Classical Cryptography

Monoalphabetic cryptanalysis

Karst Koymans

Informatics Institute

University of Amsterdam

(version 22.4, 2023/02/13 13:41:58 UTC)

Tuesday, February 14, 2023

1. Statistical Cryptanalysis
   - Frequencies
   - The index of coincidence: ϕ- and χ-tests
2. Example
3. Countermeasures against statistical cryptanalysis
   - Homophones
   - Polyalphabetic substitutions

Outline

1. Statistical Cryptanalysis
   - Frequencies
   - The index of coincidence: ϕ- and χ-tests
2. Example
3. Countermeasures against statistical cryptanalysis
   - Homophones
   - Polyalphabetic substitutions

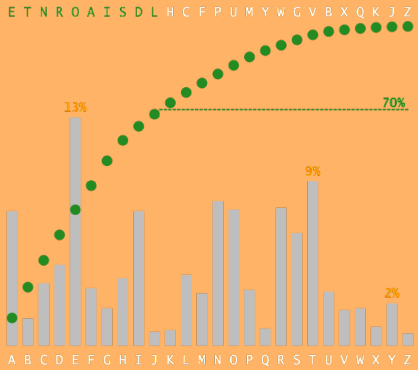
Outline

1. Statistical Cryptanalysis
   - Frequencies
   - The index of coincidence: ϕ- and χ-tests
2. Example
3. Countermeasures against statistical cryptanalysis
   - Homophones
   - Polyalphabetic substitutions

Letter frequencies

- A simple method to attack monoalphabetic ciphers
  - **Letter frequency analysis**
- Some letters occur more (or less) than others
  - This is (somewhat) language dependent

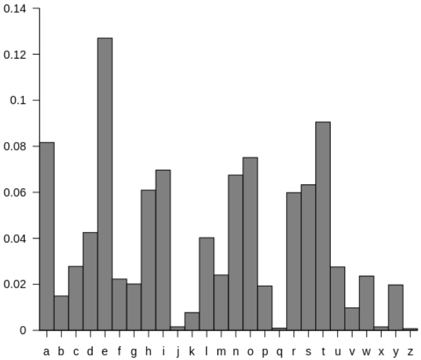
Letter frequency diagram



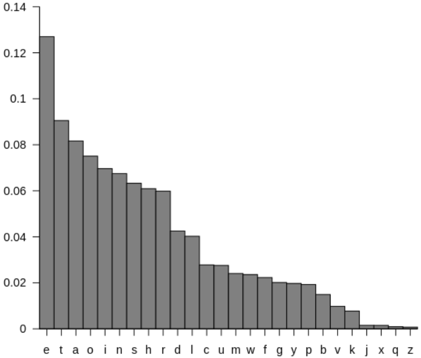
Source: Slides Hans van der Meer

Unknown language or text source

English letter frequency



Ordered by alphabet



Ordered by frequency

Source: `https://en.wikipedia.org/wiki/Letter_frequency`

Outline

1. Statistical Cryptanalysis
   - Frequencies
   - The index of coincidence: ϕ- and χ-tests
2. Example
3. Countermeasures against statistical cryptanalysis
   - Homophones
   - Polyalphabetic substitutions

William Friedman

William Friedman

The index of coincidence (IoC)

- Introduced by **William Friedman**
- Probability that two letters chosen randomly from a text, based on an alphabet of n letters, are the same
- Given probabilities of occurrence $p_0, \ldots, p_{n-1}$ for the n letters
  - $\text{IoC} = \sum_{i=0}^{n-1} p_i^2$
- For text with a (uniformly) random frequency distribution this reduces theoretically (obviously) to $1/n$ (≈ 0.038 for $n = 26$)
- For an English text (with the English frequency distribution) this amounts to ≈ 0.066, or $\approx 1/15$, found by doing experiments

The ϕ-test

- The IoC clearly distinguishes English text from random text
- Friedman observed that the IoC is **invariant under monoalphabetic substitution**
- Using the IoC to check for monoalphabeticity is called the ϕ-test
- For an unknown ciphertext of length $M > 1$ this test calculates
  - $\text{IoC} = \sum_{t=A}^{Z} f_t(f_t - 1)/M(M-1)$
  - Here f_t is the number of occurrences of the letter t
  - For small texts the -1 is used to avoid counting identity as equality
    - Hence letters that occur only once don't contribute to the IoC

Breaking Caesar (by hand and automatically)

- Brute force 26 keys and see if you get plaintext (we did this before)
- Match (visually) the frequency distribution of the cryptogram to standard English by shifting the frequency graph
- To automate this the ϕ-test doesn't help, use the χ-test instead
  - The χ-test is also called cross-product sum
  - Consider two texts f and g of length M and N and calculate $\chi = \sum_{t=A}^{Z} f_t g_t / MN$
  - Find the highest χ value after comparing the shifted frequency diagram of the cryptogram with that of normal English text

Breaking general monoalphabetic substitutions

- First use the ϕ-test to check for monoalphabeticity
- Order the ciphertext letter distribution by frequency and

  try to match this with the standard English letter distribution

  or whatever language you may suspect is being used
- Look at digraph – or even trigraph – frequencies
- Look at beginning and ending of words (each has a different frequency distribution)
- Check vowels versus consonants and other letter patterns
- Look at keywords for alphabet construction
- Try to find cribs

Outline

1 Statistical Cryptanalysis
- Frequencies
- The index of coincidence: ϕ- and χ-tests

2 Example

3 Countermeasures against statistical cryptanalysis
- Homophones
- Polyalphabetic substitutions

Math of Secrets: 2.2 monoalphabet

QBVDL WXTEQ GXOKT NGZJQ GKXST RQLYR
XJYGJ NALRX OTQLS LRKJQ FJYGJ NGXLK
QLYUZ GJSXQ GXSLQ XNQXL VXKOJ DVJNN
BTKJZ BKPXU LYUNZ XLQXU JYQGX NTYQG
XKXQJ KXULK QJNQN LQBYL OLKKX SJYQG
XNGLU XRSBN XOFUL YDSXU GJNSX DNVTY
RGXUG JNLEE SXLYU ESLYY XUQGX NSLTD
GQXKB AVBKX JYYBR XYQNQ GXKXZ LNYBS
LRPBA VLQXK JLSOB FNGLE EXYXU LSBYD
XWXKF SJQQS XZGJS XQGXF RLVXQ BMXXK
OTQKX VLJYX UQBZG JQXZL NG

Exercise 1

Exercise 1

- Count letters and make a table of frequencies
- Generate a frequency diagram, using a spreadsheet
- Calculate the Index of Coincidence
- Is it an additive cipher?
- Try to solve the cryptogram by assuming it is affine

Outline

1. Statistical Cryptanalysis
   - Frequencies
   - The index of coincidence: ϕ- and χ-tests
2. Example
3. Countermeasures against statistical cryptanalysis
   - Homophones
   - Polyalphabetic substitutions

Outline

1. Statistical Cryptanalysis
   - Frequencies
   - The index of coincidence: ϕ- and χ-tests
2. Example
3. Countermeasures against statistical cryptanalysis
   - Homophones
   - Polyalphabetic substitutions

Homophones

- Homophones
  - A classic way to flatten frequency distributions
  - Introduce more than one ciphertext letter option for (some of) the plaintext letters
    - Especially for plaintext letters with high frequency
    - Needs a larger ciphertext alphabet
  - This is an example where the encryption function may be randomised (to a small extent)
  - The Zodiac Killer used homophones in both Z408 and Z340
    - But could have done a better job in randomisation

Math of Secrets: 2.2 homophones

```
IW*CI W@G*L &H&L( ASN*A E)U&V $CNPC
SIW*E DDSA@ LTCIH !(A#C V%EIW *!#HA
*IW@N TAEHR $CI(C JTS!C SHDS# SIW@S
DVW@R G$HH* SIW*W )JH@( CUGDC IDUIW
*&AIP GWTUA TLS$L CIW*D IWTG! #HATW
TRG$H H*SQT U$G*I W@S)D GHWTR APBDG
*S%EI W@WDB @HIG@ IRWWX H&CV+ XHWVG
*LLXI WW#HE G)VG@ HHI#A AEGTH @CIAN
W*L!H Q%I!L )DAAN R)BTI B)K#C VXC#I
HDGQX ILXIW IW@VA *&B!C SIWTH E**S$
UA(VW I
```

Exercise 2

Exercise 2

- Count symbols and make a table of frequencies
- Generate a frequency diagram, using a spreadsheet
- Calculate the Index of Coincidence for all symbols
- Calculate the Index of Coincidence for only the letters
- Is it a monoalphabetic cipher?
- Identify homophones and solve the cryptogram

Outline

1. Statistical Cryptanalysis
   - Frequencies
   - The index of coincidence: ϕ- and χ-tests
2. Example
3. Countermeasures against statistical cryptanalysis
   - Homophones
   - Polyalphabetic substitutions

Polyalphabetic substitutions

Definition (polygraphic)

A **polygraphic substitution** is the replacement of groups of letters by other groups of letters according to one big substitution table

Definition (polyliteral)

A **polyliteral substitution** is the replacement of single letters by groups of letters according to one big substitution table

Definition (polyalphabetic)

A **polyalphabetic substitution** is the replacement of single letters by other letters by using a varying ciphertext alphabet for encrypting each plaintext letter