12 Sinkov's Formula

Let's apply the approximative formulas for $\kappa_q(c)$ from Section 9 to the coincidence index of a periodically polyalphabetically encrypted text $c = f(a)$ with $a \in M$ of length r. In the case $l|r$ we get:

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

since $\frac{r}{l} - 1$ summands scatter around κ_M, the other $r - \frac{r}{l}$ ones around κ_{Σ^*}.

In the same way for $l \nmid r$ we get:

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

that is the same approximative formula in both cases. Note that this is a weighted mean.

$$\boxed{\varphi(c) \approx \frac{1}{l} \cdot \kappa_M + \frac{l-1}{l} \cdot \kappa_{\Sigma^*}}$$

For the example $M =$ "German" and $l = 7$ we therefore expect

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

and this is in accordance with the empirical value from the former example. In general Table 30 and Figure 16 show the connection between period and expected coincidence index for a polyalphabetically encrypted German text. The situation for English is even worse.

If we solve the above formula for the period length l, we get Sinkov's formula:

$$\begin{aligned}
l \cdot \varphi(c) &\approx \kappa_M + (l-1) \cdot \kappa_{\Sigma^*}, \\
l \cdot [\varphi(c) - \kappa_{\Sigma^*}] &\approx \kappa_M - \kappa_{\Sigma^*},
\end{aligned}$$

Table 30: *Coincidence index and period length (for German)*

period	1	2	3	4	5
Coinc. index	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
period	10	20	30	40	50
Coinc index	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

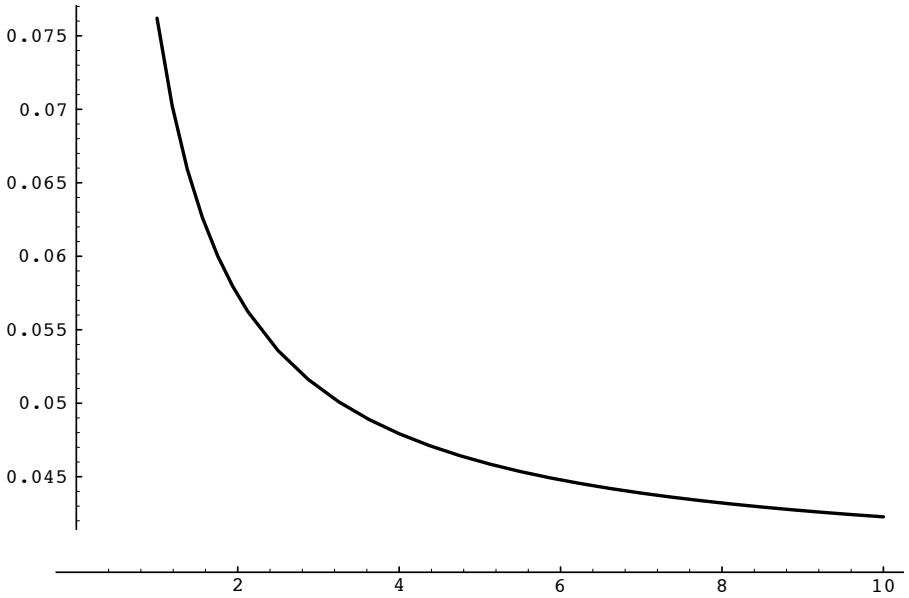

Figure 16: *Coincidence index and period length (for German)*

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

Remark. There are "more exact" versions of this formula. But these don't give better results due to the variation of $\varphi(c)$ and the numerical instability of the small denominator.

For our sample cryptanalysis we get

$$l \approx \frac{0.0762 - 0.0385}{0.0440 - 0.0385} \approx 6.85.$$

This is also evidence for 7 being the length of the period.

The problem with SINKOV's formula is the lack of numerical stability: the larger the period, the closer the coincidence index is to the value for random texts, as the table shows, that is, the closer the denominator in the formula is to 0.

Therefore the autocoincidence spectrum usually yields a better guess of the period. In fact SINKOV himself in his book [8] uses "his" formula—or rather the English equivalents of Table 30 and Figure 16—only for distinguishing between monoalphabetic and polyalphabetic ciphertexts. For determining the period he gives a very powerful test, see Section 13.