College Cryptografie

Cursusjaar 2003

Statistische Analyse

22 februari 2003

Monoalfabetische distributie

Phi-test monoalfabeticiteit

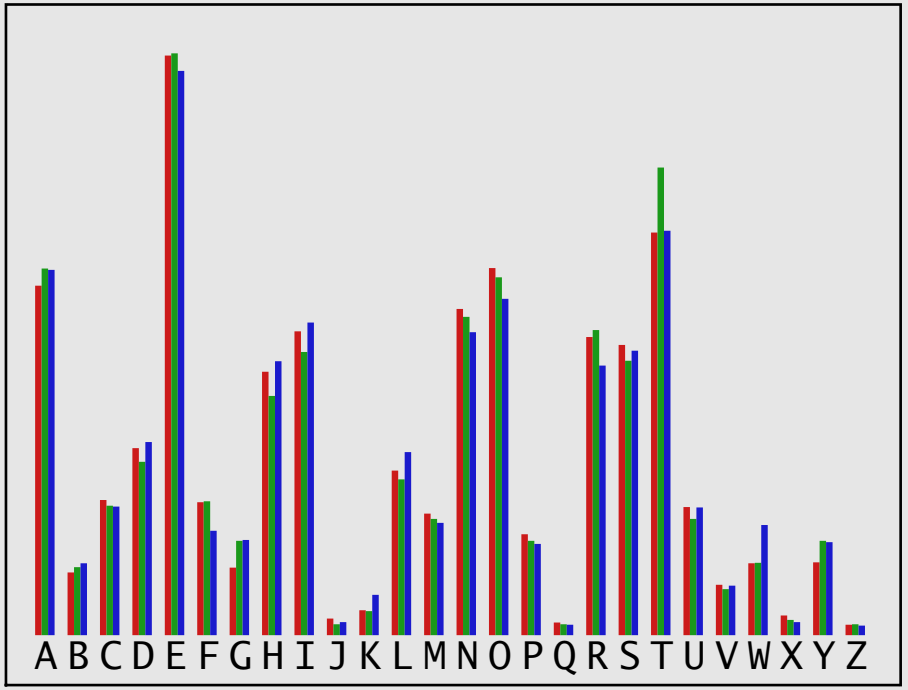
Periode bepalen

Aantal letters bepalen

Coïncidenties

Chi-test

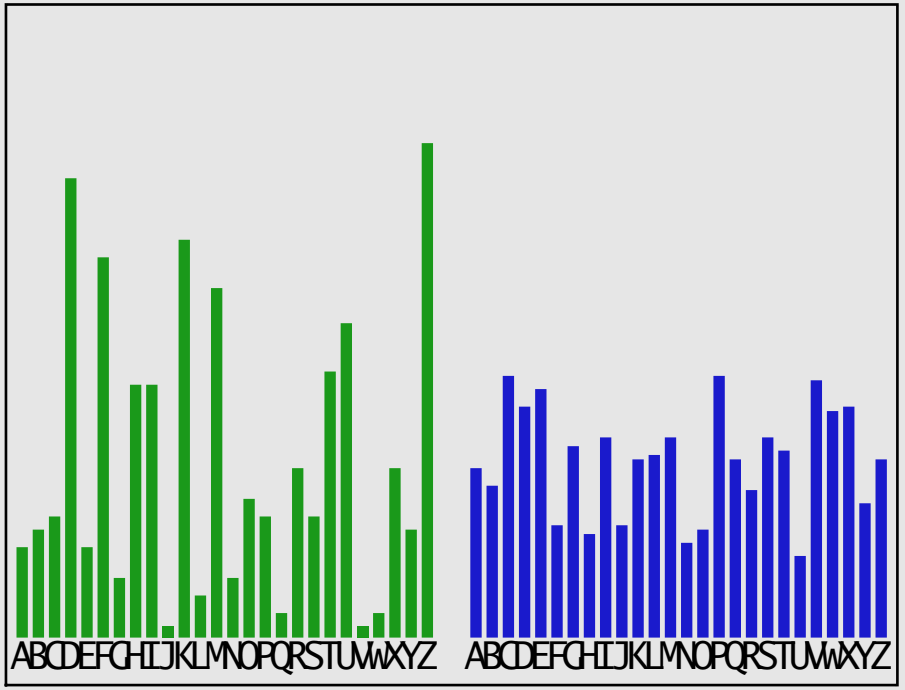
- letterdistributie en tweede moment
- phi-test voor bepaling monoalfabeticiteit
- bepaling periode polyalfabeet
- bepaling aantal letters
- coïncidenties
- chi-test voor matching distributies



Bronnen:

- 1933 (Gaines)
- 1942 (SI-Course)
- 1982 (Beker)

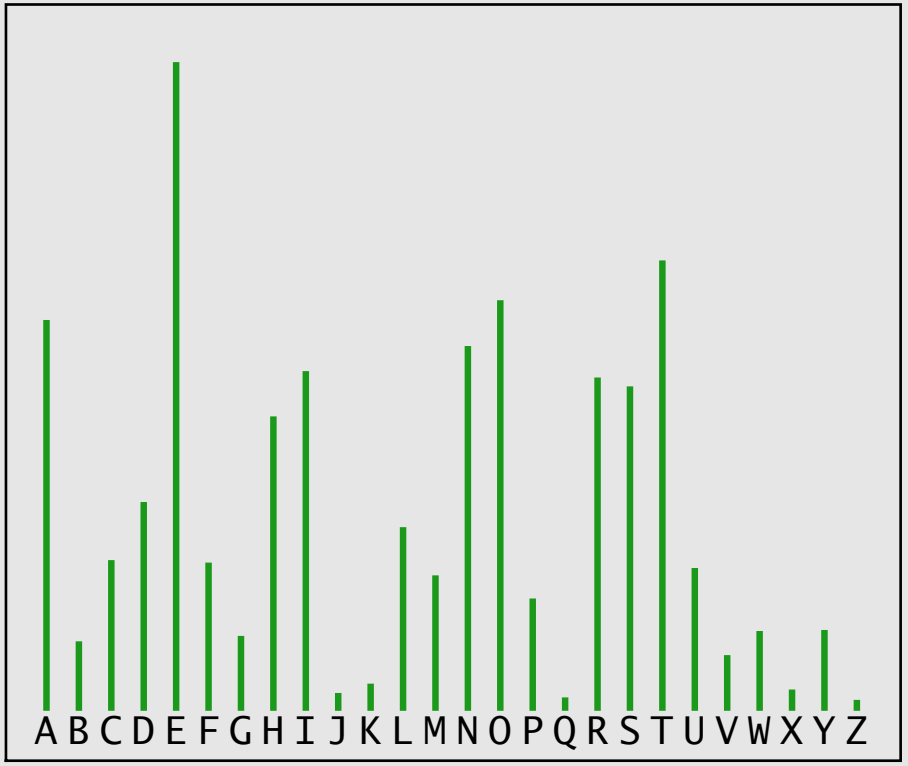
Monoalfabetische distributie



Vercijfering:

- Monoalfabetisch
- Polyalfabetisch

Polyalfabetische distributie



$$S_2 = \sum_{i=A}^{Z} p_i^2$$

taal	S_2
Engels	0.066
Frans	0.078
Duits	0.076
Russisch	0.053
Romaji	0.082
random	0.038

Tweede moment

Kans op bepaalde letter:

$$p_i = \lim_{N \to \infty} \frac{f_i}{N} \quad \forall i = A, \ldots, Z \quad \text{met } N = \sum_{i=A}^{Z} f_i$$

Tweede moment:

$$S_2 = \sum_{i=A}^{Z} p_i^2$$

Zuivere schatter Φ voor S_2:

$$\Phi = \frac{\sum_{i=A}^{Z} f_i(f_i - 1)}{N(N-1)} \quad \text{met } N = \sum_{i=A}^{Z} f_i$$

reken verwachtingswaarde $E(\Phi)$ uit:

$$E(\Phi) = E\left(\sum_{i=A}^{Z} f_i(f_i - 1)\right) = \sum_{i=A}^{Z} E(f_i(f_i - 1)) = \sum_{i=A}^{Z} \left(E(f_i^2) - E(f_i)\right)$$

multinomiale verdeling voor kansen:

$$p(f_A, f_B, \ldots, f_Z) = \frac{N!\, p_A^{f_A} p_B^{f_B} \ldots p_Z^{f_Z}}{f_A! f_B! \ldots f_Z!}$$

maak gebruik van:

$$E(f_i) = p_i \mid var(f_i) = N p_i (1 - p_i) \mid var(f_i) = E(f_i^2) - E(f_i)^2$$

eliminatie van $E(f_i^2)$:

$$E(\Phi) = \sum_{i=A}^{Z} \left(var(f_i) + E(f_i)^2 - E(f_i) \right)$$

substitutie van $E(f_i) = p_i$ en $var(f_i) = Np_i(1 - p_i)$:

$$E(\Phi) = \sum_{i=A}^{Z} \left(Np_i(1 - p_i) + (Np_i)^2 - Np_i \right)$$

resultaat:

$$E(\Phi) = N(N-1) \sum_{i=A}^{Z} p_i^2 = N(N-1)S_2$$

Afleiding Phi (vervolg)

monoalfabetische substitutie:

ORDE LEIDT TOT ALLE DEUGDEN

WHZM VMQZU UWU YVVM ZMLXZMG

```
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
        1 1       1 5       1       3 3 2 1 1 4
```

$\Phi = 46$ (gevonden)

$\Phi_m = 23 \times 22 \times 0,076 = 38,4 \pm 14,0$ (monoalfabetisch)

$\Phi_r = 23 \times 22 \times 0,038 = 19,2 \pm 6,1$ (random)

$$var(\phi) = 4N^3(S_3 - S_2^2) + 2N^2(5S_2^2 + S_2 - 6S_3) + 2N(4S_3 - S_2 - 3S_2^2)$$

klaartekst autoclaaf cryptogram:

VYAND NADER TWATE RLINI ESTOP FORTA SPERE NPARA AT
KCNGD TOQZP TJDGE UPZGE ELXFA NBZXS LDTWS EIAJP EK

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
2	1	1	3	5	1	3		1	2	2	2		2	1	3	1		2	3	1		1	2		3

Φ = 64 (gevonden)

$\Phi_m = 42 \times 41 \times 0,076 = 131 \pm 32$ (monoalfabetisch)

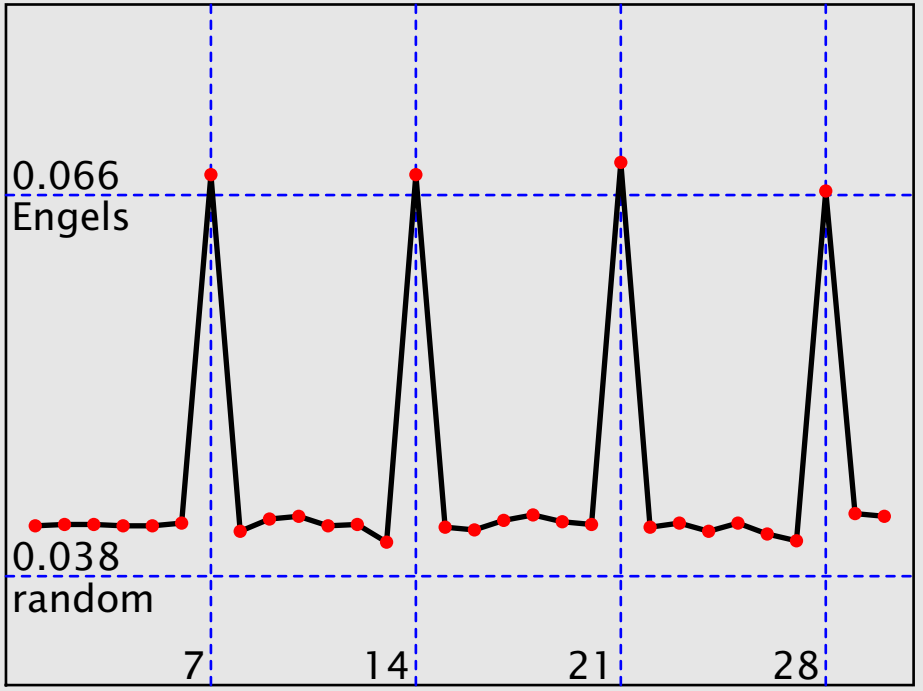
$\Phi_r = 42 \times 41 \times 0,038 = 65 \pm 11$ (random)

Niet monoalfabetisch

Recept:

- veronderstel periode p
- verdeel in p groepen met $c_i \in \{k\}$ iff $i \bmod p \equiv k$
- bereken Φ_j voor alle p groepen
- bereken gemiddelde $\overline{\Phi} = \sum \Phi_j / p$
- test $\overline{\Phi}$ tegen Φ_{mono}
- herhaal voor $p = 1, 2, \ldots$

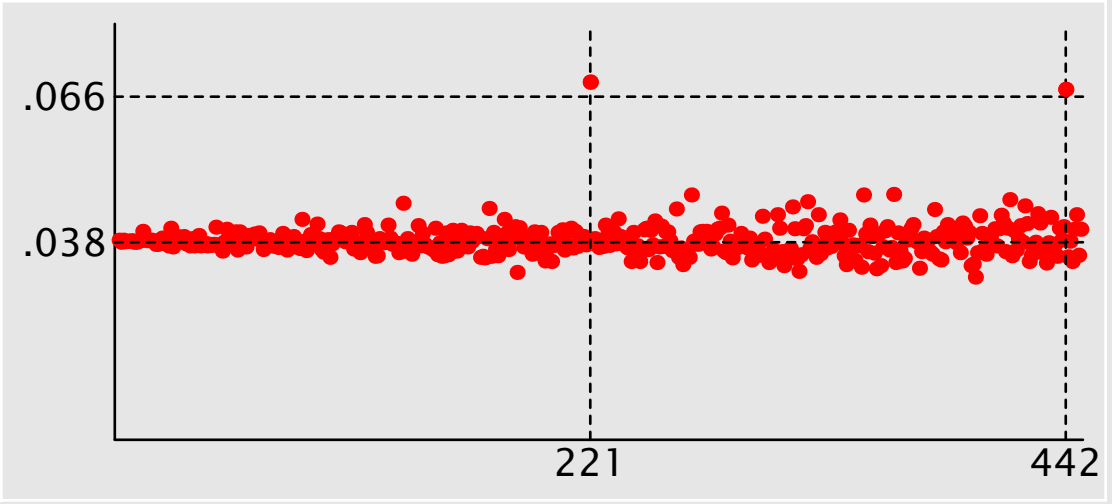
Periode bepalen



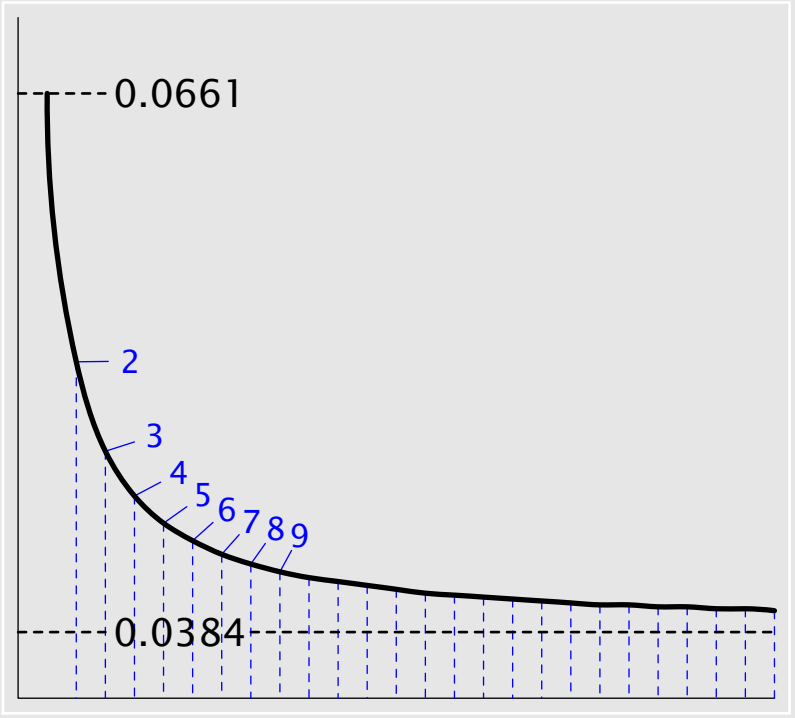
Phi-test op
Vigenère
Engels
1673 letters
periode 7

Voorbeeld periode bepalen

Phi-test polyalfabeet, Engels,
2272 letters, periode 221



Ander voorbeeld



voor 1 → 26 letters
geleidelijke overgang
$S_2^{mono} \rightarrow S_2^{26} > S_2^{random}$

als alle k letters in
sleutel verschillend:
$E(\Phi) =$

$$\frac{N(N-1)}{k}\left(S_2 + \frac{k-1}{26}\right)$$

Aantal letters bepalen

XWVVQBRTDDGZMBDEQPFMPMTUZQ· · ·FVATFCKYZPJBMLFEPXZDODIUBILOAM

LBTNTIFYOIVTOZWHEBTMTDTINV· · ·OVSANHWLQQDUFHXMIOUFFGMGXTQGON

PLANNENVANDEZEERQVERSVIELD· · ·ORTEAFSTANDVANHETEILANDGEWORPE

ITTENVOUWDEKAARTOPENENLEGH· · ·AATHETNOORDENMISSCHIENRONDKYKE

100 letters → kans = $100 \times S_2$:

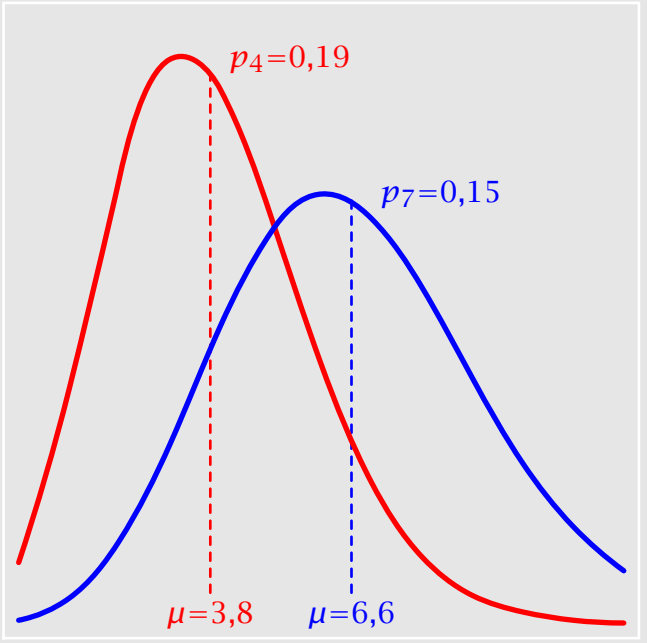
$$m_{random} = 4 \pm 2$$

$$m_{mono} = 7 \pm 3$$

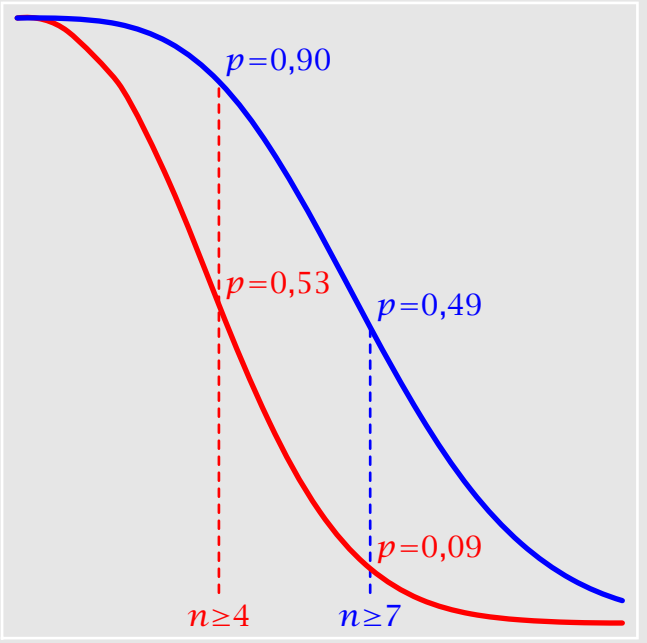
Poisson verdeling met verwachting m en $\sigma = \sqrt{m}$:

$$P(0) = e^{-m},\ P(1) = me^{-m},\ P(2) = \frac{m^2 e^{-m}}{2!}\ \cdots$$

kans en cumulatieve kans

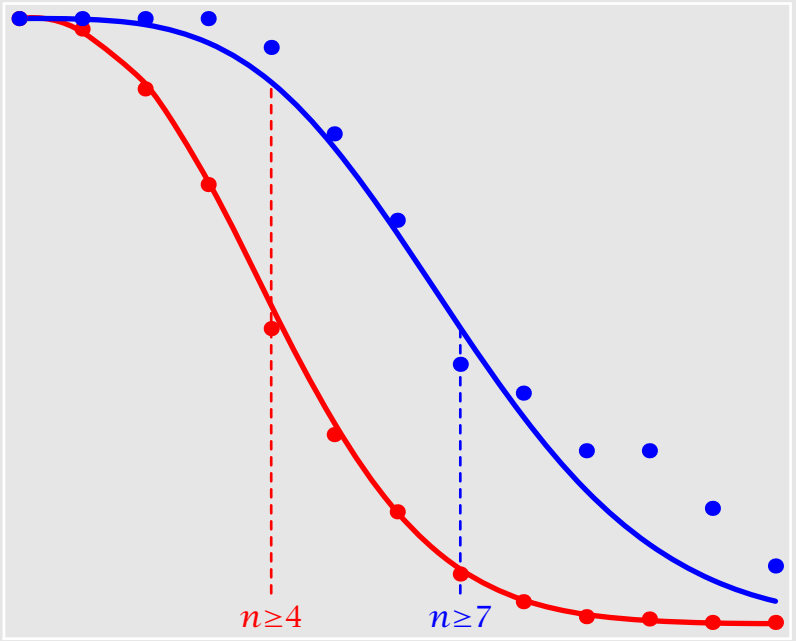


kans op n events



kans op n of meer events

Poisson verdeling



Coïncidenties
cumulatief geteld
Vigenère 1673 letters
opgedeeld naar periode 7
100 letters per telling

	525 geen match	
	21 match	
# ≤ 4:	361	4
# ≥ 7:	43	9
# ≥ 12:	1	2

Voorbeeld Poisson

Meet overeenkomst tussen twee distributies:

$$\chi = \sum_{i=A}^{Z} f_i f_i' \quad \text{met } N = \sum_{i=A}^{Z} f_i \text{ en } N' = \sum_{i=A}^{Z} f_i'$$

Overeenkomende distributies:

$$E(\chi) = NN' \sum_{i=A}^{Z} \left(\frac{f_i}{N}\right)\left(\frac{f_i'}{N'}\right) = NN' \sum_{i=A}^{Z} p_i^2 = NN'S_2$$

Verschillende distributies:

$$E(\chi) = \frac{NN'}{n} \text{ met } n \text{ het aantal symbolen (26)}$$

Vigenère, 1673 letters, periode 7:

- 21 matchende distributies
  $E_{21}(\chi) = 3776 \pm 162$
  gevonden $\chi = 3819 \pm 149$ met range $3599 \cdots 4121$

- 525 niet matchende distributies
  $E_{525}(\chi) = 2197 \pm 361$
  gevonden $\chi = 2132 \pm 303$ met range $1432 \cdots 2909$