The Cipher Challenge

Cryptanalysis for classical cryptography

Izaak van Dongen

October 12, 2018

Hills Road SFC

Outline

Statistics

Strategy

We take the English letter distribution [1] to be roughly invariant:

L	E	T	A	O	I	N	S	H	R	D	...
$P(L)/\%$	13	9.1	8.2	7.5	7	6.7	6.3	6.1	6	4.3	...

[1] http://en.algoritmy.net/article/40379/Letter-frequency-English

Index Of Coincidence

This implies that the IOC will also be invariant. IOC can be calculated as:

$$\text{IOC} = \sum_{r=0}^{25} \frac{n_r(n_r-1)}{N(N-1)}$$

```
1  def ioc(count: collections.Counter):
2      total = sum(count.values())
3      return (sum(freq ** 2 - freq for freq in count.values())
4              / (total ** 2 - total))
```

In English, this is expected to be around 0.06654.

Fitness metrics

We can measure the closeness of some text to English (its "fitness") in several ways.

One is to take its letter distribution as a vector in 26 dimensions. It is normalised such that the sum of its components is equal to 1. We can then take the total deviance of this distribution from the expected English distribution as a measure of fitness.

Another approach is by quadgrams. We take every four adjacent letters and look up the expected frequency of this combination in a large table of every 26^4 quadgrams. We use this to judge how similar we think the text is to English.

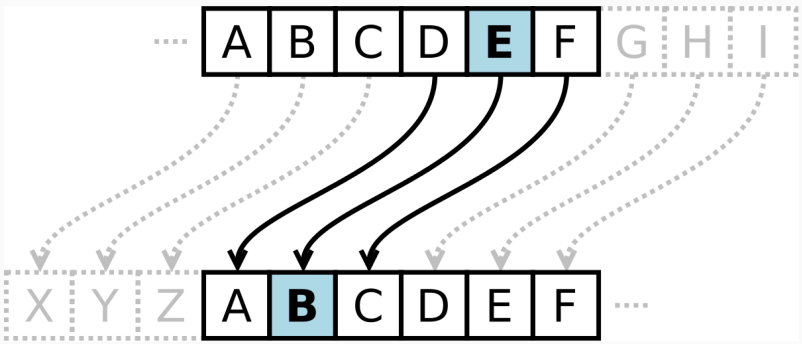
Monoalphabetic substitution ciphers

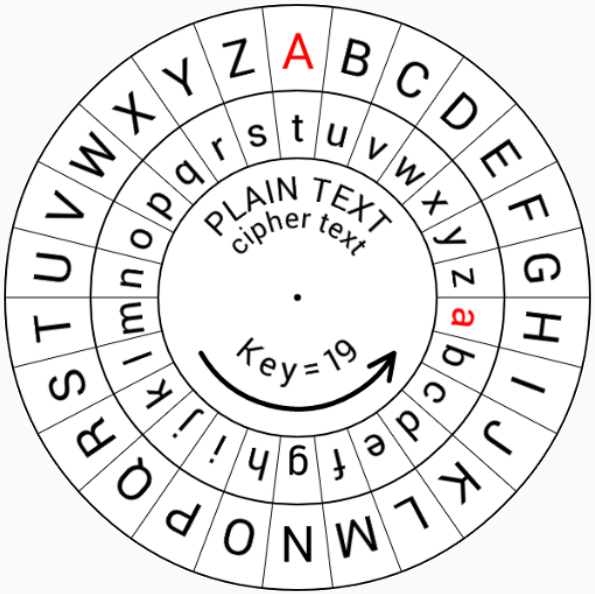
Monoalphabetic substitution ciphers

The Caesar cipher

The Caesar cipher

The Caesar cipher is normally thought of as a simple shift, or rotation, of each letter.





Mathematical notation

First, we consider each character as an integer: $A..Z$ maps to $0..25$.

For now, just consider the plaintext as having all uppercase characters.

We denote the encryption function of some Caesar cipher of key (shift) s as: $E_s(x) = x + s \pmod{26}$

Here $a \pmod{b}$ denotes the remainder when a is divided by b. This makes it "wrap around" the way we want.

Decryption

If $E_s(x) = x + s \pmod{26}$, the asssociated decryption function is $D_s(x) = x - s \pmod{26}$.

The recipient of the ciphertext already knows s, so they simply plug it in to find out what the plaintext was.

Monoalphabetic substitution ciphers

The affine shift cipher

Affine shift

We now let the encryption function $E_{(a,b)}(x) = ax + b \pmod{26}$. Therefore,

$$\begin{aligned} D_{(a,b)}(x) &= a^{-1}(x - b) \pmod{26} \\ &= E_{(a^{-1}, -a^{-1}b)}(x) \end{aligned}$$

Note that an affine shift is invertible iff a is coprime to 26 ($gcd(a, 26) = 1$) (eg consider that $E_{(2,b)}(0) = E_{(2,b)}(13) = b$).

Polygraph substitution ciphers

Polygraph substitution

Instead of a monograph substitution, where we have $E(x)$, we can have for example a digraph substitition.

This means that the cipher splits the text into pairs of letters (digraphs), and we have a function on these: $E\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)$

Polygraph substitution ciphers

Vigenère cipher

Vigenère cipher

These have a key which is a sequence of letters. Each letter encodes a Caesar shift for its corresponding letter, in essence.

An n-length Vigenère cipher with key k_i is:

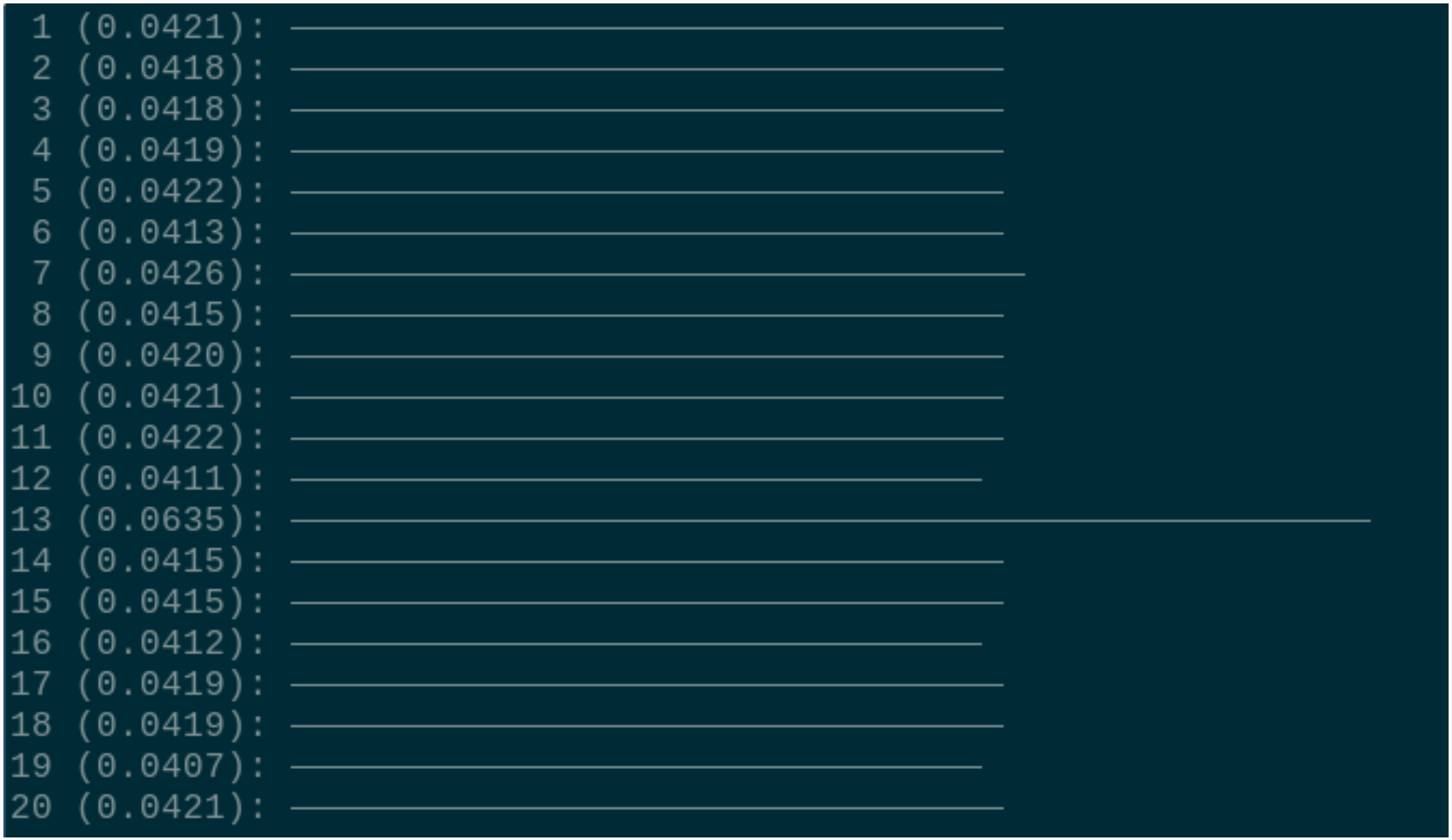
$$E_{\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix}} \left(\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \right) = \begin{bmatrix} c_1 + k_1 \\ c_2 + k_2 \\ \vdots \\ c_n + k_n \end{bmatrix} \pmod{26}$$

Cryptanalysis

Note that this basically leaves the shape of the distribution of every nth letter, every $n + 1$th letter, and so on untouched.

If we hypothesise some key length n, we can therefore find the average IOC of each of these intervals. We expect to see a reasonable spike if this is correct (this spike will also happen for multiples of the correct n).

Analysis of 2017-5b



Polygraph substitution ciphers

Other polygraph ciphers

Other variants

There are a lot of these. Really, you'll need to do some research on how to spot each.

Some examples are:

- Playfair
- Hill
- Bifid

Stateful ciphers

Ciphers with state

This is most modern ciphers. These can be quite tricky. Some examples are:

- Autokey - this can be attacked with the techniques in this presentation.
- Enigma (simplified). This is really mean so don't come to me for help.