

0.1 Rappels et prérequis — matrices et suites

Avant de nous pencher sur ce qui sera le coeur de ce chapitre (les modèles matriciels de population), il est nécessaire d'introduire (ou rappeler) de quelques notions importantes sur les matrices et les suites.

Définition 0.1.1: Base de l'espace vectoriel \mathbb{R}^n

Soient e_1, \dots, e_n , n vecteurs de \mathbb{R}^n . On dit que $\mathcal{B} = \{e_1, \dots, e_n\}$ est une base de \mathbb{R}^n si elle vérifie l'une des assertions équivalentes suivantes

- quelque soit $u \in \mathbb{R}^n \setminus \{0_n\}$, il existe une unique façon d'écrire u comme combinaison linéaire des $(e_i)_{i=1, \dots, n}$. C'est à dire qu'il existe un unique choix de $(c_i)_{i=1, \dots, n} \in \mathbb{R}^n$ tel que $\sum_{i=1}^n c_i e_i = u$.
- aucun des e_i ne peut s'écrire comme combinaison linéaire des autres $(e_j)_{j \neq i}$ (i.e. $e_i \neq \sum_{j \neq i} c_j e_j$). On dit que les vecteurs $(e_i)_{i=1, \dots, n}$ sont linéairement indépendants.

Géométriquement, cela signifie que les $(e_i)_{i=1, \dots, n}$ forment un repère de l'espace \mathbb{R}^n . Ainsi, dans la base $(e_i)_{i=1, \dots, n}$, n'importe quel élément de \mathbb{R}^n est caractérisé par un unique vecteur de coordonnées.

On appelle **base canonique** de \mathbb{R}^n celle formée par les vecteurs

$$e_i = \underbrace{(0, \dots, 0, 1, 0, \dots, 0)}_{i\text{-ème position}}$$

De façon générale, si E est un espace vectoriel, le nombre minimal de vecteurs nécessaires pour l'engendrer s'appelle la **dimension** de E .

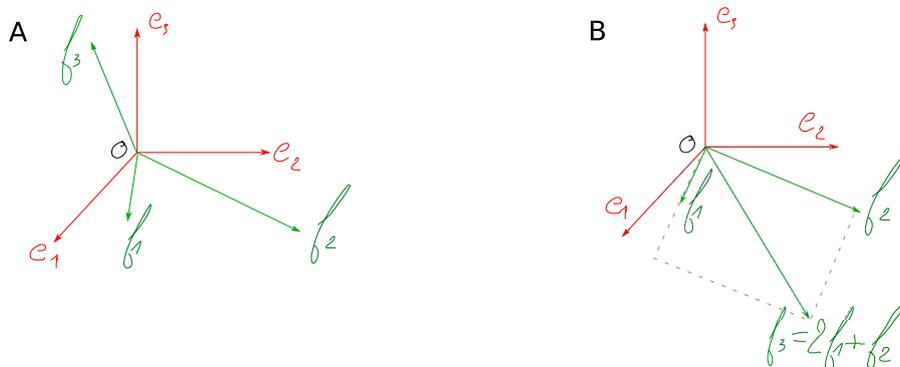


FIGURE 1 – Représentation graphique de la base canonique (e_1, e_2, e_3) de \mathbb{R}^3 et **A)** trois vecteurs f_1, f_2 et f_3 linéairement indépendants formant une autre base de \mathbb{R}^3 , **B)** trois vecteurs f_1, f_2 et f_3 linéairement dépendants et ne constituant donc pas une base de \mathbb{R}^3 .

Définition 0.1.2: Application linéaire

On dit qu'une fonction (ou "application") $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est linéaire si quelque soient $\lambda \in \mathbb{R}$ et $u, v \in \mathbb{R}^n$,

$$f(u + v) = f(u) + f(v) \text{ et } f(\lambda u) = \lambda f(u) \quad (1)$$

Ou encore, si $f(u + \lambda v) = f(u) + \lambda f(v)$ (ces définitions sont strictement équivalentes).

Exemples

1. L'application $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ définie pour tout $(x, y) \in \mathbb{R}^2$ par $f(x, y) = (y, 2x + y, 3y - 5x)$ est linéaire.
2. Si $A \in \mathbb{M}_{n,m}(\mathbb{R})$, l'application $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ définie pour tout $U \in \mathbb{R}^m$ par $f(U) = AU$ est linéaire. On pourrait en fait aller un peu plus loin et montrer que toute application linéaire peut s'écrire sous la forme matricielle $f(U) = AU$.
3. Toutes les fonctions de type *division* ($f(x) = 2/x$), *polynôme* ($f(x, y) = 2 + x * y$), *puissance* ($f(x) = 2^x$) ne sont pas linéaires.

0.1.1 Rappels sur les matrices

La notion de valeur propre/vecteur propre sera essentielle dans les différents chapitre de ce cours.

Définition 0.1.3: Valeur propre, vecteur propre

Soit $A \in \mathbb{M}_n(\mathbb{R})$. On dit que $\lambda \in \mathbb{R}$ est une **valeur propre** de A si il existe $W \in \mathbb{R}^n \setminus \{0_n\}$ tel que $AW = \lambda W$. On dit que W est un **vecteur propre** de A associé à la valeur propre λ .

Géométriquement, W est un vecteur propre de A si la multiplication de W par A ne fait que changer sa longueur, sans changer sa direction.

Proposition 0.1.1: Vecteur propre à gauche

Soit $A \in \mathbb{M}_n(\mathbb{R})$ et λ une valeur propre de A . Alors il existe $V \in \mathbb{R}^n$ tel que ${}^tVA = \lambda {}^tV$. On dit que V est un **vecteur propre à gauche** de A associé à λ . Notons que cela est équivalent à dire que V est un vecteur propre de tA .

Proposition 0.1.2: Multiple d'un vecteur propre

Soit $A \in \mathbb{M}_n(\mathbb{R})$, λ une valeur propre de A et W est un vecteur propre de A associé à λ . Alors quelque soit $c \in \mathbb{R} \setminus \{0\}$, cW est également un vecteur propre de A associé à λ . Autrement dit, tout multiple d'un vecteur propre est encore un vecteur propre (associé à la même valeur propre). Le résultat tient aussi pour les vecteurs propres à gauche.

Démonstration. La démonstration est laissée en exercice. □

Remarque 0.1.1. Parfois, il est possible de trouver, pour une même valeur propre λ , plusieurs vecteurs propres linéairement indépendants. Dans ce cas, toute combinaison linéaire de ces vecteurs propres est encore un vecteur propre. Autrement dit, si W_1, W_2, \dots, W_k sont des vecteurs propres de A associés à une valeur propre λ , alors $c_1W_1 + c_2W_2 + \dots + c_kW_k$ est encore un vecteur propre associé à λ , quelles que soient les constantes $(c_i)_{i=1, \dots, k}$.

Définition 0.1.4: Matrice diagonalisable

Une matrice $A \in \mathbb{M}_n(\mathbb{R})$ est dite diagonalisable si il existe une matrice $P \in \mathbb{M}_n(\mathbb{R})$ inversible et une matrice $D \in \mathbb{M}_n(\mathbb{R})$ diagonale telles que $A = PDP^{-1}$.

Proposition 0.1.3: Condition suffisante pour être diagonalisable

Soit $A \in \mathbb{M}_n(\mathbb{R})$. Si A possède n valeurs propres distinctes, alors A est diagonalisable.

Attention, il s'agit d'une condition suffisante, mais non nécessaire. Autrement dit, il existe des matrices diagonalisables qui possède moins de n valeurs propres distinctes.

Proposition 0.1.4: Condition nécessaire et suffisante pour être diagonalisable

$A \in \mathbb{M}_n(\mathbb{R})$ est diagonalisable si et seulement si il existe une base de \mathbb{R}^n formée de vecteurs propres de A .

Proposition 0.1.5: Matrice de changement de base

Soit $A \in \mathbb{M}_n(\mathbb{R})$ diagonalisable et $(W_i)_{i=1, \dots, n}$ une base de \mathbb{R}^n formée de vecteurs propres de A associés aux valeurs propres à $(\lambda_i)_{i=1, \dots, n}$ (non nécessairement distinctes, *i.e.* une même valeur peut apparaître plusieurs fois).

Alors la matrice P dont les colonnes sont formées par les vecteurs $(W_i)_{i=1, \dots, n}$ est inversible et $A = PDP^{-1}$ avec D une matrice diagonale dont les éléments diagonaux sont $d_{i,i} = \lambda_i$.

Exercice 1

Soit

$$A = \begin{pmatrix} 0 & 2 & 2 \\ -3 & 7 & 1 \\ -3 & 1 & 7 \end{pmatrix}, \quad W_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \text{ et } W_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \text{ et } W_3 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.$$

Vérifier que W_1 et W_2 sont des vecteurs propres (linéairement indépendants) de A associés à la même valeur propre, à déterminer. Vérifier que W_3 est un vecteur propre de A pour une valeur propre à déterminer.

Vérifier que $A = PDP^{-1}$ avec

$$P = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{pmatrix}, \quad P^{-1} = \frac{1}{4} \begin{pmatrix} -1 & -1 & 3 \\ -1 & 3 & -1 \\ 3 & -1 & -1 \end{pmatrix}, \quad \text{et } D = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

0.1.2 Rappels sur les suites

Définition 0.1.5: Suite

Une **suite** est une fonction définie sur \mathbb{N} . On note généralement une suite sous la forme $(u_n)_{n \in \mathbb{N}}$ ou, quand il n'y a pas d'ambiguïté, simplement $(u_n)_n$, ou (u_n) . Les éléments de la suite s'appellent ses **termes**. Si les termes appartiennent à un même ensemble E on parle de suite dans E .

Parfois, notamment quand on s'intéresse à une suite de vecteurs, il est plus commode d'utiliser la notation $u(n)$. Ainsi, le n -ième d'une suite dans \mathbb{R}^m pourra être noté $u(n) = (u_1(n), \dots, u_m(n))$ plutôt que $u_n = (u_{1,n}, \dots, u_{m,n})$ pour qu'il n'y ait pas d'ambiguïté entre les indices.

Attention aux notations : $(u_n)_{n \in \mathbb{N}}$ désigne la suite, et u_n désigne le n -ième terme de la suite $(u_n)_n$, pour un n donné.

Définition 0.1.6: Convergence et divergence dans \mathbb{R}

Soit $(u_n)_{n \in \mathbb{N}}$ une suite dans \mathbb{R} .

- On dit que $(u_n)_n$ converge si il existe $\ell \in \mathbb{R}$ tel que pour tout $\epsilon > 0$ il existe $N \in \mathbb{N}$ tel que si $n \geq N$, alors

$$|u_n - \ell| < \epsilon.$$

On dit que ℓ est la limite de $(u_n)_n$ et note $\lim_{n \rightarrow +\infty} u_n = \ell$.

- On dit que $(u_n)_n$ diverge si elle ne converge pas.

Remarque 0.1.2. Si pour tout $A \in \mathbb{R}$ il existe N_A tel que $n \geq N_A \Rightarrow u_n > A$ (resp. $u_n < -A$) on dit que $(u_n)_n$ diverge vers $+\infty$ (resp. $-\infty$) et on note $\lim_{n \rightarrow +\infty} u_n = +\infty$ (resp. $-\infty$).

Une suite divergente ne diverge pas nécessairement vers $\pm\infty$. Par exemple, la suite de terme général $u_n = (-1)^n$ diverge.

Exercice 2: Quelques calculs de limites

Les suites suivantes sont-elles convergentes ou divergentes ?

1. a) $u_n = \frac{n}{n+1}$, b) $u_n = \frac{n^2}{n+1}$, c) $u_n = (-1)^n e^{-n}$, d) $u_n = \frac{n(n+1)}{n^2-1}$.
2. $u_{n+1} = u_n + \frac{9}{10^{n+1}}$ avec $u_0 = 9$.

Soit $(u_n)_{n \in \mathbb{N}}$ une suite de \mathbb{R} . On définit $(u_{2n})_n$ et $(u_{2n+1})_n$ les suites des termes pairs et impairs de $(u_n)_n$. Les affirmations suivantes sont-elles vraies ?

1. Si $(u_n)_n$ converge vers $l \in \mathbb{R}$ alors $(u_{2n})_n$ et $(u_{2n+1})_n$ convergent vers l .
2. $(u_{2n})_n$ et $(u_{2n+1})_n$ sont convergentes, alors $(u_n)_n$ est convergente.
3. $(u_{2n})_n$ et $(u_{2n+1})_n$ sont convergent vers la même limite $l \in \mathbb{R}$, alors $(u_n)_n$ converge vers l .

Définition 0.1.7: Suite géométrique

La suite réelle $(u_n)_{n \in \mathbb{N}}$ est dite géométrique si il existe $a \in \mathbb{R}$ tel que pour tout $n \in \mathbb{N}$, $u_{n+1} = au_n$. Le paramètre a s'appelle la *raison* de la suite.

Si $u_0 \neq 0$ alors : si $|a| > 1$ (resp. $|a| < 1$), la suite est divergente (resp. convergente). Elle converge si $a = 1$ et diverge si $a = -1$.

Définition 0.1.8: Convergence et divergence dans \mathbb{R}^m

Soit $(u_n)_{n \in \mathbb{N}}$ une suite dans \mathbb{R}^m , pour un $m \in \mathbb{N}^*$. Si on note $u_n = (u_{1,n}, \dots, u_{m,n})$. On dit que $(u_n)_n$ converge vers $L = (\ell_1, \dots, \ell_m)$ si pour tout $1 \leq i \leq m$, $(u_{i,n})_n$ converge vers ℓ_i . Autrement dit, si chaque composante de la suite de vecteurs $(u_n)_n$ converge vers la composante correspondante du vecteur L .

Dans le cas contraire on dit qu'elle diverge.

Exercice 3: Suite de Fibonacci

Au XIII^{ème} siècle, Fibonacci a considéré la situation suivante

- Un couple de lapins adultes produit chaque mois un couple de jeunes lapins.
- Un couple de jeunes lapins est adulte au début de son deuxième mois.
- Les lapins ne meurent jamais.

On note $(u_n)_n$ la suite telle que u_n est le nombre de couples de lapins adultes au n -ième mois et $u_0 = u_1 = 1$.

1. Déterminer la relation définissant u_{n+1} en fonction de u_n et u_{n-1} .
2. On pose $R_n = \frac{u_{n+1}}{u_n}$. Que représente R_n ?
3. Que vaut R_0 ? Montrez que la double récurrence sur $(u_n)_n$ se transforme en récurrence simple sur $(R_n)_n$ (c'est à dire que l'on peut exprimer R_{n+1} en fonction de uniquement R_n)
4. Montrer que $(R_n)_n$ admet un unique point fixe positif (on dit que a est un point fixe de $(R_n)_n$ si $R_n = a \Rightarrow R_{n+1} = a$).
5. On admet que la suite $(R_n)_n$ converge vers son point fixe. La suite (u_n) est-elle convergente ?

Chapitre 1

Modèles matriciels de population

1.1 Introduction

La **dynamique des populations** est un domaine des mathématiques visant à étudier l'évolution temporelle du nombre d'individus dans une *population* au sens large : généralement animale ou végétale, mais il peut s'agir de populations cellulaires, bactériennes, etc. L'objectif d'un modèle en dynamique des populations est généralement de mieux comprendre le fonctionnement d'un phénomène (modèles mécanistiques) ou de faire des prédictions sur son état futur (modèles prédictifs).

Vous avez déjà rencontré un champ important de la dynamique des populations : les **systèmes différentiels autonomes**, pour lesquels l'évolution d'une population X (X peut être un vecteur) est donnée par une équation différentielle $X'(t) = F(X(t))$. Le temps t s'écoule de façon continue, on parle donc de modèles *continus*, et la solution X est une *fonction*.

Vous avez notamment étudié des **systèmes différentiels linéaires (SDL)**, c'est à dire de la forme $X'(t) = MX(t)$ où M est une matrice, pour lesquels vous avez discuté de l'influence des valeurs propres de M sur l'évolution de la solution X .

Dans ce cours, nous étudierons un formalisme différent, malgré quelques similarités avec les SDL : les **modèles matriciels de population (MMP)**. A vrai dire, les termes *matriciel* et *linéaire* sont pratiquement interchangeable, on pourrait tout aussi bien parler de *modèles linéaires de population*. La différence avec les SDL réside plutôt dans la description de la variable temporelle t . Ici, on ne s'intéresse plus à tous les instants d'un *intervalle temporel*, mais uniquement à certains *instants isolés* $t_0, t_1, t_2 \dots$. On parle de modèles **discrets**, dont la solution est une **suite** $(X(t_k))_{k \in \mathbb{N}}$. Les MMP constituent une classe particulière de modèles discrets : ceux de la forme $X(t_{k+1}) = AX(t_k)$.

	Modèle continu	Modèle discret
Définition générale	<p>Système différentiel (autonome)</p> $X'(t) = F(X(t))$	<p>Suite récurrente</p> $X(t_{k+1}) = G(X(t_k))$
Linéaire	<p>SDL</p> $X'(t) = MX(t)$	<p>MMP</p> $X(t_{k+1}) = AX(t_k)$

Remarque 1.1.1. Lien entre $X' = MX$ et $X(n+1) = AX(n)$ ¹

La solution de $X'(t) = MX(t)$ de condition initiale $X(0) = X_0$ est $X(t) = e^{tM} X_0$.

En particulier, pour $n \in \mathbb{N}^*$,

$$X(n+1) = e^{(n+1)M} X_0 \text{ et } X(n) = e^{nM} X_0.$$

On en déduit que

$$X(n+1) = e^M X(n).$$

Autrement dit, si l'on ne regarde la solution de $X' = MX$ qu'en un ensemble discret de temps, on regarde en réalité le MMP $X(n+1) = AX(n)$ avec $A = e^M$.

La réciproque n'est pas vraie : tout MMP ne peut pas s'écrire comme la solution d'une équation différentielle en un ensemble discret de temps. Formellement, il existe des matrices $A \in \mathbb{M}_n(\mathbb{R})$ qui ne peuvent pas s'écrire sous la forme e^M .

1.1.1 Trois hypothèses de travail

1) L'unité de temps : Les MMP servent presque toujours à décrire l'évolution d'une population à intervalles de temps réguliers (jour, année,...). Si on cherche à décrire l'évolution de X d'une année sur l'autre, on choisira naturellement l'année comme unité de temps. Ainsi, pour tout $k \in \mathbb{N}$, $t_k = k$ et l'instant $t = k$ correspond à k années depuis l'instant initial $t = 0$.

2) Modèles centrés sur les femelles : En général (pas toujours mais presque) dans les modèles de population (discrets ou continus, en dimension 1 ou supérieure, linéaires ou non), seules les femelles sont prises en compte. Deux arguments majeurs :

1. le nombre de descendants d'un mâle est plus difficile à estimer expérimentalement car la paternité des individus n'est en général pas connue,
2. on peut supposer que les mâles sont suffisamment nombreux pour que toutes les femelles fertiles soient fécondées : le facteur limitant est donc le nombre de femelles. La question ne se pose pas lorsque la reproduction n'est pas sexuées (bactéries...).

3) L'hypothèse "pre-breeding" : Les MMP sont bien adaptés à des populations qui se reproduisent à intervalles de temps périodiques (typiquement, annuellement, comme beaucoup de plantes et animaux). Le modèle s'écrira différemment selon que l'on considère que $X(k)$ correspond à la population juste avant (pre-breeding), ou juste après (post-breeding) la période des naissances. Dans ce cours, on considèrera l'hypothèse de pre-breeding. L'ensemble des naissances ayant lieu entre les instants k et $k+1$ ont donc lieu "juste après" l'instant k (voir Figure 1.1).

1. Les modèles discrets $X(t_{k+1}) = f(X(t_k))$ sont à priori beaucoup plus simples à utiliser que les continus $Y'(t) = g(Y(t))$. En effet, on peut toujours calculer explicitement tous les termes de la suite $X(t_k)$ alors que pour connaître $Y(t)$ il faut résoudre une équation différentielle, ce qui n'est pas toujours possible. Il ne sont pas pour autant plus simple à étudier, parfois au contraire. Vous pouvez par exemple consulter (sur wikipedia ou autre) et comparer la *suite logistique* (discrète) et le *modèle de Verhulst* (continu). Pour le second, on a une forme explicite de la solution et sa convergence vers un équilibre. Pour la suite logistique on ne dispose pas d'une écriture explicite de la solution et on observe de plus, pour certaines valeurs de paramètres, des comportements complexes (chaotiques) <http://experiences.math.cnrs.fr/Iterations-de-1-application.html>.

1.2 Approche naïve en dimension 1 : le modèle de Malthus

Exemple 1.2.1

Considérons l'évolution d'une population de mésanges (femelles uniquement) d'année en année, en partant d'une année de référence (l'année 0). On constate qu'une mésange vivant au printemps de l'année k a une probabilité s de survivre au moins jusqu'au printemps de l'année $k + 1$ et que, en moyenne, en un an chaque mésange pond f œufs de femelles viables (voir remarque ci-dessous). Naturellement, on utilise l'année comme unité de temps. L'évolution de la population de mésanges entre une année k donnée et l'année $k + 1$ est alors décrite par

$$m(k+1) = \underbrace{sm(k)}_{\text{ survivants }} + \underbrace{fm(k)}_{\text{ naissances }} = (s+f)n(k) = am(k).$$

Remarque 1.2.1. Sur le taux de naissance f : Un "œuf viable" signifie que l'œuf donnera un oisillon qui survivra au moins un an. En effet, $m(k)$ est le nombre d'oiseau au temps k . Si un oisillon né durant l'année k meure durant cette même année, il ne sera pas présent à l'année $k+1$. Autrement dit $f = s_0b$ où b est le nombre moyen d'œufs pondus par une mésange en un an et s_0 le taux de survie lors de la première année (potentiellement différent de s).

Attention : sous l'hypothèse de pre-breeding, les individus les plus jeunes présents au temps $k + 1$ sont nés "juste après" le temps k . Au temps $k + 1$ ils ont donc (quasiment) 1 an.

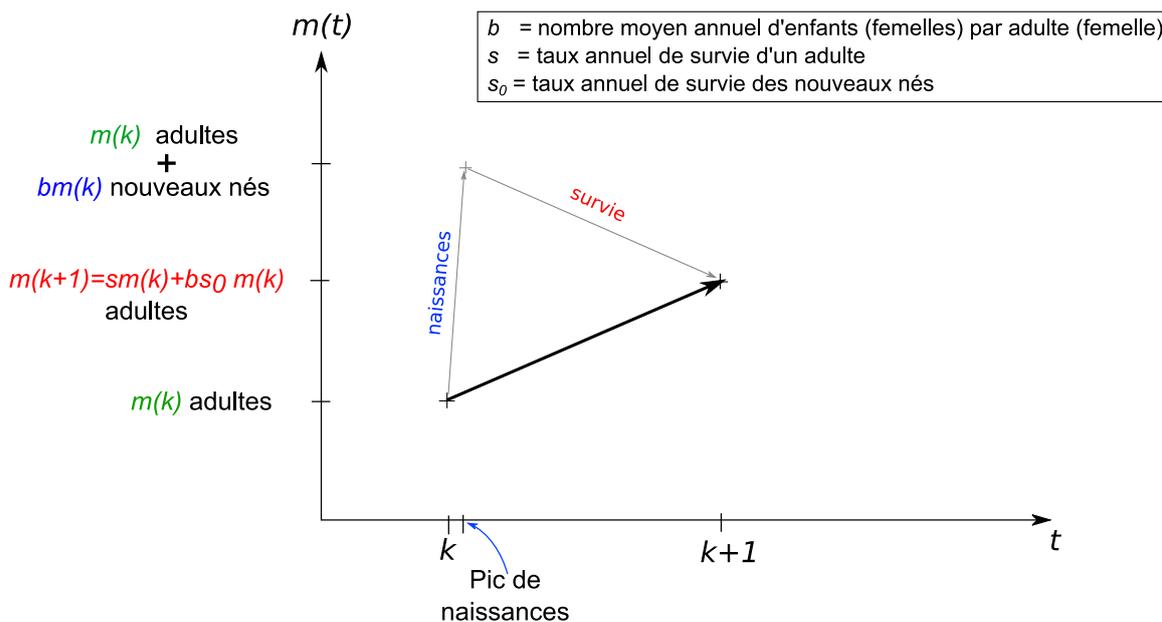


FIGURE 1.1 – Illustration du taux de naissance f sous l'hypothèse *pre-breeding*. Autre fait important : les individus nés juste après le temps k on désormais quasiment 1 an et sont donc comptés parmi les adultes (ici adulte = 1 an ou plus).

Ce modèle correspond à la situation biologique où une population multiplie sa taille par un même facteur a à intervalles de temps périodiques (c'est un MMP où a est une matrice de

taille 1). Ce modèle particulièrement simple s'appelle le *modèle géométrique de Malthus*, qui s'écrit

$$m(k+1) = am(k), \quad m(0) = m_0 \text{ donné} \quad (1.1)$$

où m_0 (resp. $m(t)$) correspond à la taille de la population à l'instant initial (resp. à l'instant t) et a est appelé taux de croissance géométrique de la population.

On reconnaît ici une suite géométrique de raison a . On montre alors (par récurrence immédiate) que $m(k) = a^k m_0$. On en déduit que, quel que soit $m_0 > 0$:

si $a \in [0, 1[$: la population s'éteint ($m(k) \rightarrow 0$ quand $k \rightarrow +\infty$),

si $a=1$: la population ne change pas ($m(k) = m(0)$ pour tout k),

si $a > 1$: la population croît à l'infini à vitesse exponentielle ($\lim_{k \rightarrow +\infty} m(k) = +\infty$),

si $a < 0$: on retrouve des conclusions similaires (essayez par vous-même), qui ne sont pas pertinentes d'un point de vue de la modélisation car elles conduisent à des populations de taille négative.

Quelques limites des modèles linéaires L'avantage des MMP (ici en dimension 1, mais cette remarque restera valide pour les dimensions supérieures) est qu'ils sont relativement simple à concevoir et à analyser. En revanche ils présentent plusieurs défauts dont il faut avoir conscience.

1. Il n'y a **pas d'interaction** (compétition, coopération,...) entre les individus.
2. Ce sont des modèles **déterministes**. Ils ne permettent pas de prendre en compte les effets aléatoires qui surviennent nécessairement en pratique, particulièrement importants lorsque la population est petite (risque d'extinction, proba de rencontre d'un partenaire sexuel,...). La taille de la population n doit davantage être vue comme l'espérance de la taille réelle de la population.

Par conséquent, les MMP ne peuvent décrire qu'une variété assez restreinte de types d'évolution de la population. Toutefois nous verrons que, utilisés dans un cadre adapté, on peut tout de même extraire des informations intéressantes de ces modèles.

Exemple 1.2.2: Un exemple non linéaire et un peu plus subtile

Bien souvent, la règle d'évolution change en fonction de la taille m de la population. Ce phénomène ne peut pas être décrit par un modèle matriciel (*i.e.* linéaire). Pour rendre le modèle plus précis, on pourrait par exemple prendre en compte la saturation du milieu en considérant un taux de croissance dépendant de la valeur de m . C'est par exemple le cas du modèle suivant, appelé *suite logistique*

$$m(k+1) = \underbrace{\left(a - \frac{m(k)}{K}\right)}_{\tilde{a}(m(k))} (a-1) m(k), \quad m(t_0) = m_0 \quad (1.2)$$

Sans entrer dans les détails, le taux de croissance $\tilde{a}(m)$ de l'équation (1.2) décroît avec m pour modéliser l'effet de la compétition entre individus. Plus précisément, ici

- si m est petit, alors $\tilde{a}(m) \approx a$ (on retrouve presque Malthus),
- si $m = K$, alors $\tilde{a}(m) = 1$ (la population cesse de croître).

Le paramètre K , appelé capacité de charge du milieu, correspond donc au nombre d'individus que le milieu peut accueillir. Au delà de K , le nombre de mort dépasse le nombre de naissances.

Pertinence du modèle de matriciel (Malthus) Si l'on se place dans un **environnement très riche en ressource**, alors la compétition inter-individus est beaucoup moins importante et le modèle linéaire de Malthus est raisonnable. Exemples :

- Croissance de populations bactériennes ou cellulaires dans une boîte de pétri.
- Dans l'exemple de la population de mésanges, si l'environnement est très grand et riche en ressource, alors la capacité de charge K du modèle logistique (1.2) est très grande et le modèle de Malthus est raisonnable durant un certain temps.

Exercice 4: Application numérique

Écrire un programme pour calculer les premiers termes de la suite définie par (1.2) pour une population initiale petite par rapport à K (par exemple $K = 50000$, $n_0 = 5$ et $a = 1.2$). Tracer l'évolution de n en fonction de k . Comparer avec la solution du modèle de Malthus (1.1).

1.3 Modèles matriciels de dimension supérieure — Populations structurées

Dans la section précédente, la population décrite était **homogène** : tous les individus sont considérés identiques les uns aux autres et obéissent exactement aux mêmes règles. Pour la plupart des systèmes, cette hypothèse trouve rapidement ses limites. Nous allons voir que les MMP s'adaptent facilement aux situations où la population peut être découpée en classes **hétérogènes**.

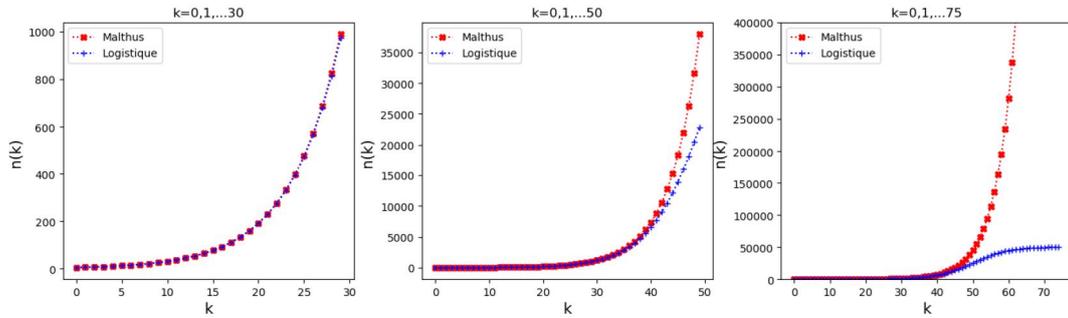


FIGURE 1.2 – Comparaison du modèle de Malthus et de la suite logistique pour les paramètres $K = 50000$, $n_0 = 5$ et $a = 1.2$. On constate une bonne adéquation des deux modèles tant que $k \leq 50$, mais en temps long la suite logistique converge vers K tandis que le modèle de Malthus connaît une croissance exponentielle de taux géométrique a .

1.3.1 Définition générale

Définition 1.3.1: Matrice positive

Une matrice $A = (a_{i,j})_{i \in \{1, \dots, m\}, j \in \{1, \dots, n\}} \in \mathbb{M}_{m,n}(\mathbb{R})$ est dite *positive*, et on note $A \geq 0$, si $a_{i,j} \geq 0$ pour tout couples (i, j) . Autrement dit si $A \in \mathbb{M}_{m,n}(\mathbb{R}_+)$.

Si les inégalités sont stricte ($a_{i,j} > 0$) on dit que A est *strictement positive* ($A > 0$).

Définition 1.3.2: Modèle matriciel de population (MMP)

Un **modèle matriciel de population** est défini par une expression de la forme

$$X(k+1) = AX(k), \quad X(0) = X_0 \in \mathbb{R}_+^2 \text{ donné} \quad (1.3)$$

où $A \in \mathbb{M}_n(\mathbb{R}_+)$ est une matrice carrée positive appelée **matrice de projection**.

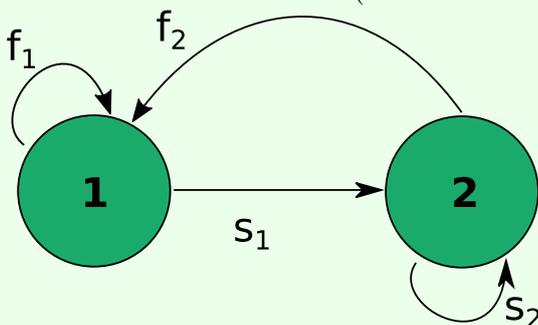
$X = (x_1, x_2, \dots, x_n)$ est un vecteur dont les différentes composantes x_i sont appelées **classes** de la population.

Exemple d'introduction

Exemple 1.3.1: Structure en âge

Reprenons l'exemple de la population de mésanges (femelles uniquement) de la section précédente, mais cette fois distinguons les mésanges juvéniles ($1 \leq \text{âge} < 2$) des mésanges âgées ($\text{âge} \geq 2$). On note

- $X_1(k)$ le nombre de mésanges juvéniles et $X_2(k)$ le nombre de mésanges âgées.
- s_1 et s_2 la probabilité de survie d'une mésange de chacune des catégories.
- b_1 et b_2 le nombre moyen d'œufs (femelles) engendrés par une mésange juvénile et adulte chaque année.
- s_0 la probabilité de survie des oisillons ($\text{âge} < 1$).
- f_1 et f_2 le nombre d'œufs femelles viables (survivant au moins un an).



Exercice 5

1. Après avoir précisé quelle unité de temps est la plus naturelle à adopter, exprimer f_1 et f_2 en fonction de b_1 , b_2 et s_0 .
2. Écrire le modèle matriciel correspondant sous la forme $X(k+1) = AX(k)$, où $X = (X_1, X_2)$ et A est une matrice à déterminer.
3. Essayer d'en déduire la relation liant $X(0)$ à $X(k)$ pour un $k \in \mathbb{N}^*$ quelconque.

Correction: (succincte)

L'année comme unité de temps car on dispose de données biologiques pour des classes qui durent un an ($\text{âge des individus de la première classe} = 1$ an et durée de la première classe = 1 an). Pour $i = 1, 2$, $f_i = s_0 b_i$.

$$A = \begin{pmatrix} f_1 & f_2 \\ s_1 & s_2 \end{pmatrix}.$$

On peut en déduire que $X(1) = AX(0)$ puis (par récurrence immédiate) que $X(k) = A^k X(0)$, on reverra ça dans la section suivante (Proposition 1.3.1).

1.3.2 Calcul de la solution exacte

Définition 1.3.3: Puissance de matrice

Soit $A = (a_{i,j})_{i,j \in \{1, \dots, n\}} \in \mathbb{M}_n(\mathbb{R})$. On appelle puissance k -ième de A et on note A^k la matrice définie par $A^k = \underbrace{A \times A \times \dots \times A}_{k \text{ fois}}$.

En général les coefficients de A^k sont notés $(a_{i,j}^{(k)})_{1 \leq i,j \leq n}$

Proposition 1.3.1: Solution explicite de (1.3)

Soit $A \in \mathbb{M}_n(\mathbb{R})$, $X_0 \in \mathbb{R}^n$ et X solution de (1.3), alors pour tout $k \in \mathbb{N}$, $X(k) = A^k X_0$.

Démonstration. Preuve par récurrence. a) Initiation : $X(1) = AX(0) = AX_0$. b) Hérédité : on suppose que $X(j) = A^j X_0$ pour un $j \in \mathbb{N}$. Alors $X(j+1) = AX(j) = AA^j X_0 = A^{j+1} X_0$. Donc la proposition est vraie. \square

Proposition 1.3.2

Soit $A = (a_{i,j})_{1 \leq i,j \leq n} \in \mathbb{M}_n(\mathbb{R})$.

- si A est diagonale, alors pour tout $k \geq 0$, $a_{i,j}^{(k)} = a_{i,j}^k$,
- si A est diagonalisable ($A = PDP^{-1}$), alors $A^k = PD^k P^{-1}$.

Démonstration. exercice \square

Corollaire 1.3.3

Si $A \in \mathbb{M}_n(\mathbb{R})$ est diagonalisable ($A = PDP^{-1}$), $X_0 \in \mathbb{R}_+^n$ et X solution de (1.3), alors pour tout $k \in \mathbb{N}$, $X(k) = PD^k P^{-1} X_0$.

Proposition 1.3.4: Inversion de matrices de $M_2(\mathbb{R})$

Soient $a, b, c, d \in \mathbb{R}$. Alors la matrice

$$P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

est inversible si et seulement si $ad - bc \neq 0$ et alors

$$P^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Exercice 6: Mésanges charbonnières

On reprend l'exemple 1.3.1 des mésanges (femelles uniquement) et on considère les hypothèses suivantes :

Seules 10% des mésanges parviennent à l'âge d'un an. Une mésange juvénile ($1 \leq \text{âge} < 2$) a une probabilité annuelle de survie de 0.9 puis à partir de 2 ans la probabilité de survie annuelle passe à 0.4. On suppose qu'une mésange juvénile pond en moyenne 20 œufs par an, pour les mésanges plus âgées ce nombre passe à 16. La moitié des œufs contiennent des femelles. À l'instant initial, il y a α mésanges juvéniles et β âgées.

1. En déduire les valeurs de s_0, s_1, s_2, f_1 et f_2 et écrire, en précisant quelle est l'unité de temps, le modèle matriciel de population associé sous la forme $X(k+1) = AX(k)$.
2. De quelle taille est le vecteur X ? que représentent chacune de ses composantes? que vaut $X(0)$?
3. Calculer $X(1)$ et $X(2)$.
4. Quelle est la trace de A ? Quel est le déterminant de A ? Quel est la relation entre trace, déterminant et valeurs propres?
5. Trouver les valeurs propres de A et un vecteur propre associé à chacune d'elles. Indication : il faut trouver les valeurs propres $\lambda_1 = 1,6$ et $\lambda_2 = -0,2$ et, pour les vecteurs propres, n'importe quel multiple de $v_1 = (4, 3)$ et $v_2 = (-2, 3)$.
6. Vérifier que

$$\begin{pmatrix} 4 & -2 \\ 3 & 3 \end{pmatrix}^{-1} = \frac{1}{18} \begin{pmatrix} 3 & 2 \\ -3 & 4 \end{pmatrix}$$

7. Justifier que A est diagonalisable puis donner P, D et P^{-1} telles que $A = PDP^{-1}$ avec D diagonale.
8. Soit $k \geq 0$, écrire D^k en fonction de λ_1 et λ_2 .
9. Exprimer $X(k)$ en fonction de P, P^{-1}, D^k et $X(0)$.

Correction: (succincte)

1) En choisissant l'année comme unité de temps, $s_0 = 0.1$, $f_1 = 1$, $f_2 = 0.8$, $s_1 = 0.9$, $f_2 = 0.4$ et $A = \begin{pmatrix} f_1 & f_2 \\ s_1 & s_2 \end{pmatrix} = \begin{pmatrix} 1 & 0.8 \\ 0.9 & 0.4 \end{pmatrix}$. On a alors $X(k+1) = AX(k)$. 2) X est de taille 2, X_1 est le nombre de mésanges juvéniles et X_2 le nombre de mésanges âgées, $X(0) = (\alpha, \beta)$. 3) Il faut calculer $X(1) = AX(0)$, $X(2) = AX(1) = A^2X(0)$. 4) $Tr(A) = 1.4$, $det(A) = -0.32$. Or trace=somme des valeurs propres et déterminant=produit des valeurs propres. 5) Si on note λ_1 et λ_2 les deux valeurs propres de A , la question précédente induit que $\lambda_1 + \lambda_2 = 1.4$ et $\lambda_1\lambda_2 = -0.32$. En injectant $\lambda_2 = -0.32/\lambda_1$ dans la première équation on obtient un polynôme de degré 2 dont les racines sont les valeurs propres de A . Autre méthode équivalente : calculer les racines de $det(A - \lambda I)$. Puis on trouve les vecteurs propres en résolvant un système de taille 2 (tout multiple de v_1 ou de v_2 est une bonne solution). 6) le produit donne bien la matrice identité. 7) $A = PDP^{-1}$ où les colonnes de P sont les vecteurs propres v_1 et v_2 . La question 6 nous donne sa matrice inverse. D est la matrice diagonale avec λ_1 et λ_2 sur sa diagonale. 8) D^k est la matrice diagonale avec λ_1^k et λ_2^k sur sa diagonale. 9) $X(k) = PD^kP^{-1}X(0)$.
Remarque/analyse : Si on effectue le calcul de la dernière question, on montre que

$$X(k) = \frac{1}{18} \begin{pmatrix} 4(3\alpha + 2\beta)\lambda_1^k + 2(-3\alpha + 4\beta)\lambda_2^k \\ 3(3\alpha + 2\beta)\lambda_1^k + 3(-3\alpha + 4\beta)\lambda_2^k \end{pmatrix} \quad (1.4)$$

$$= \lambda_1^k \frac{3\alpha + 2\beta}{18} \begin{pmatrix} 4 \\ 3 \end{pmatrix} + \lambda_2^k \frac{(-3\alpha + 4\beta)}{18} \begin{pmatrix} -2 \\ 3 \end{pmatrix}. \quad (1.5)$$

Or, $\lambda_2 = -0.8$, donc $\lambda_2^k \xrightarrow[k \rightarrow +\infty]{} 0$ alors que $\lambda_1^k \xrightarrow[k \rightarrow +\infty]{} +\infty$. Autrement dit, lorsque k est grand,

$$X(k) \approx \lambda_1^k \frac{3\alpha + 2\beta}{18} \begin{pmatrix} 4 \\ 3 \end{pmatrix}.$$

En résumé :

- a) A est diagonalisable et sa diagonalisation permet d'obtenir une expression explicite de $X(k)$ qui reste relativement simple.
- b) On observe que lorsque k devient grand, chaque composante de $X(k)$ croît à la vitesse de λ_1^k .
- c) la solution X finit par s'aligner avec le vecteur propre correspondant à la plus grande valeur propre λ_1 . En particulier, ici cela signifie que, bien que la population juvénile $X_1(k)$ et la population adulte $X_2(k)$ tendent vers l'infini, la population juvénile est 4/3 fois plus grande que la population adulte (quelque soit la population initiale (α, β) !).

On va voir que ce comportement (points **b** et **c**) aurait pu être prédit !

Conclusion de la section Pour calculer la valeur de la solution X après k unités de temps, il suffit donc de calculer la puissance k -ième de la matrice A . L'inconvénient est que calcul de $A^k \in \mathbb{M}_n(\mathbb{R})$ peut être particulièrement laborieux si k et n sont grands. Le Corollaire 1.3.3 assure que lorsque la matrice A est diagonalisable, ce calcul est grandement facilité.

Nous verrons que l'on peut parfois déterminer d'autres propriétés intéressantes d'un MMP.

La session suivant introduit quelques notions importantes.

1.3.3 Le cycle de vie — correspondance graphe \Leftrightarrow matrice

Formalisons les notions entrevues jusqu'ici pour des systèmes de dimensions quelconques.

Le **cycle de vie** correspond à l'ensemble des flux entre les différentes classes d'une population. Celles-ci peuvent être de natures diverses en fonction du système étudié et du type de classes utilisées (par âge, par taille, par niveau de développement, par lieu...). On y trouve typiquement des taux de survie, de reproduction, de croissance... mais tout peut être envisagé!

L'objectif de ce chapitre est de convenir d'une représentation graphique standard, permettant de décrire de manière unique n'importe quel système susceptible d'être modélisé par un modèle matriciel de population. Une telle représentation a deux intérêts notoires

1. Il est plus commode de visualiser la structure d'une population structurée en classe à partir d'un graphe qu'à partir de sa matrice de projection.
2. Dans le chapitre suivant, nous introduirons les notions d'*irréductibilité* et de *primitivité* d'une matrice. Ces deux propriétés peuvent également être exprimées dans le langage de la théorie des graphes et sont, en général, plus simples à démontrer sous cette forme.

Graphe du cycle de vie

Le **graphe du cycle de vie** est construit de la façon suivante

- Les *classes* sont représentées par les *nœuds* (ou *sommets*) du graphe,
- Les *flux* entre deux classes sont représentés par des *arêtes orientées* (ou arcs) et pondérées reliant ces classes. Un arc peut très bien aller d'un nœud à lui-même.

Correspondance graphe \Leftrightarrow matrice

Définition 1.3.4: Matrice d'adjacence

À tout graphe \mathcal{G} à n nœuds, on fait correspondre une unique matrice $A \in \mathbb{M}_n(\mathbb{R}_+)$, appelée **matrice d'adjacence** de \mathcal{G} dont le coefficient $a_{i,j}$ est égal à au poids de l'arc allant du nœud j au nœud i , et égal à 0 en l'absence d'arc.

Voici un exemple abstrait pour illustrer cette construction.

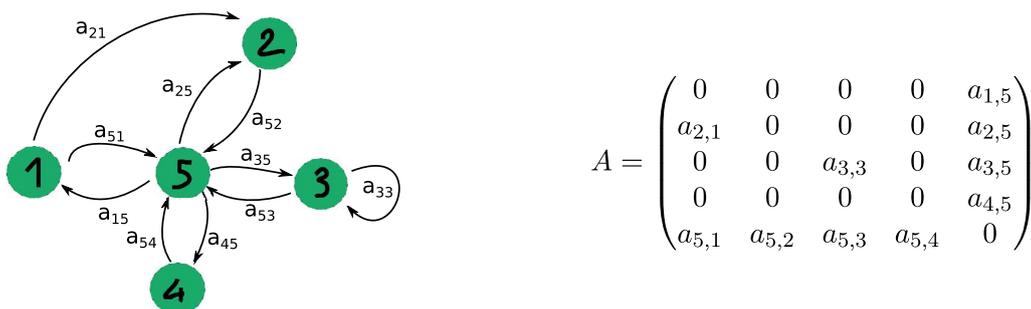


FIGURE 1.3 – Un graphe à 5 nœuds et sa matrice d'adjacence.

1.3.4 Le modèle historique de Leslie : Structuration en âges

Le plus célèbre des modèles matriciels est le modèle de Leslie.

- La population étudiée est divisée en n classes d'âge de même durée (pas nécessairement en années). On note i la classes des individus d'âge compris entre i et $i + 1$.
- L'unité de temps correspond à la durée de ces classes.
- Pour chaque classe i , on considère
 1. La **probabilité de survie** s_i pour qu'un individu de la classe i survive sur une unité de temps. Ainsi que s_0 la probabilité qu'un nouveau-né survive jusqu'à la première classe.
 2. Le **taux de reproduction** b_i correspondant au nombre moyen de descendants engendrés par chaque individu lors du pic de naissances (breeding).
 3. Le **taux de fertilité viable** f_i correspondant au nombre moyen de descendants **viabes** engendrés par chaque individu lors du pic de naissances. Les descendants viables sont ceux qui survivront jusqu'à l'âge 1. On a donc $f_i = b_i s_0$.

Remarque 1.3.1. *Rappel Dans cette définition, comme dans tout ce cours, on ne parle que des femelles, et on considère l'hypothèse de pre-breeding.*

Dans la formulation classique du modèle de Leslie, la dernière classe représente l'âge maximal pouvant être atteint par un individu. Ainsi, à chaque itération, tous les individus de la classe n meurent. Éventuellement, la dernière classe peut être utilisée pour regrouper tous les individus de plus d'un certain âge, ayant tous une même probabilité de survie s_n .

Le modèle de Leslie s'exprime donc toujours sous la forme générale illustrée par la Figure 1.4.

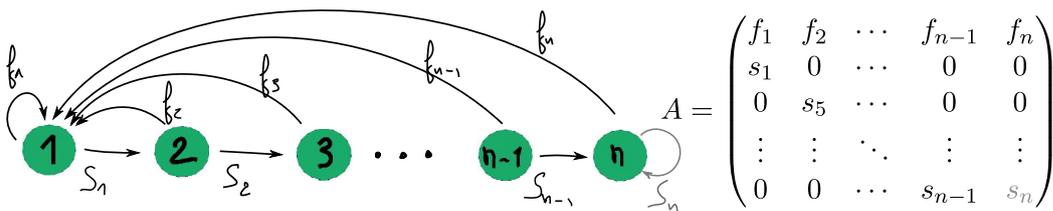


FIGURE 1.4 – Forme générale du modèle de Leslie.

1.3.5 Exemples de structurations différentes

Dans le modèle de Leslie et tous les exemples vus précédemment, les individus changent nécessairement de classe à chaque avancée du temps. Plutôt que de classer les individus par âge, il peut être plus intéressant de les classer par stade de développement, par taille, par lieux... ou encore de combiner plusieurs de ces critères. Dans ces situations, un individu peut rester dans une même classe durant plusieurs unités de temps, il peut également évoluer vers une classe quelconque. Il n'y a donc *a priori* aucune forme de matrice de projection qui ne puisse être envisagée. On présente ici deux exemples.

Classification en stades de développements

Exemple 1.3.2

Un exemple tiré d'un article de recherche (*Barot 2000, Journal of tropical Ecology*) portant sur une variété de palmiers dans la Station d'écologie de Lamto (Côte d'Ivoire).

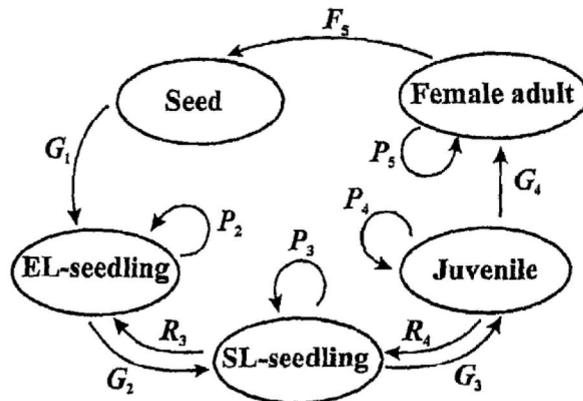


Figure 1. Life cycle diagram corresponding to the five-stage female-based matrix model used for *Borassus aethiopum* population. F_i , fecundity in stage i ; G_i , probability in stage i of survival and recruitment to the next stage; P_i , probability in stage i of survival in the same stage; R_i , probability in stage i of survival and retrogression to the precedent stage.

Légende : Seed = graine, EL-seedling = plantule à feuille entière, SL-seedling = plantule à feuilles fendues, juvénile, femelle adulte.

Par rapport au modèle de Leslie, on voit ici la possibilité de rester dans une même classe, voir de retourner dans une classe précédente. Formellement, la probabilité de survie s_i (Cf Section 1.3.4) est décomposée en $s_i = G_i + P_i + R_i$ avec

- G_i la probabilité pour un individu de la classe i de survivre sur une unité de temps et de progresser vers la classe $i + 1$.
- P_i la probabilité pour un individu de la classe i de survivre sur une unité de temps et de rester dans la classe i .
- R_i la probabilité pour un individu de la classe i de survivre sur une unité de temps et de régresser vers la classe $i - 1$.

Exercice 7: Écriture du modèle

Écrire la matrice de projection associée au cycle de vie du modèle en rangeant les 5 classes dans l'ordre : (Seed,EL,SL,Juvénile,Adulte).

Classification en stade de développement et en espace

On considère une plante à fruits (fictive) qui pousse dans un archipel formé de 3 îles. Après une phase d'un an sous forme de graine, cette espèce traverse 2 stades de développement "plante juvénile" et "plante adulte". Seule les plantes adultes sont fertiles, et les graines qu'elles sèment peuvent passer d'une île à l'autre grâce aux oiseaux qui les ingèrent. On obtient alors un cycle de vie de la forme

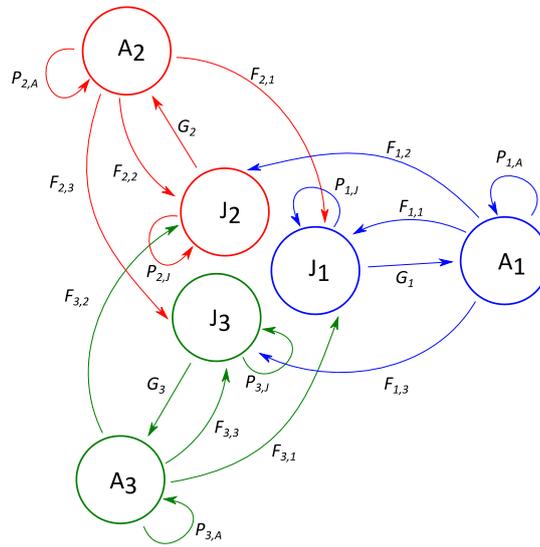


FIGURE 1.5 – Un exemple de cycle de vie structuré en stade de développement et en espace.

Exercice 8: Écriture du modèle

Écrire la matrice de projection associée au cycle de vie de la Figure 1.5 en rangeant les 6 classes dans l'ordre : $(J_1, J_2, J_3, A_1, A_2, A_3)$. Donner le sens biologique des familles de paramètres F , P et G .

1.3.6 Irréductibilité, primitivité

Définition 1.3.5: chemin, circuit

Soit \mathcal{G} un graphe à n nœuds.

On appelle **chemin** une suite d'arcs consécutifs.

La **longueur** d'un chemin correspond au nombre d'arcs qui le composent.

Un chemin reliant un nœud de \mathcal{G} à lui-même s'appelle un **circuit** (ou cycle orienté).

Irréductibilité

Définition 1.3.6: Matrice irréductible

On dit que $A = (a_{i,j})_{1 \leq i,j \leq n} \in \mathbb{M}_n(\mathbb{R})$ est irréductible si pour tout couple (i, j) il existe une puissance $k \in \mathbb{N}^*$ tel que $a_{i,j}^{(k)} \neq 0$.

Heureusement, le caractère irréductible d'une matrice est beaucoup plus simple à vérifier en utilisant son graphe.

Proposition 1.3.5: Lecture de l'irréductibilité sur le graphe

Une matrice positive est irréductible si et seulement si, dans le graphe associé, pour tout couple (i, j) , il existe un chemin reliant le nœud i au nœud j .
En particulier, pour tout $1 \leq i \leq n$ il existe un chemin partant de i et revenant à i , c'est à dire un circuit.

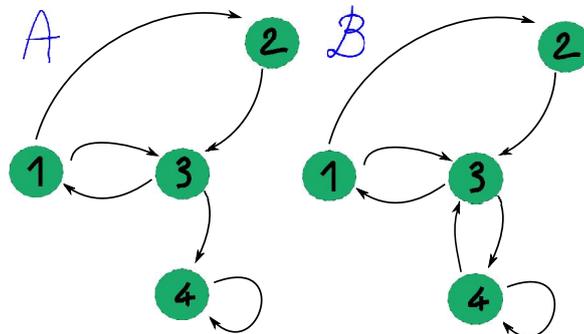


FIGURE 1.6 – Le graphe A n'est pas irréductible, mais le graphe B l'est.

Primitivité

Définition 1.3.7: Matrice primitive

On dit que $A = (a_{i,j})_{1 \leq i,j \leq n} \in \mathbb{M}_n(\mathbb{R})$ est primitive si il existe $k \in \mathbb{N}^*$ tel que pour tout couple (i, j) , $a_{i,j}^{(k)} \neq 0$.
En particulier, une matrice primitive est irréductible.

Là encore, la primitivité d'une matrice peut être déduite de l'étude de son graphe.

Proposition 1.3.6: Lecture de la primitivité sur le graphe

Une matrice positive est primitive si et seulement si le plus grand dénominateur commun (PGCD) des longueurs des circuits de son graphe est 1 (on dit qu'elle est apériodique) et elle est irréductible.

Une autre interprétation graphique, équivalente, de la primitivité est qu'il existe $k \in \mathbb{N}^*$ tel que pour tout nœuds i, j et pour toute longueur $\tilde{k} > k$, il existe un chemin de longueur \tilde{k} allant de i à j (éventuellement en repassant plusieurs fois par les mêmes nœuds).

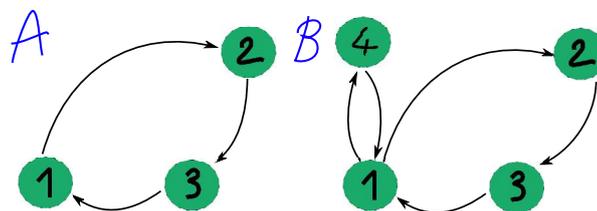


FIGURE 1.7 – Le graphe A n'est pas primitif (bien qu'irréductible), mais le graphe B l'est.

Remarquons que les Propositions 1.3.5 et 1.3.6 induisent que l'irréductibilité et la primitivité d'un graphe dépendent uniquement de la position coefficients non nuls de la matrice de projection (indépendamment de leur valeur).

Rappel : Le PGCD d'une famille d'entiers naturels est le plus grand entier naturel qui divise tous les éléments de cette famille. Exemple : le *PGCD* de $(8, 12, 36)$ est 4. Le moyen le plus rustique de déterminer le *PGCD* d'une famille d'entiers naturels est donc de faire la liste des diviseurs de chacun d'eux. Outre ce calcul, faire la liste des circuits d'un graphe peut être laborieux. On peut toutefois retenir quelques cas particuliers où le *PGCD* vaut forcément 1 :

- si il existe un arc reliant un nœud i à lui même (car on a un circuit de longueur 1).
- si on peut trouver deux circuits respectivement de longueurs d et $d + 1$.
- si on peut trouver deux circuits respectivement de longueurs d_1 et d_2 premières entrelles (*i.e.* tels que $PGCD(d_1, d_2) = 1$).

Remarque 1.3.2. Attention, pour une famille d'entiers (a_1, \dots, a_n) , si il existe a_i, a_j premiers entre eux (pas de diviseur commun > 1), alors $PGCD(a_1, \dots, a_n) = 1$. Mais la réciproque est fausse. Par exemple : le *PGCD* de la famille $(6, 10, 15)$ est 1 bien que ces entiers ne sont pas deux à deux premiers entre eux.

Exercice 9: Irréductibilité, primitivité

Les matrices suivantes sont-elles irréductibles? primitives?

$$a) \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad b) \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad c) \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}, \quad d) \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

Correction

- a) Pas irréductible car pas de chemins de la classe 2 vers la classe 1 (ni vers 3, ni de 4 vers 1 ou 3, mais un seul contre exemple suffit). Donc pas primitive, bien qu'apériodique (le *pgcd* des longueurs de circuits est bien 1).
- b) Primitive (irréductible et apériodique, on peut par exemple trouver un circuit de longueur 3 ($1 \rightarrow 4 \rightarrow 5 \rightarrow 1$) et un de longueur 2 ($2 \rightarrow 3 \rightarrow 2$)).
- c) Irréductible mais non primitive puisqu'il n'y a que des circuits de longueur 2^k , pour $k \geq 1$, donc de *pgcd*=2.
- d) Primitive (irréductible et on a un circuit de longueur 1).

1.4 Comportement asymptotique et le théorème de Perron-Frobenius

Il peut être intéressant de prévoir l'évolution de X en temps long (lorsque k tend vers l'infini). On parle du **comportement asymptotique** de X . Le théorème de Perron-Frobenius rend cela est possible lorsque la matrice de projection A est *primitive*.

Théorème 1.4.1: Perron-Frobenius

Soit $A \in \mathbb{M}_n(\mathbb{R})$ une matrice positive et primitive. Alors il existe une valeur propre λ de A telle que

- $\lambda \in \mathbb{R}$ et $\lambda > 0$
- si μ est une autre valeur propre de A , alors $\lambda > |\mu|$ (λ est la plus grande valeur propre de A , on dit qu'elle est **dominante**)
- il existe un vecteur propre $W = (w_1, \dots, w_n)$ associé à λ qui est strictement positif ($\forall i, w_i > 0$) et dont la somme des composantes vérifie $w_1 + \dots + w_n = 1$. On l'appelle **vecteur de Perron** de A ,
- l'ensemble des vecteurs propres associés à λ sont les multiples de W ,
- A admet un vecteur propre à gauche $V = (v_1, \dots, v_n)$ associé à λ qui est strictement positif ($\forall i, v_i > 0$) et tel que ${}^t V W = 1$ (i.e. $v_1 w_1 + \dots + v_n w_n = 1$). On l'appelle vecteur des **valeurs reproductives**.
- tous les vecteurs propres à gauche associés à λ sont des multiples de V ,

Remarque 1.4.1. Si $\tilde{W} = (\tilde{w}_1, \dots, \tilde{w}_n)$ est un vecteur propre de A associé à λ , il suffit de le diviser par $\tilde{w}_1 + \dots + \tilde{w}_n$ pour retrouver le vecteur de Perron W .

En pratique, pour déterminer V on cherche un vecteur propre de la transposée de A (${}^t A$) associé à la valeur propre λ , puis on le divise par $\tilde{w}_1 \tilde{v}_1 + \dots + \tilde{w}_n \tilde{v}_n$.

Dans l'étude des MMP, le théorème de Perron-Frobenius permet de montrer les résultats suivants, que nous ne démontrerons pas ici.

Proposition 1.4.2: Conséquence du théorème de Perron-Frobenius : distribution et croissance asymptotique

Soit un MMP de matrice de projection $A \in \mathbb{M}_n(\mathbb{R})$ positive et primitive. soit X sa solution de condition initiale $X(0) \in \mathbb{R}_+^n \setminus 0$. On note λ sa valeur propre dominante, $W = (w_1, \dots, w_n)$ son vecteur de Perron et V son vecteur des valeurs reproductives (garantis par le théorème de Perron-Frobenius).

Enfin, on note $N(k) = x_1(k) + \dots + x_n(k)$ la somme des composantes de X (donc $N(k)$ donne la taille totale de la population au temps k). Alors

1) Distribution asymptotique

$$\lim_{k \rightarrow +\infty} \frac{X(k)}{N(k)} = W. \quad (1.6)$$

2) Croissance asymptotique

$$\lim_{k \rightarrow +\infty} \frac{N(k+1)}{N(k)} = \lambda. \quad (1.7)$$

On peut être plus précis. Posons $c = {}^t V X(0) = v_1 x_1(0) + \dots + v_n x_n(0)$, alors

$$\lim_{k \rightarrow +\infty} \frac{N(k)}{c \lambda^k} = 1. \quad (1.8)$$

Un peu d'analyse de ces résultats :

Point 1 l'équation (1.6) signifie que la solution X du MMP s'aligne avec la direction du vecteur W . Si l'on réécrit l'équation (1.6) composante par composante, plutôt que sous forme vectorielle, on a pour tout $i = 1, \dots, n$, $\lim_{k \rightarrow +\infty} \frac{x_i(k)}{N(k)} = w_i$. Autrement dit

la fraction de la population occupée par la classe i converge vers w_i .

On peut notamment remarquer que cette distribution W est complètement indépendante de la distribution initiale $X(0)$.

Point 2 • Lorsque k est grand, le taux de croissance géométrique de N , défini par $N(k+1)/N(k)$, tend vers λ . Dans un écosystème, à moins que celui-ci n'ait récemment connu une perturbation, on s'attend à ce que λ soit proche de 1, ce qui signifie que la taille de la population évolue peu. En particulier :

— Si $\lambda > 1$, $\lim_{k \rightarrow +\infty} N(k) = +\infty$ avec un taux de croissance géométrique λ .

— Si $\lambda = 1$ alors $\lim_{k \rightarrow +\infty} N(k) = c$, et donc $\lim_{k \rightarrow +\infty} X(k) = cW$.

— Si $\lambda < 1$, $\lim_{k \rightarrow +\infty} N(k) = 0$ avec un taux de décroissance géométrique λ .

• De part la définition de $c = v_1 x_1(0) + \dots + v_n x_n(0)$, on voit que les valeurs reproductives v_i jouent un rôle de pondération de l'influence des conditions initiales $x_i(0)$ de chacune des classes dans la taille de la population en temps long. Plus v_i est grand, plus la taille initiale de la classe i a d'influence sur la taille de la population totale à long terme.

Conclusion Même si ce n'est pas parfaitement rigoureux mathématiquement, pour mieux visualiser et retenir la Proposition 1.4.2, on pourra retenir que lorsque k devient grand

$$\underbrace{N(k) \approx {}^t V X(0) \lambda^k}_{\text{croissance expo. à taux } \lambda} \text{ et } \underbrace{X(k) \approx N(k) W}_{\text{distribution stable}} \quad (1.9)$$

Remarque 1.4.2. Attention, l'hypothèse de primitivité de A est essentielle. Si A est irréductible mais non primitive, on peut avoir plusieurs (pgcd des longueurs des circuits du graphe du cycle de vie) valeurs propres de module $|\lambda|$ et des proportions $x_i(k)/N(k)$ qui oscillent périodiquement. Si A n'est même pas irréductible, on peut avoir des comportements assez divers (extinction de certaines classes, valeur reproductive nulle,...) que nous n'étudierons pas.

Exemple : Dans l'exercice 6, $\lambda = 1.6$ et le vecteur de Perron est $W = (4, 3)/(4 + 3) = ((4/7, 3/7)$.

Régime transitoire

L'évolution de la solution X durant le régime transitoire (entre l'instant initial et l'alignement de X avec le vecteur de Perron W) peut être complexe. En particulier, si A possède des valeurs propres négatives ou complexes ($\in \mathbb{C}$), on peut observer des phénomènes d'oscillation avant que la solution ne s'aligne avec le vecteur de Perron. (\Rightarrow nous aurons l'occasion de voir ça lors d'un TP).

En pratique, on retiendra simplement que la convergence de la solution X vers sa distribution asymptotique (alignement avec le vecteur de Perron et croissance en λ^k) est d'autant plus rapide que le ratio

$$r = \lambda / |\lambda_2|, \quad (1.10)$$

appelé coefficient d'amortissement est grand, où λ_2 est la deuxième plus grande valeur propre de A (en module, puisqu'il peut s'agir d'une valeur propre négative ou complexe)

1.4.1 Preuve du comportement asymptotique de X dans le cas diagonalisable **HORS PROGRAMME**

L'idée dans le cas général est la suivante.

1. Le Théorème de Perron-Frobenius justifie l'existence d'une valeur propre λ dominante réelle et strictement positive, de son vecteur de Perron et de son vecteur des valeurs reproductives.
2. À chaque multiplication par A , la direction portée par W sera la plus fortement étirée (si $\lambda > 1$), ou la moins contractée (si $\lambda < 1$),
3. Lorsque k devient grand, l'influence des autres valeurs propres est écrasée par l'influence de λ : la solution X finit par s'aligner avec W et évoluer à taux géométrique λ .

On donne ici une preuve que, si $A \in \mathbb{M}_n(\mathbb{R}_+)$ est primitive ET diagonalisable, $X(k)$ s'aligne avec w lorsque k tend vers l'infini (la preuve dans le cas où A est non-diagonalisable n'est pas très différente, mais un peu trop technique pour ce cours).

Puisque A est diagonalisable, elle admet donc n vecteurs propres $(W_i)_{i=1,\dots,n}$ qui forment une base de \mathbb{R}^n , respectivement associés aux valeurs propres $(\lambda_i)_{i=1,\dots,n}$. Le théorème de Perron-Frobenius assure que l'une de ces valeurs propres (disons λ_1) est dominante et strictement positive. Autrement dit $\lambda_1 > 0$ et $\forall i \geq 2, \lambda_1 > |\lambda_i|$.

Soit $X_0 \in \mathbb{R}^n$ une condition initiale dont toutes les composantes sont strictement positive², et $k \geq 1$. Alors que vaut $A^k X_0$?

Puisque les $(W_i)_{i=1,\dots,n}$ forment une base de \mathbb{R}^n , on peut écrire X_0 dans cette base, on a alors

$$X_0 = c_1 W_1 + c_2 W_2 + \dots + c_n W_n.$$

Ainsi

$$A X_0 = c_1 \lambda_1 W_1 + c_2 \lambda_2 W_2 + \dots + c_n \lambda_n W_n$$

puis, en itérant k fois,

$$A^k X_0 = c_1 \lambda_1^k W_1 + c_2 \lambda_2^k W_2 + \dots + c_n \lambda_n^k W_n.$$

Maintenant, en divisant cette expression par λ_1^k , on obtient

$$\frac{A^k}{\lambda_1^k} X_0 = c_1 W_1 + c_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k W_2 + \dots + c_n \left(\frac{\lambda_n}{\lambda_1}\right)^k W_n.$$

Or, puisque $\lambda_1 > |\lambda_i|$, on sait que

$$\lim_{k \rightarrow +\infty} \left(\frac{\lambda_i}{\lambda_1}\right)^k = 0.$$

2. **Pour aller plus loin :** Dans cette preuve on voit qu'il est nécessaire que $c_1 > 0$. On peut montrer que c'est forcément le cas. Avec les notations du théorème, le projecteur de Perron (projecteur sur $\text{vect}(W_1)$ parallèlement à $\text{vect}(W_2, \dots, W_n)$) est la matrice $\Pi_p = W^t V / {}^t V W$ ($= W^t V$ si on choisit de renormaliser V et W convenablement). Or W et V ont des composantes strictement positives. Donc tous les éléments de Π_p sont strictement positifs. Ainsi, dès que X_0 a ses composantes ≥ 0 avec au moins une composante non nulle, alors $\Pi_p X_0$ a toutes ses composantes strictement positives.

On en conclue que

$$\lim_{k \rightarrow +\infty} \frac{A^k}{\lambda_1^k} X_0 = c_1 W_1.$$

1.4.2 Sensibilité aux paramètres

Donnons un dernier résultat sur l'étude mathématique des MMP. On se pose la question suivante : à quel point la dynamique du modèle est-elle sensible à une perturbation des coefficients de la matrice A ? Cette question présente au moins deux intérêts : 1) on peut vouloir agir sur l'évolution du système en intervenant (physiquement) sur celui-ci, ce qui a pour effet de modifier A , 2) les paramètres $(a_{i,j})$ doivent être estimés expérimentalement et la valeur estimée n'est jamais exacte, elle dépend des expériences réalisées, on se demande alors si l'imprécision des estimations impacte sensiblement les résultats du modèle.

On a vu que lorsque A est primitive, sa valeur propre dominante λ joue un rôle essentiel dans l'évolution de la population structurée X . Une bonne question est donc : comment varie λ lorsqu'un des $a_{i,j}$ varie. La réponse est apportée par la proposition suivante

Proposition 1.4.3: Sensibilité aux paramètres

Avec les notations du Théorème de Perron-Frobenius 1.4.1, en notant W le vecteur de Perron de A et V le vecteur propre à gauche associé

$$\frac{\partial \lambda}{\partial a_{i,j}} = v_i w_j \quad (1.11)$$

Les différents paramètres $a_{i,j}$ pouvant avoir des ordre de grandeurs différents, on regarde généralement une normalisation de cette dérivée partielle

Définition 1.4.1: Élasticité d'un paramètre

On appelle élasticité du paramètre $a_{i,j}$ la quantité

$$e(a_{i,j}) = \frac{a_{i,j}}{\lambda} \frac{\partial \lambda}{\partial a_{i,j}} = \frac{a_{i,j} v_i w_j}{\lambda}.$$

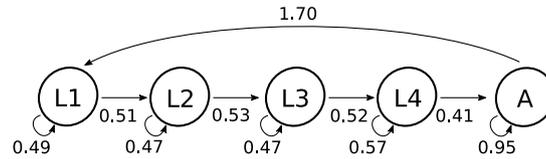
On pourrait montrer que la somme des élasticités vaut 1 ($\sum_{i,j=1}^n e(a_{i,j}) = 1$) et l'élasticité $e(a_{i,j})$ peut être interprétée comme une sensibilité proportionnelle de la valeur propre dominante à une variation du coefficient $a_{i,j}$, en ce sens que si $a_{i,j}$ varie de $n\%$ pour un petit n petit, alors λ augmente d'environ $(n \times e(a_{i,j}))\%$.

Exercice 10: Pucerons et Luzerne (Tenhumberg et al., 2009)

On s'intéresse à l'évolution d'une population de pucerons dans un champ de luzerne au printemps (seules les femelles sont considérées).

On considère 5 classes : les larves de type 1, 2, 3 et 4 (L_1, L_2, L_3, L_4) et les adultes (A). Seules les pucerons adultes peuvent se reproduire. Un jour après sa naissance, un puceron devient une larve L_1 .

On dispose des probabilités journalières pour qu'une larve de type i reste dans la classe i ou évolue vers le type supérieur (L_{i+1} ou adulte). En moyenne un puceron adulte donne naissance à 1.7 pucerons viables par jour. On obtient alors le graphe de cycle de vie suivant :



Partie théorique

1. Ecrire la matrice de projection M associée.
2. Quel est la signification de chacun des arcs sortant du nœud L_1 ? La somme de leur poids est égale à 1, ce qui n'est pas le cas pour le nœud L_3 par exemple, comment peut-on l'interpréter ?
3. Quelle est la signification de chacun des arcs sortant du nœud A ?
4. On note $X = (x_1, x_2, x_3, x_4, x_5)$ la solution associée au MMP de matrice de projection M et de condition initiale $X(0) = X_0 \in \mathbb{R}_+^5 \setminus \{0\}$. Donner l'expression de $X(k)$ en fonction de M et $X(0)$.
5. Justifier l'existence de λ : une valeur propre de M qui est dominante réelle et strictement positive, ainsi que d'un vecteur propre W associé à λ dont toutes les composantes sont positives et dont la somme vaut 1.

Correction

1.

$$M = \begin{pmatrix} 0.49 & 0 & 0 & 0 & 1.7 \\ 0.51 & 0.47 & 0 & 0 & 0 \\ 0 & 0.53 & 0.47 & 0 & 0 \\ 0 & 0 & 0.52 & 0.57 & 0 \\ 0 & 0 & 0 & 0.41 & 0.95 \end{pmatrix}$$

2. L'arc allant de la classe $L1$ vers elle-même, de poids 0.49, correspond à la probabilité journalière qu'un puceron de classe $L1$ survive et reste dans cette classe. L'arc allant de la classe $L1$ vers la classe $L2$, de poids 0.51, correspond à la probabilité journalière qu'un puceron de classe $L1$ survive et évolue vers la classe $L2$.

Pour la classe $L3$, la somme des deux arcs sortant vaut 0.99, donc une fraction 0.01 des pucerons de classe $L3$ (i.e 1%) meure chaque jour.

3. L'arc allant de A à $L1$ correspond au nombre moyen de pucerons femelles viables (qui survivront jusqu'à atteindre la classe $L1$) engendré par chaque puceron adulte chaque jour.

4. $X(k) = M^k X(0)$

5. La matrice M est positive, et elle est primitive. En effet, son graphe est irréductible et le *pgcd* des longueurs de ses circuits est 1 (par exemple, parce que le circuit $L1 \rightarrow L1$ est de longueur 1). Par conséquent, d'après le théorème de Perron-Frobenius, la valeur propre dominante λ et le vecteur W recherché (le vecteur de Perron de M) existent.

Exercice 11: Pucerons et Luzerne, partie II

Partie numérique (sur Python)

- Définir la matrice M et calculer ses valeurs et vecteurs propres à l'aide de la fonction `eig`. Identifier sa valeur propre dominante et son vecteur de Perron W .
- On note $N(k) = \sum_i x_i(k)$ la taille de la population totale au temps k . Sans calcul ni simulation : que peut-on dire sur l'évolution de $N(k)$ en temps long ?
- Sans calcul ni simulation : que peut-on dire de l'évolution de $X(k)/N(k)$ en temps long ?
- Définir la condition initiale $X_0=[100,0,0,0,0]$ et le nombre d'itérations à réaliser `n_max=30`. Puis calculer $X(k)$ pour $k = 1, \dots, n_max$ à l'aide d'une boucle `for` (stocker les résultats dans un tableau).
- Tracer l'évolution de $N(k)$ en fonction de k . Tracer l'évolution de $N(k+1)/N(k)$ en fonction de k , que représente ce ratio et vers quelle valeur converge-t-il ?
- Tracer l'évolution de $x_i(k)$ en fonction de k pour $i = 1, 2, 3, 4$ et 5.
- Tracer l'évolution de $x_i(k)/N(k)$ en fonction de k pour $i = 1, 2, 3, 4$ et 5. Que représente x_i/N et vers quelle valeur pourrait-on prédire sa convergence ?
- Déterminer le vecteur des valeurs de reproduction de M . Quelle est la classe dont la condition initiale a le plus d'influence sur la taille de la population à long terme ?
- Construire et afficher la matrice des élasticités de la matrice M (on pourra utiliser des outils de produits matriciels ou une boucle `for` pour remplir la matrice coefficient par coefficient). Quel paramètre a le plus d'influence sur la vitesse de croissance de la population en temps long.
- On introduit un insecte qui va manger chaque jour 50% des jeune larves des classes $L1$, $L2$ et $L3$ ainsi que 50% des larves âgées de moins d'un jour (qui n'apparaissent pas explicitement dans le modèle, mais dans le taux de fécondité des larves adultes). Trouver les 7 flèches du graphe du cycle de vie dont le poids se voit modifié, et définir la matrice de projection B associé à ce nouveau graphe. Recommencer la partie numérique sur la matrice B (on pourra choisir une valeur plus élevée pour `n_max`). Qu'observe-t-on ?

Correction

6. Voir fichier de code python, on trouve $\lambda \approx 1.25$, et $W = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{pmatrix} \approx \begin{pmatrix} 0.35 \\ 0.22 \\ 0.15 \\ 0.12 \\ 0.16 \end{pmatrix}$
7. On sait qu'en temps long, $N(k+1)/N(k)$ tend vers λ , autrement dit $N(k)$ évolue proportionnellement à λ^k et tend vers l'infini.
Pour aller plus loin, on pourrait dire plus précisément que $N(k)/c\lambda^k$ tends vers 1 avec $c = {}^t V X(0)$ où V est le vecteur des valeurs reproductives de M .
8. On peut dire que $X(k)/N(k) \xrightarrow[k \rightarrow +\infty]{} W$, i.e. $\forall i = 1, \dots, 5$ $x_i(k)/N(k) \xrightarrow[k \rightarrow +\infty]{} w_i$
9. voir fichier python.
10. voir fichier python.
11. voir fichier python.
12. X_i/N représente la fraction de la population occupée par la classe i , on a déjà dit que x_i/N converge vers w_i .
13. Le vecteur des valeurs reproductives de M est $V \approx \begin{pmatrix} 0.43 \\ 0.65 \\ 0.96 \\ 1.46 \\ 2.43 \end{pmatrix}$. Donc la classe adulte (A) est celle qui a le plus d'influence sur $N(k)$ à long terme.
14. La plus grande élasticité est $e(m_{5,5}) \approx 0.29$, le taux de survie des adulte est donc le paramètre ayant le plus d'influence sur la taille de la population à long terme.
15. Voir fichier python. Cette fois la population finit par s'éteindre.

Exercice 12: Un exemple de matrice non primitive

Considérer le MMP de matrice de projection $A = \begin{pmatrix} 0 & 3 \\ 2 & 0 \end{pmatrix}$. Calculer les valeurs et vecteurs propres de A . Montrer que A est diagonalisable et donner la relation matriciel reliant $X(k)$ et $X(0)$. Reprendre les question 10 11 et 12 de l'exercice 11 (avec une condition initiale quelconque).

Chapitre 2

Probabilités conditionnelles, Chaines de Markov et distribution stationnaires

Délaissions un instant les modèles matriciels pour intéresser au calcul de probabilités, nous verrons qu'il arrive que ces deux problématiques se rejoignent.

2.1 Opération sur les ensembles

Définition 2.1.1: Union, Intersection, Différence

Soit A et B deux ensembles. On appelle

- **Union** de A et B , et on note $A \cup B$, l'ensemble des éléments appartenant à A **ou** à B . En langage mathématique : $x \in A \cup B \Leftrightarrow x \in A$ ou $x \in B$. **Attention :** en mathématique le “**ou**” est *inclusif*, c'est dire que A ou B signifie “soit A , soit B , soit les deux”.
- **Intersection** de A et B , et on note $A \cap B$, l'ensemble des éléments appartenant à A **et** à B . En langage mathématique : $x \in A \cap B \Leftrightarrow x \in A$ et $x \in B$.
 - Si $A \cap B = \emptyset$ (l'ensemble vide, que l'on peut aussi noter \emptyset) on dit que A et B sont **disjoints**.
 - Si $A \cap B = A$, on dit que A est **inclus** dans B (ou que B contient A) et on note $A \subset B$.
- **Différence** de A et B , et on note $A \setminus B$, l'ensemble des éléments appartenant à A **mais pas** à B . En langage mathématique : $x \in A \setminus B \Leftrightarrow x \in A$ et $x \notin B$.

Si il est clair, d'après le contexte, que l'on ne travaille que sur un ensemble Ω donné alors plutôt que d'écrire $\Omega \setminus A$ (où $A \subset \Omega$ est un sous-ensemble de Ω), on écrit simplement \overline{A} . On dit que \overline{A} est l'ensemble **complémentaire** de A (sous-entendu “dans Ω ”). Par exemple si on travaille dans l'ensemble $\{1, 2, 3, 4, 5, 6\}$ des six faces d'un dé, alors $\overline{\{2, 3\}} = \{1, 4, 5, 6\}$.

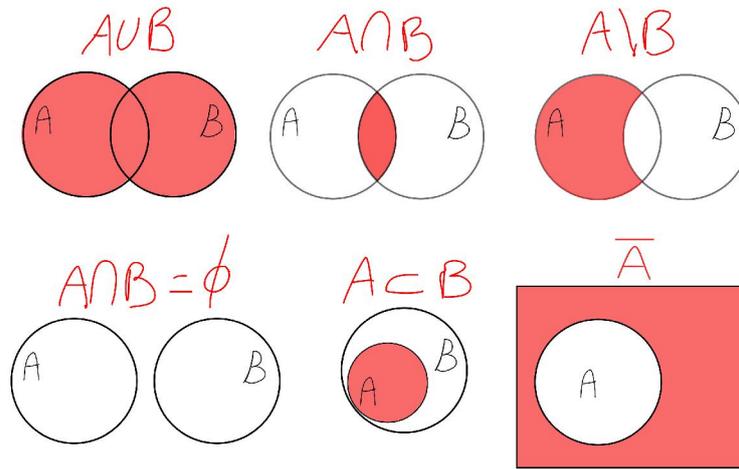


FIGURE 2.1 – Illustration de la définition 2.1.1

Définition 2.1.2: Partition

Soit Ω un ensemble et A_1, A_2, \dots, A_n des sous-ensembles de Ω (i.e. $A_i \subset \Omega$). On dit que les $(A_i)_i$ forment une partition de Ω si $\forall i \neq j, A_i \cap A_j = \emptyset$ et $A_1 \cup A_2 \dots \cup A_n = \Omega$.

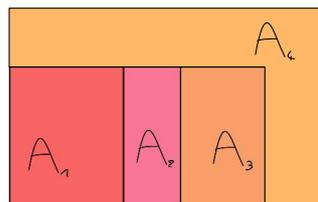


FIGURE 2.2 – Partition d'un ensemble

2.2 Lien avec les probabilités

Lors de la réalisation d'une expérience aléatoire, on note Ω l'ensemble des issues possibles de l'expérience. Ω s'appelle l'**ensemble fondamental**, ou **univers des possibles**.

Exemples

- Si la variable aléatoire X correspond au résultat d'un lancé de dé à six faces $\Omega = \{X = i\}, \text{ pour } i=1, \dots, 6\}$
- Si X correspond à la carte tirée dans un jeu de 52 cartes, alors Ω contient les 52 cartes ($\Omega = \{X = \text{As de coeur}\}, \{X = \text{As de pique}\}, \dots$).
- Si X correspond à la distance d'un lancé de javelot, alors $\Omega = \{X = x\}, x \in \mathbb{R}_+$.

Définition 2.2.1: Événement

Un sous-ensemble $A \subset \Omega$ s'appelle un **événement**. Ainsi, à chaque événement on peut attribuer une **probabilité** $\mathbb{P}(A)$.

Par exemple, pour un lancé de dé à six faces (équilibrées), on peut définir les événements $A = \{X = 2\}$, $B = \{X \in \{1, 2, 3\}\}$, $C = \{X \text{ est pair}\}$ ou encore $D = \emptyset$ (l'ensemble vide). Leurs probabilités sont $\mathbb{P}(A) = 1/6$, $\mathbb{P}(B) = 1/2$, $\mathbb{P}(C) = 4/6 = 2/3$ et $\mathbb{P}(D) = 0$.

Dans l'exemple du javelot on pourrait définir $A = \{X = 1\}$, $B = \{X \in [1, 2]\}$...

2.2.1 Propriétés sur les probabilités

Voici quelques propriétés élémentaires sur les probabilités (les trois premières sont des axiomes, c'est à dire qu'elles sont considérées comme vraie par défaut).

Une bonne façon de visualiser ces propriétés est de se représenter les événements comme des disques (ou formes quelconques), à la manière de la Figure 2.1. Ces propriétés sont assez naturelles si on pense à $\mathbb{P}(A)$ comme à la mesure de la surface de A .

Proposition 2.2.1: Propositions élémentaires des probabilités

1. $\mathbb{P}(A) \in [0, 1]$ (axiome 1),
2. $\mathbb{P}(\Omega) = 1$ (axiome 2),
3. Si $\mathbb{P}(A \cap B) = 0$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (axiome 3),
4. $\mathbb{P}(\emptyset) = 0$,
5. Si $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$,
6. $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$,
7. $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$,
8. $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Définition 2.2.2: Événements presque-sûrs, incompatibles, système exhaustif

On dit que deux événements A et B sont **incompatibles** si $\mathbb{P}(A \cap B) = 0$.

On dit que $A_1, A_2, \dots, A_n \subset \Omega$ forment un **système exhaustif** d'événements si $\forall i \neq j$, $\mathbb{P}(A_i \cap A_j) = 0$ et $\mathbb{P}(A_1 \cup A_2 \dots \cup A_n) = 1$.

Si $\mathbb{P}(A) = 1$ on dit que A est **presque sûr**.

Par exemple, au lancer de dé les événements "faire 3" ($\{X = 3\}$) et "faire un nombre pair" ($\{X \in \{2, 4, 6\}\}$) sont incompatibles.

Les événements $\{X \in \{1, 3\}\}$, $\{X = 2\}$ et $\{X \in \{4, 5, 6\}\}$ forment un système exhaustif

L'événement $\{X \in \{1, 2, 3, 4, 5, 6\}\}$ est presque sûr. Ou encore, $\{X > 0\}$ ou $\{X \neq 7\}$ sont également presque sûr.

Remarque 2.2.1. On peut remarquer qu'en probabilités, "incompatible" et "système exhaustif" sont très similaires aux notions "d'ensembles disjoints" et de "partitions" dans le langage

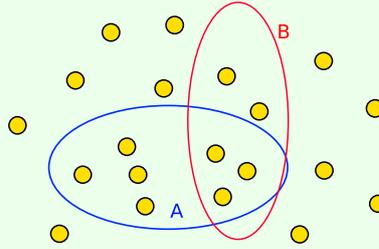
des ensembles. Ces termes sont presque interchangeables pour les variables aléatoires ne pouvant prendre qu'un nombre fini de valeurs.

Par exemple, si X suit une loi uniforme sur $[0, 10]$, alors $\mathbb{P}(X \in \{1, 2, 3, 4\}) = 0$ bien que $\{1, 2, 3, 4\} \neq \emptyset$. On a même $\mathbb{P}(X \in \mathbb{Q}) = 0$ bien que \mathbb{Q} possède une infinité de termes. Ou encore, $\mathbb{P}(\{X \in [0, 5]\} \cap \{X \in [5, 10]\}) = 0$ bien que $[0, 5] \cap [5, 10] = \{5\} \neq \emptyset$.

2.3 Probabilité conditionnelle, Théorème de Bayes et événements indépendants

Exemple 2.3.1: Approche naïve

Il y a 20 billes jaunes dans le dessin ci-dessous. On choisit l'une d'elles au hasard (avec loi de probabilité uniforme).



Quelle est la probabilité que la bille appartienne : 1) à l'ensemble A ? 2) À l'ensemble A sachant qu'elle appartient à B ? À l'ensemble A sachant qu'elle n'appartient pas à B ?

Définition 2.3.1: Probabilité conditionnelle

Soient deux événements A et B tels que $\mathbb{P}(B) \neq 0$. On appelle **probabilité (conditionnelle) de A sachant B** la quantité

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Pour tout événement B fixé, l'application $\mathbb{P}_B : A \mapsto \mathbb{P}(A|B)$ est une nouvelle mesure de probabilité, qui tient compte du fait que l'événement B est vrai. En particulier, \mathbb{P}_B vérifie la Proposition 2.2.1.

Exercice 13: Les enfants d'en face

Un petit garçon apprend que ses nouveaux voisins ont deux enfants. On suppose que les deux genres ont la même probabilité et que le genre des deux enfants sont indépendants. Quelle est la probabilité que les voisins aient une fille et un garçon ?

On apprend que les voisins ont au moins une fille. Quelle est la probabilité qu'ils en aient 2 ?

Correction

Les quatre configurations $F + F$, $F + G$, $G + F$ et $G + G$ sont équiprobables ($1/4$), on peut éventuellement s'en convaincre à partir d'un arbre de probabilités. Donc la probabilité recherchée est $1/4 + 1/4 = 1/2$.

On note A l'événement "les voisins ont deux filles" et B l'événement "les voisins ont au moins une fille". Alors $A \cap B = A$ et $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{P(A)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$.

2.3.1 Événements indépendants

On dira les événements que A et B sont **indépendants** si la connaissance de l'un ne modifie pas la probabilité de l'autre. Formellement :

Définition 2.3.2: Événements indépendants

On dit que deux événements A et B sont **indépendants** si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ ou encore, de façon équivalent dès que $\mathbb{P}(B) \neq 0$, si $\mathbb{P}(A|B) = P(A)$.

En particulier, un événement de probabilité nulle est indépendant de tout autre événement.

Exercice 14

On tire une carte au hasard dans un jeu de 32 cartes classique. Vérifier que les événements A = "La carte est une dame" et B = "La carte est un trèfle" sont indépendants. Mais que l'événement A n'est pas indépendant de C = "La carte est une tête (c'est à dire valet, reine ou roi)".

Dans l'exercice précédent, on peut avoir facilement l'intuition de l'indépendance, mais ce n'est pas toujours le cas.

Exercice 15: Charlatans

On dispose du nombre de guérisons observée parmi une cohorte de 250 individus atteint de migraine après deux jours de traitement homéopathique (Groupe H) ou placebo (Groupe P).

Issue \ Traitement	H	P	
Guéri (G)	120	80	
Malade (M)	30	20	

- L'industriel qui propose ce traitement assure que 80% des patients ont guéri après avoir reçu le traitement homéopathique. Vérifier cette information en utilisant la définition de la probabilité conditionnelle.
- Sachant qu'un patient a guéri, quelle est la probabilité qu'il ait suivi le traitement homéopathique.
- Quelle est la probabilité de guérir sachant que l'on a reçu le placebo ?
- Quelle est la probabilité de guérir ?
- Que peut-on dire de l'efficacité du traitement ?

Correction

- Il faut vérifier que $\mathbb{P}(G|H) = 0,8$. En effet, $\mathbb{P}(G|H) = \frac{\mathbb{P}(H \cap G)}{\mathbb{P}(H)} = \frac{120/250}{150/250} = 4/5 = 0,8$. Donc l'industriel dit vrai.
- $\mathbb{P}(H|G) = \frac{\mathbb{P}(H \cap G)}{\mathbb{P}(G)} = \frac{120/250}{200/250} = 0,6$.
- $\mathbb{P}(G|P) = \frac{\mathbb{P}(P \cap G)}{\mathbb{P}(P)} = \frac{80/250}{100/250} = 8/10 = 0,8$.
- $\mathbb{P}(G) = \frac{200}{250} = 0,8$.
- D'après les questions a) et c), les taux de guérisons avec le traitement ou le placebo sont les mêmes, donc le traitement n'a pas plus d'intérêt qu'un placebo. On aurait aussi pu montrer que $\mathbb{P}(P \cap G) = \mathbb{P}(P)\mathbb{P}(G)$, donc les événements "guérir" et "être dans le groupe homéopathie" sont indépendants.

Exercice 16: Le paradoxe de Simpson — tirée de Charig et al 1986 — Cf wikipedia

Le tableau suivant décrit le nombre de réussite (R) ou d'échec (E) d'extraction de calculs rénaux par deux techniques : la néphrolithotomie percutanée (N) qui consiste à introduire un appareil dans le rein pour fragmenter les calculs ; et la chirurgie ouverte (C) qui nécessite une large incision de la peau.

Issue \ Traitement	C	N	Total
R	273	289	562
E	77	61	138
Total	350	350	

a) Calculer et comparer les taux de succès de chacune des méthodes (probabilité de réussite).

Si l'on distingue (parmi les cas précédents) les opérations pour de petits (< 2cm) ou de gros (> 2cm) calculs, on obtient les tableaux suivant :

Petits calculs				Gros calculs			
	C	N	Total		C	N	Total
R	81	234	315	R	192	55	247
E	6	36	42	E	71	25	96
Total	87	270		Total	263	80	

b) Calculer les taux de succès de chacune des méthodes pour de petits calculs.

c) Calculer les taux de succès de chacune des méthodes pour de gros calculs.

d) Quelle méthode présente les meilleurs taux de réussites pour chacun des types de calculs ? Comment expliquer le résultat de la question a) ?

Correction

a) 0.78 (ou 78%) pour le traitement *C*, contre 83% pour le traitement *N*. On a donc un plus grand pourcentage de réussites avec le traitement *N*.

b) 93% pour *C* contre 87% pour *N*.

c) 73% pour *C* contre 69% pour *N*.

d) C'est donc avec la chirurgie que l'on obtient les meilleurs résultats sur de petits comme sur de gros calculs. Le résultat de la question a) donne l'impression de *N* et plus efficace, mais en réalité la méthode *N* a été davantage utilisée sur de petits calculs et la méthode *C* est plus utilisée pour les gros calculs, qui sont plus difficiles à extraire (taux de réussite de $247/(247 + 96) = 0.72$, contre $315/(315 + 42) = 0.88$ pour les petits calculs). La méthode *C* est donc plus efficace mais, puisqu'elle est en général utilisée sur des cas plus difficiles, elle a un taux de réussite global qui est inférieur à celui de la méthode *N*.

Exercice 17: Tiré du polycopié de S. Lemaire, DIFFICILE / HORS PROGRAMME

Une famille a n enfants. On suppose que les probabilités d'avoir une fille ou un garçon sont équiprobables ($\mathbb{P} = 1/2$). Montrer que les événements $A = \{\text{avoir au plus une fille}\}$ et $B = \{\text{avoir des enfants de chaque sexe}\}$ sont indépendants si et seulement si $n = 3$.

Correction

Remarquons déjà que $A \cap B$ est l'événement "avoir exactement une fille", en effet B implique qu'il y a au moins une fille et par définition A signifie qu'il y a au plus une fille. Il y a 2^n possibilités (chaque enfant peut être soit un garçon soit une fille).

La probabilité d'avoir exactement une fille est $\mathbb{P}(A \cap B) = n/2^n$ (la fille peut être n'importe lequel des n enfants). La probabilité d'avoir 0 fille est $1/2^n$ (avoir uniquement des garçons). Donc $\mathbb{P}(A) = \frac{n+1}{2^n}$.

On a $\bar{B} = \{\text{avoir uniquement des garçons ou uniquement des filles}\}$. Donc $\mathbb{P}(B) = 1 - 1/2^n - 1/2^n = (2^n - 2)/2^n$.

Finalement, $\mathbb{P}(A \cap B) = n/2^n$ et $\mathbb{P}(A)\mathbb{P}(B) = (n+1)(2^n - 2)/2^{2n}$.

Ainsi

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \Leftrightarrow n = (n+1)(2^n - 2)/2^n \Leftrightarrow 2n + 2 = 2^n$$

et on montre que $n = 3$ est l'unique solution de cette équation (on peut par dresser le tableau de variation de $f(x) = 2^x - (2x + 2)$ pour montrer qu'il n'existe qu'une seule racine > 0 , et que 3 en est une).

Attention à ne pas confondre événements *indépendants* (la connaissance de l'un ne nous renseigne pas sur l'autre) et *incompatibles* (les deux événements ne peuvent être vrais simultanément, $\mathbb{P}(A \cap B) = 0$).

2.3.2 Le théorème de Bayes

Remarquons que, puisque $\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$, les probabilités des causes et des conséquences peuvent être inversées par la formule suivante

$$\mathbb{P}(A|B) = \mathbb{P}(B|A) \frac{\mathbb{P}(A)}{\mathbb{P}(B)} \tag{2.1}$$

L'équation précédent s'appelle **Théorème de Bayes**.

On peut également en donner une versions plus développée à l'aide de la formule des probabilités totales.

Remarquons que soit A est vrai, soit \bar{A} est vrai, mais jamais les deux à la fois. Autrement dit A et \bar{A} forment un système exhaustif de Ω . On en déduit que $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap \bar{A})$. Plus généralement,

Proposition 2.3.1: Formule des probabilités totales

Soit $n \in \mathbb{N}^*$ et $(A_i)_{1 \leq i \leq n}$ un système exhaustif tel que $\forall i, \mathbb{P}(A_i) \neq 0$ et soit B un événement. Alors

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B \cap A_i) = \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

Démonstration. La première égalité est une conséquence de l'axiome 3 de la Proposition 2.2.1. La seconde égalité correspond à la définition de la probabilité conditionnelle. \square

On en déduit donc la forme développée du théorème de Bayes :

Théorème 2.3.2: Théorème de Bayes

Soit $n \in \mathbb{N}^*$ et $(A_i)_{1 \leq i \leq n}$ un système exhaustif tel que $\forall i, \mathbb{P}(A_i) \neq 0$ et soit B un événement. Alors

$$\mathbb{P}(A_i|B) = \mathbb{P}(B|A_i) \frac{\mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^n \mathbb{P}(B \cap A_j)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^n \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

En particulier, si A est tel que $0 < \mathbb{P}(A) < 1$, alors

$$\mathbb{P}(A|B) = \mathbb{P}(B|A) \frac{\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B \cap A) + \mathbb{P}(B \cap \bar{A})} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\bar{A})\mathbb{P}(\bar{A})}$$

Exercice 18: Forêt du Bugey

Une forêt du Bugey se compose de 30% de chênes, 50% d'épicéas et de 20% de hêtres. Après le passage d'une tempête, une maladie se propage et touche 10% des chênes, 4% des épicéas et 25% des hêtres. Sachant qu'un arbre est malade, quelle est la probabilité que ce soit un chêne ? un épicéa ? un hêtre ?

Correction

On note M l'événement "l'arbre en question est malade", puis C , E et H les événements "l'arbre est un chêne", "un épicéa" et "un hêtre".

(C, E, H) forme un système exhaustif, ainsi d'après la formule de Bayes :

$$\mathbb{P}(C|M) = \mathbb{P}(M|C) \frac{\mathbb{P}(C)}{\mathbb{P}(M)} = \frac{\mathbb{P}(M|C)\mathbb{P}(C)}{\mathbb{P}(M|C)\mathbb{P}(C) + \mathbb{P}(M|E)\mathbb{P}(E) + \mathbb{P}(M|H)\mathbb{P}(H)} = 0.3$$

Selon le même raisonnement on trouve que $\mathbb{P}(E|M) = 0.02/0.1 = 0.2$ puis enfin $\mathbb{P}(H|M) = 0.05/0.1 = 0.5$. Remarque, pour les hêtres on peut aussi utiliser le fait que $\mathbb{P}(H|M) = 1 - \mathbb{P}(E|M) - \mathbb{P}(C|M)$ puisque la somme des probabilités fait 1.

Exercice 19: Groupe sanguin

Certains antigènes présents à la surface des globules rouges permettent de classer le sang humain en 4 groupes (A, B, AB, O). Indépendamment du groupe, les individus porteur de l'antigène dit "grand D" sur leurs globules rouges sont dit de rhésus positif (Rh^+). Dans le cas contraire ils sont de rhésus négatif (Rh^-).

D'après l'Institut National de la Transfusion Sanguine, dans la population française il a 45% d'individus de groupe sanguin A, 43% de groupe O, 9% de groupe B et 3% de AB. Par ailleurs, pour chaque groupe sanguin, la proportion d'individus de rhésus positifs ou négatifs est la suivante

Rhésus \ groupe	A	O	B	AB
Rh^+	85%	85%	85%	86%
Rh^-				

a) Compléter la dernière ligne du tableau.

On choisit un individu au hasard, calculer la probabilité que cet individu

b) soit de groupe sanguin O.

c) soit donneur universel (groupe O-).

d) ait un rhésus négatif.

e) soit de groupe O sachant qu'il est de rhésus négatif.

Correction

a) Chaque colonne doit faire 100% donc on a dans l'ordre : 15, 15, 15 et 14%.

b) $\mathbb{P}(O) = 0.43$ d'après l'énoncé.

c) $\mathbb{P}(O \cap Rh^-) = \mathbb{P}(Rh^-|O)\mathbb{P}(O) = 0.15 \times 0.43 = 0.0645$.

d) Formule des probabilités totales :

$$\mathbb{P}(Rh^-) = \mathbb{P}(Rh^-|A)\mathbb{P}(A) + \mathbb{P}(Rh^-|O)\mathbb{P}(O) + \mathbb{P}(Rh^-|B)\mathbb{P}(B) + \mathbb{P}(Rh^-|AB)\mathbb{P}(AB).$$

Donc $\mathbb{P}(Rh^-) = 0.15 \times 0.45 + 0.15 \times 0.43 + 0.15 \times 0.09 + 0.14 \times 0.03 = 0.1497$.

e) On utilise la formule de Bayes $\mathbb{P}(O|Rh^-) = \mathbb{P}(Rh^-|O) \frac{\mathbb{P}(O)}{\mathbb{P}(Rh^-)} = 0.15 \frac{0.43}{0.1497} \approx 0.4308$.

Exercice 20: Test de dépistage

On dispose d'un test permettant de dépister une maladie. On note P et M les événements "le test est positif" et "l'individu est malade". On suppose que la maladie touche un individu sur 1000 ($\mathbb{P}(M) = 1/1000$). Le test n'étant pas infaillible, on suppose que $\mathbb{P}(P|M) = 0.99$ et $\mathbb{P}(P|\bar{M}) = 0.03$ (faux positif).

a) Quelle est la probabilité $\mathbb{P}(\bar{P}|M)$ que le test soit négatif pour un individu malade (faux négatif).

b) Utiliser la formule des probabilités totales pour déterminer la probabilité $\mathbb{P}(P)$ qu'un test soit positif.

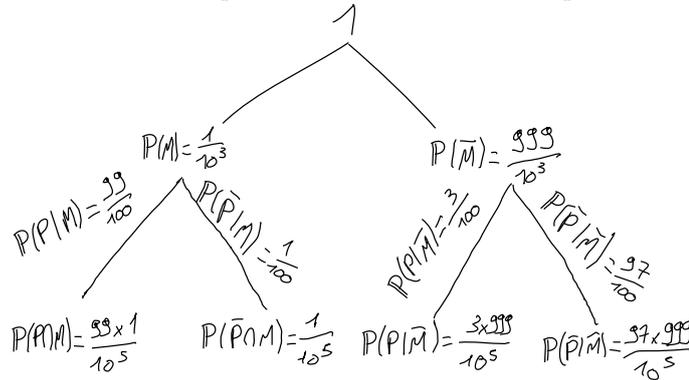
c) Vous faites le test. Celui-ci est positif. Quelle est la probabilité que vous soyez malade ?

Correction

- a) $\mathbb{P}(\bar{P}|M) + \mathbb{P}(P|M) = 1$, donc $\mathbb{P}(\bar{P}|M) = 1 - 0.99 = 0.01$. Le risque de faux négatif est plutôt faible (0.01 est très petit, mais les faux négatifs ont de très graves conséquences).
- b) $\mathbb{P}(P) = \mathbb{P}(P|M)\mathbb{P}(M) + \mathbb{P}(P|\bar{M})\mathbb{P}(\bar{M}) = 0.99/1000 + 0.03(1 - 1/1000) = 0.03096$
- c) $\mathbb{P}(M|P) = \mathbb{P}(P|M) \frac{\mathbb{P}(M)}{\mathbb{P}(P)} = \frac{\mathbb{P}(P|M)\mathbb{P}(M)}{\mathbb{P}(P|M)\mathbb{P}(M) + \mathbb{P}(P|\bar{M})\mathbb{P}(\bar{M})} = \frac{0.99/1000}{0.03096} \approx 0.032$

Analyse : Ce résultat peut être assez déroutant puisque la probabilité d'être positif sur un individu sain est très faible (0.03=3%). Par conséquent le pourcentage de faux-positifs parmi les individus sains est petit (3%), mais les individus sains sont très nombreux (99.9% de la population), donc le nombre de faux-positif est grand par rapport au nombre de malades. On voit bien dans le dénominateur du le théorème de Bayes que le test serait plus efficace si le taux de faux-positifs $\mathbb{P}(P|\bar{M})$ était plus faible, ou l'incidence de la maladie $\mathbb{P}(M)$ plus grande.

Éventuellement, on peut s'aider d'un arbre de probabilités



$$\mathbb{P}(M|P) = \frac{\mathbb{P}(M \cap P)}{\mathbb{P}(P)} = \frac{\mathbb{P}(M \cap P)}{\mathbb{P}(M \cap P) + \mathbb{P}(\bar{M} \cap P)} = \frac{99}{99 + 3 \times 999} \approx 0,032$$

2.4 Chaines de Markov

Une chaîne de Markov est une suite $(X_k)_{k \in \mathbb{N}}$ dont chaque terme est une variable aléatoire et telle que la prédiction de la valeur de X_{k+1} lorsque l'on connaît X_k n'est pas rendue plus précise par la connaissance de $X_0, X_1 \dots X_{k-1}$. On dit parfois que les chaînes de Markov sont *sans mémoire*. Formellement

Définition 2.4.1: Chaîne de Markov

On dit qu'une suite de variables aléatoires $(X_k)_n \geq 0$, prenant ses valeurs dans un espace E quelconque, est une **chaîne de Markov** si pour tout $k \geq 0$, et pour tous $a_0, a_1, \dots, a_{k+1} \in E$,

$$\mathcal{P}(X_{k+1} = a_{k+1} | X_0 = a_0, X_1 = a_1, \dots, X_k = a_k) = \mathcal{P}(X_{k+1} = a_{k+1} | X_k = a_k)$$

Attention : cela ne signifie pas que X_{k+1} est indépendant de son passé. Mais uniquement que dès lors que l'on connaît X_k , la connaissance des X_i pour $k < i$ est sans importance. Autrement dit, l'événement connu le plus récent contient tout le pouvoir prédictif des événements qui le précèdent.

Proposition 2.4.1: Probabilité d'une trajectoire

Soit $(X_k)_n \geq 0$ une chaîne de Markov à valeurs dans un ensemble E et soient $a_0, a_1, \dots, a_n \in E$. Alors

$$\mathbb{P}(X_0 = a_0 \text{ et } X_1 = a_1, \dots \text{ et } X_n = a_n) = \mathbb{P}(X_0 = a_0) \prod_{i=0}^{n-1} \mathbb{P}(X_{i+1} = a_{i+1} | X_i = a_i).$$

Remarque 2.4.1. C'est tout simplement ce que l'on fait lorsque l'on calcule la probabilité d'une branche dans un arbre de probabilité.

Définition 2.4.2: Chaîne de Markov homogène

On dit qu'une chaîne de Markov est **homogène** si, pour tout $k \geq 1$,

$$\mathbb{P}(X_{k+1} = b | X_k = a) = \mathbb{P}(X_1 = b | X_0 = a).$$

Autrement dit, si les probabilités de transition d'un état a vers un état b ne dépendent que de a et b (et non de k). Nous ne considérerons que des chaînes de Markov homogènes.

Un exemple simple : On suit l'activité d'un poisson rouge qui ne connaît que trois occupations $E = \{\text{manger, dormir, promener}\}$. Connaissant l'activité du poisson à un instant donné, on connaît les probabilités qu'il s'occupe à chacune des activités une heure plus tard. Par exemple, si le poisson dort, il se promènera avec probabilité 0.1 une heure plus tard. Les probabilités sont reportées sur le graphe ci-dessous.

la suite $(X_i)_i$ où X_n est l'activité du poisson à l'heure n est une chaîne de Markov homogène. En effet, connaissant l'activité du poisson à l'heure n , on connaît les probabilités de son activité à l'heure $n + 1$ (quelles qu'aient pu être ses activités dans le passé), et les probabilités de transition ne dépendent pas de n mais uniquement de l'activité X_n .

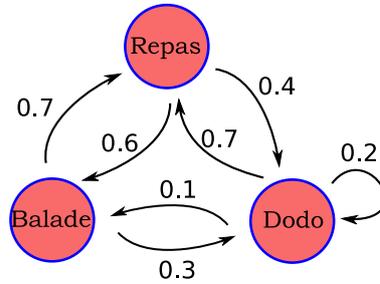


FIGURE 2.3 – Graphe de transition des activités d'un poisson rouge.

Exemple 2.4.1: Quelques exemples un peu plus tordus

Des couples d'oiseaux migrateurs arrivent progressivement dans une forêt et choisissent un arbre pour y faire leur nid et pondre.

- On suppose que deux nids ne sont jamais établis dans le même arbre. X_n désigne la position de l'arbre dans lequel le n -ième nid est construit. Alors la suite (X_n) n'est pas une chaîne de Markov
- On note Y_n désigne le nombre de couples d'oiseaux qui s'installent dans la forêt le n -ème jour, et $Z_n = Y_0 + Y_1 + \dots + Y_n$ le nombre total de couples installés (depuis le premier jour). Si les $(Y_i)_i$ sont identiquement distribués, alors Z_n est une chaîne de Markov homogène. Sinon, c'est une chaîne de Markov inhomogène (l'augmentation de la population entre le jour 0 et le jour 1 ne suit pas la même loi que celle entre le jour 10 et le jour 11, car la loi des Y_i dépend de i).

Définition 2.4.3: Matrice de probabilités de transition

Soit $(X_k)_k \geq 0$ une chaîne de Markov homogène pouvant prendre un nombre fini d'états. On appelle **matrice de transition** de $(X_k)_k \geq 0$ la matrice $\Pi = (\pi_{i,j})_{1 \leq i,j \leq n} \in \mathbb{M}_n(\mathbb{R})$ telle que pour tout $1 \leq i, j \leq n$ et tout $k \geq 0$,

$$\pi_{i,j} = \mathbb{P}(X_{k+1} = E_i | X_k = E_j).$$

Dans l'exemple du poisson rouge, en rangeant les activités dans l'ordre 1 =repas, 2 =dormir, 3 =balade, on obtient la matrice des probabilités de transition suivante

$$A = \begin{pmatrix} 0 & 0.7 & 0.7 \\ 0.4 & 0.2 & 0.3 \\ 0.6 & 0.1 & 0 \end{pmatrix}$$

remarque : si la chaîne n'était pas homogène, la matrice de transition changerait à chaque itération ($A(k)$).

Définition 2.4.4: Vecteur de probabilités

On dit que $L = (l_1, \dots, l_n) \in \mathbb{R}^n$ est un vecteur de probabilités si ses éléments sont positifs et leur somme vaut 1 ($l_i \geq 0$ et $\sum l_i = 1$).

2.4.1 Lien avec les MMP

Dans toutes la suite de cette section,

- $(X_k)_{k \geq 0}$ désigne une chaîne de Markov homogène prenant ses valeurs dans un ensemble fini $E = \{E_1, E_2, \dots, E_n\}$
- $L(k) = (l_1(k), \dots, l_n(k))$ désigne le vecteur de taille n donnant la loi de $X(k)$, c'est à dire $l_i(k) = \mathbb{P}(X(k) = E_i)$. En particulier $L(0)$ est la loi initiale de la chaîne $(X_k)_{k \geq 0}$.
- $\Pi \in \mathbb{M}_n(\mathbb{R})$ désigne la matrice des probabilités de transition de la chaîne $(X_k)_{k \geq 0}$.

Proposition 2.4.2: Matrice stochastique

Pour tout $k \geq 0$,

1. Si U est un vecteur de probabilités, alors $\Pi^k U$ est un vecteur de probabilités.
2. Chaque colonne de Π^k est un vecteur de probabilités.
3. $\forall k \geq 0, L(k+1) = \Pi L(k)$.
4. La loi $L(k)$ de X_k est donnée par

$$L(k) = \Pi^k L(0).$$

En particulier

$$\pi_{i,j}^{(k)} = \mathbb{P}(X_k = E_i | X_0 = E_j)$$

c'est à dire la probabilité de passer de j à i en k itérations.

Démonstration.

1. Preuve par récurrence. Initiation : si $k = 0, \Pi^k L = L$ qui est bien un vecteur de probabilité (par hypothèse). Hérédité : soit $k \geq 1$. Notons $Q = \Pi^k L$ et supposons qu'il s'agisse d'un vecteur de probabilité, c'est à dire que $\forall i, Q_i \geq 0$ et $\sum_{j=1}^n Q_j = 1$.

Alors, pour tout $1 \leq i \leq n, (\Pi Q)_i = \sum_{j=1}^n \pi_{i,j} Q_j$. Ainsi $(\Pi Q)_i \geq 0$, comme somme et produit de termes positifs. Et

$$\sum_{i=1}^n (\Pi Q)_i = \sum_{i=1}^n \sum_{j=1}^n \pi_{i,j} Q_j = \sum_{j=1}^n \sum_{i=1}^n \pi_{i,j} Q_j = \sum_{j=1}^n Q_j \underbrace{\sum_{i=1}^n \pi_{i,j}}_{=1} = \sum_{j=1}^n Q_j = 1.$$

Donc $\Pi Q = \Pi^{k+1} L$ est un vecteur de probabilité.

2. $\forall 1 \leq i \leq n$, le vecteur E ayant sa i -ème composante égale à 1 et les autres à 0 est un vecteur de probabilité. On vérifie aisément que ΠE est égal à la i -ème colonne de Π . En vertu du premier point, c'est un vecteur de probabilité.
3. admis (démonstrable par récurrence)
4. admis (démonstrable par récurrence)

□

Conclusion : une chaîne de Markov à nombre d'états finis n'est rien de plus qu'un MMP particulier, dont la matrice de projection est la matrice de transition d'une chaîne de Markov. En particulier, on peut définir le graphe associé à Π , et tous les résultats des MMP s'appliquent.¹

La différence est surtout conceptuelle : ici $L(k)$ est le vecteur des probabilités que X_k se trouve dans chacun des états de E , alors que dans le chapitre 1, $X(k)$ (l'équivalent de $L(k)$ ici, et non de la chaîne $(X_k)_{k \geq 0}$) désignait une quantité qui pouvait prendre des valeurs arbitrairement grandes.

2.4.2 Particularité des chaînes de Markov

Par définition, les colonnes d'une matrice de transition sont des vecteurs de probabilité (la somme de leurs éléments vaut 1). Cette particularité induit quelques résultats propres aux chaînes de Markov.

Définition 2.4.5: Probabilité stationnaire

On dit qu'un vecteur de probabilité $Q = (q_1, \dots, q_n) \in \mathbb{R}^n$ est une probabilité stationnaire de la chaîne de Markov $(X_k)_{k \geq 0}$ si $\Pi Q = Q$. Autrement dit, Q est un vecteur propre de Π associé à la valeur propre 1.

Théorème 2.4.3: Résultats essentiels sur les probabilités stationnaires

Soit Π la matrice de transition d'une chaîne de Markov homogène à n états.

1. La plus grande valeur propre de Π est 1 (en particulier, il existe toujours au moins une probabilité stationnaire).
2. Si Π est primitive, alors
 - il existe une unique probabilité stationnaire $Q = (q_1, q_2, \dots, q_n)$, elle correspond au vecteur de Perron de Π
 - Tous les coefficients Q sont strictement positifs ($\forall i, q_i > 0$)
 - Quelque soit la loi initiale $L(0)$, $L(k)$ converge vers Q lorsque $k \rightarrow +\infty$ (c'est la Proposition 1.4.2).
 - Π^k converge vers la matrice de $\mathbb{M}_n(\mathbb{R})$ dont toutes les colonnes sont identiques, et égales au vecteur Q .
 - Pour tout $1 \leq i \leq n$, on note T_{ii} le temps moyen entre deux visites de l'état i par la chaîne $(X_k)_k$ (c'est à dire $T_{ii} = \mathbb{E}(\min\{k \in \mathbb{N}^* \text{ tel que } X_k = i \text{ sachant que } X_0 = i\})$), alors

$$T_{ii} = 1/q_i.$$

Remarque 2.4.2. Les trois premiers tirets du point 2 sont ni plus ni moins les conséquences du Théorème de Perron-Frobenius pour un MMP (Théorème 1.4.1 et Proposition 1.4.2)

1. **Attention si vous consultez d'autres ressources** : traditionnellement, on définit $\Pi_{i,j}$ comme la probabilité de passage de l'état E_i à l'état E_j (et non de E_j à E_i), on écrit alors $L(k)$ comme un vecteur ligne (et non colonne) et on a $L(k) = L(0)\Pi^k$. Les deux choix sont *strictement équivalents* (il suffit d'utiliser la *transposée* de Π et de L). Le choix de définir Π comme dans la définition 2.4.3 permet de retomber sur le formalisme des MMP.

Que signifie que $L(k)$ converge vers une probabilité stationnaire Q ? La convergence de $L(k)$ vers Q peut s'écrire : $\forall i, j \in \{1, \dots, n\}, \lim_{k \rightarrow +\infty} \mathbb{P}(X_k = i | X_0 = j) = q_i$. Si on attend suffisamment longtemps, le système "oublie" sa condition initiale $L(0)$ et le vecteur Q décrit avec quelles probabilités le système se trouve dans chacun des états possibles. Une autre façon de le dire est que, après une première phase de régime transitoire (dépendant des conditions initiales), le système étudié passe une fraction q_i de son temps dans l'état i (pour chaque i).

Démonstration des trois premiers points. 1. Soit λ une valeur propre de Π , notons $V = (v_1, \dots, v_n)$ un vecteur propre à gauche associé. Soit i_0 tel que $|v_{i_0}| = \max_{1 \leq i \leq n} |v_i|$. Alors

$$|(V\Pi)_{i_0}| = \left| \sum_{1 \leq i \leq n} v_i \pi_{i, i_0} \right| \leq \sum_{1 \leq i \leq n} |v_i \pi_{i, i_0}| = \sum_{1 \leq i \leq n} \pi_{i, i_0} |v_i| \leq |v_{i_0}| \sum_{1 \leq i \leq n} \pi_{i, i_0} = |v_{i_0}|,$$

où la première inégalité résulte de l'inégalité triangulaire, et la dernière égalité du caractère stochastique de Π .

En outre, par définition de V , $V\Pi = \lambda V$. En particulier $(V\Pi)_{i_0} = \lambda v_{i_0}$. On a donc

$$|(V\Pi)_{i_0}| = |\lambda v_{i_0}| \leq |v_{i_0}|,$$

c'est à dire $|\lambda| \leq 1$.

2. D'après le théorème de Perron-Frobenius, Π admet une valeur propre dominante λ . Or d'après 1) on sait que $\lambda \leq 1$. On constate (il suffit de calculer le produit) que le vecteur $(1, 1, \dots, 1)$ est un vecteur propre à gauche de Π associé à la valeur propre 1. Ainsi $\lambda = 1$ est la valeur propre dominante de Π . En nommant Q le vecteur de Perron associé, le point 2) est une conséquence immédiate de la Proposition 1.4.2). □

Exercice 21

Considérons un gène codant pour la couleur de la robe d'un bovin (vache ou taureau) et pouvant être représenté par deux allèles différents a et A . Ce gène étant présent sur les deux chromosomes d'une même paire, un bovin peut avoir les génotypes (phénotypes) aa (pelage blanc), AA (noir) ou aA (tacheté). On ne fait pas de différence entre aA et Aa .

1. Un éleveur possède une vache blanche qu'il fait saillir par un taureau tacheté pour obtenir un (et un seul) veau. Lorsque ce veau devient fertile, il le fait s'accoupler avec un bovin tacheté, et ainsi de suite : à chaque génération, il apparie systématiquement l'individu de la n -ème génération avec un bovin tacheté. Modéliser par une chaîne de Markov $(X_k)_k$ le phénotype observé à chaque génération : quels états peuvent être pris par la chaîne ? En rangeant les génotypes dans l'ordre AA, aA, aa , quelle est la matrice de transition Π ? la loi initiale $L(0)$?
2. Quelle est la probabilité que la veau de deuxième génération (dont la couleur est donnée par X_2) soit noir ?
3. Reprendre la question 1 si l'éleveur décide de faire appel à des taureaux/vaches reproducteurs noirs plutôt que tachetés, on notera \tilde{P}_i la matrice de transition et $(\tilde{X}_k)_k$ la chaîne.
4. Comparer l'évolution des deux chaînes, que peut-on dire de la distribution stationnaire des différents phénotypes ?
5. Pour la première chaîne, on peut montrer que la distribution stationnaire est $Q = (1/4, 1/2, 1/4)$: combien de génération faut-il attendre en moyenne, si l'on part avec une vache blanche, pour qu'un veau blanc naisse ?
6. Quel effet cela aurait-il eu si l'éleveur avait commencé ce processus avec une vache noire ?

Correction

1. On note X_k la variable aléatoire correspondant à la robe du bovin de k -ème génération, cette chaîne prend ses valeurs dans $E = \{AA, aA, aa\}$ et on sait que $X_0 = aa$. En rangeant les états dans l'ordre AA, aA, aa , la matrice de transition et la loi initiale sont

$$\Pi = \begin{pmatrix} 1/2 & 1/4 & 0 \\ 1/2 & 1/2 & 1/2 \\ 0 & 1/4 & 1/2 \end{pmatrix}, \quad L(0) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Pour trouver ces probabilités il faut envisager toutes les recombinaisons possible entre les gènes d'un bovin tacheté et d'un bovin de chacun des autres types. on peut faire des arbres de proba

2. $\mathbb{P}(X_2 = AA|X_0 = aa)$ est donnée par $\pi_{1,3}^{(2)}$ (le coefficient 1, 3 de Π^2). Ou encore (c'est la même chose) on peut calculer $L(2) = \Pi^2 L(0)$ avec $L(0) = (0, 0, 1)$ et regarder le premier coefficient. On trouve

$$\Pi^2 = \begin{pmatrix} 3/8 & 1/4 & 1/8 \\ 1/2 & 1/2 & 1/2 \\ 1/8 & 1/4 & 3/8 \end{pmatrix} \text{ donc } L(2) = \begin{pmatrix} 1/8 \\ 1/2 \\ 3/8 \end{pmatrix} \text{ et } \mathbb{P}(X_2 = AA|X_0 = aa) = 1/8.$$

3. Dans ce cas là, la loi initiale est la même et la matrice de transition est

$$\tilde{\Pi} = \begin{pmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

4. On vérifie que la matrice Π est primitive, donc la loi de X_k converge vers une loi stationnaire $Q = (q_1, q_2, q_3)$ avec $q_i > 0$ pour tout i (on pourrait montrer que $Q = (1/4, 1/2, 1/4)$). Donc, sur un grand nombre de génération, on a 1/2 tacheté, 1/4 blanches et 1/4 noires.

En revanche, en dessinant le graphe associé à la matrice $\tilde{\Pi}$ on se rend compte que tôt ou tard, il n'y aura plus que des bovins noirs (AA). Pour s'en convaincre par le calcul, on peut calculer $\tilde{L}(1) = (0, 1, 0)$, donc $P(X_1 = aA) = 1$, puis pour $k \geq 1$ on a $\tilde{l}_1(k)$ qui croit et $\tilde{l}_2(k)$ qui tend vers 0 (cf théorie des états absorbants que nous n'avons pas eu le temps de voir cette année).

5. D'après le dernier théorème du cours, il faut attendre en moyenne $T_{3,3} = 4$ générations en dessous de la vache originelle.)

Exercice 22: (inspiré d'un TD de Vladimir Kostov)

Une certaine maladie touchant l'Homme se transmet par pique d'insecte. Chaque individu peut être dans l'un des trois états suivants : ni malade ni immunisé (R), malade (M) ou immunisé (I). D'un mois sur l'autre, son état peut changer selon les règles suivantes :

S'il est immunisé, il a une probabilité 0,8 de le rester et 0,2 de passer à l'état R . S'il est malade, il a une probabilité 0,25 de le rester et 0,75 de devenir immunisé. Enfin, s'il est dans l'état R , il a une probabilité 0,6 de le rester et 0,4 de tomber malade.

1. Donner la matrice de transition Π de la chaîne de Markov homogène modélisant la proportion d'individus dans chacun des états de l'ensemble $E = \{I, M, R\}$ (en rangeant les états dans l'ordre I, M, R).
2. Calculer la proportion d'individus malades dans la population après un mois si 4% des individus sont malades au départ et les autres sont dans l'état R .
3. Si l'on calcule avec un ordinateur la puissance Π^{100} de la matrice de transition on trouve (en arrondissant)

$$\Pi^{100} = \begin{pmatrix} 0,566 & 0,566 & 0,566 \\ 0,151 & 0,151 & 0,151 \\ 0,283 & 0,283 & 0,283 \end{pmatrix}.$$

Peut-on en déduire que la matrice de transition de cette chaîne de Markov est une matrice primitive ? Pourquoi ?

4. On peut vérifier que $\Pi^{101} = \Pi^{100}$. Quelle sera la proportion d'individus malades dans cette population à long terme ? Justifier votre réponse.
5. Selon ce modèle, que se passerait-il à long terme dans une région où la proportion de malades serait de 48% au départ (et les autres dans l'état R) ?

Correction

1. $\Pi = \begin{pmatrix} 0,8 & 0,75 & 0 \\ 0 & 0,25 & 0,4 \\ 0,2 & 0 & 0,6 \end{pmatrix}$ est la matrice de transition associée à la chaîne de Markov (X_k) telle que $X_k \in \{I, M, R\}$ donne l'état d'un individu tiré au hasard dans la population.
2. On calcule $L(1) = \Pi L(0)$ avec $L(0) = (0, 0.04, 0.96)$. En fait on ne s'intéresse qu'à la seconde composante de $L(1)$ et on peut se dispenser de calculer les autres. Donc $L(1) = (\dots, 0.04 \times 0.25 + 0.96 \times 0.6, \dots) = (\dots, 0.394, \dots)$ et donc $\mathbb{P}(X_1 = M) = 0.394$.
3. Si on se ramène à la définition d'une matrice positive (il existe k tel que Π^k est strictement positive), on voit que Π est primitive (pour $k = 100$ par exemple). Sinon, on peut faire l'étude du graphe associé.
4. On sait que Π^k converge, quand k tend vers l'infini, vers la matrice dont chaque colonne correspond à la probabilité stationnaire Q . Comme $\Pi^{101} = \Pi^{100}\Pi = \Pi^{100}$, on sait que pour tout $k > 100$, $\Pi^k = \Pi^{100}$, donc la suite des $(\Pi^k)_k$ a convergé, et sa limite est donnée par Π^{100} . Donc la loi stationnaire est $Q = (0.566, 0.151, 0.283)$: en temps long il y a 15.1% de malades.
5. Le théorème du cours assure que la loi $L(k)$ converge vers la loi stationnaire Q quelle que soit la loi initiale. Donc la proportion de malade convergera vers 15.1% également.

Exercice 23: (emprunté au cours de Vladimir Kostov)

On veut étudier l'effet de la présence d'un couple de lions dans une portion de savane dans laquelle cohabitent trois populations d'animaux dont les lions se nourrissent. On modélise les animaux mangés, antilopes (a), gnous (g) et zèbres (z) comme les états d'une chaîne de Markov dont les trajectoires sont des successions de proies mangées par les lions, comme par exemple ($gzzagga$). On fait l'hypothèse que la probabilité qu'un lion mange une proie a (ou g ou z) après avoir mangé une proie g (ou a ou z) ne dépend pas de ce qu'il avait mangé avant g (ou a ou z) et que cette probabilité est invariante au cours du temps. D'où la modélisation par une chaîne de Markov homogène d'espace d'états $E = \{a, g, z\}$ et dont on propose la matrice de transition suivante

$$\Pi = \begin{pmatrix} 0,5 & 0,2 & 0,2 \\ 0,1 & 0,3 & 0,2 \\ 0,4 & 0,5 & 0,6 \end{pmatrix}.$$

1. Quelle est, selon ce modèle, la probabilité qu'un lion mangent un zèbre après avoir mangé une antilope?
2. Des deux trajectoires suivantes, ($gazg$) et ($gzag$), quelle est la plus probable?
3. La distribution $L(0) = (6/21, 4/21, 11/21)$ est-elle une distribution stationnaire pour la chaîne de Markov? Justifier votre réponse.
4. Si au départ la nourriture des lions se composait pour $3/4$ de zèbres, cette proportion dans leur régime alimentaire va-t-elle, selon ce modèle, diminuer, augmenter ou rester constante? Expliquer.

Correction

1. $\mathbb{P}(X_{k+1} = z | X_k = a) = \Pi_{3,1} = 0.4$.
2. D'un côté, $\mathbb{P}(X_0 = g, X_1 = a, X_2 = z, X_3 = g) = \mathbb{P}(X_0 = g)\mathbb{P}(X_1 = a | X_0 = g)\mathbb{P}(X_2 = z | X_1 = a)\mathbb{P}(X_3 = g | X_2 = z) = \mathbb{P}(X_0 = g) \times 0.2 \times 0.4 \times 0.2 = 0.016\mathbb{P}(X_0 = g)$.

De l'autre, $\mathbb{P}(X_0 = g, X_1 = z, X_2 = a, X_3 = g) = \mathbb{P}(X_0 = g)\mathbb{P}(X_1 = z | X_0 = g)\mathbb{P}(X_2 = a | X_1 = z)\mathbb{P}(X_3 = g | X_2 = a) = \mathbb{P}(X_0 = g) \times 0.5 \times 0.2 \times 0.1 = 0.01\mathbb{P}(X_0 = g)$.

Donc, à moins que $\mathbb{P}(X_0 = g) = 0$, la trajectoire $gzag$ est 1.6 fois plus probable que la trajectoire $gazg$.

3. On calcule le produit

$$\Pi L(0) = \frac{1}{21} \begin{pmatrix} 0,5 & 0,2 & 0,2 \\ 0,1 & 0,3 & 0,2 \\ 0,4 & 0,5 & 0,6 \end{pmatrix} \begin{pmatrix} 6 \\ 4 \\ 11 \end{pmatrix} = \frac{1}{21} \begin{pmatrix} 3 + 0.8 + 2.2 \\ 0.6 + 1.2 + 2.2 \\ 2.4 + 2 + 6.6 \end{pmatrix} = \begin{pmatrix} 6/21 \\ 4/21 \\ 11/21 = L(0) \end{pmatrix}$$

Donc $L(0)$ est une mesure stationnaire de la chaîne.

4. La matrice Π est primitive. On peut le montrer via les propriétés sur les graphes, mais surtout on peut remarquer que tous les coefficients de Π sont strictement positifs donc elle est primitive par définition (Définition 1.3.7). Ainsi, la théorie de Perron-Frobenius permet d'affirmer que la loi $L(k)$ de la chaîne au temps k converge, lorsque k tend vers l'infini vers l'unique loi stationnaire $(6/21, 4/21, 11/21)$. Conclusion : en temps long la proportion de zèbre va passer de $3/4$ à $11/21$, elle va donc décroître (multiplication par $44/63 < 1$).