Vigenère Cipher

Math 4175

§2.4.1. Vigenère Cipher

In cryptosystems we learned so far, each fixed key maps every alphabet in the plain text to a unique alphabet in the cipher text. So these cryptosystems are called monoalphabetic cryptosystems.

Now we will consider other cryptosystems called polyalphabetic cryptosystems where a key maps an alphabetic character in the plain text to $m > 1$ possible alphabets in the cipher text.

We will start with a well known polyalphabetic cryptosystem called Vigenère Cipher.

In Vigenère Cipher, each key $k \in \mathcal{K}$ is a key word, which is an alphabetic string of length m, for any fixed positive integer m.

§2.4.1. Vigenère Cipher

We will begin with an example. Suppose that the key word is CIPHER with $m = 6$.

Using the correspondence A $= 0$, B $= 1$, $\cdots$, Z $= 25$ as described earlier, this corresponds to the numerical equivalent $K = (2, 8, 15, 7, 4, 17)$.

Suppose that the plaintext is the string

`hereishowitworks`

We convert the plaintext alphabetic characters to numerals modulo 26. To encrypt the message using the given above key, we shift the first letter of the plaintext by 2. Then shift the second letter by 8, the third by 15, and so on. Once we get to the end of the key, we start back at its first entry, so the seventh letter is shifted by 2, the eight letter by 4, etc. and then convert it back to alphabets as follows:

§2.4.1. Vigenère Cipher

h	e	r	e	i	s	h	o	w	i	t	w	o	r	k	s
7	4	17	4	8	18	7	14	22	8	19	22	14	17	10	18
2	8	15	7	4	17	2	8	15	7	4	17	2	8	15	7
9	12	6	11	12	9	9	22	11	15	23	13	16	25	25	25
J	M	G	L	M	J	J	W	L	P	X	N	Q	Z	Z	Z

To decrypt, we can use the same keyword, but we would subtract it modulo 26 from the cipher text instead of adding.

§2.4.1. Vigenère Cipher

Vigenère Cipher is a cryptosystem with $\mathcal{P} = \mathcal{C} = \mathcal{K} = \left(\mathbb{Z}_{26}\right)^m$ where m is a positive integer. For a key $K = (k_1, k_2, \cdots, k_m)$, we define

$$e_K(x_1, x_2, \cdots, x_m) = (x_1 + k_1, x_2 + k_2, \cdots, x_m + k_m)$$

$$d_K(y_1, y_2, \cdots, y_m) = (y_1 - k_1, y_2 - k_2, \cdots, y_m - k_m)$$

where all operations are performed in $\mathbb{Z}_{26}$.

- For a Vigenère Cipher with key word length m, the number of possible key is 26^m.
- For example, if $m = 5$, the number of possible keys exceeds 11 millions. So it is not feasible to do exhaustive key search by hand without the help of a computer.
- In general, the cryptanalysis is more difficult for polyalphabetic than monoalphabetic cryptosystem.

§2.4.2. Cryptanalysis of Vigenère Cipher

Cipher text only attack: As a cryptanalyst of Vigenère Cipher, we need to find the key word from the ciphered text in two steps.

- Find the length of the key word, which is a tricky part.
- Find the actual key word.

We will give three methods to find the length of the key word.

Consider the following ciphered text:

```
DBMOKWWODJPAOPEPHGQAHWVUJNXDZA
IAQMRPNEMHGPERZOVKRVLWUGAEOPXE
HMXDVBXDDVKOYQHJJBLWKXIJDVUQDB
IPCMAWTBLATKSIZJEYFBSIZBSZVG
```

§2.4.2. Cryptanalysis of Vigenère Cipher

Method 1: Displacement-Coincidences

Write the ciphered text on a long strip of paper, and again on another long strip. Put one strip just above the other, but displaced by a certain number of places (the potential key length). For example,

Displacement of two units with coincidences highlighted

*	*	D	B	M	O	K	W	W	O	D	J	P	A	O	P	E	P	H	G	Q	A	H	W	V	U	J	N	X	D
D	B	M	O	K	W	W	O	D	J	P	A	O	P	E	P	H	G	Q	A	H	W	V	U	J	N	X	D	Z	A

Z	A	I	A	Q	M	R	P	N	E	M	H	G	P	E	R	Z	O	V	K	R	V	L	W	U	G	A	E	O	P
I	A	Q	M	R	P	N	E	M	H	G	P	E	R	Z	O	V	K	R	V	L	W	U	G	A	E	O	P	X	E

X	E	H	M	X	D	V	B	X	D	D	V	K	O	Y	Q	H	J	J	B	L	W	K	X	I	J	D	V	U	Q
H	M	X	D	V	B	X	D	D	V	K	O	Y	Q	H	J	J	B	L	W	K	X	I	J	D	V	U	Q	D	B

D	B	I	P	C	M	A	W	T	B	L	A	T	K	S	I	Z	J	E	Y	F	B	S	I	Z	B	S	Z	V	G
I	P	C	M	A	W	T	B	L	A	T	K	S	I	Z	J	E	Y	F	B	S	I	Z	B	S	Z	V	G	*	*

§2.4.2. Cryptanalysis of Vigenère Cipher

Displacement of three units with coincidences highlighted

*	*	*	D	B	M	O	K	W	W	O	D	J	P	A	O	P	E	P	H	G	Q	A	H	W	V	U	J	N	X
D	B	M	O	K	W	W	O	D	J	P	A	O	P	E	P	H	G	Q	A	H	W	V	U	J	N	X	D	Z	A

D	Z	A	I	A	Q	M	R	P	N	E	M	H	G	P	E	R	Z	O	V	K	R	V	L	W	U	G	A	E	O
I	A	Q	M	R	P	N	E	M	H	G	P	E	R	Z	O	V	K	R	V	L	W	U	G	A	E	O	P	X	E

P	X	E	H	M	X	D	V	B	X	D	D	V	K	O	Y	Q	H	J	J	B	L	W	K	X	I	J	D	V	U
H	M	X	D	V	B	X	D	D	V	K	O	Y	Q	H	J	J	B	L	W	K	X	I	J	D	V	U	Q	D	B

Q	D	B	I	P	C	M	A	W	T	B	L	A	T	K	S	I	Z	J	E	Y	F	B	S	I	Z	B	S	Z	V
I	P	C	M	A	W	T	B	L	A	T	K	S	I	Z	J	E	Y	F	B	S	I	Z	B	S	Z	V	G	*	*

G
*

§2.4.2. Cryptanalysis of Vigenère Cipher

Displacement of four units with coincidences highlighted

*	*	*	*	D	B	M	O	K	W	W	O	D	J	P	A	O	P	E	P	H	G	Q	A	H	W	V	U	J	N	X	D	Z	A
D	B	M	O	K	W	W	O	D	J	P	A	O	P	E	P	H	G	Q	A	H	W	V	U	J	N	X	D	Z	A				

X	D	Z	A	I	A	Q	M	R	P	N	E	M	H	G	P	E	R	Z	O	V	K	R	V	L	W	U	G	A	E
I	A	Q	M	R	P	N	E	M	H	G	P	E	R	Z	O	V	K	R	V	L	W	U	G	A	E	O	P	X	E

O	P	X	E	H	M	X	D	V	B	X	D	D	V	K	O	Y	Q	H	J	J	B	L	W	K	X	I	J	D	V
H	M	X	D	V	B	X	D	D	V	K	O	Y	Q	H	J	J	B	L	W	K	X	I	J	D	V	U	Q	D	B

U	Q	D	B	I	P	C	M	A	W	T	B	L	A	T	K	S	I	Z	J	E	Y	F	B	S	I	Z	B	S	Z
I	P	C	M	A	W	T	B	L	A	T	K	S	I	Z	J	E	Y	F	B	S	I	Z	B	S	Z	V	G	*	*

V	G
*	*

§2.4.2. Cryptanalysis of Vigenère Cipher

Number of coincidences for cipher text is displacements:

displacements	1	2	3	4	5	6	7	8	9	10	11
coincidences	3	2	3	10	8	1	1	9	2	2	3

We have most coincidences for the displacement of 4. So, this is the best guess for the length of the key by this method.

This method works very quickly, even without a computer, and usually (but may not always) yields the key length.

§2.4.2. Cryptanalysis of Vigenère Cipher

Method 2: Kasiski Test (Friedrich Kasiski - 1863)

The Kasiski test involves looking for repeated digrams, trigrams, etc in the cipher text.

- Two identical segments of plaintext will be encrypted to the same cipher text whenever their occurrence in the plaintext is δ positions apart, where $\delta \equiv 0 \mod m$.

- If we observe two identical segments of cipher text, each of length at least two, there is a good chance that they correspond to identical segments in plaintext, particularly when they are repeated many times.

§2.4.2. Cryptanalysis of Vigenère Cipher

More precisely, we search for pairs of identical segments of lengths at least two, and record the distance between the starting positions of the two segments. If we obtain several such distances, say $\delta_1, \delta_2, \cdots$, then we would assume that m divides all of the $\delta_i's$, and hence m divides the greatest common divisor of the $\delta_i's$.

Example: Consider the cipher text:

OPKVZQYEOPORBPKVZ

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
O	P	K	V	Z	Q	Y	E	O	P	O	R	B	P	K	V	Z

The digram OP appears twice, initiating in positions 1 and 9. The trigram KVZ appears twice initiating in positions 3 and 15. The most reasonable guess for the length of the keyword is $\gcd(8, 12) = 4$.

§2.4.2. Cryptanalysis of Vigenère Cipher

Method 3: Friedman Test (William Friedman (1920)

Further evidence for the value of m can be obtained by the index of coincidence, denoted by $I_c(\mathbf{x})$ as follows:

Definition: Suppose $\mathbf{x} = x_1 x_2 \cdots x_n$ is a string of n alphabetic characters. Then $I_c(x)$ is defined to be the probability that two randomly selected characters (without replacement) in $\mathbf{x}$ are equal.

Recall: The number of ways to select r elements from a set of n elements (without replacement) is given by the formula:

$$nCr = \frac{n!}{r!(n-r)!}$$

In particular, $nC2 = \dfrac{n(n-1)}{2}$.

§2.4.2. Cryptanalysis of Vigenère Cipher

Let us denote the frequencies of A, B, C, $\cdots$, Z in a string **x** of n characters by $f_0, f_1, \cdots, f_{25}$ respectively. Suppose that we select two alphabets randomly without replacement.

Probability that both are A $= \frac{\text{Number of ways to select two A's}}{\text{Number of ways to select any two alphabets}}$.

$$Pr[\text{Both } A] = \frac{f_0 C2}{nC2} = \frac{f_0(f_0-1)}{n(n-1)}.$$

Similarly, $Pr[\text{Both } B] = \frac{f_1 C2}{nC2} = \frac{f_1(f_1-1)}{n(n-1)}$ and so on.

Hence, we have the following formula: $I_c(\mathbf{x}) = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)}$

§2.4.2. Cryptanalysis of Vigenère Cipher

Example 1: Find the index of coincidence for the string

$$\mathbf{x} = \texttt{testinglettersincipher}$$

Solution: There are 4 e's, 3 i's, 2 n's, 2 r's, 2 s's, 4 t's, and one each of c, g, p, h, and l. So:

$$I_c(\mathbf{x}) = \frac{4(3) + (3)(2) + (2)(1) + (2)(1) + (2)(1) + 4(3)}{22(21)} = \frac{6}{77} \approx 0.078$$

Example 2: Given a text $\mathbf{x}$ with 2600 characters where each character appears 100 times, find the index of coincidence of $\mathbf{x}$?

$$\text{Answer: } I_c(\mathbf{x}) = \frac{26(100)(99)}{2600(2599)} = \frac{99}{2599} \approx 0.038$$

§2.4.2. Cryptanalysis of Vigenère Cipher

Limiting Case: If we are dealing with a text **x** containing a large number of letters, then

$$\frac{f_i}{n} \approx \frac{f_i - 1}{n - 1} \approx p_i$$

where p_i's are probabilities of alphabets given in the frequency table and so

$$I_c(\mathbf{x}) \approx \sum_{i=0}^{25} p_i^2 \approx 0.065$$

Observation: A completely random string in English will have the index of coincidence:

$$I_c = \sum_{i=0}^{25} \left(\frac{1}{26}\right)^2 = 26\left(\frac{1}{26}\right)^2 = \frac{1}{26} \approx 0.038$$

§2.4.2. Cryptanalysis of Vigenère Cipher

Step to recover key word length: Guess the keyword length m as $1, 2, 3, 4, \cdots$ and then use statistical tests given before to check which guess looks good.

Example: Suppose that our cipher text is $\mathbf{y} = y_1 y_2 y_3 \cdots$ and we guess that $m = 4$. We create four substrings of $\mathbf{y}$, denoted $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_4$; by taking every 4^{th} letter of the ciphertext starting at positions $1, 2, 3, 4$. That is,

$$\begin{aligned}
\mathbf{y}_1 &= y_1 y_5 y_9 \cdots \\
\mathbf{y}_2 &= y_2 y_6 y_{10} \cdots \\
\mathbf{y}_3 &= y_3 y_7 y_{11} \cdots \\
\mathbf{y}_4 &= y_4 y_8 y_{12} \cdots
\end{aligned}$$

We then calculate the index of coincidence of each substring; that is, $I_c(\mathbf{y}_i)$ for $i = 1, 2, 3, 4$.

§2.4.2. Cryptanalysis of Vigenère Cipher

- If $m = 4$ is correct, then each value $I_c(\mathbf{y}_i)$ should be roughly equal to 0.065.
- If $m = 4$ is wrong, then the substrings $\mathbf{y}_i$ will look much more random, since they will have been obtained by shift encryption with different keys. In this case, the indices of coincidences of the substrings will be lower than 0.065.

Important: The two values 0.065 and 0.038 are sufficiently far apart that we will often be able to determine the correct keyword length by this method (or confirm a guess that has already been made using previous tests)

§2.4.2. Cryptanalysis of Vigenère Cipher

Recall the ciphered text:

DBMOKWWODJPAOPEPHGQAHWVUJNXDZA
IAQMRPNEMHGPERZOVKRVLWUGAEOPXE
HMXDVBXDDVKOYQHJJBLWKXIJDVUQDB
IPCMAWTBLATKSIZJEYFBSIZBSZVG

We deduced last time that $m = 4$. So we have

$$\mathbf{y}_1 = \texttt{DKDOHHJZQNGZRUOHVDYJKDDCTTZFZV}$$
$$\mathbf{y}_2 = \texttt{BWJPGWNAMEPOVGPMBVQBXVBMBKJBBG}$$
$$\mathbf{y}_3 = \texttt{MWPEQVXIRMEVLAXXXKHLIUIALSESS}$$
$$\mathbf{y}_4 = \texttt{OOAPAUDAPHRKWEEDDOJWJQPWAIYIZ}$$

Exercise: Calculate the index of coincidences for the above ciphered text by assuming $m = 1, 2, 3, 4, 5$ and verify whether this method supports our guess that $m = 4$.

§2.4.2. Cryptanalysis of Vigenère Cipher

For $m = 1$:

Substring ($\mathbf{y}_1$):
DBMOKWWODJPAOPEPHGQAHWVUJNXDZAIAQMRPNEMHGPERZOVKRVLWUGAEOP
XEHMXDVBXDDVKOYQHJJBLWKXIJDVUQDBIPCMAWTBLATKSIZJEYFBSIZBSZVG

Index of Coincidence: 0.037520

For $m = 2$:

Substring ($\mathbf{y}_1$):
DMKWDPOEHQHVJXZIQRNMGEZVRLUAOXHXVXDKYHJLKIDUDICATLTSZEFSZSV

Index of Coincidence: 0.033314

Substring ($\mathbf{y}_2$):
BOWOJAPPGAWUNDAAMPEHPROKVWGEPEMDBDVOQJBWXJVQBPMWBAKIJYBIBZG

Index of Coincidence: 0.050263

§2.4.2. Cryptanalysis of Vigenère Cipher

For $m = 3$:

Substring ($\mathbf{y}_1$):
DOWJOPQWJDIMNHEORWAPHDXVYJLXDQIMTASJFISG

Index of Coincidence: 0.035897

Substring ($\mathbf{y}_2$):
BKOPPHAVNZAREGRVVUEXMVDKQJWIVDPABTIEBZZ

Index of Coincidence: 0.039136

Substring ($\mathbf{y}_3$):
MWDAEGHUXAQPMPZKLGOEXBDOHBKJUBCWLKZYSBV

Index of Coincidence: 0.029690

§2.4.2. Cryptanalysis of Vigenère Cipher

For $m = 4$:

Substring ($\mathbf{y}_1$):
DKDOHHJZQNGZRUOHVDYJKDDCTTZFZV

Index of Coincidence: 0.055172

Substring ($\mathbf{y}_2$):
BWJPGWNAMEPOVGPMBVQBXVBMBKJBBG

Index of Coincidence: 0.080460

Substring ($\mathbf{y}_3$):
MWPEQVXIRMEVLAXXXKHLIUIALSESS

Index of Coincidence: 0.051724

Substring ($\mathbf{y}_4$):
OOAPAUDAPHRKWEEDDOJWJQPWAIYIZ

Index of Coincidence: 0.051724

§2.4.2. Cryptanalysis of Vigenère Cipher

For $m = 5$:

Substring ($\mathbf{y}_1$): DWPPHNIPGOLEHBKJKVIWTJSZ
Index of Coincidence: 0.028986

Substring ($\mathbf{y}_2$): BWAHWXANPVWOMXOJXUPTKEIV
Index of Coincidence: 0.036232

Substring ($\mathbf{y}_3$): MOOGVDQEEKUPXDYBIQCBSYZG
Index of Coincidence: 0.025362

Substring ($\mathbf{y}_4$): ODPQUZMMRRGXDDQLJDMLIFB
Index of Coincidence: 0.047431

Substring ($\mathbf{y}_5$): KJEAJARHZVAEVVHWDBAAZBS
Index of Coincidence: 0.071146

The closest indices of coincidence to .065 occur when our guess for m is 4. Therefore, $m = 4$ is the correct guess!

§2.4.2. Cryptanalysis of Vigenère Cipher

Assuming we have guessed the correct value of m, now we will describe an effective method to determine the actual key $K = (k_1, k_2, \cdots, k_m)$.

- Let $\vec{p} = (p_0, p_1, \cdots, p_{25}) = (0.082, 0.015, \cdots, 0.001) \in \mathbb{R}^{26}$ where p_i's are as given in frequency table.
- For each string $\mathbf{y}_i$ $(1 \leq i \leq m)$, let

$$\vec{q} = (q_0, q_1, \cdots, q_{25}) = \left(\frac{f_0}{N}, \frac{f_1}{N}, \cdots, \frac{f_{25}}{N}\right),$$

  where N is the length of the string $\mathbf{y}_i$, and f_i's are the number of times A, B, $\cdots$, Z occurs respectively in the string $\mathbf{y}_i$.
- For each $0 \leq g \leq 25$, let $\vec{v}_g$ be $\vec{q}$ cyclically shifted by g steps to the right, i.e.

$$\vec{v}_g = (q_g, q_{1+g}, \cdots, q_{25+g}) = \left(\frac{f_g}{N}, \frac{f_{1+g}}{N}, \cdots, \frac{f_{25+g}}{N}\right)$$

§2.4.2. Cryptanalysis of Vigenère Cipher

- Compute the dot product $M_g = \vec{p} \cdot \vec{v}_g$ for $0 \leq g \leq 25$.
- By assuming our guess for m is correct, the characters in the string $\mathbf{y}_i$ are obtained from a plaintext by shift encryption using a shift k_i. So we expect that for $g = k_i$, $M_g \approx 0.065$ and M_g will be significantly smaller for $g \neq k_i$,
- Then one can decrypt with this key to check whether the key is a valid one.

Let us compute M_g's for our example to determine the key.

§2.4.2. Cryptanalysis of Vigenère Cipher

g	1 M_g	2 M_g	3 M_g	4 M_g
0	3.41	2.45	4.61	4.78
1	3.47	4.7	3.33	3.62
2	4.1	4.46	2.67	3.36
3	5.15	3.26	3.89	4.32
4	2.07	3.4	6.77	3.74
5	3.64	3.33	3.42	3.12
6	4.73	3	3.19	3.22
7	4.21	3.39	4.21	4.47
8	2.68	6.29	4.55	4.06
9	3.13	4.58	3.36	3.67
10	4.48	3.93	3.53	4.01
11	4.02	2.79	3.87	4.59
12	4	3.28	3.8	4.24
13	3.23	4.75	3.23	3.16
14	4.08	4.23	3.38	3.4
15	3.95	3.37	4.18	4.29
16	4.23	3.07	4.18	4.14
17	4.56	3.2	3.89	3.24
18	3.8	3.54	4.22	4.26
19	2.95	5.11	5.09	2.86
20	2.96	3.57	3.69	2.8
21	5.85	3.95	2.67	4.26
22	3.96	3.99	3.7	6.78
23	2.85	5.53	4.52	3.78
24	2.88	3.99	2.99	2.96
25	5.81	3.07	3.28	3.09

§1.2.4. Cryptanalysis of Vigenère Cipher

Hence, the keyword K should be $K = (21, 8, 4, 22)$. That is,

$$K = (v, i, e, w).$$

Using the keyword $K = (v, i, e, w)$ as indicated by our coincidence index calculations, we get:

```
itispossiblethatmymemoryoftheseeventswill
havegrownhazywithtimethatthingsdidnothappen
inquitethewaytheycomebacktometoday
```

More precisely,

```
it is possible that my memory of these events will
have grown hazy with time that things did not happen
in quite the way they come back to me today
```