DS 6 Mathématiques

Problème d'informatique

On s'intéresse à déterminer la plus longue sous-chaine commune à deux séquences ADN. C'est un indicateur simple de la proximité de deux espèces lors d'une étude phylogénétique.
Nos chaines de caractères sont donc formées avec des lettres uniquement prises parmi `A`, `C`, `G`, `T`. Pour une telle chaine de longueur n, on écrit $S = a_0 a_1 \dots a_{n-1}$. Une sous-chaine de longueur k est une chaine du type $a_{i_0} a_{i_1} \dots a_{i_{k-1}}$ où les indices vérifient $0 \leqslant i_0 < i_1 < \dots < i_{k-1} \leqslant n-1$. Cela signifie que les caractères de la sous-chaine se retrouvent dans le même ordre dans S, **mais ne sont pas nécessairement consécutifs**. Si T est une autre chaine, de longueur p et notée $T = b_0 b_1 \dots b_{p-1}$, alors une sous-chaine commune est une chaine qui est à la fois une sous-chaine de S et une sous-chaine de T.
Par exemple, pour les chaines `S = "ATAGA"` et `T = "TAACA"`, alors `"AAA"` est une sous-chaine commune, et `"TAA"` est aussi une sous-chaine commune.
Pour les chaines `S = "AATGCG` et `T = "TATTAGC"`, la plus longue sous-chaine commune est `"ATGC"`.
Le but est de déterminer la longueur d'une plus longue sous-chaine commune.

1. Étude d'exemples

1. Quelle est la longueur d'une plus longue sous-chaine commune à `"AATGCG"` et à `"TATTAGC"` ? Justifier.
2. Déterminer une plus longue sous-chaine commune aux chaines `"AATGCG"` et `"TATTAGC"` qui soit autre que `"ATGC"`.
3. Déterminer toutes les plus longues sous-chaines communes aux chaines `"TCGTA"` et `"CTG"`.

2. Fonctions Python préliminaires

On rappelle que les chaines de caractères en Python se manipulent à de nombreux égards comme les listes ; pour un telle chaine `S`, `len(S)` donne la longueur et `S[i]` le i-ème caractère ; la chaine vide est `""`.

4. Écrire une fonction `maximum(p, q)` qui renvoie le maximum des deux nombres `p` et `q`.
5. Écrire une fonction `est_ADN(S)` qui prend en argument une chaine de caractères, et renvoie `True` si `S` est effectivement bien composée uniquement des caractères parmi `"A"`, `"C"`, `"G"`, `"T"`, et `False` sinon.
6. Écrire une fonction `compte(S)` qui prend en argument une chaine de caractères (supposée être une séquence ADN valide) et qui renvoie un dictionnaire de 4 clés qui sont exactement les lettres `"A"`, `"C"`, `"G"`, `"T"`, et qui donne combien de fois chaque lettre apparait dans `S`.
7. Il y a une description récursive assez simple de la condition d'être une sous-chaine : la chaine $C = c_0 c_1 \dots c_{k-1}$ est une sous-chaine de $S = a_0 a_1 \dots a_{n-1}$ si et seulement si C est vide, ou bien, il existe un indice i pour lequel $c_0 = a_i$ et alors $c_1 \dots c_{k-1}$ est une sous-chaine de $a_{i+1} \dots a_{n-1}$.
   Écrire une fonction récursive `est_sous_chaine(C, S)` qui renvoie `True` si la chaine `C` est une sous-chaine de `S`, et `False` sinon. Dans cette question on pourra utiliser les tranches de chaines : `S[u:v]` est la chaine formée des éléments (consécutifs) d'indice i tel que $u \leqslant i < v$.

3. Résolution par programmation dynamique

Le problème semble compliqué à résoudre autrement qu'en testant toutes les possibilités de sous-chaines, et la fonction pour tester la condition de sous-chaine est elle-même compliquée. Pour résoudre le problème, on calcule *plus* d'information que nécessaire, mais cela est malgré tout plus efficace !
Reprenant deux chaines $S = a_0 a_1 \dots a_{n-1}$ et $T = b_0 b_1 \dots b_{p-1}$ de longueurs respectives n et p, on note $\ell_{i,j}$ la longueur d'une plus longue sous-chaine commune à $a_0 a_1 \dots a_{i-1}$ (ceci est une chaine vide si $i = 0$) et à $b_0 b_1 \dots b_{j-1}$ (vide si $j = 0$). Ainsi on a un tableau $(\ell_{i,j})_{(i,j) \in [\![0,n]\!] \times [\![0,p]\!]}$ dont le coefficient $\ell_{n,p}$ est la réponse à notre problème.

8. Justifier que (à défaut, on pourra admettre ces relations et passer à la suite)

$$\forall i \in [\![0, n]\!], \quad \ell_{i,0} = 0 \tag{1}$$

$$\forall j \in [\![0, p]\!], \quad \ell_{0,j} = 0 \tag{2}$$

$$\forall (i,j) \in [\![1, n]\!] \times [\![1, p]\!], \quad \ell_{i,j} = \begin{cases} 1 + \ell_{i-1,j-1} & \text{si } a_{i-1} = b_{j-1} \\ \mathrm{Max}(\ell_{i-1,j}, \ell_{i,j-1}) & \text{sinon} \end{cases} \tag{3}$$

9. Remplir sur sa copie un tableau donnant les coefficients $\ell_{i,j}$ pour les chaines `S = "CTG"` et `T = "TCGT"`.

10. Écrire une fonction Python `longueur_plus_longue_sous_chaine(S, T)` qui prend en argument deux chaines de caractères, supposées représenter des séquences ADN valides, et qui renvoie la longueur de la plus longue sous-chaine commune, en implémentant la méthode ici décrite. On rangera les valeurs des $\ell_{i,j}$ dans un tableau `L` qu'on représentera comme une liste de listes, en donnant la liste de ses lignes.
11. La méthode décrite nous donne bien la longueur de la plus longue sous-chaine commune, mais ne nous dit pas *quelle* est cette sous-chaine. Comment pourriez-vous modifier la fonction précédente pour calculer en même temps la plus longue sous-chaine ?

Problème 1

On s'intéresse au système de deux équations différentielles suivantes, portant sur deux fonctions $x(t)$ et $y(t)$ définies et dérivables sur $\mathbb{R}$:

$$\forall t \in \mathbb{R}, \quad \begin{cases} x'(t) = x(t) + 2y(t) + 5t^2 \\ y'(t) = 3y(t) - x(t) + \cos(t) \end{cases} \tag{4}$$

1. Justifier que x vérifie une équation différentielle (E) de la forme

$$\forall t \in \mathbb{R}, \quad x''(t) - 4x'(t) + 5x(t) = P(t) + 2\cos(t) \tag{E}$$

   où P est une fonction polynomiale à déterminer.
2. Donner toutes les solutions de l'équation homogène associée à (E).
3. Chercher une solution particulière à l'équation (E_1)

$$\forall t \in \mathbb{R}, \quad x''(t) - 4x'(t) + 5x(t) = P(t) \tag{E_1}$$

   sous la forme d'un polynôme.
4. Chercher une solution particulière à l'équation (E_2)

$$\forall t \in \mathbb{R}, \quad x''(t) - 4x'(t) + 5x(t) = 2\cos(t) \tag{E_2}$$

   sous forme $x(t) = A\cos(t) + B\sin(t)$, où $(A, B) \in \mathbb{R}^2$.
5. En déduire toutes les fonctions x solutions, puis y.

Problème 2

On pose $A_n = \displaystyle\sum_{k=1}^{n} \frac{1}{k^2}$ pour $n \in \mathbb{N}^*$.

1. Soit $k \geqslant 1$. Justifier que $\forall x \in [k, k+1]$, $\dfrac{1}{(k+1)^2} \leqslant \dfrac{1}{x^2}$ puis que $\dfrac{1}{(k+1)^2} \leqslant \displaystyle\int_k^{k+1} \frac{\mathrm{d}x}{x^2}$.
2. En déduire que $A_n \leqslant 2 - \frac{1}{n}$.
3. Conclure que la suite $(A_n)_{n\in\mathbb{N}^*}$ converge, vers une limite strictement positive.

On note $\alpha = \displaystyle\lim_{n\to+\infty} \sum_{k=1}^{n} \frac{1}{k^2}$. Pour tout couple d'entiers (n, p) tels que $n \geqslant 1$ et $p \geqslant 1$, on pose

$$S_{n,p} = \sum_{j=1}^{p} \sum_{i=1}^{n} u_{i,j} \quad \text{où} \quad u_{i,j} = \begin{cases} \dfrac{1}{i^2 - j^2} & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases} \tag{5}$$

Nous allons montrer que

$$\lim_{n\to+\infty} \left(\lim_{p\to+\infty} S_{n,p} \right) \neq \lim_{p\to+\infty} \left(\lim_{n\to+\infty} S_{n,p} \right) \tag{6}$$

On prendra particulièrement garde aux variables utilisées, lesquelles sont fixées ou bien liées à la somme...

4. En utilisant un encadrement du terme sous la somme, montrer que $\forall j \geqslant 1$, $\displaystyle\lim_{n\to+\infty} \sum_{k=n-j+1}^{n+j} \frac{1}{k} = 0$.

5. On fixe un entier $j \geqslant 1$.
   (a) Montrer que $2ju_{i,j} = \frac{1}{i-j} - \frac{1}{i+j}$ pour tout $i \neq j$.
   (b) En déduire que pour tout entier $n > 2j$

$$2j \sum_{i=1}^{n} u_{i,j} = \sum_{i=1}^{j-1} \frac{1}{i-j} - \sum_{i=1}^{j-1} \frac{1}{i+j} + \sum_{i=j+1}^{n} \frac{1}{i-j} - \sum_{i=j+1}^{n} \frac{1}{i+j} \tag{7}$$

   (c) À l'aide de changements d'indices, exprimer les 4 sommes ci-dessus comme des sommes de termes $\sum_k \frac{1}{k}$ (avec des bornes à déterminer).
   (d) En déduire $2j \sum_{i=1}^{n} u_{i,j} = \frac{3}{2j} - \sum_{k=n-j+1}^{n+j} \frac{1}{k}$
6. À l'aide des résultats précédents, exprimer la limite de $S_{n,p}$ quand $n \to +\infty$ en fonction de A_p pour tout entier $p \geqslant 1$ puis en déduire que

$$\lim_{p\to+\infty} \left(\lim_{n\to+\infty} S_{n,p} \right) = \frac{3}{4}\alpha \tag{8}$$

7. En remarquant que $u_{i,j} = -u_{j,i}$ pour tout $(i,j) \in \mathbb{N}^2$ et en utilisant le résultat précédent, exprimer $\lim_{n\to+\infty} (\lim_{p\to+\infty} S_{n,p})$ en fonction de α et conclure.
8. Que vaut $\lim_{n\to+\infty} S_{n,n}$?

Exercice

Les questions de cet exercice sont indépendantes. On se place dans l'espace $\mathbb{R}^3$.

1. Soient $\mathcal{D}$ la droite et $\mathcal{P}$ le plan de l'espace définis par les équations

$$\mathcal{D} : \begin{cases} x+y+z=1 \\ x-2y-z=0 \end{cases} \qquad \mathcal{P} : x+3y+2z=6 \tag{9}$$

   (a) Donner un vecteur directeur de $\mathcal{D}$ et deux vecteurs directeurs de $\mathcal{P}$.
   (b) La droite $\mathcal{D}$ est-elle orthogonale au plan $\mathcal{P}$?
   (c) Montrer que $\mathcal{D} \cap \mathcal{P}$ est un unique point et donner ce point.
2. Soient les points $A = (1,0,2)$ et $B = (0,1,0)$. Montrer que l'ensemble des points M tels que $AM = BM$ est un plan (*plan médiateur*).
3. Soit $m \in \mathbb{R}$. On considère les 3 plans suivants :

$$\mathcal{P}_m : mx+y+z=1 \tag{10}$$
$$\mathcal{Q}_m : x+my+z=m \tag{11}$$
$$\mathcal{R}_m : x+y+mz=m^2 \tag{12}$$

   Discuter selon m de la nature (vide, point, droite, ...) de l'intersection $\mathcal{P}_m \cap \mathcal{Q}_m \cap \mathcal{R}_m$.