Classical Cryptography

Monoalphabetic cryptanalysis

Karst Koymans

Informatics Institute

University of Amsterdam

(version 19.2, 2020/02/08 13:18:13 UTC)

Monday, February 10, 2020

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

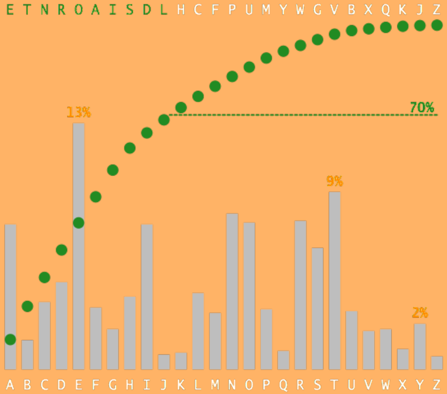
Outline

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

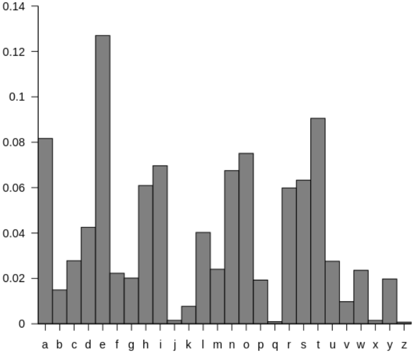
Letter frequency diagram



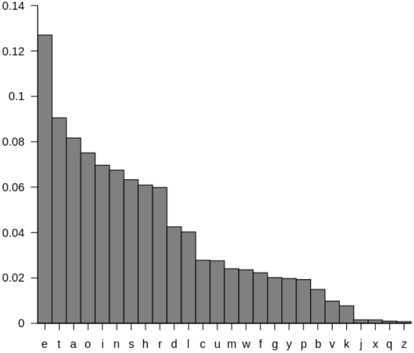
Source: Slides Hans van der Meer

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

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 $p_0, \ldots, p_{n-1}$ for the n letters
  - 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, 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 this test calculates
  - 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, respectively and calculate $\sum_{t=A}^{Z} f_t g_t / MN$
    - Find highest χ value for comparison between shifted frequency diagram of cryptogram and English text

Breaking monoalphabetic substitutions

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

  and try to match this with standard English

  (or whatever language you may suspect is being used)
- Look at digraph (or even trigraph) frequencies
- Look at beginning and ending of words (different frequencies)
- 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 randomized (to a small extent)

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

A **polyalphabetic substitution** is the replacement of letters by other letters by using a (possibly) different alphabet for each plaintext letter

- Poly**alphabetic** uses different alphabets per plaintext letter
- Poly**graphic** uses a larger alphabet for plaintext and ciphertext
- Poly**literal** uses a larger alphabet for ciphertext only