CRYPTO I: CREATING–AND CRACKING–CODES AND CIPHERS

MATH CIRCLE (BEGINNERS) 01/15/2012

1. Caesar Cipher

Before Julius Caesar was Emperor, he was a General in the Roman Army. He needed to send secret messages to other generals, but he was worried that the soldier carrying the messages might be caught by the enemy. So he would encrypt them by shifting the letters of the original message by a certain amount.

For instance, with a shift of +3 (also known as an a-D shift), a $\rightarrow$ D, b $\rightarrow$ E, c $\rightarrow$ F, and so on. When you get to the end of the alphabet, it wraps around, so x $\rightarrow$ A, y $\rightarrow$ B, and z $\rightarrow$ C. The value "3" is the key for this version of the Caesar Cipher, and it gives a codetable (a chart that helps with encrypting and decrypting) that looks like this:

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

(1) What are the different possible keys (shifts) for a Caesar cipher? How many are there?

(2) Encrypt the message "attack at dawn" using a Caesar cipher with a shift of your choice:

Shift value:

Ciphertext (encrypted message):

a	t	t	a	c	k		a	t		d	a	w	n

(3) The following riddles have answers encrypted with a Caesar cipher. Can you figure out the answers? HINT: Try to make a good guess about what the shift is, then you won't have to try lots of keys before you get the answer.

(a) What do you call a dog at the beach?

E L S X H S K.

(b) Three birds were sitting on a fence. A hunter shot one. How many were left?

V W V M. B P M W B P M Z A N T M E

I E I G.

(c) What animal keeps the best time?

K G K D M R N Y Q.

(4) The Italian Mafia boss Bernardo Provenzano used a secret code similar to the Caesar cipher to hide the names of his associates in messages. He was captured in 2006, in part because the police broke his cipher. Here is one example of something he actually wrote (except he wrote it in Italian):

"I met **5,12,15,15,22 19,12,12,15,4** and we agreed that we will see each other after the holidays."

The Italian alphabet has fewer letters than the English alphabet–it's missing J, K, W, X, Y. You can use the following Italian-style codetable to help you:

a	b	c	d	e	f	g	h	i	l	m	n	o	p	q	r	s	t	u	v	z

Can you help the police crack the code and figure out the name of Provenzano's associate?

2. KEYWORD CIPHERS

Caesar ciphers are not very secure. Someone who wants to crack them only needs to try 25 different possibilities for the key! Caesar ciphers are an example of a more general type of cipher known as a *(monoalphabetic) substitution cipher*. This means that each letter of the plaintext alphabet is always transformed to a specific letter in the ciphertext. Another kind of substitution cipher is a *keyword cipher*. Here's how they work:

- Find a friend with whom you'll want to exchange secret messages. Agree on a secret keyword, for instance "DOG", and a secret key letter, for instance "H". If your keyword has any repeated letters, use only the first of each one. For example, the keyword "APPLES" would become "APLES" and "BANANAS" would become just "BANS".

- Write down the alphabet in lowercase–these represent plaintext letters. Starting just below the key letter, write the keyword, then continue with the remaining letters of the alphabet. Be sure to skip the letters you already used in the keyword, when you get to them in the alphabet!

Example: The keyword DOG and key letter H gives a codetable that looks like this:

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

5) With a partner, create a keyword cipher and fill in the the encoding/decoding table below.

KEYWORD:

KEY 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

Use your table to encode:

(1) “the hideout is not safe”

(2) “meet me at six pm”

(3) any message you like! When you’re done encoding it, transfer the ciphertext to another sheet of paper and have your partner decrypt it using their codetable.

3. Frequency Analysis

Now let's try to break keyword ciphers. Each group will get a different piece of plain English text to analyze.

Work together to count the number of times each letter appears in your text:

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M
Frequency													

Letter	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency													

(This table is not like the others so far–it won't be used to encrypt or decrypt messages, it's just used to keep track of how often each letter occurs.)

When each group is done, we'll combine the answers to get an overall idea of how frequent each letter is in English.

OVERALL ENGLISH LETTER FREQUENCY CHART:

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M
Frequency													
Percentage													

Letter	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency													
Percentage													

Now try to decrypt this text, which uses a keyword substitution cipher. The first step is to count the frequency of each letter–work together to do this more quickly!

F U R H J U F E P Q J I K Q J O J N W X E N J

K O C R W P N Q J N F R I X W J H F W K

G J U R J Y J R V Z X N N R G U J K Q R W P N R

I F O J N F C C X E Q F Y J W K Q F I V E H Q

Z O F H K R H J N F R I K Q J M E J J W A Q J W R

A F N C X E W P J O R F U A F C N I R I R K L X O

Q F U L F W Q X E O F I F C A Q C N X V J K R V J N

R Y J G J U R J Y J I F N V F W C F N N R B

R V Z X N N R G U J K Q R W P N G J L X O J

G O J F T L F N K – U J A R N H F O O X U U L O X V

F U R H J R W A X W I J O U F W I.

4. VIGENÈRE CIPHERS

The Vigenère cipher is an example of a *polyalphabetic substitution cipher*–it uses different alphabets to encipher different letters of the plaintext message. Many people believed it was a completely secure way to send secret messages; the French called it *le chiffre indechiffrable*, "the undecipherable cipher." Well, the Titanic was supposed to be unsinkable...

To see how it works, imagine using a Caesar cipher, but using a **different** shift to encode different letters of the plaintext. The pattern of the shifts is determined by a keyword. (But Vigenère ciphers are not the same as keyword ciphers.) For example, suppose the keyword is JUICE. To encode a message, start by writing JUICE above the letters of the message, over and over, like this:

J	U	I	C		E	J	U	I	C	E	J	
t	h	i	s		m	e	s	s	a	g	e	

U	I	C	E		J	U	I		C	E	J	U
u	s	e	s		t	h	e		v	i	g	e

I	C	E	J		U	I	C	E	J	U		
n	e	r	e		c	i	p	h	e	r	.	

Now for each letter with a J above it, encrypt that letter using a shift that takes **a** to **J**. Likewise for each letter with a U above it, encrypt that letter using a shift that takes **a** to **U**. For letters with an I above them, use an **a** to **I** shift, for letters with C, use an **a** to **C** shift, and for letters with E, use an **a** to **E** shift. (Side note: Unlike with keyword ciphers, here it's completely fine to leave repeated letters in the keyword. So we could use a keyword APPLE without first making it APLE.) Once you are done encrypting, write the ciphertext down on a new sheet of paper and you're ready to send your secret message!

Decrypting a message, if you know the key, just means doing things in reverse. Write down the keyword above each letter of the ciphertext, then use each letter of the keyword to tell you the shift to go back to the plaintext. The difference is that if the keyword letter is J (for exampe), you will decode using the reverse shift that takes **J** to **a** (since that letter was originally encoded by a shift that took **a** to **J**.)

To make the process of encrypting/decrypting easier, it helps to have a Vigenère Square. The first letter of each row corresponds to the keyword letter used to encipher that particular plaintext letter. So to encrypt a plaintext letter, first go to the row indexed by the keyword letter above that plaintext letter. Then find the plaintext letter on the top row of the square (lowercase), and follow its column down until it meets the row. That will be the ciphertext letter!

To decrypt a ciphertext letter which appears under a certain keyword letter, go to the row that starts with that keyword letter. Find where the ciphertext letter appears in that row, then follow its column back up to the top row (lowercase)–that's your plaintext letter!

As an example, we've encrypted the first couple of words of "this message uses the vigenere cipher" under the keyword JUICE–**except we made one mistake.** Find which letter in the first row below wasn't enciphered correctly, cross it out and replace with the correct letter. Then finish enciphering the last two rows.

J	U	I	C		E	J	U	I	C	E	J	
t	h	i	s		m	e	s	s	a	g	e	
C	B	Q	U		T	N	M	A	C	K	N	

U	I	C	E		J	U	I		C	E	J	U
u	s	e	s		t	h	e		v	i	g	e

I	C	E	J		U	I	C	E	J	U		
n	e	r	e		c	i	p	h	e	r	.	

Now try encrypting your name using a keyword of your choice. You can use the blank grids on the next page to help.

Your Name:

Keyword:

Encryption:

Now try decoding the following quotations from Mark Twain, each of which was each encoded with the given key. We've started you off on the first one.

(1) Encrypted using the keyword CAR:

C	A	R	C	A	R		C	A		R	C	A
a	l	w	a	y	s		d	o		r	i	g
C	L	N	C	Y	J		F	O		I	K	G

R	C			A	R	C	A		R	C	A	R
h	t	.										
Y	V	.		T	Y	K	S		N	K	L	C

I	R	R	V	I	W	A		S	F	O	E	

G	G	O	G	N	E		R	P	D		R	U

T	F	P	I	J	J		T	Y	G		R	V

U	T	.										

(2) Encrypted using the keyword TWAIN:

T	W		A	I	N		T	W	A	I		N
B	B		Y	W	H		M	A	L	T		G

A	A		T	Z	H	M	D		Y	W	H	

W	K	N	’	B		U	T	R	E		B	B

K	A	M	M	Z	U	A	R		I	A	R	P

H	Q	A	Z	.								

5. Cracking the Vigenere Cipher

Even though many people thought the Vigenere cipher was unbreakable, eventually people figured out how to crack it. Let's try it ourselves!

First let's assume we know the length of the keyword used to encrypt the message (in a little while we'll see how we could figure this out if we didn't already know it). Here's the first ciphertext we'll try to decrypt:

K V X D O G R U X I O M R P H R H - K V B M R Z

V F B V V G L Z C G N S G K H H K V X V R Z V

C Y K V X C O D V O G U M X C Z X U, P H R H

G L A U V F 99, V F A X S O V B. M H L F M Z A X

Z G N G. P N K H A V P H R H W Z R G ' K F X K I K E.

P H R H G L A U V F 99, A V V H C Z X I S W R U T Z B.

K V H N I B M F H A V R H T Y B D A X U W T K S E P

C K Z ' Z E T V T I U X P C N F J X I H B D S.

L F A X K V B E U B J K K F B Z, S C L J, V B J

O L J W L K O G K G T Z R. P V C G C M A R J X

75 S C T K G. M Y S K V W L E C G L A U V F

99. M Y S F R B T X S K K V H L U A K T H I O

F F A X E H. U F O M E I F S S K 66, Y S R V Z E V R.

T I S R F I A R J B E U M I C N S Z X F I M K V X I S?

Now I am going to tell you that the length of the keyword used to encrypt this message is 3. This means that every third letter of the message is encrypted under a Caesar cipher with the same shift. So the 1st, 4th, 7th, 10th, 13th, etc. letters of the message use one shift; the 2nd, 5th, 8th, 11th, 14th, etc. letters use a second shift; and the 3rd, 6th, 9th, 12th, 15th letters use a third shift.

To distinguish which ciphertext letters are encoded under each shift, write the pattern 123123123 under the letters of the ciphertext (ignore numbers, spaces, and punctuation!!!). Then each letter with a "1" underneath it was encrypted using the same Caesar cipher, so we can try to figure out which shift was used for those letters.

Frequency chart ONLY for letters with "1" underneath:

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M
Frequency													

Letter	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency													

Frequency chart ONLY for letters with "2" underneath:

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

Frequency chart ONLY for letters with "3" underneath:

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

Now, try to guess which shift was used on each of the three types of letters. Try decoding using those shifts and see if it makes sense. If not, try another guess!

6. How to Figure Out the Key Length

We decoded the previous cipher assuming we already knew the key length? What if we didn't? Do you have any ideas about what to do?

Here is one way to learn the key length, known as the Kasiski Examination, after Friedrich Kasiski, who published about it in 1863.

The idea is to look for repeated strings of letters in the ciphertext–you want at least 3 characters in a row to be repeated, but the more the better. For instance, in the previous ciphertext the string "GLAUVF 99" appears three times.

Question: How do you think it happened that the same string repeated itself in three different places?

(1) Circle the three places where GLAUVF appears in the ciphertext.

(2) Count the distance (in letters ONLY–**don't count the numbers!**) between the G in the first GLAUVF, and the G in the second GLAUVF. (So if you start with the first GLAUVF, count L-1, A-2, U-3, V-4, F-5, V-6, F-7, A-8, X-9, etc. until you get to the G in the second GLAUVF–what did you count for that G?

What are the factors of this number? Does this by itself give you any hints about the key length?

(3) Now do the same thing, but count the distance between the G in the **second** GLAUVF and the G in the **third** GLAUVF. What do you get?

What are the factors of this number? Does it have any factors in common with the number you got in (1)? How can this help you find the key length for that cipher (if you hadn't already known it)?

As a challenge, see if you can use this technique to determine the keylength for the following cipher, then decrypt it! Try working in groups with people nearby to better divide up the work. There are empty frequency tables below for you to use if it helps. Your first step should be to find repeated substrings of at least 3 characters, and count the distances between occurrences...

A VNNS SGIAV GVDJRJ! WG OOF AB

GZS UAZYK PRZWAV HUW HESRVFU

CGGG GB GZS AADVYCA JWIWF NL

H U W B B J H U W F A L W C G T Y S Y R

K I C W F V G F. J Z W Y W V V C W A Y, W S G I A V

G B E S F Z W A Q G G G B R K. Z N L S E A

P E G I T Z H G Z S Z L C N E S G S Z R P D R J H

G G V N N S G Z S Z S D C J O V K S Q. K I E W

S A G I T Z, H U W M N J S F A Z I W F—V F O I W F L

H I E W T B J A. G Z S E W A H K H O W A B J S—V

O W Y D F R L I E F O A V G G S Y R S Q Y S W Z.

#1

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M
Frequency													

Letter	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency													

#2

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M
Frequency													

Letter	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency													

#3

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M
Frequency													

Letter	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency													

#4

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M
Frequency													

Letter	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency													

#5

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M
Frequency													

Letter	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency													