Key to breaking substitution: freq analysis
- attributed to Al Kindi circa 900s.

Thwart freq analysis: 00 to 99
E → to 12/13 symbols

Vigenere Key: KNIGHTS
10, ~~14~~ 13, 8, 6, 7, 19, 18

Plain: GOOD MORNING
6, 14, 14, 3, 12, 14, 17, 13, 8, 13, 6

\+ KEY 10, 13, 8, 6, 7, 19, 18, 10, 13, 8, 6

16, 27, 22, 9, 19, 33, 35, 23, 21, 21, 12

Ciphertext: Q B W J T H J X V V M

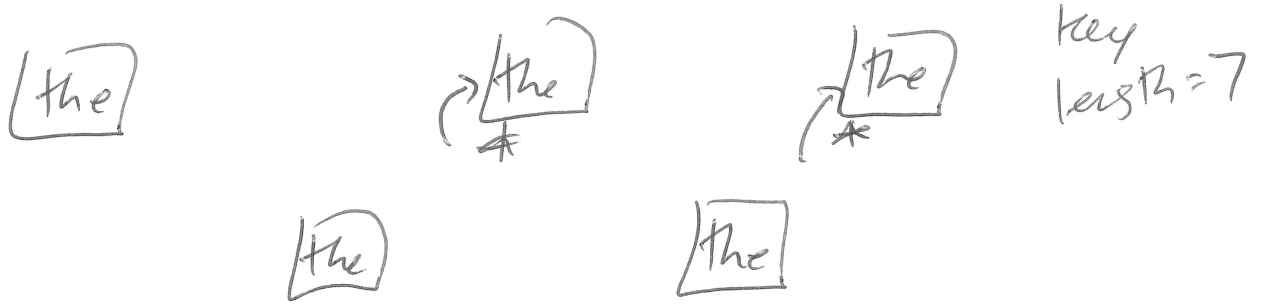
freq + infreq ⟹ same ciphertext
same letter ⟹ diff ciphertext

Paradigm Shift

If I knew the keyword length, L, then I could create L bins of ciphertext letters, where in each bin, the letters were shifted the same way.

How to find the keyword length?

1800s - Kasiski Test (identifies the keyword length)



$p(\text{2nd the diff 1st}) = \frac{6}{7}$

$p(\text{3rd diff than 1st, 2nd}) = \frac{5}{7}$

$\frac{6}{7} \times \frac{5}{7} \times \frac{4}{7} \times \frac{3}{7} = \frac{360}{2401}$

$\sim$ 15%

look for repeated strings in ciphertext (length 3 or greater)

↳ probability of accidental hit is low.

take their index values in ciphertext

CZP: 135, 233, 282

$L \mid (233-135) \rightarrow L \mid 98$

and $L \mid (282-233)$ $\quad L \mid 49$

$\gcd(98,49) = 49$

Second method to find Key Word Length

Index of Coincidence Test

$IC(S)$ = the probability that 2 randomly selected items from S are the same.

(S: multiset)

$S = \{10\text{ As}, 20\text{ Bs}, 5\text{ Cs}, 15\text{ Ds}\}$

$IC(S) = prob(2\text{ As}) + prob(2\text{Bs}) + prob(2\text{Cs}) + prob(2\text{Ds})$

$$\frac{10}{50}\times\frac{9}{49} + \frac{20}{50}\times\frac{19}{49} + \frac{5}{50}\times\frac{4}{49} + \frac{15}{50}\times\frac{14}{49}$$

freq of letter i is f_i, $n = \sum_{i=1}^{k} f_i$ (k unique letters)

$$IC(S) = \frac{\sum_{i=1}^{k} f_i(f_i-1)}{n(n-1)}$$

$IC(\text{reg Eng text}) \approx 0.0676$

$IC(\text{random letters}) = \frac{1}{26} = 0.0378$

(1) Guess key word length (L)

(2) Split ~~bins into~~ ciphertext into L bins.

(3) for each bin calculcle its Index of Coincidence.