Unit 39
Finding the period: Sinkov method

A ciphertext that has a period *m* is like mixing *m* different ciphertexts, each enciphered with a different key. As a result, the chance of picking two characters randomly and obtaining identical letters is reduced, as compared to an unencrypted English text. Without proving it, we state that the measured index of coincidence of the entire ciphertext is related to the period in this way:

$$\text{IoC} = \text{IoC}_{\text{random}} + (\text{IoC}_{\text{English}} - \text{IoC}_{\text{random}}) / m$$

Solving for the period and using $\text{IoC}_{\text{random}} = 1$ and $\text{IoC}_{\text{English}} = 1.75$ (in the normalization that we use), we obtain

$$m = \frac{0.75}{\text{IoC} - 1}$$

In tabular form:

IoC	period
1.75	1 (monoalphabetic)
1.38	2
1.25	3
1.19	4
1.15	5
1.13	6

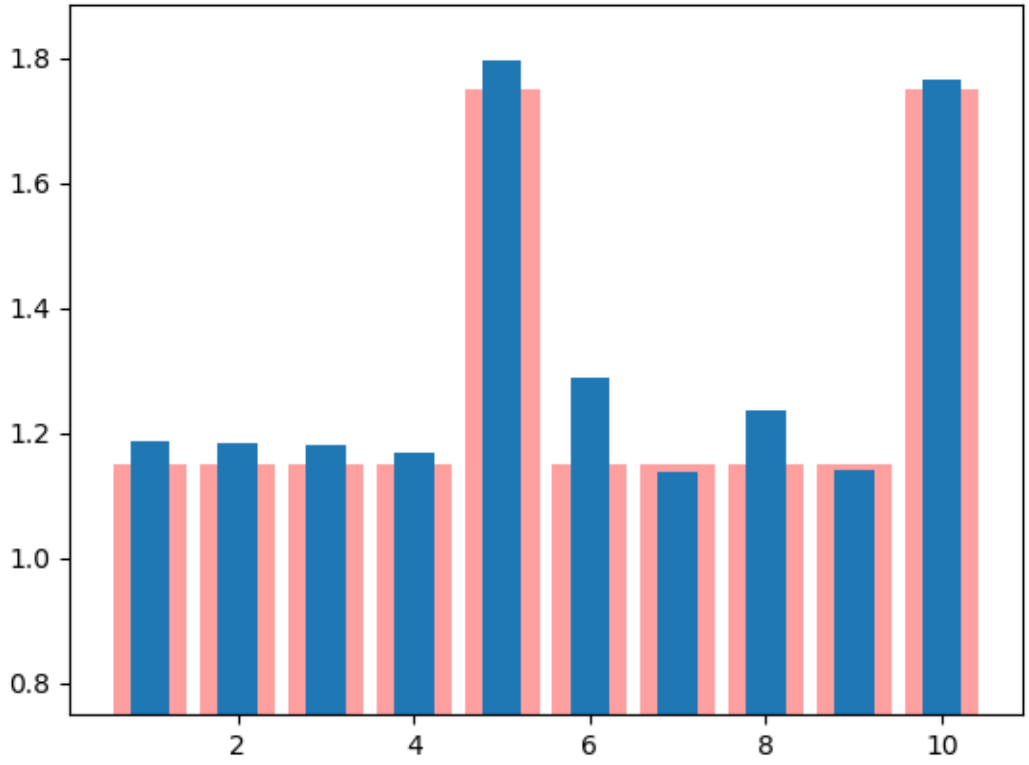
As we can see, as the period increases, the values of the IoC get closer together. Since these numbers are approximate and there is a lot of variability in the IoC, *we should not rely on this method for finding the period*.

A better way to use the IoC to find the period is to guess a value for the period *m* and then slice the ciphertext into *m* slices and find the IoC of the slices. The n^{th} slice is composed of every m^{th} character, starting with the n^{th} character. Partitioning the text this way gives us slices that would each be encrypted with one key alphabet, if we guessed the period correctly. The average of the IoCs of the slices serves as a good measure of the IoC of the plaintext that you would obtain by deciphering the ciphertext with the guessed period. If this IoC is not close to the IoC of typical English text, then the guess is a bad one. By guessing periods until the IoC calculated in this way is close to that of English, we can find the period, and we are usually correct.

Let’s look at this ciphertext, for example:

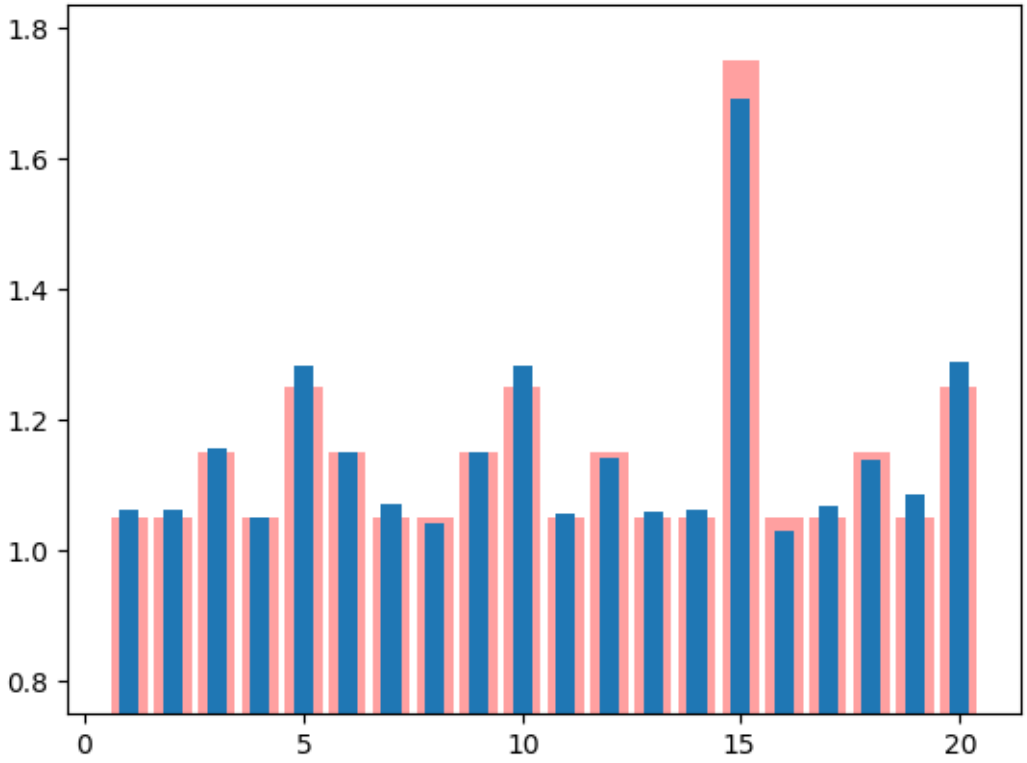
```
BUHMKLRASCKBLRZQQHRZMVVZBZLXWBNHOMKKEBTQWTUEMPLWLBQAGI
UWSFSFVPLHBVPHGXVOHYPMQWSCQAGXEMCHVFWQRJXXRUMLLVFTLTLD
PMPNEIQQPVYYAGLXRBVRRZQCKIOBUHFBAGZEVBBXWBBUHMKLRASCKB
LRZQQHRZMGRJFVQWLBXRUMLLVVXLKHWXEMPLTEMEWIUBVQXLAYLGBA
NQHXDRUEDMGKIFWPRJQPRVPFKRVMCBUHESMEDKBQBFMPKYRWBBBWLX
BBIIKOYLWEBUHRQPRQYJJRUSCAYLGBAVVPFSROCQWOHXEMCHVFWQ
```

And here is a graph of the IoC averaged over the slices, for periods one through ten. The measured IoC is in blue and the theoretical values for a period of five is in pink. The peak at ten is due to the fact that when we take ten slices of the ciphertext, then each slice is half of one of the slices that we took for period five. We conclude that the period of the cipher is five.



But the astute reader will notice that five is a prime number. What happens if the period is not prime? Below is a similar graph for a text with a period of 15. Again, the measured IoC is in blue and the theoretical values are in pink. The secondary peaks at periods 5 and 10 and the tertiary peaks at 3, 6, 9, and 12 are due to the fact that when we take a number of slices that shares a factor with the true period, then each slice contains some letters that were enciphered with the same key alphabet, i.e., that are in the same “true” slice. The expected theoretical value at period *n* when the true period is *m* is

$$IoC_{theory} = IoC_{random} + (IoC_{English} - IoC_{random}) \cdot \gcd(n, m) / m$$



Reading and references

William F. Friedman, The Index of Coincidence and Its Applications in Cryptography, Riverbank Laboratories Department of Ciphers Publication 22, Geneva, Illinois, 1920, www.marshallfoundation.org/library/methods-solution-ciphers

William F. Friedman and Lambros D. Callimahos, Military cryptanalytics, Part I, Volume 2, Aegean Park Press, 1956, reprinted 1985.

M. Mountjoy (1963) The bar statistics, NSA Technical Journal VII (2, 4).

Abraham Sinkov, *Elementary Cryptanalysis: A Mathematical Approach*, 2nd edition, revised by Todd Feil, published by Mathematical Association of America, 2009; www.jstor.org/stable/10.4169/j.ctt19b9krf; section 3.3.

Practical Cryptography: practicalcryptography.com/cryptanalysis/stochastic-searching/cryptanalysis-vigenere-cipher

David Kahn, *The Codebreakers: The Story of Secret Writing*, New York: Simon & Schuster, 1967, revised and updated 1996, pages 376-380.

Programming tasks

1. Write a function that cuts a ciphertext into a number of slices as described above. If the number of slices is *n*, then the first slice contains every n^{th} letter, starting with the first letter of the text; the second slice contains every n^{th} letter, starting with the second letter of the text; etc.

2. Write a function to find the period of a ciphertext with the method described above. You will need to decide on an appropriate cut-off, which you can base on the results of Exercise 1 in Unit 12.

Exercises

1. Use your function to find the period of the example ciphertexts from this unit and the previous unit.