Vigenere Cipher

- Key is of the form $K = (k_0,k_1,\ldots,k_{n-1})$
  - Where each $k_i \in \{0,1,2,\ldots,25\}$
- Encryption

  $$c_i = p_i + k_{i \pmod n} \pmod{26}$$
- Decryption

  $$p_i = c_i - k_{i \pmod n} \pmod{26}$$
- Nothing tricky here!
- Just a repeating sequence of (shift by n) simple substitutions

Vigenere Cipher

- For example, suppose key is MATH
  - That is, $K = (12,0,19,7)$, since M is letter 12, and so on
- Plaintext: SECRETMESSAGE
- Ciphertext: EEVYQTFLESTNQ
- Encrypt:

```
  S   E   C   R   E   T   M   E   S   S   A   G   E
 18   4   2  17   4  19  12   4  18  18   0   6   4
+12   0  19   7  12   0  19   7  12   0  19   7  12
---------------------------------------------------
  4   4  21  24  16  19   5  11   4  18  19  13  16  (mod 26)
  E   E   V   Y   Q   T   F   L   E   S   T   N   Q
```

Vigenere Cipher

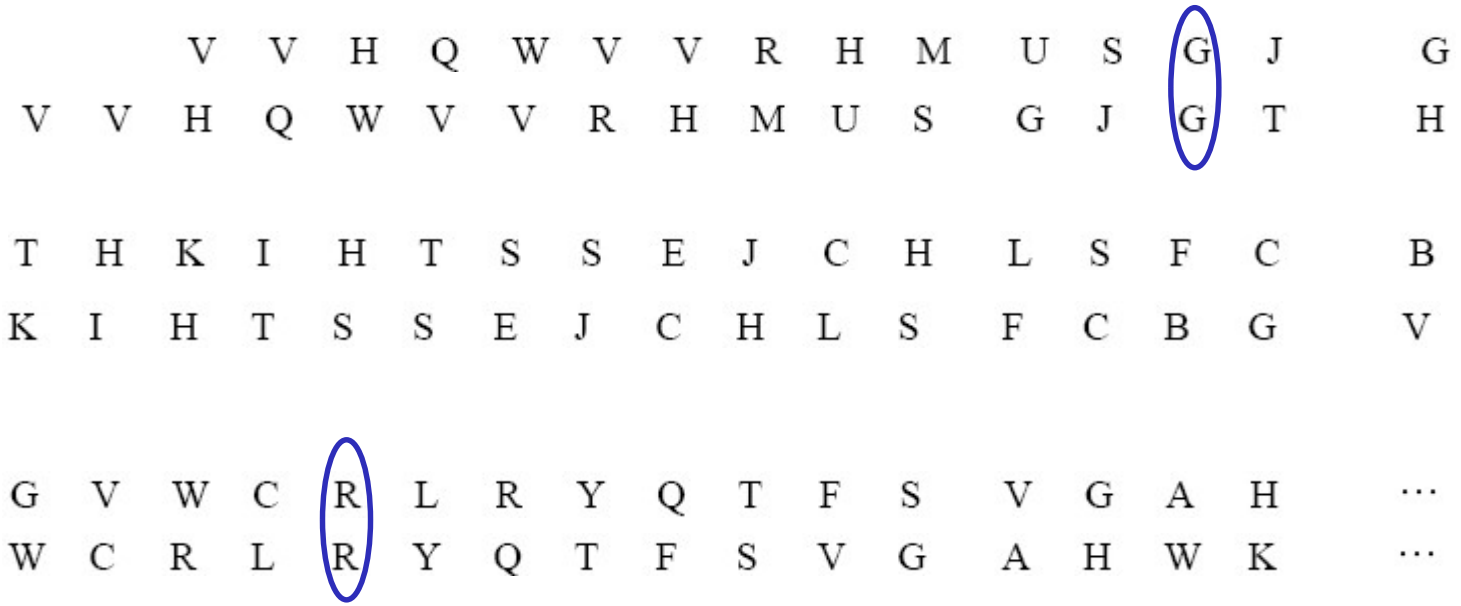
- Vigenere is just a series of k simple substitution ciphers
- Should be able to do k simple substitution attacks
  - Provided enough ciphertext
- But how to determine k (key length)?
- Index of coincidence…

Determining the key length(1)

- Ciphertext:

VVHQWVVRHMUSGJGTHKIHTSSEJCHLSFCBGVWCRLRYQTFSVGAHW
KCUHWAUGLQHNSLRLJSHBLTSPISPRDXLJSVEEGHLQWKASSKUWE
PWQTWVSPGOELKCQYFNSVWLJSNIQKGNRGYBWLWGOVIOKHKAZKQ
KXZGYHCECMEIUJOQKWFWVEFQHKIJRCLRLKBIENQFRJLJSDHGR
HLSFQTWLAUQRHWDMWLGUSGIKKFLRYVCWVSPGPMLKASSJVOQXE
GGVEYGGZMLJCXXLJSVPAIVWIKVRDRYGFRJLJSLVEGGVEYGGEI
APUUISFPBTGNWWMUCZRVTWGLRWUGUMNCZVILE

- Ideal: More coincidences appear for the same key letter.



Dispacement:	1	2	3	4	5	6
Coincidences:	14	14	16	14	24	12

Why?

Table 1.3: English Letter Frequencies as Percentages

Letter	Relative Frequency	Letter	Relative Frequency
A	8.399	N	6.778
B	1.442	O	7.493
C	2.527	P	1.991
D	4.800	Q	0.077
E	12.15	R	6.063
F	2.132	S	6.319
G	2.323	T	8.999
H	6.025	U	2.783
I	6.485	V	0.996
J	0.102	W	2.464
K	0.689	X	0.204
L	4.008	Y	2.157
M	2.566	Z	0.025

V=(0.08399, 0.01442, …, 0.00025)

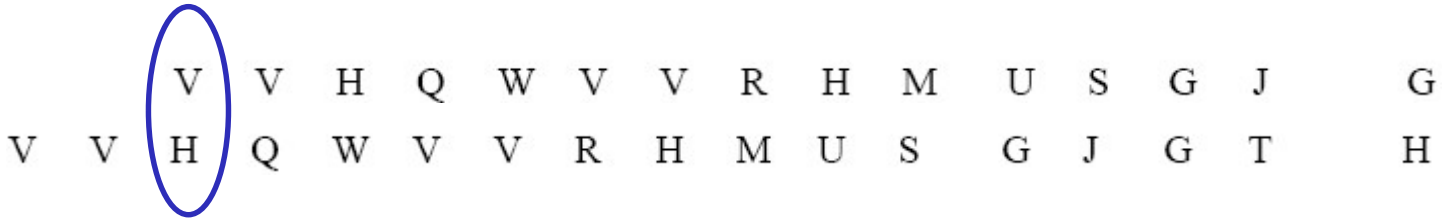
V_0=V

V_1= (0.01442, …, 0.00025, 0.08399)

V_2= (0.02527, …, 0.08399, 0.01442)

…

V=(0.08399, 0.01442, …, 0.0025)



The probability of coincidence is $V_i \cdot V_j$

$V_i \cdot V_j \leq V \cdot V$

In other words, when $V_i = V_j$, $V_i \cdot V_j$ is maximal.

When the key length is 3, the distributions of ciphertexts:

$V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ \ldots\ldots$

Dispacement of 1, 2, and 3:

$\phantom{V_{i1}\ }V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ \ldots\ldots$

$V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ \ldots\ldots$

$\phantom{V_{i1}\ V_{i2}\ }V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ \ldots\ldots$

$V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ \ldots\ldots$

$\phantom{V_{i1}\ V_{i2}\ V_{i3}\ }V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ \ldots\ldots$

$V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ V_{i1}\ V_{i2}\ V_{i3}\ \ldots\ldots$

When the dispacement number is 3 or the multiple of 3, distributions are the same, therefore the probability of coincidence is maximal.

Determining the key length(2) Index of Coincidence

- Assume ciphertext is English letters
- Let n_0 be number of As, n_1 number of Bs, …, n_{25} number of Zs in ciphertext
- Let $n = n_0 + n_1 + \ldots + n_{25}$
- Define index of coincidence

$$I = \frac{\binom{n_0}{2} + \binom{n_1}{2} + \cdots + \binom{n_{25}}{2}}{\binom{n}{2}} = \frac{1}{n(n-1)} \sum_{i=0}^{25} n_i(n_i - 1)$$

- What does this measure?

Index of Coincidence

- Gives the probability that 2 randomly selected letters are the same
- For plain English, prob. 2 letter are same:
  - $p_0^2 + p_1^2 + \ldots + p_{25}^2 \approx$ **0.065**, where p_i is probability of i^{th} letter
- Then for simple substitution, $\mathrm{I} \approx 0.065$
- For random letters, each $p_i = 1/26$
  - Then $p_0^2 + p_1^2 + \ldots + p_{25}^2 \approx$ **0.03846**
- Then $\mathrm{I} \approx 0.03846$ for poly-alphabetic substitution with a very long keyword

Index of Coincidence

- How to use this to estimate length of keyword in Vigenere cipher?
- Suppose keyword is length k, message is length n
  - Ciphertext in matrix with k columns, n/k rows
- Select 2 letters from same columns
  - Like selecting from simple substitution
- Select 2 letters from different columns
  - Like selecting random letters

Index of Coincidence

- Suppose k columns and n/k rows
- Approximate number of matching pairs from same column, but 2 different rows:

$$0.065\binom{\frac{n}{k}}{2}k = 0.065\,\frac{1}{2}\left(\frac{n}{k}\right)\left(\frac{n}{k}-1\right)k = 0.065\left(\frac{n(n-k)}{2k}\right)$$

- Approximate number of matching pairs from 2 different columns, and any two rows:

$$0.03846\binom{k}{2}\left(\frac{n}{k}\right)^2 = 0.03846\,\frac{n^2(k-1)}{2k}$$

Index of Coincidence

- Approximate index of coincidence by:

$$I \approx \frac{0.03846\,\frac{n^2(k-1)}{2k} + 0.065\left(\frac{n(n-k)}{2k}\right)}{\binom{n}{2}}$$
$$= \frac{0.03846n(k-1) + (0.065)(n-k)}{k(n-1)}$$

- Solve for k to find:

$$k \approx \frac{0.02654n}{(0.065 - I) + n(I - 0.03846)}$$

- Use n and I (known from ciphertext) to approximate length of Vigenere keyword

Index of Coincidence: Bottom Line

- A crypto breakthrough when invented
  - By William F. Friedman in 1920s
- Useful against classical and WWII-era ciphers
- Index of coincidence is a well-known statistical test
  - Many other statistical tests exists

Finding the key (1)

V V H Q W
V V R H M
U S G J G
T H K I H
T S S E J
C H L S F
C B G V W
C R L R Y
Q T F S V
G A H W K
C U H W A
U G L Q H

A	B	C	D	E	F	G	H	I	J	K	L	M
0	0	7	1	1	2	9	0	1	8	8	0	0

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
3	0	4	5	2	0	3	6	5	1	0	1	0

G, J, K, C

e→J => x →C

e→K => j →P, k →Q

e→C => x →V

e→G

The key is: **2 14 3 4 18**

c o d e s

Plaintext:

themethodusedforthepreparationandreadingofcodemessagesissimpleinthe
extremeandatthesametimeimpossibleoftranslationunlessthekeyisknownth
eeasewithwhichthekeymaybechangedisanotherpointinfavoroftheadoptiono
fthiscodebythosedesiringtotransmitimportantmessageswithouttheslight
estdangeroftheirmessagesbeingreadbypoliticalorbusinessrivalsetc

Finding the key (2)

A	B	C	D	E	F	G	H	I	J	K	L	M
0	0	7	1	1	2	9	0	1	8	8	0	0

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
3	0	4	5	2	0	3	6	5	1	0	1	0

$$(0, 0, 7, 1, 1, 2, 9, 0, 1, 8, 8, 0, 0, 3, 0, 4, 5, 2, 0, 3, 6, 5, 1, 0, 1, 0)$$

$$W = (0, 0, 0.1045, 0.0149, 0.0149, 0.0299, \cdots, 0.0149, 0)$$

The key j gives the maximum of $W \cdot V_j$

0.0250, 0.0391, 0.0713, 0.0388, 0.0275, 0.0380, 0.0512, 0.0301, 0.0325,
0.0430, 0.0338, 0.0299, 0.0343, 0.0446, 0.0356, 0.0402, 0.0434, 0.0502,
0.0392, 0.0296, 0.0326, 0.0392, 0.0366, 0.0316, 0.0488, 0.0349