

1 Theory

Indices of letters:

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Measure of Roughness (**MR**) is a measure how much a distribution differs from a uniform distribution.

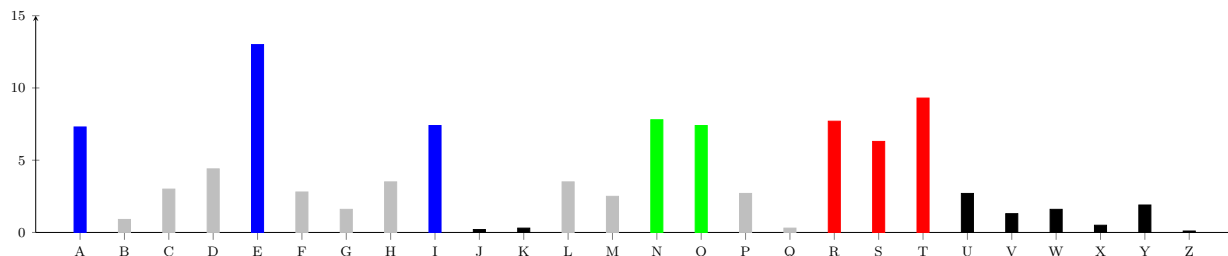
$$\mathbf{MR} = \sum_i \left(p_i - \frac{1}{26} \right)^2 = \sum_i p_i^2 - 2 \underbrace{\frac{1}{26} \sum_i p_i}_{=1} + \underbrace{\sum_i \left(\frac{1}{26} \right)^2}_{=26 \cdot \frac{1}{26^2}} = \sum_i p_i^2 - \frac{1}{26} \approx \sum_i p_i^2 - 0.038 .$$

Index of coincidence **IC** is an approximation to $\sum_i p_i^2$. For a ciphertexts $Y = y_1 y_2 \dots y_N$ and $Y' = y'_1 y'_2 \dots y'_M$ with characteristic frequency of letter i denoted as f_i , the index of coincidence (**IC**) is

$$\mathbf{IC}(Y) = \frac{\sum_i f_i(f_i - 1)}{N(N - 1)} , \quad \mathbf{IC}(Y, Y') = \sum_i \frac{f_i f'_i}{NM} .$$

I.C. approximates the probability that any two letters randomly sampled from a distribution (or from two different distributions) will be the same. Since IC approximates $\sum_i p_i^2$, it has the same range of variation 0.038 to 0.066. The lower bound corresponds to a uniform distribution, and the upper bound corresponds to monoalphabeticity. The number 0.066 is obtained by summing up the squared characteristic frequencies of English letters. On average, in a 1000 letter long sample of English text, the letters are distributed as follows:

A	73	B	9	C	30	D	44	E	130	F	28
G	16	H	35	I	74	J	2	K	3	L	35
M	25	N	78	O	74	P	27	Q	3	R	77
S	63	T	93	U	27	V	13	W	16	X	5
Y	19	Z	1								



The same picture would result from the examination of any reasonably long plain language text. Relative frequencies may vary slightly, but the basic facts remain the same:

- Evenly spaced vowels A E I with high frequency are evenly spaced 4 letters apart.

- Letter E is the most frequent of all the letters
- Consecutive part N,O have high frequency
- Consecutive triplet R,S,T has high frequency
- The pair J,K has low frequency
- The string U,V,W,X,Y,Z has low frequency.

2 Tasks

1. An additive cipher maps plaintext G to ciphertext X . What is the encryption key? Which decryption key will allow to reconstruct the plaintext?
2. We know that a ciphertext was produced by a shift cipher, and that the encryption key was 17. What is the decryption key?
3. We know that the plaintext word THE is encrypted by an affine cipher into trigam NHM. What is the encryption key? What is the decryption key?
4. A ciphertext obtained by an affine cipher with key $(3, 17)$. Which key will you use to decrypt it?
5. What is the I.C. of the ciphertext EPYEPOPDSZUFPO?
6. Encrypt the word MORNING using a shift cipher with key 11.
7. Encrypt the word SYMBOL using an affine cipher with key $(3, 2)$.
8. Encrypt the word PARADOX using a Vigenère cipher with key YESTERDAY.
9. Given two ciphertexts $Y : M \text{ BQDEAZ ITA ZQHGD YMPQ M YUEFMWQ ZQHGD FDUQP MZKFTUZZS ZQI}$ and $Y' : GZVMI AMJH TZNOZMYVT GDQZ AJM OJYVT CJKZ AJM OJHJMMJR$, which are two different messages encrypted by shift cipher using different keys z and z' . Find $z - z' \pmod{26}$.
10. Decipher the following messages:

ESPCP TD L ETOP TY ESP LQQLTCD ZQ XPY HSTNS ELVPY LE ESP QWZZO WPLOD ZY EZ QZCEFYP

SV SQ VNY OWMWR KWRTYL WD EHH MYR NWK EMWRI QW MERU MSHHSWRQ WD DEOYQ VNYLY QNWAHT

BDSTGC WXHIDGN AXZT P STPU BPC PCHLTGH FJTHIXDCH CD DCT PHZTS

JDI HVANGNKFKJS JDGJ EI MGS PGKF KT G EAVJDS UGQCI KC TAJ CQPPKUKITJ RQCJKPKUGJKAT PAV AQV VIPQCKTW JA CQHHAJ KJ