Monoalphabetic Substitution

Assume that we have intercepted N letters of a ciphertext message that was encoded using a Monoalphabetic substitution and that the entropy of english is 2 bits.

Length of text	5	10	15	20	30	40	50
# of distinct letters	4	8	11	12	14	16	18

For instance, a typical english sample of 30 letters contains about 14 di erent letters. Thus the key for a Monoalphabetic substitution only permutes 14 letters. Therefore the number of keys is

$$26 \times 25 \times \cdots \times 13$$

and not 26!. if n=22 $26 \times 25 \times \cdots \times 14$ OR 15

Assuming that each key is equally likely, we have

$$H(K) = \log_2(26 \times 25 \times \cdots \times 13) \approx 59.54$$

Assuming that each of the 26^N ciphertexts is equally likely, we have

$$H(C) = \log_2 26^N = N \log_2 26 \approx 4.7N$$

Therefore,

F=2 $\quad 59.54 = 4.7N - 2N \Rightarrow N \approx 22.05$ ✓

F=1.2 $\quad 59.54 = 4.7N - 1.2N \Rightarrow \approx 17.01$

KEY 1 KEY KEY

OBSCENE OBSCENE OBSCENE ....

HOUSEHOUSEHOUSEHOUSE ....

VP

APCTDFOFPCTDFO ....

GNTRDGNTRDGNT ...

VP

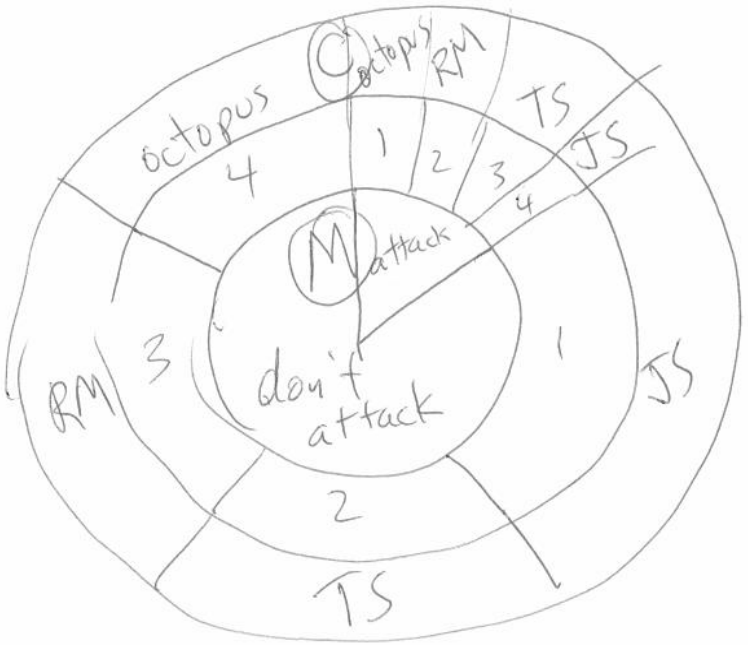
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
P Q R S T U V W X Y Z A B C D E F G H I J K L M N O
G H I J K L M N O P Q R S T U V W X Y Z A B C D E F

# of keys for Vigenere $p=7$ 26^7

# of keys for Vigenere $q=5$ 26^5

# of keys for Vernam $p=7$ & $q=5$ $\frac{26^7 \cdot 26^5}{26} = 26^{11}$

M & C are independent.



$H($  $)$ = information that I learn if I don't know the ciphertext

$H($  $)$ = information given that ciphertext = Roger Moore

4. You cast a pair of dice: die #1 shows X and #2 shows Y but you only record $X+Y$.
   a) How much information have you lost in doing this?
   b) What is the expected number of bits you need to store 300 samples of $X+Y$?

6. Verbally explain why your intuition tells you that we must have H(Y|X)<=H(X) and when should we have equality.

(2) Random words are created by choosing the first letter using the following table of single letter statistics then each subsequent letter is chosen using the table of biletter statistics shown to the right.

A	4
B	1
C	2
D	3

	A	B	C	D
A	4	0	1	5
B	3	1	0	6
C	7	3	0	0
D	1	0	9	0

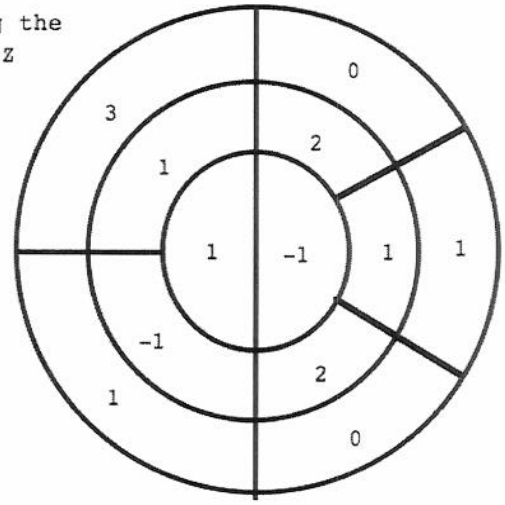
(a) How much information do you learn on average when you are told a two letter word given that you know that the word begins with the letter A?
(b) How much information do you learn on average when you are told a two letter word given that you know that the word ends with the letter D?
(c) How much information do you learn on average when you are told a two letter word given that you know at least one of the letters is a D?

3. Let X,Y,Z be produced by spinning the attached fortune wheel, with X,Y,Z given by the inner, middle and outer circles respectively. Calculate:

   a) H(Y|X,Z)

   b) H(X x Z)

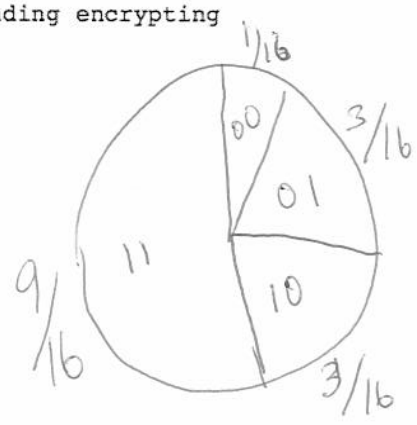
   c) The expected amount of information about Z you get when you are told the values of X and Y.



4. Suppose a random cryptosystem is obtained as follows: The message consists of two independent random variables X1 , X2 both generated by spinning a wheel with arcs labelled 0,1 of lengths 1/4 and 3/4 respectively. Let there be four equiprobable keys with the following corresponding encrypting transformations:

   k1: (X1,X2) ---> (X1,X2)
   k2: (X1,X2) ---> (X2,X1)
   k3: (X1,X2) ---> (X1, (X1+X2) mod 2)
   k4: (X1,X2) ---> ((X1+X2) mod 2 , X1)

   Calculate the entropy of the cyphertext.



Not perfectly secret