Tony Majestro

Problem Set 2

1. The following was encrypted by the Vigenere method. Decrypt it. Carefully document your work – in general, more details are better than fewer details.

XKJUROWMLLPXWZNPIMBVBQJCNOWXPCCHHVVFVSLLFVXHAZITYXOHULX

QOJAXELXZXMYJAQFSTSRULHHUCDSKBXKNJQIDALLPQSLLUHIAQFPBPC

IDSVCIHWHWEWTHBTXRLJNRSNCIHUVFFUXVOUKJLJSWMAQFVJWJSDYLJ

OGJXDBOXAJULTUCPZMPLIWMLUBZXVOODYBAFDSKXGQFADSHXNXEHSAR

UOJAQFPFKNDHSAAFVULLUWTAQFRUPWJRSZXGPFUTJQIYNRXNYNTWMHC

To decrypt the cipher text, we must begin by finding the length of the key used to encrypt the plaintext. I will use the Kasiski test. First I must find the indexes of repeating trigraphs in the cipher text. The following code calculates us the trigraphs that repeat more than once in the cipher text:



This gives us the following trigraphs:

'HSA', 'WML', 'LLP', 'LLU', 'OJA', 'DSK', 'AQF', 'QFP', 'CIH', 'JQI', 'SLL', 'XVO', 'JAQ'

We can now get the distances between each trigraph and try to find the greatest common divisor between them. The following code gives the indexes of a trigraph in the cipher text:



The trigraph ‘HSA' appears at 216 and 231, which gives a distance of 15. The trigraph ‘JAQ’ appears at 67 and 222, which gives a distance of 155. The trigraph ‘XVO’ appears at 142 and 192, which gives a distance of 50. The greatest common divisor of all of these numbers is 5, so this would lead us to believe that 5 is a multiple of the key length. Since 5 is prime, the key length is probably 5.

Now that we know the key length, we can use it to find the key used to encrypt the plaintext. We will begin by getting the columns obtained by stacking the cipher text into rows of length 5. Since each column was encrypted using the same letter, we can say that it was encrypted using a monoalphabetic shift cipher. To find the shift for each column, we’ll perform frequency analysis on each column. Each column should have a letter frequency distribution similar to English, which we can then exploit to find the shift amount.



This gives us the following 5 column strings and their letter frequencies:

'XOPPBOCFFZOQEMFUCXIPUFIIETNIFUSFSOBUPIBOFGSEUFDFUFJGJRT', 'KWXIQWHVVIHOLYSLDKDQHPDHWXRHUKWVDGOLZWZDDQHHOPHVWRRPQXW', 'JMWMJXHSXTUJXJTHSNASIBSWTRSUXJMJYJXTMMXYSFXSJFSUTUSFINM', 'ULZBCPVLHYLAZASHKJLLAPVHHLNVVLAWLXAUPLVBKANAAKALAPZUYYH', 'RLNVNCVLAXXXXQRUBQLLQCCWBJCFOJQJJDJCLUOAXDXRQNALQWXTNNC'

Counter({'F': 10, 'U': 6, 'I': 5, 'O': 5, 'P': 4, 'B': 3, 'E': 3, 'S': 3, 'C': 2, 'G': 2, 'J': 2, 'T': 2, 'X': 2, 'D': 1, 'M': 1, 'N': 1, 'Q': 1, 'R': 1, 'Z': 1})

Counter({'H': 8, 'W': 7, 'D': 6, 'Q': 4, 'V': 4, 'K': 3, 'L': 3, 'O': 3, 'P': 3, 'R': 3, 'X': 3, 'I': 2, 'Z': 2, 'G': 1, 'S': 1, 'U': 1, 'Y': 1})

Counter({'S': 9, 'J': 8, 'X': 7, 'M': 6, 'T': 5, 'U': 4, 'F': 3, 'I': 2, 'H': 2, 'N': 2, 'W': 2, 'Y': 2, 'A': 1, 'B': 1, 'R': 1})

Counter({'A': 10, 'L': 10, 'H': 5, 'V': 5, 'P': 4, 'K': 3, 'U': 3, 'Y': 3, 'Z': 3, 'B': 2, 'N': 2, 'C': 1, 'J': 1, 'S': 1, 'W': 1, 'X': 1})

Counter({'X': 7, 'C': 6, 'L': 6, 'Q': 6, 'J': 5, 'N': 5, 'A': 3, 'R': 3, 'B': 2, 'D': 2, 'O': 2, 'U': 2, 'W': 2, 'V': 2, 'F': 1, 'T': 1})

To find the shift amounts, we’ll try to find which letter decrypts to the letter E. We can do this by looking at the most frequently occurring letters in each column. For the first column, the letter F appears 10 times. Since the next most frequent letter appears only 6 times, we can assume that E encrypts to F. This corresponds to a shift amount of 1.

For the second column, H is the most frequently occurring letter. This means E could encrypt to H, which gives a shift of 3. The second most frequently occurring letter in the column is W. Since the second most frequently occurring English letter is T, we can guess that T encrypts to W. This also gives a shift of 3, so we can safely guess that the second column is encrypted with a shift of 3.

For the third column, S appears 9 times and J appears 8 times. This probably means that E encrypts either to S or J. This gives shifts of either 14 or 5.

For the fourth column, both A and L appear 10 times. If E encrypts to one of these letters, we could have shifts of either 14 or 7. If we guess that T encrypts to A, and that E encrypts to L, we get a shift of 7. So 7 is probably the shift for the fourth column.

For the fifth column, X appears 7 times. We can guess that E encrypts to X, which gives a shift of 19.

We’ll begin by guessing that the shift amount for the third column is 14 instead of 5. Using the shift amounts, we can decrypt each column and read the plaintext.



Using the following shifts, this decryption gives us the following plaintext:

WHVNYNTYESOUISUOFYUCANVVUNTJIJBETOCESEESESJAHYFFRENEGEEPLVTEDIJSELVVTXEPFLYTITABBAEDIWHZCXHAMESONEESTEUTXEMNIJHAEOJHEIADDTFAISUDEQMOEGJHEGOMERJOVTHVEQRTYTXESVPQRAKEQNDVQKALJTQTIFNJOWYISHTYEBAWJOVNAKUHEAEDEFNRTKREJGEDEETYTLVTXEMRDUCEETHESGESTTFTXEOGIDIOESEFMRNAINURUQUZRUSTYAJ

This is not really readable, so we’ll try out other option of shift equal to 5 for the third column. This gives the following plaintext:

WHENYNTHESOURSUOFHUCANEVUNTSIJBECOCESNESESSAHYFORENEPEEPLETEDISSELVETXEPOLYTICABBANDIWHICXHAVESONNESTEDTXEMWIJHANOJHERADDTOAISUMEQMONGJHEPOMERSOVTHEEQRTHTXESEPQRATEQNDEQKALSTQTIONJOWHISHTHEBAWSOVNATUHEANDEFNATKRESGEDENTYTLETXEMADUCENTHESPESTTOTXEOPIDIONSEFMANAINDRUQUIRUSTHAJ

This is close to English, but one of the columns still has the wrong shift amount. In the plaintext, we can spot the word ‘becoces’, which probably is supposed to be ‘becomes’. We can also spot the word ‘peeple’, which should probably be people. The incorrect letters C and E occur at positions 35 and 55, which means that the fifth column has the wrong shift. The 35th and 55th letters in the cipher text are V and X. We want M to encrypt to V and O to encrypt to X. This gives a shift of 9. If we decrypt again using 9 as the shift for the fifth column, we get the following plaintext:

WHENINTHECOURSEOFHUMANEVENTSITBECOMESNECESSARYFORONEPEOPLETODISSOLVETHEPOLITICALBANDSWHICHHAVECONNECTEDTHEMWITHANOTHERANDTOASSUMEAMONGTHEPOWERSOFTHEEARTHTHESEPARATEANDEQUALSTATIONTOWHICHTHELAWSOFNATUREANDOFNATURESGODENTITLETHEMADECENTRESPECTTOTHEOPINIONSOFMANKINDREQUIRESTHAT

This plaintext is part of the introduction to the Declaration of Independence.

2. Suppose you have a language with only the three letters a, b, c, and they occur with frequencies 0.7, 0.2 and 0.1, respectively. The following ciphertext was encrypted by the Vigenere method (shifts are mod 3 instead of mod 26): ABCBABBBAC

You are told that the key length is 1, 2, or 3. Show that the key length is probably 2, and determine the most probable key. Again, document your work.

To determine the key length, we can make guesses at the key length in order to find the columns of cipher text that give Indexes of Coincidence closest to that of the language. If we say A = 0, B = 1, and C = 2, then the IC for the language is:

Now we need to find the key length. If we assume that the key length is 1, then the only column is the cipher text. A occurs 3 times, B occurs 5 times, and C occurs 2 times. The IC for this column is

This is not very close to the IC of the language. If we guess the key length to be 2, we get the columns ‘acaba’ and ‘bbbbc’. For the first column, A occurs 3 times, B occurs 1 time, and C occurs 1 time. Using the above formula, this gives an IC of 0.3. For the second column, A occurs 0 times, B occurs 4 times, and C occurs 1 time. This gives an IC of 0.6. This is closer to the IC of the language.

If we guess the key length to be 3, then we get the columns ‘abbc’, ‘bab’, and ‘cba’. The IC of ‘abbc’ is 0.167. The IC of ‘bab’ is 0.33. The IC of ‘cba’ is 0. This is not very close to the IC of the language.

The guess for the key length that gave us the closest Index of Coincidence for the language was 2, so the key length is probably 2.

Since we know the key length is 2, then we can figure out the key by analyzing the frequency distribution of each of the 2 columns: ‘acaba’ and ‘bbbbc’. Since the most frequently occurring letter in the first column is A and the most frequently occurring letter in the language is A, then we can guess that A encrypts to A for the first column. This gives a shift of 0.

For the second column, the most frequently occurring letter is B, which means that A encrypts to B for the second column. This means the shift is 1.

This gives us the key AB, since the A = 0 and B = 1.

3. The ciphertext GEXZDS was encrypted by a Hill cipher with a 2x2 encryption matrix M. The plaintext is solved. Find the encryption matrix M.

We can begin by converting the ciphertext and plaintext to numbers:

* The ciphertext GEXZDS is 6, 24, 23, 25, 3, 18.
* The plaintext SOLVED is 18, 14, 11, 21, 4, 3.

This gives us the following equations:

|  |  |  |
| --- | --- | --- |
|  |  |  |

We can combine pairs of the following equations to solve for M. We can eliminate the first equation because both of the numbers are even, which would lead to an even determinate. This would mean the combined matrix is not invertible, since no even numbers are coprime with 26.

If we combine the last two equations, we can get the congruence:

|  |
| --- |
|  |
|  |

To solve for M, we must first prove that the matrix is invertible. We can do this by calculating its determinate mod 26. If the determinate of the matrix is coprime with 26, then it is invertible.

Since the determinate of the matrix is 1, which is coprime with 26, then the matrix is invertible. We can now calculate :

Now to calculate the encryption matrix M:

|  |
| --- |
|  |
|  |
|  |
|  |

Therefore, the encryption matrix is .