1 Analyse fréquentielle

Les lettres les plus fréquences en français (les valeurs changent selon le corpus étudié) sont le E (14, 715%), le S (7, 948%), le A (7, 636%) et le I (6, 55%). Dans un texte on considère que chaque lettre apparaît avec une probabilité égale à sa fréquence. On considère un texte de n caractères chiffré par chiffrement de Polybe, chaque lettre est remplacée par une autre. On appelle M le texte clair et C le texte chiffré

1. Soient l_1 et l_2 les lettres codant respectivement E et S dans M. Quelle est la probabilité qu'il y ait d'occurrence de lettres l_2 que l_1 ?

2. Déduire une borne sur la probabilité de l'événement "La lettre ayant le plus grand nombre d'occurrences dans C correspond à la lettre E dans M".

3. A partir de quelle longueur de texte est-il possible d'identifier la lettre qui encode E ?

4. Les lettres S et A ont des fréquences d'apparition proche. Quelle taille de texte faut-il pour différencier les deux lettres par leur nombre d'occurrences ? Comment contourner ce problème en pratique ?

5. On considère maintenant un texte codé par la méthode de Vigenère avec une clef de taille k (on suppose la longueur k connue). A partir de quelle longueur un texte est-il vulnérable à l'analyse fréquentielle ?

2 Registre à décalage à rétroaction linéaire

Code ascii Le code ASCII est une méthode d'encodage des caractères. On considère un encodage sous format ASCII 8 bits. Décodez le message suivant.
01000001 01101101 01100101 01110010 01101001 01100011 01100001 01101110 00100000 01010011 01110100 01100001 01101110 01100100 01100001 01110010 01100100 00100000 01000011 01101111 01100100 01100101 00100000 01100110 01101111 01110010 00100000 01001001 01101110 01100110 01101111 01110010 01101101 01100001 01110100 01101001 01101111 01101110 00100000 01001001 01101110 01110100 01100101 01110010 01100011 01101000 01100001 01101110 01100111 01100101

Chiffrement On considère un registre sur 3 bits. On suppose tout d'abord qu'il est défini par la suite $u_{n+3} = u_{n+2} + u_{n+1} + u_n \mod 2$.

1. Quelle est la période maximale d'un registre sur 3 bits ? Pourquoi est-ce une mauvaise idée d'utiliser un registre avec cette période sur du code ascii ?

2. Quelle est la période maximale de ce registre ?

3. Encodez le message "DANGER" en ascii puis avec le registre sur 4 bits $u_{n+4} = u_{n+3} + u_{n+1} + u_n \mod 2$ avec la clef 0100.

4. Quelle est la période maximal d'un registre sur n bits ? Prouver qu'un tel registre ne contient pas n bits à 0 consécutifs. Que déduisez-vous ?