MA080G Cryptography Assignment Block 1

Viktor Rosvall

April 20, 2019

Question 1

(a)

Explain how a substitution cipher works.

Answer (a)

A **Substitution cipher** is done by permuting (scrambling) the letters of the alphabet. For example: replacing the letter "a" $\rightarrow$ "K". The key to a substitution cipher is a table of each permutation.

(b)

What does it mean for a substitution cipher to be
(i) monoalphabetic?
(ii) polyalphabetic?

Answer (b)

(i)
In a monoalphabetic substitution cipher, the letters are always encrypted the same.

For example: "a" will always be permuted the same, no matter the position in the plaintext.

(ii)
In a polyalphabetic substitution cipher, the latter may be encrypted differently depending on it's position in the plaintext.

For example: In a Vigenère cipher with a key length of 3, the letter "a" will be permuted differently depending on if it's in position 0,1 or 2 in the plaintext cycle. Like "a_0" $\rightarrow$ "K", "a_1" $\rightarrow$ "D" and "a_2" $\rightarrow$ "Q"

(c)

Give a non-trivial example of a monoalphabetic substitution cipher and use it to encrypt the plaintext

Teach thy necessity to reason thus
There is no virtue like necessity

Answer (c)

We first need to define a key, which looks like a table of sorts.
Using this key:

a	b	c	d	e	f	g	h	i	j
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
Q	W	E	R	T	Y	U	I	O	P

k	l	m	n	o	p	q	r	s	t
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
A	S	D	F	G	H	J	K	L	Z

u	v	w	x	y	z
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
X	C	V	B	N	M

our plaintext

Teach thy necessity to reason thus
There is no virtue like necessity

is encoded as:

ZTQEI ZIN FTETLLOZN ZG KTQLGF ZIXL
ZITKT OL FG COKZXT SOAT FTETLLOZN

(d)

Explain how to break a monoalphabetic substitution cipher.

Answer (d)

By looking at the *letter frequencies* in the ciphertext, we can make guesses on what common letters they map to. By replacing the letters of digrams and trigrams in the ciphertext, to letters that's common in the English alphabet, we may start to see English words begin to form.

Question 4

(a)

Define Euler's Φ-function.

Answer (a)

Euler's Φ function on the natural numbers $n \geq 2$ given by:

$$\Phi(n) = \#\text{ congruence classes } [a] \in Z_n \text{ such that } gcd(a, n = 1)$$

Counts the number of invertible elements of Z_n.

(b)

Compute $\Phi(17)$, $\Phi(289)$ and $\Phi(221)$.

Answer (b)

To compute $\Phi(n)$, we first need to prime factorize n, and then use the following formula:

$$\Phi(n) = p_1^{a_1-1}(p_1 - 1)p_2^{a_2-1}(p_2 - 1)...p_r^{a_r-1}(p_r - 1)$$

Where $n = p_1^{a_1}p_2^{a_2}...p_r^{a_r}$ and $p_1, p_2, ..., p_r$ are distinct primes and $a_1, a_2, ..., a_r > 0$.

17 is already prime so we can calculate $\Phi(17)$ as:

$$\Phi(17) = 17^{1-1}(17 - 1) = 16$$

289 need to be factorized before we can calculate Φ.

$$289 = 17^2$$

$$\Phi(289) = 17^{2-1}(17 - 1) = 17 * 16 = 272$$

And the same thing for 221:

$$221 = 13 + 17$$

$$\Phi(221) = 13^{1-1}(13 - 1) + 17^{1-1}(17 - 1) = 12 * 16 = 192$$

Question 5

The following has been enciphered with a Vigenère cipher using a keyword of length less than 12. Find the length of the keyword, hence the keyword and finally decrypt at least the first line of the text.

HRMWNUURSEAE UCSMNKTYQWNC GNHEEPQUVWAI UIGGMVVOFMRP EAIKAIPOXLOG

Answer 5

We can use the fact that common trigrams and digrams will occur to find the *gcd* of their positions to determine the key length.

Counting trigrams we get 3 "KAI"s at positions 52, 244, 358. And 3 "EAI"s at 49, 241, 355. Counting digrams we get 5 "SR"s at positions 89, 149, 185,...

Calculating the gcd:

$$\begin{aligned} \text{KAI} &= 244 - 52 = 192 \\ &= 358 - 52 = 306 \\ &= 358 - 244 = 114 \end{aligned}$$

$$\begin{aligned} \text{EAI} &= 241 - 49 = 192 \\ &= 355 - 49 = 306 \\ &= 355 - 241 = 114 \end{aligned}$$

$$\begin{aligned} \text{SR} &= 185 - 89 = 93 \\ &= 185 - 149 = 36 \\ &= 149 - 89 = 60 \end{aligned}$$

$$\begin{aligned} gcd(192, 306) &= \mathbf{6} \\ gcd(192, 114) &= \mathbf{6} \\ gcd(93, 192) &= \mathbf{3} \\ gcd(36, 114) &= \mathbf{6} \end{aligned}$$

There are a few $gcd(a, b) = 3$, but the majority has $gcd(a, b) = 6$, so I make a guess that the key length is 6.

The next step in solving the ciphertext is to divide it into 6 substrings / columns

(based on the key length), and start counting the frequency of letters.

Column 1
RRCYNUIOAOAMIMITMOTRHBOBCTLUTURBSETRSAEYANDELFSEOUHANTLNLOEAFL

Column 2
RRCYNUIOAOAMIMITMOTRHBOBCTLUTURBSETRSAEYANDELFSEOUHANTLNLOEAFL

Column 3
MSSQHVGFIXMXPIZIXHMIXSPIELIWSGAMMVAMJRZLIWMYIFEVVWSFWLESIQEIY

Column 4
WEMWEWGMKLKZLFWJZAFVZFWOWWTZDSSLXWSWSVGSKOLFSJFWTAFDGWDMEWCKF

Column 5
NANNEAMRAOEEHDSTESTWEETISNRADESIISSVUGUTAEHDVUDSRSOEAYLREIIAE

Column 6
UEKCPIVPIGYVRFRYXFVZZJZKRFLKPJRFKFRFCIJYIIVVVKKKLRLDIRYREJEII

Now each column individually, is the same as Ceasar Chipers. The 4 most common letters of each column is:

Column 1

Letter	Amount	Frequency (%)
T	8	12.9
V	7	11.29
Q	5	8.06
U	5	8.06

Column 2

Letter	Amount	Frequency (%)
O	6	9.68
A	6	9.68
T	6	9.68
R	5	8.06

Column 3

Letter	Amount	Frequency (%)
I	10	16.39
M	7	11.48
S	6	9.84
E	4	6.56

Column 4

Letter	Amount	Frequency (%)
W	12	19.67
F	7	11.48
S	6	9.84
K	4	6.56

Column 5

Letter	Amount	Frequency (%)
E	11	18.03
S	8	13.11
A	7	11.48
I	5	8.2

Column 6

Letter	Amount	Frequency (%)
I	8	13.11
R	8	13.11
K	7	11.48
V	6	9.84

The most common letter in the English alphabet is "e", so I make the guess that the columns with the largest difference in letter frequency to be "e".

So column 3: "I" → "e" (shift 4, key "E"), column 4: "W" → "e" (shift 18, key "S"). And column 5: "E" → "e" (shift 0, key "A"), even though that would mean it's not encrypted.

So far the first 6 letters of the ciphertext looks like this: "HRienU". I make the guess that the letter "U" → "d", as I think it's a common word ending, and it fits with "-Riend". Thus column 6: "I" → "r" (shift 17, key "R").

At this point I have a theory that the first 6 letters form the word "friend". To test this theory I begin by shifting the letter "R" → "r". Thus column 2: "O" → "o" (shift 0, key "A").

At this point words are almost complete in many parts of the ciphertext, so I try shifting "H" → "f". Thus column 1: "T" → "r" (shift 2, key "C").

Key:

Column	Shift	Letter
1	2	C
2	0	A
3	4	E
4	18	S
5	0	A
6	17	R

This gives me the first line of the plaintext:

friends romans countrymen lend me your ears i come to bury caesar not top

Question 6

(a)

Explain what Friedman's Index of Coincidence measures and compute it for the ciphertext below, hence explain how you can see that the cipher used to encrypt it was not monoalphabetic.

Answer (a)

The *Index of Coincidence* (IOC) I, measures the likelihood of picking 2 identical letters, from a text. If I(English) $\approx I$(ciphertext) then the encryption form is monoalphabetic.

We can calculate the IOC I, by using the counted letters, given in the assignment. Where the letters: a,b,c...,z are represented as $n_0, n_1, ..., n_{25}$, and n is the total number of letters.

$$\begin{aligned} I &= \frac{\sum_0^{25} n_i(n_i - 1)}{n(n-1)} \\ &= \frac{26(26-1) + 20(20-1) + ... + 25(25-1)}{444 * 443} \\ &= \frac{8638}{196692} \\ &\approx 0.44 \end{aligned}$$

(b)

Actually the ciphertext has been encrypted by using a Vigenère cipher with keylength 5. Explain how the χ^2-test for goodness of fit can be used to guess that the first letter of the keyword is **f**. Include in your explanation full calculations to show that a guess of **f** for first letter of the key is a better guess than the letter **z**.

Answer (b)

The χ^2-test can be used to determine the best shift. This is determined calculating χ^2(shift n) for each shift, where the lowest value (deviation) is the best shift amount.

We begin by extracting the first letter of each column in the cipher text. This gives us the cipher text: KYXYZMJGYHMFNYNXTYXFDSQJMTPJYXFQYTF-SJNIMJJJWNYLWRNFXHXMQJNAFYMMFZJYXPJXNMMJJMINJXTXJKWHYI

We want to calculate χ^2 when "a" is shifted to an "f" (shift 5). And we want to calculate χ^2 when "a" is shifted to an "z" (shift 25).

χ^2 is calculated with the following formula

$$\chi^2 = \sum \frac{(a_i - e_i)^2}{e_i}$$

where a_i is the number of letters: a,b,c,...,z. And e_i is the expected frequency of the letter in the English alphabet (Pia's Block 1 notes, P.56).

Letter	Expected freq. e_i	Observed frequency a_i	Shift 5	Shift 25
A	5.30	1	7	2
B	0.89	0	1	1
C	1.44	0	3	0
D	2.98	1	3	0
E	8.20	0	14	1
F	1.21	7	2	0
G	1.53	1	1	7
H	4.45	3	10	1
I	4.53	3	8	3
J	0.09	14	0	3
K	0.70	2	2	14
L	2.84	1	3	2
M	1.27	10	1	1
N	4.24	8	2	10
O	4.90	0	4	8
P	0.91	2	0	0
Q	0.12	3	0	2
R	3.26	1	3	3
S	3.93	2	10	1
T	6.45	4	11	2
U	2.09	0	2	4
V	0.51	0	1	0
W	1.62	3	0	0
X	0.08	10	0	3
Y	1.37	11	1	10
Z	0.05	2	0	11

$$\begin{aligned}
\chi^2(\text{shift } 5) &= \sum \frac{(a_i - e_i)^2}{e_i} \\
&= \frac{(7 - 5.3)^2}{5.3} + \frac{(1 - 0.89)^2}{0.89} + ... + \frac{(0 - 0.05)^2}{0.05} \\
&\approx 36.51
\end{aligned}$$

$$\begin{aligned}\chi^2(\text{shift } 25) &= \sum \frac{(a_i - e_i)^2}{e_i} \\ &= \frac{(2-5.3)^2}{5.3} + \frac{(1-0.89)^2}{0.89} + ... + \frac{(11-0.05)^2}{0.05} \\ &\approx 2992.16\end{aligned}$$

This means that χ^2(shift 5) is the preferred shift amount for column 1, as it's lower than the value of χ^2(shift 11).

(c)

Decrypt the first line of the ciphertext below, given that it has been encrypted by using a Vigenère cipher with the keyword **fishy**.

KQJZR YPWMG XPEBQ YJWJY ZOZAR MILPQ JIKFY GITFG YPAUI HWMSB MINLA FCYOR

Answer (c)

Using the formula we can get the ciphertext y_i:

$$y_i = (x_i + k_{iMODn})MOD26$$

In this case, we are looking for the plaintext letters x_i.

The key: **fishy**, corresponds to the sequence (5,8,18,7,24) as y_i.

If we want to decrypt the cipher text "KQJZR", which in numeric corresponds to "10,16,9,25,17", we start by using $y_0 = 10$ to decrypt K:

$$\begin{aligned}10 &= (x_0 + 5_{0MOD5})MOD26 \\ 10 &= (x_0 + 5_0)MOD26 \\ 5 &= x_0 MOD26 \\ x_0 &= 5\end{aligned}$$

So the first letter "K" $\rightarrow$ "f". Using this formula for the remaining 4 ciphertext letters, we get "KQJZR" $\rightarrow$ "first".

The next 5 letters "YPWMG" (24,15,22,12,6), is decrypted the same way:

$$\begin{aligned}24 &= (x_0 + 5_{0MOD5})MOD26 \Rightarrow x_0 = 19 \text{ (t)} \\ 15 &= (x_0 + 8_{1MOD5})MOD26 \Rightarrow x_1 = 7 \text{ (h)} \\ 22 &= (x_0 + 18_{2MOD5})MOD26 \Rightarrow x_2 = 4 \text{ (e)} \\ 12 &= (x_0 + 7_{3MOD5})MOD26 \Rightarrow x_3 = 5 \text{ (f)} \\ 6 &= (x_0 + 24_{4MOD5})MOD26 \Rightarrow x_4 = 8 \text{ (i)}\end{aligned}$$

Continuing the same way for the ramainder of the ciphertext, we get the plaintext:

first thefi shmus tbeca ughtt hatis easya babyi think could havec aught

first the fish must be caught that is easy a baby i think could have caught