# A More Efficient Method of Breaking a Vigenère Cipher

D. R. Stinson

The cryptanalysis of a Vigenère Cipher has two phases: first, the keyword length, $m$ is determined; and then the keyword $(k_1, \ldots, k_m)$ is found. The second step is accomplished in the textbook by using the idea of mutual index of coincidence of two strings to compute the relative shifts of the $m$ substrings of ciphertext, $\mathbf{y}_1, \ldots, \mathbf{y}_m$. It was pointed out to me by Dan Velleman that mutual index of coincidence can be used in a more direct way to determine the keyword. We describe this method now.

Suppose that $\mathbf{z}$ is a string of typical English text. Suppose that $\mathbf{z}$ has length $n$ and frequency distribution $f_0, \ldots, f_{25}$. Then the probability distribution of the 26 letters in $\mathbf{z}$ is $f_0/n, \ldots, f_{25}/n$. We expect that the values $f_i/n$ $(0 \leq i \leq 25)$ will be "close to" the values $p_i$ that were tabulated in Table 1.1 of the textbook, if $n$ is reasonably large.

Now suppose that $\mathbf{y}$ is a string having length $n'$ and frequency distribution ${f'}_0, \ldots, {f'}_{25}$. Then the mutual index of coincidence of $\mathbf{z}$ and $\mathbf{y}$ is

$$MI_c(\mathbf{z}, \mathbf{y}) = \frac{\displaystyle\sum_{i=0}^{25} f_i {f'}_i}{nn'} \approx \frac{\displaystyle\sum_{i=0}^{25} p_i {f'}_i}{n'}.$$

Let the $m$ substrings of ciphertext be denoted $\mathbf{y}_1, \ldots, \mathbf{y}_m$, and define the shifted strings $\mathbf{y}_j^g$ $(1 \leq j \leq m,\ 0 \leq g \leq 25)$ as was done in the textbook.

If $\mathbf{y}_j$ has frequency distribution ${f'}_0, \ldots, {f'}_{25}$ and length $n$, then

$$MI_c(\mathbf{z}, \mathbf{y}_j^g) \approx \frac{\displaystyle\sum_{i=0}^{25} p_i {f'}_{i-g}}{n'}.$$

We expect that

$$MI_c(\mathbf{z}, \mathbf{y}_j^g) \approx .065$$

if $g = -k_j \bmod 26$, and

$$MI_c(\mathbf{z}, \mathbf{y}_j^g) < .045,$$

otherwise.

This suggests the following algorithm to determine the value of $k_j$, which can be applied for $1 \leq j \leq m$.

1. Suppose that $\mathbf{y}_j$ has frequency distribution ${f'}_0, \ldots, {f'}_{25}$ and length $n$. Let $p_0, \ldots, p_{25}$ be the probabilities tabulated in Table 1.1 of the textbook. Then, for $0 \leq g \leq 25$, compute
$$M_g = \frac{\sum_{i=0}^{25} p_i {f'}_{i-g}}{n'}.$$
2. Find $h$ such that
$$M_h = \max\{M_g : 0 \leq g \leq 25\}.$$
3. Set
$$k_j = -h \bmod 26.$$

Let's try this method on the ciphertext in Example 1.11, where the keyword length is $m = 5$. For $1 \leq j \leq 5$, we tabulate in Table 1 the values $M_0, \ldots, M_{25}$ that are computed. The boxed values indicate the maximum $M_g$ values, for $1 \leq j \leq 5$.

From the data in Table 1, the following is obtained:

- For $j = 1$, we find that $h = 17$, $M_{17} = 0.06084$, and threfore $k_1 = 9$.
- For $j = 2$, we find that $h = 0$, $M_0 = 0.06881$, and threfore $k_1 = 0$.
- For $j = 3$, we find that $h = 13$, $M_{13} = 0.06551$, and threfore $k_1 = 13$.
- For $j = 4$, we find that $h = 22$, $M_{22} = 0.05969$, and threfore $k_1 = 4$.
- For $j = 5$, we find that $h = 7$, $M_7 = 0.07244$, and threfore $k_1 = 19$.

This yields the correct keyword, *JANET*.

Table 1: Mutual Indices of Coincidence

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | $j = 1$ | | | |
| 0.03517 | 0.03589 | 0.04162 | 0.04344 | 0.04886 | 0.03333 | 0.03648 |
| 0.04333 | 0.04205 | 0.02998 | 0.03597 | 0.04471 | 0.03771 | 0.03843 |
| 0.03971 | 0.03156 | 0.03903 | $\boxed{0.06084}$ | 0.04790 | 0.02773 | 0.02798 |
| 0.03916 | 0.03524 | 0.03735 | 0.03603 | 0.03148 | | |
| | | | $j = 2$ | | | |
| $\boxed{0.06881}$ | 0.04686 | 0.02622 | 0.03159 | 0.04311 | 0.03363 | 0.03241 |
| 0.03737 | 0.03387 | 0.03198 | 0.03719 | 0.04225 | 0.04581 | 0.04489 |
| 0.03960 | 0.04494 | 0.04163 | 0.03133 | 0.02933 | 0.03275 | 0.03571 |
| 0.03427 | 0.04425 | 0.03489 | 0.03232 | 0.04397 | | |
| | | | $j = 3$ | | | |
| 0.04797 | 0.03981 | 0.04589 | 0.03519 | 0.03649 | 0.03433 | 0.03438 |
| 0.03467 | 0.02740 | 0.04451 | 0.03559 | 0.03790 | 0.03546 | $\boxed{0.06551}$ |
| 0.03460 | 0.03089 | 0.03470 | 0.04879 | 0.03233 | 0.03513 | 0.03768 |
| 0.03394 | 0.04381 | 0.04321 | 0.04224 | 0.02859 | | |
| | | | $j = 4$ | | | |
| 0.04490 | 0.03453 | 0.03784 | 0.04365 | 0.03898 | 0.03398 | 0.03410 |
| 0.05032 | 0.03708 | 0.04424 | 0.04032 | 0.03550 | 0.02852 | 0.03295 |
| 0.04024 | 0.04255 | 0.03319 | 0.03315 | 0.04989 | 0.03439 | 0.03419 |
| 0.03427 | $\boxed{0.05969}$ | 0.03782 | 0.03260 | 0.03210 | | |
| | | | $j = 5$ | | | |
| 0.03403 | 0.03713 | 0.03555 | 0.04811 | 0.03073 | 0.02710 | 0.03673 |
| $\boxed{0.07244}$ | 0.04431 | 0.02919 | 0.03166 | 0.04545 | 0.03603 | 0.03706 |
| 0.03555 | 0.03176 | 0.03318 | 0.03708 | 0.04210 | 0.03847 | 0.04263 |
| 0.03697 | 0.04737 | 0.04424 | 0.03518 | 0.03097 | | |