Prof. Dr. Rudolf Mathar, Dr. Arash Behboodi, Markus Rothe

Exercise 3
- Proposed Solution -

Friday, May 4, 2018

Solution of Problem 1

The first step is to perform a frequency analysis. The frequency analysis consists of counting the number of appearances of each letter and the number of ordered pairs as follows:

i	Character	k_i	p_i
0	A	1	0
1	B	4	6
2	C	12	66
3	D	4	6
4	E	2	1
5	F	1	0
6	G	6	15
7	H	2	1
8	I	7	21
9	J	2	1
10	K	14	91
11	L	7	21
12	M	1	0

i	Character	k_i	p_i
13	N	3	3
14	O	4	6
15	P	4	6
16	Q	4	6
17	R	10	45
18	S	4	6
19	T	0	0
20	U	0	0
21	V	9	36
22	W	6	15
23	X	8	28
24	Y	3	3
25	Z	0	0

- k_i the total number of ocurrences of each letter
- p_i the number of ordered pairs calculated as $p_i = \binom{k_i}{2}$

The next step is to calculate the total length of the ciphertext using that the text is divided in blocks:

$$n = 14 \cdot 8 + 6 = 118 \tag{1}$$

Now we have to use the formula of the index of coincidence as follows *(Def: 3.1, page 13, lecture notes):*

$$I_C = \frac{|\{i,j\}|C_i = C_j, 1 \leq i < j \leq n\}|}{\binom{n}{2}} = \frac{\sum_{i=0}^{25}\binom{k_i}{2}}{\binom{n}{2}} \tag{2}$$

For 2, we calculate all the possible pair of combinations of n symbols as:

$$\binom{n}{2} = \frac{n!}{(n-2)!\,2!} = \frac{n(n-1)}{2!} = \frac{118 \cdot 117}{2} = 6903 \tag{3}$$

Therefore, substituting in 2, we have:

$$I_C = \frac{6 \cdot 0 + 3 \cdot 1 + 2 \cdot 3 + 6 \cdot 6 + 2 \cdot 15 + 2 \cdot 21 + 1 \cdot 28 + 1 \cdot 36 + 1 \cdot 45 + 1 \cdot 66 + 1 \cdot 91}{6093}$$

$$= \frac{383}{6903} = 0.055483 \tag{4}$$

Using the hint of the exercise, we can assume that the text is monoalphabetic and probably in English, because the index of coincidence obtained is close to its value.

The plaintext is the following:

All the world's a stage and all the men and women merely players: They have their exits and their entrances, and one man in his time plays many parts, ...

(William Shakespeare, As You Like It, Acti II, Scene 7) - Key: Key.

Solution of Problem 2

a) We have the auto-key cryptosystem:

$$c_i = \begin{cases} m_i + k_i \pmod{26} & 0 \leq i \leq n-1 \\ m_i + c_{i-n} \pmod{26} & n \leq i \leq l-1 \end{cases}$$

Using a ciphertext only attack, we can compute the message as follows:

$$c_n = m_n + c_0 \iff m_n = c_n - c_0$$

$$c_{n+1} = m_{n+1} + c_1 \iff m_{n+1} = c_{n+1} - c_1$$

$$\implies m_{n+j} = c_{n+j} - c_j$$

Therefore, the next task is to determine n

b) Using the result from a) we decipher the following text, just shifting the ciphertext along itself:

For $n = 1$

D	L	G	V	T	Y	O	A	C	O	U	V	C	E	Z	A	
	D	L	G	V	T	Y	O	A	C	O	U	V	C	E	Z	A
	I	V	P													

For $n = 2$

D	L	G	V	T	Y	O	A	C	O	U	V	C	E	Z	A		
		D	L	G	V	T	Y	O	A	C	O	U	V	C	E	Z	A
		D	K	N													

For $n = 3$

Only the first characters are missing in the message. For these characters, we guess them. Message: THIS IS THE AUTOKEY

D	L	G	V	T	Y	O	A	C	O	U	V	C	E	Z	A			
			D	L	G	V	T	Y	O	A	C	O	U	V	C	E	Z	A
			S	I	S	T	H	E	A	U	T	O	K	E	Y			

c) Consider:

$$\hat{c}_i = \begin{cases} m_i + k_i \pmod{26} & 0 \leq i \leq n-1 \\ m_i + m_{i-n} \pmod{26} & n \leq i \leq l-1 \end{cases}$$

In this case, we know the keylength n, and we know that the message m is used to generate $\hat{c}_i$. Therefore, we can obtain the message by frequency analysis on:

$$\hat{c}_i = m_i + m_{i-n} \tag{5}$$

With a Friedmann attack, using the most common characters in the English language, we derive the most common $\hat{c}_i$'s. The message can be deciphered with a high probability then. Here, we can say 'e' is the most common english language which means that 'e'+'e'='i' is the most common $\hat{c}_i$.

d)

Q	E	X	Y	I	R	V	E	S	I	U	X	X	K	Q	V	F	L	H	K	G
T		E		E		R		B		T		E		M		T		O		S
	H		R		A		E		E		T		R		E		H		D	

The plaintext is: THERE ARE BETTER METHODS

Solution of Problem 3

a) Claim: $\mathrm{Var}(I_C) = \frac{1}{\binom{n}{2}}\left(\kappa(1-\kappa) + 2(n-2)(\beta-\kappa^2)\right)$ with $\kappa = \sum_{l=1}^{m} p_l^2$ and $\beta = \sum_{l=1}^{m} p_l^3$.

Recall that the index of coincidence can be written as $I_C = \frac{1}{\binom{n}{2}} \sum\limits_{i<j} Y_{ij}$, where $Y_{ij} = \begin{cases} 1 & C_i = C_j \\ 0 & \text{otherwise} \end{cases}$ and $E(Y_{ij}) = \sum_{l=1}^{m} p_l^2 = \kappa$.

Then it holds

$$\begin{aligned}
\mathrm{Var}(I_C) &= \mathrm{Var}\left(\frac{1}{\binom{n}{2}^2} \sum_{i<j} Y_{ij}\right) \\
&= \frac{1}{\binom{n}{2}^2}\left(\sum_{i<j} \mathrm{Var}(Y_{ij}) + \sum_{i<j} \sum_{\substack{k<l \\ (i,j)\neq(k,l)}} \mathrm{Cov}(Y_{ij}, Y_{kl})\right) \\
&\overset{(1)}{=} \frac{1}{\binom{n}{2}^2}\left(\sum_{i<j} \mathrm{Var}(Y_{ij}) + 2\sum_{i<j} \sum_{\substack{k<l \\ (i,j)<(k,l)}} \mathrm{Cov}(Y_{ij}, Y_{kl})\right) \\
&\overset{(2),(3)}{=} \frac{1}{\binom{n}{2}^2}\left(\sum_{i<j} \kappa(1-\kappa) + n(n-1)(n-2)(\beta-\kappa^2)\right) \\
&= \frac{1}{\binom{n}{2}}\left(\kappa(1-\kappa) + 2(n-2)(\beta-\kappa^2)\right)
\end{aligned}$$

(1) This equality holds for the definition: $(i,j) < (k,l) \Leftrightarrow i < k \vee ((i = k) \wedge (j < l))$.

(2) It holds

$$\mathrm{Var}(Y_{ij}) = E(Y_{ij}^2) - E(Y_{ij})^2 = 1^2 \cdot P(Y_{ij} = 1) - \kappa^2 = E(Y_{ij}) - \kappa^2 = \kappa(1-\kappa).$$

(3) For the covariance it holds

$$\mathrm{Cov}(Y_{ij}, Y_{kl}) = E(Y_{ij} \cdot Y_{kl}) - E(Y_{ij}) \cdot E(Y_{kl}) = E(Y_{ij} \cdot Y_{kl}) - \kappa^2$$

Investigate: $Y_{ij} \cdot Y_{kl} = 1 \Leftrightarrow Y_{ij} = Y_{kl} = 1 \Leftrightarrow C_i = C_j \wedge C_k = C_l$.

There are four disjoint cases (i)-(iv):

(i) $\mathbf{i = k}$: In that case $j < l$ must hold with respect to (1). Hence,

$$\begin{aligned}
&Y_{ij} \cdot Y_{il} = 1 \Leftrightarrow Y_{ij} = Y_{il} = 1 \Leftrightarrow C_i = C_j \wedge C_i = C_l \Leftrightarrow C_i = C_j = C_l \\
\Rightarrow &E(Y_{ij} \cdot Y_{il}) = 1 \cdot P(Y_{ij} \cdot Y_{il} = 1) = P(C_i = C_j = C_l) = \sum_{n=1}^{m} p_n^3 = \beta \\
\Rightarrow &\mathrm{Cov}(Y_{ij}, Y_{il}) = E(Y_{ij} \cdot Y_{il}) - \kappa^2 = \beta - \kappa^2 = \alpha
\end{aligned}$$

(ii) $\mathbf{i < k \wedge j = l}$: In that case it holds

$$\begin{aligned}
&Y_{ij} \cdot Y_{kj} = 1 \Leftrightarrow Y_{ij} = Y_{kj} = 1 \Leftrightarrow C_i = C_j \wedge C_k = C_j \Leftrightarrow C_i = C_j = C_k \\
\Rightarrow &E(Y_{ij} \cdot Y_{kj}) = 1 \cdot P(Y_{ij} \cdot Y_{jl} = 1) = P(C_i = C_j = C_l) = \sum_{n=1}^{m} p_n^3 = \beta \\
\Rightarrow &\mathrm{Cov}(Y_{ij}, Y_{jl}) = E(Y_{ij} \cdot Y_{kj}) - \kappa^2 = \beta - \kappa^2 = \alpha
\end{aligned}$$

(iii) $\mathbf{i < k \wedge j = k}$: In that case it holds

$$\begin{aligned}
&Y_{ij} \cdot Y_{jl} = 1 \Leftrightarrow Y_{ij} = Y_{jl} = 1 \Leftrightarrow C_i = C_j \wedge C_l = C_j \Leftrightarrow C_i = C_j = C_l \\
\Rightarrow &E(Y_{ij} \cdot Y_{jl}) = 1 \cdot P(Y_{ij} \cdot Y_{kj} = 1) = P(C_i = C_j = C_k) = \sum_{n=1}^{m} p_n^3 = \beta \\
\Rightarrow &\mathrm{Cov}(Y_{ij}, Y_{kj}) = E(Y_{ij} \cdot Y_{kj}) - \kappa^2 = \beta - \kappa^2 = \alpha
\end{aligned}$$

(iv) $\mathbf{i < k \wedge j \neq l}$: In that case it holds that the indices are pairwise unequal. Therefore,

$$\begin{aligned}
&Y_{ij} \cdot Y_{kl} = 1 \Leftrightarrow Y_{ij} = Y_{kl} = 1 \Leftrightarrow C_i = C_j \wedge C_k = C_l \\
\Rightarrow &E(Y_{ij} \cdot Y_{kl}) = 1 \cdot P(Y_{ij} \cdot Y_{kl} = 1) = P(Y_{ij} = 1) \cdot P(Y_{kl} = 1) = \kappa^2 \\
\Rightarrow &\mathrm{Cov}(Y_{ij}, Y_{kl}) = E(Y_{ij} \cdot Y_{kl}) - \kappa^2 = \kappa^2 - \kappa^2 = 0
\end{aligned}$$

It Rightarrow:

$$\begin{aligned}
&2 \sum_{i<j} \sum_{k<l,(i,j)<(k,l)} \mathrm{Cov}(Y_{ij}, Y_{kl}) \\
&= 2 \sum_{i<j} \left(\sum_{l=j+1}^{n} \mathrm{Cov}(Y_{ij}, Y_{il}) + \sum_{l=j+1}^{n} \mathrm{Cov}(Y_{ij}, Y_{jl}) \right) \\
&+ \sum_{k=i+1}^{j-1} \mathrm{Cov}(Y_{ij}, Y_{kj}) + \sum_{k=i+1}^{n} \sum_{l=k+1,k,l\neq j}^{n} \mathrm{Cov}(Y_{ij}, Y_{kl}) \Bigg) \\
&= 2 \sum_{i<j} \left(\sum_{l=j+1}^{n} \alpha + \sum_{k=i+1}^{n} \left(\alpha + \sum_{l=k+1,l\neq j}^{n} 0 \right) \right) \\
&\overset{(4)(5)}{=} 2\alpha \left(\frac{1}{6} n(n-1)(n-2) + \frac{1}{6} n(n-1)(n-2) + \frac{1}{6} n(n-1)(n-2) \right) \\
&= (\beta - \kappa^2) n(n-1)(n-2)
\end{aligned}$$

(4) It holds

$$\begin{aligned}
&\sum_{i<j} \sum_{l=j+1}^{n} \alpha = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{l=j+1}^{n} \alpha = \alpha \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (n-j) = \alpha \sum_{i=1}^{n-1} \sum_{j=1}^{n-i-1} j \\
&= \frac{\alpha}{2} \sum_{i=1}^{n-1} (n-i)(n-i-1) = \frac{\alpha}{2} \sum_{i=1}^{n-2} (n^2 - n\,i - n - n\,i + i^2 + i) \\
&= \frac{\alpha}{2} \left[(n-2)(n^2-n) + (1-2n) \sum_{i=1}^{n-2} i + \sum_{i=1}^{n-2} i^2 \right] \\
&= \frac{\alpha}{2} \left[(n-2)(n-1)n + (1-2n) \frac{1}{2}(n-1)(n-2) + \frac{1}{3}(n-1)^3 - \frac{1}{2}(n-1)^2 + \frac{1}{6}(n-1) \right] \\
&= \frac{\alpha}{12} (n-1) \left[6n^2 - 12n - 6n^2 + 15n - 6 + 2n^2 - 4n + 2 - 3n + 3 + 1 \right] \\
&= \frac{\alpha}{12} (n-1) \left[2n^2 - 4n \right] = \frac{\alpha}{6} n(n-1)(n-2)
\end{aligned}$$

(5) It holds

$$\begin{aligned}
\sum_{i<j} \sum_{k=i+1}^{j-1} \alpha &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=i+1}^{j-1} \alpha = \alpha \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (j-1-i) \\
&= \alpha \sum_{i=1}^{n} \sum_{j=1}^{n-i-1} j \overset{(4)}{=} \frac{\alpha}{6} n(n-1)(n-2)
\end{aligned}$$