-\{0\}$, where $\mathbb{F}$ is a field, then $f(c) = 0$ iff $(x-c)|f(x)$.

Proof: By the division algorithm $f(x) \overset{!}{=} q(x)(x-c) + r(x)$ where $r(x)$ is zero or has degree less than 1. In any case $r(x)$ is a constant $k \in \mathbb{F}$. When we evaluate at $c = 0$ we find $f(c) = k$. Thus $f(c) = 0$ iff $r = 0$. ■

As a corollary (provable by induction) we have

Theorem: (Lagrange again) If $\mathbb{F}$ is a field and $f(x) \in \mathbb{F}[x]$ has degree n, then f has at most n roots in $\mathbb{F}$.

We are mostly interested in the extreme behavior with respect to the previous theorem. For example one can show that $x^2 + x + 1$ has no roots in $\mathbb{Z}_2[x]$. On the other hand when p is prime $x^p - x$ has p distinct roots in $\mathbb{Z}_p[x]$ by Fermat's Little Theorem. When a polynomial f of degree n has exactly n distinct roots in $\mathbb{F}$ we say that $\mathbb{F}$ <u>splits</u> f, or that f splits in $\mathbb{F}$.

A polynomial $f \in \mathbb{F}[x]$ is <u>irreducible</u> over $\mathbb{F}$ if $f = gh$ implies either $g \in \mathbb{F}$ or $h \in \mathbb{F}$. Notice that an irreducible polynomial has no roots in $\mathbb{F}$. The converse is false as demonstrated by $x^4 + x^3 + x + 2 = (x^2 + x + 2)(x^2 + 1)$ in $\mathbb{Z}_3[x]$. However the following lemma is true

Lemma: If $\mathbb{F}$ is a field, f is an irreducible monic polynomial in $\mathbb{F}[x]$, g is monic and $g|f$, then $g = 1$ or $g = f$.

Thus irreducible monic polynomials in $\mathbb{F}[x]$, where $\mathbb{F}$ is a field, are analogous to primes in $\mathbb{Z}$.

We are now ready to generalize modular arithmetic. In $\mathbb{F}[x]$ we write $g \equiv h(\text{mod } f)$ in case $f|(g-h)$. This relation is called congruence modulo f.

<u>Fact:</u> If $f(x) \in \mathbb{F}[x] - \{0\}$, then congruence modulo f is an equivalence relation on $\mathbb{F}[x]$.

We can therefore mimic the definition of $+_m$ and $\times_m$ for $\mathbb{Z}_m$ to build binary operations on the set of equivalence classes modulo f in $\mathbb{F}[x]$, where $\mathbb{F}$ is a field. If f has degree n then by the division algorithm $\mathbb{K} = \{\sum_{i=0}^{n-1} c_i x^i | c_i \in \mathbb{F}\}$ is a system of distinct representatives for the equivalence classes modulo f. We define the binary operations $+_f$ and $\times_f$ on $\mathbb{K}$ by $g \times_f h = r$

when $gh \overset{!}{=} fq + r$ and $r = 0$ or $deg(r) < deg(f)$ and $g +_f h = r$ when $(g + h) \overset{!}{=} fq + r$ and $r = 0$ or $deg(r) < deg(f)$.

When $\mathbb{F}$ is a field, $\mathbb{K}$ inherits much of the structure of $\mathbb{F}[x]$, and is always a commutative ring. The following theorem is the natural generalization of $\mathbb{Z}_m$ is a field iff m is prime.

Theorem: The set $\mathbb{K}$ as described above is a field with respect to the operations $+_f$ and $\times_f$ iff f is irreducible in $\mathbb{F}[x]$.

Notice that when $\mathbb{F} = \mathbb{Z}_p$, where p is prime and f is an irreducible polynomial of degree n in $\mathbb{Z}_p[x]$, $\mathbb{K}$ is a finite field of order p^n usually denoted $GF(p^n)$.

We sometimes use the notation $\mathbb{K} = \mathbb{Z}_p[x]/\langle f \rangle$.

Another natural generalization to draw is

Theorem: (Fermat's Larger Theorem) If $a \in GF(q) - \{0\}$ (where $GF(q)$ is the residues modulo f in $\mathbb{Z}_p[x]$ and $q = p^n$), then $a^{p^n - 1} \equiv 1(\text{mod } f)$.

This can be restated as: If $a \in GF(q)$, then $a^{p^n} \equiv a(\text{mod } f)$.

Another restatement is: If $a \in GF(q)$, then a is a root of $x^{p^n} - x$ in $GF(q)[x]$.

Yet another restatement is: $x^{p^n} - x$ splits in $GF(p^n)[x]$.

Finally we may generalize the primitive root theorem to

Theorem: (Primitive Root Theorem) If $\mathbb{F}$ is a finite field of order $q = p^n$ where p is prime, then there exactly $\varphi(q - 1)$ elements α in $\mathbb{F}$ with $ord_f \alpha = q - 1$. These elements are called primitive roots.

If $h(x) \in \mathbb{Z}_p[x]$ is monic and irreducible of degree n and α is a root of h, then we will call h a primitive polynomial, if α is a primitive root. Otherwise h is not primitive.

For example when $p = 3$ and $n = 2$, $h(x) = x^2 + x + 2$ is primitive. Indeed h is irreducible in $\mathbb{Z}_3[x]$ since it has no roots, and therefore no linear factors. Moreover if $h(\alpha) = 0$, then $\alpha^2 + \alpha + 2 = 0$, or equivalently $\alpha^2 = 2\alpha + 1$. Thus the powers of α are

power of α	element
α^0	1
α^1	α
α^2	$2\alpha + 1$
α^3	$2\alpha + 2$
α^4	2
α^5	2α
α^6	$\alpha + 2$
α^7	$\alpha + 1$
α^8	1

On the other hand when $p = 3$ and $n = 2, g(x) = x^2 + 1$ is not primitive. Here g is irreducible in $\mathbb{Z}_3[x]$ since it has no roots, and therefore no linear factors. However if $g(\beta) = 0$, we get the condition $\beta^2 = -1 = 2$. Whence $\beta^4 = (-1)^2 = 1$. So β does not have order 8.

This does not mean that we cannot use g to build a field of order 9, it just means that no root of g will generate the multiplicative group of the field. However one can show that $\beta + 1$ will be a primitive root.

In general to build $GF(q)$, where $q = p^n$ we factor $x^q - x$ over $\mathbb{Z}_p[x]$. Every irreducible monic polynomial of degree n over $\mathbb{Z}_p[x]$ will appear in this factorization since such a polynomial splits completely in $GF(q)$. Moreover it can be shown that every monic irreducible degree n polynomial in $\mathbb{Z}_p[x]$ has n distinct roots all of which have the same order in $GF(q)^*$. Therefore there will be in general $\varphi(q-1)/n$ monic irreducible primitive polynomials of degree n in $\mathbb{Z}_p[x]$. Any resulting monic irreducible degree n polynomial can be used to build $GF(q)$.

For example when $p=2, n=3$ and $q=2^3=8$, we have $q-1=7$ and $\varphi(q-1)=6$. So there are $6/3=2$ monic irreducible cubics in $\mathbb{Z}_2[x]$ both of which are primitive. We have in $\mathbb{Z}_2[x]$ that $x^8-x=x(x-1)(x^3+x+1)(x^3+x^2+1)$.

As another example, when $p=3, n=2$ and $q=9$, $\varphi(q-1)=4$ and there are $4/2=2$ monic irreducible primitive quadratic polynomials in $\mathbb{Z}_3[x]$. We have $x^9-x=x(x-1)(x+1)(x^2+1)(x^2+x+2)(x^2+2x+2)$.

The following helpful facts can be used to minimize the amount of work that it takes to factor x^q-x into irreducibles in $\mathbb{Z}_p[x]$.

Fact: If f is an irreducible factor of $x^{p^n}-x$ in $\mathbb{Z}_p[x]$, then *deg*(f) divides n. Especially *deg*$(f)\leq n$.

Proof: Let $\mathbb{F}=GF(q)$. Denote the degree of f by m. Then $\mathbb{K}=\mathbb{Z}_p[x]/\langle f\rangle$ is a finite sub-field of $\mathbb{F}$. If β is a primitive root for $\mathbb{K}$, then it has order p^m-1 in $\mathbb{K}$, and therefore in $\mathbb{F}$. Therefore p^m-1 divides p^n-1, which by the following lemma is equivalent to $m|n$. ■

Lemma: If a, m and n are positive integers with $a>1$, then $(a^m-1)|(a^n-1)$ iff $m|n$.

Proof: First $m\leq n$ is necessary, so write $n=mq+r$, where $0\leq r<m$ and $q,r\in\mathbb{Z}$.

Then $a^n-1=(a^m-1)[a^{(q-1)m+r}+a^{(q-2)m+r}+...+a^m+r]+a^r-1$ where $0\leq a^r-1<a^m-1$.

So $(a^m-1)|(a^n-1)$ iff $a^r-1=0$ iff $r=0$ iff $m|n$. ■

So by the way if $m|n$ every irreducible monic polynomial of degree m in $\mathbb{Z}_p[x]$ will be an irreducible factor of $x^{p^n}-x$ in $\mathbb{Z}_p[x]$.

For example $x^{27}-x$ will have 3 monic irreducible linear factors (x, $x-1$, and $x-2=x+1$). Every other irreducible monic factor must be a cubic. There are therefore 8 monic irreducible cubics in $\mathbb{Z}_3[x]$, only 4 of which are primitive.

Finally if α is a root of the monic polynomial $f(x)=x^m+a_{m-1}x^{m-1}+...+a_1x+a_0$ where $a_0\neq 0$, then α^{-1} is a root of $1+a_{m-1}x+...+a_{m-i}x^i+...+a_0x^m$. Equivalently α^{-1} is a root of $x^m+\frac{a_1}{a_0}x^{m-1}+...+\frac{a_{m-i}}{a_0}x^i+...+\frac{1}{a_0}=g(x)$. We call $g(x)$ the reciprocal polynomial of $f(x)$. A polynomial can be self-reciprocal, i.e. equal to its own reciprocal. Since the multiplicative group of a finite field is cyclic, and the order of any element in a cyclic group is the same as its inverse's order we have that if f is a monic irreducible factor of x^q-x, then so is g.

§5.3 Latin Squares

Suppose we want to conduct an experiment to determine which of 5 varieties of seed gives the best yield. As a first pass we might test each variety in each of 5 different soil types. By collecting the seeds in blocks this really only requires 5 separate tests, although 25 data sets must be maintained. For a larger number of seeds and/or a larger number of soil types, this could get expensive. Also this plan doesn't take other possible factors into account. Still it gives the idea of a so-called factorial design - just try every single possibility.

As a second pass we might want to test our 5 varieties not only in 5 soil types, but also using 5 different brands of pesticide. To ensure no variety of seed receives preferential treatment - that they should all be treated the same, we should test each variety with each possible ordered pair (soil type, pesticide brand). Similarly every pesticide brand should be involved in a test with each ordered pair (seed variety, soil type). It turns out that in this case what we need to organize our experiment is a Latin Square.

A Latin Square of order n is an $n \times n$ array with entries from a set of size n, so that every entry appears exactly once in each row and exactly once in each column.

If we label our pesticide brands A, B, C, D and E, label 5 columns with our seed varieties, and label 5 rows with the soil types, we might wind up with

$$\begin{array}{ccccc} A & B & C & D & E \\ B & C & D & E & A \\ C & D & E & A & B \\ D & E & A & B & C \\ E & A & B & C & D \end{array}$$

So now with our 5 experiments we can take more effects into account.

But suppose that we also want to take into account the brand of herbicide used. Say we have five herbicide brands. Then we can build another Latin Square of order 5 with columns labelled by seed varieties, and rows labelled by soil types, where the entries are the herbicide brands. But is it possible to preserve fairness? That is can we arrange it so that this table in concert with the previous table has the property that in our tests we can give every ordered pair of herbicide and pesticide a fair shake. The answer is yes, by using a pair of orthogonal Latin squares.

Two Latin squares of order n are orthogonal if every possible ordered pair of entries occurs exactly once. A set of Latin squares of order n is mutually orthogonal if they are pairwise orthogonal. We say we have a set of MOLS.

To save ourselves some trouble we will also use A, B, C, D and E to label our pesticide brands. The following Latin square of order 5 is orthogonal to the one above

$$\begin{array}{ccccc} A & B & C & D & E \\ C & D & E & A & B \\ E & A & B & C & D \\ B & C & D & E & A \\ D & E & A & B & C \end{array}$$

In general we will want to know the answers to the following questions.

1) Given v and r can we construct a set of r MOLS of order v?

2) Given v, what is the maximal number of MOLS in a set?

3) How can we verify that two Latin squares of order v are orthogonal?

4) What if the number of varieties is not equal to the number of soil types?

To answer the first two questions we begin by adopting the convention that all of our squares of order n will have entries from just one n-set

Theorem: *If there are r MOLS of order n, then $r \leq n-1$.*

Proof: Suppose that $A^{(1)}, A^{(2)}, ..., A^{(r)}$ are r MOLS of order n. For the purpose of proving the theorem let's suppose the entries are from $\{1, 2, ..., n\}$. Denote by $a_{i,j}^{(p)}$ the i, jth entry of $A^{(p)}$, $p = 1, 2, ..., r$.

In $A^{(1)}$ permute $\{1, 2, ..., n\}$ using the transposition $(1k)$ if necessary, so that $a_{1,1}^{(1)} = 1$. Interchanging 1 and k throughout keeps $A^{(1)}$ a Latin square on $\{1, 2, ..., n\}$ and it is still orthogonal with the remaining squares in the set: If before $(a_{i,j}^{(1)}, a_{i,j}^{(p)})$ was (k, l), it's now $(1, l)$, and if it was $(1, l)$, it's now (k, l). This process is called normalization.

Continuing, we can normalize so that $a_{1,k}^{(j)} = k$, for $k = 1, 2, ..., n$, and $j = 1, 2, ..., r$. At which point we say that the set is in standard order.

Now since every square has a 1 in the 1,1 position, $a_{2,1}^{(p)} \neq 1$, for $p = 1, 2, ..., r$. Also because $(i, i) = (a_{1,i}^{(p)}, a_{1,i}^{(q)})$ we must have that $a_{2,1}^{(p)} \neq a_{2,1}^{(q)}$ for $p \neq q$. So $\{a_{2,1}^{(1)}, a_{2,1}^{(2)}, ..., a_{2,1}^{(r)}\}$ is an r-subset of $\{2, 3, 4, ..., n\}$. Therefore $r \leq n-1$. ■

A set of $n-1$ MOLS of order n is called a complete set.

Theorem: *There is a complete set of MOLS of order p^k when p is prime and $k > 0$.*

Proof: Put $n = p^d$, where p is prime and $d > 0$. Let $\mathbb{F} = GF(p^d)$. Say $\mathbb{F} = \{b_0 = 0, b_1 = 1, ..., b_{p^d-1}\}$. For $\gamma \in \mathbb{F}^*$ build $A^{(\gamma)}$ by $a_{i,j}^{(\gamma)} = \gamma \cdot b_i + b_j$.

Lemma: $A^{(\gamma)}$ so constructed is a Latin square of order n, for all $\gamma \in F^*$.

Proof of lemma: 1) $a_{i,j}^{(\gamma)} = a_{i,k}^{(\gamma)}$ iff $\gamma \cdot b_i + b_j = \gamma \cdot b_i + b_k$ iff $b_j = b_k$, by additive cancellation, iff $j = k$.
2) $a_{i,j}^{(\gamma)} = a_{k,j}^{(\gamma)}$ iff $\gamma \cdot b_i + b_j = \gamma \cdot b_k + b_j$ iff $\gamma \cdot b_i = \gamma \cdot b_k$ iff $b_i = b_k$, by multiplicative cancellation since $\gamma \neq 0$, iff $i = k$.
3) Let $x \in \mathbb{F}$ a) For i given, $x = a_{i,j}^{(\gamma)}$, where $b_j = x - \gamma \cdot b_i$. b) For j given, $x = a_{i,j}^{(\gamma)}$, where $b_i = \gamma^{-1}(x - b_j)$. ■

Lemma: *If $\gamma, \delta \in \mathbb{F}^*$, with $\gamma \neq \delta$, then $A^{(\gamma)}$, and $A^{(\delta)}$ are orthogonal.*

Proof: Let $(x, y) \in \mathbb{F} \times \mathbb{F}$. Then $(x, y) = (a_{i,j}^{(\gamma)}, a_{i,j}^{(\delta)})$ iff $x = \gamma \cdot b_i + b_j$, and $y = \delta \cdot b_i + b_j$ iff $(x - y) = \gamma \cdot b_i - \delta \cdot b_i = (\gamma - \delta)b_i$ iff $b_i = \frac{x-y}{\gamma-\delta}$. So i is completely determined given x and y, and j is now determined by the previous lemma. Therefore every ordered pair from $\mathbb{F} \times \mathbb{F}$ occurs at least once. None can occur more than once by tightness. ■

The theorem is proved. ■

So for prime powers the Latin square problem is nicely closed. For every prime power we have as many Latin squares as we could hope to have, and no more.

For integers which are not prime powers, the story is quite different. We begin here with the problem of the 36 officers: We are given 36 officers, six officers from each of six different ranks, and also six officers from each of six different regiments. We are to find a 6 × 6 square formation so that each row and column contains one and only one officer of each rank and one and only one officer from each regiment, and there is only one officer from each regiment of each rank. So the ranks give a Latin square, and the regiments, give an orthogonal Latin square.

In 1782, Leonhard Euler conjectured that this could not be done. In fact he conjectured that there was not a pair of MOLS for any positive whole number which was twice an odd number.

In 1900, a mathematician named John Tarry systematically checked all 9,408 pairs of normalized Latin squares of order 6 and showed that Euler was correct for this case. Normalization was an important part of Tarry's proof, since there are 812,851,200 pairs of Latin squares of order 6.

However in 1960, a trio of mathematicians proved that Euler was wrong in general.

Theorem: (Bose, Shrikhande, Parker) *If $n > 6$ is twice an odd number, then there is a pair of MOLS of order n.*

The proof of this theorem goes beyond the scope of our course, but we can discuss some of the tools that they had at hand, and why this was an important problem.

One of the most important construction techniques is due to MacNeish. In 1922 he used what is commonly called the Kronecker product of matrices to build larger Latin squares from smaller ones. Given two square arrays, A of side m and B of side n, we can build a square array of side mn whose entries are ordered pairs (a, b). The first coordinate is the same for the $n \times n$ block which is the intersection of the first n rows, and the first n columns. The second coordinates in this block are simply the elements of B. In general the first coordinate for the $n \times n$ block at the intersection of the $(km+1)$th through $(km+n)$th columns with the $(jm+1)$th through $(jm+n)$th rows is the j,kth entry of A, while the second coordinates are the entries of B. We will label the new square $A \otimes B$.

Lemma: *If A is a Latin square of order m, and B is a Latin square of order n, then $A \otimes B$ is a Latin square of order mn.*

Proof: Exercise. ■

Theorem:(MacNeish) *If there are r MOLS of order m and another r MOLS of order n, then there are r MOLS of order mn.*

Proof: Let $A^{(1)}, A^{(2)}, ..., A^{(r)}$ be r MOLS of order m, and let $B^{(1)}, B^{(2)}, ..., B^{(r)}$ be r MOLS of order n. Put $C^{(i)} = A^{(i)} \otimes B^{(i)}$, $i = 1, 2, ..., r$. Then the $C^{(i)}$'s form a set of r MOLS of order mn. The remainder of the proof is left as an exercise. ■

Corollary: *Suppose that $n > 1$ has prime factorization $p_1^{e_1} p_2^{e_2} ... p_s^{e_s}$, where all exponents are positive whole numbers. Let r be the smallest of quantities $p_i^{e_i} - 1, i = 1, 2, ..., s$. Then there are at least r MOLS of order n.*

Corollary: *If $n > 1$ has prime factorization $n = 2^e p_1^{e_1} p_2^{e_2} ... p_s^{e_s}$, where $e \neq 1$ and the p_i's are odd primes whose exponents are positive integers, then there is a pair of MOLS of order n.*

So Bose, Shrikhande and Parker really had only to show existence of a pair of MOLS of order $2p$, for all odd primes $p > 3$. This noteworthy accomplishment touched off a flurry of research in this area which lasted a good twenty years.

There are however many unanswered questions here. For example, for $n = 10$ are there three MOLS of order 10? What is the largest value of r so that there is a set of r MOLS of order 22?

Finally we mention the following, which may have been proven recently:

Conjecture: *If there is a complete set of MOLS of order n, then n is a prime power.*

The next section deals with our fourth question: What if the number of varieities is not equal to the number of soil types?

§5.4 Introduction to Balanced Incomplete Block Designs

In the preceding section we considered running experiments where we were given a certain number of varieties on which we wanted to run experiments under basic conditions. Almost naively we had the number of varieties equal to the number of conditions, for each type of condition considered. If we consider the possibility of wanting to generate statistics for tread wear for automotive tires (vehicles with four wheels only) then we might have 8 different brands of tires and 10 different automobiles. Most importantly on any given car we could run experiments using at most four brands of tires. That is, when we separate our varieties into blocks we can no longer have all varieties per block. So we need to generalize the idea of orthogonal Latin squares. The result of this abstraction will be objects called Balanced Incomplete Block Designs, or BIBDs for short.

A block design on a set V with $|V| \geq 2$ consists of a collection $\mathcal{B}$ of non-empty subsets of V called blocks. The elements of V are called varieties, or points.

A block design is balanced if all blocks have the same size k, every variety appears in exactly the same number of blocks r, and any two varieties are simultaneously in exactly λ blocks.

A block design is incomplete in case $k < v$.

If V is a v-set of varieties and $\mathcal{B}$ a b-set of k-subsets of V, which are the blocks of a BIBD where every point is in r blocks, and every pair of points is in λ blocks we say that V and $\mathcal{B}$ form a BIBD with parameters (b, v, r, k, λ). Equivalently a (b, v, r, k, λ)-BIBD means a BIBD with parameters (b, v, r, k, λ).

Example: Let $V = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and put $\mathcal{B} = \{\{1, 3, 4, 5, 9\}, \{2, 4, 5, 6, 10\}, \{0, 3, 5, 6, 7\}, \{1, 4, 6, 7, 8\}, \{2, 5, 7, 8, 9\}, \{3, 6, 8, 9, 10\}, \{0, 4, 7, 9, 10\}, \{0, 1, 5, 8, 10\}, \{0, 1, 2, 6, 9\}, \{1, 2, 3, 6, 10\}, \{0, 2, 3, 4, 8\}\}$. Then V and $\mathcal{B}$ form an $(11, 11, 5, 5, 2)$-BIBD.

This example helps raise a number of pertinent questions

1) How do we know that this example really works, i.e. how can we verify that all pertinent conditions are actually satisfied?

2) What are necessary conditions, if any, which must be satisfied for a BIBD to exist?

3) What are sufficient conditions?

4) When a BIBD exists, is it essentially unique?, If so, why?, If not unique, how many essentially different configurations are there?

The remainder of this section will deal the second of these questions. Subsequent sections will address the remaining questions.

Lemma: *In a BIBD with parameters* (b, v, r, k, λ) *we have* $bk = vr$.

Proof: Count occurences of varieties in blocks two ways. On the one hand there are b blocks each consisting of k varieties. On the other hand there are v varieties each occuring in exactly r blocks. ■

Lemma: *In a BIBD with parameters* (b, v, r, k, λ) *we have* $\lambda(v-1) = r(k-1)$.

Proof: For a particular variety count pairs of varieties it occurs in blocks with two ways. On the one hand, there are $v-1$ other varieties that it must appear with in exactly λ blocks each. On the other hand this variety appears in exactly r blocks and in each block there are $k-1$ other varieties. ■

Example: The parameter set $(12, 9, 4, 3, 2)$ satisfies the first lemma since $12 \cdot 3 = 9 \cdot 4$, but it fails to satisfy the second condition $2 \cdot 8 \neq 4 \cdot 2$. Therefore there can be no $(12, 9, 4, 3, 2)$-BIBD.

Example: The parameter set $(43, 43, 7, 7, 1)$ satisfies both lemmata. However, as we'll see, there is also no $(43, 43, 7, 7, 1)$-design. So we stress that the conditions from the lemmata are necessary, but not sufficient.

It is useful when investigating BIBDs to represent them via incidence matrices similar to what we did for graphs. An incidence matrix for a (b, v, r, k, λ)-BIBD is a $v \times b$ $0-1$ matrix whose rows are labelled by the varieties, columns by blocks, and where the i, j entry is 1 in case the ith variety is in the jth block, and 0 if not.

Example: An incidence matrix for our example of an $(11, 11, 5, 5, 2)$-BIBD where the rows are labelled in order $0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ and the blocks are in reverse order except for the first one

$$\begin{bmatrix}
0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1
\end{bmatrix}$$

We henceforth assume that the reader is familiar with the basics of matrix multiplication and determinants.

An incidence matrix for a BIBD with parameters (b, v, r, k, λ) must have all column sums equal to k, all row sums equal to r, and for $i \neq j$, the dot product of the ith row with the jth row counts the number of columns in which both rows have a 1, which must equal λ. If we denote the identity matrix of side n by I_n and the all 1's matrix of size m by J_m we have

Lemma: *An incidence matrix A for a (b, v, r, k, λ)-BIBD must satisfy the matrix equation $AA^T = (r - \lambda)I_v + \lambda J_v$.*

Proof: The i, j entry of AA^T is the dot product of the ith row of A with the jth column of A^T, which is the dot product of the ith row of A and the jth row of A. This value is λ when $i \neq j$ and r when $i = j$. ■

Lemma: *If A is an incidence matrix for a BIBD with parameters (b, v, r, k, λ), then $det(AA^T) = [r + (v - 1)\lambda](r - \lambda)^{v-1}$.*

Proof:

$$det(AA^T) = det \begin{bmatrix} r & \lambda & \lambda & ... & \lambda \\ \lambda & r & \lambda & ... & \lambda \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \lambda & ... & \lambda & r & \lambda \\ \lambda & ... & \lambda & \lambda & r \end{bmatrix}$$

Adding the last $v - 1$ rows to the first row does not change the value of the determinant so

$$det(AA^T) = det \begin{bmatrix} r + \lambda(v-1) & r + \lambda(v-1) & r + \lambda(v-1) & ... & r + \lambda(v-1) \\ \lambda & r & \lambda & ... & \lambda \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \lambda & ... & \lambda & r & \lambda \\ \lambda & ... & \lambda & \lambda & r \end{bmatrix}$$

Next we factor $r + \lambda(v - 1)$ out of the first row. So

$$det(AA^T) = [r + \lambda(v-1)]det \begin{bmatrix} 1 & 1 & 1 & ... & 1 \\ \lambda & r & \lambda & ... & \lambda \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \lambda & ... & \lambda & r & \lambda \\ \lambda & ... & \lambda & \lambda & r \end{bmatrix}$$

Subtract λ times the first row from each of the last $v - 1$ rows to get

$$det(AA^T) = [r + \lambda(v-1)]det \begin{bmatrix} 1 & 1 & 1 & ... & 1 \\ 0 & r - \lambda & 0 & ... & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & ... & 0 & r - \lambda & 0 \\ 0 & ... & 0 & 0 & r - \lambda \end{bmatrix}$$

From which the result follows. ■

A first application for the previous theorem follows from the fact that for an incomplete design we have $k < v$. The constraint $\lambda(v - 1) = r(k - 1)$ then implies that $\lambda < r$. So the previous theorem shows that $det(AA^T) > 0$. Which led R.A. Fisher to the following theorem.

Theorem: (Fisher's Inequality) *For a BIBD with parameters (b, v, r, k, λ), $v \leq b$.*

Proof: If $v > b$ we can pad an incidence matrix A for the design by adding $v - b$ columns of zeroes resulting in a matrix B with $BB^T = AA^T$. But since B has a column of zeroes $det(B) = det(B^T) = 0$. So $0 = det(B)det(B^T) = det(BB^T) > 0$, a contradiction. ■

When a BIBD with parameters (b, v, r, k, λ) has $v = b$, then the equation $bk = vr$ implies that $k = r$ too. We call a BIBD with parameters (v, v, k, k, λ) a (v, k, λ)-symmetric design, or a symmetric design with parameters (v, k, λ). The integer $n = k - \lambda$ is called the order of the design. A second application of the determinant theorem explicitly for symmetric designs is due to Schutzenberger.

Theorem: (Schutzenberger) *For a symmetric design with parameters (v, k, λ), if v is even, then $n = k - \lambda$ is a square.*

Proof: Let D be a (v, k, λ)-symmetric design and A an incidence matrix for D. Since $r = k$ the equation $\lambda(v-1) = r(k-1)$ now reads $\lambda(v-1) = k(k-1)$. So $r + \lambda(v-1) = k + k(k-1) = k + k^2 - k = k^2$. Thus

$$[det(A)]^2 = det(A)det(A^T) = det(AA^T) = [r + \lambda(v-1)](k-\lambda)^{v-1} = k^2 n^{v-1}.$$

Since the left hand side is a square, the right hand side must be too. But the exponent on n is odd, which means that n must be a square. ■

Example: Consider a putative symmetric design with parameters $(22, 7, 2)$. These parameters satisfy both earlier lemmata. However $v = 22$ is even, and $n = 5$ is not a square. Therefore no such design can exist.

§5.5 Sufficient Conditions for, and Constructions of BIBDs

A design with block size $k = 3$ is called a triple system. If in addition $\lambda = 1$, we have what we call a Steiner Triple System, or STS. Notice that for an STS, given v, we can find b and r from the lemmata of the previous section. So the parameter set of an STS is completely determined as soon as we know v. We call a Steiner Triple System on v varieties a STS(v).

Ten years before Steiner considered these objects the Rev. T.P. Kirkman considered them and proved:

Theorem: *There exist an STS(v) with*
1) $v = 6n + 1, b = nv$, and $r = 3n$ for all $n \in \mathbb{Z}^+$
2) $v = 6n + 3, b = (3n + 1)(2n + 1)$, and $r = 3n + 1$, for all $n \in \mathbb{Z}^+$

This theorem is a corollary of the following theorem, and known constructions of smaller STS's.

Theorem: *If $\mathcal{D}_1$ is an STS(v_1) and $\mathcal{D}_2$ is an STS(v_2), then there exists an STS($v_1 v_2$).*

Proof: Let the varieties of $\mathcal{D}_1$ be $a_1, a_2, ...a_{v_1}$, and the varieties of $\mathcal{D}_2$ be $b_1, b_2, ..., b_{v_2}$. Let c_{ij} $1 \le i \le v_1, 1 \le j \le v_2$ be a set of $v_1 \cdot v_2$ symbols. Define the blocks of an STS($v_1 v_2$) by $\{c_{ir}, c_{js}, c_{kt}\}$ is a block iff one of the following conditions holds

1) $r = s = t$ and $\{a_i, a_j, a_k\}$ is a block of $\mathcal{D}_1$

2) $i = j = k$ and $\{b_r, b_s, b_t\}$ is a block of $\mathcal{D}_2$

3) $\{a_i, a_j, a_k\}$ is a block of $\mathcal{D}_1$, and $\{b_r, b_s, b_t\}$ is a block of $\mathcal{D}_2$.

Now check that all properties are satisfied. ∎

To find small STS's, and other designs there are a number of construction techniques. The first way is to find what is called a difference set. A difference set D with parameters (v, k, λ), aka a (v, k, λ)-difference set, is a k-subset of a group G of order v, written multiplicatively, so that every $g \in G - \{1\}$ is writable as $d_2^{-1}d_1$ for exactly λ pairs $d_1, d_2 \in D$, with $d_1 \neq d_2$.

Example: Let G be the integers modulo 13 written additively. $D = \{0, 1, 3, 9\}$ is a $(13, 4, 1)$-difference set in G as can be seen by the following table of differences.

$-$	0	1	3	9
0	0	12	10	4
1	1	0	11	5
3	3	2	0	7
9	9	8	6	0

For a difference D in a group G the set $Dg = \{dg | d \in d\}$ is called a translate of the difference set. From the previous example $D + 5 = \{5, 6, 8, 1\}$.

Theorem: *If D is a (v, k, λ)-difference set in a group G, then the set of translates $\{Dg | g \in G\}$ form the blocks of a symmetric (v, k, λ)-design where G is the set of varieties.*

Proof: First notice that the number of translates is v. Also each translate has cardinality k by the group cancellation law. A group element g is in the translate Dg_i iff there is $d_i \in D$ with $g = d_i g_i$. There are k chioces for d_i, there are therefore k choices for $g_i = d_i^{-1}g$. So we have v varieties, v blocks, each block of cardinality k, and each variety in k blocks. It remains to show that any pair of varieties is in exactly λ blocks together. So let $g_1 \neq g_2$ be two group elements. Put $x = g_2 g_1^{-1}$. x appears exactly λ times as $d_2^{-1}d_1$ for $d_1, d_2 \in D$ with $d_1 \neq d_2$. There are therefore exactly λ solutions in d_1 and d_2 to the equation $d_2^{-1}d_1 = g_2 g_1^{-1} = x$. That is, exactly λ times we have $d_1 g_1 = d_2 g_2$, with $d_1, d_2 \in D$. Thus $|Dg_1 \cap Dg_2| = \lambda$. ∎

Example: Starting with the block $D + 0 = \{0, 1, 3, 9\}$, we build a $(13, 4, 1)$-design with varieties from $\mathbb{Z}_{13} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ and remaining blocks $D + 1 = \{1, 2, 4, 10\}$, $D + 2 = \{2, 3, 5, 11\}, D + 3 = \{3, 4, 6, 12\}, D + 4 = \{0, 4, 5, 7\}, D + 5 = \{1, 5, 6, 8\}$, $D + 6 = \{2, 6, 7, 9\}, D + 7 = \{3, 7, 8, 10\}, D + 8 = \{4, 8, 9, 11\}, D + 9 = \{5, 9, 10, 12\}$, $D + 10 = \{0, 6, 10, 11\}, D + 11 = \{1, 7, 11, 12\}$, and $D + 12 = \{0, 2, 8, 12\}$.

For a (v, k, λ)-difference set, the number $n = k - \lambda$ is called the order. It is often the case that the difference set is left invariant by a multiplier. In fact frequently a prime dividing n will be a multiplier.

Example: From our $(13, 4, 1)$-difference set in $\mathbb{Z}_{13}$ we have $n = 3$ which is prime. We can decompose $\mathbb{Z}_{13}$ into orbits under multiplication by 3. For $g \in \mathbb{Z}_{13}$ its orbit will be $[g] = \{3^k g | k \in \mathbb{Z}\}$. So we have $[0] = \{0\}$, $[1] = \{1, 3, 9\}$, $[2] = \{2, 6, 5\}$, $[4] = \{4, 12, 10\}$, $[7] = \{7, 8, 11\}$. Notice that the difference set is the union of orbits. Therefore 3 is a multiplier for this difference set.

The term difference set comes from the fact that the first examples of this kind of behavior were discovered by Gauss. In his famous book *Disquisitiones Arithmeticae* Gauss essentially proves the following theorem which we present without proof.

Theorem: *Let p be a prime congruent to* 3 *modulo* 4. *The set of quadratic residues modulo p, which are those numbers $x \in \mathbb{Z}_p^*$ so that there is a $y \in \mathbb{Z}_p^*$ with $x = y^2$, form a $(p, (p-1)/2, (p-3)/4)$-difference set in $\mathbb{Z}_p$.*

There are many other results on difference sets. It is also true that this method generalizes to what is called the method of mixed differences. This more general method is often used to generate designs which are not symmetric designs by starting with a collection of starter blocks. For the method above, we have only one starter block - namely the difference set.

We close this section with a short list of methods which allow us to build new designs from a given design.

Replication: By simply repeating every block of a (b, v, r, k, λ)-design l times we can build a $(bl, v, rl, k, \lambda l)$-design.

Complementation: For a (b, v, r, k, λ)-design on V with blocks $\{a_1, ..., a_b\}$ the complementary design has blocks $\{V - a_1, ..., V - a_b\}$ and parameters $(b, v, b - r, v - k, b - 2r + \lambda)$. So up to complementation we can take $2k < v$.

Set Difference: Starting with a symmetric (v, k, λ)-design with blocks $a_1, ..., a_v$ select some block a_i to delete, and remove all points on that block from the remaining blocks. So we get a $(v - 1, v - k, k, k - \lambda, \lambda)$-design on $V - a_i$ with blocks $a_j - a_i$, for $i \neq j$ and $1 \leq j \leq v$.

Restriction: Again start with a symmetrtic (v, k, λ)-design with blocks $a_1, ..., a_v$. Now select a block a_i and for $j \neq i$ form new blocks $a_j \cap a_i$. This gives a $(v - 1, k, k - 1, \lambda, \lambda - 1)$-design as long as $\lambda > 1$.

§5.6 Finite Plane Geometries

The last major source of designs we wish to discuss are those coming from finite geometries. This is an incredibly rich area for research questions, especially if one does not assume that a finite field is used to coordinatize the space. We will stick to the relatively safe realm where we assume that we have a finite field $\mathbb{F}$ of order $q = p^n$ for some prime integer p. We will also deal mainly with low dimensional geometries, namely planes.

Similar to continuous mathematics we can consider $\mathbb{F}^2$ as a Cartesian product coordinatizing a plane. It's just that in this case because $\mathbb{F}$ is finite, there will be only finitely many points in our plane, which we will denote $EG(2, q)$. EG stands for Euclidean Geometry, and 2 corresponds to the dimension.

In fact we clearly have exactly q^2 points in our plane, since any point can be coordinatized by an ordered pair (x, y), where $x, y \in \mathbb{F}$.

The other natural objects from Euclidean geometry to consider are lines, which over a field satisfy equations of the form $Ax + By = D$, where not both A and B are zero. In fact if $B \neq 0$ we can re-write our line in point-intercept form $y = mx + b$, where m is the slope, and b is the y-intercept. If $B = 0$, then $A \neq 0$ and we get the so-called vertical lines with equations $x = c$. There are q^2 lines of the first form since $m, b \in \mathbb{F}$. And there are q lines of the form $x = c$ since $c \in \mathbb{F}$.

By the properties of a field each line will contain exactly q points. Also every point will be on exactly $q+1$ lines – one for each slope in $\mathbb{F}$, and one with infinite slope (vertical).

Finally, any two points determine a unique line. So the points and lines of $EG(2,q)$ form a $(q^2+q, q^2, q+1, q, 1)$-design.

This design is slightly different from previous examples in that it is resolvable. This just means that blocks come in "parallel" classes - blocks from a parallel class partition the variety set.

If we extend every line of slope m to ∞, we get the effect of looking down a pair of parallel railroad tracks - which we perceive as intersecting eventually. Let the common point at infinity which is the intersection of all lines of slope m be labelled (m). Include the point (∞) as the intersection of all of the vertical lines. Finally connect all of these points at infinity with a line at infinity. The result will be a new structure with $q+1$ new points and 1 new line. We call this $PG(2,q)$. PG stands for projective geometry.

The points and lines of $PG(2,q)$ form a symmetric $(q^2+q+1, q+1, 1)$-design. It can be shown that the existence of such a design is equivalent to a complete set of MOLS of order q. It can also be shown that the result of performing the set difference construction starting with a projective plane of order q and using the line at infinity gives an affine plane of order q. In fact any line (not just the one at inifinity) can be used.

Another way to construct $PG(2,q)$ is to define its points to be the one-dimensional subspaces of $\mathbb{F}^3$, and its lines to be the two-dimensional subspaces. A point is on a line, if the corresponding one-dimensional subspace is included in the two-dimensional subspace.

Now any "point" is completely determined by a direction vector $v \neq \langle 0,0,0\rangle$. We assign the components of v as the homogeneous coordinates of the point, with the convention that two sets of homogeneous coordinates determine the same point iff they are in a common ratio, that is $[a:b:c] = [d:e:f]$ iff there is $\alpha \in \mathbb{F}^*$ with $\langle a,b,c\rangle = \alpha\langle d,e,f\rangle$.

Given a point with homogeneous coordinates $[a:b:c]$ and $c \neq 0$, the point also has coordinates $[a/c : b/c : 1] := [x:y:1]$. We call such a point a "finite" point and let it correspond to $(x,y) \in \mathbb{F}^2$. If $c=0$, then say $x \neq 0$, which means $[x:y:0] = [1:y/x:0] := [1:m:0]$. So the points at infinity labelling non-infinite slopes of lines are $(m) = [1:m:0]$. Finally, if both $a=c=0$, then $b \neq 0$ and $[0:b:0] = [0:1:0] = (\infty)$.

We can also use homogeneous coordinates for the lines, but commonly represent them as column vectors. Any line is really a plane through $(0,0,0)$ in $\mathbb{F}^3$, and so is completely determined by its normal vector. The point $[x:y:z]$ lies on the line $[A:B:C]^T$ in case $Ax+By+Cz=0$.

A line $[A:B:C]^T$ with $B \neq 0$ is the same as the line $[-m:1:-b]^T$, where $m=-A/B$ and $b=-C/B$. These have equation $y=mx+b$ in the finite part of the plane. A line with $B=0$ but $A \neq 0$ is $[1:0:-c]^T$, where $-c=-C/A$. These have equation $x=c$ in the finite part of the plane. Finally, the line at infinity l_∞ has coordinates $[0,0,1]^T$.

Chapter 5 Exercises

1. Show that x^2+1 is irreducible mod 3. Show that if $\alpha^2+1=0 \pmod 3$, then α has order 4 in GF(9). Hence x^2+1 is not a primitive polynomial mod 3.

2. Show that x^2+2x+2 is irreducible mod 3. Show that if $\alpha^2+2\alpha+2=0$ mod 3, then α has order 8 in GF(9). Hence x^2+2x+2 is a primitive polynomial modulo 3.

3. Find the multiplicative inverse of 8 in a) GF(11), b) GF(13), c) GF(17).

4. In GF(7) solve the equations

$$\begin{aligned} x+2y+3z&=2\\ 2x+4y+5z&=6\\ 3x+y+6z&=4 \end{aligned}$$

5. Show that 5 is a primitive root modulo 23.

6. Find all monic irreducible quadratic polynomials mod 5.

7. Let $p=2$ and $n=4$

a) How many monic irreducible quartics are there in $\mathbb{Z}_p[x]$?
b) How many primitive monic irreducible quartics are there in $\mathbb{Z}_p[x]$?
c) How many self-reciprocal monic irreducible quartics are there in $\mathbb{Z}_p[x]$?
Use a-c to factor $x^{p^4}-x$ into monic irreducible factors in $\mathbb{Z}_p[x]$.

8. Show that $q(x)=x^2+1$ is reducible mod 5. Repeat mod 7. Is $q(x)$ primitive mod 7?

9. Show that the probability that a monic irreducible cubic is primitive in $\mathbb{Z}_5[x]$ is $\frac{1}{2}$.

10. Show that x^4+x+1 is primitive in $\mathbb{Z}_2[x]$.

11. Find a primitive root modulo 31.

12. For each value of n determine how many mutually orthogonal Latin squares can be constructed using MacNeish's Theorem.

a) $n=12$	b) $n=13$	c) $n=21$	d) $n=25$
e) $n=35$	f) $n=36$	g) $n=39$	h) $n=75$
i) $n=140$	j) $n=185$	k) $n=369$	l) $n=539$

13. Describe how you would construct 10 MOLS of order 275.

14. Construct a complete set of MOLS of order 5.

15. Repeat exercise 3, with squares of order 7.

16. Repeat exercise 3, with squares of order 8.

17. For each of the following block designs, determine if the design is a BIBD and if so compute its parameters b, v, r, k, and λ.

a) Varieties: $\{1,2,3\}$
Blocks: $\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}$

b) Varieties: $\{1,2,3,4,5\}$
Blocks: $\{1,2,3\},\{2,3,4\},\{3,4,5\},\{1,4,5\},\{1,2,5\}$

c) Varieties: $\{1,2,3,4,5\}$

Blocks: $\{1,2,3,4\},\{1,3,4,5\},\{1,2,4,5\},\{1,2,3,5\},\{2,3,4,5\}$

d) Varieties: $\{1,2,3,4,5,6,7,8,9\}$

Blocks: $\{1,2,3\},\{4,5,6\},\{7,8,9\},\{1,4,7\},\{2,5,8\},\{3,6,9\}$
$\{1,5,9\},\{2,6,7\},\{3,4,8\},\{1,6,8\},\{2,4,9\},\{3,5,7\}$

18. Find an incidence matrix for each design from exercise 17.

19. If a BIBD has parameters $v=15, k=10$, and $\lambda=9$, find b and r.

20. If a BIBD has parameters $v=47=b$, and $r=23$, find k and λ.

21. If a BIBD has parameters $b=14, k=3$, and $\lambda=2$, find v and r.

22. Show that there is no $(7,5,4,3,2)-$BIBD.

23. Show that there is no $(22,22,7,7,1)$-BIBD.

24. For a Steiner triple system with $v=9$, find b and r.

25. The following nine blocks form part of a Steiner triple system on $\{a,b,c,d,e,f,g,h,i\}$:
$\{a,b,c\},\{d,e,f\},\{g,h,i\},\{a,d,g\},\{c,e,h\},\{b,f,i\},\{a,e,i\},\{c,f,g\},\{b,d,h\}$
Add the remaining blocks to complete the design.

26. Four of the blocks of a symmetric (7,3,1)-design are $\{1,2,3\},\{2,4,6\},\{3,4,5\},\{3,6,7\}$. Find the remaining blocks.

27. Find a $(11,5,2)$-difference set in $\mathbb{Z}_{11}$. Display the table of differences.

28. Find a $(13,4,1)$-difference set in $\mathbb{Z}_{13}$. Display the table of differences.

29. Find a $(21,5,1)$-difference set in $\mathbb{Z}_{21}$. Display the table of differences.

30. Find a $(16,6,2)$-difference set in $\mathbb{Z}_2^4$. Display the table of differences.

31. Use your answer from exercise 30 to construct a 16×16 matrix, H, whose entries are all ± 1, and for which $H=H^T$, and $H^2=16I_{16}$.

32. If α is the primitive root of $GF(9)=\mathbb{Z}_3/<x^2+2x+2>$, find all points on the line $\alpha x+\alpha^3 y=\alpha^7$ in $EG(2,9)$.

33. Find the point of intersection of the lines $2x+y=5$ and $3x+4y=6$ in $EG(2,7)$.

34. For the geometry $EG(2,4)$ find equations for every line and determine which points are incident with each line.

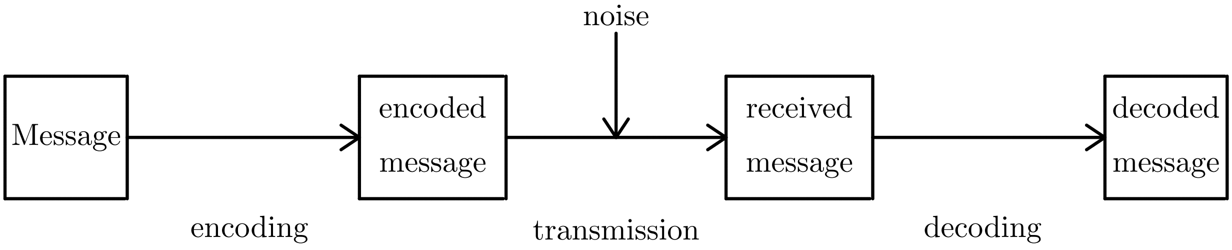
Chapter 6: Introductory Coding Theory

Our goal in this chapter is to introduce the rudiments of a body of theory whose purpose for being is reliable and efficient communication of information. Applications include minimization of noise on CD recordings, data transfer between computers, transmission of information via phone line, radio signal...

All of the applications have in common that there is a medium, called the channel through which the information is sent.

Disturbances, called noise, may cause what is received to differ from what was sent. Noise may be caused by sunspots, lightning, meteor showers, random radio interference (dissonant waves), poor typing, poor hearing,....

The classic diagram for coding theory is



If there is no noise, there is no need for the theory. Engineering tactics and/or choice of channel may be able to combat certain types of noise. For any remaining noise we'll need coding theory.

Specifically we want a scheme which

1. allows fast encoding of information
2. allows for easy transmission
3. allows fast decoding
4. allows us to detect and correct any errors caused by noise
5. has maximal transfer of information per unit time interval

§6.1 Basic Background

Humans have built-in decoding. We can determine from context and proximity where an error has occurred, and correct the mistake with very high reliability.

We won't assume that any of our schemes are going to be able to correct syntactic errors, or even to correct errors based on context. We shall concentrate on trying to correct errors only by using proximity. That is, if we know that an error has occurred in transmission, we will conclude that the most likely message sent is the one which is "closest" to the message received.

In general we will need two alphabets A, and B, one for inputs to the encoder, one for outputs. For this course we will take $A = B = \{0, 1\}$, and we call our scheme a binary code. The generalizations to non-binary code we leave to a later course.

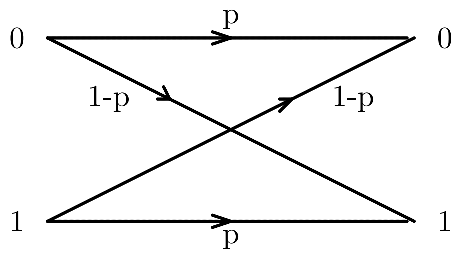
A message will simply be a string over A. We will generally be considering what are called block codes, where messages are formatted into blocks each of which have the same length. Similarly for B, all images under encoding will customarily have the same length.

Historically this was not the case. The most noteworthy example of a binary code which is not a block code is Morse code. A is the set of English letters, $B = \{\cdot, -\}$. a is encoded as $\cdot-$, while e is encoded as $\cdot$. SOS is encoded $\cdots - - - \cdots$ etc.

Our assumption is that any message block is in A^k, while the encoded block is in B^n. A code C is simply the set of words in B^n which are images of elements of A^k under encoding. We want to recapture the original message after decoding. So if we think of $E : A^k \longrightarrow B^n$ as a function, then $D : C = imE \longrightarrow A^k$ will be E's inverse function. So when $A = B = \{0, 1\}$ we will need $k \leq n$.

We also need to make some assumptions about the channel. First that no information is lost - if a message of n symbols is sent, then n symbols are received. Second we assume that noise is randomly distributed - as opposed to occurring in clumps (this type of noise is known as "burst" errors). More specifically, the probability of any message symbol being affected by noise is the same for each symbol, regardless of its position in the message. Also what happens to one symbol is independent to what happens to symbols nearby.

For a binary channel we will further assume symmetry. That is that p the probability of a 1 being sent and a 1 being received, is the same as the probability of a 0 being sent and a 0 being received. So we have the following diagram for a binary symmetric channel (BSC).



The quantity p for a BSC is called the <u>reliability</u> of the code. A BSC with $p = 1$ is <u>perfect</u>. (Contact the author if you find one.) A BSC with $p = 0$ can be turned into one with $p = 1$ simply by toggling the bits prior to transmission. Similarly any BSC with $0 < p < 1/2$ can be turned into a BSC with $1/2 < p' = 1 - p < 1$ via the same trick. A BSC with $p = 1/2$ would only have value as a random number generator (let me know if you find one of these too). Henceforth we assume that $1/2 < p < 1$.

§6.2 Error Detection, Error Correction, and Distance

If a word c is sent, and the word r is received, then for the receiving person to detect an error, it is necessary that the received word is not a codeword. If it is a codeword, they won't realize anything is wrong. Therefore the image, C, of A^k under E needs to be a proper subset of B^n.

For example a trivial code which has $A = B$ and $n = k$ and uses $E = id$ is not capable of detecting any error that occurs. If the code cannot detect an error, it certainly is incapable of correcting any error.

Example: Let $n = 1$, $k = 3$ and $E(0) = 000$, with $E(1) = 111$. So are set of codewords is $C = \{000, 111\}$. If no more than two errors occur in transmitting a single word, then we can always detect this, since three errors are required to change 000 into 111 and vice versa. We therefore call C a 2-error detecting code. C is also a 1-error correcting code, since if we suppose that no more than one error occurs, any received word is either a codeword, or is uniquely the cause of a single error affecting one of our two codewords. This ability comes at the price that our rate of information is essentially 1/3, 1 bit of information being conveyed for every 3 bits sent.

The essential property of the code in the last example which allows it to detect and correct errors is distance. Given two vectors of the same length, their Hamming distance is the number of positions where they differ. The distance of a code C is the minimum distance among all distinct pairs of codewords. Notice that for binary codes the Hamming distance $d(x, y)$ is the number of 1's in the string $x + y$ where addition is componentwise mod 2. The number of 1s in a binary string, z, is called its weight and is denoted by $wt(z)$. So $d(x, y) = wt(x + y)$ for a binary code.

If we now further assume that the probability of fewer errors occuring is higher than the probability of a large number of errors occuring we arrive at the principle of Maximum Likelihood Decoding: Suppose that a word r is received and it is detected that r is not a codeword. In case there is a unique codeword c that is closest to r (the Hamming distance from r to c is minimum), decode r as c.

In our preceeding example the code C clearly has distance 3 since 000 and 111 differ in all three positions. We write $d(000, 111) = 3$. From our previous discussion we get.

Theorem: *For a binary code C, let $\delta = min_{v\neq w\in C} d(v, w)$. Then C can detect up to $\delta - 1$ errors, but there is a way that δ errors can occur which cannot be detected.*

More importantly

Theorem: *If C is a code with minimum distance δ and $t = \lceil(\delta/2) - 1\rceil$, then C can correct up to t errors by using Maximum Likelihood Decoding, but there will be a way for $t + 1$ errors to occur and correction is not possible.*

Notice that the probability of errors occuring comes into play in our assumptions. From the binomial theorem we have the following corollary.

Theorem: *In a BSC with reliability p, the probability that exactly r errors will occur in transmitting a bit string of length n is given by $\binom{n}{r}(1-p)^r p^{n-r}$.*

Example: When $n = 7$ and $p = 0.99$ the probability that exactly one error occurs is $\binom{7}{1}(0.01)^1(0.99)^6 \approx 0.0659036$. The probability that no error occurs is $\binom{7}{0}(0.01)^0(0.99)^7 \approx 0.9320653$. The probability that exactly two errors occur is $\binom{7}{2}(0.01)^2(0.99)^5 \approx 0.0019971$. So the probability that more than two errors occur is about 0.00003397.

Example: If $n = 11$ and $p = 1 - 10^{-8}$, and the rate of transmission is 10^7 digits/second the probability that a word of length $n = 11$ is transmitted incorrectly is about $11p^10(1 - p)$. Since $10^7/11$ words are transmitted each second, we can expect $\frac{11}{10^8} \cdot \frac{10^7}{11} \approx 0.1$ words/second to be transmitted incorrectly, which translates into 8640 words transmitted incorrectly each day. If we add a single parity check digit at the end of each word, to make the number of ones in the word even, then the probability of at least 2 errors occuring per codeword is $1 - p^{12} - 12p^{1}1(1-p) \approx 66/10^{16}$. Now we find that we can expect to wait about 2000 days before an undetectable set of errors occurs.

R.W. Hamming was one of the original contributors to coding theory. Among other things he is credited with the following bound on the size of a code given its minimum distance.

Theorem: (Hamming Bound) *If C is a binary code of length n and minimum distance δ, then* $|C| \leq \dfrac{2^n}{\binom{n}{0} + \binom{n}{1} + ... + \binom{n}{t}}$, *where* $t = \lceil (\delta/2) - 1 \rceil$.

Proof: Let $s \in C$. Denote by $B_r(s)$ those bit strings of length n which have Hamming distance exactly r from s. Note that $|B_r(s)| = \begin{pmatrix} n \\ r \end{pmatrix}$ since we simply choose the r positions from s to change.

Now denote by $B'_m(s)$ the set of bit strings of length n which are at distance at most m from s (inclusive). Now $|B'_m(s)| = \begin{pmatrix} n \\ 0 \end{pmatrix} + \begin{pmatrix} n \\ 1 \end{pmatrix} + ... + \begin{pmatrix} n \\ m \end{pmatrix}$.

If $t = \lceil (\delta/2) - 1 \rceil$, then $B'_t(s_1) \cap B'_t(s_2) = \emptyset$ for all $s_1, s_2 \in C$. Thus any bit string of length n is in at most one set $B'_t(s)$.

Therefore $|C||B'_t(s)| = \sum_{s \in C} |B'_t(s)| \leq | \cup_{s \in C} B'_t(s)| \leq 2^n$. ∎

§6.3 Linear Codes

Almost all coding schemes in use are what are known as linear codes. To say that a binary code C is linear, means that if $x, y \in C$, then $x + y \in C$. So a linear binary code is a k-dimensional subspace of $\mathbb{Z}_2^n$. Thus there is a set of codewords $\{b_1, ..., b_k\}$ (a basis), so that every codeword $c \in C$ is (uniquely) a linear combination of the basis vectors, i.e. $c = \alpha_1 b_1 + \alpha_2 b_2 + ... + \alpha_k b_k$.

Notice, that every binary code contains the all 0s word since if $c \in C$, $c + c = 0 \in C$. Also notice that the minimum distance in a linear code coincides with the minimum weight taken over all non-zero code words. This is true because $d(x, y) = wt(x + y)$, and for a linear code $x + y$ is always a code word when x and y are.

The $k \times n$ matrix M whose rows are $b_1, ..., b_k$ is a generating matrix for C. In general any matrix whose rows form a basis for C is a generating matrix for C. Most linear codes have a generating matrix which takes standard form. The generating matrix for a code in standard form consists of a $k \times k$ identity matrix appended with a $k \times (n - k)$ matrix G.

For a binary linear code C, encoding is accomplished by vector-matrix multiplication. If $v \in \{0, 1\}^k$ is a message word, $E(v) = vM$, where M is a generating matrix for C, and all arithmetic is performed mod 2. When M is in standard form $E(v) = (v|vG)$, so the information bits of a word are preserved in the first k bits of the code word. Thus if no errors occur decoding amounts to stripping off the first k bits of every codeword received. Life is more interesting when errors do occur.

To detect whether errors have occurred we use what's called a parity check matrix. A parity check matrix for a binary linear code C of dimension k and length n, is an $(n - k) \times n$ matrix H, so that $Hc^T = 0$, for all $c \in C$.

If we label the entries of a binary vector x as $x_1, ..., x_n$, then the condition $Hx^T = 0$, that a binary vector is in a linear code is equivalent to a set of equations that the coordinates x_i must satisfy. These are the so-called parity check equations.

Luckily when C has a generating matrix M in standard form, it also has a parity check matrix in standard form, namely H is G^T appended by an $(n-k) \times (n-k)$ identity matrix. This works modulo 2 since $HM^T = G^T + G^T = 0$ modulo 2. So $Hb^T = 0$ on for every vector b in a basis for C. Which means $Hc^T = 0$ for all $c \in C$ since H is linear, and every codeword is a linear combination of basis vectors.

When C has a parity check matrix H in standard for we can also standardize the parity check equations by solving for the last $n-k$ components in terms of the first.

Example If C has generating matrix

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Then

$$G = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Then we may take

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

So the parity check equations are

$$x_1 + x_2 + x_4 + x_5 = 0$$
$$x_1 + x_3 + x_4 + x_6 = 0$$
$$x_2 + x_3 + x_4 + x_7 = 0.$$

Which are equivalent to

$$x_5 = x_1 + x_2 + x_4$$
$$x_6 = x_1 + x_3 + x_4$$
$$x_7 = x_2 + x_3 + x_4.$$

If the likelihood of multiple errors is low enough then MLD can be performed by using the parity check matrix.

We start with a linear binary code with minimum distance of at least three and a generating matrix M in standard form. we assume that either no errors occur in transmitting a single codeword, or one error occurs when transmitting a codeword. So at most one error occurs per codeword which is a relatively safe assumption if the reliability of our channel is high enough.

Suppose that a word v is encoded as $c = vM = (v|vG)$ and the word r is received. If no error occured, $Hr^T = 0$, and we simply decode by taking the first k bits of r. If one error has occured in the ith position, then $r = c + e_i$, where e_i is the bit string of length n with a 1 in the ith position and zeroes everywhere else. Now $Hr^T = H(c + e_i)^T = Hc^T + He_i^T = He_i^T$ must then be the ith column of H. Thus we can identify the location i of the error from Hr^T. We toggle this bit and decode as if no error had occurred.

Notice that this only corrects single errors, and heavily relies on the relaibility of the channel. To correct more errors, requires more intricate procedures, but utilizes similar thinking.

Chapter 6 Exercises

1. In each case find the Hamming distance between x and y.

a) $x = 1110010$, $y = 0101010$

b) $x = 10001000$, $y = 10100101$

c) $x = 111111000$, $y = 001001001$

d) $x = 111000111000$, $y = 101010101010$

2. For each of the following codes C, find the number of errors that could be detected, and the number of errors that could be corrected using maximum-likelihood decoding.

a) $C = \{0000000, 1111110, 1010100, 0101010\}$

b) $C = \{0000000, 1111111, 1111000, 0000111\}$

c) $C = \{00011, 00101, 01001, 10001, 00110, 01010, 10010, 01100, 10100, 11000\}$

3. Find an upper bound on the number of codewords in a code where each codeword has length 8 and the minimum distance of the code is 5.

4. Let C be a binary code where every codeword has even weight. Show that the minimum distance of the code is an even number.

5. In each case C is a linear code with the given generating matrix. Encode each message word.

a) $M = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$, *i)* $v = 11$, *ii)* $v = 10$

b) $M = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$, *i)* $v = 111$, *ii)* $v = 100$, *iii)* $v = 000$

c) $M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$, *i)* $v = 1110$, *ii)* $v = 1010$

6. Find the linear code generated by M in each case for M from exercise 6.

7. Find the minimum distance for each of the following linear codes.

a) $C = \{000000, 001001, 010010, 100100, 011011, 101101, 110110, 111111\}$

b) $C = \{000000000, 111111111, 111100000, 000011111, 101010101, 010101010\}$

c) $C = \{11111111, 10101010, 11001100, 10011001, 11110000, 10100101, 11000011, 10010110, 00000000, 01010101, 00110011, 01100110, 00001111, 01011010, 00111100, 01101001\}$

8. Find the standard parity check matrix for each part of exercise 6.

9. Find the standard parity check equations for each part of exercise 9.

10. If C is a linear code with standard parity check matrix

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

find a generating matrix for C in standard form.

11. Assume that relatively few errors occur and that $r = 101001$ is received using the code generated by the matrix from exercise 6 b). What is the most likely codeword sent?

12. Repeat exercise 11 with $r = 001001$.

13. Suppose we are using the code generated by M from exercise 6c) and $r = 1110110$ is received. Determine whether an error occured in transmission, and if so determine the most likely codeword transmitted.

14. Repeat exercise 13 with $r = 1010011$.

Appendix 1: Pólya's Theorem and Structure in Cyclic Groups

A. Proof of Pólya's Theorem:

Theorem: (Pólya Version 2) *If a group G acts on a set D whose elements are colored by elements of R, which are weighted by w, then the expression*

$$P_G\Big[\sum_{r\in R} w(r), \sum_{r\in R}(w(r)^2), ..., \sum_{r\in R}(w(r)^k)\Big]$$

generates the pattern inventory of distinct colorings by weight, where $P_G[x_1, x_2, ..., x_k]$ is the cycle index of G.

Proof: We will proceed by lemmata supposing throughout that $R = \{1, 2, ..., m\}$.

Lemma: *Suppose that the sets $D_1, D_2, ..., D_p$ partition D and let C be the subset of $C(D, R)$ which consists of all colorings, f, which are constant on the parts. That is, if $a, b \in D_i$, for some i, then $f(a) = f(b)$. Then the pattern inventory of the set C is*

$$\prod_{i=1}^{m}[w(1)^{|D_i|} + w(2)^{|D_i|} + ... + w(m)^{|D_i|}] \tag{1}$$

Proof: The terms in the expansion of (1) are of the form $w(i_1)^{|D_1|}w(i_2)^{|D_2|}...w(i_m)^{|D_m|}$. This is exactly the weight given to the coloring which assigns the color i_1 to every element of D_1, assigns the color i_2 to every element of D_2, etc. Thus (1) gives the sums of the weights of colorings which are constant on each part D_i. ■

Lemma: *Suppose that $G^* = \{\pi_1^*, \pi_2^*, ...\}$ is a group of permutations of $C(D, R)$. For each $\pi^* \in G^*$, let $sw(\pi^*)$ be the sum of the weights of all colorings f which are invariant under π^*. Suppose that $C_1, C_2, ...$ are the equivalence classes of colorings and denote by $wt(C_i)$, the common weight of all f in C_i. Then*

$$wt(C_1) + wt(C_2) + ... = \frac{1}{|G^*|}[sw(\pi_1^*) + sw(\pi_2^*) + ...] \tag{2}$$

Proof: For a particular coloring f, $w(f)$ is added in the sum part of the right-hand side exactly as many times as a group element leaves f invariant. Thus $w(f)$ is accounted for $|St(f)|$ times in the summation part of the right-hand side. But $|St(f)| = |G^*|/|C(f)|$ by the second theorem from section 4.3, where $C(f)$ is the equivalence class of f under G^*, and $St(f)$ is the stabilizer of f under the action. So by substitution and simplification the right-hand side of (2) is the same as

$$\frac{1}{|G^*|}\sum_{f_i\in C(D,R)} w(f_i)|St(f_i)| = \frac{1}{|G^*|}\sum_{f_i\in C(D,R)} w(f_i)\frac{|G^*|}{|C(f_i)|} = \sum_{f_i\in C(D,R)}\frac{w(f_i)}{|C(f_i)|} \tag{3}$$

Similar to the proof of THE LEMMA, we now add up the terms $w(f_i)/|C(f_i)|$ for all f_i in an equivalence class C_j. The result is $wt(C_j)$ for each class C_j, since all colorings in the class have a common weight, and the number of terms is exactly $|C_j| = |C(f_i)|$. So the total is the left-hand side of (2). ■

Notice that the left-hand side of (2) is the pattern inventory.

Let π be a permutation whose cycle decomposition is $\gamma_1\gamma_2\gamma_3...\gamma_p$, where γ_i is a cyclic permutation of D_i, for $i = 1, 2, ..., p$. A coloring f is invariant under π^* iff $f(a) = f(b)$ whenever a and b are in the same part D_i.

By the first lemma above (1) gives the inventory of the set of colorings left invariant by π^*. Each term in (1) is of the form

$$\sum_{r\in R}[w(r)]^j, \text{ where } j = |D_i| \tag{4}$$

So a term like (4) occurs in (1) as many times as $|D_i| = j$, that is. as many times as π has a cycle of length j in its cycle decomposition. We called this e_j in chapter 4. So $wt(\pi^*)$ can be rewritten

$$\prod_{j\geq 1}\left[\sum_{r\in R}[w(r)]^j\right]^{e_j}$$

So the right-hand side of (2) becomes

$$P_G\left[\sum_{r\in R} w(r), \sum_{r\in R}(w(r)^2), ..., \sum_{r\in R}(w(r)^k)\right]. \qquad \blacksquare$$

B. Structure in Cyclic Groups

When $|G| < \infty$, and $a \in G$, $o(a) = |\langle a\rangle|\Bigg|\,|G|$.

Especially if $b \in \langle a\rangle$, then $o(b)|o(a)$. Also $b = a^i$, for some $0 \leq i < o(a)$.

Fact: If $b^j = e$, then $o(b)|j$.

Proof: Let $l = o(b)$, so $b^l = e$ and l is the least positive integer with this property. Write $j = ql + r$ with $0 \leq r < l$. If $e = b^j = b^{ql+r} = b^{ql}b^r = (b^l)^q b^r = e^q b^r = b^r$. $r = 0$ or we reach a contradiction to the minimality of l. ■

Theorem: $o(a^i) = \dfrac{o(a)}{\text{gcd }(i, o(a))}$.

Proof: Put $d = \text{gcd }(i, o(a))$ and $k = o(a)$. Write $k = db$ and $i = dc$, where $b, c \in \mathbb{Z}$. Notice that gcd $(b, c) = 1$, and that $b = \dfrac{k}{d} = \dfrac{o(a)}{\text{gcd }(i, o(a))}$.

Now $(a^i)^b = (a^{dc})^b = (a^{db})^c = (a^k)^c = e^c = e$. So $o(a^i)|b$.

On the other hand $e = (a^i)^{o(a^i)} = a^{i\cdot o(a^i)}$, so $k|i \cdot o(a^i)$. That is $db|dc \cdot o(a^i)$. Thus $b|c \cdot o(a^i)$. Since gcd $(b, c) = 1$, we conclude $b|o(a^i)$. ■

Appendix 2: Some Linear Algebra

One way to describe a field is as an algebraic object in which we can perform all of the standard arithmetic operations of subtraction, addition, multiplication, and division (except by 0). One way to define a vector space is as the collection of ordered n-tuples whose entries lie in a field, and for which we define addition componentwise, and scalar multiplication by multilplying each component by the scalar. This will be sufficient for our purposes.

We call our generic field $\mathbb{F}$, and denote its elements by lowercase Greek letters. A vector $x = [x_1, x_2, ..., x_n] \in \mathbb{F}^n$ has components x_i and is usually denoted as a row vector, as opposed to a column vector.

$$x + y = [x_1, x_2, ..., x_n] + [y_1, y_2, ..., y_n] = [x_1 + y_1, x_2 + y_2, ..., x_n + y_n]$$

$$\alpha x = \alpha[x_1, x_2, ..., x_n] = [\alpha x_1, \alpha x_2, ..., \alpha x_n]$$

For two vectors x and y of the same length their dot product is $x_1 \cdot y_1 + x_2 \cdot y_2 + ... + x_n \cdot y_n$.

A linear combination of a set of vectors $S = \{a, b, c, ..., e\}$ is any vector of the form $\alpha a + \beta b + \gamma c + ... + \epsilon e$. The span of a set of vectors is the set of all possible linear combinations of those vectors. The span of the empty set is taken to be the zero vector $0 = [0, 0, ..., 0]$.

A set $S = \{a, b, c, ..., e\}$ of vectors is linearly dependent if there exist scalars $\alpha, \beta, \gamma, ..., \epsilon$, not all zero, so that $\alpha a + \beta b + ... + \epsilon e = 0 = [0, 0, ..., 0]$.

A set $S = \{a, b, c, ..., e\}$ of vectors is linearly independent if it is not linearly dependent. So S is linearly independent means that if $\alpha a + \beta b + ... + \epsilon e = 0 = [0, 0, ..., 0]$, then $\alpha = \beta = \gamma = ... = \epsilon = 0$.

Notice that every superset of a linearly dependent set of vectors is linearly dependent, and that every subset of a linearly independent set of vectors is linearly independent.

A subset of a vector space which is closed with respect to vector addition and scalar multiplication is called a subspace (it can be viewed as a vector space in its own right after making suitable adjustments in notation if necessary).

A linearly independent subset S of a vector space whose span is all of the vector space is called a basis. Equivalently (for our purposes) a basis is a maximally sized linearly independent set. Every basis for a vector space has the same cardinality called the dimension of the space.

The standard basis for $\mathbb{F}^n$ is $e_1, e_2, ..., e_n$, where e_i is the vector which has a one in the ith position and zeroes everywhere else. The dimension of $\mathbb{F}^n$ is n.

To test whether a set of vectors is linearly dependent or linearly independent we place them in a matrix as column vectors, and row reduce. The rank of the corresponding matrix is the number of linearly independent column vectors = the number of linearly independent row vectors = the dimension of the vector subspace spanned by the set of vectors. The rank of the matrix is read off the reduced echelon form after row reduction.

A matrix is in reduced echelon form if 1) all zero rows are below any non-zero rows, and 2) the left-most entry in a row is a 1 (called the leading 1), and 3) the leading 1 in each row is strictly to the right of any leading one from a row above the given row. The rank is the number of leading ones.

Matrix multiplication, AB, is defined whenever the length of the rows of A coincides with the length of the columns of B. The i,j th entry of AB is the dot product of the ith row of A with the jth column of B.

Appendix 3: Some Handy Mathematica Commands

A. PólyaCounting

When we have the cycle index of a permutation group we can enter it into Mathematica as a function which we can evaluate and manipulate to more easily take advantage of the pattern inventories via weight functions.

For example we might determine that the cycle index of the symmetric group on 4 letters acting on the edges of K_4 is

$$P_{S_4}\bigg[x_1, x_2, x_3, x_4\bigg] = \frac{1}{24}(x_1^6 + 6x_2x_4 + 8x_3^2 + 9x_1^2x_2^2)$$

In Mathematica we would enter

$$K4[x\_, y\_, z\_, w\_] := (1/24)(x^6 + 6yw + 8z^2 + 9x^2y^2)$$

The _'s after x, y, z, w let Mathematica know that these are being declared as variables. After defining this function if we enter $K4[2, 2, 2, 2]$ the output would be the number of ways of coloring the edges of K_4 using two colors (such as there and not there). If we next give weights 1 for not there, and x for there, then entering $K4[1+x, 1+x^2, 1+x^3, 1+x^4]$ and $Expand[\%]$, the result is the pattern inventory which is the enumeration of isomorphism classes of simple graphs on four vertices. Another useful command is one such as $Sum[Coefficient[\%, x^i], \{i, 3, 6\}]$, which would sum all coefficients of x^3 through x^6 in the previous output.

B. Finite Fields

To factor a polynomial modulo an integer m we use $Factor[x^{11} - x, Modulus- > m]$. Of course we would probably prefer to define the function $g[x\_] = x^{81} - x$; and then $Factor[g[x], Modulus- > 3]$.

This would allow us to find all monic irreducible quartics mod 3 which might be used to build $GF(81)$.

To work in $GF(81)$ we need commands such as

$$Reddus[expr\_] := PolynomialMod[PolynomialMod[expr, m/.x- > \alpha], p]$$

$$Add[f\_, g\_] := Reddus[f + g];, \text{and } Mult[f\_, g\_] := Reddus[fg];$$

To determine if (mod 3) a monic irreducible quartic is primitive we would set say $p = 3; m = x^4 + 2x + 2$; and then generate the powers of a root of m, to see if it's order was 80 or not.

$$nonzero = Column[Table[a^i, Reddus[a^i], i, 0, 80]]$$

The above logarithmic table would allow us to do computations in the field more efficiently.

C. Matrices

A matrix can be entered into Mathematica as a list of lists, or vector of vectors. For example

$A := \{\{1, 0, 0, 0, 1, 1, 1\}, \{0, 1, 0, 0, 0, 1, 1\}, \{0, 0, 1, 0, 1, 0, 1\}, \{0, 0, 0, 1, 1, 1, 0\}\}$; enters the generating matrix for a Hamming code of length 7.

Matrix muliplication is done by ., so if A and B are appropriately sized matrices there product is $A.B$. This works for matrix-vector multiplication as well. Output can be seen in matrix form using the command $MatrixForm[A.B]$. Other useful commands are $Transpose[M]$, $Inverse[A]$, and $RowReduce[M, Modulus- > 2]$.

Appendix 4: GNU Free Documentation License

Version 1.3, 3 November 2008

Copyright 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc. <<http://fsf.org/>>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document ”free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of ”copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software. We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The ”Document”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as ”you”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A ”Modified Version” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A ”Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The ”Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The ”Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A ”Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not ”Transparent” is called ”Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The ”Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, ”Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

The ”publisher” means any person or entity that distributes copies of the Document to the public.

A section ”Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that t ranslates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as ”Acknowledgements”, ”Dedications”, ”Endorsements”, or ”History”.) To ”Preserve the Title” of such a section when you modify the Document means that it remains a section ”Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3. You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document’s license notice requires Cover

Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission. B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement. C. State on the Title page the name of the publisher of the Modified Version, as the publisher. D. Preserve all the copyright notices of the Document. E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices. F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below. G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice. H. Include an unaltered copy of this License. I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the

Modified Version as stated in the previous sentence. J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission. K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein. L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles. M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version. N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section. O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles. You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various partiesfor example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard. You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one. The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail. If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not

three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

back to top Check out other Free Software Foundation campaigns

Defective by Design <http://defectivebydesign.org/>, a campaign against Digital Restrictions Management (DRM) Windows 7 Sins <http://windows7sins.org/>, the case against Microsoft and proprietary software PlayOgg <http://playogg.org/> support free media formats Please send FSF & GNU inquiries to gnu@gnu.org <mailto:gnu@gnu.org>. There are also other ways to contact </contact/> the FSF.

Please send broken links and other corrections or suggestions to webmasters@gnu.org <mailto:webmasters@gnu.org>.

Please see the Translations README </server/standards/README.translations.html> for information on coordinating and submitting translations of this article.

51 Franklin Street, Fifth Floor, Boston, MA 02110, USA

Updated: $Date: 2009/06/1719:20:31$

Translations of this page

English </licenses/fdl-1.3.html> [en]