CONGRUENCES AND CRYPTOGRAPHY – SATURDAY 24TH FEBRUARY, 2018

Barry J. Griffiths

The theory of congruences was introduced by Gauss in 1801, and is an extension of ideas involving divisibility and remainders.

To give an example, take the numbers 5 and 2. If we divide each by 3, then the remainders are both equal to 2. We therefore say that 5 and 2 are congruent modulo 3, i.e. they have the same remainder when divided by 3, and write $5 \equiv 2 \pmod 3$.

Definition: Let n be a positive integer, and let a and b be any integers. We say that a is congruent to b, modulo n, and write $a \equiv b \pmod n$, if a and b have the same reminder when divided by n.

Ex. $28 \equiv 0 \pmod 7$

Ex. $100 \equiv 2 \pmod 7$

Ex. What day of the week will it be 100 days from now? Using the previous example (Since there are 7 days in a week), asking what day it will be 100 days from now is equivalent to asking what day it will be 2 days from now, and so the answer is Monday.

We will look at more advanced calendar problems shortly, but we first need to understand the greatest integer function.

Definition: Given any real number x, we denote by [x] the largest integer less than or equal to x.

Ex. $[\ 5.87\ ] = 5$

Ex. $[\ \pi\ ] = 3$

Ex. $[\ \sqrt{2}\ ] = 1$

Ex. $[\ -3/2\ ] = -2$

We now turn our attention to problems concerning the calendar, or more precisely the Gregorian calendar, which although we take for granted (even though there are aspects to it you may be unfamiliar with) has still not been universally adopted. Burton's textbook "Elementary Number Theory" gives an excellent historical description:

"Our familiar calendar, the Gregorian calendar, goes back as far as the second half of the 16thcentury. The earlier Julian calendar, introduced by Julius Caesar, was based on a year of 365.25 days, with a leap year every fourth year. This was not precise enough measure, because the length of a solar year – the time required for the earth to complete an orbit about the sun – is approximately 365.2422 days. This small error meant that the Julian calendar receded a day from its astronomical norm every 128 years.

By the 16th century, the accumulating inaccuracy caused the vernal equinox (the first day of spring) to fall on March 11 instead of its proper day, March 21. The calendar's inaccuracy naturally persisted

throughout the year, but meant that the Easter festival was celebrated at the wrong astronomical time. Pope Gregory XIII rectified the discrepancy in a new calendar, imposed on the predominately Catholic countries of Europe. He decreed that 10 days were to be omitted from the year 1582, by having October 15 of that year immediately follow October 4. At the same time, the German mathematician Christopher Clavius amended the scheme for leap years: those would be years divisible by 4, except for those marking centuries. Century years would be leap years only if they were divisible by 400. (For example, the century years 1600 and 2000 are leap years, but 1700, 1800, 1900, and 2100 are not)

Because the edict came from Rome, Protestant England and her possessions – including the American colonies – resisted. They did not officially adopt the Gregorian calendar until 1752. By then it was necessary to drop 11 days in September [the 3rd to the 13th] from the Julian calendar. Other nations gradually adopted the reformed calendar: Russia in 1918, and China as late as 1949."

Our goal now is to construct the formula (first derived by Christian Zeller in 1882) which will calculate the day of the week for any given date since the adoption of the Gregorian calendar in 1582.

We start by making the following definitions:

Let d denote the day of the month

Let m denote the month of the year, starting with March

Let y denote the last two digits of the year, using the previous year for January and February

Let c denote the first two digits of the year

Theorem: Given the definitions above, where $c \geq 16$, and $0 \leq y < 100$, the day of the week w is given by

$$\mathbf{w \equiv d + [2.6m - 0.2] - 2c + y + \left[\frac{c}{4}\right] + \left[\frac{y}{4}\right] \pmod 7}$$

where 0 represents Sunday, 1 represents Monday etc.

Ex. Suppose we wanted to calculate the day of the week that my first son was born – 6th of January, 2008.

Using the formula, with d = 6, m = 11, y = 7, c = 20 we get

$w \equiv 6 + [26(11) - 0.2] - 40 + 7 + [20/4] + [7/4] \pmod 7$

$\equiv 6 + 28 - 40 + 7 + 5 + 1 \pmod 7$

$\equiv 0 \pmod 7 \rightarrow$ Sunday

Q. On what day of the week will July 4th, 2037 fall?

We will now discuss the subject of cryptography, which is the science of creating and breaking codes. Many codes are based on number theory, the simplest being shift ciphers, where we assign numbers to each letter:

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

We then shift each letter (modulo 26) by a fixed number, called the key, which we denote by k. This method was used by Julius Caesar, who chose k = 3, to send messages to his military officers.

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓

D E F G H I J K L M N O P Q R S T U V W X Y Z A B C

Q. Decode Caesar's message: KDYHD QLFHG DB

Note: Coded text is often grouped into blocks of five to prevent quickly spotting common words.

Now obviously we can choose any value for the key from 0 to 25, and in general we can write:

$$C \equiv P + k \pmod{26}$$

where C represents cipher (coded) text, P represents plain text (A ≡ 0, B ≡ 1 etc.), and k is the key.

Cracking such a code is not an exact science. However instead of randomly guessing at the value of k, we use the relative frequency of the letters in arbitrary text, and then compare with that with the most frequently used letter in the coded text.

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Rel. Freq. (%)	8	2	3	4	12	2	2	5	7	0	1	4	3	7	8	2	0	6	7	9	3	1	2	0	2	0

Ex. Dicipher the following text, assuming it was created using a shift cipher:

EREXM SRGER RSXTV SWTIV PSRKA LIRMX JEZSV WSRPC XLITV SWTIV SYW

We first construct a frequency table for the letters in the ciphertext:

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency	1	0	1	0	4	0	1	0	4	1	1	2	2	0	0	2	0	7	8	4	0	5	4	4	1	1

Our first guess might be that S in the cipher text corresponds with E in plain text, since they occur most often, implying a key value of 14. However this means that the first five letters become QDQJY.

Continuing with this approach we might then guess that S corresponds with T, then A, then O, with the latter yielding ANATI for the first five letters. This looks plausible, and if we decode the remaining letters we get:

ANATI ONCAN NOTPR OSPER LONGW HENIT FAVOR SONLY THEPR OSPER OUS

This can be quickly recognized as meaning "A NATION CANNOT PROSPER LONG WHEN IT FAVORS ONLY THE PROSPEROUS", a quote from President Obama's inauguration speech.

In the example above, there are 26 possibilities for the value of k. We can increase the number of possible codes to 312 by considering the more general affine ciphers, which have the form

$$C \equiv aP + k \pmod{26}, \text{ where the greatest common divisor of a and 26 is 1}$$

Q. If a = 7 and k = 3, encode GAUSS and decode MFSJDG.