

Étude du modèle d'Ehrenfest

Ophélie ROUBY

Rapport de stage, 31 août 2010

Table des matières

Introduction	5
Présentation du stage	5
Présentation du rapport	5
Paul EHRENFEST	7
1 Les chaînes de Markov	9
1.1 Processus de Markov	9
1.2 Classification des états	13
1.3 Classification des chaînes	16
1.4 Distribution stationnaire et théorème limite	18
1.4.1 Distribution stationnaire	18
1.4.2 Théorème limite	24
1.5 Réversibilité des chaînes de Markov	26
1.6 Chaînes ayant un nombre fini d'étapes	29
1.7 Bilan	29
2 La théorie des grandes déviations	31
2.1 La théorie des grandes déviations	31
2.1.1 Le principe des grandes déviations	31
2.1.2 Autour du principe des grandes déviations	32
2.1.3 Détermination de la fonction taux	33
2.1.4 Le théorème de Cramér	36
2.1.5 Propriétés de la fonction génératrice des cumulants échelonnés et de la fonction taux	39
2.1.6 Le principe de contraction	42
2.2 Application aux Chaînes de Markov	43
2.3 Bilan	44
3 Le modèle d'Ehrenfest	45
3.1 Description du modèle	45
3.1.1 Description originale donnée par les époux Ehrenfest	45
3.1.2 Description mathématique et physique	45
3.2 Étude probabiliste du modèle	46
3.2.1 Un processus de Markov	46
3.2.2 Temps de récurrence	49
3.2.3 Espérance et Variance	51
3.2.4 Convergence en probabilité	54
3.2.5 Grandes Déviations	57
3.3 Bilan	61
Conclusion	63
A Les probabilités conditionnelles	65
A.1 Probabilité conditionnelle	65
A.2 Distribution conditionnelle et espérance conditionnelle, dans le cas discret	65
A.3 Distribution conditionnelle et espérance conditionnelle, dans le cas continu	66

B Programmes Scilab	69
B.1 Programme 1 :	69
B.2 Programme 2 :	69
B.3 Programme 3 :	69
Bibliographie	71

Introduction

Présentation du stage

Mon stage a été effectué au sein du laboratoire de Probabilités et Modèles aléatoires, rattaché aux Universités de Paris VI et VII, pendant une durée de cinq semaines, sous la direction de Raphaël Lefevre.

Le sujet du stage est : l'étude du modèle d'Ehrenfest.

Présentation du rapport

Le présent rapport se découpe en plusieurs parties. Tout d'abord, nous nous pencherons sur la biographie de Paul Ehrenfest ; elle figure dans ce rapport car il est intéressant d'essayer de connaître un peu l'auteur du dit modèle. Nous considérerons ensuite, dans un premier temps, la théorie des Chaînes de Markov en nous concentrant sur des notions-clés et des théorèmes que nous utiliserons par la suite lors de l'étude du modèle d'Ehrenfest. Dans un second temps, nous aborderons la théorie des Grandes Déviations, ce chapitre ayant pour vocation de nous donner les bases de cette théorie afin de considérer le modèle d'Ehrenfest à travers le prisme de celle-ci. Enfin, nous nous intéresserons à l'étude détaillée du modèle d'Ehrenfest en nous appuyant sur les deux théories vues dans les chapitres précédents, c'est-à-dire en l'observant d'un point de vue probabiliste mais également d'un point de vue physique puisqu'il modélise un phénomène d'une telle nature. Figurent également dans ce rapport deux annexes, l'une sur des notions de base sur les probabilités conditionnelles et l'autre donnant les programmes Scilab écrits afin de simuler la variable aléatoire associée au modèle d'Ehrenfest.

Paul EHRENFEST



Né à Vienne, le 18 janvier 1880.
Mort à Amsterdam, le 25 septembre 1933.

Paul Ehrenfest est issu d'une famille juive et assez pauvre de cinq enfants, dont il est le plus jeune.

Il étudie la physique théorique à Vienne, où il valide son doctorat en 1904 avec pour sujet de thèse : "the extension of Hertz's mechanics to problems in hydrodynamics", thèse supervisée par Boltzman, personne qui influença beaucoup Ehrenfest.

Le 21 décembre 1904, Ehrenfest se marie avec Tatyana Alexeyevna Afanassjewa (une étudiante en mathématiques russe) qu'il rencontre à Göttingen en 1902 au cours de ses études. À cause de la loi austro-hongroise sur le mariage (le mariage d'une catholique et d'un juif n'est possible que si les deux partenaires renoncent à leurs religions), Paul et Tatyana Ehrenfest renoncent aux leurs.

Durant les premières années après leur mariage, Paul et Tatyana Ehrenfest font ensemble une série de publications qui clarifie des obscurités sur la mécanique statistique de Boltzman et de Gibbs. Ils relèvent, entre autres, une erreur fondamentale dans la mécanique statistique de Gibbs. Ils sont par la suite invités par Felix Klein, pour rédiger un paragraphe sur les fondements de la mécanique statistique dans "die Encyklopädie der mathematischen Wissenschaften".

En 1907, Paul Ehrenfest et sa famille s'installent à Saint Petersburg, où il publie plusieurs articles, dont une étude du principe Le Chatelier-Braun et deux sur le principe de la relativité. Dès lors, Paul Ehrenfest part à la recherche d'un poste dans une académie, mais il rencontre des difficultés principalement à cause de son statut religieux. C'est finalement en 1912, qu'on lui propose d'occuper la chaire de physique théorique à Leiden, en tant que successeur de Lorentz.

En 1907, Ehrenfest propose un modèle théorique simple, le modèle des urnes d'Ehrenfest, qui montre comment les lois des probabilités peuvent produire une tendance moyenne vers l'équilibre, alors que le comportement du modèle est réversible dans le temps et que chacun de ses états peut être récurrent. Ce qui a permis de montrer que le H-théorème de Boltzman (montrant que

la collision des molécules produit, lorsqu'on approche de l'équilibre, une augmentation d'entropie) interprété statistiquement, n'est pas nécessairement contradictoire avec les lois réversibles de la mécanique, comme Poincaré, Zermelo ou Loschmidt le pensaient.

Un trait de caractère particulier chez Ehrenfest est sa capacité de critique et de remise en question. L'approche critique d'Ehrenfest l'amène à sa plus grande contribution en physique : le principe adiabatique. Ehrenfest est également l'un des premiers à avoir essayé de comprendre le concept du quanta d'énergie, introduit par Planck en 1900. Son étude majeure : "Wich features of the quantum hypothesis play an essential role in the theory of heat radiation?" donne l'essentiel de la théorie quantique et compte beaucoup de résultats démontrés. Dans les années 1920-1930, il démontre un résultat qui porte aujourd'hui le nom de théorème d'Ehrenfest, en mécanique quantique.

En septembre 1933, Paul Ehrenfest met fin à ses jours.

Chapitre 1

Les chaînes de Markov

Les chaînes de Markov ont été introduites par Andreï Andreïevitch Markov (1856-1922) et appelées ainsi en son honneur. C'est en 1902 que Markov commença son étude sur un nouveau type de processus aléatoire, étant alors perçu comme une extension des suites de variables aléatoires indépendantes. Dans ces processus, l'issue d'une expérience donnée peut affecter l'issue de la prochaine expérience.

1.1 Processus de Markov

Commençons, tout d'abord, par introduire la notion de processus.

On se place sur un ensemble de probabilités $(\Omega, \mathcal{A}, \mathbb{P})$.

On rappelle qu'une variable aléatoire X , à valeurs dans un ensemble mesurable I est une fonction : $X : \Omega \rightarrow I$.

Si on pose $\lambda_i = \mathbb{P}(X = i)$, avec $\lambda = (\lambda_i, i \in I)$ telle que $\sum_{i \in I} \lambda_i \delta_i$ soit une mesure sur l'ensemble I vérifiant $\sum_{i \in I} \lambda_i = 1$, alors λ est appelée la distribution de X .

Un processus aléatoire est une famille $\{X_t, t \in T\}$ de variables aléatoires indexées par un ensemble T , définies sur un même espace de probabilités et à valeurs dans le même ensemble.

Par exemple :

- Si $T = \mathbb{N}$, on parle de processus discret.
- Si $T = \mathbb{R}$, on parle de processus continu dans le temps.

Intéressons-nous, tout d'abord à des processus discrets.

Soit $\{X_0, X_1, \dots\}$ une suite de variables aléatoires à valeurs dans un certain ensemble \mathcal{S} mesurable (pour une mesure donnée) appelé *espace d'états*. Chaque X_n étant une variable aléatoire discrète qui peut prendre l'une des N valeurs possibles, si $N = \text{cardinal}(\mathcal{S})$.

Définition 1.1 *Le processus X est une chaîne de Markov, s'il vérifie la condition de Markov, ie :*

$$\mathbb{P}(X_n = s | X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = s | X_{n-1} = x_{n-1}) \quad (1.1)$$

pour tout $n \geq 1$ et pour tous $s, x_0, x_1, \dots, x_{n-1} \in \mathcal{S}$.

Remarque : La condition de Markov peut aussi s'écrire, pour tout $s \in \mathcal{S}$ et pour toute suite $\{x_i, i \geq 0\}$ à valeurs dans \mathcal{S} :

$$\mathbb{P}(X_{n+1} = s | X_{n_1} = x_{n_1}, X_{n_2} = x_{n_2}, \dots, X_{n_k} = x_{n_k}) = \mathbb{P}(X_{n+1} = s | X_{n_k} = x_{n_k}) \quad (1.2)$$

ou encore, pour tous $m, n \geq 0$:

$$\mathbb{P}(X_{m+n} = s | X_0 = x_0, X_1 = x_1, \dots, X_m = x_m) = \mathbb{P}(X_{m+n} = s | X_m = x_m) \quad (1.3)$$

Définition 1.2 La chaîne X est dite homogène si :

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i)$$

pour tous $n \in \mathbb{N}$ et $i, j \in \mathcal{S}$.

La matrice de transition $\mathcal{P} = (p_{ij})$ est la matrice $|\mathcal{S}| \times |\mathcal{S}|$ de probabilités de transition, ie pour tous $i, j \in \mathcal{S}$, $p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$.

Dans toute la suite, on considèrera que les chaînes de Markov sont homogènes.

Théorème 1.1 La matrice de transition \mathcal{P} est une matrice stochastique, ie :

- (i) $p_{ij} \geq 0$ pour tous $i, j \in \mathcal{S}$.
- (ii) $\sum_j p_{ij} = 1$ pour tout $i \in \mathcal{S}$.

DÉMONSTRATION

Montrons que \mathcal{P} est une matrice stochastique.

Pour tous $i, j \in \mathcal{S}$, on a :

$$p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) \geq 0$$

car une probabilité est toujours positive ou nulle.

Soit $i \in \mathcal{S}$,

$$\begin{aligned} \sum_j p_{ij} &= \sum_j \mathbb{P}(X_{n+1} = j | X_n = i) \\ &= \sum_j \frac{\mathbb{P}(X_{n+1} = j \cap X_n = i)}{\mathbb{P}(X_n = i)} \end{aligned}$$

par définition d'une probabilité conditionnelle.

De plus, on remarque que $\{X_{n+1} = j\}_{j \in \mathcal{S}}$ est une partition disjointe de l'ensemble \mathcal{S} , donc :

$$\mathbb{P}(X_n = i) = \sum_j \mathbb{P}(X_n = i \cap X_{n+1} = j)$$

D'où $\sum_j p_{ij} = 1$.

Remarque : Ce théorème caractérise les matrices de transition. En effet, une matrice stochastique peut être vue comme une matrice de transition car ses coefficients sont compris entre 0 et 1.

Nous nous intéressons ici à l'évolution de X selon deux échelles de temps, le court-terme décrit par la matrice de transition \mathcal{P} et le long-terme décrit par la matrice des n -ièmes probabilités de transition définie plus loin.

Exemple : On considère la matrice suivante :

$$\begin{pmatrix} 0.5 & 0.1 & a \\ b & 0.2 & 0.4 \\ 0.3 & c & 0.1 \end{pmatrix}$$

Déterminons a , b et c afin que notre matrice soit une matrice de probabilités de transition, ie stochastique.

On utilise pour cela le théorème (1.1), on sait que $\sum_{j=1}^3 p_{ij} = 1$, ainsi :

$$\begin{cases} 0.5 + 0.1 + a = 1 \\ b + 0.2 + 0.4 = 1 \\ 0.3 + c + 0.1 = 1 \end{cases}$$

Donc, pour que notre matrice soit une matrice de probabilités de transition, il faut que :

$$\begin{cases} a = 0.4 \\ b = 0.4 \\ c = 0.6 \end{cases}$$

Définition 1.3 La n -ième matrice de transition $\mathcal{P}(m, m+n) = (p_{ij}(m, m+n))$ est la matrice des n -ièmes probabilités de transition, avec : $p_{ij}(m, m+n) = \mathbb{P}(X_{m+n} = j | X_m = i)$ pour tous $i, j \in \mathcal{S}$ et $m, n \in \mathbb{N}$.

Remarque :

- La n -ième matrice de transition est une matrice stochastique.
- $\mathcal{P}(m, m+1) = \mathcal{P}$.
- $\mathcal{P}(m, m+n)$ ne dépend pas de m .

La dernière assertion découle du théorème suivant :

Théorème 1.2 (Les équations de Chapman-Kolmogorov) *Pour tous $i, j \in \mathcal{S}$ et $m, n, r \in \mathbb{N}$, on a :*

$$p_{ij}(m, m+n+r) = \sum_k p_{ik}(m, m+n)p_{kj}(m+n, m+n+r)$$

Ainsi, on a également :

$$\mathcal{P}(m, m+n+r) = \mathcal{P}(m, m+n)\mathcal{P}(m+n, m+n+r)$$

et

$$\mathcal{P}(m, m+n) = \mathcal{P}^n$$

où \mathcal{P} est la matrice de transition.

DÉMONSTRATION

Soient $i, j \in \mathcal{S}$ et $m, n, r \in \mathbb{N}$. Revenons à la définition des coefficients d'une matrice de transition :

$$\begin{aligned} p_{ij}(m, m+n+r) &= \mathbb{P}(X_{m+n+r} = j | X_m = i) \\ &= \sum_k \mathbb{P}(X_{m+n+r} = j \cap X_{m+n} = k | X_m = i) \quad \text{car } \{X_{m+n} = k\}_{k \in \mathcal{S}} \text{ partition disjointe de } \mathcal{S} \\ &= \sum_k \mathbb{P}(X_{m+n+r} = j | X_{m+n} = k \cap X_m = i) \mathbb{P}(X_{m+n} = k | X_m = i) \\ &= \sum_k \mathbb{P}(X_{m+n+r} = j | X_{m+n} = k) \mathbb{P}(X_{m+n} = k | X_m = i) \quad \text{d'après la condition de Markov (1.2)} \\ &= \sum_k p_{ik}(m, m+n)p_{kj}(m+n, m+n+r) \end{aligned}$$

On peut justifier les calculs effectués ci-dessus par la propriété suivante, pour tous événements A, B, C :

$$\mathbb{P}(A \cap B | C) = \mathbb{P}(A | B \cap C) \mathbb{P}(B | C) \quad (1.4)$$

Par conséquent, on obtient bien les équations de Chapman-Kolmogorov.

Quant aux deux autres résultats du théorème, ils découlent immédiatement des équations de Chapman-Kolmogorov.

Remarque La propriété (1.4) utilisée ci-dessus se démontre en utilisant la définition d'une probabilité conditionnelle, ie pour deux événements A et B d'un espace de probabilité : $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Une conséquence de théorème ci-dessus est que : $\mathcal{P}(m, m+n) = \mathcal{P}(0, n)$. Nous écrirons désormais \mathcal{P}_n pour $\mathcal{P}(m, m+n)$ et $p_{ij}(n)$ pour $p_{ij}(m, m+n)$.

L'étude d'une chaîne de Markov peut se réduire à l'étude des propriétés algébriques de sa matrice de transition.

Exemple 1 : On considère deux urnes. Une urne rouge contenant deux boules rouges et trois boules bleues, et une urne bleue contenant une boule rouge et quatre boules bleues. Une boule est choisie dans une urne : on note sa couleur et on la remet dans son urne. La boule suivante est extraite de l'urne dont la couleur coïncide avec celle de la boule tirée à l'étape précédente. L'espace des états correspond à la couleur des urnes dans lesquelles sont effectuées les tirages.

1. La matrice de transition de la chaîne de Markov ainsi définie est :

$$\mathcal{P} = \begin{pmatrix} 4/5 & 1/5 \\ 3/5 & 2/5 \end{pmatrix}$$

en considérant que la première ligne correspond à je choisis une boule dans l'urne bleue, la deuxième ligne à je choisis une boule dans l'urne rouge et la première colonne à je vais (ou reste) dans l'urne bleue, la deuxième colonne à je vais (ou reste) dans l'urne rouge.

2. Déterminons maintenant la probabilité que la troisième boule soit extraite de l'urne bleue, sachant que la première boule est extraite de l'urne rouge.

On souhaite déterminer : $\mathbb{P}(X_3 = \text{bleue} | X_1 = \text{rouge}) = p_{rb}(1, 1 + 2)$, il nous faut donc déterminer la matrice des 2-ièmes probabilités de transition, ie \mathcal{P}^2 .

$$\mathcal{P}^2 = \begin{pmatrix} 19/25 & 6/25 \\ 18/25 & 7/25 \end{pmatrix}$$

Ici la réponse à la question est donnée par le coefficient ligne rouge, colonne bleue, ie $\mathbb{P}(X_3 = \text{bleue} | X_1 = \text{rouge}) = \frac{18}{25}$.

On aurait pu également traiter cet exercice en faisant un arbre.

Exemple 2 : On considère cinq urnes numérotées de un à cinq et disposées dans l'ordre croissant. L'urne numéro i contient i boules noires et $5 - i$ boules blanches. Lorsqu'on tire au hasard une boule dans l'urne numéro i , on note sa couleur et on la remet dans son urne :

- si la boule obtenue était noire, le tirage suivant se fait dans l'urne $i + 1$;
- si la boule était blanche, le tirage suivant se fait à nouveau dans l'urne i .

Après avoir extrait une boule dans l'urne numéro 5, on recommence à l'urne numéro 1.

Soit $(X_n)_{n \in \mathbb{N}}$, la variable indiquant le numéro de l'urne dans laquelle se fait le n -ième tirage.

1. Déterminons la matrice de transition de la chaîne de Markov correspondante.

L'espace d'états ici est : $\mathcal{S} = \{1, 2, 3, 4, 5\}$ et la matrice de transition :

$$\mathcal{P} = \begin{pmatrix} 4/5 & 1/5 & 0 & 0 & 0 \\ 0 & 3/5 & 2/5 & 0 & 0 \\ 0 & 0 & 2/5 & 3/5 & 0 \\ 0 & 0 & 0 & 1/5 & 4/5 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

2. Sachant que l'on part de l'urne numéro 1, calculons la probabilité que le 3-ième tirage se fasse dans l'urne 1, puis 2, puis 3.

Pour cela, il suffit de déterminer la matrice des 2-ièmes probabilités de transition :

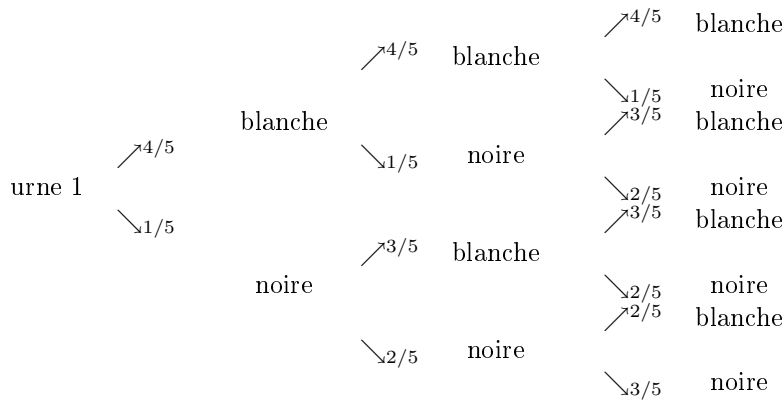
$$\mathcal{P}^2 = \begin{pmatrix} 16/25 & 7/25 & 2/5 & 0 & 0 \\ 0 & 9/25 & 2/5 & 6/25 & 0 \\ 0 & 0 & 4/25 & 9/25 & 12/25 \\ 4/5 & 0 & 0 & 1/25 & 4/25 \\ 4/5 & 1/5 & 0 & 0 & 0 \end{pmatrix}$$

Donc :

$$\begin{cases} \mathbb{P}(X_3 = 1 | X_1 = 1) = 16/25 \\ \mathbb{P}(X_3 = 2 | X_1 = 1) = 7/25 \\ \mathbb{P}(X_3 = 3 | X_1 = 1) = 2/5 \end{cases}$$

3. Déterminons, pour finir, la probabilité que la troisième boule tirée soit noire, sachant que l'on commence le tirage dans l'urne 1.

Pour ce faire, construisons un arbre :



On en déduit la probabilité \mathbb{P} d'avoir une boule noire au troisième tirage :

$$\mathbb{P} = \binom{4}{5}^2 \frac{1}{5} + \frac{4}{5} \frac{1}{5} \frac{2}{5} + \frac{1}{5} \frac{3}{5} \frac{2}{5} + \frac{1}{5} \frac{2}{5} \frac{3}{5} = \frac{36}{125}$$

1.2 Classification des états

Définition 1.4 Soit $i \in \mathcal{S}$, \mathcal{S} étant l'ensemble des états.

L'état i est dit persistant (ou récurrent) si pour un certain $n \geq 1$, on a : $\mathbb{P}(X_n = i | X_0 = i) = 1$, ce qui signifie que la probabilité d'un éventuel retour à l'état i est de 1, sachant que le processus a commencé à l'état i .

Si cette probabilité est strictement inférieure à 1, on parle d'état passager.

Une chaîne de Markov est dite persistante, si elle admet au moins un état persistant et si de chaque état, il est possible d'aller à cet état persistant (pas nécessairement en une étape).

On pose pour $i, j \in \mathcal{S}$ et $n \geq 1$: $f_{ij}(n) = \mathbb{P}(X_1 \neq j \cap X_2 \neq j \cap \dots \cap X_{n-1} \neq j \cap X_n = j | X_0 = i)$ la probabilité d'une première visite à l'état j , à la n -ième étape, sachant que le processus a commencé à l'état i .

On pose aussi :

$$f_{ij} = \sum_{n=1}^{+\infty} f_{ij}(n)$$

la probabilité que la chaîne atteigne l'état j sachant qu'elle est partie de l'état i .

On a ainsi que j est un état persistant si $f_{jj} = 1$.

Le but est maintenant de déterminer un critère de persistance pour les n -ièmes probabilités de transition. On définit pour ce faire, les deux fonctions génératrices suivantes :

$$P_{ij}(s) = \sum_{n=0}^{+\infty} s^n p_{ij}(n)$$

$$F_{ij}(s) = \sum_{n=0}^{+\infty} s^n f_{ij}(n)$$

avec $p_{ij}(0) = \delta_{ij}$ et $f_{ij}(0) = 0$, pour tous $i, j \in \mathcal{S}$.

On a bien entendu : $f_{ij} = F_{ij}(1)$.

On considèrera que $|s| < 1$, afin que les fonctions génératrices soient bien définies.

Théorème 1.3 Pour tous $i, j \in \mathcal{S}$, on a :

(i) $P_{ii}(s) = 1 + F_{ii}(s)P_{ii}(s)$.

(ii) $P_{ij}(s) = F_{ij}(s)P_{jj}(s)$ si $i \neq j$.

DÉMONSTRATION

Soient $i, j \in \mathcal{S}$ fixés.

Soient $A_m = \{X_m = j\}$ et $B_m = \{X_r \neq j, 1 \leq r < m \cap X_m = j\}$. Les B_m sont disjoints, ainsi on peut écrire :

$$\mathbb{P}(A_m | X_0 = i) = \sum_{r=1}^m \mathbb{P}(A_m \cap B_r | X_0 = i)$$

On utilise maintenant la propriété (1.4), puis la condition de Markov :

$$\begin{aligned}\mathbb{P}(A_m \cap B_r | X_0 = i) &= \mathbb{P}(A_m | B_r \cap X_0 = i) \mathbb{P}(B_r | X_0 = i) \\ &= \mathbb{P}(A_m | X_r = j) \mathbb{P}(B_r | X_0 = i)\end{aligned}$$

D'où

$$p_{ij}(m) = \sum_{r=1}^m f_{ij}(r) p_{jj}(m-r)$$

On multiplie cette égalité par s^m et pour $N > 0$, on considère :

$$\begin{aligned}\sum_{m=1}^N s^m p_{ij}(m) &= \sum_{m=1}^N s^m \sum_{r=1}^m f_{ij}(r) p_{jj}(m-r) \\ &= \sum_{r=1}^N \sum_{m=r}^N s^m s^r s^{-r} f_{ij}(r) p_{jj}(m-r) \\ &= \sum_{r=1}^N s^r f_{ij}(r) \sum_{m=r}^N s^{m-r} p_{jj}(m-r) \\ &= \sum_{r=1}^N s^r f_{ij}(r) \sum_{m=0}^{N-r} s^m p_{jj}(m) \\ &= \sum_{r=0}^N s^r f_{ij}(r) \sum_{m=0}^{N-r} s^m p_{jj}(m) \quad \text{car } f_{ij}(0) = 0\end{aligned}$$

Donc :

$$\sum_{m=0}^N s^m p_{ij}(m) - \delta_{ij} = \sum_{r=0}^N s^r f_{ij}(r) \sum_{m=0}^{N-r} s^m p_{jj}(m) \quad \text{car } p_{ij}(0) = \delta_{ij}$$

On passe à la limite quand $N \rightarrow \infty$ et on obtient le résultat :

$$P_{ij}(s) - \delta_{ij} = F_{ij}(s) P_{jj}(s)$$

Corollaire 1.3.1 Soient $j \in \mathcal{S}$ et $n \in \mathbb{N}$.

- (i) L'état j est dit persistant si $\sum_n p_{jj}(n) = \infty$ et dans ce cas, on a $\sum_n p_{ij}(n) = \infty$ pour tout $i \in \mathcal{S}$ tel que $f_{ij} > 0$.
- (ii) L'état j est dit passager si $\sum_n p_{jj}(n) < \infty$ et dans ce cas, on a $\sum_n p_{ij}(n) < \infty$ pour tout $i \in \mathcal{S}$.

DÉMONSTRATION

Montrons que j est persistant si et seulement si $\sum_n p_{jj}(n) = \infty$.

D'après l'assertion (i) du théorème (1.3), on a pour s , tel que $|s| < 1$:

$$P_{jj}(s) = 1 + F_{jj}(s) P_{jj}(s)$$

D'où :

$$\begin{aligned}P_{jj}(s)(F_{jj}(s) - 1) &= -1 \\ P_{jj}(s) &= \frac{1}{1 - F_{jj}(s)}\end{aligned}$$

Ainsi si $s \rightarrow 1$, $P_{jj}(s) \rightarrow \infty \Leftrightarrow f_{jj} = F_{jj}(1) = 1$.

Un théorème d'Abel (voir [GS01]) nous donne alors :

$$\lim_{s \nearrow 1} P_{jj}(s) = \sum_n p_{jj}(n)$$

Par conséquent, comme nécessairement $F_{jj}(1) = 1$ car j est passager, on a montré que : j est passager si et seulement si $\sum_n p_{jj}(n) = \infty$.

De plus, d'après l'assertion (ii) du théorème (1.3), on a pour tout $i \in \mathcal{S}$ tel que $f_{ij} > 0$:

$$\lim_{s \nearrow 1} P_{ij}(s) = \sum_n p_{ij}(n) = \infty$$

On a ainsi démontré le premier point du corollaire, le deuxième point en découle.

Corollaire 1.3.2 Soit $j \in \mathcal{S}$.

Si j est un état passager, alors $p_{ij}(n) \rightarrow 0$, quand $n \rightarrow \infty$, $\forall i \in \mathcal{S}$.

DÉMONSTRATION

L'assertion (ii) du corollaire (1.3.1) nous livre le résultat, car le terme général d'une série convergente, converge vers 0.

Remarque : Un état est soit persistant, soit passager.

Soit $N(i)$ le nombre de fois qu'une chaîne visite son point de départ i , on a alors :

$$\mathbb{P}(N(i) = \infty) = \begin{cases} 1 & \text{si } i \text{ est persistant} \\ 0 & \text{si } i \text{ est passager} \end{cases}$$

De plus le retour à l'état i est assuré $\Leftrightarrow f_{ii} = 1$.

Afin de faire une autre classification des états, on pose :

$$T_j = \min \{n \geq 1, X_n = j\}$$

Cette quantité représente le temps correspondant à la première visite à l'état j , avec $T_j = \infty$ si cette visite n'a jamais lieu.

De plus : $\mathbb{P}(T_i = \infty | X_0 = i) > 0 \Leftrightarrow i$ est passager. Et dans ce cas, on aura : $\mathbb{E}(T_i | X_0 = i) = \infty$.

Définition 1.5 Le temps moyen de récurrence μ_i de l'état $i \in \mathcal{S}$ est défini par :

$$\mu_i = \mathbb{E}(T_i | X_0 = i) = \begin{cases} \sum_n f_{ii}(n) & \text{si } i \text{ persistant} \\ \infty & \text{si } i \text{ passager} \end{cases}$$

Remarque : On peut avoir $\mu_i = \infty$, dans le cas où i est persistant.

Définition 1.6 Pour un état persistant $i \in \mathcal{S}$, on dit que :

- i est nul, si $\mu_i = \infty$.
- i est non-nul (ou positif), si $\mu_i < \infty$.

Remarque : Cette définition nous donne un simple critère de nullité en terme de probabilités de transition.

Théorème 1.4 Un état $i \in \mathcal{S}$ persistant est nul si et seulement si $p_{ii}(n) \rightarrow 0$ quand $n \rightarrow \infty$, et dans ce cas $p_{ji}(n) \rightarrow 0$, $\forall j \in \mathcal{S}$.

Remarque : Ce théorème sera démontré plus loin.

Définition 1.7 La période $d(i)$ d'un état $i \in \mathcal{S}$ est définie par : $d(i) = \text{pgcd} \{n, p_{ii}(n) > 0\}$. On dira que i est périodique, si $d(i) > 1$ et apériodique si $d(i) = 1$.

Remarque : $p_{ii}(n) = 0$, sauf si n est un multiple de $d(i)$.

Définition 1.8 Un état est dit ergodique, s'il est persistant, non-nul et apériodique.

Exemple : On considère une chaîne de Markov sur l'espace d'états $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, dont la matrice de transition est :

$$\begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

Déterminons le type de chacun des états de \mathcal{S} .

On remarque tout d'abord que si la chaîne se trouve à l'état 1, elle ne peut plus qu'aller à l'état 2 ou retourner à l'état 1, il en est de même si elle se trouve à l'état 5, dans ce cas elle ne peut plus qu'aller à l'état 6 ou rester à l'état 5. Ainsi les états 1, 2, 5 et 6 apparaissent comme des états persistants et non-nuls, car $\mu_i, i=1,2,5,6 < \infty$.

Quant aux états 3 et 4, ils sont passagers car par exemple partant des état 3 ou 4, au bout d'un certain temps, on va se retrouver dans les états 1 et 2 ou 5 et 6, et à partir de ce moment, la chaîne ne pourra plus retourner en 3 ou 4.

De plus, pour chaque état, on a $p_{ii}(1) > 0$, donc la période $d = 1$, par conséquent la chaîne est apériodique.

Au final, les états 3 et 4 sont passagers et apériodiques et les états 1, 2, 5 et 6 sont ergodiques.

1.3 Classification des chaînes

Définition 1.9 Soient $i, j \in \mathcal{S}$.

On dit que i communique avec j , noté $i \longrightarrow j$, si la chaîne visite toujours l'état j , avec une probabilité strictement positive, en commençant par l'état i . Donc $i \longrightarrow j$, si $p_{ij}(m) > 0$, pour un certain $m \geq 0$.

On dit que i et j intercommuniquent si $i \longrightarrow j$ et $j \longrightarrow i$. On note cela $i \longleftrightarrow j$.

Remarques :

- Si $i \neq j$, alors $i \longrightarrow j \Leftrightarrow f_{ij} > 0$.
- $i \longrightarrow i$, tant que $p_{ii}(0) = 1$.

Ce dont on peut déduire que \longleftrightarrow est une relation d'équivalence. Ainsi, l'espace d'états \mathcal{S} , peut être partitionné en classes d'équivalence.

Théorème 1.5 Soient $i, j \in \mathcal{S}$.

Si $i \longleftrightarrow j$, alors :

- (i) i et j ont la même période.
- (ii) i est passager $\Leftrightarrow j$ est passager.
- (iii) i est nul et persistant $\Leftrightarrow j$ est nul et persistant.

DÉMONSTRATION

Démontrons le (ii).

Soient $i, j \in \mathcal{S}$, tels que $i \longrightarrow j$, alors il va exister $m, n \geq 0$, tels que $\alpha = p_{ij}(m)p_{ji}(n) > 0$, par définition de $i \longrightarrow j$ et de $j \longrightarrow i$.

D'après les équations de Chapman-Kolmogorov, on a $\forall r \geq 0$:

$$p_{ii}(m+r+n) \geq p_{ij}(m)p_{jj}(r)p_{ji}(n) = \alpha p_{jj}(r)$$

On somme sur r et on obtient :

$$\sum_r p_{jj}(r) < \infty \quad \text{si} \quad \sum_r p_{ii}(r) < \infty$$

Ainsi, d'après le corollaire (1.3.1), on a j passager, si i l'est.

En intervertissant les rôles de i et j , on trouve que i est passager si j l'est, d'où l'assertion (ii).

Pour démontrer le point (i) on utilise des arguments similaires à ceux utilisés ci-dessus et la définition de période.

L'assertion (iii) sera démontrée plus loin.

Définition 1.10 Un ensemble \mathcal{C} d'états est dit :

- fermé, si $p_{ij} = 0$, pour tous $i \in \mathcal{C}$ et $j \notin \mathcal{C}$.
- irréductible, si $i \longleftrightarrow j$, pour tous $i, j \in \mathcal{C}$.

Remarque : Si une chaîne prend une valeur dans un ensemble d'états fermé, alors elle ne peut plus en sortir.

Exemple 1 : Promenade sur un tore On considère une marche aléatoire sur l'ensemble $\mathcal{S} = \{0, 1, 2, \dots, 9\}$, avec condition aux frontières périodique, c'est-à-dire que depuis l'état 9 on peut passer à l'état 0 en se déplaçant vers la droite. À chaque étape on se déplace vers la droite avec une probabilité p et on reste sur le même site avec une probabilité $1 - p$, $p \in [0, 1]$.

1. Déterminons la matrice de transition associée à notre marche aléatoire.

$$P = \begin{pmatrix} 1-p & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-p & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-p & p & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-p & p & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p & p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1-p & p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-p & p & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p & p & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p & p \\ p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p \end{pmatrix}$$

2. Montrons maintenant que la chaîne est irréductible.

D'après la définition (1.10), on sait que la chaîne est irréductible si $i \longleftrightarrow j$, $\forall i, j \in \mathcal{S}$. Et $i \longrightarrow j$ si, pour un certain m , $p_{ij}(m) > 0$.

Grâce à la matrice de probabilités de transition, on voit que chaque état i communique avec lui-même et avec l'état $i + 1$. Or \longleftrightarrow est une relation d'équivalence, donc chaque état communique avec tous les autres états, donc notre chaîne est bien irréductible.

Exemple 2 : Une chaîne de Markov est donnée par la matrice de transition suivante :

$$\begin{pmatrix} 1/3 & 2/3 & 0 \\ 3/4 & 1/4 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix}$$

Déterminons si la chaîne est irréductible.

Grâce à la matrice de transition, on se rend compte que si l'on se trouve à l'état 1 ou à l'état 2 on ne pourra jamais atteindre l'état 3, par conséquent 1 et 2 ne communiquent pas avec 3, donc la chaîne n'est pas irréductible.

Théorème 1.6 (Théorème de décomposition) *L'espace des états \mathcal{S} peut être partitionné de manière unique de la façon suivante :*

$$\mathcal{S} = T \cup C_1 \cup C_2 \cup \dots$$

avec T l'ensemble des états passagers et C_i des ensembles fermés irréductibles d'états persistants.

DÉMONSTRATION

Soient C_1, C_2, \dots les classes d'équivalence pour la relation \longleftrightarrow d'états persistants.

Montrons que les C_r sont fermés.

On raisonne par l'absurde. Supposons qu'il existe $i \in C_r$ et $j \notin C_r$ tels que $p_{ij} > 0$.

On a j ne communique pas avec i , donc :

$$\mathbb{P}(X_n \neq i, \forall n \geq 1 | X_0 = i) \geq \mathbb{P}(X_1 = j | X_0 = i) = p_{ij} > 0$$

ce qui contredit le fait que i est persistant, car si i était persistant alors pour un certain $n \geq 1$:

$$\mathbb{P}(X_n = i | X_0 = i) = 1$$

D'où :

$$\mathbb{P}(X_n \neq i, \forall n \geq 1 | X_0 = i) = 0$$

On a ainsi montré que tous les ensembles irréductibles d'états persistants, sont fermés. De plus, comme \longleftrightarrow est une relation d'équivalence, on a directement la partition avec l'ensemble d'états passagers et les ensembles irréductibles d'états persistants.

Remarque : Le théorème de décomposition nous indique que si $X_0 \in C_r$, alors les états de la chaîne ne quittent pas C_r . Par contre, si $X_0 \in T$, alors la chaîne peut changer d'ensemble d'états. Ainsi, chaque chaîne prend des valeurs dans l'ensemble des états passagers ou reste dans un ensemble irréductible fermé d'états persistants. Dans le cas où \mathcal{S} est fini, il en est différemment.

Lemme 1.1 *Si \mathcal{S} est fini, alors au moins l'un des états de la chaîne est persistant et tous les états persistants sont non-nuls.*

DÉMONSTRATION

On raisonne par l'absurde. Supposons que tous les états d'une chaîne, définis sur un ensemble d'états fini, sont passagers. On aurait alors :

$$1 = \lim_{n \rightarrow \infty} \sum_j p_{ij}(n)$$

par définition de la matrice de probabilités de transition \mathcal{P}^n .
Or, d'après le corollaire (1.3.2), on a :

$$\lim_{n \rightarrow \infty} \sum_j p_{ij}(n) = 0$$

On obtient donc une contradiction, par conséquent notre chaîne passe par au moins un état persistant.

Maintenant, supposons qu'un état persistant est nul.

On a toujours, l'égalité :

$$1 = \lim_{n \rightarrow \infty} \sum_j p_{ij}(n)$$

Cependant, si on utilise le théorème (1.4), on a aussi :

$$\lim_{n \rightarrow \infty} \sum_j p_{ij}(n) = 0$$

Ce qui nous fournit encore une fois une contradiction. Ainsi, tous les états persistants sont non-nuls.

1.4 Distribution stationnaire et théorème limite

1.4.1 Distribution stationnaire

Définition 1.11 *Un vecteur π est appelé distribution stationnaire (ou mesure invariante) d'une chaîne, si $\pi = (\pi_j, j \in \mathcal{S})$ vérifie :*

- (i) $\pi_j \geq 0, \forall j \in \mathcal{S}$ et $\sum_j \pi_j = 1$.
- (ii) $\pi = \pi\mathcal{P}$, ie $\pi_j = \sum_i \pi_i p_{ij}, \forall j \in \mathcal{S}$.

Remarque : Une telle distribution est dite stationnaire car si on itère l'assertion (ii) de la définition (1.11), on a : $\pi\mathcal{P}^2 = (\pi\mathcal{P})\mathcal{P} = \pi\mathcal{P} = \pi$. Autrement dit, on a : $\pi\mathcal{P}^n = \pi, \forall n \geq 0$.
Intéressons-nous maintenant à l'existence de distributions stationnaires.

Théorème 1.7 *Une chaîne irréductible admet une distribution stationnaire $\pi \Leftrightarrow$ tous ses états sont non-nuls et persistants.*

Dans ce cas, π est l'unique distribution stationnaire et est donnée par : $\pi_i = \frac{1}{\mu_i}, \forall i \in \mathcal{S}$, où μ_i est le temps moyen de récurrence de l'état i .

Fixons un état $k \in \mathcal{S}$.

Soit $\rho_i(k)$ le nombre minimal de visite de la chaîne à l'état i entre deux visites successives de l'état k , alors on peut écrire :

$$\rho_i(k) = \mathbb{E}(N_i | X_0 = k) \text{ où } N_i = \sum_{n=1}^{\infty} I_{\{X_n=i\} \cap \{T_k \geq n\}}$$

où T_k est le temps du premier retour à l'état k .

On peut remarquer que $N_k = 1$ et $\rho_k(k) = 1$.

On a également :

$$\rho_i(k) = \sum_{n=1}^{\infty} \mathbb{P}(X_n = i \cap T_k \geq n | X_0 = k)$$

On pose : $\rho(k) = (\rho_i(k))_{i \in \mathcal{S}}$.

On a aussi : $T_k = \sum_{i \in \mathcal{S}} N_i$ et donc :

$$\mu_k = \sum_{i \in \mathcal{S}} \rho_i(k)$$

Ainsi les termes du vecteur $\rho(k)$ une fois sommés sont égaux au temps moyen de récurrence μ_k .

Lemme 1.2 Soit $k \in \mathcal{S}$ un état d'une chaîne irréductible et persistante. Alors le vecteur $\rho(k)$ vérifie :

$$\rho_i(k) < \infty, \forall i \quad \text{et} \quad \rho(k) = \rho(k)\mathcal{P}$$

DÉMONSTRATION

Soit k un état d'une chaîne irréductible et persistante.

Montrons, tout d'abord, que $\rho_i(k) < \infty$ pour $i \neq k$.

On pose $l_{ki}(n) = \mathbb{P}(X_n = i \cap T_k \geq n | X_0 = k)$ la probabilité que la chaîne soit à l'état i au bout de n étapes, sachant qu'elle n'est jamais repassée par l'état k .

On a :

$$f_{kk}(m+n) \geq l_{ki}(m)f_{ik}(n)$$

puisque le premier retour à l'état k a lieu au bout de $m+n$ étapes si :

- $X_m = i$ et
- il n'y a pas de retour à l'état k avant la m -ième étape et
- la prochaine visite à l'état k a lieu au bout de n autres étapes.

Comme la chaîne est irréductible, il va exister un n tel que $f_{ik}(n) > 0$ donc, avec ce choix pour n , on a :

$$l_{ki}(m) \leq \frac{f_{kk}(m+n)}{f_{ik}(n)}$$

Donc

$$\begin{aligned} \rho_i(k) &= \sum_{m=1}^{\infty} l_{ki}(m) \leq \frac{1}{f_{ik}(n)} \sum_{m=1}^{\infty} f_{kk}(m+n) \\ &= \frac{1}{f_{ik}(n)} \sum_{m=1+n}^{\infty} f_{kk}(m) \\ &\leq \frac{1}{f_{ik}(n)} \sum_{m=1}^{\infty} f_{kk}(m) \\ &= \frac{1}{f_{ik}(n)} \quad \text{car} \quad f_{kk} = \sum_{m=1}^{\infty} f_{kk}(m) = 1 \end{aligned}$$

Donc

$$\rho_i(k) \leq \frac{1}{f_{ik}(n)} < \infty$$

Montrons maintenant la deuxième partie du lemme.

On sait que : $\rho_i(k) = \sum_{n=1}^{\infty} l_{ki}(n)$. On a alors :

$$l_{ki}(1) = \mathbb{P}(X_1 = i \cap T_k \geq 1 | X_0 = k) = p_{ki}$$

Par récurrence sur $n \geq 2$, on peut montrer que :

$$l_{ki}(n) = \sum_{j, j \neq k} l_{kj}(n-1)p_{ji}$$

On somme sur n :

$$\begin{aligned}
\rho_i(k) &= \sum_{n=1}^{\infty} l_{ki}(n) = p_{ki} + \sum_{n=2}^{\infty} \sum_{j, j \neq k} l_{kj}(n-1) p_{ji} \\
&= p_{ki} + \sum_{j, j \neq k} \left(\sum_{n=2}^{\infty} l_{kj}(n-1) \right) p_{ji} \\
&= p_{ki} + \sum_{j, j \neq k} \left(\sum_{n=1}^{\infty} l_{kj}(n) \right) p_{ji} \\
&= \rho_k(k) p_{ki} + \sum_{j, j \neq k} \rho_j(k) p_{ji} \quad \text{car } \rho_k(k) = 1 \\
&= \sum_{j \in \mathcal{S}} \rho_j(k) p_{ji}
\end{aligned}$$

L'interversion des deux sommes est possible : il suffit de considérer une somme finie sur n , puis de passer à la limite. On obtient bien de cette façon le résultat du lemme.

Théorème 1.8 *Si une chaîne est irréductible et persistante, alors il existe une unique racine $x \geq 0$, à constante multiplicative près, de l'équation $x = x\mathcal{P}$.*

La chaîne est non-nulle si $\sum_i x_i < \infty$ et est nulle si $\sum_i x_i = \infty$.

Démontrons maintenant le théorème (1.7).

DÉMONSTRATION

Supposons que π est une distribution stationnaire de notre chaîne irréductible. Supposons que tous les états de notre chaîne sont passagers, alors $p_{ij}(n) \rightarrow 0$, quand $n \rightarrow \infty$, $\forall i, j \in \mathcal{S}$, d'après le corollaire (1.3.2).

Par définition de distribution stationnaire, on a : $\pi \mathcal{P}^n = \pi$, ie

$$\pi_j = \sum_i \pi_i p_{ij}(n) \rightarrow 0, \quad \text{quand } n \rightarrow \infty \quad \forall i, j \in \mathcal{S} \quad (1.5)$$

ce qui contredit l'assertion (i) de la définition (1.11) donc tous les états de la chaîne sont persistants. Démontrons, l'assertion (1.5).

Soit $F \subset \mathcal{S}$ un ensemble fini. On a alors :

$$\sum_i \pi_i p_{ij}(n) \leq \sum_{i \in F} \pi_i p_{ij}(n) + \sum_{i \notin F} \pi_i$$

On passe à la limite quand $n \rightarrow \infty$:

$$\sum_i \pi_i p_{ij}(n) \rightarrow \sum_{i \notin F} \pi_i$$

On passe à la limite quand $F \nearrow \mathcal{S}$:

$$\sum_i \pi_i p_{ij}(n) \rightarrow 0$$

Montrons maintenant que l'existence de π implique que tous les états persistants sont non-nuls et que $\forall i \in \mathcal{S}$, $\pi_i = \frac{1}{\mu_i}$.

Supposons que X_0 a pour distribution π , alors : $\mathbb{P}(X_0 = i) = \pi_i$, $\forall i \in \mathcal{S}$.

On a aussi :

$$\begin{aligned}
\pi_j \mu_j &= \sum_{n=1}^{\infty} \mathbb{P}(T_j \geq n | X_0 = j) \mathbb{P}(X_0 = j) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(T_j \geq n \cap X_0 = j)
\end{aligned}$$

Or : $\mathbb{P}(T_j \geq 1 \cap X_0 = j) = \mathbb{P}(X_0 = j)$ et $\forall n \geq 2$:

$$\begin{aligned} \mathbb{P}(T_j \geq n \cap X_0 = j) &= \mathbb{P}(X_0 = j \cap X_m \neq j, 1 \leq m \leq n-1) \\ &= \mathbb{P}(X_m \neq j, 1 \leq m \leq n-1) - \mathbb{P}(X_m \neq j, 0 \leq m \leq n-1) \\ &= \mathbb{P}(X_m \neq j, 0 \leq m \leq n-2) - \mathbb{P}(X_m \neq j, 0 \leq m \leq n-1) \\ &= a_{n-2} - a_{n-1} \end{aligned}$$

avec $a_n = \mathbb{P}(X_m \neq j, 0 \leq m \leq n)$.

On somme sur n et on obtient :

$$\begin{aligned} \pi_j \mu_j &= \mathbb{P}(X_0 = j) + \mathbb{P}(X_0 \neq j) - \lim_{n \rightarrow \infty} a_n \\ &= 1 - \lim_{n \rightarrow \infty} a_n \\ &= 1 \quad \text{car } a_n \rightarrow \mathbb{P}(X_m \neq j, \forall m) = 0 \end{aligned}$$

On a donc montré que $\mu_j = \frac{1}{\pi_j} < \infty$ si $\pi_j > 0$.

Pour démontrer que $\pi_j > 0, \forall j \in \mathcal{S}$, on suppose le contraire, ie qu'il existe j , tel que $\pi_j = 0$.

Alors $\pi_j = 0 = \sum_i \pi_i p_{ij}(n) \geq \pi_i p_{ij}(n), \forall i \in \mathcal{S}, \forall n \in \mathbb{N}$, d'où $\pi_i = 0$ lorsque $i \rightarrow j$.

Mais alors $\pi_i = 0, \forall i \in \mathcal{S}$, ce qui est impossible car contredisant le fait que $\sum_i \pi_i = 1$. Donc $\forall j \in \mathcal{S}, \pi_j > 0$, ie $\mu_j < \infty$, ie les états persistants sont non nuls.

Par conséquent, si π existe, elle est unique (car uniquement déterminée) et tous les états de la chaîne sont non-nuls et persistants.

Inversement, si les états de la chaîne sont non-nuls et persistants, alors la chaîne admet une distribution stationnaire uniquement déterminée par le lemme (1.2).

On peut maintenant démontrer l'assertion (iii) du théorème (1.5).

DÉMONSTRATION

Montrons que si i et j intercommuniquent, alors i est nul et persistant si et seulement si j l'est.

Soit $C(i)$ une classe d'équivalence d'états irréductibles et fermés contenant l'état persistant, non-nul i .

Supposons que $X_0 \in C(i)$, alors $X_n \in C(i), \forall n \in \mathbb{N}$.

Ainsi d'après le Lemme (1.2), quel que soit k , état dans $C(i)$, le vecteur $\rho(k)$, défini précédemment sera une distribution stationnaire de notre chaîne X_n . Mais alors, d'après le théorème (1.7), on sait que tous les états de la chaîne seront non-nuls et persistants, d'où j non-nul et persistant.

Les rôles de i et j peuvent s'inverser, donc on a le résultat voulu.

Exemple 1 : Promenade sur un tore On reprend le premier exemple de la partie 1.3 et nous allons maintenant déterminer une distribution stationnaire de la chaîne.

On cherche π telle que $\pi_j \geq 0, \forall j \in \mathcal{S}$ et $\sum_j \pi_j = 1$. On résout alors le système suivant :

$$\begin{cases} \pi_0 = p\pi_9 + (1-p)\pi_0 \\ \pi_1 = p\pi_0 + (1-p)\pi_1 \\ \pi_2 = p\pi_1 + (1-p)\pi_2 \\ \dots \\ \pi_9 = p\pi_8 + (1-p)\pi_9 \end{cases}$$

D'où :

$$\begin{cases} p\pi_0 = p\pi_9 \\ p\pi_1 = p\pi_0 \\ p\pi_2 = p\pi_1 \\ \dots \\ p\pi_9 = p\pi_8 \end{cases}$$

Ainsi, si $p \neq 0$, on a : $\pi_0 = \pi_1 = \pi_2 = \dots = \pi_9$. Donc, comme $\sum_j \pi_j = 1$, on a $\pi_j = \frac{1}{10}, \forall j \in \mathcal{S}$.

Exemple 2 : On reprend l'exemple 2 de la partie 1.3, et nous allons aussi déterminer une distribution stationnaire pour cette chaîne.

On doit résoudre :

$$\begin{cases} \pi_0 = 1/3\pi_0 + 3/4\pi_1 \\ \pi_1 = 2/3\pi_0 + 1/4\pi_1 + 1/2\pi_2 \\ \pi_2 = 1/2\pi_2 \end{cases}$$

D'où :

$$\begin{cases} \pi_2 = 0 \\ 2/3\pi_0 = 3/4\pi_1 \end{cases}$$

Or $\pi_0 + \pi_1 + 0 = \pi_0 + 8/9\pi_0 = 1$, donc $\pi_0 = 9/17$ et $\pi_1 = 8/17$.

On trouve donc une seule distribution stationnaire, bien que la chaîne ne soit pas irréductible, qui se trouve être $\pi = (9/17, 8/17, 0)$.

Exemple 3 : On reprend ici l'exemple de la partie 2 et on souhaite déterminer les temps moyens de récurrence des états 1 et 2.

On considère l'ensemble $\mathcal{C} = \{1, 2\}$ qui est un ensemble d'états irréductible et fermé. Si $X_0 \in \mathcal{C}$, alors on peut résoudre l'équation $\pi = \pi\mathcal{P}_{\mathcal{C}}$, avec :

$$\mathcal{P}_{\mathcal{C}} = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$$

On résout :

$$\begin{cases} \pi_1 = 1/2\pi_1 + 1/4\pi_2 \\ \pi_2 = 1/2\pi_1 + 3/4\pi_2 \end{cases}$$

d'où : $1/4\pi_2 = 1/2\pi_1$, ie $\pi_1 = 1/2\pi_2$.

Or $\pi_1 + \pi_2 = 1/2\pi_2 + \pi_2 = 1$, donc $\pi_2 = 2/3$ et $\pi_1 = 1/3$.

On en déduit les temps moyens de récurrence : $\mu_1 = 3$ et $\mu_2 = 3/2$.

Théorème 1.9 Soit $s \in \mathcal{S}$ un état d'une chaîne irréductible.

La chaîne est passagère si et seulement si il existe une solution non nulle $\{y_j, j \neq s\}$ telle que $|y_j| \leq 1, \forall j$ aux équations :

$$y_i = \sum_{j, j \neq s} p_{ij}y_j, \quad i \neq s \quad (1.6)$$

DÉMONSTRATION

La chaîne est passagère si et seulement si s est passager.

Supposons que s est passager.

On pose $\tau_i(n) = \mathbb{P}(X_m \neq s, 1 \leq m \leq n | X_0 = i)$.

Alors, on a :

$$\begin{aligned} \tau_i(1) &= \mathbb{P}(X_1 \neq s | X_0 = i) \\ &= 1 - \mathbb{P}(X_1 = s | X_0 = i) \\ &= \sum_{j \in \mathcal{S}} p_{ij} - \mathbb{P}(X_1 = s | X_0 = i) \quad \text{par définition de } \mathcal{P} \\ &= \sum_{j \in \mathcal{S}} \mathbb{P}(X_{n+1} = j | X_n = i) - \mathbb{P}(X_1 = s | X_0 = i) \\ &= \sum_{j \in \mathcal{S}} \mathbb{P}(X_1 = j | X_0 = i) - \mathbb{P}(X_1 = s | X_0 = i) \quad \text{car la chaîne est homogène} \\ &= \sum_{j \in \mathcal{S}, j \neq s} p_{ij} \end{aligned}$$

On peut alors également montrer que :

$$\tau_i(n+1) = \sum_{j \in \mathcal{S}, j \neq s} p_{ij}\tau_j(n)$$

Ainsi, par récurrence sur $n > 0$, on a :

$$\tau_i(n) \geq \tau_i(n+1)$$

Si on pose $\tau_i = \lim_{n \rightarrow \infty} \tau_i(n)$ quand $n \rightarrow \infty$, on a :

$$\begin{aligned}\tau_i &= \lim_{n \rightarrow \infty} \mathbb{P}(X_m \neq s, 1 \leq m \leq n | X_0 = i) \\ &= \lim_{n \rightarrow \infty} (1 - \mathbb{P}(X_m = s, 1 \leq m \leq n | X_0 = i)) \\ &= 1 - f_{is} \quad \text{par définition de } f_{is}\end{aligned}$$

Montrons que τ_i vérifie l'équation (1.6) :

$$\begin{aligned}\sum_{j \in \mathcal{S}, j \neq s} p_{ij} \tau_j &= \sum_{j \in \mathcal{S}, j \neq s} p_{ij} (1 - f_{js}) \\ &= \sum_{j \in \mathcal{S}, j \neq s} p_{ij} - \sum_{j \in \mathcal{S}, j \neq s} p_{ij} f_{js} \\ &= \sum_{j \in \mathcal{S}} p_{ij} - p_{is} - \sum_{j \in \mathcal{S}} p_{ij} f_{js} + p_{is} f_{ss} \\ &= 1 - \sum_{j \in \mathcal{S}} p_{ij} f_{js} \quad \text{car } f_{ss} = 1 \text{ et } \mathcal{P} \text{ est stochastique} \\ &= 1 - f_{is} \\ &= \tau_i\end{aligned}$$

Donc τ_i vérifie bien l'équation (1.6).

De plus, $\tau_i > 0$, pour un certain $i \in \mathcal{S}$, puisque sinon on aurait $f_{is} = 1, \forall i \neq s$ et alors :

$$\begin{aligned}f_{ss} &= p_{ss} + \sum_{i \in \mathcal{S}, i \neq s} p_{si} f_{is} \\ &= p_{ss} + \sum_{i \in \mathcal{S}, i \neq s} p_{si} \\ &= 1\end{aligned}$$

Ce qui est impossible car contredisant le fait que s est passager.

Par conséquent $(\tau_i)_{i \neq s}$ satisfait les conditions du théorème.

Inversement, supposons qu'il existe $y = \{y_j, j \neq s\}$, vérifiant $|y_j| \leq 1, \forall j$ et l'équation (1.6). Alors :

$$\begin{aligned}|y_i| &= \left| \sum_{j \in \mathcal{S}, j \neq s} p_{ij} y_j \right| \\ &\leq \sum_{j \in \mathcal{S}, j \neq s} p_{ij} |y_j| \quad \text{par inégalité triangulaire} \\ &\leq \sum_{j \in \mathcal{S}, j \neq s} p_{ij} \\ &= \tau_i(1)\end{aligned}$$

Ainsi $|y_i| \leq \tau_i(1)$. Mais alors, on a aussi :

$$\begin{aligned}|y_i| &\leq \sum_{j \in \mathcal{S}, j \neq s} p_{ij} \tau_j(1) \\ &= \tau_i(2)\end{aligned}$$

On continue ainsi et on obtient que : $|y_i| \leq \tau_i(n), \forall n$.

Quand $n \rightarrow \infty$, on a vu que : $\lim_{n \rightarrow \infty} \tau_i(n) = \tau_i = 1 - f_{is} > 0$, donc $f_{is} < 1$. Or on sait que la chaîne est irréductible, donc nécessairement s sera passager. (En effet, dans le cas où on a une chaîne irréductible et persistante, $f_{ij} = 1, \forall i, j \in \mathcal{S}$).

Remarque : Ce théorème nous donne une condition nécessaire et suffisante : une chaîne irréductible est persistante si et seulement si l'unique solution de (1.6) est la solution nulle.

Intéressons-nous maintenant au cas où \mathcal{S} est infini.

Théorème 1.10 Soit $s \in \mathcal{S}$ un état quelconque d'une chaîne irréductible, avec $\mathcal{S} = \{0, 1, 2, \dots\}$. La chaîne est persistante s'il existe une solution $\{y_j, j \neq s\}$ aux inégalités :

$$y_i \geq \sum_{j, j \neq s} p_{ij} y_j, i \neq s, \text{ tel que } \lim_{i \rightarrow \infty} y_i = \infty$$

1.4.2 Théorème limite

Étudions maintenant le lien entre les distributions stationnaires et le comportement aux bords des probabilités $p_{ij}(n)$ quand $n \rightarrow \infty$.

Théorème 1.11 Pour une chaîne apériodique et irréductible on a :

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \frac{1}{\mu_j}, \forall i, j \in \mathcal{S}$$

Remarque :

- Si la chaîne est persistante et nulle ou passagère, alors $p_{ij}(n) \rightarrow 0$, car $\mu_j = \infty$.
- Si la chaîne est persistante et non-nulle, alors $p_{ij}(n) \rightarrow \pi_j = \frac{1}{\mu_j}$, où π est l'unique distribution stationnaire.
- Du théorème (1.11), on déduit que $\lim_{n \rightarrow \infty} p_{ij}(n)$ ne dépend pas du point de départ $X_0 = i$.
- Si $X = \{X_n\}$ une chaîne irréductible de période d , alors $Y = \{Y_n = X_{nd}, n \geq 0\}$ est une chaîne apériodique. Il s'ensuit que $p_{ij}(nd) = \mathbb{P}(Y_n = j | Y_0 = i) \rightarrow \frac{d}{\mu_j}$.

On peut maintenant démontrer le théorème (1.4).

DÉMONSTRATION

Soit $C(i)$ un ensemble d'états irréductible et fermé contenant l'état persistant i .

Si $C(i)$ est apériodique, alors le résultat découle du théorème (1.11).

Si $C(i)$ est périodique, on peut se ramener au cas apériodique, via la dernière remarque faite ci-dessus.

Effectuons maintenant, la démonstration du théorème (1.11).

DÉMONSTRATION

- Si la chaîne est passagère, le résultat est immédiat, avec le corollaire (1.3.2).

- Sinon, on construit un couple de chaîne $Z = (X, Y)$ étant une paire ordonnée, avec : $X = \{X_n, n \geq 0\}$ et $Y = \{Y_n, n \geq 0\}$ des chaînes de Markov indépendantes ayant même espace d'états \mathcal{S} et même matrice de probabilités de transition \mathcal{P} . Alors $Z = \{Z_n = (X_n, Y_n), n \geq 0\}$ est à valeur dans $\mathcal{S} \times \mathcal{S}$ et on peut montrer que Z est une chaîne de Markov.

Soient $i, j, k, l \in \mathcal{S}$:

$$\begin{aligned} p_{ij,kl} &= \mathbb{P}(Z_{n+1} = (k, l) | Z_n = (i, j)) \\ &= \mathbb{P}(X_{n+1} = k | X_n = i) \mathbb{P}(Y_{n+1} = l | Y_n = j) \quad \text{car } X \text{ et } Y \text{ sont indépendantes} \\ &= p_{ik} p_{jl} \end{aligned}$$

Donc Z est une chaîne de Markov, car X et Y le sont.

Comme X est irréductible et apériodique pour les états i, j, k, l , alors il va exister $N = N(i, j, k, l)$ tel que $p_{ik}(n) p_{jl}(n) > 0, \forall n \geq N$. Ainsi Z est aussi irréductible (car il existe un N tel que $p_{ij,kl}(n) > 0, \forall n \geq N$).

On découpe la suite en deux cas.

Cas 1 : X est non-nulle et persistante.

Alors X admet une unique distribution stationnaire π d'après le théorème (1.7) et ainsi Z admet aussi une distribution stationnaire $\nu = (\nu_{ij}, i, j \in \mathcal{S})$, avec $\nu_{ij} = \pi_i \pi_j$, par conséquent Z est aussi non-nulle et persistante, d'après le théorème (1.7).

Maintenant, supposons que $X_0 = i$ et $Y_0 = j$, ie $Z_0 = (i, j)$. On choisit un état $s \in \mathcal{S}$ et on pose :

$$T = \min \{n \geq 1, Z_n = (s, s)\}$$

quantité nous donnant le temps du premier passage à l'état (s, s) .

Comme Z est persistante, on a $\mathbb{P}(T < \infty) = 1$.

L'idée centrale de la preuve se trouve dans la remarque suivante :

Si $m \leq n$ et $X_m = Y_m$, alors X_n et Y_n sont identiquement distribuées, puisque les distributions de X_n et Y_n dépendent seulement de la matrice de transition \mathcal{P} et de la valeur à la m -ième étape. Donc si $T \leq n$, X_n et Y_n ont même distribution.

Nous allons maintenant utiliser le fait que T est fini pour montrer que les distributions X et Y sont indépendantes de leurs points de départ.

On a $Z_0 = (i, j)$, alors :

$$\begin{aligned} p_{ik}(n) &= \mathbb{P}(X_n = k | X_0 = i) \\ &= \mathbb{P}(X_n = k | X_0 = i \cap T \leq n) + \mathbb{P}(X_n = k | X_0 = i \cap T > n) \\ &= \mathbb{P}(Y_n = k | Y_0 = j \cap T \leq n) + \mathbb{P}(X_n = k | X_0 = i \cap T > n) \\ &\leq \mathbb{P}(Y_n = k | Y_0 = j) + \mathbb{P}(T > n) \\ &= p_{jk}(n) + \mathbb{P}(T > n) \end{aligned}$$

On peut écrire une équation similaire en inversant les rôles de i et j . De cette manière, on obtient :

$$|p_{ik}(n) - p_{jk}(n)| \leq \mathbb{P}(T > n) \rightarrow 0, \text{ quand } n \rightarrow \infty$$

car $\mathbb{P}(T < \infty) = 1$.

D'où

$$\lim_{n \rightarrow \infty} p_{ik}(n) - p_{jk}(n) = 0, \forall i, j, k \in \mathcal{S}$$

Donc si la limite de $p_{ik}(n)$ existe, elle ne dépend pas de i .

Montrons maintenant que cette limite existe. On écrit pour cela :

$$\pi_k - p_{jk}(n) = \sum_i \pi_i (p_{ik}(n) - p_{kj}(n)) \rightarrow 0 \quad (1.7)$$

d'où l'existence de la limite et le résultat du théorème est alors démontré dans ce cas.

Pour montrer la convergence de (1.7), on considère un ensemble fini $F \subset \mathcal{S}$, et on a :

$$\begin{aligned} \sum_{i \in \mathcal{S}} \pi_i (p_{ik}(n) - p_{jk}(n)) &\leq \sum_{i \in F} p_{ik}(n) - p_{jk}(n) + 2 \sum_{i \notin F} \pi_i \\ &\rightarrow 2 \sum_{i \notin F} \pi_i \quad \text{quand } n \rightarrow \infty \\ &\rightarrow 0 \quad \text{quand } F \rightarrow \mathcal{S} \end{aligned}$$

Cas 2 : Supposons maintenant que X est nulle et persistante.

Si Z est passagère, alors d'après le corollaire (1.3.2) appliqué à Z , on a :

$$\mathbb{P}(Z_n = (j, j) | Z_0 = (i, i)) = p_{ij}(n)^2 \rightarrow 0$$

d'où le résultat.

Si Z est non-nulle et persistante, commençons par le cas où $Z_0 = (i, i)$.

La période T_{ii}^Z du premier retour de Z à l'état (i, i) n'est pas plus petite que la période T_i du premier retour à l'état i de X . Cependant $\mathbb{E}(T_i) = \infty$ et $\mathbb{E}(T_{ii}^Z) < \infty$, ce qui est en contradiction.

Supposons, finalement que Z est nulle et persistante.

Le même argument qu'en (1.7) fonctionne et on obtient $p_{ij}(n) \rightarrow 0, \forall i, j \in \mathcal{S}$, sinon, il existerait une extraction $\{n_r, r \geq 0\}$, telle que :

$$\lim_{r \rightarrow \infty} p_{ij}(n_r) = \alpha_j, \forall i, j \quad (1.8)$$

avec α_j ne contenant pas que des zéros et indépendant de i .

L'assertion (1.8) impliquerait, alors que pour tout ensemble d'états fini F :

$$\sum_{j \in F} \alpha_j = \lim_{r \rightarrow \infty} \sum_{j \in F} p_{ij}(n_r) \leq 1$$

et donc $\alpha = \sum_j \alpha_j$ satisfierait $0 < \alpha \leq 1$.

De plus,

$$\sum_{k \in F} p_{ik}(n_r) p_{kj} \leq p_{ij}(n_r + 1) = \sum_k p_{ik} p_{kj}(n_r)$$

On fait $r \rightarrow \infty$:

$$\sum_{k \in F} \alpha_k p_{kj} \leq \sum_k p_{ik} \alpha_j = \alpha_j$$

On fait $F \rightarrow \mathcal{S}$:

$$\sum_k \alpha_k p_{kj} \leq \alpha_j, \forall j \in \mathcal{S}$$

De plus l'égalité peut avoir lieu, puisque si on a seulement l'inégalité stricte pour tout j , alors :

$$\sum_k \alpha_k = \sum_{k,j} \alpha_k p_{kj} < \sum_j \alpha_j$$

ce qui est impossible. Donc pour un certain $j \in \mathcal{S}$:

$$\sum_k \alpha_k p_{kj} = \alpha_j$$

Mais alors $\pi = \left\{ \frac{\alpha_j}{\alpha}, j \in \mathcal{S} \right\}$ est une distribution stationnaire pour X , ce qui est impossible, d'après le théorème (1.7). Par conséquent, on a bien démontré le résultat voulu.

Théorème 1.12 *Pour tout état apériodique j d'une chaîne de Markov :*

$$\lim_{n \rightarrow \infty} p_{jj}(n) = \frac{1}{\mu_j}$$

De plus, si i est un autre état, alors :

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \frac{f_{ij}}{\mu_j}$$

Corollaire 1.12.1 *Soit $\tau_{ij}(n) = \frac{1}{n} \sum_{m=1}^n p_{ij}(m)$ la proportion moyenne de temps écoulé jusqu'à la n -ième étape durant laquelle la chaîne est à l'état j sachant qu'elle est partie de l'état i .*

Si j est apériodique, alors $\lim_{n \rightarrow \infty} \tau_{ij}(n) = \frac{f_{ij}}{\mu_j}$.

1.5 Réversibilité des chaînes de Markov

Supposons que $\{X_n, 0 \leq n \leq N\}$ est une chaîne de Markov irréductible, non-nulle et persistante avec \mathcal{P} pour matrice de transition et π comme distribution stationnaire. Supposons de plus que X_n admet la distribution π comme distribution stationnaire pour tout n .

On définit la chaîne inverse par :

$$Y_n = X_{N-n}, \quad 0 \leq n \leq N$$

On définit bien ainsi une chaîne de Markov.

Théorème 1.13 *Soient $i, j \in \mathcal{S}, n \in [0; N]$. La suite Y est une chaîne de Markov, vérifiant :*

$$\mathbb{P}(Y_{n+1} = j | Y_n = i) = \frac{\pi_j}{\pi_i} p_{ji}$$

DÉMONSTRATION

Soient $i, j \in \mathcal{S}$, et $n \in [0; N]$.

$$\begin{aligned} \mathbb{P}(Y_{n+1} = j | Y_n = i) &= \mathbb{P}(X_{N-n-1} = j | X_{N-n} = i) \\ &= \mathbb{P}(X_{m-1} = j | X_m = i) \quad \text{on pose } m = N - n \\ &= \mathbb{P}(X_m = i | X_{m-1} = j) \frac{\mathbb{P}(X_{m-1} = j)}{\mathbb{P}(X_m = i)} \\ &= p_{ji} \frac{\pi_j}{\pi_i} \end{aligned}$$

Remarque : On a utilisé dans la démonstration, le fait que pour A, B deux événements, on a :

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

cette formule se démontre en écrivant la définition d'une probabilité conditionnelle.

On appelle la chaîne Y , l'inverse temporel de la chaîne X et on dit que X est réversible, si X et Y ont les mêmes probabilités de transition.

Définition 1.12 Soit $X = \{X_n, 0 \leq n \leq N\}$ une chaîne de Markov irréductible, telle que X_n ait la distribution stationnaire π , pour tout n .

La chaîne est dite réversible, si les matrices de transition de X et de son inverse temporel Y sont les mêmes, ie :

$$\pi_i p_{ij} = \pi_j p_{ji}, \forall i, j \in \mathcal{S} \quad (1.9)$$

Remarque : Les équations (1.9) sont appelées les équations détaillées de l'équilibre et sont fondamentales dans l'étude des chaînes réversibles.

Plus généralement, la matrice de transition \mathcal{P} et la distribution λ sont en équilibre détaillé, si :

$$\lambda_i p_{ij} = \lambda_j p_{ji}, \forall i, j \in \mathcal{S}$$

Une chaîne irréductible X ayant une distribution stationnaire π est dite réversible à l'équilibre si \mathcal{P} est en équilibre détaillé avec π .

Notons qu'une chaîne ayant une matrice de transition tridiagonale et une distribution stationnaire est réversible à l'équilibre. En effet :

DÉMONSTRATION

Soit X une chaîne de Markov ayant une matrice de transition de la forme :

$$\begin{pmatrix} p_{11} & p_{12} & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ p_{21} & p_{22} & p_{23} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & p_{32} & p_{33} & p_{34} & 0 & \cdot & \cdot & 0 \\ 0 & 0 & p_{43} & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & p_{n-1,n} \\ 0 & 0 & \cdot & \cdot & 0 & 0 & p_{n,n-1} & p_{nn} \end{pmatrix}$$

Montrons que X est réversible.

Soit π une distribution stationnaire de X , ie $\pi_j = \sum_i \pi_i p_{ij}$, ie

$$\begin{cases} \pi_j = \pi_{j-1} p_{j-1,j} + \pi_j p_{jj} + \pi_{j+1} p_{j+1,j} & 2 \leq j \leq n-1 \\ \pi_1 = \pi_1 p_{11} + \pi_2 p_{21} \\ \pi_n = \pi_{n-1} p_{n-1,n} + \pi_n p_{nn} \end{cases}$$

$$\begin{cases} \pi_j(1 - p_{jj}) = \pi_{j-1} p_{j-1,j} + \pi_{j+1} p_{j+1,j} & 2 \leq j \leq n-1 \\ \pi_1(1 - p_{11}) = \pi_2 p_{21} \\ \pi_n(1 - p_{nn}) = \pi_{n-1} p_{n-1,n} \end{cases}$$

Or $\sum_j p_{ij} = 1$, d'où :

$$\begin{cases} \pi_j(p_{j-1,j} + p_{j+1,j}) = \pi_{j-1} p_{j-1,j} + \pi_{j+1} p_{j+1,j}, & 2 \leq j \leq n-1 \\ \pi_1 p_{12} = \pi_2 p_{21} \\ \pi_n p_{n,n-1} = \pi_{n-1} p_{n-1,n} \end{cases}$$

On procède maintenant par récurrence sur $i \in \mathcal{S}$, avec $|\mathcal{S}| = n$.

- Si $i = 1$,

On a : $\pi_1 p_{12} = \pi_2 p_{21}$ d'après ce qui précède.

De plus : $\pi_1 p_{1j} = 0$, si $j > 2$ et $\pi_j p_{j1} = 0$, si $j > 2$.

Donc $\forall j \in \mathcal{S}$, $\pi_1 p_{1j} = \pi_j p_{j1}$.

- On suppose que la relation est vraie pour $k \in \mathcal{S}$. Montrons qu'elle est vraie au rang $k + 1$.
D'après les égalités ci-dessus, on a :

$$\pi_{k+1}(p_{k+1,k} + p_{k+1,k+2}) = \pi_k p_{k,k+1} + \pi_{k+2} p_{k+2,k+1}$$

D'après l'hypothèse de récurrence on a $\pi_{k+1} p_{k+1,k} = \pi_k p_{k,k+1}$, donc si on considère ceci dans l'équation ci-dessus on a : $\pi_{k+1} p_{k+1,k+2} = \pi_{k+2} p_{k+2,k+1}$.

De plus $\pi_{k+1} p_{k+1,j} = 0$ pour $j > k + 2$ et $j < k$ et $\pi_j p_{j,k+1} = 0$ pour les mêmes valeurs de j .
Donc on obtient bien le résultat voulu.

Par conséquent, une chaîne de Markov ayant une matrice de transition tridiagonale et admettant une distribution stationnaire, est réversible, car sa distribution stationnaire vérifie l'équation (1.9).

Théorème 1.14 Soit \mathcal{P} la matrice de transition d'une chaîne irréductible X .

Supposons qu'il existe une distribution π telle que $\pi_i p_{ij} = \pi_j p_{ji}, \forall i, j \in \mathcal{S}$.

Alors π est une distribution stationnaire de X .

De plus X est réversible à l'équilibre.

DÉMONSTRATION

Soit π une distribution satisfaisant les conditions du théorème, alors :

$$\begin{aligned} \sum_i \pi_i p_{ij} &= \sum_i \pi_j p_{ji} \\ &= \pi_j \sum_i p_{ji} \\ &= \pi_j \end{aligned}$$

Ainsi $\pi = \pi \mathcal{P}$ donc π est une distribution stationnaire.

La réversibilité à l'équilibre découle de la définition (1.12).

Remarque : La définition (1.12) peut être étendue aux chaînes infinies, mais dans ce cas, pour que X_n ait pour distribution π pour tout n il ne suffit pas que X_0 l'ait.

Exemple : Promenade sur un tore On considère une marche aléatoire sur l'ensemble $\mathcal{S} = \{0, 1, 2, \dots, 9\}$, avec condition aux frontières périodique. À chaque étape, on se déplace vers :

- la droite avec une probabilité p ;
- la gauche avec une probabilité $1 - p$.

avec $p \in [0, 1]$.

Cette chaîne admet donc comme matrice de probabilités de transition :

$$\begin{pmatrix} 0 & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p \\ 1-p & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-p & 0 & p & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-p & 0 & p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p & 0 & p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1-p & 0 & p & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p & 0 & p \\ p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p & 0 \end{pmatrix}$$

Cette chaîne est irréductible car chaque état communique avec les autres.

Déterminons maintenant si cette chaîne est réversible.

On cherche une distribution π telle que $\pi_i p_{ij} = \pi_j p_{ji}$.

Ici :

$$\begin{cases} p_{i,i+1} = p, & 0 \leq i \leq 8 \\ p_{i,i-1} = 1-p, & 1 \leq i \leq 9 \\ p_{09} = 1-p \\ p_{90} = p \\ p_{ij} = 0 & \text{sinon} \end{cases}$$

On se place dans le cas où : $1 \leq i \leq 8$.

On a :

$$\pi_i p_{ij} = \begin{cases} \pi_i p & \text{si } j = i + 1 \\ \pi_i(1 - p) & \text{si } j = i - 1 \end{cases}$$

et

$$\pi_j p_{ji} = \begin{cases} \pi_j p & \text{si } i = j + 1, \text{ ie } j = i - 1 \\ \pi_j(1 - p) & \text{si } i = j - 1, \text{ ie } j = i + 1 \end{cases}$$

Par conséquent, on veut :

$$\begin{cases} \pi_i p = \pi_j(1 - p) \\ \pi_j p = \pi_i(1 - p) \end{cases}$$

ie si on suppose $p \neq 0$, $\pi_j(1 - 2p) = 0$.

Deux cas se présentent alors :

Cas 1 : $\pi_j = 0$.

Alors pour tout $i \in \mathcal{S}$, on a $\pi_i = 0$, ce qui est impossible, car contredit le fait que $\sum_j \pi_j = 1$. **Cas**

2 : $1 - 2p = 0$, ie $p = \frac{1}{2}$.

Alors pour tout $i, j \in \mathcal{S}$, on a $\pi_i = \pi_j = \frac{1}{10}$.

Par conséquent, on en déduit que dans le cas où $p = \frac{1}{2}$, la chaîne est réversible.

1.6 Chaînes ayant un nombre fini d'étapes

On se place de nouveau dans le cas où \mathcal{S} est fini. On pose $N = |\mathcal{S}|$. On s'intéresse ici à une méthode pratique pour déterminer la matrice des n-ièmes probabilités de transition.

Théorème 1.15 (Théorème de Perron-Frobenius) *Si \mathcal{P} est une matrice de transition d'une chaîne finie et irréductible, avec une période d , alors :*

- (i) $\lambda_1 = 1$ est une valeur propre de \mathcal{P} .
- (ii) Les d racines de l'unité : $\lambda_1 = \omega^0, \lambda_2 = \omega^1, \dots, \lambda_d = \omega^{d-1}$, où $\omega = e^{2i\pi/d}$, sont valeurs propres de \mathcal{P} .
- (iii) Les valeurs propres $\lambda_{d+1}, \dots, \lambda_N$ vérifient $|\lambda_j| < 1$.

Deux cas se présentent alors si l'on veut déterminer la matrice $\mathcal{P}^n, \forall n > 0$.

- Si les valeurs propres de \mathcal{P} sont distinctes, alors il va exister $B \in \mathcal{M}_{N \times N}$ et $\Lambda \in \mathcal{M}_{N \times N}$ matrice diagonale contenant les valeurs propres de \mathcal{P} telles que :

$$\mathcal{P} = B^{-1}\Lambda B$$

Ainsi, on aura :

$$\mathcal{P}^n = B^{-1}\Lambda^n B$$

- Si les valeurs propres de \mathcal{P} sont non distinctes on écrit \mathcal{P} sous sa forme de Jordan.

Remarque : On peut ainsi utiliser le théorème de Perron-Frobenius pour déterminer des propriétés de \mathcal{P}^n .

1.7 Bilan

Dans ce chapitre, nous avons vu des notions de base sur la théorie des Chaînes de Markov, celle qui vont nous servir par la suite dans l'étude du modèle d'Ehrenfest sont : l'irréductibilité, la périodicité, la réversibilité, la notion de distribution invariante, celle de chaîne persistante et de temps de récurrence. Également, plusieurs théorèmes démontrés dans cette partie seront utiles. Nous allons maintenant nous intéresser à la théorie des Grandes Déviations.

Chapitre 2

La théorie des grandes déviations

La théorie des grandes déviations se penche sur le comportement asymptotique des suites de variables aléatoires. C'est seulement en 1966, que cette théorie fut posée proprement par Varadhan, elle formalise principalement l'idée de concentration de la mesure et élargit la notion de convergence.

2.1 La théorie des grandes déviations

2.1.1 Le principe des grandes déviations

On se place dans un espace de probabilités $(\Omega, \mathcal{A}, \mathbb{P})$. Souvent, une approximation du type : $P_n \simeq e^{-nI}$, avec P_n une certaine probabilité, n un paramètre assez grand et I une constante positive ou nulle, fait référence au principe des grandes déviations. Définissons ce principe :

Définition 2.1 Soit A_n une variable aléatoire indéxée par $n \in \mathbb{N}$. Soit $\mathbb{P}(A_n \in B)$ la probabilité que A_n soit à valeurs dans l'ensemble B . On dit que $\mathbb{P}(A_n \in B)$ satisfait au principe des grandes déviations, avec un taux I_B , si la limite suivante existe :

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log(\mathbb{P}(A_n \in B)) = I_B$$

Remarques :

- On comprend désormais que la notation $P_n \simeq e^{-nI}$ signifie que le comportement de $\mathbb{P}(A_n \in B)$ est une décroissance exponentielle en n .
- En utilisant la notation o , on peut réécrire la définition de la façon suivante :

$$-\log(\mathbb{P}(A_n \in B)) = nI_B + o(n)$$

avec $I_B \geq 0$.

- Pour extraire la constante I_B , il suffit de diviser par n :

$$-\frac{1}{n} \log(\mathbb{P}(A_n \in B)) = I_B + o(1)$$

On passe ensuite à la limite quand $n \rightarrow \infty$, afin de se débarrasser de la contribution de $o(1)$, on retrouve ainsi la définition (2.1).

- De plus, si $\mathbb{P}(A_n \in B)$ a un comportement dominant exponentiel en n , alors la limite I_B va exister.
- Par contre, si la limite n'existe pas, alors soit $\mathbb{P}(A_n \in B)$ a un comportement trop singulier, soit $\mathbb{P}(A_n \in B)$ décroît plus rapidement qu'en e^{-na} , avec $a > 0$. Dans ce cas on parle de décroissance "super-exponentielle" et $I = \infty$.
- La limite de la définition (2.1) peut être nulle, dans ce cas $\mathbb{P}(A_n \in B)$ décroît "sous-exponentiellement" en n , ie $\mathbb{P}(A_n \in B)$ décroît moins rapidement qu'en e^{-na} , avec $a > 0$.

Le cas qui nous intéresse ici se produit quand la limite de la définition (2.1) ne vaut ni zéro, ni l'infini.

Dans le cas d'une variable aléatoire continue, on a une définition similaire du principe des grandes déviations :

Définition 2.2 Soit A_n une variable aléatoire continue.

Soit $a \in \mathbb{R}$.

On dit que $\mathbb{P}(A_n = a)$ satisfait au principe des grandes déviations si la limite suivante existe :

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log(\mathbb{P}(A_n = a)) = I(a)$$

où I est une fonction continue appelée la fonction taux.

Remarque : La différence principale avec une variable aléatoire discrète est que l'on considère plutôt la densité à la place de la distribution de la variable aléatoire.

Le principe des grandes déviations en terme de densité peut être traduit de la façon suivante :

$$\mathbb{P}(S_n \in [s, s + ds]) = f_{S_n}(s) ds$$

avec f_{S_n} la densité de S_n .

On peut alors écrire :

$$\mathbb{P}(S_n \in [s, s + ds]) \simeq e^{-nJ(s)} ds$$

On retrouve alors la fonction taux J , comme dans le cas d'une distribution discrète, en prenant la limite de la façon suivante :

$$\begin{aligned} \lim_{n \rightarrow \infty} -\frac{1}{n} \log(\mathbb{P}(S_n \in [s, s + ds])) &= \lim_{n \rightarrow \infty} -\frac{1}{n} \log(e^{-nJ(s)} ds) \\ &= \lim_{n \rightarrow \infty} -\frac{1}{n} \left(\log(e^{-nJ(s)}) + \log(ds) \right) \\ &= \lim_{n \rightarrow \infty} -\frac{1}{n} (-nJ(s) + \log(ds)) \\ &= J(s) + \lim_{n \rightarrow \infty} -\frac{1}{n} \log(ds) \\ &= J(s) \end{aligned}$$

Pour justifier la dernière égalité on considèrera ici que ds est un élément non nul, infinitésimal.

2.1.2 Autour du principe des grandes déviations

La définition du principe des grandes déviations présentée dans la partie précédente est une définition pratique plus que rigoureuse. En réalité plusieurs précautions doivent être prises pour pouvoir définir la limite d'une probabilité, nous allons maintenant en donner deux.

- La limite peut ne pas exister. Dans ce cas, on est capable de trouver une borne inférieure et une borne supérieure de $\mathbb{P}(A_n \in B)$ et toutes deux sont exponentielles en n :

$$e^{-nI_B^-} \leq \mathbb{P}(A_n \in B) \leq e^{-nI_B^+}$$

Ces deux bornes donnent un sens précis à l'affirmation : $\mathbb{P}(A_n \in B)$ décroît exponentiellement en n , et nous donne deux principes de grandes déviations ; l'un défini en terme de limite inférieure donnée par I_B^- et l'autre en terme de limite supérieure donnée par I_B^+ .

Dans toute la suite, on considèrera que $I_B^- = I_B^+$, afin de retrouver la définition (2.1).

- Les variables aléatoires discrètes sont souvent traitées comme si elles devenaient continues lorsque $n \rightarrow \infty$.

Le remplacement de variables aléatoires discrètes par des variables aléatoires continues se justifie mathématiquement, avec la notion de convergence faible.

Soit A_n une variable aléatoire discrète avec pour distribution $\mathbb{P}(A_n = a)$ définie sur un sous-ensemble de valeurs $a \in \mathbb{R}$.

Soit \widehat{A}_n une variable aléatoire continue de densité $f_{\widehat{A}_n}$ définie sur \mathbb{R} .

Dire que A_n converge faiblement vers \widehat{A}_n signifie que, quand $n \rightarrow \infty$:

$$\sum_a f(a) \mathbb{P}(A_n = a) \rightarrow \int f(a) f_{\widehat{A}_n}(a) da$$

avec f une fonction continue et bornée sur \mathbb{R} .

Finalement, la plupart des variables aléatoires considérées ici seront soit des variables aléatoires discrètes qui convergent faiblement vers des variables aléatoires continues, soit des variables aléatoires continues. Pour l'étude de ces deux cas de figure, nous utiliserons les mêmes notations, à savoir celle avec les distributions plutôt qu'avec la densité, ie $\mathbb{P}(A_n \in [a, a + da])$. On ne distinguera plus dans la suite A_n et \widehat{A}_n . Désormais, on écrira :

$$\mathbb{P}(A_n \in [a, a + da]) \sim e^{-nI(a)} da$$

que A_n soit discrète ou continue, elle vérifie le principe des grandes déviations.

Pour finir, on introduit deux notations :

- on utilisera $\mathbb{P}(A_n \in da)$ à la place de $\mathbb{P}(A_n \in [a, a + da])$;
- deuxièmement, on remplacera \sim par \asymp quand on parlera de grandes déviations.

Finalement, dans la suite, on écrira :

$$\mathbb{P}(A_n \in da) \asymp e^{-nI(a)} da$$

qui signifie que A_n satisfait au principe des grandes déviations, défini en (2.1), avec pour fonction taux I . Le signe \asymp est utilisé pour signifier que, lorsque $n \rightarrow \infty$, la partie dominante de $\mathbb{P}(A_n \in da)$ est la décroissance exponentielle en $e^{-nI(a)}$.

2.1.3 Détermination de la fonction taux

La théorie des grandes déviations peut être vue, d'un point de vue pratique, comme une série de méthodes permettant de résoudre principalement deux problèmes :

- établir que le principe des grandes déviations existe pour des variables aléatoires données ;
- déterminer la fonction taux associée.

Nous allons maintenant présenter le théorème de Gärtner-Ellis, qui permet de répondre à ces deux problèmes.

Le théorème de Gärtner-Ellis

Définition 2.3 Soit A_n une variable aléatoire, paramétrée par $n \in \mathbb{N}$.

On définit la fonction génératrice des cumulants échelonnés de A_n par la limite :

$$\lambda(k) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \langle e^{nkA_n} \rangle$$

avec $k \in \mathbb{R}$ et :

$$\langle e^{nkA_n} \rangle = \int_{\mathbb{R}} e^{nka} \mathbb{P}(A_n = a) \quad (2.1)$$

Théorème 2.1 (Théorème de Gärtner-Ellis) Soit A_n une variable aléatoire, paramétrée par $n \in \mathbb{N}$.

Si $\lambda(k)$ existe et est différentiable pour tout $k \in \mathbb{R}$, alors A_n satisfait au principe des grandes déviations, ie :

$$\mathbb{P}(A_n = a) \asymp e^{-nI(a)} da$$

avec la fonction taux donnée par :

$$I(a) = \sup_{k \in \mathbb{R}} \{ka - \lambda(k)\} \quad (2.2)$$

I est alors appelée la Transformée de Legendre-Fenchel.

DÉMONSTRATION

Nous nous contenterons ici de donner les principales idées de la démonstration.

On suppose que la fonction génératrice des cumulants échelonnés existe et est différentiable pour tout $k \in \mathbb{R}$.

Montrons que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(A_n \in da) = -I(a)$$

On pose

$$p = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} a \mathbb{P}(A_n \in da)$$

Tout d'abord, on peut remarquer que $I(p) = 0$. En effet, on sait que :

$$\begin{aligned} \lambda(k) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \int_{\mathbb{R}} e^{nka} \mathbb{P}(A_n \in da) \\ &\geq \lim_{n \rightarrow \infty} \frac{1}{n} \int_{\mathbb{R}} nka \mathbb{P}(A_n \in da) \quad \text{d'après l'inégalité de Jensen} \\ &= \lim_{n \rightarrow \infty} k \int_{\mathbb{R}} a \mathbb{P}(A_n \in da) \\ &= k \lim_{n \rightarrow \infty} \int_{\mathbb{R}} a \mathbb{P}(A_n \in da) \end{aligned}$$

Par conséquent, $\forall k \in \mathbb{R}$, $\lambda(k) \geq kp$, ie $\forall k \in \mathbb{R}$, $kp - \lambda(k) \leq 0$, d'où $I(p) \leq 0$.

Or la fonction I est positive (nous verrons une démonstration de ceci, dans le paragraphe 4.1.5) donc nécessairement $I(p) = 0$.

Montrons maintenant que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(A_n \in da) \leq -I(a)$$

Nous allons distinguer deux cas :

Cas 1 : $a \geq p$

Comme

$$\lambda(k) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}(e^{nkA_n})$$

alors : $\forall \epsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0,$

$$\log \mathbb{E}(e^{nkA_n}) \leq n(\lambda(k) + \epsilon) \quad (2.3)$$

De plus, on a :

$$\begin{aligned} \mathbb{P}(A_n \in da) &\leq \mathbb{P}(A_n \in [a, +\infty[) \\ &= \mathbb{P}(A_n \geq a) \\ &= \mathbb{P}(e^{nkA_n} \geq e^{nka}) \quad \text{car l'exponentielle est bijective} \\ &\leq e^{-nka} \mathbb{E}(e^{nkA_n}) \quad \text{par inégalité de Tchebytchev} \\ &\leq e^{-nka} e^{n(\lambda(k) + \epsilon)} \quad \text{d'après (2.3)} \end{aligned}$$

$\forall n \geq n_0, \forall k \geq 0.$

Par conséquent, $\forall k \geq 0,$

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(A_n \in da) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(A_n \in da) \\ &\leq -\sup_{k \geq 0} \{ka - \lambda(k)\} + \epsilon \end{aligned}$$

Or, comme $a \geq p$, alors $I(a) = \sup_{k \geq 0} \{ka - \lambda(k)\}$.

Donc :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(A_n \in da) \leq -I(a) + \epsilon$$

On a pris ϵ quelconque, par conséquent on obtient :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(A_n \in da) \leq -I(a)$$

Cas 2 : $a \leq p$

On obtient dans ce cas des résultats analogues, en procédant comme dans le cas précédent. Montrons maintenant que :

$$-I(a) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(A_n \in da)$$

On se place tout d'abord, dans le cas où il existe $a^* \in \mathbb{R}$, tel que :

$$\sup_{k \in \mathbb{R}} \{ka - \lambda(k)\} = a^*a - \lambda(a^*) = I(a)$$

c'est-à-dire que le supremum de la fonction $f : k \mapsto ka - \lambda(k)$ est atteint.

Cas 1 : Supposons que $a^* \geq 0$.

On définit une variable aléatoire $(Z_n, n \geq 0)$ telle que :

$$\mathbb{P}(Z_n \in dx) = \frac{e^{a^*x}}{\mathbb{E}(e^{a^*nA_n})} \mathbb{P}(A_n \in dx)$$

Ainsi :

$$\begin{aligned} \mathbb{P}(A_n \in da) &= \mathbb{P}(A_n \in [a, a + da]) \\ &= \int_a^{a+da} \mathbb{P}(A_n \in dx) \\ &= \int_a^{a+da} \frac{\mathbb{E}(e^{a^*nA_n})}{e^{a^*x}} \mathbb{P}(Z_n \in dx) \\ &= \mathbb{E}(e^{na^*A_n}) \int_a^{a+da} e^{-a^*x} \mathbb{P}(Z_n \in dx) \\ &\geq \mathbb{E}(e^{na^*A_n}) e^{-a^*(a+da)} \mathbb{P}(Z_n \in da) da \\ &= \mathbb{E}(e^{na^*A_n}) e^{-a^*a} e^{-a^*da} \mathbb{P}(Z_n \in da) da \\ &\geq \mathbb{E}(e^{na^*A_n}) e^{-a^*an} e^{-a^*da} \mathbb{P}(Z_n \in da) da \\ &= e^{\log(\mathbb{E}(e^{na^*A_n}))} e^{-a^*an} e^{-a^*da} \mathbb{P}(Z_n \in da) da \end{aligned}$$

D'où par croissance du logarithme :

$$\log(\mathbb{P}(A_n \in da)) \geq \log(\mathbb{E}(e^{na^*A_n})) - a^*an - a^*da + \log(\mathbb{P}(Z_n \in da)) + \log(da)$$

On divise par $n > 0$ dans l'inégalité :

$$\frac{1}{n} \log(\mathbb{P}(A_n \in da)) \geq \frac{1}{n} \log(\mathbb{E}(e^{na^*A_n})) - a^*a - \frac{a^*da}{n} + \frac{\log(\mathbb{P}(Z_n \in da))}{n} + \frac{\log(da)}{n}$$

On passe à la limite quand $n \rightarrow \infty$:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log(\mathbb{P}(A_n \in da)) &\geq -a^*a + \lim_{n \rightarrow \infty} \frac{1}{n} \log(\mathbb{E}(e^{na^*A_n})) \\ &= -(a^*a - \lim_{n \rightarrow \infty} \frac{1}{n} \log(\mathbb{E}(e^{na^*A_n}))) \\ &= -(a^*a - \lambda(a^*)) \\ &= -I(a) \end{aligned}$$

Donc, on obtient bien que :

$$-I(a) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log(\mathbb{P}(A_n \in da))$$

Cas 2 : Supposons que $a^* < 0$.

Dans ce cas, on raisonne de manière similaire à celui effectué dans le cas précédent nous donne le résultat.

Supposons maintenant que le supremum de la fonction $f : k \mapsto ka - \lambda(k)$ n'est pas atteint, autrement dit, supposons que $I(a) > a^*a - \lambda(a^*)$. On considère de nouveau deux cas :

Cas 1 : $a \geq p$

Alors, on sait que la fonction I sera croissante sur $[p, +\infty[$, par conséquent il va exister une suite $(x_l)_l$ croissant vers $+\infty$ et telle que $(ax_l - \lambda(x_l))_l$ croisse vers $I(a)$. Dans ce cas, on pourra conclure en utilisant le théorème de convergence dominée.

Cas 2 : $a < p$

Un raisonnement analogue à celui effectué dans le cas 1 permet de conclure.

Ce qui termine la preuve.

Remarque : Le théorème de Gärtner-Ellis nous donne lorsque la fonction λ associée à A_n est différentiable, que A_n obéit au principe des grandes déviations, avec une fonction taux I donnée par la transformée de Legendre-Fenchel.

Dans la prochaine partie, nous verrons comment calculer les fonction taux car il est important de savoir que toutes les fonctions taux ne peuvent pas se calculer avec ce théorème. Nous allons donc voir quels arguments nous permettent de savoir qu'on pourra déterminer la fonction taux avec la transformée de Legendre-Fenchel.

Conséquences du théorème de Gärtner-Ellis

Considérons l'équation (2.2), sachant que A_n satisfait au principe des grandes déviations. Nous allons voir les conséquences du théorème de Gärtner-Ellis.

On sait que :

$$\mathbb{P}(A_n = a) \asymp e^{-nI(a)} da$$

Si on insère cela dans l'équation (2.1), on a :

$$\langle e^{nkA_n} \rangle \asymp \int_{\mathbb{R}} e^{nka} e^{-nI(a)} da = \int_{\mathbb{R}} e^{n(ka - I(a))} da$$

Maintenant, nous allons approximer cette intégrale par son intégrant le plus grand, qui a lieu d'être lorsque $ka - I(a)$ atteint son maximum. Cette approximation est connue sous le nom d'approximation du point selle ou approximation de Laplace. On peut considérer une telle approximation ici car l'erreur associée à celle-ci est du même ordre de grandeur que l'erreur associée à l'approximation des grandes déviations elle-même.

Ainsi, en admettant que le maximum de $ka - I(a)$ existe et est unique, on peut écrire :

$$\langle e^{nkA_n} \rangle \asymp e^{n \sup_{a \in \mathbb{R}} \{ka - I(a)\}}$$

Mais alors :

$$\lambda(k) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \langle e^{nkA_n} \rangle = \sup_{a \in \mathbb{R}} \{ka - I(a)\}$$

Pour obtenir $I(a)$ en fonction de $\lambda(k)$, nous utiliserons le fait que la transformée de Legendre-Fenchel est réversible lorsque λ est différentiable en tout point. Dans ce cas la transformée de Legendre-Fenchel est son propre inverse, ie :

$$I(a) = \sup_{k \in \mathbb{R}} \{ka - \lambda(k)\}$$

ce qui donne le résultat de l'équation (2.2).

Ce qu'on vient de faire illustre deux points importants de la théorie des grandes déviations. Le premier est que la transformée de Legendre-Fenchel apparaît comme une conséquence naturelle de l'approximation de Laplace. Le deuxième est que le théorème de Gärtner-Ellis est essentiellement une conséquence du principe des grandes déviations combiné avec l'approximation de Laplace.

2.1.4 Le théorème de Cramér

Le théorème de Cramér est une application du théorème de Gärtner-Ellis, dans le cas d'une variable aléatoire de la forme suivante : $S_n = \frac{1}{n} \sum_{i=1}^n X_i$, avec les X_i des variables aléatoires

indépendantes et identiquement distribuées.

Dans ce cas, la fonction génératrice des cumulants échelonnés s'écrit :

$$\begin{aligned}\lambda(k) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \langle e^{nk \frac{1}{n} \sum_i X_i} \rangle \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \prod_{i=1}^n \langle e^{kX_i} \rangle \\ &= \log \langle e^{kX} \rangle\end{aligned}$$

avec X l'un des X_i .

Par conséquent, on peut étendre le principe des grandes déviations à S_n simplement en calculant la fonction génératrice des cumulants échelonnés d'un seul terme de la somme composant S_n , puis en prenant sa transformée de Legendre-Fenchel, on en déduit la fonction taux associée.

Proposition 2.2 (Théorème de Cramér) *Soit S_n une variable aléatoire de la forme $S_n = \frac{1}{n} \sum_{i=1}^n X_i$, avec les X_i des variables aléatoires indépendantes et identiquement distribuées, ayant pour fonction génératrice des cumulants échelonnés pour tout $k \in \mathbb{R}$:*

$$\lambda(k) = \log \langle e^{kX} \rangle$$

avec X l'un des X_i .

Alors S_n satisfait au principe des grandes déviations avec pour fonction taux :

$$I(a) = \sup_{k \in \mathbb{R}} \{ka - \lambda(k)\}$$

DÉMONSTRATION

Montrons que la fonction génératrice des cumulants échelonnés est différentiable.

On sait que pour tout $k \in \mathbb{R}$:

$$\begin{aligned}\lambda(k) &= \log \langle e^{kX} \rangle \\ &= \log \int_{\mathbb{R}} e^{ka} \mathbb{P}(X = a) \\ &= \log \int_{\mathbb{R}} \sum_{n=0}^{\infty} \frac{(ka)^n}{n!} \mathbb{P}(X = a) \\ &= \log \sum_{n=0}^{\infty} \int_{\mathbb{R}} \frac{(ka)^n}{n!} \mathbb{P}(X = a)\end{aligned}$$

Pour finir, comme la fonction logarithme est développable en série entière, on a que la fonction génératrice des cumulants échelonnés est développable en série entière. D'où la différentiabilité de λ .

On peut maintenant appliquer le théorème de Gärtner-Ellis à la variable aléatoire S_n et on obtient le théorème de Cramér.

Exemple 1 : On considère :

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i$$

avec les X_i des variables aléatoires indépendantes et identiquement distribuées de loi normale $\mathcal{N}(\mu, \sigma^2)$.

Déterminons la fonction génératrice des cumulants échelonnés associée à S_n :

$$\begin{aligned}
\lambda(k) &= \log \langle e^{kX} \rangle \\
&= \log \left(\int_{\mathbb{R}} e^{kx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right) \\
&= \log \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{k(x+\mu)} e^{-\frac{x^2}{2\sigma^2}} dx \right) \quad \text{par changement de variable } x \mapsto x + \mu \\
&= \log \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{k\mu} e^{-\frac{(x-\sigma^2 k)^2}{2\sigma^2}} e^{\frac{1}{2}\sigma^2 k^2} dx \right) \\
&= \log \left(e^{k\mu} e^{\frac{1}{2}\sigma^2 k^2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\sigma^2 k)^2}{2\sigma^2}} dx \right) \\
&= k\mu + \frac{1}{2}\sigma^2 k^2 \quad \text{car on a la densité de } \mathcal{N}(\sigma^2 k, \sigma^2)
\end{aligned}$$

Par conséquent, λ est différentiable pour tout $k \in \mathbb{R}$, donc d'après le théorème de Cramér, on sait que S_n satisfait au principe des grandes déviations, ie que :

$$\mathbb{P}(S_n \in ds) \asymp e^{-nI(s)}$$

avec :

$$I(s) = \sup_{k \in \mathbb{R}} \{ks - \lambda(k)\}$$

Déterminons, plus précisément la fonction taux. On pose f la fonction définie sur \mathbb{R} , par :

$$f(k) = ks - k\mu - \frac{1}{2}\sigma^2 k^2$$

$$\forall k \in \mathbb{R}, \quad f'(k) = s - \mu - k\sigma^2$$

Ainsi, $f'(k) = 0$ lorsque $k = \frac{s - \mu}{\sigma^2}$. Le maximum de la fonction f est atteint en ce point, donc :

$$I(s) = \frac{(s - \mu)^2}{2\sigma^2}$$

Exemple 2 : On considère :

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i$$

avec les X_i des variables aléatoires indépendantes et identiquement distribuées de loi exponentielle $\mathcal{E}(\mu)$, avec $\mu > 0$.

Déterminons la fonction génératrice des cumulants échelonnés associée à S_n :

Cas 1 : $k = \mu$

$$\lambda(k) = \log \left(\int_0^{\infty} \mu da \right) = \infty$$

Cas 2 : $k \neq \mu$

$$\begin{aligned}
\lambda(k) &= \log \langle e^{kX} \rangle \\
&= \log \left(\int_0^{\infty} e^{ka} \mu e^{-\mu a} da \right) \\
&= \log \left(\mu \int_0^{\infty} e^{a(k-\mu)} da \right) \\
&= \log \left(\frac{\mu}{k - \mu} \left(\lim_{a \rightarrow \infty} (e^{a(k-\mu)}) - 1 \right) \right)
\end{aligned}$$

Donc, on obtient :

$$\lambda(k) = \begin{cases} \log \left(\frac{\mu}{\mu - k} \right) & \text{si } k < \mu \\ \infty & \text{si } k \geq \mu \end{cases}$$

Déterminons maintenant la fonction taux associée dans le cas où $k < \mu$.
Soit $s \in \mathbb{R}$,

$$\begin{aligned} I(s) &= \sup_{k \in \mathbb{R}} \{ks - \lambda(k)\} \\ &= \sup_{k \in \mathbb{R}} \left\{ ks - \log \left(\frac{\mu}{\mu - k} \right) \right\} \end{aligned}$$

On définit la fonction f sur \mathbb{R} par :

$$f(k) = ks - \log \left(\frac{\mu}{\mu - k} \right)$$

On a :

$$f'(k) = s + \frac{-1}{\mu - k}$$

On distingue alors deux cas :

Cas 1 : $s \leq 0$

La fonction f n'admet pas de maximum.

Cas 2 : $s > 0$

La fonction admet un maximum en $k = \mu - \frac{1}{s}$, on trouve alors :

$$I(s) = \mu s - 1 - \log(\mu + s)$$

Au final, pour $k < \mu$, on a :

$$I(s) = \begin{cases} \mu s - 1 - \log(\mu + s) & \text{si } s > 0 \\ \infty & \text{si } s \leq 0 \end{cases}$$

Remarque : On peut déduire de ce qu'on vient de faire que le théorème de Gärtner-Ellis peut s'appliquer pour différents types de variables aléatoires, à partir du moment où l'on peut déterminer la fonction génératrice λ et qu'elle vérifie les conditions du théorème.

2.1.5 Propriétés de la fonction génératrice des cumulants échelonnés et de la fonction taux

Nous allons maintenant énoncer des propriétés sur la fonction génératrice des cumulants échelonnés λ et sur la fonction taux I dans le cas où I est donnée par le théorème de Gärtner-Ellis. Les propriétés seront valables pour A_n variable aléatoire vérifiant les conditions du dit théorème et pas seulement dans le cas d'une somme de variables aléatoires indépendantes et identiquement distribuées.

Propriétés de la fonction λ lorsque $k = 0$

Comme λ est une mesure de probabilités alors elle vérifie : $\lambda(0) = 0$.

De plus

$$\lambda'(0) = \lim_{n \rightarrow \infty} \frac{\langle A_n e^{nkA_n} \rangle}{\langle e^{nkA_n} \rangle} \Big|_{k=0} = \lim_{n \rightarrow \infty} \langle A_n \rangle$$

si $\lambda'(0)$ existe.

Par exemple, dans le cas d'une somme de variables aléatoires indépendantes et identiquement distribuées de loi normale $\mathcal{N}(\mu, \sigma^2)$, on obtient :

$$\lambda'(0) = \langle X \rangle = \mu$$

De la même façon, on peut calculer :

$$\lambda''(0) = \lim_{n \rightarrow \infty} n(\langle A_n^2 \rangle - \langle A_n \rangle^2) = \lim_{n \rightarrow \infty} n \text{Var}(A_n)$$

ce qui nous donne pour une somme de variables aléatoires identiquement distribuées de loi normale :

$$\lambda''(0) = \sigma^2$$

Convexité de λ

Proposition 2.3 *La fonction génératrice des cumulants échelonnés λ est convexe.*

DÉMONSTRATION

Par définition pour $k, a \in \mathbb{R}$ et A_n variable aléatoire, on a :

$$\lambda(k) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \int_{\mathbb{R}} e^{nka} \mathbb{P}(A_n \in da)$$

Soit $l \in [0, 1]$ et soient $k, h \in \mathbb{R}$. On considère :

$$\begin{aligned} \log \int_{\mathbb{R}} e^{nalk + na(1-l)h} \mathbb{P}(A_n \in da) &= \log \int_{\mathbb{R}} e^{nalk} e^{na(1-l)h} \mathbb{P}(A_n \in da) \\ &\leq \log \left(\left(\int_{\mathbb{R}} e^{nalk} \mathbb{P}(A_n \in da) \right)^l \left(\int_{\mathbb{R}} e^{na(1-l)h} \mathbb{P}(A_n \in da) \right)^{1-l} \right) \\ &= l \log \left(\int_{\mathbb{R}} e^{nalk} \mathbb{P}(A_n \in da) \right) + (1-l) \log \left(\int_{\mathbb{R}} e^{na(1-l)h} \mathbb{P}(A_n \in da) \right) \end{aligned}$$

L'inégalité ci-dessus est une application de l'inégalité de Hölder.

Par conséquent, si on multiplie par $\frac{1}{n}$ et qu'on passe à la limite, on obtient :

$$\lambda(lk + (1-l)h) \leq l\lambda(k) + (1-l)\lambda(h)$$

Remarque : La convexité de λ implique que λ est continue à l'intérieur de son domaine de définition et est différentiable partout, sauf éventuellement en un nombre dénombrable de points.

Théorème de Varadhan

Précédemment, on a vu que si A_n satisfait au principe des grandes déviations, avec la fonction taux I , alors la fonction λ est donnée par le transformée de Legendre-Fenchel de I , ie :

$$\lambda(k) = \sup_{a \in \mathbb{R}} \{ka - I(a)\}$$

Remplaçons maintenant le produit kA_n par une fonction arbitraire, continue, f de A_n .

Proposition 2.4 (Théorème de Varadhan) *Soit A_n une variable aléatoire indexée par $n \in \mathbb{N}$. Soit f une fonction continue.*

On suppose que A_n satisfait au principe des grandes déviations, avec une fonction taux I , alors la fonction génératrice des cumulants échelonnés s'écrit :

$$\lambda(f) = \lim_{n \rightarrow \infty} \frac{1}{n} \langle e^{nf(A_n)} \rangle = \sup_{a \in \mathbb{R}} \{f(a) - I(a)\}$$

Remarque : Ce théorème est aussi une conséquence du principe des grandes déviations pour A_n et de l'approximation de Laplace. Cependant, le théorème de Varadhan est une conséquence de l'approximation de Laplace seulement dans le cas où A_n est une variable aléatoire réelle, tandis que pour les autres variables aléatoires le théorème de Varadhan s'applique encore et on étend donc ainsi l'approximation de Laplace à ces autres types de variables aléatoires.

Positivité et fonction taux

Proposition 2.5 *Les fonctions taux sont positives.*

DÉMONSTRATION

Comme la fonction génératrice des cumulants échelonnés λ peut s'écrire comme la transformée de Legendre-Fenchel de la fonction taux I , ie :

$$\lambda(k) = \sup_{a \in \mathbb{R}} \{ka - I(a)\}$$

et comme $\lambda(0) = 0$, alors :

$$\begin{aligned}\lambda(0) &= \sup_{a \in \mathbb{R}} \{-I(a)\} \\ &= -\inf_{a \in \mathbb{R}} I(a) \\ &= 0\end{aligned}$$

Donc nécessairement, la fonction taux I est positive.

Convexité et fonction taux

Proposition 2.6 *Les fonctions taux, données par le théorème de Gärtner-Ellis, sont convexes.*

DÉMONSTRATION

Par définition, pour tout $a \in \mathbb{R}$, on a :

$$I(a) = \sup_{k \in \mathbb{R}} \{ka - \lambda(k)\}$$

Soit $l \in [0, 1]$, soient $a, b \in \mathbb{R}$, on considère :

$$\begin{aligned}I(la + (1-l)b) &= \sup_{k \in \mathbb{R}} \{k(la + (1-l)b) - \lambda(k)\} \\ &= \sup_{k \in \mathbb{R}} \{lka + (1-l)kb - \lambda(k)\} \\ &= \sup_{k \in \mathbb{R}} \{lka + (1-l)kb - l\lambda(k) - (1-l)\lambda(k)\} \\ &= \sup_{k \in \mathbb{R}} \{l(ka - \lambda(k)) + (1-l)(kb - \lambda(k))\} \\ &\leq \sup_{k \in \mathbb{R}} \{l(ka - \lambda(k))\} + \sup_{k \in \mathbb{R}} \{(1-l)(kb - \lambda(k))\} \\ &\leq lI(a) + (1-l)I(b)\end{aligned}$$

D'où la convexité de la fonction I .

Remarque : La fonction I donnée par le théorème de Gärtner-Ellis est même strictement convexe, ie convexe sans parties linéaires.

La loi des grands nombres

Si la fonction taux I admet un unique minimum global et zéro a^* , alors :

$$a^* = \lambda'(0) = \lim_{n \rightarrow \infty} \langle A_n \rangle$$

De plus, si I est différentiable en a^* , alors :

$$I'(a^*) = k(a^*)a^* - \lambda(k(a^*)) = 0$$

Le minimum global et zéro a^* a la propriété particulière de correspondre, s'il est unique, à la seule valeur pour laquelle $\mathbb{P}(A_n \in da)$ ne décroît pas exponentiellement. Ainsi autour de cette valeur, on a un effet de concentration et à cause de cet effet on a :

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n \in da^*) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n \in [a^*, a^* + da]) = 1$$

On appelle a^* la valeur la plus probable ou la valeur typique de A_n . L'existence de cette valeur typique est une expression de la loi faible des grands nombres, qui donne que A_n converge vers a^* avec une probabilité de 1.

Par conséquent la théorie des grandes déviations étend la loi des grands nombres en nous donnant la vitesse de convergence de A_n . Plus précisément, soit B un ensemble de valeurs :

$$\mathbb{P}(A_n \in B) = \int_B \mathbb{P}(A_n \in da) \asymp \int_B e^{-nI(a)} da \asymp e^{-n \inf_{a \in B} I(a)}$$

en appliquant l'approximation de Laplace.

Donc $\mathbb{P}(A_n \in B) \rightarrow 0$ exponentiellement vite en n , si $a^* \notin B$, ce qui signifie que $\mathbb{P}(A_n \in B) \rightarrow 1$ exponentiellement vite en n si $a^* \in B$.

Remarque : En général, lorsqu'on peut appliquer la loi des grands nombres à une variable aléatoire, il y a de forte chance pour qu'on puisse également lui appliquer le principe des grandes déviations.

Fluctuation gaussienne et théorème Central-Limite

Le théorème Central-Limite apparaît dans la théorie des grandes déviations lorsque la fonction I possède un seul minimum global et zéro a^* et est différentiable en a^* , car si on approxime $I(a)$ avec le premier terme quadratique :

$$I(a) \simeq \frac{1}{2} I''(a^*) (a - a^*)^2$$

alors, avec l'approximation gaussienne, on obtient :

$$\mathbb{P}(A_n \in da) \simeq e^{-n I''(a^*) (a - a^*)^2} da$$

qu'on peut voir comme une forme faible du théorème Central-Limite.

L'approximation gaussienne précise que les valeurs de A_n autour de a^* sont de l'ordre de $\mathcal{O}(1/\sqrt{n})$, ou les valeurs de nA_n autour de a^* sont de l'ordre de $\mathcal{O}(\sqrt{n})$. Ceci explique le sens du terme grande déviation. D'un côté une petite déviation de A_n est une valeur $A_n = a$ pour laquelle l'approximation quadratique de $I(a)$ est une bonne approximation et pour laquelle, donc, le théorème Central-Limite nous donne la même information que le principe des grandes déviations. D'un autre côté, une grande déviation est la valeur $A_n = a$ pour laquelle $I(a)$ part sensiblement proche de son approximation quadratique et pour laquelle, alors, le théorème Central-Limite ne nous donne pas d'information sur la grande fluctuation de A_n . En ce sens, on peut voir le principe des grandes déviations comme une généralisation du théorème Central-Limite caractérisant les petites aussi bien que les grandes fluctuations d'une variable aléatoire. La théorie des grandes déviations généralise donc le théorème Central-Limite dans le cas où $I(a)$ existe, mais n'admet pas de décomposition de Taylor quadratique autour de son minimum.

2.1.6 Le principe de contraction

Nous avons vu deux théorèmes classiques de la théorie des grandes déviations. Le premier, celui de Gärtner-Ellis, est utilisé pour prouver que le principe des grandes déviations s'applique et calculer la fonction taux associée à la fonction λ . Le second, celui de Varadhan, est utilisé pour calculer λ à partir de I . Le dernier résultat que nous allons voir, appelé principe de contraction, permet de calculer une fonction taux à partir d'une autre.

Proposition 2.7 (Principe de contraction) *Soit A_n une variable aléatoire satisfaisant au principe des grandes déviations, avec pour fonction taux I_A .*

Soit B_n une variable aléatoire telle que $B_n = f(A_n)$, avec f une application continue.

Alors le principe des grandes déviations s'applique à B_n , avec pour fonction taux, pour $b \in \mathbb{R}$:

$$I_B(b) = \inf_{\{a \in \mathbb{R}, f(a)=b\}} I_A(a)$$

DÉMONSTRATION

On sait que :

$$\mathbb{P}(B_n \in db) = \int_{\{a \in \mathbb{R}, f(a)=b\}} \mathbb{P}(A_n \in da)$$

Par conséquent l'approximation de Laplace nous donne alors :

$$\mathbb{P}(B_n \in db) \asymp \exp\left(-n \inf_{\{a \in \mathbb{R}, f(a)=b\}} I_A(a)\right) da$$

Ce qui montre que B_n satisfait au principe des grandes déviations, car A_n satisfait à celui-ci, donc si on applique un tel principe on sait qu'il va exister une fonction taux I_B , telle que :

$$\mathbb{P}(B_n \in db) \asymp e^{-n I_B(b)} db$$

avec, nécessairement :

$$I_B(b) = \inf_{\{a \in \mathbb{R}, f(a)=b\}} I_A(a)$$

Remarque :

- Si la fonction f , appelée contraction de A_n est bijective d'inverse f^{-1} , alors : $I_B(b) = I_A(f^{-1}(b))$.
- $I_B(b) = \infty$, s'il n'existe pas de réel a tel que $f(a) = b$.

On peut interpréter le principe de contraction de la façon suivante : puisque les probabilités de la théorie des grandes déviations sont mesurées sur une échelle exponentielle, la probabilité de toute grande fluctuation peut être approximée par la probabilité de l'événement le plus probable (même s'il est improbable) provoquant une fluctuation.

2.2 Application aux Chaînes de Markov

Considérer, comme nous l'avons fait dans les exemples, des variables aléatoires indépendantes et identiquement distribuées est le cas le plus simple d'application de la théorie des grandes déviations. Il est donc logique de considérer la classe suivante de variables aléatoires : les processus de Markov.

La théorie des grandes déviations pour les chaînes de Markov a principalement été étudiée par Donsker et Varadhan, qui établirent à travers leurs travaux, la base de la théorie des grandes déviations telle qu'on la connaît aujourd'hui. Pour plus de facilité, on considèrera ici des chaînes de Markov finies.

On considère une suite $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ de n variables aléatoires à valeurs dans un ensemble fini \mathcal{S} . Étudions :

$$S_n = \frac{1}{n} \sum_{i=1}^n f(\omega_i)$$

avec f une fonction arbitraire telle que $f : \mathcal{S} \rightarrow \mathbb{R}^d$, avec $d \geq 1$.

La différence avec les variables aléatoires indépendantes et identiquement distribuées est l'ajout de la fonction f et le fait que les ω_i forment une chaîne de Markov définie par :

$$\mathbb{P}(\omega) = \mathbb{P}(\omega_1, \omega_2, \dots, \omega_n) = \rho(\omega_1) \prod_{i=2}^n \mathbb{P}(\omega_i | \omega_{i-1})$$

avec $\rho(\omega_1)$ la probabilité de l'état initial ω_1 et $\mathbb{P}(\omega_i | \omega_{i-1})$ la probabilité conditionnelle de ω_i sachant ω_{i-1} .

On considère ici que la chaîne est homogène.

Pour déterminer si on peut appliquer le principe des grandes déviations calculons, comme dans la partie précédente, la fonction λ . La fonction génératrice de la variable aléatoire S_n peut s'écrire :

$$\begin{aligned} \langle e^{nkS_n} \rangle &= \sum_{\{\omega_1, \omega_2, \dots, \omega_n\}} \rho(\omega_1) e^{kf(\omega_1)} \mathbb{P}(\omega_2 | \omega_1) e^{kf(\omega_2)} \dots \mathbb{P}(\omega_n | \omega_{n-1}) e^{kf(\omega_n)} \\ &= \sum_{\{\omega_1, \omega_2, \dots, \omega_n\}} \mathbb{P}_k(\omega_n | \omega_{n-1}) \dots \mathbb{P}_k(\omega_2 | \omega_1) \rho_k(\omega_1) \end{aligned}$$

avec $\rho_k(\omega_1) = \rho(\omega_1) e^{kf(\omega_1)}$ et $\mathbb{P}_k(\omega_i | \omega_{i-1}) = \mathbb{P}(\omega_i | \omega_{i-1}) e^{kf(\omega_i)}$.

On reconnaît dans la seconde équation une série de produits matriciels faisant intervenir le vecteur $\rho_k(\omega_1)$ et la matrice de transition $\mathbb{P}_k(\omega_i | \omega_{i-1})$. Plus explicitement, si on pose : le vecteur ρ_k de coordonnées $(\rho_k)_i = \rho_k(\omega_1 = i)$ et la matrice \mathcal{P}_k formée des $\mathbb{P}_k(\omega_i | \omega_{i-1})$, ie $(\mathcal{P}_k)_{ij} = \mathbb{P}(j|i)$, alors on peut écrire :

$$\langle e^{nkS_n} \rangle = \sum_{j \in \mathcal{S}} (\mathcal{P}_k^{n-1} \rho_k)_j$$

La fonction λ se déduit de cette expression en déterminant le comportement asymptotique du produit $\mathcal{P}_k^{n-1} \rho_k$ en utilisant l'étude des matrices positives, avec le théorème de Perron-Frobenius. Dépendant de la forme de \mathcal{P} , trois cas se présentent :

Cas 1 : \mathcal{P} est ergodique (irréductible et apériodique) et admet donc une unique distribution stationnaire π telle que $\pi \mathcal{P} = \pi$.

Dans ce cas \mathcal{P}_k a une unique valeur propre dominante ou principale $\zeta(\mathcal{P}_k)$ telle que :

$$\langle e^{nkS_n} \rangle \asymp \zeta(\mathcal{P}_k)^n$$

Donc :

$$\lambda(k) = \log \zeta(\mathcal{P}_k)$$

avec $\zeta(\mathcal{P}_k)$ analytique en k .

D'après le théorème de Gärtner-Ellis, on peut alors conclure que le principe des grandes déviations s'applique, avec la fonction taux :

$$I(s) = \sup_{k \in \mathbb{R}} \{ks - \log \zeta(\mathcal{P}_k)\}$$

Cas 2 : \mathcal{P} n'est pas irréductible, ie admet plusieurs distributions stationnaires.

Dans ce cas λ existe en général mais dépend de la distribution initiale. De plus λ peut être non différentiable et alors on ne peut plus appliquer le théorème de Gärtner-Ellis. Ceci se produit, par exemple lorsque deux valeurs propres ou plus sont en compétition pour être la valeur propre dominante pour différentes distributions initiales.

Cas 3 : \mathcal{P} n'admet pas de distribution stationnaire.

Dans ce cas, le principe des grandes déviations est en général non applicable. En fait, la plupart du temps, même la loi des grands nombres ne s'applique pas.

2.3 Bilan

Nous venons de voir un aperçu de la théorie des grandes déviations. Dans la dernière section, le lien entre grandes déviations et chaînes de Markov est fait, cependant, on remarque qu'il faut avoir une chaîne de Markov ayant des caractéristiques assez remarquables pour qu'on puisse immédiatement lui appliquer le principe des grandes déviations. Nous allons voir, par la suite, que malheureusement ces conditions ne sont pas vérifiées par la chaîne de Markov qui va nous intéresser, cependant nous aboutirons tout de même à des résultats intéressants grâce à la théorie générale des grandes déviations. Nous sommes maintenant en mesure de faire l'étude du modèle d'Ehrenfest.

Chapitre 3

Le modèle d'Ehrenfest

C'est en 1907 que Paul et Tatyana Ehrenfest introduisent le modèle qui porte aujourd'hui leur nom. C'est un modèle probabiliste simple, qui permet de décrire l'évolution de la pression d'un gaz, évolution macroscopique irréversible dans le temps, par l'évolution microscopique réversible des molécules composant ce gaz. Ils proposent ce modèle en réponse aux vives critiques essuyées par la théorie cinétique des gaz proposée par Boltzmann, qui énonçait des équations physiques qui, elles, étaient inchangées si on inverse le cours du temps, alors qu'elles devaient décrire un phénomène irréversible.

3.1 Description du modèle

3.1.1 Description originale donnée par les époux Ehrenfest

On considère N boules numérotées de 1 à N et réparties dans deux urnes. L'urne A contient au départ un nombre P_0 de boules et l'urne B $Q_0 = N - P_0$. On prend ensuite un sac de lotterie dans lequel on met sur des bouts de papiers les numéros de 1 à N . À chaque fois qu'on tire un numéro de ce sac, on prend la boule, dont le numéro correspond, de l'urne dans laquelle elle se trouve et on la met dans la deuxième urne (il est intéressant de noter au cours des tirages les numéros sortis). Il faut ensuite remettre le papier contenant le numéro qu'on a tiré dans le sac de lotterie. On continue ainsi, selon le nombre de tirages qu'on souhaite effectuer.

3.1.2 Description mathématique et physique

On considère $2R$ balles numérotées de 1 à $2R$ et réparties dans deux urnes A et B . Supposons qu'au départ, on a $R + n$ balles dans le compartiment A , avec $-R \leq n \leq R$. On choisit ensuite aléatoirement un entier entre 1 et $2R$ de manière uniforme, puis la boule dont le numéro est choisi est déplacée dans l'urne où elle n'était pas auparavant. Ce processus est répété s fois.

Cette formulation nous donne un modèle simple et pratique de l'échange de chaleur entre deux corps isolés à des températures différentes. En effet, ce modèle peut être considéré à la fois pour l'évolution de la pression d'un gaz ou pour sa température (les deux quantités étant liées). La température étant symbolisée par le nombre de balles dans les urnes et l'échange de chaleur est ici considéré comme un processus aléatoire, comme dans la théorie cinétique de la matière et non pas comme un processus ordonné, comme dans la théorie classique de la thermodynamique.

C'est pourquoi ce modèle a donné lieu à des discussions sur le paradoxe suivant : vouloir expliquer une évolution thermodynamique avec la théorie cinétique. En effet, la thermodynamique classique nous donne que l'évolution de la chaleur entre deux corps à des températures différentes est irréversible, alors que la théorie cinétique considère les corps comme des systèmes dynamiques, dont chaque état sera atteint au moins une fois, d'après la théorie de Poincaré.

Le problème pointé par Zermelo, entre autres, est que l'irréversibilité en thermodynamique et les effets récurrents des systèmes dynamiques sont incompatibles.

Boltzmann a cependant fait remarquer, par la suite, que les cycles de Poincaré (un cycle commençant à un état i du système et finissant quand le système retourne à cet état) étaient très long par rapport à la durée moyenne d'une expérience et donc qu'ainsi la théorie cinétique pouvait s'appliquer. Cette affirmation a beaucoup été contestée et c'est seulement après un long travail

d'Ehrenfest et de Smoluchovski, que l'irréversibilité en thermodynamique a été interprétée correctement d'un point de vue statistique.

3.2 Étude probabiliste du modèle

3.2.1 Un processus de Markov

On considère l'espace d'états $\mathcal{S} = \{0, 1, 2, 3, \dots, 2R\}$ donnant les différents états de l'urne A . En effet dans cette urne on peut avoir aucune balle ou tout aussi bien $2R$ balles.

Soit $(X_n)_n$ une famille de variables aléatoires à valeurs dans \mathcal{S} déterminant à chaque instant le nombre de balles dans l'urne A .

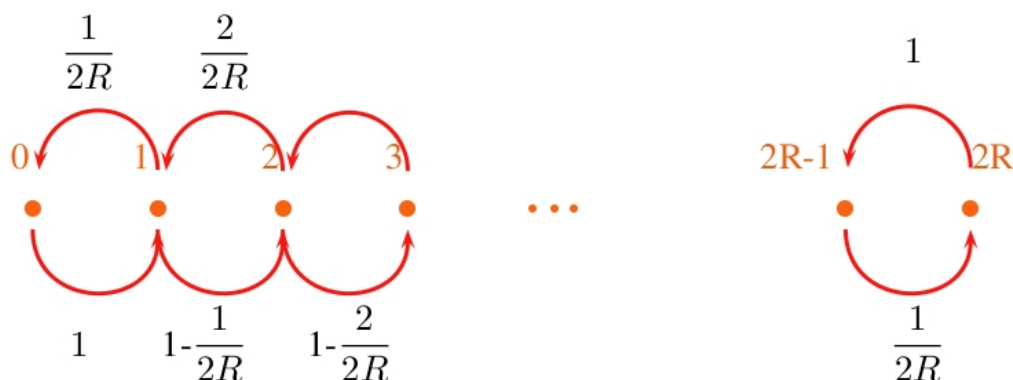
Proposition 3.1 *Le processus $(X_n)_n$ est une chaîne de Markov de matrice de transition :*

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1/2R & 0 & (2R-1)/2R & 0 & \dots & 0 & 0 & 0 \\ 0 & 2/2R & 0 & (2R-2)/2R & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & (2R-1)/2R & 0 & 1/2R \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}$$

En d'autres termes, si on note $\mathcal{P} = (p_{ij})$, pour $i, j \in \mathcal{S}$, on a :

$$p_{ij} = \begin{cases} 1 - \frac{i}{2R}, & \text{si } j = i + 1 \\ \frac{i}{2R}, & \text{si } j = i - 1 \\ 0 & \text{sinon} \end{cases}$$

On peut représenter la chaîne de Markov de la façon suivante, avec les probabilités de passages d'un état à l'autre en noir et le nombre de balles dans l'urne A à un moment donné en orange.



Voyons maintenant les propriétés de cette chaîne de Markov.

Proposition 3.2 *La chaîne $(X_n)_n$ est irréductible, récurrente et périodique de période 2.*

DÉMONSTRATION

La chaîne est irréductible car $\forall i, j \in \mathcal{S}$, i et j intercommuniquent. En effet, si on part d'un état i , on peut atteindre tous les autres états, au bout d'un certain temps. Par exemple, la matrice \mathcal{P} , nous donne que 0 communique avec 1, car $p_{01} > 0$, mais également que 1 communique avec 0 et avec 2, ainsi 0 communique avec lui-même, 1 et 2. Par conséquent, on obtient bien l'irréductibilité de la chaîne.

Montrons que la chaîne est récurrente.

On considère la matrice \mathcal{P}^2 , ie la matrice des 2-ièmes probabilités de transition.

$$\mathcal{P}^2 = \begin{pmatrix} 1/2R & 0 & (2R-1)/2R & 0 & \cdots & 0 & 0 & 0 \\ 0 & (6R-2)/4R^2 & 0 & (4R^2-6R+2)/4R^2 & \cdots & 0 & 0 & 0 \\ 2/4R^2 & 0 & (10R-8)/4R^2 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & (6R-2)/4R^2 & 0 \\ 0 & 0 & 0 & 0 & \cdots & (2R-1)/2R & 0 & 1/2R \end{pmatrix}$$

On voit que $\mathbb{P}(X_2 = i | X_0 = i) = 1, \forall i \in \mathcal{S}$, donc la chaîne est bien récurrente, car tous ses états le sont et communiquent entre eux.

Montrons que la chaîne est périodique de période 2.

On le fait dans le cas où l'urne contient 4 balles au départ, ie $\mathcal{S} = \{0, 1, 2, 3, 4\}$ pour simplifier les calculs.

On procède par récurrence sur $n > 0$ pour montrer que \mathcal{P}^{2n} admet des chiffres sur sa diagonale et que \mathcal{P}^{2n+1} admet une diagonale de zéros, puis une sur et une sous-diagonale de nombres et ainsi de suite, en alternant diagonale de zéros et diagonale de nombres.

- Si $n = 1$, on remarque qu'en effet \mathcal{P} et \mathcal{P}^2 sont bien de la forme voulue :

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\mathcal{P}^2 = \begin{pmatrix} 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 5/8 & 0 & 3/8 & 0 \\ 1/8 & 0 & 6/8 & 0 & 1/8 \\ 0 & 3/8 & 0 & 5/8 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \end{pmatrix}$$

- Supposons maintenant que \mathcal{P}^{2n} admet des nombres sur sa diagonale et que \mathcal{P}^{2n+1} admet des zéros sur sa diagonale, puis une alternance de diagonales de zéros et de diagonales de nombres. Montrons que \mathcal{P}^{2n+2} et \mathcal{P}^{2n+3} sont de la forme voulue.

$\mathcal{P}^{2n+2} = \mathcal{P}^{2n}\mathcal{P}^2$, ie si on note les coefficients de \mathcal{P}^{2n} par α_{ij} , on a :

$$\mathcal{P}^{2n+2} = \begin{pmatrix} \alpha_{00} & \alpha_{01} & \alpha_{02} & \alpha_{03} & \alpha_{04} \\ \alpha_{10} & \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{20} & \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{30} & \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \\ \alpha_{40} & \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} \end{pmatrix} \times \begin{pmatrix} 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 5/8 & 0 & 3/8 & 0 \\ 1/8 & 0 & 6/8 & 0 & 1/8 \\ 0 & 3/8 & 0 & 5/8 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \end{pmatrix}$$

$$\mathcal{P}^{2n+2} = \begin{pmatrix} 1/4\alpha_{00} + 1/8\alpha_{02} & * & * & * & * \\ * & 5/8\alpha_{11} + 3/8\alpha_{13} & * & * & * \\ * & * & 3/4\alpha_{20} + 6/8\alpha_{22} + 3/4\alpha_{24} & * & * \\ * & * & * & 3/8\alpha_{31} + 5/8\alpha_{33} & * \\ * & * & * & * & 1/8\alpha_{42} + 1/4\alpha_{44} \end{pmatrix}$$

Or, on sait que $\alpha_{ii} \neq 0, \forall i \in \mathcal{S}$ (par hypothèse de récurrence) et que $\alpha_{ij} \geq 0, \forall i, j \in \mathcal{S}$, donc les coefficients diagonaux de \mathcal{P}^{2n+2} sont non-nuls.

Maintenant, considérons \mathcal{P}^{2n+3} , sous la forme : $\mathcal{P}^{2n+1}\mathcal{P}^2$, ie si on note les coefficients de \mathcal{P}^{2n+1} par β_{ij} , on a :

$$\mathcal{P}^{2n+3} = \begin{pmatrix} 0 & \beta_{01} & 0 & \beta_{03} & 0 \\ \beta_{10} & 0 & \beta_{12} & 0 & \beta_{14} \\ 0 & \beta_{21} & 0 & \beta_{23} & 0 \\ \beta_{30} & 0 & \beta_{32} & 0 & \beta_{34} \\ 0 & \beta_{41} & 0 & \beta_{43} & 0 \end{pmatrix} \times \begin{pmatrix} 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 5/8 & 0 & 3/8 & 0 \\ 1/8 & 0 & 6/8 & 0 & 1/8 \\ 0 & 3/8 & 0 & 5/8 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \end{pmatrix} = \begin{pmatrix} 0 & * & 0 & * & 0 \\ * & 0 & * & 0 & * \\ 0 & * & 0 & * & 0 \\ * & 0 & * & 0 & * \\ 0 & * & 0 & * & 0 \end{pmatrix}$$

Donc \mathcal{P}^{2n+3} est bien de la forme voulue. Par conséquent, on sait que $\forall n > 0$, \mathcal{P}^{2n} admet des nombres sur sa diagonale et \mathcal{P}^{2n+1} admet des zéros sur sa diagonale. Ainsi, comme la période est définie par : $d(i) = \text{pgcd}\{n, p_{ii}(n) > 0\}$, alors nécessairement $d = 2$.

Pour la démonstration dans le cas général, il suffit de procéder exactement de la même façon que dans le cas où $N = 4$.

Proposition 3.3 *La mesure invariante de la chaîne de Markov est la loi binomiale $\mathcal{B}(2R, 1/2)$ et la chaîne est réversible.*

DÉMONSTRATION

Montrons que $\pi = \left\{ \pi_i = \frac{1}{2^{2R}} C_{2R}^i, i \in \mathcal{S} \right\}$ est une distribution stationnaire pour la chaîne de Markov associée au modèle d'Ehrenfest.

Tout d'abord, on a bien : $\forall i \in \mathcal{S}, \pi_i \geq 0$.

Ensuite :

$$\begin{aligned} \sum_{i \in \mathcal{S}} \pi_i &= \sum_{i=0}^{2R} \frac{1}{2^{2R}} C_{2R}^i \\ &= \frac{1}{2^{2R}} \sum_{i=0}^{2R} C_{2R}^i \\ &= \frac{1}{2^{2R}} 2^{2R} = 1 \end{aligned}$$

Pour finir, montrons que $\pi_j p_{ji} = \pi_i p_{ij}$, pour tout $i, j \in \mathcal{S}$, car comme la chaîne est irréductible, alors on aura démontré qu'à la fois π est une distribution stationnaire et que la chaîne est réversible, d'après le théorème (1.14).

$$\pi_i p_{ij} = \frac{1}{2^{2R}} C_{2R}^i p_{ij} \begin{cases} \frac{1}{2^{2R}} C_{2R}^i \left(1 - \frac{i}{2R}\right) & \text{si } j = i + 1 \\ \frac{1}{2^{2R}} C_{2R}^i \frac{i}{2R} & \text{si } j = i - 1 \\ 0 & \text{sinon} \end{cases}$$

et :

$$\pi_j p_{ji} = \frac{1}{2^{2R}} C_{2R}^j p_{ji} \begin{cases} \frac{1}{2^{2R}} C_{2R}^j \left(1 - \frac{j}{2R}\right) & \text{si } j = i + 1 \\ \frac{1}{2^{2R}} C_{2R}^j \frac{j}{2R} & \text{si } i = j - 1 \\ 0 & \text{sinon} \end{cases}$$

ie

$$\pi_j p_{ji} = \begin{cases} \frac{1}{2^{2R}} C_{2R}^{i-1} \left(1 - \frac{i-1}{2R}\right) & \text{si } j = i - 1 \\ \frac{1}{2^{2R}} C_{2R}^{i+1} \frac{i+1}{2R} & \text{si } j = i + 1 \\ 0 & \text{sinon} \end{cases}$$

Ainsi, si $j = i - 1$, on a :

$$\begin{aligned} \pi_j p_{ji} &= \frac{1}{2^{2R}} C_{2R}^{i-1} \left(1 - \frac{i-1}{2R}\right) \\ &= \frac{1}{2^{2R}} \frac{(2R)!}{(i-1)!(2R-i+1)!} \frac{2R-i+1}{2R} \\ &= \frac{1}{2^{2R}} \frac{(2R-1)!}{(i-1)!(2R-i)!} \\ &= \frac{1}{2^{2R}} \frac{(2R-1)! 2Ri}{(i-1)!(2R-i)! 2Ri} \\ &= \frac{1}{2^{2R}} C_{2R}^i \frac{i}{2R} \end{aligned}$$

D'où $\pi_{i+1}p_{i+1,i} = \pi_i p_{i,i+1}$.

De la même façon, si on considère le cas où $j = i + 1$, on trouve que $\pi_{i-1}p_{i-1,i} = \pi_i p_{i,i-1}$. De plus, pour tout $j \neq i + 1$ et $j \neq i - 1$, on a $\pi_i p_{ij} = 0 = \pi_j p_{ji}$.

Donc la distribution π vérifie bien les conditions voulues qui font d'elle une distribution stationnaire et de la chaîne associée, une chaîne réversible.

3.2.2 Temps de récurrence

On s'intéresse ici au temps au bout duquel, si on part d'un état où l'urne A contient $R + n$ balles, l'urne A se retrouve à cet état. En déterminant le temps de récurrence, on va voir si le modèle d'Ehrenfest est bien valable, dans le sens où on souhaite trouver un temps de récurrence très grand dans le cas où on part avec l'urne A pleine et un temps de récurrence faible, dans le cas où on part avec autant de balles dans l'urne A que dans l'urne B .

Proposition 3.4 *Le temps moyen de récurrence de l'état initial, si on part de l'état $R + n$, avec $-R \leq n \leq R$ pour l'urne A est :*

$$\mu_{R+n} = 2^{2R} \frac{(R+n)!(R-n)!}{(2R)!}$$

Remarque : Si $R + n$ et $R - n$ diffèrent de beaucoup, alors le temps moyen de récurrence sera énorme, par exemple si $R = 10000$ et $n = 10000$, on trouve un temps d'environ 10^{6000} années.

Par contre si $R + n$ et $R - n$ sont proches, ce sera le contraire, par exemple si $R = 10000$ et $n = 0$, on trouve un temps moyen de récurrence d'environ 175 secondes.

C'est Smoluchovski qui avait observé ce phénomène et en avait déduit que si on commençait avec un état ayant un temps moyen de récurrence long, alors le processus pouvait être considéré comme irréversible, alors que si le temps de récurrence de l'état de départ est petit, parler d'irréversibilité n'a plus aucun sens.

DÉMONSTRATION

D'après le théorème (1.7) sur les chaînes de Markov, on sait qu'une chaîne irréductible ayant une distribution stationnaire π va alors admettre π comme unique distribution stationnaire et que son temps moyen de récurrence vaut alors, pour tout $i \in \mathcal{S}$,

$$\mu_i = \frac{1}{\pi_i} = 2^{2R} \frac{1}{C_{2R}^i} = 2^{2R} \frac{i!(2R-i)!}{(2R)!}$$

Donc si au départ l'urne A contient $R + n$ balles, alors pour cet état le temps de récurrence est :

$$\mu_{R+n} = 2^{2R} \frac{(R+n)!(R-n)!}{(2R)!}$$

On peut maintenant simuler la chaîne de Markov associée au modèle d'Ehrenfest et observer le temps de récurrence de différents points de départ pour l'urne A . On le fait dans le cas où on a 200 balles et dans le cas où dans l'urne A , on a au départ 100 balles puis 200 balles.

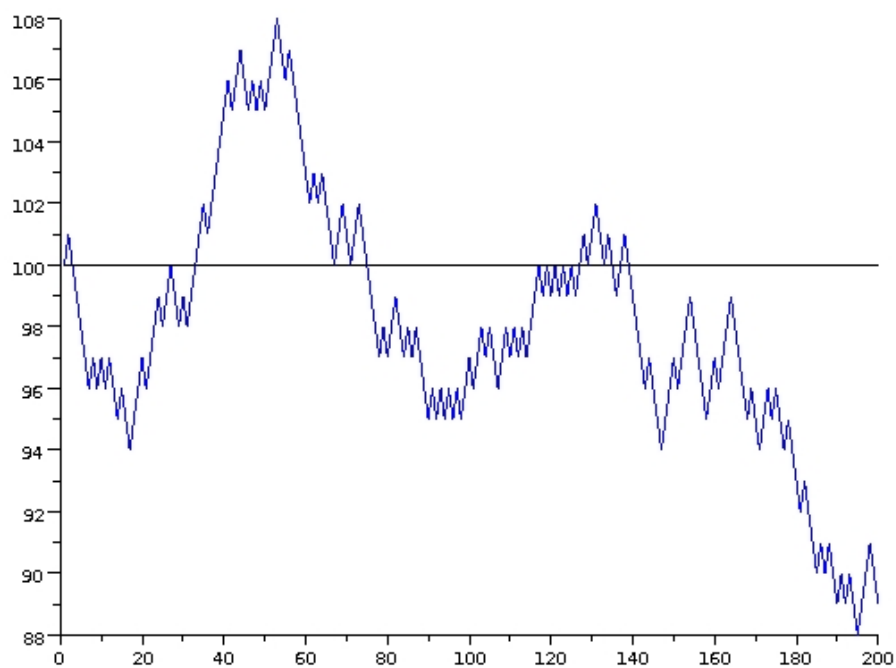


FIG. 3.1 – Cas où on a 100 balles au départ (1/2)

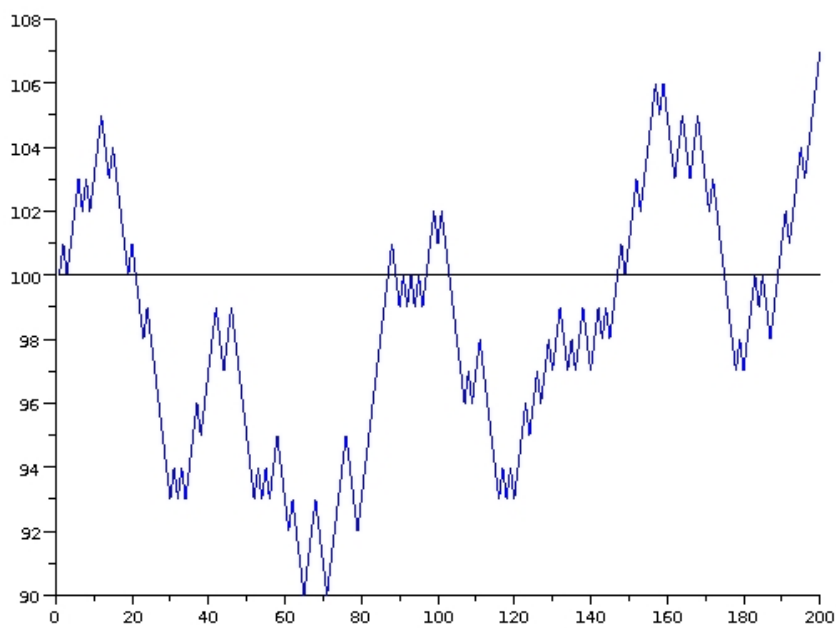


FIG. 3.2 – Cas où on a 100 balles au départ (2/2)

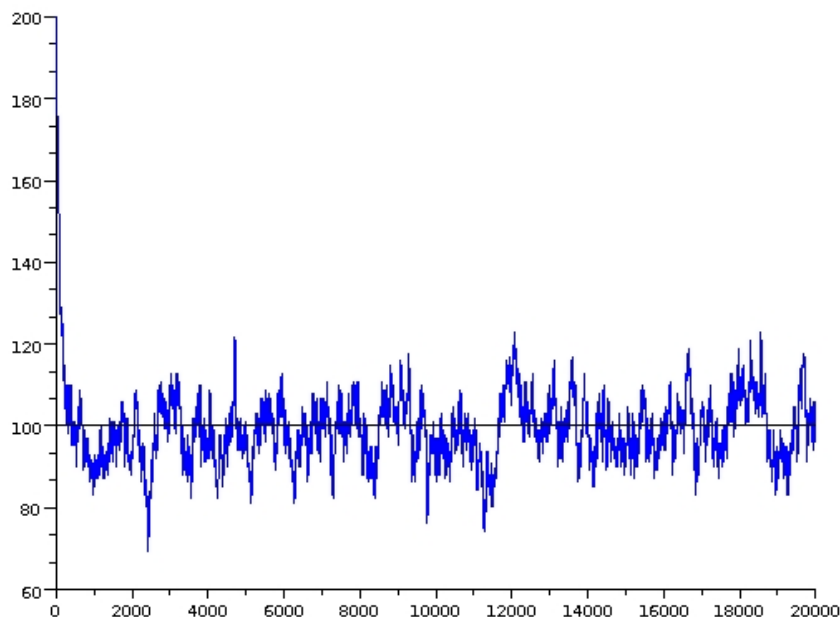


FIG. 3.3 – Cas où on a 200 balles au départ

On observe que dans le cas où au départ les urnes A et B ont le même nombre de balles, le temps de récurrence de l'état de départ est très faible, alors que dans le cas où au départ toutes les balles sont dans l'urne A , le temps de récurrence n'est toujours pas atteint au bout de 20000 étapes. Ces graphiques nous illustrent la formule du temps de récurrence et les propos de Boltzman. Ainsi, le modèle des urnes d'Ehrenfest modélise bien le phénomène d'irréversibilité d'un gaz, bien que ce modèle soit réversible.

3.2.3 Espérance et Variance

La pression au bout de n étapes est de l'ordre de $\frac{X_n}{2R}$. Etant donné que la pression d'un gaz, normalement, diminue jusqu'à trouver un état d'équilibre, on souhaite vérifier, en déterminant les deux premiers moments de la variable aléatoire $\frac{X_n}{2R}$, que notre variable subit bien une décroissance.

Proposition 3.5 On pose $\alpha = 1 - 2/2R$ et $P_n = \frac{X_n}{2R}$. Pour $n \in \mathbb{N}$, on a :

$$\mathbb{E}(P_n) = \frac{1}{2} + \left(\mathbb{E}(P_0) - \frac{1}{2} \right) \alpha^n$$

DÉMONSTRATION

On sait que l'événement X_{n+1} sachant X_n a une probabilité de $1 - \frac{X_n}{2R}$ de devenir l'événement $X_n + 1$ et une probabilité $\frac{X_n}{2R}$ de devenir l'événement $X_n - 1$, ce qui signifie que :

$$\begin{aligned} \mathbb{E}(X_{n+1}|X_n) &= (X_n + 1) \left(1 - \frac{X_n}{2R} \right) + (X_n - 1) \frac{X_n}{2R} \\ &= X_n - \frac{X_n^2}{2R} + 1 - \frac{X_n}{2R} + \frac{X_n^2}{2R} - \frac{X_n}{2R} \\ &= X_n \left(1 - \frac{2}{2R} \right) + 1 \\ &= \alpha X_n + 1 \end{aligned}$$

On raisonne par conditionnement (voir l'annexe A, pour la démonstration du théorème utilisé ici) :

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X_{n+1}|X_n)) &= \mathbb{E}(X_{n+1}) \\ &= \alpha\mathbb{E}(X_n) + 1\end{aligned}$$

Ainsi, par linéarité de l'espérance, on a :

$$\mathbb{E}(P_{n+1}) = \alpha\mathbb{E}(P_n) + \frac{1}{2R}$$

On obtient donc une relation de récurrence sur $n \in \mathbb{N}$, qui nous permet de déterminer l'espérance de P_n .

Montrons par récurrence sur $n \in \mathbb{N}$, que

$$\mathbb{E}(P_n) = \frac{1}{2} + \left(\mathbb{E}(P_0) - \frac{1}{2}\right) \alpha^n$$

- Si $n = 0$,

$$\begin{aligned}\mathbb{E}(P_1) &= \alpha\mathbb{E}(P_0) + \frac{1}{2R} \\ &= \alpha\mathbb{E}(P_0) + \frac{1}{2R} + \frac{1}{2} - \frac{1}{2} \\ &= \alpha\mathbb{E}(P_0) - \frac{\alpha}{2} + \frac{1}{2} \\ &= \alpha \left(\mathbb{E}(P_0) - \frac{1}{2}\right) + \frac{1}{2}\end{aligned}$$

On obtient bien la relation voulue dans le cas où $n = 0$.

- Soit $n \geq 0$, supposons que la relation est vraie au rang n , montrons qu'elle est alors vraie au rang $n + 1$.

$$\begin{aligned}\mathbb{E}(P_{n+1}) &= \alpha\mathbb{E}(P_n) + \frac{1}{2R} \\ &= \alpha \left(\frac{1}{2} + \left(\mathbb{E}(P_0) - \frac{1}{2}\right) \alpha^n\right) + \frac{1}{2R} \quad \text{par hypothèse de récurrence} \\ &= \alpha^{n+1} \left(\mathbb{E}(P_0) - \frac{1}{2}\right) + \alpha \frac{1}{2} + \frac{1}{2R} \\ &= \alpha^{n+1} \left(\mathbb{E}(P_0) - \frac{1}{2}\right) + \left(1 - \frac{2}{2R}\right) \frac{1}{2} + \frac{1}{2R} \\ &= \alpha^{n+1} \left(\mathbb{E}(P_0) - \frac{1}{2}\right) + \frac{1}{2}\end{aligned}$$

D'où le résultat.

Proposition 3.6 On pose $\beta = 1 - \frac{4}{2R}$ et $P_n = \frac{X_n}{2R}$. Pour tout $n \in \mathbb{N}$, on a :

$$\mathbb{V}(P_n) = \frac{1}{8R} + \left(\mathbb{V}(P_0) - \frac{1}{8R}\right) \beta^n + \left(\mathbb{E}(P_0) - \frac{1}{2}\right)^2 (\beta^n - \alpha^{2n})$$

DÉMONSTRATION

On raisonne de la même façon que dans la démonstration précédente.

Tout d'abord, on sait que :

$$\begin{aligned}
\mathbb{E}(X_{n+1}^2|X_n) &= (X_n + 1)^2 \left(1 - \frac{X_n}{2R}\right) + (X_n - 1)^2 \frac{X_n}{2R} \\
&= (X_n^2 + 2X_n + 1) \left(1 - \frac{X_n}{2R}\right) + (X_n^2 - 2X_n + 1) \frac{X_n}{2R} \\
&= X_n^2 + 2X_n + 1 - \frac{X_n^3}{2R} - \frac{2X_n^2}{2R} - \frac{X_n}{2R} + \frac{X_n^3}{2R} - \frac{2X_n^2}{2R} + \frac{X_n}{2R} \\
&= X_n^2 + 2X_n + 1 - \frac{4X_n^2}{2R} \\
&= X_n^2 \left(1 - \frac{4}{2R}\right) + 2X_n + 1
\end{aligned}$$

On raisonne ensuite par conditionnement :

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(X_{n+1}^2|X_n)) &= \mathbb{E}(X_{n+1}^2) \\
&= \left(1 - \frac{4}{2R}\right) \mathbb{E}(X_n^2) + 2\mathbb{E}(X_n) + 1
\end{aligned}$$

Par définition : $\mathbb{V}(X_n) = \mathbb{E}(X_n^2) - \mathbb{E}(X_n)^2$, d'où :

$$\mathbb{V}(X_{n+1}) + \mathbb{E}(X_{n+1})^2 = \left(1 - \frac{4}{2R}\right) (\mathbb{V}(X_n) + \mathbb{E}(X_n)^2) + 2\mathbb{E}(X_n) + 1$$

Or $P_{n+1} = \frac{X_{n+1}}{2R}$, donc $\mathbb{V}(P_{n+1}) = \frac{\mathbb{V}(X_{n+1})}{4R^2}$ et $\mathbb{E}(P_{n+1}) = \frac{\mathbb{E}(X_{n+1})}{2R}$, alors :

$$\begin{aligned}
4R^2\mathbb{V}(P_{n+1}) + 4R^2\mathbb{E}(P_{n+1})^2 &= \left(1 - \frac{4}{2R}\right) (4R^2\mathbb{V}(P_n) + 4R^2\mathbb{E}(P_n)^2) + 4R\mathbb{E}(P_n) + 1 \\
&= \beta 4R^2(\mathbb{V}(P_n) + \mathbb{E}(P_n)^2) + 4R\mathbb{E}(P_n) + 1
\end{aligned}$$

D'où :

$$\mathbb{V}(P_{n+1}) + \mathbb{E}(P_{n+1})^2 = \beta(\mathbb{V}(P_n) + \mathbb{E}(P_n)^2) + \frac{2}{2R}\mathbb{E}(P_n) + \frac{1}{4R^2}$$

Or, d'après la proposition précédente, on a : $\mathbb{E}(P_{n+1}) = \alpha\mathbb{E}(P_n) + \frac{1}{2R}$, donc :

$$\mathbb{V}(P_{n+1}) + \left(\alpha\mathbb{E}(P_n) + \frac{1}{2R}\right)^2 = \mathbb{V}(P_{n+1}) + \alpha^2\mathbb{E}(P_n)^2 + \frac{1}{4R^2} + \frac{2\alpha}{2R}\mathbb{E}(P_n)$$

Ainsi :

$$\begin{aligned}
\mathbb{V}(P_{n+1}) &= \beta\mathbb{E}(P_n) + \mathbb{E}(P_n)^2(\beta - \alpha^2) + \frac{2}{2R}\mathbb{E}(P_n)(1 - \alpha) \\
&= \beta\mathbb{E}(P_n) + \mathbb{E}(P_n)^2(\beta - \alpha^2) + \frac{2}{2R}\mathbb{E}(P_n) \left(1 - 1 + \frac{2}{2R}\right) \\
&= \beta\mathbb{E}(P_n) + \mathbb{E}(P_n)^2(\beta - \alpha^2) + \frac{4}{4R^2}\mathbb{E}(P_n) \\
&= \beta\mathbb{E}(P_n) + \mathbb{E}(P_n)^2 \left(1 - \frac{4}{2R} - \left(1 + \frac{4}{4R^2} - \frac{4}{2R}\right)\right) + \frac{4}{4R^2}\mathbb{E}(P_n) \\
&= \beta\mathbb{E}(P_n) - \frac{4}{4R^2}\mathbb{E}(P_n)^2 + \frac{4}{4R^2}\mathbb{E}(P_n) \\
&= \beta\mathbb{E}(P_n) + \frac{4}{4R^2}\mathbb{E}(P_n)(1 - \mathbb{E}(P_n)) \\
&= \beta\mathbb{E}(P_n) + \frac{4}{4R^2} \left(\frac{1}{2} + \left(\mathbb{E}(P_0) - \frac{1}{2}\right) \alpha^n\right) \left(1 - \frac{1}{2} - \alpha^n \left(\mathbb{E}(P_0) - \frac{1}{2}\right)\right) \quad \text{d'après la proposition (3.5)} \\
&= \beta\mathbb{V}(P_n) + \frac{4}{4R^2} \left(\frac{1}{2} + \alpha^n \left(\mathbb{E}(P_0) - \frac{1}{2}\right)\right) \left(\frac{1}{2} - \alpha^n \left(\mathbb{E}(P_0) - \frac{1}{2}\right)\right) \\
&= \beta\mathbb{V}(P_n) + \frac{4}{4R^2} \left(\frac{1}{4} - \alpha^{2n} \left(\mathbb{E}(P_0) - \frac{1}{2}\right)^2\right)
\end{aligned}$$

On obtient donc une relation de récurrence sur $n \in \mathbb{N}$ et par récurrence sur n on peut montrer que

$$\mathbb{V}(P_n) = \frac{1}{8R} + \left(\mathbb{V}(P_0) - \frac{1}{8R} \right) \beta^n + \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 (\beta^n - \alpha^{2n})$$

En effet :

- Si $n = 0$,

$$\begin{aligned} \frac{1}{8R} + \left(\mathbb{V}(P_0) - \frac{1}{8R} \right) \beta^0 + \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 (\beta^0 - \alpha^0) &= \frac{1}{8R} + \mathbb{V}(P_0) - \frac{1}{8R} \\ &= \mathbb{V}(P_0) \end{aligned}$$

Donc la formule est vérifiée au rang $n = 0$.

- Soit $n \geq 0$, supposons que la relation est vraie au rang n et montrons-la au rang $n + 1$.

$$\begin{aligned} \mathbb{V}(P_{n+1}) &= \beta \mathbb{V}(P_n) + \frac{4}{4R^2} \left(\frac{1}{4} - \alpha^{2n} \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 \right) \\ &= \beta \left(\frac{1}{8R} + \left(\mathbb{V}(P_0) - \frac{1}{8R} \right) \beta^n + \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 (\beta^n - \alpha^{2n}) \right) + \frac{4}{4R^2} \left(\frac{1}{4} - \alpha^{2n} \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 \right) \\ &= \beta \frac{1}{8R} + \left(\mathbb{V}(P_0) - \frac{1}{8R} \right) \beta^{n+1} + \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 (\beta^{n+1} - \beta \alpha^{2n}) + \frac{4}{4R^2} \left(\frac{1}{4} - \alpha^{2n} \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 \right) \\ &= \beta \frac{1}{8R} + \left(\mathbb{V}(P_0) - \frac{1}{8R} \right) \beta^{n+1} + \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 \left(\beta^{n+1} - \beta \alpha^{2n} - \frac{4}{4R^2} \alpha^{2n} \right) + \frac{4}{4R^2} \frac{1}{4} \\ &= \left(1 - \frac{4}{2R} \right) \frac{1}{8R} + \frac{1}{4R^2} + \left(\mathbb{V}(P_0) - \frac{1}{8R} \right) \beta^{n+1} + \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 \left(\beta^{n+1} - \alpha^{2n} \left(\beta + \frac{4}{4R^2} \right) \right) \\ &= \frac{1}{8R} + \left(\mathbb{V}(P_0) - \frac{1}{8R} \right) \beta^{n+1} + \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 \left(\beta^{n+1} - \alpha^{2n} \left(1 - \frac{4}{2R} + \frac{4}{4R^2} \right) \right) \\ &= \frac{1}{8R} + \left(\mathbb{V}(P_0) - \frac{1}{8R} \right) \beta^{n+1} + \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 (\beta^{n+1} - \alpha^{2n} \alpha^2) \\ &= \frac{1}{8R} + \left(\mathbb{V}(P_0) - \frac{1}{8R} \right) \beta^{n+1} + \left(\mathbb{E}(P_0) - \frac{1}{2} \right)^2 (\beta^{n+1} - \alpha^{2n+2}) \end{aligned}$$

D'où le résultat.

3.2.4 Convergence en probabilité

On s'intéresse maintenant à la variable : $\frac{X_{2Rt}}{2R}$, où t est un entier non nul, lorsque $2R \rightarrow \infty$. Le but est de déterminer vers quelle quantité converge cette variable aléatoire.

Proposition 3.7 *La quantité $\frac{X_{2Rt}}{2R}$, pour $t > 0$, converge en probabilité vers la fonction h définie sur \mathbb{R} , par : $h(t) = \frac{1}{2} + \frac{1}{2}e^{-2t}$.*

DÉMONSTRATION

Déterminons tout d'abord la limite de l'espérance d'une telle quantité. D'après le paragraphe précédent, on sait que :

$$\mathbb{E}(P_n) = \frac{1}{2} + \left(\mathbb{E}(P_0) - \frac{1}{2} \right) \left(1 - \frac{2}{2R} \right)^n$$

Alors, pour $t \in \mathbb{R}$, on a :

$$\mathbb{E} \left(\frac{X_{2Rt}}{2R} \right) = \frac{1}{2} + \left(\mathbb{E} \left(\frac{X_0}{2R} \right) - \frac{1}{2} \right) \left(1 - \frac{2}{2R} \right)^{2Rt}$$

On s'intéresse au cas où au départ toutes les balles se trouvent dans l'urne A . Ainsi $\mathbb{P}(X_0 = 2R) = 1$, ie $\mathbb{E}\left(\frac{X_0}{2R}\right) = 1$, ie :

$$\mathbb{E}\left(\frac{X_{2Rt}}{2R}\right) = \frac{1}{2} + \frac{1}{2}\left(1 - \frac{2}{2R}\right)^{2Rt}$$

Donc

$$\lim_{2R \rightarrow \infty} \mathbb{E}\left(\frac{X_{2Rt}}{2R}\right) = \frac{1}{2} + \frac{1}{2}e^{-2t}$$

Remarque : Ce résultat rappelle la loi de refroidissement de Newton pour les fluides.

Déterminons maintenant la limite de la variance de $\frac{X_{2Rt}}{2R}$.

D'après le paragraphe précédent, on sait également que :

$$\mathbb{V}\left(\frac{X_{2Rt}}{2R}\right) = \frac{1}{8R} + \left(\mathbb{V}\left(\frac{X_0}{2R}\right) - \frac{1}{8R}\right)\left(1 - \frac{4}{2R}\right)^{2Rt} + \left(\mathbb{E}\left(\frac{X_0}{2R}\right) - \frac{1}{2}\right)^2 \left(\left(1 - \frac{4}{2R}\right)^{2Rt} - \left(1 - \frac{2}{2R}\right)^{4Rt}\right)$$

Comme $\mathbb{E}(X_0) = a$, alors $\mathbb{V}(X_0) = 0$, d'où :

$$\mathbb{V}\left(\frac{X_{2Rt}}{2R}\right) = \frac{1}{8R} - \frac{1}{8R}\left(1 - \frac{4}{2R}\right)^{2Rt} + \frac{1}{4}\left(\left(1 - \frac{4}{2R}\right)^{2Rt} - \left(1 - \frac{2}{2R}\right)^{4Rt}\right)$$

Donc :

$$\lim_{2R \rightarrow \infty} \mathbb{V}\left(\frac{X_{2Rt}}{2R}\right) = 0$$

Remarque : On qualifie alors notre chaîne de Markov de quasi-déterministe car sa variance est toujours nulle.

Montrons maintenant la convergence de $\frac{X_{2Rt}}{2R}$:

Soit $\epsilon > 0$,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{X_{2Rt}}{2R} - h(t)\right| > \epsilon\right) &\leq \frac{\mathbb{E}\left(\left|\frac{X_{2Rt}}{2R} - h(t)\right|^2\right)}{\epsilon^2} \quad \text{d'après l'inégalité de Tchebychev} \\ &= \frac{\mathbb{E}\left(\left|\frac{X_{2Rt}}{2R} - \mathbb{E}\left(\frac{X_{2Rt}}{2R}\right) + \mathbb{E}\left(\frac{X_{2Rt}}{2R}\right) - h(t)\right|^2\right)}{\epsilon^2} \\ &\leq \frac{\mathbb{E}\left(\left|\frac{X_{2Rt}}{2R} - \mathbb{E}\left(\frac{X_{2Rt}}{2R}\right)\right|^2\right)}{\epsilon^2} + \frac{\mathbb{E}\left(\left|\mathbb{E}\left(\frac{X_{2Rt}}{2R}\right) - h(t)\right|^2\right)}{\epsilon^2} \\ &\leq \frac{\mathbb{V}\left(\frac{X_{2Rt}}{2R}\right)}{\epsilon^2} + \frac{\mathbb{E}\left(\left|\mathbb{E}\left(\frac{X_{2Rt}}{2R}\right) - h(t)\right|^2\right)}{\epsilon^2} \quad \text{par définition de la variance} \\ &\leq \frac{\mathbb{E}\left(\left|\mathbb{E}\left(\frac{X_{2Rt}}{2R}\right) - h(t)\right|^2\right)}{\epsilon^2} \end{aligned}$$

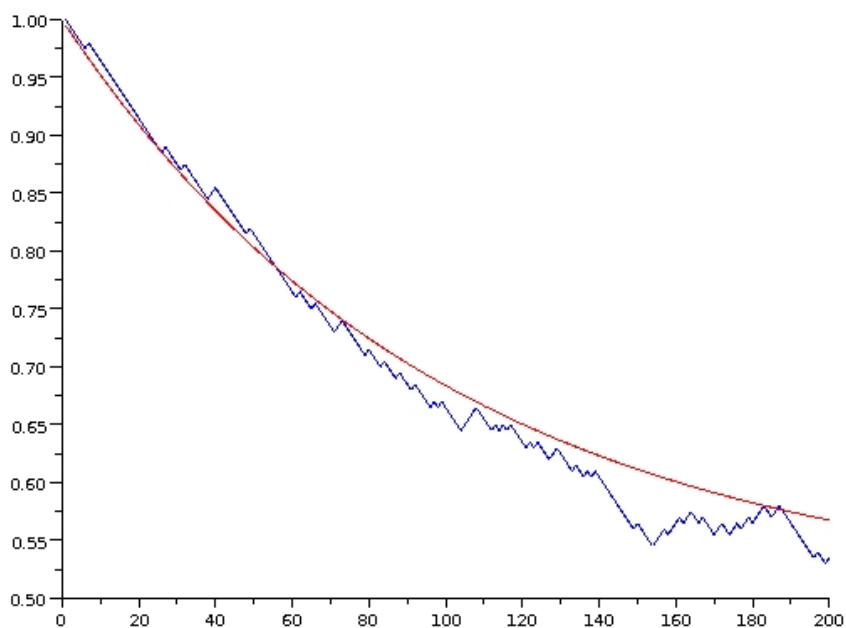
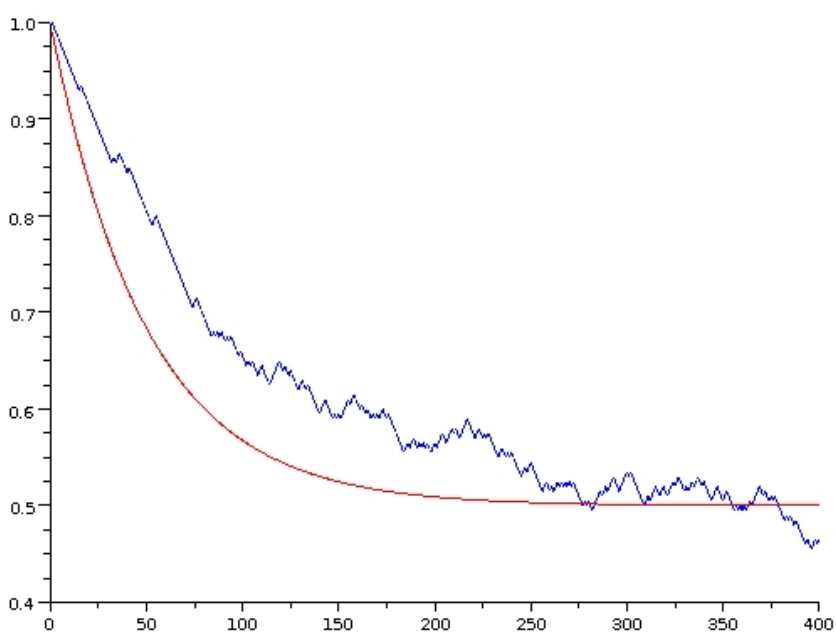
Donc :

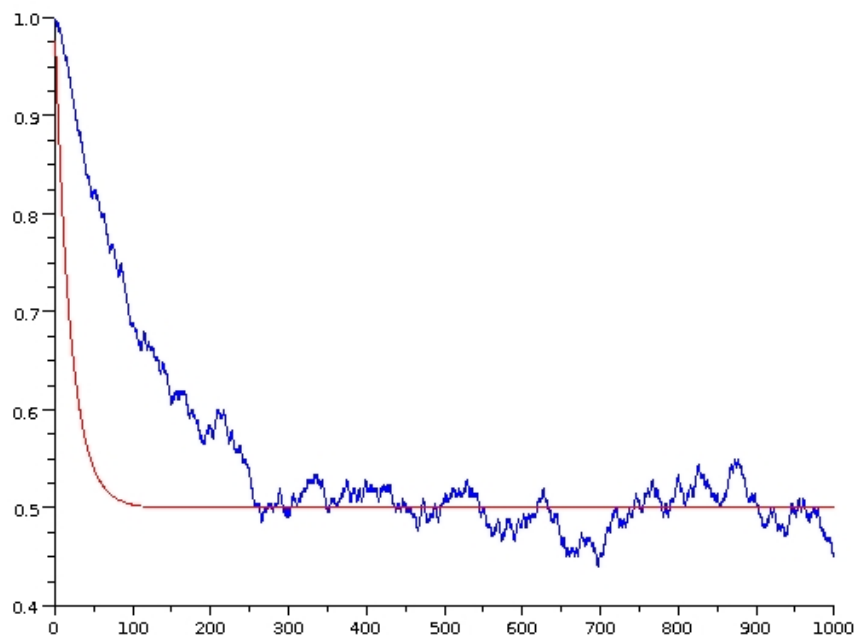
$$\lim_{2R \rightarrow \infty} \mathbb{P}\left(\left|\frac{X_{2Rt}}{2R} - h(t)\right| > \epsilon\right) = 0$$

car $\mathbb{E}\left(\frac{X_{2Rt}}{2R}\right) \rightarrow h(t)$.

On a donc bien démontré la convergence en probabilité de $\frac{X_{2Rt}}{2R}$ vers la fonction h .

On peut observer, à l'aide d'une simulation, la convergence de la variable aléatoire $\frac{X_{2Rt}}{2R}$ vers la fonction $h(t) = \frac{1}{2} + \frac{1}{2}e^{-2t}$. Dans le cas où on a 200 balles et qu'au départ il y a toutes les balles dans l'urne A , on observe pour les valeurs de $t = 1$, $t = 2$ et $t = 5$:

FIG. 3.4 – Cas où $t = 1$ FIG. 3.5 – Cas où $t = 2$

FIG. 3.6 – Cas où $t = 5$

avec en rouge la courbe de la fonction h et en bleu la variable aléatoire. Le but est maintenant de déterminer la vitesse d'une telle convergence.

3.2.5 Grandes Déviations

Nous allons déterminer la vitesse de convergence de la quantité $\frac{X_{2R}}{2R}$ à l'aide de la théorie des grandes déviations. Déterminons tout d'abord, la fonction génératrice des cumulants échelonnés de notre variable aléatoire.

Proposition 3.8 *La fonction génératrice des cumulants échelonnés de $\frac{X_{2R}}{2R}$ est, pour tout $k \in \mathbb{R}$:*

$$\begin{aligned} \lambda(k) &= \lim_{R \rightarrow \infty} \frac{1}{2R} \log \langle e^{2Rk \frac{X_{2R}}{2R}} \rangle \\ &= \log(1 + e^k) - \log(2) \end{aligned}$$

DÉMONSTRATION

Procédons par étapes :

étape 1 : Déterminons la fonction génératrice pour la variable aléatoire $\frac{X_{2n}}{n}$.

Soit $k \in \mathbb{R}$,

$$\begin{aligned} \langle e^{nk \frac{X_{2n}}{n}} \rangle &= \mathbb{E}(e^{kX_{2n}}) \\ &= \sum_{j=0}^{2R} e^{kj} \mathbb{P}(X_{2n} = j) \end{aligned}$$

Or, on sait que la chaîne de Markov X_{2n} est apériodique car issue de la chaîne de Markov X_n de période 2.

On peut donc lui appliquer le théorème limite (1.11) pour une chaîne de Markov irréductible et apériodique, qui nous donne que :

$$\lim_{n \rightarrow \infty} p_{ij}(2n) = \frac{2}{\mu_j}$$

avec μ_j le temps de récurrence de la chaîne de Markov X_{2n} . Or le temps de récurrence de la chaîne de Markov X_{2n} vaut nécessairement, au vu du temps de récurrence de la chaîne X_n :

- Si $X_0 = (\alpha_0, 0, \alpha_2, 0, \dots, 0, \alpha_{2R})$, alors :

$$\frac{2}{\mu_j} = \begin{cases} 2\pi_j & \text{si } j \in \{0, 2, 4, \dots, 2R\} \\ 0 & \text{si } j \in \{1, 3, \dots, 2R-1\} \end{cases}$$

- Si $X_0 = (0, \alpha_1, 0, \alpha_3, \dots, \alpha_{2R-1}, 0)$, alors :

$$\frac{2}{\mu_j} = \begin{cases} 2\pi_j & \text{si } j \in \{1, 3, \dots, 2R-1\} \\ 0 & \text{si } j \in \{0, 2, 4, 6, \dots, 2R\} \end{cases}$$

Considérons le cas où $X_0 = (\alpha_0, 0, \alpha_2, 0, \dots, 0, \alpha_{2R})$.

Déterminons la limite de $\mathbb{P}(X_{2n} = j)$ lorsque $n \rightarrow \infty$, pour cela, on remarque que :

$$\begin{aligned} \mathbb{P}(X_{2n} = j) &= \sum_{k \in \mathcal{S}} \mathbb{P}(X_{2n} = j \cap X_0 = k) \\ &= \sum_{k \in \mathcal{S}} \mathbb{P}(X_{2n} = j | X_0 = k) \mathbb{P}(X_0 = k) \\ &= \sum_{k \in \mathcal{S}} p_{kj}(n) \mathbb{P}(X_0 = k) \end{aligned}$$

Par conséquent :

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_{2n} = j) = \sum_{k \in \mathcal{S}} \frac{2}{\mu_j} \mathbb{P}(X_0 = k) = \frac{2}{\mu_j}$$

Alors, on a :

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}(e^{kX_{2n}}) &= \sum_{j=0}^{2R} e^{kj} \frac{2}{\mu_j} \\ &= \sum_{j=0,2,4,\dots,2R} e^{kj} 2\pi_j \\ &= \sum_{j=0,2,4,\dots,2R} e^{kj} 2C_{2R}^j \frac{1}{2^{2R}} \\ &= \frac{1}{2^{2R}} \sum_{j=0,2,4,\dots,2R} e^{kj} 2C_{2R}^j \\ &= \frac{1}{2^{2R}} ((1+e^k)^{2R} + (1-e^k)^{2R}) \end{aligned}$$

Ainsi, on peut écrire que :

$$\begin{aligned} \frac{1}{2R} \log(\lim_{n \rightarrow \infty} \mathbb{E}(e^{kX_{2n}})) &= -\log(2) + \frac{1}{2R} \log((1+e^k)^{2R} + (1-e^k)^{2R}) \\ &= -\log(2) + \frac{1}{2R} \log\left((1+e^k)^{2R} \left(1 + \left(\frac{1-e^k}{1+e^k}\right)^{2R}\right)\right) \end{aligned}$$

Maintenant, on passe à la limite quand $R \rightarrow \infty$ et on obtient :

$$\lim_{R \rightarrow \infty} \frac{1}{2R} \log\left(\lim_{n \rightarrow \infty} \mathbb{E}(e^{kX_{2n}})\right) = -\log(2) + \log(1+e^k)$$

On admettra ici que passer à la limite sur n , puis sur R revient au même que de passer à la limite sur n pour la variable $\frac{1}{n} \log(\mathbb{E}(e^{kX_{2n}}))$.

Donc, on obtient :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log(\mathbb{E}(e^{kX_{2n}})) = \log(1+e^k) - \log(2)$$

étape 2 : Déterminons la fonction génératrice pour la variable aléatoire $\frac{X_{2n+1}}{n}$.

Soit $k \in \mathbb{R}$,

On raisonne par conditionnement, comme dans les démonstrations précédentes :

$$\begin{aligned}\mathbb{E}(e^{kX_{2n+1}}|X_{2n}) &= e^{k(X_{2n+1})} \left(1 - \frac{X_{2n}}{2R}\right) + e^{k(X_{2n-1})} \frac{X_{2n}}{2R} \\ &= e^{kX_{2n}} e^k \left(1 - \frac{X_{2n}}{2R}\right) + e^{kX_{2n}} e^{-k} \frac{X_{2n}}{2R} \\ &= e^k e^{kX_{2n}} + \frac{1}{2R} (e^{-k} - e^k) X_{2n} e^{kX_{2n}}\end{aligned}$$

Donc :

$$\mathbb{E}(\mathbb{E}(e^{kX_{2n+1}}|X_{2n})) = \mathbb{E}(e^{kX_{2n+1}}) = e^k \mathbb{E}(e^{kX_{2n}}) + \frac{1}{2R} (e^{-k} - e^k) \frac{d}{dk} \mathbb{E}(e^{kX_{2n}})$$

Par conséquent, on peut écrire :

$$\begin{aligned}\mathbb{E}(e^{kX_{2n+1}}) &= e^k \mathbb{E}(e^{kX_{2n}}) + \frac{1}{2R} (e^{-k} - e^k) \frac{d}{dk} \mathbb{E}(e^{kX_{2n}}) \\ &= e^k \mathbb{E}(e^{kX_{2n}}) + \frac{1}{2R} (e^{-k} - e^k) \frac{d}{dk} \sum_{j \in \mathcal{S}} e^{kj} \mathbb{P}(X_{2n} = j) \\ &= e^k \mathbb{E}(e^{kX_{2n}}) + \frac{1}{2R} (e^{-k} - e^k) \sum_{j \in \mathcal{S}} j e^{kj} \mathbb{P}(X_{2n} = j)\end{aligned}$$

Donc quand $n \rightarrow \infty$, on a :

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathbb{E}(e^{kX_{2n+1}}) &= e^k \sum_{j=0,2,4,\dots,2R} \frac{2}{2^{2R}} C_{2R}^j e^{kj} + \frac{1}{2R} (e^{-k} - e^k) \sum_{j=2,4,\dots,2R} \frac{2}{2^{2R}} e^{kj} j C_{2R}^j \\ &= \frac{e^k}{2^{2R}} ((1+e^k)^{2R} + (1-e^k)^{2R}) + \frac{1}{2R} (e^{-k} - e^k) \frac{4R}{2^{2R}} \sum_{j=2,4,\dots,2R} e^{kj} C_{2R-1}^{j-1} \\ &= \frac{e^k}{2^{2R}} ((1+e^k)^{2R} + (1-e^k)^{2R}) + \frac{1}{2R} (e^{-k} - e^k) \frac{4R}{2^{2R}} \sum_{j=1,3,\dots,2R-1} e^{k(j+1)} C_{2R-1}^j \\ &= \frac{e^k}{2^{2R}} ((1+e^k)^{2R} + (1-e^k)^{2R}) + \frac{1}{2R} (1-e^{2k}) \frac{4R}{2^{2R}} \sum_{j=1,3,\dots,2R-1} e^{kj} C_{2R-1}^j \\ &= \frac{e^k}{2^{2R}} ((1+e^k)^{2R} + (1-e^k)^{2R}) + (1-e^{2k}) \frac{1}{2^{2R}} ((1+e^k)^{2R-1} - (1-e^k)^{2R-1}) \\ &= \frac{e^k}{2^{2R}} ((1+e^k)^{2R} + (1-e^k)^{2R}) + (1-e^k)(1+e^k) \frac{1}{2^{2R}} ((1+e^k)^{2R-1} - (1-e^k)^{2R-1}) \\ &= \frac{e^k}{2^{2R}} ((1+e^k)^{2R} + (1-e^k)^{2R}) + \frac{1}{2^{2R}} ((1+e^k)^{2R}(1-e^k) - (1-e^k)^{2R}(1+e^k)) \\ &= \frac{1}{2^{2R}} (1+e^k)^{2R} \left(e^k + \left(\frac{1-e^k}{1+e^k} \right)^{2R} e^k + 1 - e^k - (1+e^k) \left(\frac{1-e^k}{1+e^k} \right)^{2R} \right) \\ &= \frac{1}{2^{2R}} (1+e^k)^{2R} \left(1 - \left(\frac{1-e^k}{1+e^k} \right)^{2R} \right)\end{aligned}$$

Par conséquent :

$$\lim_{R \rightarrow \infty} \frac{1}{2R} \log \lim_{n \rightarrow \infty} \mathbb{E}(e^{kX_{2n+1}}) = -\log(2) + \log(1+e^k)$$

ie :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}(e^{kX_{2n+1}}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}(e^{kX_{2n}})$$

étape 3 : Conclusion

Ainsi, on peut en déduire que $\forall k \in \mathbb{R}$,

$$\lambda(k) = \lim_{R \rightarrow \infty} \frac{1}{2R} \log \langle e^{2Rk \frac{X_{2R}}{2R}} \rangle = \log(1+e^k) - \log(2)$$

Remarque : Pour le calcul de la fonction λ pour X_{2n} , l'état initial n'a pas d'importance, on retrouve le même résultat si on considère que $X_0 = (0, \alpha_1, 0, \dots, \alpha_{2n-1}, 0)$.

Comme la fonction génératrice des cumulants échelonnés λ existe et est différentiable pour tout $k \in \mathbb{R}$, alors on peut appliquer le théorème de Gärtner-Ellis à notre chaîne de Markov. Par conséquent $\frac{X_{2R}}{2R}$ satisfait au principe des grandes déviations avec pour fonction taux, pour $a \in \mathbb{R}$:

$$I(a) = \sup_{k \in \mathbb{R}} \{ka - \lambda(k)\}$$

Déterminons la fonction taux I .

Soit $a \in \mathbb{R}$, on pose

$$f(k) = ka - \lambda(k) = ka - \log(1 + e^k) + \log(2)$$

Plusieurs cas se présentent :

Cas 1 : $0 < a < 1$

Alors $\forall k \in \mathbb{R}$,

$$f'(k) = a - \frac{e^k}{1 + e^k}$$

Donc $f'(k) = 0$, quand $k = \log\left(\frac{a}{1-a}\right)$. La borne supérieure est atteinte en cette valeur de k , on obtient alors :

$$I(a) = a \log(a) + (1-a) \log(1-a) + \log(2)$$

Cas 2 : $a = 0$

Alors $\forall k \in \mathbb{R}$,

$$f'(k) = -\frac{e^k}{1 + e^k} < 0$$

Donc la borne supérieure de f est sa limite en $-\infty$ et on obtient :

$$I(0) = \log(2)$$

Cas 3 : $a = 1$

Alors $\forall k \in \mathbb{R}$,

$$f'(k) = 1 - \frac{e^k}{1 + e^k} > 0$$

Donc la borne supérieure de f est sa limite en $+\infty$ et on obtient :

$$I(1) = \log(2)$$

Cas 4 : $a > 1$

Alors $\forall k \in \mathbb{R}$,

$$f'(k) = a - \frac{e^k}{1 + e^k} > 0$$

Donc la borne supérieure de f est sa limite en $+\infty$ et on obtient :

$$I(a) = +\infty$$

Cas 5 : $a < 0$

Alors $\forall k \in \mathbb{R}$,

$$f'(k) = a - \frac{e^k}{1 + e^k} < 0$$

Donc la borne supérieure de f est sa limite en $-\infty$ et on obtient :

$$I(a) = +\infty$$

Par conséquent la fonction taux associée à la chaîne de Markov $\frac{X_{2R}}{2R}$ vaut :

$$I(a) = \begin{cases} a \log(a) + (1-a) \log(1-a) + \log(2) & \text{si } 0 < a < 1 \\ \log(2) & \text{si } a = 0 \text{ et } a = 1 \\ +\infty & \text{si } a > 1 \text{ et } a < 0 \end{cases}$$

De la sorte, on a démontré que la convergence en probabilité de la variable aléatoire $\frac{X_{2Rt}}{2R}$ vers $h(t) = \frac{1}{2} + \frac{1}{2}e^{-2t}$ est une convergence qui est exponentielle d'après le principe des grandes déviations.

3.3 Bilan

Finalement, le modèle probabiliste proposé par les époux Ehrenfest, se révèle être, malgré sa réversibilité, un bon modèle pour expliquer l'irréversibilité de la propagation d'un gaz dans un récipient. Nous avons en effet, vu que le temps de récurrence de chaque état de l'urne A varie selon le nombre de balles présent dans cet état. Plus le nombre de balles est proche de la moitié du nombre de balles total et plus le temps de récurrence sera faible, alors qu'à l'inverse le temps de récurrence sera beaucoup plus élevé, ce qui est en accord avec le principe de l'irréversibilité. Nous avons également vu l'interprétation mathématique de ce modèle au travers de la théorie des chaînes de Markov d'une part et de celles des grandes déviations d'autre part. En effet, l'étude du modèle est aussi passée par l'étude du comportement asymptotique de la chaîne de Markov considérée, ce qui nous a amené à étudier la convergence exponentielle de la chaîne avec la théorie des grandes déviations. Finalement, cette étude du modèle d'Ehrenfest, exemplifie la conception de modèles mathématiques qui sont en mesure d'expliquer des phénomènes physiques.

Conclusion

Ce stage effectué dans le laboratoire de Probabilités et Modèles aléatoires, m'a permis de prendre un premier contact avec le monde de la recherche. Au cours de mon stage, j'ai pu découvrir d'autres domaines des probabilités, qui m'étaient inconnus, sur lesquels j'ai pu effectuer un travail approfondi, en parcourant différents livres sur un même sujet, afin de voir différents points de vue et en faisant des simulations informatiques sur des exemples concrets afin de cerner au mieux à quoi pouvait servir de telles théories. Outre l'apprentissage de la théorie des grandes déviations et des chaînes de Markov, ce stage m'a également permis de faire un lien entre la physique et les probabilités, avec l'étude du modèle d'Ehrenfest, modèle mathématique créé pour comprendre un phénomène physique. Pour finir, je remercie le laboratoire de Probabilités et Modèles aléatoires d'avoir bien voulu m'accueillir et plus particulièrement mon maître de stage, Raphael Lefevre, qui m'a permis d'effectuer ce stage et m'a suivi pendant toute la durée de celui-ci.

Annexe A

Les probabilités conditionnelles

Dans ce chapitre, se trouvent des notions de base sur les probabilités conditionnelles qui ont été utilisées lors de démonstrations dans les chapitres qui précèdent.

A.1 Probabilité conditionnelle

Définition A.1 Soient A, B deux événements d'un espace de probabilités $(\Omega, \mathcal{A}, \mathbb{P})$.

Si $\mathbb{P}(B) > 0$, alors on définit la probabilité conditionnelle que l'événement B soit réalisé sachant que A l'est, par :

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

On dit aussi probabilité de A sachant B .

Lemme A.1 Soient A, B deux événements. Alors :

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$$

Plus généralement, soient B_1, B_2, \dots, B_n une partition de Ω , alors :

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

DÉMONSTRATION

On peut écrire A sous la forme : $A = (A \cap B) \cup (A \cap B^c)$. On a une union disjointe, donc :

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) \\ &= \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c) \end{aligned}$$

par définition de la probabilité conditionnelle.

A.2 Distribution conditionnelle et espérance conditionnelle, dans le cas discret

Soient X et Y deux variables aléatoires discrètes sur $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition A.2 La fonction distribution conditionnelle de Y sachant $X = x$, notée $F_{Y|X}(\cdot|x)$ est définie par :

$$F_{Y|X}(y|x) = \mathbb{P}(Y \leq y|X = x) \quad \forall x \text{ tel que } \mathbb{P}(X = x) > 0$$

La densité conditionnelle de la probabilité de Y sachant $X = x$, notée $f_{Y|X}(\cdot|x)$ est définie par :

$$f_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x) \quad \forall x \text{ tel que } \mathbb{P}(X = x) > 0$$

Remarques :

- La formule de la densité peut aussi être écrite de la façon suivante : $f_{Y|X} = \frac{f_{X,Y}}{f_X}$.
- Les notions ci-dessus ne sont pas définies si $\mathbb{P}(X = x) = 0$.
- X et Y sont indépendants, alors $\Leftrightarrow f_{Y|X} = f_Y$.
- La densité de probabilité de la distribution de Y , $f_{Y|X}(y|x)$ est une fonction de y .
- La quantité $\sum_y y f_{Y|X}(y|x)$ est appelé espérance conditionnelle de Y sachant $X = x$ et on pose $\psi(x) = \mathbb{E}(Y|X = x)$.
- L'espérance dépend de x .

Définition A.3 Soit $\psi(x) = \mathbb{E}(Y|X = x)$, alors $\psi(X)$ est appelé l'espérance conditionnelle de Y sachant X et est notée $\mathbb{E}(Y|X)$.

Remarque : L'espérance conditionnelle est une variable aléatoire et non un nombre.

Théorème A.1 L'espérance conditionnelle $\psi(X) = \mathbb{E}(Y|X)$ satisfait :

$$\mathbb{E}(\psi(X)) = \mathbb{E}(Y)$$

DÉMONSTRATION

$$\begin{aligned} \mathbb{E}(\psi(X)) &= \sum_x \psi(x) f_X(x) \\ &= \sum_{x,y} y f_{Y|X}(y|x) f_X(x) \\ &= \sum_{x,y} y f_{X,Y}(x,y) \quad \text{car } f_{Y|X} = \frac{f_{X,Y}}{f_X} \\ &= \sum_y f_Y(y) \\ &= \mathbb{E}(Y) \end{aligned}$$

D'où le résultat.

A.3 Distribution conditionnelle et espérance conditionnelle, dans le cas continu

Soient X, Y deux variables aléatoires ayant pour densité f . On cherche à déterminer la distribution conditionnelle de Y sachant $X = x$. Cependant, on ne peut pas la définir comme dans la section précédente. Nous procédons alors comme suit :

Si $f_X(x) > 0$, alors :

$$\begin{aligned} \mathbb{P}(Y \leq y | X = x) &= \frac{\mathbb{P}(Y \leq y \cap x \leq X \leq x + dx)}{\mathbb{P}(x \leq X \leq x + dx)} \\ &\simeq \frac{\int_{\nu=-\infty}^y f(x, \nu) dx d\nu}{f_X(x) dx} \\ &= \int_{\nu=-\infty}^y \frac{f(x, \nu)}{f_X(x)} d\nu \end{aligned}$$

Quand $dx \searrow 0$, le côté gauche de l'équation approche la probabilité de $Y \leq y$ sachant $X = x$. D'où la définition suivante :

Définition A.4 La fonction distribution conditionnelle de Y sachant $X = x$, notée $F_{Y|X}(\cdot|x)$ est définie pour tout x , tel que $f_X(x) > 0$, par :

$$F_{Y|X}(y|x) = \int_{\nu=-\infty}^y \frac{f(x, \nu)}{f_X(x)} d\nu$$

A.3. DISTRIBUTION CONDITIONNELLE ET ESPÉRANCE CONDITIONNELLE, DANS LE CAS CONTINU 67

La fonction densité conditionnelle de Y sachant $X = x$, notée $f_{Y|X}(\cdot|x)$ est définie pour tout x , tel que $f_X(x) > 0$, par :

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$$

Remarque : On a toujours $f_X(x) = \int_{\mathbb{R}} f(x, y) dy$ et on en déduit que :

$$f_{Y|X}(y|x) = \frac{f(x, y)}{\int_{\mathbb{R}} f(x, y) dy}$$

On peut également retenir la deuxième définition de la façon suivante : $f_{Y|X} = \frac{f_{X,Y}}{f_X}$.

Théorème A.2 L'espérance conditionnelle $\psi(X) = \mathbb{E}(Y|X)$ satisfait :

$$\mathbb{E}(\psi(X)) = \mathbb{E}(Y)$$

Remarque : La démonstration du théorème (A.2) est similaire à celle du théorème (A.1).

Théorème A.3 L'espérance conditionnelle $\psi(X)$ satisfait :

$$\mathbb{E}(\psi(X)g(X)) = \mathbb{E}(Yg(X))$$

avec g , fonction pour laquelle les deux espérances considérées ci-dessus sont bien définies.

Annexe B

Programmes Scilab

Les programmes détaillés ci-dessous ont été élaborés en se plaçant d'un point de vue microscopique, c'est-à-dire en comptant à chaque itération le nombre de balles se trouvant dans l'urne A .

B.1 Programme 1 :

Ce programme permet de donner le nombre de balles se trouvant dans l'urne A après une étape de la chaîne de Markov associée au modèle d'Ehrenfest.

```
function y = suivant(V)
N= length(V)
u=grand(1,1,'uin',1,N )
    if V(u)==1
        V(u)=0
    else
        V(u)=1
    end
y=V
```

B.2 Programme 2 :

Ce programme permet de donner la marche de la chaîne, c'est-à-dire le nombre de balles se trouvant dans l'urne A à chaque étape sachant que cette fois on fait un nombre d'étapes n qu'on choisit.

```
function y = marche(A,n)
getf suivant.sci;
    for i=1 :n
        V(i)=sum(A);
        A=suivant(A);
    end;
y=V
```

B.3 Programme 3 :

Ce programme est celui qui a permis de faire tous les graphiques du chapitre sur le modèle d'Ehrenfest.

```
clf;
A=ones(1,200);
n=200
getf marche.sci;
N=length(A);
X=linspace(1,n,n);
Y=marche(A,n);
plot2d(X,Y,2)
```

```
plot2d(X,Y/N,2)  
plot2d(X,(1/2+1/2*exp(-2*X/N)),5)
```

Bibliographie

- [Cha00] Cheng Shang Chang. *Performance Guarantee in Communication Networks*. Springer, 2000.
- [dH08] Frank den Hollander. *Large Deviations*. American Mathematical Society, 2008.
- [DS89] Jean-Dominique Deuschel and Daniel W. Stroock. *Large Deviations*. American mathematical Society, 1989.
- [EE07] Paul Ehrenfest and Tatyana Ehrenfest. Über zwei bekannte einwände gegen das botzmannsche h-theorem. *Physikalische Zeitschrift*, 9, 1907.
- [GS92] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes, second edition*. Clarenton Press, 1992.
- [GS01] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes, third edition*. Clarenton Press, 2001.
- [Kac47] Mark Kac. Random walk and the theory of brownian motion. *The American Mathematical Monthly*, 54, 1947.
- [Mal09] Florent Malrieu. Urnes et particules. *Epreuve de l'agrégation externe de mathématique*, 2009.
- [son71] Charles Scribner's sons. *Dictionary of scientific biography IV*. Charles Coulston Gillispie, 1971.
- [Tou09] Hugo Touchette. The large deviation approach to statistical mechanics. 2009.