République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOULOUD MAMMERI DE TIZI-OUZOU
FACULTE DE SCIENCES - DEPARTEMENT DE MATHEMATIQUES

Mémoire de Master de Mathématiques Appliquées

Option: Processus Aléatoires et Statistique de la Décision

Le théorème ergodique et les chaînes de Markov Itératives

présenté par

HAMROUN KAHINA ,

sous la Direction de Mme ZIDI YAMINA,

le 30/10/2013, devant le jury:

BOUDIBA MOHAND AREZKI, MCA, UMMTO, Président;
Mme ZIDI YAMINA, Chargée de Recherche, UMMTO, Rapporteur;
MAMOU MOHAMED, Chargé de Recherche, UMMTO, Examinateur.

Remérciements

J'adresse mes vifs remérciements à Monsieur BOUDIBA Mouhand Arezki, pour l'honneur qu'il me fait en présidant ce jury, et pour ses critiques et ses encouragements pour faire aboutir ce mémoire.

Je remércie également MAMOU Mohamed pour avoir accepté de faire partie du jury.

Mes remérciements vont aussi à Mme LADJIMI Fetima pour ses conseils et ses remarques sur le théorème ergodique et sur la méthode de construction du mouvement brownien standard.

Je présente mes sinsères remérciements à ma directrice, Mme ZIDI Yamina, pour m'avoir proposer ce thème et aussi pour ses orientations, ses critiques et conseils qui ont permis l'aboutissement de ce projet.

Dédicaces

Je dédie ce mémoire à :

– Mes chers parents qui m'ont toujours soutenus;
– Mon cher mari Arezki et ma petite fille adorée Thanina ;
– Mes chers frères et soeurs ainsi que ma belle famille.

Table des matières

Introduction

Nous nous intéréssons aux applications du théorème ergodique aux chaînes de Markov itératives. Une chaîne de Markov itérative est la donnée d'un espace mesurable $(E,\mathcal{E})$, d'une suite $(Y_n)_n$ de variables aléatoires independantes identiquement distribuées (i.i.d) à valeurs dans E, d'une fonction ou procédure $x \mapsto f_y(x)$ de E dans E, pour $y \in E$ fixé et d'un processus (chaîne de Markov) $(X_n^x)_n$ défini par $X_0^x = x$ avec $x \in E$, et pour $n > 0$,

$$X_n^x = f_{Y_n} \circ \cdots \circ f_{Y_1}(x)$$

$$i.e. \;\; X_n^x = f_{Y_n} \circ X_{n-1}^x \qquad (1)$$

L'étude des chaînes de Markov itératives du type $(X_n^x)_n$ est assez intense ces dernières années. De nombreux travaux y sont consacrés: Mirek [11], Diaconis et Freedman [4], O.Lee [8],[9]... Cela s'explique par leurs applications à l'étude de nombreux phénomènes et leur utilisation dans d'autres branches des mathémathiques appliquées en particulier les systèmes dynamiques, la simulation exacte,...

Partant du fait que si X est une variable aléatoire et T une transformation qui préserve la mesure P, on peut considérer le processus $(X_n)_n$ défini par

$$X_n = T^{n-1} oX \qquad (2)$$

Les sommes de Birkhoff associées sont les variables aléatoires

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

Si on considère la chaîne de Markov itérative (1), par analogie on peut considérer les sommes de Birkhoff associées

$$\frac{1}{n}\sum_{k=1}^{n} X_n^x.$$

Le processus $(X_n^x)_n$ est l'analogue du processus $(X_n)_n$. Nous avons, pour les processus stationnaires, dans des conditions raisonables, le théorème ergodique de Birkhoff et ses differentes variantes dont la loi forte des grands nombres est un cas particulier. Nous voulons savoir si des propriétés similaires, dans des conditions appropriées, ont lieu pour des chaînes de Markov itératives.

Une autre motivation pour l'application du théorème ergodique aux chaînes de Markov itératives est que le théorème ergodique s'applique aussi aux chaînes de Markov à probabilités de transition stationnaires et admettant une mesure stationnaire. Dans le cas dénombrable, si $(\pi_x)_x$ est la loi stationnaire, nous connaissons la formule élémentaire

$$\frac{1}{n}\sum_{1}^{n} p_{xy}^{(k)} \to \pi_y$$

qui est le théorème ergodique dans ce cas.

Notre problème est de voir dans quelles conditions on a des résultats similaires pour les chaînes de Markov itératives. Les problèmes liés à l'étude de ces chaînes de Markov sont de plusieurs types:
1. Conditions pour l'existance et l'unicité d'une loi stationnaire;
2. Etude de la récurence;
3. Existance d'une limite (prèsque sûrement, en probabilité, en loi ou dans un sens);
4. Enfin l'étude des processus d'itération à droite qui présente un intérêt en particulier pour leurs applications aux algorithmes de simulation exacte (algorithmes de Propp-Wilson).

Nous nous limitons à l'étude des sommes de Birkhoff associées, pour les chaîne de Markov itératives qui admettent une loi stationnaire. Dans le premier chapitre nous faisons le point sur les théorèmes ergodiques. Dans le deuxième chapitre nous synthétisons quelques propriétés des chaînes de Markov et des noyaux de transition. Enfin dans le troisième chapitre, nous introduisons les chaînes itératives et l'application du théorème ergodique à ce type de chaînes de Markov.

Chapitre 1

Eléments sur le théorème ergodique

Les théorèmes ergodiques de Birkhoff et de Von Neumann sont les premiers résultats importants à avoir été démontré en théorie ergodique. Nous fixons ci-après la terminologie et donnons les définitions et propriétés des sommes de Birkhoff.Cf[3]

1.1 Processus stationnaire

L'espace $(\mathbb{R}^\infty,\mathcal{B}_{\mathbb{R}^\infty})$:
Si T est un ensemble non vide, un processus aléatoire est une famille $(X_t)_{t\in T}$ de variables aléatoires définies sur le même espace probabilisé $(\Omega,\mathcal{A},P)$, à valeurs dans un espace mesurable $(E,\mathcal{E})$, qui peut être par exemple $(\mathbb{R}^n,\mathcal{B}_{\mathbb{R}^n})$, auquel cas on dira que c'est un processus vectoriel. Si $E = \mathbb{R}$ (respectivement $E = \mathbb{C}$), on dira que c'est un processus réel (respectivement complexe). En général $T = \mathbb{R}_+$. Souvent $T = \mathbb{N}$ ou $T = \mathbb{Z}$. Le paramétre t est en général interprété comme le temps discrêt ou non. On parlera alors de processus à temps discrêt ou continu.
On considère l'ensemble $\mathbb{R}^N$, des suites de nombres réels, noté encore $\mathbb{R}^\infty$, i.e. $\mathbb{R}^\infty = \{x = (x_n)_n \quad i.e. \; x : \mathbb{N} \longrightarrow \mathbb{R}, x(n) = x_n \in \mathbb{R}\}$. On dit que R est un rectangle si R est de la forme $R = I_1 \times I_2 \times \cdots \times I_n$, où les I_k sont des intervalles de $\mathbb{R}$. Si R est l'ensemble de tous les rectangles, soit $\mathcal{B}_\infty$, la tribu engendrée par R.

Définition 1. *Soit $X = (X_n)_n$ un processus aléatoire à valeurs dans $\mathbb{R}$, alors*

$$\forall B \in \mathcal{B}_\infty, \ \ \{X \in B\} \in \mathcal{A}.$$

Construction de $\hat{P}$:

Si $X = (X_1, X_2,..., X_n,...)$ est un processus aléatoire défini sur $(\Omega,\mathcal{A},P)$, pour B de la forme $B = B_1 \times B_2 \times ... \times B_n$ avec $B_k \in \mathcal{B}_\mathbb{R}$, on peut définir $\hat{P}$ sur $\mathcal{B}_{\mathbb{R}^n}$, on posant

$$\hat{P}[B] = P[X_1 \in B_1,\ X_2 \in B_2,...,\ X_n \in B_n]$$

(B est appelé ensemble cylindriques). $\hat{P}$ est une loi sur $(\mathbb{R}^n,\mathcal{B}_{\mathbb{R}^n})$ la loi fini dimensionnelle du processus $(X_n)_n$ c'est la loi définie sur les rectangles qui étendue aux ensembles cylindriques. l'extention de $\hat{P}$ à $(\mathbb{R}^\infty,\mathcal{B}_\infty)$ est par définition la loi du processus $(X_n)_n$.

Théorème 1. *Avec les notations ci-dessus, il existe une seule mesure de probabilité $\hat{P}$ restreinte à $\mathcal{B}_{\mathbb{R}^n}$, vérifie pour tout ensemble cylindrique B*

$$\hat{P}[B] = P[X \in B] = P[X_1 \in B_1,\ X_2 \in B_2,...,\ X_n \in B_n]$$

Définition 2. *Soit $(X_n)_n$ un processus aléatoire. On dit que $(X_n)_n$ est un processus stationnaire si $\forall k \in \mathbb{N}, \ \ \forall B \in \mathcal{B}_\infty$*

$$P[(X_1,\ X_2,\ X_n,...) \in B] = P[(X_{k+1},\ X_{k+2},\ X_{k+n},...) \in B]$$

Propsition 1. *Soit $(X_n)_n$ un processus aléatoite et $\varphi : \mathbb{R}^\infty \longrightarrow \mathbb{R}$ une application mesurable. Si $(Y_k)_k$ est le processus défini par*

$$\forall k \geq 1, \ \ Y_k = \varphi(X_k,\ X_{k+1},...)$$

alors $(Y_k)_k$ est un processus stationnaire.

Remarques:

1. La définition de la stationnarité d'un processus exprime que la loi du processus est invariante par translation dans le temps, si l'indice du processus est interprété comme le temps.
2. Il résulte immédiatement de cette définition que si $(X_n)_n$ un processus aléatoire stationnaire, alors les variables aléatoires X_n ont la même loi.
3. Le plus simple processus stationnaire est le processus $(X_n)_n$ formé par une suite $(X_n)_n$ de variables aléatoires indépendantes et de même loi.

1.2 Définitions et propriétés:

Définition 3. *Dans un espace probabilisé* $(\Omega,\mathcal{A},P)$ *une transformation* T *est une transformation mesurable si,* $T : \Omega \longrightarrow \Omega$

$$T^{-1}A \; = \; \{w, \, Tw \in A\} \; \in \mathcal{A}, \quad \forall A \in \mathcal{A}$$

Définition 4. ***Transformation préservant une mesure***
Soit $(\Omega,\mathcal{A},P)$ *un espace probabilisé,* $T : \Omega \longrightarrow \Omega$ *,* T *est une transformation mesurable elle est dite transformation préservant la mesure* P *si:*

$$\forall A \in \mathcal{A}, \;\; P[T^{-1}A] = P[A].$$

Définition 5. ***Tribu des événement invariants:***
Soit T *la transformation qui préserve* P *et* A *un événement dans* $\mathcal{A}$*. On dit que* A *est un événement invariant si* $T^{-1}A = A$

Propsition 2. *Soit* $\mathfrak{I}$ *l'ensemble des événements invariants par rapport à* T*. Alors* $\mathfrak{I}$ *est dite tribu des événement invariants.*

Démonstration. Si $\mathfrak{I} = \varnothing$ il a rien à démontrer.
Si $\mathfrak{I} \neq \varnothing$ on a alors $\varnothing$ et $\Omega \in \mathfrak{I}$

1. De plus
$$\forall A \in \mathfrak{I}, (T^{-1}A)^c = A^c = T^{-1}A^c$$
$$T^{-1}A^c = A^c \Rightarrow A^c \in \mathfrak{I};$$

2. Si (A_n) est une suite d'événements invariants on a alors $T^{-1}A_n = A_n$ et
$$T^{-1}(\bigcup_n A_n) = \bigcup_n (T^{-1}A_n) = \bigcup_n (A_n)$$
$$\Rightarrow \bigcup_n A_n \in \mathfrak{I}.$$

$\mathfrak{I}$ est stable par passage de complémentaire et par réunion dénombrable. Donc c'est une tribu. □

Définition 6. *Soit* T *une transformation préservant la mesure* P*.*
T *sera dite transformation ergodique si*

$$\forall A \in \mathfrak{I} \; , \; P[A] = 0 \;\; ou \;\; 1.$$

Définition 7. ***(fonction invariante:)***
Soit T une transformation préservant la mesure P, f une fonction mesurable sur l'espace probabilisé $(\Omega,\mathcal{A},P)$, $f : \Omega \longrightarrow \mathbb{R}$, f est dite invariante si :

$$\forall w, f(w) = foT(w).$$

Propsition 3. *f une fonction invariante $\Leftrightarrow$ f est $\mathfrak{I}$-mesurable i.e $f \in \mathfrak{I}$*
Démonstration. Soit f une fonction invariante, montrons que $f \in \mathfrak{I}$ c'est-à-dire $\{f < a\} \in \mathfrak{I}$, $\forall a \in \mathbb{R}$ On a

$$\begin{aligned}
f^{-1}(]-\infty,a]) &= \{w, f(w) \leq a\} \\
&= \{w, foT(w) \leq a\} \\
&= (foT)^{-1}(]-\infty,a]) \\
&= T^{-1}of^{-1}(]-\infty,a]) \\
&= T^{-1}(f^{-1}(]-\infty,a]))
\end{aligned}$$

$\Rightarrow f^{-1}(]-\infty,a])$ est un événement invariant donc
$f^{-1}(]-\infty,a]) \in \mathfrak{I} \Rightarrow \{f < a\} \in \mathfrak{I}$, alors $f \in \mathfrak{I}$.
Soit maintenant $f \in \mathfrak{I}$, montrons que f est une fonction invariante. Si $A \in \mathfrak{I}$ en posant $f = \mathbb{I}_A$ on a $\mathbb{I}_A$ pour $A \in \mathfrak{I}$ est invariante, car

$$\begin{aligned}
f(Tw) &= \mathbf{1}_A(Tw) = \mathbb{I}_{T^{-1}A}(w) \\
&= \mathbb{I}_A(w) = f(w)
\end{aligned}$$

$T(w) \in A \Leftrightarrow w \in T^{-1}A$
Donc de toute combinaison linéaire fini d'indicateurs $\sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i}$ pour $A_i \in \mathfrak{I}$ est invariante. On a si $X \in \mathfrak{I}$ donc X est $\mathfrak{I}-$ mesurable$\Rightarrow$ $\exists$ une suite croissante de variables aléatoires simples dans $\mathfrak{I}$ tel que

$$X = \lim_n X_n$$

$$X(w) = \lim_n X_n(w)$$

$$X(Tw) = \lim_n X_n(Tw) = \lim_n X_n(w) = X(w)$$

Par convergence monotone. Ceci montre que f est bien une fonction invariante. $\square$

Propsition 4. *Soit T la transformation qui préserve la mesure P sur l'espace probabilisé $(\Omega,\mathcal{A},P)$, si T est ergodique alors les seules fonctions invariantes sont les fonctions constantes prèsque sûrement.*

Propsition 5. *Cf[15] Soit $(\Omega,\mathcal{A},P)$ un espace probabilisé, $T : \Omega \longrightarrow \Omega$ la transformation qui préserve la mesure P, alors les assertions sont équivalentes.*

(1) *T est ergodique*

(2) *f une fonction mesurable et $\forall x \in \Omega$, $(f \circ T)(x) = f(x) \;\; \Rightarrow f$ est une fonction constante présque partout.*

(3) *f une fonction mesurable et pour présque tout $x \in \Omega$,*
$(f \circ T)(x) = f(x) \;\; \Rightarrow f$ est une fonction constante présque partout.

(4) *$f \in \mathbb{L}^2(\Omega,\mathcal{A},P)$ et $\forall x \in \Omega$, $(f \circ T)(x) = f(x) \;\; \Rightarrow f$ est une fonction constante présque partout.*

(5) *$f \in \mathbb{L}^2(\Omega,\mathcal{A},P)$ et pour présque tout $x \in \Omega$, $(f \circ T)(x) = f(x) \;\; \Rightarrow f$ est une fonction constante présque partout.*

Propsition 6. *Soit T la transformation qui préserve la mesure P sur l'espace probabilisé $(\Omega,\mathcal{A},P)$, $f \in L^1_P(\Omega,\mathcal{A})$ mesurable alors:*

$$\mathbb{E}[f] = \mathbb{E}[foT]$$

Démonstration. En utilisant le théorème de transfert, soient$(X,\mathcal{A},\mu)$ et$(Y,\mathcal{B},\sigma)$ deux espaces mesurables. $f : X \longrightarrow Y$ mesurable et σ la mesure définie par $\sigma = (\mu f)$ qui est l'image de μ par f. $B \in \mathcal{B}$, $\mu(B) = \sigma(f^{-1}(B))$, μ bornée et σ aussi. $g \in L^1_\sigma(Y,\mathcal{B})$ on a

$$\int_Y g(y)\sigma(dy) = \int_X gof(x)\mu(dx)$$

avec $f^{-1}(Y) = X$ Si on prend $f = T : \Omega \longrightarrow \Omega$, alors $g \in L^1_P(\Omega,\mathcal{B})$ donc

$$\int_\Omega g(w)TP[dw] = \int_\Omega goT(w)P[dw]$$

$$\Rightarrow \int_\Omega g(w)P[T^{-1}dw] = \int_\Omega goT(w)P[dw]$$

Comme T est la transformation qui préserve P alors $P[T^{-1}dw] = P[dw]$

$$\Rightarrow \int_\Omega g(w)P[T^{-1}dw] = \int_\Omega g(w)P[dw]$$

$$\Rightarrow \int_\Omega g(w)P[dw] = \int_\Omega goT(w)P[dw]$$
$$\Rightarrow \mathbb{E}[g] = \mathbb{E}[goT]$$

□

1.3 Rappels sur les espaces de Hilbert et $\mathbb{L}^2_\mu(X,\mathcal{A})$

Définition 8. *Soit E un espace vectoriel réel de dimension finie ou infinie, E est un espace euclidien si E est muni d'un produit scalaire $< .|. >$; i.e une forme bilinéaire symetrique définie positive. Si E est un espace vectoriel complexe de dimension finie ou infinie E est un espace préhilbertien si E est muni d'un produit scalaire $< .|. >$ i.e d'une forme hérmitienne, définie positive.*
Si E est un espace vectoriel réel ou complexe de dimension finie ou infinie muni d'un produit scalaire $< .|. >$, soit $\| \, . \, \|$ la norme associée au produit scalaire.
Si E est complet, on dit que E est un espace de Hilbert.
Définition 9. $(E, \| \, . \, \|)$ *est un espace complet si*

$$\forall (x_n)_n \subset E \quad tel \ que \quad \| x_n - x_m \| \longrightarrow 0 \quad n,m \longrightarrow \infty$$

c'est-à-dire $(x_n)_n$ est une suite de Cauchy; $(x_n)_n$ est convergente i.e

$$\exists x \ \in E \ tel \ que \quad \| x_n - x \| \longrightarrow 0 \quad n \longrightarrow \infty$$

Définition 10. *Soit $(X,\mathcal{A},\mu)$ un espace mesuré. $\mathbb{L}^2_\mu(X,\mathcal{A})$ ou simplement $\mathbb{L}^2_\mu$ est l'espace fonctionnel défini par*

$$\mathbb{L}^2_\mu(X,\mathcal{A}) = \{f : X \to \mathbb{R}, ou\ f\ est\ une\ fonction\ mesurable\ et\ \int_X f^2 d\mu < \infty\}$$

Si pour $f,g \in \mathbb{L}^2_\mu(X,\mathcal{A})$, nous définissons la forme bilinéaire symétrique (f.b.s), $< f|g >$, en posant

$$< f|g > = \int_X f(x)g(x)d\mu \quad (1)$$

alors $< .|. >$ définit un produit scalaire sur $\mathbb{L}^2_\mu(X,\mathcal{A})$ car la f.b.s $< .|. >$ est définie positive.

Notons que la formule (1) a toujours un sens, compte tenu de l'inégalité de Holder (car ici $\mathbb{L}^2_\mu$ est en dualité avec lui-même).
La norme sur $\mathbb{L}^2_\mu$ est définie par la formule

$$\| f \| = \sqrt{< f|f >} = \sqrt{\int_X f^2(x) d\mu}$$

Par définition c'est la norme euclidienne de $\mathbb{L}^2_\mu$. La convergence d'une suite des fonctions au sens de cette norme est la convergence en moyenne quadratique.

Définition 11. *Soit $(f_n)_n$ une suite de fonctions dans $\mathbb{L}^2_\mu$. $(f_n)_n$ converge en moyenne quadratique vers f (i.e. au sens de $\mathbb{L}^2_\mu$), si*

$$\| f_n - f \| \longrightarrow 0$$

quand $n \to \infty$.

Définition 12. *Pour $\mu(X) < \infty$, $\mathbb{L}^2_\mu(X,\mathcal{A})$ est un espace complet. Donc l'espace $\mathbb{L}^2_\mu(X,\mathcal{A})$ est un espace de Hilbert.*

La convergence en moyenne quadratique et sa liaison avec d'autres types de convergence des suites de fonctions :

1-Si une suite $(f_n)_n$ de fonctions de l'espace $\mathbb{L}^2_\mu(X,\mathcal{A})$ est convergente pour la métrique de $\mathbb{L}^2_\mu(X,\mathcal{A})$, elle est convergente aussi pour la métrique de $\mathbb{L}^1_\mu(X,\mathcal{A})$.
2-Si une suite $(f_n)_n$ est uniformement convergente, elle est aussi convergente en moyenne quadratique.
3-Si une suite de fonctions mesurables $(f_n)_n$ est convergente en moyenne quadratique, elle est aussi convergente en mesure sur X
4-Si une suite $(f_n)_n$ est convergente en moyenne quadratique , en peut extraire une sous-suite $(f_{nk})_{nk}$, convergente présque partout sur X.

Définition 13. ***Orthogonalité:***
Soit E un espace euclidien, $< .,. >$ son produit scalaire et $||.||$ la norme associée. x et $y \in E$ sont orthogonaux si leur produit scalaire est nul.

Si x et $y \in E$ sont orthogonaux, nous avons le théorème de Pythagore, i.e.

$$< x|y >= 0 \ \Leftrightarrow \| x + y \|^2 = \| x \|^2 + \| y \|^2$$

Un système de vecteurs $\{\varphi_i\}_{i\in X}$ *est orthonormé si on a* $<\varphi_i|\varphi_j>=\ \delta_{ij}$ *Où* δ_{ij} *est le symbole de Kronecker.*
Si $(\varphi_n)_n$ *est un système orthogonal alors* $\frac{\varphi_n}{\|\varphi_n\|}$ *est un système orthonormé.*
Théorème 2. *Soient* f_1, f_2,..., f_n,... (1) *un système liénairement indépendant d'éléments d'un espace euclidien* E. *Alors dans* E *il existe un système d'éléments* φ_1,φ_2,...φ_n,... (2) *satisfaisant les conditions suivantes:*

1. *Le système* (2) *est orthonormé;*
2. *Chaque élément* φ_n *est une combinaison linéaire des éléments* f_1, f_2,...,f_n, ...

$$\varphi_n = \ a_{n1}f_1 + \ a_{n2}f_2 + ... + a_{nn}f_n \quad avec \quad a_{nn}\neq 0$$

3. *Chaque élément* f_n *peut-être mis sous la forme*

$$f_n = \ b_{n1}\varphi_1 + \ b_{n2}\varphi_2 + ... + b_{nn}\varphi_n \quad avec \quad b_{nn}\neq 0$$

Le passage du système (1) au système (2) vérifiant les conditions 1 et 3 s'appelle procédé d'orthogonalisation.
Inégalité de Bessel:systèmes orthogonaux fermés:

Définition 14. *Soit* E *un espace euclidien,* $<.|.>$ *son produit scalaire et* $\|\ .\ \|$ *la norme associée, Soit* $(\varphi_n)_n$ *un système orthonormé de l'espace* E, *et*

$$\forall f\in E,\ c_k =< f|\varphi_k > \qquad k=1,2...$$

que nous appelons coéfficients de Fourier de l'élément f *par rapport au système* φ_{nn} ,

$$\sum_k c_k\varphi_k$$

est la serie de Fourier de l'élément f *par rapport au système* $(\varphi_n)_n$.

La question qui se pose: la serie de Fourier est-elle convergente? c'est-à-dire la suite des sommes partielles de cette serie converge-t-elle (au sens de la métrique de E)et si elle est convergente, alors sa somme coîncide-t-elle avec l'élément initial f?

Propsition 7. *Dans un espace euclidien* E *on a pour toute fonction* f *à valeurs dans E, la serie* $\sum_{k=1}^\infty c_k^2$ *est convergente et*

$$\sum_{k=1}^{\infty} c_k^2 \leq\| f \|^2$$

Où les c_k sont les coéfficients de Fourier, cette inégalité s'appelle inégalité de Bessel.

Démonstration. Pour cela pour n donné choisir les coéfficients α_k $(k = 1,2,...,n)$ de façon que la distance entre f et la somme $S_n = \sum_{k=1}^{n} \alpha_k \varphi_k$ soit minimale.

$$\begin{aligned}
\| f - S_n \|^2 &= < f - \sum_{k=1}^{n} \alpha_k \varphi_k | f - \sum_{k=1}^{n} \alpha_k \varphi_k > \\
&= < f|f > -2 < f| \sum_{k=1}^{n} \alpha_k \varphi_k > + < \sum_{k=1}^{n} \alpha_k \varphi_k | \sum_{i=1}^{n} \alpha_i \varphi_i > \\
&= \| f \|^2 -2 \sum_{k=1}^{n} \alpha_k c_k + \sum_{k=1}^{n} \alpha_k \\
&= \| f \|^2 + \sum_{k=1}^{n} (\alpha_k - c_k)^2 - \sum_{k=1}^{n} c_k^2.
\end{aligned}$$

Il est clair que pour $\| f - S_n \|^2$ soit minimale il faut que

$$\sum_{k=1}^{n} (\alpha_k - c_k)^2 = 0$$

c'est-à-dire $\alpha_k = c_k \quad \forall k = 1,2...,n$ dans ce cas on a :

$$\| f - S_n \|^2 = \| f \|^2 - \sum_{k=1}^{n} c_k^2$$

$$or \quad \| f - S_n \|^2 \geq 0 \quad \Rightarrow \| f \|^2 \geq \sum_{k=1}^{n} c_k^2.$$

ici n est arbitraire et le second membre ne dépend pas de n; par conséquant, la serie $\sum_{k=1}^{\infty} c_k^2$ est convergente et on a

$$\sum_{k=1}^{\infty} c_k^2 \leq \| f \|^2$$

□

Introduisons la notion importante suivante:

Définition 15. *le système orthonormé $\varphi_1,\varphi_2,...\varphi_n,...$ est dit fermé, si pour tout $f \in E$ on a l'égalité suivante:*

$$\sum_{k=1}^{\infty} c_k^2 = \| f \|^2$$

appelé relation de Parseval.

Remarque 1. *Le systhème orthonormé $\varphi_1,\varphi_2,...\varphi_n,...$ est fermé si et seulement si pour toute fonction $f \in E$ la serie de Fourier $\sum_{k=1}^{\infty} c_k\varphi_k$ par rapport à f converge vers f i.e*

$$\forall f \in E, \quad f \longrightarrow \sum_{k=1}^{\infty} c_k\varphi_k$$

Démonstration. Soit E un espace euclidien, $f \in E$, soit $S_n = \sum_{k=1}^{n} c_k\varphi_k$ la serie de Fourier par rapport à f on a $f \longrightarrow \sum_{k=1}^{n} c_k\varphi_k$ si $\| f - S_n \|^2 \longrightarrow 0$. Or

$$\| f - S_n \|^2 \geq 0 \quad \Rightarrow \| f \|^2 \geq \sum_{k=1}^{n} c_k^2$$

Donc

$$\| f - S_n \|^2 \rightarrow 0 \quad \Rightarrow \| f \|^2 \backsimeq \sum_{k=1}^{n} c_k^2$$

□

La notion de système orthonormé fermé est étroitement liée à la notion de système complet.

Théorème 3. ***Théorème de Riesz-Fischer***

Soit $(\varphi_n)_n$ un système orthonormé quelconque dans un espace euclidien complet E soient alors $c_1,c_2,...,c_n,...$ des nombres, tels que la serie

$$\sum_{k=1}^{\infty} c_k^2 < \infty$$

Alors il existe un élément $f \in E$ qui vérifie

$$c_k =< f|\varphi_k > \quad et \quad \sum_{k=1}^{\infty} c_k^2 =< f|f >=\| f \|^2 .$$

Series de fourier par rapport à un système orthogonal:
On a $\mathbb{L}^2_\mu$ un espace euclidien complet de dimension infinie , c'est-à-dire un espace de Hibert donc il va exister des systèmes orthogonaux en particulier des systèmes orthonormées de fonctions d'aprés le théorème 3 et le théorème de Riesz-Fischer.
Alors la serie de Fourier par rapport au système orthogonal $(\varphi_n)_n$ converge vers f μ p.p i.e

$$f(x) = \sum_{n=1}^{+\infty} c_n \varphi_n \quad \mu \ p.p$$

c_n: coéfficients de Fourier de la fonction f par rapport au système $(\varphi_n)_n$, sont définis par :

$$c_n = \frac{1}{\| \varphi_n \|^2} \int_X f(x)\varphi_n(x) d\mu$$

Ci-dessous, nous citons les exemples les plus usuels de systèmes orthogonaux dans $\mathbb{L}^2_\mu$.
1.système trigonométrique

1.Considerons l'espace $\mathbb{L}^2_\mu[-\pi,\pi]$, la mesure sur ce segment étant la mesure de Lebesgue alors,

$$1, \ \cos nx, \ \sin nx \quad , n \geq 1$$

forment un système orthogonal complet, dit système trigonomètrique.
Donc les coéfficients de Fourier par rapport à la base orthogonale sontréspectivement a_0, a_n *et* b_n

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)dx$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx$$

la série de Fourier correspondante est:

$$\frac{a_0}{2} + \sum_{n=1}^{+\infty} (a_n \cos nx + b_n \sin nx)$$

Généralement si $f \in \mathbb{L}^2([-l,l])$, on fait un changement de variable $x = \frac{n\pi}{l}$, $n = \frac{lx}{\pi}$ conduit à une fonction $f^*(\frac{lx}{\pi})$, définie sur $[-\pi,\pi]$
2.Soit $[-\pi,0]$ et $[0,\pi]$ deux segments , si la mesure sur ces deux segments est la mesure Lebesgue alors,

$$1,\ \cos x, \cos 2x,..., \qquad et \quad \sin x,\ \sin 2x,...,$$

forment dans leur enssemble un système orthogonal sur $[-\pi,\pi]$.
- Sur $[-\pi,0]$ f est une fonction paire, donc

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx dx = \frac{2}{\pi}\int_0^{\pi} f(x)\cos nx dx$$

Car $x \mapsto f(x)\cos nx$ est paire.
$b_n = 0$ car $x \mapsto f(x)\sin nx$ est impaire d' où la serie de Fourier est

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx.$$

- Sur $[0,\pi]$ f est une fonction impaire, donc

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx dx = \frac{2}{\pi}\int_0^{\pi} f(x) sinnx dx$$

Car $x \mapsto f(x)\sin nx$ est paire.
$a_n = 0$ car $x \mapsto f(x)\cos nx$ est impaire d' où la serie de Fourier est

$$f(x) = \sum_{n=1}^{\infty} b_n \sin nx.$$

En général $f \in [0,l)$ *ou* $[-l,0)$ on fait un changement de variable $x = \frac{2\pi t}{l} \Rightarrow t = \frac{xl}{2\pi}$ conduit à remplacer $f(t)$ par $f(\frac{xl}{2\pi})$ définit sur $[-\pi,\pi]$.
2.Serie de Fourier sous forme complexe
L'ecriture d'une serie trigonométrique peut-être condensée, si l'on fait l'usage des formules d'Euler

$$\cos nx = \frac{e^{inx} + e^{-inx}}{2} \ , \quad \sin nx = \frac{e^{inx} - e^{-inx}}{2i}$$

En remplaçons dans la serie de Fourier on obtient

$$\frac{a_0}{2} + \sum_{n=0}^{+\infty}(a_n \cos nx + b_n \sin nx) = \sum_{-\infty}^{\infty} c_n e^{inx}$$

avec

$$c_0 = \frac{a_0}{2} \ , \ \ c_n = \frac{a_n - ib_n}{2} \ , \ \ c_{-n} = \frac{a_n + ib_n}{2}$$

Donc

$$f(x) = \sum_{-\infty}^{\infty} c_n e^{inx}$$

s'appelle serie trigonométrique de Fourier sous la forme complexe.
Il est plus facile de calculer les coéfficients de Fourier c_n. En effet, un calcul immédiat montre que

$$\int_{-\pi}^{\pi} e^{inx} e^{-imx} dx = \begin{cases} 0 & \text{si } \ n \neq m; \\ 2\pi & \text{si } \ n = m. \end{cases}$$

Donc

$$\int_{-\pi}^{\pi} f(x) e^{-imx} dx = 2\pi c_m$$

$$i.e \quad c_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-imx} dx \ \ , m \in \mathbb{Z}$$

Remarque 2. *Le développement de f est valable également pour les fonctions dans $\mathbb{L}^2_\mu[-\pi,\pi]$. Autrement dit , les fonctions e^{inx} forment une base de $\mathbb{L}^2_\mu[-\pi,\pi]$ et les c_m représentent les produits scalaires de f et e^{imx}.*
En remplaçant les fonctions e^{imx} par $e^{\frac{in\pi x}{l}}$, on peut étendre tout ce qui a été dit à l'espace $\mathbb{L}^2_\mu[-l,l]$
En remplaçant les fonctions e^{imx} par $e^{\frac{2in\pi x}{l}}$, on peut étendre tout ce qui a été dit à l'espace $\mathbb{L}^2_\mu[-l,0]$ ou $\mathbb{L}^2_\mu[0,l]$

Système orthogonaux de fonctions dans $\mathbb{L}^2_\mu$:
1.Polynômes de Legendre:
Soit 1, x, x^2, ... (1) des fonctions dans $\mathbb{L}^2_\mu$, le système formé par les combinaisons linéaires de (1) est complet dans $\mathbb{L}^2_\mu$.
En orthogonalisant les systèmes sur le segment $[-1,1]$ relativement au produit scalaire $< f|g >= \int_{-1}^{1} f(x)g(x)dx$, nous obtiendrons un système orthogonal complet $Q_0(x)$, $Q_1(x)$, ..., Q_n de polynômes de degrés n avec $Q_n(x)$ coïncide, a un facteur prés, avec le polynôme $R_n(x) = \frac{d^n}{dx^n}(x^2-1)^2$. Les polynomes définis par

$$P_n(x) = \frac{1}{n!2n} \ \frac{d^n}{dx^n}(x^2-1)^2$$

sont appelé polynômes de Legendre.

2.Polynômes orthogonaux par rapport à un point donné:

Soit maintenant μ une autre mesure sur $[-1,1]$ tel que les combinaisons de (1) avec le produit scalaire $\int_{-1}^{1} f(x)g(x)d\mu$ soient linéairement independants, en orthogonalisant (1) nous obtiendrons un nouveau système de polynômes Q_n qui dépend du choix de la mesure μ. μ est mesurable au sens de Lebesgue $\mu(X) = \int_X h(x)dx$, $h(x)$ est appelée fonction poids. La condition d'orthonormalité devient

$$< Q_m|Q_n >= \int_{-1}^{1} Q_m(x)Q_n(x)h(x)dx = \begin{cases} 1 & \text{si} \quad n = m; \\ 0 & \text{si} \quad n \neq m. \end{cases}$$

Dans ce cas on dit que Q_m et Q_n $\forall n$, m, sont orthogonaux par rapport au poids h. Le choix du poids conduit à un nouveau système de polynômes orthogonaux, pour $h(x) = \frac{1}{\sqrt{1-x^2}}$ on obtient des polynômes qui coïncide, à un facteur prés, avec les polynômes de Tchebychev

$$T(x) = \cos n \arccos x \qquad (n \geq 1)$$

Base orthogonale dans l'espace $\mathbb{L}^2_\mu(-\infty,\infty)$

Fonction d'Hermite

On doit chercher parmis des fonctions qui décroissent suffisament vite à l'infini pour constuire une base. En particulier les fonctions de la forme $x^n e^{\frac{-x^2}{2}}$. En leurs appliquant le procédé d'orthogonalisation on obtient

$$\varphi_n(x) = H_n(x)e^{\frac{-x^2}{2}}, \quad n \in \mathbb{N}$$

où H_n est un polynôme de degrés n ces polynômes s'appellent polynômes d'Hermite et les fonctions φ_n s'appellent fonctions d'Hermite.

les polynômes d'Hermite coîncident, à un facteur prés, avec les polynômes $H_n^*(x) = (-1)^n e^{x^2} \frac{d^n e^{x^2}}{dx^n}$ la condition d'orthogonalité est que

$$\int_{-\infty}^{\infty} H_n^*(x)H_m^*(x)e^{-x^2}dx = 0 \qquad n = m$$

Le résultat obtenu admet encore l'intérprétation suivante: considérons μ de densité e^{x^2}, cette mesure est finie d'où le produit scalaire

$$< f|g >= \int_{-\infty}^{\infty} f(x)g(x)e^{-x^2}dx$$

alors les polynômes d'Hermite forment un système orthogonal dans $\mathbb{L}^2_\mu(-\infty,\infty)$

Base orthogonale dans l'espace $\mathbb{L}^2_\mu(0,\infty)$
Prenons le système de fonction $x^n e^{-x}$, $n \in \mathbb{N}$. En appliquant le procédé d'orthogonalisation on obtient le système de fonctions $L_n(x)e^{-x}$ que l'on appellent fonctions de Laguerre, les polynômes correspondant L_n s'appellent polynômes de Laguerre qui forment une base orthogonal sur $\mathbb{L}^2_\mu(0,\infty)$ par rapport à la mesure $d\mu = e^{-x}dx$.
Pour plus de précisions, Cf. [6].

1.4 Exemple de transformation ergodique:

(Cf.[3][15]) Soit $(\Omega,\mathcal{A},P)$ un espace probabilisé. On pose $\Omega = [0,\ 1[$ et $\mathcal{A} = \mathcal{B}_{[0,\ 1[}$ et P la mesure de Lebesgue. Soit

$$T : [0,\ 1[\longrightarrow [0,\ 1[$$

$$x \longmapsto (\lambda + x)[1]$$

Alors T est une transformation ergodique pour λ irrationnel.
Démonstration. Montrons d'abord que T est mesurable.
T mesurable $\Leftrightarrow \forall a \in [0,\ 1],\quad \{0 \leq T(x) \leq a\} \in \mathcal{B}_{[0,\ 1[}$.
Précisons donc pour $a \in [0,\ 1],\ \{0 \leq T(x) \leq a\} = T^{-1}([0,\ a[)$

$$\begin{aligned}\{0 \leq T(x) \leq a\} &= [\{0 \leq x+\lambda \leq a\} \cap \{0 \leq x+\lambda < 1\} \cap \{0 \leq x \leq 1\}] \\ &\cup\ [\{0 \leq x \leq 1\} \cap \{0 \leq x+\lambda-1 \leq a\} \cap \{1 \leq x+\lambda < 2\}].\end{aligned}$$

$x+\lambda < 2$ car $x \in [0,\ 1[\ \ et\ \ \lambda \in [0,\ 1[$
$A_1 = \{0 \leq x+\lambda \leq a\} \cap \{0 \leq x+\lambda < 1\} \cap \{0 \leq x \leq 1\}$
Si $a-\lambda < 0$ i.e $a < \lambda$, alors $A_1 = \varnothing$
Si $a-\lambda \geq 0$ i.e $a \geq \lambda$, alors $A_1 = [0,\ a-\lambda] \cap [0,\ 1-\lambda] \cap [0,\ 1[$
$A_2 = \{0 \leq x \leq 1\} \cap \{0 \leq x+\lambda-1 \leq a\} \cap \{1 \leq x+\lambda < 2\}$
Si $a-\lambda < 0$ i.e $a < \lambda$, alors $A_2 = [1-\lambda,\ 1-\lambda+a]$
Si $a-\lambda \geq 0$ i.e $a \geq \lambda$, alors $A_2 = [1-\lambda,\ 1]$
Donc si $a < \lambda$, $\{0 \leq T(x) \leq a\} = \varnothing \cup [1-\lambda,\ 1-\lambda+a] = [1-\lambda,\ 1-\lambda+a] \in \mathcal{B}_{[0,\ 1[}$
Si $a \geq \lambda$, $\{0 \leq T(x) \leq a\} = [0,\ a-\lambda] \cup [1-\lambda,\ 1] \in \mathcal{B}_{[0,\ 1[}$.
Donc T est mesurable.

Montrons maintenant que T préserve la mesure de Lebesgue i.e
$P[T^{-1}A] = P[A]$ pour tout événement A soit $[0, a[$ donc $P[[0, a[[= P[A] = a$
1.Pour $a < \lambda$ on a $T^{-1}A = [1-\lambda, 1-\lambda+a]$ donc $P[T^{-1}A] = a$
2.Pour $a \geq \lambda$ on a $T^{-1}A = [0, a-\lambda] \cup [1-\lambda, 1]$ donc $P[T^{-1}A] = a$
Dans les deux cas on a $P[T^{-1}A] = P[A]$ d'où T préserve la mesure de Lebesgue.

Montrons maintenant que T est une transformation ergodique.
Soit f une fonction mesurable, supposons que $f \in \mathbb{L}^2_\mu[0, 1[$ donc $\int f^2 dx < \infty$ et $\sum \mid c_n \mid^2 < \infty$ et d'aprés les notations précédentes on a f admet un développement en serie de Fourier tel que

$$f(x) = \sum_{-\infty}^{\infty} c_n e^{2\pi inx} \qquad P\ p.p$$

De plus on a

$$f(Tx) = \sum_{-\infty}^{\infty} c_n e^{2\pi in\lambda} e^{2\pi inx} \qquad P\ p.p$$

Pour f une fonction invariante on a $f(x) = f(Tx)$ donc,

$$\sum_{-\infty}^{\infty} c_n e^{2\pi inx} = \sum_{-\infty}^{\infty} c_n e^{2\pi in\lambda} e^{2\pi inx}$$

$\Rightarrow c_n(1 - e^{2\pi in\lambda}) = 0$ donc soit $e^{2\pi in\lambda} = 1$ ou bien $c_n = 0$
Si λ est irrationnel, $e^{2\pi in\lambda} \neq 1$ pour $n \neq 0$. Donc $c_n = 0$, $\forall n \neq 0$ et de là on a $f(x) = c_0$, $Pp.p$ i.e f est une constante pour λ irrationnel d'où T est ergodique pour λ irrationnel d'aprés la proposition 5. $\square$

1.5 Théorèmes ergodiques

Les premiers théorèmes fondamentaux en théorie ergodique sont les théorèmes ergodiques de Birkhoff(1931)(Cf[3], Cf[5]) et Von Neumann(1932)(Cf[15]). Les deux théorèmes concernent la convergence de la suite $\frac{1}{n}\sum_{k=1}^{n} f_k$.

Théorème 4. ***Théorème ergodique de Von Neumann***
Soit T une transformation qui préserve la mesure P sur $(\Omega,\mathcal{A},P)$,$1 \leq p < \infty$.

Si $f \in \mathbb{L}^p_P(\Omega,\mathcal{A})$*, donc il va exister une fonction* $f^* \in \mathbb{L}^p_P(\Omega,\mathcal{A})$ *tel que* $f^* o T = f^*$ *p.s Si*

$$f_1(w) = f(T^0 w) = f(w)$$

$$f_k(w) = f(T^{k-1} w)$$

Alors on a

$$\parallel \frac{1}{n}\sum_{k=1}^{n} f_k(x) - f^*(x) \parallel_p \longrightarrow 0.$$

Avec $f^* = \mathbb{E}[f/\mathcal{I}]$

Lemme ergodique maximal:(Cf[3],[15])
Soit T la transformation qui préserve P sur $(\Omega,\mathcal{A},P)$, $f \in L^1_P(\Omega,\mathcal{A})$, Soit

$$f_k(w) = f(T^{k-1}w), \quad k \geq 1$$

$$S_k(w) = f_1(w) + f_2(w) + ... + f_k(w), \quad f_1(w) = f(w)$$

$$M_n(w) = Max(0,S_1(w),S_2(w),...,S_n(w))$$

Alors on a,

$$\int_{\{M_n>0\}} f_k dp \geq 0$$

Démonstration.

On a $f \in L^1_P(\Omega,\mathcal{A})$ et $f_k(w) = f(T^{k-1}w)$, pour $k \geq 1$

$$S_k(w) = f_1(w) + ... + f_k(w) \;\; et \;\; M_k(w) = Max\{0,S_1(w),...,S_k(w)\}.$$

$$\begin{aligned} Pour \;\; k \leq n, \;\; S_k(Tw) &= f_1(Tw) + ... + f_k(Tw) \\ &= f_2(w) + ... + f_{k+1}(w) \end{aligned}$$

Donc $M_n(Tw) \geq S_k(Tw)$

$$\begin{aligned} &\Rightarrow \;\; f(w) + M_n(Tw) \geq f(w) + S_k(Tw) \\ &\Rightarrow \;\; f(w) + M_n(Tw) \geq S_{k+1}(w) \\ &\Rightarrow \;\; f(w) \geq S_{k+1}(w) - M_n(Tw), \quad \forall k \in \{1...n\} \end{aligned}$$

$$\begin{array}{rcl} Donc \quad f(w) & \geq & Max\{S_1(w),S_2(w),...,S_k(w)\} - M_n(Tw) \\ & \geq & M_n(w) - M_n(Tw) \end{array}$$

Car sur $\{M_n(w) > 0\}$

$$\begin{array}{rcl} Max\{0,S_1(w),S_2(w),...,S_k(w)\} & = & Max\{S_1(w),S_2(w),...,S_k(w)\} \\ & = & M_n(w) \end{array}$$

$$\Rightarrow \int_{M_n>0} f(w)dP \geq \int_{M_n>0} (M_n(w) - M_n(Tw))dP$$

$$\Rightarrow \int_{\Omega} f(w)\mathbf{1}_{\{M_n>0\}}dP \geq \int_{\Omega} (M_n(w) - M_n(Tw))\mathbf{1}_{\{M_n>0\}}dP$$

On a $\mathbb{E}[M_n] = \mathbb{E}[M_n oT]$

$$\Rightarrow \mathbb{E}[M_n(w) - M_n(Tw)] = 0$$

$$\Rightarrow \int_{\Omega} (M_n(w) - M_n(Tw))dP = 0$$

$$\Rightarrow \int_{\{M_n>0\}} (M_n(w) - M_n(Tw))dP + \int_{\{M_n\leq 0\}} (M_n(w) - M_n(Tw))dP = 0$$

$$\Rightarrow \int_{\{M_n>0\}} (M_n(w)dP - \int_{\{M_n=0\}} M_n(Tw))dP = 0$$

Car on a $M_n \geq 0$, $\forall n \in \mathbb{N}$, donc $\int_{\{M_n<0\}}(M_n(w) - M_n(Tw))dP = 0$ et $\int_{\{M_n=0\}} M_n(w)dP = 0$

$$\Rightarrow \int_{\{M_n>0\}} (M_n(w) - M_n(Tw))dP = \int_{\{M_n=0\}} M_n(Tw)dP$$

$$\Rightarrow \int_{\{M_n>0\}} f(w) \geq \int_{\{M_n>0\}} (M_n(w) - M_n(Tw))dP \geq 0$$

$$\Rightarrow \int_{\{M_n>0\}} f(w)dP \geq 0$$

Car $M_n(Tw) \geq 0 \Rightarrow \int_{\{M_n=0\}} M_n(Tw)dP \geq 0$ $\square$

Théorème 5. ***Théorème ergodique de Birkhoff***
Soit T une transformation qui préserve la mesure P sur $(\Omega,\mathcal{A},P)$, $f \in L^1_P(\Omega,\mathcal{A})$. Si

$$f_1(w) = f(T^0 w) = f(w)$$
$$f_k(w) = f(T^{k-1} w)$$

Alors on a pour $n \to \infty$

$$\frac{1}{n}\sum_{k=1}^{n} f_k \longrightarrow \mathbb{E}[f/\mathfrak{I}], \quad P-ps.$$

Où $\mathfrak{I}$: la tribu des événements invariants.

Remarque 3. *Si de plus T est ergodique alors on a*

$$\lim_n \frac{1}{n}\sum_{k=1}^{n} f_k(x) = \mathbb{E}[f] \quad p.s$$

Démonstration. $\mathbb{E}[f/\mathfrak{I}]$ est une variable $\mathfrak{I}$-mesurable par définition de l'espérence conditionnelle. Comme T est ergodique, d'aprés la proposition 4, $\mathbb{E}[f/\mathfrak{I}]$ est p.s constante, donc

$$\mathbb{E}[f/\mathfrak{I}] = \mathbb{E}[f] \quad p.s$$

□

Remarque 4. *Si $T: \quad \Omega \longrightarrow \Omega$ est une transformation préservant la mesure et ergodique, X la variable aléatoire définie par $X = \mathbb{I}_A$ avec $A \in \mathcal{A}$, $X_k(w) = \mathbb{I}_A(T^{k-1}(w))$ et donc*

$$\frac{1}{n}\sum_{k=1}^{n} X_k(w) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{I}_A(T^{k-1}(w))$$

est la fréquence des poids $T^{k-1}(w)$ qui tombent dans A, les premiers n instants. Le théorème ergodique de Birkhoff pour T ergodique assure que cette fréquence

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{I}_A(T^{k-1}(w)) \longrightarrow P[A] \quad quand \;\; n \longrightarrow \infty$$

Propsition 8. *Soit T la transformation qui préserve P sur l'espace probabilisé $(\Omega,\mathcal{A},P)$, soit X une variables aléatoires sur $(\Omega,\mathcal{A})$. Si $(X_n)_n$ le processus aléatoire défini par*

$$X_n(w) = X(T^{n-1}(w)), \qquad avec \ \ T^0 = \mathbb{I}_d$$

Alors $(X_n)_n$ est un processus stationnaire.

Définition 16. *Dans $(\mathbb{R},\mathcal{B}_{\mathbb{R}^\infty})$, on appelle schift ou opérateur de translation, l'application S définie pour $x = (x_1,\ x_2,...) \in \mathbb{R}^\infty$ par*

$$x \longmapsto S(x_1,\ x_2,...) = (x_2,\ x_3,...)$$

Propsition 9. *Si $(X_n)_n$ est un processus stationnaire, l'opérateur de translation S, prèserve la mesure de probabilité $\hat{P}$, c'est-à-dire la loi du processus $(X_n)_n$*

Définition 17. *Soit $(\Omega,\mathcal{A},P)$ un espace probabilisé, soit $(X_n)_n$ un processus. Le processus $(X_n)_n$ est ergodique si la mesure P est invariante.*

Théorème 6. *Soit $X = (X_n)_n$ un processus stationnaire. Si $(X_n)_n$ est ergodique, alors on a*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow \mathbb{E}[X_1] \quad p.s \ \ et \ \ dans \ \ \mathbb{L}^1$$

Démonstration. La démonstration est assez longue car elle nécessite de passez à l'espace de représentation du processus $(X_n)_n$ i.e définir $(\hat{X}_n)_n$. Une bonne démonstration est dans [3] □

Conséquence:

Si $(X_n)_n$ est une suite de variables aléatoires independantes de même loi dans $\mathbb{L}^1$ alors

$$\frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow \mathbb{E}[X]$$

qui est la loi des grands nombres.

1.6 Une application à la théorie des nombres: (Théorème de Borel):

Propsition 10. *Pour tout réel x de $]0,1[$, la fréquence d'apparition du chiffre 1 dans l'écriture binaire de x est $\frac{1}{2}$.*

Démonstration. La démonstration est faite dans Walters.P Soit $T[0,1) \longrightarrow [0,1)$ définie par: $T(x) = 2x[1]$ une transformation qui préserve la mesure de Lebesgue et aussi ergodique. Supposons que x admette comme écriture binaire unique $\frac{a_1}{2} + \frac{a_2}{2^2} + ...$ Alors

$$T(x) = T(\frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + ...) = \frac{a_2}{2} + \frac{a_3}{2^2} + ...$$

Posant $f(x) = \mathbb{I}_{[\frac{1}{2},1[}$. Alors

$$f(T^i(x)) = f(\frac{a_{i+1}}{2} + \frac{a_{i+2}}{2^2} + ...) = \begin{cases} 1 & \text{si } a_{i+1} = 1 \\ 0 & \text{si } a_{i+1} = 0 \end{cases}$$

Ainsi le nombre de 1 parmi les n premier chiffres de l'écriture diadique de x est

$$\sum_{n=0}^{n-1} f(T^i(x)).$$

En appliquant le théorème ergodique de Birkhoff, on obtient

$$\lim \frac{1}{n} \sum_{n=0}^{n-1} f(T^i(x)) = \int \mathbb{I}_{[\frac{1}{2},1[} dp = \frac{1}{2} \quad p.s$$

□

Chapitre 2

Chaîne de Markov et noyaux de transitions

2.1 Définitions et propriétés

Nous faisons une synthèse sur les noyaux de transition et les chaînes de Markov à partir de (Cf[13])

Définition 1. *Soit* $(E,\mathcal{E})$ *un espace mesurable. Un noyau de transition sur* E *est une application*

$$N : E \times \mathcal{E} \rightarrow [0,1]$$

tel que :

(1) $\forall x \in E$, *l'application* $A \mapsto N(x,A)$*: de* $\mathcal{E}$ *dans* $[0,1]$ *est une mesure de probabilité sur* $(E,\mathcal{E})$.

(2) $\forall A \in \mathcal{E}$, *l'application* $x \mapsto N(x,A)$*: de* E *dans* $[0,1]$ *est une fonction mesurable.*

$N(x,A)$ *représente la probabilité de transition de* x *vers* A

Propsition 1. *Si* $f : E \longrightarrow \mathbb{R}$ *est une fontion positive bornée et mesurable, soit* N *le noyau de transition, alors* Nf *de* E *dans* $\mathbb{R}$ *définie par la formule :*

$$Nf(x) = \int_E N(x,dy)f(y) =< N(x,.), f > , \forall x \in E$$

Alors Nf *est une fonction positive, bornée et mesurable .*

Propsition 2. *Si* ν *est une mesure sur* $(E,\mathcal{E})$ *et si on définit* νN *sur* $(E,\mathcal{E})$ *par :*

$$\forall A \in \mathcal{E}, \nu N(A) = \int_E \nu(dx)N(x,A) =< \nu|N(.,A) >$$

avec N le noyau de transition alors νN est aussi une mesure sur $(E, \mathcal{E})$.
Démonstration. 1- $A \longrightarrow \nu N(A)$ est une fonction mesurable sur $\mathcal{E}$.
2- νN est σ-additive. En effet

$$\begin{aligned} \nu N(\cup_i A_i) &= \int_E \nu(dx) N(x,\ \cup_i A_i) \\ &= \sum_i \int_E \nu(dx) N(x,\ A_i) \\ &= \sum_i \nu N(x,\ A_i) \end{aligned}$$

$\forall (A_i)_i \in \mathcal{E}$ disjoints.
3- $\nu N \neq \infty$ car $\nu N(\varnothing) = \int_E \nu(dx) N(x,\ \varnothing) = 0$ □

Propsition 3. *La composition ou le produit de deux noyaux N et M sur $(E, \mathcal{E})$, est défini par :*

$$MN(x, A) = \int_{y \in E} M(x, dy) N(y, A) \quad \forall x \in E, \forall A \in \mathcal{E}$$

est aussi un noyau de transition sur $(E, \mathcal{E})$.

Démonstration. On montre que la composition de deux noyaux de transition N et M est un noyau de transition, ce qui revient à montrer que:
(1) $\forall x \in E$, l'application $A \mapsto MN(x,A)$: de $\mathcal{E}$ dans [0,1] est une mesure de probabilité sur $(E,\mathcal{E})$.
(2) $\forall A \in \mathcal{E}$, l'application $x \mapsto MN(x,A)$:de E dans [0,1] est une fonction mesurable.
Si on pose $f(x) = N(x,A)$ pour A fixé, alors d'aprés la proposition précédente
$\forall A \in \mathcal{E}$, l'application $x \longmapsto MN(x,A)$: de E dans [0, 1] est une fonction mesurable.
On a aussi $\forall x \in E$, lapplication $A \longmapsto M(x,A)$: de $\mathcal{E}$ dans [0, 1] est une mesure de probabilité sur $(E,\mathcal{E})$.
Si on pose $\nu = M(x,A)$ pour x fixé, alors d'aprés la proposition précédente, $\forall x \in E$, l'application $A \longmapsto MN(x,A)$: de $\mathcal{E}$ dans [0, 1] est une mesure de probabilité sur $(E,\mathcal{E})$.
Ce qui résulte que MN est un noyau de transition.

□

Remarque 1. *La composition de deux noyaux de transition est une opération associative.*

Définition 2. *Soit N et M deux noyaux, on dit que M est plus petit que N et on écrit $M \leq N$ si, pour toute fonction f bornée, positive et mesurable, on a*

$$Mf \leq Nf.$$

Définition 3. *Soit N un noyau. Si $N(x,A) < 1$ pour tout $x \in E$ alors N est appelé probabilité de transition ou un noyau submarkovien. Il est appelé noyau markovien si $N(x,A) = 1$ pour tout $x \in E$*

On notera N^n le noyau de transition obtenu en éffectuant $n-1$ produits de N avec lui même.

Propsition 4. *Si N est un noyau de transition sur l'espace mesurable $(E, \mathcal{E})$, alors pour tout $n \geq 1$,*

$$N^{n+1}(x, A) = NN^n(x, A) \quad \forall x \in E, \forall A \in \mathcal{E}$$

Et comme N et N^n sont deux noyaux de transitions et d'aprés la proposition précédente on obtient

$$N^{n+1}(x, A) = \int_{y \in E} N(x, dy) N^n(y, A).$$

On a aussi $N^0(y, A)$ est la mesure de Dirac au point y

Démonstration. (Par réccurence)
Pour $n \geq 1$, on montre que

$$N^{n+1}(x, A) = \int_{y \in E} N(x, dy) N^n(y, A), \quad \forall x \in E, \forall A \in \mathcal{E}. \quad (1)$$

Pour $n = 1$

$$N^2(x, A) = N \cdot N(x, A) = \int_{y \in E} N(x, dy) N(y, A), \quad \forall x \in E, \forall A \in \mathcal{E}.$$

Par définition de composition de deux noyaux de transition, (1) est vérifiée pour $n = 1$.
Pour $n = 2$

$$\begin{aligned} N^3 = N \cdot N \cdot N &= N^2 \cdot N \\ &= N \cdot N^2 \quad \text{(Par associativité)} \end{aligned}$$

D'après la proposition (3), N^2 est un noyau de transition sur $(E, \mathcal{E})$

$$N \cdot N^2(x, A) = \int_{y \in E} N(x, dy) N^2(y, A), \quad \forall x \in E, \forall A \in \mathcal{E}.$$

Donc, (1) est vérifiée pour $n = 2$

. Hypothèse de réccurence: on suppose que (1) est vérifiée pour n et qu'elle reste vraie pour $n + 1$.

$$\begin{aligned} N^{n+1} &= \overbrace{N \cdot N \ldots \cdot N}\, n+1 \text{ fois} \\ &= N^n \cdot N \\ &= N \cdot N^n \quad \text{(Par associativité)} \end{aligned}$$

$$N \cdot N^n = \int_{y \in E} N(x, dy) N^n(y, A), \quad \forall x \in E, \forall A \in \mathcal{E}.$$

Donc

$$N^{n+1}(x, A) = \int_{y \in E} N(x, dy) N^n(y, A), \quad \forall x \in E, \forall A \in \mathcal{E}$$

□

2.2 Chaîne de Markov à espace des états E quelconque:

Définition 4. *Soient $(E, \mathcal{E})$ un espace mesurable et μ une mesure de probabilité sur E, $(X_n)_n$ une suite de variables aléatoires à valeurs dans E est une chaîne de Markov d'espace des états E de loi initiale $\mu = \mathcal{L}(X_0)$ si pour toute fonction f $\mathcal{E}$-mesurable et bornée sur E*

$$\forall n > 0, \mathbb{E}[f(X_{n+1})|X_n, \ \ldots \ , X_1, X_0] = \mathbb{E}[f(X_{n+1})|X_n] - p.s.$$

Cette définition peut être reformulée en termes de probabilité conditionnelle de la manière suivante:

Définition 5. *Soient $(E, \mathcal{E})$ un espace mesurable et $(X_n)_{n \geq 0}$ une suite de variables aléatoires à valeurs dans E, de loi initiale $\mu = \mathcal{L}(X_0)$, On dit que $(X_n)_{n \geq 0}$ est une chaîne de Markov d'espace des états E si:*

$$\forall A \in \mathcal{E}, \forall n \geq 0, P[X_{n+1} \in A \,|\, X_n, \ \ldots \ , X_1, X_0)] = P[X_{n+1} \in A | X_n] - p.s.$$

i.e $\quad \forall A \in \mathcal{E}, \forall n \geq 0, P[X_{n+1} \in A \,|\, \sigma(X_n, \ \ldots \ , X_1, X_0)] = P[X_{n+1} \in A | \sigma(X_n)] - p.s.$

Le présent du processus est déterminé par les événements de la tribu engendrée par X_n, le passé est déterminé par les événements de $\mathcal{F}_n = \sigma(X_0, \dots, X_{n-1})$ et le futur est déterminé par les événements de la tribu engendrée par X_k, $k \geq n+1$. Si la chaîne de Markov ne dépend pas de n alors la chaîne de Markov est homogène.

Remarque 2. *La définition* (4) *est équivalente à la définition* (5)*.*

Démonstration.

$\Rightarrow$.

$(X_n)_n$ une chaîne de Markov de loi initiale μ, alors d'après La définition (4) $\forall f$ une fonction $\mathcal{E}$-mesurable et bornée sur E :

$$\forall n > 0, \mathbb{E}[f(X_{n+1})|X_n, \dots, X_1, X_0] = \mathbb{E}[f(X_{n+1})|X_n].$$

Nous avons $\forall A \in \mathcal{E}$, $\mathbb{I}_A$ mesurable et bornée, donc pour $f = \mathbb{I}_A$, avec $\forall A \in \mathcal{E}$:

$$\mathbb{E}[\mathbb{I}_A(X_{n+1})|X_n, \dots, X_1, X_0] = \mathbb{E}[\mathbb{I}_A(X_{n+1})|X_n] \; \forall n > 0.$$

$i.e.$ $\forall n \geq 0, P[X_{n+1} \in A \,|\, X_n, \dots, X_1, X_0] = P[X_{n+1} \in A \,|\, X_n] - p.s \; \forall n > 0.$

par définition de la probabilité conditionnelle.

$\Leftarrow$.

$(X_n)_n$ une chaîne de Markov de loi initiale μ, alors d'après La définition (5)

$$\forall A \in \mathcal{E}, \forall n \geq 0, P[X_{n+1} \in A \,|\, X_n, \dots, X_1, X_0] = P[X_{n+1} \in A \,|\, X_n] - p.s.$$
$$i.e \quad \mathbb{E}[\mathbb{I}_A(X_{n+1})|X_n, \dots, X_1, X_0] = \mathbb{E}[\mathbb{I}_A(X_{n+1})|X_n] \; \forall n > 0.$$

par définition de la probabilité conditionnelle.

Si f une fonction mesurable est bornée, alors f est une limite d'une suite croissante de fonctions élémentaires $(f_k)_k$

$$i.e \; f = \lim_{k \to \infty} f_k$$

$$avec \quad f_k = \sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i}, \text{ pour tout } k \in \mathbb{N}$$

D'où α_i est un nombre réel pour tout $i = 1, \dots, k$ et $\forall i, A_i \in \mathcal{E}$.

$Ona \;\; \mathbb{E}[\mathbb{I}_{A_i}(X_{n+1})|X_n, \dots, X_1, X_0] = \mathbb{E}[\mathbb{I}_{A_i}(X_{n+1})|X_n], \forall i = 1, \dots, k$ Pour tout k *de* $\mathbb{N}$

$$\begin{aligned}
D'ou \quad & \alpha_i \mathbb{E}[\mathbb{I}_{A_i}(X_{n+1})|X_n, \ldots, X_1, X_0] = \alpha_i \mathbb{E}[\mathbb{I}_{A_i}(X_{n+1})|X_n] \\
i.e. \quad & \mathbb{E}[\alpha_i \mathbb{I}_{A_i}(X_{n+1})|X_n, \ldots, X_1, X_0] = \mathbb{E}[\alpha_i \mathbb{I}_{A_i}(X_{n+1})|X_n] \\
et \quad & \sum_{i=1}^{k} \mathbb{E}[\alpha_i \mathbb{I}_{A_i}(X_{n+1})|X_n, \ldots, X_1, X_0] = \sum_{i=1}^{k} \mathbb{E}[\alpha_i \mathbb{I}_{A_i}(X_{n+1})|X_n] \\
Donc \quad & \mathbb{E}[\sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i}(X_{n+1})|X_n, \ldots, X_1, X_0] = \mathbb{E}[\sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i}(X_{n+1})|X_n] \\
Alors \quad & \lim_{k\to+\infty} \mathbb{E}[\sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i}(X_{n+1})|X_n, \ldots, X_1, X_0] = \lim_{k\to+\infty} \mathbb{E}[\sum_{i=1}^{k} \alpha_i \mathbb{I}_{A_i}(X_{n+1})|X_n] \\
i.e \quad & \lim_{k\to+\infty} \mathbb{E}[f_k(X_{n+1})|X_n, \ldots, X_1, X_0] = \lim_{k\to+\infty} \mathbb{E}[f_k(X_{n+1})|X_n]
\end{aligned}$$

Par convergence monotone, on a

$$\mathbb{E}[\lim_{k\to+\infty} f_k(X_{n+1})|X_n, \ldots, X_1, X_0] = \mathbb{E}[\lim_{k\to+\infty} f_k(X_{n+1})|X_n, \ldots, X_1, X_0]$$

. On établit la validité du passage á la limite par la convergence monotone. D'une part nous avons $f_k \in \mathbb{L}^1$, $\forall k$ $(f_k)_k$ est une suite croissante vers f, f est bornée, i.e $\exists M$ tel que $f(X) \leq M$, $\forall X$. Comme f_k croit vers f alors, $\forall k$, $\int_E f_k(X_{n+1})dP \leq \int_E f(X_{n+1})dP \leq M \int_E dP = M$,
Donc par la convergence monotone, nous avons

$$\lim_k \int_E f_k(X_{n+1})dP = \int_E \lim_k f_k(X_{n+1})dP$$

D'autre part, nous avons $\mathbb{E}[f_k(X_{n+1})|X_n] \in \mathbb{L}^1$, $\forall k$ $(\mathbb{E}[f_k(X_{n+1})|X_n])_k$ est une suite croissante, donc $\exists M$ tel que $\forall A \in \sigma(X_n)$

$$\begin{aligned}
\int_A \mathbb{E}[f_k(X_{n+1})|X_n]dP &= \int_A f_k(X_{n+1})dP \quad \text{par définition de l'espérance conditionnelle} \\
&\leq \int_A f(X_{n+1})dP \quad \text{Par hypothèse, } f_k \; croit\ vers\ f \\
&\leq M \int_A dP = M \cdot P(A) \leq M \cdot 1 = M
\end{aligned}$$

Donc, par convergence monotone nous avons

$$\lim_k \int_E \mathbb{E}[f_k(X_{n+1})|X_n] = \int_E \lim_k \mathbb{E}[f_k(X_{n+1})|X_n]$$

On a alors pour tout k et pour tout $A \in \sigma(X_n)$

$$\begin{aligned}
\int_A \lim_k \mathbb{E}[f_k(X_{n+1})|X_n] &= \lim_k \int_A \mathbb{E}[f_k(X_{n+1})|X_n] \\
&= \int_A \lim_k f_k(X_{n+1})dP \quad \text{par définition de l'espérance conditionnelle} \\
&= \lim_k \int_A f_k(X_{n+1})dP \\
&= \int_A \mathbb{E}[\lim_k f_k(X_{n+1})|X_n]dP \quad \text{par définition de l'espérance conditionnelle}
\end{aligned}$$

Par unicité présque sûre de l'espérance conditionnelle, nous avons

$$\mathbb{E}[\lim_k f_k(X_{n+1})|X_n] = \lim_k \mathbb{E}[f_k(X_{n+1})|X_n]$$

D'une manière analogue, on peut montrer que

$$\mathbb{E}[\lim_k f_k(X_{n+1})|X_0, \dots ,X_n] = \lim_k \mathbb{E}[f_k(X_{n+1})|X_0, \dots ,X_n]$$

Par conséquent

$$\mathbb{E}[f(X_{n+1})|X_0, \dots ,X_n] = \mathbb{E}[f(X_{n+1})|X_n]$$

□

Propsition 5. *Si $(X_n)_n$ une chaîne de Markov sur l'espace mesurable $(E, \mathcal{E})$, de noyau de transition N et de loi initiale μ, alors $N^n(x, A)$ s'interprète comme la probabilité de l'événement $\{X_{n+1} \in A\}$ conditionnelle à $\{X_n = x\}$ i.e.*

$$N^n(x, A) = P[X_{n+1} \in A|\, X_n = x\,].$$

Si $N(x,A)$ ne dépend pas de n , on dit que la chaîne de Markov est à probabilités de transitions stationnaires.

Remarque 3. *Une généralisation utile est la relation de Chapman-Kolmogorov:* $\forall n \in N, 0 \leq m \leq n, \forall x \in E, \forall A \in \mathcal{E}$

$$N^n(x, A) = \int_{y \in E} N^m(x, dy) N^{n-m}(y, A).$$

Interprétation:
Pour que la chaîne de Markov atteigne l'événement A en n étapes, partant de x, la chaîne doit nécessairement passer par une valeur y en m étapes, puis on suppose comme si la chaîne commence de y , pour atteindre A en $n-m$ étapes.

Propsition 6. *Soit N le noyau de transition sur l'espace mesurable $(E, \mathcal{E})$ associé à la chaîne de Markov $(X_n)_{n\geq 0}$, f est une fonction positive et mesurable alors on a $\forall x \in E, \forall A \in \mathcal{E}$*

$$N^n(x,A)f(x) = \int_{y\in E} f(y)N^n(x,dy) = \mathbb{E}_x[f(X_n)]$$

Propsition 7. *Soit N un noyau de transition sur l'espace mesurable $(E, \mathcal{E})$ associé à la chaîne de Markov homogène $(X_n)_{n\geq 0}$, alors si f positive, bornée et mesurable et $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ Pour deux entiers m,n tel que $m < n$*

$$\mathbb{E}[f(X_m)|\mathcal{F}_n] = N^{n-m}f(X_m) - p.s.$$

Propsition 8. *Soient $(E,\mathcal{E})$ un espace mesurable, $(X_n)_n$ une chaîne de Markov sur $(E,\mathcal{E})$, de noyau de transition N et de loi initiale $\mathcal{L}(X_0) = \pi$, alors $\forall A_0, A_1, \ldots, A_n \in \mathcal{E}$*

$$P[X_n \in A_n, X_{n-1} \in A_{n-1}, \ldots, X_0 \in A_0] =$$

$$\int_{A_{n-1}}\int_{A_{n-2}}\cdots\int_{A_0} N^{n-1}(A_n, x_{n-1})N^{n-2}(dx_{n-1}, x_{n-2})\ldots N^1(dx_1, x_0)\pi dx_0.$$

Définition 6. *Soit $(X_n)_n$ une chaîne de Markov sur $\mathbb{R}$, de noyau de transition N. Pour n un entier, soit $\mathcal{F}_n = \sigma(X_0,\ X_1, ...,\ X_n)$. Un temps d'arrêt ou temps optionnel ou temps markovien τ est une variable aléatoire à valeurs dans N, tel que $\forall n,\ \ \{\tau \leq n\} \in \mathcal{F}_n$ si,*

$$\mathcal{F}_\tau = \{A \in \mathcal{A},\ \forall n,\ A \cap \{\tau \leq n\}\} \in \mathcal{F}_n$$

alors $\mathcal{F}_\tau$ est une tribu dite tribu des événements antérieurs à τ.

Propsition 9. ***Propriété de Markov forte:***
Avec les notation ci-dessus, soit $(X_n)_n$ une chaîne de Markov sur $\mathbb{R}$ de noyau de transition N et τ un temps d'arrêt relativement à $(\mathcal{F}_n)_n$. Pour $B \in \mathcal{B}_\mathbb{R}$,

$$P[(X_\tau,\ \ X_{\tau+1}\ ,...) \in B/\mathcal{F}_\tau] = P[(X_0,\ \ X_1,...) \in B/X_\tau].$$

Définition 7. *Une mesure π sur l'espace mesurable $(E, \mathcal{E})$ est une mesure stationnaire pour la chaîne de Markov $(X_n)_n$ de noyau de transition N, si*

$$\forall A \in \mathcal{E},\, \pi(A) = \int_E N(A, x)\pi(dx).$$

Cette définition signifie que si π est la loi de X_0, c'est aussi la loi de X_1 . Remarquons que si cela est le cas alors π est la loi de X_n, $\forall n$.

Propsition 10. *Si π est une mesure stationnaire pour la chaîne de Markov $(X_n)_n$ de noyau de transition N , alors π vérifie*

$$\forall A \in \mathcal{E},\, \pi(A) = \int_E N^n(x, A)\pi(dx),\, \forall n.$$

Démonstration. Soit π la loi initiale de la chaîne de Markov $(X_n)_n$

$$\begin{aligned}
\pi(A) &= \int_E \pi(dw)N(w, A) \\
&= \int_E [\int_E \pi(dx)N(x, dw)]N(w, A) \\
&= \int_E \pi(dx) \int_E N(x, dw)N(w, A) \\
&= \int_E \pi(dx)N^2(x, A) \\
&\vdots \\
&= \int_E \pi(dx)N^n(x, A) = P(X_n \in A) \quad \textit{pour } n > 0,\, A \in \mathcal{E}
\end{aligned}$$

$\Rightarrow \forall n \;\; , \;\; \forall A \in \mathcal{E} \;\; : P[X_n \in A] = \pi(A),$ □

Donc d'aprés les deux dérnières propositions on a $N^n(x, A) = N(x, A)$ ce qui signifie que la chaîne de Markov est de probabilité de transition stationnaire.

Définition 8. *Soit $(E, \mathcal{E})$ un espace mesurable Une chaîne de Markov d'espace des états E, de noyau de transition N vérifie la condition d'équilibre (detailed balance condition) s'il existe une fonction π tel que*

$$N(y, x)\pi(y) = N(x, y)\pi(x), \quad \forall x, y \in E.$$

Propsition 11. *Si la chaîne de Markov vérifie la condition d'équilibre alors π est une loi stationnaire.*

Démonstration.

$$\begin{aligned}
\int_E \pi(y)N(y,B)dy &= \int_E \pi(y)[\int_B N(y,x)dx\,]dy \\
&= \int_E\int_B \pi(y)N(y,x)dx\,dy \\
&= \int_E\int_B \pi(x)N(x,y)dx\,dy \\
&= \int_B \pi(x)[\int_E N(x,y)dy\,]dx \\
&= \int_B \pi(x)dx \\
&= \pi(B)
\end{aligned}$$

□

Théorème 1. ***Théorème de Fubini:***

Soient $(\Omega_1,\mathcal{A}_1,\mu_1)$, $(\Omega_2,\mathcal{A}_2,\mu_2)$ *deux espaces mesurés,* μ_1 *et* μ_2 $\sigma-$*finies. On note* $\mu = \mu_1 \bigotimes \mu_2$ *sur* $(\Omega_1\times\Omega_2,\mathcal{A}_1\bigotimes\mathcal{A}_2)$.
Soit f *une fonction mesurables positives ou integrable définie sur* $(\Omega_1\times\Omega_2,\mathcal{A}_1\bigotimes\mathcal{A}_2)$, *alors on a*

$$\begin{aligned}
\int_{\Omega_1\times\Omega_2} f d\mu &= \int_{\Omega_1}[\int_{\Omega_2} f(x,y)d\mu_2]d\mu_1 \\
&= \int_{\Omega_2}[\int_{\Omega_1} f(x,y)d\mu_1]d\mu_2
\end{aligned}$$

2.3 Exemple de chaîne de Markov: La marche aléatoire sur $\mathbb{R}$

Soit $(Y_n)_n$ une suite de variables aléatoires à valeurs dans $\mathbb{R}$ indépendantes et de même loi μ. Si X_0 est une variable aléatoire à valeurs dans $\mathbb{R}$ indépendante des Y_n, soit $(X_n)_n$ la suite de variablers aléatoires définie par

$$X_0 \text{ et } \forall n>0,\, X_n = \sum_{k=1}^{n} Y_k.$$

Avec ces données, nous avons

Propsition 12. *Le processus $(X_n)_n$ est une chaîne de Markov sur $\mathbb{R}$, de noyau de transition N défini pour tout $x \in \mathbb{R}$ et $A \in \mathcal{B}_{\mathbb{R}}$ par*

$$N(A, x) = \mu(A - x)$$

Où $(A - x) = \{y \in \mathbb{R}, \quad tel \quad qu'il \quad \exists\, z \in A \quad tel \quad que \quad y = z - x\}$.

Démonstration. Nous avons :

$$\forall A \in \mathbb{R}, \forall n,\, P[X_{n+1} \in A | X_n, X_{n-1}, \ \ldots\ , X_0] = P[X_{n+1} \in A \,|\, X_n]$$

En effet, par définition de la probabilité conditionnelle,

$$\forall A_n, A_{n-1}, \ \ldots\ , A_0, \quad P[X_{n+1} \in A \,|\, X_n, X_{n-1}, \ \ldots\ , X_0]$$

vérifie

$$\begin{aligned}
P[X_{n+1} \in A, X_n \in A_n, \ \ldots\ , X_0 \in A_0] &= \int_{\{X_n \in A_n, \ldots, X_0 \in A_0\}} P[X_{n+1} \in A \,|\, X_n, \ \ldots\ , X_0] dP \\
&= \int_{\{X_n \in A_n, \ldots, X_0 \in A_0\}} \mathbb{E}[\mathbb{I}_A(X_{n+1}) \,|\, X_n, \ \ldots\ , X_0] dP \\
&= \int_{\{X_n \in A_n, \ldots, X_0 \in A_0\}} \mathbb{I}_A(X_{n+1}) dP \\
&= \int_{\{X_n \in A_n, \ldots, X_0 \in A_0\}} \mathbb{I}_A(Y_{n+1} + X_n) dP \\
&= \int_{\{X_n \in A_n, \ldots, X_0 \in A_0\}} \int_{\mathbb{R}} \mathbb{I}_A(x + y)\ P[X_n \in dx, Y_{n+1} \in dy]
\end{aligned}$$

Par indépendance de Y_{n+1} et X_n et on posons que $\{X_n \in A_n, \ \ldots\ , X_0 \in A_0\} = B$

$$\int_B \int_{\mathbb{R}} \mathbb{I}_A(x + y)\ P[X_n \in dx, Y_{n+1} \in dy] = \int_B \int_{\mathbb{R}} \mathbb{I}_A(x + y)\ p[X_n \in dx\ ]\ P[Y_{n+1} \in dy]$$

En appliquant le théorème de Fubini

$$\int_B \int_{\mathbb{R}} \mathbb{I}_A(x + y)\ p[X_n \in dx\ ]\ P[Y_{n+1} \in dy] = \int_B P[X_n \in dx] \int_{\mathbb{R}} \mathbb{I}_A(x + y) P[Y_{n+1} \in dy]$$

Si $\mu = \mathcal{L}(Y_n)$ et en posant

$$\begin{aligned}
\int_B P[X_n \in dx] \int_{\mathbb{R}} \mathbb{I}_A(x+y) P[Y_{n+1} \in dy] &= \int_B P[X_n \in dx\,] \int_{\mathbb{R}} \mathbb{I}_A(x+y)\, \mu\,(dy) \\
&= \int_B P[X_n \in dx\,] \int_{\mathbb{R}} \mathbb{I}_{A-x}(y)\, \mu\,(dy) \\
&= \int_B P[X_n \in dx\,]\, \mu\,(A-x) \\
&= \int_B P[Y_{n+1} \in A - x\,]\, P[X_n \in dx\,] \\
&= \int_B \; P[Y_{n+1} \in A - X_n] Pd(x).
\end{aligned}$$

Soit $\phi(X_n, X_{n-1}, \ldots, X_0)$ une fonction mesurable de $\mathbb{R}$ dans $[0,1]$. Si

$$\phi(X_n, X_{n-1}, \ldots, X_0) = P[X_{n+1} \in A \,|\, X_n, \ldots, X_0]$$

Alors par unicité de la probabilité conditionnelle on a

$$\phi(X_n, X_{n-1}, \ldots, X_0) = \mu\,(A - X_n)\text{-p.s.}$$

Donc

$$P[X_{n+1} \in A \,|\, X_n, X_{n-1}, \ldots, X_0\,] = P[X_{n+1} \in A \,|\, X_n\,]$$

qui est la propriètè de Markov donc $(X_n)_n$ est bien une chaîne de Markov sur $\mathbb{R}$.

. Pour caractériser $(X_n)_n$, on a besoin de préciser le noyau de transition associé N qu'on définit par :

$$\begin{aligned}
N(A, x) &= P[X_{n+1} \in A \,|\, X_n = x\,] \\
&= P[Y_{n+1} \in A - x\,] \\
&= \mu\,(A - x)
\end{aligned}$$

En effet le calcul fait ci-dessus montre que le noyau de transition N de la chaîne de Markov $(X_n)_n$ sur $\mathbb{R}$ est défini par $x \in \mathbb{R}$ et $A \in \mathcal{B}_R$ par $N(A, x) = \mu\,(A - x)$.

Avec $(A - x) = \{y \in \mathbb{R}, \;\; tel \;\; qu'il \;\; \exists\, z \in A \;\; tel \;\; que \;\; y = z - x\}$.

$\square$

Chapitre 3

Théorème ergodique pour les chaînes de Markov itératives

3.1 Chaînes de Markov itératives

Nous donnons ci-dessous quelques définitions et caractérisations des chaînes de Markov itératives.

Soit $(E,\mathcal{E})$ un espace mesurable, $f(x,y)$ une fonction à valeurs dans E est définie sur $E \times E$. Nous considerons $(Y_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs dans E de loi μ et X_0 une variable aléatoire à valeurs dans E indépendante des Y_n. Considérons le processus $(X_n^x)_n$ défini pour $x \in E$ par

$$X_0^x = x \text{ et pour tout } n > 0, X_{n+1}^x = f_{Y_{n+1}} \circ \cdots \circ f_{Y_1}(x)$$

Nous montrons ci-dessous que ce processus itératif est une chaîne de Markov.

Propsition 1. *Avec les notations ci-dessus, si μ est la loi des Y_n, alors $(X_n)_n$ est une chaîne de Markov de noyau de transition défini pour $x \in E$ et $A \in \mathcal{E}$, par*

$$N(x,A) = \mu\{f_{Y_1}(x) \in A\}$$

Démonstration.
La démonstration est basée sur les formules données par Diaconis et Freedman dans [4] et le calcul éffectué par Arezki dans [1]. Nous reprenons ce calcul.

Par définition de la probabilité conditionnelle $\forall A \in \mathcal{E}$, $P[X_{n+1} \in A \,|\, X_n]$ est l'unique variable aléatoire $\sigma(X_n)$-mesurable qui vérifie :

$$\forall B \in \mathcal{E},\, P[X_{n+1} \in A,\, X_n \in B] = \int_{\{X_n \in B\}} P[X_{n+1} \in A \,|\, X_n] dP$$

Cette variable aléatoire étant $\sigma(X_n)$-mesurable, il existe une fonction $\varphi : E \to [0,1]$ mesurable tel que

$$P[X_{n+1} \in A \,|\, X_n] = \varphi(X_n).$$

Donc, avec cette notation, pour A et $B \in \mathcal{B}$:

$$\begin{aligned}
P[X_{n+1} \in A, X_n \in B] &= \int_{\{X_n \in B\}} \varphi(X_n) dP \\
&= \int_{\{X_n \in B\}} \mathbb{E}[\mathbb{I}_A(X_{n+1}) \,|\, X_n] dP \\
&= \int_{\{X_n \in B\}} \mathbb{I}_A(X_{n+1}) dP \\
&= \int_{\{X_n \in B\}} \mathbb{I}_{\{X_{n+1} \in A\}}(w) dP \\
&= \int_{\{X_n \in B\}} \mathbb{I}_{\{f_{Y_{n+1}}(X_n) \in A\}}(w) dP \\
&= \int_E \int_B \mathbb{I}_{\{f_y(x) \in A\}} P[X_n \in dx,\, Y_{n+1} \in dy]
\end{aligned}$$

Par indépendance de Y_{n+1} et X_n, nous avons

$$\int_E \int_B \mathbb{I}_{\{f_y(x) \in A\}} P[X_n \in dx,\, Y_{n+1} \in dy] = \int_E \int_B \mathbb{I}_{\{f_y(x) \in A\}} P[X_n \in dx] \cdot P[Y_{n+1} \in dy]$$

En appliquant le théorème de Fubini, nous aurons :

$$\begin{aligned}
\int_E \int_B \mathbb{I}_{\{f_y(x) \in A\}} P[X_n \in dx] \cdot P[Y_{n+1} \in dy] &= \int_E \int_B \mathbb{I}_{\{f_y(x) \in A\}} P[X_n \in dx] \mu(dy) \\
&= \int_B P[X_n \in dx] \int_E \mathbb{I}_{\{f_y(x) \in A\}} \mu(dy) \\
&= \int_B P[X_n \in dx] \mu\{y,\, f_y(x) \in A\}
\end{aligned}$$

Par unicité de φ

$$\mathbb{P}(x, A) = \mu\{y, f_y(x) \in A\}, \quad x \in E, A \in \mathcal{B}_E$$

Il reste à montrer que X_n^x vérifie la propriété de Markov *i.e.*

$$\forall B \in \mathcal{B}_{\mathbf{E}}, \forall n, [X_{n+1} \in B | X_n, X_{n-1}, \ \dots \ , X_0] = P[X_{n+1} \in B | X_n]$$

Calculons $P[X_{n+1} \in A | X_n, \ \dots \ , X_0]$
Par définition de la probabilité conditionnelle $P[X_{n+1} \in B | X_n, X_{n-1}, \dots, X_0]$ est l'unique variable aléatoire $\sigma[X_n, \ X_{n-1} \dots, \ X_0]$ mesurable qui vérifie $\forall \ A_n, A_{n-1}, , \ \dots \ , A_0 \in \mathcal{B}_E$.

$$\begin{aligned}
P[X_{n+1} \in B, X_n \in A_n, \ \dots \ , X_0 \in A_0] &= \int_{\{X_n \in A_n, \dots, X_0 \in A_0\}} P[X_{n+1} \in B | X_n, X_{n-1}, \ \dots \ , X_0] dP \\
&= \int_{\{X_n \in A_n, \dots, X_0 \in A_0\}} \mathbb{E}[\mathbb{I}_B(X_{n+1}) | X_n, X_{n-1}, \ \dots \ , X_0] dP \\
&= \int_{\{X_n \in A_n, \dots, X_0 \in A_0\}} \mathbb{I}_B(X_{n+1}) dP \\
&= \int_{\{X_n \in A_n, \dots, X_0 \in A_0\}} \mathbb{I}_{\{f_{Y_{n+1}}(X_n) \in B\}}(w) dP \\
&= \int_{\{X_n \in A_n, \dots, X_0 \in A_0\}} N(B, X_n) dP
\end{aligned}$$

Donc si $\varphi(X_n, X_{n-1}, \ \dots \ , X_0)$ est une fonction mesurable dans $\mathbb{R}$ tel que

$$\varphi(X_n, X_{n-1}, \ \dots \ , X_0) = P[X_{n+1} \in B | X_n, X_{n-1}, \dots, X_0]$$

Alors

$$\varphi(X_n, X_{n-1}, \ \dots \ , X_0) = N(B, X_n) \text{ p.s}$$

Par unicité de la probabilité conditionnelle

$$P[X_{n+1} \in B | X_n, X_{n-1}, \dots, X_0] = P[X_{n+1} \in B | X_n].$$

□

3.2 Sommes de Birkhoff pour les chaînes de Markov itératives:

Définition 1. *Un espace métrique est un couple* (S, ρ) *constitué d'un espace d'éléments (points)* S *et d'une distance(métrique)* ρ. *Il est dit séparable si dans* S *il existe un ensemble dénombrable et partout dense.*

Définition 2. *Soit* (E, ρ) *un espace métrique ,* $\forall y \in E$, $f_y : E \to E$, *et dite une fonction lipschitizienne de rapport* K_y *si :*

$$\forall x, x' \in E, \quad \rho(f_y(x),\, f_y(x')) \leq K_y \rho(x, x').$$

Soit (S, ρ) un espace métrique séparable complet et $\mathcal{B}_S$ la tribu Borélienne de S. Soit Γ un enssemble de fonctions continues de S dans S. Soit $\mathcal{C}$ la tribu engendrée par Γ tel que la fonction $(\gamma, x) \longrightarrow \gamma(x)$ est une fonction mesurable sur $(\Gamma \times S, \mathcal{C} \bigotimes \mathcal{B}_S)$. Soit alors P une mesure de probabilité sur $(\Gamma,\mathcal{C})$. Nous avons alors la proposition suivante:

Propsition 2. *Dans un espace probabilisé* $(\Omega,\mathcal{F},Q)$, *soit* $(\Gamma_n)_n$ *une suite de variables aléatoires indépendantes et identiquement distribuées de loi* P *et* X_0 *une variable aléatoire dans* S *independante de* Γ_n. *Soit* $(X_n)_n$ *le processus défini par*

$$X_1 = \Gamma_1 X_0, \ldots,\ X_n = \Gamma_n X_{n-1} = \Gamma_n \cdots \Gamma_1 X_0$$

Alors $(X_n)_n$ *est une chaîne de Markov, de probabilité de transition* $N(x,\, B)$ *donné par*

$$N(x,B) = P[\{\gamma \in \Gamma : \gamma(x) \in B\}] \quad \forall x \in S,\ \ \forall B \in \mathcal{B}_S$$

Démonstration. La démonstration découle de la proposition précédente; Le calcul est fait à partir de Diaconis dans[4]. □

Notons que $\Gamma_n\ \cdots\ \Gamma_1\ x$ a la même distribution que $\Gamma_1\ \cdots\ \Gamma_n\ x$.
Soit $\Gamma^m = \Gamma \times \cdots \times \Gamma$, et P^m la probabilité produit sur $(\Gamma,\mathcal{C}^m)$, et $\Gamma^{(m)}$ l'essemble de toutes les combinaisons de $\gamma_m\ \cdots\ \gamma_1$ avec $\gamma_i \in \Gamma$, $1 \leq i \leq m$ et

$$\mathcal{C}^{(m)} = \{B \in \mathcal{B}_S \text{ tel que } (\gamma_m \cdots \gamma_1)^{-1} \in \mathcal{C}^m\}$$

et $P^{(m)}$ la probabilité induite.

Notations: Soit $X_0 \in S$, m_0 un entier positif et G une fonction positive définie par $G : \quad \Gamma^{(m_0)} \longrightarrow \mathbb{R}$ tel que

(A1) $\forall \gamma \in \Gamma^{(m_0)} : \rho(\gamma x,\, \gamma y) \leq G(\gamma)\rho(x,y), \forall x,y \in S$

(A2) $\lambda = \mathbb{E}[G(\Gamma_{(m_0)}...\Gamma_1)] < 1$

(A3) Pour tout $r \geq 0$, $\sup_{\rho(x,x_0)\leq r} \mathbb{E}[\rho(X_n(x),x_0)] < \infty \quad n = 1,...,m$

Le résultat ci-dessous précise des conditions pour que le système itératif $(X_n)_n$ admette une loi stationaire. Nous devons signaler que ces conditions ont été largement approffondies dans [4]. Nous nous en limitons car elles sont sufisantes pour l'étude des sommes de Birkhoff.

Théorème 1. *On suppose pour tout x_0 dans S et m_0 un entier naturel et G une fonction positive, mesurable sur $\Gamma^{(m_0)}$, de plus **(A1)** et **(A3)** sont verifiés. Alors il existe une mesure de probabilité invariante et unique $\pi(dy)$ donc pour $\mathbb{N}(x,dy)$ et $\mathbb{N}^n(x,dy)$ convergent faiblement vers $\pi(dy)$, pour tout $x \in S$.*

Théorème 2. *Avec les hypothèses du théorème ci-dessus et si π est l'unique probabilité invariante, alors:*

$$\frac{1}{n}\sum_{i=1}^{n-1} f(X_i) \longrightarrow \int f d\pi \qquad p.s$$

Pour toute fonction Lipschitizienne .

Rappelons qu'un processus $(B_t)_{t\geq 0}$ est un mouvement Brownien s'il vérifie les conditions de la définition suivante. Pour un peu plus de précisions, voir un exemple de construction dans Annexe, que nous avons adapté à partir du mémoire de Magister de LADJIMI (Cf.[7]).

Définition 3. *Un processus aléatoire $(B_t)_{t\geq 0}$ est un processus de Wiener ou mouvement Brownien si*

(1) *les accroissements $B_{t+s} - B_s$ sont stationnaires et*

$$\mathcal{L}(B_{t+s} - B_s) = N(0,\, \sigma^2 t)$$

(2) *Si $\forall\ 0 \leq< t_1 < t_2 < \cdots < t_n$ alors les accroissements sont independants c'est-à-dire les $B_{t_i} - B_{t_j}$ sont sont indépendants;*

(3) *$B(0) = 0$, B_t est continu en 0.*

Supposons que les notations de théorème 1 sont vérifiées et π est l'unique mesure de probabilité invariante pour $N(x,\, dy)$.

Soit T l'opérateur de transition dans $\mathbb{L}^2(S,\pi)$ défini tel que

$$Tf(x) = \int f(y)N(x,dy), \quad f \in \mathbb{L}^2(S,\pi)$$

$$Alors \;\; (T^n f)(x) = \int f(y)N^n(x,dy)$$

Pour $f \in \mathbb{L}^2(S,\pi)$, soit $\bar{f} = \int f d\pi$, et pour tout entier n soit $(Y_n(t))_t$ le processus défini par

$$Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=0}^{[nt]} (f(X_i) - \bar{f}) + (t - \frac{[nt]}{n})(f(X_{[nt]+1}) - \bar{f}) \;\; pour \;\; tout \;\; t \geq 0.$$

Pour $f \in \mathbb{L}^2(S,\pi)$ fixé, nous avons le résultat principal est que $Y_n(.)$ converge en loi vers un mouvement Brownien, c'est-à-dire

$$Y_t \longrightarrow B_t \;\; en \;\; loi.$$

Théorème 3. *Si les conditions du théorème 1 sont vérifiées, avec les mêmes notations, nous avons :*

(1) *Si la distribution initial de X_0 est π, alors pour toute fonction Lipchitizienne on a la convergence de Y_t vers B_t.*

(2) *Si, en outre $\int \rho^2(x,x_0)d\pi < \infty$, alors pour toute fonction Lipchitizienne on a la convergence de Y_t vers B_t.*

(3) *Les convergences dans **(1)** et **(2)** sont vérifié sans utiliser la distribution initial.*

Conclusion

Dans cette étude, nous avons essayé de donner une synthése sur le théorème ergodique et son application aux chaînes de Markov itératives. Les éléments pour assoir l'appareillage théorique sont assez élaborés et variés. Nous avons essayé d'aller vers l'essentiel, dans le cadre de ce mémoire de Master. Ce qui a retenu notre attention c'est surtout l'approximation donnée par Oesook Lee d'une chaîne de Markov itérative par un processus convergeant faiblement vers un mouvement brownien. Il est à noter que cettte approximation est valide sous l'hypothèse que les fonctions aléatoires considérées sont lipschitziennes. Peut être peut-on affaiblir cette hypothése? Cela constitue une première direction de travail. D'autres approximations sont à explorer pour les sommes de Birkhoff en liaison, en particulier, avec les travaux de Diaconis et Mirek. Ce qui fournit une deuxième direction de travail. Il est à signaler que beaucoup de travaux dans la litérature sont consacrés aux chaînes de Markov itératives. Cela s'explique sans doute par les nombreuses applications de ce modèle.

Annexe: Marches Aléatoires et Mouvement Brownien

Considérons une particule qui se déplace sur l'axe $\mathbb{Z}$, partant de 0 à l'instant 0, en effectuant à chaque unité de temps un saut d'une unité de longeur vers la droite ou vers la gauche avec la même probabilité. Soit Y_n le saut effectué à l'instant n. Avec ces hypothéses nous avons $Y_n = \pm 1$. Nous supposons de plus que les Y_n sont des variables aléatoires indépendantes et de même loi. Soit X_n la position de la particule à l'instant n. Nous avons

$$\forall n > 0,\, X_n = X_{n-1} + Y_n$$

avec $X_0 = 0$. $(X_n)_n$ est la marche aléatoire symétrique sur $\mathbb{Z}$.

Si on note $P_n(x) = P[X_n = x | X_0 = 0]$, nous avons

$$P_n(x) = C_n^{\frac{n+x}{2}} \frac{1}{2^n} \quad \text{avec} \quad C_n^k = 0 \quad \text{si} \quad k \notin \mathbb{Z}$$

Les P_n satisfont alors à l'equation aux différences:

$$P_{n+1}(x) = \frac{1}{2} P_n(x-1) + \frac{1}{2} P_n(x+1),$$

avec les conditions aux limites

$$P_0(x) = \begin{cases} 0 & \text{si } x \neq 0 \\ 1 & \text{si } x = 0 \end{cases}$$

Supposons maintenant que la marche soit accélérée en prenant des sauts de plus en plus petits, dans des intervalles de temps de plus en plus petits. Y a t'il un processus limite, obtenu en faisant tendre vers 0 la longeur des sauts et l'intervalle de temps?

De façon précise, supposons que la particule se deplace d'une longeur Δx à droite ou à gauche avec la même probabilité à chacun des instants $0, \Delta t, 2\Delta t...$ Notons S_t sa position à l'instant $t > 0$ avec l'hypothèse que $S_0 = 0$. On a alors

$$S_t = X_1 + X_2 + ... + X_{[\frac{t}{\Delta t}]},$$

où (X_n) est une suite de variables aléatoires indépendantes de même loi définies par

$$X_n = \pm \Delta x$$

avec

$$P(X_n = \Delta x) = P(X_n = -\Delta x) = 1/2 \quad pour \quad n = 1,2,...,[\frac{t}{\Delta t}].$$

Nous avons $\mathbb{E}(X_n) = 0$ et $\mathbb{E}(X_n)^2 = \Delta x^2$, d'ou

$$\mathbb{E}(S_t) = 0 \quad et \quad Var(S_t) = [\frac{t}{\Delta t}](\Delta x)^2.$$

Il s'agit maintenant de passer à la limite, en faisant tendre Δx et Δt vers 0, de sorte que le processus limite ne soit pas trivial. Si, par exemple, on prend $\Delta x = \Delta t$ et on fait tendre Δt vers 0, on obtient $\mathbb{E}(S_t) = 0$ et $Var(S_t) \longrightarrow 0$. Le processus $(S_t)_{t>0}$ serait alors presque sûrement nul.

Faisant tendre Δx et Δt vers 0, en imposant la condition

$$\frac{(\Delta x)^2}{\Delta t} \longrightarrow 2D$$

où D est une constante appelée coefficient de diffusion, Einstein a montré que $D = \frac{2RT}{Nf}$.
N represente le nombre d'Avogadro, T la temperature, R la constante des gaz parfaits et f le coefficient de friction. Avec cette condition, l'equation aux differences devient:

$$P_{t+\Delta t}(x) = \frac{1}{2}P_t(x - \Delta x) + \frac{1}{2}P_t(x + \Delta x)$$

D'où

$$\frac{(P_{t+\Delta t}(x) - P_t(x))}{\Delta t} \times \frac{\Delta t}{(\Delta x)^2} = \frac{1}{2} \times \frac{P_t(x - \Delta x) - P_t(x) + P_t(x + \Delta x) - P_t(x)}{(\Delta x)^2}. \tag{1}$$

Quand $(\Delta t,\Delta x) \longrightarrow (0{,}0)$ l'equation (1) devient equation dite de Fokker-Planck,

$$\frac{1}{D}\frac{\partial P}{\partial t} = \frac{\partial^2 P}{\partial^2 x} \tag{2}$$

Avec les conditions aux limites l'equation (2) devient:

$$(E)\begin{cases} \frac{\partial P}{\partial t} - D\frac{\partial^2 P}{\partial^2 x} = 0, & (t,x) \in \,]0,+\infty[\times\mathbb{R} \\ \\ P(0,x) = P_0(x) = \delta_x(0), & x \in \mathbb{R}. \end{cases}$$

C'est une equation du type de l'equation de la chaleur ou equation de diffusion. Pour plus de détail, on confère [**?**].

La résolution de l'equation (E) consiste à chercher une fonction $P(t,x)$ de classe C^1 par rapport à t et de classe C^2 par rapport à x, pour $(t,x) \in \,]0;+\infty[\times\mathbb{R}$.

$$P(0,x) = P_0(x) \text{ signifie que } \lim_{t\to 0} P(t,x) = P_0(x)$$

On suppose que

$$\int_{-\infty}^{+\infty} |P(t,x)|dx < \infty,\ \int_{-\infty}^{+\infty} |\frac{\partial P}{\partial t}(t,x)|dx < \infty \quad et \quad \int_{-\infty}^{+\infty} |\frac{\partial^2 P}{\partial^2 t}(t,x)|dx < \infty$$

de sorte que les fonctions

$$x \longmapsto P(t,x),\ \ x \longmapsto \frac{\partial P}{\partial t}(t,x) \ \ et \ \ x \longmapsto \frac{\partial^2 P}{\partial^2 x}(t,x)$$

ont pour chaque t une transformée de Fourier par rapport à la variable x.

Nous désignerons la transformée de Fourier de $x \longmapsto P(t,x)$ par

$$\mathcal{F}[P](t,k) = \widehat{P}(t,k) = \int_{\mathbb{R}} P(t,x)e^{-ikx}dx$$

On suppose de plus que pour tout $t > 0$, on a

$$\int_{-\infty}^{+\infty} \frac{\partial P}{\partial t}(t,x)e^{-ikx}dx = \frac{\partial}{\partial t}\left(\int_{\mathbb{R}} P(t,x)e^{-ikx}dx\right)$$

En appliquant la transformée de fourier à l'equation (2), on obtient, pour tout $k \in \mathbb{R}$:

$$\begin{aligned}
\mathcal{F}\left[\frac{\partial P}{\partial t} - D\frac{\partial^2 P}{\partial^2 x}\right](t,k) &= \mathcal{F}\left[\frac{\partial P}{\partial t}\right](t,k) - D\mathcal{F}\left[\frac{\partial^2 P}{\partial^2 x}\right](t,k) \\
&= \int_{-\infty}^{+\infty} \frac{\partial P}{\partial t}(t,x)e^{-ikx}dx - D\int_{-\infty}^{+\infty} \frac{\partial^2 P}{\partial^2 x}(t,x)e^{-ikx}dx. \\
&= \frac{\partial}{\partial t}\left(\int_{\mathbb{R}} P(t,x)e^{-ikx}dx\right) + Dk^2\int_{\mathbb{R}} P(t,x)e^{-ikx}dx \\
&= \frac{\partial \widehat{P}}{\partial t}(t,k) + Dk^2\widehat{P}(t,k).
\end{aligned}$$

Car

$$\int_{-\infty}^{+\infty} \frac{\partial^2 P}{\partial^2 x}(t,x)e^{-ikx}dx = (ik)^2\mathcal{F}[P](t,k).$$

Ainsi, Pour tout $k \in \mathbb{R}$, $\widehat{P}$ est solution de l'équation différentielle par rapport au temps t suivante :

$$\frac{\partial \widehat{P}}{\partial t} + Dk^2\widehat{P} = 0.$$

Par ailleurs $\widehat{P}(0,k) = \widehat{P_0}(k)$, d'où

$$\widehat{P}(t,k) = \widehat{P}_0(k)e^{-Dk^2t}.$$

On sait que pour tout $t > 0$ fixé,

$$e^{-Dk^2t} = \mathcal{F}[G](t,k), \text{ avec } G(x,t) = \frac{1}{\sqrt{4\pi Dt}}\exp(\frac{-x^2}{4Dt}).$$

Pour $t > 0$ La fonction $G : x \longmapsto \frac{1}{\sqrt{4\pi Dt}}\exp(\frac{-x^2}{4Dt})$ est applée noyau de Gauss ou noyau de la chaleur.

La fonction $k \longmapsto \widehat{P}_0(k)e^{-Dk^2t}$ est dans $\mathbb{L}^1(\mathbb{R})$, donc

$$\widehat{P}(t,k) = \mathcal{F}[P_0 * G](t,k).$$

Comme $\mathcal{F}$ est injective sur $\mathbb{L}^1(\mathbb{R})$ alors,

$$P(t,x) = (P_0 * G)(t,x) = \frac{1}{\sqrt{4\pi Dt}}\int_{\mathbb{R}} \exp(-\frac{(x-y)^2}{4Dt})P_0(y)dy.$$

si à l'instant $t = 0$ la particule est on $x = 0$, alors $P_0(x) = \delta_x(0)$ et la solution de (E) est:

$$P(t,x) = \frac{1}{\sqrt{4\pi Dt}}\ e^{-\frac{x^2}{4\sigma t}}.$$

Remarquons que le même résultat peut être obtenu en appliquant le théorème de la limite centrale. En effet: Quand $(\Delta t,\Delta x) \longrightarrow (0,0)$, $\mathbb{E}(S_t) = 0$ et $Var(S_t) \to 2Dt$, donc

$$\frac{S_t}{\sqrt{2Dt}} \quad \text{tend vers la loi normale} \quad \mathcal{N}(0,1)$$

puisque $S(t)$ est la somme de variables aléatoire indépendantes, et de même loi. Ce qui ce traduit par la relation

$$P\Big[\frac{S_t}{\sqrt{2Dt}} < \alpha\Big] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-\frac{x^2}{2}} dx,$$

ce qui donne

$$P[S_t < \beta] = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\beta} e^{-\frac{x^2}{4tD}} dx.$$

Si par exemple , en prend $\Delta x = \sqrt{\Delta t}$ et $\Delta t \longrightarrow 0$, alors $D = \frac{1}{2}$, d'où

$$\mathbb{E}(S_t) = 0, Var(S_t) \longrightarrow t, \ \text{ et } \ P(t,x) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{x^2}{2t}}.$$

On obtient un processus limite $(S_t)_{t>0}$ admettant les propriétés suivantes:

1. Pour $t > 0$ fixé, la variable aléatoire S_t suit la loi $\mathcal{N}(0,\sqrt{t})$

2. Le processus $(S_t)_{t>0}$ est à accroissements indépendants.
   En effet ce resultat est du au fait que dans la marche aléatoire, les sauts ayant lieu dans des intervalles de temps disjoints sont indépendants.

3. Le processus $(S_t)_{t>0}$ est à accroissements stationnaires.
   Cette propriété résulte du fait que dans la marche aléatoire le changement de position dans un intervalle de temps ne dépend que de la longueur de cet intervalle.

Le Processus obtenu est le Mouvement Brownien standard.

Bibliographie

[1] AREZKI O.(2013). *Itération de fonctions aléatoires et systèmes dynamiques* Memoire de Master "Maths. Appli. Processus aléatoires et Statistique de la Décision" UMMTO.

[2] BARRIERA L. (2011).*Ergodic theory, hyperbolic dynamique and dimension theory.* Universitext, Ed.,Springer, London Dordrecht Heidelberg New York.

[3] BREIMAN L.(1992). *Probability.* American sociaty for industrial and applied mathematics Philadelphia, University of California, Berkley.

[4] DIACONIS P. et FREEDMAN D. (1999). *Iterated Random Functions.* SIAM Review, 41, 45-53.

[5] EINIEDLER M. et WARD T. (2010). *Ergodic theory with a view towards number theory.* Graduate texts in mathematics, Ed., Springer, London Dordrecht Heidelberg New York.

[6] KOLMOGOROV A. et FOMINE S.(1973) *Elément de la théorie des fonctions et de l'analyse fonctionnelle*,2ème Ed., Mir, Moscou.

[7] LADJIMI F. (2011). *Contribution à l'étude de l'integral stochastique et du mouvement Brownien* Memoire de Magister " UMMTO.

[8] LEE O. (1996). *Limit theorems for Markov processes generated by iterations of random maps.* Journal Korean Mathematics Soc., 33(4), 983-992.

[9] LEE O. (1998). *Weak convergence for iterated random maps.* Bulltin Korean Mathematics Soc., 35(3), 485-490.

[10] MANE R. (1983). *Ergodic theory and differentiable dynamics.* Ed., Springer-Verlag Berlin Heidelberg New York London Paris Tokyo.

[11] MIREK M. (2011). *Heavy tail phenomenon and convergence to stable laws for iterarted Lipschitz maps.* Probability Theory and Related Fields, 151, 705-734.

[12] POLLICOT M. et YURI M.*Dynamical systems and ergodic theory*Department of mathematics, Manchester university. Department of business administration, Sapporo university.

[13] REVUZ D. (1984). *Markov chains.* North-Holland mathematical library, Ed., North Holland, Amsterdam, New York, Oxford. Volume 11.

[14] SHIRYAEV A. N. *Probability.* Stelkov mathematical institue Moscow , Ed.,Springer. Second edition.

[15] WALTERS P. (1982).*An introduction to ergodic theory.*Graduate texts in mathematics, Ed., Springer,Verlag New York Heidelberg .