Cryptography – Math Circle

November 10, 2019

Since ancient times, people have wanted to be able to send messages to their friends or allies privately. Sending a message to your general giving new orders for your military campaign is a lot safer if they are encrypted so that if your messenger is captured, the enemy can't read the orders, for example. Today's worksheet will explore examples of **substitution ciphers**, how to use them, and how to break them.

Any scheme for taking a message and making it 'impossible' to read without a key is called a **cipher**. The process of obscuring the content of the message is called **encoding** or **enciphering**, while the reverse process to obtain the original message is **decoding** or **deciphering**. A message that has not been encrypted is called the **plaintext** and once it has been encryped it becomes the **ciphertext**.

1 Caesar Cipher

Perhaps the simplest way to encode text is to use a **Caesar cipher**. In a Caesar cipher, each letter is shifted by some fixed amount to get a new letter. Here is an example Caesar cipher with a shift of 3:

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

We can encrypt by using this substitution: A $\mapsto$ D, B $\mapsto$ E, etc.

Plaintext:	I		L	O	V	E		M	A	T	H
Ciphertext:	L		O	R	Y	H		P	D	W	K

1. The following message was encoded with a Caesar cipher using shift 3. Decode it.

   FDHVDU FLSKHUB DUH QRW VHFXUH

2. Why did the mathematical tree fall over? (Shift of 7)

   ILJHBZL PA OHK UV YLHS YVVAZ

3. What did one math book say to the other? (Shift of 17)

UFE'K SFKYVI DV Z'MV XFK DP FNE GIFSCVDJ!

2 Breaking Caesar Ciphers

If you know a Caesar cipher was used to encode an intercepted ciphertext, you can just try all possible shifts until the message makes sense.

4. If you use the set $\{A, B, C, \ldots, Z\}$, how many different Caesar ciphers are there? In other words, if you want to break a Caesar cipher, how many tries might it take if you are really unlucky and try the correct shift last?

Try to break the following codes that were encrypted with a Caesar cipher.

5. ESP DXLWW HZCOD LCP GPCJ SPWAQFW.

6. GUVFVFUNEQREORPNHFRGURERNERABFCNPRF

3 Affine Ciphers

We would like to use some mathematics in our encryption. In order to do this, we need to first change our letters to numbers. We will do it this way:

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
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 can use this scheme for Caesar ciphers as well as more complicated ciphers. For example, let's encode "A GREEN AND PURPLE COW" using a Caesar cipher with a shift of 12. We can just add 12 to each number. The only problem with this scheme is that sometimes the result is larger than 25 – for these numbers, we take the remainder after division by 26. This called doing arithmetic modulo 26.

Plaintext:	A		G	R	E	E	N		A	N	D		P	U	R	P	L	E		C	O	W
Numbers x:	0		6	17	4	4	13		0	13	3		15	20	17	15	11	4		2	14	22
$x + 12$:	12		18	29	16	16	25		12	25	15		27	32	29	27	23	16		14	26	34
$x + 12 \bmod 26$:	12		18	3	16	16	25		12	25	15		1	6	3	1	23	16		14	0	8
Ciphertext:	M		S	D	Q	Q	Z		M	Z	P		B	G	D	B	X	Q		O	A	I

Once we have replaced the letters with numbers, there are many other things we can do. For example, we could encode our message with a linear equation like $x \mapsto 3x+7 \bmod 26$.

7. Encrypt the message A PURPLE AND GREEN COW using various formulas. The $3x + 7$ code has been done for you.

Plaintext:	A		G	R	E	E	N		A	N	D		P	U	R	P	L	E		C	O	W
Numbers x:	0		6	17	4	4	13		0	13	3		15	20	17	15	11	4		2	14	22
$3x + 7 \bmod 26$:	7		25	6	19	19	20		7	20	16		0	15	6	0	14	19		13	23	21
$3x + 7$ Code:	H		Z	G	T	T	U		H	U	Q		A	P	G	A	O	T		N	X	V
$2x + 1 \bmod 26$:																						
$2x + 1$ Code:																						
$7x + 2 \bmod 26$:																						
$7x + 2$ Code:																						

8. Our next problem is decrypting. We want to work out a formula to undo the affine shift we used to encrypt the message. You can decrypt the $3x + 7$ cipher using the map $x \mapsto 9x - 11$. Try it.

Ciphertext:	W	O	B	F	U	Z		R	X	U	L	T	B	J
Ciphertext Numbers:														
$9x - 11 \bmod 26$:														
Plaintext:														

9. Can you find the decryption formula $x \mapsto Ax+B$ for the encryption formula of $7x+2$? Use your formula to decode the answer to this question.

Which word in the dictionary is spelled incorrectly?

Ciphertext:	G	P	Q	W	R	R	E	Q	F	B	O
Ciphertext Numbers:											
Decrpyted Numbers:											
Plaintext:											

10. Look closely at the line for the $2x + 1$ code in Problem 7. There is a big problem with this code. Can you find it? Why did this happen? How can we avoid this problem in the future?

11. If you use the set $\{A, B, C, \ldots, Z\}$ how many different affine ciphers are there? How many general substitution ciphers where you can send any letter to any other letter are there?

4 Frequency Analysis

General substitution ciphers like those we have looked at today can be broken by **frequency analysis**. This method takes advantage of the fact that text that has been encrypted with a simple substitution cipher still looks like English test, just with different characters. For example, the letters 'e','t','a','o', and 'i' are much more common than the letters 'j','x','q', or 'z'. The pairs 'th', 'er', 'on', and 'an' are the most common pairs of letters, etc.

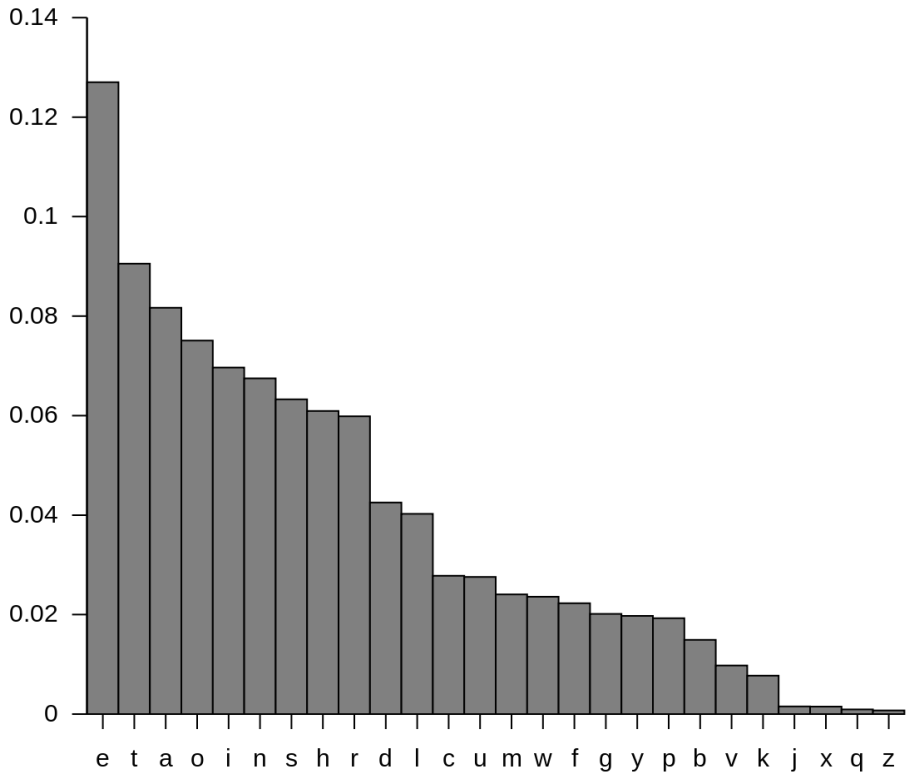

Figure 1: Frequencies of letters in English

If we have a large body of ciphertext, we can attempt to decode by looking for which letters appear the most often and guessing that they will be 'e', 't', 'a', etc. Once we have made a few guesses, we can look for fragments of words we recognize and make more guesses from there. One thing that can make this tricky is that our sample ciphertext may not have the same frequencies as the overall English language, so we might have to backtrack and try different guesses. The longer our ciphertext is, the more accurate this method will be.

12. Here are a bunch of encrypted messages you have intercepted. You believe they were encrypted with a substitution cipher, but you don't know what the substitutions are. Use frequency analysis to decode the messages.

- XCF MR YTLDQH CNGD HYQTIDH? HR YCDF MRAY LDY HIRYYDM.
- XCNY MTM YCD GRSVNAR HNF YR CTH XTUD? T SNGN FRP HR JPVC!
- XCNY MTM RAD XNSS HNF YR YCD RYCDQ XNSS? T'SS JDDY FRP NY YCD VRQADQ.
- XCNY MR FRP VNSS YXR MTARHNPQH YCNY CNGD EDDA TA NA NVVT-MDAY? YFQNAARHNPQPH XQDVBH.
- XCDQD HCRPSM N UTGD CPAMQDM IRPAM NSTDA LR? RA N MTDY.
- XCF MTM YCD ITVYPQD LR YR KNTS? EDVNPHD TY XNH UQNJDM.
- XCNY MTM YCD INIDQ HNF YR YCD IDAVTS? XQTYD RA!
- XCNY LDYH XDYYDQ YCD JRQD TY MQTDH? N YRXDS.
- XCF MR MQNLRAH HSDDI MPQTAL YCD MNF? HR YCDF VNA UTLCY BATL-CYH!
- XCNY MTM YCD HYNJI HNF YR YCD DAGDSRID? HYTVB XTYC JD NAM XD XTSS LR ISNVDH!
- XCNY MTM RAD DSDGNYRQ HNF YR YCD RYCDQ DSDGNYRQ? T YCTAB T'J VRJTAL MRXA XTYC HRJDYCTAL!
- XCF XNH YCD EDSY NQQDHYDM? EDVNPHD TY CDSM PI HRJD INAYH!
- XCTVC CNAM TH TY EDYYDQ YR XQTYD XTYC? ADTYCDQ, TY'H EDHY YR XQTYD XTYC N IDA!
- XCF VNA'Y FRPQ ARHD ED YXDSGD TAVCDH SRAL? EDVNPHD YCDA TY XRPSM ED N URRY!
- XCNY CNH URPQ XCDDSH NAM USTDH? N LNQENLD YQPVB!
- XCF MTM YCD QREEDQ YNBD N ENYC EDURQD CD HYRSD UQRJ YCD ENAB? CD XNAYDM YR JNBD N VSDNA LDY NXNF!

Here are the counts of each letter in the text.

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
41	8	60	105	15	20	9	45	14	12	1	16	39	69	0	20	37	58	31	51	9	21	0	40	92	0