DTTF/NB479: Dszquphsbqiz Day 5

- Announcements:
  - Please pass in Assignment 1 now.
  - Assignment 2 posted (when due?)
- Questions?
- Roll Call
- Today: Vigenere ciphers

Shift, Affine, and Substitution ciphers are related

1. How many possibilities to brute force?
2. What idea is new?

Shift

Affine

Substitution

Vigenere ciphers

- Invented in 1553 by Bellaso
- A different type of complexity

Vigenere Ciphers

- Idea: the key is a *vector* of shifts
  - The key and its length are unknown to Eve
- Encryption:
  - Repeat the vector as many times as needed to get the same length as the plaintext
  - Add this repeated vector to the plaintext.
- Example:

Key = *hidden (7 8 3 3 4 13)*.

```
The recent development of various methods
19 7 4  17 4  2 4 13 19  3…
```

Key

```
7 8 3  3 4 13 7 8 3  3 4 13 7 8 3 3 4 13 7 8  3 3  4 13 7 8 3 3 4  13 7 8 3 3 4 13
0 15 7  20 8 15 1121 22  6 8 811191718161720 1  17 8  25132416172322  2511 11017 7 5
aph uiplvw giiltrsqrub ri znyqrxw zlbkrhf
```

Security

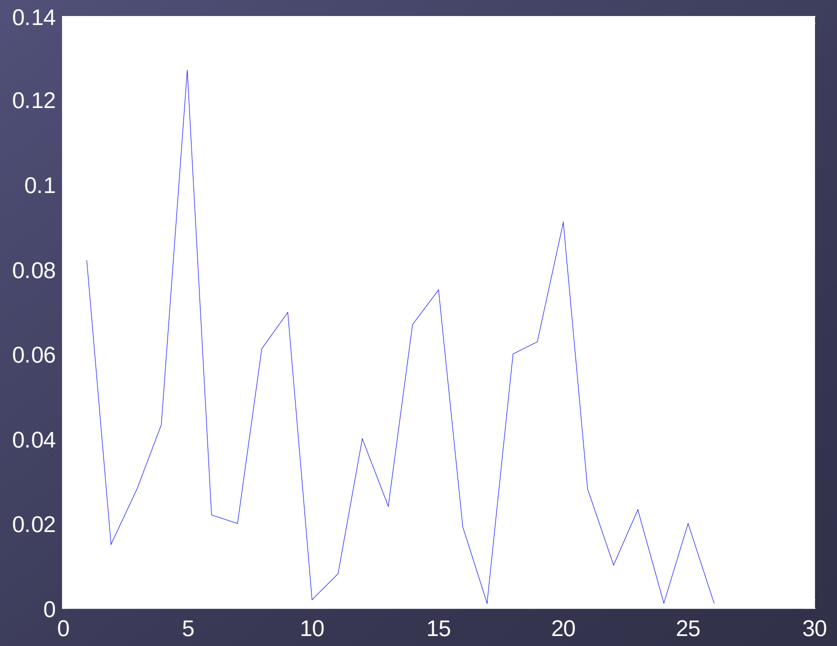
- The shift vector isn't known (of course)
  1. With shift ciphers, the most frequent cipher letter is probably e.

     But here, e maps to H, I, L, … (**spread out!**)
  2. The vector's length isn't even known!
- Consider 4 attacks:
  - Known plaintext?
  - Chosen plaintext?
  - Chosen ciphertext?
  - Ciphertext only? (most interesting)

English letter frequencies

A 0.082	H 0.061	O 0.075	U 0.028
B 0.015	I 0.070	P 0.019	V 0.010
C 0.028	J 0.002	Q 0.001	W 0.023
D 0.043	K 0.008	R 0.060	X 0.001
E 0.127	L 0.040	S 0.063	Y 0.020
F 0.022	M 0.024	T 0.091	Z 0.001
G 0.020	N 0.067		

Graph:



Ciphertext-only attack

- Assume you know the key length, L.
- Make any other assumptions you need.
- Take 5 min with a partner and devise a method to break Vigenere.

Perhaps yours looks something like this?

- Assume we know the key length, L, …
  - We'll see how to find it shortly
- Method 1:
  - Parse out the characters at positions p = *i* (mod L)
    - These have all been shifted the same amount
    - Do a frequency analysis to find shift
      - The most frequent letter should be e, given enough text. Can verify to see how shift affects other letters.
  - This gives the first letter of the key
  - Repeat for positions p = 1, p = 2, … p = L-1
  - Problem: involves some trial and error.
  - For brute force to work, would need to brute force all letters of key simultaneously: _____ possibilities

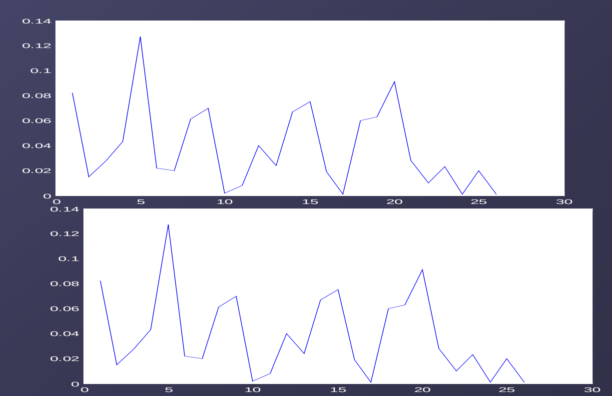
Using the whole frequency distribution is more robust than using a single letter

- Do this via dot products of frequency vectors.

Dot products

$$A \cdot B = A.*B = \sum_i A_i B_i$$

- Consider A = (0.082 0.015 0.028 0.043 0.127 0.022 0.020 0.061 0.070 0.002 0.008 0.040 0.024 0.067 0.075 0.019 0.001 0.060 0.063 0.091 . 0.028 0.010 0.023 0.001 0.020 0.001);
- $A_i = A$ displaced i positions to the right
- A_0 = (0.082 0.015 0.028 … 0.001 0.020 0.001)
- A_1 = (0.001 0.082 0.015 0.028 … 0.023 0.001 0.020)
- A_2 = (0.020 0.001 0.082 0.015 0.028 … 0.023 0.001)
- A_0 .* A_1 = 0.039
- A_0 .* A_0 = 0.066
- A_i .* A_j depends on _____ only.
- Max occurs when _____.
  Why?



Towards another method

- Method 1
  - Parse out the characters at positions p = 0 (mod L)
    - These have all been shifted the same amount
    - Do a frequency analysis to find shift
      - The most frequent letter should be e, given enough text. Can verify to see how shift affects other letters.
  - This gives the first letter of the key
  - Repeat for positions p = 1, p = 2, … p = L-1

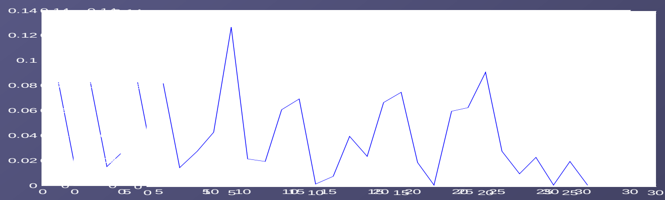
Another method

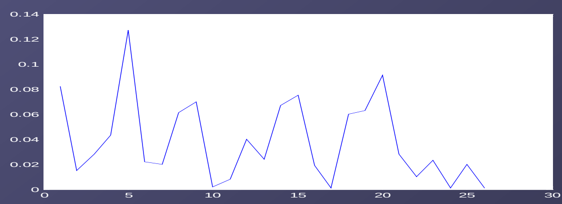
- Method 2
  - Parse out the characters at positions p = 0 (mod L)
    - These have all been shifted the same amount
    - Get the whole freq. distribution **W = (0.05, 0.002, …)**
      - W approximates A. Calculate $W \cdot A_i \;\; for \; 0 \le i \le 25$
      - Max occurs when we got the shift correct.
  - This gives the first letter of the key
  - Repeat for positions p = 1, p = 2, … p = L-1
  - Demo

Method 2 is more robust since it uses the whole letter distribution

- Find dot product of A_i:

  and W:





More robust than just using 1 letter (‘e’)…

…but harder to compute by hand.

Finding the key length also uses dot products

- Just displace the ciphertext by various amounts and look for the maximum dot product

Finding the key length

- What if the frequency of letters in the plaintext approximates A?
- Then for each k, the frequency of each group of letters in position p = k (mod L) in the ciphertext approximates A.
- Then loop, displacing the ciphertext by i, and counting the number of matches.
  - Get max when displace by correct key length
  - So just look for the max number of matches!

```
                                       displacement
APHUIPLVWGIILTRSQRUBRIZNYQRXWZLBKRHFVN (0)
NAPHUIPLVWGIILTRSQRUBRIZNYQRXWZLBKRHFV (1)  1 match
VNAPHUIPLVWGIILTRSQRUBRIZNYQRXWZLBKRHF (2)  0 matches
…
KRHFVNAPHUIPLVWGIILTRSQRUBRIZNYQRXWZLB (6)  5 matches
…
```

Key length: an example

Take any random pair in the ciphertext:

- The letter in the top row is shifted by i (say 0)
- The letter in the bottom row is shifted by j (say 2)

  - Prob(both ‘A’) = P(‘a’)*P(‘y’) = 0.082 * 0.020
  - Prob(both ‘B’) = P(‘b’)*P(‘z’) = 0.015 * 0.001
  - Prob (both same (any letter)) is ___ or generally ___
  - Recall, this is maximum when ______
  - When are each letter in the top and bottom rows shifted by same amount?

A_0 = (0.082 0.015 0.028 … 0.001 0.020 0.001)

A_2 = (0.020 0.001 0.082 0.015 0.028 … 0.023 0.001)

The text helps with implementation

- Read it. Implement it. You’ll own it.
  - You’ll do this on Homework 2:
  - Week 3 programming test: use your program to decrypt a vigenere-encrypted message

Exceptions

- Consider *Gadsby* by Ernest Vincent Wright, February 1939:
  - http://www.spinelessbooks.com/gadsby/01.html
- What do you notice about it?