
CHAINES DE MARKOV

PROJET TUTORE

Auteurs :

Garance QUERIC

Mélissa EVEILLARD

Encadrant :

Mr Antoine MARCHINA

4 décembre 2021

Table des matières

Introduction	2
1 Chaînes de Markov	3
1.1 Définitions et propriétés élémentaires	3
1.2 Caractérisation d'une chaîne de Markov	4
1.3 Propriétés	5
1.4 Classification des états	6
1.5 Lois de probabilités stationnaires	7
2 Théorème de convergence	7
2.1 Théorème ergodique	7
3 Estimation de la matrice de transition	8
3.1 Estimateur du maximum de vraisemblance	8
3.2 Propriétés de l'estimateur	9
4 Statistiques des chaînes de Markov	9
4.1 Simulation de chaînes de Markov	10
4.1.1 Simulation	10
4.1.1.1 Exemple avec une chaîne de Markov à deux états	10
4.1.2 Illustration de la loi stationnaire	12
4.1.2.1 Exemple avec un algorithme implémenté	12
4.1.3 Simulation de la matrice de transition	14
5 Applications des chaînes de Markov sur des données temporelles	15
5.1 Description	15
5.2 Analyse descriptive	15
5.3 Modélisation	17
5.4 Evaluation de la performance	18
5.5 Prévision	19
6 Conclusion	21
7 Fonctions R	22

Introduction

Le but de ce projet tutoré est d'étudier un modèle aléatoire basé sur les chaînes de Markov. Les chaînes de Markov sont des séquences de variables aléatoires qui ne sont ni indépendantes, ni identiquement distribuées. La principale caractéristique de cet objet est que le passé n'apporte pas plus d'information que le présent dans la prédiction du futur.

Les chaînes de Markov ont été décrites pour la première fois par Andrei Markov en 1906, date à laquelle il a publié ses premiers résultats concernant les chaînes de Markov à espace d'état fini. Une généralisation à un espace d'états infini dénombrable a été publiée par Kolmogorov en 1936. On se sert des chaînes de Markov pour modéliser l'évolution dynamique de systèmes aléatoires : on peut voir les réalisations d'une chaîne de Markov comme des trajectoires au cours du temps. Elles sont très utilisées dans divers domaines : Andrei Markov a eu recours à des chaînes de Markov pour analyser les successions de lettres dans l'alphabet russe. Aujourd'hui, on les retrouve dans l'analyse de l'ADN ou encore pour définir la popularité d'une page Web.

Au cours de ce projet nous allons, définir ce qu'est une chaîne de Markov, et regarder ses différentes propriétés. Ensuite, on établira les théorèmes de convergences et on simulera sur R des chaînes de Markov afin d'étudier leur comportement.

Enfin, on verra comment, à partir des chaînes de Markov, on peut étudier et modéliser une série temporelle, en utilisant pour cela des données concernant les éruptions d'un geyser.

1 Chaînes de Markov

Dans cette première partie nous allons définir formellement les chaînes de Markov et proposer les théorèmes et notions dont nous allons avoir besoin par la suite.

1.1 Définitions et propriétés élémentaires

Définition 1.1. Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires à valeurs dans un ensemble E , supposé au plus dénombrable ; on appelle E espace d'états de la chaîne. Cette suite est une chaîne de Markov, si pour tout $n \geq 0$ et toute suite $(x_0, \dots, x_{n-1}, x, y)$ d'éléments de E , telle que $\mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) > 0$, on a l'égalité suivante :

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

Remarque. Dans la suite on considérera les probabilités sur E comme un vecteur ligne.

Remarque. Cette propriété traduit le fait que la chaîne de Markov est définie telle que l'information utile pour la prédiction du futur ne dépend que de l'état actuel et ne dépend pas de l'état précédent. Ce phénomène est appelé sans mémoire.

Définition 1.2. Une chaîne de Markov est homogène si la probabilité de passage de l'état x à l'état y ne dépend pas de n , c'est-à-dire du temps. Ainsi, on note pour tout $n \geq 0$,

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

la probabilité de passer de l'état $x \in E$ à l'état $y \in E$. Ces probabilités sont appelées probabilités de transition, et celles-ci sont indépendantes du temps.

Dans la suite, nous nous intéresserons uniquement au cas de chaînes de Markov homogènes d'espace d'états E .

Définition 1.3. Soit (X_n) ($n \geq 0$) une chaîne de Markov. On appelle distribution initiale le vecteur $\mu^{(0)}$ donné par $\mu^{(0)} = (\mu_x^{(0)})_{x \in E}$ où $\mu_x^{(0)} = \mathbb{P}(X_0 = x)$ pour tout $x \in E$.

On note de la même manière, $\forall n \geq 0$, $\mu^{(n)} = (\mu_x^{(n)})_{x \in E}$ les probabilités d'états c'est-à-dire : $\forall x \in E$, $\mu_x^{(n)} = \mathbb{P}(X_n = x)$.

Définition 1.4. La matrice P définie par $P = (p_{x,y})_{x,y \in E}$ est la matrice de transition de la chaîne de Markov.

Définition 1.5. Une matrice de $M_n(\mathbb{R})$ est une matrice stochastique si tous ses coefficients sont positifs et si la somme de chaque ligne est égale à 1.

Proposition 1.6. La matrice de transition d'une chaîne de Markov est une matrice stochastique.

Démonstration.

1. Soient $x, y \in E$. $p_{x,y} \geq 0$ car $p_{x,y}$ est une probabilité.
2. Soit $x \in E$. Pour tout $n \geq 0$:

$$\begin{aligned} \sum_{y \in E} p_{x,y} &= \sum_{y \in E} \mathbb{P}(X_{n+1} = y | X_n = x) \\ &= \mathbb{P}(X_{n+1} \in \bigcup_{y \in E} \{y\} | X_n = x) \\ &= \mathbb{P}(X_{n+1} \in E | X_n = x) \\ &= 1. \end{aligned}$$

1.2 Caractérisation d'une chaîne de Markov

Théorème 1.7. Soit (X_n) une chaîne de Markov de distribution initiale $\mu^{(0)}$ et de matrice de transition P . On a pour tout n

$$\mu^{(n)} = \mu^{(0)} P^n \quad (1)$$

Remarque. Ce théorème permet d'affirmer qu'une chaîne de Markov est complètement définie si l'on connaît sa matrice de transition P et sa distribution initiale $\mu^{(0)}$.

Démonstration. Montrons cette égalité par récurrence. On range arbitrairement les éléments de l'espace d'états E et on les note $e_k, k \in \mathbb{N}$.

Initialisation : Prenons le cas $n = 1$. On a :

$$\begin{aligned} \mu_{e_k}^{(1)} &= \mathbb{P}(X_1 = e_k) \\ &= \sum_{l \in \mathbb{N}} \mathbb{P}(X_0 = e_l, X_1 = e_k) \quad (\text{par la formule des probabilités totales}) \\ &= \sum_{l \in \mathbb{N}} \mathbb{P}(X_0 = e_l) \mathbb{P}(X_1 = e_k | X_0 = e_l) \quad (\text{définition probabilité conditionnelle}) \\ &= \sum_{l \in \mathbb{N}} \mu_{e_l}^{(0)} p_{e_l, e_k} = (\mu^{(0)} P)_{e_k} \end{aligned}$$

où $(\mu^{(0)} P)_{e_k}$ est le $k^{\text{ième}}$ élément du vecteur ligne $\mu^{(0)} P$. D'où $\mu^{(1)} = \mu^{(0)} P$.

Hérédité : Soit $n \in \mathbb{N}$. On suppose que l'égalité (1) est vraie en n . En $n + 1$, on a

$$\begin{aligned} \mu_{e_k}^{(n+1)} &= \mathbb{P}(X_{n+1} = e_k) = \sum_{l \in \mathbb{N}} \mathbb{P}(X_n = e_l, X_{n+1} = e_k) \\ &= \sum_{l \in \mathbb{N}} \mathbb{P}(X_n = e_l) \mathbb{P}(X_{n+1} = e_k | X_n = e_l) \\ &= \sum_{l \in \mathbb{N}} \mu_{e_l}^{(n)} p_{e_l, e_k} = (\mu^{(n)} P)_{e_k} \end{aligned}$$

Ainsi, $\mu^{(n+1)} = \mu^{(n)} P$. Comme on a $\mu^{(n)} = \mu^{(0)} P^n$ par hypothèse,

$$\mu^{(n+1)} = \mu^{(n)} P = \mu^{(0)} P^n P = \mu^{(0)} P^{n+1}.$$

On a bien montré par récurrence le théorème ci-dessus.

Théorème 1.8. (Condition nécessaire et suffisante)

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires. $(X_n)_{n \geq 0}$ est une chaîne de Markov de matrice de transition $P = (p_{x,y})_{x,y \in E}$ et de loi initiale $\mu^{(0)}$ si et seulement si pour tout $n \in \mathbb{N}$ la loi du vecteur aléatoire (X_0, \dots, X_n) est donnée par :

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mu_{x_0}^{(0)} p_{x_0, x_1} \dots p_{x_{n-1}, x_n}$$

Démonstration. Pour simplifier les notations on notera $X_{0:n}$ au lieu de (X_0, \dots, X_n) et $x_{0:n}$ au lieu de (x_0, \dots, x_n) .

\Rightarrow Supposons que $\{X_n, n \geq 0\}$ soit une chaîne de Markov de matrice de transition P et de loi initiale $\mu^{(0)}$. Soit $n \in \mathbb{N}$. On a :

$$\begin{aligned} \mathbb{P}(X_{0:n} = x_{0:n}) &= \mathbb{P}(X_n = x_n | X_{0:n-1} = x_{0:n-1}) \mathbb{P}(X_{0:n-1} = x_{0:n-1}) \\ &= \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \mathbb{P}(X_{0:n-1} = x_{0:n-1}) \quad \text{par définition d'une chaîne de Markov} \\ &= p_{x_{n-1}, x_n} \mathbb{P}(X_{0:n-1} = x_{0:n-1}) \\ &= p_{x_{n-1}, x_n} \dots p_{x_0, x_1} \mathbb{P}(X_0 = x_0) \quad \text{par itérations} \\ &= p_{x_{n-1}, x_n} \dots p_{x_0, x_1} \mu_{x_0}^{(0)} \end{aligned}$$

\Leftarrow) Supposons maintenant que pour tout $n \in \mathbb{N}$ la loi de (X_0, \dots, X_n) est donnée par :

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mu_{x_0}^{(0)} p_{x_0, x_1} \dots p_{x_{n-1}, x_n}$$

.

Soient $k \in \mathbb{N}$ et $(x_0, \dots, x_{k-1}, x, y) \in \mathbb{E}^{k+2}$.

$$\begin{aligned} \mathbb{P}(X_{k+1} = y | X_{0:k} = (x_0, \dots, x_k)) &= \frac{\mathbb{P}(X_{0:k+1} = (x_0, \dots, x_k, y))}{\mathbb{P}(X_{0:k} = (x_0, \dots, x_k))} \\ &= \frac{\mu_{x_0}^{(0)} p_{x_0, x_1} \dots p_{x_{k-1}, x} p_{x, y}}{\mu_{x_0}^{(0)} p_{x_0, x_1} \dots p_{x_{k-1}, x}} \\ &= p_{x, y} \\ &= \mathbb{P}(X_{k+1} = y | X_k = x) \end{aligned}$$

.

Donc, $\{X_n, n \geq 0\}$ est une chaîne de Markov homogène de matrice de transition P . L'hypothèse nous donne aussi avec $n = 0$ et $x \in \mathbb{E}$: $\mathbb{P}(X_0 = x) = \mu_x^{(0)}$. Donc, $\{X_n, n \geq 0\}$ est une chaîne de Markov de matrice de transition P et de loi initiale $\mu^{(0)}$.

1.3 Propriétés

Définition 1.9. Pour $n \geq 0$ et $x, y \in E$, on désigne par $p_{x,y}^{(n)}$ la probabilité, partant de l'état x en l'instant 0, d'être à l'état y en l'instant n . C'est-à-dire

$$p_{x,y}^{(n)} = \mathbb{P}(X_n = y | X_0 = x)$$

On définit également la matrice $P^{(n)} := (p_{x,y}^{(n)})_{(x,y) \in E^2}$.

Théorème 1.10. (Relation de Chapman-Kolmogorov) Pour tout $n \geq 0$, la matrice de transition en n étapes est égale à la puissance $n^{\text{ième}}$ de la matrice de transition en une étape :

$$P^{(n)} = P^n$$

Démonstration. Soit $\mu^{(0)}$ la distribution initiale d'une chaîne de Markov et $k \in \mathbb{N}$ (en reprenant les notations précédentes). On a :

$$\begin{aligned} (\mu^{(0)} P^{(n)})_{e_k} &= \sum_{l \in \mathbb{N}} \mu_{e_l}^{(0)} p_{e_l, e_k}^{(n)} \\ &= \sum_{l \in \mathbb{N}} \mathbb{P}(X_0 = e_l) \mathbb{P}(X_n = e_k | X_0 = e_l) \\ &= \mathbb{P}(X_n = e_k) \sum_{l \in \mathbb{N}} \mathbb{P}(X_0 = e_l | X_n = e_k) \\ &= \mathbb{P}(X_n = e_k) \\ &= (\mu^{(n)})_{e_k} \\ &= (\mu^{(0)} P^n)_{e_k} \quad (\text{d'après le théorème (1)}) \end{aligned}$$

Cette relation est vraie pour toute loi initiale $\mu^{(0)}$ donc, $P^{(n)} = P^n$.

Proposition 1.11. $\forall p \in \mathbb{N}, \forall n \in \mathbb{N}, \quad \mathbb{P}(X_{n+p} = y | X_p = x) = p_{x,y}^{(n)}$

Il en résulte que $\{Y_n, n \geq 0\}$ dont les variables aléatoires Y_1, Y_2, \dots sont données par $Y_n = X_{n+p}$, est encore une chaîne de Markov de matrice de transition P et de loi initiale $\mu^{(p)}$.

1.4 Classification des états

Définition 1.12. On dit que l'état y est **accessible** à partir de l'état x s'il existe un entier $n \geq 0$ tel que $p_{x,y}^{(n)} > 0$. On écrit $x \rightarrow y$.

Définition 1.13. On dit que deux états x et y **communiquent** entre eux et on écrit $x \leftrightarrow y$ si on a à la fois : $x \rightarrow y$ et $y \rightarrow x$.

Définition 1.14. Une chaîne de Markov (X_n) ($n \geq 0$), avec un espace des états E et une matrice de transition P est **irréductible** si pour tout $x, y \in E$, on a que $x \leftrightarrow y$. (Tous les états communiquent entre eux)

Pour tout état y , désignons par T_y l'instant du premier retour en y . Autrement dit,

$$T_y = \inf \{n \geq 1, X_n = y\}$$

Définition 1.15. On dit que l'état y est **récurrent** si

$$\mathbb{P}(T_y < +\infty | X_0 = y) = 1$$

C'est à dire la probabilité d'un éventuel retour à l'état y sachant que la chaîne a commencé à l'état y vaut 1. Sinon, lorsque $\mathbb{P}(T_y < +\infty | X_0 = y) < 1$, l'état y est transitoire.

Définition 1.16. L'état y est **récurrent positif** si

$$\mathbb{E}[T_y | X_0 = y] < +\infty$$

C'est-à-dire si l'espérance du temps de premier retour en y , partant de y , est finie.

Définition 1.17 Pour tout couple (x, y) d'états et tout $n \geq 0$, on pose :

$$f_{x,y}^{(n)} := \mathbb{P}(T_y = n | X_0 = x)$$

Ainsi, $f_{x,y}^{(n)}$ ($n \geq 1$) est la probabilité pour que le processus, partant de l'état x , atteigne l'état y , (pour la première fois), à l'instant n . Pour tout couple d'états (x, y) avec $x \neq y$, on pose, par convention, $f_{x,y}^{(0)} = 0$.

Définition 1.18. La probabilité $f_{x,y}$ de passer au moins une fois en y en partant de x (ou de retourner au moins une fois en x si $x = y$) vaut

$$f_{x,y} = \mathbb{P}(T_y < +\infty | X_0 = x) = \sum_{n \geq 1} f_{x,y}^{(n)}.$$

Remarque. On peut donc affirmer qu'un état x est récurrent si $f_{x,x} = 1$ et transitoire si $f_{x,x} < 1$.

Proposition 1.19. Pour $n \in \mathbb{N}$, $p_{x,y}^{(n)} = \sum_{k=0}^n p_{yy}^{(n-k)} f_{x,y}^{(k)}$.

En effet, passer de x à y en n étapes est équivalent à passer de x à y pour la première fois en k étapes avec $0 \leq k \leq n$ (terme $f_{x,y}^{(k)}$) puis de passer de y à y en $(n - k)$ étapes (terme $p_{yy}^{(n-k)}$).

Théorème 1.20. L'état y est récurrent si et seulement si $\sum_{n \geq 0} p_{y,y}^{(n)} = +\infty$. L'état y est transitoire si et seulement si $\sum_{n \geq 0} p_{y,y}^{(n)} < +\infty$

Remarque. Intuitivement, si l'on note $N_y := \sum_{n \geq 0} \mathbb{1}_{\{X_n = y\}}$ le nombre de passages dans l'état y après l'instant 0, alors le nombre moyen $\mathbb{E}[N_y | X_0 = y] = \sum_{n \geq 0} p_{y,y}^{(n)}$ est infini si et seulement si y est récurrent :

$$\mathbb{E}[N_y | X_0 = y] = \mathbb{E}[\sum_{n \geq 0} \mathbb{1}_{X_n = y} | X_0 = y] = \sum_{n \geq 0} \mathbb{E}[\mathbb{1}_{X_n = y} | X_0 = y] = \sum_{n \geq 0} \mathbb{P}(X_n = y | X_0 = y) = \sum_{n \geq 0} p_{yy}^{(n)}$$

1.5 Lois de probabilités stationnaires

Définition 1.21. Soit $\pi = (\pi_x)_{x \in E}$, une loi de probabilité sur l'espace d'états E d'une chaîne de Markov $X = (X_n)_{n \geq 0}$ de matrice de transition $P = (p_{x,y})_{(x,y) \in E^2}$. Elle est dite stationnaire si

$$\pi = \pi P$$

ou de manière équivalente si pour tout $y \in E$, on a

$$\pi_y = \sum_{x \in E} \pi_x p_{x,y}$$

Proposition 1.22. Si la loi initiale de la chaîne de Markov $(X_n)_{n \geq 0}$ (ie la loi de X_0) est une loi de probabilité stationnaire π , alors pour tout $n \geq 0$, la loi de X_n est encore π .

Démonstration. On procède par récurrence.

Initialisation Prenons le cas $n = 1$. On a $\mu^{(1)} = \pi P = \pi$. La loi de X_1 est bien π .

Hérédité On suppose pour n fixé que la loi de X_n est π . La loi de X_{n+1} est : $\mu^{(n+1)} = \pi P^{n+1} = \pi P P^n = \pi P^n = \pi$.

On a bien montré la proposition ci-dessus.

Remarque. Plus généralement, la chaîne de Markov est un processus stationnaire si et seulement si sa loi initiale est stationnaire.

Théorème 1.23. Si une chaîne de Markov est irréductible alors il existe au plus une probabilité stationnaire. Il y a de plus équivalence entre :

- i) il existe une unique probabilité stationnaire ;
- ii) il existe un état récurrent positif

Donc dans ce cas, $\pi_x = \frac{1}{\mathbb{E}[T_x | X_0 = x]}$.

Proposition 1.24. Si une chaîne de Markov est irréductible et si son espace d'états E est fini, alors tous ses états sont récurrents positifs. Il existe donc une unique loi de probabilité stationnaire.

2 Théorème de convergence

2.1 Théorème ergodique

Théorème 2.1. On considère une chaîne de Markov irréductible d'espace d'états fini. On note π son unique loi de probabilité invariante. On a alors pour toute fonction f sur E

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \sum_{x \in E} f(x) \pi(x) \quad p.s$$

Remarque. Le théorème central limite dans le cas des chaînes de Markov s'écrit ainsi : Soit $\{X_n, n \geq 0\}$ une chaîne de Markov irréductible et récurrente positive de matrice de transition P et de mesure invariante π . Soit g une fonction définie sur E^2 à valeurs dans \mathbb{R} telle que $(\pi, Pg^2) = \sum_{x,y \in E} \pi_x p_{x,y} g^2(x,y) < \infty$.

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n g(X_{k-1}, X_k) - Pg(X_{k-1}) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

sous $\mathbb{P}(\cdot | X_0 = x)$. Avec $\sigma^2 = (\pi, Pg^2) - (\pi, (Pg)^2)$.

3 Estimation de la matrice de transition

3.1 Estimateur du maximum de vraisemblance

On considère une chaîne de Markov $X_n, n \geq 0$ sur l'espace d'états E de matrice de transition $P = (p_{x,y})_{x,y \in E}$. On suppose que la chaîne est irréductible et on note π , sa loi stationnaire et $\mu^{(0)}$, sa distribution initiale. Supposons que l'on dispose d'une réalisation (x_0, \dots, x_n) du vecteur aléatoire (X_0, \dots, X_n) . Ne connaissant pas la matrice de transition P , on cherche à l'estimer par la méthode du maximum de vraisemblance. Pour tous $(x, y) \in E^2$, nous noterons

$$N_n^x = \sum_{p=0}^{n-1} \mathbb{1}_{\{X_p=x\}} \quad \text{et} \quad N_n^{xy} = \sum_{p=0}^{n-1} \mathbb{1}_{\{X_p=x, X_{p+1}=y\}}$$

Lemme 3.1. (Maximum de vraisemblance) L'estimateur du maximum de vraisemblance $\hat{P} = (\hat{p}_{x,y})_{x,y \in E}$ de P est donné pour tous $x, y \in E$ par

$$\hat{p}_{x,y} = \begin{cases} \frac{N_n^{xy}}{N_n^x} & \text{si } N_n^x \geq 1 \\ \frac{1}{\#E} & \text{sinon.} \end{cases}$$

Démonstration.

$$\begin{aligned} L(P) &= \prod_{x,y \in E} (\mu_x^{(0)})^{\mathbb{1}_{\{X_0=x\}}} p_{x,y}^{\mathbb{1}_{\{X_1=x, X_2=y\}}} \dots p_{x,y}^{\mathbb{1}_{\{X_{n-1}=x, X_n=y\}}} \\ &= \prod_{x,y \in E} (\mu_x^{(0)})^{\mathbb{1}_{\{X_0=x\}}} p_{x,y}^{N_n^{xy}} \end{aligned}$$

La log-vraisemblance s'écrit donc :

$$l(P) = \log(L(P)) = \sum_{x \in E} \mathbb{1}_{\{X_0=x\}} \log(\mu_x^{(0)}) + \sum_{x,y \in E} N_n^{xy} \log(p_{x,y})$$

On souhaite maximiser $\log(L)$ sous la contrainte $\sum_{y \in E} p_{x,y} = 1$ pour tout $x \in E$. On reprend la notation qu'on a déjà utilisé plus haut : $E = \{e_k, k \in \mathbb{N}\}$. On remarque que pour tout $k \in \mathbb{N}$, $p_{e_k, e_1} = 1 - \sum_{l \geq 2} p_{e_k, e_l}$. Donc, $\sum_{k,l \in \mathbb{N}} N_n^{e_k e_l} \log(p_{e_k, e_l})$ peut se réécrire $\sum_{k \in \mathbb{N}} N_n^{e_k e_1} \log(p_{e_k, e_1}) + \sum_{\{k,l\} \in \mathbb{N} \times \mathbb{N} \setminus \{1\}} N_n^{e_k e_l} \log(p_{e_k, e_l})$. On a donc, pour tout k , pour tout $l \neq 1$:

$$\begin{aligned} \frac{\partial l(P)}{\partial p_{e_k, e_l}} &= 0 \\ \Leftrightarrow \frac{-N_n^{e_k e_1}}{1 - \sum_{l \geq 2} p_{e_k, e_l}} + \frac{N_n^{e_k e_l}}{p_{e_k, e_l}} &= 0 \\ \Leftrightarrow \frac{-N_n^{e_k e_1}}{p_{e_k, e_1}} + \frac{N_n^{e_k e_l}}{p_{e_k, e_l}} &= 0 \\ \Leftrightarrow \frac{N_n^{e_k e_1}}{p_{e_k, e_1}} &= \frac{N_n^{e_k e_l}}{p_{e_k, e_l}} \\ \Leftrightarrow \widehat{p_{e_k, e_l}} &= N_n^{e_k e_l} \times \frac{\widehat{p_{e_k, e_1}}}{N_n^{e_k e_1}} \end{aligned}$$

Le terme $\widehat{\frac{p_{e_k, e_1}}{N_n^{e_k e_1}}}$ ne dépend pas de e_l . Donc, en utilisant à nouveau les notations simplifiées, $\widehat{p_{x,y}}$ est proportionnelle à N_n^{xy} . Le coefficient dépendra uniquement de x , notons le K_x . En reprenant la condition $\sum_y \widehat{p_{x,y}} = 1$, on a du coup que :

$$\sum_y K_x N_n^{xy} = 1$$

$$K_x \sum_y \sum_{p=0}^{n-1} \mathbb{1}_{\{X_p=x, X_{p+1}=y\}} = 1$$

Donc, K_x vaut l'inverse du nombre de passage de x à y pour tout y . Donc K_x est le nombre de passage en x de la chaîne. C'est à dire $K_x = \frac{1}{N_n^x}$. On obtient donc que $\hat{p}_{x,y} = \frac{N_n^{xy}}{N_n^x}$ si $N_n^x > 0$. Si $N_n^x = 0$, on n'a pas d'informations supplémentaires sur $\hat{p}_{x,y}$. On fait le choix suivant : si $N_n^x = 0$ alors $\hat{p}_{x,y} = \frac{1}{\#E}$ pour tout $y \in E$.

Remarque. L'estimateur \hat{P}_n est intuitif puisqu'il estime $p_{x,y}$ par la proportion de transitions de l'état x vers l'état y parmi le nombre de fois que la chaîne passe par l'état x .

3.2 Propriétés de l'estimateur

Théorème 3.2. (Consistance et normalité asymptotique) Pour tout $x \in E$, sachant $X_0 = x$, l'estimateur \hat{P}_n est convergent et asymptotiquement normal : pour tous $x, y \in E$,

$$\hat{p}_{x,y} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} p_{x,y}$$

Ce qui garantit la convergence en probabilité. De plus,

$$\sqrt{n\pi_x}(\hat{p}_{x,y} - p_{x,y}) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, p_{x,y}(1 - p_{x,y}))$$

Démonstration. D'après le lemme ergodique, on a pour tout $x \in E$

$$\frac{1}{n} N_n^x = \frac{1}{n} \sum_{p=0}^{n-1} \mathbb{1}_{\{X_p=x\}} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \pi_x$$

On remarque que $(X_n, X_{n+1})_{n \geq 0}$ est encore une chaîne de Markov. Ainsi, pour tous $x, y \in E$

$$\frac{1}{n} N_n^{xy} = \frac{1}{n} \sum_{p=0}^{n-1} \mathbb{1}_{\{X_p=x, X_{p+1}=y\}} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \sum_{a,b \in E} \pi_a p_{a,b} \mathbb{1}_{\{a=x, b=y\}} = \pi_x p_{x,y}$$

donc,

$$\hat{p}_{x,y} = \frac{N_n^{xy}}{N_n^x} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \frac{\pi_x p_{x,y}}{\pi_x} = p_{x,y}$$

4 Statistiques des chaînes de Markov

Après avoir établi les différents théorèmes et propriétés pour les chaînes de Markov, nous allons, à présent, voir comment simuler des chaînes de Markov puis comment estimer des matrices de transitions. Par la suite nous vérifierons numériquement les théorèmes de convergence.

4.1 Simulation de chaînes de Markov

4.1.1 Simulation

Soit $X_n, n > 0$ une chaîne de Markov irréductible à valeur dans E discret de loi initiale μ et de matrice de transition P .

Nous avons défini plus en amont qu'il n'y a pas d'indépendance entre les X_n , il faut donc en tenir compte lors de la simulation. Or, nous connaissons la loi de X_0 (la loi initiale μ) mais on ne connaît pas la loi de X_n pour $n \geq 1$. En revanche, on connaît la loi de x_n sachant X_{n-1} via la matrice de transition P .

Ainsi pour simuler (X_0, \dots, X_p) avec $p > 1$ on procède de la façon suivante :

- On commence par simuler $X_0 = x_0$, de loi discrète $\mu^{(0)}$.
- Pour $n = 1, \dots, p$, on utilise, la réalisation x_{n-1} obtenue à l'itération précédente, et la loi de $X_n = x_n$ qui est la ligne de P donnant les probabilités de transition entre l'état x_{n-1} et x_n .

4.1.1.1 Exemple avec une chaîne de Markov à deux états

Pour cette simulation nous allons considérer une chaîne de Markov à deux états. Nous avons implémenté sous R la fonction `fct_CM` qui simule une chaîne de Markov à deux états 0 ou 1, de longueur p , de loi initiale μ et de matrice de transition P . Ces 3 derniers éléments sont les arguments pris en entrée par la fonction, puis celle-ci exécute les instructions suivantes :

- On crée un vecteur de taille p dans lequel on va stocker les x_1, \dots, x_p simulés ;
- On tire une variable aléatoire U qui suit une loi uniforme $(0, 1)$: si $U \leq \mu(0)$ alors X_0 prend la valeur 0, sinon 1 ;
- Pour tout $i = 2, \dots, p$, on tire de nouveau U selon une loi $U \sim (0, 1)$; si $U \leq P(x_{i-1}, 0)$ alors x_i prend la valeur 0, sinon x_i prend la valeur 1 ;
- La fonction retourne en sortie le vecteur x créé.

On utilise donc cette fonction afin de longueur $p = 100$, de loi initiale $\mu = (0.5, 0.5)$ pour différents choix de matrices de transition P qui sont :

$$P_1 = \begin{pmatrix} 0.1 & 0.9 \\ 0.8 & 0.2 \end{pmatrix}, P_2 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, P_3 = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}.$$

On représente la trajectoire des différentes chaînes créées à l'aide d'un nuage de points $(i, X_i)_{1 \leq i \leq p}$. On appelle changement de régime, la transition d'un état à un autre ; afin de mieux illustrer ces changements on relie les points (i, X_i) .

Les figures 1, 2 et 3 donnent les trajectoires obtenues pour les trois cas. On remarque que plus les coefficients diagonaux de la matrice de transition P sont grands, moins les changements d'états sont fréquents, cela s'explique par le fait que les coefficients diagonaux de la matrice sont les probabilités de rester au même état.

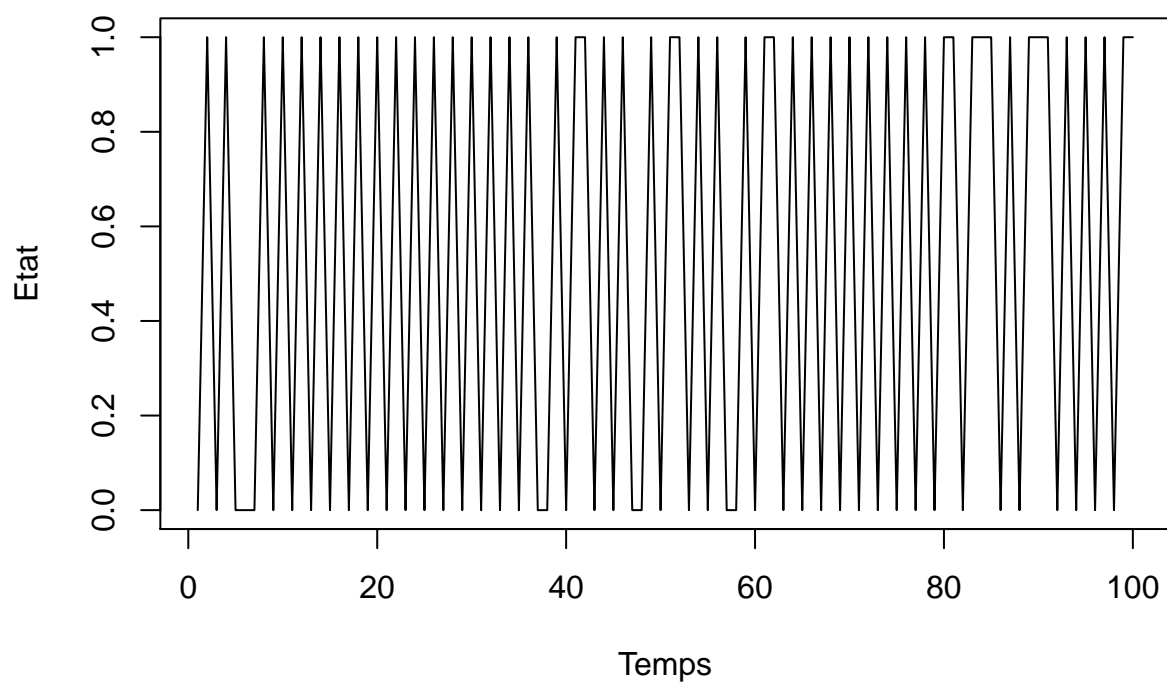


FIGURE 1 – Trajectoire de x pour $P = P_1$

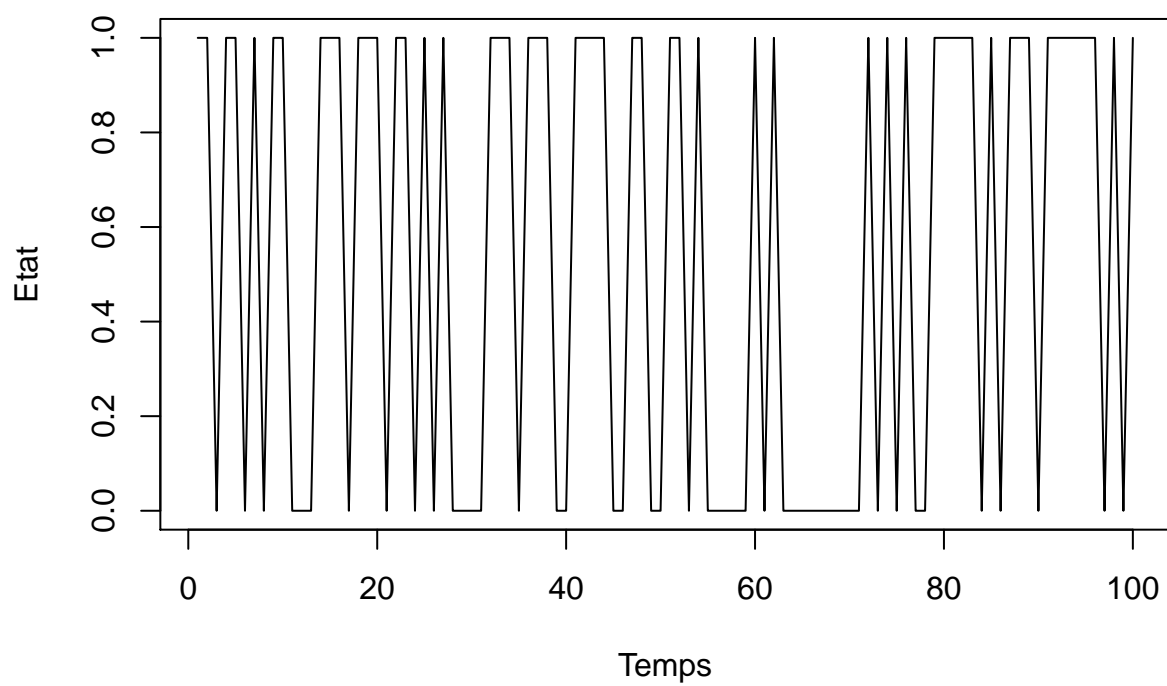


FIGURE 2 – Trajectoire de x pour $P = P_2$

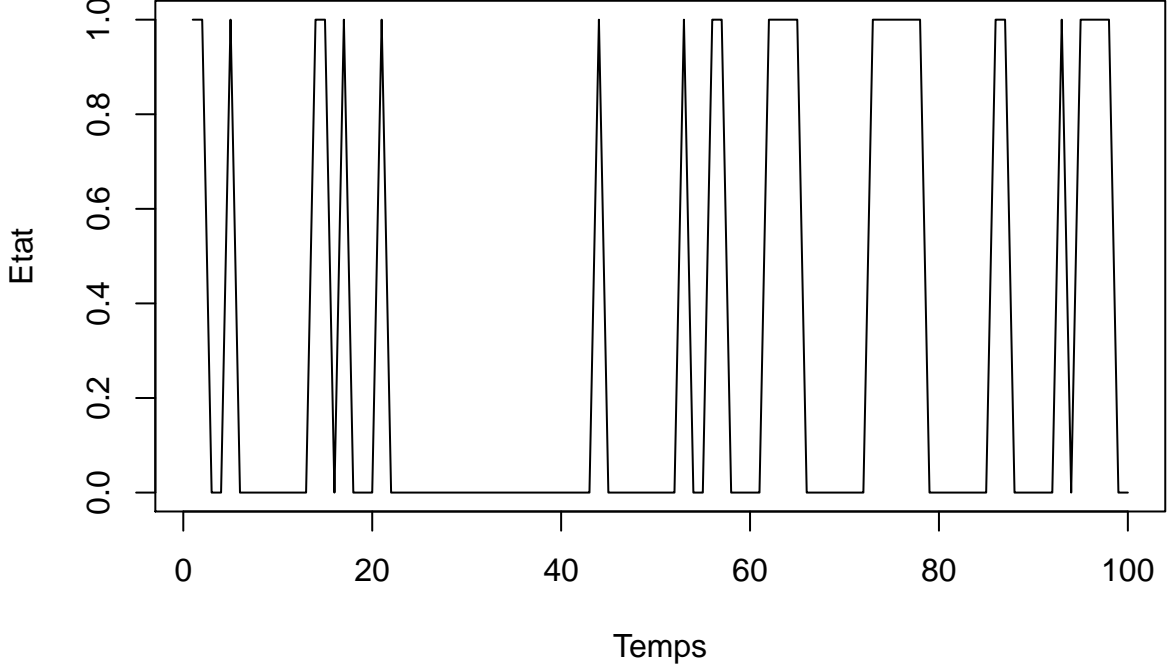


FIGURE 3 – Trajectoire de x pour $P = P_3$

4.1.2 Illustration de la loi stationnaire

Soit la chaîne $(X_n)_{n \geq 0}$ homogène d'espace d'états fini de cardinal m et de matrice de transition P . Comme la chaîne est supposée irréductible d'espace d'états fini, elle est donc récurrente positive ce qui implique qu'elle admet une unique probabilité invariante ou loi stationnaire que nous allons noter $\pi \in \mathbb{R}^{1 \times m}$. Nous allons donc déterminer cette loi ; pour cela on a implémenté une fonction `dist_stationnaire` qui calcule la loi stationnaire π en résolvant le système d'équation suivant :

$$\begin{aligned} & \begin{cases} \pi P = \pi \\ \sum_{k=1}^m \pi(k) = 1 \end{cases} \\ & \Leftrightarrow \begin{cases} P^t \pi^t - \pi^t = 0 \\ (1, \dots, 1) \pi^t = 1 \end{cases} \\ & \Leftrightarrow \begin{cases} \begin{pmatrix} P^t - I_m \\ (1, \dots, 1) \end{pmatrix} \pi^t = \begin{pmatrix} 0_m \\ 1 \end{pmatrix} \end{cases} \end{aligned}$$

Ainsi notre fonction créée procède de la façon suivante :

- On stocke la taille m de la matrice de transition P ;
- On crée la matrice $(P^t - I_m)$ et le vecteur $(0, \dots, 1)$;
- On détermine π en résolvant le système ... avec la fonction `solve()` ;
- Pour finir on retourne π .

4.1.2.1 Exemple avec un algorithme implémenté

On se place à nouveau dans le cas d'une chaîne de Markov à deux états 0 ou 1. On regarde ici pour les différentes matrices de transitions P que l'on a défini précédemment leurs loi stationnaires qu'on note $\pi_i, i \in \{1, 2, 3\}$.

Ainsi pour la matrice P_1 on obtient : $\pi_1 = (0.4705882, 0.5294118)$;

Pour la matrice P_2 on observe : $\pi_2 = (0.5, 0.5)$;

Et la loi stationnaire de la matrice P_3 est : $\pi_3 = (0.6, 0.4)$.

Maintenant qu'on a vu comment simuler des chaînes de Markov et déterminer leurs lois stationnaires, on va vérifier numériquement le théorème ergodique.

On se place à nouveau dans le cas d'une chaîne de Markov à 2 états 0 et 1 de loi initiale μ et de matrice de transition, cette fois-ci, $P = P_3$ que nous avons défini dans la section précédente.

On note π la loi stationnaire de cette chaîne. En utilisant la fonction `loi_stationnaire()` on trouve $\pi = (0.6, 0.4)$. On veut montrer que $\forall k \in \{0, 1\}$, on a :

$$\frac{1}{l} \sum_{i=1}^l \mathbb{1}_{X_i=k} \xrightarrow[l \rightarrow +\infty]{p.s.} \pi(k)$$

Etant donné que l'on travaille avec une chaîne de Markov à deux états, par définition d'une loi de probabilité, il suffit de montrer que :

$$\frac{1}{l} \sum_{i=1}^l \mathbb{1}_{X_i=0} \xrightarrow[l \rightarrow +\infty]{p.s.} \pi(0)$$

Pour notre simulation on pose $l = 10000$ et pour $n \in \{100, l\}$ on simule des chaînes de Markov de longueur n et pour chacune on calcule le nombre moyen de passage à l'état 0, qu'on stocke dans un vecteur t . on trace ensuite le nuage de points dans lequel on superpose la droite d'équation $y = \pi(0) = 0.6$. La figure 4 présente le résultat obtenu. Pour une chaîne de longueur assez grande, le nombre moyen de visites à l'état 0 est proche de 0.6, c'est-à-dire de $\pi(0)$.

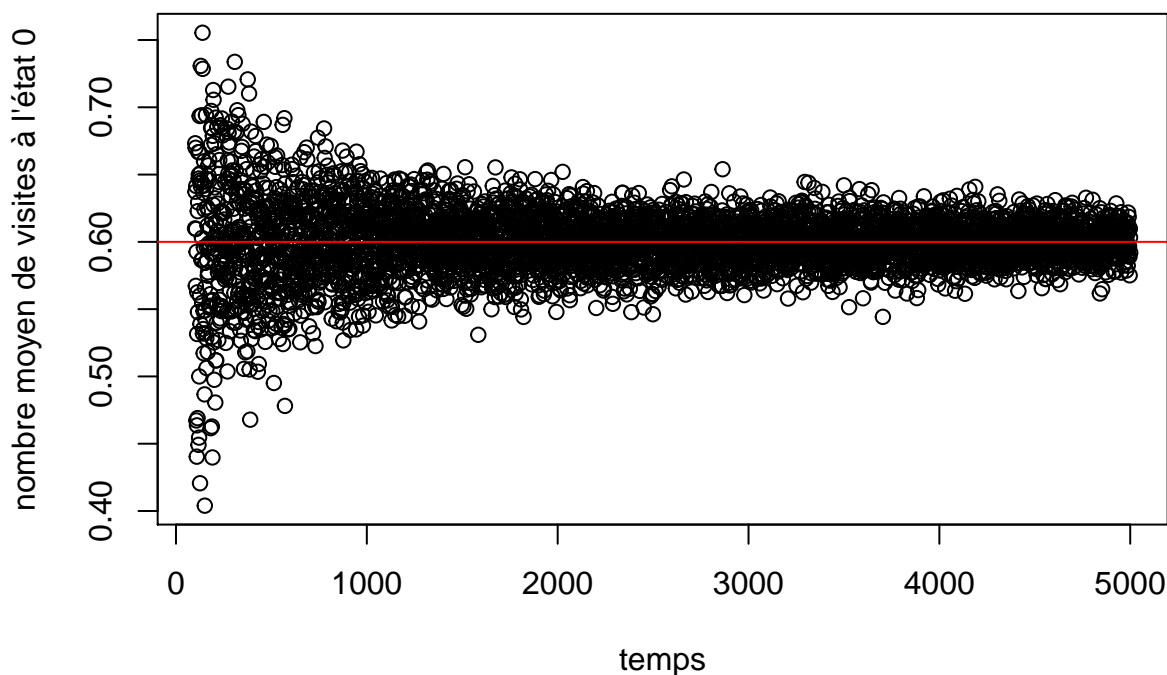


FIGURE 4 – Nombre moyen de visites à l'état 0 selon la longueur de la chaîne simulé

4.1.3 Simulation de la matrice de transition

Maintenant que l'on sait simuler une chaîne de Markov à partir de sa matrice de transition, nous allons à présent estimer la matrice de transition à partir d'une de ses réalisations. Pour cela, on implémente une fonction `matrice_estime()` qui estime la matrice de transition d'une chaîne de Markov $\{X_n, n > 0\}$, à valeurs dans E fini à partir d'une de ses réalisations et de son espace d'états en utilisant les formules établies dans la section 3.1.

Pour faire ces estimations, on utilise de nouveau, pour matrice de transition P les matrices de la sous-section.

Pour chacune de ces trois matrices, on calcule les estimations obtenues avec des réalisations de longueur $n = 1000$ (simulées grâce à la fonction `fct_CM()`). On crée ensuite une fonction \hat{P} .

La fonction consiste à utiliser l'estimateur du maximum de vraisemblance afin de calculer les $\hat{p}_{(x,y)}$, où les éléments $n_{x,y}$ sont le nombre de séquences tel que $(X_t = x, X_{t+1} = y)$. Ainsi :

$$\hat{p}_{(x,y)} = \frac{N_{n,x,y}}{N_n}.$$

Les résultats obtenus sont les suivants :

$$\hat{P}_1 = \begin{pmatrix} 0.803 & 0.197 \\ 0.308 & 0.692 \end{pmatrix}$$

$$\hat{P}_2 = \begin{pmatrix} 0.521 & 0.479 \\ 0.469 & 0.531 \end{pmatrix}$$

$$\hat{P}_3 = \begin{pmatrix} 0.813 & 0.187 \\ 0.276 & 0.724 \end{pmatrix}$$

On s'intéresse également à la convergence des estimateurs, en prenant la matrice P_1 pour référence. Pour mieux visualiser la convergence de l'estimateur de P_3 , on trace pour ses quatre coefficients, la valeur estimée en fonction de la longueur n de la réalisation de la chaîne correspondante. On obtient les graphiques suivants :

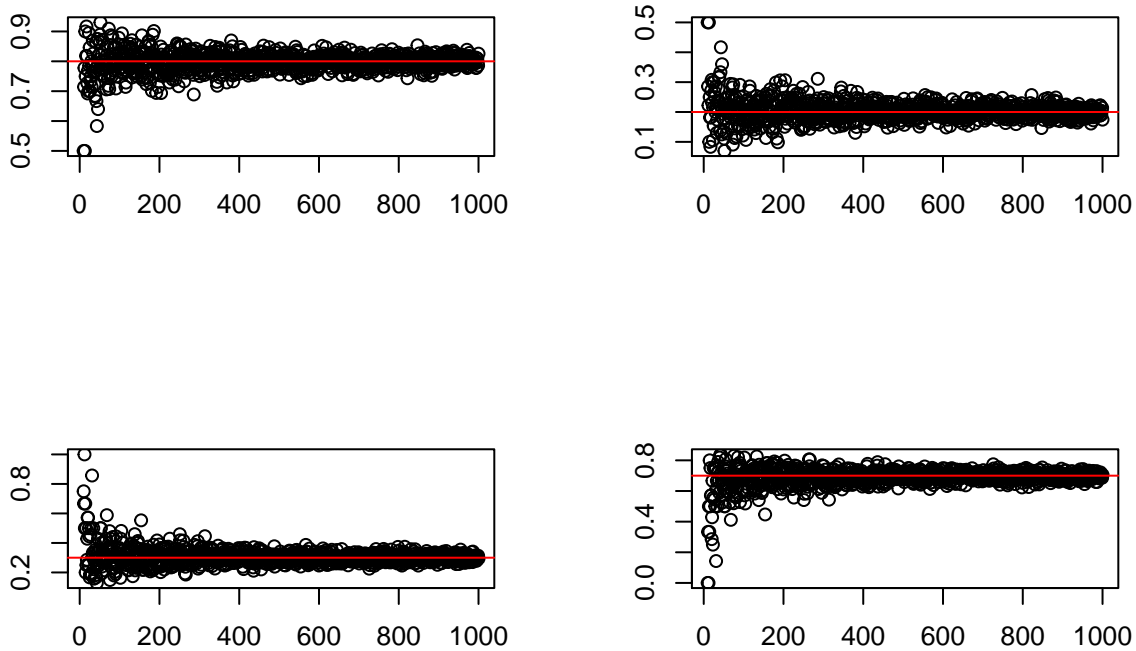


FIGURE 5 – Convergence de l'estimateur de P_3

On observe ainsi que plus la longueur de la chaîne est importante et plus l'estimateur converge vers la vraie valeur.

TABLE 1 – Durée d_t et temps d'attente w_t du Old Faithful Geyser

attente	durée
80	4.016667
71	2.150000
57	4.000000
80	4.000000
75	4.000000
77	2.000000
60	4.383333
86	4.283333
77	2.033333
56	4.833333
81	1.833333
50	5.450000
89	1.616667
54	4.866667
90	4.383333
73	1.766667
60	4.666667
83	2.000000
65	4.733333
82	4.216667

5 Applications des chaînes de Markov sur des données temporelles

On s'intéresse maintenant à la modélisation d'une série temporelle binaire en utilisant les chaînes de Markov.

5.1 Description

Nous allons étudier les données du geyser d'Old Faithful situé dans le Parc National de Yellowstone, aux États-Unis ; il est l'un des geysers les plus connus au monde. La cyclicité de ses éruptions est étudiée depuis les années soixantes afin de comprendre sa dynamique. En effet, le caractère bimodal de la fréquence de ses éruptions intriguent les scientifiques qui cherchent à en connaître les causes. Azzilini et Bowman (1990) ont présenté une série temporelle de ces données c'est celle-ci que nous utilisons dans cette partie. Notre jeu de données contient 299 paires de mesures correspondant au temps d'intervalles entre deux éruptions successives ainsi que la durée d'une éruption. Les données ont été récoltées en continu du 1^e au 15 août 1985. Les paires sont nommées (w_t, d_t) , avec w_t étant le temps d'attente (en minutes) entre deux éruptions successives et d_t , la durée d'une éruption (en minutes).

5.2 Analyse descriptive

La table 1 nous montre les vingt premières lignes du jeu de données.

On se concentre pour notre application, seulement sur la durée d'éruption d'un geyser. La figure 5 nous montre la distribution de la durée d'une éruption. Nous remarquons que la durée d'une éruption est comprise généralement dans un intervalle d'1 minute à 5 minutes, mais on distingue deux sous-groupes dans ce jeu de données, centrées respectivement à 2 minutes et 4 minutes.

Il est important de regarder le comportement, car si l'on considère seulement les données de statistiques (cf. table 2) cela ne nous permet pas de bien discerner le comportement bimodale de cette série ; par exemple

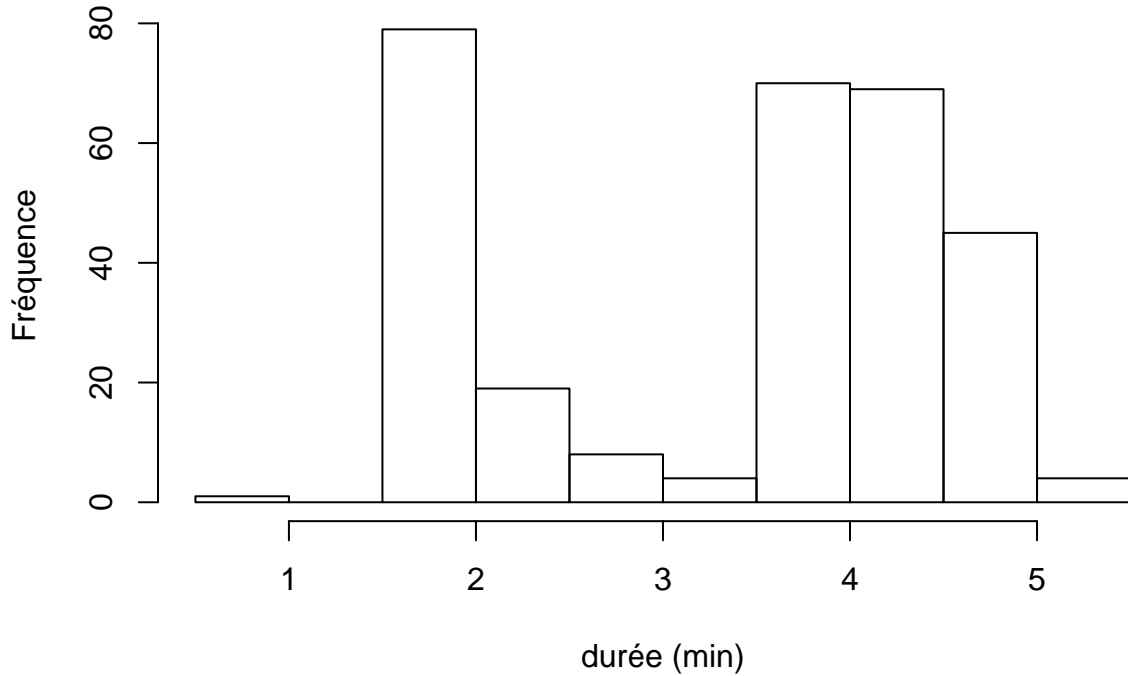


FIGURE 6 – Histogramme de la durée d’une éruption

TABLE 2 – Proportion d’éruptions courtes et longues

Type	Total
0	105
1	194

on voit que la moyenne de durée d’une éruption est de 3.46 minutes mais cela ne décrit pas un résultat typique que l’on retrouve dans un des deux groupes.

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.8333  2.0000  4.0000  3.4608  4.3833  5.4500
```

Du fait de la particularité de ce jeu de données on décide donc de discrétiser le jeu de données en symbolisant par 0 les éruptions de courte durée (moins de 3 minutes) et par 1 les séries de longue durée (plus de 3 minutes). La fréquence des valeurs obtenues est donnée dans la table 2.

On applique sur ce jeu de données une chaîne de Markov. Les motivations sont d’abord d’ordre géologique (cf. Azzalini et Bowman, 1990), mais également cela semble justifié lorsque l’on calcule l’ACF de la série.

```
##
## Autocorrelations of series 'st', by lag
##
## 0.0000 0.0502 0.1003 0.1505 0.2007 0.2508 0.3010 0.3512 0.4013 0.4515 0.5017
## 1.000 -0.657 0.553 -0.398 0.346 -0.282 0.223 -0.172 0.152 -0.172 0.166
## 0.5518 0.6020 0.6522 0.7023 0.7525 0.8027 0.8528 0.9030 0.9532 1.0033 1.0535
## -0.208 0.208 -0.178 0.125 -0.103 0.074 -0.026 0.052 -0.010 0.014 -0.019
## 1.1037 1.1538 1.2040
## 0.031 -0.086 0.097
```

L’autocorrélation d’une série X pour un décalage d’ordre k se calcule de la façon suivante :

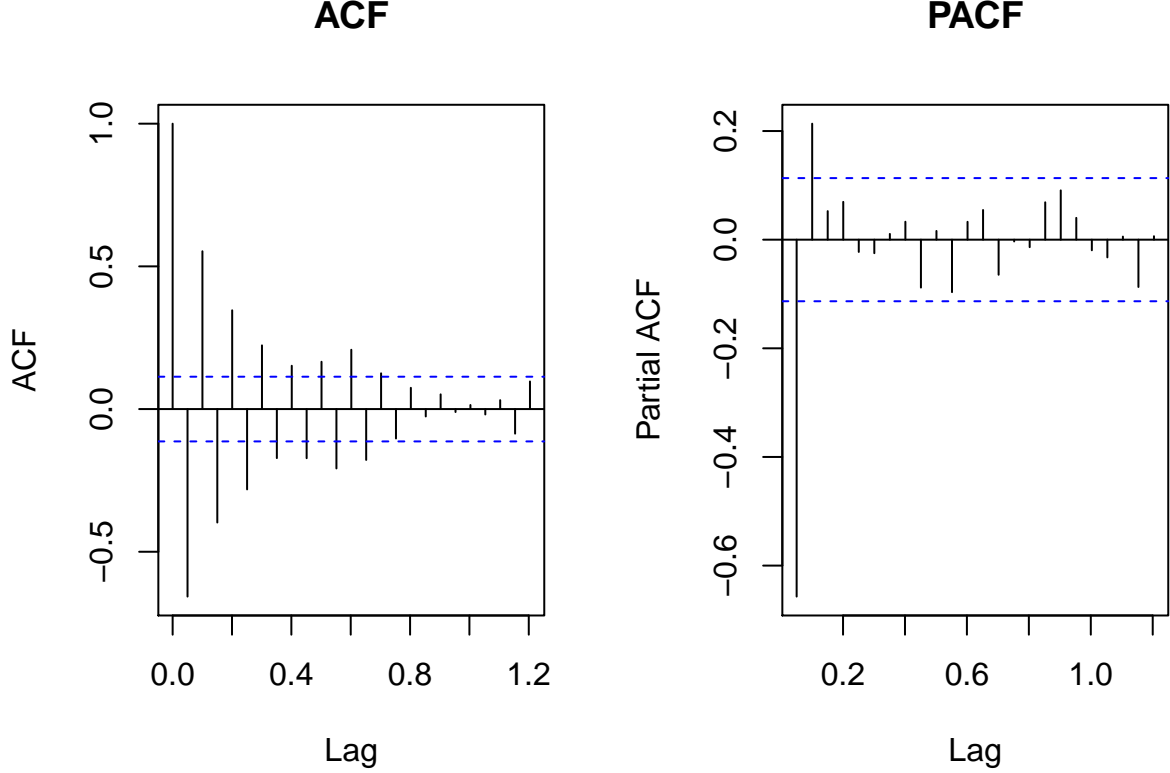


FIGURE 7 – ACF et PACF de la série

$\rho_T(h) : h \rightarrow \frac{\sigma_T(h)}{\sigma_T(0)}$, avec $\sigma_T(h) = \frac{1}{T} \sum_{t=h+1}^T (X_t - \bar{X}_T)(X_{t-h} - \bar{X}_T)$ et $\sigma_T(0) = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}_T)^2$.

Pour une série stationnaire, l'autocorrélogramme empirique, graphe de $h \rightarrow \rho(h)$ décroît exponentiellement vers 0, avec éventuellement des oscillations. Inversement, un graphe d'autocorrélogramme empirique qui ne montre pas de décroissance rapide est l'indice d'une nonstationnarité. Ainsi l'examen du corrélogramme empirique d'une série permet de se faire une idée de sa stationnarité.

Sur le graphique **3** la ligne pointillée bleue indique le niveau en dessous duquel les autocorrélations empiriques simples ne sont pas statistiquement significatives. Ici, on constate que le premier ordre de l'ACF est -0.657 ce qui est significativement négatif. Il semble y avoir une forte corrélation au décalage 1 et 2. Nous utilisons donc les outils des chaînes de Markov pour analyser cette série de données.

5.3 Modélisation

Pour procéder à la modélisation par une chaîne de Markov on sépare notre jeu de données en deux ; le premier correspond au jeu d'apprentissage et le second au jeu de test (validation). Ces derniers vont nous permettre de construire nos différents modèles et d'évaluer la qualité du modèle de prédiction.

Nous décidons de séparer les données avec les proportions suivantes : 80% pour le jeu d'apprentissage et 20% pour le jeu de test.

On commence donc par modéliser une chaîne de Markov de 1^e ordre. A l'aide du jeu de données d'apprentissage, nous commençons par estimer la matrice de transition \hat{P}_{train} .

On note les probabilités de la matrice de transition tel que $p_D = P(D_t = j | D_{t-1} = i)$ avec $i = (0, 1)$ et $j = (0, 1)$.

On calcule l'estimateur de la matrice de transition avec la méthode de l'estimation du maximum de vraisemblance comme cela a été vu précédemment.

La matrice \hat{P}_{train} obtenue est la suivante : $\hat{P}_{train} = \begin{pmatrix} 0 & 1 \\ 0.552 & 0.442 \end{pmatrix}$,

On remarque qu'ici il n'y a pas de transitions 0 à 0, c'est-à-dire qu'une éruption courte est toujours suivi d'une éruption longue.

La loi stationnaire, qu'on note π_{train} est la suivante : (0.3556485, 0.6443515).

Avec le graphique 7 nous avons vu que les coefficients devenaient nuls à partir du retard $h = 2$, ce qui justifie la construction d'une chaîne de markov de second ordre. Etant donné que la séquence (0, 0) ne se produit pas, les 3 états nécessaires pour exprimer la chaîne de markov sont : (0, 1), (1, 0) et (1, 1). La matrice de transition $\hat{P}_{(2,train)}$ est donc :

$$\hat{P}_{(2,train)} = \begin{pmatrix} 0 & 0.690 & 0.31 \\ 1 & 0 & 0 \\ 0 & 0.382 & 0.618 \end{pmatrix}.$$

La distribution stationnaire $\pi_{(2,train)}$ est la suivante : (0.3559322, 0.3559322, 0.2881356).

5.4 Evaluation de la performance

On compare maintenant ces deux modèles sur la base de la log vraisemblance, qu'on note LL , utilisée pour calculer les critères d'AIC et du BIC.

Dans la section 3.1 nous avons vu que la log-vraisemblance s'écrivait de la façon suivante :

$$LL = \sum_{x \in E} \mathbb{1}_{\{X_0=x\}} \log(\mu_x^{(0)}) + \sum_{x,y \in E} N_n^{xy} \log(p_{x,y})$$

apres avoir redéfini cela, on peut estimer l'AIC et le BIC ainsi :

$AIC = 2 \times k - 2 \log(LL)$, avec k le nombre d'états. Le BIC quant à lui est égal à : $BIC = -2 \ln(L) + k \times \ln(N)$

Pour la chaîne de Markov de premier ordre, la log-vraisemblance équivaut à :

$$LL(\hat{P}_{train}) = \log\left(\frac{85}{239}\right) + 85 \times \log\left(\frac{85}{154}\right) + 68 \times \log\left(\frac{68}{154}\right) = -107.14$$

Ainsi, on peut donc en déduire les valeurs de l'AIC et du BIC :

$$AIC = 2 \times 2 - 2 \times (-107.14) = 218.28$$

$$BIC = -2 \times (-107.14) + 2 \times \log(239) = 225.23$$

La log vraisemblance pour la chaîne de de markov de second ordre est égale à :

$$LL = \log(85/237) + 58 \times \log\left(\frac{58}{84}\right) + 26 \log\left(\frac{26}{84}\right) + 26 \times \log\left(\frac{26}{68}\right) + 42 \times \log\left(\frac{42}{68}\right) = -98.23$$

On en déduit donc que :

$$AIC = 2 \times 2 - 2 \times (-98.23) = 200.46$$

$$BIC = -2 \times (-98.23) + 2 \times \log(239) = 207.41.$$

On compare également ces modèles avec une suite de variables aléatoires X_n indépendante de loi de bernouilli p et de taille $n = 239$. On étudie le cas d'indépendance, bien que l'on ait vu que la durée des geysers étaient clairement dépendants entre eux.

TABLE 4 – Table de contingence des valeurs réelles et prédites pour la chaîne de Markov 1e ordre

Prédiction	Réalité	
	0	1
0	0.12	0.28
1	0.22	0.38

TABLE 5 – Table de contingence des valeurs réelles et prédites pour la chaîne de Markov 2nd ordre

Prédiction	Réalité	
	0	1
0	0.08	0.42
1	0.25	0.25

La log-vraisemblance de cette série est :

$$LL = 49 * \log\left(\frac{49}{112}\right) + 63 * \log\left(\frac{63}{112}\right) + 63 * \log\left(\frac{63}{126}\right) + 63 * \log\left(\frac{63}{126}\right) = -164.09$$

Ainsi, on obtient : $AIC = 2 \times 2 - 2 \times (-164.09) = 332.18$ et $BIC = -2 \times (-164.09) + 2 \times \log(239) = 339.13$.

La table 3 presente une comparaison des 3 modèles, en incluant le cas où les variables aléatoires seraient indépendantes. Etant donné la forte dépendance que l'on a remarqué dans ce jeu de donnée il est normal de voir pour ce cas-ci un AIC extrêmement élevé. Ainsi, plus l'ordre du modèle augmente plus l'AIC diminue, le second ordre de la chaîne de markov semble etre un meilleur modèle pour décrire le jeu de données, cela était également observée sur le graphique de l'ACF.

TABLE 3 - Log-vraisemblance et critères de qualité pour les 3 modèles

Modèle	-LL	AIC	BIC
Indépendance	164.09	332.18	339.13
Chaîne de Markov 1 ^e ordre	107.14	218.28	225.23
Chaîne de Markov 2 nd ordre	98.23	200.46	207.41

5.5 Préviation

Avec ces différents modèles nous les utilisons pour faire des prévisions et le jeu de donnée test nous permettra d'évaluer la performance de chacun.

On utilise le modèle basé sur une chaîne de markov de premier ordre pour prédire les 60 valeurs suivantes. On note ici que la dernière observation du jeu d'apprentissage, D_{239} est 1, ainsi la loi initiale de la chaîne de prévision est $\mu = (0, 1)$.

Nous pouvons ainsi comparer notre modele avec les véritables données et ainsi déterminer si notre modèle est efficace pour prédire le type d'éruption qui surviendra.

La table nous montre la proportion de valeurs bien classées.

Pour calculer la performance de notre modèle , on utilise le critère d'accuracy (ou justesse) qui désigne la proportion des prédiction correctes effectuées par le modèle.

Ainsi *accuracy* équivaut à : $\frac{11+30}{11+10+9+30} = 0.683$. Nous obtenons une valeur de justesse égale à 68% (sur 100 prédiction, 68 sont correctes). Notre modèle de prédiction affiche donc un résultat assez satisfaisant.

On utilise à présent le modèle construit sur la base d'une chaîne de Markov de second ordre, afin de prédire des valeurs. La table de contingence obtenue entre les valeurs réelles et prédites est la suivante :

Le critère de justesse pour le modèle de second ordre est de 33% (accuracy : $\frac{5+15}{5+15+25+15} = 0.33$) ; on constate que la qualité de prévision est inférieure par rapport au modèle précédent bien que nous ayons vu que la qualité d'ajustement du modèle était plus élevé.

6 Conclusion

Au cours de ce projet tutoré, nous avons premièrement étudié les chaînes de Markov, défini leurs différentes propriétés et établi les deux principaux théorèmes de convergence : le théorème ergodique ainsi que le théorème central limite. Nous avons également regardé comment construire l'estimateur d'une matrice de transition et étudié ses propriétés. Nous avons ainsi abordé un nouvel outil de probabilités. Ce travail nous a permis par la suite de simuler des chaînes de Markov nous permettant d'illustrer numériquement ces différents théorèmes. Nous avons implémenté sur R différentes fonctions nous permettant de simuler des chaînes de Markov, de trouver leur lois stationnaires ainsi que leurs matrices de transitions. Nous avons ensuite appliqué ces fonctions dans le cas des chaînes à deux états dans le but de vérifier les théorèmes de convergence et la convergence de l'estimateur.

Nous nous sommes ensuite intéressées à une application des chaînes de Markov à une situation réelle : nous avons modélisé la durée d'une éruption. Nous avons travaillé ici avec les données du geyser d'Old Faithful situé dans le Parc National de Yellowstone, aux États-Unis.

La cyclicité de ses éruptions est étudiée depuis les années soixantes afin de comprendre sa dynamique car il a été constaté que la fréquence des éruptions présentaient un caractère bimodal ; elles ne sont donc pas indépendantes entre elles. On utilise donc les chaînes de Markov afin de modéliser ces fréquences, en les considérant comme étant une chaîne de Markov à deux états. Notre but était, maintenant, à partir d'une séquence d'éruption données, de réaliser des prévisions. Dans un premier temps, nous avons séparé notre jeu de données en deux ; l'un nous a servi pour construire nos modèles et le second pour évaluer la qualité des prévisions faites grâce au modèle. Nous avons décidé de comparer différents modèles : Un modèle dans lequel on considère les variables indépendantes, une chaîne de Markov de premier ordre et une de second ordre.

Grâce aux simulations faites précédemment, nous avons pu déterminer pour chacun de ces modèles les éléments qui les caractérisaient c'est-à-dire leurs matrices de transition et leur lois stationnaires. Cela nous a donc permis, pour finir, de réaliser des prévisions et d'évaluer la qualité de performance des modèles ; nous nous sommes donc aperçus que c'est la chaîne de Markov de second ordre qui semble être la plus pertinente pour étudier ces données.

7 Fonctions R

Section 4

- Fonction simulant une chaîne de Markov a 2 états (0 et 1) de longueur n et de loi initiale mu et de matrice de transition P

```
fct_CM<-function(n,mu,P){
  X<-c(length=n)
  U<-sample(runif(1), size = 1)
  X[1]<-1*(U>mu[1])
  for (i in 2:n){
    U<-sample(runif(1), size = 1)
    X[i]<-1*(U>P[X[i-1]+1,1])
  }
  return(X)
}
```

- Fonction calculant la loi stationnaire d'une chaine de Markov

```
dist_stationnaire <- function(P){
  n = dim(P)[1]
  one = diag(x = c(rep(1,n)), nrow = n, ncol = n)
  P_trans = t(P) - one
  A = matrix(nrow = n, ncol = n)
  A[1:n-1,] = P_trans[1:n-1,]
  A[n,] = rep(1,n)

  b = c(rep(0,n-1),1)
  return(solve(A,b))
}
```

- Calcul du nombre moyen de visites à l'état 0

```
P = matrix(c(0.8,0.2,0.3,0.7), ncol = 2, nrow = 2, byrow = T)

l = 1000
t = c()
for (i in 100:l){
  x = fct_CM(n = i, mu = c(0.5,0.5), P = P)
  t[i] = sum(x == 0) / i
}
t = t[100:l]
```

- Fonction calculant la matrice de transition par estimateur de maximum de vraisemblance

```
matrix_estimate = function(x){
  etats = rownames(table(x))
  nb_etat = length(etats)

  # Total de la somme de chaque états dans la chaîne
  N_i = sum(x[1:length(x)-1] == etats[1])
  N_j = sum(x == etats[2])

  # Calcul du nombre de transition

  N_ii = 0
```

```

N_ij= 0
N_ji= 0
N_jj= 0

for (i in seq(1, length(x)-1)){
  if (x[i] == etats[1] & lead(x)[i] == etats[1]){
    N_ii = N_ii+1
  }
  if (x[i] == etats[1] & lead(x)[i] == etats[2]){
    N_ij = N_ij +1
  }
  if (x[i] == etats[2] & lead(x)[i] == etats[1]){
    N_ji = N_ji +1
  }
  if (x[i] == etats[2] & lead(x)[i] == etats[2]){
    N_jj = N_jj + 1
  }
}

P_ii = N_ii / N_i
P_ij = N_ij / N_i

P_ji = N_ji / N_j
P_jj = N_jj / N_j

P_estimate = matrix(c(P_ii, P_ij, P_ji, P_jj), nrow = 2, ncol = 2, byrow = T)

return(P_estimate)
}

```

Section 5

— Création d'un dataset d'apprentissage et d'un dataset de validation

```

nb_lignes <- floor(length(tab$duration) * 0.8)
duree.train = tab$duration[1:nb_lignes]
duree.test = tab$duration[(nb_lignes+1):length(tab$duration)]

cm_duree.train = ifelse(duree.train < 3, 0,1)
cm_duree.test = ifelse(duree.test <3,0,1)

```

— Calcul de la matrice de transition d'une CM d'ordre 2

```

d22 = embed(cm_duree.train,3)
colnames(d22) <- c("passe2", "passe", "present")

counts = table(as.data.frame(d22))
probabilities2 = counts

probabilities2[] <- as.vector(counts) / rep(as.vector(apply(counts, 2:3, sum)), each=dim(counts)[1])

mat2 = as.matrix(probabilities2)

```



```
cm22_matrice = matrix(data = NA, nrow = 3, ncol = 3)

cm22_matrice[1,1] = 0
cm22_matrice[1,2] = mat2[3]
cm22_matrice[1,3] = mat2[4]

cm22_matrice[2,1] = 1
cm22_matrice[2,2] = 0
cm22_matrice[2,3] = 0

cm22_matrice[3,1] = 0
cm22_matrice[3,2] = mat2[7]
cm22_matrice[3,3] = mat2[8]
```

Bibliographie

- [1] Gall, Le Jean-François. *Intégration, Probabilités et Processus Aléatoires (French Edition)*. Independently published, 2019,
www.dropbox.com/sh/19kmitf51i4wzxr/AAA4HgDeCgpZDwXGy7J5zokXa/IPPA2%281%29.pdf?dl=0.
- [2] Chafaï, Djilil, et Florent Malrieu. *Recueil de Modèles Aléatoires (Mathématiques et Applications, 78) (French Edition)*. 1ère éd. 2016, Springer, 2016, doi:10.1007/978-3-662-49768-5.
- [3] Anderson, Mike. « A Closer Look at Old Faithful ». *A Closer Look at Old Faithful*, 13 avril 2016, faculty.business.utsa.edu/manderso/R-examples/Geyser/Geyser.html#the_data.
- [4] Azzalini, A., et A. W. Bowman. « A Look at Some Data on the Old Faithful Geyser ». *Applied Statistics*, vol. 39, no 3, 1990, p. 357. *Crossref*, doi:10.2307/2347385.
- [5] « Estimating Markov Transition Probabilities Between Health States in the Social Security Malaysia (SOCSE) Dataset ». *International Journal of Innovative Technology and Exploring Engineering*, septembre 2019, www.ijitee.org/wp-content/uploads/papers/v8i11S2/K104309811S219.pdf.
- [6] « Markov Chains explained visually ». *Explained Visually*, setosa.io/ev/markov-chains. Consulté le 3 octobre 2020.