Exercices
Semaine 7

Cours Turing

1 Chiffre de César et analyse des fréquences

a) Ecrire un programme qui attend en entrée :

- Soit le mot "ENCRYPTER" soit "DECRYPTER" en fonction de ce que le programme doit faire

- une lettre majuscule (celle utilisée comme clé de décalage dans le chiffrement de César)

- une chaîne de caractères (au format str), composée uniquement de lettres majuscules et d'espaces

et dont la sortie soit la chaîne de caractères chiffrée (ou respectivement déchiffrée)(attention que les espaces doivent rester des espaces).

b) Ecrire ensuite un programme qui effectue une analyse des fréquences de la chaîne cryptée pour retrouver le texte d'origine. Pour cela, vous aurez besoin des ingrédients suivants :

- le calcul des fréquences (en %) de chaque lettre du texte crypté, à enregistrer dans une liste `frequences_empiriques`

- la comparaison, pour chaque décalage, avec les fréquences théoriques des lettres dans la langue française :

```
frequences_theoriques = [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, 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]
```

- et la sélection du décalage qui minimise la différence entre les deux listes (au fait, vous avez plusieurs choix pour calculer cette différence : nous vous laissons réfléchir comment faire au mieux...)

2 Chiffre de Vigenère et indice de coïncidence

a) Ecrire un programme qui attend en entrée :

- Soit le mot "ENCRYPTER" soit "DECRYPTER" en fonction de ce que le programme doit faire

- une courte chaîne de caractères (celle utilisée comme clé dans le chiffrement de Vigenère), composée de k caractères.

- une chaîne de caractères (au format str), composée uniquement de lettres majuscules et d'espaces.

et dont la sortie soit la chaîne de caractères chiffrée (attention à nouveau que les espaces doivent rester des espaces).

Note : Vous pouvez également utiliser ici le texte de l'exercice 1. Il est également conseillé de réutiliser le code écrit dans l'exercice 1.

b) Le décryptage du chiffre de Vigenère est plus délicat, notamment parce qu'a priori, on ne connaît pas le nombre de lettres k utilisé pour chiffrer le message.

La première étape consiste donc à estimer ce nombre de lettres k de la clé de Vigenère en utilisant un *indice de coïncidence*. Pour un texte donné, l'indice de coïncidence du texte est défini ainsi : si $n_A, \ldots, n_Z$ sont les nombres de lettres $A, \ldots, Z$ apparaissant dans le texte et n est le nombre total de lettres du texte (on oublie ici les espaces), alors

$$IC = \sum_{i=A}^{Z} \frac{n_i(n_i - 1)}{n(n-1)}$$

A quoi peut bien servir un tel indice ? Pour un texte sans signification composé de lettres tirées complètement au hasard, on peut montrer que IC vaut à peu près 0,0385. Tandis qu'en français, les lettres sont tout sauf tirées au hasard, et comme on l'a déjà vu, certaines lettres sont bien plus fréquentes que d'autres. Le calcul de IC pour un texte français (avec les fréquences théoriques de l'exercice précédent) donne un nombre bien plus grand, à savoir 0,0746.

Une autre remarque très importante est que l'indice de coïncidence d'un texte chiffré avec un chiffre de César est le même que celui du texte original ! En effet, seules comptes les fréquences d'apparition des lettres dans le calcul de IC ; peu importe la valeur de ces lettres.

La stratégie de décryptage du chiffre de Vigenère est maintenant la suivante : tous les k caractères, le chiffre de Vigenère utilise le même décalage pour chiffrer une lettre. Ainsi, l'indice de coïncidence du sous-texte composé des lettres numéro $1, k+1, 2k+1, 3k+1, \ldots$ sera indentique à celui du français. Idem pour le sous-texte composé des lettres numéro $2, k+2, 2k+2, 3k+2, \ldots$ et ainsi de suite jusqu'au k^e sous-texte composé des lettres numéro $k, 2k, 3k, \ldots$

En calculant, pour chaque valeur de k, l'indice de coïncidence de chacun des k sous-textes, et en faisant la moyenne de ces indices, on obtient une bonne indication de la longueur k de la clé : c'est celle qui donne le plus grand indice de coïncidence moyen. En effet, si k n'est pas la

bonne valeur, chacun des sous-textes ressemblera à un texte écrit au hasard.

Une fois la longueur de la clé k trouvée, il ne reste "qu'à" effectuer l'analyse de l'exercice 1 pour retrouver chacun des k décalages utilisés, et ainsi retrouver le texte d'origine.