Die Zeichenkoinzidenz zweier Texte[1]

Definition

Sei Σ ein endliches Alphabet. Seien $a = (a_0, \ldots, a_{r-1})$ und $b = (b_0, \ldots, b_{r-1}) \in \Sigma^r$ zwei gleich lange Texte. Dann heißt

$$\kappa(a,b) := \frac{1}{r} \cdot \#\{j \mid a_j = b_j\} = \frac{1}{r} \cdot \sum_{j=0}^{r-1} \delta_{a_j b_j}$$

die **Zeichenkoinzidenz** von a und b. [δ = KRONECKER-Symbol.]

Sie misst also die Übereinstimmung der beiden Texte und definiert für jedes $r \in \mathbb{N}_1$ eine Abbildung

$$\kappa\colon \Sigma^r \times \Sigma^r \longrightarrow \mathbb{Q}.$$

Bemerkungen

1. $0 \leq \kappa(a,b) \leq 1$ stets.
2. $\kappa(a,b) = 1 \Longleftrightarrow a = b$.
3. Als Konvention setzt man $\kappa(\emptyset, \emptyset) = 1$.

Invarianzen

Die Zeichenkoinzidenz ist eine Invariante der polyalphabetischen Verschlüsselung (bei gleichem Schlüssel):

Satz 1 (Invarianzsatz) (i) *Sei $f : \Sigma^* \longrightarrow \Sigma^*$ eine polyalphabetische Verschlüsselungsfunktion. Dann ist*

$$\kappa(f(a), f(b)) = \kappa(a,b)$$

für alle gleich langen $a, b \in \Sigma^*$.

(ii) *Ebenso ist die Zeichenkoinzidenz invariant unter einer Transposition.*

[1] Klaus Pommerening, Kryptologie; 27. November 1999, letzte Änderung: 20. Mai 2002

Mittelwerte

Wir bestimmen für festes $a \in \Sigma^r$ den Mittelwert von $\kappa(a,b)$ über alle $b \in \Sigma^r$:

$$\begin{aligned}
\frac{1}{n^r} \cdot \sum_{b\in\Sigma^r} \kappa(a,b) &= \frac{1}{n^r} \cdot \sum_{b\in\Sigma^r} \left[\frac{1}{r} \cdot \sum_{j=0}^{r-1} \delta_{a_j b_j} \right] \\
&= \frac{1}{rn^r} \cdot \sum_{j=0}^{r-1} \underbrace{\left[\sum_{b\in\Sigma^r} \delta_{a_j b_j} \right]}_{n^{r-1}} \\
&= \frac{1}{rn^r} \cdot r \cdot n^{r-1} = \frac{1}{n},
\end{aligned}$$

da bei festem $b_j = a_j$ noch n^{r-1} Möglichkeiten für b bestehen.

Analog bestimmen wir den Mittelwert von $\kappa(a, f_\sigma(b)$ für feste $a, b \in \Sigma^r$ über alle Permutationen $\sigma \in \mathcal{S}(\Sigma)$:

$$\begin{aligned}
\frac{1}{n!} \cdot \sum_{\sigma\in\mathcal{S}(\Sigma)} \kappa(a, f_\sigma(b)) &= \frac{1}{n!} \cdot \frac{1}{r} \sum_{\sigma\in\mathcal{S}(\Sigma)} \#\{j \mid \sigma b_j = a_j\} \\
&= \frac{1}{rn!} \cdot \#\{(j,\sigma) \mid \sigma b_j = a_j\} \\
&= \frac{1}{rn!} \cdot \sum_{j=0}^{r-1} \#\{\sigma \mid \sigma b_j = a_j\} \\
&= \frac{1}{rn!} \cdot r \cdot (n-1)! = \frac{1}{n},
\end{aligned}$$

denn genau $(n-1)!$ Permutationen bilden a_j nach b_j ab.

Damit ist gezeigt:

Satz 2 (i) *Der Mittelwert von $\kappa(a,b)$ über alle gleichlangen Texte b ist $\frac{1}{n}$ für alle $a \in \Sigma^*$.*

(ii) *Der Mittelwert von $\kappa(a,b)$ über alle $a, b \in \Sigma^r$ ist $\frac{1}{n}$ für alle $r \in \mathbb{N}_1$.*

(iii) *Der Mittelwert von $\kappa(a, f_\sigma(b))$ über alle monoalphabetischen Substitutionen ist $\frac{1}{n}$ für beliebige gleich lange Texte $a, b \in \Sigma^*$.*

(iV) *Der Mittelwert von $\kappa(f_\sigma(a), f_\tau(b))$ über alle Paare von monoalphabetischen Substitutionen ist $\frac{1}{n}$ für beliebige gleich lange Texte $a, b \in \Sigma^*$.*

Deutung.

- Für einen gegebenen Text a und einen „zufälligen“ gleich langen Text b ist $\kappa(a,b) \approx \frac{1}{n}$.
- Für „zufällige“ gleich lange Texte a und b ist $\kappa(a,b) \approx \frac{1}{n}$.

- Für gegebene gleich lange Texte a und b und eine „zufällige“ monoalphabetische Substitution f_σ ist $\kappa(a, f_\sigma(b)) \approx \frac{1}{n}$.
- Für gegebene gleich lange Texte a und b und zwei „zufällige“ monoalphabetische Substitution f_σ, f_τ ist $\kappa(f_\sigma(a), f_\tau(b)) \approx \frac{1}{n}$.
- Das gleiche gilt, da sich die Zählung der Koinzidenzen bezüglich beliebiger Zerlegungen der Texte additiv verhält, auch für „zufällige“ polyalphabetische Substitutionen.

Werte, die deutlich davon abweichen, erwecken im Kryptoanalytiker den Verdacht, *nicht* zufällig zustande gekommen zu sein.