MESURES D'EWENS

L'objectif de ce devoir est d'étudier la répartition en diverses espèces des individus d'un écosystème ; par exemple, la répartition par espèces des arbres d'une forêt tropicale. Dans la première partie, on introduit un modèle markovien qui prend en compte les phénomèmes de spéciation et d'extinction des espèces, et qui permet de décrire l'évolution d'un écosystème. Dans la seconde partie du devoir, on s'intéresse à l'évolution d'une espèce fixée au sein d'un écosystème, et on montre une borne sur le temps d'extinction de cette espèce. Enfin, dans la dernière partie, on s'intéresse au régime stationnaire de la chaîne de Markov introduite dans la première partie.

1. Évolution d'un écosystème

Dans ce qui suit, on fixe un entier $N \geq 1$, qui représentera le nombre total d'individus dans l'écosystème que l'on considère, toutes espèces confondues. La répartition des individus en différentes espèces est encodée par une *composition* de taille N, c'est-à-dire une suite

$$c = (N_1, N_2, \ldots, N_R)$$

d'entiers telle que $N_i \geq 0$ pour tout i, et telle que $\sum_{i=1}^R N_i = N$. Le nombre N_i représente le nombre d'individus appartenant à l'espèce i. Par exemple, si l'on considère une forêt avec 10 arbres dont 5 chênes, 2 peupliers et 3 sapins, alors on pourra représenter cette répartition par la composition $c = (5, 2, 3)$.

On note $\mathfrak{C}_N$ l'ensemble des compositions de taille N :

$$\mathfrak{C}_N = \left\{ (N_1, \ldots, N_R) \mid R \geq 1,\ N_i \geq 0 \text{ pour tout } i \in [\![1, R]\!],\ \sum_{i=1}^R N_i = N \right\}$$

On fixe également un paramètre $\nu \in (0, 1)$. On va introduire une chaîne de Markov $(c(t))_{t\geq 0}$ à valeurs dans $\mathfrak{C}_N$ et qui modélise l'évolution de la répartition entre espèces. La taille N de l'écosystème restera fixe au cours du temps. On fixe une composition de départ $c(0) \in \mathfrak{C}_N$. Si $c(t) = (N_1(t), N_2(t), \ldots, N_{R(t)}(t))$ est construite, pour obtenir $c(t + 1)$, on procède comme suit :

(E1) On choisit au hasard un individu de la population qui meurt. La probabilité que cet individu soit issu de l'espèce i avec $i \in [\![1, R(t)]\!]$ est égale à $\frac{N_i(t)}{N}$.

(E2) Avec probabilité ν, l'individu de type i qui est mort à l'étape (E1) est remplacé par un nouvel individu d'une nouvelle espèce (phénomène de spéciation). Dans ce cas, la nouvelle composition $c(t + 1)$ s'écrit :

$$c(t + 1) = (N_1(t), \ldots, N_{i-1}(t), N_i(t) - 1, N_{i+1}(t), \ldots, N_{R(t)}(t), 1).$$

(E3) Avec probabilité $1-\nu$, l'individu de type i qui est mort à l'étape (E1) est remplacé par un individu d'une espèce j avec $j \in [\![1, R(t)]\!]$. La probabilité de choix de j est $\frac{N_j(t)}{N}$. Les indices i et j étant choisis, la nouvelle composition $c(t + 1)$ s'écrit :

$$c(t+1) = \begin{cases} (N_1(t), \ldots, N_i(t) - 1, \ldots, N_j(t) + 1, \ldots, N_{R(t)}(t)) & \text{si } j \neq i, \\ c(t) & \text{si } j = i. \end{cases}$$

On suppose que les choix faits dans les étapes (E1), (E2) et (E3) sont indépendants entre eux, et indépendants pour des temps distincts. L'algorithme écrit ci-dessus définit alors une chaîne de Markov à valeurs dans $\mathfrak{C}_N$. Notons que l'ensemble $\mathfrak{C}_N$ est infini, car il contient toutes les compositions $c = (0, 0, \ldots, 0, N)$ avec un nombre arbitraire de 0. Ceci n'empêche pas de travailler avec la chaîne $(c(t))_{t\in\mathbb{N}}$, car $\mathfrak{C}_N$ est dénombrable.

Q1. On note $\mathfrak{C}_N^*$ l'ensemble des compositions de taille N sans part égale à 0 ; par exemple, $\mathfrak{C}_4^* = \{(4), (3, 1), (2, 2), (1, 3), (2, 1, 1), (1, 2, 1), (1, 1, 2), (1, 1, 1, 1)\}$. Montrer que pour tout $N \geq 1$, $\mathfrak{C}_N^*$ contient 2^{N-1} compositions. On pourra procéder par récurrence sur N.

Q2. Dans ce qui suit, on suppose que $c(0)$ appartient à $\mathfrak{C}_N^*$. Les autres compositions $c(t \geq 1)$ pourront en revanche avoir des parts $N_i(t)$ égales à 0, qui représenteront les espèces i éteintes au temps t. Écrire un programme qui simule $(c(0), c(1), \ldots, c(T))$, avec par exemple $N = 10$, $\nu = 0.1$, $c(0) = (5, 2, 3)$ et $T = 20$. Donner un résultat de cette simulation.

Q3. On note $R(t)$ le nombre total de parts de $c(t)$. Montrer qu'on peut écrire $R(t)$ sous la forme

$$R(t) = R(0) + V_1 + V_2 + \cdots + V_t,$$

où les V_i sont des variables indépendantes et de même loi, que l'on précisera. Décrire le comportement des variables aléatoires $R(t)$ et $\frac{R(t)}{t}$ lorsque t tend vers l'infini. Que peut-on dire du nombre

$$E(t) = \operatorname{card}\{i \in [\![1, R(t)]\!] \mid N_i(t) = 0\}$$

d'espèces qui sont éteintes au temps t, en particulier lorsque t tend vers l'infini ?

2. Évolution d'une espèce fixée

Dans cette partie, on s'intéresse à l'évolution d'une espèce fixée dans l'écosystème, par exemple celle d'indice $i = 1$. Les paramètres N et ν sont fixés comme précédemment, avec $\nu > 0$.

Q4. Montrer que $(N_1(t))_{t\geq 0}$ est une chaîne de Markov sur $[\![0, N]\!]$, dont les probabilités de transition sont

$$\begin{aligned} P(k, k+1) &= \frac{(N-k)k}{N^2}\,(1-\nu); \\ P(k, k) &= \frac{N-k}{N}\,\nu + \frac{k^2 + (N-k)^2}{N^2}\,(1-\nu); \\ P(k, k-1) &= \frac{k}{N}\,\nu + \frac{(N-k)k}{N^2}\,(1-\nu). \end{aligned}$$

Q5. Montrer que 0 est l'unique état absorbant de la chaîne $(N_1(t))_{t\geq 0}$. En déduire que $N_1(t) \to_{t\to\infty} 0$ presque sûrement.

Q6. On cherche à déterminer à quelle vitesse l'espèce d'indice $i = 1$ s'éteint. Pour $k \in [\![0, N]\!]$, calculer l'espérance conditionnelle $\mathbb{E}[N_1(t+1) \mid N_1(t) = k]$, et en déduire par récurrence sur t que

$$\mathbb{E}[N_1(t)] = N_1(0)\left(1 - \frac{\nu}{N}\right)^t.$$

Q7. Soit $T = \inf\{t \geq 0 \mid N_1(t) = 0\}$. Montrer que $\mathbb{P}[T > t] \leq \mathbb{E}[N_1(t)]$. En déduire que

$$\mathbb{E}[T] \leq \sum_{t=0}^{\infty} \mathbb{E}[N_1(t)] = \frac{N\,N_1(0)}{\nu}.$$

Commenter ce résultat.

3. Régime stationnaire de la chaîne de Markov

Si $c \in \mathfrak{C}_N$ est une composition de taille N, on note $p = \pi(c)$ la *partition* de taille N qui est obtenue en retirant de c les parts de taille 0, et en réordonnant ses parts de manière décroissante. Par exemple, si $c = (1,5,1,0,2,1,0)$, alors $p = \pi(c) = (5,2,1,1,1)$. On notera $\mathfrak{P}_N$ l'ensemble des partitions de taille N (compositions sans part égale à 0, et avec des parts classées par ordre décroissant) :

$$\mathfrak{P}_N = \left\{ (N_1 \geq N_2 \geq \cdots \geq N_R) \mid R \geq 1,\ N_i \geq 1 \text{ pour tout } i \in [\![1,R]\!],\ \sum_{i=1}^{R} N_i = N \right\}.$$

Par exemple,

$$\mathfrak{P}_4 = \{(4),(3,1),(2,2),(2,1,1),(1,1,1,1)\}.$$

Si $(c(t))_{t\geq 0}$ est la chaîne de Markov introduite dans la première partie, alors $p(t) = \pi(c(t))$ définit encore une chaîne de Markov, cette fois-ci à valeurs dans l'ensemble fini $\mathfrak{P}_N$ (on ne demande pas de démonstration de ce fait). De plus, pour construire la chaîne $(p(t))_{t\geq 0}$, on peut utiliser les mêmes étapes (E1), (E2) et (E3) que pour la chaîne $(c(t))_{t\geq 0}$; la seule différence est qu'il faut à chaque fois supprimer les éventuelles parts 0 qui ont été créées, et reclasser par ordre décroissant les parts de $p(t)$.

Q8. Si $N = 4$, écrire en fonction de ν la matrice de transition de la chaîne de Markov $(p(t))_{t\in\mathbb{N}}$ (comme $\operatorname{card}\mathfrak{P}_4 = 5$, cette matrice est de taille 5×5). On ne demande pas le détail des calculs pour tous les coefficients.

Q9. Pour tout entier $N \geq 1$, montrer que la chaîne $(p(t))_{t\in\mathbb{N}}$ est irréductible sur $\mathfrak{P}_N$, et qu'elle est apériodique.

Q10. Montrer qu'il existe une mesure de probabilité $\mu_{N,\nu}$ sur $\mathfrak{P}_N$ telle que

$$\lim_{t\to\infty} \mathbb{P}[p(t) = p] = \mu_{N,\nu}(p)$$

pour toute partition $p \in \mathfrak{P}_N$.

Étant donnée une partition $p \in \mathfrak{P}_N$ et un entier $i \in [\![1,N]\!]$, on note $m_i(p)$ le nombre de parts de p égales à i. Par exemple, si $p = (5,2,1,1,1) \in \mathfrak{P}_{10}$, alors la suite de ses multiplicités est :

$$(m_1, m_2, \ldots, m_{10}) = (3,1,0,0,1,0,0,0,0,0).$$

Si θ est un paramètre positif, la mesure d'Ewens de paramètre θ sur $\mathfrak{P}_N$ est la mesure positive définie par

$$\rho_{N,\theta}(p) = \frac{N!}{\prod_{i=1}^{N} i^{m_i}\,(m_i)!}\,\frac{\theta^{m_1+m_2+\cdots+m_N}}{\theta(\theta+1)(\theta+2)\cdots(\theta+N-1)}$$

où $(m_1, m_2, \ldots, m_N)$ sont les multiplicités de la partition p.

Q11. Montrer que pour tout $\theta > 0$, $\rho_{4,\theta}$ est une mesure de probabilité sur $\mathfrak{P}_4$. Si $\theta = \frac{4\nu}{1-\nu}$, montrer que $\rho_{4,\theta}$ est laissée invariante par la matrice de transition calculée à la question Q8, et donc que $\rho_{4,\theta} = \mu_{4,\nu}$.

On admet dans la suite que pour tout entier $N \geq 1$, si $\theta = \frac{N\nu}{1-\nu}$, alors la mesure d'Ewens $\rho_{N,\theta}$ est la mesure invariante $\mu_{N,\nu}$ introduite à la question Q10. Les dernières questions vont permettre de simuler la loi $\rho_{N,\theta}$ sur l'ensemble des partitions $\mathfrak{P}_N$.

On note $\mathfrak{S}_N$ l'ensemble des $N!$ permutations de taille N, c'est-à-dire les bijections $\sigma : [\![1, N]\!] \to [\![1, N]\!]$. Un cycle $(a_1, a_2, \ldots, a_p)$ est une permutation qui envoie a_1 sur a_2, a_2 sur a_3, *etc.* jusqu'à a_p qui est envoyé sur a_1 ; et tous les autres éléments de $[\![1, N]\!]$ sont laissés invariants. On rappelle que toute permutation $\sigma \in \mathfrak{S}_N$ peut s'écrire sous la forme

$$\sigma = (a_{1,1}, \ldots, a_{1,p_1}) \circ (a_{2,1}, \ldots, a_{2,p_2}) \circ \cdots \circ (a_{\ell,1}, \ldots, a_{\ell,p_\ell}),$$

les cycles $(a_{i,1}, \ldots, a_{i,p_i})$ étant disjoints et recouvrant $[\![1, N]\!]$. Par exemple, la permutation $\sigma = 914362857$ qui envoie 1 sur 9, 2 sur 1, 3 sur 4, *etc.* se décompose en le produit de cycles $(1, 9, 7, 8, 5, 6, 2) \circ (3, 4)$. Le type cyclique d'une permutation $\sigma \in \mathfrak{S}_N$ est la partition $p(\sigma) \in \mathfrak{P}_N$ dont les parts sont les tailles des cycles de σ. Par exemple, la permutation 914362857, qui est de taille 9, a pour type cyclique $(7, 2)$.

Q12. Montrer que si $p \in \mathfrak{P}_N$, alors le nombre de permutations $\sigma \in \mathfrak{S}_N$ dont le type cyclique est p est égal à

$$\frac{N!}{\prod_{i=1}^{N} i^{m_i}\,(m_i)!},$$

où $(m_1, m_2, \ldots, m_N)$ sont les multiplicités de la partition p.

Q13. Montrer par récurrence sur N que toute permutation $\sigma \in \mathfrak{S}_N$ s'écrit de manière unique sous la forme

$$\sigma = (1, n_1) \circ (2, n_2) \circ (3, n_3) \circ \cdots \circ (N, n_N),$$

où chaque n_i appartient à $[\![1, i]\!]$, et où par convention, si $n_i = i$, alors (i, i) est la permutation identité. On pourra dans la récurrence poser $n_N = \sigma^{-1}(N)$. Montrer que de plus, le nombre de cycles disjoints de σ est

$$\ell(\sigma) = \mathrm{card}\,\{i \in [\![1, N]\!] \mid n_i = i\}.$$

Un paramètre $\theta > 0$ étant fixé, pour chaque $i \in [\![1, N]\!]$, on note m_i la variable aléatoire de loi

$$\mathbb{P}[m_i = j] = \frac{1}{\theta + i - 1} \text{ si } j \in [\![1, i-1]\!] \quad ; \quad \mathbb{P}[m_i = i] = \frac{\theta}{\theta + i - 1}.$$

Q14. Si $\sigma_N = (1, m_1) \circ (2, m_2) \circ \cdots \circ (N, m_N)$, montrer que la permutation aléatoire σ_N a pour loi

$$\mathbb{P}[\sigma_N = \sigma] = \frac{\theta^{\ell(\sigma)}}{\theta(\theta+1)(\theta+2)\cdots(\theta+N-1)},$$

où $\ell(\sigma)$ est le nombre de cycles de σ (les points fixes étant comptés comme des cycles de longueur 1). En déduire que $p_N = p(\sigma_N)$ a pour loi la mesure d'Ewens $\rho_{N,\theta}$.

Q15. Utiliser la question précédente pour écrire un programme qui simule une partition aléatoire de loi $\rho_{N,\theta}$, par exemple avec $N = 1000$ et $\theta = 10$. Dans `Sage`, on pourra utiliser la classe `Permutation` et ses méthodes, en particulier la méthode `.cycle_type()` qui affiche le type cyclique d'une permutation. Donner un résultat de cette simulation.

On obtient ainsi une méthode simple pour simuler la répartition par espèces des individus d'un écosystème. Le paramètre θ est appelé constante de biodiversité du modèle, et il régit le nombre moyen d'espèces distinctes et non éteintes dans une population.

Corrigé

Q1. Pour montrer que $\mathfrak{C}_N^*$ a pour cardinal 2^{N-1}, on peut procéder par récurrence sur N, le cas $N=1$ étant trivial puisque $\mathfrak{C}_1^* = \{(1)\}$. Supposons établi le résultat au rang N, et considérons une composition $c = (N_1, N_2, \ldots, N_l)$ de taille

$$N_1 + N_2 + \cdots + N_l = N+1.$$

Si $N_l = 1$, on peut associer à c la composition $c' = (N_1, N_2, \ldots, N_{l-1})$ de taille N. D'autre part, si $N_l \geq 2$, on peut associer à c la composition $c' = (N_1, N_2, \ldots, N_l - 1)$ de taille N. L'application $c \in \mathfrak{C}_{N+1}^* \mapsto c' \in \mathfrak{C}_N^*$ est surjective, et elle atteint toute composition $c' \in \mathfrak{C}_N^*$ deux fois : on peut retrouver c à partir de c' soit en rajoutant 1 à la dernière part de c', soit en rajoutant une part de taille 1 à la fin de c'. On conclut que

$$\operatorname{card} \mathfrak{C}_{N+1}^* = 2(\operatorname{card} \mathfrak{C}_N^*) = 2^N,$$

d'où le résultat au rang $N+1$.

Q2. Dans la suite, les programmes écrits fonctionnent dans `Sage`, et à quelques modifications près, ils devraient également fonctionner dans `Python`. On utilisera le programme suivant pour choisir une part d'une composition $c = (N_1, N_2, \ldots, N_r)$ de taille N, la part i étant choisie avec probabilité $\frac{N_i}{N}$:

```
def choose_part(c):
    N = sum(c)
    count = 0
    res = -1
    alea = random()
    while count < alea:
        res +=1
        count += c[res]/N
    return res
```

Le programme suivant prend en argument la taille N de la population, le paramètre nu, un temps T, et en argument optionnel une répartition initiale c0 (le programme utilise sinon comme répartition initiale une composition aléatoire de taille N). Il calcule toutes les compositions $c(t)$ pour $t \in [\![0, T]\!]$.

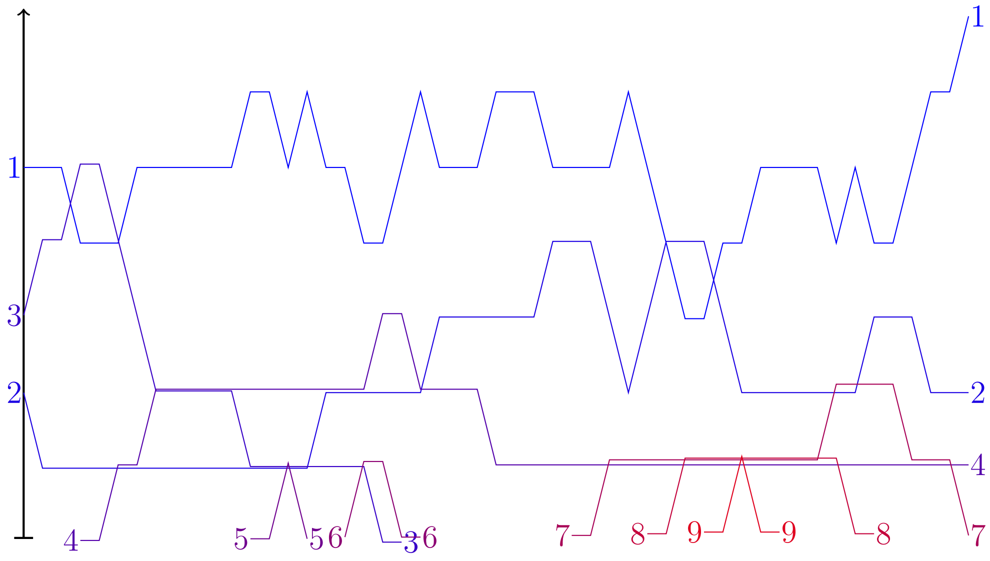
```
def markov_composition(N,nu,T,c0=None):
    res = []
    #composition initiale
    if c0 == None:
        comp = []
        count = 1
        while sum(comp) < N:
        if sum(comp) + count == N:
            comp.append(count)
        elif random() < 0.5:
            comp.append(count)
            count = 1
        else:
            count +=1
    else:
        comp = copy(c0)
    res.append(copy(comp))
    #transitions markoviennes
    for t in range(0,T):
```

```
        i = choose_part(comp)
        if random() < nu:
            comp[i] -= 1
            comp.append(1)
        else:
            j = choose_part(comp)
            comp[i] -=1
            comp[j] +=1
        res.append(copy(comp))
    return res
```

Par exemple, avec $N = 10$, $\nu = 0.1$, $T = 20$ et $c(0) = (5, 2, 3)$, on obtient comme résultat :

```
[[5, 2, 3],
 [4, 3, 3],
 [3, 3, 3, 1],
 [3, 4, 2, 1],
 [2, 4, 2, 1, 1],
 [2, 3, 2, 2, 1],
 [2, 3, 2, 2, 1],
 [1, 4, 2, 2, 1],
 [1, 4, 2, 2, 1],
 [1, 5, 2, 1, 1],
 [1, 4, 2, 2, 1],
 [1, 3, 2, 3, 1],
 [0, 3, 2, 4, 1],
 [0, 3, 2, 3, 2],
 [0, 3, 2, 3, 2],
 [0, 2, 2, 3, 3],
 [0, 2, 1, 3, 4],
 [0, 2, 1, 2, 5],
 [0, 2, 1, 2, 4, 1],
 [0, 3, 1, 2, 4, 0],
 [0, 2, 1, 3, 4, 0]]
```

On a représenté ci-dessous une évolution de la répartition entre espèces avec $N = 10$, $\nu = 0.1$, $c(0) = (5, 2, 3)$ et pour $t \in [\![0, 50]\!]$ — le dessin a été produit par un programme `Python` qui a écrit le code `LaTeX/TikZ` à partit du programme précédent.



On note l'apparition de nouvelles espèces $4, 5, 6, \ldots$, et aussi la disparition de certaines espèces (y compris parmi celles qui sont apparues après le temps $t = 0$).

Q3. Le processus $(R(t))_{t\in\mathbb{N}}$ est une marche aléatoire sur $\mathbb{N}$ avec

$$\mathbb{P}[R(t+1) - R(t) = 1] = \nu$$
$$\mathbb{P}[R(t+1) - R(t) = 0] = 1 - \nu,$$

et les variables $V_t = R(t+1) - R(t),\ t \geq 0$ qui sont indépendantes (et de même loi). En effet, on rajoute une part à $c(t)$ pour obtenir $c(t+1)$ lorsqu'on applique l'étape 2 de l'algorithme, ce qui arrive avec probabilité ν ; et sinon $c(t)$ et $c(t+1)$ ont le même nombre de parts.

Par la loi des grands nombres, $\frac{R(t)-R(0)}{t} = \frac{1}{t}\sum_{i=1}^{t} V_i \to \nu$ presque sûrement, et de même, $\frac{R(t)}{t} \to \nu$ presque sûrement lorsque t tend vers l'infini. Par conséquent, $R(t)$ tend presque sûrement vers l'infini. Or,

$$R(t) - E(t) = \text{nombre de parts non nulles de } c(t) \leq N,$$

donc $E(t)$ tend aussi presque sûrement vers l'infini : la plupart des nouvelles espèces qui apparaissent lors du processus disparaissent rapidement après.

Q4. Fixons une composition $c = (N_1, \ldots, N_l)$ telle que $N_1 = k$, et fixons une trajectoire $(c_0, \ldots, c_{t-1}, c)$ telle que $c(0) = c_0, \ldots, c(t) = c$. Alors, la valeur de N_1 au temps $t+1$ est :

— soit $k+1$, ce qui arrive si l'indice 1 n'est pas choisi à l'étape 1, s'il n'y a pas de nouvelle espèce qui apparaît, et si l'indice 1 est choisi dans l'étape 3 de l'algorithme. Ceci arrive avec probabilité

$$\frac{N-k}{N}\,\frac{k}{N}\,(1-\nu)$$

conditionnellement à l'événement $\{c(0) = c_0, \ldots, c(t) = c\}$.

— soit k, ce qui arrive si l'indice 1 n'est pas choisi à l'étape 1 et si une nouvelle espèce apparaît, ou si l'indice 1 est choisi à l'étape 1 et à l'étape 3, ou si l'indice 1 n'est choisi ni à l'étape 1 ni à l'étape 3. Conditionnellement à $\{c(0) =$

$c_0, \ldots, c(t) = c\}$, ceci arrive avec probabilité

$$\frac{N-k}{N}\,\nu + \left(\frac{k}{N}\right)^2 (1-\nu) + \left(\frac{N-k}{N}\right)^2 (1-\nu).$$

— soit $k-1$, ce qui arrive si l'indice 1 est choisi à l'étape 1 et si une nouvelle espèce apparaît, ou si l'indice 1 est choisi à l'étape 1 et n'est pas choisi à l'étape 3. Conditionnellement à $\{c(0) = c_0, \ldots, c(t) = c\}$, ceci se produit avec probabilité

$$\frac{k}{N}\,\nu + \frac{N-k}{N}\,\frac{k}{N}\,(1-\nu).$$

Notons $k_0, \ldots, k_{t-1}, k$ les tailles des premières parts des compositions $c_0, \ldots, c_{t-1}, c$. Si $N_1(t)$ est la taille de la première part de $c(t)$, alors on peut écrire :

$$\begin{aligned}
&\mathbb{P}[N_1(t+1) = k+1 \mid N_1(0) = k_0, \ldots, N_1(t-1) = k_{t-1},\ N_1(t) = k] \\
&= \sum \left(\begin{smallmatrix} \mathbb{P}[N_1(t+1)=k+1 \ \mid\ c(0)=c_0,\ldots,c(t)=c] \\ \times\, \mathbb{P}[c(0)=c_0,\ldots,c(t)=c \ \mid\ N_1(0)=k_0,\ldots,N_1(t)=k] \end{smallmatrix} \right) \\
&= \frac{N-k}{N}\,\frac{k}{N}\,(1-\nu) \sum \mathbb{P}[c(0) = c_0, \ldots, c(t) = c \mid N_1(0) = k_0, \ldots, N_1(t) = k] \\
&= \frac{N-k}{N}\,\frac{k}{N}\,(1-\nu),
\end{aligned}$$

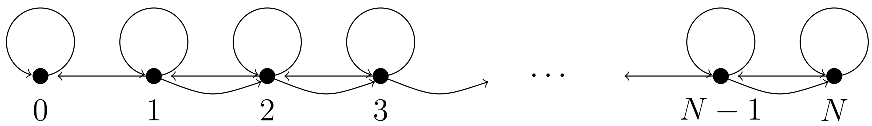
les sommes intermédiaires portant sur les suites de compositions $c_0, \ldots, c_{t-1}, c$ avec premières parts $k_0, \ldots, k_{t-1}, k$. On montre de même que

$$\mathbb{P}[N_1(t+1) = k \mid N_1(0) = k_0, \ldots, N_1(t) = k] = \frac{N-k}{N}\,\nu + \left(\frac{k}{N}\right)^2 (1-\nu) + \left(\frac{N-k}{N}\right)^2 (1-\nu)$$

$$\mathbb{P}[N_1(t+1) = k-1 \mid N_1(0) = k_0, \ldots, N_1(t) = k] = \frac{k}{N}\,\nu + \frac{N-k}{N}\,\frac{k}{N}\,(1-\nu),$$

de sorte que $(N_1(t))_{t\in\mathbb{N}}$ est une chaîne de Markov sur $[\![0, N]\!]$ avec les probabilités de transition indiquées ci-dessus.

Q5. Si $k \geq 1$, alors $P(k, k-1) \geq \frac{k}{N}\,\nu > 0$, et si $1 \leq k \leq N-1$, alors $P(k, k+1) = \frac{N-k}{N}\,\frac{k}{N}\,(1-\nu) > 0$. Par ailleurs, toutes les probabilités de transition $P(k, k)$ sont positives. Le graphe des transitions possibles de la chaîne $(N_1(t))_{t\in\mathbb{N}}$ est donc



En particulier, 0 est bien l'unique état absorbant. S'il y avait un autre état $k \geq 1$ récurrent pour la chaîne $(N_1(t))_{t\in\mathbb{N}}$, alors comme cet état k communique avec 0 (on a $P^k(k, 0) > 0$) et comme la relation de communication est une équivalence sur les états récurrents, on devrait avoir 0 qui communique avec k ; mais ce n'est pas le cas puisque 0 est un état absorbant. Donc, tous les autres états $k \in [\![1, N]\!]$ sont transients, et la chaîne $(N_1(t))_{t\in\mathbb{N}}$ n'occupe plus ses états pour t assez grand (presque sûrement), donc est presque sûrement stationnaire à l'état 0.

Q6. On calcule facilement :

$$\begin{aligned}
&\mathbb{E}[N_1(t+1) \mid N_1(t)=k] \\
&= k + \left(\frac{N-k}{N}\,\frac{k}{N}\,(1-\nu)\right) - \left(\frac{k}{N}\,\nu + \frac{N-k}{N}\,\frac{k}{N}\,(1-\nu)\right) \\
&= k\left(1-\frac{\nu}{N}\right).
\end{aligned}$$

On en déduit :

$$\begin{aligned}
\mathbb{E}[N_1(t+1)] &= \sum_{k=0}^{N} \mathbb{P}[N_1(t)=k]\,\mathbb{E}[N_1(t+1) \mid N_1(t)=k] \\
&= \left(1-\frac{\nu}{N}\right)\sum_{k=0}^{N} \mathbb{P}[N_1(t)=k]\,k = \left(1-\frac{\nu}{N}\right)\,\mathbb{E}[N_1(t)],
\end{aligned}$$

d'où la formule souhaitée par une récurrence immédiate sur t.

Q7. On a $\mathbb{E}[T] = \sum_{t=1}^{\infty} \mathbb{P}[T \geq t] = \sum_{t=0}^{\infty} \mathbb{P}[T > t]$, et par ailleurs, si $T > t$, alors $N_1(t) \geq 1$, donc

$$\mathbb{E}[N_1(t)] = \mathbb{E}[N_1(t)\,1_{N_1(t)\geq 1}] \geq \mathbb{P}[N_1(t) \geq 1] = \mathbb{P}[T > t].$$

On en déduit avec la formule de la question précédente :

$$\mathbb{E}[T] \leq \sum_{t=0}^{\infty} \mathbb{E}[N_1(t)] = N_1(0)\,\sum_{t=0}^{\infty}\left(1-\frac{\nu}{N}\right)^t = \frac{N_1(0)\,N}{\nu}.$$

La borne supérieure croît linéairement avec $N_1(0)$, la taille de l'espèce 1 au départ. C'est un comportement plausible pour $\mathbb{E}[T]$, puisque chaque descente $k \to k-1$ avec $k \in [\![1, N_1(0)]\!]$ augmente le temps nécessaire pour atteindre 0.

Q8. On ordonne les éléments de $\mathfrak{P}_4 = \{(4),(3,1),(2,2),(2,1,1),(1,1,1,1)\}$ comme dans l'énoncé. La matrice de transition de $(p(t))_{t\in\mathbb{N}}$ s'écrit alors :

$$P = \begin{pmatrix}
1-\nu & \nu & 0 & 0 & 0 \\
\frac{3(1-\nu)}{16} & \frac{5-3\nu}{8} & \frac{3(1-\nu)}{16} & \frac{3\nu}{4} & 0 \\
0 & \frac{1-\nu}{2} & \frac{1-\nu}{2} & \nu & 0 \\
0 & \frac{1-\nu}{4} & \frac{1-\nu}{8} & \frac{5-\nu}{8} & \frac{\nu}{2} \\
0 & 0 & 0 & \frac{3(1-\nu)}{4} & \frac{1+3\nu}{4}
\end{pmatrix}$$

Détaillons par exemple le calcul de $P((3,1),(3,1))$. Si $p(t)=(3,1)$, alors $c(t)$ est une composition avec une part de taille 3, une part de taille 1 et éventuellement des parts de taille 0. Pour que $p(t+1)=(3,1)$, il faut :

— soit qu'à l'étape 1 ce soit un individu dans l'espèce de taille 3 au temps t qui soit choisi, et que cet individu mort soit remplacé par un individu dans la même espèce ; ceci se produit avec probabilité

$$\frac{3}{4}\left(\frac{3}{4}(1-\nu)\right).$$

— soit qu'à l'étape 1 ce soit un individu dans l'espèce de taille 1 au temps t qui soit choisi, et que cet individu mort soit remplacé par un individu dans la même espèce, ou par un individu d'une nouvelle espèce ; ceci se produit avec probabilité

$$\frac{1}{4}\left(\frac{1}{4}(1-\nu)+\nu\right).$$

La somme de ces deux quantités donne bien $\frac{5-3\nu}{8}$. Les autres coefficients se calculent de la même façon.

Q9. L'entier N étant fixé, on peut passer en un nombre fini d'étapes de n'importe quelle partition $p = (N_1 \geq N_2 \geq \cdots \geq N_\ell)$ de taille N à la partition $(1, 1, \ldots, 1)$ en choisissant à chaque transition de diminuer d'une unité une part de p plus grande que 2, et en rajoutant une nouvelle part égale à 1 (apparition d'une nouvelle espèce). Réciproquement, partant de la partition $(1, 1, \ldots, 1)$, pour obtenir une partition arbitraire $p = (N_1 \geq N_2 \geq \cdots \geq N_\ell)$, on peut à chaque étape supprimer une part N_i de taille 1 et augmenter une part N_j d'une unité, jusqu'à obtenir la partition souhaitée. Ainsi, $(p(t))_{t\geq 0}$ est irréductible sur l'espace d'états $\mathfrak{P}_N$. Comme à chaque étape il y a une probabilité non nulle de ne pas changer la partition (ne pas rajouter d'espèces et choisir $i = j$), la période de chaque état est égale à 1, c'est-à-dire qu'on a apériodicité.

Q10. On a une chaîne de Markov irréductible, récurrente positive (l'espace d'états est fini) et apériodique, donc par le théorème de convergence vers la mesure stationnaire, il existe une unique mesure de probabilité $\mu_{N,\nu}$ sur $\mathfrak{P}_N$ telle que $\mu_{N,\nu} P = \mu_{N,\nu}$, et telle que $\mathbb{P}[p(t) \to p] \to_{t\to+\infty} \mu_{N,\nu}(p)$ pour toute partition $p \in \mathfrak{P}_N$.

Q11. La mesure $\rho_{4,\theta}$ s'écrit :

$$\rho_{4,\theta} = \frac{1}{Z_\theta} \left(6, 8\theta, 3\theta, 6\theta^2, \theta^3\right)$$

avec $Z_\theta = (\theta + 1)(\theta + 2)(\theta + 3) = \theta^3 + 6\theta^2 + 11\theta + 6$, donc c'est bien une mesure de probabilité sur $\mathfrak{P}_4$. En posant $\theta = \frac{4\nu}{1-\nu}$, on obtient

$$\rho_{4,\theta} = \frac{1}{Z_\theta} \left(6, \frac{32\nu}{1-\nu}, \frac{12\nu}{1-\nu}, \frac{96\nu^2}{(1-\nu)^2}, \frac{64\nu^3}{(1-\nu)^3}\right),$$

et on peut vérifier avec un logicel de calcul formel que la matrice de la question Q8. laisse invariant ce vecteur (pour l'action de la matrice à droite).

Q12. Un type cyclique $(p_1, p_2, \ldots, p_\ell) = p \in \mathfrak{P}_N$ étant fixé, considérons l'application $s_p : \mathfrak{S}(N) \to \mathfrak{S}(N)$ définie par :

$$s_p(\sigma) = (\sigma(1), \sigma(2), \ldots, \sigma(p_1)) \circ (\sigma(p_1+1), \ldots, \sigma(p_1+p_2)) \circ \cdots \circ (\sigma(N-p_\ell+1), \ldots, \sigma(N)).$$

L'application s_p est une surjection sur l'ensemble des permutations de type cyclique p. De plus, chaque permutation avec ce type cyclique est atteinte $\prod_{i=1}^N (m_i)!\, i^{m_i}$ fois, puisque dans une écriture

$$(x_1, x_2, \ldots, x_{p_1})(x_{p_1+1}, \ldots, x_{p_1+p_2}) \cdots (x_{N-p_\ell+1}, \ldots, x_N),$$

d'une permutation τ de type p comme produit de cycles disjoints :

- on peut permuter les cycles de même taille sans changer la permutation, ce qui représente $\prod_{i=1}^N (m_i)!$ permutations possibles des cycles ;
- on peut dans chaque cycle $(a_1, a_2, \ldots, a_i)$ de longueur i effectuer une permutation cyclique des entrées a_j sans changer le cycle, ce qui représente $\prod_{i=1}^N i^{m_i}$ possibilités.

On conclut que l'ensemble C_p des permutations de type cyclique p vérifie :

$$N! = \operatorname{card} \mathfrak{S}_N = \left(\prod_{i=1}^N (m_i)!\, i^{m_i}\right) \operatorname{card} C_p,$$

d'où la formule pour $\mathrm{card}\, C_p$.

Q13. On veut montrer que toute permutation $\sigma \in \mathfrak{S}_N$ s'écrit de manière unique sous la forme $\sigma = (1, n_1) \circ (2, n_2) \circ \cdots \circ (N, n_N)$ avec $n_i \leq i$ pour tout i. Pour des raisons de cardinalité, il suffit de montrer l'existence d'une telle écriture, ce que l'on peut faire par récurrence sur N. Si le résultat est vrai au rang $N-1$, étant donnée une permutation $\sigma \in \mathfrak{S}_N$, la permutation $\sigma' = \sigma \circ (N, \sigma^{-1}(N))$ envoie N sur N, donc elle appartient à $\mathfrak{S}_{N-1}$. Il existe donc des entiers $n_1, \ldots, n_{N-1}$ tels que $\sigma' = (1, n_1) \circ (2, n_2) \circ \cdots (N-1, n_{N-1})$, et en posant $n_N = \sigma^{-1}(N)$, on conclut que $\sigma = (1, n_1) \circ (2, n_2) \circ \cdots (N, n_N)$.

Montrons maintenant que $\ell(\sigma) = \mathrm{card}\,\{i \in [\![1, N]\!] \mid n_i = i\}$. On procède de nouveau par récurrence sur N, le cas $N = 1$ étant trivial. Si le résultat est vrai au rang $N-1$, considérons $\sigma = (1, n_1) \circ \cdots \circ (N, n_N)$ dans $\mathfrak{S}_N$. On a par hypothèse de récurrence

$$\begin{aligned}\sigma' &= (1, n_1) \circ (2, n_2) \circ \cdots \circ (N-1, n_{N-1}) \\ &= (a_{1,1}, \ldots, a_{1,p_1}) \circ (a_{2,1}, \ldots, a_{2,p_2}) \circ \cdots \circ (a_{l,1}, \ldots, a_{l,p_l})\end{aligned}$$

où $l = \mathrm{card}\,\{i \in [\![1, N-1]\!] \mid n_i = i\}$, et où les cycles du produit de la seconde ligne sont disjoints et recouvrent $[\![1, N-1]\!]$. Supposons d'abord $n_N = N$. Alors,

$$\sigma = (a_{1,1}, \ldots, a_{1,p_1}) \circ (a_{2,1}, \ldots, a_{2,p_2}) \circ \cdots \circ (a_{l,1}, \ldots, a_{l,p_l}) \circ (N),$$

donc $\ell(\sigma) = l + 1 = \mathrm{card}\,\{i \in [\![1, N-1]\!] \mid n_i = i\} + 1 = \mathrm{card}\,\{i \in [\![1, N]\!] \mid n_i = i\}$. Inversement, si $n_N < N$, on peut supposer à une permutation des cycles près que n_N appartient au premier cycle $(a_{1,1}, \ldots, a_{1,p_1})$ de σ', et à une permutation cyclique près des entrées de ce premier cycle, on peut même supposer $n_N = a_{1,p_1}$. Alors,

$$\sigma = \sigma' \circ (N, n_N) = (a_{1,1}, \ldots, a_{1,p_1}, N) \circ (a_{2,1}, \ldots, a_{2,p_2}) \circ \cdots \circ (a_{l,1}, \ldots, a_{l,p_l}),$$

donc $\ell(\sigma) = l = \mathrm{card}\,\{i \in [\![1, N-1]\!] \mid n_i = i\} = \mathrm{card}\,\{i \in [\![1, N]\!] \mid n_i = i\}$. On a donc aussi le résultat vrai au rang N.

Q14. Dans l'algorithme décrit pour la construction de σ_N, on a :

$$\mathbb{P}[m_i = j] = \frac{\theta^{1_{(j=i)}}}{\theta + i - 1}$$

pour tout $j \leq i$, donc la loi de σ_N s'écrit :

$$\mathbb{P}[\sigma_N = \sigma] = \frac{\theta^{\mathrm{card}\,\{i \,|\, m_i = i\}}}{\prod_{i=1}^{N} \theta + i - 1} = \frac{\theta^{\ell(\sigma)}}{\prod_{i=1}^{N} \theta + i - 1} = \rho_{N,\theta}(\sigma)$$

d'après la question précédente.

Q15. Le programme suivant simule une partition aléatoire sous la mesure d'Ewens $\rho_{N,\theta}$ (on utilise pour ce programme la classe **Permutation** de `Sage` ; dans `Python`, il faudrait reprogrammer quelques méthodes pour les permutations, par exemple le produit et le type cyclique).

```
def ewens(N,theta):
    perm = Permutation(range(1,N+1))
    for i in range(2,N+1):
        if random() > theta/(theta+i-1):
            j = floor(1+(i-1)*random())
            perm = perm * Permutation((i,j))
    return perm.cycle_type()
```

Un résultat de ce programme avec $N = 1000$ et $\theta = 10$ est la partition

```
[125, 110, 106, 98, 78, 74, 74, 41, 40, 38, 29, 25, 23, 21, 19, 14, 11, 9,
7, 6, 5, 5, 4, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1].
```