Correction du devoir surveillé du samedi 23 novembre 2024

Problème 1. Distance de Jukes-Cantor

1) $(A_{n+1}, C_{n+1}, G_{n+1}, T_{n+1})$ est un système complet d'événements et P_{A_n} est une probabilité donc

$$\begin{aligned} P_{A_n}(A_{n+1}) &= 1 - P_{A_n}(C_{n+1}) - P_{A_n}(G_{n+1}) - P_{A_n}(T_{n+1}) \\ &= 1 - \alpha - \alpha - \alpha \qquad \textit{(car } h = 1 \textit{)} \end{aligned}$$

$$\boxed{P_{A_n}(A_{n+1}) = 1 - 3\alpha}$$

2) (A_n, C_n, G_n, T_n) est un système complet d'événements donc *(formule des probabilités totales)*

$$\begin{aligned} P(A_{n+1}) &= P(A_n).P_{A_n}(A_{n+1}) + P(C_n).P_{C_n}(A_{n+1}) + P(G_n).P_{G_n}(A_{n+1}) + P(T_n).P_{T_n}(A_{n+1}) \\ &= (1-3\alpha)P(A_n) + \alpha P(C_n) + \alpha P(G_n) + \alpha P(T_n) \end{aligned}$$

En raisonnant de même pour $P(C_{n+1})$, $P(G_{n+1})$, $P(T_{n+1})$ il vient

$$\boxed{\text{pour tout } n\text{, on a } X_{n+1} = MX_n, \quad \text{où } M = \begin{pmatrix} 1-3\alpha & \alpha & \alpha & \alpha \\ \alpha & 1-3\alpha & \alpha & \alpha \\ \alpha & \alpha & 1-3\alpha & \alpha \\ \alpha & \alpha & \alpha & 1-3\alpha \end{pmatrix}}$$

3) a) Soit $X \in \mathscr{M}_{4,1}(\mathbb{R})$

$$\begin{aligned} MX = (1-4\alpha)X &\iff \begin{pmatrix} \alpha & \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha & \alpha \end{pmatrix} X = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\ &\iff x_1 + x_2 + x_3 + x_4 = 0 \qquad \textit{car } \alpha \neq 0 \end{aligned}$$

$$\boxed{\text{L'ensemble des solutions de } MX = (1-4\alpha)X \text{ est Vect} < \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} >}$$

b) On remarque qu' $\boxed{\text{en prenant } X_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \text{ on a : } MX_1 = X_1}$.

4) a) Soit $A \in \mathscr{M}_n(\mathbb{R})$ une matrice symétrique et inversible

On a $A = A^T$ et A inversible donc $A^{-1} = \left(A^T\right)^{-1}$, de plus on sait que $\left(A^{-1}\right)^T = \left(A^T\right)^{-1}$ *(inverse de la transposée)*

donc $A^{-1} = \left(A^{-1}\right)^T$, autrement dit A^{-1} est symétrique.

$\boxed{\text{Si une matrice est symétrique et inversible alors son inverse est elle aussi symétrique}}$

b) *(Avec la "bonne" matrice)*

$$\begin{aligned} P\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} &\iff \left\{ \begin{array}{rcl} x+y+z+t &=& a \\ x-y &=& b \\ x-z &=& c \\ x-t &=& d \end{array} \right. \\ &\iff \left\{ \begin{array}{rcl} x+(x-b)+(x-c)+(x-d) &=& a \qquad \textit{(substition)} \\ y &=& x-b \\ z &=& x-c \\ t &=& x-d \end{array} \right. \\ &\iff \left\{ \begin{array}{rcl} x &=& \frac{1}{4}(a+b+c+d) \\ y &=& \frac{1}{4}(a-3b+c+d) \\ z &=& \frac{1}{4}(a+b-3c+d) \\ t &=& \frac{1}{4}(a+b+c-3d) \end{array} \right. \\ &\iff \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 \\ 1 & 1 & -3 & 1 \\ 1 & 1 & 1 & -3 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \end{aligned}$$

$$\boxed{P \text{ est inversible et } P^{-1} = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 \\ 1 & 1 & -3 & 1 \\ 1 & 1 & 1 & -3 \end{pmatrix}}$$

(Avec la matrice de l'énoncé)

$$P\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \iff \left\{\begin{array}{rcl} x+y+z+t & = & a \\ x-y & = & b \\ x-z & = & c \\ x+t & = & d \end{array}\right.$$

$$\iff \left\{\begin{array}{rcll} x+(x-b)+(x-c)+(d-x) & = & a & \textit{(substition)} \\ y & = & x-b & \\ z & = & x-c & \\ t & = & d-x & \end{array}\right.$$

$$\iff \left\{\begin{array}{rcl} x & = & \frac{1}{2}(a+b+c-d) \\ y & = & \frac{1}{2}(a-b+c-d) \\ z & = & \frac{1}{2}(a+b-c-d) \\ t & = & \frac{1}{2}(-a-b-c+3d) \end{array}\right.$$

$$\iff \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

$$\boxed{P \text{ est inversible et } P^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix}}$$

c) *Avec le cours sur la diagonalisation vous auriez pu voir mon erreur* 8(

$$M\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = 1\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad M\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} = (1-4\alpha)\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} \qquad M\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} = (1-4\alpha)\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} \qquad M\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} = (1-4\alpha)\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

ce qui entraîne $MP = PD$ avec $D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & (1-4\alpha) & 0 & 0 \\ 0 & 0 & (1-4\alpha) & 0 \\ 0 & 0 & 0 & (1-4\alpha) \end{pmatrix}$

$$\boxed{P^{-1}MP = D \text{ avec } D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & (1-4\alpha) & 0 & 0 \\ 0 & 0 & (1-4\alpha) & 0 \\ 0 & 0 & 0 & (1-4\alpha) \end{pmatrix}}$$

d) La question précédente donne la relation $M = PDP^{-1}$ et permet d'établir par récurrence que

$$\boxed{\forall n \in \mathbb{N},\ M^n = PD^nP^{-1}}$$

Remarque : C'est une question très classique. On rédige la récurrence en se souvenant bien que produit matricielle n'est pas une opération commutative.

5) On a vu à la question 2) que $\forall n \in \mathbb{N}, X_{n+1} = MX_n$, ce qui nous permet d'établir par récurrence que :

$$\boxed{\forall n \in \mathbb{N},\ X_n = M^n X_0}$$

Remarque : C'est une question très classique. On rédige la récurrence en faisant bien attention à la taille des fonctions $X_0.M$ est opération impossible.

6) On suppose qu'au temps 0 , le nucléotide observé est un A donc $X_0 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}$ et alors :

$$\begin{aligned}
X_n &= M^n\begin{pmatrix}1\\0\\0\\0\end{pmatrix}\\
&= \frac{1}{4}\begin{pmatrix}1&1&1&1\\1&-3&1&1\\1&1&-3&1\\1&1&1&-3\end{pmatrix}\begin{pmatrix}1&0&0&0\\0&(1-4\alpha)^n&0&0\\0&0&(1-4\alpha)^n&0\\0&0&0&(1-4\alpha)^n\end{pmatrix}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}\\
&= \frac{1}{4}\begin{pmatrix}1&1&1&1\\1&-3&1&1\\1&1&-3&1\\1&1&1&-3\end{pmatrix}\begin{pmatrix}1\\(1-4\alpha)^n\\(1-4\alpha)^n\\(1-4\alpha)^n\end{pmatrix}
\end{aligned}$$

donc

$$\boxed{\forall n\in\mathbb{N},\quad P(A_n)=\frac{1}{4}\left(1+3(1-4\alpha)^n\right)}$$

Partie B. Un modèle continu

Préliminaires

7) a) C'est une équation différentielle linéaire d'ordre 1 dont l'ensemble de ses solutions est :

$$\boxed{S=\left\{\begin{array}{ccc}\mathbb{R}&\longrightarrow&\mathbb{R}\\ t&\longmapsto&ke^{-4\lambda t}\end{array}\;\middle|\; k\in\mathbb{R}\right\}}$$

b) En cherchant une solution constante on trouve que $t\longmapsto\dfrac{1}{4}$ est une solution *(particulière)* de cette équation on peut alors conclure que l'ensemble des solutions de l'équation $y'=\lambda-4\lambda y$ est

$$\boxed{S=\left\{\begin{array}{ccc}\mathbb{R}&\longrightarrow&\mathbb{R}\\ t&\longmapsto&ke^{-4\lambda t}+\dfrac{1}{4}\end{array}\;\middle|\; k\in\mathbb{R}\right\}}$$

8) on note $g=\ln\circ f$,

• comme ln est croissante sur $]0;+\infty[$ et f est à valeurs postives alors
si f est croissante sur I alors g est croissante sur I; si f est décroissante sur I alors g est décroissante sur I;

• On remarque que $f=\exp\circ g$, et comme exp est croissante sur $\mathbb{R}$ on a :
si g est croissante sur I alors f est croissante sur I; si g est décroissante sur I alors f est décroissante sur I;

$$\boxed{\text{Montrer que } f \text{ et } \ln\circ f \text{ ont les mêmes variations sur cet intervalle.}}$$

9) Soit $d\in\mathbb{R}_+$

$$\begin{aligned}
4m-3n\left(1-e^{-\frac{4d}{3}}\right)\geqslant 0 &\iff 1-e^{-\frac{4d}{3}}\leqslant\frac{4m}{3n} && \textit{car } n>0\\
&\iff e^{-\frac{4d}{3}}\geqslant 1-\frac{4m}{3n}\\
&\iff -\frac{4d}{3}\geqslant\ln\left(1-\frac{4m}{3n}\right) && \textit{car on a supposé } m<\frac{3}{4}n\\
&\iff d\leqslant-\frac{3}{4}\ln\left(1-\frac{4m}{3n}\right)
\end{aligned}$$

$$\boxed{\text{L'ensemble des solutions de cette inéquation est }\left[0;-\frac{3}{4}\ln\left(1-\frac{4m}{3n}\right)\right]}$$

B-I.

10) La première ligne de la relation matricielle $X(t+h)=R(h)X(t)$ donne :

$$P(A(t+h))=(1-3\alpha h)P(A(t))+\alpha\left(P(C(t))+P(G(t))+P(T(t))\right)$$

et $(A(t), C(t), G(t), T(t))$ est un système complet donc

$$P(C(t)) + P(G(t)) + P(T(t)) = 1 - P(A(t))$$

et ainsi

$$\boxed{P(A(t+h)) = (1-3\alpha h)P(A(t)) + \alpha h\big(1 - P(A(t))\big)}$$

11) **a)** La relation précédente donne :

$$\begin{aligned} P(A(t+h)) - P(A(t)) &= (1-3\alpha h)P(A(t)) + \alpha h\big(1-P(A(t))\big) - P(A(t)) \\ &= -4\alpha h P(A(t)) + \alpha h \end{aligned}$$

$$\boxed{\frac{\varphi_A(t+h) - \varphi_A(t)}{h} = \alpha - 4\alpha\varphi_A(t)}$$

b) La question précédente montre qu'en tout point t de $\mathbb{R}^+$ le taux d'accroissement admet une limite réel quand h tend vers 0,

$$\boxed{\varphi_A \text{ est dérivable sur } \mathbb{R}^+ \text{ et } \forall t \in \mathbb{R}^+, \quad \varphi_A'(t) = \alpha - 4\alpha\varphi_A(t)}$$

12) On reconnait l'équation différentielle de la question **7)b)** donc
il existe $k \in \mathbb{R}$ telle que $\varphi_A : t \longmapsto ke^{-4\alpha t} + \dfrac{1}{4}$
et en utilisant $P(A(0))$ il vient :

$$\boxed{P(A(t)) = \left(P(A(0)) - \frac{1}{4}\right)e^{-4\alpha t} + \frac{1}{4}}$$

13) $p_{AA}(t)$ est la probabilité que le nucléotide passe de A à A en t unités de temps.
Si le nucléotide est en A à $t = 0$ on a $P(A(0)) = 1$ et alors pour $t > 0$, $P(A(t)) = \dfrac{3}{4}e^{-4\alpha t} + \dfrac{1}{4}$
donc

$$\boxed{p_{AA}(t) = \frac{1}{4}\left(1 + 3e^{-4\alpha t}\right)}$$

Si le nucléotide est en x $\not{A}$ à $t = 0$ on a $P(A(0)) = 0$ et alors pour $t > 0$, $P(A(t)) = -\dfrac{1}{4}e^{-4\alpha t} + \dfrac{1}{4}$
donc

$$\boxed{\text{si } x \neq A, \quad p_{xA}(t) = \frac{1}{4}\left(1 - e^{-4\alpha t}\right)}$$

B-II.

14) **a)**
```
def p_distance(X, Z):
    n = len(X)
    m = 0
    for k in range(n):
        if X[k] != Z[k]:        # on compte de différence entre les deux listes.
            m += 1
    return m/n
```

b) La fonction suivante ne renvoie rien, elle modifie la liste passée en argument.

```
def mutation(L):
    k = rd.randrange(len(L))             # On choisit un nucléotide au hasard dans la séquence
    urne = [ i for i in range(4) if i != L[k]] # On fait une liste avec les autres nucléotides
    i = rd.randrange(len(urne))
    L[k] = urne[i]                       # on remplace le nucléotide par un autre pris au hasar
```

c) **i.** `espece` contient une liste contenant n réalisations indépendantes d'une variable aléatoire suivant la loi uniforme sur $[\![0, 3]\!]$.
Cette liste représente une séquence quelconque.
$\boxed{\text{La liste n'est pas modifiée}}$ grâce à la copie `X=espece[:]`

ii. L'instruction : `plt.plot(x, x, 'b--')` trace en pointillé bleu la droite d'équation $y = x$.

iii. Au fur à mesure des mutations on calcule la p-distance avec la séquence de l'espèce de départ et affiche le point (nb de mutation/longueur, p-distance)

$$\boxed{\text{Cette courbe trace la } p\text{-distance en fonction la distance évolutive réelle.}}$$

15) Dans le cadre de ce modèle l'observation des n nucléotides est une succession de n épreuves de Bernoulli identiques et indépendantes le succès étant le nucléotide a été modifié de probabilité noté $q(d)$.
Le nombre de succès suit une loi binomiale de paramètre $(n, q(d))$

donc la probabilité d'avoir exactement m nucléotides différents est $\boxed{L(d) = \binom{n}{m} q(d)^m \big(1 - q(d)\big)^{n-m}}$,

où $q(d)$ est la probabilité qu'un nucléotide est modifié pendant $t = \dfrac{d}{3\alpha}$.

on a donc $q(d) = 1 - p_{AA}\left(\dfrac{d}{3\alpha}\right)$ et enfin avec le résultat de la question 13) on obtient $\boxed{q(d) = \dfrac{3}{4}\left(1 - e^{-\frac{4d}{3}}\right)}$

16) a) En notant $g := \ln\circ L$, $\forall d \in \mathbb{R}^+,\ g(d) = \ln\binom{n}{m} + m\ln(q(d)) + (n-m)\ln\big(1 - q(d)\big)$

g est dérivable et

$$\begin{aligned}
g'(d) &= m\frac{q'(d)}{q(d)} - (n-m)\frac{q'(d)}{1-q(d)} \\
&= \frac{mq'(d) - mq(d)q'(d) + (m-n)q(d)q'(d)}{q(d)(1-q(d))} \\
&= q'(d)\frac{m - nq(d)}{q(d)(1-q(d))}
\end{aligned}$$

or $q'(d) = e^{-\frac{4d}{3}} > 0$ et $0 < q(d) < 1$ donc $0 < q(d)(1-q(d)) \leqslant 1$

donc $g'(d)$ est du signe de $m - nq(d) = n - m\dfrac{3}{4}\left(1 - e^{-\frac{4d}{3}}\right)$ et ainsi

$$\boxed{(\ln\circ L)'(d) \text{ est de même signe que } 4m - 3n\left(1 - e^{-\frac{4d}{3}}\right)}$$

b) Le résultat précédent avec le celui de la question 9) nous donne les variations de g qui sont les mêmes que celles de f d'après la question 8).

x	0		$-\dfrac{3}{4}\ln\left(1 - \dfrac{4m}{3n}\right)$		$+\infty$
f		$\nearrow$		$\searrow$	

(je ne complète pas ce tableau)

c) La distance de Jukes-Cantor est la valeur qui maximalise L donc on a bien :

$$\boxed{\widehat{d} = -\frac{3}{4}\ln\left(1 - \frac{4m}{3n}\right).}$$

d) On sait que : $\forall x > -1,\ \ln(1+x) \leqslant x$ donc $\ln\left(1 - \dfrac{4m}{3n}\right) \leqslant -\dfrac{4m}{3n}$

donc

$$\boxed{\frac{m}{n} \leqslant \widehat{d}}$$

Interprétation : la p-distance sous-estime la distance évolutive réelle car les mutations peuvent être invisibles, par exemple un nucléotide A qui mute en C puis en A.

e) *(Difficile sans calculatrice)*

Sur l'exemple donné en introduction $\widehat{d} = -\dfrac{3}{4}\ln\left(1 - \dfrac{4\times 3}{3\times 14}\right) \approx 0{,}25$ et $d = \dfrac{12}{14} \approx 0{,}85$ sont "très différents".

En effet : Sur l'exemple on montre différentes sortes de substitutions elles n'ont pas toute la même probabilité d'apparaitre dans une série de mutations.

17) a) $\widehat{d} = -\dfrac{3}{4}\ln\left(1 - \dfrac{4m}{3n}\right)$. donc $\boxed{\dfrac{m}{n} = \dfrac{3}{4}\left(1 - e^{-\frac{4\widehat{d}}{3}}\right)}$

b) la ligne permettant de tracer la courbe en rouge est : `z.append(3/4* ( 1-exp(-4/3*(i+1)/n)) )`

c) Au vue de ces simulations et en restant dans ce modèle, La distance de Jukes-Cantor semble un bon outil pour estimer la distance évolutive réelle en observant la p-distance entre les deux séquences.

18) *(A vous de m'en dire plus ..., mais il ne faut pas en faire trop)*
Le modèle de Jukes-Cantor suppose que tous les types de substitutions se produisent au même taux, ce qui simplifie le calcul des distances évolutives. Il ne prend en compte que les substitutions de bases et néglige complètement les insertions et délétions, qui jouent pourtant un rôle clé dans l'évolution des séquences.

Problème 2. Modèle de Galton-Watson

Partie A. Préliminaires :

Suites récurrentes de la forme $u_{n+1} = f(u_n)$

1) Montrons par récurrence sur n que pour tout $n \in \mathbb{N}$, $u_n \in I$.

- Pour $n = 0$, on a bien $u_0 \in I$.
- Soit $n \in \mathbb{N}$ tel que $u_n \in I$,

  on sait que $u_n \in I$, comme I est stable par f on a : $u_{n+1} = f(u_n) \in I$,

  En conclusion : $\boxed{\text{Pour tout } n \in \mathbb{N}, \quad u_n \in I}$

2) On sait que : $\forall n \in \mathbb{N},\ u_{n+1} = f(u_n)$, (u_n) converge vers un réel $\ell \in I$ et $\underbrace{\lim_{x \to \ell} f(x) = f(\ell)}_{f \text{ continue en } \ell}$

On a d'une part : $\lim_{n \to +\infty} u_n = \ell$ et $\lim_{x \to \ell} f(x) = f(\ell)$ donc $\lim_{n \to +\infty} f(u_n) = f(\ell)$,

d'autre part : $\lim_{n \to +\infty} u_n = \ell$ donc $\lim_{n \to +\infty} u_{n+1} = \ell$.

et comme : $\forall n \in \mathbb{N},\ u_{n+1} = f(u_n)$, on obtient en passant à la limite : $\boxed{f(\ell) = \ell}$

3) ❶ Si $\boldsymbol{u_0 \leqslant u_1}$ et montrons par récurrence sur n que pour tout n, $u_n \leqslant u_{n+1}$.

- Pour $n = 0$, on a supposé que $u_0 \leqslant u_1$,
- Soit $n \in \mathbb{N}$ tel que $u_n \leqslant u_{n+1}$,

  on sait de plus que $\begin{cases} f \text{ est croissante sur } I \\ u_n \in I \text{ et } u_{n+1} \in I \end{cases}$ on en déduit que : $f(u_n) \leqslant f(u_{n+1})$

  et ainsi : $u_{n+1} \leqslant u_{n+2}$ *(ce qui achève la récurrence)*

Pour tout $n \in \mathbb{N}$, $u_n \leqslant u_{n+1}$. ((u_n) est croissante)

❷ Si $\boldsymbol{u_0 \geqslant u_1}$ et montrons par récurrence sur n que pour tout n, $u_n \geqslant u_{n+1}$.

- Pour $n = 0$, on a supposé que $u_0 \geqslant u_1$,
- Soit $n \in \mathbb{N}$ tel que $u_n \geqslant u_{n+1}$,

  on sait de plus que $\begin{cases} f \text{ est croissante sur } I \\ u_n \in I \text{ et } u_{n+1} \in I \end{cases}$ on en déduit que : $f(u_n) \geqslant f(u_{n+1})$

  et ainsi : $u_{n+1} \geqslant u_{n+2}$ *(ce qui achève la récurrence)*

Pour tout $n \in \mathbb{N}$, $u_n \geqslant u_{n+1}$. ((u_n) est décroissante)

En conclusion : $\boxed{\text{On a bien montré que lorsque } f \text{ est croissante alors la suite est monotone}}$

4)
```
def suite(u0, n, f):
    u = u0
    L = [u]
    for k in range(n):
        u = f(u)          # u contient successivement les valeurs de u_n
        L.append(u)       # on stocke les termes de u_n dans la liste L
    return L
```

les fonctions convexes

5) L'équation réduite de la tangente en a est

$$\boxed{y = \varphi'(a)(x-a) + \varphi(a)}$$

6) **a)** On suppose φ strictement convexe sur I, et on fixe un réel $a \in I$.

i. Théorème des accroissements finis : φ est continue sur $[a,x]$ et dérivable sur $]a,x[$ donc il existe $c \in]a,x[$ tel que $\varphi(x) - \varphi(a) = \varphi'(c)(x-a)$

$$\boxed{\text{il existe un réel } c \in]a,x[\text{ tel que : } \varphi(x) = \varphi(a) + \varphi'(c)(x-a).}$$

ii. Soit $x \in]a,+\infty[\cap I$,

on a $x > a$, on applique le résultat de la question précédente pour définir c

φ est strictement convexe sur I donc $\varphi'(c) > \varphi'(a)$, de plus $x - a > 0$ on obtient alors :

$$\varphi(x) > \varphi(a) + \varphi'(a)(x-a)$$

$$\boxed{\text{sur }]a,+\infty[\cap I, \text{ la courbe de } \varphi \text{ est strictement au-dessus de la tangente à la courbe en } a}$$

iii. On montre de même pour $x \in]-\infty, a[\cap I$ que

$$\boxed{\text{il existe un réel } c \in]x,a[\text{ tel que : } \varphi(x) = \varphi(a) + \varphi'(c)(x-a).}$$

et comme φ' est croissante on obtient

$$\varphi(x) > \varphi(a) + \varphi'(a)(x-a)$$

$$\boxed{\text{la courbe de } \varphi \text{ est strictement au-dessus de la tangente à la courbe en } a \text{ sur }]-\infty, a[\cap I}$$

b) **i.** Le point de la courbe d'abscisse b est strictement au dessus de la tangente en a donc $\varphi(b) > \varphi'(a)(b-a) + \varphi(a)$ et comme $b - a > 0$, il vient :

$$\boxed{\varphi'(a) < \frac{\varphi(b) - \varphi(a)}{b-a}}$$

ii. Le point de la courbe d'abscisse a est strictement au dessus de la tangente en b donc $\varphi(a) > \varphi'(b)(a-b) + \varphi(b)$ et comme $b - a > 0$, il vient :

$$\boxed{\varphi'(b) > \frac{\varphi(b) - \varphi(a)}{b-a}}$$

iii. On vient de montrer que si $a < b$ alors $\varphi'(a) < \dfrac{\varphi(b) - \varphi(a)}{b-a} < \varphi'(b)$, en particulier $\varphi'(a) < \varphi'(b)$

donc $\boxed{\varphi' \text{ est strictement croissante sur } I}$, ce qui est la définition de $\boxed{\varphi \text{ est strictement convexe sur } I}$

Partie B. Un exemple

7) Il s'agit de l'espérance de la loi

x_i	0	1	2
$P(X = x_i)$	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{2}{8}$

c'est à dire $E(X) = 0 + 1 \times \dfrac{5}{8} + 2 \times \dfrac{2}{8}$

$$\boxed{\text{le nombre moyen d'enfants par individu est égal à } \frac{9}{8}}$$

8) p_1 est la probabilité que le premier individu n'est pas d'enfant donc $\boxed{p_1 = \dfrac{1}{8}}$

9) On note Y_n le nombre de naissances de l'année n .

$(Y_1 = 0), (Y_1 = 1), (Y_1 = 2)$ forment un système complet donc *(formule des probabilités totales)*.

$$p_2 = P(Y_2 = 0) = P(Y_1 = 0)P_{(Y_1=0)}(Y_2 = 0) + P(Y_1 = 1)P_{(Y_1=1)}(Y_2 = 0) + P(Y_1 = 2)P_{(Y_1=2)}(Y_2 = 0)$$

Si $(Y_1 = 0)$ est réalisée personne ne pourra avoir d'enfant l'année suivante donc on est certain que $(Y_2 = 0)$,

$$P_{(Y_1=0)}(Y_2 = 0) = 1$$

Si $(Y_1 = 1)$ est réalisée alors au début de l'année 1 il n'y a qu'un individu, $(Y_2 = 0)$ est réalisé lorsque cet individu a 0 enfant ,

$$P_{(Y_1=1)}(Y_2 = 0) = q_0$$

Si $(Y_1 = 2)$ est réalisée alors au début de l'année 1 il y a 2 individus, $(Y_2 = 0)$ est réalisé lorsqu'ils ont tous les deux 0 enfant de manière indépendantes,

$$P_{(Y_1=2)}(Y_2 = 0) = q_0^2$$

Il vient

$$\boxed{p_2 = \frac{1}{8} + \frac{5}{8}q_0 + \frac{2}{8}q_0^2}$$

10) On note toujours Y_n le nombre de naissances de l'année n .
$(Y_1 = 0), (Y_1 = 1), (Y_1 = 2)$ forment un système complet donc *(formule des probabilités totales)*.

$$p_{n+1} = P(Y_{n+1} = 0) = P(Y_1 = 0)P_{(Y_1=0)}(Y_{n+1} = 0) + P(Y_1 = 1)P_{(Y_1=1)}(Y_{n+1} = 0) + P(Y_1 = 2)P_{(Y_1=2)}(Y_{n+1} = 0)$$

Si $(Y_1 = 0)$ est réalisée alors on est certain que $(Y_2 = 0)$,

$$P_{(Y_1=0)}(Y_{n+1} = 0) = 1$$

Si $(Y_1 = 1)$ est réalisée alors on se retrouve dans la situation du début avec un décacalage d'un an, donc

$$P_{(Y_1=1)}(Y_{n+1} = 0) = p_n$$

Si $(Y_1 = 2)$ est réalisée alors au début de l'année 1 il y a 2 individus, on observe deux expériences indépendantes identiques à celle du début donc

$$P_{(Y_1=2)}(Y_{n+1} = 0) = p_n^2$$

Il vient en conclusion :

$$\boxed{p_{n+1} = \frac{1}{8} + \frac{5}{8}p_n + \frac{2}{8}p_n^2}$$

11) L'étude de f ne pose pas de problème :

x	0		1
f	$\frac{1}{8}$	$\nearrow$	1

f est croissante sur $[0,1]$ donc d'après la partie A, (p_n) est une suite monotone,

or $p_0 = 0$, $p_1 = \dfrac{1}{8}$ donc $p_0 \leqslant p_1$ et ainsi : $\boxed{\text{La suite } (p_n) \text{ est croissante}}$

12) (p_n) est croissante et majorée par 1 (p_n est *la valeur d'*une probabilité)
donc *(théorème de convergence monotone)* (p_n) converge.

13) Comme $f(1/2) = 1/2$ on a :

x	0		$\frac{1}{2}$
f	$\frac{1}{8}$	$\nearrow$	$\frac{1}{2}$

donc $[0;1/2]$ est stable par f et ainsi *(Partie A)* (p_n) est à valeurs dans $[0,1/2]$

$$\boxed{\text{Pour tout } n \in \mathbb{N}, \text{ on a } p_n \leqslant \frac{1}{2}}$$

14) D'après la partie A, les limites possibles de (p_n) sont les solutions de $f(x) = x$ sur $[0,1]$, or

$$\begin{aligned} f(x) = x &\iff \frac{1}{8} + \frac{5}{8}x + \frac{2}{8}x^2 = x \\ &\iff \frac{1}{8} - \frac{3}{8}x + \frac{2}{8}x^2 = 0 \\ &\iff 2x^2 - 3x + 1 = 0 \\ &\iff (2x-1)(x-1) = 0 \end{aligned}$$

Les deux limites possibles sont donc 1 et $\frac{1}{2}$, mais comme $\forall n \in \mathbb{N},\ p_n \leqslant \frac{1}{2}$,

$$\boxed{(p_n) \text{ converge vers } \frac{1}{2}}$$

Interprétation : Au bout d'un temps long l'espèce aura disparu avec une probabilité proche de $\frac{1}{2}$.

Partie C. Le cas général

15) Si $q_0 = 1$ on est certain que le premier individu n'est pas d'enfant donc

$$\boxed{\text{Si } q_0 = 1\text{, la probabilité d'extinction de l'espèce est égale à } 1}$$

Si $q_0 = 0$ on est certain que chaque individu est au moins un enfant donc

$$\boxed{\text{Si } q_0 = 0\text{, la probabilité d'extinction de l'espèce est égale à } 0}$$

16) Le nombre moyen d'enfants par individu est : $\boxed{m = \sum_{k=0}^{N} k q_k}$ *(je décide de exceptionnellement de ne pas rédiger)*

17) On raisonne comme à la question 10) en utilisant la formule des probabilités totales avec le système complet $(Y_n = i)_{0 \leqslant i \leqslant N}$

$$P_{(Y_1=i)}(Y_{n+1} = 0) = p_n^i$$

il vient

$$\boxed{\text{pour tout } n \in \mathbb{N},\ p_{n+1} = f\left(p_n\right) \text{ où } f : x \longmapsto \sum_{k=0}^{N} q_k x^k}$$

18) **a)**
```
def construction_f(q):
    def f(x):                    # On définit une fonction f dans la fonction construction_f .
        s = 0
        for k in range(len(q)):
            s += q[k]*x**k
        return s
    return f
```

b)
```
n = 100
u0 = 0
q = [1/8, 5/8 , 2/8]
f = construction_f(q)
u = suite(u0, n, f)
plt.plot(u,'k+')
```

19) La fonction f est dérivable et sa dérivée est $f' : x \longmapsto \sum_{k=1}^{N} k q_k x^{k-1}$ est positive sur $[0, 1]$,

donc $\boxed{f \text{ est croissante sur } [0,1]}$,

On est dans les conditions pour pouvoir appliquer les résultats de la partie A.

La suite (p_n) est donc monotone et à valeurs dans $[0, 1]$ donc

$$\boxed{(p_n) \text{ est convergente}}$$

20) $f(1) = \sum_{k=0}^{N} q_k$ et (q_k) est une loi de probabilité donc $\boxed{f(1) = 1}$.

21) La fonction f' est dérivable et sa dérivée est $f'' : x \longmapsto \sum_{k=2}^{N} k(k-1) q_k x^{k-2}$ est strictement positive sur $[0, 1]$ car on a supposé $q_0 + q_1 < 1$,

donc f' est strictement croissante sur $[0, 1]$ et ainsi f est ainsi $\boxed{f \text{ est strictement convexe sur } [0,1]}$

L'équation de la tangente en 1 est : $y = f'(1)(x-1) + f(1)$ or $f(1) = 1$ et on remarque que $m = f'(1)$ donc

$$\boxed{y = m(x-1) + 1 \text{ est une équation de la tangente à } C_f \text{ en } 1}$$

22) **a)** on suppose que $m \leqslant 1$.

i. Pour $x \in [0, 1]$, $m(x-1) + 1 \geqslant x - 1 + 1$ car $m \geqslant 1$ donc $m(x-1) + 1 \geqslant x$

$$\boxed{\text{la tangente à la courbe de } f \text{ en } 1 \text{ est au-dessus de } \Delta : y = x \text{ sur } [0,1]}$$

ii. f est strictement convexe donc $f(x) > m(x-1)+1$ sur $[0, 1[$ et avec le résultat précédent il vient $f(x) > x$. de plus $f(1) = 1$ donc

$$\boxed{\text{l'équation } f(x) = x \text{ admet une unique solution sur } [0,1]}$$

iii. On sait (p_n) converge (19)) et on est dans les conditions permettant d'utiliser les résultats de 3) il vient que (p_n) converge vers une solution de $f(x) = x$ sur $[0, 1]$ et on vient de montrer 1 est la seule solution donc

$$\boxed{(p_n) \text{ converge vers } 1}$$

b) on suppose $m > 1$.

i. on a $f'(0) = q_1 < 1$, $f'(1) = m > 1$ et f est continue sur $[0,1]$, donc *(théorème des valeurs intermédiaires)*

$$\boxed{\text{il existe un réel } c \in \left]0,1\right[\text{ tel que } f'(c) = 1}$$

ii. *(je n'utilise pas de raisonnement par l'absurde)*,
f est strictement convexe donc en 1 le point de la courbe est strictement au-dessus de la tangente en $c \in]0,1[$, autrement dit : $f(1) > f'(c)(1-c) + f(c)$ et comme $f'(c) = 1$ et $f(1) = 1$ il vient

$$\boxed{f(c) < c}$$

iii. On pose : $g : x \longmapsto f(x) - x$
$g(0) = q_1 > 0$ et $g(c) = f(c) - c < 0$, donc *(théorème des valeurs intermédiaires)*

$$\boxed{\text{il existe un réel } \alpha \in \left]0,c\right] \text{ tel que } f(\alpha) = \alpha}$$

iv. *(rapidement)*
Le tableau de variations de g est

x	0	c	1
g	$q_0 > 0$ ↘	$f(c) - c < 0$	↗ 0

$$\boxed{\alpha \text{ est l'unique point fixe de } f \text{ sur } \left]0,1\right[}$$

v. l'intervalle $\left[0,\alpha\right]$ est stable par f et p_0 est dans cet intervalle donc

$$\boxed{(p_n) \text{ est à valeurs dans } [0,\alpha]}$$

vi. Comme α est l'unique point fixe de $[0,\alpha]$ et que (p_n) converge vers un point fixe

$$\boxed{\text{La suite } \left(p_n\right) \text{ converge vers } \alpha}$$

23) l'extinction de l'espèce est-elle certaine quand $m \leqslant 1$ (quand le nombre moyen d'enfants par individu est inférieur ou égal à 1)

Partie D. Vitesse d'extinction.

24) Le programme compléter est

```
def simul():
    a = rd.random()
    if a <= 1/8:
        return 0
    if a <= 6/8:
        return 1
    return 2

m = 9/8
```

25) On voit 11 courbes sur 20 pour lesquelles l'espèce ne disparait pas

$$\boxed{\text{la probabilité d'extinction de cette espèce semble proche de } \frac{1}{2}}$$

26) On remarque que lorsque l'espèce ne s'éteint pas la suite $\left(\dfrac{a_n}{m^n}\right)$ semble constante après un temps long.

27) Les limites possibles semblent être 0 ou $+\infty$.

28) *(Réponse non personnel)*
Le modèle de Galton-Watson est adapté pour étudier la dynamique des populations, la propagation de gènes ou d'épidémies, et les processus de branchement. Cependant, il présente des limites, comme l'absence d'interactions entre individus, des probabilités de reproduction fixes et une incapacité à modéliser des phénomènes stabilisateurs ou continus. Pour l'améliorer, on peut introduire des interactions, des probabilités variables, des conditions environnementales, des modèles multi-types ou des mécanismes de régulation, rendant le modèle plus réaliste pour des populations complexes.

FIN DE LA CORRECTION