Cryptanalyse des chiffrements par substitution
Master DID, I131

Les deux méthodes classiques sont :

— celle qui consiste à calculer l'indice de coïncidence, proposée par Wolfe Friedmann (1881-1869, né en Moldavie et cryptologue de l'armée américaine) ;

— le test de Kasiski (Friedrich Kasiski, 1805-1881, officier cryptologue, au sein de l'armée prussienne).

Ce qui suit est reproduit de "Cryptographie, théorie et pratique" de D. Stinson ($2^{\text{ème}}$ édition).

1 Indice de Coïncidence - Définition

Soit $x = x_1x_2 \dots x_n$ une chaîne composée de n caractères alphabétiques. L'*indice de coïncidence* est la probabilité que deux caractères aléatoires de x soient identiques.

On note respectivement $f_0, f_1, \dots, f_{25}$ les fréquences absolues de $A, B, \dots, Z$. Pour chaque $i, 0 \leq i \leq 25$, il y a $\binom{f_i}{2}$ [1] façons de choisir deux caractères i parmi les n caractères du texte. Donc, on a

$$Ic(x) = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)} \tag{1}$$

En français, la probabilité d'apparition des lettres est donnée figure 1. Dans ce cas, en notant $p_0, p_1, \dots, p_{25}$ les probabilités d'apparition des lettres $A, B, \dots, Z$, on peut s'attendre à ce que l'indice de coïncidence soit

$$Ic(x) = \sum_{i=0}^{25} p_i^2 = 0,0778.$$

Dans le cas d'une suite de lettre aléatoire, on aurait

$$Ic(x) = 26 \cdot \left(\frac{1}{26}\right)^2 = 0,038.$$

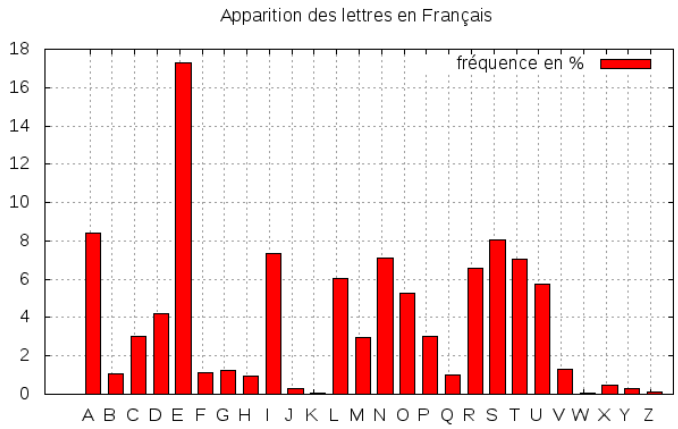
C'est ce biais que l'on exploite.

2 Cas du chiffrement de de Vigenère

Ce test peut être utilisé dans le cas du chiffrement de de Vigenère, à condition de l'adapter à l'hypothèse d'une longueur de clé.

— on découpe le texte chiffré en sous-chaînes correspondant aux suites de lettres chiffrées par le même caractère de la clé. Ceci revient, pour une hypothèse de clé de longueur l, à prendre un caractère tout les l caractères à partir du premier pour constituer la première

1. On rappelle le coefficient binomial $\binom{n}{k} = \frac{n!}{k!(n-k)!}$



A	B	C	D	E	F	G	H	I	J	K	L	M
8.4	1.06	3.03	4.18	17.26	1.12	1.27	0.92	7.34	0.31	0.05	6.01	2.96

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
7.13	5.26	3.01	0.99	6.55	8.08	7.07	5.74	1.32	0.04	0.45	0.3	0.12

FIGURE 1 – Fréquence d'apparition des lettres en français.

sous-chaîne, puis à recommencer à partir du deuxième caractère pour la deuxième sous-chaîne, et à poursuivre ainsi jusqu'à former les l sous-chaînes (la dernière commençant par le $(l-1)^{\text{ème}}$caractère ;
— on calcule l'indice de coïncidence de chacune de ces sous-chaînes et on en fait la moyenne pour avoir l'indice de coïncidence global ;
— lorsque l'hypothèse sur la longueur de la clé est la bonne, l'indice de coïncidence global doit être sensiblement supérieur à la valeur pour une suite de lettre aléatoire $(0,038)$ et se rapprocher de $0,0778$ pour un texte en français.

3 Test de Kasiski

Friedrich Kasiski est un officier prussien (1805-1881), et cryptologue, au sein de l'armée prussienne. Il propose une méthode de cryptanalyse du chiffrement de Vigenère basée sur l'observation suivante :
— deux segments identiques du texte clair sont chiffrés de façon identiques dès qu'ils sont décalés de $x \equiv 0 \bmod m$ (m est la longueur de la clé) ;
— inversement, il en découle que deux segments identiques d'une longueur suffisante (3 suffit souvent) ont de bonnes chances de provenir d'un segment identique du texte clair.

Ceci donne lieu au test suivant :
— on collecte les paires de segments identiques sur le texte chiffré ;
— on note les distances entre leurs premiers caractères ;
— on obtient ainsi une suite de distances : $d_1, d_2, \cdots$, on peut conjecturer que m est le plus grand diviseur commun des valeurs d_i.

Ce test est complémentaire du précédent, et permet de trouver la longueur de la clé.