

NOM :  
PRENOM :

Couige

Date :  
Groupe :

**Mathématiques Appliquées à la Biologie : Feuille-réponses du TD 3**  
Initiation au calcul matriciel

On répondra aux questions posées aussi clairement que possible dans les espaces prévus et on remettra cette *feuille-réponses* en fin de séance à l'enseignant chargé du Cours/TD.

Exercice 1. :

1. Soit une chaîne de Markov à 2 états de matrice de transition  $\mathbb{P}$  telle que  $p_{11} = \frac{3}{4}$  et  $p_{22} = \frac{1}{2}$ . Calculer l'image  $\pi_1$  par la chaîne de Markov de la distribution initiale  $\pi_0 = (0 ; 1)$ .

On déduit du fait que toutes les lignes de  $\mathbb{P}$  sont de somme égale à 1 que  $\mathbb{P} = \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{pmatrix}$ .

$$\pi_1 = \pi_0 \mathbb{P} = (0 \quad 1) \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{pmatrix} = \left( 0 \left( \frac{3}{4} \right) + 1 \left( \frac{1}{2} \right) \quad 0 \left( \frac{1}{4} \right) + 1 \left( \frac{1}{2} \right) \right) = \left( \frac{1}{2} \quad \frac{1}{2} \right)$$

2. Vérifier que  $\pi_0^* = \left( \frac{2}{3} ; \frac{1}{3} \right)$  est une distribution stationnaire pour cette chaîne de Markov.

Il faut vérifier que l'on a bien  $\pi_0^* \mathbb{P} = \pi_0^*$ .

$$\begin{aligned} \pi_0^* \mathbb{P} &= \left( \frac{2}{3} \quad \frac{1}{3} \right) \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{pmatrix} = \left( \frac{2}{3} \left( \frac{3}{4} \right) + \frac{1}{3} \left( \frac{1}{2} \right) \quad \frac{2}{3} \left( \frac{1}{4} \right) + \frac{1}{3} \left( \frac{1}{2} \right) \right) \\ &= \left( \frac{1}{2} + \frac{1}{6} \quad \frac{1}{6} + \frac{1}{6} \right) = \left( \frac{4}{6} \quad \frac{2}{6} \right) = \left( \frac{2}{3} \quad \frac{1}{3} \right). \end{aligned}$$

3. En déduire que  $\lambda = 1$  est une valeur propre à gauche de la matrice  $\mathbb{P}$ ; expliquez.

$\lambda = 1$  est une valeur propre à gauche de  $\mathbb{P}$  s'il existe un vecteur  $V$  tel que  $V \mathbb{P} = \lambda V = V$ . Ici, si l'on prend  $V = \pi_0^*$ , on a bien  $\pi_0^* \mathbb{P} = \lambda \pi_0^* = \pi_0^*$  pour  $\lambda = 1$ .

4. Calculer le carré  $\mathbb{P}^2$ . En déduire l'image  $\pi_2$  de la distribution initiale  $\pi_0 = (0 ; 1)$  après deux étapes.

$$\begin{aligned} \mathbb{P}^2 = \mathbb{P} \cdot \mathbb{P} &= \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} \frac{3}{4} \left( \frac{3}{4} \right) + \frac{1}{4} \left( \frac{1}{2} \right) & \frac{3}{4} \left( \frac{1}{4} \right) + \frac{1}{4} \left( \frac{1}{2} \right) \\ \frac{1}{2} \left( \frac{3}{4} \right) + \frac{1}{2} \left( \frac{1}{2} \right) & \frac{1}{2} \left( \frac{1}{4} \right) + \frac{1}{2} \left( \frac{1}{2} \right) \end{pmatrix} \\ &= \begin{pmatrix} \frac{11}{16} & \frac{5}{16} \\ \frac{5}{8} & \frac{3}{8} \end{pmatrix} = \frac{1}{8} \begin{pmatrix} 11/2 & 5/2 \\ 5 & 3 \end{pmatrix} \end{aligned}$$

$$\pi_2 = \pi_0 \mathbb{P}^2 = (0 \quad 1) \begin{pmatrix} \frac{11}{16} & \frac{5}{16} \\ \frac{5}{8} & \frac{3}{8} \end{pmatrix} = \left( \frac{5}{8} \quad \frac{3}{8} \right) = \frac{1}{8} \begin{pmatrix} 5 & 3 \end{pmatrix}$$

**Exercice 2.** : Un modèle simpliste<sup>1</sup> d'un brin d'ADN est celui d'un enchaînement "désordonné" de nucléotides de l'un des 4 types adénine (a), cytosine (c), guanine (g) et thymine (t). Le brin est considéré alors comme une trajectoire d'une chaîne de Markov  $X_n$  à quatre états  $S = \{a, c, g, t\}$  dont la matrice de transition  $\mathbb{P}$  fournit les probabilités que l'un de ces états succède à un autre. Ainsi le brin  $aagc$  est la trajectoire  $X_0 = a, X_1 = a, X_2 = g, X_3 = c$ .

1. En supposant que la matrice  $\mathbb{P}$  est donnée par  $\mathbb{P} = \begin{pmatrix} 0 & 0,3 & 0 & 0,7 \\ 0,6 & 0 & 0,4 & 0 \\ 0 & 0,8 & 0 & 0,2 \\ 0,3 & 0 & 0,7 & 0 \end{pmatrix}$  calculer la matrice

$\mathbb{P}^2$  et en déduire la probabilité de trouver dans ce gène une succession de 3 nucléotides  $cxt$  où  $x \in \{a, c, g, t\}$  est l'un quelconque des 4 nucléotides.

$$\mathbb{P}^2 = \mathbb{P} \cdot \mathbb{P} = \begin{pmatrix} 0 & 0,3 & 0 & 0,7 \\ 0,6 & 0 & 0,4 & 0 \\ 0 & 0,8 & 0 & 0,2 \\ 0,3 & 0 & 0,7 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0,3 & 0 & 0,7 \\ 0,6 & 0 & 0,4 & 0 \\ 0 & 0,8 & 0 & 0,2 \\ 0,3 & 0 & 0,7 & 0 \end{pmatrix} = \begin{pmatrix} 0,39 & 0 & 0,61 & 0 \\ 0 & 0,5 & 0 & 0,5 \\ 0,54 & 0 & 0,46 & 0 \\ 0 & 0,65 & 0 & 0,35 \end{pmatrix}$$

La proba de passer de c à t en 2 étapes  $P(X_{t+2}=t/X_t=c) = 0,5$ .

2. La distribution initiale  $\pi_0 = (\frac{1}{2}; 0; \frac{1}{2}; 0)$  est-elle une distribution stationnaire pour cette chaîne de Markov?

Ce serait le cas si  $\pi_0 \mathbb{P} = \pi_0$ , donc en particulier si  $C_1$  est la 1<sup>ère</sup> colonne de  $\mathbb{P}$ , on devrait avoir  $\pi_0 C_1 = \pi_0(a) = \frac{1}{2}$ .  
Or  $\pi_0 C_1 = (\frac{1}{2}, 0, \frac{1}{2}, 0) \begin{pmatrix} 0 \\ 0,6 \\ 0 \\ 0,3 \end{pmatrix} = 0 \neq \frac{1}{2}$ . Ce n'est donc pas le cas.

3. Compléter les valeurs manquantes et en déduire la distribution  $\pi_3$  si  $\pi_0 = (\frac{1}{4}; \frac{1}{4}; \frac{1}{4}; \frac{1}{4})$ .

$$\mathbb{P}^3 = \begin{pmatrix} 0 & 0,605 & 0 & 0,395 \\ 0,45 & 0 & 0,55 & 0 \\ 0 & 0,53 & 0 & 0,47 \\ 0,495 & 0 & 0,505 & 0 \end{pmatrix} \begin{matrix} 1 - 0,395 = 0,605 \\ 1 - 0,53 = 0,47 \\ 1 - 0,505 = 0,495 \end{matrix}$$

→ Le coefficient est le produit scalaire de la 2<sup>ème</sup> ligne par la 2<sup>ème</sup> colonne de  $\mathbb{P}$ , il est donc nul. D'où  $\downarrow 1 - 0,45 = 0,55$ .

On trouve :

$$\pi_3 = \pi_0 \mathbb{P}^3 = (0,23625 \quad 0,28375 \quad 0,26375 \quad 0,21625)$$

4. Supposons que vous puissiez observer un très long brin d'ADN. Indiquer quelle méthode vous choisiriez pour estimer les probabilités de transition figurant dans la matrice  $\mathbb{P}$ .

Sur un long brin, on peut compter la proportion de paires  $aa$  parmi toutes les paires  $aa, ag, at, ac$ . Cette proportion donne la probabilité qu'un a suive un a. On peut faire de même pour les 9 paires  $ga, gt, \dots$  et en déduire les coefficients de la matrice de transition  $\mathbb{P}$ .

<sup>1</sup>Bien que ce modèle soit très peu réaliste il est cependant à la base de nombreuses techniques utilisées en génétique pour reconnaître par exemple les parties codantes d'un gène (exons) des parties non (introns), ces dernières présentant des successions de bases bien plus désordonnées, ou pour comparer une séquence de gène dont on ne connaît pas encore la fonction avec une séquence déjà étudiée.