Die Formel von SINKOV[1]

Wendet man die Näherungsformeln aus Anwendung 2 im letzten Absatz weiter auf den Koinzidenzindex eines periodisch polyalphabetisch verschlüsselten Textes $c = f(a)$ mit $a \in M$ an, so folgt im Fall $l|r$:

$$\begin{aligned}
\varphi(c) &= \frac{1}{r-1} \cdot [\kappa_1(c) + \cdots + \kappa_{r-1}(c)] \\
&\approx \frac{1}{r-1} \cdot \left[(\frac{r}{l} - 1) \cdot \kappa_M + (r - \frac{r}{l}) \cdot \kappa_{\Sigma^*} \right] \\
&= \frac{r-l}{r-1} \cdot \frac{1}{l} \cdot \kappa_M + \frac{r(l-1)}{l(r-1)} \cdot \kappa_{\Sigma^*} \\
&\approx \frac{1}{l} \cdot \kappa_M + \frac{l-1}{l} \cdot \kappa_{\Sigma^*},
\end{aligned}$$

denn $\frac{r}{l} - 1$ Summanden streuen um κ_M, die übrigen $r - \frac{r}{l}$ um κ_{Σ^*}.

Im Fall $l \not| r$ folgt ebenso:

$$\begin{aligned}
\varphi(c) &\approx \frac{1}{r-1} \cdot \left[\frac{r-1}{l} \cdot \frac{q \cdot \kappa_{\Sigma^*} + (r-q) \cdot \kappa_M}{r} \right. \\
&\quad \left. + \frac{r-1}{l} \cdot \frac{q \cdot \kappa_M + (r-q) \cdot \kappa_{\Sigma^*}}{r} + (r-1) \cdot (1 - \frac{2}{l}) \cdot \kappa_{\Sigma^*} \right] \\
&= \frac{1}{l} \cdot \frac{r \cdot \kappa_{\Sigma^*} + r \cdot \kappa_M}{r} + (1 - \frac{2}{l}) \cdot \kappa_{\Sigma^*} \\
&= \frac{1}{l} \cdot \kappa_M + \frac{l-1}{l} \cdot \kappa_{\Sigma^*},
\end{aligned}$$

also in beiden Fällen die gleiche Näherungsformel.

Im Beispiel $M =$ „Deutsch“ erwartet man also

$$\varphi(c) \approx \frac{1}{7} \cdot 0.0762 + \frac{6}{7} \cdot 0.0385 \approx 0.0439,$$

und das ist fast genau der Wert aus aus dem früheren Beispiel. Allgemein geben die folgende Tabelle und Grafik den Zusammenhang zwischen Periode und erwartetem Koinzidenzindex für einen polyalphabetisch verschlüsselten deutschen Text:

Periode	1	2	3	4	5
Koinzidenzindex	0.0762	0.0574	0.0511	0.0479	0.0460
	6	7	8	9	10
	0.0448	0.0439	0.0432	0.0427	0.0423
Periode	10	20	30	40	50
Koinzidenzindex	0.0423	0.0404	0.0398	0.0394	0.0393
	60	70	80	90	100
	0.0391	0.0390	0.0390	0.0389	0.0389

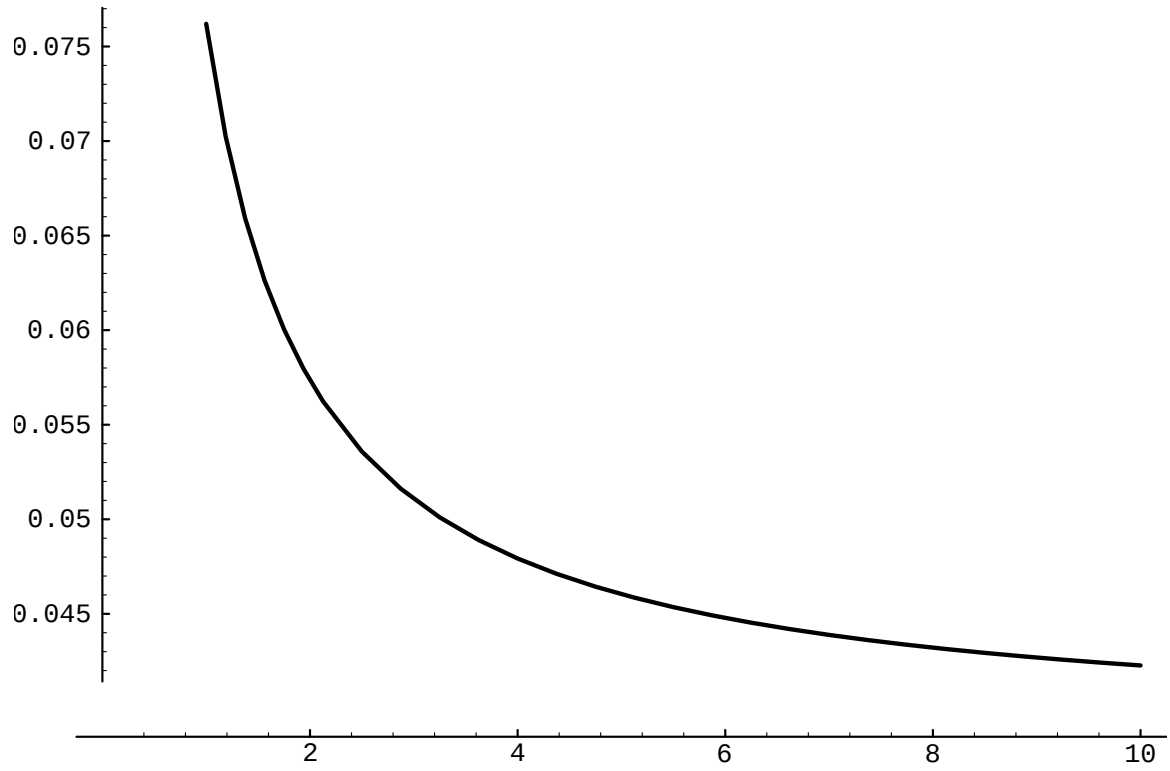

Abbildung 1: Koinzidenzindex und Periodenlänge

Anwendung: Ein (langer) Geheimtext c habe den Koinzidenzindex $\varphi(c) \approx 0.0762$. Dann lässt sich vermuten, dass c ein monolphabetisch verschlüsselter deutscher Text ist.

Anmerkung. Daraus lässt sich ein statistischer Test für die Alternativhypothese „monolphabetisch verschlüsselter deutscher Text“ gegen die Nullhypothese „zufällige Zeichenfolge“ herleiten.

Löst man die obige Formel nach der Periodenlänge l auf, so erhält man die **Näherungsformel von** SINKOV:

$$l \cdot \varphi(c) \approx \kappa_m + (l-1) \cdot \kappa_{\Sigma^*},$$

$$l \cdot [\varphi(c) - \kappa_{\Sigma^*}] \approx \kappa_M - \kappa_{\Sigma^*},$$

$$l \approx \frac{\kappa_M - \kappa_{\Sigma^*}}{\varphi(c) - \kappa_{\Sigma^*}}.$$

Anmerkung. Es gibt „genauere“ Formeln, die aber wegen der Unschärfe in $\varphi(c)$ auch keine besseren Ergebnisse liefern.

Für die als Beispiel durchgeführte Kryptoanalyse ergibt sich

$$l \approx \frac{0.0762 - 0.0385}{0.0440 - 0,0385} \approx 6.85;$$

auch hiermit würden wir auf die Periodenlänge 7 schließen.

Das Problem der Formel von SINKOV ist die mangelnde numerische Stabilität: Je größer die Periode, desto näher ist, wie die Tabelle zeigt, der Koinzidenzindex am Wert für Zufallstexte, der Nenner der Formel also an 0.

[1] Klaus Pommerening, Kryptologie; 27. November 1999, letzte Änderung: 26. Mai 2002