

# **Théorème central limite pour les chaînes de Markov**

Sous la direction de Damien Thomine

**STURMA Thomas**  
**MARION Adrien**

L3 Magistère  
Université Paris-Saclay  
2021

## Introduction

Les chaînes de Markov peuvent être vu comme une suite de variables aléatoires  $(X_n)_{n \in \mathbb{N}}$  à "mémoire instantanée" : la loi de la variable  $X_{n+1}$  est entièrement déterminée par le résultat de la variable au temps  $n$ . Elles constituent donc une extension des suites de variables aléatoires indépendantes, dans le sens où on peut bien sûr rajouter une telle mémoire à ces dernières (qui ne serait bien sûr d'aucune utilité).

On s'intéresse dans ce dossier aux propriétés "sur des temps longs" que les chaînes de Markov ont en commun avec les suites de variables identiquement distribuées (I.I.D.).

### Différences apparentes entre chaînes de Markov et suite de variables aléatoires I.I.D

Remarquons pour commencer que pour les variables aléatoires (I.I.D), un argument fréquent réside dans la stabilité de l'espérance par produit :

#### Proposition 0.0.1.

Soit  $(X_n)_{n \in \mathbb{N}}$  est une suite de variable I.I.D à valeurs dans un ensemble  $\Omega$ , et  $(f_n)_{n \in \mathbb{N}}$  une famille de fonctions de  $\Omega$  à valeurs complexes, alors pour tout  $n \in \mathbb{N}$ , on a

$$\mathbb{E} \left( \prod_{k=1}^n f_k(X_k) \right) = \prod_{k=1}^n \mathbb{E} (f_k(X_1))$$

Malheureusement, cet argument (crucial pour la démonstration du théorème central limite et la loi faible des grands nombres) ne peut pas s'appliquer tel quel pour les chaînes de Markov. Il en est de même pour de nombreux cas

Cependant, les simulations informatiques réalisés (voir les sections [3.1.3](#) et [3.2.2](#)) montrent une forte ressemblance de comportement sur des temps longs entre les deux types de suites de variables aléatoires : il semble en effet que le théorème centrale limite et la loi faible des grands nombres s'applique dans certains cas aux chaînes de Markov. Notre objectif dans ce dossier sera donc de nous diriger vers ces résultat.

### Trouver des alternatives avec le formalisme des chaîne de Markov

Notre première étape consiste à trouver une alternative à la proposition [0.0.1](#) pour les chaînes de Markov. Pour cela, on encode les chaînes de Markov dans un cadre matriciel : les lois de probabilité correspondent à une matrice ligne et la définition de la loi au temps  $n + 1$  en fonction du résultat de l'expérience aléatoire au temps  $n$  s'encode dans une matrice de transition (ou stochastique)  $P$  (i.e une

matrice strictement positive, dont la somme des lignes vaut toujours 1). Sous ce formalisme, on détermine une formule pour l'espérance d'un produit :

**Théorème 0.0.2.**

Soit  $(X_n)_{n \in \mathbb{N}}$  une chaîne de Markov à valeurs dans un ensemble  $\Sigma$  de matrice de transition  $P$  et de loi initiale  $\mu_0$ . Soit  $n \in \mathbb{N}$  et  $(f_i)_{0 \leq i \leq n}$  des fonctions de  $\Sigma$  dans  $\mathbb{C}$ . Alors, pour tout  $\sigma \in \Sigma$  :

$$\mathbb{E}\left(\prod_{i=0}^n f_i(X_i)\right) = \mu_0(f_0 \cdot P(f_1 \cdot P(f_2 \cdot P(\dots(f_{n-1} \cdot P f_n)))) \dots) \quad (1)$$

où  $\cdot$  représente le produit terme à terme de vecteurs (i.e le produit de fonctions).

Ce théorème nous conduira assez vite (en remplaçant  $f_1 \dots f_{n-1}$  par la fonction constante égale à 1, que l'on notera  $\mathbf{1}$ ) à nous poser la question du comportement de  $P^n$  quand  $n$  devient grand. C'est l'objet du théorème de Perron Frobenius

**Le théorème de Perron-Frobenius et les premiers résultats asymptotiques**

Pour des matrices  $P \in M_m(\mathbb{C})$  stochastiques spécifiques (nous nous intéresserons ici principalement au cas des matrices dites "primitives"), le théorème de Perron Frobenius permet de dégager deux sous-espaces stables pour  $P$ , l'un de dimension 1 et de vecteur propre  $\mathbf{1}$ , et l'autre de dimension  $n - 1$  et de spectre strictement inférieur à 1.

Cela conduit ( $P^n$ ) à converger vers une matrice stochastique dont toutes les lignes sont constantes, ou en d'autres termes vers une matrice de transition associée à une suite de variables aléatoires I.I.D. Nous montrerons ainsi une forme de décorrélation entre la loi initiale et la loi au temps  $n$  quand  $n$  devient grand :

**Proposition 0.0.3.**

Soit  $f, g$  des fonctions de  $\Sigma$  dans  $\mathbb{C}$ ,

$$\mathbb{E}(f(X_0)g(X_n)) \longrightarrow \mathbb{E}(f(X_0))\mathbb{E}(g(X_n)) \quad (2)$$

**Le théorème centrale limite pour les chaînes de Markov**

Le théorème de Perron-Frobenius, outre la première conséquence d'"indépendantisation" sera un outil précieux, associé à d'autres outils d'analyse dont le théorème des fonctions implicite, pour pouvoir enfin démontrer le théorème centrale limite :

**Théorème 0.0.4.** (*Théorème central limite pour les chaînes de Markov*)

Soit  $(X_n)$  une chaîne de Markov de matrice stochastique  $P$  associée primitive. Soit  $F$  une application de  $\Sigma$  dans  $\mathbb{R}$ , soit  $\pi$  l'unique mesure stationnaire de  $P$  (donnée par le théorème de Perron Frobenius) et supposons que  $\pi F = 0$ . On note :

$$\sigma^2(F) = \pi(F^2) + 2 \sum_{k=1}^{\infty} \pi(F \cdot P^k F)$$

Posons :

$$S_n(F) = \sum_{k=0}^{n-1} F(X_k).$$

La suite  $(S_n(F))$  converge en loi vers  $X \sim \mathcal{N}(0, \sigma^2(F))$ .

### Organisation du dossier

Ce dossier se divise en trois parties : la première constitue une introduction plus formelle aux chaînes de Markov, dont la finalité est de démontrer la formule de l'espérance d'un produit et d'en tirer quelques exemples pratiques d'utilisation. La seconde partie, axée autour du théorème de Perron-Frobenius étudie les premières conséquences "sur des temps longs" de ce théorème fondamental dans l'étude des matrices stochastiques. Enfin, la dernière section s'axe sur les grands théorèmes de convergence : on démontre ici le théorème central limite pour les chaînes de Markov, après l'avoir fait pour le cas I.I.D. On peut ainsi saisir les similarités entre les deux démonstrations, et les complications induites par le cadre Markovien.

# Table des matières

<b>1</b>	<b>Introduction aux chaînes de Markov</b>	<b>5</b>
1.1	Définitions et propriétés	5
1.2	Formule de l'espérance	9
1.2.1	Formule de l'espérance	9
1.2.2	Quelques conséquences	10
<b>2</b>	<b>Mesures stationnaires</b>	<b>13</b>
2.1	Classification	13
2.2	Matrice positive	14
2.3	Théorème de Perron-Frobenius	16
2.4	Applications aux chaînes de Markov	19
2.5	Le cas $A$ irréductible	21
2.6	Conséquences sur des "temps longs"	22
<b>3</b>	<b>Théorème central limite</b>	<b>26</b>
3.1	Le cas des v.a. indépendantes	26
3.1.1	Loi des grands nombres	26
3.1.2	Théorème central limite	27
3.1.3	Expérimentation	31
3.2	Le cas des chaînes de Markov	31
3.2.1	Théorème central limite	32
3.2.2	Expérimentation	36

# Chapitre 1

## Introduction aux chaînes de Markov

Le présent chapitre vise à introduire formellement les chaînes de Markov, ainsi qu'à exposer certains résultats importants qui nous serviront maintes fois par la suite. On s'attardera notamment sur les possibilités d'utilisation du calcul matriciel dans les chaînes de Markov.

Nous n'étudierons ici que des chaînes de Markov ayant un nombre fini d'états. Le lecteur pourra consulter [3] pour étudier des chaînes de Markov plus complexes.

Dans la suite,  $\Sigma = \{\sigma_1, \dots, \sigma_m\}$  sera un ensemble fini qui représentera l'ensemble des états possibles pour la chaîne de Markov  $(X_n)$ . Les fonctions de  $\Sigma$  dans  $\mathbb{C}$  seront représentées par un vecteur colonne (identifiées à un élément de  $M_{m,1}(\mathbb{C})$ ). On les distinguera des mesures de probabilités, définies de  $\Sigma$  dans  $[0, 1]$  qui seront représentées par un élément de  $M_{1,m}(\mathbb{R})$ , c'est-à-dire un vecteur ligne. Grâce à ces notations, on remarquera que, par exemple :

$$\mu f = \sum_{\sigma \in \Sigma} \mu(\sigma) f(\sigma)$$

est l'espérance de la variable aléatoire  $f$  sous la loi  $\mu$ .

### 1.1 Définitions et propriétés

#### Définition 1.1.1.

Soit  $(X_n)_{n \in \mathbb{N}}$  une chaîne de variables aléatoires à valeurs dans  $\Sigma$ . On dit que  $(X_n)$  est une *chaîne de Markov* lorsque, pour tout  $n \in \mathbb{N}$  et  $(i_k) \in \Sigma^{n+1}$  :

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n). \quad (1.1)$$

La chaîne est dite *homogène* (en temps) si pour tout  $n \in \mathbb{N}$  et  $i, j \in \Sigma$  :

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_0 = i) := p(i, j). \quad (1.2)$$

Si les  $X_i$  ne sont pas deux à deux indépendantes, elles n'ont cependant, comme annoncé au début qu'une relation de dépendance restreinte : la variable aléatoire  $X_{n+1}$  ne se "souvient" que du résultat de la variable aléatoire  $X_n$ . Notre objectif est de montrer une forme de ressemblance asymptotique entre chaînes de Markov et suites variables aléatoires indépendantes et identiquement distribuées.

Dans la suite de l'exposé, nous ne considérerons que des chaînes de Markov homogènes.

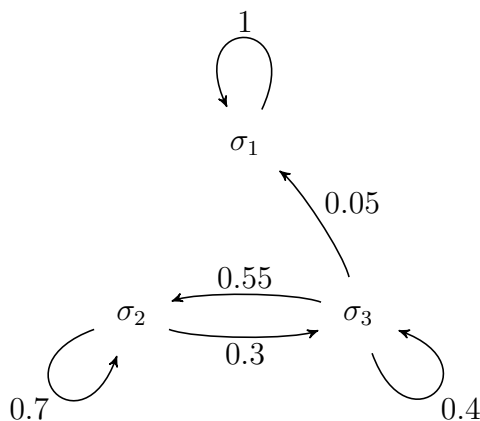
**Définition 1.1.2.**

Soit  $(X_n)$  une chaîne de Markov homogène. On définit  $P$  la *matrice associée* à la chaîne comme :

$$P = (p(i, j))_{i, j \in \Sigma}. \quad (1.3)$$

*Remarque 1.1.3.*

Une chaîne de Markov peut-être vu comme un graphe orienté pondéré, dont sa matrice d'adjacence est exactement la matrice associée à la chaîne.



$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.7 & 0.3 \\ 0.05 & 0.55 & 0.4 \end{pmatrix}$$

**Définition 1.1.4.**

Soit  $P \in M_n(\mathbb{R})$ . Une matrice  $P$  associée à une chaîne de Markov est dite *stochastique*. Elle vérifie les propriétés suivantes :

$$\forall i, j \in \Sigma, 0 \leq P_{i,j} \leq 1 \quad (1.4)$$

$$\forall i \in \Sigma, \sum_{j \in \Sigma} P_{i,j} = 1. \quad (1.5)$$

**Définition 1.1.5.**

Soit  $(X_n)$  une chaîne de Markov. Pour tout  $i, j \in \Sigma$ , posons :

$$p_n(i, j) := \mathbb{P}(X_n = j | X_0 = i). \quad (1.6)$$

On note de plus la matrice  $P_n = (p_n(i, j))_{i, j \in \Sigma}$ .

**Définition 1.1.6.**

On définit la *distribution initiale*, qu'on notera  $\mu_0 \in M_{1,m}(\mathbb{R})$  comme la loi de  $X_0$ . Plus généralement, on note  $\mu_i$  la loi de  $X_i$ .

*Exemple 1.1.7.*

Posons  $\Sigma = \{1, 2, 3, 4, 5, 6\}$ . On invente un dé magique : Il se comporte comme un dé normal au premier lancer, puis par la suite ne peut pas prendre 2 fois la même valeur d'affiler (et tombe de manière équiprobable sur les cinq autres valeurs à chaque lancer). On note la distribution initiale  $\mu_0$  :

$$\mu_0 = (1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6).$$

Si on pose  $(X_n)$  la valeur du dé du  $(n + 1)$ -ème lancer, alors  $(X_n)$  est une chaîne de Markov homogène et la matrice associée à cette chaîne de Markov est :

$$\begin{pmatrix} 0 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0 \end{pmatrix}.$$

**Proposition 1.1.8.**

Soit  $(X_n)$  une chaîne de Markov, alors pour tout  $k \in \mathbb{N}$  :

$$P_k = P^k. \quad (1.7)$$

*Démonstration.*

On va montrer la proposition par récurrence :

- Pour  $k = 0$ , on a bien  $P_0 = P^0 = Id$
- Supposons que la propriété est vraie pour un certain  $k \in \mathbb{N}$ , montrons alors



$$P^{k+1} = P_{k+1} \quad : \text{ soit } i, j \in \Sigma^2$$

$$\begin{aligned} (P_{k+1})_{i,j} &= \mathbb{P}(X_{k+1} = j \mid X_0 = i) \\ &= \sum_{\sigma \in \Sigma} \mathbb{P}(X_1 = \sigma \mid X_0 = i) \mathbb{P}(X_{k+1} = j \mid X_1 = \sigma) \text{ par les probabilités totales} \\ &= \sum_{\sigma \in \Sigma} p(i, \sigma) p_k(\sigma, j) \\ &= (P P_k)_{i,j} \\ &= (P^{k+1})_{i,j} \text{ par hypothèse de récurrence} \quad \square \end{aligned}$$

**Proposition 1.1.9.**

Soit  $(X_n)$  une chaîne de Markov. Alors, pour tout  $k \in \mathbb{N}$  :

$$\mu_k = \mu_0 P^k. \quad (1.8)$$

*Démonstration.*

On raisonne par récurrence :

- Pour  $k = 0$ , on a  $\mu_0 = \mu_0 Id = \mu_0 P^0$  par définition.
- On suppose la propriété vraie pour  $k \in \mathbb{N}$ . On veut alors montrer que  $\mu_{k+1} = \mu_0 P^{k+1}$ , i.e que  $\mu_{k+1} = \mu_k P$ . On remarque que :

$$\begin{aligned} \mu_k P &= \left( \sum_{\sigma \in \Sigma} \mathbb{P}(X_k = \sigma) \times \mathbb{P}(X_{k+1} = p \mid X_k = \sigma) \right)_{p \in \Sigma} \\ &= (\mathbb{P}(X_{k+1} = p))_{p \in \Sigma} \\ &= \mu_{k+1} \end{aligned}$$

et on a bien le résultat souhaité □

*Remarque 1.1.10.* Les suites de variables aléatoires indépendantes et identiquement distribuées constituent une famille particulière de chaîne de Markov. En considérant une telle suite de loi  $\mu$ , on a avec les notations précédentes

$$\begin{aligned} \text{— } \mu_0 &= \mu \\ \text{— } P &= \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}. \end{aligned}$$

En particulier, on vérifie que  $\mu_1 = \mu_0 * P = \mu$ , et d'une manière plus générale, on vérifie que pour tout  $n \in \mathbb{N}$ , on a bien  $P_n = P$ , donc que  $\mu_n = \mu$ .

## 1.2 Formule de l'espérance

La formule de l'espérance présentée ici permettra, en reformulant astucieusement certains problèmes, d'appliquer des résultats de la théorie matricielle aux chaînes de Markov. Elle nous servira tout au long de cet exposé et nous en donnons dès la fin de cette section des applications pratiques.

### 1.2.1 Formule de l'espérance

**Théorème 1.2.1.** (*Formule de l'espérance*)

Soit  $n \in \mathbb{N}$  et  $(f_i)_{0 \leq i \leq n}$  des fonctions de  $\Sigma$  dans  $\mathbb{C}$ . Alors, pour tout  $\sigma \in \Sigma$  :

$$\mathbb{E}\left(\prod_{i=0}^n f_i(X_i)\right) = \mu_0(f_0 \cdot P(f_1 \cdot P(f_2 \cdot P(\dots(f_{n-1} \cdot P f_n)))) \dots) \quad (1.9)$$

où  $\cdot$  représente le produit terme à terme de vecteurs (i.e le produit de fonctions).

*Démonstration.*

Tout d'abord on remarque que pour tout  $n \in \mathbb{N}$  et  $(f_i)_{0 \leq i \leq n}$  des fonctions de  $\Sigma$  dans  $\mathbb{C}$ , alors

$$\mathbb{E}\left(\prod_{i=0}^n f_i(X_i)\right) = \sum_{\sigma \in \Sigma} \mu_0(\sigma) \mathbb{E}\left(\prod_{0 \leq i \leq n} f_i(X_i) \mid X_0 = \sigma\right). \quad (1.10)$$

On démontre alors le résultat par récurrence :

- pour  $n = 0$ , le théorème de transfert indique que l'espérance de  $(f_0(X_0))$  est simplement le produit du vecteur ligne  $\mu_0$  par le vecteur colonne  $f_0(X_0)$ . On a bien le résultat attendu.
- pour  $n = 1$ , on a pour  $\sigma \in \Sigma$  fixé,

$$\begin{aligned} \mathbb{E}(f_0(X_0)f_1(X_1) \mid X_0 = \sigma) &= \sum_{j \in \Sigma} p(\sigma, j) (f_0(\sigma)f_1(j)) \\ &= \sum_{j \in \Sigma} f_0(\sigma)p(\sigma, j)f_1(j) \\ &= f_0 \cdot (P f_1)(\sigma) \end{aligned}$$

et la propriété est bien vérifiée grâce au résultat (1.10).

- soit  $n \geq 1$ , supposons que le résultat est vrai au rang  $n - 1$ , alors en posant  $x_0 = \sigma$

$$\begin{aligned}
 \mathbb{E}\left(\prod_{i=0}^n f_i(X_i) \mid X_0 = \sigma\right) &= \sum_{x_1, \dots, x_n \in \Sigma} f_0(\sigma) \cdot \prod_{i=1}^n f_i(x_i) p(x_{i-1}, x_i) \text{ avec le lemme 1.1.9} \\
 &= f_0(\sigma) \cdot \sum_{x_1, \dots, x_{n-1} \in \Sigma} \left( \prod_{i=1}^{n-1} p(x_{i-1}, x_i) f_i(x_i) \cdot \sum_{x_n \in \Sigma} p(x_{n-1}, x_n) f_n(x_n) \right) \\
 &= f_0(\sigma) \cdot \sum_{x_1, \dots, x_{n-1} \in \Sigma} \prod_{i=1}^{n-1} p(x_{i-1}, x_i) f_i(x_i) \cdot P f_n(x_{n-1}).
 \end{aligned}$$

On pose pour tout  $i \in \{0, \dots, n\}$

$$g_i = \begin{cases} f_n & \text{si } i < n-1 \\ f_{n-1} \cdot (P f_n) & \text{si } i = n-1 \end{cases}$$

on a alors par hypothèse de récurrence :

$$\begin{aligned}
 \mathbb{E}\left(\prod_{i=0}^n f_i(X_i) \mid X_0 = \sigma\right) &= \mathbb{E}\left(\prod_{i=0}^{n-1} g_i(X_i) \mid X_0 = \sigma\right) \\
 &= g_0(\sigma) P(g_1 \cdot P(g_2 \cdot P(\dots (g_{n-2} \cdot P g_{n-1})))) \dots (\sigma) \\
 &= f_0(\sigma) P(f_1 \cdot P(f_2 \cdot P(\dots (f_{n-2} \cdot P(f_{n-1} \cdot P f_n)))) \dots) (\sigma)
 \end{aligned}$$

et on conclut avec la formule (1.10). □

## 1.2.2 Quelques conséquences

### Proposition 1.2.2.

Soit  $n \in \mathbb{N}$ , et  $\sigma_0, \dots, \sigma_n \in \Sigma$ , alors :

$$\mathbb{P}(X_0 = \sigma_0, \dots, X_n = \sigma_n) = \mu(\sigma_0) P_{\sigma_0, \sigma_1} \dots P_{\sigma_{n-1}, \sigma_n}. \quad (1.11)$$

*Démonstration.*

On pose, pour tout  $i \in \llbracket 0, n \rrbracket$ ,  $f_i = \mathbb{1}_{\sigma_i}$ . On remarque que

$$\mathbb{P}(X_0 = \sigma_0, \dots, X_n = \sigma_n) = \mathbb{E}\left(\prod_{i=0}^n \mathbb{1}_{\sigma_i}(X_i)\right).$$

On peut alors appliquer le théorème 1.2.1 :

$$\mathbb{P}(X_0 = \sigma_0, \dots, X_n = \sigma_n) = \mu(\sigma_0) \times P(\mathbb{1}_{\sigma_1} \cdot P(\mathbb{1}_{\sigma_2} \cdot P(\dots(\mathbb{1}_{\sigma_{n-1}} \cdot P\mathbb{1}_{\sigma_n}))) \dots)(\sigma_0).$$

Or, on remarque que  $\mathbb{1}_{\sigma_n} \cdot P\mathbb{1}_{\sigma_{n+1}} = P_{\sigma_n, \sigma_{n+1}} \mathbb{1}_{\sigma_n}$ . On montre alors par récurrence que :

$$P(\mathbb{1}_{\sigma_1} \cdot P(\mathbb{1}_{\sigma_2} \cdot P(\dots(\mathbb{1}_{\sigma_{n-1}} \cdot P\mathbb{1}_{\sigma_n}))) \dots)(\sigma_0) = \prod_{i=0}^{n-1} P_{\sigma_i, \sigma_{i+1}}.$$

1. Initialisation : L'équation (1.2.2) est vérifiée pour  $n = 0$ ,
2. Hérédité : Supposons (1.2.2) vraie pour un certain  $n$ , alors :

$$\begin{aligned} P(\mathbb{1}_{\sigma_1} \cdot \dots(\mathbb{1}_{\sigma_n} \cdot P\mathbb{1}_{\sigma_{n+1}}) \dots)(\sigma_0) &= P(\mathbb{1}_{\sigma_1} \cdot \dots(P_{\sigma_n, \sigma_{n+1}} * \mathbb{1}_{\sigma_n}) \dots)(\sigma_0) \\ &= P_{\sigma_n, \sigma_{n+1}} \times P(\mathbb{1}_{\sigma_1} \cdot \dots(\mathbb{1}_{\sigma_n}) \dots)(\sigma_0) \\ &= P_{\sigma_0, \sigma_1} \dots P_{\sigma_n, \sigma_{n+1}}. \end{aligned} \quad \square$$

**Proposition 1.2.3.**

Pour toutes fonctions  $f, g$  de  $\Sigma$  dans  $\mathbb{C}$  et tout  $n \in \mathbb{N}$ ,

$$\mathbb{E}(f(X_0)g(X_n)) = \mu_0(f \cdot (P^n g)) \quad (1.12)$$

où  $\cdot$  encode le produit de fonctions.

*Démonstration.*

On pose  $f_0 = f$ ,  $f_i = \mathbb{1}$  pour  $1 \leq i \leq n-1$  et  $f_n = g$ . On a alors

$$\mathbb{E}(f(X_0)g(X_n)) = \mathbb{E}(f_0(X_0)f_1(X_1)\dots f_n(X_n)).$$

Donc avec le théorème 1.2.1 :

$$\begin{aligned} \mathbb{E}(f(X_0)g(X_n)) &= \mathbb{E}(f_0(X_0)f_1(X_1)\dots f_n(X_n)) \\ &= \sum_{i=0}^m \mu_0(\sigma_i) \mathbb{E}(f_0(X_0)f_1(X_1)\dots f_n(X_n) \mid X_0 = \sigma_i) \\ &= \sum_{i=0}^m \mu_0(\sigma_i) f_0(\sigma_i) P(f_1 \cdot P(f_2 \cdot P(\dots(f_{n-1} \cdot P f_n))) \dots)(\sigma_i) \\ &= \sum_{i=0}^m \mu_0(\sigma_i) (f \cdot (P^n g))(\sigma_i) \\ &= \mu_0 f \cdot (P^n g). \end{aligned}$$

□

**Proposition 1.2.4.**

Pour toutes fonctions  $g$  de  $\Sigma$  dans  $\mathbb{C}$ , et tout  $n \in \mathbb{N}$ ,

$$\mathbb{E}(g(X_n)) = \mu_0(P^n g). \quad (1.13)$$

*Démonstration.*

Il suffit d'appliquer la proposition précédente avec  $f = \mathbb{1}$ .

□

*Exemple 1.2.5.*

On peut utiliser le théorème pour avoir une formule explicite pour de nombreux calcul de probabilité : on cherche par exemple à connaître la probabilité  $E$  d'avoir les tirages suivants avec le dé magique :  $X_0$  pair,  $X_1$  impair et  $X_2 \geq 4$ .

On pose alors  $f_0 = \mathbb{1}_{X \text{ pair}}$ ,  $f_1 = \mathbb{1}_{X \text{ impair}}$  et  $f_2 = \mathbb{1}_{X > 3}$ . On a

$$\begin{aligned} E &= \mathbb{P}(f_0(X_0) = 1, f_1(X_1) = 1, f_2(X_2) = 1) \\ &= \mathbb{E}(f_0(X_0)f_1(X_1)f_2(X_2)) \\ &= \mu_0 f_0 \cdot P (f_1 \cdot P f_2) \\ &= \mu_0 \begin{pmatrix} 0 \\ 0.32 \\ 0 \\ 0.32 \\ 0 \\ 0.32 \end{pmatrix} \\ &= 0.16 \end{aligned}$$

avec  $\mu_0 = \frac{1}{6}\mathbb{1}$  la distribution initiale du dé magique et  $P$  la matrice de transition associée à l'expérience.

# Chapitre 2

## Mesures stationnaires, théorème de Perron-Frobenius

Quand on s'intéresse aux comportements des chaînes de Markov sur des temps longs, plusieurs cas de figures peuvent apparaître. Donnons deux exemples :

1. En reprenant l'idée du dé magique exposé en première partie, on devine intuitivement que le résultat du premier lancé influera peu sur le résultat du dixième.
2. Si on considère un dé magique  $B$ , qui ne peut pas donner deux nombres de même parité à la suite, les choses changent considérablement. La parité du nombre tiré au premier lancer est alors déterminante pour la suite.

Si l'on veut dégager une ressemblance asymptotique entre chaînes de Markov et suites de variables aléatoires I.I.D, il est donc nécessaire de se restreindre à une sous-famille de chaîne de Markov, associées à des matrices dites primitives.

Dans ce chapitre, nous nous pencherons en détail sur le théorème de Perron Frobenius dans le cas primitif. Élément centrale de la théorie des matrices à coefficients positifs, ce théorème se révèlera très utile à appliquer aux chaînes de Markov pour montrer une forme "d'indépendantisation" des variables aléatoires sur des temps longs

### 2.1 Classification

#### **Proposition-Définition 2.1.1.**

*Soit  $i, j \in \Sigma$ . On dit que les états  $i$  et  $j$  communiquent entre eux, noté  $i \sim j$ , s'il existe  $k, l \in \mathbb{N}$  tel que  $p_k(i, j) > 0$  et  $p_l(j, i) > 0$ . C'est une relation d'équivalence.*

*Démonstration.*

On doit montrer les 3 propriétés d'une relation d'équivalence :

1. Réflexivité : On a bien  $i \sim i$  car  $p_0(i, i) = 1$ ,
2. Symétrie : Supposons qu'on a  $i \sim j$ , alors  $p_l(j, i) > 0$  et  $p_k(i, j) > 0$ , donc  $j \sim i$ ,
3. Transitivité : Soit  $x, y, z \in \Sigma$  tel que  $i \sim j$  et  $j \sim h$ . Il existe  $k, k'$  tel que  $p_k(i, j), p_{k'}(j, h) > 0$ . Or :

$$\begin{aligned}
 p_{k+k'}(i, h) &= \mathbb{P}(X_{k+k'} = h | X_0 = i) \\
 &\geq \mathbb{P}(X_{k+k'} = h, X_k = \sigma | X_0 = i) \\
 &\geq \mathbb{P}(X_{k+k'} = h | X_m = j) \mathbb{P}(X_k = j | X_0 = i) \\
 &\geq p_{k'}(j, h) p_k(i, j) > 0.
 \end{aligned}$$

Par le même calcul avec les indices  $l, l'$  tel que  $p_l(j, i), p_{l'}(h, j) > 0$ , on obtient bien  $p_{l+l'}(k, i) > 0$ . Donc, on a  $i \sim k$ .

□

### Définition 2.1.2.

Soit  $(X_n)$  une chaîne de Markov. On dit que la chaîne est *irréductible* lorsque pour tout  $i$  et  $j$  des éléments de  $\Sigma$  on a  $i \sim j$ .

### Exemple 2.1.3.

Reprenons l'exemple du dé magique.  $(X_n)$  est une chaîne irréductible. En effet, soit  $i, j \in \{1, \dots, 6\}$ , alors si  $i \neq j$  on a  $p(i, j) > 1$ , et si  $i = j$  alors  $p_0(i, j) = 1$ .

### Définition 2.1.4.

Soit  $T \in \Sigma / \sim$  et notons  $V_T = \text{Card}(n | X_n \in T)$  On dit que  $T$  est transitoire lorsque  $\mathbb{P}(V_T = \infty) = 0$ . Au contraire,  $T$  est dite récurrente si elle n'est pas transitoire.

### Remarque 2.1.5.

Au bout d'un certain temps, une chaîne de Markov se retrouve "bloquée" dans une classe récurrente. Dans la suite, on supposera que la chaîne est irréductible.

## 2.2 Matrice positive

### Définition 2.2.1.

Soit  $A$  une matrice. On dit que  $A$  est *positive* (resp. strictement positive) si toutes les composantes de  $A$  sont positives (resp. strictement positives). On note  $A \geq 0$  (resp.  $A > 0$ ) lorsque c'est le cas. Cette définition s'étend pour les vecteurs, en les voyant comme élément de  $M_{m,1}(\mathbb{R})$  par exemple.

**Définition 2.2.2.**

Soit  $A$  une matrice de  $M_m(\mathbb{C})$ . On dit que  $A$  est une matrice *primitive* si  $A$  est positive et qu'il existe  $k \in \mathbb{N}$  tel que  $A^k > 0$ .

De plus,  $A$  est une matrice *irréductible* lorsque pour tout couple  $(i, j) \in \Sigma$ , il existe  $k \in \mathbb{N}$  tel que  $(A^k)_{i,j} > 0$ .

Faisons le lien entre ces définitions et les chaînes de Markov :

**Proposition 2.2.3.**

Soit  $(X_n)$  une chaîne de Markov et  $P$  sa matrice associée, alors  $P$  est une matrice irréductible si et seulement si la chaîne  $(X_n)$  est irréductible.

*Démonstration.*

Soit  $i, j \in \Sigma$ , alors par la proposition 1.1.8 on a les équivalences suivantes :

$$(P^n)_{i,j} > 0 \iff (P_n)_{i,j} > 0 \iff p_n(i, j) > 0.$$

□

*Exemple 2.2.4.*

Dans le cas du dé magique, la matrice associée  $A$  est primitive car  $A^2 > 0$ . Par contre, le dé magique B, dont la définition est donné en introduction à ce chapitre, la matrice associée n'est pas primitive, mais seulement irréductible.

**Lemme 2.2.5.**

Soit  $A \in M_n(\mathbb{R})$ . Il y a équivalence entre les 2 assertions suivantes :

1.  $A$  est strictement positive
2. Pour tout vecteur  $x$  positif non nul,  $Ax > 0$

*Démonstration.*

Si  $A$  est strictement positive, alors pour tout  $x \in (\mathbb{R}^+)^n$  non nul et  $j \in \{1, \dots, n\}$  :

$$(Ax)_j = \sum_{i=1}^n a_{i,j}x_i > 0$$

car il existe  $k \in \{1, \dots, n\}$  tel que  $x_k$  est non nul, donc  $a_{k,j}x_k > 0$ . Montrons alors le sens indirect : soit  $i, j \in \{1, \dots, n\}$ , notons  $x = \mathbf{1}_j$ , alors, par le calcul :

$$\begin{aligned} (Ax)_i &= \sum_{k=1}^n A_{i,k}x_k \\ &= A_{i,j} \end{aligned}$$

or, puisque  $Ax > 0$ , alors  $A_{i,j} > 0$ . Donc  $A > 0$ .

□



## 2.3 Théorème de Perron-Frobenius

*Remarque 2.3.1.*

La première version de ce théorème fut formulée (dans le cas primitif) en 1907 par Oskar Perron, alors qu'il étudiait les fractions continues. Georg Frobenius l'étendit en 1912 aux matrices à coefficients positifs. C'est durant la même période, en 1908 qu'Andreï Markov introduisait les chaînes qui portent aujourd'hui son nom.

Cependant, peut être à cause de leurs nationalités différentes, leurs travaux ne semblent pas s'être croisés et Markov démontra une version restreinte du même théorème. Ce ne sera que dans les années 1930 que le rapprochement entre les deux branches aura lieu[1].

**Définition 2.3.2.**

Soit  $A \in M_n(\mathbb{R})$ . On définit le *rayon spectral* de  $A$ , noté  $\rho(A)$ , par :

$$\rho(A) = \sup\{|\lambda| \mid \lambda \in Sp(A)\}.$$

**Théorème 2.3.3.** (*Théorème de Perron-Frobenius, cas primitif*)

Soit  $A$  une matrice primitive, alors :

1. Il existe un vecteur  $x^0$  strictement positif tel que  $Ax^0 = \rho(A)x^0$ .
2. Pour toute valeur propre  $\lambda$  différent de  $\rho(A)$ ,  $|\lambda| < \rho(A)$ .
3.  $\rho(A)$  est une valeur propre simple.

*Démonstration.*

Supposons d'abord que  $A > 0$ .

*Étape 1* : On va tout d'abord montrer qu'il existe bien un vecteur  $x^0$  tel que  $Ax^0 = rx^0$  avec  $r$  le rayon spectral de  $A$ . Notons  $X = \{\lambda \in \mathbb{R} \mid \exists x \geq 0, \|x\|_1 = 1, \lambda x \leq Ax\}$  où la norme  $\|\cdot\|_1$  est définie par :

$$\|x\|_1 = \sum_{k=1}^n |x_k|.$$

On remarque que  $X$  est non vide, et est majoré par  $\|A\|_\infty$ . On note  $r$  sa borne supérieure, on montrera plus tard que  $r = \rho(A)$ . Soit  $(\lambda_i)$  une suite de  $X$  qui converge vers  $r$ , et  $x^i$  des vecteurs positifs tel que  $Ax^i \geq \lambda_i x^i$  et  $\|x^i\|_1 = 1$ . Par compacité de  $[0, 1]^n$ , on peut extraire une sous-suite convergente, par Bolzano-Weirstrass, qui converge vers  $x^0$ . On a  $\|x^0\|_1 = 1$ ,  $x^0 \geq 0$  et  $Ax^0 \geq rx^0$ .

Supposons que  $Ax^0 \neq rx^0$  et posons  $y^0 = Ax^0$ . On remarque que le vecteur  $(y^0 - rx^0)$  est positif non nul. En appliquant le lemme 2.2.5, on obtient  $A(y^0 - rx^0) > 0$ , donc  $Ay^0 > ry^0$ . Il existe donc  $\epsilon > 0$  tel que  $Ay^0 > (r + \epsilon)y^0$ . Alors :

$$A \frac{y^0}{\|y^0\|_1} > (r + \epsilon) \frac{y^0}{\|y^0\|_1}.$$

Dans ce cas,  $r + \epsilon$  est dans  $X$  puisque  $y^0$  est positif. Cela contredit le fait que  $r$  soit la borne supérieure de  $X$  car  $r + \epsilon \in X$ . Dans ce cas  $Ax^0 = rx^0$ . De ce fait, puisque le lemme 2.2.5 nous donne  $Ax^0 = \lambda x^0 > 0$ , donc  $x^0 > 0$ .

*Étape 2* : Maintenant, on va montrer que  $r = \rho(A)$  et que pour toute autre valeur propre de  $A$ , celle-ci est de module strictement inférieur à  $\rho(A)$ . Soit  $i \in \llbracket 1, n \rrbracket$ , soit  $\lambda$  une valeur propre différent de  $r$  et  $z$  un vecteur propre associé à  $\lambda$ . Alors :

$$\sum_{j=1}^n a_{i,j} z_j = \lambda z_i$$

on peut appliquer le module à l'égalité et l'inégalité triangulaire :

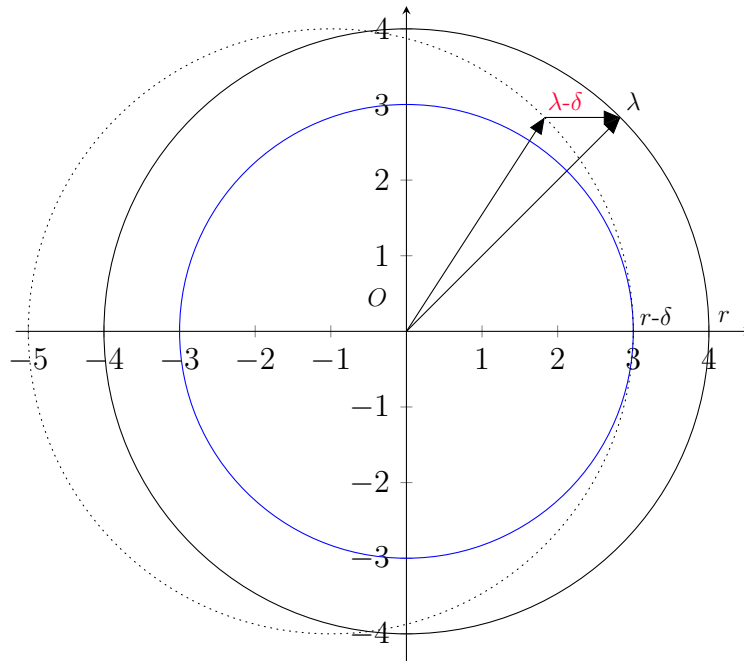
$$\sum_{j=1}^n a_{i,j} |z_j| \geq |\lambda| |z_i|.$$

Donc  $A|z| \geq |\lambda| |z|$ , alors  $|\lambda| \in X$ , donc  $|\lambda| \leq r$  puisque  $r$  est la borne supérieure de  $X$ . On a alors que le rayon spectral de  $A$  est égal à  $r$ .

Soit  $\delta$  assez petit tel que  $A_\delta = A - \delta Id > 0$ . On a alors que  $r - \delta$  est la plus grande valeur propre positive de  $A_\delta$ . Dans ce cas, on a  $|\lambda - \delta| \leq r - \delta$ . Alors :

$$|\lambda| = |\lambda - \delta + \delta| \leq |\lambda - \delta| + \delta \leq r.$$

Dans le cas où  $|\lambda| = r$ , alors  $|\lambda| = |\lambda - \delta| + \delta$ . Cela implique que  $\lambda$  est réel et positif, donc  $\lambda = r$ . On a bien  $|\lambda| < r$ .



De manière géométrique, lorsque  $\lambda$  parcourt le cercle de centre  $O$  et de rayon  $r$  (représenté par le cercle noir en trait plein), alors  $\lambda - \delta$  parcourt le cercle translaté de  $-\delta$  (représenté par le cercle en pointillée). Or, vu que  $\lambda - \delta$  est une valeur propre de  $A_\delta$  et que  $r - \delta$  est la valeur spectrale de  $A_\delta$  alors  $|\lambda - \delta| \leq r - \delta$ . On s'intéresse donc aux points du cercle en pointillée qui se trouve dans le cercle bleu de rayon  $r - \delta$ . On remarque que cela ne concerne qu'un seul point : lorsque  $\lambda - \delta = r - \delta$ . Donc on trouve bien que  $\lambda = r$ .

*Étape 3* : Pour finir, montrons que  $\rho(A) = r$  est valeur simple. Soit  $y$  un vecteur propre associé à  $r$  et qui n'est pas colinéaire à  $x^0$ . Comme  $A$  est réelle, alors  $\Re(y)$  et  $\Im(y)$  sont aussi vecteurs propres associés à  $r$ . Puisque  $y$  n'est pas colinéaire à  $x^0$ , alors  $\Re(y)$  ou  $\Im(y)$  n'est pas colinéaire à  $x^0$ . Quitte à prendre ce vecteur non colinéaire à  $x^0$ , on peut supposer  $y$  réel.

Puisque  $x^0 > 0$  alors il existe  $\mu$  tel que  $z = x^0 - \mu y$  admette une coordonnée nulle. On peut prendre

$$|\mu| = \min \frac{x_i^0}{|y_i|}$$

avec  $y_i$  non nul. On prend aussi  $\mu$  de signe convenable. On a donc  $Az = rz$ , or par le lemme 2.2.5, on obtient  $z > 0$ , il y a donc une contradiction puisque  $z$  admet une coordonnée nul. Donc il ne peut pas exister un tel  $y$ , ce qui montre bien que  $r$  est une valeur propre simple.

Maintenant supposons  $A$  primitive, et posons  $m \in \mathbb{N}$  tel que  $A^m > 0$ . On va montrer les 3 points du théorème :

*Étape 1* : Premièrement, on trouve  $x^0$  tel que  $Ax^0 \geq rx^0$ . Supposons  $Ax^0 - rx^0$  est non nul, alors, en multipliant par  $A^m$ , on obtient  $A^{m+1}x^0 > rA^m x^0$ . Posons  $y = A^m x^0 > 0$ , alors  $Ay > ry$  mais, de même que tout à l'heure, cela contredit la définition même de  $r$ . Donc on a bien  $Ax^0 = rx^0$ .

De plus, on peut en déduire que  $A^m x^0 = r^m x^0$ . Puisque  $A^m > 0$  et  $x^0$  positif non nul, alors par le lemme 2.2.5, on a  $r^m x^0$  et donc  $x^0 > 0$ . On a montré le point 1.

*Étape 2* : On a toujours  $|\lambda| \leq r$  car, précédemment, on a juste utilisé le fait que  $A$  soit positive. Donc, même pour une matrice primitive, on a bien  $\rho(A) = r$ . Supposons que  $|\lambda| = r$  et notons  $z$  un vecteur propre associé à  $\lambda$ . On a donc  $A^m z = \lambda^m z$  et  $|\lambda^m| = r^m$ . On veut montrer que  $r^m$  est la plus grande valeur propre de  $A^m$ , pour réutiliser l'argument dans le cas  $A > 0$  :

Notons  $s$  la plus grande valeur propre de  $A^m$ , que l'on suppose différent de  $r^m$  et  $x^1$  un vecteur propre associé. On peut supposer que  $x^1 > 0$  car  $A^m > 0$ . On remarque que  $A^m$  admet 2 valeurs propres positives, et  $x^0, x^1 > 0$ . Soit  $\mu$  tel que  $y = x^0 - \mu x^1 \geq 0$  mais admette une composante nulle. Puisque  $A^m > 0$ , alors  $A^m y > 0$  par le lemme 2.2.5. Or :

$$A^m y = r^m x^0 - \mu s x^1 = r^m (x^0 - \mu x^1) - (s - r^m) \mu x^1.$$

Or, dans le terme de droite, le premier vecteur n'est pas strictement positif, alors que le deuxième l'est : cela contredit le fait que  $A^m y > 0$ . Donc  $s = r^m$ .

On sait alors que  $r^m$  est la plus grande valeur propre de  $A^m$ . On peut alors appliquer les arguments d'avant pour montrer que  $\lambda^m = r^m$ . Or puisque  $x^0$  et  $z$  ne sont pas colinéaires, alors  $r^m$  ne serait pas simple, c'est donc en contradiction avec le résultat d'avant car  $A^m > 0$ .

*Étape 3* : Pour finir, puisque tout vecteur propre de  $A$  pour  $r$  est vecteur propre de  $A^m$  pour  $r^m$ , alors :  $\dim(E_r(A)) \leq \dim(E_{r^m}(A^m)) = 1$ . Donc  $r$  est une valeur propre simple. □

## 2.4 Applications aux chaînes de Markov

Cette section va faire le lien entre le théorème de Perron-Frobenius et les chaînes de Markov. Pour commencer, calculons le rayon spectral d'une matrice stochastique :

**Proposition 2.4.1.**

Soit  $A$  une matrice stochastique. Alors 1 est valeur propre de  $A$ .

*Démonstration.*

En effet, le vecteur  $\mathbf{1}$  est vecteur propre de  $A$  associé à la valeur propre 1. □

**Proposition 2.4.2.**

Soit  $A$  une matrice stochastique. Alors le rayon spectral de  $A$  est 1.

*Démonstration.*

On a déjà vu que 1 est valeur propre de  $A$ , donc  $\rho(A) \geq 1$ . Or, soit  $\lambda$  une valeur propre de  $A$  et  $x$  un vecteur propre associé à  $\lambda$ . Alors  $Ax = \lambda x$ , donc

$$\lambda x_i = \sum_{j=1}^n a_{i,j} x_j.$$

Donc, en appliquant le module :

$$|\lambda| |x_i| \leq \sum_{j=1}^n a_{i,j} |x_j|.$$

Posons  $k$  tel que  $x_k = \max(|x_i|)$ , alors :

$$|\lambda| |x_k| \leq \sum_{j=1}^n a_{k,j} |x_k| = |x_k|.$$

On en conclut que  $|\lambda| \leq 1$  □

**Définition 2.4.3.**

Soit  $A$  une matrice stochastique. On dit que  $\pi \in \mathbb{R}^n$  est une mesure de probabilité stationnaire de  $A$  si :

$$\pi = \pi A. \tag{2.1}$$

*Remarque 2.4.4.*

On remarque qu'une mesure stationnaire  $\pi$  est un vecteur propre à gauche de valeur propre 1 tel que  $\|\pi\|_1 = 1$  et  $\pi \geq 0$ .

Grâce au théorème de Perron-Frobenius, on va montrer l'existence et l'unicité d'une telle mesure stationnaire pour une certaine catégorie de matrice :

**Théorème 2.4.5.**

Soit  $A$  une matrice stochastique primitive. Alors il existe une unique mesure de probabilité stationnaire  $\pi$  de  $A$ .

*Démonstration.*

Tout d'abord, remarquons que  $\rho({}^tA) = \rho(A)$ . Puisque  $A$  est stochastique, alors on a vu par la proposition 2.4.2 que  $\rho(A) = 1$ . De plus, on sait que  $A$  est primitive, on peut alors appliquer le théorème de Perron-Frobenius à la matrice  ${}^tA$ . Donc il existe un unique vecteur positif  $x$  tel que  ${}^tAx = x$ , on pose alors :

$$\pi = \frac{x}{\|x\|_1}. \quad \square$$

*Exemple 2.4.6.* Notons  $A$  la matrice associée au dé magique (1.1.7). Par le calcul, on trouve que :

$$\text{Sp}(A) = \{1, -0.2\}$$

dont -0.2 est valeur propre de multiplicité 5.

On verra l'utilité des mesures stationnaires dans la partie 2.6

## 2.5 Le cas $A$ irréductible

Dans cette partie, nous parlerons du théorème de Perron-Frobenius dans le cas irréductible et de quelques conséquences. On se privera de démonstration. Le lecteur pourra consulter [5, chapitre III] pour plus de détails.

**Théorème 2.5.1.** (*Théorème de Perron-Frobenius, cas irréductible*)

*Soit  $A$  une matrice irréductible, alors :*

1. *Il existe un vecteur  $x^0$  strictement positif tel que  $Ax^0 = \rho(A)x^0$ .*
2. *Pour toute autre valeur propre  $\lambda$  que  $\rho(A)$ ,  $|\lambda| \leq \rho(A)$ .*
3.  *$\rho(A)$  est une valeur propre simple.*

*Remarque 2.5.2.* La seule différence entre les deux théorèmes de Perron-Frobenius est le fait que  $\rho(A)$  ne soit pas forcément strictement supérieure à toutes les autres, mais seulement supérieure. Dans le cas d'une matrice stochastique, on peut même donner le résultat suivant :

**Proposition 2.5.3.** *Soit  $P$  une matrice stochastique irréductible et  $\lambda$  une valeur propre de  $P$  de module 1. Alors  $\lambda$  est une racine de l'unité.*

Pour autant, on peut encore prouver l'existence et l'unicité d'une mesure stationnaire dans le cas des matrices stochastiques irréductibles :

**Théorème 2.5.4.**

*Soit  $A$  une matrice stochastique irréductible. Alors il existe une unique mesure de probabilité stationnaire  $\pi$  de  $A$ .*

*Démonstration.*

La démonstration est semblablement la même que celle du théorème 2.4.5, puisque le théorème de Perron-Frobenius dans le cas irréductible assure que 1 est toujours une valeur simple.  $\square$

## 2.6 Conséquences sur des "temps longs"

Dans le cas d'une matrice stochastique  $P$ , on a vu que 1 était valeur propre (de vecteur propre  $\mathbf{1}$ ). En réalisant une réduction de Jordan, on en déduit qu'il existe une matrice  $D$  de la forme

$$D = \begin{pmatrix} 1 & (0) \\ (0) & N \end{pmatrix} \quad (2.2)$$

et une matrice  $Q$  inversible telle que  $P = QDQ^{-1}$ . La première ligne de  $Q^{-1}$  est alors l'unique mesure stationnaire de  $\pi$  de  $P$ .

On peut traduire ces remarques en termes de sous espaces stables :

### Proposition 2.6.1.

*Soit  $P$  une matrice stochastique irréductible et  $\pi$  son unique mesure de probabilité stationnaire. Alors  $\pi$  scinde l'espace des fonctions de  $\Sigma$  dans  $\mathbb{C}$  en deux sous espaces supplémentaires stables par  $P$ , i.e. :*

$$\mathbb{C}^\Sigma = \text{Ker}(\pi) \oplus \text{Vect}(\mathbf{1}). \quad (2.3)$$

*Démonstration.*

Procédons par analyse-synthèse :

- Analyse : Soit  $f \in \mathbb{C}^\Sigma$  et supposons qu'il existe  $g \in \text{Ker}(\pi)$  et  $h \in \text{Vect}(\mathbf{1})$  tel que  $f = g + h$ . Alors  $\pi(f) = \pi(h)$ , donc  $h = \pi(f)\mathbf{1}$ .

- Synthèse : Soit  $f \in \mathbb{C}^\Sigma$ , posons  $h = \pi(f)\mathbf{1}$  et  $g = f - h$ . On a bien  $h \in \text{Vect}(\mathbf{1})$  et :

$$\begin{aligned} \pi(g) &= \pi(f) - \pi(h) \\ &= 0 \end{aligned}$$

$\square$

### Proposition 2.6.2.

*Soit  $A$  une matrice carrée complexe de rayon spectral  $\rho(A)$ . Alors pour  $\|\cdot\|$  une norme matricielle quelconque :*

$$\forall \epsilon > 0, \exists C > 0, \forall n \in \mathbb{N}, \|A^n\| \leq C(\rho(A) + \epsilon)^n$$

Cette proposition s'appuie principalement sur le théorème suivant, dont une démonstration peut être trouvée dans [6]

**Théorème 2.6.3.** (*Householder*)

Soit  $A$  une matrice carrée à coefficients complexes, si l'on désigne par  $\mathcal{N}$  l'ensemble des normes matricielles, alors

$$\rho(A) = \inf_{\|\cdot\| \in \mathcal{N}} \|A\|.$$

*Démonstration.* (De la proposition) Soit  $\|\cdot\|$  une norme matricielle. Soit  $\epsilon > 0$ , soit  $N$  une norme donnée par le théorème de Householder, vérifiant  $N(A) \leq \rho(A) + \epsilon$ . Comme en dimension finie les normes sont équivalentes, il existe une constante  $C$  telle que  $\|\cdot\| \leq CN$ . On en déduit

$$\forall n \in \mathbb{N}, \|A^n\| \leq CN(A^n) \leq CN(A)^n \leq C(\rho(A) + \epsilon)^n$$

□

Appliquons cette proposition aux matrices stochastiques : en reprenant les notations précédentes, et si l'on suppose en outre que  $P$  est primitive, on sait avec le théorème de Perron-Frobenius que le rayon spectral de  $P$  est 1 et que 1 est valeur propre simple.

De plus,  $\text{Sp}(P) \setminus \{1\}$  ne comporte que des nombres strictement inférieurs à 1 en module. On note  $A = P|_{Ker(\pi)}^{Ker(\pi)}$  la restriction et corestriction de  $P$  à  $Ker(\pi)$ . On pose  $\lambda = \rho(A) < 1$ .

Soit  $\epsilon > 0$  tel que  $\rho(A) + \epsilon > 1$ . 'après la proposition précédente, il existe  $C > 0$  tel que

$$\forall f \in Ker(\pi), \forall n \in \mathbb{N}, \|P^n f\| \leq C(\lambda + \epsilon)^n \|f\| \quad (2.4)$$

Donc  $P^n f$  converge de manière exponentielle vers (0). On en déduit la

**Proposition 2.6.4.**

Soit  $P$  une matrice stochastique primitive. Quelque soit la loi initiale  $\mu_0$  choisit pour définir une chaîne de Markov, la suite  $(\mu_n)_{n \in \mathbb{N}}$  converge à vitesse exponentielle vers l'unique mesure stationnaire  $\pi$  de  $P$ .

**Lemme 2.6.5.**

Soit  $(\mu_n)_{n \in \mathbb{N}}$  une suite de mesures de probabilités et  $\mu$  une mesure de probabilité. On a l'équivalence

$$\mu_n \xrightarrow{n \rightarrow \infty} \mu \iff \forall f \in \Sigma^{\mathbb{C}}, (\mu_n(f))_{n \in \mathbb{N}} \xrightarrow{n \rightarrow \infty} \mu(f)$$



*Démonstration.*

Rappelons que la norme naturelle pour les mesures (signées) est la "norme en variation totale", que l'on peut identifier à la norme duale sur les formes linéaires de  $\Sigma$  dans  $\mathbb{C}$  :

$$\|\nu\| := \sum_{\sigma \in \Sigma} |\nu(\sigma)|$$

( $\Rightarrow$ ) Soit  $f \in \mathbb{C}^\Sigma$  et  $n \in \mathbb{N}$ ,

$$\begin{aligned} |(\mu_n - \mu)(f)| &= \left| \sum_{\sigma \in \Sigma} (\mu_n(\sigma) - \mu(\sigma))f(\sigma) \right| \\ &\leq \sum_{\sigma \in \Sigma} (|\mu_n(\sigma) - \mu(\sigma)|) |f(\sigma)| \\ &\leq \|\mu_n - \mu\| \sum_{\sigma \in \Sigma} |f(\sigma)| \end{aligned}$$

Comme  $\Sigma$  est fini, on en déduit le sens direct.

( $\Leftarrow$ ) On considère la famille de fonctions  $(\mathbf{1}_\sigma)_{\sigma \in \Sigma}$ . On remarque alors qu'une mesure de probabilité  $\nu$  se réécrit

$$\nu = (\nu(\mathbf{1}_\sigma))_{\sigma \in \Sigma}$$

Soit donc  $n \in \mathbb{N}$ ,

$$\begin{aligned} \|\mu_n - \mu\| &= \sum_{\sigma \in \Sigma} |\mu_n(\sigma) - \mu(\sigma)| \\ &= \sum_{\sigma \in \Sigma} |(\mu_n - \mu)(\mathbf{1}_\sigma)| \end{aligned}$$

Mais comme par hypothèse,  $(\mu_n - \mu)(\mathbf{1}_\sigma) \xrightarrow{n \rightarrow \infty} 0$  et qu'on travaille sur  $\Sigma$  fini, on en déduit le sens indirect  $\square$

*Démonstration de la proposition 2.6.4 :* On montre en vertu du lemme précédent que  $\forall f \in \mathbb{C}^\Sigma$ ,  $\mu_n(f) \xrightarrow{n \rightarrow \infty} \mu(f)$  : Soit  $f \in \mathbb{C}^\Sigma$ , on décompose  $f$  en somme d'un unique élément de  $\text{Vect}(\mathbf{1})$  et d'un unique élément  $K$  de  $\text{Ker}(\pi)$  : on a

$$f = \pi(f)\mathbf{1} + (f - \pi(f)\mathbf{1}) = \pi(f)\mathbf{1} + K$$

On réécrit alors pour  $n \in \mathbb{N}$

$$\begin{aligned} \mu_n(f) &= \mu_n(\pi(f)\mathbf{1}) + \mu_n(K) \\ &= \pi(f) * \mu_0 P^n \mathbf{1} + \mu_0 P^n(K) \\ &= \pi(f) + r_n \end{aligned}$$

On aimerait donc montrer que  $r_n \xrightarrow{n \rightarrow \infty} 0$ . En considérant comme norme pour l'espace des mesures  $\|\cdot\|$  la "norme en variation totale", que l'on peut identifier à la norme duale sur les formes linéaires de  $\mathbb{C}$  dans  $\Sigma$ , et comme norme pour l'espace des fonctions la norme de la convergence uniforme définie pour une fonction  $f$  par  $\|f\| = \sup_{\sigma \in \Sigma} |f(\sigma)|$ , on remarque par identification aux cas des espaces vectoriels normés que pour une mesure  $\mu$  et une fonction  $f$ ,

$$|\mu(f)| \leq \|\mu\| \|f\|$$

Avec les remarques préliminaires, on en déduit que pour  $n \in \mathbb{N}$ ,

$$|r_n| = |\mu_n(f)| \leq \lambda^n \|\mu_0\| \|f\|$$

Donc  $|r_n| = |\pi_n(f) - \pi(f)|$  converge de manière exponentielle vers 0. □

**Corollaire 2.6.6.**

Soit  $(\mu_0, P)$  définissant une chaîne de Markov avec  $P$  primitive. soit  $i, j \in \Sigma$ , on a

$$P(X_0 = i, X_n = j) \xrightarrow{n \rightarrow \infty} \mu_0(i)\pi(j) \tag{2.5}$$

.

*Démonstration.*

Soit  $n \in \mathbb{N}$ . On a en vertu du théorème 1.2.1

$$\begin{aligned} P(X_0 = i, X_n = j) &= \mathbb{E}(\mathbf{1}_i(X_0)\mathbf{1}_j(X_n)) \\ &= \mu_0(\mathbf{1}_i \cdot P^n \mathbf{1}_j). \end{aligned}$$

Avec le théorème précédent, on a

$$P^n \mathbf{1}_j \xrightarrow{n \rightarrow \infty} (\pi \mathbf{1}_j) \mathbf{1}.$$

Or comme la multiplication de fonction est bilinéaire continue et que l'espérance est aussi linéaire continue, on en déduit que

$$P(X_0 = i, X_n = j) \xrightarrow{n \rightarrow \infty} \pi(j) \mathbb{E}(\mathbf{1}_i \cdot \mathbf{1}) = \pi(j)\mu_0(i).$$

□

Comme annoncé au départ, dans le cas primitif, quand  $n$  est grand, alors  $X_0$  et  $X_n$  deviennent "de plus en plus indépendant".

# Chapitre 3

## Théorème central limite

### 3.1 Le cas des v.a. indépendantes

Avant d'étudier le théorème central limite pour les chaînes de Markov, étudions d'abord ce théorème dans un cas particulier : les variables aléatoires indépendantes et identiquement distribuées (i.i.d.).

#### 3.1.1 Loi des grands nombres

**Théorème 3.1.1.** (*Loi faible des grands nombres, version 1*)

Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires réelles indépendantes de même loi et de variance finie. Posons :

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k$$

Alors pour tout  $\varepsilon > 0$  :

$$\mathbb{P}(|Y_n - \mathbb{E}(X_1)| > \varepsilon) \rightarrow 0.$$

*Remarque 3.1.2.*

L'hypothèse de la variance finie est utile lorsque l'on utilise l'inégalité de Bienaymé-Tchebychev. On peut cependant l'enlever, ce qu'on montrera avec la proposition 3.1.9 grâce au théorème de Lévy.

*Démonstration.*

Par linéarité de l'espérance, on trouve que  $\mathbb{E}(Y_n) = \mathbb{E}(X_1)$ . De plus, par les propriétés de la variance ( $V(aX) = A^2V(X)$ ), on obtient  $V(Y_n) = V(X_1)/n$ . On applique alors l'inégalité de Bienaymé-Tchebychev :

$$\forall n \in \mathbb{N}^* \quad \mathbb{P}(|Y_n - \mathbb{E}(X_1)| > \varepsilon) \leq \frac{V(X)}{n \varepsilon^2}. \quad \square$$

Le théorème précédent est appelé la loi faible des grands nombres. Sans démonstration, on énoncera sa version forte, mais tout d'abord rappelons la définition de convergence presque sûre :

**Définition 3.1.3.**

Soit  $(X_n)$  une suite de variables aléatoires réelles et  $X$  une variable aléatoire réelle, définies sur  $(\Omega, \mathcal{F}, \mathbb{P})$ . On dit que converge  $X_n$  *presque sûrement* vers  $X$  lorsque :

$$\mathbb{P}(\{\omega \in \Omega \mid X_n(\omega) \rightarrow X(\omega)\}) = 1.$$

**Théorème 3.1.4.** (*Loi forte des grands nombres*)

Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires réelles indépendantes intégrables de même loi. Par les mêmes notations que le théorème précédent, on obtient :

$$Y_n \longrightarrow \mathbb{E}(X_1) \text{ presque sûrement.}$$

### 3.1.2 Théorème central limite

**Définition 3.1.5.**

Soit  $(X_n)$  une suite de variables aléatoires indépendantes et de même loi. On dit que  $(X_n)$  converge en loi vers  $X$  si pour toutes fonction continue bornée  $\varphi$  sur  $E$  à valeurs réelles :

$$\mathbb{E}(\varphi(X_n)) \rightarrow \mathbb{E}(\varphi(X)).$$

**Définition 3.1.6.**

Soit  $X$  une variable aléatoire réelle. On définit alors la fonction caractéristique  $\varphi_X$  de  $X$  :

$$\forall t \in \mathbb{R}, \varphi_X(t) = \mathbb{E}(e^{itX}).$$

Le théorème central limite énonce une convergence en loi d'une certaine variable aléatoire, c'est pour cela que le théorème suivant est crucial pour les démonstrations du théorème central limite :

**Théorème 3.1.7.** (*Théorème de Lévy*)

Soit  $(X_n)$  une suite de variables aléatoires réelles indépendantes et de même loi et  $X$  une variable aléatoire réelle. Il y'a équivalence entre les 2 assertions suivantes :

1.  $(X_n)$  converge en loi vers  $X$
2.  $\varphi_{X_n}$  converge simplement vers  $\varphi_X$

*Remarque 3.1.8.*

Pour la démonstration, on ramène le lecteur à [4, théorème 10.3.4].

Une application du théorème de Lévy est une version plus forte de la loi faible des grands nombres

**Théorème 3.1.9.** (*Loi faible des grands nombres, version 2*)

Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires réelles indépendantes de même loi et d'espérance  $\mu$  finie. Posons :

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Alors pour tout  $\varepsilon > 0$  :

$$\mathbb{P}(|Y_n - \mathbb{E}(X_1)| > \varepsilon) \rightarrow 0.$$

*Démonstration.*

Par le calcul : pour tout  $t \in \mathbb{R}$ ,

$$\begin{aligned} \varphi_{S_n/n}(t) &= \mathbb{E} \left( e^{it \frac{S_n}{n}} \right) \\ &= \left( \mathbb{E} \left( e^{it \frac{X_1}{n}} \right) \right)^n \\ &= \left( \varphi_{X_1} \left( \frac{t}{n} \right) \right)^n. \end{aligned}$$

On peut développer la fonction caractéristique par Taylor-Young à l'ordre en 1 en 0 :

$$\varphi_{X_1} \left( \frac{t}{n} \right) = 1 + i \frac{\mu t}{n} + o\left(\frac{t}{n}\right).$$

Par propriété de l'exponentielle, on montre donc :

$$\varphi_{S_n/n}(t) \longrightarrow e^{i\mu t} = \varphi_{\mu}(t).$$

On peut donc appliquer le théorème 3.1.7 pour montrer la convergence en loi de  $S_n/n$ . Puisque  $S_n/n$  converge en loi vers une constante  $\mu$ , alors la variable converge en probabilité. □

**Définition 3.1.10.**

Soit  $X$  une variable aléatoire réelle. On dit que  $X$  suit la loi normale d'espérance  $\mu$  et d'écart-type  $\sigma$ , noté  $X \sim \mathcal{N}(\mu, \sigma^2)$  si  $X$  admet comme densité de probabilité la fonction  $f$  définie par :

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right)$$

Avant d'énoncer le théorème central limite dans le cas des variables aléatoires i.i.d., on va montrer quelques lemmes :

**Lemme 3.1.11.**

Soit  $X$  une variable aléatoire qui suit la loi gaussienne  $\mathcal{N}(0, 1)$ . Alors, pour tout  $\xi \in \mathbb{R}$  :

$$\varphi_X(\xi) = \exp\left(-\frac{\xi^2}{2}\right).$$

*Démonstration.*

Par définition des fonctions caractéristiques :

$$\varphi_X(\xi) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} + i\xi x\right) dx$$

Par argument de parité, on a  $\Im(\varphi_X) = 0$ . On obtient donc :

$$\varphi_X(\xi) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(\xi x) dx$$

Grâce à l'inégalité  $|x \sin(\xi x) e^{-x^2/2}| \leq |x| e^{-x^2/2}$ , on peut dériver sous le signe intégrale :

$$\varphi'_X(\xi) = - \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} \sin(\xi x) dx.$$

En intégrant par partie, on a :

$$\varphi'_X(\xi) = - \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \xi \cos(\xi x) dx = -\xi \varphi_X(\xi).$$

Donc  $\varphi_X$  est solution de l'équation différentielle  $f' + \xi f = 0$  avec  $f(0) = 1$ . On a bien :

$$\varphi_X(\xi) = \exp\left(-\frac{\xi^2}{2}\right).$$

□

**Lemme 3.1.12.**

Soit  $(z_n)_{n \in \mathbb{N}}$  une suite de nombres complexes convergeant vers  $\alpha$ , alors :

$$\left(1 + \frac{z_n}{n}\right)^n \rightarrow e^\alpha.$$

*Démonstration.*

Soit  $N \in \mathbb{N}$  tel que :

$$\forall n \geq N, \left| \frac{z_n}{n} \right| < 1$$

On peut donc en déduire par propriété de l'exponentielle et du logarithme : soit  $n \geq N$ ,

$$\begin{aligned} \left(1 + \frac{z_n}{n}\right)^n &= \exp\left(n \ln\left(1 + \frac{z_n}{n}\right)\right) \\ &= \exp\left(n\left(\frac{z_n}{n} + o\left(\frac{z_n}{n}\right)\right)\right) \\ &= \exp(z_n + o(z_n)). \end{aligned} \quad \square$$

**Théorème 3.1.13.** (*Théorème central limite*)

Soit  $(X_n)$  une suite de variables aléatoires réelles indépendantes, de carrés intégrables et de même loi. On note  $\mu = \mathbb{E}(X_1)$  et  $\sigma^2 = V(X_1)$ . Posons :

$$S_n = \sum_{k=1}^n X_k$$

De plus, posons :

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

La suite  $(Z_n)$  converge en loi vers  $X \sim \mathcal{N}(0, 1)$ .

*Démonstration.*

Supposons que  $X_i$  est une variable centrée réduite, quitte à la remplacer par  $(X_i - \mu)/\sigma$ . On va alors calculer la fonction caractéristique de  $S_n/\sqrt{n}$  :

$$\begin{aligned} \varphi_{S_n/\sqrt{n}}(t) &= \mathbb{E}\left(e^{it\frac{S_n}{\sqrt{n}}}\right) \\ &= \left(\mathbb{E}\left(e^{it\frac{X_1}{\sqrt{n}}}\right)\right)^n \text{ car les } X_i \text{ sont indépendantes} \\ &= \left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n. \end{aligned}$$

Puisque les  $X_i$  sont de carrés intégrables, on peut faire un développement de Taylor-Young à l'ordre 2 de  $\varphi_{X_1}$  au voisinage de 0 :

$$\varphi_{X_1}(t) = 1 - \frac{t^2}{2} + o(t^2) \text{ lorsque } t \rightarrow 0.$$

On obtient donc par le calcul :

$$\varphi_{S_n/\sqrt{n}}(t/\sqrt{n}) = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n$$

Donc par le lemme 3.1.12, on obtient :

$$\varphi_{S_n/\sqrt{n}}(t) \longrightarrow e^{-\frac{t^2}{2}}$$

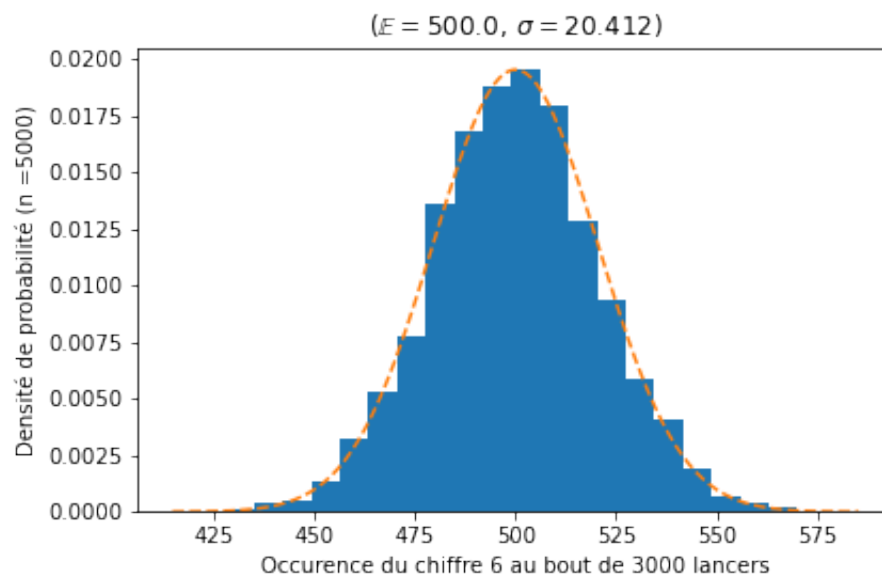
On a bien montré que  $\varphi_{S_n/\sqrt{n}} \rightarrow \varphi_{\mathcal{N}(0,1)}$  grâce au lemme 3.1.11. On applique alors le théorème 3.1.7 pour montrer la convergence en loi.

□

### 3.1.3 Expérimentation

Dans cette section, nous allons illustrer la loi des grands nombres et le théorème central limite dans le cas des variables aléatoires i.i.d. . Toutes les images ont été obtenues grâce au module `matplotlib.pyplot` de Python.

Nous allons prendre pour exemple des lancers de dé. On prend alors un dé normal à 6 faces et on le lance  $k$  fois. On compte alors le nombre de 6, que l'on a obtenu lors de ces  $k$  lancers, on note le résultat  $S_k$  (*i.e.*  $S_k \sim \mathcal{B}(k, 1/6)$ ), et on répète cette expérience  $n$  fois, avec  $n$  assez grand. Voici les résultats de notre expérience avec notre code Python :



Pour notre modélisation, on a pris  $k = 3000$  et  $n = 5000$ . Le graphe nous montre la proportion de chaque résultat possible. On peut alors voir que les  $S_k$  se concentrent vers 500, qui est la valeur de l'espérance de  $X_1$  comme nous énonçons la loi des grands nombres. De plus, on remarque que cette concentration s'apparente à une gaussienne, que nous avons tracé en pointillés sur le graphe. Cette gaussienne est la prédiction faite par le théorème central limite, avec ici  $\sigma = 20.412$ .

## 3.2 Le cas des chaînes de Markov

Dans toute cette partie, on notera  $F$  une application de  $\Sigma$  dans  $\mathbb{R}$ .



### 3.2.1 Théorème central limite

#### Définition 3.2.1.

Soit  $P$  une matrice stochastique. Pour tout  $w \in \mathbb{R}$  et  $x \in \mathbb{R}^n$ , on définit l'opérateur  $P_w$  défini par la formule suivante :

$$P_w(x) = e^{iwF} \cdot (P(x))$$

De plus, on définit l'application :

$$G = \begin{cases} \mathbb{R} \times \mathbb{R}^{n+1} & \longrightarrow & \mathbb{R}^{n+1} \\ (w, x, \lambda) & \longmapsto & \begin{pmatrix} P_w(x) - \lambda x \\ \pi(x) \end{pmatrix} \end{cases}$$

#### Théorème 3.2.2. (Théorème des fonctions implicites)

Soit  $U$  un ouvert de  $\mathbb{R}^p \times \mathbb{R}^q$  et  $f : (x, y) \in U \mapsto f(x, y) \in \mathbb{R}^q$  une application de classe  $\mathcal{C}^k$ . De plus, soit  $(a, b) \in \mathbb{R}^p \times \mathbb{R}^q$  et posons  $c = f(a, b)$ . Supposons que  $D_y f(a, b) \in GL_q(\mathbb{R})$ , alors il existe un voisinage  $\mathcal{V} \subset \mathbb{R}^p$  un voisinage de  $a$ ,  $\mathcal{W} \subset \mathbb{R}^q$  un voisinage de  $b$  et une application  $\psi : \mathcal{V} \rightarrow \mathcal{W}$  de classe  $\mathcal{C}^k$  tel que  $\mathcal{V} \times \mathcal{W} \subset U$  et :

$$\forall (x, y) \in \mathcal{V} \times \mathcal{W}, f(x, y) = c \Leftrightarrow y = \psi(x).$$

*Remarque 3.2.3.* La démonstration sera omise dans ce mémoire, le lecteur curieux pourra s'orienter vers [2, théorème 8.18].

#### Corollaire 3.2.4.

Il existe  $\mathcal{V} \subset \mathbb{R}$  un voisinage de 0,  $\mathcal{W} \subset \mathbb{R}^{n+1}$  un voisinage de  $(\mathbb{1}, 1)$  et une application  $\psi : \mathcal{V} \mapsto \mathcal{W}$  de classe  $\mathcal{C}^1$  tel que :

$$\forall (w, f, \lambda) \in \mathcal{V} \times \mathcal{W}, G(w, f, \lambda) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \Leftrightarrow (f, \lambda) = \psi(w)$$

*Démonstration.* On veut appliquer le théorème des fonctions implicites à la fonction  $G$ . Pour cela, on va calculer  $DG(w, f, \lambda)$ . On sait que les applications  $f \mapsto P_w(f) - \lambda f$  et  $\pi(\cdot)$  sont linéaires, donc on peut facilement calculer les différentielles suivantes :

$$\begin{aligned} \partial_f G_1(w, f, \lambda) &= P_w - \lambda Id \\ \partial_f G_2(w, f, \lambda) &= \pi \end{aligned}$$

On peut, par le calcul, trouver les autres différentielles :

$$\begin{aligned}\partial_\lambda G_1(w, f, \lambda) &= -f \\ \partial_\lambda G_2(w, f, \lambda) &= 0\end{aligned}$$

donc :

$$DG(w, f, \lambda) = \begin{pmatrix} \partial_w G_1(w, f, \lambda) & \boxed{P_w - \lambda Id} & -f \\ \partial_w G_2(w, f, \lambda) & \boxed{\pi} & 0 \end{pmatrix}.$$

La matrice extraite encadrée est celle qui nous intéresse pour vérifier les hypothèses du théorème des fonctions implicites, que l'on notera  $\partial_2 G(w, f, \lambda)$ .

Par ailleurs, puisque l'on cherche les solutions de  $P_w(f) = \lambda f$ , on va utiliser le résultat suivant :

$$G(0, \mathbb{1}, 1) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Ensuite, on a le calcul suivant :

$$\partial_2 G(0, \mathbb{1}, 1) = \begin{pmatrix} P - Id & -\mathbb{1} \\ \pi & 0 \end{pmatrix} = M.$$

On veut vérifier que la matrice obtenue est inversible. Pour cela, on prend  $\begin{pmatrix} h \\ \eta \end{pmatrix}$  un élément de  $\text{Ker}(M)$ . Alors :

$$M \begin{pmatrix} h \\ \eta \end{pmatrix} = \begin{pmatrix} Ph - h - \eta \mathbb{1} \\ \pi(h) \end{pmatrix}.$$

Puisque le vecteur est un élément du noyau de  $M$ , on remarque que :

$$\begin{aligned}\pi(Ph - h - \eta \mathbb{1}) &= \pi(Ph) - \pi(h) - \eta \pi(\mathbb{1}) \\ &= \pi(Ph) - \eta \text{ car } \pi(h) = 0 \\ &= (\pi P)h - \eta \\ &= \pi(h) - \eta \\ &= -\eta.\end{aligned}$$

Donc on a bien  $\eta = 0$ , cela implique que  $Ph = h$ . Or, puisque 1 est valeur propre simple de  $P$ , on sait qu'il existe un unique vecteur de  $E_1(P)$  tel que  $\pi(h) = 0$  et c'est le vecteur nul. On en conclut que  $\text{Ker}(M) = \{0\}$ , on peut donc appliquer le théorème des fonctions implicites.  $\square$

**Théorème 3.2.5.** (Théorème central limite pour les chaînes de Markov)

Soit  $(X_n)$  une chaîne de Markov et supposons que  $\pi F = 0$ . On note :

$$\sigma^2(F) = \pi(F^2) + 2 \sum_{k=1}^{\infty} \pi(F \cdot P^k F)$$

Posons :

$$S_n(F) = \sum_{k=0}^{n-1} F(X_k).$$

La suite  $(S_n(F))$  converge en loi vers  $X \sim \mathcal{N}(0, \sigma^2(F))$ .

*Démonstration.* De la même manière que la démonstration du TCL pour les variables aléatoires réelles i.i.d., on fait intervenir les fonctions caractéristiques : soit  $w \in \mathbb{R}$ ,

$$\begin{aligned} \varphi_{S_n(F)/\sqrt{n}}(w) &= \mathbb{E} \left( \exp \left( i \frac{S_n(F)}{\sqrt{n}} w \right) \right) \\ &= \mathbb{E} \left( \prod_{k=0}^{n-1} \exp \left( i \frac{F}{\sqrt{n}} w \right) (X_k) \right) \\ &= \mu_0 \left( \exp \left( i \frac{F}{\sqrt{n}} w \right) \cdot P \left( \dots \left( \exp \left( i \frac{F}{\sqrt{n}} w \right) \cdot P \mathbf{1} \right) \dots \right) \right) \text{ grâce à 1.2.1} \\ &= \mu_0 P_{w/\sqrt{n}}^n(\mathbf{1}). \end{aligned}$$

Par 3.2.4, il existe les fonctions  $w \mapsto f_w$  et  $w \mapsto \lambda_w$  définies sur  $\mathcal{V}$ , un voisinage de 0 tel que  $\forall w \in \mathcal{V} P_w f_w = \lambda_w f_w$  et  $\pi(f_w) = 1$ . Par un argument de transposité, on peut de même définir  $w \mapsto \pi_w$  tel que  $\pi_w P_w = \lambda_w \pi_w$  et  $\pi_w(f_w) = 1$ . Posons, pour tout  $w \in \mathcal{V}$  :  $P_w = \lambda_w f_w \pi_w + Q_w$ . Posons, pour tout  $n \in \mathbb{N}$ ,  $P_n : "P_w^n = \lambda_w^n f_w \pi_w + Q_w^n"$ . On remarque que  $P_0$  et  $P_1$  est vérifiée. Supposons que  $P_n$  soit vérifiée, alors :

$$\begin{aligned} P_w^{n+1} &= (\lambda_w f_w \pi_w + Q_w)(\lambda_w^n f_w \pi_w + Q_w^n) \\ &= \lambda_w^{n+1} f_w \pi_w f_w \pi_w + \lambda_w f_w \pi_w Q_w^n \\ &\quad + \lambda_w^n Q_w f_w \pi_w + Q_w^{n+1} \end{aligned}$$

Or, on remarque que  $\pi_w Q_w = \pi_w(P_w - \lambda_w f_w \pi_w) = 0$ . De même pour  $Q_w f_w = 0$ . Donc  $P_{n+1}$  est vérifiée.

D'un autre côté, remarquons que, puisque  $\text{Sp}(Q_0) < 1$ , alors  $Q_0^n$  tend vers la matrice nulle. On en déduit l'existence d'un  $n \in \mathbb{N}$  et  $\rho' < 1$  tel que  $\|Q_0^n\| \leq \rho'^n$ . Puisque l'application  $w \mapsto Q_w$  est  $\mathcal{C}^\infty$ , alors il existe  $\rho'' < 1$  tel que, pour tout  $w$  dans un certain voisinage de 0 :  $\|Q_w^n\| \leq \rho''^n$ . Désormais, soit  $N \in \mathbb{N}$  et posons

$N = np + r$  sa division euclidienne. Par équivalence des normes,  $\|\cdot\|$  sera une norme subordonnée. Alors :

$$\begin{aligned} \|Q_w^N\| &\leq \|Q_w^n\|^p \|Q_w^r\| \\ &\leq C' \rho'^{np} \text{ avec } C' = \max_{0 \leq r < n} \|Q_w^r\| \\ &\leq C \rho'^{mN} \end{aligned}$$

Alors, par le calcul :

$$\begin{aligned} \varphi_{S_n(F)/\sqrt{n}}(w) &= \mu_0 P_{w/\sqrt{n}}^n(\mathbf{1}) \\ &= \mu_0 \lambda_{w/\sqrt{n}}^n f_{w/\sqrt{n}} \pi_{w/\sqrt{n}} \mathbf{1} + \mu_0 Q_{w/\sqrt{n}}^n \mathbf{1} \\ &= \mu_0 \lambda_{w/\sqrt{n}}^n f_{w/\sqrt{n}} \pi_{w/\sqrt{n}} \mathbf{1} + o(\rho'^{mN}) \\ &= \lambda_{w/\sqrt{n}}^n + o(1)(?) \end{aligned}$$

On veut calculer les deux premières dérivées de  $w \mapsto \lambda_w$  en 0. Pour cela, on dérive l'équation  $P_w f_w = \lambda_w f_w$  :

$$P'_w f_w + P_w f'_w = \lambda'_w f_w + \lambda_w f'_w \quad (3.1)$$

Puis on évalue en 0 le résultat précédent :

$$P'_0 \mathbf{1} + P f'_0 = \lambda'_0 \mathbf{1} + f'_0$$

Puis on applique  $\pi$  :

$$\pi P'_0 \mathbf{1} = \lambda'_0$$

Or, par la définition de  $P_w$ , on a  $P'_w = iF e^{iwF} \cdot P$ . Donc  $P'_0(f) = iF \cdot P(f)$  et  $\pi P'_0 \mathbf{1} = i\mathbb{E}(F)$ . Dans la suite, on supposera que  $\mathbb{E}(F) = 0$ . De plus :

$$P'_0 \mathbf{1} + P f'_0 = f'_0 \implies f'_0 = (I - P)^{-1}(iF)$$

$(I - P)$  est inversible si on la considère comme une application de  $\text{Ker}(\pi)$  dans lui-même, qui elle est inversible. On a bien  $iF \in \text{Ker}(\pi)$  par hypothèse sur  $F$  donc  $(I - P)^{-1}(iF)$  existe bien. On veut calculer la dérivée seconde en dérivant l'équation 3.1 :

$$P''_0 \mathbf{1} + 2P'_0 f'_0 + P_0 f''_0 = \lambda''_0 \mathbf{1} + 2P'_0 f'_0 + P_0 f''_0$$

Donc en multipliant par  $\pi$  :

$$-\pi(F^2) - 2\pi F \cdot (P(I - P)^{-1}(F)) = \lambda''_0$$

Or  $(I - P)^{-1} = \sum_{k=0}^{\infty} P^k$ . Donc :

$$\lambda''_0 = -\pi(F^2) - 2 \sum_{k=1}^{\infty} \pi(F \cdot P^k F) = -\sigma^2(F)$$

Donc par formule de Taylor-Young, lorsque  $w$  tend vers 0 :

$$\lambda(w) = 1 - \frac{\sigma^2(F)}{2}w^2 + o(w^2)$$

On peut donc en déduire un développement limité de  $\varphi_{S_n/\sqrt{n}}$  lorsque  $n$  tend vers  $\infty$  :

$$\begin{aligned} \varphi_{S_n(F)/\sqrt{n}}(w) &= \left(1 - \frac{\sigma^2(F)w^2}{2n} + o\left(\frac{w^2}{n}\right)\right)^n \\ &\rightarrow \exp\left(-\frac{\sigma^2(F)w^2}{2}\right) \end{aligned}$$

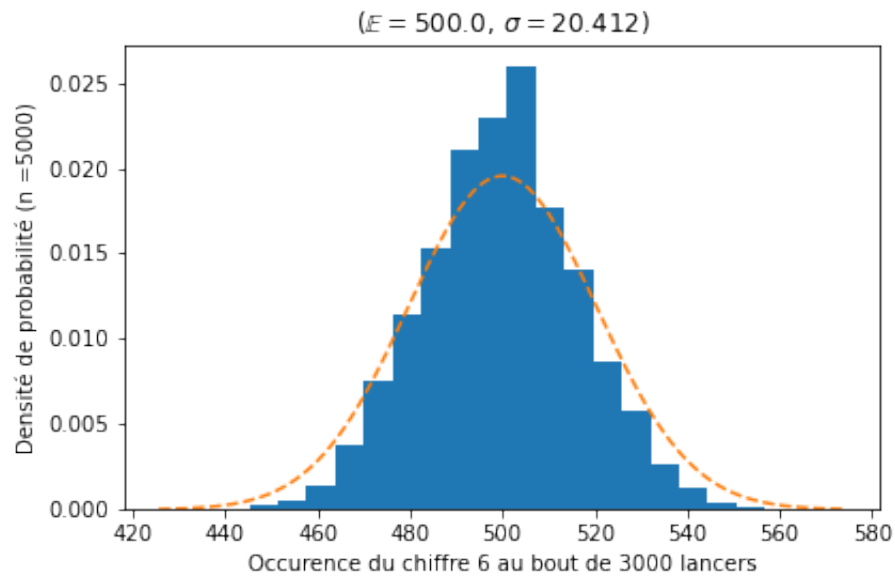
Or le terme de droite est la fonction caractéristique de  $\mathcal{N}(0, \sigma^2(F))$ . Donc par le théorème de Lévy, on a la convergence en loi.  $\square$

*Exemple 3.2.6.* On veut calculer  $\sigma^2(F)$  pour le dé magique. On note  $A$  la matrice associée à la chaîne de Markov du dé magique. Soit  $F$  une fonction tel que  $\pi F = 0$ , on remarque que cela implique  $AF = \mu F$  avec  $\mu = -1/5$ . Donc :

$$\begin{aligned} \sigma^2(F) &= \pi(F^2) + 2 \sum_{n \geq 1} \pi(F \cdot P^n F) \\ &= \pi(F^2) + 2\pi(F^2) \sum_{k \geq 1} \mu^k \\ &= \pi(F^2) \left(\frac{1 + \mu}{1 - \mu}\right) \end{aligned}$$

### 3.2.2 Expérimentation

On veut illustrer, comme on l'a fait pour des variables aléatoires i.i.d., le théorème central limite pour des chaînes de Markov. Ce coup-ci, on lancera des dés magiques. Faisons alors  $k$  lancers dont on notera les résultats dans les variables  $X_1, \dots, X_k$ . On pose aussi  $F$  la fonction  $\mathbb{1}_{X=6}$ . On remarquera que  $F$  ne suit pas les hypothèses du théorème central limite :  $\pi F \neq 0$ . Cette hypothèse est surtout utile pour que la gaussienne limite soit centrée, mais ici ce ne sera pas le cas. En gardant les notations du théorème central limite, on obtient alors une valeur de  $S_k(F)$ . On répète alors  $n$  fois cette expérience pour obtenir plusieurs valeurs de  $S_k(F)$  que l'on affiche sur ce graphe (on gardera  $k = 3000$  et  $n = 5000$ ) :



Les résultats se concentrent une nouvelle fois comme une gaussienne centrée vers 500. La fonction gaussienne tracée en pointillé est celle qui correspond dans le cadre du dé normal. On remarque alors qu'elle ne colle pas bien à l'expérience, il faut donc changer  $\sigma$ . C'est ce que nous donne le théorème. Le programme nous donne un écart-type d'environ 16, contrairement au 20 pour le dé non magique.

# Bibliographie

- [1] Thomas HAWKINS. « Continued fractions and the origins of the Perron-Frobenius theorem ». eng. In : *Archive for history of exact sciences* 62.6 (2008), p. 655-717. ISSN : 0003-9519.
- [2] Dominique HULIN. *Topologie, Calcul différentiel*. 2015.
- [3] Gregory F. LAWLER. *Introduction to Stochastic Processes, 2nd Edition*. 2006.
- [4] Jean-François LE GALL. *Intégration, Probabilités et Processus Aléatoires*. 2006.
- [5] Philippe MALBOS. « Algèbre appliquée, partie 2 : Théorème de Perron-Frobenius et les moteurs d'indexation du web ». In : (2012).
- [6] Denis SERRE. *Matrices : Theory and Applications*. T. 216. Graduate texts in mathematics. Springer New York, 2010. ISBN : 9781441976826.