

Statistiques des processus 3A

2 novembre 2015

Table des matières

1	Les chaînes de Markov	5
1.1	Rappels sur les chaînes de Markov	5
1.1.1	Définition et premiers exemples	5
1.2	Irréductibilité, transience et récurrence	10
1.2.1	Récurrence positive et probabilité invariante	13
1.2.2	Exemple : les chaînes de naissance et de mort	17
1.3	Chaînes de Markov absorbantes	19
1.3.1	Quelques résultats	19
1.3.2	Exemple : le modèle de Wright-Fisher	20
1.4	Théorèmes limites et application à la statistique des chaînes de Markov	22
1.4.1	Le théorème ergodique	22
1.4.2	Convergence vers la loi stationnaire	23
1.4.3	Un théorème limite centrale pour les chaînes de Markov	24
1.4.4	Estimation de la matrice de transition par la méthode du maximum de vraisemblance	27
1.4.5	Construction d'un test d'adéquation.	29
1.5	Application à la recherche des mots de fréquence exceptionnelle dans les séquences ADN	30
1.5.1	Objectifs et formalisation du problème	30
1.5.2	Loi exacte du nombre d'occurrences	32
1.5.3	Utilisation de l'approximation gaussienne	33
1.5.4	Loi des petits nombres	34
1.6	Exercices	35
2	Processus de Poisson	39
2.1	Le processus de Poisson simple	39
2.1.1	Estimation de l'intensité par maximum de vraisemblance	43
2.2	Le processus de Poisson non homogène	45
2.2.1	Définition et propriétés fondamentales	45
2.2.2	Estimation par maximum de vraisemblance	48
2.3	Processus de Poisson composé et modèle de Cramér-Lundberg	51
2.3.1	Définition	51
2.3.2	Modèle de Cramér-Lundberg et probabilité de ruine	52

3	Processus Markoviens de sauts	53
3.1	Propriété de Markov d'une fonction aléatoire de sauts	54
3.2	Le générateur infinitésimal	57
3.3	Existence et construction d'un processus de Markov de générateur donné	60
	3.3.1 Trois exemples de construction	60
	3.3.2 Conditions de non-explosion	62
3.4	Estimation du générateur par maximum de vraisemblance	63
3.5	Classification des états et Théorème ergodique	64
3.6	Exemples	66
	3.6.1 Processus à deux états	66
	3.6.2 Processus de naissance et de mort	67
	3.6.3 Dynamique d'une épidémie	68
	3.6.4 Processus de Galton-Watson à temps continu	69
3.7	Exercices	72

Chapitre 1

Les chaînes de Markov

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Dans la suite nous utiliserons les deux abréviations suivantes : i.i.d pour indépendantes et identiquement distribuées et p.s pour presque sûrement.

1.1 Rappels sur les chaînes de Markov

1.1.1 Définition et premiers exemples

Dans tout ce chapitre E désignera un ensemble fini ou infini dénombrable (par exemple $E = \{a, b, c, d\}$, $E = \mathbb{N}$ ou $E = \mathbb{Z}$).

Définition 1 On dit qu'une suite de variables aléatoires $\mathbf{X} = (X_n)_{n \in \mathbb{N}}$, toutes à valeurs dans E , est une chaîne de Markov si pour tout $n \in \mathbb{N}$ et tout $(x_0, x_1, \dots, x_{n+1}) \in E^{n+2}$ tel que $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$,

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n).$$

Autrement dit, une chaîne de Markov est une suite de variables aléatoires pour laquelle la loi conditionnelle de la $n + 1$ -ième coordonnée sachant toutes les coordonnées précédentes ne dépend que de la n -ième, et ce pour tout n . L'indice n représente souvent le temps mais pas toujours.

Remarque. On dit que trois variables aléatoires discrètes A, B, C sont telles que A est indépendante de B conditionnellement à C si

$$\mathbb{P}(A = a, B = b | C = c) = \mathbb{P}(A = a | C = c) \times \mathbb{P}(B = b | C = c),$$

pour tout $c \in E$ tel que $\mathbb{P}(C = c) > 0$. On pourra vérifier que cette dernière propriété est équivalente à avoir $\mathbb{P}(A = a | B = b, C = c) = \mathbb{P}(A = a | C = c)$ pour tout $(b, c) \in E^2$ tel que $\mathbb{P}(B = b, C = c) > 0$. Ainsi, une chaîne de Markov est une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ telle que pour tout $n \in \mathbb{N}$, X_{n+1} est indépendante de X_{n-1}, \dots, X_0 conditionnellement à X_n . Dans un contexte temporel, on peut donc interpréter cette définition en disant que "conditionnellement au présent, le futur est

indépendant du passé".

Dans la plupart des exemples considérés dans ce cours, la loi conditionnelle de X_{n+1} sachant X_n sera indépendante de n , ce qui donne lieu à la définition suivante.

Définition 2 1. On dit qu'une "matrice" $(P(x, y))_{(x, y) \in E^2}$ est stochastique si $P(x, y) \geq 0$ pour tout $(x, y) \in E^2$ et si $\sum_{y \in E} P(x, y) = 1$ pour tout $x \in E$.

2. On dit qu'une chaîne de Markov \mathbf{X} est homogène s'il existe une matrice stochastique P telle que pour tout $(n, x, y) \in \mathbb{N} \times E \times E$ tel que $\mathbb{P}(X_n = x) > 0$, alors

$$\mathbb{P}(X_{n+1} = y | X_n = x) = P(x, y).$$

Convention. Lorsque $\mathbb{P}(X_n = x) = 0$, on conviendra que

$$\mathbb{P}(X_{n+1} = y | X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \mathbb{P}(X_{n+1} = y | X_n = x) = P(x, y).$$

La matrice P est souvent appelée matrice de transition ou plus simplement transition.

La figure 1.1.1 compare une trajectoire obtenue à partir de 100 variables aléatoires i.i.d toutes de loi de Bernoulli de paramètre $\frac{1}{2}$ (graphe de gauche) et une trajectoire de longueur 100 d'une chaîne

de Markov sur $E = \{0, 1\}$ telle que $P = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$ (graphe de droite). Le cas des Bernoulli

indépendantes correspond en fait à $P = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$.

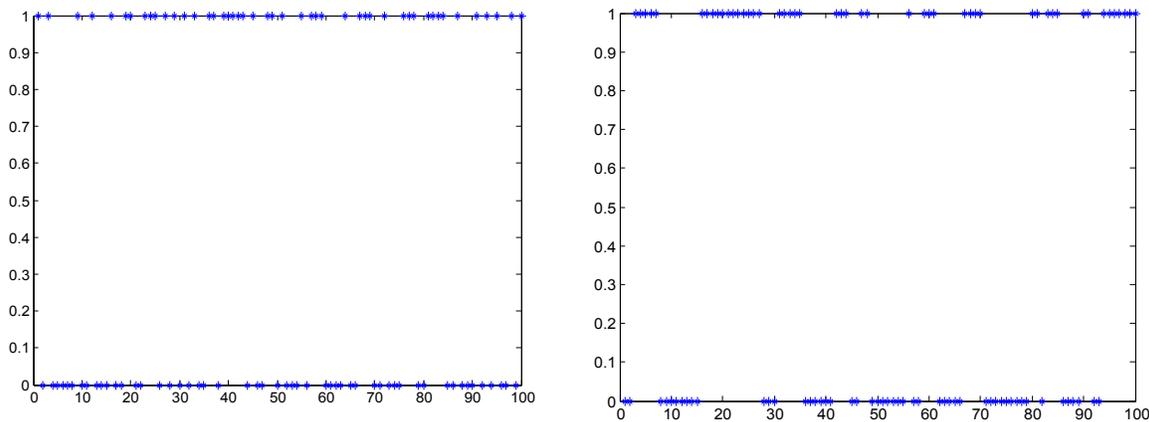


FIGURE 1.1 – Comparaison de trajectoires de deux chaînes sur $E = \{0, 1\}$

Exemples

1. L'exemple classique est celui de la marche aléatoire sur le réseau \mathbb{Z}^d . A partir d'une suite $(\xi_n)_{n \geq 1}$ de variables aléatoires i.i.d à valeurs dans \mathbb{Z}^d et indépendante d'une autre variable aléatoire X_0 (elle aussi à valeurs dans \mathbb{Z}^d), on définit une suite de variables aléatoires à l'aide

des relations $X_{n+1} = X_n + \xi_{n+1}$, $n \in \mathbb{N}$. On obtient alors une chaîne de Markov homogène dont la transition P est donnée par

$$P(x, y) = \mathbb{P}(\xi_1 = y - x), \quad (x, y) \in \mathbb{Z}^d.$$

Dans le cas où $d = 1$ et $\mathbb{P}(\xi_1 = 1) = \mathbb{P}(\xi_1 = -1) = \frac{1}{2}$, on parle de marche aléatoire simple symétrique. Dans ce cas, on obtient $P(x, y) = \begin{cases} 0 & \text{si } |x - y| \neq 1 \\ \frac{1}{2} & \text{si } |x - y| = 1 \end{cases}$

2. On considère la situation suivante. Entre deux instants donnés notés n et $n + 1$, $q \in \mathbb{N}^*$ pièces sont fabriquées en usine. On note D_{n+1} le nombre de pièces (aléatoire) que les clients achètent entre les instants n et $n + 1$. Si X_n désigne le nombre de pièces en stock au temps n , alors on a la relation

$$X_{n+1} = \max(X_n + q - D_{n+1}, 0), \quad n \in \mathbb{N}.$$

En supposant les variables aléatoires D_1, D_2, \dots i.i.d, on obtient un chaîne de Markov homogène à valeurs dans $E = \mathbb{N}$ et dont la matrice de transition P est donnée par $P(x, 0) = \mathbb{P}(D_1 \geq x + q)$ et $P(x, y) = \mathbb{P}(D_1 = x + q - y)$ si $0 < y \leq x + q$.

3. Le processus de Galton-Watson est un processus de branchement simple qui permet de modéliser l'évolution du nombre de descendants d'un individu donné. Initialement, il a été proposé par F. Galton et H.W. Watson vers 1873 pour étudier le problème suivant.

Soit $p_0, p_1, p_2 \dots$ les probabilités respectives pour qu'un homme ait 0, 1, 2, ... enfants et supposons qu'il en soit de même pour ses fils, les fils de ses fils, etc. Quelle est la probabilité que la descendance mâle s'éteigne au bout de r générations ?

L'objectif initial était l'étude de l'extinction des noms de familles nobles en Angleterre. Dans ce problème, le nombre de descendants peut être représenté de la façon suivante. On pose $X_0 = 1$ et pour tout entier n , on définit récursivement $X_{n+1} = \sum_{i=1}^{X_n} Y_{n,i} \mathbb{1}_{X_n > 0}$ où $\{Y_{n,i} : (n, i) \in \mathbb{N}^* \times \mathbb{N}^*\}$ est une famille de variables aléatoires i.i.d et toutes de loi discrète $(p_k)_{k \in \mathbb{N}}$. La suite $(X_n)_{n \in \mathbb{N}}$ est alors une chaîne de Markov dont la matrice de transition P est donnée par

$$P(x, y) = \mathbb{P}\left(\sum_{i=1}^x Y_{1,i} = y\right) = \sum_{j_1 + \dots + j_x = y} p_{j_1} \cdots p_{j_x}, \quad (x, y) \in \mathbb{N}^* \times \mathbb{N},$$

et $P(0, y) = \mathbb{1}_{y=0}$.

4. En génétique, le modèle de Wright-Fisher est utilisé pour étudier l'évolution de la proportion de deux allèles A et B dans une population haploïde (i.e chaque individu possède un seul exemplaire de chaque double-brin d'ADN) et dont la taille reste à peu près constante d'une génération sur l'autre. Plus précisément, si N désigne la taille de cette population et X_n le nombre d'individus possédant l'allèle A à la génération $n \in \mathbb{N}$, on estime que la loi conditionnelle $X_{n+1} | X_n, X_{n-1}, \dots, X_0$ est une loi binomiale de paramètres N et $\frac{X_n}{N}$. Schématiquement, cela revient à considérer que tout individu i de la génération $n + 1$ provient, indépendamment des autres, du parent j de la génération n avec probabilité $\frac{1}{N}$ et hérite de l'allèle correspondant. La suite $(X_n)_{n \in \mathbb{N}}$ est alors une chaîne de Markov par construction.

Plus généralement, nous avons le résultat suivant.

Proposition 1 Soient $(U_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires i.i.d à valeurs dans un espace mesurable G , indépendante d'une variable aléatoire X_0 à valeurs dans E et $F : E \times G \rightarrow E$ une fonction mesurable. Alors si

$$X_{n+1} = F(X_n, U_{n+1}), \quad n \in \mathbb{N},$$

la suite $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov homogène.

Preuve. Pour $n \in \mathbb{N}$ et $x_0, x_1, \dots, x_n \in E$, posons $A_n = \{X_0 = x_0, \dots, X_n = x_n\}$. Alors

$$\begin{aligned} \mathbb{P}(X_{n+1} = x_{n+1} | A_n) &= \mathbb{P}(F(x_n, U_{n+1}) = x_{n+1} | A_n) \\ &= \mathbb{P}(F(x_n, U_{n+1}) = x_{n+1}) \\ &= \mathbb{P}(F(x_n, U_1) = x_{n+1}). \end{aligned}$$

En posant $P(x_n, x_{n+1}) = \mathbb{P}(F(x_n, U_1) = x_{n+1})$, on voit que la suite $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov de matrice de transition P . \square

Notations. Dans la suite, P^n désignera la puissance n de la matrice de transition P . De plus, si μ est une mesure sur E , la mesure μP^n désignera la mesure sur E définie par

$$\mu P^n(x) = \sum_{z \in E} \mu(z) P^n(z, x), \quad x \in E.$$

Proposition 2 Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov de loi initiale μ (i.e μ est la loi de X_0) et de matrice de transition P . On a alors les formules suivantes.

1. Pour tout $(x_0, x_1, \dots, x_n) \in E^{n+1}$, on a

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mu(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

2. Pour tout $(n, p) \in \mathbb{N}^2$, $\mathbb{P}(X_{n+p} = y | X_p = x) = P^n(x, y)$.

3. $\mathbb{P}(X_n = y) = \mu P^n(y)$.

D'après le point 1 de la Proposition 2, on voit que la loi initiale μ et la matrice de transition P détermine complètement les lois fini-dimensionnelles de la chaîne de Markov. On admettra que étant donné une mesure μ sur E et une matrice markovienne P , il est toujours possible de considérer un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ sur lequel est définie une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ qui soit une chaîne de Markov de loi initiale μ et de matrice de transition P .

Preuve de la Proposition 2

1. Il suffit d'appliquer la formule des conditionnements successifs et d'utiliser la définition des chaînes de Markov. Nous avons, en notant $A_i = \{X_0 = x_0, \dots, X_i = x_i\}$,

$$\begin{aligned} & \mathbb{P}(X_0 = x_0, \dots, X_n = x_n) \\ &= \mathbb{P}(X_n | A_{n-1}) \mathbb{P}(X_{n-1} = x_{n-1} | A_{n-2}) \cdots \mathbb{P}(X_1 = x_1 | A_0) \mathbb{P}(X_0 = x_0) \\ &= \mathbb{P}(X_0 = x_0) \mathbb{P}(X_1 = x_1 | X_0 = x_0) \cdots \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \\ &= \mu(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n). \end{aligned}$$

2. Toujours en utilisant la formule des conditionnements successifs, nous avons

$$\begin{aligned} & \mathbb{P}(X_{n+p} = x_{n+p} | X_p = x_p) \\ &= \sum_{x_{p+1}, \dots, x_{n+p-1} \in E} \mathbb{P}(X_{p+1} = x_{p+1}, \dots, X_{n+p} = x_{n+p} | X_p = x_p) \\ &= \sum_{x_{p+1}, \dots, x_{n+p-1} \in E} \mathbb{P}(X_{n+p} = x_{n+p} | X_p = x_p, \dots, X_{n+p-1} = x_{n+p-1}) \cdots \mathbb{P}(X_{p+1} = x_{p+1} | X_p = x_p) \\ &= \sum_{x_{p+1}, \dots, x_{n+p-1} \in E} P(x_p, x_{p+1}) \cdots P(x_{n+p-1}, x_{n+p}) \\ &= P^n(x_p, x_{n+p}). \end{aligned}$$

3. C'est une conséquence du point précédent en remarquant que

$$\mathbb{P}(X_n = y) = \sum_{x \in E} \mathbb{P}(X_n = y, X_0 = x) = \sum_{x \in E} \mathbb{P}(X_n = y | X_0 = x) \mathbb{P}(X_0 = x). \square$$

Exercice 1

Soit \mathbf{X} une chaîne de Markov homogène. Montrer les propriétés suivantes.

1. $\mathbb{P}(X_{n+p} = x_{n+p}, \dots, X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0)$ ne dépend pas de x_0, \dots, x_{n-1} .
2. Déterminer la loi du vecteur $Y_n = (X_{i_1}, X_{i_2}, \dots, X_{i_n})$ si $0 \leq i_1 < i_2 < \dots < i_n$.

Exercice 2

Soit \mathbf{X} une suite de variables aléatoires à valeurs dans E et telle que pour $n \geq p$,

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_{n-p} = x_{n-p}).$$

Montrer que la suite $(Y_n)_{n \in \mathbb{N}}$ définie par $Y_n = (X_n, X_{n+1}, \dots, X_{n+p})$ est une chaîne de Markov.

La propriété suivante que nous admettrons est appelée propriété de Markov simple. Il est parfois utile de considérer des fonctions qui dépendent de tout le futur.

Proposition 3 Si $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov homogène et $g : E^{\mathbb{N}} \rightarrow \mathbb{R}$ une fonction mesurable bornée (pour simplifier) alors

$$\mathbb{E}(g(X_n, X_{n+1}, \dots) | X_n = x_n, \dots, X_0 = x_0) = \mathbb{E}(g(X_0, X_1, \dots) | X_0 = x_0).$$

1.2 Irréductibilité, transience et récurrence

Définition 3 Une chaîne de Markov \mathbf{X} est dite irréductible si pour tout $(x, y) \in E^2$, il existe un entier $n = n(x, y) \geq 1$ tel que $P^n(x, y) = \mathbb{P}(X_n = y | X_0 = x) > 0$.

La définition précédente signifie que la chaîne \mathbf{X} peut passer de tout état x à tout état y avec une probabilité positive. Remarquons que

$$P^n(x, y) = \sum_{x_1, \dots, x_{n-1} \in E} P(x, x_1)P(x_1, x_2) \cdots P(x_{n-1}, y).$$

La définition précédente signifie alors que pour tout $(x, y) \in E^2$, il existe $n \in \mathbb{N}^*$ ainsi que $x_1, \dots, x_{n-1} \in E$ tel que $P(x_i, x_{i+1}) > 0$, pour $i = 0, \dots, n-1$ (en convenant que $x_0 = x$ et $x_n = y$).

Exemples

1. La marche aléatoire $X_{n+1} = X_n + \xi_n$ sur \mathbb{Z}^d telle que $\mathbb{P}(\xi_1 = e_i), \mathbb{P}(\xi_1 = -e_i) > 0$ où e_1, \dots, e_d sont les vecteurs de la base canonique de \mathbb{R}^d est irréductible.
2. La chaîne à trois états de matrice de transition $P = \begin{pmatrix} 0.5 & 0 & 0.5 \\ 1 & 0 & 0 \\ 0 & 0.1 & 0.9 \end{pmatrix}$ est irréductible.
3. La chaîne de Galton-Watson n'est pas irréductible. En effet l'état 0 ne peut mener qu'à lui-même. Lorsque pour une chaîne de Markov de transition P , on a $P(x, x) = 1$, on dit que l'état x est absorbant.

La notion suivante joue un rôle fondamental pour analyser le comportement en temps long d'une chaîne de Markov. Elle fait intervenir le temps de retour en un état $x \in E$ défini par

$$T_x = \inf \{n \geq 1 : X_n = x\}.$$

Remarquer que T_x est une variable aléatoire à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$ (on convient que $\inf \emptyset = +\infty$). En effet, on a

$$\{T_x = k\} = \{X_1 \neq x, \dots, X_{k-1} \neq x, X_k = x\}.$$

Dans la suite, nous aurons également besoin de la loi d'une chaîne de Markov de transition P lorsque X_0 est constante et égale à x . Cette loi, notée \mathbb{P}_x , correspond à la loi conditionnelle $\mathbb{P}(\cdot | X_0 = x)$ lorsque sous \mathbb{P} , \mathbf{X} est une chaîne de Markov de transition P et de loi initiale quelconque.

Définition 4 Un état $x \in E$ est dit récurrent lorsque $\mathbb{P}_x(T_x < +\infty) = 1$. Sinon, on dit que x est transitoire.

On interprète la définition précédente de la façon suivante. Si $x \in E$ est récurrent, on est sûr la chaîne qui part de x repassera par cet état. Si x est transitoire, il existe un évènement de probabilité positive sur lequel toute trajectoire de la chaîne de repasse jamais en x .

Remarquons que la notion de récurrence ne dépend que de la transition P de la chaîne (et pas de la loi initiale).

Dans la suite pour un état $x \in E$, nous noterons

$$U_x = \sum_{n \geq 1} \mathbb{1}_{X_n=x},$$

le nombre de retours de la chaîne au point x . La notation \mathbb{E}_x désignera l'espérance sous la probabilité \mathbb{P}_x .

Proposition 4 *Soit \mathbf{X} une chaîne de Markov irréductible de loi initial μ quelconque. On a les deux cas de figures suivants.*

1. *Tous les états sont récurrents. Dans ce cas, p.s, pour tout $x \in E$, la chaîne de Markov passe x une infinité de fois. On dit que la chaîne est récurrente.*
2. *Tous les états sont transitoires. Dans ce cas, p.s, pour tout $x \in E$, la chaîne ne passe qu'un nombre fini de fois par x . On dit que la chaîne est transiente.*

Remarques

1. On a alors la conséquence suivante. **Lorsque E est fini, une chaîne irréductible est automatiquement récurrente.** En effet, si la chaîne était transitoire, elle ne passerait qu'un nombre fini de fois par chaque état, ce qui est incompatible avec la finitude de E .
2. Si $E = \mathbb{N}$ ou $E = \mathbb{Z}$, une chaîne transiente vérifie $\lim_{n \rightarrow +\infty} |X_n| = +\infty$ p.s.

Preuve de la Proposition 4. Pour $x \in E$, posons

$$U(x) = \sum_{n \geq 1} \mathbb{1}_{\{X_n=x\}}.$$

$U(x)$ représente le nombre de visites au point $x \in E$ (on exclut l'état initial). Remarquons que pour tout $y \in E$,

$$\mathbb{E}_y(U(x)) = \sum_{n \geq 1} \mathbb{P}(X_n = x | X_0 = y) = \sum_{n \geq 1} P^n(y, x).$$

La preuve est en partie basée sur le lemme suivant.

Lemme 1 *On considère une chaîne de Markov irréductible \mathbf{X} . Soient $x \in E$ un état de la chaîne. L'état $x \in E$ est récurrent (resp. transitoire) si et seulement si $\mathbb{P}_x(U(x) = +\infty) = 1$ (resp. $= 0$). Dans ce cas, on a $\mathbb{E}_x(U(x)) = +\infty$ (resp. $< +\infty$).*

Preuve du Lemme 1. Remarquons qu'en posant $F_n = \{X_n = x, X_k \neq x; k \geq n + 1\}$ pour tout entier $n \geq 1$ et $G = \{X_k \neq x; k \geq 1\}$, on a

$$\{U(x) < +\infty\} = \cup_{n \geq 1} F_n \cup G.$$

Les F_n étant disjoints deux à deux et disjoints de G , la formule des probabilités totales entraîne l'égalité

$$\mathbb{P}_x(U(x) < +\infty) = \mathbb{P}_x(G) + \sum_{n \geq 1} \mathbb{P}_x(F_n).$$

De plus, en utilisant la propriété de Markov et les propriétés des lois conditionnelles, on a

$$\mathbb{P}_x(F_n) = \mathbb{P}(X_n = x | X_0 = x) \mathbb{P}(X_k \neq x; k \geq 1 | X_0 = x) = P^n(x, x) \mathbb{P}_x(G)$$

et donc

$$\mathbb{P}_x(U(x) < +\infty) = \left(1 + \sum_{n \geq 1} P^n(x, x) \right) \mathbb{P}_x(G) = (1 + \mathbb{E}_x(U(x))) \mathbb{P}_x(G).$$

– Supposons d'abord x transient. Alors, par définition, $\mathbb{P}_x(G) > 0$ et donc

$$\mathbb{E}_x(U(x)) = \frac{\mathbb{P}_x(U(x) < +\infty)}{\mathbb{P}_x(G)} - 1 < +\infty.$$

On en déduit que $\mathbb{P}_x(U(x) = +\infty) = 0$.

– Supposons ensuite x récurrent. Dans ce cas $\mathbb{P}_x(G) = 0$. On déduit des calculs précédents que $\mathbb{P}_x(U(x) < +\infty) = 0$ et donc que $\mathbb{E}_x(U(x)) = +\infty$. \square

Fin de la preuve de la Proposition 4. Nous allons utiliser le Lemme 1, en montrant d'abord que pour deux états distincts x et y , les quantités $\mathbb{E}_x(U(x))$ et $\mathbb{E}_y(U(y))$ sont ou bien toutes les deux finies, ou bien toutes les deux infinies. Cela revient à montrer que les séries $\sum_{n \geq 1} P^n(x, x)$ et $\sum_{n \geq 1} P^n(y, y)$ sont de même nature. Par irréductibilité de la chaîne, il existe deux entiers r et s tels que $P^r(x, y) > 0$ et $P^s(y, x) > 0$. En utilisant les deux inégalités

$$P^{n+r+s}(x, x) \geq P^r(x, y) P^n(y, y) P^s(y, x), \quad P^{n+r+s}(y, y) \geq P^s(y, x) P^n(x, x) P^r(x, y),$$

on conclut aisément que les séries $\sum_{n \geq 1} P^n(x, x)$ et $\sum_{n \geq 1} P^n(y, y)$ sont ou bien toutes les deux convergentes ou bien toutes les deux divergentes. D'après le Lemme 1, on conclut que les deux points sont soit tous les deux transitoires, soit tous les deux récurrents. Tous les états de la chaîne sont donc de même nature. Nous allons ensuite étudier la quantité $\mathbb{P}_y(U(x) < +\infty)$.

Supposons d'abord la chaîne transiente. Comme $P^{n+r}(y, y) \geq P^n(y, x) P^r(x, y)$ et $\sum_{n \geq 1} P^n(x, x) < +\infty$, on en déduit que

$$\sum_{n \geq 1} P^n(y, x) = \mathbb{E}_y(U(x)) < +\infty$$

et donc que $\mathbb{P}_y(U(x) < +\infty) = 1$.

Supposons maintenant la chaîne récurrente. Nous allons prouver que $\mathbb{P}_y(U(x) = +\infty) = 1$.

– Nous commençons par montrer que $\mathbb{P}_y(T_x = +\infty) = 0$. Posons $T_y^{(1)} = T_y$ et si ≥ 1 ,

$$T_y^{(k+1)} = \inf \{ n > T_y^{(k)} : X_n = y \}.$$

La suite $(T_y^{(k)})_{k \geq 1}$ désigne la suite des temps de retour successifs en y . D'après ce qui précède, sous \mathbb{P}_y , tous ces temps sont finis. Notons $c = \mathbb{P}_y(T_x > T_y)$. Remarquons que $c < 1$. En

effet, par irréductibilité, il existe $n \in \mathbb{N}^*$ et $z_1, \dots, z_{n-1} \in E \setminus \{y\}$ tels que $P(z_i, z_{i+1}) > 0$ si $0 \leq i \leq n-1$ (en posant $z_0 = y$ et $z_n = x$). Mais $1 - c \geq P(z_0, z_1) \cdots P(z_{n-1}, z_n) > 0$ et donc $c < 1$. On a

$$\mathbb{P}_y(T_x = \infty) = \lim_{k \rightarrow +\infty} \mathbb{P}_y(T_x > T_y^{(k)}).$$

Nous allons montrer que $\mathbb{P}_y(T_x > T_y^{(k)}) = c^k$ ce qui conduira au résultat. On a

$$\begin{aligned} \mathbb{P}_y(T_x > T_y^{(k)}) &= \sum_{m_1 < \dots < m_k} \sum_{z_1, \dots, z_{m_k} \neq x, z_{m_i} = y} P(y, z_1) P(z_1, z_2) \cdots P(z_{m_1-1}, y) \cdots P(y, z_{m_{k-1}+1}) \cdots P(z_{m_k-1}, y) \\ &= \left(\sum_{h \in \mathbb{N}^*} \sum_{z_1, \dots, z_{h-1} \neq x} P(y, z_1) \cdots P(z_{h-1}, y) \right)^k \\ &= c^k. \end{aligned}$$

- Ensuite, en procédant comme dans la preuve du Lemme 1 (voir les notations des évènements F_n et G), nous pouvons montrer que

$$\mathbb{P}_y(U(x) < +\infty) = \mathbb{E}_y(U(x)) \mathbb{P}_x(G) + \mathbb{P}_y(G).$$

Comme $\mathbb{P}_x(G) = 0$ par définition de la récurrence et $\mathbb{P}_y(G) = \mathbb{P}_y(T_x = +\infty) = 0$ d'après ce qui précède, on conclut que $\mathbb{P}_y(U(x) = +\infty) = 1$.

Pour terminer la preuve de la proposition, considérons une loi initiale μ arbitraire. On a

$$\mathbb{P}(U(x) = +\infty) = \sum_{y \in E} \mu(y) \mathbb{P}_y(U(x) = +\infty).$$

Lorsque la chaîne est transiente (resp. récurrente), nous avons montré précédemment que pour tout $y \in E$, $\mathbb{P}_y(U(x) = +\infty) = 0$ (resp. 1). On a donc $\mathbb{P}(U(x) = +\infty) = 0$ ou 1 suivant que la chaîne soit transiente ou récurrente. Ceci termine la preuve. \square

Exercice 3

Pour une chaîne de Markov \mathbf{X} récurrente irréductible de loi initiale μ , on considère la suite des temps de retour successifs au point $x \in E$, soit

$$T_x^{(1)} = \inf \{n \geq 1 : X_n = x\},$$

$$T_x^{(k)} = \inf \{n > T_x^{(k-1)} : X_n = x\}, \quad k \geq 2.$$

On convient que $T_x^{(0)} = 0$. Montrer que les variables aléatoires $S_k = T_x^{(k)} - T_x^{(k-1)}$, $k \geq 1$, sont indépendantes et que pour tout entier $k \geq 2$, la loi de S_k correspond à la loi de conditionnelle de S_1 sachant $X_0 = x$. On pourra se contenter de prouver le résultat pour les deux premiers délais, en détaillant l'évènement $\{S_1 = p, S_2 = q\}$ et en utilisant la propriété de Markov.

1.2.1 Récurrence positive et probabilité invariante

Pour une chaîne de Markov irréductible, le nombre moyen de visites de X_1, \dots, X_n au point x converge p.s. On a plus précisément le résultat suivant.

Proposition 5 Soit X une chaîne de Markov irréductible. Alors

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{X_k=x\}} = \frac{1}{\mathbb{E}_x(T_x)}, \text{ p.s.}$$

On a alors les deux cas suivants.

- Soit $\mathbb{E}_x(T_x) < +\infty$ pour tout $x \in E$ (dans ce cas, tous les états sont récurrents). On dit alors que la chaîne est récurrente positive.
- Soit $\mathbb{E}_x(T_x) = +\infty$ pour tout $x \in E$. La chaîne peut être alors récurrente ou transiente. Si elle est récurrente, on dit qu'elle est récurrente nulle.

Preuve de la Proposition 5. Posons $m(n) = \sum_{k=1}^n \mathbb{1}_{\{X_k=x\}}$. La preuve est claire dans le cas où la chaîne est transiente puisque d'après la Proposition 4, on a $\mathbb{P}_x(T_x = +\infty) > 0$ pour tout $x \in E$ et donc $\mathbb{E}_x(T_x) = +\infty$ pour tout $x \in E$. De plus $\frac{m(n)}{n} \rightarrow 0$ p.s. car $m(n)_\omega$ s'annule pour n suffisamment grand.

Supposons ensuite la chaîne récurrente et considérons la variable aléatoire $T_x^{(n)} = \sum_{k=1}^n S_k$ qui représente le temps (ou l'indice) du n -ème retour au point x (voir l'exercice précédent pour la définition des variables $T_x^{(n)}$). En utilisant le résultat de l'exercice précédent ainsi que la loi des grands nombres, on sait que p.s., $\frac{S_n}{n} \rightarrow \mathbb{E}_x(T_x)$, la limite étant éventuellement égale à $+\infty$. Pour simplifier les notations, posons $V_n = T_x^{(n)}$. Comme par récurrence, $m(n) \rightarrow +\infty$ p.s., on a aussi $\frac{V_{m(n)}}{m(n)} \rightarrow \mathbb{E}_x(T_x)$ p.s. Remarquons aussi que si $m(n)_\omega = \ell$, on a $V_\ell(\omega) \leq n < V_{\ell+1}(\omega)$. En effet par définition de $m(n)$, on retourne ℓ fois en x avant l'indice n et le $(\ell + 1)$ -ème retour en x se fait pour un indice k de la chaîne tel que $k > n$. On a donc les inégalités entre variables aléatoires $V_{m(n)} \leq n < V_{m(n)+1}$ et donc

$$\frac{V_{m(n)}}{m(n)} \leq \frac{n}{m(n)} < \frac{V_{m(n)+1}}{m(n)} = \frac{V_{m(n)+1}}{m(n)+1} \cdot \frac{m(n)+1}{m(n)}.$$

Lorsque $n \rightarrow +\infty$, on voit alors que $\lim_{n \rightarrow +\infty} \frac{n}{m(n)} = \mathbb{E}_x(T_x)$ p.s. ce qui conduit à la convergence annoncée.

Enfin, posons $\pi(x) = \frac{1}{\mathbb{E}_x(T_x)}$ et vérifions que si $\mathbb{E}_x(T_x) = +\infty$ (i.e $\pi(x) = 0$) pour un état $x \in E$, alors $\mathbb{E}_y(T_y) = +\infty$ pour tout autre état $y \in E$. En utilisant le résultat de convergence démontré ci-dessus et le théorème de convergence dominée, on a pour tous les états $z \in E$,

$$\lim_{n \rightarrow +\infty} \mathbb{E}_z \left(\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{X_k=z\}} \right) = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n P^k(z, z) = \pi(z). \quad (1.1)$$

L'irréductibilité de la chaîne garantit l'existence de deux entiers r et s tels que $P^r(x, y) > 0$ et $P^s(y, x) > 0$. Comme de plus pour tout entier k , $P^{k+r+s}(x, x) \geq P^r(x, y)P^k(y, y)P^s(y, x)$, on déduit de (1.1) et de l'égalité $\pi(x) = 0$ que

$$\mathbb{E}_y \left(\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{X_k=y\}} \right) = \frac{1}{n} \sum_{k=1}^n P^k(y, y) \leq \frac{1}{nP^r(x, y)P^s(y, x)} \sum_{k=1}^n P^{k+r+s}(x, x) \rightarrow_{n \rightarrow +\infty} 0. \quad (1.2)$$

Comme on a aussi $\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n P^k(y, y) = \pi(y)$ p.s., on a nécessairement $\pi(y) = 0$ ou encore $\mathbb{E}_y(T_y) = +\infty$ ce qui achève la démonstration. \square

Définition 5 Soit P une matrice de transition sur E . Une probabilité π sur E est dite invariante si $\pi P = \pi$.

Remarques

1. La notion de probabilité invariante est liée à la notion de stationnarité. On pourra remarquer que si la loi initiale $\mu = \pi$ est une probabilité invariante, alors $\mu P^n = \mu$, autrement dit la loi de X_n est indépendante de n . A titre d'exercice, on vérifiera que pour tous entiers n, k , la loi du vecteur (X_n, \dots, X_{n+k}) est aussi la loi du vecteur (X_0, \dots, X_k) . On dit que la suite $(X_n)_{n \in \mathbb{N}}$ est stationnaire.
2. Une mesure π sur E qui vérifie l'équation $\pi P = \pi$ n'est pas unique puisque $\lambda \pi$ vérifie encore cette égalité pour tout réel λ . De plus si $\sum_{x \in E} \pi(x) < +\infty$ (automatique si E est fini), il est toujours possible de considérer une probabilité invariante : $\frac{\pi}{\sum_{x \in E} \pi(x)}$.

Corollaire 1 Une chaîne (irréductible) transiente ou récurrente nulle n'admet pas de probabilité invariante. Une chaîne récurrente positive admet une unique probabilité invariante π sur E donnée par $\pi(x) = \frac{1}{\mathbb{E}_x(T_x)}$.

Remarques

1. Lorsque E est fini, les résultats se simplifient considérablement. En effet, dans ce cas, la limite (que nous noterons $\pi(x)$) de la Proposition 5 ne peut être nulle pour tous les $x \in E$. Il existe donc au moins un état $x \in E$ tel que $\pi(x) > 0$. Mais cela signifie que la chaîne est récurrente positive (d'après la Proposition 5, la récurrence nulle et la transience correspondent à une limite identiquement nulle). On retiendra donc que **toute chaîne de Markov irréductible sur un espace d'état fini est récurrente positive**.
2. Lorsque E est infini, on peut déduire du Corollaire 1 un critère de récurrence positive pour les chaînes irréductibles. Une chaîne irréductible est récurrente positive si et seulement si l'équation $\pi P = \pi$ a une solution qui soit une probabilité.

Preuve du Corollaire 1. Remarquons tout d'abord que la convergence obtenue dans la Proposition 5 ainsi que le théorème de convergence dominée assure que

$$\frac{1}{n} \sum_{k=1}^n \mathbb{P}(X_k = x) = \frac{1}{n} \sum_{k=1}^n \mu P^k(x) \xrightarrow{n \rightarrow +\infty} \pi(x) = \frac{1}{\mathbb{E}_x(T_x)},$$

pour toute loi initiale μ et tout état $x \in E$. En particulier, pour tous états $x, y \in E$,

$$\frac{1}{n} \sum_{k=1}^n P^k(y, x) \xrightarrow{n \rightarrow +\infty} \pi(x).$$

1. Considérons le cas d'une chaîne transiente ou récurrente nulle. Supposons qu'il existe une probabilité invariante μ et que la loi de X_0 soit égale à μ . En utilisant la remarque ci-dessus, on a pour tout $x \in E$,

$$\mathbb{E} \left(\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{X_k=x} \right) = \frac{1}{n} \sum_{k=1}^n \mathbb{P}(X_k = x) \xrightarrow{n \rightarrow +\infty} 0.$$

Mais μ étant invariante, on a $\mathbb{P}(X_k = x) = \mu(x)$ et on obtiendrait $\mu(x) = 0$ pour tout x , ce qui est impossible pour une probabilité.

2. Supposons maintenant la chaîne récurrente positive. Si μ est une probabilité invariante alors, en utilisant les mêmes arguments que dans le point précédent, on montre que nécessairement $\mu = \pi$. La probabilité invariante est donc unique si elle existe. Reste à montrer que π est bien une probabilité invariante (pour l'instant, on sait juste que $\pi(x) \in [0, 1]$). Nous commençons par montrer que $\sum_{x \in E} \pi(x) \leq 1$. On a, en utilisant la Proposition 5 ainsi que le lemme de Fatou,

$$\begin{aligned} \sum_{x \in E} \pi(x) &= \sum_{x \in E} \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n \mu P^k(x) \\ &\leq \liminf_{n \rightarrow +\infty} \sum_{x \in E} \frac{1}{n} \sum_{k=1}^n \mu P^k(x) \\ &= \liminf_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n \sum_{x \in E} \mu(x) \\ &= 1. \end{aligned}$$

En utilisant les mêmes arguments, nous avons aussi

$$\begin{aligned} \sum_{x \in E} \pi(x) P(x, y) &= \sum_{x \in E} P(x, y) \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n \mathbb{P}(X_k = x) \\ &\leq \liminf_{n \rightarrow +\infty} \sum_{x \in E} P(x, y) \cdot \frac{1}{n} \sum_{k=1}^n \mu P^k(x) \\ &= \liminf_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n \mu P^{k+1}(y) \\ &= \pi(y). \end{aligned}$$

On en déduit que pour tout $y \in E$, $\pi P(y) \leq \pi(y)$. De plus, comme

$$\sum_{y \in E} \pi P(y) = \sum_{y \in E} \pi(y)$$

et $\sum_{y \in E} \pi(y) < +\infty$, on en déduit que $\pi P(y) = \pi(y)$ pour tout $y \in E$. Il reste à vérifier que π

est une probabilité. On a pour tout $n \geq 1$,

$$\begin{aligned}\pi(x) &= \frac{1}{n} \sum_{k=1}^n \pi P^k(x) \\ &= \sum_{y \in E} \pi(y) \frac{1}{n} \sum_{k=1}^n P^k(y, x).\end{aligned}$$

Si $n \rightarrow +\infty$, on peut appliquer le théorème de convergence dominée et on obtient

$$\pi(x) = \sum_{y \in E} \pi(y) \cdot \pi(x).$$

Ceci montre que $\sum_{y \in E} \pi(y) = 1$ et la probabilité π est bien l'unique probabilité invariante de la chaîne. \square

1.2.2 Exemple : les chaînes de naissance et de mort

On considère la chaîne de Markov à espace d'états $E = \mathbb{N}$ et dont la matrice de transition P est donnée par

$$P(n, n+1) = p_n, \quad P(n, n-1) = q_n \mathbb{1}_{n>0}, \quad P(n, n) = r_n = 1 - p_n - q_n.$$

On peut voir cette chaîne comme une modélisation simplifiée du nombre de personnes présentes dans une file d'attente ou du nombre de personnes connectées à un serveur au cours du temps (discrétisé). Des modèles plus réalistes (à temps continu) seront présentés dans le chapitre sur les files d'attentes mais les résultats pour cet exemple discret seront très utiles pour la suite.

Lorsque $p_n, q_{n+1} > 0$ pour tout $n \in \mathbb{N}$ (condition supposée par la suite), il est immédiat de vérifier que la chaîne est irréductible. Concernant les propriétés de récurrence de la chaîne, nous avons la proposition suivante.

Proposition 6 *La chaîne est récurrente si et seulement si $\sum_{n=1}^{+\infty} \gamma_n = +\infty$ avec $\gamma_n = \frac{q_1 \cdots q_n}{p_1 \cdots p_n}$. La chaîne est récurrente positive si et seulement si $s = \sum_{n=1}^{+\infty} \frac{p_0 \cdots p_{n-1}}{q_1 \cdots q_n} < +\infty$. Dans ce dernier cas, l'unique probabilité invariante π est donnée par*

$$\pi(n) = \frac{1}{s+1} \left(\mathbb{1}_{n=0} + \frac{p_0 \cdots p_{n-1}}{q_1 \cdots q_n} \mathbb{1}_{n \geq 1} \right).$$

On pourra vérifier que la condition $s < +\infty$ entraîne bien la condition $\sum_{n=1}^{+\infty} \gamma_n = +\infty$. Dans le cas où $p_n = p, q_n = q$, alors la chaîne est récurrente si et seulement si $p \leq q$ et récurrente positive si et seulement si $p < q$.

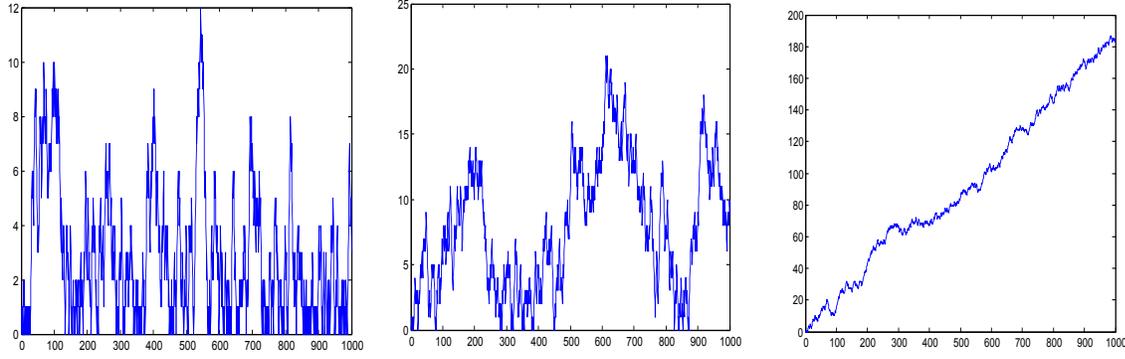


FIGURE 1.2 – Simulation de trajectoires de la chaîne de naissance et de mort lorsque $p_n = 1 - q_n = 0.4$, $p_n = 1 - q_n = 0.5$ et $p_n = 1 - q_n = 0.6$

Preuve de la Proposition 6.

1. Il est en fait plus facile de commencer par trouver la condition de récurrence positive. Il suffit pour cela de résoudre l'équation $\pi P = \pi$ et de regarder s'il existe une probabilité π solution. On obtient alors le système d'équations suivant.

$$\begin{cases} \pi(0)r_0 + \pi(1)q_1 = \pi(0) \\ \pi(n)r_n + \pi(n+1)q_{n+1} + \pi(n-1)p_{n-1} = \pi(n), \quad n \geq 1 \end{cases}$$

On trouve alors que

$$\pi(1) = \frac{p_0}{q_1}\pi(0), \quad \pi(n+1) = \frac{p_n + q_n}{q_{n+1}}\pi(n) - \frac{p_{n-1}}{q_{n+1}}\pi(n-1), \quad n \geq 1.$$

Par récurrence, on pourra vérifier que

$$\pi(n) = \pi(0) \frac{p_0 \cdots p_{n-1}}{q_1 \cdots q_n}, \quad n \geq 1.$$

Il existe alors une probabilité invariante si et seulement si $s < +\infty$. En utilisant l'égalité $s\pi(0) = 1 - \pi(0)$, on voit que cette (unique) probabilité est bien de la forme annoncée.

2. La récurrence est plus délicate à étudier. Nous allons chercher une condition qui assure que $\mathbb{P}_0(T_0 < +\infty) = 1$, où T_0 est le temps de retour en 0. Mais

$$\begin{aligned} \mathbb{P}_0(T_0 < +\infty) &= \mathbb{P}_0(T_0 < +\infty, X_1 = 0) + \mathbb{P}_0(T_0 < +\infty, X_1 = 1) \\ &= \mathbb{P}_0(X_1 = 0) + \mathbb{P}(T_0 < +\infty | X_1 = 1, X_0 = 0) \cdot \mathbb{P}(X_1 = 1 | X_0 = 0) \\ &= r_0 + \mathbb{P}_1(T_0 < +\infty) p_0. \end{aligned}$$

Dans la dernière égalité, nous avons utilisé la propriété de Markov simple. On voit alors que $\mathbb{P}_0(T_0 < +\infty) = 1$ si et seulement si $\mathbb{P}_1(T_0 < +\infty) = 1$. Remarquons que $\mathbb{P}_1(T_0 < +\infty) = \mathbb{P}_1(\tau_0 < +\infty)$ où pour entier m , $\tau_m = \inf \{k \geq 0 : X_k = m\}$ désigne le temps d'atteinte de l'état m (l'utilisation du temps d'atteinte sera plus commode dans ce qui suit). L'idée est

alors d'évaluer $u_b(n) = \mathbb{P}_n(\tau_0 < \tau_b)$ pour $0 \leq n \leq b$ et de faire tendre b vers l'infini. On sait déjà que $u_b(0) = 1$ et $u_b(b) = 0$. En utilisant les propriétés des lois conditionnelles et la propriété de Markov simple, nous avons si $0 < n < b$,

$$\begin{aligned} u_b(n) &= p_n \mathbb{P}_n(\tau_0 < \tau_b | X_1 = n+1) + r_n \mathbb{P}_n(\tau_0 < \tau_b | X_1 = n) + q_n \mathbb{P}_n(\tau_0 < \tau_b | X_1 = n-1) \\ &= p_n u_b(n+1) + r_n u_b(n) + q_n u_b(n-1). \end{aligned}$$

Posons $v_b(n) = u_b(n) - u_b(n-1)$ pour $1 \leq n \leq b$. Les relations précédentes deviennent alors

$$v_b(n+1) = \frac{q_n}{p_n} v_b(n), \quad b-1 \geq n \geq 1, \quad \sum_{j=1}^b v_b(j) = u_b(b) - u_b(0) = -1.$$

On en déduit que si $2 \leq n \leq b$, $v_b(n) = \gamma_{n-1} v_b(1)$ puis que $v_b(1) = \frac{-1}{1 + \sum_{j=1}^{b-1} \gamma_j}$. Il vient alors

$$u_b(n) = - \sum_{j=n+1}^b v_b(j) = \frac{\sum_{j=n}^{b-1} \gamma_j}{1 + \sum_{j=1}^{b-1} \gamma_j}.$$

On voit alors que $\lim_{b \rightarrow +\infty} u_b(n) = 1$ si et seulement si $\sum_{n \geq 1} \gamma_n$ est divergente. Comme $\lim_{b \rightarrow +\infty} u_b(n) = \mathbb{P}_n(\tau_0 < +\infty) = \mathbb{P}_n(T_0 < +\infty)$ si $n \geq 1$, la condition $\sum_{n \geq 1} \gamma_n = +\infty$ est bien une condition nécessaire et suffisante de récurrence pour cette chaîne. \square

Exercice 4

On considère maintenant la chaîne de naissance et de mort mais lorsque $E = \{0, 1, \dots, N\}$. Il suffit de poser $p_n = r_n = 0$ pour tout entier $n \geq N$ dans la définition de la chaîne initiale.

1. Que peut-on dire de cette nouvelle chaîne ?
2. Quelle est sa probabilité invariante ?

1.3 Chaînes de Markov absorbantes

1.3.1 Quelques résultats

Définition 6 Un état x est dit absorbant si $P(x, x) = 1$. Une chaîne est dite absorbante si pour tout état, un état absorbant est atteignable depuis cet état.

Formalisons un peu. Si $E = \{1, \dots, k\}$ et $1, 2, \dots, m$ sont les états absorbants, alors on a $P = \begin{pmatrix} I & 0 \\ A & T \end{pmatrix}$, avec I matrice identité de taille m . On peut alors vérifier par récurrence que

$$P^n = \begin{pmatrix} I_m & 0 \\ (I + T + \dots + T^{n-1})A & T^n \end{pmatrix}.$$

Proposition 7 La matrice $I - T$ est inversible, d'inverse $F = I + T + T^2 + \dots$. On a alors

$$F(i, j) = \mathbb{E}_i \left(\sum_{n \geq 0} \mathbb{1}_{X_n = j} \right), \quad i, j \geq m+1.$$

De plus, $\sum_{j=m+1}^k F(i, j)$ est le nombre moyen de pas avant absorption, lorsque l'état initial est i .

Preuve. Pour montrer que T est inversible, nous allons montrer que

$$\lim_{n \rightarrow \infty} T^n = 0. \quad (1.3)$$

Supposons d'abord (1.3). Dans ce cas, $I - T$ est injective puisque $(I - T)x = 0$ entraîne $Tx = x$ et donc $T^n x = x$ et donc $x = 0$ en faisant tendre n vers l'infini. En notant $F = (I - T)^{-1}$, l'égalité $(I - T)(I + T + \dots + T^n) = I - T^{n+1}$ entraîne $I + T + \dots + T^n = F(I - T^{n+1})$, série qui converge donc vers F . De plus si i et j sont non absorbants (donc $\geq m + 1$), on a $P(i, j) = T(i, j)$ et entre deux états non absorbants, on ne peut avoir que des états non absorbants ; ceci entraîne donc $T^n(i, j) = P^n(i, j) = \mathbb{E}_i(\mathbb{1}_{X_n=j})$ pour tout entier n . La formule donnée pour $F(i, j)$ est donc claire. Enfin on a

$$\sum_{j=m+1}^k F(i, j) = \mathbb{E}_i \left(\sum_{n \geq 0} \mathbb{1}_{X_n \geq m+1} \right),$$

qui représente le temps moyen écoulé avant le premier état absorbant. Il reste à prouver (1.3). Pour cela, on introduit

$$\tau = \inf \{n \geq 0 : 1 \leq X_n \leq m\},$$

le temps d'atteinte de l'ensemble des points absorbants. Si $i \geq m + 1$, il existe $1 \leq \ell \leq m$ et un entier n tel que $P^n(i, \ell) > 0$ (traduction de la définition de la chaîne absorbante). Notons t_i le plus petit de ces entiers n . On a donc $p_i = \mathbb{P}_i(X_{t_i} \geq m + 1) < 1$. Notons aussi que $p_i = \mathbb{P}_i(\tau > t_i)$. Ensuite, on observe que

$$\begin{aligned} \mathbb{P}_i(\tau > (\ell + 1)t_i) &= \mathbb{P}_i(X_1 \geq m + 1, \dots, X_{(\ell+1)t_i} \geq m + 1) \\ &= \mathbb{P}_i(X_{\ell t_i} \geq m + 1, X_{(\ell+1)t_i} \geq m + 1) \\ &= \sum_{j=m+1}^k \mathbb{P}_i(X_{\ell t_i} = j, X_{(\ell+1)t_i} \geq m + 1) \\ &= \sum_{j=m+1}^k \mathbb{P}(X_{(\ell+1)t_i} \geq m + 1 | X_{\ell t_i} = j) \mathbb{P}_i(X_{\ell t_i} = j) \\ &= \sum_{j=m+1}^k \mathbb{P}_j(X_{t_i} \geq m + 1) \mathbb{P}_i(X_{\ell t_i} = j) \\ &\leq p \mathbb{P}_i(X_{\ell t_i} \geq m + 1) \\ &= p \mathbb{P}_i(\tau > \ell t_i), \end{aligned}$$

où p est le plus grand des p_j . On obtient $\mathbb{P}_i(\tau > \ell t_i) \leq p^\ell$ et comme $p < 1$, $\mathbb{P}(\tau = +\infty) = 0$ si $\ell \rightarrow +\infty$. Comme $T^n(i, j) = \mathbb{P}_i(X_n = j) \leq \mathbb{P}_i(\tau > n)$ pour $m + 1 \leq i, j \leq k$, on a bien $\lim_{n \rightarrow +\infty} T^n(i, j) = 0$, ce qui prouve (1.3). La preuve est complète. \square

1.3.2 Exemple : le modèle de Wright-Fisher

On reprend le modèle de Wright-Fisher introduit au début de ce chapitre. On dispose d'une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ à valeurs dans $\{0, 1, \dots, N\}$ telle que la loi conditionnelle de $X_{n+1} | X_n = i$

soit une loi binomiale de paramètres N et $\frac{i}{N}$. La chaîne est absorbante et le temps

$$\tau = \inf \{k \geq 0 : X_k \in \{0, N\}\}$$

est fini p.s d'après la preuve de la proposition précédente (pour tout point de départ). Nous allons montrer que $\mathbb{P}_i(X_\tau = N) = \frac{i}{N}$. Ceci signifie que la diversité génétique disparaît selon ce modèle simplifié puisque un des deux allèles prend le dessus sur l'autre. Des extensions de ce modèle prenant en compte les problèmes de mutations et qui préservent une certaine diversité ont aussi été étudiés (on pourra consulter le Chapitre 7 de [1]).

Comme τ est fini p.s, on a $\lim_{n \rightarrow +\infty} X_n = X_\tau$ p.s (une fois arrivé en 0 ou N , la chaîne y reste). Le théorème de convergence dominé assure que

$$\mathbb{E}_i(X_\tau) = \lim_{n \rightarrow +\infty} \mathbb{E}_i(X_n). \quad (1.4)$$

Comme la loi conditionnelle de X_n sachant X_{n-1}, \dots, X_0 est $N \times \frac{X_{n-1}}{N} = X_{n-1}$ (moyenne d'une loi binomiale). Donc $\mathbb{E}_i(X_n) = \mathbb{E}_i(X_{n-1})$ pour tout n , ce qui donne $\mathbb{E}_i(X_n) = \mathbb{E}_i(X_0) = i$. Comme $\mathbb{E}_i(X_\tau) = N\mathbb{P}_i(X_\tau = N)$, on déduit de (1.4) que $\mathbb{P}_i(X_\tau = N) = \frac{i}{N}$ ce qui montre le résultat annoncé.

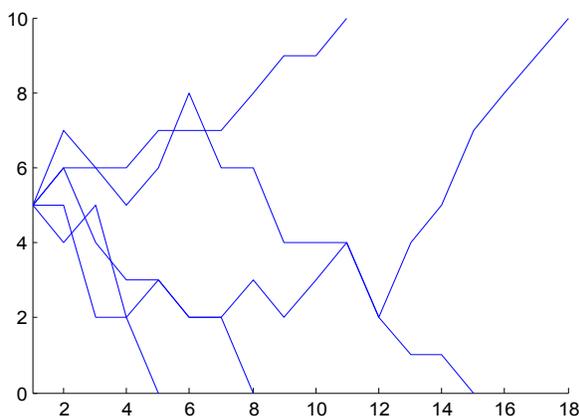


FIGURE 1.3 – 5 trajectoires de la chaîne de Wright-Fisher lorsque $X_0 = 5$ et $N = 10$. Dans ce cas, l’allèle A l’emporte sur l’allèle B avec probabilité $\frac{1}{2}$

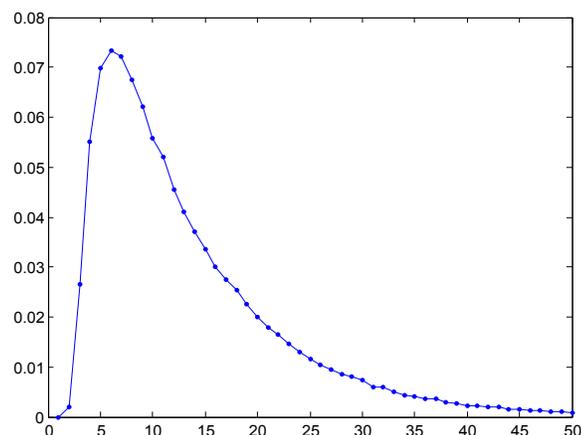


FIGURE 1.4 – Distribution de probabilité de τ lorsque $X_0 = 5$ et $N = 10$.

1.4 Théorèmes limites et application à la statistique des chaînes de Markov

1.4.1 Le théorème ergodique

Le théorème ergodique pour les chaînes de Markov est l’analogie de la loi forte des grands nombres pour les variables indépendantes et identiquement distribuées. Nous nous limiterons au cas E fini pour la preuve.

Théorème 1 Soient X une chaîne de Markov irréductible et récurrente positive sur E et π son unique probabilité invariante. Alors si $f : E \rightarrow \mathbb{R}$ vérifie $\int |f| d\pi = \sum_{x \in E} |f(x)| \pi(x) < +\infty$, on a

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n f(X_k) = \int f d\pi \text{ p.s.}$$

Preuve. Lorsque l’espace d’état E est fini, on peut prouver ce théorème par une simple application de la Proposition 5. En effet, on a

$$\frac{1}{n} \sum_{k=1}^n f(X_k) = \sum_{x \in E} f(x) \cdot \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{X_k=x\}}$$

qui est une somme finie de variables aléatoires convergentes p.s et on obtient bien la limite annoncée. Lorsque E est infini, il est plus délicat d’utiliser cet argument car la somme sur les états x est

infinie et il n'est plus possible d'intervertir limite et somme sans autre hypothèse. Nous admettrons le résultat. \square

Le résultat de l'exercice suivant sera fondamental pour la suite.

Exercice 5

Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov, irréductible, récurrente positive, de transition P et de probabilité invariante π . Pour un entier $m \geq 1$, on pose $Z_k = (X_k, X_{k+1}, \dots, X_{k+m})$, $k \in \mathbb{N}$.

1. Montrer que la chaîne de Markov $(Z_n)_{n \in \mathbb{N}}$ est irréductible sur l'ensemble

$$F = \{(x_1, x_2, \dots, x_{m+1}) \in E^{m+1} : P(x_i, x_{i+1}) > 0, 1 \leq i \leq m\}.$$

2. Déterminer la matrice de transition Q de \mathbf{Z} .
3. Montrer que cette chaîne admet une unique probabilité invariante ν définie par

$$\nu(x_1, x_2, \dots, x_{m+1}) = \pi(x_1)P(x_1, x_2) \cdots P(x_m, x_{m+1}), \quad (x_1, \dots, x_{m+1}) \in E^{m+1}.$$

Le corollaire suivant sera utile pour la suite.

Corollaire 2 Soit ν la mesure sur E^{m+1} définie à l'exercice précédent. Si $g : E^{m+1} \rightarrow \mathbb{R}$ est une fonction positive où telle que $\int |g| d\nu < +\infty$, alors

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n g(X_k, X_{k+1}, \dots, X_{k+m}) = \int g d\nu, \text{ p.s.}$$

1.4.2 Convergence vers la loi stationnaire

Moyennant l'utilisation du théorème de convergence dominée, nous avons vu que la Proposition 5 assure que pour toute chaîne de Markov irréductible et récurrente positive, la suite $(\mu P^n(x))_{n \in \mathbb{N}}$ converge en moyenne de Cesaro vers $\pi(x)$ (c'est à dire que $\lim_{n \rightarrow +\infty} \frac{\mu P(x) + \dots + \mu P^n(x)}{n} = \pi(x)$). Mais que dire de la convergence de $\mu P^n(x) = \mathbb{P}(X_n = x)$ au sens usuel ? En fait une autre hypothèse est nécessaire pour garantir cette convergence. Il faut éviter l'existence d'une décomposition de l'espace d'état E en une partition à d sous-ensembles C_1, \dots, C_d telle que $\mathbb{P}(X_1 \in C_k | X_0 \in C_{k-1}) = 1$ (en convenant que $C_0 = C_d$). Dans ce cas, la convergence de la loi de X_n vers la probabilité invariante est impossible. Lorsqu'une telle décomposition est impossible, on dit que la chaîne est apériodique. On peut alors montrer que pour toute chaîne de Markov irréductible, récurrente positive et apériodique, on a $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = x) = \pi(x)$ et que la convergence se fait à vitesse exponentielle si E est fini.

L'utilisation de la convergence de la loi marginale de la chaîne de Markov vers sa probabilité invariante est particulièrement importante pour les problèmes de simulation (e.g algorithme de Metropolis-Hastings, échantillonneur de Gibbs). Ces aspects ne seront pas nécessaires pour ce cours. Notons seulement qu'une chaîne de transition P pour laquelle il existe un état $x \in E$ tel que $P(x, x) > 0$ est forcément apériodique et lorsque E est fini la loi de X_n converge vers la probabilité invariante à vitesse exponentielle, ce qui minimisera le rôle de la loi initiale de la chaîne (les séquences ADN que nous considérerons dans la section 1.5 rentreront dans ce cas de figure).

1.4.3 Un théorème limite centrale pour les chaînes de Markov

Le TLC pour les chaînes de Markov est plus délicat. Nous admettrons le résultat suivant.

Théorème 2 Soit \mathbf{X} une chaîne de Markov irréductible, récurrente positive et de probabilité invariante π . Soit également $f : E \rightarrow \mathbb{R}$ une fonction de carré intégrable par rapport à π , i.e

$$\int f^2 d\pi = \sum_{x \in E} f(x)^2 \pi(x) < +\infty,$$

et telle que $\int f d\pi = 0$. Alors il existe, à une constante additive près, une unique fonction g telle que $\int |g| d\pi < +\infty$ et $g - Pg = f$. De plus, pour toute loi initiale μ , on a

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n f(X_k) \xrightarrow{\text{Loi}} \mathcal{N}(0, \sigma^2),$$

où

$$\sigma^2 = \int g^2 d\pi - \int (Pg)^2 d\pi.$$

Remarques

1. Contrairement au cas des variables aléatoires i.i.d, ce TLC est plus difficile à utiliser puisque la variance asymptotique σ^2 dépend d'une fonction g vérifiant $(I - P)g = f$ (équation de Poisson). Il est en général difficile de trouver g explicitement et de calculer σ^2 .
2. Remarquons que pour toute fonction $g : E \rightarrow \mathbb{R}$ telle que $\int |g|^k d\pi < +\infty$ pour un entier donnée $k \geq 1$, on a aussi $\int |Pg|^k d\pi < +\infty$. En effet, l'inégalité de Jensen assure que $|Pg|^k \leq P|g|^k$ et de plus, l'invariance de π assure que

$$\int P|g|^k d\pi = \sum_{x,y \in E} P(x,y) |g(y)|^k \pi(x) = \int |g|^k d\pi.$$

3. Lorsque E est fini, toutes les conditions d'intégrabilité qui apparaissent dans l'énoncé du Théorème 2 sont superflues puisque toute fonction $f : E \rightarrow \mathbb{R}$ est bornée.
4. Sous les hypothèses du théorème précédent, nous avons

$$\begin{aligned} \sum_{k=1}^n f(X_k) &= \sum_{k=1}^n [g(X_k) - Pg(X_k)] \\ &= \sum_{k=1}^n [g(X_k) - Pg(X_{k-1})] + \sum_{k=1}^n [Pg(X_{k-1}) - Pg(X_k)] \\ &= \sum_{k=1}^n [g(X_k) - Pg(X_{k-1})] + Pg(X_0) - Pg(X_n). \end{aligned}$$

En posant $\mathcal{F}_k = \sigma(X_0, \dots, X_k)$ pour tout $k \in \mathbb{N}$, la suite $(M_n)_{n \in \mathbb{N}}$ définie par $M_0 = 0$ et

$$M_n = \sum_{k=1}^n [g(X_k) - Pg(X_{k-1})], \quad n \geq 1,$$

est une martingale (notion non abordée dans ce cours), c'est à dire que $\mathbb{E}(M_n | \mathcal{F}_{n-1}) = M_{n-1}$. A titre d'exercice, on pourra le vérifier en montrant que $\mathbb{E}(h(X_k) | \mathcal{F}_{k-1}) = Ph(X_{k-1})$ pour toute fonction h qui satisfait une condition d'intégrabilité appropriée et justifier aussi que M_n est une somme de variables aléatoires décorrélatées deux à deux. On pourra aussi en déduire que lorsque la loi initiale est la probabilité stationnaire π , on a

$$\mathbb{E}(M_n) = 0, \quad \text{Var} \left(\frac{M_n}{\sqrt{n}} \right) = \sigma^2,$$

ce qui explique la moyenne et la variance de la loi gaussienne limite dans le Théorème 2. Il est alors possible de montrer que $\frac{1}{\sqrt{n}} \sum_{k=1}^n f(X_k)$ a la même loi asymptotique $\frac{M_n}{\sqrt{n}}$ (qui existe bien en utilisant des résultats pour les martingales).

5. En gardant les notations de la remarque précédente, il faut noter que si $Pf = 0$ alors $(I - P)f = f$ et il faut prendre $g = f$ dans le Théorème 2. Dans ce cas, c'est directement $M_n = \sum_{k=1}^n f(X_k)$ qui est une martingale. Dans ce cas la variance asymptotique σ^2 est plus explicite. Cette dernière remarque est à la base du prochain corollaire.

Le résultat suivant est un corollaire presque immédiat du TLC précédent. Il concerne la chaîne de Markov construite à partir de deux variables successives d'une chaîne de Markov sur E (voir Exercice 5). Ce résultat permettra de montrer la normalité asymptotique de l'estimateur du maximum de vraisemblance de la matrice de transition P . Introduisons d'abord la notation suivante. Si $h : E \times E \rightarrow \mathbb{R}$ est telle que pour tout $x \in E$, $\sum_{y \in E} P(x, y)|h(x, y)| < +\infty$, on définit la fonction $Ph : E \rightarrow \mathbb{R}$ par

$$Ph(x) = \sum_{y \in E} P(x, y)h(x, y), \quad x \in E.$$

Corollaire 3 Soient \mathbf{X} une chaîne de Markov irréductible, récurrente positive, de matrice de transition P et $h : E \times E \rightarrow \mathbb{R}$ une fonction telle que

$$\int P(h^2) d\pi = \sum_{x, y \in E} \pi(x)P(x, y)h^2(x, y) < +\infty, \quad (1.5)$$

où π désigne l'unique probabilité invariante de \mathbf{X} . Alors, on a la convergence en loi

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n [h(X_{k-1}, X_k) - Ph(X_{k-1})] \xrightarrow{Loi} \mathcal{N}(0, \sigma^2),$$

avec

$$\sigma^2 = \int P(h^2) d\pi - \int (Ph)^2 d\pi.$$

Preuve du Corollaire 3. Pour $k \in \mathbb{N}$, posons $Z_k = (X_k, X_{k+1})$. On rappelle que la suite \mathbf{Z} est une chaîne de Markov irréductible, récurrente positive, que sa transition $Q : E^2 \times E^2 \rightarrow [0, 1]$ est donnée par

$$Q((x, y), (u, v)) = \mathbb{1}_{y=u}P(y, v)$$

et que son unique mesure invariante est la mesure ν définie sur $E \times E$ par $\nu(x, y) = \pi(x)P(x, y)$. On pourra vérifier que $Qh(x, y) = Ph(y)$ et $Q((u, v) \mapsto Ph(u))(x, y) = Ph(y)$, ce qui entraîne que $Q((x, y) \mapsto h(x, y) - Ph(x)) = 0$. D'après la dernière remarque qui précède ce corollaire, on a bien la normalité asymptotique annoncée avec

$$\begin{aligned} \sigma^2 &= \int (h(x, y) - Ph(x))^2 d\nu \\ &= \sum_{x, y \in E} \pi(x)P(x, y) (h(x, y) - Ph(x))^2 \\ &= \int P(h^2) d\pi + \int (Ph)^2 d\pi - 2 \int (Ph)^2 d\pi \\ &= \int P(h^2) d\pi - \int (Ph)^2 d\pi. \square \end{aligned}$$

Donnons enfin l'analogie multivarié du corollaire précédent.

Corollaire 4 *Conservons les notations et les hypothèses du Corollaire 3, mais en supposant maintenant que $h = (h_1, \dots, h_d)'$ est une fonction définie sur E^2 , à valeurs dans \mathbb{R}^d et telle que $\int P(h_i^2) d\pi < +\infty$ pour $i = 1, \dots, d$. En notant Ph la fonction à valeurs vectorielles $(Ph_1, \dots, Ph_d)'$, nous avons, lorsque $n \rightarrow +\infty$,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n [h(X_{k-1}, X_k) - Ph(X_{k-1})] \xrightarrow{\text{Loi}} \mathcal{N}_d(0, \Sigma),$$

$$\text{où } \Sigma_{i,j} = \int P(h_i h_j) d\pi - \int Ph_i \cdot Ph_j d\pi.$$

Preuve du Corollaire 4. Pour $\lambda \in \mathbb{R}^d$, soit $h_\lambda = \sum_{i=1}^d \lambda_i h_i = \lambda' h$. Alors, d'après le Corollaire 3, on a, lorsque $n \rightarrow +\infty$,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n [h_\lambda(X_{k-1}, X_k) - Ph_\lambda(X_{k-1})] \xrightarrow{\text{Loi}} \mathcal{N}(0, \sigma_\lambda^2)$$

avec

$$\begin{aligned} \sigma_\lambda^2 &= \int P(h_\lambda^2) d\pi - \int (Ph_\lambda)^2 d\pi \\ &= \lambda' \cdot \int P(hh') d\pi \cdot \lambda - \lambda' \cdot \int P_h P_h' d\pi \cdot \lambda \\ &= \lambda' \Sigma \lambda. \end{aligned}$$

On en déduit facilement le résultat en utilisant la caractérisation de la convergence en loi à l'aide des fonctions caractéristiques. \square

1.4.4 Estimation de la matrice de transition par la méthode du maximum de vraisemblance

On supposera ici que E est fini et que la chaîne est récurrente irréductible. Supposons que l'on dispose d'une réalisation du vecteur aléatoire (X_1, \dots, X_n) . Pour tout $(x, y) \in E^2$, nous noterons $N_n(x, y) = \sum_{i=1}^{n-1} \mathbb{1}_{\{X_i=x, X_{i+1}=y\}}$ et $N_n(x) = \sum_{i=1}^{n-1} \mathbb{1}_{\{X_i=x\}}$. Le théorème ergodique assure que $\lim_{n \rightarrow +\infty} \frac{N_n(x, y)}{N_n(x)} = P(x, y)$ p.s.

Remarquons que comme $\pi(x) > 0$, le théorème ergodique assure que pour presque tout ω , il existe un entier $n_0(\omega)$ tel que $N_n(x)_\omega > 0$ pour tout $n \geq n_0(\omega)$. Donc le ratio $\frac{N_n(x, y)_\omega}{N_n(x)_\omega}$ est bien défini (p.s) lorsque n est assez grand. Par convention, on peut poser $\frac{N_n(x, y)_\omega}{N_n(x)_\omega} = 0$ si $N_n(x)_\omega = 0$.

Nous allons vérifier que l'estimateur $\left(\frac{N_n(x, y)}{N_n(x)}\right)_{(x, y) \in E^2}$ correspond à quelque chose près à l'estimateur du maximum de vraisemblance de la matrice de transition P . La vraisemblance que nous allons considérer est la vraisemblance conditionnelle à la première observation. On a

$$\mathbb{P}(X_2 = x_2, \dots, X_n = x_n | X_1 = x_1) = P(x_1, x_2)P(x_2, x_3) \cdots P(x_{n-1}, x_n),$$

ce qui amène à définir la vraisemblance

$$L(X_1, \dots, X_n; P) = \prod_{(x, y) \in E^2} P(x, y)^{N_n(x, y)}.$$

En utilisant la convention $0 \times \ln(0) = 0$, la log-vraisemblance s'écrit

$$\ell(X_1, \dots, X_n; P) = \sum_{x \in E} \sum_{y \in E} N_n(x, y) \ln(P(x, y)).$$

Il suffit alors de maximiser à x fixé. Nous utiliserons la notation $N_n(x+) = \sum_{i=1}^{n-1} \mathbb{1}_{X_i=x}$.

- Remarquons que si $N_n(x+)_\omega = 0$, alors pour tout $y \in E$, $N_n(x, y)_\omega = 0$ et n'importe quel choix de la probabilité $(P(x, y))_{y \in E}$ conduit à un maximum. Toutefois ce cas disparaît lorsque n est suffisamment grand.
- Si maintenant $N_n(x+)_\omega > 0$. Alors le Lemme 2 énoncé ci-dessous assure que

$$\sum_{y \in E} \frac{N_n(x, y)_\omega}{N_n(x+)_\omega} \ln(P(x, y)) \leq \sum_{y \in E} \frac{N_n(x, y)_\omega}{N_n(x+)_\omega} \ln\left(\frac{N_n(x, y)_\omega}{N_n(x+)_\omega}\right).$$

On en déduit que l'estimateur du maximum de vraisemblance est donné par

$$\hat{P}(x, y) = \frac{N_n(x, y)}{N_n(x+)}, \quad x \in E.$$

Lemme 2 Soient $(p_y)_{y \in E}$ et $(q_y)_{y \in E}$ deux probabilités sur E . Alors

$$\sum_{y \in E} q_y \ln(p_y) \leq \sum_{y \in E} q_y \ln(q_y)$$

et l'égalité a lieu si et seulement si $p = q$.

Preuve du Lemme 2. Nous allons utiliser l'inégalité $\ln(x) \leq x - 1$ qui est valable pour tout $x \geq 0$, inégalité qui est stricte si et seulement si $x \neq 1$. Soit $y \in E$ tel que $q_y > 0$. Alors

$$q_y (\ln(p_y) - \ln(q_y)) = q_y \ln\left(\frac{p_y}{q_y}\right) \leq q_y \left(\frac{p_y}{q_y} - 1\right) \leq p_y - q_y.$$

Remarquons aussi que l'inégalité précédente reste valable si $q_y = 0$ et qu'elle est stricte uniquement si $p_y \neq q_y$. On obtient alors le résultat en sommant sur y . \square

La proposition suivante donne les propriétés asymptotiques de l'estimateur du maximum de vraisemblance \hat{P} de P .

Proposition 8 Soit \mathbf{X} une chaîne de Markov sur un ensemble E fini, irréductible (donc récurrente positive), de transition P et de mesure invariante π . Alors \hat{P} est un estimateur fortement consistant de P . De plus, le vecteur $\left(\sqrt{n} \left[\frac{N_n(x,y)}{N_n(x+)} - P(x,y)\right]\right)_{(x,y) \in E^2}$ converge en loi lorsque $n \rightarrow +\infty$ vers un vecteur gaussien $(Z_{x,y})_{x,y \in E}$ centré et tel que

$$\text{Cov}(Z_{x,y}, Z_{x',y'}) = \frac{1}{\pi(x)} P(x,y) \left[\mathbb{1}_{y=y'} - P(x',y') \right] \mathbb{1}_{x=x'}.$$

On pourra remarquer l'indépendance asymptotique entre $\hat{P}(x,y)$ et $\hat{P}(x',y')$ lorsque $x \neq x'$.

Preuve de la Proposition 8. Nous allons appliquer le Corollaire 4. Pour tout $(x,y) \in E^2$, soit $h_{x,y} : E^2 \rightarrow \mathbb{R}$ la fonction définie par $h_{x,y}(u,v) = \mathbb{1}_{x=u,y=v}$ pour $(u,v) \in \mathbb{R}^2$. Remarquons alors que $N_n(x,y) = \sum_{k=2}^n h_{x,y}(X_{k-1}, X_k)$ et que

$$Ph_{x,y}(u) = \sum_{v \in E} \mathbb{1}_{x=u,y=v} P(u,v) = P(x,y) \mathbb{1}_{x=u}.$$

En posant

$$S_n(x,y) = \frac{1}{\sqrt{n}} \sum_{k=2}^n \left[h_{x,y}(X_{k-1}, X_k) - Ph_{x,y}(X_{k-1}) \right],$$

on a

$$\begin{aligned} S_n(x,y) &= \frac{1}{\sqrt{n}} \left[N_n(x,y) - \sum_{k=2}^n P(x,y) \mathbb{1}_{X_{k-1}=x} \right] \\ &= \frac{1}{\sqrt{n}} [N_n(x,y) - P(x,y)N_n(x+)] \\ &= \frac{N_n(x+)}{n} \times \sqrt{n} \left[\frac{N_n(x,y)}{N_n(x+)} - P(x,y) \right]. \end{aligned}$$

En utilisant le théorème ergodique, on a $\frac{N_n(x+)}{n} \xrightarrow{n \rightarrow +\infty} \pi(x)$ p.s. De plus, d'après le Corollaire 4, la suite $(S_n)_n$ converge en loi vers un vecteur gaussien de dimension $|E|^2$, de moyenne 0 et de matrice

de variance-covariance $\Sigma^{(1)}$ définie par

$$\begin{aligned}\Sigma^{(1)}((x, y), (x', y')) &= \sum_{u, v \in E} \pi(u)P(u, v)\mathbb{1}_{u=x, v=y}\mathbb{1}_{u=x', v=y'} - \sum_{u \in E} \pi(x)P(x, y)P(x', y')\mathbb{1}_{u=x}\mathbb{1}_{u=x'} \\ &= \pi(x)P(x, y)\mathbb{1}_{x=x', y=y'} - \pi(x)P(x, y)P(x', y')\mathbb{1}_{x=x'} \\ &= \pi(x) \left[P(x, y)\mathbb{1}_{y=y'} - P(x, y)P(x', y') \right] \mathbb{1}_{x=x'}.\end{aligned}$$

Comme

$$\sqrt{n} \left[\frac{N_n(x, y)}{N_n(x+)} - P(x, y) \right] = \frac{n}{N_n(x+)} S_n(x, y),$$

la convergence en loi annoncée découle des propriétés de la convergence en loi (limite en loi d'une fonction d'un vecteur aléatoire qui converge en loi et d'une matrice aléatoire qui converge en probabilité vers une matrice déterministe). \square

1.4.5 Construction d'un test d'adéquation.

Nous pouvons appliquer les calculs de convergence effectués précédemment pour tester l'adéquation d'un échantillon à une chaîne de Markov homogène de transition P donnée.

Théorème 3 *Pour une chaîne de Markov irréductible sur un espace d'état E fini et de matrice de transition P , la statistique*

$$T_n = \sum_{(x, y) \in E^2: P(x, y) > 0} \frac{(N_n(x, y) - P(x, y)N_n(x+))^2}{P(x, y)N_n(x+)}$$

converge en loi vers une loi du χ^2 à $k - |E|$ degrés de liberté où k désigne le nombre de paires (x, y) telle que $P(x, y) > 0$. Par conséquent, si on rejette l'hypothèse d'une transition égale à P lorsque $T_n(\omega) \geq \chi_{k-|E|}^2(1 - \alpha)$ (quantile d'ordre $1 - \alpha$ d'une loi du χ^2 à $k - |E|$ degrés de liberté), on obtient un test convergent et de niveau asymptotique α .

Preuve du Théorème 3. Dans la preuve de la Proposition 8, nous avons montré que la statistique S_n définie par

$$S_n(x, y) = \frac{1}{\sqrt{n}} [N_n(x, y) - P(x, y)N_n(x+)]$$

converge en loi vers un vecteur gaussien centré et de covariance $\Sigma^{(1)}$ définie par

$$\Sigma^{(1)}((x, y), (x', y')) = \pi(x)P(x, y) \left[\mathbb{1}_{y=y'} - P(x, y') \right] \mathbb{1}_{x=x'}.$$

Par conséquent la statistique S_n^* définie par

$$S_n^*(x, y) = \frac{S_n(x, y)}{\sqrt{P(x, y)\pi(x)}}, \quad (x, y) \in A = \{(u, v) \in E^2 : P(u, v) > 0\},$$

converge en loi vers un vecteur gaussien centré et de covariance Σ^* donnée par

$$\Sigma^* ((x, y), (x', y')) = \left[\mathbb{1}_{y=y'} - \sqrt{P(x, y)} \sqrt{P(x, y')} \right] \mathbb{1}_{x=x'}.$$

Il en est alors de même de pour la statistique

$$S_n^{**}(x, y) = \frac{[N_n(x, y) - P(x, y)N_n(x+)]}{\sqrt{P(x, y)N(x+)}} , \quad (x, y) \in A = \{(u, v) \in E^2 : P(u, v) > 0\},$$

En considérant pour $x \in E$, le vecteur gaussien Z_x de dimension $p_x = \sum_{y \in E} \mathbb{1}_{P(x, y) > 0}$ de moyenne 0 et de covariance Γ_x définie par

$$\Gamma_x(y, y') = \mathbb{1}_{y=y'} - \sqrt{P(x, y)} \sqrt{P(x, y')}, \quad y, y' \in A_x = \{z \in E : P(x, z) > 0\},$$

on en déduit que $\|S_n^{**}(x, y)\|^2$ converge en loi vers $\sum_{x \in E} \|Z_x\|^2$. Les vecteurs aléatoires Z_x , $x \in E$ sont indépendants et correspondent à des projections de vecteurs aléatoires indexés par E sur l'orthogonal du vecteur $\left(\sqrt{P(x, y)} \right)_{y \in A_x}$ qui est normé. On en déduit que pour tout $x \in E$, $\|Z_x\|^2$ suit une loi du χ^2 à $p_x - 1$ degrés de liberté et que $\sum_{x \in E} \|Z_x\|^2$ suit une loi du χ^2 à $\sum_{x \in E} p_x - |E|$ degrés de liberté. Ceci montre la loi asymptotique de T_n . L'application au test d'hypothèses est alors évidente. \square

1.5 Application à la recherche des mots de fréquence exceptionnelle dans les séquences ADN

1.5.1 Objectifs et formalisation du problème

Cette section est entièrement basée sur deux références données à la fin de ce chapitre. La référence [3] donne une introduction aux problèmes biologiques associés à la recherche de certains mots sur les séquences ADN et décrit les différentes approches probabilistes pour détecter les mots ayant une fréquence exceptionnelle. La référence [1], Chapitre 6, traite également de ces problèmes et donne quelques arguments mathématiques qui permettent de justifier les méthodes utilisées. Nous renvoyons à ces deux références pour les simulations effectuées sur des jeux de données réelles ainsi que pour des comparaisons numériques entre les différentes méthodes.

Un brin d'ADN peut être vu comme une suite de nucléotides de 4 types (A pour adénine, C pour cytosine, G pour guanine et T pour thymine). Les interactions entre l'ADN et son milieu passent par la reconnaissance de sites spécifiques sur l'ADN par de nombreux agents. Par exemple, des enzymes de restriction (endonucléases) sont capables de couper l'ADN à l'endroit où apparaît une séquence (ou mot) bien précise (cet endroit est appelé site de restriction), ce qui permet à des bactéries de se protéger d'un virus en fractionnant son ADN. Mais ce mécanisme de défense peut se retourner contre la bactérie elle-même et on s'attend à ce que l'ADN de la bactérie ne dispose que d'un très petit nombre de ces sites de restriction.

Il existe aussi d'autres enzymes (exonucléases) qui dégradent l'ADN à partir d'une extrémité du

brin. Cette dégradation s'arrête dès que les enzymes rencontrent un motif particulier appelé Chi. Dans le cas particulier de la bactérie *Escherichia coli* qui possède un génome circulaire, la présence d'une extrémité libre d'ADN est anormale. Un exonucléase est alors chargé de dégrader le brin jusqu'à la rencontre du mot 'GCTGGTGG', le brin pouvant alors être recombéné avec un fragment complémentaire. On s'attend ici à ce que le motif Chi apparaisse assez fréquemment sur l'ADN. Plus généralement, isoler des mots exceptionnellement rares ou fréquents permet au biologiste de rechercher de nouvelles fonctions biologiques d'intérêt. Il revient alors au statisticien de définir cette notion d'exceptionnalité qui sera toujours relative au modèle considéré.

Introduisons quelques notations. La suite des nucléotides est supposée être la réalisation d'une séquence (X_1, X_2, \dots, X_n) de variables aléatoires à valeurs dans l'alphabet $E = \{A, C, G, T\}$. Les tests d'indépendance rejettent habituellement l'hypothèse d'indépendance pour les variables aléatoires qui modélisent la séquence : les nucléotides ne semblent donc pas agencés de façon aléatoire. De plus, lorsqu'on utilise un sens de lecture pour la séquence (habituellement le sens de la répllication), il apparait que la probabilité de trouver un nucléotide donné à une position i dépend du nucléotide à la position $i - 1$, d'où l'utilisation fréquente des chaînes de Markov. Toutes les transitions $P(x, y)$ semblent strictement positives et la chaîne est donc supposée irréductible (donc récurrente positive) et apériodique (ce qui justifie l'utilisation de la mesure invariante π pour loi initiale). Cependant la notion de modèle est à interpréter différemment ici. Il est illusoire de penser qu'une chaîne de Markov modélise correctement la séquence et la chaîne est plutôt utilisée ici comme un modèle de référence qui permet de détecter des phénomènes atypiques sur la séquence. Un mot $w = w_1 w_2 \dots w_k$ de longueur k est une séquence de k lettres. Pour $i \geq 1$, soit $Y_i(w) = \mathbb{1}_{\{X_i=w_1, \dots, X_{i+k-1}=w_k\}}$. Soit alors $N_n(w) = \sum_{i=1}^{n-k+1} Y_i$ le nombre d'occurrences du mot w le long de la séquence. Pour détecter des mots w exceptionnellement rares ou fréquents par rapport au modèle que l'on s'est donné (ici un modèle de chaîne de Markov), l'idée est d'évaluer des " p -valeurs" du type $p_n(w) = \mathbb{P}(N_n(w) > N_n(w)^{obs})$ où $N_n(w)^{obs}$ désigne le nombre d'occurrences du mot w sur la séquence observée. Si $p_n(w)$ est élevé (resp. faible), on dira que le mot est exceptionnellement rare (resp. fréquent). Dire que $p_n(w)$ est élevé (resp. faible) signifie qu'il est très probable (resp. peu probable) d'observer un nombre d'occurrences plus élevé sous le modèle considéré. Nous ferons l'abus de notation suivant. Le nombre $\pi(w)$ désignera la valeur en w de la mesure invariante pour la chaîne vectorielle $Z_n = (X_n, X_{n+1}, \dots, X_{n+k-1})$ et ce quelle que soit la longueur k du mot w . On a ainsi d'après l'Exercice 5, $\pi(w) = \pi(w_1)P(w_1, w_2) \dots P(w_{k-1}, w_k)$.

Ci-dessous, nous décrivons trois approches pour détecter des mots exceptionnels. Seule la méthode basée sur la loi exacte du nombre d'occurrences sera justifiée, les développements techniques nécessaires à une présentation rigoureuse des deux autres méthodes dépassent le cadre de ce cours. Les simulations et estimations faites dans [3] semblent montrer que la méthode utilisant l'approximation gaussienne (resp. poissonnienne) est plutôt indiquée pour les mots de fréquence importante (resp. très faible) mais que les trois méthodes classent les p -valeurs des différents mots w dans le même ordre. Il existe aussi des modèles de Markov à l'ordre m (voir Exercice 2) qui s'ajustent mieux aux séquences ADN et auxquels les trois méthodes décrites ci-dessous peuvent être adaptées.

1.5.2 Loi exacte du nombre d'occurrences

Il est possible d'obtenir des formules pour la loi du nombre d'occurrences $N_n(w)$. Pour cela, nous supposons que la loi initiale correspond à la probabilité invariante. La suite $(X_n)_{n \in \mathbb{N}^*}$ est donc une chaîne de Markov stationnaire. Soit T_1, T_2, \dots les positions successives du mot w (ou de façon équivalente les temps de retour successifs de la chaîne stationnaire $(Y_i)_i$ au point w). On a alors $\{N_n(w) \geq j\} = \{T_j \leq n - k + 1\}$ et donc en posant $p(t, j) = \mathbb{P}(T_j = t)$, il vient

$$\mathbb{P}(N_n(w) = j) = \sum_{t=1}^{n-k+1} p(t, j) - \sum_{t=1}^{n-k+1} p(t, j+1). \quad (1.6)$$

- Commençons par déterminer une formule de récurrence pour $p(t, 1)$. Remarquons d'abord que

$$p(t, 1) = \mathbb{P}(Y_1(w) = Y_2(w) = \dots = Y_{t-1}(w) = 0, Y_t(w) = 1).$$

Par la formule des probabilités totales puis la propriété de Markov, on a

$$\begin{aligned} \pi(w) &= \mathbb{P}(Y_t(w) = 1) \\ &= \sum_{s=1}^{t-1} \mathbb{P}(Y_1(w) = \dots = Y_{s-1}(w) = 0, Y_s(w) = 1, Y_t(w) = 1) + p(t, 1) \\ &= p(t, 1) + \sum_{s=1}^{t-1} p(s, 1) \mathbb{P}(Y_t(w) = 1 | Y_s(w) = 1). \end{aligned} \quad (1.7)$$

On initialise avec $p(1, 1) = \pi(w)$.

- Regardons ensuite $p(t, j)$ pour $j \geq 2$. Dire que w est observé en t signifie que soit w apparaît en t pour la i -ième fois ($i \leq j$), soit que w apparaît en t et w est apparu pour la j -ième fois avant t . Par la formule des probabilités totales puis la propriété de Markov, on a

$$\begin{aligned} \pi(w) &= \mathbb{P}(Y_t(w) = 1) \\ &= \sum_{i=1}^j \mathbb{P}(T_i = t) + \sum_{s=1}^{t-1} \mathbb{P}(T_j = s, Y_t(w) = 1) \\ &= \sum_{i=1}^j p(t, i) + \sum_{s=1}^{t-1} p(s, j) \mathbb{P}(Y_t(w) = 1 | Y_s(w) = 1). \end{aligned} \quad (1.8)$$

- Il reste à calculer $\mathbb{P}(Y_t(w) = 1 | Y_s(w) = 1)$ pour $t > s$. On a $\mathbb{P}(Y_s(w) = 1) = \pi(w)$. Ensuite examinons $\mathbb{P}(Y_s(w) = 1, Y_{s+d}(w) = 1)$ lorsque $d < k$. Cette dernière probabilité vaut 0 sauf s'il existe un chevauchement de $k - d$ lettres (c'est à dire que les $k - d$ premières lettres de w correspondent aux $k - d$ dernières, donc $w_{d+1} = w_1, w_{d+2} = w_2, \dots, w_k = w_{k-d}$). On a dans ce cas

$$\begin{aligned} &\{Y_s(w) = 1, Y_{s+d}(w) = 1\} \\ &= \{X_s = w_1, \dots, X_{s+k-1} = w_k, X_{s+d} = w_1, \dots, X_{s+k-1} = w_{k-d}, X_{s+k} = w_{k-d+1}, \dots, X_{s+d+k-1} = w_k\} \\ &= \{X_s = w_1, \dots, X_{s+k-1} = w_k, X_{s+d} = w_{d+1}, \dots, X_{s+k-1} = w_k, X_{s+k} = w_{k-d+1}, \dots, X_{s+d+k-1} = w_k\} \\ &= \{X_s = w_1, \dots, X_{s+k-1} = w_k = w_{k-d}, X_{s+k} = w_{k-d+1}, \dots, X_{s+d+k-1} = w_k\}. \end{aligned}$$

On a donc

$$\mathbb{P}(Y_s(w) = 1, Y_{s+d}(w) = 1) = \pi(w) \prod_{j=k-d+1}^k P(w_{j-1}, w_j).$$

Si maintenant, $d \geq k$, on a

$$\begin{aligned} \mathbb{P}(Y_s(w) = 1, Y_{s+d}(w) = 1) &= \mathbb{P}(X_s = w_1, \dots, X_{s+k-1} = w_k, X_{s+d} = w_1, \dots, X_{s+d+k-1} = w_k) \\ &= \pi(w) P^{d-k+1}(w_k, w_1) P(w_1, w_2) \cdots P(w_{k-1}, w_k) \\ &= P^{d-k+1}(w_k, w_1) \frac{\pi(w)^2}{\pi(w_1)}. \end{aligned}$$

Nous en déduisons la formule

$$\mathbb{P}(Y_t(w) = 1 | Y_s(w) = 1) = \mathbb{1}_{t-s \geq k} P^{t-s-k+1}(w_k, w_1) \frac{\pi(w)}{\pi(w_1)} + \mathbb{1}_{t-s < k} \varepsilon_{k-t+s}(w) \prod_{j=k-t+s+1}^k P(w_{j-1}, w_j), \quad (1.9)$$

où $\varepsilon_\ell(w)$ vaut 1 si w admet un chevauchement de ℓ lettres et 0 sinon.

On voit alors que la loi de $N_n(w)$ peut s'obtenir à partir des relations (1.6), (1.7), (1.8) et (1.9). Toutefois le calcul numérique de cette loi obtenue de façon récursive est assez long si la séquence ADN est longue. De plus, en pratique, il faut remplacer la loi invariante et les transitions par les valeurs estimées dans les équations récursives précédentes. L'accumulation d'erreur peut conduire à une instabilité.

1.5.3 Utilisation de l'approximation gaussienne

L'utilisation de l'approximation gaussienne permet de calculer les p -valeurs de façon approchée. D'après le Théorème 2, on sait que

$$\frac{1}{\sqrt{n}} \left(\frac{N_n(w)}{n-k+1} - \pi(w) \right) \rightarrow_{n \rightarrow +\infty} \mathcal{N}(0, \sigma^2),$$

mais $\pi(w)$ doit être remplacée par un estimateur. Evidemment, on ne prendra pas $\frac{N_n(w)}{n-k+1}$ mais plutôt un estimateur qui utilise l'expression de π en fonction de la matrice transition $\pi(w) = \pi(w_1) P(w_1, w_2) \cdots P(w_{k-1}, w_k)$. On utilisera

$$\tilde{\pi}(w) = \frac{N_n(w_1+)}{n-1} \cdot \frac{N_n(w_1 w_2)}{N_n(w_1+)} \cdots \frac{N_n(w_{k-1}, w_k)}{N_n(w_{k-1}+)},$$

où si $a \in E$, $N_n(a+) = N_n(aA) + N_n(aC) + N_n(aG) + N_n(aT)$. En posant

$$U_n = \left(\frac{N_n(w)}{n-k+1}, \frac{N_n(w_1+)}{n-1}, \dots, \frac{N_n(w_{k-1}+)}{n-1}, \frac{N_n(w_1 w_2)}{n-1}, \dots, \frac{N_n(w_{k-1} w_k)}{n-1} \right)',$$

on sait d'après le Corollaire 4 que $\sqrt{n}(U_n - \mathbb{E}(U_n))$ est asymptotiquement gaussien. De plus il existe une fonction F , différentiable au point $\mathbb{E}(U_n)$, telle que $F(\mathbb{E}(U_n)) = 0$ et

$$\frac{N_n(w)}{n-k+1} - \tilde{\pi}(w) = F(U_n).$$

En utilisant un résultat connu sous le nom de δ -méthode, on peut alors montrer que $\sqrt{n}F(U_n)$ est aussi asymptotiquement gaussien. Mais c'est le calcul de la variance asymptotique qui pose problème. Cette variance asymptotique possède une expression compliquée due à la dépendance entre les variables et aux chevauchements possibles du mot w . Nous admettrons le résultat suivant (une preuve est donnée dans l'article [2] donné en référence).

Théorème 4 *La statistique $S_n = \sqrt{n} \left(\frac{N_n(w)}{n-k+1} - \tilde{\pi}(w) \right)$ converge en loi vers une variable gaussienne de moyenne 0 et de variance*

$$\begin{aligned} \sigma^2 &= \pi(w) + 2 \sum_{d=1}^{k-2} \varepsilon_{k-d}(w) \pi(w^{(d)}w) \\ &+ \pi(w)^2 \left(\sum_{y \in E} \frac{n_y(w-)^2}{\pi(y)} - \sum_{y,z \in E} \frac{n_{yz}(w)^2}{\pi(yz)} - \frac{2n_{w_1}(w-) - 1}{\pi(w_1)} \right), \end{aligned}$$

avec

$$w- = w_1 w_2 \cdots w_{k-1}, \quad w^{(d)}w = w_1 \cdots w_d w_1 \cdots w_k$$

et $n_u(v)$ désigne le nombre d'occurrences du mot u dans le mot v .

En remplaçant la mesure invariante π par un estimateur, on obtient un estimateur $\hat{\sigma}$ de σ et $Z_n = \frac{S_n}{\hat{\sigma}}$ converge vers une $\mathcal{N}(0, 1)$. Au lieu d'évaluer, $\mathbb{P}(N_n(w) > N_n(w)^{obs})$, on s'intéresse ici à

$$\mathbb{P}(Z_n > Z_n^{obs}) \approx \Phi(Z_n^{obs}),$$

où $\Phi(x) = \int_x^{+\infty} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du$. L'interprétation de ces quantités reste la même. Si pour un mot w donné, $\mathbb{P}(Z_n > Z_n^{obs})$ est grand (resp. petit), cela signifie que la fréquence de w observée $\frac{N_n(w)^{obs}}{n-k+1}$ est petite (resp. grande) par rapport à sa prévision $\tilde{\pi}(w)^{obs}$ pour le modèle de chaîne de Markov. On dira alors que le mot w est exceptionnellement rare (resp. fréquent) pour le modèle considéré.

1.5.4 Loi des petits nombres

Lorsque un mot w est très rare, l'approximation gaussienne précédente n'est plus vraiment utilisable. Pour comprendre pourquoi, nous nous plaçons dans un cadre volontairement simplifié. Supposons que X_1, \dots, X_n soient des variables aléatoires i.i.d toutes de loi de Bernoulli de paramètre $0 < p < 1$. Cette situation correspond au comptage d'un mot d'une seule lettre pour une chaîne formée de variables indépendantes. Supposons que p soit de l'ordre de $\frac{1}{n}$ en supposant que $p = p_n$ vérifie $np_n \rightarrow \lambda > 0$. Alors, il est connu que $\sum_{i=1}^n X_i$ (qui suit une loi binomiale de paramètres n et p_n) converge en loi vers la loi de Poisson de paramètres λ . Ceci met en évidence que c'est une approximation poissonnienne et non-gaussienne qui est pertinente ici. Ce type d'approximation est encore valable pour étudier l'occurrence d'un mot de taille quelconque à l'aide d'une chaîne de Markov plus générale, sous réserve que $m_n = \mathbb{E}(N_n(w)) = (n-k+1)\pi(w)$ vérifie $\sup_n m_n < +\infty$. Vu que $\pi(w) \geq \alpha^k$ pour $\alpha = \min\{\pi(x), P(x, y); (x, y) \in E^2\} > 0$, on peut montrer que cette dernière condition entraîne $k \geq c \log(n)$ pour un certain $c > 0$ (donc le mot w doit être suffisamment long). On a alors les deux conclusions suivantes.

- Si le mot w n'a pas de chevauchements, alors la loi de $N_n(w)$ s'approche par une loi de Poisson de paramètre $\mathbb{E}(N_n(w))$.
- Si le mot w admet des chevauchements, la situation est plus compliquée. On note $a(w)$ la probabilité que le mot w soit recouvrable par la gauche (par exemple ATA est recouvrable par la gauche ainsi que $CGATCG$ par exemple). On appelle alors train toute occurrence de w dans la séquence qui n'est pas recouverte à gauche par une autre occurrence de w . Si $\tilde{N}_n(w)$ désigne le nombre de trains dans la séquence et $K_i(w)$ la longueur du train n° i , on a

$$N_n(w) = \sum_{i=1}^{\tilde{N}_n(w)} K_i(w)$$

et on peut montrer que $K_i(w)$ suit une loi géométrique de paramètre $a(w)$. Comme les trains ne se chevauchent pas par définition, on peut montrer que $\tilde{N}_n(w)$ s'approche par une loi de Poisson et que les $K_i(w)$ deviennent asymptotiquement indépendantes. Ce type de loi limite est appelée loi de Poisson composé (une loi de Poisson composée est de la forme $\sum_{i=1}^Q U_i$ où les U_i sont i.i.d et indépendantes de Q qui suit une loi de Poisson).

1.6 Exercices

Exercice 6 On pose $X_0 = x \in \mathbb{Z}$ et

$$X_n = x + \sum_{i=1}^n Y_i, \quad n \in \mathbb{N}^*.$$

où Y désigne une suite de variables aléatoires i.i.d avec

$$\mathbb{P}(Y_n = 1) = 1 - \mathbb{P}(Y_n = -1) = p, \quad 0 < p < 1.$$

A partir d'un échantillon (X_1, X_2, \dots, X_n) de cette marche aléatoire, on souhaite estimer p .

1. Justifier l'irréductibilité de cette chaîne. Cette chaîne peut-elle être récurrente positive ?
2. En utilisant les accroissements Y_j de la marche, déterminer un estimateur de p . Donner les propriétés asymptotiques de cet estimateur.
3. Pourquoi cet estimateur correspond-t-il à l'estimateur du maximum de vraisemblance pour la chaîne de Markov X ?

Exercice 7 On considère une chaîne de Markov à deux états 0 et 1. Sa matrice de transition est donnée par

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

avec $0 < p, q < 1$. Déterminer l'expression du maximum de vraisemblance du couple (p, q) . Donner également les propriétés asymptotiques de l'estimateur.

Exercice 8 On considère la chaîne de Markov

$$X_{n+1} = X_n + Y_{n+1} - Z_{n+1} \mathbb{1}_{X_n > 0}, \quad n \in \mathbb{N}$$

où Y et Z sont deux suites de variables aléatoires i.i.d indépendantes entre elles. On suppose que la loi de Y_1 (resp. Z_1) une Bernoulli de paramètre $p \in (0, 1)$ (resp. $q \in (0, 1)$). On peut voir cette dynamique comme l'évolution du nombre d'individus dans une file d'attente ($Y_{n+1} = 1$ signifie qu'un client rentre dans la file et $Z_{n+1} = 1$ signifie qu'un client en sort).

1. Montrer que la chaîne de Markov X est irréductible.
2. En utilisant les résultats obtenus pour les chaînes de naissance et de mort, trouver une condition nécessaire et suffisante de récurrence positive. Quelle est alors la probabilité invariante π ?
3. Que vaut $\mathbb{E}_\pi(X_n)$?
4. On suppose que le système est initialisé avec la probabilité invariante π . On introduit un client dans la file à l'instant n . Soit T le temps de séjour de ce client dans le système. Calculer l'espérance de T . On pourra montrer que

$$\mathbb{E}_\pi(T) = \frac{1 + \mathbb{E}_\pi(X_n)}{q}.$$

Exercice 9 On reprend le cas du problème de stock de pièces détachées introduit dans le cours. Si X_n représente l'état du stock au temps n , on définit

$$X_{n+1} = \max(X_n + q - D_{n+1}, 0),$$

où $(D_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires i.i.d. On suppose dans cet exercice que $q = 1$ et $p_k = \mathbb{P}(D_1 = k)$ est strictement positive uniquement si $k = 0, 1, 2$. Montrer que la chaîne est une chaîne de naissance et de mort et déterminer la nature de la chaîne en fonction de $\mathbb{E}(D_1)$. Lorsque $\mathbb{E}(D_1) > 1$, calculer la limite du stock moyen $\frac{1}{n} \sum_{k=1}^n X_k$.

Exercice 10 Soient $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov irréductible et récurrente positive sur E et $x \in E$. On notera par $(T_n)_{n \geq 1}$ les temps de retour successifs au point x et $S_n = T_n - T_{n-1}$ le délai avant le n -ième retour. On s'intéresse aux excursions hors de l'état x , c'est à dire aux variables aléatoires

$$Y_n = (S_n, X_{T_{n-1}}, X_{S_{n-1}+1}, \dots, X_{T_n}), \quad n \in \mathbb{N}^*.$$

L'application $\omega \mapsto Y_n(\omega)$ est une variable aléatoire à valeurs dans $\cup_{k \geq 1} \{k\} \times E^{k+1}$. Remarquons que X_{T_n} est en fait constante et égale à x pour tout entier n (de façon générale, si T est une variable aléatoire à valeurs entières, X_T est définie par $X_T(\omega) = X_{T(\omega)}(\omega)$ et $X_T = X_k$ sur l'évènement $\{T = k\}$). Montrer que les variables aléatoires $(Y_n)_{n \geq 1}$ sont indépendantes et que pour tout $n \geq 2$, Y_n a la même loi de que la loi de Y_1 sous \mathbb{P}_x .

Cette propriété peut être utilisée pour prouver le théorème ergodique et une version du TLC en considérant des sommes du type $\sum_{k=T_{n-1}+1}^{T_n} f(X_k)$. Elle généralise le résultat de l'Exercice 3.

Bibliographie

- [1] Delmas, J-F, Jourdain, B. (2006) *Modèles aléatoires. Applications aux sciences de l'ingénieur et du vivant*. Springer.
- [2] Prum, B., Rodolphe, F., De Turckheim, E. *Finding words with unexpected frequencies in deoxyribonucleic acid sequences*. J.R. Statist. Soc. B, 57(1) : 205-220, 1995.
- [3] Robin, S., Rodolphe, F., Schbath, S. (2003) *ADN, mots et modèles*. Belin.

Chapitre 2

Processus de Poisson

2.1 Le processus de Poisson simple

Définition 7 Soit $(X_t)_{t \geq 0}$ un processus stochastique à valeurs réelles. On dit $(X_t)_{t \geq 0}$ est un processus de comptage si pour \mathbb{P} -presque tout $\omega \in \Omega$, la trajectoire $t \mapsto X_t(\omega)$ est croissante par sauts de 1, continue à droite et telle que $X_0 = 0$ p.s.

Définition 8 Un processus de comptage $(N_t)_{t \geq 0}$ est appelé un processus de Poisson simple si :

1. $N_0 = 0$ p.s.
2. le processus est à accroissements indépendants, c'est à dire $N_{t+s} - N_s$ est indépendante de $\sigma(N_u : u \leq s)$ pour tous $t, s \geq 0$.
3. Pour tous s et t positifs, la variable aléatoire $N_{t+s} - N_s$ suit une loi de Poisson de paramètre λt où λ est un réel strictement positif.

Remarque

- Les accroissements d'un processus de Poisson sont aussi stationnaires au sens où la loi de $N_{t+s} - N_s$ est la même pour toutes les valeurs de s .
- Le paramètre λ est appelé l'intensité du processus (on rappelle que la moyenne d'une loi de Poisson de paramètre λ vaut λ).

Des exemples d'utilisation

- N_t peut représenter le nombre de clients arrivés à un guichet avant l'instant t (ou le nombre d'appels vers un central téléphonique). On peut également s'intéresser au nombre de pannes d'une machine à l'instant t , la machine étant réparée immédiatement après chaque panne.
- En physique nucléaire, un compteur Geiger permet de détecter des rayonnements α ou γ émis par des particules radioactives. Sous l'action d'un rayonnement électromagnétique, un gaz basse pression produit des ions de façon aléatoire. Supposons que les ionisations se produisent indépendamment les unes des autres et notons par N_t le nombre d'ionisations au cours de l'intervalle $[0, t]$. Soit alors $p_n(t) = \mathbb{P}(N_t = n)$ et supposons que $p_1(h) \sim \lambda h$

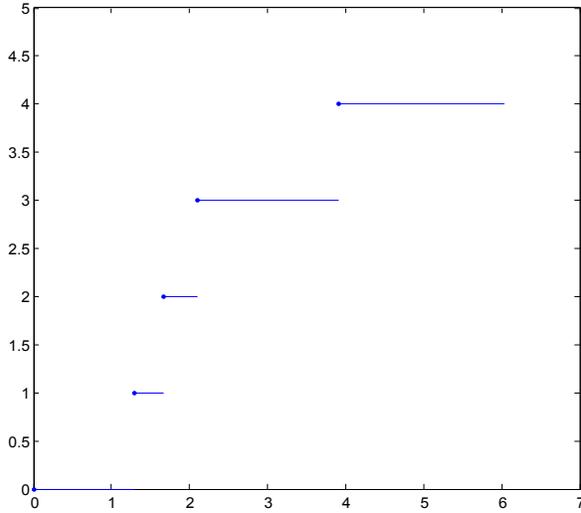


FIGURE 2.1 – Trajectoire lorsque $\lambda = 1$

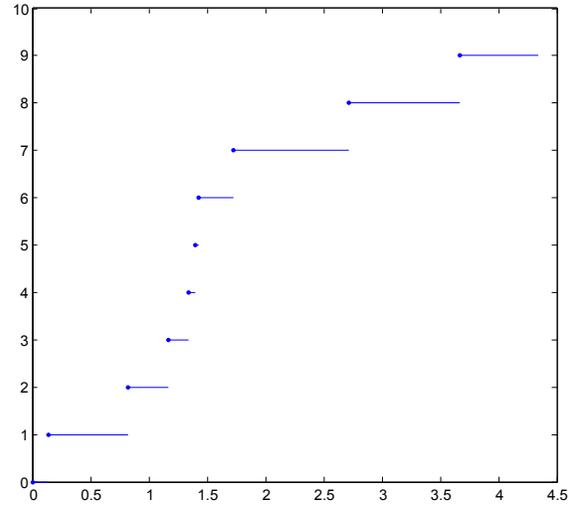


FIGURE 2.2 – Trajectoire lorsque $\lambda = 2$

et $\mathbb{P}(N_h \geq 2) = o(h)$ lorsque h tend vers 0. Si les accroissements du processus $(N_t)_t$ sont supposés stationnaires, on peut alors écrire

$$p_n(t+h) = p_n(t)p_0(h) + p_{n-1}(t)p_1(h) + \dots + p_0(t)p_n(h) = p_n(t)p_0(h) + p_{n-1}(t)p_1(h) + o(h).$$

Comme $p_0(h) + p_1(h) = 1 - \mathbb{P}(N_h \geq 2) = 1 + o(h)$, on obtient l'équation différentielle

$$p'_n(t) = \lambda[p_{n-1}(t) - p_n(t)].$$

En écrivant $p_n(t) = u_n(t) \exp(-\lambda t)$, on a nécessairement $u'_n(t) = \lambda u_{n-1}(t)$, équation dont la solution est $u_n(t) = \frac{(\lambda t)^n}{n!}$. On voit alors que N_t suit une loi de Poisson de paramètre λt . Le processus $(N_t)_{t \geq 0}$ est alors un processus de Poisson.

Exercice 11 Montrer que la somme de deux processus de Poisson indépendants reste un processus de Poisson. Quelle est alors l'intensité du processus ?

Proposition 9 Soit N un processus de Poisson simple. Alors N est un processus stationnaire d'événements rares :

- $\mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h)$.
- $\mathbb{P}(N_{t+h} - N_t > 1) = o(h)$.

Preuve. On a

$$\begin{aligned} \mathbb{P}(N_{t+h} - N_t = 1) &= \lambda h \exp(-\lambda h) \\ &= \lambda h (1 + o(1)) \\ &= \lambda h + o(h). \end{aligned}$$

De même, on a

$$\begin{aligned}\mathbb{P}(N_{t+h} - N_t > 1) &= 1 - \lambda h \exp(-\lambda h) - \exp(-\lambda h) \\ &= o(h). \square\end{aligned}$$

Proposition 10 Si $0 = T_0 < T_1 < \dots < T_n < \dots$ sont les instants de sauts du processus de Poisson simple alors

$$N_t = \sum_{n \geq 1} \mathbb{1}_{\{T_n \leq t\}}.$$

De plus les instants $(T_n - T_{n-1})_{n \geq 1}$ d'inter-arrivées sont des variables aléatoires indépendantes et toutes de loi exponentielle de paramètre λ . La densité de (T_1, \dots, T_n) est donnée par

$$f_{(T_1, \dots, T_n)}(t_1, \dots, t_n) = \lambda^n \exp(-\lambda t_n) \mathbb{1}_{0 < t_1 < \dots < t_n}.$$

Preuve. Soient $t_1, h_1, \dots, t_n, h_n$ des nombres réels positifs tels que

$$0 < t_1 < t_1 + h_1 < t_2 < t_2 + h_2 < \dots < t_n < t_n + h_n.$$

On a alors en utilisant l'indépendance et la loi des accroissements

$$\begin{aligned}&\mathbb{P}(t_1 < T_1 < t_1 + h_1, \dots, t_n < T_n < t_n + h_n) \\ &= \mathbb{P}(N_{t_1} = 0, N_{t_1+h_1} - N_{t_1} = 1, N_{t_2} - N_{t_1+h_1} = 0, \dots, N_{t_n} - N_{t_{n-1}+h_{n-1}} = 0, N_{t_n+h_n} - N_{t_n} \geq 1) \\ &= \exp(-\lambda t_n) \lambda^n h_1 \cdots h_{n-1} (1 - \exp(-\lambda h_n)).\end{aligned}$$

En divisant par le produit $h_1 \cdots h_n$ et en faisant tendre h vers 0, on obtient la loi annoncée pour le vecteur (T_1, \dots, T_n) . La formule du changement de variables permet alors de montrer que la densité du vecteur $(T_i - T_{i-1})_{1 \leq i \leq n}$ est bien celle d'un n -uplet de variables aléatoires indépendantes toutes de loi exponentielle de paramètre λ . \square

Remarque. Ce résultat permet d'envisager une définition du processus de Poisson à partir d'une suite i.i.d ξ de variables aléatoires toutes de loi exponentielle de paramètre λ . En effet en posant $T_n = \sum_{i=1}^n \xi_i$ le processus N défini par $N_t = \sum_{n \geq 1} \mathbb{1}_{T_n \leq t}$ si $t \in \mathbb{R}_+$ a la même loi qu'un processus de Poisson (si ce processus de Poisson existe vraiment). Il faut donc vérifier que le processus obtenu avec cette suite i.i.d de variables aléatoires vérifient bien la définition du processus de Poisson. Nous l'admettrons.

Proposition 11 Conditionnellement à $\{N_t = n\}$, le vecteur (T_1, \dots, T_n) a même loi que les statistiques d'ordre de n variables aléatoires indépendantes et toutes de loi uniforme sur l'intervalle $[0, t]$.

Preuve. Si U_1, \dots, U_n sont des variables aléatoires i.i.d toutes de loi uniforme sur $[0, t]$, alors le vecteur des statistiques d'ordre $(U_{(1)}, \dots, U_{(n)})$ admet pour densité la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$ définie par

$$f(t_1, \dots, t_n) = \frac{n!}{t^n} \mathbb{1}_{0 < t_1 < \dots < t_n \leq t}.$$

En utilisant la loi du $(n + 1)$ -uplet (T_1, \dots, T_{n+1}) , on obtient pour tout borélien $B \in \mathcal{B}(\mathbb{R}^n)$

$$\begin{aligned} \mathbb{P}((T_1, \dots, T_n) \in B, N_t = n) &= \mathbb{P}((T_1, \dots, T_n) \in B, T_n \leq t < T_{n+1}) \\ &= \int \lambda^{n+1} \exp(-\lambda t_{n+1}) \mathbb{1}_B(t_1, \dots, t_n) \mathbb{1}_{0 < t_1 < \dots < t_n \leq t < t_{n+1}} dt_1 \cdots dt_{n+1} \\ &= \int \lambda^n \exp(-\lambda t) \mathbb{1}_B(t_1, \dots, t_n) \mathbb{1}_{0 < t_1 < \dots < t_n \leq t} dt_1 \cdots dt_n. \end{aligned}$$

Etant donné que $\mathbb{P}(N_t = n) = \exp(-\lambda t) \frac{\lambda^n t^n}{n!}$, le résultat est obtenue en divisant. \square

Au niveau du comportement asymptotique de N , nous avons les deux résultats suivants.

Théorème 5 1. On a $\lim_{t \rightarrow \infty} \frac{N_t}{t} = \lambda$ p.s.

2. On a de plus la convergence en loi

$$\sqrt{\frac{t}{\lambda}} \left(\frac{N_t}{t} - \lambda \right) \xrightarrow{t \rightarrow +\infty} \mathcal{N}(0, 1).$$

Preuve.

1. Nous avons l'encadrement $T_{N_t} \leq t < T_{N_t+1}$ p.s et donc

$$\frac{N_t}{T_{N_t+1}} \leq \frac{N_t}{t} \leq \frac{N_t}{T_{N_t}}. \quad (2.1)$$

Il est facile de montrer que $\lim_{t \rightarrow \infty} N_t = +\infty$ (exercice). De plus, comme $T_n = \sum_{i=1}^n (T_i - T_{i-1})$, la loi des grands nombres assure que $\lim_{n \rightarrow \infty} \frac{T_n}{n} = \frac{1}{\lambda}$ p.s (on rappelle que les intervalles entre les sauts sont i.i.d de loi exponentielle de paramètre λ). Il est alors facile d'en déduire qu'avec probabilité 1, les bornes de l'encadrement 2.1 tendent vers λ lorsque n tend vers l'infini.

2. Une première solution est d'utiliser le critère des fonctions caractéristiques (exercice). Une autre solution consiste à utiliser le théorème de la limite centrale en utilisant la décomposition

$$\sqrt{\frac{t}{\lambda}} \left(\frac{N_t}{t} - \lambda \right) = \sqrt{\frac{t}{\lambda}} \left(\frac{N_{[t]}}{[t]} - \lambda \right) + A_t.$$

– Pour le premier terme, on a

$$\sqrt{\frac{t}{\lambda}} \left(\frac{N_{[t]}}{[t]} - \lambda \right) = \sqrt{\frac{t}{[t]}} \sqrt{\frac{[t]}{\lambda}} \left(\frac{\sum_{i=1}^{[t]} (N_i - N_{i-1})}{[t]} - \lambda \right).$$

On obtient alors la convergence en loi de ce premier terme vers la loi $\mathcal{N}(0, 1)$ en utilisant le théorème central limite et le lemme de Slutsky (car $\lim_{t \rightarrow \infty} t/[t] = 1$).

– Pour le deuxième terme, on a

$$\begin{aligned} |A_t| &= \sqrt{\frac{t}{\lambda}} \left| \frac{N_{[t]}}{[t]} - \frac{N_t}{t} \right| \\ &\leq \sqrt{\frac{t}{\lambda}} \left(\frac{N_t - N_{[t]}}{[t]} + \frac{N_t}{[t]t} \right). \end{aligned}$$

On pourra montrer que le premier de deux termes majorants tend vers 0 en probabilité en utilisant l'inégalité de Markov. Comme le deuxième tend vers 0 p.s en utilisant le premier point du théorème, on conclut que $\lim_{t \rightarrow \infty} A_t = 0$ en probabilité.

D'après ce qui précède et en utilisant le lemme de Slutsky, on obtient la convergence annoncée. \square

2.1.1 Estimation de l'intensité par maximum de vraisemblance

Nous allons considérer trois cas d'échantillonnage.

1. Considérons d'abord le cas où le processus est observé jusqu'à l'instant t . La variable aléatoire $\frac{N_t}{t}$ est en fait l'estimateur du maximum de vraisemblance du paramètre λ . Dans ce chapitre, nous calculerons la vraisemblance de façon informelle ; le calcul sera justifié au chapitre suivant. L'idée est que lorsqu'on observe $N_t = n$, les temps de sauts T_1, \dots, T_n détermine complètement la trajectoire du processus sur $[0, t]$. Nous définirons donc la vraisemblance en calculant la loi de (T_1, \dots, T_n, N_t) . Ceci peut être fait en utilisant la Proposition 11. En effet, la loi de $(T_1, \dots, T_n) | N_t = n$ a pour densité

$$f : (t_1, \dots, t_n) \rightarrow \frac{n!}{t^n} \mathbb{1}_{0 < t_1 < \dots < t_n \leq t}.$$

La fonction de vraisemblance sera alors définie pour une observation (t_1, \dots, t_n, n) par

$$L(t_1, \dots, t_n, n, \lambda) = f(t_1, \dots, t_n) \mathbb{P}(N_t = n) = \lambda^n \exp(-\lambda t).$$

Le maximum de vraisemblance est alors $\hat{\lambda}(\omega) = \frac{n}{t} = \frac{N_t(\omega)}{t}$. On a vu que cet estimateur était consistant et asymptotiquement gaussien. De plus l'utilisation du lemme de Slutsky permet de montrer la convergence en loi

$$\sqrt{\frac{t^2}{N_t}} \left(\frac{N_t}{t} - \lambda \right) \xrightarrow{t \rightarrow +\infty} \mathcal{N}(0, 1).$$

Un intervalle de confiance de niveau asymptotique $1 - \alpha$ est alors

$$I_n = \left[-\sqrt{\frac{N_t}{t^2}} q_{1-\frac{\alpha}{2}} + \frac{N_t}{t}, +\sqrt{\frac{N_t}{t^2}} q_{1-\frac{\alpha}{2}} + \frac{N_t}{t} \right],$$

où $q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la $\mathcal{N}(0, 1)$.

Remarque. Il existe un moyen de construire des intervalles de confiance de niveau exact lorsqu'on observe une loi de Poisson. Ici N_t suit une loi de Poisson de paramètre λt . De plus, si $n \in \mathbb{N}$,

$$\begin{aligned} F_{\lambda}(n) &= \mathbb{P}(N_t \leq n) \\ &= \mathbb{P}(N_t < n + 1) \\ &= \mathbb{P}(T_{n+1} > t) \\ &= 1 - \int_0^{\lambda t} \frac{1}{n!} x^n \exp(-x) dx. \end{aligned}$$

Soit $\alpha \in (0, 1)$. Par continuité et décroissance stricte de $\lambda \mapsto F_{\lambda}(n)$, il existe pour tout entier n deux réels positifs $\lambda_{-}(n)$ et $\lambda_{+}(n)$ tels que

$$F_{\lambda_{+}(n)}(n) = \frac{\alpha}{2}, \quad F_{\lambda_{-}(n)}(n-1) = 1 - \frac{\alpha}{2}.$$

On convient que $\lambda_{-}(0) = 0$. L'intervalle $[\lambda_{-}(N_t), \lambda_{+}(N_t)]$ est alors un intervalle de confiance de niveau $1 - \alpha$ pour λ .

2. On observe uniquement les temps de sauts T_1, \dots, T_n du processus. Dans ce cas la loi de (T_1, \dots, T_n) est donnée par la Proposition 10 et a pour densité

$$(t_1, \dots, t_n) \rightarrow \lambda^n \exp(-\lambda t_n) \mathbb{1}_{0 < t_1 < \dots < t_n},$$

par rapport à la mesure de Lebesgue sur \mathbb{R}^n . Il est alors facile de voir que la maximum de vraisemblance est donné par $\hat{\lambda} = \frac{n}{T_n} = \frac{N_{T_n}}{T_n}$. $\hat{\lambda}$ converge p.s vers λ (voir la preuve du Théorème 5). De plus, en décomposant $T_n = \sum_{i=0}^{n-1} (T_{i+1} - T_i)$ (somme de v.a indépendantes toutes de loi exponentielle de paramètre λ), une utilisation du théorème limite centrale permet d'obtenir la convergence en loi

$$\lim_{n \rightarrow +\infty} \sqrt{n} \left(\lambda \frac{T_n}{n} - 1 \right) = \mathcal{N}(0, 1).$$

A titre d'exercice, on pourra construire un intervalle de confiance de niveau donné pour λ à partir de cette convergence.

3. Supposons enfin que l'on observe le processus uniquement aux temps $t_1 < \dots < t_n$. On calcule alors le maximum de vraisemblance en utilisant l'indépendance des accroissements. La vraisemblance associée à l'échantillon $(N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}})$ est alors donnée par

$$L(d_1, \dots, d_n, \lambda) = \exp(-\lambda t_n) \lambda^{\sum_{i=1}^n d_i} \frac{t_1^{d_1} \cdots (t_n - t_{n-1})^{d_n}}{d_1! \cdots d_n!}.$$

On obtient alors

$$\hat{\lambda} = \arg \max_{\lambda > 0} \left\{ \exp(-\lambda t_n) \lambda^{N_{t_n}} \right\} = \frac{N_{t_n}}{t_n}.$$

Remarquons que l'estimateur est identique à celui qui aurait été obtenu si toute la trajectoire avait été observée sur $[0, t_n]$. Ceci n'est pas étonnant car N_t est une statistique exhaustive du modèle (elle contient toute l'information).

2.2 Le processus de Poisson non homogène

Le processus de Poisson simple possède des accroissements stationnaires. Cette caractéristique est assez réductrice en pratique. Par exemple, en fiabilité, le cumul des défaillances d'une machine peut faire apparaître une tendance au vieillissement. Dans ce cas, on s'attend à avoir une valeur de l'intensité λ plus importante à la fin de l'étude qu'au début. Le processus des arrivées aux caisses d'un supermarché ou des appels entrants vers un central téléphonique peut aussi connaître des changements au cours du temps. Il est alors naturel de remplacer l'intensité λ par une fonction $\lambda(\cdot)$.

2.2.1 Définition et propriétés fondamentales

Définition 9 Soit $\lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ une fonction localement intégrable et $\Lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ la fonction définie par

$$\Lambda(t) = \int_0^t \lambda(s) ds, \quad t \in \mathbb{R}_+.$$

On dit qu'un processus de comptage $(N_t)_{t \geq 0}$ est appelé un processus de Poisson d'intensité λ si

1. $N_0 = 0$ p.s.
2. Le processus est à accroissements indépendants.
3. $N_t - N_s$ suit une loi de Poisson de paramètre $\Lambda(t) - \Lambda(s)$, $0 \leq s < t$.

Remarques

- On a $\mathbb{E}(N_t) = \text{Var}(N_t) = \Lambda(t)$. De plus Λ est une fonction continue.
- Le processus de Poisson simple correspond à une intensité constante au cours du temps (i.e $\lambda(t) = \lambda$ pour tout $t \geq 0$).

Construction. La construction d'un processus de Poisson non homogène peut se faire à partir d'un processus de poisson simple. Si N est un processus de Poisson simple d'intensité 1, il est facile de vérifier que le processus \tilde{N} défini par $\tilde{N}_t = N_{\Lambda(t)}$ est un processus de Poisson non homogène d'intensité λ . Notons

$$\Lambda^{-1}(t) = \inf\{x \geq 0 : \Lambda(x) \geq t\}, \quad t \geq 0,$$

la pseudo-inverse continue à gauche de Λ (qui correspond à l'inverse si Λ est strictement croissante et de limite $+\infty$ en $+\infty$). On adopte la convention $\inf \emptyset = +\infty$. On pose alors $S_n = \Lambda^{-1}(T_n)$ pour $n \in \mathbb{N}^*$, où T_n est le temps du n -ième saut de N . La définition de Λ^{-1} assure que

$$\Lambda(x) \geq t \Leftrightarrow x \geq \Lambda^{-1}(t).$$

Ainsi on a

$$\tilde{N}_t = \sum_{n \geq 1} \mathbb{1}_{S_n \leq t}, \quad t \geq 0.$$

Applications

- On peut considérer que l'intensité des arrivées à un guichet subit une modification brusque au cours du temps : $\lambda(t) = \lambda_1$ si $t \leq t_0$ et $\lambda(t) = \lambda_2$ si $t > t_0$.
- En fiabilité, lorsque un matériel subit une succession de pannes et lorsque chaque panne est réparée instantanément, on utilise souvent un processus de Weibull, c'est à dire un processus de Poisson non homogène d'intensité :

$$\lambda(s) = \frac{\beta}{\alpha} \left(\frac{s}{\alpha} \right)^{\beta-1}, \quad s \geq 0.$$

On peut alors montrer que le premier instant de saut $S_1 = \Lambda^{-1}(T_1)$ (voir le paragraphe sur la construction) a pour loi la loi de Weibull de paramètres α et β dont la densité est donnée par

$$f(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha} \right)^{\beta-1} \exp\left(-\left(\frac{t}{\alpha}\right)^\beta\right) \mathbb{1}_{\mathbb{R}_+}(t).$$

La loi de Weibull est souvent utilisé en fiabilité car le taux de défaillance d'un matériel dont la durée de vie T a une densité sur \mathbb{R}_+ est défini par

$$\lim_{h \rightarrow 0^+} \frac{1}{h} \mathbb{P}(t < T \leq t + h | T > t)$$

et la valeur de ce dernier correspond bien à celui obtenu pour la loi de Weibull, c'est à dire $\lambda(t)$ (plus précisément le taux de défaillance est souvent une fonction affine du temps dans un graphe log-log, d'après des études empiriques). De plus si $\beta = 1$ (taux de panne constant autour du temps) on retrouve la loi exponentielle de paramètre $\lambda = \frac{1}{\alpha}$ et donc le processus de Poisson simple. Si $\beta > 1$, le taux de défaillance augmente avec le temps (usure du système) alors que si $\beta < 1$, le taux de défaillance diminue avec le temps (les éléments défectueux qui fragilisent le système tombent en panne rapidement). Les figures 2.3, 2.4 et 2.5 montrent des trajectoires dans les trois cas.

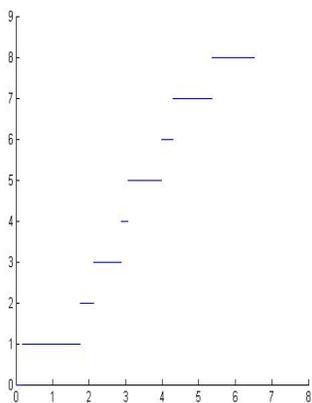


FIGURE 2.3 – $\alpha = 1, \beta = 1$

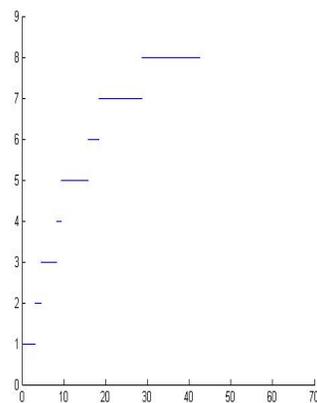


FIGURE 2.4 – $\alpha = 1, \beta = 0.5$

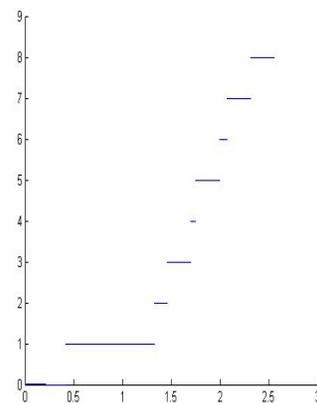


FIGURE 2.5 – $\alpha = 1, \beta = 2$

Proposition 12 *Supposons l'intensité λ continu au voisinage de t . Alors*

$$\mathbb{P}(N_{t+h} - N_t = 1) = \lambda(t)h + o(h),$$

$$\mathbb{P}(N_{t+h} - N_t \geq 2) = o(h).$$

Preuve. Il suffit de faire un développement limité comme pour le processus de Poisson simple. On a

$$\begin{aligned} \mathbb{P}(N_{t+h} - N_t = 1) &= \exp(\Lambda(t) - \Lambda(t+h)) (\Lambda(t+h) - \Lambda(t)) \\ &= (1 + o(1)) (\lambda(t)h + o(h)) \\ &= \lambda(t)h + o(h). \end{aligned}$$

On obtient ensuite

$$\begin{aligned} \mathbb{P}(N_{t+h} - N_t \geq 2) &= 1 - \mathbb{P}(N_{t+h} - N_t = 0) - \mathbb{P}(N_{t+h} - N_t = 1) \\ &= 1 - \exp(\Lambda(t) - \Lambda(t+h)) - (\lambda(t)h + o(h)) \\ &= \Lambda(t+h) - \Lambda(t) + o(\Lambda(t+h) - \Lambda(t)) - (\lambda(t)h + o(h)) \\ &= o(h), \end{aligned}$$

en utilisant le fait que $\Lambda(t+h) - \Lambda(t) = \lambda(t)h + o(h) = O(h)$. \square

Nous pouvons également généraliser certains résultats sur la loi des temps de saut. **Pour simplifier nous supposons dans la suite de ce chapitre que l'intensité du processus $\lambda(\cdot)$ est une fonction strictement positive, continue sur $]0, +\infty[$ et que $\lim_{t \rightarrow +\infty} \Lambda(t) = +\infty$.** Le processus de Weibull vérifie ces conditions.

Proposition 13 *Soit $(T_n)_{n \geq 1}$ désigne les temps de sauts d'un processus de Poisson non homogène d'intensité λ . Soit $n \in \mathbb{N}^*$.*

1. *Alors pour $n \in \mathbb{N}^*$, la loi de (T_1, \dots, T_n) a pour densité*

$$(t_1, \dots, t_n) \rightarrow \exp\left(-\int_0^{t_n} \lambda(u) du\right) \prod_{i=1}^n \lambda(t_i) \mathbb{1}_{0 < t_1 < \dots < t_n},$$

par rapport à la mesure de Lebesgue.

2. *Soit $t > 0$. Alors la loi conditionnelle de $(T_1, \dots, T_n) | N_t = n$ a pour densité*

$$(t_1, \dots, t_n) \rightarrow \frac{n!}{\Lambda(t)^n} \prod_{i=1}^n \lambda(t_i) \mathbb{1}_{0 < t_1 < \dots < t_n < t}.$$

Remarque. Rappelons que si X_1, \dots, X_n sont n variables aléatoires i.i.d de loi à densité f , le réarrangement croissant $(X_{(1)}, \dots, X_{(n)})$ de ces variables admet une loi dont la densité est

$$(y_1, \dots, y_n) \rightarrow n! \prod_{i=1}^n f(y_i) \mathbb{1}_{0 < y_1 < \dots < y_n}$$

par rapport à la mesure de Lebesgue. Ainsi la proposition ci-dessus assure que la loi conditionnelle de $(T_1, \dots, T_n) | N_t = n$ est celle du réarrangement croissant d'un n -échantillon dont la densité f est définie par $f(x) = \frac{\lambda(x)}{\Lambda(t)} \mathbb{1}_{[0,t]}(x)$.

Preuve

1. Λ réalise un C^1 difféomorphisme de \mathbb{R}_+^* sur lui même, d'inverse Λ^{-1} . D'après la construction du processus de Poisson non homogène, les temps $S_1 = \Lambda(T_1), \dots, S_n = \Lambda(T_n)$ sont les n premiers temps de saut d'un processus de Poisson homogène d'intensité 1. En appliquant la formule du changement de variable et le résultat de la Proposition 10, on a donc pour une fonction $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ mesurable bornée :

$$\begin{aligned} \mathbb{E}(\phi(T_1, \dots, T_n)) &= \mathbb{E}(\phi(\Lambda^{-1}(S_1), \dots, \Lambda^{-1}(S_n))) \\ &= \int \phi(\Lambda^{-1}(s_1), \dots, \Lambda^{-1}(s_n)) \exp(-s_n) \mathbb{1}_{0 < s_1 < \dots < s_n} ds_1 \cdots ds_n \\ &= \int \phi(t_1, \dots, t_n) \exp(-\Lambda(t_n)) \mathbb{1}_{0 < t_1 < \dots < t_n} \lambda(t_1) \cdots \lambda(t_n) dt_1 \cdots dt_n. \end{aligned}$$

2. On en déduit le deuxième point exactement comme pour le processus de Poisson simple (voir la preuve de la Proposition 11). \square

2.2.2 Estimation par maximum de vraisemblance

Comme pour le cas du processus de Poisson simple, on peut utiliser la Proposition 13 pour calculer la fonction de vraisemblance du modèle. Nous nous placerons dans un modèle paramétrique où l'intensité du processus appartient à une famille $\{\lambda_\theta : \theta \in \Theta\}$. Nous supposons ici uniquement le cas d'une trajectoire observée sur l'intervalle $[0, t]$. Si on observe $N_t = n$ et $(T_1, \dots, T_n) = (t_1, \dots, t_n)$, on a alors

$$L(t_1, \dots, t_n, n; \theta) = \prod_{i=1}^n \lambda_\theta(t_i) \exp(-\Lambda_\theta(t)).$$

Nous allons effectuer les calculs lorsque le processus est un processus de Weibull. Dans ce cas $\theta = (\alpha, \beta)$ et

$$\lambda_\theta(s) = \frac{\beta}{\alpha} \left(\frac{s}{\alpha}\right)^{\beta-1}, \quad s > 0.$$

Ainsi la log-vraisemblance s'écrit

$$\ell(t_1, \dots, t_n, n; \theta) = (\ln(\beta) - \beta \ln(\alpha)) n + (\beta - 1) \sum_{i=1}^n \ln(t_i) - \left(\frac{t}{\alpha}\right)^\beta.$$

L'unique point (α, β) qui annule les deux dérivées partielles est donné par

$$\frac{1}{\beta} = \ln(t) - \frac{1}{n} \sum_{i=1}^n \ln(t_i),$$

$$\ln(\alpha) = \ln(t) - \frac{1}{\beta} \ln(n).$$

On peut voir que ce point correspond au maximum de la vraisemblance en procédant comme suit. Posons $\lambda = \frac{1}{\alpha^\beta}$, $s = \sum_{i=1}^n \ln(t_i)$ et

$$g(\lambda, \beta) = n \ln(\beta) + n \ln(\lambda) + \beta s - \lambda t^\beta.$$

Pour $\beta > 0$ donné, la fonction concave $\lambda \rightarrow g(\lambda, \beta)$ atteint son maximum en $\lambda = nt^{-\beta}$, maximum qui vaut

$$h(\beta) = n \ln(\beta) + n \ln(n) - \beta n \ln(t) + \beta s - n.$$

La fonction h est concave sur $]0, +\infty[$ et atteint son maximum en $\beta = \frac{n}{n \ln(t) - s}$. On en déduit un maximum de la vraisemblance au point (α, β) annoncé. Ainsi le maximum de vraisemblance $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ est donné par

$$\frac{1}{\hat{\beta}} = \ln(t) - \frac{1}{N_t} \sum_{i=1}^{N_t} \ln(T_i),$$

$$\ln(\hat{\alpha}) = \ln(t) - \frac{1}{\hat{\beta}} \ln(N_t).$$

Nous n'étudierons pas les propriétés de convergence de ces estimateurs dans ce cours. Il est néanmoins possible de donner des intervalles de confiance pour le paramètre β , en utilisant les propriétés que nous avons vues.

Proposition 14 *Dans le cas de l'observation d'un processus de Weibull sur $[0, t]$, la loi de $\frac{2n\beta}{\hat{\beta}}$ sachant $N_t = n$ ($n \neq 0$) est la loi du χ^2 à $2n$ degrés de liberté. Ainsi, chacun des intervalles I_n suivants est un intervalle de confiance de niveau γ , au sens où $\mathbb{P}(\beta \in I_n | N_t = n) = \gamma$:*

- $I_n = [0, \frac{\hat{\beta}}{2n} \chi_\gamma^2(2n)]$,
- $I_n = [\frac{\hat{\beta}}{2n} \chi_{1-\gamma}^2(2n), +\infty[$,
- $I_n = [\frac{\hat{\beta}}{2n} \chi_{\frac{1-\gamma}{2}}^2(2n), \frac{\hat{\beta}}{2n} \chi_{\frac{1+\gamma}{2}}^2(2n)]$.

Le nombre $\chi_\gamma^2(2n)$ désigne le quantile d'ordre γ de la loi du χ^2 à $2n$ degrés de liberté.

Preuve. En utilisant l'expression de $\hat{\beta}$ et la Proposition 13, on peut vérifier que la loi conditionnelle de $\frac{2n\beta}{\hat{\beta}} | N_t = n$ est la même que celle de

$$Z_n = 2n\beta \ln(t) - 2\beta \sum_{i=1}^n \ln(X_i),$$

où (X_1, \dots, X_n) est un n -échantillon de loi de densité $s \rightarrow \frac{\lambda(s)}{\Lambda(t)} \mathbb{1}_{[0,t]}(s)$ (en utilisant aussi le fait que la somme de n variables aléatoires correspond à la somme des statistiques d'ordre). En effectuant le changement de variables $u = \ln(x)$, on peut montrer que la fonction caractéristique de $\ln(X_1)$ est donnée par

$$\phi_{\ln(X_1)}(s) = \frac{\beta \exp(\ln(t)(is + \beta))}{\Lambda(t)\alpha^\beta(is + \beta)}.$$

En utilisant les propriétés d'indépendance des X_i et l'expression de l'intensité λ , on en déduit

$$\phi_{Z_n}(u) = \frac{1}{(1 - i2u)^n}.$$

La fonction caractéristique de Z_n est la même que celle de la loi du χ^2 à $2n$ degrés de liberté. Les propriétés des intervalles de confiance sont alors immédiates. \square

Application à des problèmes de test sur le paramètre β . Soit $\beta_0 > 0$. Le cas $\beta_0 = 1$ est le plus important dans ce qui suit.

- Pour tester $H_0 : \beta = \beta_0$ contre $\beta \neq \beta_0$ au niveau γ , on rejette l'hypothèse nulle si

$$\beta_0 \notin \left[\frac{\hat{\beta}}{2N_t} \chi_{\frac{\gamma}{2}}^2(2N_t), \frac{\hat{\beta}}{2N_t} \chi_{1-\frac{\gamma}{2}}^2(2N_t) \right].$$

La zone de rejet s'écrit donc

$$\mathcal{R}_t = \left\{ \frac{2N_t\beta_0}{\hat{\beta}} \notin [\chi_{\frac{\gamma}{2}}^2(2N_t), \chi_{1-\frac{\gamma}{2}}^2(2N_t)] \right\}.$$

D'après le résultat de la Proposition 14, on a pour tout entier $n \geq 1$,

$$\mathbb{P}_{\beta_0}(\mathcal{R}_t | N_t = n) = \gamma,$$

ce qui entraîne que

$$\mathbb{P}_{\beta_0}(\mathcal{R}_t | N_t \geq 1) = \gamma.$$

Lorsque $\beta_0 = 1$, on obtient un test de niveau γ pour tester si le processus de Weibull est un processus de Poisson simple.

- Pour tester $H_0 : \beta \leq \beta_0$ contre $H_1 : \beta > \beta_0$, on définit

$$\mathcal{R}_t = \left\{ \beta_0 \leq \frac{\hat{\beta}}{2N_t} \chi_{\gamma}^2(2N_t) \right\} = \left\{ \frac{2N_t\beta_0}{\hat{\beta}} \leq \chi_{\gamma}^2(2N_t) \right\}.$$

D'après la Proposition 14, on a, par les mêmes arguments que pour le premier test,

$$\mathbb{P}_{\beta_0}(\mathcal{R}_t | N_t \geq 1) = \gamma.$$

Mais on a aussi

$$\mathbb{P}_{\beta}(\mathcal{R}_t | N_t \geq 1) \leq \gamma, \quad \beta \leq \beta_0. \tag{2.2}$$

Pour prouver (2.2), on utilise un argument de monotonie du rapport de vraisemblance. Plus précisément, pour $n \geq 1$, on a

$$\mathbb{P}_\beta(\mathcal{R}_t | N_t = n) = \mathbb{P}_\beta \left(\sum_{i=1}^n \log(T_i) \geq n \log(t) - \frac{\chi_\gamma^2(2n)}{2\beta_0} | N_t = n \right).$$

Mais en notant $f_\beta(n, \cdot)$ la densité de la loi de $(T_1, \dots, T_n) | N_t = n$, nous avons pour $\beta' > \beta$,

$$\frac{f_{\beta'}(n, T_1, \dots, T_n)}{f_\beta(n, T_1, \dots, T_n)} = \left(\frac{\beta'}{\beta} \right)^n t^{(\beta - \beta')n} \exp \left((\beta' - \beta) \sum_{i=1}^n \log(T_i) \right).$$

Le rapport des deux densités est donc une fonction strictement croissante en $\sum_{i=1}^n \log(T_i)$. Les résultats sur le rapport de vraisemblance monotone garantissent que (2.2) est vérifié dès lors que

$$\mathbb{P}_{\beta_0} \left(\sum_{i=1}^n \log(T_i) \geq n \log(t) - \frac{\chi_\gamma^2(2n)}{2\beta_0} | N_t = n \right) = \gamma.$$

On en déduit que pour tout $\beta \leq \beta_0$,

$$\mathbb{P}_\beta(\mathcal{R}_t | N_t \geq 1) \leq \gamma.$$

– On obtient de même un test de niveau γ pour tester $H_0 : \beta \geq \beta_0$ contre $H_1 : \beta < \beta_0$, en posant

$$\mathcal{R}_t = \left\{ \frac{2N_t\beta_0}{\hat{\beta}} \geq \chi_{1-\gamma}^2(2N_t) \right\}.$$

2.3 Processus de Poisson composé et modèle de Cramér-Lundberg

2.3.1 Définition

Définition 10 On appelle processus de Poisson composé, tout processus de la forme $X_t = \sum_{i=1}^{N_t} W_i$ où $(N_t)_{t \geq 0}$ est un processus de Poisson simple d'intensité $\lambda > 0$ et $(W_i)_{i \in \mathbb{N}^*}$ une suite i.i.d de variables aléatoires, indépendante de $(N_t)_{t \geq 0}$.

Exercice 12 Montrer qu'un processus de Poisson composé est à accroissements indépendants et stationnaires. Pour simplifier, on pourra montrer que pour $t > s$, les variables aléatoires X_s et $X_t - X_s$ sont indépendantes et que $X_t - X_s$ a même loi que X_{t-s} .

Exercice 13 Si $(X_t)_{t \geq 0}$ est un processus de Poisson composé, montrer les deux formules suivantes.

$$\mathbb{E}(X_t) = \lambda t \cdot \mathbb{E}(W_1), \quad \text{Var}(X_t) = \lambda t \cdot \mathbb{E}(W_1^2).$$

2.3.2 Modèle de Cramér-Lundberg et probabilité de ruine

La théorie de la ruine étudie les réserves financières d'une compagnie d'assurance au temps t . La réserve initiale de cette compagnie est $u > 0$ et les primes d'assurances sont supposées déterministes, de valeur c unités de compte par unité de temps. Les occurrences des sinistres sont modélisées par un processus de Poisson de paramètre $\lambda > 0$ et le montant des sinistres par des variables aléatoires i.i.d W_1, W_2, \dots de loi non dégénérée sur \mathbb{R}_+ . Le montant cumulé des sinistres est décrit par le processus $(S_t)_{t \geq 0}$ tel que $S_t = \sum_{i=1}^{N_t} W_i$ et le montant des réserves par le processus $(R_t)_{t \geq 0}$ définie par $R_t = u + ct - S_t$. Une des quantités essentielles est la probabilité de ruine (faire un dessin)

$$\psi(u) = \mathbb{P}(R_t < 0 \text{ pour un } t \in \mathbb{R}_+) = \mathbb{P}\left(\sum_{i=1}^n (W_i - cX_i) > u \text{ pour un } n \geq 0\right),$$

où X_1, X_2, \dots désignent les délais entre les sauts du processus de Poisson (donc des variables i.i.d de loi exponentielle). L'évaluation de $\psi(u)$ se ramène à un calcul de probabilité de franchissement de barrière d'une marche aléatoire. Le premier paramètre important est le chargement de sécurité $\rho = \frac{c}{\mathbb{E}(Y_1)\lambda} - 1$. Nous admettrons le résultat suivant.

Théorème 6 *On a $\psi(u) = 1$ si et seulement si $\rho \leq 0$.*

Remarquons que $\rho < 0$ entraîne $\mathbb{E}(W_1 - cX_1) > 0$. Dans ce cas, la loi des grands nombres assure que $M_n = \sum_{i=1}^n (W_i - cX_i)$ tend p.s vers $+\infty$ lorsque $n \rightarrow +\infty$. On a donc $\psi(u) = 1$. Lorsque $\rho = 0$, la marche aléatoire $(M_n)_n$ est centrée. Comme pour la marche aléatoire symétrique sur \mathbb{Z} , on peut montrer que $\limsup_n M_n = +\infty$ et $\liminf_n M_n = -\infty$. On a encore $\psi(u) = 1$. Lorsque $\rho > 0$, la marche aléatoire converge vers $-\infty$ p.s et on peut montrer que $\psi(u) < 1$. Remarquons que ces propriétés ne dépendent pas du caractère poissonien du processus de comptage. Elles restent valables lorsque les délais entre les sinistres X_1, X_2, \dots sont des variables i.i.d intégrables.

Chapitre 3

Processus Markoviens de sauts

Nous avons vu que les chaînes de Markov ont le désavantage de modéliser des phénomènes temporels en utilisant une discrétisation du temps. Cela ne correspond pas toujours au phénomène physique sous-jacent qui exigerait que des événements surviennent en des temps aléatoires (un peu comme le processus de Poisson qui compte le nombre d'événements survenus avant une date donnée). Par exemple, pour un instant $t \geq 0$ donné, $X_t(\omega)$ peut correspondre au nombre d'individus infectés par une maladie, au temps $t \geq 0$ et au sein d'une population donnée. Un autre exemple est le cas d'une machine qui est en fonctionnement ou en panne. Il est alors légitime de supposer les temps de marche ou de réparation aléatoires.

Ces exemples assez naturels amènent à définir des fonction aléatoire de sauts. Une fonction aléatoire de saut $t \mapsto X_t(\cdot)$ est de la forme

$$X_t(\omega) = \sum_{n=0}^{+\infty} Z_n(\omega) \mathbb{1}_{T_n(\omega) \leq t < T_{n+1}(\omega)} \mathbb{1}_{T_n(\omega) < +\infty}, \quad (3.1)$$

où les variables aléatoires Z_0, Z_1, \dots sont toutes à valeurs dans un ensemble E fini ou infini dénombrable (comme pour les chaînes de Markov, on dire que E est l'espace des états) et les variables aléatoires T_1, T_2, \dots sont des temps aléatoires tels que $T_n(\omega) < T_{n+1}(\omega)$ pour tous ω, n . On exigera souvent une condition naturelle pour la modélisation, à savoir que

$$\lim_{n \rightarrow +\infty} T_n = +\infty, \text{ p.s.} \quad (3.2)$$

Si la condition (3.2) n'était pas vérifiée, la fonction aléatoire pourrait changer d'état une infinité de fois avant une date donnée (accumulation de sauts en temps fini). La condition (3.2) est appelée condition de non-explosion. On autorise en fait la valeur $+\infty$ pour les temps aléatoires T_n , en convenant que si $T_n(\omega) = +\infty$, alors $T_{n+1}(\omega) = T_n(\omega)$ est $Z_{n+1}(\omega) = Z_n(\omega)$. Cette convention est naturelle si l'on veut autoriser des points absorbants, par analogie avec les chaînes de Markov. Par exemple, l'extinction d'une population après contamination est irréversible et l'état 0 est absorbant si X_t désigne le nombre de personnes infectés au temps t .

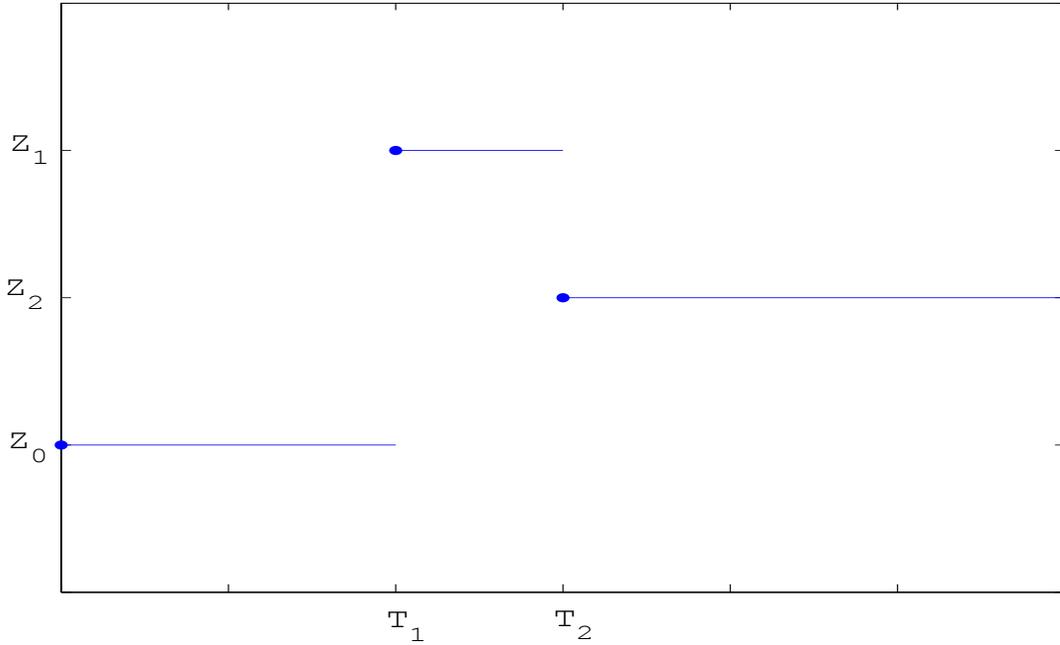


FIGURE 3.1 – Trajectoire d'une fonction aléatoire de sauts

3.1 Propriété de Markov d'une fonction aléatoire de sauts

Définition 11 Une fonction aléatoire de sauts $(X_t)_{t \geq 0}$ à valeurs dans E est un processus de Markov de sauts si

$$\mathbb{P}(X_t = y | X_{s_1} = x_1, \dots, X_{s_n} = x_n, X_s = x) = \mathbb{P}(X_t = y | X_s = x),$$

pour tous $0 \leq s_1 < \dots < s_n < s < t$ et tous $(x_1, \dots, x_n, x) \in E^{n+1}$ tels que $\mathbb{P}(X_{s_1} = x_1, \dots, X_{s_n} = x_n, X_s = x) > 0$.

On pourra alors remarquer que si $0 \leq t_0 < t_1 \dots$, alors la suite $(X_{t_i})_{i \in \mathbb{N}}$ est une chaîne de Markov. Par analogie avec les chaînes de Markov, on dira que le processus de Markov à temps continu $(X_t)_{t \geq 0}$ est homogène s'il existe une famille de matrices stochastiques $P_h : E \times E \rightarrow [0, 1]$ telles que

$$\mathbb{P}(X_t = y | X_s = x) = P_{t-s}(x, y),$$

pour tous $(s, t, x, y) \in \mathbb{R}_+ \times \mathbb{R}_+ \times E \times E$ tels que $s < t$ et $\mathbb{P}(X_s = x) > 0$. Remarquons alors que pour tout $h > 0$, la suite $(X_{kh})_{k \in \mathbb{N}}$ est une chaîne de Markov homogène de matrice de transition P_h . Dans la suite, nous ne considérerons que le cas homogène.

Proposition 15 Soit μ la loi initiale du processus de Markov de sauts $(X_t)_{t \geq 0}$ (c'est à dire que μ est la loi de X_0).

1. Pour tous $0 < t_1 < \dots < t_n$ et $x_0, \dots, x_n \in E$, on a

$$\mathbb{P}(X_0 = x_0, X_{t_1} = x_1, \dots, X_{t_n} = x_n) = \mu(x_0) P_{t_1}(x_0, x_1) P_{t_2 - t_1}(x_1, x_2) \cdots P_{t_n - t_{n-1}}(x_{n-1}, x_n).$$

$$2. \mathbb{P}(X_t = y) = \sum_{x \in E} \mu(x) P_t(x, y) = \mu P_t(y).$$

Proposition 16 (*Equations de Chapman-Kolmogorov*)

Pour tous $s, t \geq 0$, on a $P_{t+s} = P_t \cdot P_s = P_s \cdot P_t$.

Preuve. Prouvons l'égalité $P_{t+s} = P_s \cdot P_t$, l'autre égalité s'obtenant en échangeant le rôle de t et s . Si $x, y \in E$, nous avons

$$\begin{aligned} P_{t+s}(x, y) &= \mathbb{P}(X_{t+s} = y | X_0 = x) \\ &= \sum_{z \in E} \mathbb{P}(X_{t+s} = y, X_s = z | X_0 = x) \\ &= \sum_{z \in E} \mathbb{P}(X_{t+s} = y | X_s = z) \mathbb{P}(X_s = z | X_0 = x) \\ &= \sum_{z \in E} P_t(z, y) P_s(x, z) \\ &= (P_s \cdot P_t)(x, y), \end{aligned}$$

ce qui conduit au résultat. \square

Exemples

1. Si $x_0 \in \mathbb{N}$, et $X_t = x_0 + N_t$ où $(N_t)_{t \geq 0}$ est un processus de Poisson simple d'intensité $\lambda > 0$, alors $(X_t)_{t \geq 0}$ est un processus de Markov de sauts. En effet, on a pour $y \geq x$,

$$\begin{aligned} &\mathbb{P}(X_t = y | X_{s_1} = x_1, \dots, X_{s_n} = x_n, X_s = x) \\ &= \mathbb{P}(N_t = y - x_0 | N_{s_1} = x_1 - x_0, \dots, N_{s_n} = x_n - x_0, N_s = x - x_0) \\ &= \mathbb{P}(N_t - N_s = y - x | N_{s_1} = x_1 - x_0, \dots, N_{s_n} = x_n - x_0, N_s = x - x_0) \\ &= \mathbb{P}(N_t - N_s = y - x) \\ &= \exp(-\lambda(t-s)) \frac{\lambda^{y-x} (t-s)^{y-x}}{(y-x)!}. \end{aligned}$$

De plus, si $y < x$, la probabilité calculée précédemment vaut 0. La propriété de Markov est bien vérifiée et la transition associée est donnée par

$$P_h(x, y) = \exp(-\lambda h) \frac{\lambda^{y-x} h^{y-x}}{(y-x)!} \mathbb{1}_{y \geq x}, \quad h \geq 0.$$

2. Processus du télégraphe. Le processus défini par $X_t = x_0 \cdot (-1)^{N_t}$ où $x_0 \in \{-1, 1\}$ et $(N_t)_{t \geq 0}$ est un processus de Poisson d'intensité $\lambda > 0$ est un processus de Markov de sauts à valeurs dans $E = \{-1, 1\}$. Dans ce cas, on a

$$P_h(-1, -1) = P_h(1, 1) = \mathbb{P}((-1)^{N_t - N_s} = 1) = \sum_{k=0}^{+\infty} \exp(-\lambda(t-s)) \frac{[\lambda(t-s)]^{2k}}{(2k)!}.$$

De même,

$$P_h(-1, 1) = P_h(1, -1) = \mathbb{P}\left((-1)^{N_t - N_s} = 1\right) = \sum_{k=0}^{+\infty} \exp(-\lambda(t-s)) \frac{[\lambda(t-s)]^{2k+1}}{(2k+1)!}.$$

On peut généraliser les deux exemples précédents. La proposition suivante donne un procédé de construction assez général de processus de Markov de sauts.

Proposition 17 Soient $(Z_k)_{k \in \mathbb{N}}$ une chaîne de Markov à valeurs dans E , homogène de transition Q et $(N_t)_{t \geq 0}$ un processus de Poisson homogène d'intensité λ et indépendant de $(Z_k)_{k \in \mathbb{N}}$. Pour tout $t \geq 0$, posons $X_t = Z_{N_t}$. Alors $(X_t)_{t \geq 0}$ est un processus de Markov de sauts dont la transition est donnée par

$$P_t = \exp(\lambda t(Q - I)), \quad t \geq 0.$$

Remarque. Lorsque E est infini, la définition de l'exponentielle de matrice s'étend à l'ensemble \mathcal{F} des matrices A telles que $\|A\| = \sup_{x \in E} \sum_{y \in E} |A(x, y)| < +\infty$. En effet, l'ensemble \mathcal{F} est muni d'une structure d'espace vectoriel normé complet et $\|\cdot\|$ est une norme d'opérateur sur cet espace (c'est à dire que $\|AB\| \leq \|A\| \cdot \|B\|$). La série $\sum_{k=0}^{+\infty} \frac{A^k}{k!}$ est alors normalement convergente, de norme $\leq \exp(\|A\|)$. De plus, on garde la formule $\exp(A + B) = \exp(A) \cdot \exp(B) = \exp(B) \cdot \exp(A)$ lorsque $AB = BA$.

Preuve de la Proposition 17. Soient $0 = t_0 < t_1 < \dots < t_n$ et $x_0, x_1, \dots, x_n \in E$. On a

$$\begin{aligned} & \mathbb{P}(X_{t_0} = x_0, \dots, X_{t_n} = x_n) \\ &= \sum_{h_1, \dots, h_n \in \mathbb{N}} \mathbb{P}(N_{t_1} = h_1, N_{t_1} = h_1 + h_2, \dots, N_{t_n} = h_1 + \dots + h_n, Z_0 = x_0, Z_{h_1} = x_1, \dots, Z_{h_1 + \dots + h_n} = x_n) \\ &= \sum_{h_1, \dots, h_n \in \mathbb{N}} \mathbb{P}(N_{t_1} = h_1, N_{t_1} = h_1 + h_2, \dots, N_{t_n} = h_1 + \dots + h_n) \mathbb{P}(Z_0 = x_0) Q^{h_1}(x_0, x_1) \cdots Q^{h_n}(x_{n-1}, x_n) \\ &= \sum_{h_1, \dots, h_n \in \mathbb{N}} \mathbb{P}(N_{t_1} = h_1, N_{t_2} - N_{t_1} = h_2, \dots, N_{t_n} - N_{t_{n-1}} = h_n) \mathbb{P}(Z_0 = x_0) Q^{h_1}(x_0, x_1) \cdots Q^{h_n}(x_{n-1}, x_n) \\ &= \sum_{h_1, \dots, h_n \in \mathbb{N}} \exp(-\lambda t_1) \frac{(\lambda t_1)^{h_1}}{h_1!} \cdots \exp(-\lambda(t_n - t_{n-1})) \frac{(\lambda(t_n - t_{n-1}))^{h_n}}{h_n!} \mathbb{P}(Z_0 = x_0) Q^{h_1}(x_0, x_1) \cdots Q^{h_n}(x_{n-1}, x_n). \end{aligned}$$

On voit alors que la propriété de Markov est satisfaite et que la transition est donnée par

$$P_h(x, y) = \sum_{h \geq 0} \exp(-\lambda s) \frac{(\lambda s)^h}{h!} Q^h(x, y) = \exp(-\lambda s) [\exp(\lambda s Q)](x, y).$$

On en déduit le résultat, en remarquant que

$$\exp(\lambda s(Q - I)) = \exp(-\lambda s I) \cdot \exp(\lambda s Q) = \exp(-\lambda s) \cdot \exp(\lambda s Q). \quad \square$$

3.2 Le générateur infinitésimal

Les équations de Chapman-Kolmogorov permettent d'écrire pour $t > 0$ donné et tout entier positif n , $P_t = \left(P_{\frac{t}{n}}\right)^n$. Ainsi, P_t est connu lorsque il est connu en temps petit. En fait, nous allons voir qu'il est uniquement déterminé par sa dérivée à droite en 0 que l'on appelle générateur infinitésimal.

Théorème 7 *Si $(X_t)_{t \geq 0}$ est un processus de Markov de sauts, il existe une matrice $A : E \times E \rightarrow \mathbb{R}$ telle que*

1. *pour tous $x \neq y$, $A(x, y) \geq 0$ pour tout $(x, y) \in E^2$ tel que $x \neq y$ et $A(x, x) = -\sum_{z \neq x} A(x, z)$ (on parle de générateur markovien pour A si ces deux conditions sont vérifiées),*
2. *si $x \neq y$, $P_h(x, y) = hA(x, y) + o(h)$ et $P_h(x, x) = 1 + A(x, x)h + o(h)$.*

Le comportement de $h \mapsto P_h(x, y)$ au voisinage de 0 est lié à l'étude de la loi de (T_1, Z_1) (on rappelle que T_1 désigne le premier instant de saut et $Z_1 = X_{T_1}$). Le lemme suivant sera utile pour la suite. Si $x \in E$ et A est un évènement, nous noterons $\mathbb{P}_x(A)$ la probabilité conditionnelle $\mathbb{P}(A|X_0 = x)$.

Lemme 3 *Soit $h, t > 0$ et $m = m(h, t) \in \mathbb{N}$ tels que $mh \searrow_{h \searrow 0} t$ (par exemple $m = \left\lfloor \frac{t}{h} \right\rfloor + 1$). Alors pour tous $x \neq y$, on a les deux limites suivantes.*

$$\mathbb{P}_x(T_1 > t) = \lim_{h \searrow 0} P_h(x, x)^m, \quad \mathbb{P}_x(T_1 \leq t, Z_1 = y) = \lim_{h \searrow 0} \frac{1 - P_h(x, x)^m}{1 - P_h(x, x)} \cdot P_h(x, y).$$

Preuve du Lemme 3. Remarquons tout d'abord que tout évènement B satisfait à l'inclusion

$$B \subset (B \cap \{T_2 - T_1 > h\}) \cup \{T_2 - T_1 \leq h\}.$$

Nous utiliserons ce type d'inclusion dans ce qui suit. Prouvons d'abord la première limite. Nous avons les inclusions

$$\{T_1 > mh\} \subset B = \{X_0 = X_h = \dots = X_{mh}\} \subset \{T_1 > mh\} \cup \{T_2 - T_1 \leq h\}.$$

La première inclusion est évidente. De plus, pour la deuxième, on peut remarquer que sur l'évènement $B \cap \{T_2 - T_1 > h\}$ on ne peut avoir $T_1(\omega) \leq mh$ car sinon on aurait un deuxième saut avant $T_1(\omega) + h$. Comme $\lim_{h \searrow 0} \mathbb{P}_x(T_2 - T_1 \leq h) = \mathbb{P}_x(T_2 = T_1) = 0$, on obtient

$$\begin{aligned} \mathbb{P}_x(T_1 > t) &= \lim_{h \searrow 0} \mathbb{P}_x(T_1 > mh) \\ &= \lim_{h \searrow 0} \mathbb{P}_x(X_0 = X_h = \dots = X_{mh}) \\ &= \lim_{h \searrow 0} P_h(x, x)^m. \end{aligned}$$

Pour la deuxième limite, posons

$$A = \{\exists 1 \leq \ell \leq m : X_0 = X_h = \dots = X_{(\ell-1)h} = x, X_{\ell h} = y\}.$$

On a les inclusions

$$A \subset \{X_0 = x, Z_1 = y, T_1 \leq mh\} \cup \{T_2 - T_1 \leq h\},$$

$$\{X_0 = x, Z_1 = y, T_1 \leq mh\} \subset A \cup \{T_2 - T_1 \leq h\}.$$

En effet sur $A \cap \{T_2 - T_1 > h\}$, on a $T_1(\omega) \leq mh$ et on ne peut avoir $Z_1(\omega) \neq y$ car sinon le processus aurait effectué deux sauts dans un intervalle de longueur h . La même raison conduit à la deuxième inclusion.

On voit alors que

$$\begin{aligned} \mathbb{P}_x(T_1 \leq t, Z_1 = y) &= \lim_{h \searrow 0} \mathbb{P}_x(T_1 \leq mh, Z_1 = y) \\ &= \lim_{h \searrow 0} \mathbb{P}_x(\exists \ell \leq m : x = X_h = \dots = X_{(\ell-1)h}, X_{\ell h} = y) \\ &= \lim_{h \searrow 0} \sum_{\ell=1}^m \mathbb{P}(x = X_h = \dots = X_{(\ell-1)h}, X_{\ell h} = y) \\ &= \lim_{h \searrow 0} \sum_{\ell=1}^m P_h(x, x)^{m-\ell} P_h(x, y) \\ &= \lim_{h \searrow 0} \frac{1 - P_h(x, x)^m}{1 - P_h(x, x)} P_h(x, y). \square \end{aligned}$$

Preuve du Théorème 7. La preuve utilise le Lemme 3. En gardant les notations de ce lemme, nous avons

$$\ln(\mathbb{P}_x(T_1 > t)) = \lim_{h \searrow 0} m \ln(P_h(x, x)) \in [-\infty, 0].$$

Comme $\lim_{h \searrow 0} P_h(x, x) = 1$, on a les équivalents (lorsque $h \searrow 0$)

$$\frac{\ln(\mathbb{P}_x(T_1 > t))}{t} \sim \frac{hm \ln(P_h(x, x))}{th} \sim \frac{\ln(P_h(x, x))}{h} \sim \frac{P_h(x, x) - 1}{h}.$$

On en déduit que le quotient $\frac{P_h(x, x) - 1}{h}$ possède une limite notée $-a(x)$ avec $a(x) \in [0, +\infty]$. On a de plus $a(x) = -\frac{\ln(\mathbb{P}_x(T_1 > t))}{t}$ pour tout $t > 0$. Remarquons que le cas $a(x) = +\infty$ est impossible, sinon on aurait $\mathbb{P}_x(T_1 > t) = 0$ pour tout $t > 0$ et donc $\mathbb{P}_x(T_1 = 0) = 1$, ce qui contredit la définition de la fonction aléatoire de sauts. De plus, si $a(x) = 0$, alors $\mathbb{P}_x(T_1 > t) = 1$, ce qui entraîne $\mathbb{P}_x(T_1 = +\infty) = 1$ (on dit alors que x est un état absorbant). En posant $A(x, x) = -a(x)$, le développement limité est justifié lorsque $y = x$.

Supposons ensuite $y \neq x$ et $a(x) \neq 0$. D'après le Lemme 3,

$$\mathbb{P}_x(T_1 \leq t, Z_1 = y) = \lim_{h \searrow 0} h \frac{1 - P_h(x, x)^m}{1 - P_h(x, x)} \cdot \frac{P_h(x, y)}{h}.$$

D'après la première partie de la preuve, on a

$$\lim_{h \searrow 0} h \frac{1 - P_h(x, x)^m}{1 - P_h(x, x)} = \frac{1 - \exp(-a(x)t)}{a(x)}.$$

On en déduit que le quotient $\frac{P_h(x,y)}{h}$ possède une limite quand $h \searrow 0$. Notons $A(x,y)$ cette limite. Nous avons alors

$$\mathbb{P}_x(T_1 \leq t, Z_1 = y) = (1 - \exp(-a(x)t)) \frac{A(x,y)}{a(x)}.$$

En sommant sur $y \neq x$, il vient

$$\mathbb{P}_x(T_1 \leq t) = (1 - \exp(-a(x)t)) \frac{\sum_{y \neq x} A(x,y)}{a(x)}.$$

Comme $\mathbb{P}_x(T_1 > t) = 1 - \exp(-a(x)t)$, on en déduit que $\sum_{y \neq x} A(x,y) = a(x) = -A(x,x)$. Lorsque $a(x) = 0$, on a $P_h(x,y) = 0$ pour tout $y \neq x$ (puisque $\mathbb{P}_x(T_1 = +\infty) = 1$). La preuve du théorème est complète. \square

D'après la preuve du Théorème 7, on a, lorsque $a(x) \neq 0$,

$$\mathbb{P}_x(T_1 > t, Z_1 = y) = (1 - \exp(-a(x)t)) \frac{A(x,y)}{a(x)}, \quad y \neq x.$$

On en déduit le résultat suivant.

Proposition 18 *Supposons $a(x) > 0$ (i.e x n'est pas un état absorbant). Alors, conditionnellement à $X_0 = x$, les variables aléatoires T_1 et Z_1 sont indépendantes. De plus, le premier temps de saut T_1 suit une loi exponentielle de paramètre $a(x) = -A(x,x)$ et la loi de Z_1 est donnée par $\mathbb{P}_x(Z_1 = y) = \frac{A(x,y)}{a(x)}$, $y \neq x$.*

La proposition suivante montre que la famille de transitions $\{P_t : t \geq 0\}$ est solution de deux équations différentielles.

Proposition 19 *On a $P'_t = AP_t$ (équation rétrograde ou backward) et $P'_t = P_t A$ (équation progressive ou forward).*

Lorsque E est fini, il est connu que l'équation $P'_t = AP_t$ avec condition initiale $P_0 = I$ a pour unique solution $P_t = \exp(tA)$. Les matrices de transition s'écrivent donc comme une exponentielle du générateur. On peut aussi prouver que cette représentation reste valable lorsque E est infini, en particulier lorsque $\sup_{x \in E} a(x) < +\infty$. Au passage, remarquons que $\sup_{x \in E} a(x) < +\infty$ est équivalent pour un générateur markovien à $\sup_{x \in E} \sum_{y \in E} |A(x,y)| < +\infty$. En pratique, c'est plutôt par les développements limités du Théorème 7 que l'on spécifie le générateur A car les $A(x,y)$ pour $y \neq x$ s'interprètent comme des taux de transition. Le processus passera de la valeur x à la valeur y au bout d'un temps h petit d'autant plus facilement que $A(x,y)$ est élevé.

Preuve. Nous ne prouverons ce résultat que lorsque E est fini. En partant de $P_{t+h} = P_t P_h$ (équations de Chapman-Kolmogorov), on a $\frac{P_{t+h} - P_t}{h} = P_t \cdot \frac{P_h - I}{h}$, où I désigne la matrice identité ($I(x,y) = 1$ si $x = y$ et 0 sinon). Ainsi, pour tout $(x,y) \in E^2$,

$$\frac{P_{t+h}(x,y) - P_t(x,y)}{h} = \sum_{z \in E} P_t(x,z) \frac{P_h(z,y) - I(z,y)}{h}.$$

Il suffit alors de prendre la limite lorsque $h \searrow 0$ en utilisant le Théorème 7. Ceci prouve l'équation forward. L'équation backward se prouve de la même façon à partir de l'égalité $\frac{P_{t+h}-P_t}{h} = \frac{P_t-I}{h} P_t$. \square

Nous admettrons le résultat suivant qui caractérise la loi des délais entre les sauts et des états. Dès que $T_n(\omega) = +\infty$, nous conviendrons que $T_{n+j}(\omega) - T_{n+j-1}(\omega) = +\infty$ pour tout $j \geq 1$. De plus, toujours par convention, une loi exponentielle de paramètre 0 correspondra à la loi d'une variable aléatoire égale à $+\infty$ p.s.

Proposition 20 Soit $(X_t)_{t \geq 0}$ un processus de Markov de sauts de générateur A .

1. La suite $(Z_n = X_{T_n})_{n \in \mathbb{N}}$ est une chaîne de Markov de transition Q définie par

$$Q(x, y) = \begin{cases} \frac{A(x, y)}{a(x)} \mathbb{1}_{y \neq x} & \text{si } a(x) > 0, \\ \mathbb{1}_{x=y} & \text{sinon} \end{cases}$$

On dit que $(Z_n)_{n \in \mathbb{N}}$ est la chaîne immergée ou induite du processus.

2. Conditionnellement à Z_0, Z_1, \dots les délais $T_1, T_2 - T_1, \dots$ sont indépendants et suivent des lois exponentielles de paramètres $a(Z_0), a(Z_1), \dots$

On a déjà calculé la loi du couple (T_1, Z_1) conditionnellement à $Z_0 = x$ (voir Proposition 18). Nous admettrons que la propriété de Markov du processus garantit que la loi conditionnelle $(T_2 - T_1, Z_2) | Z_1 = x, T_1 = t$ coïncide avec la loi conditionnelle $(T_1, Z_1) | Z_0 = x$. Plus généralement, conditionnellement à $Z_1 = x_1, \dots, Z_n = x_n, T_1 = s_1, \dots, T_n - T_{n-1} = s_n$, la loi du couple $(T_{n+1} - T_n, Z_{n+1})$ ne dépend que de x_n et coïncide avec la loi $(T_1, Z_1) | Z_0 = x_n$. Une façon de résumer le résultat de la Proposition 20 est la suivante. Pour tout entier $n \geq 1$, tous $s_1, \dots, s_n \in \mathbb{R}_+$ et tous $x_1, \dots, x_n \in E$,

$$\mathbb{P}_{x_0}(S_1 > s_1, \dots, S_n > s_n, Z_1 = x_1, \dots, Z_n = x_n) = \prod_{i=0}^{n-1} Q(x_i, x_{i+1}) \exp(-a(x_{i+1})s_{i+1}).$$

3.3 Existence et construction d'un processus de Markov de générateur donné

3.3.1 Trois exemples de construction

Supposons donnée une matrice $A : E \times E \rightarrow \mathbb{R}$ telle que $A(x, y) \geq 0$ pour $x \neq y$ et $A(x, x) = -\sum_{y \neq x} A(x, y)$. Ci-dessous, nous donnons trois constructions qui permettent de définir un processus de Markov de sauts de générateur A . La condition de non-explosion sera étudiée dans la sous-section suivante.

Construction d'une fonction aléatoire de sauts Soit $(Z_n)_{n \in \mathbb{N}}$ une chaîne de Markov de matrice de transition Q définie par

$$Q(x, y) = \begin{cases} \frac{A(x, y)}{a(x)} & \text{si } a(x) \neq 0, \\ \mathbb{1}_{x=y} & \text{sinon} \end{cases}$$

Si $0 = \tilde{T}_0 < \tilde{T}_1 < \tilde{T}_2 < \dots$ sont les temps de saut d'un processus de Poisson simple d'intensité $\lambda = 1$ indépendant de la suite $(Z_n)_{n \in \mathbb{N}}$, on pose pour tout $n \in \mathbb{N}$,

$$S_n = \begin{cases} \frac{\tilde{T}_n - \tilde{T}_{n-1}}{a(Z_{n-1})} & \text{si } a(Z_{n-1}) \neq 0, \\ +\infty & \text{si } a(Z_{n-1}) = 0 \end{cases}, \quad T_n = S_1 + \dots + S_n$$

en convenant que $T_0 = 0$. La fonction aléatoire de sauts associée est un alors processus de Markov de générateur A . En toute rigueur, il faudrait démontrer que cette fonction aléatoire de sauts vérifie la propriété de Markov. Nous l'admettons.

Simulation dynamique Supposons que $X_0 = x$. On simule des variables aléatoires indépendantes $\{U_y : y \in E\}$ telles que pour tout $y \in E$, U_y suive une loi exponentielle de paramètre $A(x, y)$. La plus petite des valeurs des U_y donne le délai à attendre avant le prochain saut et la valeur $y_0 \in E$ (qui dépend de $\omega \in \Omega$) qui réalise ce minimum est l'état suivant. On recommence alors ce procédé pour définir un nouveau délai ainsi qu'un nouvel état. Justifions ce type de construction. Si l'état courant est x , soit I la variable aléatoire à valeurs dans $\{y : y \in E \setminus \{x\}\}$ telle que $U_I(\omega) \leq U_y(\omega)$ pour tout $y \neq x$. Alors on a

$$\begin{aligned} \mathbb{P}(I = y, U_I > t) &= \mathbb{P}(E_z > E_y > t; z \neq x, y) \\ &= \int_t^{+\infty} A(x, y) e^{-A(x, y)v} e^{-\sum_{z \neq x, y} A(x, z)v} dv \\ &= \int_t^{+\infty} A(x, y) e^{-a(x)v} dv \end{aligned}$$

On obtient

$$\mathbb{P}(I = y, U_I > t) = \begin{cases} \frac{A(x, y)}{a(x)} e^{-a(x)t} & \text{si } a(x) \neq 0, \\ 0 & \text{si } a(x) = 0 \end{cases}$$

Ainsi le couple (I, U_I) a bien la loi du couple (Z_1, T_1) d'un processus de Markov de sauts de générateur A et issu de $X_0 = x$. En itérant cette construction, on définit une fonction aléatoire de sauts qui a la même loi que la fonction de sauts du paragraphe précédent.

Construction à partir d'une chaîne de Markov et d'un processus de Poisson Si $(Y_n)_{n \in \mathbb{N}}$ est une chaîne de Markov sur E de transition Q et indépendante d'un processus de Poisson simple $(N_t)_{t \geq 0}$ d'intensité $\lambda > 0$, la Proposition 17 garantit que le processus défini par $X_t = Y_{N_t}$ est un processus de Markov dont les transitions sont données par $P_t = \exp(\lambda(Q - I))$. On obtiendra un processus de Markov de générateur A en posant $Q = \frac{A}{\lambda} + I$. Cette construction n'est possible que si $\frac{A}{\lambda} + I$ est une matrice de transition (matrice stochastique) ce qui revient à avoir $\frac{A(x, x)}{\lambda} + 1 \geq 0$ ou encore $\sup_{x \in E} a(x) \leq \lambda$. Ainsi, lorsque $\sup_{x \in E} a(x) < +\infty$, on peut utiliser cette construction. On notera toutefois que $(Y_n)_{n \in \mathbb{N}}$ n'est pas la chaîne immergée du processus puisque $Q(x, x) \neq 0$ pour certains états $x \in E$. De plus les délais entre les sauts ne sont pas donnés par le processus de Poisson (on saute en tout instant t tel que $n = N_t(\omega) = \lim_{s \nearrow t} N_s(\omega) + 1$ uniquement si $Y_n(\omega) \neq Y_{n-1}(\omega)$).

3.3.2 Conditions de non-explosion

Proposition 21 *La condition de non-explosion $\lim_{n \rightarrow +\infty} T_n = +\infty$ p.s s'écrit aussi*

$$\sum_{n=0}^{+\infty} \frac{1}{a(Z_n)} = +\infty, \quad p.s.$$

Remarques

1. On voit alors que **si E est fini, il ne peut y avoir explosion.**
2. On a

$$\begin{aligned} \mathbb{P}\left(\lim_{n \rightarrow +\infty} T_n < +\infty\right) &= \mathbb{P}\left(\sum_{i \in \mathbb{N}} (T_{i+1} - T_i) < +\infty\right) \\ &= \mathbb{E}\left(\mathbb{P}\left(\sum_{i \in \mathbb{N}} (T_{i+1} - T_i) < +\infty \mid Z_0, Z_1, \dots\right)\right). \end{aligned}$$

On a donc non-explosion ssi

$$\mathbb{P}\left(\sum_{i \in \mathbb{N}} (T_{i+1} - T_i) < +\infty \mid Z_0, Z_1, \dots\right) = 0 \text{ p.s.}$$

On peut en fait montrer que pour des variables aléatoires indépendantes U_1, U_2, \dots toutes de loi exponentielle de paramètres respectifs $\lambda_1, \lambda_2, \dots$, $\mathbb{P}(\sum_{k \geq 1} U_k < +\infty) = 0$ ssi $\sum_{k \geq 1} \frac{1}{\lambda_k} = +\infty$. On déduit alors de cette propriété et du calcul qui précède le résultat de la Proposition.

Corollaire 5 *L'une des deux conditions suivantes garantit la non-explosion.*

1. $\sup_{x \in E} a(x) < +\infty$.
2. La chaîne immergée $(Z_n)_{n \in \mathbb{N}}$ est récurrente irréductible.

Preuve du Corollaire

1. On montre que la première condition est suffisante en notant que $\frac{1}{a(Z_n)} \geq \frac{1}{\sup_{x \in E} a(x)} > 0$ et en utilisant la Proposition 21.
2. Comme la chaîne immergée passe une infinité de fois par un état x donné, on

$$\sum_{n \in \mathbb{N}} \frac{1}{a(Z_n)} \geq \frac{K}{a(x)}$$

pour tout entier K ce qui entraîne la divergence de la série.

Donnons enfin sans démonstration une condition nécessaire et suffisante de non-explosion qui porte directement sur le générateur.

Proposition 22 *Il y a non-explosion si et seulement si $k \equiv 0$ est l'unique solution bornée et à coordonnées positives de l'équation $Ak = k$.*

3.4 Estimation du générateur par maximum de vraisemblance

Dans cette section, on suppose que l'espace des états E est fini. Soit $(X_s)_{s \geq 0}$ un processus de Markov de sauts sur E , de générateur A_0 , observé sur l'intervalle $[0, t]$ où $t > 0$. Les trajectoires $s \in [0, t] \mapsto X_s(\omega)$ peuvent être résumées par

- le nombre de sauts $\tilde{N}_t(\omega) = n \in \mathbb{N}$ avant t ,
- les délais entre les sauts $S_1 = T_1, S_2 = T_2 - T_1, \dots, S_n = T_n - T_{n-1}$,
- les différents états du processus Z_0, \dots, Z_n .

On travaillera conditionnellement à $X_0 = Z_0 = x_0$. Intuitivement le calcul de la fonction de vraisemblance se fait à partir du calcul des probabilités du type

$$\mathbb{P}_{x_0}^{(A)} \left(\tilde{N}_t = n, S_1 \in B_1, \dots, S_n \in B_n, Z_1 = x_1, \dots, Z_n = x_n \right)$$

où B_1, \dots, B_n sont des boréliens de \mathbb{R}_+ , $x_1, \dots, x_n \in E$ tels que $x_i \neq x_{i+1}$ et $\mathbb{P}^{(A)}$ désigne la mesure de probabilité sous laquelle le processus $(X_t)_{t \geq 0}$ a pour générateur A . En utilisant la Proposition 20, nous avons

$$\begin{aligned} & \mathbb{P}_{x_0}^{(A)} \left(\tilde{N}_t = n, S_1 \in B_1, \dots, S_n \in B_n, Z_1 = x_1, \dots, Z_n = x_n \right) \\ &= \mathbb{P}_{x_0}^{(A)} \left(T_n \leq t < T_{n+1}, S_1 \in B_1, \dots, S_n \in B_n, Z_1 = x_1, \dots, Z_n = x_n \right) \\ &= \prod_{i=0}^{n-1} Q(x_i, x_{i+1}) \\ &\times \int_{B_1 \times \dots \times B_n} \exp \left(-a(x_n) \left(t - \sum_{j=1}^n s_j \right) \right) \mathbb{1}_{\{\sum_{j=1}^n s_j \leq t\}} \prod_{i=0}^{n-1} a(x_i) \exp(-a(x_i) s_{i+1}) ds_1 \cdots ds_n \\ &= \int_{B_1 \times \dots \times B_n} \exp \left(-a(x_n) \left(t - \sum_{j=1}^n s_j \right) \right) \mathbb{1}_{\{\sum_{j=1}^n s_j \leq t\}} \prod_{i=0}^{n-1} A(x_i, x_{i+1}) \exp(-a(x_i) s_{i+1}) ds_1 \cdots ds_n. \end{aligned}$$

Soit alors

$$L_A(n, S_1, \dots, S_n, Z_1, \dots, Z_n) = \prod_{i=0}^{n-1} A(Z_i, Z_{i+1}) \exp(-a(Z_i) S_{i+1} - a(Z_n)(t - T_n))$$

et donc

$$\begin{aligned} & \ln L_A(n, S_1, \dots, S_n, Z_1, \dots, Z_n) \\ &= \sum_{i=0}^{n-1} A(Z_i, Z_{i+1}) - \sum_{i=0}^{n-1} \{a(Z_i) S_{i+1} + a(Z_n)(t - T_n)\} \\ &= \sum_{x \neq y} N_t(x, y) \ln(A(x, y)) - \sum_{x \in E} a(x) R_t(x) \\ &= \sum_{x \neq y} \{N_t(x, y) \ln(A(x, y)) - A(x, y) R_t(x)\}. \end{aligned}$$

où $R_t(x)$ désigne le temps passé en x dans $[0, t]$ et $N_t(x, y)$ le nombre de transitions de x vers y avant t . Il est alors facile de vérifier que le maximum de vraisemblance est donnée par $\hat{A}(x, y) = \frac{N_t(x, y)}{R_t(x)}$ pour $x \neq y$ (on conviendra que l'estimateur est nul si $R_t(x) = 0$). Remarquons que $R_t(x) = \int_0^t \mathbb{1}_{\{X_s=x\}} ds$. Nous n'étudierons pas les propriétés asymptotiques de cet estimateur.

3.5 Classification des états et Théorème ergodique

Définition 12 Un processus de Markov de sauts $(X_t)_{t \geq 0}$ est dit irréductible, récurrent ou transient si la chaîne immergée $(Z_n)_{n \in \mathbb{N}}$ l'est.

Noter que le générateur d'un processus irréductible doit nécessairement vérifier $a(x) > 0$ pour tout $x \in E$, puisque $a(x) = 0$ signifie que x est un point absorbant.

Proposition 23 Un processus de Markov $(X_t)_{t \geq 0}$ est irréductible si et seulement si $P_t(x, y) > 0$ pour tout $(t, x, y) \in \mathbb{R}_+^* \times E \times E$.

Preuve (\Rightarrow) Si $(X_t)_{t \geq 0}$ est irréductible, il existe $x_0 = x, x_1, \dots, x_n = y$ tel que $Q(x_i, x_{i+1}) > 0$. D'où

$$\begin{aligned} P_t(x, y) &= \mathbb{P}(X_t = y | X_0 = x) \\ &\geq \mathbb{P}(X_{i_n} = x_i, 1 \leq i \leq n | X_0 = x) \\ &= \prod_{i=1}^n P_{t_i}^i(x_{i-1}, x_i). \end{aligned}$$

De plus, pour tout $h > 0$, on a

$$P_h(x_i, x_{i+1}) \geq \mathbb{P}_{x_i}(Z_1 = x_{i+1}, T_1 \leq h, T_2 > h) > 0.$$

La stricte positivité de la probabilité précédente est une conséquence de l'expression de la loi de $Z_1, T_1, T_2 - T_1$ conditionnellement à $Z_0 = X_0 = x_i$. Ceci prouve que $P_t(x, y) > 0$.

(\Leftarrow) Si $x \neq y$ et $t > 0$, on a

$$0 < P_t(x, y) \leq \mathbb{P}_x(\exists n \geq 1 : Z_n = y).$$

Ceci entraîne immédiatement l'irréductibilité de la chaîne immergée. \square

Définition 13 On dit qu'une probabilité $\pi : E \rightarrow \mathbb{R}_+$ est invariante si $\pi P_t = \pi$ pour tout $t \geq 0$.

Remarques

1. Si π est une probabilité invariante et que la loi initiale μ est égale à π , alors le processus $(X_t)_{t \geq 0}$ a même loi que le processus $(X_{t+s})_{t \geq 0}$ et ce pour tout $s \geq 0$. On dit que le processus de Markov est stationnaire. En particulier, pour tout $t \geq 0$, la loi de X_t est π .

2. On prendra garde au fait qu'une mesure invariante pour $(X_t)_{t \geq 0}$ n'est pas une mesure invariante pour la chaîne immergée en général. Par exemple si $E = \{0, 1\}$, $A = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$, alors $\pi = \left(\frac{\mu}{\lambda+\mu}, \frac{\lambda}{\lambda+\mu}\right)$ est une probabilité invariante pour $(X_t)_{t \geq 0}$ (voir la dernière section de ce chapitre) alors que $Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ a pour probabilité invariante $\left(\frac{1}{2}, \frac{1}{2}\right)$.
3. On suppose E fini. Si $\pi P_t = \pi$, alors en dérivant en $t = 0$, on obtient $\pi A = 0$. Réciproquement, si $\pi A = 0$, on a

$$\pi P_t = \pi \exp(tA) = \pi I + \sum_{k=1}^{+\infty} \frac{t^k}{k!} \pi A^k = \pi$$

et π est invariante pour $(X_t)_{t \geq 0}$. Ce résultat se généralise au cas $|E| = +\infty$.

Proposition 24 Une probabilité π est invariante si et seulement si $\pi A = \pi$.

Définition 14 Soient $x \in E$ et

$$R_x = \inf \{t \geq T_1 : X_t = x\}.$$

L'état x est dit récurrent positif si $\mathbb{E}_x(R_x) < +\infty$ et récurrent nul si $\mathbb{E}_x(R_x) = +\infty$.

Théorème 8 Soit $(X_t)_{t \geq 0}$ un processus de Markov irréductible.

1. Si $(X_t)_{t \geq 0}$ est transient alors il n'existe pas de probabilité invariante.
2. Si $(X_t)_{t \geq 0}$ est récurrent alors tous les états sont soit récurrents nuls soit récurrents positifs. On dit alors que le processus est récurrent nul ou récurrent positif. S'il est récurrent nul, il n'existe pas de probabilité invariante. S'il est récurrent positif, il existe une unique probabilité invariante π donnée par $\pi(x) = \frac{1}{a(x)\mathbb{E}_x(R_x)}$ (en particulier, la probabilité invariante charge tous les points).

Remarques

1. Le premier point du théorème précédent est facile à justifier. Si $(X_t)_{t \geq 0}$ est transient alors $(Z_n)_{n \in \mathbb{N}}$ est transiente. Donc $\lim_{t \rightarrow +\infty} \mathbb{1}_{\{X_t=y\}} = 0$ p.s pour tout $y \in E$ et pour tout point de départ $x \in E$. S'il existe une probabilité invariante π , alors par convergence dominée, on aurait

$$\pi(y) = \mathbb{P}_\pi(X_t = y) = \mathbb{E}_\pi(\mathbb{1}_{\{X_t=y\}}) = \sum_{x \in E} \pi(x) \mathbb{E}_x(\mathbb{1}_{\{X_t=y\}}) \rightarrow_{t \rightarrow +\infty} 0.$$

Ceci est une contradiction.

2. Une probabilité invariante pour $(X_t)_{t \geq 0}$ n'est pas une probabilité invariante pour la chaîne induite $(Z_n)_{n \in \mathbb{N}}$. Supposons le processus irréductible (donc $a(x) > 0$ pour tout $x \in E$). On a vu qu'une probabilité invariante pour $(X_t)_{t \geq 0}$ vérifie $\pi A = 0$. Comme

$$\sum_{x \in E} \pi(x) A(x, y) = \sum_{x \neq y} a(x) \pi(x) Q(x, y) - \pi(y) a(y),$$

on voit que π est une probabilité invariante pour $(X_t)_{t \geq 0}$ ssi $a\pi$ est une mesure invariante pour $(Z_n)_{n \in \mathbb{N}}$. On voit alors que la récurrence positive de la chaîne $(Z_n)_n$ est acquise ssi $\sum_{x \in E} a(x)\pi(x) < +\infty$.

3. **Lorsque E est fini, tout processus de Markov irréductible est récurrent positif.** En effet, dans ce cas la chaîne immergée est récurrente positive et possède une unique probabilité invariante m . En posant $\pi(x) = \frac{m(x)}{a(x)}$, la remarque précédente assure que $\pi A = \pi$ et donc que π est une probabilité invariante pour le processus de Markov qui est alors récurrent positif.

Théorème 9 Soit $(X_t)_{t \geq 0}$ un processus de Markov de sauts irréductible, récurrent positif. Si π est son unique probabilité invariante et $f : E \rightarrow \mathbb{R}$ une fonction telle que $\int |f| d\pi < +\infty$, alors

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \int_0^t f(X_s) ds = \int f d\pi \text{ p.s.}$$

En particulier, le temps moyen passé en x , $\frac{1}{t} \int_0^t \mathbb{1}_{\{X_s=x\}} ds$ converge p.s vers $\pi(x)$ sous les hypothèses du théorème précédent. Concernant la convergence de la loi de X_t vers la probabilité invariante, on le résultat suivant. Contrairement au cas des chaînes de Markov, il n'y a pas besoin de condition d'apériodicité (en fait les chaînes du type $(X_{nh})_{n \in \mathbb{N}}$ sont apériodiques car $P_h(x, y) > 0$ pour tout $h > 0$ et tout $(x, y) \in E^2$).

Théorème 10 Pour un processus irréductible, récurrent positif, on a $\lim_{t \rightarrow +\infty} \mu P_t(x) = \pi(x)$ pour tout $x \in E$ et toute loi initiale μ .

3.6 Exemples

3.6.1 Processus à deux états

Supposons $E = \{0, 1\}$ et $A = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$ avec $\lambda, \mu > 0$. Si on considère l'état d'une machine à l'instant t (état 0 si elle est en réparation, état 1 si elle est en service), les délais de réparation suivent des lois exponentielles de paramètre λ et les temps de fonctionnement des loi exponentielles de paramètre μ . Sur cet exemple, on peut calculer explicitement les transitions

$P_t = \exp(tA)$ en diagonalisant la matrice A . Comme $A = PDP^{-1}$ avec $D = \begin{pmatrix} 0 & 0 \\ 0 & -(\lambda + \mu) \end{pmatrix}$, $P = \begin{pmatrix} 1 & \lambda \\ 1 & -\mu \end{pmatrix}$, $P^{-1} = \begin{pmatrix} \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \\ \frac{1}{\lambda + \mu} & -\frac{1}{\lambda + \mu} \end{pmatrix}$, il vient

$$P_t = \begin{pmatrix} \frac{\mu + \lambda e^{-(\lambda + \mu)t}}{\lambda + \mu} & \frac{\lambda - \lambda e^{-(\lambda + \mu)t}}{\lambda + \mu} \\ \frac{\mu - \mu e^{-(\lambda + \mu)t}}{\lambda + \mu} & \frac{\lambda + \mu e^{-(\lambda + \mu)t}}{\lambda + \mu} \end{pmatrix}.$$

En résolvant l'équation $\pi A = \pi$, l'unique probabilité invariante est donnée par $\pi = \left(\frac{\mu}{\lambda + \mu}, \frac{\lambda}{\lambda + \mu} \right)$.

3.6.2 Processus de naissance et de mort

On suppose $E = \mathbb{N}$ et on considère le processus de Markov de sauts dont le générateur A est défini par

$$A(n, n-1) = \beta_n > 0 \text{ si } n \geq 1, \quad A(n, n+1) = \alpha_n > 0 \text{ si } n \geq 0.$$

La matrice de transition Q de la chaîne immergée est donnée par

$$Q(n, n-1) = \frac{\beta_n}{\alpha_n + \beta_n} = q_n \text{ si } n \geq 1, \quad Q(n, n+1) = \frac{\alpha_n}{\alpha_n + \beta_n} = p_n \text{ si } n \geq 0,$$

en convenant que $\alpha_0 = 0$. La chaîne immergée $(Z_n)_{n \in \mathbb{N}}$ est donc une chaîne de naissance et de mort et $(X_t)_{t \geq 0}$ est irréductible. La récurrence de la chaîne $(Z_n)_{n \in \mathbb{N}}$ est valable ssi

$$\sum_{n=1}^{+\infty} \frac{\beta_1 \cdots \beta_n}{\alpha_1 \cdots \alpha_n} = +\infty. \quad (3.3)$$

Sous la condition (3.3) le processus de Markov est bien défini puisque les trajectoires n'explosent pas en temps fini. Sous la condition (3.3), le processus est récurrent. On peut chercher l'expression de la probabilité invariante et la condition de récurrence positive en résolvant l'équation $\pi A = 0$. Mais on peut aussi regarder les mesures invariantes pour $(Z_n)_{n \in \mathbb{N}}$. Dans le Chapitre 1, nous avons vu qu'une mesure m est invariante pour $(Z_n)_{n \in \mathbb{N}}$ ssi

$$m(n) = m(0) \frac{p_0 \cdots p_{n-1}}{q_1 \cdots q_n} = \frac{\alpha_1 \cdots \alpha_{n-1}}{\beta_1 \cdots \beta_n} (\alpha_n + \beta_n), \quad n \geq 1$$

Toute mesure invariante π pour le processus s'écrit donc

$$\pi(n) = \frac{m(n)}{\alpha_n + \beta_n} \text{ si } n \geq 1, \quad \pi(0) = \frac{m(0)}{\alpha_0}.$$

On voit alors que π est une probabilité ssi

$$S = 1 + \sum_{n \geq 1} \frac{\alpha_0 \cdots \alpha_{n-1}}{\beta_1 \cdots \beta_n} < +\infty. \quad (3.4)$$

Dans ce cas

$$\pi(0) = \frac{1}{S}, \quad \pi(n) = \frac{1}{S} \frac{\alpha_0 \cdots \alpha_{n-1}}{\beta_1 \cdots \beta_n}, n \geq 1.$$

Exemple. Si $\beta_n = \beta n$ et $\alpha_n = \alpha n$ pour $n \geq 1$, on a récurrence ssi $\alpha \leq \beta$ et récurrence positive ssi $\alpha < \beta$.

Remarque. Le processus $(X_t)_{t \geq 0}$ est récurrent positif ssi les deux conditions (3.3) et (3.4) sont vérifiées. Contrairement au cas des chaînes de naissance et de mort, la condition (3.4) n'entraîne pas la condition (3.3). Par exemple si $\beta_n = 3^n$ et $\alpha_n = 2 \cdot 3^n$, on a $S < +\infty$ (il existe une probabilité invariante π telle que $\pi A = 0$) et pourtant (3.3) n'est pas vérifiée. Le problème vient de la condition de non-explosion qui n'est pas vérifiée ici. La proposition suivante donne une CNS de non-explosion. Cette condition peut être déduite de la Proposition 22.

Proposition 25 *Le processus de naissance et de mort est non-explosif ssi*

$$\sum_{n \geq 1} r_n = +\infty, \quad r_n = \sum_{k=0}^n \frac{\beta_{k+1} \cdots \beta_n}{\alpha_k \cdots \alpha_n}.$$

3.6.3 Dynamique d'une épidémie

Un cas de maladie non curable. On considère une population formée de m individus. On suppose qu'à l'instant initial, un individu est atteint d'une maladie non curable. Cette maladie est contagieuse. On admet que dans un intervalle de temps $[t, t + h]$, tout individu est susceptible de contaminer quelqu'un avec probabilité $p_h = \alpha h + o(h)$ et ce indépendamment des autres individus. On note X_t le nombre de personnes infectées à l'instant t . Pour définir un processus de Markov associé, il faut définir un générateur. Si $1 \leq x \leq m - 1$ et $x + 1 \leq y \leq m$, $P_h(x, y)$ s'interprète intuitivement comme la probabilité qu'une variable de loi binomiale de paramètres $(m - x, xp_h)$ prenne la valeur $y - x$. On en déduit après quelques calculs que

$$P_h(x, y) = \begin{cases} o(h) & \text{si } y \geq x + 2, \\ \alpha x(m - x)h + o(h) & \text{si } y = x + 1 \end{cases}$$

De plus, $P_h(x, x)$ (probabilité de n'avoir aucune contamination) est

$$P_h(x, x) = (1 - xp_h)^{m-x} = 1 - \alpha x(m - x)h + o(h).$$

Soit alors T le temps nécessaire pour que toute la population soit contaminée. On a $T = T_{m-1}$ et en utilisant l'expression du générateur ainsi que la loi des délais entre les sauts, on trouve

$$\mathbb{E}_1(T) = \sum_{x=1}^{m-1} \frac{1}{\alpha x(m-x)} \sim \frac{2 \ln(m)}{\alpha m}, \quad m \rightarrow +\infty.$$

Un cas de maladie curable. La population est formée de m individus. On suppose que le générateur est donné par $A(n, n - 1) = \beta_n > 0$ pour $1 \leq n \leq m$, $A(n, n + 1) = \alpha_n > 0$ si $1 \leq n \leq m - 1$. On suppose que $A(0, 1) = 0$. Il s'agit alors d'un processus de naissance et de mort mais avec un nombre fini d'états et 0 est un point absorbant. La chaîne de Markov immergée est absorbante et d'après les résultats du premier chapitre (paragraphe 1.3), 0 est atteint au bout d'un nombre fini de pas. Donc le temps d'atteinte τ de 0 est fini p.s. Montrons également que $\mathbb{E}_{i_0}(\tau) < +\infty$, pour $1 \leq i_0 \leq m$. Pour cela introduisons $\tilde{\tau}$, le nombre de pas nécessaire pour que la chaîne immergée

atteigne 0. On rappelle que $\mathbb{E}_{i_0}(\bar{\tau}) < +\infty$ pour tout i_0 . Notons $\kappa = \max_{1 \leq i \leq m} \frac{1}{\alpha_i + \beta_i}$. On a alors

$$\begin{aligned}
\mathbb{E}_{i_0}(\tau) &= \sum_{k=1}^{+\infty} \mathbb{E}_{i_0}(\tau \mathbb{1}_{\bar{\tau}=k}) \\
&= \sum_{k=1}^{+\infty} \sum_{i_1, \dots, i_{k-1} \neq 0} \mathbb{E}_{i_0}(\tau \mathbb{1}_{Z_1=i_1, \dots, Z_{k-1}=i_{k-1}, Z_k=0}) \\
&= \sum_{k=1}^{+\infty} \sum_{i_1, \dots, i_{k-1} \neq 0} \mathbb{E}_{i_0} \left[\left(\frac{1}{\alpha_{i_0} + \beta_{i_0}} + \dots + \frac{1}{\alpha_{i_{k-1}} + \beta_{i_{k-1}}} \right) \mathbb{1}_{Z_1=i_1, \dots, Z_{k-1}=i_{k-1}, Z_k=0} \right] \\
&\leq \sum_{k=1}^{+\infty} \sum_{i_1, \dots, i_{k-1} \neq 0} k \kappa \mathbb{P}_{i_0}(Z_1 = i_1, \dots, Z_{k-1} = i_{k-1}, Z_k = 0) \\
&= \sum_{k=1}^{+\infty} k \kappa \mathbb{P}_{i_0}(\bar{\tau} = k) \\
&= \kappa \mathbb{E}_{i_0}(\bar{\tau}).
\end{aligned}$$

Ceci prouve que le temps moyen de la durée de l'épidémie est fini p.s. Donnons alors son expression. Si $1 \leq i_0 \leq m-1$, nous avons en utilisant la propriété de Markov :

$$\begin{aligned}
\mathbb{E}_{i_0}(\tau) &= \mathbb{E}_{i_0}(\tau \mathbb{1}_{Z_1=i_0+1}) + \mathbb{E}_{i_0}(\tau \mathbb{1}_{Z_1=i_0-1}) \\
&= \mathbb{E}(\tau | Z_0 = i_0, Z_1 = i_0 + 1) \mathbb{P}_{i_0}(Z_1 = i_0 + 1) + \mathbb{E}(\tau | Z_0 = i_0, Z_1 = i_0 - 1) \mathbb{P}_{i_0}(Z_1 = i_0 - 1) \\
&= \mathbb{E}_{i_0+1}(\tau) \frac{\alpha_{i_0}}{\alpha_{i_0} + \beta_{i_0}} + \mathbb{E}_{i_0-1}(\tau) \frac{\beta_{i_0}}{\alpha_{i_0} + \beta_{i_0}} + \frac{1}{\alpha_{i_0} + \beta_{i_0}}.
\end{aligned}$$

De plus, on a aussi $\mathbb{E}_m(\tau) = \mathbb{E}_{m-1}(\tau) + \frac{1}{\beta_m}$. On voit alors que les temps moyens sont solutions du système linéaire :

$$[A(i, j)]_{1 \leq i, j \leq m} \begin{pmatrix} \mathbb{E}_1(\tau) \\ \vdots \\ \mathbb{E}_m(\tau) \end{pmatrix} = -\mathbb{1}.$$

3.6.4 Processus de Galton-Watson à temps continu

On part d'un ancêtre qui au bout d'un temps exponentiel de paramètre $a > 0$ donne naissance à k enfants avec probabilité p_k , $k \in \mathbb{N}$ et meurt ensuite. Les k enfants évoluent ensuite de la même manière, indépendamment du passé et des autres enfants. On note X_t le nombre d'individus à l'instant t . On supposera $p_1 < 1$. Calculons le générateur de ce processus informellement. Nous utiliserons la propriété suivante : si U suit une loi exponentielle de paramètre a et S est une variable aléatoire positive, indépendante de U , alors

$$\mathbb{P}(U + S > t + h | U + S > t, S \leq t) = \exp(-ah).$$

autrement dit tout temps U de loi exponentielle non achevé à l'instant $t-S$ aura une durée résiduelle $U - (t - S)$ qui suivra toujours une loi exponentielle. On parle d'absence de mémoire pour la loi

exponentielle. Si $X_t = 0$ alors $X_{t+h} = 0$ pour tout $h > 0$. On a donc $A(0, y) = 0$ pour tout $y \in \mathbb{N}$. Si $X_t = x > 0$, la probabilité que $X_{t+h} = y \in \{x-1, x+1, x+2, \dots\}$ conditionnellement à toutes les valeurs du processus jusqu'au temps t se calcule en distinguant les deux cas suivants.

- On a une fin de vie avant $t+h$ et une seule. La probabilité correspondante est

$$x\mathbb{P}(U_1, \dots, U_{x-1} > h, U_x \leq h, V_1, \dots, V_{y-x+1} > h - U_x) p_{y-x+1}$$

où $U_1, \dots, U_x, V_1, \dots, V_{y-x+1}$ sont des variables aléatoires i.i.d de loi exponentielle de paramètre a . Cette probabilité s'écrit $axp_{y-x+1}h + o(h)$ et ne dépend que de x .

- On a au moins deux fins de vie avant $t+h$. Comme la probabilité $\mathbb{P}(U_1, V_1 \leq h) = o(h)$ pour deux variables indépendantes U_1, V_1 toutes deux de loi exponentielle de paramètre $a > 0$, cette probabilité est aussi un $o(h)$.

Ensuite, si $y = x$, on ne peut avoir $X_{t+h} = x$ que si aucune fin de vie n'a lieu avant $t+h$ ou si une seule fin de vie a lieu et 1 seul enfant est né ou si on est revenu à l'état x après deux fins de vies au moins. La probabilité se décompose donc sous la forme

$$\exp(-axh) + xp_1ah + o(h) = 1 + ax(p_1 - 1)h + o(h).$$

Le générateur vaut donc $A(x, y) = axp_{y-x+1}\mathbb{1}_{\{y \geq x-1, y \neq x\}} + ax(p_1 - 1)\mathbb{1}_{y=x}$. On peut voir que les délais entre les sauts suivent des lois exponentielles de paramètres $ax(1 - p_1)$.

Pour que le processus soit bien défini, il faut faire attention à l'explosion. Montrons que l'explosion n'a pas lieu dès lors que $m = \sum_{k \in \mathbb{N}} kp_k < +\infty$ (moyenne de la loi de reproduction finie). La chaîne immergée $(Z_n)_{n \in \mathbb{N}}$ a pour matrice de transition Q telle que

$$Q(0, 0) = 1, \quad Q(x, x+k) = \frac{p_{k+1}}{1-p_1}, \quad Q(x, x-1) = \frac{p_0}{1-p_1} \text{ si } x, k \geq 1.$$

Cette chaîne a la même loi que la chaîne $(Y_n)_{n \in \mathbb{N}}$ définie par $Y_0 = 1$ et

$$Y_{n+1} = (Y_n + \xi_{n+1}) \mathbb{1}_{Y_n \geq 1}, \quad n \in \mathbb{N},$$

où $(\xi_n)_{n \in \mathbb{N}}$ est une suite i.i.d telle que

$$\mathbb{P}(\xi_1 = k) = \frac{p_0 \mathbb{1}_{k=-1} + p_{k+1} \mathbb{1}_{k \geq 1}}{1-p_1}.$$

Remarquons que $\mathbb{E}(\xi_1) = \frac{m-1}{1-p_1}$. Soit $\omega \in \Omega$. S'il existe $n \in \mathbb{N}$ tel que $Y_n(\omega) = 0$ alors $Y_m(\omega) = 0$ pour tout $m \geq n$ et donc $\sum_{j \geq 0} \frac{1}{a(Y_j(\omega))} = +\infty$. Sinon, on a $\frac{Y_n(\omega)}{n}$ converge vers $\frac{m-1}{1-p_1}$ et donc il existe $\epsilon > 0$ tel que $Y_n(\omega) < \epsilon n$ à partir d'un certain rang, ce qui entraîne de nouveau que $\sum_{j \geq 0} \frac{1}{a(Y_j(\omega))} = +\infty$. On voit que la condition de non-explosion est vérifiée. De plus, si $m < 1$, on voit que $Y_n(\omega)$ est nul à partir d'un certain rang (on dit qu'il y a extinction).

Supposons ensuite $m \geq 1$ et $p_0 > 0$. Définissons la probabilité d'extinction par

$$q_i = \mathbb{P}_i(\exists t > 0 : X_{t+s} = 0, s \geq 0).$$

En utilisant les égalités entre évènements

$$\{\exists t > 0 : X_{t+s} = 0, s \geq 0\} = \cup_{n \in \mathbb{N}} \{X_{n+s} = 0, s \geq 0\},$$

on trouve que

$$q_i = \lim_{n \rightarrow +\infty} \mathbb{P}_i(X_{n+s} = 0, s \geq 0) = \lim_{n \rightarrow +\infty} P_n(i, 0).$$

Comme $t \mapsto P_t(i, 0) = \mathbb{P}_i(X_{t+s} = 0, s \geq 0)$ est croissante, $\lim_{t \rightarrow +\infty} P_t(i, 0)$ existe et coïncide évidemment avec $\lim_{n \rightarrow +\infty} P_n(i, 0)$. Comme un processus de Galton-Watson démarré avec i ancêtres se comporte comme i processus de Galton-Watson indépendants démarrés avec un seul ancêtre, on obtient au final

$$q_i = \lim_{t \rightarrow +\infty} P_t(i, 0) = q_1^i = \left(\lim_{t \rightarrow +\infty} P_t(1, 0) \right)^i.$$

Pour calculer q_1 , on utilise l'équation backward. On a

$$\begin{aligned} \frac{\partial}{\partial t} P_t(1, 0) &= \sum_{j \geq 0} A(1, j) P_t(j, 0) \\ &= \sum_{j \neq 1} a p_j P_t(1, 0)^j + a(p_1 - 1) P_t(1, 0) \\ &= u(P_t(1, 0)), \end{aligned}$$

où $u(s) = a(f(s) - s)$ et $f(s) = \sum_{k \geq 0} p_k s^k$. On notera que f est la fonction génératrice de la loi de reproduction $(p_k)_{k \in \mathbb{N}}$. On rappelle que f est convexe continu et comme $f(0) > 0$ et $f(1) = 1$. Soit $\eta > 0$ la plus petite racine de l'équation $f(s) = s$ (si $m = f'(1^-) > 1$ il existe un unique $0 < \eta < 1$ tel que $f(\eta) = \eta$, sinon $m = 1$ et $\eta = 1$). Comme $P_0(1, 0)$, on a $P_t(1, 0) = \int_0^t u(P_s(1, 0)) ds$. Comme $t \mapsto P_t(1, 0)$ est croissante, on $t \mapsto u(P_s(1, 0))$ qui est positive. Ceci entraîne que $q_1 = \lim_{t \rightarrow +\infty} P_t(1, 0) \leq \eta$. De plus u est continue. On ne peut avoir $q_1 > \eta$ sinon $u(q_1) > 0$ et donc

$$\lim_{t \rightarrow +\infty} \int_0^t u(P_s(1, 0)) ds = +\infty = \lim_{t \rightarrow +\infty} P_t(1, 0),$$

ce qui est impossible. On en déduit que $u(q_1) = 0$ et donc que $q_1 = \eta$.

Exemple de modélisation. Une molécule d'ADN composée de k nucléotides se réplique au bout d'un temps exponentiel de paramètre a . Durant la phase de réplication, la molécule est détruite avec probabilité w . Si elle n'est pas détruite, elle se réplique correctement avec probabilité q^k (q proba qu'un nucléotide soit copié correctement). On obtient donc 0 copies correctes avec probabilité $p_0 = w$, 1 copie de cette molécule avec probabilité $(1-w)(1-p^k)$ et 2 copies avec probabilité $(1-w)p^k$. Si $(X_t)_{t \geq 0}$ est le processus de Galton-Watson associé, X_t désigne le nombre de molécules identiques à l'ancêtre au temps t .

3.7 Exercices

Exercice 14 Soit $(X_t)_{t \geq 0}$ un processus de Markov de sauts à trois états dont le générateur est donné par

$$A = \begin{pmatrix} -\lambda & \lambda & 0 \\ 0 & -\lambda & \lambda \\ \lambda & 0 & -\lambda \end{pmatrix}.$$

Donner une expression du maximum de vraisemblance λ .

Exercice 15 On considère X un processus de Markov à sauts à 3 états dont le générateur est donné par

$$A = \begin{pmatrix} -\lambda & \lambda & 0 \\ 0 & -\mu & \mu \\ \nu & 0 & -\nu \end{pmatrix},$$

où λ, μ et ν sont des réels strictement positifs. On conviendra que l'espace d'états est $E = \{0, 1, 2\}$.

1. Préciser la matrice de transition de la chaîne immergée.
2. En supposant que $X_0 = 0$, décrire la dynamique du processus.
3. Justifier la récurrence positive du processus X et donner sa probabilité invariante.
4. Si $x \in E$, on note

$$R_x = \inf \{t \geq T_1 : X_t = x\},$$

le temps de retour au point x . Ici T_1 désigne le délai avant le premier saut. Montrer que $\mathbb{E}_x(R_x)$ ne dépend pas de x et est égal à une valeur que l'on précisera. Vérifier alors la formule donnée en cours

$$\pi(x) = \frac{1}{a(x)\mathbb{E}_x(R_x)}, \quad x \in E,$$

où π est la probabilité invariante déterminée à la question précédente et $a(x) = -A(x, x)$ pour tout $x \in E$.